# Sparse Based Maintaining and Extending of Case-Based Reasoning Using a Competence and Dense Based Algorithm

**Changjian Yan[1,2], Chaojian Shi[1], Hamido Fujita[3], Nan Ma[4]**

[1] Merchant Marine College, Shanghai Maritime University, Shanghai, China

[2] Navigation College, Jimei University, Xiamen, China

[3] Iwate Prefectural University, Iwate 020-0693, Japan

[4] College of Information Technology, Beijing Union University, Beijing, China

*Abstract: Case-based Reasoning (CBR), an approach for analogical reasoning, has recently emerged as a major reasoning methodology in the field of artificial intelligence. The knowledge contained in a case base is crucial to solve problem for a CBR system and thus, there is always a tradeoff between the number of cases and the retrieval performance. Although many people attempt to deal with this issue these years, constructing a well compact competent case base needs much effort. In this paper, a new approach is proposed to maintain the size of case base. The maintenance process is divided into two separate stages. The former focuses on overcoming the competence of sparse cases and the latter emphasizes the dense cases. Using this strategy, we could appropriately maintain the size of the case base by extending the competence without losing significant information. We illustrate our approach by applying it to a range of standard UCI data sets. Experimental results show that the proposed technique outperforms current traditional approaches.*

*Keywords: Case Base; competence; hybridization ;density; sparse*

# 1 Introduction

Case Based Reasoning (CBR) is a problem solving paradigm utrilizing the solutions of similar problems stored as cases, in a Case Base CB and adapting them based on problem differences[1-3]. CBR is able to find a solution to a problem by employing the luggage of knowledge, in the form of cases. Usually, the case is represented in a pair as a "problem" and "solution" and cases are divided into groups.

* Corresponding author. E-mail: xxtmanan@buu.edu.cn, chjyan@jmu.edu.cn

Each case describes one particular situation and all cases are independent from each other. In the process of case adapting, the CBR system usually learns by storing the new case in the CB after solving a new problem. Generally speaking, if all new cases are retained, problem-solving speed will be inevitably impaired by the retrieval cost[4]. Recently, the case base maintenance issue has drawn more and more attention, and the major concern is how to select the cases to retain[5]. Some scholars have made great efforts in exploring "competence-preserving" case deletion which intends to delete cases whose loss may cause the least harm to the overall competence of the CB[6]. In this paper we concentrate on two maintenance scenarios: CNN (Condensed Nearest Neighbor Rule) and the adaptation of cases.

CNN is based on the k-nearest-neighbor (k-NN) algorithm which selects the reduced sets of cases[7]. The methods built on refinement of CNN framework such as IB2[8], which starts from an empty training set and adds misclassified instances and their IB3 to address the problem of keeping noisy instances. Smyth[5], Delany and Cunningham[9] as well as Angiulli[10] proposed the relative coverage considering how many other cases in the CB can solve the cases in the coverage set. Craw[11]discarded cases with extremely low complexity (redundant cases) or high complexity local case based on complexity. Other approaches to instance pruning systems are those that take into account the order in which instances are removed [24].

Adaptation-guided case base maintenance is another direction for maintaining the CB[3]. Hanney and Keane attempted to learn adaptation rules from CB with a difference heuristic approach[12]. Jalali and Leake implemented the case difference heuristic for lazy generation of adaptation rules[13], then they extended the case adaptation with automatically-generated ensembles of the adaptation rules[14].

Although the methods above maintain CBS with appropriate competence to some extent, there are still some problems. First, these methods rarely take into account the sparse distribution in the case space where the competence of the CB trends to deteriorate for lack of useful information. Second, they selectively delete and retain cases coming from different clusters. To this end, a new approach is proposed in tghis work, to maintain the size of case base. The maintaining process is divided into two separated stages. The former focuses on overcoming competence of the sparse cases, and the latter emphasizes the dense cases. We call the two separate stages hybridization, which tries to provide competence among case base and then provide condense base adaptions. Using this strategy, we can appropriately maintain the size of the case base by extending the competence without losing significant information and consequently provide an efficient solution to give impetus to these issues.

The paper is organized as follows. Section 2 presents the formalization representation of CB and introduces the competence and performance of CB for

evaluating. In section 3, the proposed approach to maintaining the size of case base is implemented. In Section 4, the performance of the proposed approach is compared with other traditional classifiers, using some well selected UCI datasets. Finally the conclusions on the proposed approach for adaptation and an outline of future works on the suitability for different adaptation tasks are summarized.

# 2 Theory Basis and Formalization Representation

To describe information and knowledge about CB, some general assumptions and corresponding definitions about the basic CB are introduced in this section.

## 2.1 Basic concepts about CB

**Definition 1** ***Case Model***: A case model is a finite, ordered list of attributes $c = (A_1, A_2, .....A_n)$ , where n > 0 and $A_i$ is an attribute denoted with a pair $A_i \in [(A_{name}, A_{range})]^l$ . The basic value type of $A_i$ can be the real symbolic type, numeric type, temporal type like Date and Time, etc. The symbol $\tilde{\mathbb{C}} = \{c_1, c_2, ...c_i ...c_{n-1}, c_n\}$ denotes the space of case models.

**Definition 2** ***Case Base:*** A case base *CB* for a given case model $\hat{\mathbb{C}}$ is a finite set of cases $\{c_1, c_2, ...c_i ...c_n\}$ with $c_i \in C'$ where $C'$ is the subset of case space $\hat{\mathbb{C}}$ .

**Definition 3** ***Cases Similarities Measure***(*CSM*): the *CSM* can be defined as a function $Sim_c = \Re_c \times \Re_c \to [0,1]$ measuring the degree of similarity between two different cases $c_i$ and $c_j$ . Generally, the *CSM* is a dual notation to distance measures and the reason is very obvious that a given *CSM* can be transformed to a distance measure by some transition function $f : Sim(X_i, X_j) = f(d(X_i, X_j))$ .

**Definition 4** ***Cases Cluster***: A cases cluster $\mathbb{Q}$ is a non-empty subset of the whole case base CB satisfying the following conditions:

(1) $\forall c_i, c_j$ , if $c_i \in \mathbb{C}$ and $c_j$ are density-reachable (see Definition 7) from $c_i$ , then $c_j \in \mathbb{C}$ .

(2) $\forall c_i, c_j \in \mathbb{C}$ , $c_i$ is density connected to $c_j$ .

**Definition 5** ***Case Density(CD)*** [15]: The density of an individual case can be defined as the average similarity between the cases $C_i$ and other clusters of cases called competence groups(see Equation 1)[16].

$$Density(c_i, \vec{C}) = \frac{\sum_{c_j \in \vec{C} - c_i} Sim(c_i, c_j)}{\left|\vec{C}\right| - 1} \quad (1)$$

Where $Sim(c_i, c_j)$ is the *CSM* value of different cases $c_i$ and $c_j$ and $\vec{C}$ is some cluster of cases satisfying Definition 4. And $\left|\vec{C}\right|$ is the number of cases in the group $\vec{C}$.

**Definition 6** ***Case Cluster Density(CCD)***: The density of some case cluster $\vec{C}$ can be measured as a whole as the average density of all cases in $\vec{C}$ (see Equation 2)

$$Density(\vec{C}) = \frac{\sum_{c \in \vec{C}} Density(c, \vec{C})}{\left|\vec{C}\right|} \quad (2)$$

**Definition 7** ***Density reachable***: A case $\zeta$ is density reachable from another case $\varsigma$ if there exists a case chain containing $L = \{C_1, C_2, ...C_i ...C_n\}$ where $\zeta = C_1, \varsigma = C_n$ such that $C_{t+1}$ is directly density-reachable from $C_t$.

## 2.2 Competence and Performance of CB:Criteria for Evaluating

Generally speaking, an effective case base with high quality should produce as many solutions as possible to queries for users. Reference [6] and [17-18] defined such criteria as competence and performance to judge the quality and effectiveness of a given case base.

- *Competence* is the range of target problems that can be successfully solved.
- *Performance* is the answer time that is necessary to compute a solution for case targets. This measure is bound directly to adaptation and result costs.

And to better understand the competence criteria above, two important properties are given as follows:

**Definition 8** ***Coverage***: given a case base $\mathbb{C} = \{c_1, c_2, ...c_i ...c_n\}$, for $c \in \mathbb{C}$, $Coverage(c) = \{\hat{c} \in \mathbb{C} : adaptable(c, \hat{c})\}$.

Obviously, the *Coverage* of a case is the set of target problems that it can be used to solve.

**Definition 9** ***Reachability***: given a case base $\mathbb{C} = \{c_1, c_2, ...c_i ...c_n\}$, for $c \in \mathbb{C}$, $Reachable(c) = \{\hat{c} \in \mathbb{C} : adaptable(\hat{c}, c)\}$

And from definition 9, we get get that the *Reachability* of a target problem is the set of cases that can be used to provide a solution for the target.

In a case base, all cases are not equal, i.e., some cases contribute more to the competence of the case base and others may contribute less to its competence. And it's also true for the performance criteria. Four different types of cases are defined as follows.

**Definition 10** $Pivot\_base(c^o)$ iff $Reachable(c^o) - (c^o) = \varnothing$

**Definition 11** $Support\_base(c^o)$ iff

$\exists c \in Reachable(c^o) - \{c^o\} : Coverage(c) \subset Coverage(c^o)$

**Definition 12** $c^o \in Span\_case(c)$ iff

$Pivot\_case(c) \wedge Coverage(c) \cap \mathrm{U}_{c^o \in Reachable(c)-\{c\}} Coverage(c) \neq \varnothing,$

**Definition 13** $Auxiliary\_base(c^o)$ iff

$c^o \in Reachable(c) - \{c\} : Coverage(c) \not\subset Coverage(c^o)$

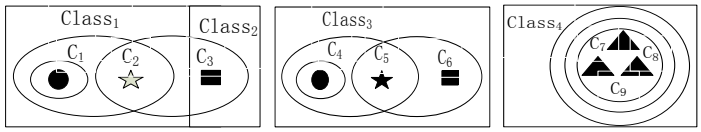



Figure 1 An example of different types of cases

From definition 11 to definition 13, we can easily classify cases $C_1$ and $C_4$ as $Auxiliary\_base$, $C_2$ and $C_5$ as $Span\_case$, $C_3$ and $C_6$ as $Pivot\_base$, $C_7$, $C_8$ and $C_9$ as $Support\_base$. And the case categories described above provide a benchmark for deletion order according to their competence contributions. This reduction technique Footprint first deletes auxiliary problems. Then it supports problems, and finally pivotal problems [19]. The approach is better than traditional deletion policies in view of preserving competence, however, the competence of case base is not always guaranteed to be preserved [20]. Furthermore, although this strategy keeps some CBs of suitable size, with good competence, the crucial relationship between local and global competence is ignored, for a long time, especially in a sparse region and dense region.

# 3 Maintaining and Extending of Case-Based Reasoning

## 3.1 Case Distribution

To better understand the competence and adaptation of the case base, in this section, we introduce the case distribution for some CB. By the way, for the sake of simplicity, we concentrate our attention on the cases distribution in the binary scenario.

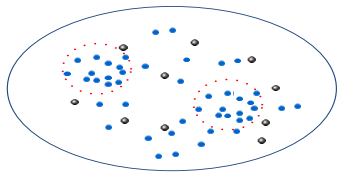


Figure 2 An example of case distribution

For a binary classification problem, as illustrated in Figure 2, the dots in blue are the cases pertaining to the dense cases called the majority, and the ones in black pertaining to the sparse cases called the minority. Obviously, the former type of cases have a larger contribution to the competence than the latter, i.e., the sparse ones. But we cannot ignore the fact, that more cases, as the ones surrounded by the important consideration, that too few cases will tend to a lack of coverage and red circles, may inevitably tend to redundancy, and we also cannot ignore one adaptation for the whole case base. In extreme circumstances, a target from a sparse region is likely to be unsolvable. Considering the fact that the distribution of the solutions of the case bases is a crucial factor affecting competence[21], we maintain the case base in two combined strategies as shown in the following sections.

## 3.2 Case Density Based Ranking

Algorithm 1 Case Density Matrix Ranking ***CDM-Ranking*** (*CB*)
Input (*n*, *k*, ***CB***)
output(*n*, *CB'*) (∗ ***CB*** is the orginal case base and *CB'* is produced with the density-ordered cases. ∗)
(1)Compute *distance* between case $c_i$ and its *k* Nearest neighbors respectively

$$distance(c_i, c_j) = 1 - similarity(c_i, c_j) \quad (3)$$

where

$$similarity(c_i, c_j) = \frac{\sum_{t=1}^{l} w_{c_{it}} \cdot w_{c_{it}}}{\sqrt{\sum_{t=1}^{l} (w_{c_i})^2 \cdot \sum_{t=1}^{l} (w_{c_{jt}})^2}} \quad (4)$$

(2)Compute the sum of $distance(c_i, c_j)$ denoted $Dis_{sum}$for case $c_i$

$$Dis_{sum}(c_i) = \sum_{j=1}^{k} \frac{1}{distance(c_i, c_j)} \quad (5)$$

(3)Normalize $Dis_{sum}(c_i)$ to get the density of case $c_i$

$$density(c_i) = \frac{Dis_{sum}(c_i)}{\sum_{i=1}^{n} Dis_{sum}(c_i)} \quad (6)$$

(4)Rank all the cases in CB according to the value of their density
(5) Return *CB'* (∗ End of *CDM-Ranking* ∗)

To effectively delete special cases in dense regions and add or generate new cases in sparse regions, we first rank the cases according to their density distribution. Algorithm 1, i.e., *CDM-Ranking*, is to generate the density-ordered. For some case *c*, its density depends on value of the total distance between itself and all its *k* nearest neighbors, i.e., the bigger the $distance(c_i, c_j)$ is, the less value of $density(c_i)$ gets and vice versa. The density of the cases provides fundamental

basis for generating new cases and deleting existing cases, as in section 3.3 and section 3.4, respectively.

## 3.3 Synthetic Addition for the Sparse Case

The proposed methodology here, is concerned with synthetic additions for the sparse case, namely SASC, originated from the idea of SMOTE[22], and generating new cases for the sparse cases according to their distribution density. Just as illustrated in Figure 3, the cases in the spatial distribution are sparse, such that their competence is not sufficient to new queries.

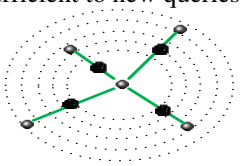


Figure 3 Strategy of generating new synthetic cases

To enhance the spatial distribution for the sparse cases, we implement algorithm 2 to generate the new synthetic cases to support more information for the queries.

Algorithm *2* ***NextSmote***(*CB*) (∗ Function to generate the synthetic cases. ∗)

（1） *while* $n \geq 0$

（2） Choose a random number $\lambda$ between 1 and *k*. (∗ *k* is the number of neighours of some case ∗)

（3） *for attr* ← 1 to $\lambda$

（4） *diff*= case[ $\lambda$ ]][*attr*] − case[*i*][*attr*]

（5） *gap* = random number between 0 and 1

（6） Synthetic case *c*[*newindex*][*attr*] = case[*i*][*attr*] + *gap* ∗ *diff*

（7） *endfor*

（8） *newindex*++

（9） $n = n - 1$

（10） *endwhile*

（11） return *c* with new cases (∗ End of ***NextSmote***. ∗)

New cases generated with *NextSmote* rely on the number of the nearest neighors of some cases, as this algorithm generates new cases between his case

and all its $k$ neighbours respectively. In the 4th step, the parameter *gap* is a random number between 0 and 1, and the algorithm ends when all cases in CB are smoted [22] and such that the CB is filled with newly generated cases. Consequently, the distribution space can provide more useful information for case queries and analysis.

Now we can turn to the important step of SASC to guide *NextSmote* process, see algorithm 3.

SASC's basic algorithm starts as the cluster idea[24].On the premise of threshold value of the cases, we deploy the *nextsmote* algorithm to generate synthetic cases. It should be noted that only those cases that can enhance the total competence are concentrated rather than all new generated cases. Details of the Estimation Error can be referenced from reference[3].

Algorithm *3* SASC's basic algorithm

**Input** ($n$, $\rho$ ,$CB$) ($*n$ is the number of cases to maintain, and $\rho$ is the maxum number of the cases, and $CB$ is the case base $*$)

**Output** ($CB'$) ($* CB'$ is a condensed set of $CB$ consisting of $n'$ cases, $n' \le n$ $*$)

ClusteredCases using *KNN* to get $m$ Clusters ;

*for* $i$=1 to $m$

*CDM-Ranking*($t$,$Cluster_i$) ($*t$ is the number of cases of the *ith* $Cluster_i$ $*$)

*for j=1 to k* ($*k$ is the number of cases in the *ith* cluster $Cluster_i$ $*$)

while *size*( $CB'$ ) $<$ $\rho$ do

*while density(c)*$<$ $\kappa$

$c$=*NextSmote*($Cluster_i$)

*EstimationError Abs(Value(c) -FindSol(c, CB))*

*if EstimationError* $<$ $\gamma$ ($*$ $\gamma$ is the threshold of *Error* $*$)

*Add*( $CB'$ , c)

*endif*

*endwhile*

*endwhile*

*endfor*

*return* $CB'$

## 3.4 Selective Deletion for the Dense Case

The proposed methodology that is concerned with selective deletions in the dense case, namely *SDDC*, is based on a categorization method as definition 10 to definition 13, and then delete cases in the order as proposed as section 3.2.

Algorithm *4 SDDC*'s basic algorithm

**Input** (*CB*)

O**utput**(condensed case set $CB'$)

$CB' = \varnothing$

*for* i=1 to *n*

*if* $c_i \in Pivot\_base$

*if* $density(c_i) < \xi$

$CB' \leftarrow c_i$;

*else return*;

*elseif* $c_i \in Support\_base$

*if* $density(c_i) < \alpha$

$CB' \leftarrow c_i$;

*elseif* $c_i \in Span\_base$

*if* $density(c_i) < \beta$

$CB' \leftarrow c_i$;

*elseif* $c_i \in Auxiliary\_base$

*if* $density(c_i) < \gamma$

$CB' \leftarrow c_i$;

*else return*;

*endif*

*endfor*

*Return* $CB'$

The most obvious points exist in two aspects. In the first, the *SDDC* algorithm deletes auxiliary cases, then supports cases and finally pivotal cases, i.e., we retain selectively the pivotal cases firstly, then support cases and span cases, finally the auxiliary cases. In the second place, we delete cases according to their density within their spatial distribution. Considering the contribution to the total coverage of the CB, we set the threshold value of the

density of the pivotal cases, i.e., $\xi$ , to the maximum and then one of the support cases and span cases and finally the auxiliary cases. In other words, we set $\xi > \alpha > \beta > \gamma$ .

# 4 Experimental Results

## 4.1 Datasets

We evaluated the performance of the proposed approach (i.e., SASC& SDDC) on four case domains Housing, MPG, Computer Hardware (Hardware) and Automobile (Auto) from the UCI repository[25]. These datasets were chosen in order to provide a wide variety of application areas, sizes, and difficulty as measured by the accuracy achieved by the current algorithms. The choice was also made with the goal of having enough data points to extract conclusions. First, for all data sets, the records with missing values were removed. And for sake of accordance with assessment of similarity by Euclidean distance in our proposed method, we selectively removed those cases, with numeric value of features. In order to facilitate comparison, values of each feature were standardized by subtracting that feature's mean value from the feature value and the result was divided by the standard deviation of that feature.

## 4.2 Experimental method

A ten-fold cross-validation was used for all experiments, and each fold was used as an independent test set, in turn, while the remaining nine folds were used as the training set. Then the mean absolute error (*MAE*), accuracy and resulting size were calculated. Experimental accuracy for all algorithms was measured by mean absolute error at different compression rates, defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|f_i - y_i\right| = \frac{1}{n}\sum_{i=1}^{n}\left|e_i\right| \tag{7}$$

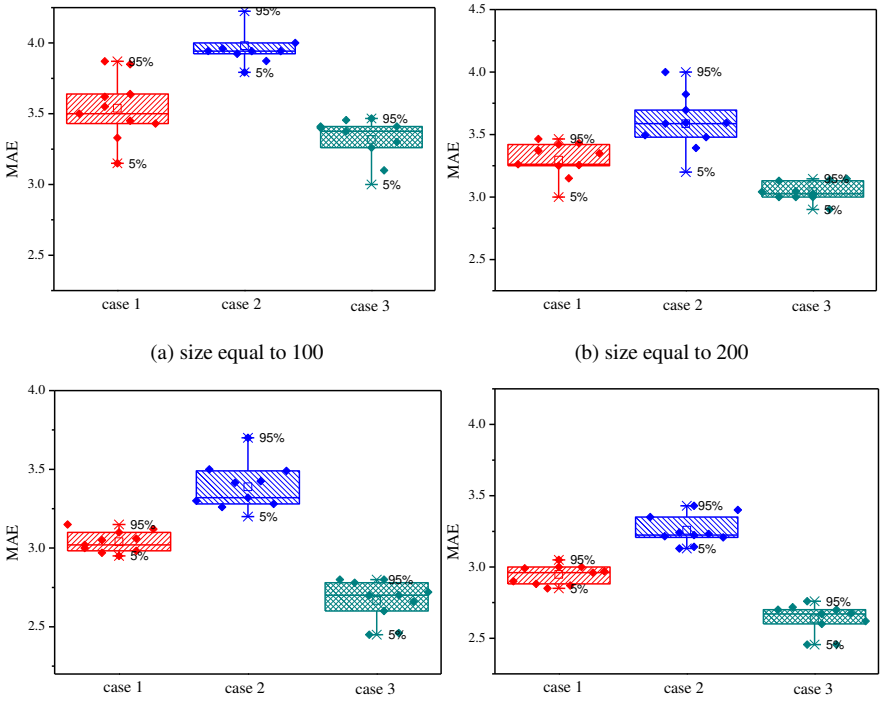
As the name suggests, the mean absolute error is an average of the absolute errors $\left|e_i\right| = \left|f_i - y_i\right|$ , where $f_i$ is the prediction and $y_i$ the true value.

**Experiment 1: Parameters Choice**

The performance of the proposed methodology concerned with selective deletion for the dense case, i.e., SDDC, is affected by the parameters $\xi, \alpha, \beta$ and $\gamma$ in

algorithm 4. More analysis concerning the unique fashion of the selection of the parameters, is made to try to achieve better results for SDDC here.

We first set the $\xi, \alpha, \beta$ and $\gamma$ equal, i.e., $\xi=\alpha=\beta=\gamma=1/n$ ,where $n$ is the size of the case base. Then we set $\xi = 1/n$ , $\alpha = 2/n$ , $\beta = 4/n$ , $\gamma = 8/n$ and $\xi = 8/n, \alpha = 4/n, \beta = 2/n, \gamma = 1/n$ respectively. The final results of *MAE* with ten-fold cross-validation on data set Housing are reported as follows when the sizes of case base are set to 100, 200, 300 and 400 respectively.




(a) size equal to 100 (b) size equal to 200

(c) size equal to 300 (d) size equal to 400

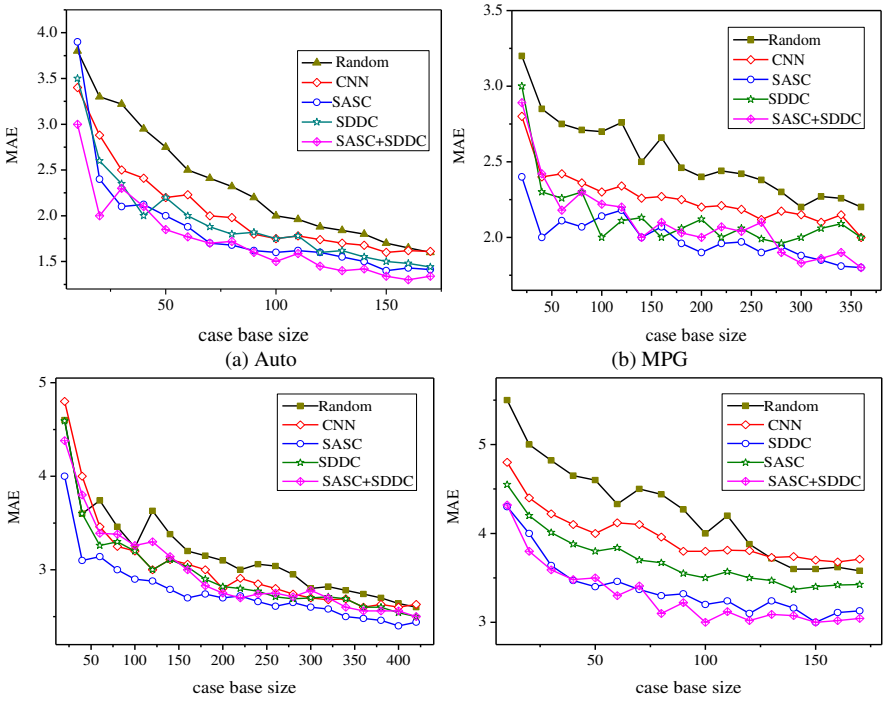
Figure 3 *MAE* of the *SDDC* methods with different size and delete threshold

In Figure 3, case 1, case 2 and case 3 represent three different parameters arrangement of $\xi, \alpha, \beta$ and $\gamma$ respectively in the order described above. What impressed us most exist two aspects. In the first place, the values of *MAE* of all the four scenarios gradually decrease with the increase of size of case base. And in another place, in every scenario the performance of the SDDC is the best in the case 3, then in the case 1 and at last the case 2. The reason these three cases achieved good distinction in such an arrangement is that the pivotal cases are most important, next support cases and then span cases and finally the auxiliary cases.

With the ordered selective ratio, information can be more useful when supported by the more powerful cases.

**Experiment 2: Error Rates**

In this experiment, the proposed *SASC*, *SDDC* and their hybrid model *SASC +SDDC* are compared to two, case base, maintenance methods that are standard in the current literature: Random Deletion[26](Random) and CNN[27]. The final results of *MAE* with ten-fold cross-validation are illustrated in figure 4.




(a) Auto (b) MPG

(c) Housing (d) Hardware

Figure 4 *MAE* of the candidate methods on four samples CB

In the four domains, the curves illustrate that the *SASC*, *SDDC* and their hybrid model *SASC+SDDC* out perform the other methods, i.e., the *Random* and the *CNN* methods. For example, in the MPG case base, for three case base sizes, relative coverage of *SASC+SDDC* notably outperforms *Random* by 21.62%, 20%, 20.21% and by 3.6%, 10% and 17.48%, when the number of maintained cases are equal to 100, 200 and 300 respectively. In general, experimental results show that the Random method displays the lowest performance. Moreover, for smaller case base sizes, for different domains, relative coverage shows the *SASC* method has the highest performance and depends on more information provided by the *smoted* [22] cases in the sparse space.

**Experiment 3: Performance Comparison**

To finish the empirical research, the second comparative study for the four methods (*CM*, *SASC*,*SDDC* as well as their hybrid model *SASC*+ *SDDC*), was carried out on the remaining 18 datasets, taken from the UCI repository[28]. Details are described in table 1 below:

Table 1
The classification accuracy and storage requirements for each dataset.

| dataset | *CM* | | *SASC* | | *SDDC* | | *SASC*+ *SDDC* | | The best method |
|---|---|---|---|---|---|---|---|---|---|
| | Storage (%) | Accuracy (%) | Storage (%) | Accuracy (%) | Storage (%) | Accuracy (%) | Storage (%) | Accuracy (%) | |
| anneal | 20.05 | 100 | 25.12 | 100 | 18.26 | 100 | 20.02 | 100 | *SASC+SDDC* |
| balance-scale | 13.78 | 95.83 | 14.5 | 98.24 | 12.9 | 96.05 | 13.48 | 96.88 | *SASC+SDDC* |
| breast-cancer-l | 4.02 | 96.56 | 6.1 | 97.28 | 4.01 | 96.71 | 4.42 | 96.87 | *SDDC* |
| breast-cancer-w | 5.29 | 93.24 | 5.43 | 92.22 | 5.22 | 95.1 | 5.02 | 93.24 | *SASC+SDDC* |
| Cleveland | 6 | 91.01 | 6.8 | 94.07 | 5.1 | 92.21 | 5.4 | 91.27 | *SASC+SDDC* |
| credit | 9.3 | 88.76 | 9.99 | 92.96 | 7.7 | 90.29 | 8.75 | 89.86 | *SASC+SDDC* |
| glass | 13.08 | 72.51 | 14.76 | 84.55 | 12.23 | 76.58 | 12.54 | 84.51 | *SASC+SDDC* |
| hepatitis | 11.03 | 90.03 | 13.04 | 88.05 | 10.01 | 92.37 | 11.56 | 95.05 | *SASC+SDDC* |
| iris | 10.66 | 86.99 | 12.69 | 87.66 | 9.55 | 88.82 | 9.72 | 90.39 | *SASC+SDDC* |
| lymphography | 18.92 | 96.31 | 19.42 | 94.33 | 14.52 | 96.66 | 14.86 | 96.69 | *SASC+SDDC* |
| mushrooms | 14.65 | 98.22 | 14.96 | 92.51 | 14.66 | 93.55 | 14.93 | 95.71 | *CM* |
| Pima-indians | 8 | 93.09 | 8.95 | 90.42 | 8.14 | 93 | 8.22 | 91.89 | *CM* |
| post-operative | 3.33 | 83.46 | 5.38 | 90 | 3 | 89.66 | 3.33 | 88.86 | *SASC+SDDC* |
| thyroid | 18.3 | 86.16 | 20.6 | 89.31 | 13.59 | 88.19 | 15.38 | 89.16 | *SASC+SDDC* |
| voting | 2.5 | 100 | 4.9 | 97.5 | 2.52 | 98.4 | 2.59 | 99.06 | *CM* |
| waveform | 18.53 | 96.87 | 22.7 | 97.2 | 13.55 | 97.6 | 15.59 | 97.93 | *SASC+SDDC* |
| wine | 3.66 | 92.94 | 6.46 | 96.2 | 3.1 | 92.9 | 4.01 | 94.04 | *SASC+SDDC* |
| zoo | 18.81 | 100 | 21.32 | 100 | 12.25 | 100 | 14.87 | 100 | *SASC+SDDC* |
| Average | 11.11 | 92.33 | 12.95 | 93.47 | 9.46 | 93.23 | 10.26 | 93.97 | *SASC+SDDC* |

From the results, we can make several observations and conclusions. Generally speaking, *SASC+SDDC* obtains a balanced behavior, with good storage reduction and generalization accuracy among all the 18 data sets, where the *SASC+SDDC*

method is better than the other three approaches over 14 data sets and *CM* is superior in 3 of the data sets. For the single methods, i.e., the *CM*, *SASC* and *SDDC* methods, the *SASC* has the best accuracy, although it has a bigger size. The reason is obvious since it generates more cases in sparse space such that, this method will provide more information for new cases. In the second place, the *SDDC* provides good performance and the least size, e.g., the average accuracy and size of the *SDDC* is 10.26 and 93.97, respectively, while the counterparts of *CM* and *SASC* are 11.11, 92.33 and 12.95, 93.47, respectively.

## Conclusions and Future Work

Case Based maintenance is one of the most important issues in the Artificial Intelligence field. In this paper, we have introduced the *SASC, SDDC* and their hybrid model *SASC+SDDC,* as an approach to maintaining the size of case base, as well as, make an effort to enhance the competence of the CB. In contrast to current methods, all these innovations derive from the distribution of cases in the space and the importance of different cases. Comparative experiments began with the parameters choice for *SDDC* followed by the performance analysis between the proposed *SASC*, *SDDC*, *SASC+SDDC* and the two standard methods as the Random Deletion and CNN. Finally, 18 datasets were used from the UCI repository to study the validity and performance of the proposed methods.

However, before closing we would like to emphasize that this research has spotlighted the current modeling state of CB competence and represents the tip of the iceberg for case-base maintenance in complex scenarios. Obviously, our experiments need to be extended to include a broader range of traditional maintenance techniques, such as, the typical Wilson-editing methods [28]. Much remains to be done in refining this approach and providing a richer model. Such work will include refining the performance metrics, considering both retrieval and adaptation costs and combining performance/size metrics to achieve metrics that balance both factors in a desirable way.

Another future direction includes application of the model in other CBR domains. We believe that, ultimately, the hybrid approach to maintaining a CB, will inevitably incorporate a range of ideas from a variety of maintenance approaches.

## Acknowledgements

This work is partially supported by the National Natural Science Foundation of China under grants 61300078 and 61175048.

## References

[1] A.Aamodt, and E.Plaza, “Case-based reasoning: Foundational issues, methodological variation, and systems”, AI Communication, Vol. 7, No. 1, pp. 36-59, 1994

[2] L.-D.Mantaras, R. McSherry, D.Bridge, et al., “Retrieval, Reuse, Revision, and Retention in CBR”, Knowledge Engineering Review, Vol. 20, No. 3, pp. 215-240, 2005

[3] V.Jalali, and D.Leake, "Adaptation-Guided Case Base Maintenance", Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, pp.1875-1881, 2014
[4] B. Smyth, and P.Cunningham, "The utility problem analysed: A case-based reasoning perspective", In Proceedings of the Third European Workshop on Case-Based Reasoning, pp. 392-399, 1996
[5] D.-B. Leake and D.-C. Wilson, "Maintaining Case-Based Reasoners: Dimensions and Directions", Computational Intelligence, Vol. 17, pp. 196-213, 2001
[6] B. Smyth, M.Keane, M.San, "Remembering to forget: A competence-preserving case deletion policy for case-based reasoning systems", In Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, pp. 377-382, 1995
[7] C. -H. Chou, B. -H Kuo and F. Chang, "The Generalized Condensed Nearest Neighbor Rule as A Data Reduction Method", 18th International Conference on Pattern, pp.556-559, 2006
[8] D.Aha, D.Kibler, and M.Albert, "Instance-based learning algorithms", Machine Learning, Vol. 6, No.1, pp. 37-66, 1991
[9] S. Delany, and P.Cunningham, "An analysis of case base editing in a spam filtering system", In Advances in Case-Based Reasoning, pp. 128-141, 2004
[10] F.Angiulli, "Fast condensed nearest neighbour rule", In Proceedings of the twenty-second international conference on Machine learning, pp. 25-32, 2005
[11] S.Craw, S.Massie, and N.Wiratunga, "Informed case base maintenance: A complexity profiling approach", In Proceedings of the Twenty-Second National Conference on Artificial Intelligence, pp. 1618-1621, 2007
[12] K.Hanney, and M.Keane, "Learning adaptation rules from a case-base", In Proceedings of the Third European Workshop on Case-Based Reasoning", pp. 179-192, 1996
[13] V.Jalali, and D.Leake, "A context-aware approach to selecting adaptations for case-based reasoning", Lecture Notes in Computer Science, Vol. 8175, pp. 101-114, 2013
[14] V.Jalali, and D.Leake, "Extending case adaptation with automatically-generated ensembles of adaptation rules", In Case-Based Reasoning Research and Development, ICCBR2013, pp. 188-202, 2013
[15] B. Smyth, and E.McKenna, "Building compact competent case-bases", In Proceedings of the Third International Conference Case-Based Reasoning, pp.329-342, 1999
[16] A.Smiti and Z.Elouedi, "Competence and Performance-Improving approachfor maintaining Case-Based Reasoning Systems", In Proceedings on Computational Intelligence and Information Technology, pp.231-236, 2012
[17] K.Racine, and Q.Yang, "On the consistency Management of Large Case Bases: the Case for Validation", AAAI Technical Report-Verification and Validation Workshop, 1996

[18] M. -K. Haouchine, B. -CMorello and N. Zerhouni, "Competence-Preserving Case-Deletion Strategy for Case-Base Maintenance", 9th European Conference on Case-Based Reasoning, pp.171-184, 2008
[19] B.Smyth, "Case-Base Maintenance", Proceedings of the 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, pp.507-516,1998
[20] J. Zhu, "Similarity Metrics and Case Base Maintenance", the School of Computing Science, University of British Columbia, 1998
[21] K.Bradley, B.Smyth, "An architecture for case-based personalised search", Lecture Notes in Computer Science, Vol.3155, pp. 518-532, 2004
[22] N.-V. Chawla, K.-W. Bowyer, L -O Hall, and W. -P.Kegelmeyer, "SMOTE Synthetic Minority Over-sampling Technique", Journal of Artificial Intelligence Research", Vol.16, pp. 321–357, 2002
[23] D.-R. Wilson and T.-R. Martinez. "Reduction techniques for Instance-Based Learning Algorithms", Machine Learning, Vol.38, pp. 257-286, 2000
[24] A.-J. Parkes. "Clustering at the phase transition", Proceedings of the Fourteenth National Conference on Artificial Intelligence, pp. 340-345, 1997
[25] D.-L.Wilson, "Asymptotic Properties of Nearest Neighbour Rules Using Edited Data", IEEE Transactions on Systems, Man, and Cybernetics, Vol.2, pp.408-421, 1972
[26] S.Markovitch, and P.Scott, "Information filtering: Selection mechanisms in learning systems", Machine Learning, Vol. 10, No.2, pp. 113–151,1993
[27] P. -E. Hart, "The condensed nearest neighbour rule", IEEE Transactions on Information Theory, Vol. 14, pp.515–516, 1968
[28] C. Blake, E.Keogh, C.-J. Merz, UCI Repository of machine learning databases, http://www.ics.uci.edu/~mlearn/MLRepository.html, 1998

# Reorthogonalization Methods in ABS Classes

**József Abaffy**

John von Neumann Faculty of Informatics, Óbuda University
Bécsi út 96/b, H-1034 Budapest, Hungary
abaffy.jozsef@nik.uni-obuda.hu

**Szabina Fodor**

Department of Computer Science, Corvinus University of Budapest
Fővám tér 13-15, H-1093 Budapest, Hungary
szabina.fodor@uni-corvinus.hu

*Abstract: In this paper we analyze two subclasses of ABS class of methods which produce orthogonal projection vectors. We theoretically prove that the "twice is enough" selective reorthogonalization criterion of Parlett-Kahan [14] and of Hegedüs [8] can be used in the various ABS classes. Here we also provide a detailed numerical analysis of these ABS-based algorithms. We revealed that the ABS-based algorithm combined with the modified Parlett-Kahan criterion by Hegedüs provided more accurate results in the three considered cases (the rank of the coefficient matrix, the determination of the orthogonal bases, and the QR factorization) than the built-in rank and qr MATLAB functions.*

*Keywords: ABS methods; orthogonalization; reorthogonalization; Gram-Schmidt; Parlett-Kahan*

# 1 Introduction

Orthogonalization is an important step of matrix calculation and produces matrices that are much easier to work with. Let $A=(a_1,...,a_n)$ be a $m\times n$ matrix ($m\geq n$) with full column rank ($rank(A)=n$). Orthogonalization yields an orthogonal basis $Q=(q_1,...,q_n)$ of $span(A)$ such that $A=QR$, where $R$ is an upper triangular matrix. There are several approaches and algorithms for the orthogonalization of a matrix, including the Gram-Schmidt (GS) orthogonalization algorithm, as well as conceptually different approaches such as Householder transformations or Givens rotations [2], [5], [9], [10], [11].

Though the above approaches provide a theoretical foundation of orthogonalization, the calculation of the actual orthogonal basis can be problematic. Indeed, while it is critical that the computed vector $\overline{Q} = (\overline{q}_1,...,\overline{q}_n)$ is as close to the theoretical vector $Q = (q_1,...,q_n)$ as possible, this is often limited by the precision level of the computer used. The issue of limited accuracy of computing $Q$ is known as the orthogonal basis problem [11].

Several investigators have also proposed to perform the orthogonalization several times (called "reorthogonalization"), though the benefit of repeated reorthogonalization is questionable and it seems that one reorthogonalization (i. e. two consecutive orthogonalization steps) is sufficient to ensure the orthogonality of the computed vectors at high accuracy [14], [3], and [4].

Reorthogonalization can also be performed as an optional step, depending on a certain criterion applied during the execution of the algorithm. This situation is called "selective reorthogonalization" and is based on a criterion dependent on the quality of the computed vector. A further assessment of this selective step is provided by Parlett [14] who analyzed the use of two vectors attributed to Kahan (Parlett-Kahan algorithm). Parlett showed that while two consecutive orthogonalization steps improved the accuracy of the computation, further orthogonalization steps failed to provide additional benefit, establishing the principle of "twice is enough". Recently, Hegedüs [8] provided a new theoretical basis for Parlett-Kahan's "twice is enough" algorithm and a modified reorthogonalization criterion.

In this paper we apply the classical Parlett-Kahan (PK) criterion and its modification by Hegedüs (referred to as the modified Parlett-Kahan or MPK criterion) on the ABS class of methods [1]. We considered the S1 and S4 subclasses of ABS, which generate orthogonal directions. Here we also summarize the characteristics of these ABS-based algorithms, and describe their implementation using Matlab R2007b. The numerical experiments revealed that the calculations of the rank of a matrix and of the orthogonal vectors, as well as the QR factorization are more accurate with the usage of our ABS based algorithms than the functions implemented in Matlab.

Finally we should emphasize that the ABS-based algorithms are easy to parallelize. This feature expands the practical usefulness of our algorithms. The results presented in this paper may provide important novel aspects of the efficient parallel implementation of matrix calculations.

# 2 The Parlett-Kahan and Modified Parlett-Kahan Algorithms

The twice-is-enough algorithm of Parlett and Kahan is based on the following orthogonalization step. Let $z$ be the vector to be orthogonalized. The equation

$$p = \left( I - \frac{yy^T}{\|y\|^2} \right) z = \text{orth}(y, z) \tag{1}$$

is the exact orthogonalization of $z$. An undesirable feature of the Gram-Schmidt method is that in the presence of rounding errors the vector $p$ can be far from orthogonal to $y$. This loss of orthogonality is signaled by cancellation, which magnifies previous rounding errors, which in turn will generally contain components in $y$ . A solution for this problem is to repeat the procedure on the vector $p$ .

It has been observed that one reorthogonalization is usually sufficient to produce a vector that is orthogonal to working accuracy - i.e., "twice is enough".

Obviously in reality we have only a numerical approximation of $p$ , say $x'$. Let the error $e' \equiv x' - p$ satisfy $\|e'\| = \varepsilon_M \|z\|$, where $\varepsilon_M$ is the machine precision unit and let $\kappa$ be any fixed value in the range $\left[ \frac{1}{0.83 - \varepsilon_M}, \frac{0.83}{\varepsilon_M} \right]$[14].

**Algorithm-1** ***Parlett-Kahan algorithm - (PK)*** **[14]**

Calculate $x' = \text{orth}(y, z)$ , where $\text{orth}(*)$ is given in (1).

**Case 1**: If $\|x'\| \geq \|z\| / \kappa$ then accept $x = x'$ and $e = e'$ ,

otherwise compute $x'' = \text{orth}(y, x')$ with error

$$e'' \equiv x'' - \left( I - \frac{yy^T}{\|y\|^2} \right) x'$$

satisfying $\|e''\| = \varepsilon_M \|x'\|$ and go to **Case 2.**

**Case 2***:* If $\|x''\| \geq \|x'\| / \kappa$ then accept $x = x''$ and $e = e'' - p.$

**Case 3***:* If $\left\|x^{''}\right\| < \left\|x^{'}\right\| / \kappa$ then accept $x = 0$ and $e = -p.$

**Remark 2.1** *The vector x computed by the algorithm ensures that* $\|e\| \le (1 + \frac{1}{\kappa})\varepsilon_M \|z\|$ *and* $\left|y^T x\right| \le \kappa \cdot \varepsilon_M \|y\|\|x\|$ *[14].*

Hegedüs [8] reformulated the original Parlett-Kahan algorithm and proved the improvement of orthogonality and gave new criteria for the cancellation due to calculation and estimation for the accuracy of computation.

**Algorithm-2** ***modified Parlett-Kahan algorithm - (MPK)*** *[8]*

Here we consider one step of Gram-Schmidt orthogonalization with respect to cancellation.

Let $Q = (q_1, q_2, \ldots, q_{k-1}) \in \Re^{n\times(k-1)}$ and $a \in \Re^n$ be known and accurate. Vector $a$ is orthogonalized to the subspace spanned by the orthonormal columns of matrix $Q$ with one Gram-Schmidt step

$$\theta_k q_k = (I - QQ^T)a, \tag{2}$$

where $\theta_k$ is the norm of $\|a\|$. Compute

$$\eta = \frac{\theta_k}{\|a\|}.$$

If $\eta < \eta_{\min}$ then linear dependency has been detected, i.e. vectors $a, q_1, q_2, \ldots, q_{k-1}$ are linearly dependent at least computationally

If $\eta \ge \eta_{\max}$ then $q_k$ is accepted, otherwise perform a reorthogonalization step.

$$\theta_k \hat{q}_k = (I - QQ^T)q_k$$

where $\theta_k$ is the norm of $\|q_k\|$. The vector $\hat{q}_k$ is accepted, and update $\eta_{\min} = \|Q^T q_k\|$.

**Remark 2.2** *The initial value for* $\eta_{\min}$ *is* $4 \cdot \varepsilon_M$ *and Hegedüs proved that* $-\log_{10}\eta_{\min}$ *is the number of accurate digits.*

**Remark 2.3** *Hegedüs showed that when the incoming vectors are exact and there are accurate digits in the computation, then one may expect the fulfillment of condition* $\eta_{\max} \le \eta$ *after the second orthogonalization step at most. The largest*

*choice of* $\eta_{\max}$ *is* $1/\sqrt{2}$ *to fulfill the condition. Hence the resulting vector* $\hat{q}_k$ *can be considered orthogonal to* $q_1, q_2, \ldots, q_{k-1}$ *up to computational accuracy [8].*

# 3 The Symmetric Scaled ABS Algorithm with Reorthogonalization

In this section, we briefly present the scaled symmetric ABS algorithm [1] and we apply reorthogonalization on the projection vectors ( $p_i$ ) of these subclasses. The ABS algorithm was developed to solve systems of linear and non-linear equations. However, the ABS-based algorithms can be used for many other purposes, for example we can use these algorithms to compute an orthogonal basis of $span(A)$.

Instead of the original equation $Ax = b$, where $A \in \Re^{m,n}$, $b \in \Re^m$, $x \in \Re^n$, consider the scaled equations

$$V^T Ax = V^T b \tag{3}$$

where, $V = (v_1, \cdots, v_m) \in \Re^{m,m}$,is a nonsingular matrix. The set of the solutions of the equations (3) is the same as the set of the solution of $Ax = b$. Applying the non-scaled basic symmetric ABS algorithm for solving (3), we can obtain the symmetric scaled ABS algorithm. Denote the residual vector by $r(x) = Ax - b$. We remark here that not all orthogonalization processes need the residual vector.

**Algorithm-3** ***Symmetric scaled ABS algorithm with reprojection***

**Step1 -** Initialization

Let $x_1 \in \Re^n$ arbitrary, $H_1 = I \in \Re^{n,n}$, where $I$ is the unit matrix, $i = 1$ , and $iflag = 0$.

**Step2**

Let $v_i \in \Re^m$ be arbitrary provided that $v_1, \cdots, v_i$ are linearly independent. Compute the residual error vector $r_i$.

If $r_i = 0$ then stop; $x_i$ solves the equations

otherwise compute

$$\tau_i = v_i^T r_i$$
$$s_i = H_i A^T v_i .$$

**Step3** - Linear dependency check

If $s_i \neq 0$ then go to **Step4.**

If $s_i = 0$ and $\tau_i = 0$ then set

$$x_{i+1} = x_i$$
$$H_{i+1} = H_i$$
$$iflag = iflag + 1$$

and if $i < m$ then go to **Step2** otherwise stop; $x_i$ solves the equations.

If $s_i = 0$ and $\tau_i \neq 0$ , stop set $iflag = -i.$ (*Incompatibility*).

**Step4** - Compute the search direction $p_i$ by

$$p_i = H_i^T z_i$$

where $z_i \in \Re^n$ is arbitrary saving for $z_i^T H_i A^T v_i \neq 0$. Compute the reorthogonalization step

$$p_i = H_i^T p_i .$$

**Step5 -** Update the approximate solution by

$$x_{i+1} = x_i - \alpha_i p_i$$

where the step size $\alpha_i$ is given by $\alpha_i = \frac{\tau_i}{v_i^T A p_i}$.

If $i = m$ then stop; $x_{m+1}$ is the solution of the equations.

**Step6** - Update the projection matrix $H_i$

$$H_{i+1} = H_i - \frac{H_i A^T v_i \cdot v_i^T A H_i}{v_i^T A H_i \cdot H_i A^T v_i} . \tag{4}$$

**Step7 -** Set $i = i + 1$ and go to **Step2**.

**Remark 3.1** *The projection vectors $p_i$ are orthogonal [1].*

**Remark 3.2** *Observe that **Step5** is not always needed if we only wish to solve the linear system of equations.*

**Remark 3.3** *Note that the denominator of the projection matrix ( $H_i$ ) is non-zero because of **Step3**.*

**Theorem 3.1** *Define $\overline{A}^i$ as*

$$\overline{A}^i = \left( \frac{H_1 A^T v_1}{\left\| H_1 A^T v_1 \right\|}, \ldots, \frac{H_i A^T v_i}{\left\| H_i A^T v_i \right\|} \right). \tag{5}$$

*The $\overline{A}^i$ matrices are full rank. Moreover, if $H_1 = I$ , then the columns of $\overline{A}^i$ are mutually orthogonal [1].*

**Remark 3.4** *Using the notation of (5) we have the following alternative formula for (4)*

$$H_{i+1} = I - \overline{A}^i \cdot \overline{A}^{iT}. \tag{6}$$

**Remark 3.5** *Note that the choices of the $z_i \in \Re^n$ and the $v_i \in \Re^m$ are arbitrary saving for $z_i^T H_i A^T v_i \neq 0$. We considered two subclasses of the symmetric, scaled ABS algorithm, designated the **S1** and **S4** subclasses where the search vectors $p_i$ are orthogonal.*

## 3.1 Reorthogonalization in Symmetric, Scaled ABS Algorithm using the Idea of the PK and Modified PK Algorithms

In this section we use the original Parlett-Kahan and the modified Parlett-Kahan algorithms in the scaled symmetric ABS algorithms. We only describe the **Step4** of symmetric scaled ABS algorithm where we determine the searching vectors as the other steps correspond to the steps of the original algorithm.

The ABS Parlett-Kahan algorithm is based on the following orthogonalization step.

Let $z$ be a vector to be orthogonalized. Then

$$p_0 = \left( H - \frac{p \cdot p^T}{\left\| p \right\|^2} \right) z \tag{7}$$

where $p$ is the accurate vector computed at the $ith$ step and $p_0$ is the exact

orthogonalization of $z$. Obviously we can have an approximation of $p_0$ only, say $x'$. Let the error $e' \equiv x' - p_0$ satisfy $\|e'\| = \varepsilon_1 \|z\|$ [16].

**Algorithm-4** ***Reorthogonalization in ABS with the Parlett-Kahan algorithm (ABS-PK)***

**Step4-PK** Compute the search direction $p_i$ by

$$p_i = H_i^T z_i$$

where $z_i \in \Re^n$ is arbitrary saving for $z_i^T H_i A^T v_i \neq 0.$

If $\|p_i\| \geq \frac{\|z_i\|}{\kappa}$ then accept $p_i$ and using the notation of the original Parlett-Kahan $x = p_i$

otherwise compute

$$\hat{p}_i = H_i^T p_i$$

If $\|\hat{p}_i\| \geq \frac{\|p_i\|}{\kappa}$

then accept $\|\hat{p}_i\|$ i.e. $p_i = \hat{p}_i$ and using the notation of the original Parlett-Kahan $x = p_i = \hat{p}_i$

otherwise linear dependency is detected

$$\begin{aligned} p_i &= 0 \\ x_{i+1} &= x_i \\ H_{i+1} &= H_i \\ iflag &= iflag + 1 \end{aligned}$$

go to **Step2.**

**Lemma 3.1.1** *The vector $x$ computed by the algorithm ensures that* $\|e\| \leq (1+1/\kappa)\varepsilon_M \|z\|$ *and* $|p^T x| \leq \kappa \cdot \varepsilon_M \|p\| \|x\|.$

**Proof** We recall Parlett's proof for the case of our orthogonalization algorithm. Differences are only in those steps where the orthogonalization expression is explicitly used. We consider the two cases:

*Case 1*:

$$\|e\| = \|e^{'}\| \leq \varepsilon_1 \|z\|$$

$$\left|p^T x\right| = \left|p^T x^{'}\right| = \left|p^T e^{'}\right| \leq \|p\| \|e^{'}\| \leq \varepsilon_1 \|p\| \|z\| \leq \kappa \varepsilon_1 \|p\| \|x\|$$

because $\left|p^T p_0\right| = 0$ and it follows $\|z\| \leq \kappa \|x^{'}\| = \kappa \|x\|$

*Case 2*:

$$\left|p^T x\right| = \left|p^T x^{"}\right| = \left|p^T e^{"}\right| \leq \|p\| \|e^{"}\| \leq \varepsilon_2 \|p\| \|x^{'}\| = \varepsilon_2 \|p\| \|x\|$$

by the definition of $x^{'}$. Now

$$\|e\| = \|x^{"} - p\| = \|e^{"} + (H - p \cdot p^T / \|p\|^2) x^{'} - p_0\|$$

where we used the definition of $x^{"}$ From this we get using the definition of $x^{'}$ and that $\left|p^T p_0\right| = 0$ that

$$\|e\| \leq \|e^{"}\| + \|H\| \|(I - p \cdot p^T / \|p\|^2)(e^{'} + p_0) - p_0\|$$

$$\leq \varepsilon_2 \|x^{'}\| + \|(I - p \cdot p^T / \|p\|^2) e^{'}\|$$

therefore

$$\leq \varepsilon_2 \|x^{'}\| + \|e^{'}\| \leq \varepsilon_2 \|x^{'}\| + \varepsilon_1 \|z\| \leq \max(\varepsilon_1, \varepsilon_2)(1 + 1/\kappa) \|z\|$$

because the reorthogonalization has to be performed when $\|x^{'}\| < \|z\| / \kappa$. ■

**Theorem 3.1.2** *The* $\|x^{"} - p_o\| \leq \|x^{'} - p_o\|$, *where* $p_0$ *is the exact orthogonalization of* $z$.

**Proof** Using the definition of $x^{"}$ we get

$$\|x^{"} - p_o\| = \|H\left(I - pp^T / \|p\|^2\right) x^{'} - p_o\|$$

$$= \left\| H\left(I - pp^T / \|p\|^2\right)(x' - p_o) \right\| \le \left\| (x' - p_o) \right\|$$

where we used the orthogonality of $p$ and $p_o$ and

$$\left\| H\left(I - pp^T / \|p\|^2\right) \right\| \le \|H\| \left\| I - pp^T / \|p\|^2 \right\| = 1. \blacksquare$$

**Theorem 3.1.3** $\|x'' - p\| \le \|x' - p\|$ *where* $p$ *is the accurate orthogonal vector of the Parlett-Kahan algorithm.*

**Proof** If $H = I$ and $p = y$ of ABS Parlett-Kahan algorithm, then the theorem follows from the Parlett-Kahan algorithm. ■

**Remark 3.1.1** *It is worthwhile to emphasize that the ABS-Parlett-Kahan algorithm is different from the original one, because of the existence of the projection matrix* $H$ *. The projection matrix ensures the new orthogonal vector* $p_i$ *is orthogonal for all previous orthogonal vectors* $p_1$ *, ...,* $p_{i-1}$ *.*

For this algorithm the same lemma is valid as for the original Parlett-Kahan algorithm. It is enough to note that the Euclidean norm of the $H_i$ projection matrices is equal to one. Therefore we recall only the lemma of the algorithm (see [14]) without proof.

**Lemma 3.1.4** *The vector* $x$ *computed by the algorithm ensures that* $\|e\| \le (1 + 1/\kappa)\varepsilon\|z\|$ *and* $\left|p^T x\right| \le \kappa\varepsilon\|p\|\|x\|$.

**Algorithm-5** ***Reorthogonalization in ABS with the modified Parlett-Kahan algorithm (ABS-MPK)***

**Step4-MPK** Compute the search direction $p_i$ by

$$\theta_i p_i = H_i^T z_i$$

where $z_i \in \Re^n$ is arbitrary provided that $z_i^T H_i A^T v_i \neq 0$ and $\theta_i = \left\| H_i^T z_i \right\|$.

Compute

$$\eta = \frac{\theta_i}{\|z_i\|}$$

and

$$\eta_{\min} = \frac{\left\| \overline{A}^{i} p_i \right\|}{\left\| p_i \right\|}$$

If $\eta < \eta_{\min}$

then linear dependency is detected

$$p_i = 0$$
$$x_{i+1} = x_i$$
$$H_{i+1} = H_i$$
$$iflag = iflag + 1$$

go to **Step2.**

If $\eta \geq \eta_{\max}$

then $p_i$ is accepted

otherwise a reorthogonalization is needed

$$\theta_i p_i = H_i^T p_i$$

where $\theta_i = \left\| H_i^T p_i \right\|$.

**Remark 3.1.2** *We should emphasize that the computational demands of checking the computational linear dependency is very low. It is sufficient to check the value of the computed* $\eta$ *and* $\eta_{\min}$.

# 4 Numerical Experiments

Next we were interested in numerical features of the different ABS algorithms. To this end, two variants of the scaled symmetric ABS class of algorithms were implemented in MATLAB version R2007b [12].

**S1 (Subclass-S1 of ABS)**: $z_i$ is selected such that

$$z_i = A^T v_i, \text{ where } v_i = e_i \text{ the } ith \text{ unit vector}.$$

**S4 (Subclass-S4 of ABS)**: $z_i$ is selected such that

$$z_i = r_i, \text{ where } r_i \text{ is the } ith \text{ residual vector}.$$

As we did not want to check the accuracy of the solution we defined it for the sake of minimizing the error of the residual vector calculation in the S4 subclass as $(1,1,...1)$.

The below experiments were performed on a Toshiba personal computer with Intel i7-2620M CPU with integrated graphics, 8 GB RAM and 450 GB hard disk drive, running Microsoft Windows 7 Professional and MATLAB version R2007b. No software other than the operating system tasks, MATLAB and ESET NOD32 antivirus were running during the experiments.

The experiments testing the calculated rank of the coefficient matrix $A$, the orthogonal deviation
$(-\log_{10}(\max(\max(\ abs(I-QQ^T))))$, and the error of $QR$ factorization
$(-\log_{10}(\max(\max(\ abs(A-QR))))$ were performed.

First we tested the different ABS based algorithms on randomly generated dense, symmetric, positive definite, full rank matrices (*SPD*). The random matrices were generated using MATLAB.

We use the following annotations:

***ABS-S1 (AREO)***: symmetric scaled ABS algorithm with S1 selection, reprojection in every step, linear dependency check according to Hegedüs.

***ABS-S1 (PK)***: symmetric scaled ABS algorithm with S1 selection and reorthogonalization using the Parlett-Kahan algorithm.

***ABS-S1 (MPK)***: symmetric scaled ABS algorithm with S1 selection and reorthogonalization using the modified Parlett-Kahan algorithm.

***ABS-S4 (AREO)***: symmetric scaled ABS algorithm with S4 selection, reprojection in every step, linear dependency check according Hegedüs.

***ABS-S4 (PK)***: symmetric scaled ABS algorithm with S4 selection and reorthogonalization using the Parlett-Kahan algorithm.

***ABS-S4 (MPK)***: symmetric scaled ABS algorithm with S4 selection and reorthogonalization using the modified Parlett-Kahan algorithm.

As shown in Fig. 1, there were no significant differences between the accuracy of the calculations for S1 and S4 variants of ABS algorithms. This was in contrast to our expectations based on the fact that the $z_i$ vectors in the selection S4 were calculated as a function of the row vectors of the matrix $A$. The likely reason for the discrepancy is that the well-chosen solution vector minimizes the rounding error even if the approximation of the solution in step $i$ is not exact. The difference between the variants of ABS based algorithms is in the calculated rank.

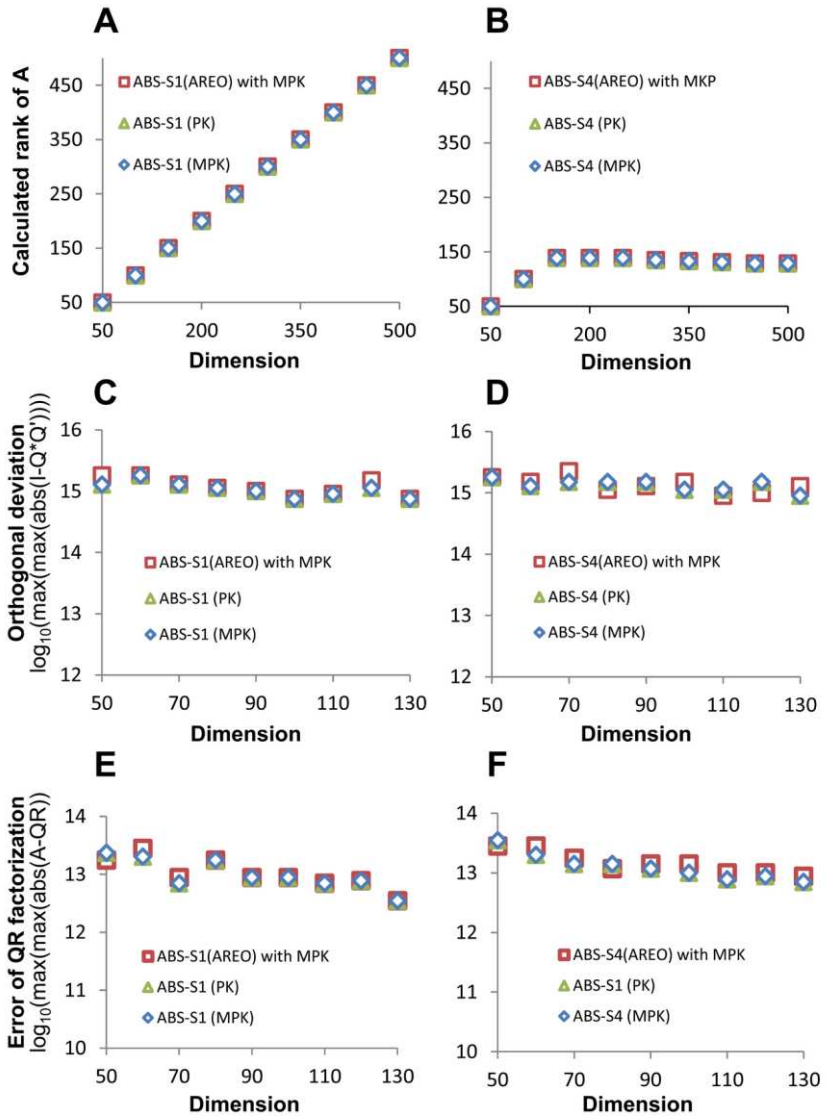



Figure 1
Comparison of different ABS-based algorithms on randomly generated, dense, symmetric, positive definite, full rank matrices

All algorithms based on S4 selection detect incorrect linear dependency in the projection vectors because of the loss of accurate digits in the computation of orthogonal vectors. If the judgment is made on aspects of finding an orthogonal basis between the two subclasses of ABS based algorithms then algorithms with S1 selection outperform the subclass S4. We should note that the subclass S4 has many valuable numerical features [1] and further studies are needed to test them.

Next we compared our best-performing ABS-based algorithm (**ABS-S1 (AREO)**) with other well-known algorithms. We selected the built-in **qr** function of MATLAB and the classic Gram-Schmidt orthogonalization (**CGS**) algorithm [2]. We implemented the CGS algorithm with reorthogonalization in every step and the linear dependency check proposed by Hegedüs.

We should mention that the formula (6) gives a connection between the CGS and the ABS-S1 methods. The ABS-based algorithm with S1 selection is practically the block version of the CGS method. Numerically we could not see a significant difference between the block and the original version of update of the projection matrix $H_i$ (data not shown). However, it is worthwhile to do further such research in the future with alternative formulas of updating the projection matrices.

We tested the algorithms on randomly generated dense, symmetric, positive definite, full rank matrices (*SPD*), Pascal (*Pascal*), normalized Pascal (*normalized Pascal*), Vandermonde (*Vandermonde*), and normalized Vandermonde (*normalized Vandermonde*) matrices. The matrices were generated using built-in MATLAB functions.

It should be mentioned that the ABS-based algorithms use the row of the matrices in the *ith* step while the **CGS** and the **qr** algorithms use the columns of the $A$. To correct for that discrepancy, the transposed matrix $A$ has been applied on the latter two algorithms.

As shown in Fig. 2, we compared the computed rank of the different matrices. Note that when the condition number of the coefficient matrix ($A$) is not extremely high, like for the SPD matrices, then all our algorithms calculated the same, exact rank of matrices (data not shown). However, we could see a significant difference on the computed rank of special (like Pascal, normalized Pascal) matrices, and we wanted to verify the actual rank of the matrices represented in double precision. We used the Multiprecision Computing Toolbox of MATLAB [13] and we recomputed the rank of the matrices originally represented in double precision into quadruple precision. We could not see any loss of the ranks because of the non-exact representations.

After all, we conclude that the MATLAB rank function performed poorly on computing the rank of special matrices, and only the ABS-S1 algorithm gave the exact rank for the normalized matrices $A$. Therefore, it would be worth using the ABS-S1 algorithm for calculating the rank of matrices.

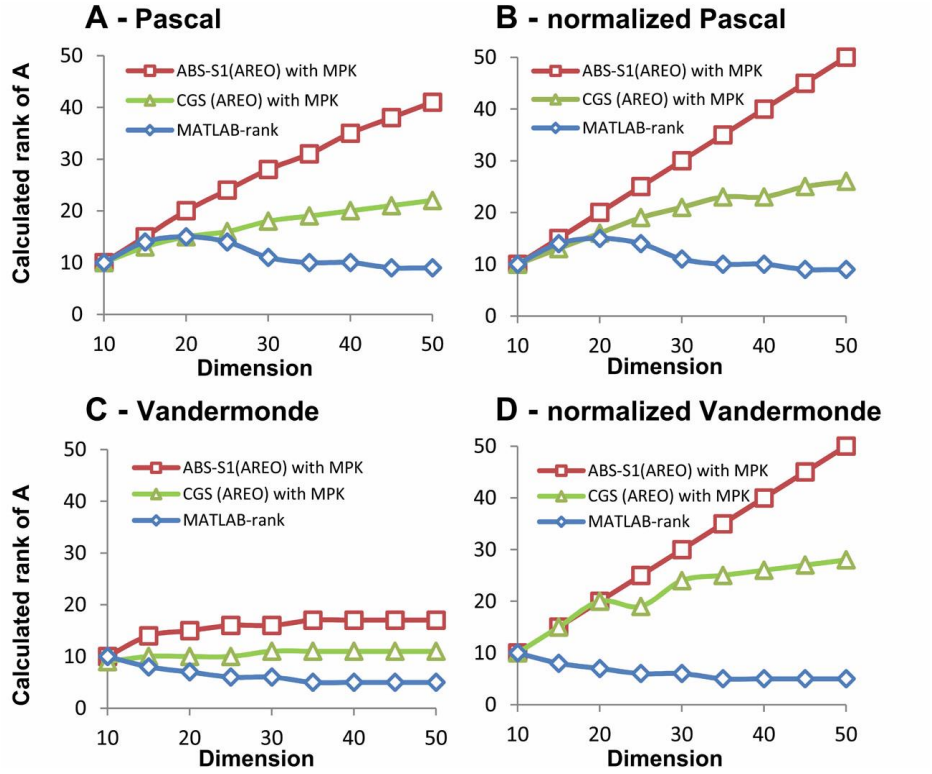



Figure 2
Comparison of the calculated rank of the coefficient matrices

Then we compared the accuracy of the computed orthogonal basis with our algorithms. To describe the deviation of orthogonal deviation, we use the $-\log_{10}(\max(\max(\ abs(I - Q \cdot Q^T))))$ formula where $I$ is the unit matrix and $Q$ is the orthogonal basis. Hegedüs proved [8] that this formula gives the number of accurate digits. Using double precision number presentation (IEEE 754 standard [15]), the maximum value of the formula is less than *16*.

As shown in Fig. 3, the MATLAB qr function computed the orthogonal basis the least accurately for every test case. There is no significant difference between the ABS-S1 and the CGS algorithms for the difficult test problems (like Pascal, Vandermonde). However, the CGS algorithm surpasses the ABS-S1 on well-conditioned matrices like our SPD matrices.

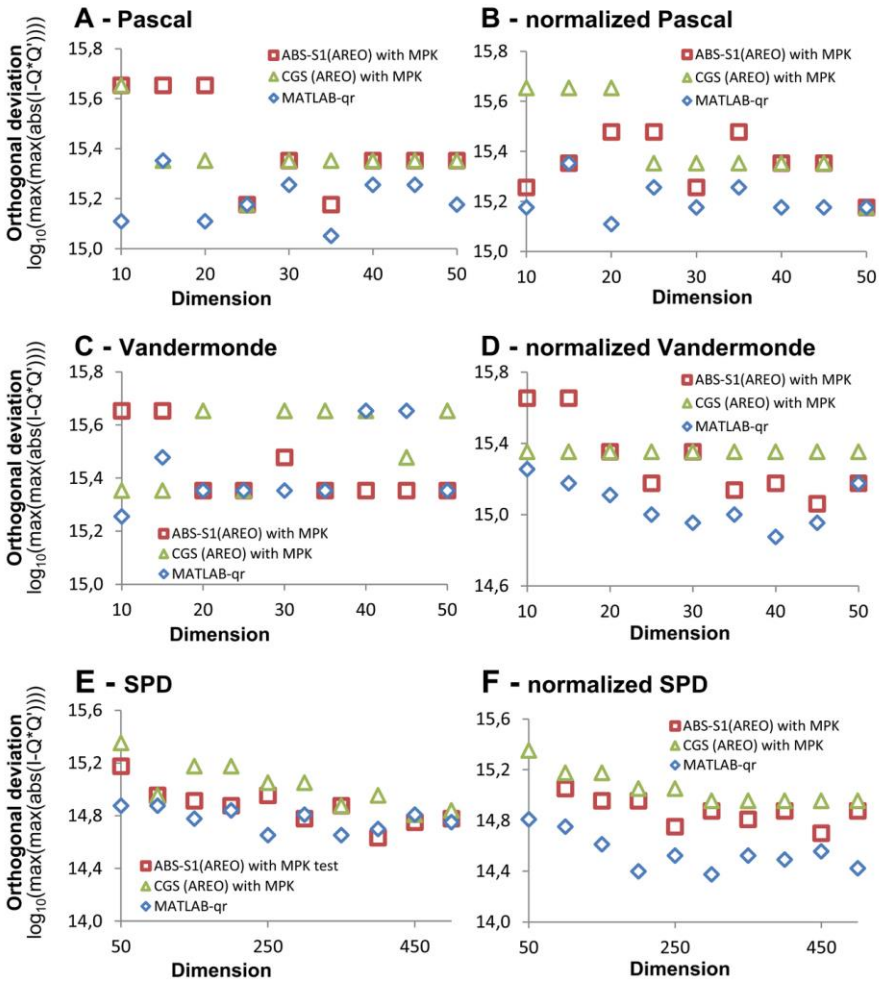



Figure 3
Comparison of the orthogonal deviation

Last we tested the accuracy of the $QR$ factorization. We used a formula $(-\log_{10}(\max(\max(\ abs(A-QR))))$ for describing the precision of $QR$ factorization, which gives the number of accurate digits of the factorization. The maximum value of the accurate digits is less than 16 in floating point representation and larger values mean more precise results.

As shown in Fig. 4 the CGS algorithm surpasses the ABS-S1 and the **qr** algorithms on every case, being the most salient on the normal test cases like SPD and normalized SPD matrices. No significant difference can be seen between the ABS-S1 algorithm and the MATLAB **qr** function.

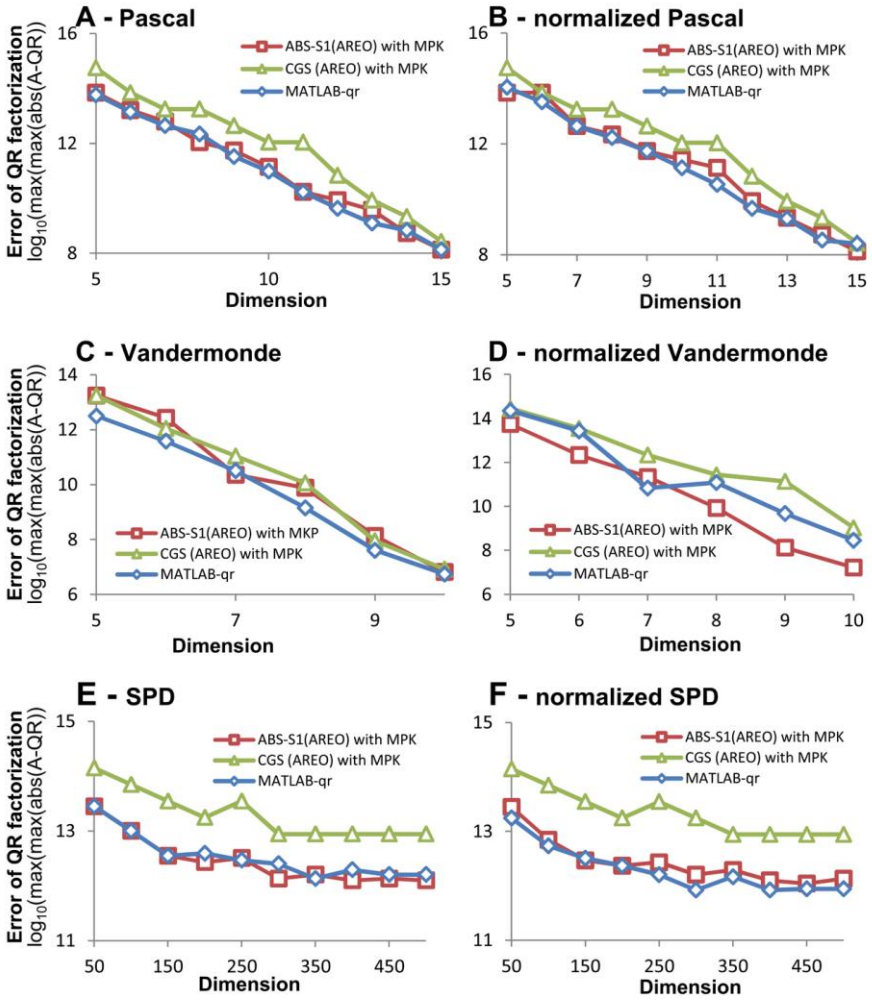



Figure 4
Comparison of the $QR$ factorization

However, we should mention that we compared the accuracy of $QR$ factorization in the range where neither of the tested algorithms found linear dependency. This strictly narrowed down the examined dimension of matrices.

It should be noted that we did not test the algorithms with pivoting the coefficient matrices ($A$). The different pivoting strategies are numerically very expensive and the accuracy of the calculations (at least $10^{-14}$) did not substantiate these efforts. However, further analyses will need to be performed for testing the effects of the pivoting strategies.

**Discussion and Conclusions**

Here we presented new ABS based algorithms for creating orthogonal basis. We numerically verified that the modified Parlett-Kahan (MPK) reorthogonalization criterion by Hegedüs can be successfully used. Our numerical experiments revealed that the ABS-based algorithm combined with the modified Parlett-Kahan provided more accurate results in the three considered cases (the rank of the coefficient matrix, the determination of the orthogonal bases, and the ***QR*** factorization) than the built-in MATLAB functions.

**Acknowledgement**

The authors thank Csaba Hegedüs for inspiring discussions and helpful suggestions.

**References**

[1] Abaffy, J., and Spedicato, E., "*ABS Projection Algorithms: Mathematical Techniques for Linear and Nonlinear Equations*", Ellis Horwood Limited, John Wiley and Sons, Chichester, England (1989)

[2] Björck, A Numerical Methods for Least Squares Problems. SIAM, Philadelphia, PA, 1996

[3] Giraud, L.,Langou, J., and Rozloznik, M "On the Round-off Error Analyisis of the Gram-Schmidt Algorithm with Reorthogonalization, CERFACS Technical Report No. TR/PA/02/33, pp. 1-11 (2002)

[4] Giraud, L., Langou, J., and Rozloznik, M. "The Loss of Orthogonality in the Gram-Schmidt Orthogonalization Process", Computers and Mathematics with Applications, Vol. 51, pp. 1069-1075, 2005

[5] Golub, G., H, and Van Loan, C., F Matrix Computations, 3rd ed. John Hopkins Univ. Press, Baltimore, MD (1996)

[6] Daniel, J., W., Gragg, W., B., Kaufman, L., and Stewart, G., W. *"Reorthogonalization and Stable Algorithms for Updating the Gram-Scmidt QR Factorization"*, Mathematics of Computation, Vol. 30, No. 136, pp. 772-795 (1976)

[7] Hegedüs, C. J. "*Short Proofs for the Pseudoinverse in Linear Algebra",* Annales Univ. Sci. Budapest, 44, 115-121 (2001)

[8] Hegedüs, C. J. "*Reorthogonalization Methods Revisited*", Acta Polytechnica, submitted

[9] Hegedüs, C. J. "Generating Conjugate Directions for Arbitrary Matrices by Matrix Equations. I, II", Computers and Mathematics with Applications 21:(1) pp. 71--85, 87--94-71. (1991)

[10] Hegedüs C. J., Bodócs L. "General Recursions for A-Conjugate Vector Pairs", Report No. 1982/56 Central Research Institute for Physics, Budapest

[11] Higham, N. J. *"Accuracy and Stability of Numerical Algorithms",* SIAM, Philadelphia, (1996)

[12] Guide MATLAB User's, The MathWorks Inc, Natick, MA, 2007

[13] Multiprecision Computing Toolbox User's Manual, Advanpix's, Yokohama, Japan.
http://www.advanpix.com/documentation/users-manual/

[14] Parlett, B., N. "*The symmetric Eigenvalue Problem"*, Englewood Cliffs, N. J. Prentice-Hall (1980)

[15] Smith W., S. "Fixed versus Floating Point". The Scientist and Engineer's Guide to Digital Signal Processing. California Technical Pub. p. 514, ISBN 0966017633, 1997 Retrieved December 31, 2012

[16] Wilkinson, J. H. "Rounding Errors in Algebraic Processes", Prentice-Hall (1963)

# A Relational Database Model for Science Mapping Analysis

**Manuel J. Cobo**

Dept. Computer Science, University of Cádiz, Avda. Ramón Puyol s/n 11202 Algeciras, Cádiz, Spain. E-mail: manueljesus.cobo@uca.es

**A. G. López-Herrera, Enrique Herrera-Viedma**

Dept. Computer Science and Artificial Intelligence, University of Granada, C/Periodista Daniel Saucedo Aranda, s/n, E-18071 Granada, Spain. E-mail: lopez-herrera@decsai.ugr.es, viedma@decsai.ugr.es

*Abstract: This paper presents a relational database model for science mapping analysis. It has been specially conceived for use in almost all of the stages of a science mapping workflow (excluding the data acquisition and pre-processing stages). The database model is developed as an entity-relations diagram using information that is typically presented in science mapping studies. Finally, several SQL queries are presented for validation purposes.*

*Keywords: Science Mapping Analysis; Bibliometric mapping; Bibliometric Studies; Bibliometric networks; Co-word; Co-citation; Database model*

# 1 Introduction

Online web-based bibliographical databases, such as ISI Web of Science, Scopus, CiteSeer, Google Scholar or NLM's MEDLINE (among others) are common sources of data for bibliometric research [24]. Many of these databases have serious problems and usage limitations since they were mainly designed for other purposes such as information retrieval, rather than for bibliometric analysis [17].

In order to overcome these limitations, one common solution consists of downloading data from online databases, cleaning it and storing it in customized ad hoc databases [22]. However, it is difficult to find information on how these customized databases are designed or constructed.

The primary procedures in bibliometrics are science mapping analysis and performance analysis. Specifically, science or bibliometric mapping is an important research topic in the field of bibliometrics [23]. It is a spatial representation of how disciplines, fields, specialties and individual documents or authors are related with each other [31]. It focuses on the monitoring of a scientific field and the delimitation of a research area in order to determine its cognitive structure and evolution [7, 11, 19, 21, 25, 26]. In other words, science mapping is aimed at displaying the structural and dynamic aspects of scientific research [4, 23].

In order to analyze the dynamic and complex aspects of scientific activities, in science mapping research, the analyst needs methods of adequately managing the many-to-many relations between data elements (publications, authors, citations, and other variables) over time. The many-to-many relationship has been a recurrent problem in science mapping [8, 9].

Although a number of software tools have been specifically developed to perform a science mapping analysis [9], to our knowledge, there have been no proposals for efficient data modelling in a science mapping context. While some research has been published on basic bibliometrics databases [20, 32], a great deal of work still needs to be done on acceptable method for managing the special needs for data used in science mapping research [15].

Thus, this paper presents the first database model for science mapping analysis. This database model has been specifically conceived for almost all of the stages of the classical science mapping workflow [9] – except for the stages of data acquisition and pre-processing. The database model is developed as an entity-relations diagram using information that is typically presented in science mapping studies [9].

This paper is structured as follows: In Section 2, the relational database model for science mapping analysis is presented. Section 3 offers a proof-of-concept for the relational database model, presenting several *SQL* and *PL/SQL* statements and procedures. Section 4 provides a practical example based on the proposed database model. Finally, some conclusions are drawn.

# 2 The Relational Database Model

Science mapping analysis is a bibliometric technique that allows for the uncovering of the conceptual, social and intellectual aspects of a research field. Furthermore, the longitudinal framework allows us to highlight the evolution of these aspects across time. By combining science mapping analysis with bibliometric indicators [8], the analyst can detect the main topics of a research field, its hot topics and those that have had a greater impact (number of citations)

and have been the most productive. As described in [8], the science mapping analysis may be divided in the following steps:

1) To determine the substructures contained (primarily, clusters of authors, words or references) in the research field by means of a bibliometric network [5, 18, 27, 30] (bibliographic coupling, journal bibliographic coupling, author bibliographic coupling, co-author, co-citation, author co-citation, journal co-citation or co-word analysis) for each studied period.
2) To distribute the results from the first step in a low dimensional space (clusters).
3) To analyze the evolution of the detected clusters based on the different periods studied, in order to detect the main general evolution areas of the research field, their origins and their inter-relations.
4) To carry out a performance analysis of the different periods, clusters and evolution areas, based on bibliometric measures.

Thus, in this section we propose a database model for the science mapping analysis. The database model presented here has been specifically designed to work with co-occurrence networks, building the map through a clustering algorithm, and categorizing the clusters in a strategic diagram. Furthermore, it allows us to perform an evolution analysis in order to reveal the dynamical aspects of the analyzed field.

The database model has been designed to store all the information needed to begin the analysis, to perform the different steps of the methodology proposed in [8] and to maintain the obtained results.

Although the model does not visualize the results (graphically), these results may be easily exported using simple *SQL* queries in order to visualize them via visualization software such as Pajek [2], Gephi [3], UCINET or Cytoscape [29].

Subsequently, we shall discuss the raw data requirements and we shall analyze the conceptual design of the database model.

## 2.1 Requirements

The main aim of science mapping analysis is to extract knowledge from a set of raw bibliographic data. Usually, the analyst has previously downloaded the data from a bibliographic source and subsequently, imported it to a specific data model. For example, the analyst could create a new spread sheet or load the data into a specific database.

Science mapping analysis uses several types of information: baseline or input data, intermediate data and results. The database model for science mapping analysis must be capable of storing these different types of information.

The baseline or input data is normally a set of documents with their set of associated units of analysis (author, references or terms). With this data, the analyst can build a bibliometric network. If the science mapping analysis is performed in a longitudinal framework, the documents should have an associated publication date.

It should be pointed out that the baseline data, especially the units of analysis, must have been pre-processed. In other words, the units of analysis need to have been previously cleaned, any errors should have been fixed, and the de-duplicating process must have been carried out previously. The data and network reduction is carried out using the database model as shown in Section 3.

As for intermediate data, the database model must be able to store a set of datasets (one per analyzed time slice or time period). Each dataset must have an associated set of documents, a set of units of analysis, the relationship between the documents and units, and finally, an undirected graph that represents the bibliometric network. In addition, the database model must be capable of generating intermediate data from the baseline data.

In regard to the resulting data, the database model must be capable of storing one set of clusters per dataset and an evolution map [8]. Also, a clustering algorithm is necessary in order to generate the map; however, since implementing this algorithm via SQL-queries could be a daunting task, it may be implemented using PL/SQL procedures.

Each cluster could have an associated set of network measures, such as Callon's centrality and density measure [6]. Furthermore, each node of a bibliometric network, or even each cluster, may have an associated set of documents, and they could be used to conduct a performance analysis [8]. For example, we may calculate the amount of documents associated with a node, the citations achieved by those documents, the h-index, etc.

Finally, regarding the functionality requirements, the database model should be capable of the following:

- Building a set of datasets from the baseline data.
- Filtering the items of the datasets based on a frequency threshold.
- Extracting a co-occurrence network.
- Filtering the edges of the network using a co-occurrence threshold.
- Normalizing the co-occurrence network based on different similarity measures.
- Extracting a set of clusters per dataset.
- Adding different network measures to each detected cluster.
- Adding a set of documents to each detected cluster.
- Building an evolution map using the clusters of consecutive time periods.

## 2.2 Conceptual Design: the EER Diagram

In this section, we describe the database model, showing the different entities and relations that are necessary to develop a complete science mapping analysis. The Enhanced Entity/Relations (EER) data modeling is used to design the database in a conceptual way [14]. Thus, the EER diagram for the proposed database is shown in Figure 1.

We should point out certain notations that are used in the EER diagram. Each entity is represented as a box and the lines between two boxes represent a relationship between them. The cardinality of a relationship is represented as a number over the line. Each box has a title, which describes the entity's name.

Thus, the proposed database model consists of five blocks: *Knowledge Base*, *Dataset*, *Bibliometric Network*, *Cluster* and *Longitudinal*. Each block is represented by a different color-shadow in Figure 1. These blocks are described below.

The *Knowledge Base* block is responsible for storing the baseline data. In order to conduct the science mapping analysis with the database model presented in this paper, the analyst should provide these entities filled out, or at least, offer the baseline data ready to be inserted in the corresponding entities, typically using *insert-into SQL* statements.

As mentioned previously, in order to perform a science mapping analysis under a longitudinal framework, the baseline data must consist, at least, of four entities: *Document*, *Publish Date*, *Period* and *Unit of Analysis*. In order to fit with different kind and format of baseline date, we should point out that the proposed structure for this block may differ in some aspects from the implementation carried out by the analyst. In that case, only the queries to fill the dataset (Query 1 to 3) should be adapted to the particular structure.

The main entity of the Knowledge Base block is *Document,* which represents a scientific document. This entity stores the primary information, such as the title, abstract or citations received. This is the minimum information to identify a scientific document, and therefore, the analyst may expand upon this entity in order to add more information.

The Knowledge Base block contains two related pieces of longitudinal information: *Publish Date*, and *Period*. The former represents the specific date or year when the document was published. A Document must have one associated Publish Date, and a Publish Date has an associated set of Documents. The latter represents a slice of time or a period of years. The entity Period shall be used to divide the documents into different subsets in order to analyze the conceptual, social or intellectual evolution (depending on the kind of unit of analysis used).

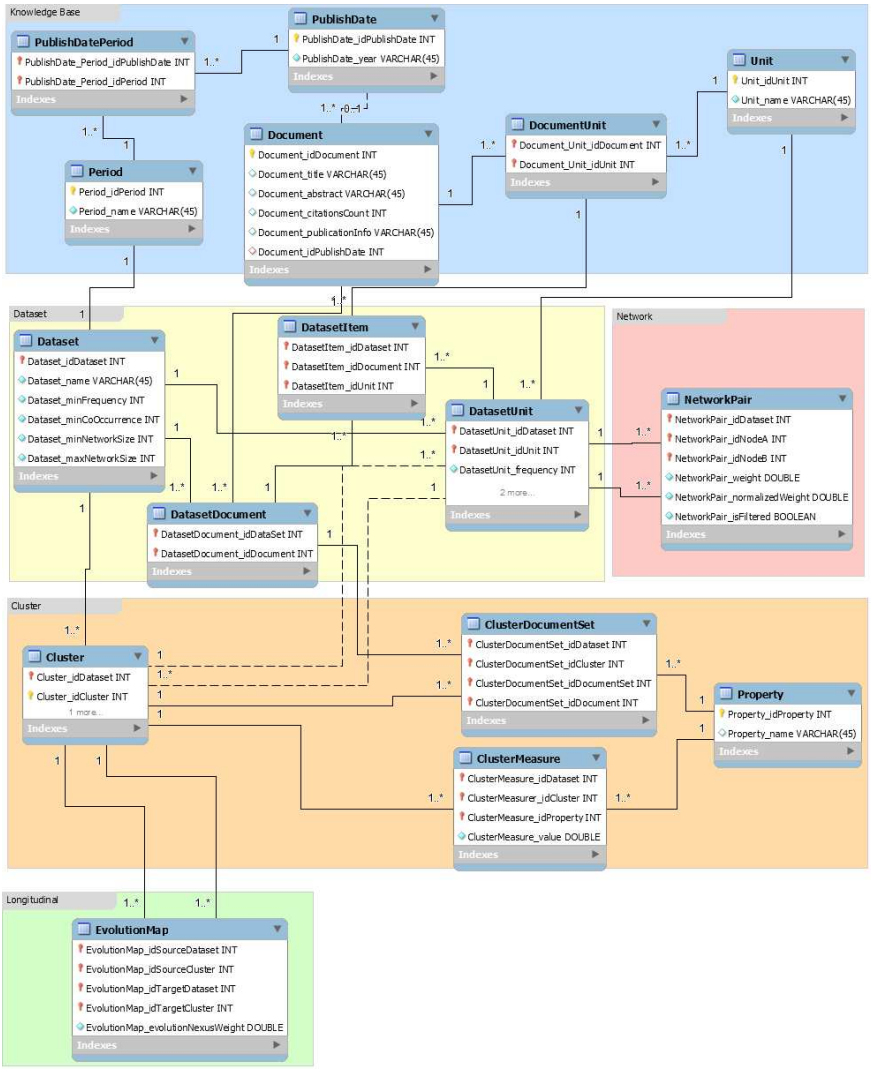



Figure 1
EER diagram

Although the periods are usually disjoint set of years, we should point out that they may not necessarily be disjoint, so a year could appear in various periods. Therefore, a Period consists of a set of Publish Dates, and a Publish Date can belong to one or more Periods.

The *Unit* represents any type of units of analysis that may be associated with a document, usually, the authors, terms or references. Our database model allows for the storage of only one type of unit of analysis at the same time. Regardless, it

shall be easy to replicate the analysis using other types of units. Since the documents usually contain a set of authors, terms and references, there is a *many-to-many* relationship between the Document and Unit entities.

When the science mapping analysis is carried out under a longitudinal framework, all of the data must be split into different slices which are usually referred to as periods. Each slice must contain a subset of documents and its related units of analysis, and therefore, the frequency of the units of analysis may differ for each slice.

Therefore, the *Dataset* block contains the entities that are responsible for storing a slice of the overall data. The Dataset block contains three entities: i) the Dataset, ii) the Documents of the dataset (DatasetDocument in Figure 1) and iii) the Unit of the dataset (DatasetUnit in Figure 1).

The *Dataset* is the main entity of this block. Since the remaining entities are related to a specific dataset, it is necessary during the followings steps. The Dataset is associated with a specific Period and it contains attributes such as a name to describe the slice, different parameters to filter the dataset and the network, and finally, the configuration of the clustering algorithm.

Each dataset consists of a set of documents and units of analysis. The documents are a subset of documents belonging to a specific period and the units are the subset of units associated with these documents. Since the units should be filtered using a minimum frequency threshold, each unit contains a Boolean attribute to indicate if it was filtered or not. Like the many-to-many relationship between the entities *Document* and *Unit*, there is a *many-to-many* relationship between the Documents of the dataset and the Units of the dataset (represented by the entity DatasetItem). It should also be noted that the same document-unit relationship may appear in different datasets.

The *Bibliometric Network* block consists of only one entity (*NetworkPair*) which stores the co-occurrence network of each dataset. Conceptually, this entity stores the network pairs. Each pair consists of a source node and a target node, with each node being a unit of a specific dataset. Each pair contains a weight (usually the co-occurrence count of both nodes in the dataset) and a normalized weight. As in Units, each pair may be filtered using a minimum co-occurrence threshold, thus there is a Boolean attribute to indicate if the pair has been filtered or not.

It is important to note that because the co-occurrence network is an undirected graph, the adjacency network representing the bibliometric network is a symmetric matrix. In other words, the source-target pair, and target-source pair should have the same weights and therefore, the same normalized weights.

The *Cluster* block contains four entities in order to represent the cluster itself and its properties. The Cluster entity belongs to a specific dataset, so there is a *one-to-many relationship* between the Dataset and Cluster entities. Furthermore, a Cluster is associated with a set of units of analysis. Moreover, the Cluster contains an

optional attribute that represents the main node of the cluster, usually the most central node of the associated sub-network [8].

As described in [9], a network analysis and a performance analysis could be applied to a set of clusters. Thus, the database model should be capable of storing the results of both analyses.

Although the results of the performance analysis are usually numerical measures (number of documents, sum of citations, average citations per document, h-index, etc.), they are calculated with a set of documents. Therefore, since each cluster contains a set of nodes (units of analysis) and each node is associated with a set of documents, the set of documents may be associated with each cluster using a *document mapper function* [10].

Thus, each cluster may contain two types of properties: i) measures and ii) a set of documents. As for the possible measures, network measures (Callon's centrality and density measure), or performance measures (citations received by the documents associated with the cluster, and the h-index) may be defined as cluster properties.

Finally, the *Longitudinal* block stores the results of the temporal or longitudinal analysis. Specifically, it contains an entity to represent an *Evolution map* [8].

The evolution map can be defined as a bipartite graph showing the evolving relationship between the clusters of two consecutive periods [8]. Therefore, the Evolution map entity consists of a cluster source and a cluster target (each one belonging to different datasets) and a weight (evolution nexus) to represent the similarity between the source and target clusters.

# 3 Proof of Concept

This section shows how to develop a science mapping analysis using the database model presented in this paper. To accomplish this, we propose using different SQL-queries in order to perform the different steps of the analysis, proving that a science mapping analysis may in fact be conducted based on our database model.

As previously mentioned, the general workflow of a science mapping analysis includes a sequence of steps: data retrieval, data pre-processing, network extraction, network normalization, mapping, analysis, visualization and interpretation. Furthermore, each science mapping software tool [9] customizes or redefines its own workflow based on the general steps.

Thus, in order to develop a science mapping analysis using the database model presented in this paper, the analyst must follow a particular workflow analysis. It is important to note that, as mentioned in Section 2.1, both the data acquisition and data pre-processing tasks must have been carried out previously. That is, the

science mapping analysis workflow described in this section begins with pre-processed data.

The workflow analysis may be divided into four stages:

1) Building the dataset.
2) Extracting the bibliometric network.
3) Applying a clustering algorithm.
4) Carrying out network, performance and longitudinal analyses.

The **first stage** includes building the dataset from the data stored in the Knowledge Base block. For this, the documents, units and their relationships must be divided into distinct subsets, corresponding to each period. The different data-slices must be inserted in the corresponding entities, and associated with a specific dataset. The first stage is divided into five steps:

1) To insert the different datasets and associate them with a period. This step may be carried out using INSERT-INTO SQL statements. Since it is dependent upon the nature of the data, we have not included the queries for this step.
2) To extract the units of analysis for each dataset (Query 1).
3) To extract the documents for each dataset (Query 2).
4) To extract the relationships between documents and units of each dataset (Query 3).
5) To filter the units using the frequency threshold specified for each dataset (Query 4).

```
INSERT INTO DatasetUnit (DatasetUnit_idDataset, DatasetUnit_idUnit, DatasetUnit_frequency)
   SELECT p.Period_idPeriod, u.Unit_idUnit, count(u.Unit_idUnit)
   FROM Period p, PublishDate_Period pup, PublishDate pu, Document d, Document_Unit du, Unit u
   WHERE
         p.Period_idPeriod = pup.PublishDate_Period_idPeriod AND
         pup.PublishDate_Period_idPublishDate = pu.PublishDate_idPublishDate AND
         pu.PublishDate_idPublishDate = d.Document_idPublishDate AND
         d.Document_idDocument = du.Document_Unit_idDocument AND
         du.Document_Unit_idUnit = u.Unit_idUnit
   GROUP BY p.Period_idPeriod, u.Unit_idUnit;
```

Query 1

Retrieving the units of analysis for each dataset

```
INSERT INTO DatasetDocument (DatasetDocument_idDataset, DatasetDocument_idDocument)
   SELECT p.Period_idPeriod, d.Document_idDocument
   FROM Period p, PublishDate_Period pup, PublishDate pu, Document d
   WHERE
         p.Period_idPeriod = pup.PublishDate_Period_idPeriod AND
         pup.PublishDate_Period_idPublishDate = pu.PublishDate_idPublishDate AND
         pu.PublishDate_idPublishDate = d.Document_idPublishDate;
```

Query 2

Retrieving the documents for each dataset

```
INSERT INTO DatasetItem (DatasetItem_idDataset, DatasetItem_idDocument, DatasetItem_idUnit)
   SELECT p.Period_idPeriod, d.Document_idDocument, du.Document_Unit_idUnit
   FROM Period p, PublishDate_Period pup, PublishDate pu, Document d, Document_Unit du
   WHERE
        p.Period_idPeriod = pup.PublishDate_Period_idPeriod AND
        pup.PublishDate_Period_idPublishDate = pu.PublishDate_idPublishDate AND
        pu.PublishDate_idPublishDate = d.Document_idPublishDate AND
        d.Document_idDocument = du.Document_Unit_idDocument;
```

Query 3

Adding the relationships between documents and units (items) of analysis

```
UPDATE DatasetUnit
   SET DatasetUnit_isFiltered = 1
   WHERE DatasetUnit_frequency < (SELECT d.Dataset_minFrequency
        FROM Dataset d
        WHERE d.Dataset_idDataset = DatasetUnit_idDataset);
```

Query 4

Filtering the units of analysis

The **second stage** consists of extracting the bibliometric network for each dataset using the co-occurrence relation between the units of analysis. Then, the co-occurrence relations must be filtered and normalized. This stage is divided into three steps:

1) Building the bibliometric network by searching the co-occurrence relationships (Query 5).

2) Filtering the bibliometric network (pairs) using the specific co-occurrence threshold defined by each dataset. (Query 6).

3) Normalizing the bibliometric network using a similarity measure [13], such as, *Salton's Cosine*, *Jaccard's Index*, *Equivalence Index*, or *Association Strength*. Query 7 shows the general query to carry out the normalization of a bibliometric network. In order to apply a specific similarity measure, the analyst must replace the text "*SIMILARITY-MEASURE*" in Query 7 with one of the formulas shown in Query 8.

```
INSERT INTO NetworkPair (NetworkPair_idDataset, NetworkPair_idNodeA, NetworkPair_idNodeB, NetworkPair_weight)
   SELECT d.DatasetDocument_idDataSet, di1.DatasetItem_idUnit AS nodeA, di2.DatasetItem_idUnit AS nodeB,
        count(DISTINCT d.DatasetDocument_idDocument) AS coOccurrence
   FROM DatasetDocument d, DatasetItem di1, DatasetUnit du1, DatasetItem di2, DatasetUnit du2
   WHERE
        d.DatasetDocument_idDataSet = di1.DatasetItem_idDataset AND
        d.DatasetDocument_idDocument = di1.DatasetItem_idDocument AND
        di1.DatasetItem_idDataset = du1.DatasetUnit_idDataset AND
        di1.DatasetItem_idUnit = du1.DatasetUnit_idUnit AND
        du1.DatasetUnit_isFiltered = 0 AND
        d.DatasetDocument_idDataSet = di2.DatasetItem_idDataset AND
        d.DatasetDocument_idDocument = di2.DatasetItem_idDocument AND
        di2.DatasetItem_idDataset = du2.DatasetUnit_idDataset AND
        di2.DatasetItem_idUnit = du2.DatasetUnit_idUnit AND
        du2.DatasetUnit_isFiltered = 0 AND
        di1.DatasetItem_idDataset = di2.DatasetItem_idDataset AND
        di1.DatasetItem_idUnit != di2.DatasetItem_idUnit
        GROUP BY d.DatasetDocument_idDataSet, di1.DatasetItem_idUnit, di2.DatasetItem_idUnit;
```

Query 5

Extracting the bibliometric network

```
UPDATE NetworkPair
   SET NetworkPair_isFiltered = 1
   WHERE NetworkPair_weight < (SELECT d.Dataset_minCoOccurrence
         FROM Dataset d
         WHERE d.Dataset_idDataset = NetworkPair_idDataset);
```

Query 6

Filtering the bibliometric network

```
UPDATE NetworkPair
   SET NetworkPair_normalizedWeight = (SELECT SIMILARITY-MEASURE
         FROM DatasetUnit du1, DatasetUnit du2
         WHERE
               NetworkPair.NetworkPair_idDataset = du1.DatasetUnit_idDataset AND
               NetworkPair.NetworkPair_idDataset = du2.DatasetUnit_idDataset AND
               NetworkPair.NetworkPair_idNodeA = du1.DatasetUnit_idUnit AND
               NetworkPair.NetworkPair_idNodeB = du2.DatasetUnit_idUnit);
```

Query 7

Normalizing the bibliometric network

```
Association strength:
   n.NetworkPair_weight / (du1.DatasetUnit_frequency * du2.DatasetUnit_frequency)
Equivalence index:
   (n.NetworkPair_weight * n.NetworkPair_weight) / (du1.DatasetUnit_frequency * du2.DatasetUnit_frequency)
Inclusion index:
   n.NetworkPair_weight / MIN(du1.DatasetUnit_frequency, du2.DatasetUnit_frequency)
Jaccard index:
   n.NetworkPair_weight / (du1.DatasetUnit_frequency + du2.DatasetUnit_frequency - n.NetworkPair_weight)
Salton index:
   (n.NetworkPair_weight * n.NetworkPair_weight) / SQRT(du1.DatasetUnit_frequency * du2.DatasetUnit_frequency)
```

Query 8

Similarity measures to normalize the bibliometric network

The **third stage** involves applying a clustering algorithm in order to divide each bibliometric network into a subset of highly connected sub-networks. Implementing a clustering algorithm using simple SQL-queries is a difficult and daunting task. Therefore, the clustering algorithm should be implemented as a store-procedure using PL/SQL.

A variety of clustering algorithms are commonly used in science mapping analysis [8, 9]. For example, we have implemented the simple centers algorithm [12] as a store-procedure (See Appendix A on: *http://sci2s.ugr.es/scimat/sma-dbmodel/AppendixA.txt*), a well-known algorithm in the context of science mapping analysis.

The **fourth stage** involves conducting several analyses of the clusters, networks and datasets. As mentioned earlier, each detected cluster may be associated with different properties: i) a set of documents and ii) network or performance measures. The process used to associate a new property to the clusters is divided into two consecutive steps:

1) Adding a new property in the table Property.

2) Calculating the property values and associate them to each cluster.

If the new property is based on a performance measure (number of documents, number of citations, etc.), an intermediate step is necessary. In this case, the analyst must first associate a set of documents to each cluster.

As an example of this, we present different queries to perform a network and a performance analysis. Based on the approach developed in [8], in order to layout the detected clusters in a strategic diagram, the Callon's centrality and density measures [8, 6] must be assessed. Callon's centrality (Query 9) measures the external cohesion of a given cluster. That is, it measures the relationship of each cluster with the remaining clusters. On the other hand, Callon's density (Query 10) measures the internal cohesion of a cluster.

```
INSERT INTO Property(Property_idProperty, Property_name) VALUES(PROPERTY_ID, 'centrality');

INSERT INTO ClusterMeasure(ClusterMeasure_idDataset, ClusterMeasure_idCluster, ClusterMeasure_idProperty,
        ClusterMeasure_value)
    SELECT n.NetworkPair_idDataset, c.Cluster_idCluster, PROPERTY_ID,
        10 * SUM(n.NetworkPair_normalizedWeight)
    FROM Cluster c, NetworkPair n, DatasetUnit du1, DatasetUnit du2
    WHERE
        c.Cluster_idDataset = n.NetworkPair_idDataset AND
        n.NetworkPair_isFiltered = 0 AND
        n.NetworkPair_idNodeA < n.NetworkPair_idNodeB AND
        n.NetworkPair_idDataset = du1.DatasetUnit_idDataset AND
        n.NetworkPair_idNodeA = du1.DatasetUnit_idUnit AND
        n.NetworkPair_idDataset = du2.DatasetUnit_idDataset AND
        n.NetworkPair_idNodeB = du2.DatasetUnit_idUnit AND
        du1.DatasetUnit_idCluster <> du2.DatasetUnit_idCluster AND
        (c.Cluster_idCluster = du1.DatasetUnit_idCluster OR
        c.Cluster_idCluster = du2.DatasetUnit_idCluster)
    GROUP BY n.NetworkPair_idDataset, c.Cluster_idCluster;
```

Query 9

Calculating the Callon's centrality measure

```
INSERT INTO Property(Property_idProperty, Property_name) VALUES(PROPERTY_ID, 'density');

INSERT INTO ClusterMeasure(ClusterMeasure_idDataset, ClusterMeasure_idCluster, ClusterMeasure_idProperty,
        ClusterMeasure_value)
    SELECT n.NetworkPair_idDataset, du1.DatasetUnit_idCluster, PROPERTY_ID,
        100 * (SUM(n.NetworkPair_normalizedWeight) / itemCluster.items)
    FROM NetworkPair n, DatasetUnit du1, DatasetUnit du2,
        (SELECT du.DatasetUnit_idCluster AS idCluster, count(*) AS items
        FROM DatasetUnit du
        GROUP BY du.DatasetUnit_idCluster) AS itemCluster
    WHERE
        du1.DatasetUnit_idCluster = itemCluster.idCluster AND
        n.NetworkPair_isFiltered = 0 AND
        n.NetworkPair_idNodeA < n.NetworkPair_idNodeB AND
        n.NetworkPair_idDataset = du1.DatasetUnit_idDataset AND
        n.NetworkPair_idDataset = du2.DatasetUnit_idDataset AND
        n.NetworkPair_idNodeA = du1.DatasetUnit_idUnit AND
        n.NetworkPair_idNodeB = du2.DatasetUnit_idUnit AND
        du1.DatasetUnit_idCluster = du2.DatasetUnit_idCluster
    GROUP BY n.NetworkPair_idDataset, du1.DatasetUnit_idCluster;
```

Query 10

Calculating the Callon's density measure

As mentioned earlier, in order to carry out a performance analysis, a set of documents must be associated to each cluster. This process may be conducted using a *document mapper function* [10]. For example, in Query 11, the *Union*

*document mapper function* [10] is shown. This function associates each cluster to all of the documents related to its nodes.

```
INSERT INTO Property(Property_idProperty, Property_name) VALUES(PROPERTY_ID, 'unionDocuments');

INSERT INTO ClusterDocumentSet(ClusterDocumentSet_idDataset, ClusterDocumentSet_idCluster,
        ClusterDocumentSet_idDocumentSet, ClusterDocumentSet_idDocument)
    SELECT DISTINCT c.Cluster_idDataset, c.Cluster_idCluster, PROPERTY_ID, dd.DatasetDocument_idDocument
    FROM Cluster c, DatasetUnit du, DatasetItem di, DatasetDocument dd
    WHERE
        c.Cluster_idDataset = du.DatasetUnit_idDataset AND
        c.Cluster_idCluster = du.DatasetUnit_idCluster AND
        du.DatasetUnit_idDataset = di.DatasetItem_idDataset AND
        du.DatasetUnit_idUnit = di.DatasetItem_idUnit AND
        di.DatasetItem_idDataset = dd.DatasetDocument_idDataset AND
        di.DatasetItem_idDocument = dd.DatasetDocument_idDocument;
```

Query 11

Retrieving documents associated with each cluster

Once a document set is added, the analyst may calculate certain performance measures, such as the number of documents associated with each cluster (Query 12) or the number of citations attained from those documents (Query 13).

```
INSERT INTO Property(Property_idProperty, Property_name) VALUES(PROPERTY_ID, 'unionDocumentsCount');

INSERT INTO ClusterMeasure(ClusterMeasure_idDataset, ClusterMeasure_idCluster, ClusterMeasure_idProperty,
        ClusterMeasure_value)
    SELECT cds.ClusterDocumentSet_idDataset, cds.ClusterDocumentSet_idCluster, PROPERTY_ID,
        COUNT(cds.ClusterDocumentSet_idDocument)
    FROM Cluster c, ClusterDocumentSet cds
    WHERE
        c.Cluster_idDataset = cds.ClusterDocumentSet_idDataset AND
        c.Cluster_idCluster = cds.ClusterDocumentSet_idCluster AND
        cds.ClusterDocumentSet_idDocumentSet = DOCUMENT_SET_ID
    GROUP BY cds.ClusterDocumentSet_idDataset, cds.ClusterDocumentSet_idCluster;
```

Query 12

Counting the number of documents associated with each cluster

```
INSERT INTO Property(Property_idProperty, Property_name) VALUES(PROPERTY_ID, 'unionDocumentCitationsCount');

INSERT INTO ClusterMeasure(ClusterMeasure_idDataset, ClusterMeasure_idCluster, ClusterMeasure_idProperty,
        ClusterMeasure_value)
    SELECT cds.ClusterDocumentSet_idDataset, cds.ClusterDocumentSet_idCluster, PROPERTY_ID,
        SUM( d.Document_citationsCount )
    FROM Cluster c, ClusterDocumentSet cds, DatasetDocument dd, Document d
    WHERE
        cds.ClusterDocumentSet_idDocumentSet = DOCUMENT_SET_ID AND
        c.Cluster_idDataset = cds.ClusterDocumentSet_idDataset AND
        c.Cluster_idCluster = cds.ClusterDocumentSet_idCluster AND
        cds.ClusterDocumentSet_idDataset = dd.DatasetDocument_idDataset AND
        cds.ClusterDocumentSet_idDocument = dd.DatasetDocument_idDocument AND
        dd.DatasetDocument_idDocument = d.Document_idDocument
    GROUP BY cds.ClusterDocumentSet_idDataset, cds.ClusterDocumentSet_idCluster;
```

Query 13

Counting the citations attained from the documents associated to each cluster

Complex bibliometric indicators, such as the h-index [1, 16], are difficult to assess using only SQL-queries. Therefore, the analyst may implement a store procedure using PL/SQL to calculate this measure (see Appendix B on:

*http://sci2s.ugr.es/scimat/sma-dbmodel/AppendixB.txt*). Query 14 shows the SQL-Query to add the h-index to each cluster.

```
INSERT INTO Property(Property_idProperty, Property_name) VALUES(PROPERTY_ID, 'unionDocumentH-Index');

CALL calculateHIndexPerCluster(PROPERTY_ID, DOCUMENT_SET_ID);
```

Query 14

Calculating the h-index of each cluster

It should be noted that, in order to assess a specific network or performance property, the analyst must replace the text "*PROPERTY_ID*" in Queries 9, 10, 11, 12, 13 and 14 with the corresponding property ID which is a positive integer that uniquely identifies a property.

The longitudinal analysis may be carried out using an evolution map [8]. To do this, it is necessary to measure the overlap between the clusters of two consecutive periods using Query 15.

```
INSERT INTO EvolutionMap(EvolutionMap_idSourceDataset, EvolutionMap_idSourceCluster,
        EvolutionMap_idTargetDataset, EvolutionMap_idTargetCluster, EvolutionMap_evolutionNexusWeight)
    SELECT c1.Cluster_idDataset, c1.Cluster_idCluster, c2.Cluster_idDataset, c2.Cluster_idCluster,
        SIMILARITY-MEASURE
    FROM Cluster c1, DatasetUnit du1, Cluster c2, DatasetUnit du2,
        (SELECT c3.Cluster_idDataset, c3.Cluster_idCluster, COUNT(du3.DatasetUnit_idCluster) AS items
        FROM Cluster c3, DatasetUnit du3
        WHERE
            c3.Cluster_idDataset = du3.DatasetUnit_idDataset AND
            c3.Cluster_idCluster = du3.DatasetUnit_idCluster
        GROUP BY c3.Cluster_idDataset, c3.Cluster_idCluster) AS freqSubq1,
        (SELECT c4.Cluster_idDataset, c4.Cluster_idCluster, COUNT(du4.DatasetUnit_idCluster) AS items
        FROM Cluster c4, DatasetUnit du4
        WHERE
            c4.Cluster_idDataset = du4.DatasetUnit_idDataset AND
            c4.Cluster_idCluster = du4.DatasetUnit_idCluster
        GROUP BY c4.Cluster_idDataset, c4.Cluster_idCluster) AS freqSubq2
    WHERE
        c1.Cluster_idDataset = c2.Cluster_idDataset - 1 AND
        c1.Cluster_idDataset = du1.DatasetUnit_idDataset AND
        c1.Cluster_idCluster = du1.DatasetUnit_idCluster AND
        c2.Cluster_idDataset = du2.DatasetUnit_idDataset AND
        c2.Cluster_idCluster = du2.DatasetUnit_idCluster AND
        du1.DatasetUnit_idUnit = du2.DatasetUnit_idUnit AND
        c1.Cluster_idDataset = freqSubq1.Cluster_idDataset AND
        c1.Cluster_idCluster = freqSubq1.Cluster_idCluster AND
        c2.Cluster_idDataset = freqSubq2.Cluster_idDataset AND
        c2.Cluster_idCluster = freqSubq2.Cluster_idCluster
    GROUP BY c1.Cluster_idDataset, c1.Cluster_idCluster, c2.Cluster_idDataset, c2.Cluster_idCluster;
```

Query 15

Building the evolution map

Query 15 must be completed using a similarity measure. For this, the analyst must replace the text "SIMILARITY MEASURE" in Query 15 with one of the formulas shown in Query 16.

```
Association strength:
    COUNT(du1.DatasetUnit_idUnit) / (freqSubq1.items * freqSubq2.items)
Equivalence index:
    (COUNT(du1.DatasetUnit_idUnit) * COUNT(du1.DatasetUnit_idUnit)) / (freqSubq1.items * freqSubq2.items)
Inclusion index:
    COUNT(du1.DatasetUnit_idUnit) / LEAST(freqSubq1.items, freqSubq2.items)
Jaccard index:
```

```
    COUNT(du1.DatasetUnit_idUnit) / (freqSubq1.items + freqSubq2.items - COUNT(du1.DatasetUnit_idUnit))
Salton index:
    (COUNT(du1.DatasetUnit_idUnit) * COUNT(du1.DatasetUnit_idUnit)) / SQRT(freqSubq1.items * freqSubq2.items)
```

Query 16

Similarity measures to build the evolution map

Finally, due to the limitations of database management systems, the visualization step is difficult to carry out, since it relies on other software tools [9]. Although the analyst may obtain a custom report (i.e., a performance measures summary, Query 17), network visualization cannot be made with SQL-Queries. To overcome this problem, the analyst could develop a store procedure using PL/SQL to export the entire network into a Pajek format file [2] (See Appendix C on: http://sci2s.ugr.es/scimat/sma-dbmodel/AppendixC.txt).

```
SELECT d.Dataset_name AS 'Period name', u.Unit_name AS 'Cluster name',
        c1.ClusterMeasure_value AS 'Number of documents',  c2.ClusterMeasure_value AS 'Number of citations',
    c3.ClusterMeasure_value AS 'h-Index'    FROM ClusterMeasure c1, ClusterMeasure c2, ClusterMeasure c3, Cluster c, DatasetUnit du,
Unit u, Dataset d      WHERE
        c1.ClusterMeasure_idProperty = DOCUMENTS_COUNT_PROPERTY_ID AND
        c2.ClusterMeasure_idProperty = CITATIONS_PROPERTY_ID AND
        c3.ClusterMeasure_idProperty = H-INDEX-PROPERTY_ID AND
        c1.ClusterMeasure_idDataset = c2.ClusterMeasure_idDataset AND
        c1.ClusterMeasure_idDataset = c3.ClusterMeasure_idDataset AND
        c1.ClusterMeasure_idCluster = c2.ClusterMeasure_idCluster AND
        c1.ClusterMeasure_idCluster = c3.ClusterMeasure_idCluster AND
        c1.ClusterMeasure_idDataset = c.Cluster_idDataset AND
        c1.ClusterMeasure_idCluster = c.Cluster_idCluster AND
        c.Cluster_idDataset = du.DatasetUnit_idDataset AND
        c.Cluster_mainNode = du.DatasetUnit_idUnit AND
        du.DatasetUnit_idUnit = Unit_idUnit AND
        c1.ClusterMeasure_idDataset = d.Dataset_idDataset
    ORDER BY c1.ClusterMeasure_idDataset ASC, c1.ClusterMeasure_value DESC;
```

Query 17

Performance measures summary

# 4 A Practical Example

In this section we present some results that could be obtained using the described queries in a real dataset. As an example, we take advantage of the dataset used by [8], which contains the keywords of the documents published by the two most important journals of the Fuzzy Set Theory research field [28, 33, 34]. The complete description and a full analysis of this dataset may be found at [8]. It is possible to download both the test dataset and the entire SQL-Queries to build the proposed database model and perform the analysis at *http://sci2s.ugr.es/scimat/sma-dbmodel/sma-dbmodel-example.zip*.

To summarize, in order to carry out a science mapping analysis, the analyst must configure the database according to his/her interests (choosing the units of analysis, periods, document mapper functions, performance measures, etc.), selecting and executing the appropriate SQL statements. So, for example, let us imagine that the analyst wishes to study the conceptual evolution of a specific

scientific field over three consecutive periods by means of a co-word analysis, using: a) keywords as units of analysis, b) union documents as document mapping function, c) documents and citations as performance measures, d) equivalence index as a normalization measure, e) simple centers algorithm as a clustering algorithm, f) Callon's centrality and density as network measures, and g) inclusion index as an overlapping function between periods. In this context, the analyst should carry out the following steps:

1) Particularize the Knowledge Base to record the correct information, i.e., the table Unit must be used to record the keywords details (keyword text and unique identification).

2) Create the needed Datasets using SQL Queries 1, 2, 3 and 4. So, if three periods are used by the analyst, three Datasets must be created / defined. For each Dataset, the correct parameters of analysis should be defined. For example, Dataset.minFrequency = 4, Dataset.minCoOccurrence = 3, Dataset.minNetworkSize = 7 and Dataset.maxNetworkSize = 12 as shown in Query 18.

3) Build the corresponding networks for each dataset using Query 5 and to filter them using the Query 6. After this, the networks must be normalized using the Equivalence index. To do this, Query 7 must be used, but replacing the text SIMILARITY-MEASURE by the Equivalence index formula shown in Query 8. The results of this step may be seen in Figure 2.

| NetworkPair_idDataset | NetworkPair_idNodeA | NetworkPair_idNodeB | NetworkPair_normalizedWeight | NetworkPair_isFiltered |
|---|---|---|---|---|
| 1 | 1 | 639 | 0.02040816326530612 | 1 |
| 1 | 1 | 628 | 0.19047619047619047 | 0 |
| 1 | 1 | 1914 | 0.02857142857142857 | 1 |
| 1 | 1 | 7838 | 0.03571428571428571 | 1 |
| 1 | 10 | 57 | 0.022222222222222223 | 1 |
| 1 | 10 | 247 | 0.022222222222222223 | 1 |
| 1 | 10 | 392 | 0.008547008547008548 | 1 |
| 1 | 10 | 438 | 0.027777777777777776 | 1 |
| 1 | 10 | 682 | 0.037037037037037035 | 1 |
| 1 | 10 | 1881 | 0.037037037037037035 | 1 |
| ... | ... | ... | ... | ... |

Figure 2
Network example output

4) Apply a clustering algorithm for dividing each network into different clusters (or themes). For example, the PL/SQL code presented in Appendix A has to be run in order to use the centers simple algorithm. The results of this step is shown in Figure 3.

| Clustre_idDataset | Cluster_idCluster | Cluster_mainNode |
|---|---|---|
| 1 | 4 | 100 |
| 1 | 3 | 148 |
| 1 | 1 | 371 |
| 1 | 5 | 826 |
| 1 | 6 | 1085 |
| 1 | 2 | 1247 |
| 1 | 7 | 4887 |
| 2 | 5 | 1 |
| 2 | 2 | 100 |
| 2 | 7 | 121 |
| ... | ... | ... |

Figure 3
Clusters example output

5) Associate documents and performance measures (Callon's centrality and density, documents, citations count, and h-index) chosen by the analyst for each cluster. To do this, Queries 9, 10, 12, 13 and 14 must be run. For example, Figure 4 shows a subset of the centrality measures associated with each cluster.

6) Build the evolution map in order to carry out a longitudinal analysis. For this, Query 15 must be run. In this example the Inclusion index formula (see Query 16) has to be set as SIMILARITY-MEASURE. For example, Figure 5 shows the results obtained for the first two consecutive periods.

After applying these steps, all of the data needed to carry out a longitudinal science mapping analysis should be recorded in the database. The data may be used directly and interpreted by the analyst or it may be exported for use by third-party software [9]. To do this, once again standard SQL statements (i.e., Query 17) or *PL/SQL* procedures may be used. An example of the output generated by this step is seen in Figure 6.

| ClusterMeasure_idDataset | ClusterMeasure_idCluser | ClusterMeasure_idProperty | ClusterMeasure_value |
|---|---|---|---|
| 1 | 1 | 1 | 2.6481481481481484 |
| 1 | 4 | 1 | 0.14814814814814814 |
| 1 | 5 | 1 | 2.5 |
| 2 | 1 | 1 | 17.717840383480777 |
| 2 | 2 | 1 | 7.060470629927034 |
| 2 | 3 | 1 | 6.0620245118848555 |
| 2 | 4 | 1 | 3.1692448390270735 |
| 2 | 5 | 1 | 7.332406699242617 |
| 2 | 7 | 1 | 3.4178775040159826 |
| 2 | 8 | 1 | 18.37574280055123 |
| ... | ... | ... | ... |

Figure 4

Centrality measure example output

| idSourceDataset | idSourceCluster | idTargetDataset | idTargetCluster | evolutionNexusWeight |
|---|---|---|---|---|
| 1 | 1 | 2 | 4 | 0.142857142 |
| 1 | 1 | 2 | 5 | 0.142857142 |
| 1 | 1 | 2 | 7 | 0.142857142 |
| 1 | 2 | 2 | 7 | 0.25 |
| 1 | 2 | 2 | 8 | 0.5 |
| 1 | 3 | 2 | 3 | 0.25 |
| 1 | 3 | 2 | 7 | 0.25 |
| 1 | 4 | 2 | 2 | 0.5 |
| 1 | 5 | 2 | 8 | 0.333333333 |
| ... | ... | ... | ... | |

Figure 5

Evolution map example output

| Period name | Cluster name | Number of documents | Number of citations | h-index |
|---|---|---|---|---|
| Subperiod 1978-1989 | DECISION-MAKING | 64 | 1131 | 14 |
| Subperiod 1978-1989 | FUZZY-CONTROL | 54 | 1648 | 18 |
| Subperiod 1978-1989 | FUZZY-RELATIONAL-EQUATIONS | 38 | 1229 | 19 |
| Subperiod 1978-1989 | FUZZY-TOPOLOGY | 36 | 382 | 13 |
| Subperiod 1978-1989 | RELATIONS | 21 | 1155 | 7 |
| Subperiod 1978-1989 | FUZZY-MAPPING | 19 | 407 | 11 |
| Subperiod 1978-1989 | SUBGROUP | 13 | 226 | 6 |
| Subperiod 1990-1994 | NEURO-FUZZY-SYSTEMS | 205 | 4135 | 34 |
| Subperiod 1990-1994 | FUZZY-NUMBERS | 194 | 3518 | 31 |
| Subperiod 1990-1994 | FUZZY-CONTROL | 172 | 6076 | 40 |
| ... | ... | ... | ... | ... |

Figure 6

Report example output

## Concluding Remarks

This paper presents a relational database model for science mapping analysis. This database model was conceived specifically for use in almost all of the stages of a science mapping workflow. The database model was developed as an entity-relations diagram based on the information that is typically present in science mapping studies [9]. The database model also allows for application of the methodology for science mapping analysis proposed in [8]. To validate the proposal, several SQL statements and three PL/SQL procedures were presented.

Finally, it is important to note that one of the most important advantages of this proposal is that it allows for the implementation of a science mapping analysis in an easy, free and cheap way, using only standard SQL statements, which are present in most database management systems.

## Acknowledgements

This work was conducted with the support of Project TIN2010-17876, the Andalusian Excellence Projects TIC5299 and TIC05991, and the Genil Start-up Projects Genil-PYR-2012-6 and Genil-PYR-2014-15 (CEI BioTIC Granada Campus for International Excellence).

## References

[1] S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, F. Herrera. “h-index: A Review Focused in Its Variants, Computation and Standardization for Different Scientific Fields”. *Journal of Informetrics*, 3(4), pp. 273-289, 2009

[2] V. Batagelj, A. Mrvar. “Pajek – Program for Large Network Analysis. Connections”. 1998

[3] M. Bastian, S. Heymann, M. Jacomy. “Gephi: an Open Source Software for Exploring and Manipulating Networks”. In: *Proceedings of the International AAAI Conference on Weblogs and Social Media*. 2009

[4] K. Börner, C. Chen, K. Boyack. “Visualizing Knowledge Domains”. *Annual Review of Information Science and Technology*, 37, pp. 179-255, 2003

[5] M. Callon, J. P. Courtial, W. A. Turner, S. Bauin. “From Translations to Problematic Networks: An Introduction to Co-Word Analysis”. *Social Science Information*, 22, pp. 191-235, 1983

[6] M. Callon, J. P. Courtial, F. Laville. “Co-Word Analysis as a Tool for Describing the Network of Interactions between Basic and Technological Research - the Case of Polymer Chemistry”. *Scientometrics*, 22, pp. 155-205, 1991

[7] M. J. Cobo, F. Chiclana, A. Collop, J. de Oña, E. Herrera-Viedma. "A Bibliometric Analysis of the Intelligent Transportation Systems Research

based on Science Mapping". IEEE Transactions on Intelligent Transportation Systems, 11(2), pp. 901-908, 2014

[8] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, F. Herrera. "An Approach for Detecting, Quantifying, and Visualizing the Evolution of a Research Field: A Practical Application to the Fuzzy Sets Theory Field". *Journal of Informetrics*, 5(1), pp. 146-166, 2011

[9] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, F. Herrera. "Science Mapping Software Tools: Review, Analysis and Cooperative Study among Tools". *Journal of the American Society for Information Science and Technology*, 62, pp. 1382-1402, 2011

[10] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, F. Herrera. "SciMAT: A New Science Mapping Analysis Software Tool". *Journal of the American Society for Information Science and Technology*, 63, pp. 1609-1630, 2012

[11] M. J. Cobo, M. A. Martínez, M. Gutiérrez-Salcedo, H. Fujita, E. Herrera-Viedma. "25 Years at Knowledge-based Systems: A Bibliometric Analysis". Knowledge-Based Systems, 80, pp. 3-13, 2015

[12] N. Coulter, I. Monarch, S. Konda. "Software Engineering as Seen through its Research Literature: A Study in Co-Word Analysis". *Journal of the American Society for Information Science*, 49, pp. 1206-1223, 1998

[13] N. J van Eck NJ, L. Waltman. "How to Normalize Co-Occurrence Data? An Analysis of Some Well-known Similarity Measures". *Journal of the American Society for Information Science and Technology*, 60, pp. 1635-1651, 2009

[14] Elmasri, R.; Navathe, S. B. (2011) Fundamentals of Database Systems (6th ed.) Pearson/Addison Wesley

[15] W. Gänzel. National Characteristics in International Scientific Coauthorship Relations. *Scientometrics*, 51, pp. 69-115, 2001

[16] J. Hirsch. "An Index to Quantify an Individuals Scientific Research Out-Put". *Proceedings of the National Academy of Sciences*, 102, pp. 16569-16572, 2005

[17] W. W. Hood, C. S. Wilson. "Informetric Studies using Databases: Opportunities and Challenges". *Scientometric*s, 58, pp. 587-608, 2003

[18] M. M. Kessler. "Bibliographic Coupling between Scientific Papers". *American Documentation*, 14, pp. 10-25, 1963

[19] A. G. López-Herrera, E. Herrera-Viedma, M. J. Cobo, M. A. Martínez, G. Kou, Y. Shi. "A Conceptual Snapshot of the First 10 Years (2002-2011) of the International Journal of Information Technology & Decision Making". International Journal of Information Technology & Decision Making 11(2), pp. 247-270, 2012

[20] N. Mallig. “A Relational Database for Bibliometric Analysis”. *Journal of Informetrics*, 4, pp. 564-580, 2010

[21] M. A. Martínez-Sánchez, M. J. Cobo, M. Herrera, E. Herrera-Viedma. "Analyzing the Scientific Evolution of Social Work Discipline using Science Mapping". Research on Social Work Practice, 5(2), pp. 257-277, 2015

[22] H. F. Moed. “The Use of Online Databases for Bibliometric Analysis”, *Proceedings of the First International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval*, pp. 133-146, 1988

[23] S. Morris, B. van Der Veer Martens. “Mapping Research Specialties”. *Annual Review of Information Science and Technology*, 42, pp. 213-295, 2008

[24] C. Neuhaus, H. Daniel. “Data Sources for Performing Citation Analysis: an Overview”. *Journal of Documentation*, 64, pp. 193-210, 2008

[25] E. C. M. Noyons, H. F. Moed, A. F. J. van Raan. “Integrating Research Performance Analysis and Science Mapping”. *Scientometrics*, 46, 591-604, 1999

[26] Orestes Appel, Francisco Chiclana, Jenny Carter. "Main Concepts, State of the Art and Future Research Questions in Sentiment Analysis". Acta Polytechnica Hungarica 12 (3), pp. 87-108, 2015

[27] H. P. F. Peters, A. F. J. van Raan. “Structuring Scientific Activities by Coauthor Analysis an Exercise on a University Faculty Level”. *Scientometrics*, 20, pp. 235-255, 1991

[28] Popescu-Bodorin, N. and Balas, V. E. "Fuzzy Membership, Possibility, Probability and Negation in Biometrics". Acta Polytechnica Hungarica 11 (4), pp. 79-100, 2014

[29] P. Shannon, A. Markiel, O. Ozier, N. Baliga, J. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks”. *Genome Research*, 13, 2498-2504, 2003

[30] H. Small. “Co-Citation in the Scientific Literature: a New Measure of the Relationship between Two Documents”. *Journal of the American Society for Information Science*, 24, pp. 265-269, 1973

[31] H. Small. “Visualizing Science by Citation Mapping”. *Journal of the American Society for Information Science*, 50, pp. 799-813, 1999

[32] H. Yu, M. Davis, C. Wilson, F. Cole. “Object-Relational Data Modelling for Informetric Databases”. *Journal of Informetrics*, 2, pp. 240-251, 2008

[33] L. Zadeh. “Fuzzy Sets”. *Information and Control*, 8, pp. 338-353, 1965

[34] L. Zadeh. “Is There a Need for Fuzzy Logic*?*” *Information Sciences*, 178, pp. 2751-2779, 2008

# Improved Accuracy Evaluation of Schema Matching Algorithms

**Balázs Villányi, Péter Martinek**

Department of Electronics Technology, Budapest University of Technology and Economics, Egry József utca 18, H-1111 Budapest, Hungary
e-mail: villanyi@ett.bme.hu, martinek@ett.bme.hu

*Abstract: Application integration is one of the most relevant topics in enterprise computing today. To support Enterprise Application Integration, various algorithms, methods and complex systems are applied. Automated schema matchers support application integration by identifying the semantically related entities of the input schemas. In this article, we present the notion of the cutting threshold problem in schema matching and propose a solution. Our approach incorporates the definition of a threshold function, which is conceptually similar to a fuzzy membership function. We have also redefined the accuracy measures most commonly used for accuracy evaluation of schema matchers. Employing the threshold function, we managed to obtain a 9.85% average accuracy improvement. The introduction of the threshold function also enables the redefinition of the schema matching accuracy maximization problem.*

*Keywords: Enterprise Application Integration; Schema Matching; Algorithm Optimization; Fuzzy Logic; Threshold Function; Accuracy Measure Improvement*

## 1 Introduction

Currently, organizations have heterogeneous software components in their system landscapes covering functions like logistics, financial, production, etc. Enterprises aim at having an integrated system landscape, in order to reduce process times and improve information availability, while reducing redundancy. On the other hand, heterogeneous systems tend to publish different interfaces. These diverse interfaces have to be matched before the communication can be launched. The matching of interfaces involves the identification of semantically related schema entities of interfaces. This related entity identification in heterogeneous system schemas is referred to as schema matching [1] and is used to be carried out manually involving a handful of schema matching experts. Needless to say, that this type of schema matching execution is cumbersome, time-consuming and is error-prone. Therefore, efficient automated schema matching solutions are required.

There are many current, published automated schema matching approaches. These approaches can be categorized as follows. The linguistic matcher evaluates the semantic relatedness of textual elements found in schemas. We treat two types of linguistic matchers, based on whether they use external dictionaries, thesauri or ontologies: First, syntactic matchers do not use any kind of external taxonomy, but assess only string relatedness (e.g. common prefix/suffix-based, substring based etc.). Second, vocabulary approaches are linked with various kinds of external ontology, dictionary or thesaurus, like the WordNet [2]. Beyond linguistic matching, another main type of schema matchers is the structural matcher. Structural matchers detect the possible resemblance in the structure of the schema. Most typically, they analyze the ascendant-descendant relations of semantically related nodes in the schemas being compared.

These types of schema matchers are not used on their own and are combined using weights. Practically, every schema matcher provide a similarity value (also called semantic distance), which expresses the relatedness of the entities being compared and subsequently these values are summed using predefined weights. A schema matcher incorporating other matchers is called hybrid schema matcher and is the prevailing type among viable schema matcher proposals.

In order to match schemas, the semantic distance serves only as a basis for classification. Another predefined value, called threshold, is required to decide whether the entities being compared will constitute a match or no (i.e. whether they represent the same real world entity/business object or not): entity pairs obtaining semantic values higher than a specific threshold are treated as "*match*" while the rest is classified as "*non-match*". Consequently, the optimal choice of the threshold is crucial for the schema matching, but it differs from the weights used to combine individual semantic distances.

The performance of schema matching solutions is expressed through accuracy. It is measured using accuracy measures such as precision, recall and F-measure [1]. Precision expresses the rate of real matches among all entity pairs classified as match, recall is the rate of retrieved true matches among all reference matches, while f-measure is the harmonic mean of the former two.

Our first problem (*Prob. I*) formulation is as follows. The schema matcher weights and threshold should be defined so that all entity pairs be classified correctly. This task is particularly difficult if the two result set (labeled "*match*" and "*non-match*") are near or overlapping. In case of the overlapping the threshold value application may cause accuracy value degradation as the classic accuracy measures (precision, recall and F-measure) assess accuracy based on crisp set membership. Entity pairs near to the threshold are excessively punished/rewarded for being on the improper/proper side of the threshold, however, their classification is not 100% certain according to the relatedness evaluation of the schema matcher. This uncertainty then may lead to a false accuracy feedback of the real schema matcher performance. See the problem in details in Section 3. We refer to this problem as the cutting threshold problem.

There is also another problem (*Prob. II*), which has encouraged us to conduct threshold related experiments: the non-continuous objective function of the accuracy measure maximization problem. It is treated in full length in [3], where we propose a methodology to optimize the parameters of schema matchers. A subgroup of our parameter optimization techniques comprises the accuracy measure maximization, whereby the objective function is a formula derived from a given accuracy measure so that it contains exclusively the schema matcher parameters (the weights and the threshold). The main drawback of the presence of a constant threshold value is that it makes the objective function non-continuous.

We provide a solution for both of these problems in Chapter 5, by proposing the introduction of schema matching threshold function. In order to be able to validate our idea of how the introduction of a threshold function resolves our accuracy measure maximization related problem (*Prob. II*), the accuracy measures are also described in Section 4.

This article is divided into six sections as follows. This first chapter provides introductory insight into the problem formulation. The second chapter presents some of the related works. The third chapter proposes a solution for the cutting threshold problem. The forth chapter introduces the reader to our revised accuracy measure maximization technique. The fifth chapter is about the results obtained with the proposed accuracy measures. Lastly, we provide some closing thoughts on the subject.

# 2 Related Works

In this chapter, we present some of our related work which have inspired us. Many schema matching has been published in the last decades like [4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. The matcher in [6] is a complex schema matcher consisting of three components which are the name similarity, the related term similarity and the attribute similarity. The name similarity is a rough, scoring based similarity assessor. The solution [4] provides context based matching, which is a very efficient technique backed by the WordNet lexical database [2]. Nevertheless, the processes of context based matching increase complexity and there are several parameters to be set before the matcher could perform optimally. The similarity flooding [10] is a key technique which involves the redistribution of concept similarities in the similarity propagation net.

Matching is also involved in [14] where authors provide a solution to reconcile the semantics of structured and semi-structured data. This solution is also geared towards flexibility and knowledge accumulation. Another solution for the matching of ontologies can be found in [15] which has linguistic and structural matching components like many other schema matchers. This solution also makes

use of the above listed lexical database [2]. Another ontology related application can be found in [16] where a special ontology system is proposed, which helps storing fuzzy information. The management of heterogeneous information is carried out using this ontology. A special ontology is introduced in [17] which is closely related to semantic networks and UML diagrams. Schema and ontology matching are closely related [18], consequently our observations and results of this current paper may apply to ontology mapping as well.

An excellent semantic integration technique can be found in [19]. Schema matching can be applied in the same context. In [19] authors propose a solution which is able to automatically translate XML schema representations of the business objects into OWL based ontology. Schema matching can be used to identify semantically related entities in the XML schemas, thus it can be applied even in those situations where highly diverse, heterogeneous schemas are to be integrated.

The solution in [20], is a generic schema matching tool, called COMA+. It is not a schema matcher by itself, but is a complex platform designed for integrating schema matching components. The individual matchers are arranged into a library. The solution provides scalability by fragmenting schemas into subsets, which is an application of the "divide and conquer" principle to the field of schema matching. This allows for the platform to be flexible. On the other hand, it does not include parameter optimization which would further improve its efficiency.

The above solutions bear in common that they have fixed threshold values. Based on our observation, this attribute may be limiting in several cases. It is also true, that the employed fixed threshold values do not take into account the cutting threshold problem introduced in the next section. Consequently, our method presented in this paper, can be used to enhance the performance evaluation of all of the above listed schema matching solution.

# 3 The Cutting Threshold Problem

As introduced in the first section (*Prob. I*), the cutting threshold problem lies in the presence of a mixed labeled - i.e. "*match*" or "*non-match*" - entity pair cluster around the threshold value. The employment of the classic sets in the schema matcher accuracy evaluation may lead to the serious distortion of the schema matching accuracy because of the followings. In certain schema matching scenarios the result sets will be overlapping even when using optimal schema matcher parameters. In other words there will always be entity pairs on the wrong side of the threshold based on the semantic value given by the schema matcher — i.e. "*match*" labeled entity pairs having semantic values lower than the threshold or "*non-match*" labeled entity pairs having semantic values higher semantic values than the threshold. See Fig. 1.

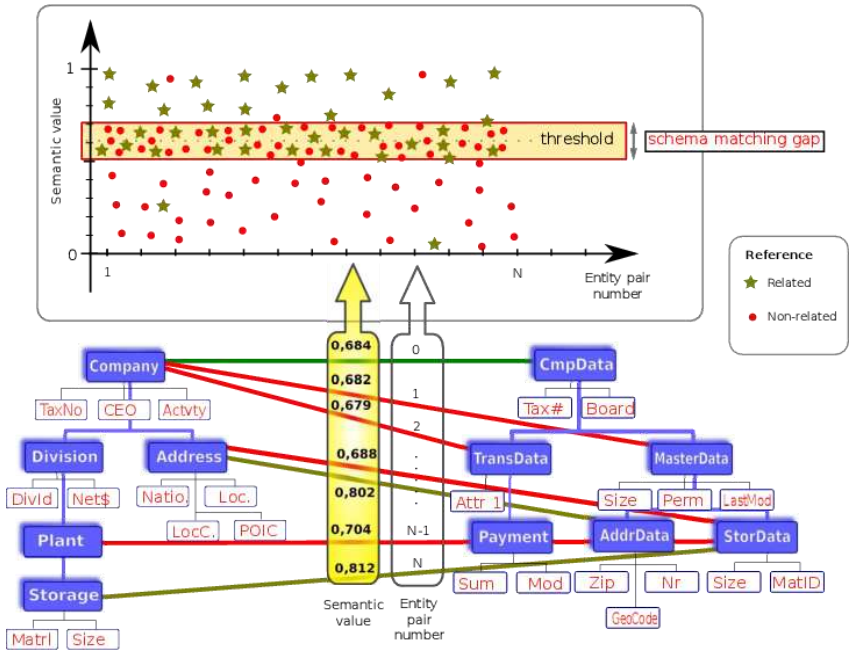



Figure 1
Depiction of the cutting threshold problem

Fig. 1 depicts the overlapping result sets by optimal schema matcher weights. On the graph, the x-axis represents the entity pairs being matched (by simply ordering them in a chosen sequence), while y-axis is the semantic value (given by the schema matcher for a given entity pair). There is also another information on this graph: the ground truth or reference match. The green star represents those entity pairs, which should be labeled "*match*" by reference, i.e. they represent the same real world entity and are thus considered to be related. Conversely, the red dot represent those entity pairs, which should be labeled "*non-match*" by reference, i.e. the red dots are those entity pairs which are considered non-related by reference.

The problem is also seen in Fig. 1. Even if the schema matcher are optimal, we cannot specify an optimal threshold because the (reference) result sets overlap. Part of this problem is the distortion effect of the threshold: should we specify a suboptimal threshold in the intersection of the result sets – we call it schema matching gap, see Fig. 1 –, the accuracy measure evaluation will not be precise since at least some of the entity pairs in the schema matching gap will be wrongly classified. This is especially true for the F-measure evaluation. As already stated, we refer to this problem as cutting threshold problem.

Stepping one step further, the classification of the entity pairs in the schema matching gap is uncertain, which is not reflected in the classic accuracy measures leading to a distortion in the accuracy assessment. Conversely, the concept of a threshold value is still useful and thus, applicable, without any modifications required, for the cases of entity pairs falling outside of the schema matching gap. This is because their schema matcher classification is firm, no matter how we specify the threshold in the schema matching gap.

There is agreement among the above mentioned observations and aspects: the concept of the threshold value can be retained if the classification is clear for semantic values outside the schema matching gap, while the cutting threshold problem can be resolved if we introduce a continuous result set transition for entity pairs in the schema matching gap.

These considerations have led us to the definition of threshold function. At the time of the threshold function formula the recurrent problem was that the result curve did not fit the needs stated above or the curve formula became so complex that it seriously hindered the practical applicability of the approach. In the end, we have found that the most appropriate threshold function is the sigmoid function as it is continuous, has a simple formula and there is a continuous transition between semantic values of matching {1} and non-matching {0} entity pairs in the schema matching gap. It is also important that this function can be easily parameterized so that it fit a large scale of schema matching scenarios. Its formula has been reformulated in order that it fit our needs, resulting in Eq. 1. We have also involved the concept of schema matching gap lsg. For further references, see Fig. 2. ($\tau = 64$ and $l_{sg} = 0.2$).In the following Eq. 1, $\tau$ denotes the threshold and $g$ denotes the granularity:

$$f(x) = \frac{1}{1 + e^{g(x-\tau)}} \tag{1}$$

where $g$ can be defined as a function of the schema matching gap $l_{sg}$ , see Eq. 2:

$$g = \frac{128}{l_{sg}} \tag{2}$$

For the practical application, the curve should be parameterized considering the followings. The transition interval from match certainty zero to one should include all semantic value of the overlapping result set intersection (by definition of the schema matching gap) and the transition interval should be centered on the original threshold value if any — see Fig. 1. Thus, entity pairs having semantic distance equal to the threshold will have uncertainty value 0.5, fairly reflecting the fact that they may as well be "*match*" or "*non-match*" according to the semantic value given by the schema matcher. Furthermore, entity pairs falling outside the schema matching gap will have semantic values {0;1} indicating that their classification is sure.

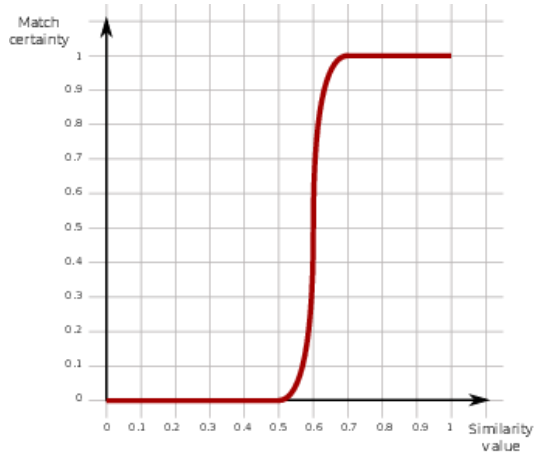



Figure 2
The sigmoid function substituting the threshold

Also the threshold function may serve for the accuracy evaluation for related entities (the "*match*" set), since on a [0, 1] scale it gives null accuracy for those related entity pairs which have semantic values lower than the lower bound of the schema matching gap, and one accuracy for those related entity pairs which have semantic values higher than the higher bound of the schema matching gap. The rest is given accuracy in-between. By the same token, the f( − x) can be used as the accuracy evaluation functions for the set "non-match". These membership functions (*f(x)* and *f( − x)*) have the common inflexion point located at the threshold value, resulting in a characteristic "X" shape.

Later we have also realized [21] that our concept of threshold function is very similar to that of the membership functions in fuzzy logic, thus we may as well use the aforementioned accuracy evaluation functions as membership functions for the result sets.

# 4 Redefinition of the Accuracy Measures

In this section, we will present our ideas on how to use the schema matching threshold function to adjust the formula of existing accuracy measures [1] to solve the accuracy measure maximization problem.

We will use the following denotations in Eq. 3 and Eq. 4. Let $M$ denote the set of matches. $M_p$ (proposed) will denote the match set returned by the schema matcher, while $M_r$ (reference) is the reference match set (ground truth). In concordance with the earlier, $g$ and $\tau$ denote the granularity and the threshold respectively. $c_w$ denotes the number of weights used by the algorithm, which is the same number

as the number of schema matcher components [3]. $w_{iu}$ shall denote $u^{th}$ weight in case of the $i^{th}$ matching pair, $c_{iu}$ is the value returned by the corresponding component.

$$\text{Precision} = \frac{\sum_{i \in M_p \cap M_r} \frac{1}{1+e^{g(\sum_{u=0}^{c_w} w_{iu} c_{iu} - \tau)}}}{\sum_{j \in M_p} \frac{1}{1+e^{g(\sum_{v=0}^{c_w} w_{iv} c_{iv} - \tau)}}} \tag{3}$$

$$\text{Recall} = \frac{\sum_{i \in M_p \cap M_r} \frac{1}{1+e^{g(\sum_{u=0}^{c_w} w_{iu} c_{iu} - \tau)}}}{\sum_{j \in M_r} \frac{1}{1+e^{g(\sum_{v=0}^{c_w} w_{jv} c_{jv} - \tau)}}} \tag{4}$$

The F-measure is simply the harmonic mean of the precision and the recall. The trade-off of the threshold function is that the correctly classified entity pairs residing in the vicinity of threshold will not be scored as full match, either. Since we have employed the threshold function also in the denominator of the accuracy measures, the maximal accuracy value can be attained in case of perfect matching.

# 5 Accuracy Measure Maximization Using the Threshold Function

As already mentioned in the introduction (*Prob. II*), the accuracy measure maximization is our schema matcher optimization technique to define the optimal parameter set of schema matcher by means of the derivation of accuracy measure formulas [22]. Unfortunately, the objective function of the accuracy measure maximization problem is non-continuous. We will now present the revised accuracy measure maximization technique using the F-measure as accuracy measure.

First of all, we introduced weights to the F-measure. The reason is that we found that the exact harmonic mean of the precision and recall may not be appropriate for all schema matching scenarios. Consequently, we have developed the generalization of the F-measure formula based on the concept of weighted harmonic mean: the weighted F-measure – see Eq. 5. Let $\pi$ denote the weight of the precision, while $\rho$ should represent the weight of the recall. $P$ is precision value, while $R$ is recall value and $F$ is the F-measure. The modified formula of the F-measure is the following:

$$F_{\pi,\rho} = (\pi + \rho)\frac{P \cdot R}{\rho \cdot P + \pi \cdot R} \tag{5}$$

Using the threshold function and the proposed weighted F-measure, the reformulation of our original accuracy measure maximization problem setting [22] became possible. Eq. 6 describes the objective function for the weighted f-measure using the already presented notation.

$$\max_{0<w_{ij}<1} \frac{\sum_{i \in M_p \cup M_r} \frac{\pi+\rho}{1+e^{g\left(\sum_{u=0}^{c_w} w_{iu} c_{iu} - \tau\right)}}}{\sum_{j \in M_p} \frac{\rho}{1+e^{g\left(\sum_{v=0}^{c_w} w_{iv} c_{iv} - \tau\right)}} + \sum_{k \in M_r} \frac{\pi}{1+e^{g\left(\sum_{z=0}^{c_w} w_{kz} c_{kz} - \tau\right)}}} \tag{6}$$

If we use the classic f-measure, the formula will be less complex (see Eq. 7) as the denominator will not contain weights:

$$\max_{0<w_{ij}<1} \frac{\sum_{i \in M_p \cap M_r} \frac{1}{1+e^{g\left(\sum_{u=0}^{c_w} w_{iu} c_{iu} - \tau\right)}}}{\sum_{j \in M_p \cup M_r} \frac{1}{1+e^{g\left(\sum_{v=0}^{c_w} w_{jv} c_{jv} - \tau\right)}}} \tag{7}$$

Furthermore, the schema matching reference values should also be specified, with the help of Eq. 8:

$$R_j = \frac{1}{1+e^{g\left(\sum_{v=0}^{c_w} w_{jv} c_{jv} - \tau\right)}} \tag{8}$$

The expression above declares that the weighted sum of the component relatedness values for a given entity pair ($j^{th}$ entity pair in the list) should be equal to the reference.

The benefit of this accuracy measure maximization problem formulation is that all contained functions are continuous, thus it can be easily solved using optimization tools. We shall note that this reformulation of the original accuracy measure maximization problem has retained its original characteristics as the sigmoid threshold function returns closely the same values outside of the transition as the original threshold value based accuracy evaluation.

# 6 Experimental Results

The goal of our experiments was to validate the benefits of the threshold function compared to the classic threshold value based approach. We have also performed several experiments to reveal the possible favorable impact of the employment of the revised accuracy measure maximization.

Our experiments had the objective of directly comparing the results of the classic accuracy measure approach with the revised method. We have tested our approach using three schema matchers: NTA [6], Similarity Flooding [10] and a WordNet based complex matcher [4]. We found these three approaches adequate to assess the possible benefits of using our proposed threshold function and refined accuracy measures. We have executed these techniques on three different test scenarios. These test scenarios are defined to contain the key challenges found in schema matching: different naming conventions, structures, etc.

Our comparison approach is as follows. We have calibrated these three techniques to scenarios, then evaluated their accuracy using the classic accuracy measures. As next, we have defined the ideal threshold function for each and every schema matching scenario and schema matcher. Lastly, we have evaluated the accuracy of every schema matcher in every scenario by means of the revised accuracy measures employing the threshold function. We have calculated the average precision, recall and F-measure in the test scenarios. The results can be seen on Fig. 3 (precision), Fig. 4 (recall) and Fig. 5 (F-measure).

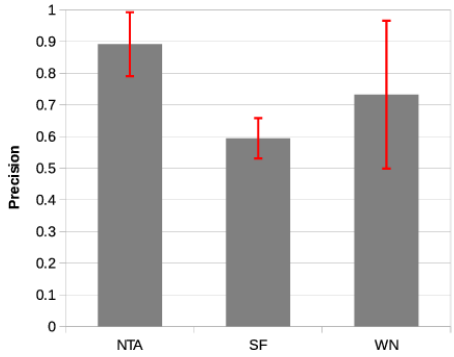



Figure 3
Average precision values with standard deviation of the analyzed schema matchers in the test scenarios

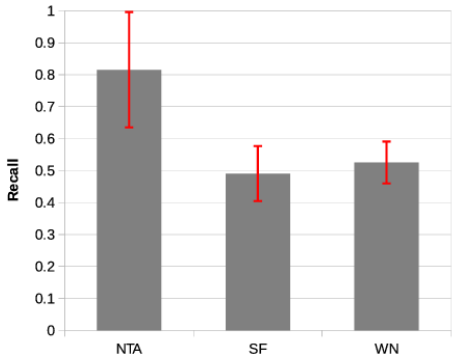



Figure 4
Average recall values with standard deviation of the analyzed schema matchers in the test scenarios

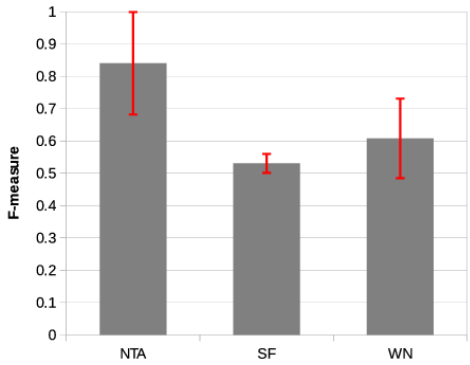



Figure 5
Average F-measure values with standard deviation of the analyzed schema matchers in the test scenarios

Fig. 3 (precision), Fig. 4 (recall) and Fig. 5 (F-measure) shows us that the standard deviation for the accuracy values are limited. We could observe less deviation in case of the recall, but precision error was also in the expected zone. The aggregated accuracy measure – the f-measure – has a less than expected deviation. In fact, the standard deviation became less after the application of the threshold function as it can be seen on Fig. 6.

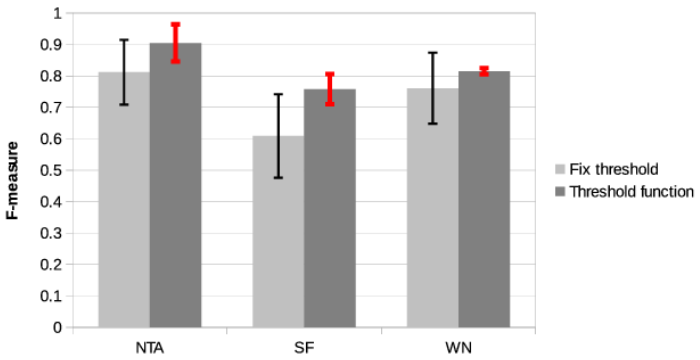



Figure 6
Average F-measure value comparison of the threshold value and the threshold function application

As it can be seen on Fig. 6, we saw an increase in the f-measure compared to the classic approach in almost every test scenario after applying adequate threshold function. We argued that this increase occurred due to the more justified evaluation of the schema matcher performance. This outcome has further strengthened our conviction that the cutting threshold problem does not just occur under very specific circumstances, but it should be treated as a ubiquitous phenomenon.

The deviation of the obtained f-measure values from the means can also be seen on Fig. 6. As we can observe it, the deviation is not substantial even in the case of the fix threshold value application, but we managed to further reduce it by means of threshold function application. In our interpretation the deviation from the average F-measure shows how much the schema matcher performance varies in different scenarios. The lesser this values is the more balanced the schema matcher performance is. This characteristic is an especially important notion in a highly variable application context, i.e. when the schema matchers are applied to significantly different test scenarios. For further details, consult Table 1 and Table 2. All in all, the application of threshold function entailed a 9.85% accuracy improvement.

Table 1

Average accuracy using fix threshold value and the threshold function (expressed with f-measure)

| | NTA | SF | WN |
|---|---|---|---|
| Threshold function | 0.905 | 0.758 | 0.815 |
| Threshold value | 0.812 | 0.609 | 0.761 |

Table 2

Standard deviation of the accuracy using fix threshold value and the threshold function (expressed with f-measure)

| | NTA | SF | WN |
|---|---|---|---|
| Threshold function | 0.059 | 0.048 | 0.08 |
| Threshold value | 0.103 | 0.133 | 0.113 |

## Conclusion

In some schema matching scenarios the result sets overlap. This phenomenon hinders the specification of an optimal threshold and also significantly distorts the accuracy evaluation of schema matchers leading to false conclusions on the real capabilities of a tested schema matcher. This problem — the cutting threshold problem — was presented in detail herein.

For the solution of cutting threshold problems, we propose the application of a threshold function and we also propose a formula which we found to be optimal. Furthermore, the revised schema matching accuracy evaluation approach also reflects the classification uncertainty — as opposed to the classic schema matching accuracy evaluation approach. We also presented revised formulas of the classic accuracy measures (precision, recall, F-measure) incorporating the proposed threshold function.

A related problem is the accuracy measure maximization, which has a non-continuous objective function. The introduction of a threshold function also resolves this problem, by presenting an approximate objective function for the original problem. Thus a revised accuracy measure maximization technique was also presented here.

Using the threshold function, we managed to attain an average accuracy improvement of 9.85%.

**References**

[1] Hong-Hai D, Rahm E. Matching Large Schemas: Approaches and Evaluation. Information Systems 2007; 32(6): pp. 857-885

[2] WordNet, http://wordnet.princeton.edu/, 2006

[3] Villanyi B, Martinek P, Szikora B. A Framework for Schema Matcher Composition. WSEAS Transactions on Computers 2010; 9(10): pp. 1235-1244

[4] Boukottaya A, Vanoirbeek C. Schema Matching for Transforming Structured Documents. In: Proceedings of the 2005 ACM symposium on Document engineering, Bristol, UK; 2005; pp. 101-110

[5] Hung NQV, Tam NT, Miklós Z, Aberer K. On Leveraging Crowdsourcing Techniques for Schema Matching Networks. In: Database Systems for Advanced Applications, Heidelberg, DE; 2012; pp. 139-154

[6] Martinek P, Szikora B. Detecting Semantically-related Concepts in a SOA Integration Scenario. Periodica Polytechnica 2008; 52(1-2): pp. 117-125

[7] Wan J, Xiaokun D. A Genetic Schema Matching Algorithm based on Partial Functional Dependencies. Journal of Computational Information Systems 2013; 9(12): pp. 4803-4811

[8] Vejber VV, Kudinov, AV, Markov NG. Algoritm sopostavlenija shem dannyh informacionnyh sistem neftegazodobyvajushhego predprijatija. Izvestija TPU 2011; (5): pp. 75-79

[9] Zhao C, Shen D, Kou Y, Nie T, Yu G. A Multilayer Method of Schema Matching based on Semantic and Functional Dependencies. In: Proceedings of Web Information Systems and Applications Conference (WISA), Haikou, China; 2012; pp. 223-228

[10] Melnik S, Garcia-Molina H, Rahm E. Similarity Flooding: a Versatile Graph Matching Algorithm and its Application to Schema Matching. In:Proceedings of the 18th International Conference on Data Engineering, San Jose, CA, USA; 2002; pp. 117-128

[11] Puttonen J, Lobov A, Lastra JLM. Semantics-based Composition of Factory Automation Processes Encapsulated by Web Services, IEEE Transactions on Industrial Informatics 2013; 9(4): pp. 2349-2359

[12] Furdík K, Tomášek M, Hreňo J. A WSMO-based Framework Enabling Semantic Interoperability in e-Government Solutions, Acta Polytechnica Hungarica 2011; 8(2): pp. 61-79

[13] Iordan V, Cicortas A. Ontologies used for Competence Management, Acta Polytechnica Hungarica 2008, 5(2): pp. 133-144

[14] Jameson M, Fei XX, Chun DS. Intelligence Benevolent Tools: A Global System Automating Integration of Structured and Semistructured Sources in One Process. International journal of intelligent systems 2004; 19(6): pp. 543-563

[15] Acampora G, Loia V, Salerno S, Vitiello A. A Hybrid Evolutionary Approach for Solving the Ontology Alignment Problem. International Journal of Intelligent Systems 2012; 27(3): pp. 189-216

[16] Blanco IJ, Vila MA, Martinez-Cruz C. The Use of Ontologies for Representing Database Schemas of Fuzzy Information. International Journal of Intelligent Systems 2008; 23(4): pp. 419-445

[17] Bíla J, Tlapák M. Ontologies and Formation Spaces for Conceptual ReDesign of Systems. Acta Polytechnica 2005; 45(4): pp. 33-38

[18] Zohra B, Bonifati A, Rahm E. Schema Matching and Mapping. Springer 2011

[19] Anicic N, Ivezic N. Semantic Web Technologies for Enterprise Application Integration. Computer Science and Information Systems 2005; 2(1): pp. 119-144

[20] Hong-Hai D, Rahm E. COMA - A System for Flexible Combination of Schema Matching Approaches. In: Proceedings of the 28th International Conference on Very Large Databases, Hong Kong, China; 2002; pp. 610-621

[21] Villanyi B, Martinek P. A Comparison of Schema Matching Threshold Function and ANFIS-generated Membership Function. In: Proceedings of IEEE 14th International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, Hungary; 2013; pp. 195-200

[22] Martinek P, Villanyi B, Szikora B. Calibration and Comparison of Schema Matchers. WSEAS Transactions on Mathematics 2009; 8(9): pp. 489-499

# Balance Assessment during the Landing Phase of Jump-Down in Healthy Men and Male Patients after Anterior Cruciate Ligament Reconstruction

**Aleksandra Melińska[1], Andrzej Czamara[2], Lukasz Szuba[2], Romuald Będziński[3], Ryszard Klempous[4]**

[1] Wrocław University of Technology, Faculty of Fundamental Problems of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland aleksandra.melinska@pwr.edu.pl

[2] College of Physiotherapy in Wroclaw, T. Kościuszki 4, 50-038 Wrocław, Poland a.czamara@wsf.wroc.pl

[2] College of Physiotherapy in Wroclaw, T. Kościuszki 4, 50-038 Wrocław, Poland lukasz.szuba@wsf.wroc.pl

[3] University of Zielona Góra, Faculty of Mechanical Engineering, Biomedical Engineering Group, ul. Licealna 9, 65-417, Zielona Góra, Poland romuald.bedzinski@pwr.edu.pl

[4] Wrocław University of Technology, Faculty of Electronics, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland, ryszard.klempous@pwr.edu.pl

*Abstract: The aim of the study was to evaluate balance during the landing phase in a control group and patients after anterior cruciate ligament reconstruction (ACLR). ACL tear is a frequent injury, particularly in athletes. It often requires proper rehabilitation, during which time the patients should be re-evaluated. We performed a jump-down assessment using a motion capture and force platform system. The tests comprised a series of time-space and dynamic parameters measurements using a motion capture system and a force plate. We tested 28 men (22 controls and 6 patients who had undergone ACLR). The tests are conducted under jump-down conditions (0.1-, 0.2-, and 0.3-m step heights). We compared horizontal components of force during the landing phase. The division of registered three-dimensional (3D) motion was demonstrated, and the fluctuations of the estimated center of gravity (eCOG) during this motion were analyzed. Compared with the controls, patients showed statistically significant differences in fluctuations in the eCOG repositioning and horizontal components of force ratios (t-test, $p<0.05$). We propose a comparative assessment of balance attainment during the landing phase of a jump. We indicate the possibility of efficient detection of ACL injuries. In summary, horizontal ground reaction forces and eCOG positioning can be used to evaluate the performance of the human biomechanical system during the jump-down phase.*

*Keywords: Ground reaction forces; Body mass center repositioning; ACLR; Motion analysis*

# 1 Introduction

Injuries of the anterior cruciate ligament (ACL) are common in athletes, but they also occur in other individuals, where permanent uneven loading of lower limbs is observed [1-4]. Mechanical failure of the ACL necessitates ligament reconstruction surgery (ACLR), often leading to lengthy rehabilitation that ensures stabilization of the knee joint and proper force distribution. Monitoring rehabilitation progress is obligatory for assessing its effectiveness. Motion analysis measurements and techniques are helpful in this regard.

There have been numerous motion analysis studies, but they are mainly concerned with gait [5-13]. Some studies considered motion analysis after ACLR [14-16]. In cases of lower limb dysfunction, such as those requiring ACLR, motion assessment is possible not only for gait conditions but also during running or jump-down [17-31]. Jump-down is a natural activity for humans (e.g., getting down a staircase), and the amount of loading differs between walking and running. Jump-down has another interesting aspect as well: the landing phase. Among the studies that considered jump-down, some of the important ones addressed only the landing strategy. The landing phase is where the subject attains balance, so the landing mechanisms should be carefully observed [32-36]. The motion capture system, force plates, and electromyography (EMG) are useful for assessing balance. It is possible to assess balance by measuring the ground reaction force, segments of the body's repositioning, the angles between segments, and muscle activity [37-40].

We suggest that there is a need to broaden studies concerning the identification of load that occurs during the landing phase of jump-down. The research scope includes analysis of 3D data derived from dynamometric platforms and the motion capture system. Unlike the existing studies, we focused here on the horizontal components of the ground reaction force and the estimated body's mass center of gravity (eCOG) repositioning during the landing phase. The aim of the study was to evaluate balance during the landing phase in a control group and patients after anterior cruciate ligament reconstruction (ACLR). Our hypothesis was: *Is the balance characteristics during the landing phase different in the normal controls and patients who underwent ACLR?*

# 2 Methods

A total of 28 men aged 20–26 years were studied. In all, 22 of them did not show any abnormality and were not diagnosed with any knee joint dysfunction. They constituted the control group. The other six men, who had undergone ACLR of the left limb, comprised the patient group. These six patients had gone through a 3-month supervised rehabilitation course and then carried out a rehabilitation program on their own. The ACLR rehabilitation program was performed according to the protocol developed by Czamara et al. [41]. The patients agreed to participate in the study during the 8th month after their surgery. The control group was matched for age, body height, and body mass. Data characterizing the groups (mean ± SD) were as follows: controls: age 25.1 ±4.3 years, height 1.78 ± 0.19 m, weight 78.5 ± 9.7 kg; patients: age 26.2 ± 2.3 years, height 1.72 ± 0.91 m, weight 74.3 ± 5.6 kg. All subjects were treated according to the tenets of the Declaration of Helsinki.

Before the examination, each person was informed about the course of the study and was requested to consent to participation. The tests comprise a series of time–space and dynamic parameters measurements. The individual tests consisted of a jump-down from three step heights (0.1, 0.2, and 0.3 m). After jump-down from 0.1 m, another 0.1 m was added to the platform until the height of 0.3 m was obtained. There was a 30-second break between the jumps. The baseline position for each participant was free standing on both legs on the step. On the examiner's command, the participant jumped down. Each subject was instructed that he should land on his toes and metatarsus on both limbs (one limb per platform). The tests were preceded by a short warm-up involving trotting (3-4 minutes), five to seven jumps, and (after a 20-s interval) four to five squats. After a short break, each subject performed trial jumps from different heights. The jumps were followed by a 2-min rest break.

A dynamometric platform and a motion capture system were used for the tests. Body center repositioning was registered using a module of the BTS Smart (BTS Bioengineering, Milan, Italy). BTS Smart-D3 is a module that performs the functions of multi-channel 3D navigation. It operates on the basis of an image composition from six infrared cameras that capture movement of markers placed on the body. For this study, we used one marker placed on the skin near the sacral bone to estimate the location of the COG. The ground reaction forces during a jump were measured by two dynamometric platforms (one for each leg) onto which the subject jumped. The output signal from each platform included the resultant force vector, component vectors, magnitude, and force direction. The measuring system is shown in Figure 1. The test also involved measuring ground reaction forces and the eCOG repositioning while free standing before each jump.

Figure 1
Data acquisition setup

The force during jump-down was measured from the first contact of the feet with the platform. Ground reaction force was measured in three directions. The motion was divided into two phases. Examples of force outputs divided into phases are shown in Figure 2 for one control group and one patient.

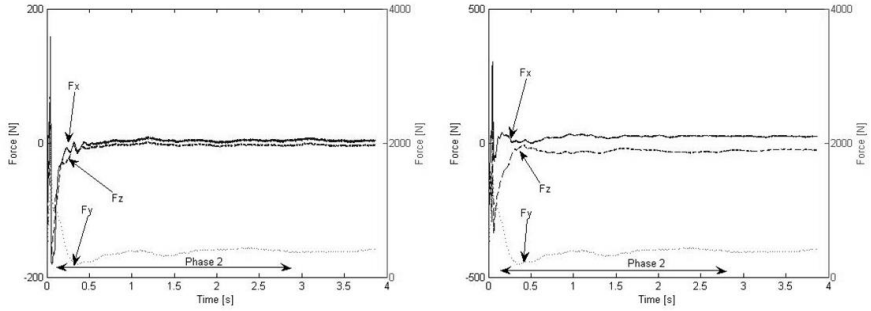



Figure 2
Force signals for one healthy control (left) and one patient (right)

The first phase covered the period from the initial contact of the foot with the surface (the time at which the vertical value of the ground reaction force, Fy, is minimum) to the time when Fy is maximum. This phase was not considered in our analysis because the meaningful phase of the jump-down is the landing phase (phase 2). The landing phase begins at the time the vertical components of force reach their maximum. It covers the period when the values of force do not vary by more than 20% of the maximum value. This occurs approximately 3 s after the jump on the platform. Note that the vertical force is not a part of this study. (We earlier reported a separate study on vertical ground reaction forces [42]). Here, the horizontal vectors were examined because they are responsible for balance. They are calculated as follows:

$$\pi_F = \frac{\overline{F_x}}{\overline{F_z}}$$

where $\bar{F} = \frac{1}{n}\sum_{p}^{n-p} F;$ $F$ is the average force component, subscripts $x$ and $z$ correspond to the horizontal components of force, $p$ is the number of observations, $F_y$ is the maximum force, and $n$ is the number of samples.

Force is not the only value that is important during the landing phase. Body repositioning is crucial as well. The eCOG position can be used to describe balance during landing. The traction is registered by cameras in the motion capture system. The eCOG position is registered in 3D space in the time domain. The 3D motion is divided into two orthogonal planes: horizontal and vertical. This concept is shown in Figure 3.

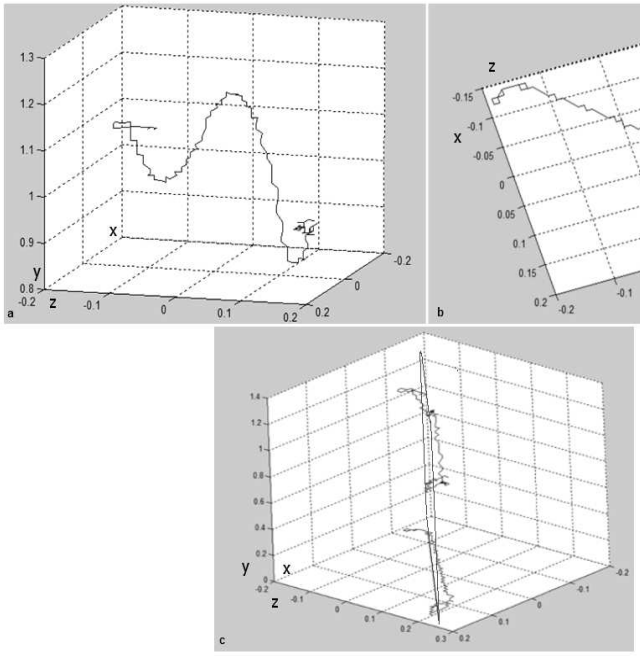



Figure 3
Concept of motion: a) 3D motion traction (XYZ), b) Projection for plane XZ c Projection for the XZ plane with a regressive trajectory

We divided the trajectory function into $f(x, y, z, t)$ for (1) projection horizontal components $p(x, z)(t)$ and (2) representation of the vertical component in the common time domain $y(t)$.

The registered trajectory in 3D space is a function of time $t$ and the marker position $x,y,z$ in axes:

$$f(x,y,z,t) = \begin{cases} p(x,z) \\ r(p,x) \end{cases} \quad (1)$$

where $p$ finds coefficients of a first-degree polynomial that best fits the input *data* $(p(x(t),z(t))$ in an orthogonal *least-squares sense*. The projected trend is given by $p = mx + n$ (2), which returns the predicted value of the polynomial, given its coefficients, to values in *x*.

Next, we found the distance $d$ from a point $(x_0, y_0,)$ to a line $(ax + by + c = p = mx + n)$ for each value in the time domain.

The parametric equation of a straight line is given as:

$$d = \frac{|mx+ny+c|}{\sqrt{m^2+n^2}} \quad (3)$$

where *m, n, c* are real constants satisfying m, n ≠ 0

Results are presented on the two-axes chart $f = (time, d;\ time, y)$.

The eCOG fluctuations in the controls and patients were analyzed. Two approaches for measuring those fluctuations are proposed:

a) Determining the average eCOG fluctuation,

$$\overline{d}_t = \frac{\sum_{t=0}^{n} d_t}{n} \quad (4)$$

where $d_t$ is the distance between the eCOG to the regressed projected trajectory registered in time *t*, where *n* is the number of samples.

b) Determining the sum of the eCOG fluctuation from the regressive trajectory.

$$d_t = \sum_{t=0}^{n} d_t \quad (5)$$

where $d_t$ is the distance between the eCOG to the regressed projected trajectory registered in time *t*, where *n* is the number of samples.

Statistical analysis was conducted using SPSS (IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp, Released 2013). We conducted t-test for assessing if there are significant differences for patients and control group in horizontal components force ratio in landing phase and also in free standing (Table 1). We used two-way ANOVA with Bonferroni correction to test if there are differences in fluctuation of eCOG in patients and control group as well as if there is interaction between step high and measured parameters (Table 2).

# 3 Results

Table 1 shows the horizontal components force ratios during the landing phase of the jump and free-standing values (mean ± SD). In case of the control group, the model horizontal components force ratio during jump-down should be similar to those observed in the free-standing condition [37]. The average force ratio increases with the step height increase. The force ratios in free-standing phase for the control group and patients are not significantly different, as well as there is no statistically significant difference between right and left leg for the control group, and there are for patients (t-test, $p > 0.05$). There are statistically significant differences between patients and the control group for right leg and for the left leg in case of 0.2 and 0.3 m step height (t-test, $p < 0.05$).

Table 1
Horizontal components ratio of ground reaction forces

| | Landing phase | | | | Free-standing phase | | |
|---|---|---|---|---|---|---|---|
| Group | Step height (m) | RL | LL | *p-value* | RL | LL | *p-value* |
| Controls | 0.1 | 1.10±0.69 | 1.17±0.87 | *,056* | 1.0±0.10 | 1.0±0.10 | *,000* |
| | 0.2 | 1.07±0.50 | 1.05±0.49 | *,058* | | | |
| | 0.3 | 1.14±0.49 | 1.05±0.45 | *,067* | | | |
| Patients | 0.1 | 0.39±2.23 | 0.023±2.43 | *,000* | 0.99±0.20 | 0.99±0.20 | *,000* |
| | 0.2 | 0.74±1.50 | 1.05±1.49 | *,000* | | | |
| | 0.3 | 1.70±1.61 | 1.30±1.59 | *,000* | | | |
| p – value for comparison of the control group and patients for both legs regard to step high | | | | | | | |
| | | RR | LL | | | | |
| 0.1 | | *,001* | *,0551* | | | | |
| 0.2 | | *,001* | *,076* | | | | |
| 0.3 | | *,003* | *,012* | | | | |

LL = left limb, RL = right limb

Fluctuations from eCOG trend motion were analyzed in the same way for both the control group and those who had undergone ACLR. An example of the results of differences between the control group and patients is shown in Figure 4, represented by the measurements for one healthy control and one ACLR patient for the 0.1-m step.

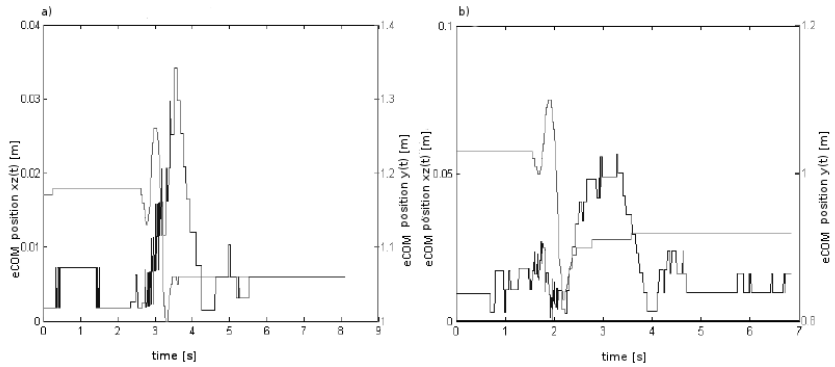



Figure 4

Example of eCOG fluctuation for a healthy control and an ACLR patient. The jump was from 0.1 m.

We propose two measures for evaluating balance: the sum and an average of the fluctuation of the eCOG from the regressed motion trajectory during the patients' jumps (equations 1, 2, 3). The results are shown in Figures 5 and 6.

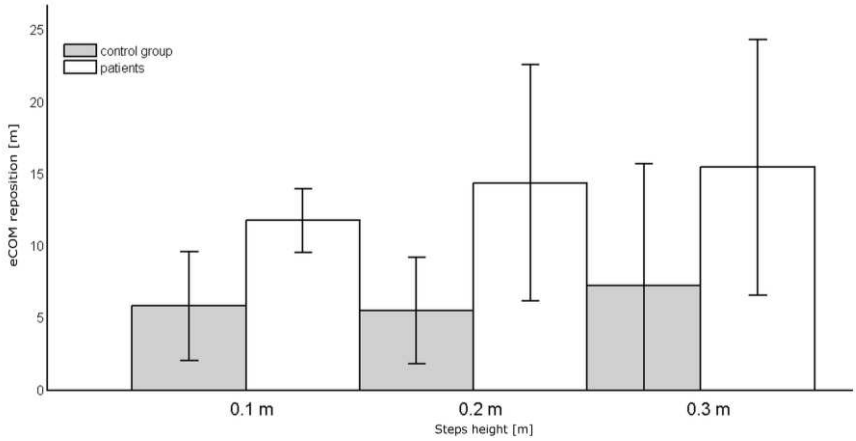



Figure 5

Sum of fluctuation of distance of the eCOG from a regressed motion trajectory in the horizontal plane

Both results show that the measured fluctuations are substantially greater for the patients than for the controls. The step height had an effect on the patients' results. The sum increased with greater step height for patients although the arithmetic mean decreases. The step height had no effect on the healthy controls. Healthy people land more stably than patients, which we proved in the final statistics. The sum of the fluctuation increased with step height for the ACLR patients. Using ANOVA test we checked if there is interaction between step height and fluctuations of eCOG (average and sum of fluctuation). There are statistical significant differences for patients for the step height 0.2 and 0.3 m.

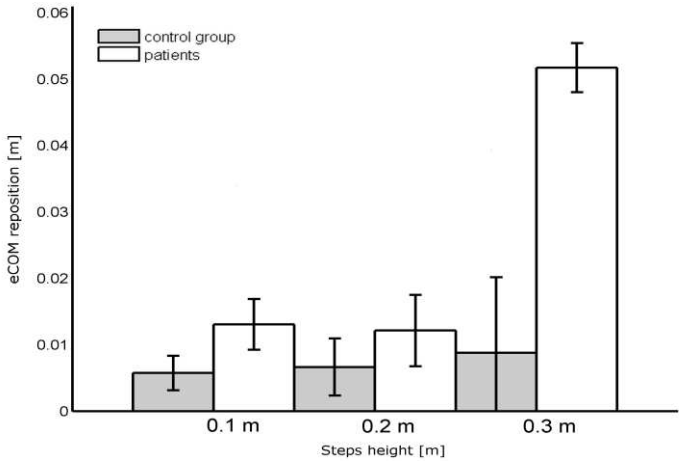



Figure 6
Average sum of the distance of the eCOG from a regressed motion trajectory in the horizontal plane

Table 2
Statistics for ANOVA test for eCOG fluctuation measures

| | | | |
|---|---|---|---|
| Sum of fluctuations of distance of the eCOG | (I) step high | (J) step high | p- value |
| | 0.1 | 0.2 | ,054 |
| | | 0.3 | |
| | 0.2 | 0.1 | ,017 |
| | | 0.3 | |
| | 0.3 | 0.1 | ,017 |
| | | 0.2 | |
| | | | |
| Average of the fluctuation of the eCOG | (I) step high | (J) step high | p- value |
| | 0.1 | 0.2 | ,066 |
| | | 0.3 | |
| | 0.2 | 0.1 | ,012 |
| | | 0.3 | |
| | 0.3 | 0.1 | ,012 |
| | | 0.2 | |

# 4 Discussion

This study evaluated the jump-down phase by analyzing horizontal components of the ground reaction force and eCOG repositioning during the landing phase. Several other studies have considered jump-down [23, 24, 38, 42-47]. Among them, some are important from the point of view of balance.

Paterno et al. [43] presented a case–control study of female athletes at a mean of 27 months following their ACLR versus healthy female subjects. All of the participants executed a task that comprised a drop vertical jump onto two force plates (one per leg). The vertical ground reaction force, which was measured during landing and takeoff, was used to calculate the landing phase loading rates. The authors concluded that female athletes who have undergone ACLR and return to their sport may continue to demonstrate biomechanical limb asymmetry 2 years or more after reconstruction. It can be identified during landing.

Similarly, using force plates, Ortiz et al. [44] demonstrated the landing mechanics in noninjured women and women with ACLR. They performed five trials of a single-leg 40-cm drop jump and two trials of a 20-cm up–down hop task. Multivariate analyses of variance were used to compare hip and knee joint kinematics, knee joint moments, ground-reaction forces, and EMG findings between the dominant leg in noninjured women and the reconstructed leg in women who had undergone ACLR. The results of their study showed that women with ACLR have neuromuscular strategies that allow them to land from a jump similar to healthy women, but they exhibit joint moments that could predispose them to future injury if they participate in sports that require jumping and landing.

Decker et al. [45] purposed their study to determine whether fully rehabilitated ACLR recreational athletes utilize adopted lower-extremity joint kinematics and kinetics during high-demand functional tasks. Subjects were compared during a 60-cm vertical drop landing. The hamstring ACLR recreational athletes used an adapted landing strategy that employed less of the hip extensor muscles and more of the ankle plantar flexor muscles. Harvesting the medial hamstring muscles for ACL reconstruction may contribute to the utilization of this protective landing strategy.

Gribble and Hertel et al. [46] designed their study to examine the role of foot type, height, leg length, and range of motion (ROM) on excursion distances while performing the Star Excursion Balance Test (SEBT). The SEBT measures dynamic postural control. Participants performed three trials of the SEBT in each of eight directions while balancing on the right and left legs, respectively. The SEBT found no statistically significant relations between foot type or ROM and the excursion distances. Significant correlations were revealed between height and excursion distance and leg length and excursion distance, with leg length having the stronger correlation. Using raw excursion measures, men were found to have substantially greater excursion distances than females. However, after normalizing

excursion distances to leg length, there were no substantial differences related to sex. Thus, when using the SEBT for experimental or clinical purposes, participants' excursion distances should be normalized to leg length to allow a more accurate comparison of performance among participants.

Use of force plates is popular for addressing biomechanics, and a number of studies have used this measurement method for balance analysis [47–50]. Torry et al. [51] described an interesting approach to drop landing analysis. They used biplane fluoroscopy to measure 3D rotations and translations of healthy knees during stiff drop landings to determine the relations between 3D rotations and anterior and lateral tibial translations. Attainment of balance can be analyzed not only via ground reaction force or EMG measurements, but by observing eCOG repositioning, as shown by Colby et al. [52]. In addition to ground reaction force, they focused on developing a functional test to measure dynamic stability that could differentiate between the injured and uninjured lower limbs, They tested two populations—individuals with ACL deficiency and those who had undergone ACLR—to establish the reliability of the test.

With regard to the current state of the art, our study provides new information on horizontal components of ground reaction force and eCOG relocation. Balance during the landing phase is different for healthy people and patients after ACLR. Our patients had greater eCOG fluctuations. The ground reaction horizontal components of the force ratio were greater for the patients than for the controls.

## Conclusions

The horizontal ground reaction forces and eCOG positioning are appropriate parameters for evaluating the performance of the human biomechanical system during jump-down. They can be used to recognize an ACL injury as well as to evaluate progress during rehabilitation after ACLR.

## References

[1] Bjordal, J. M., et al.: Epidemiology of Anterior Cruciate Ligament Injuries in Soccer, The American Journal of Sports Medicine 25.3 (1997) pp. 341-345

[2] Ireland, M. L.: Anterior Cruciate Ligament Injury in Female Athletes: Epidemiology, Journal of Athletic Training 34.2 (1999) p. 150

[3] Hootman, J. M., Dick, R., & Agel, J.: Epidemiology of Collegiate Injuries for 15 Sports: Summary and Recommendations for Injury Prevention Initiatives, Journal of Athletic Training 42.2 (2007) p. 311

[4] Frigo, C., Shiavi, R.: Applications in Movement and Gait Analysis, In: Electromyography Physiology, Engineering, and Noninvasive Applications, (2004) pp. 381-397

[5] Beard, D. J., Soundarapandian, R. S., O'Connor, J. J., & Dodd, C. A. F.: Gait and Electromyographic Analysis of Anterior Cruciate Ligament Deficient Subjects. Gait & Posture, 4.2 (1996), pp. 83-88

[6] Boulgouris, N. V., Chi, Z. X.: Gait Recognition Using Radon Transform and Linear Discriminant Analysis, IEEE Trans on Image Proc, 16.3 (2007) pp. 731-740

[7] Kale, A., Sundaresan, A., Rajagopalan, A. N., Cuntoor, N. P., Roy-Chowdhury, A. K., Krüger, V., & Chellappa, R.: Identification of Humans using Gait, IEEE Trans on Image Proc, 13.9 (2004) pp. 1163-1173

[8] Begg, R. K., Palaniswami, M., & Owen B.: Support Vector Machines for Automated Gait Classification, IEEE Trans on Biomedical Eng, 52.5 (2005) pp. 828-838

[9] Kamruzzaman, J., & Begg, R. K.: Support Vector Machines and Other Pattern Recognition Approaches to the Diagnosis of Cerebral Palsy Gait, IEEE Trans on Biomedical Eng, 53.12 (2006) pp. 2479-2490

[10] Yam, C. Y., Nixon, M. S., Carter, J. N.: Extended Model-based Automatic Gait Recognition of Walking and Running, Proc. of 3rd Int. Conf. on Audio- and Video-based Biometric Person Authentication (2001) pp. 278-283

[11] Hansen, M., Haugland, M. K., & Sinkjær, T.: Evaluating Robustness of Gait Event Detection Based on Machine Learning and Natural Sensors, IEEE Trans on Neural Systems and Rehabilitation Eng, 12.1 (2004) pp. 81-88

[12] Corazza, S., Mündermann, L., Chaudhari, A. M., Demattio, T., Cobelli, C., & Andriacchi, T. P.: A Markerless Motion Capture System to Study Musculoskeletal Biomechanics: Visual Hull and Simulated Annealing Approach. Annals of Biomedical Engineering, 34.6 (2006) pp. 1019-1029

[13] Drosou, A., Ioannidis, D., Moustakas, K., & Tzovaras, D.: Spatiotemporal Analysis of Human Activities for Biometric Authentication, Computer Vision and Image Understanding, 116.3 (2012) pp. 411-421

[14] Knoll, Z., Kiss, R. M., & Kocsis, L.: Gait Adaptation in ACL Deficient Patients before and after Anterior Cruciate Ligament Reconstruction Surgery. Journal of Electromyography and Kinesiology, 14.3 (2004) pp. 287

[15] Tashman, S., Anderst, W., Kolowich, P., Havstad, S., & Arnoczky, S. Kinematics of the ACL-Deficient Canine Knee during Gait: Serial Changes over Two Years. Journal of Orthopaedic Research, 22.5 (2004) pp. 931-941

[16] Devita, P., et al.: Gait Adaptations before and after Anterior Cruciate Ligament Reconstruction Surgery, Medicine and Science in Sports and Exercise, 29.7 (1997) p. 853

[17] Tashman, S., Kolowich, P., Collon, D., Anderson, K., & Anderst, W.: Dynamic Function of the ACL-reconstructed Knee during Running. Clinical Orthopaedics and Related Research, 454 (2007) pp. 66-73

[18] Tashman, S., Collon, D., Anderson, K., Kolowich, P., & Anderst, W.: Abnormal Rotational Knee Motion during Running after Anterior Cruciate Ligament Reconstruction, The American Journal of Sports Medicine, 32.4 (2004) pp. 975-983

[19] Waite, J. C., Beard, D. J., Dodd, C. A. F., Murray, D. W., & Gill, H. S.: In Vivo Kinematics of the ACL-Deficient Limb during Running and Cutting, Knee Surgery, Sports Traumatology, Arthroscopy, 13.5 (2005) pp. 377-384

[20] Decker, M. J., Torry, M. R., Wyland, D. J., Sterett, W. I., & Richard Steadman, J.: Gender Differences in Lower Extremity Kinematics, Kinetics and Energy Absorption during Landing. Clinical Biomechanics, 18.7 (2003) pp. 662-669

[21] Savalli, L., Hernandez-Sendin, M. I., Soille, J., & Laboute, E: Interest of the Monopodal Jump as an Indirect Means of Assessing Muscle Recovery Distance of an ACL Reconstruction, Annals of Physical and Rehabilitation Medicine, 55.1 2012, p.e76

[22] Graham-Smith, P., & Lees, A.: A Three-Dimensional Kinematic Analysis of the Long Jump Take-Off. Journal of Sports Sciences, 23.9 (2005) pp. 891-903

[23] Pflum, M. A., Shelburne, K. B., Torry, M. R., Decker, M. J., & Pandy, M. G.: Model Prediction of Anterior Cruciate Ligament Force during Drop-Landings. Medicine and Science in Sports and Exercise, 36.11 (2004) p. 1949

[24] Pain, M. T., & Challis, J. H.: The Influence of Soft Tissue Movement on Ground Reaction Forces, Joint Torques and Joint Reaction Forces in Drop Landings. Journal of Biomechanics, 39.1 (2006) pp. 119-124

[25] Kram, R., Griffin, T. M., Donelan, J. M., & Chang, Y. H.: Force Treadmill for Measuring Vertical and Horizontal Ground Reaction Forces. Journal of Applied Physiology, 85.2 (1998) pp. 764-769

[26] Kutilek, Patrik, et al.: Kinematic Quantification of Gait Asymmetry Based on Characteristics of Angle-Angle Diagrams, Acta Polytechnica Hungarica 11.5 (2014)

[27] Bíró, I., & Fekete, G.: Approximate Method for Determining the Axis of Finite Rotation of Human Knee Joint. Acta Polytechnica Hungarica, 11.9 (2014)

[28] Noyes, Frank R., et al.: The Drop-Jump Screening Test Difference in Lower Limb Control by Gender and Effect of Neuromuscular Training in

Female Athletes, The American Journal of Sports Medicine 33.2 (2005) pp. 197-207

[29] Torry, Michael R., et al.: Knee Kinematic Profiles during Drop Landings: a Biplane Fluoroscopy Study, Medicine and Science in Sports and Exercise, 43.3 (2011) pp. 533-541

[30] Horita, T., et al.: Stretch Shortening Cycle Fatigue: Interactions among Joint Stiness, Reflex, and Muscle Mechanical Performance in the Drop Jump, European Journal of Applied Physiology and Occupational Physiology 73.5 (1996) pp. 393-403

[31] Russell, Kyla A., et al.: Sex Differences in Valgus Knee Angle during a Single-Leg Drop Jump, Journal of Athletic Training 41.2 (2006) p. 166

[32] Graham-Smith, P., & Lees, A. A Three-Dimensional Kinematic Analysis of the Long Jump Take-Off. Journal of Sports Sciences, 23.9 (2005) pp. 891-903

[33] Fagenbaum, R., & Darling, W. G.: Jump Landing Strategies in Male and Female College Athletes and the Implications of Such Strategies for Anterior Cruciate Ligament Injury, The American Journal of Sports Medicine, 31.2 (2003) pp. 233-240

[34] Withrow, T. J., et al.: The Effect of an Impulsive Knee Valgus Moment on in Vitro Relative ACL Strain during a Simulated Jump Landing: Clinical Biomechanics 21.9 (2006) pp. 977-983

[35] Pappas, E., et al.: Biomechanical Differences between Unilateral and Bilateral Landings from a Jump: Gender Differences, Clinical Journal of Sport Medicine 17.4 (2007) pp. 263-268

[36] McKinley, P., & Pedotti, A.: Motor Strategies in Landing from a Jump: The Role of Skill in Task Execution. Experimental Brain Research, 90.2 (1992) pp. 427-440

[37] Horstmann, G. A., F. Huethe, and V. Dietz.: Special Treadmill for the Investigation of Standing and Walking in Research and in Clinical Medicine."Biomedizinische Technik. Biomedical Engineering 32.10 (1987) 250-254

[38] Normand, M. C., Normand, A. R., & Marchand, D.: Muscular Responses Comparison between Normal and Acl Reconstructed Knees during Drop Jump. Journal of Biomechanics, 25.7 (1992) p. 696

[39] Waite, J. C., et al.: In Vivo Kinematics of the ACL-Deficient Limb during Running and Cutting, Knee Surgery, Sports Traumatology, Arthroscopy 13.5 (2005) pp. 377-384

[40] Melińska, A. et al.: The Method of Evaluation of Biomechanical Parameters of Human Lower Limbs in the Jump from Steps of Different Heights, Fizjoterapia Polska, 11.4 (2011) pp. 327-340

[41] Czamara, A.: Evaluation of Physiotherapeutic after Endoscopic Reconstruction of the Anterior Cruciate Ligament of the Knee, The Doctoral Dissertation. University of Physical Education in Warsaw, Warsaw, Poland, 2006, pp. 39-82

[42] Melinska, A. et al.: Biomechanical Characteristics of the Jump Down of Healthy Subjects and Patients with Lower Limbs Dysfunctions, Acta of Bioengineering and Biomechanics, article in press: DOI: 10.5277/ABB-00208-2014-04

[43] Paterno, M. V., et al.: Limb Asymmetries in Landing and Jumping 2 Years Following Anterior Cruciate Ligament Reconstruction, Clinical Journal of Sport Medicine, 17.4 (2007) pp. 258-262

[44] Ortiz, A. et al.: Landing Mechanics between Noninjured Women and Women with Anterior Cruciate Ligament Reconstruction during 2 Jump Tasks, The American Journal of Sports Medicine, 36.1 (2008) pp. 149-157

[45] Decker, M. J. et al.: Landing Adaptations after ACL Reconstruction. Medicine and Science in Sports and Exercise, 34.9 (2002) p. 1408

[46] Gribble, P. A., & Hertel, J.: Considerations for Normalizing Measures of the Star Excursion Balance Test. Measurement in Physical Education and Exercise Science, 7.2 (2003) pp. 89-100

[47] Waite, J. C., et al.: In Vivo Kinematics of the ACL-Deficient Limb during Running and Cutting, Knee Surgery, Sports Traumatology, Arthroscopy 13.5 (2005) pp. 377-384

[48] Davis, Roy B., et al.: A Gait Analysis Data Collection and Reduction Technique." Human Movement Science 10.5 (1991) pp. 575-587

[49] Pain, M. T., & Challis, J. H.: The Influence of Soft Tissue Movement on Ground Reaction Forces, Joint Torques and Joint Reaction Forces in Drop Landings, Journal of Biomechanics, 39.1 (2006) pp. 119-124

[50] Kram, Rodger, et al.: Force Treadmill for Measuring Vertical and Horizontal Ground Reaction Forces, Journal of applied physiology 85.2 (1998) pp. 764-769

[51] Riley, P. O., et al.: A Kinematic and Kinetic Comparison of Overground and Treadmill Walking in Healthy Subjects, Gait & posture 26.1 (2007) pp. 17-24

[52] Colby, S. M., et al.: Lower Limb Stability with ACL Impairment, The Journal of Orthopaedic and Sports Physical Therapy 29.8 (1999) pp. 444-51

# Hot Rolling Mill Hydraulic Gap Control (HGC) thickness control improvement

**Peter Kucsera*, Zsolt Béres****

* Óbuda University, Tavaszmező u. 15-17, H-1084 Budapest, Hungary
kucsera.pete@kvk.uni-obuda.hu

**Alcoa-Köfém Kft., Verseci u. 1-15, H-8000 Székesfehérvár, Hungary,
zsolt.beres@alcoa.com

*Abstract: In this article the AGC (Automated Gauge Control) system, of a hot rolling mill is analyzed and different methods are described to improve its performance. The mill stretch compensation and other disturbances during the rolling process are compensated and a rolling speed dependent, adaptive PI controller is developed to accomplish fine thickness control. The improved system has been validated by measurements. Future improvement options are also analyzed, like eccentricity compensation, and predictive control of thickness.*

*Keywords: Hot rolling mill; AGC; Automatic Gauge Control; HGC; Hydraulic Gap Control; Mill stretch compensation; Roll eccentricity compensation; Thickness control; Adaptive; Predictive control*

# 1 Introduction

Nowadays, the competitive nature of different sectors (industry, agriculture, …) depends on technological improvements [1]. There is a growing demand for high accuracy in thickness control in metal rolling area. If the required accuracy of hot rolling mills can be achieved, then in some cases, cold rolling is not even required, which can result in a radical decrease of production costs.

The schematic structure of a typical aluminum hot reversing rolling mill can be seen in Figure 1. The aluminum slabs, which have to be rolled, are transported from a re-heat furnace via a roller table, then heated to around 420-450 $^0$C. The initial slab thickness is around 500 mm. The target thickness is typically 5 to 10 millimeters, and then the product is coiled. To reach this thickness, the strip has to be rolled several times on a reversing mill. Rolling is done by two work rolls, each supported by backup rolls. The maximum rolling force can reach 3-40000 kN. Typically, the roll gap is determined by set point of the main screws, located on both sides of the mill stand. The screws provide only a rough positioning without

load between passes, and for accurate positioning during rolling two hydraulic cylinders are installed. The position of the cylinders is controlled by servo valves and the corresponding position controllers. This hydraulic servo is often called HGC (Hydraulic Gap Control) cylinder. The AGC system provides the reference value for the HGC positioning system.

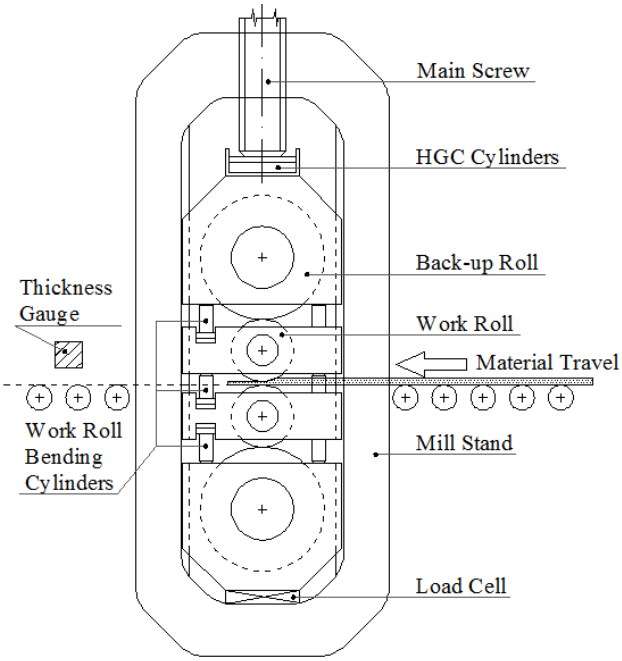



Figure 1
Schematic structure of a hot rolling mill

The thickness of the rolled material is measured by an X-Ray thickness gauge in the last pass. This measurement is possible within several meters distance from the roll gap, therefore considerable dead time has to be taken into account in thickness control.

The rolling force can be measured by a load cell or can be calculated from the pressure of the HGC cylinders. Due to several reasons (deflection of the rolls caused by the rolling force, the thermal crown and the ground crown), the roll gap is not uniform across the sheet, and therefore the cross profile usually has a parabolic shape, the center line being 0.2-1.2% thicker than the sides. Bending cylinders are used to apply a counteracting force on the work roll chocks to reduce the roll bending, thus controlling the cross profile partially.

In this article the improvement of the thickness control is discussed, focusing on the AGC system.

# 2 AGC Control Tasks

The rolling accuracy mainly depends on the automated gauge control (AGC) performance. The main tasks of an AGC system include:

- Mill stretch compensation
- Lock-on gauge control
- Work roll thermal expansion compensation
- Oil film compensation of back-up roll bearings
- Work roll bending
- Tapering
- Roll eccentricity compensation
- Thickness feedback control
- Predictive and adaptive control methods

Due to the given application and requirements, not all of the above mentioned features need to be applied to achieve acceptable thickness performance. The proper selection and combination of the various control techniques vary according to the given process and the available measurements.

## 1.1 Mill Stretch Compensation

As the strip passes between the work rolls, the thickness is reduced. The rolling force depends on the thickness reduction, on the characteristics of the rolled material, on the friction conditions, and on further process variables as well as unknown disturbances. Since the rolling force loads the mill stand, the stand suffers elastic deformation, called the mill stretch, and without compensation it would cause significant deviation in the thickness of the output product.

The mill stretch force can be measured by load cells or calculated from the measured HGC pressure and can be used to calculate the mill stretch. The mill modulus has to be measured during the commissioning of the system.

Within the operating range, the elastic deformation of a mill stand can reach a couple of millimeters. The target thickness tolerance ranges from ±0.1 to 0.3 mm, therefore the accurate measurement of the spring modulus is essential for the appropriate mill stretch compensation. The spring modulus can be considered constant, however, due to several mechanical components, there can be a slight nonlinearity in the low force range (Fig. 2).

On the unloaded mill there is no force before the slab or the strip arrives into the roll bite, therefore the gap has to be pre-set with an anticipated mill stretches. The anticipated mill stretch depends on the strip characteristics (size, material) and can be calculated using a high level system.

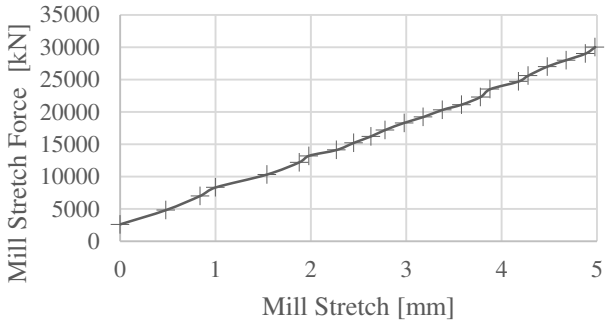



Figure 2
Measured mill stretch characteristic

When the mill is loaded, the mill stretch force is measured by load cells or calculated from the measured HGC pressure and can be used to calculate the actual mill stretch. A commonly used method is to take the elastic mill modulus into account and, using a feed forward error compensation, to compensate the mill stretch [3].

It is advised to filter the measurement errors or rapid force changes, because they cause sudden radical changes in the thickness of the output material.

Also, a proper safety protocol has to be added to the algorithm, since there is a positive feedback in the mill stand force control loop. If the force increases, the AGC closes the stand, which results in increasing force as well. If the stand works close to the maximum stand force, proper limiters have to be applied.

## 1.2 Lock-on Type Gauge Control

Under steady state rolling conditions, a linear relationship between the exit strip thickness, roll gap, and rolling force can be expressed by spring equation Eq. (1):

$$h_0 = S_0 + \frac{F_0}{M} \tag{1}$$

where $S_0$ is unloaded roll gap; $F_0$ is rolling force; M is the mill stiffness or mill modulus.

The strip thickness reference has to be created first during the lock-on process by calculating the average values of the actual roll gap and the actual rolling force. According to the spring equation, the strip thickness reference is shown in Eq. (2), where S is locked roll gap; F is locked rolling force.

$$h = S + \frac{F}{M} \tag{2}$$

The thickness deviation Δh can be calculated as follows:

$$\Delta h_0 = h_0 - h = S_0 - S + \frac{F_0 - F}{M} \tag{3}$$

The reference roll gap to eliminate thickness deviation (Δh = 0) is:

$$\Delta S_{GM} = -\left(1 + \frac{Q}{M}\right) \cdot \Delta h = -\left(1 + \frac{Q}{M}\right) \cdot \left(S_0 - S + \frac{F_0 - F}{M}\right) \tag{4}$$

where Q is the plastic modulus of the strip.

The structure of a commonly used AGC system, called the BISRA AGC, can be seen in Figure 3. The APC block represents the hydraulic servo, with the cylinders feedback and position controller.

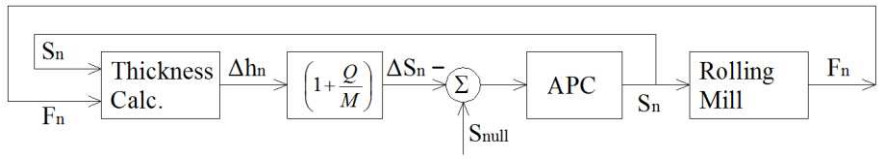



Figure 3
Schematic structure of the BISRA AGC

For accurate accurate calculations, it is first necessary to ensure that the spring curve is accurate. In most cases, linear approximation is sufficient. However, since the rolling mill has a complex mechanical structure, the ideal condition does not exist; the thickness deviation cannot be eliminated completely [2]. A good method to make the BISRA calculation more accurate is to measure the mill stretch under different pressures on an empty stand, pushing the work rolls together. This way the nonlinear mill modulus curve can be measured, and used for the calculation of mill stretch. Usually this measurement also requires some corrections, since the stand and the work rolls do not operate on working temperature in this case.

## 1.3 Work Roll Thermal Expansion Compensation

Work roll thermal expansion has a direct effect on strip thickness. Heat is generated in the strip due to deformation and conducted to the work roll when they are in contact. In addition, there is friction between the strip and the roll, which also generates heat. The work rolls are cooled by the emulsion, applied as lubrication between the rolls and the material.

Since direct measurement of the work roll's mean mass temperature is not possible, mathematical modelling can be used to estimate thermal expansion in working conditions. The measurement of the work roll surface temperature does not give the exact value of the mean-mass temperature, and this measurement is not practical, since emulsion is sprayed over the surface of the rolls [5].

Another way to determine the thermal expansion of the work roll is the modelling of the heat transfer between the work rolls and the strip. This depends on the roll speed, strip thickness, the properties of the strip, roll gap, contact time and cooling parameters. Most of researches focus on the thermal expansion contour of the work roll, and development of a proper coolant profile to compensate for the thermal expansion across the work roll [6].

From the aspect of the AGC, thermal expansion in the center line has to be calculated and compensated.

Without the above described mathematical modelling, operators manually adjust the gap to achieve the desired initial roll gap during threading.

## 1.4 Oil Film Compensation of Back-up Roll Bearings

In order to support the high rolling force, often friction bearings are applied as back-up roll bearings. In this type of bearing, an oil film separates the chock from the roll shaft. In a given bearing, the oil film thickness depends on the viscosity of the lubrication oil, the load and the rotational speed of the shaft (rolling speed), causing thickness variations. The most critical condition of the friction bearing, is the start-up under load, or operation at a very low speed.

To enhance the formation of the oil film at a low speed, hydrostatic lubrication systems can be applied. When hydrostatic lubrication is used, the oil film thickness $h_f$ can be calculated using Eq. 5 and can be adjusted during operation [7] (Fig. 4).

$$h_f = \frac{a(V/F)}{(V/F)+b} \tag{5}$$

Where, $V$ – roll speed, $F$ - roll force, $a$, $b$ – constants.

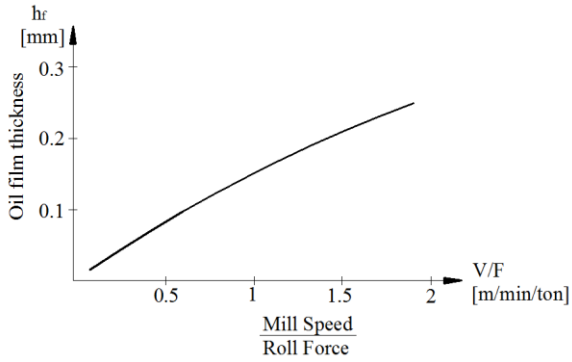



Figure 4

Relation between oil film thickness $h_f$ and ratio of roll speed $V$ to roll force $F$ *[7]*

## 1.5 Work Roll Bending

Roll bending is not directly linked to thickness control, but usually this control is incorporated in the same hardware, together with AGC.

Work roll bending forces are applied, using hydraulic cylinders, to the work roll shafts located at both roll ends. As a simplified explanation, positive or crown-in bending is applied, when the bending forces open the roll gap. Crown-in bending causes the strip profile to be concave. Negative or crown-out bending is when the bending force closes the roll gap, and results in a convex profile (Fig. 5). The positive roll bending force influences the mill stretch force, but does not have any effect on the rolling force.

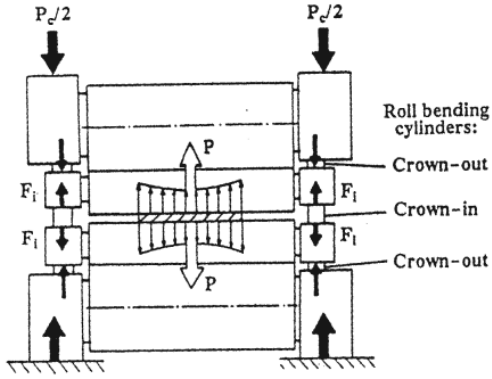



Figure 5
Work roll bending forces in 4-roll mills

## 1.6 Tapering

The hot rolled strip is later processed by a cold rolling mill, and to facilitate the threading operation, tapering is applied. Tapering is activated by a sensor which is located at a given distance from the mill stand on the rolling table at the entry side, and gives a signal when the strip end is in a given distance. This signal activates the tapering function, so the rolling thickness set-point is decreased following a given ramp.

## 1.7 Roll Eccentricity Compensation

Roll eccentricities are caused by axial deviation between the roll barrel and the roll shaft [4].

Roll eccentricity causes a periodic variation of the material thickness (Process Value PV), which can be measured with the thickness gauge (Fig. 6). This periodic disturbance is really hard to take out because of the dead time in thickness measurement.

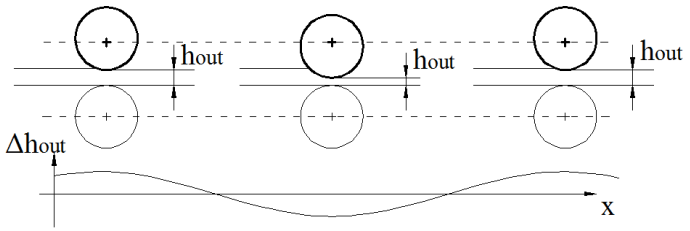



Figure 6
Effect of roll eccentricity on output strip thickness

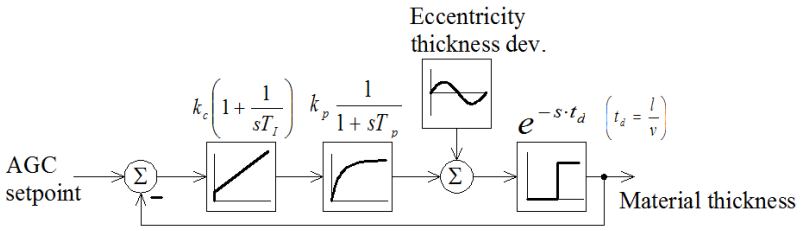



Figure 7
Thickness control loop with eccentricity thickness deviation as a disturbance

The hot rolling mill environment is quite harsh, so the X-ray thickness gauge is placed a few meters away from the roll gap. The dead time can be calculated from the distance between the work rolls and the thickness gauge (l), and the rolling speed (v) ($t_d = l / v$). Considering approximately 2 m/s rolling speed, the dead time can reach a few seconds. This amount of delay in a control loop makes accurate control difficult.

The aim of thickness control is to compensate for the slowly changing thickness deviations, using a PI controller. For this, the average of the measured thickness during one full rotation of the back-up roll is calculated and used as a process value for control. Because of the calculations and the long integration time, the control is slow, and the thickness deviation caused by eccentricity is not compensated.

The thickness deviation caused by the eccentricity has to be eliminated using other measurement and calculation methods than the X-ray thickness gauge. A possible solution is to make an eccentricity calibration measurement. After work roll change, eccentricity can be calculated on the unloaded mill, by rotating the touched work rolls (with a given force), and measuring the changing force that is loading the mill. The work roll shaft angle also has to be measured, and paired with the measured eccentricity.

Eccentricity can also be detected when the material rolls on the loaded stand, by detecting the periodic thickness deviation. The period time is known from the rotation speed (mill speed), measured by a shaft encoder or resolver.

# 2 Thickness Control

A possible HGC control scheme can be seen on Figure 8. The above described disturbances which influence the thickness of the strip are taken into account by using them for feed-forward error compensation. So, if, e.g. the mill stand force increases during rolling, the mill stretch compensation acts immediately, not waiting for the feedback thickness controller.

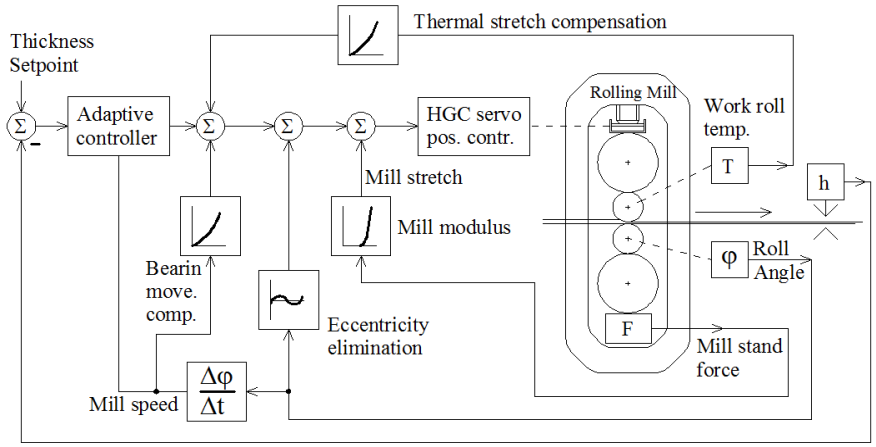



Figure 8
HGC control scheme

The simplest solution to achieve a thickness feedback control, is to choose a PI controller with dominating I effect, as it was mentioned before (Fig. 7). The rolling process can be modelled as a first order function with a delay (Eq. 6).

$$G_p(s) = k_p \frac{1}{1+sT_p} \cdot e^{-sT_d} \tag{6}$$

The controllability can be defined by the ratio of $T_p$ – process time constant to $T_d$ – dead-time. In case of a rolling mill $T_p$ is a few hundred milliseconds and $T_d$ is a few seconds; thus, the ratio shows a process which is really difficult to control. Popular recommendations are to use a pure I or PI controller (with dominating I) [8]. On Figure 9 a real measurement in a rolling mill is shown where during rolling the mill speed had to be decreased to about 30%. (It is not typical to change the mill speed during rolling, but it can happen if there is some technical problem with the tension reel or the side cutter.) The dead-time has linear connection to the mill speed (Eq. 7).

$$T_d = \frac{l}{v} \tag{7}$$

Where v is mill speed and l is the distance between the work rolls and the X-ray thickness gauge. As the mill speed changes (about 150 seconds after the rolling starts), the control becomes unstable, and causes a periodic oscillation in the output thickness.

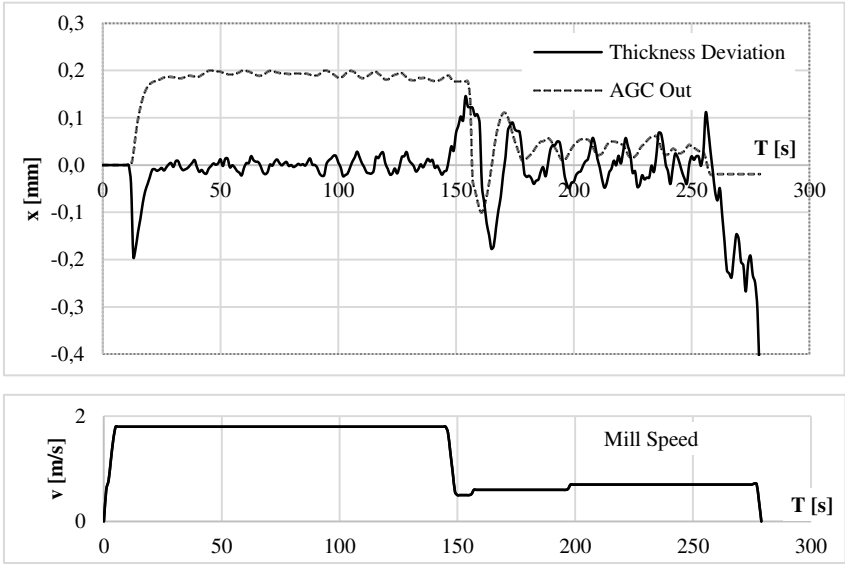



Figure 9
AGC control instability because of mill speed change

## 2.1 Adaptive PI Controller

A possible solution for the above mentioned problem is to use an adaptive PI controller, where the controller gain and the integration time are calculated by taking into consideration the mill speed.

A simple MATLAB simulation presents the structure of a possible adaptive PI controller (Fig. 10.). This simulation shows the advantages of using an adaptive controller instead of classical PI.

In the simulation the dead-time is set to 3 seconds, and after 30 seconds of rolling the speed is decreased to 1/3 of the nominal process speed, which causes 3 times more dead-time. As the mill speed is reduced, the classical PI control becomes unstable. A disturbance in thickness is modelled by a derivative, which gives a pulse signal when the speed changes, and adds it to the process output. A simple practical formula is used to set the controller parameters (Eqs. 8, 9) [10]:

$$k_c = \frac{0.3}{k_p} \tag{8}$$

$$T_i = 0.42 \cdot T_d \tag{9}$$

The process gain kp=1.5 (Tp=0.2 s) was used (estimated from Fig. 8). So the transfer function of the controller is (Eq. 10):

$$G_c(s) = 0.2 \cdot \left(1 + \frac{1}{0.42 \cdot T_d \cdot s}\right) \tag{10}$$

In classic PI, Td is constant 3 seconds, in the adaptive PI, the calculated dead time is used to set the integrator parameters.

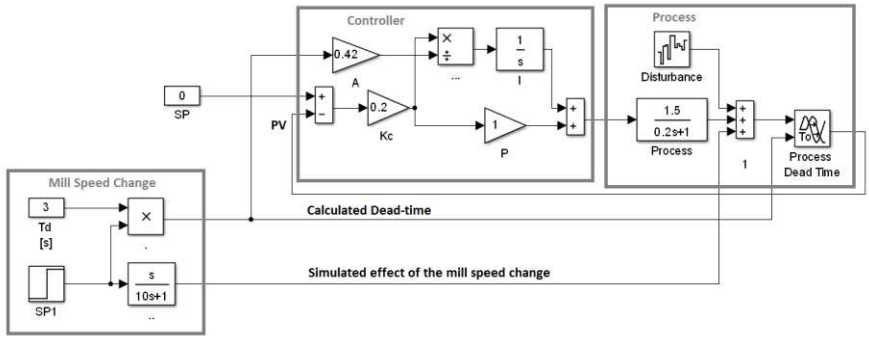



Figure 10
Structure of the adaptive PI thickness controller

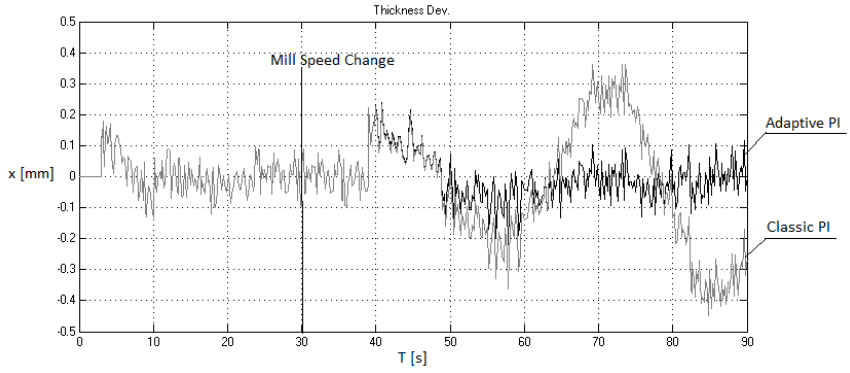



Figure 11
Comparison of the adaptive and the classic PI controller

As a result of the simulation the thickness deviation in time can be seen on Figure 11. As it could also be seen from the real measurements, the classic PI starts oscillating as the mill speed changes, but the adaptive one remains stable.

## 2.2 Predictive Controller

In the case of rolling mills, predictive control is an option, to achieve improved performance. This is because the dead-time can be calculated from the available speed measurements. However, the development of a suitable process model is a challenge.

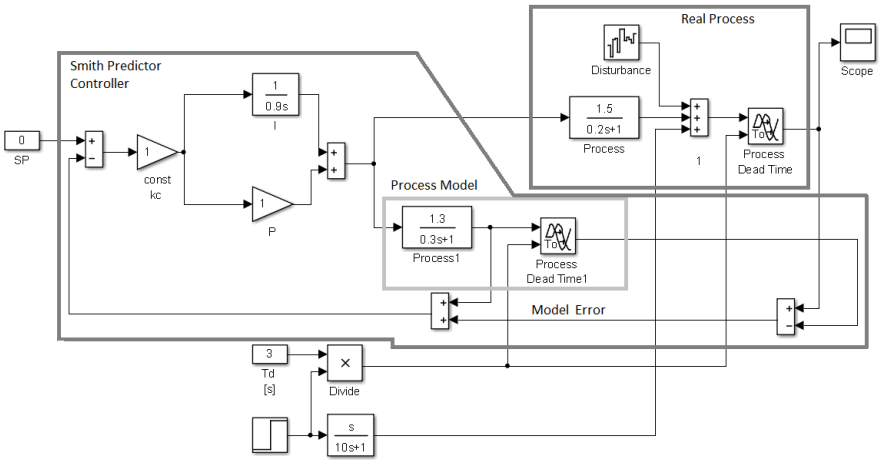



Figure 12
Structure of the Smith Predictor thickness controller

In Figure 12 a conventional Smith predictive controller MATLAB model can be seen. The PI controller output is connected to the process input, and also to the process model. The process model is divided into the first order and the dead-time parts, and the signal connecting these two blocks together is wired back to the PI controller as a simulated process value (without the dead-time). Also, the real process output and the model output are subtracted, giving the model error. If there is no model error, only the first order lag part of the model is "controlled" (without the dead-time), this way the controller can be tuned with high gains. The model error is added to the simulated process value and thus it is compensated [11].

The result of the simulation shows a better response, even if the model gain and the model first order lag time constant, were set intentionally to a different value from that of the real process to represent the effect of errors in the model. A step signal has been applied as a set point to compare the transient functions of the two controllers (Fig. 13).

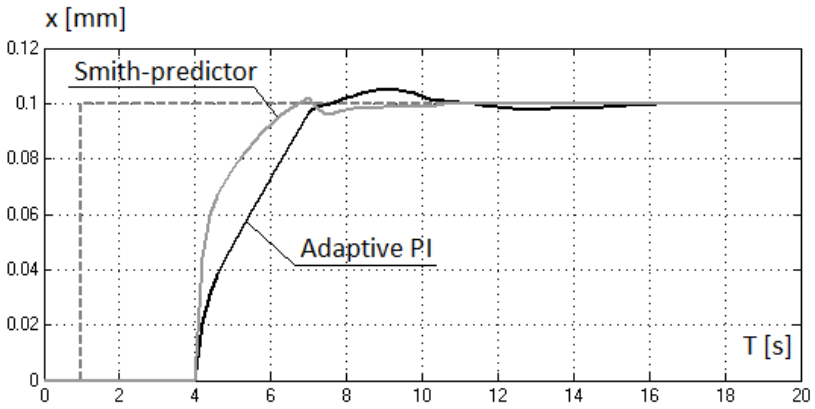



Figure 13
Comparison of adaptive PI and Smith-predictor controller transient function

# 3 Improvement

On Figure 14, the measured thickness deviation of a hot aluminum rolling mill can be seen, using a rolling speed-dependent adaptive PI thickness controller and mill stretch compensation.

The thickness deviation is within ±0.06 mm range during the rolling process. The end of the rolled strip is cooled significantly by the coiler shaft, which causes a large thickness deviation error for the last few meters of the strip. This thickness error can be corrected by the downstream cold rolling passes.

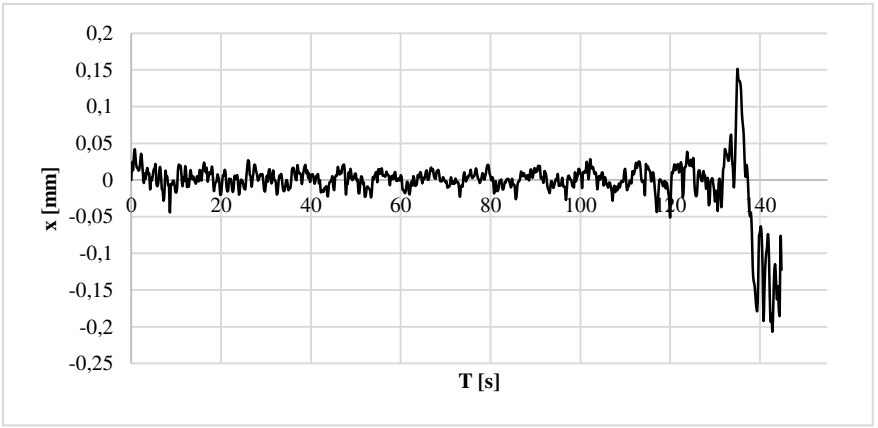



Figure 14
Measured thickness deviation using the adaptive PI controller and mill stretch compensation

## Conclusions

The structure of an existing AGC system has been analyzed and modified to introduce a thickness feedback control, within a hot rolling mill system. In order to avoid thickness oscillation at a reduced speed, a speed dependent adaptive PI regulator was applied. The improved system was validated by measurements.

In this paper, further improvement options are also analyzed, for example, eccentricity compensation and predictive control. These options promise further improvement possibilities.

## References

[1] A. Jambor – L. J. Hubbard, Changing Product Structure and Comparative Advantage: the Case of Hungarian Agri-Food Trade. (2013, Ekonomicky Casopis 61:(8) pp. 846-860

[2] Degarmo, E. Paul; Black, J T.; Kohser, Ronald A., *Materials and Processes in Manufacturing* (9th ed.) 2003 Wiley, ISBN 0-471-65653-4

[3] Ji Yafeng, Zhang Dianhua, Chen Shuzong, Sun Jie, LI Xu, Di Hongshuang, *Algorithm Design and Application of Novel GMAGC based on Mill Stretch Characteristic Curve*, Journal of Central South University March 2014, Volume 21, Issue 3, pp. 942-947

[4] Li Xua, Chen Shu-zonga, Du De-shunb, Chen Hua-xinc, Zhang Dian-huaa, *Simulation and Analyzing on Model Parameters Effect of BISRA-AGC,* 2011 International Conference on Physics Science and Technology (ICPST 2011), Volume 22, 2011, pp. 571-576

[5] Waleed I. Hameed, Khearia A. Mohamad, *Strip Thickness Control of Cold Rolling Mill with Roll Eccentricity Compensation by Using Fuzzy Neural* Network, Engineering, 2014, 6, pp. 27-33, ISSN Online 1947-394X

[6] I. Yu. Prikhod'ko, S. A. Vorobei, A. A. Sergeenko, V. V. Raznosilin, S. E. Shatokhin, *Temperature Regulation of Rollers in Broad-Strip Hot-Rolling Mills,* Steel in Translation, February 2011, Volume 40, Issue 11, pp. 985-989, ISSN 0967_0912

[7] M. Abbaspour, A. Saboonchi, *Work Roll Thermal Expansion Control in Hot Strip Mill, Applied Mathematical Modelling*, Volume 32, Issue 12, December 2008, pp. 2652-2669

[8] NIIR Board of Consultants & Engineers, *The Complete Technology Book on Hot Rolling of Steel,* National Institute of Industrial Research (2010) ISBN: 8190568582

[9] Ziegler, J. G.; Nichols, N. B. Optimum Settings for Automatic Controllers. Trans. ASME 1943, 65, 433

[10] J. Gerry, How to Control a Process with Long Dead Time, Control Engineering, 1998.03, ISSN 0010 8049, http://www.controleng.com/ 2015.03.25

[11] A. Vodencarevic, Design of PLC-based Smith Predictor for Controlling Processes with Long Dead Time, IMECS 2010, ISSN: 2078 0966

# Improvement of DC Motor Velocity Estimation Using a Feedforward Neural Network

**Miroslav Milovanović, Dragan Antić, Miodrag Spasić, Saša S. Nikolić, Staniša Perić, Marko Milojković**

Department of Control Systems, Faculty of Electronic Engineering
University of Niš
Aleksandra Medvedeva 14, 18000 Niš, Republic of Serbia
E-mail: miroslav.b.milovanovic@elfak.ni.ac.rs, dragan.antic@elfak.ni.ac.rs, miodrag.spasic@elfak.ni.ac.rs, sasa.s.nikolic@elfak.ni.ac.rs, stanisa.peric@elfak.ni.ac.rs, marko.milojkovic@elfak.ni.ac.rs

*Abstract: This paper proposes a method for improving the DC motor velocity estimations and the estimations obtained from the state observer, when the system operates with large moments of inertia. First, the state observer for estimating velocity and DC motor position, is designed. Then, the variable structure controller is formed using estimated position and velocity values. State observer and designed controller are implemented in default system control logic. Dependences between estimated velocities and moments of inertia are established and presented by experimental results. It is noted that velocity time responses of the designed controller are not as expected when the system operates with large moments of inertia on the motor shaft. The feedforward neural network is empirically designed and implemented in control logic with purpose to solve poor velocity estimations and to improve overall system performances. It is experimentally shown that an artificial network improves estimation quality of the observer and overall control of the system for different input signals.*

*Keywords: variable structure controller; neural network; state observer; servo system; DC motor; moment of inertia*

# 1 Introduction

Artificial neural networks are increasingly represented in the field of power systems control [1-4], because of their ability to operate with a large number of data. A quality training procedure is a precondition for successful neural network usage. Conventional controllers in the presence of disturbances eventually do not provide a system robust enough. Both stability and robustness can be increased by introducing neural network into the default control logic of a system. An artificial neural network could be used as a compensator, whose assignment is to bring

system dynamics to desired states [5]. The combination of neural network and fuzzy logic in the form of hybrid control can also be used for the purpose of providing better system performances [6, 7]. Inteco model of a servo system based on a DC motor will be used for experimental purposes in this paper. The artificial network for DC motor velocity control is presented in [8], where neural network control logic is formed in two parts: for estimating motor velocity and for generating control signal. Another example of well-formed neural network, which successfully controls a DC motor, is shown in [9]. A state observer will be used in this paper for estimating motor velocity. The velocity is estimated by the state observer in [10], where nonlinear control input for control of serially coupled DC motors is used. Neural networks, used for motor velocity estimations, are presented in [11-13].

The starting point of this paper is a servo system, based on a DC motor and brass inertia load. For the purpose of experimental research, the state observer is designed in Section 2, as well as a variable structure controller in Section 3. Observer design procedures are presented in [14, 15]. Poor velocity estimations of a servo system are experimentally obtained in this paper and presented in Section 4. The servo system possesses control limitations while working in sliding mode with large loads attached to the motor shaft. Those limitations directly affect unsatisfied observer velocity estimations, which are also presented in Section 4. The problem is solved by introducing neural network into the default control logic of the system. The artificial network is formed and trained with real experimental data in Section 5. Significant improvements of estimated velocities for different input signals are experimentally obtained in Section 6. Both velocity offset elimination and estimated error minimization justify the neural network implementation into default system control logic.

# 2 State Observer Design Procedure

State space coordinates of a motor are necessary for the practical implementation of the variable structure controller. A state observer represents an additional system for the state space coordinates estimation of the controlled object. The state space coordinates can be obtained at any time for a known input of the object. Often it is not possible to form an ideal model, whereas unknown and immeasurable disturbances appear on a real system. The Luenberger model is used in this paper in order to solve this problem. Comprehensive details and formulation of the Luenberger model can be found in [16]. The servo system, manufactured by Inteco, Poland [17], is powered by the Bühler 1.13.044.236 DC motor, whose characteristics and parameters are given in [18]. A mathematical model of the observer can be represented by the form:

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + B_0\left(c(t) - \hat{c}(t)\right); \qquad \hat{c}(t) = D\hat{x}(t), \tag{1}$$

where $\hat{x}(t)$ and $\hat{c}(t)$ are the state space vector and the observer output, respectively. The estimated error is defined as:

$$e(t) = x(t) - \hat{x}(t) \Rightarrow \dot{e}(t) = \dot{x}(t) - \dot{\hat{x}}(t). \tag{2}$$

It is possible to neglect viscous friction and the inductance of the rotor circuit during the design procedure of the observer, because Bühler DC motor used in this paper is a small power motor. On the basis of this possibility, a differential state equation and an output equation can be represented as:

$$\begin{bmatrix} \dot{\theta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{1}{T_s} \end{bmatrix} \begin{bmatrix} \theta \\ \omega \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{K_s}{T_s} \end{bmatrix} u_r(t); \qquad c(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \theta \\ \omega \end{bmatrix}, \tag{3}$$

where $K_s$ and $T_s$ can be calculated as: $K_s = K/\beta R + K^2$ and $T_s = RJ/\beta R + K^2$. Desired parameters can be calculated using the data from the engine specifications: $a = 1/T_s = 10,526$ and $b = K_s/T_s = 2273,68$. The characteristic equation of the motor model is:

$$\det[SI - A] = \det\begin{bmatrix} s & 0 \\ 0 & s+a \end{bmatrix} = s(s+a) = s^2 + sa. \tag{4}$$

The poles of this system are $s_1 = 0$ and $s_2 = -a$ [19]. A desired range of the observer poles is such that the pole at zero will be moved to the new position $s_1 = -20$, while the pole at $-a$ will be moved to $s_2 = -22a$. The characteristic equation of the motor now takes the form: $(s+20)(s+22a) = 0$.

The observer matrix $A_0$ is designed using a rule:

$$\det[SI - A_0] = (s+20)(s+22a) = 0. \tag{5}$$

It is known that: $A_0 = \begin{bmatrix} 0 & 1 \\ 0 & -a \end{bmatrix}$, $B_0 = \begin{bmatrix} l_1 \\ l_2 \end{bmatrix}$, $D = \begin{bmatrix} 1 & 0 \end{bmatrix}$, so the characteristic observer equation can be found from:

$$\det[SI - A_0] = \begin{bmatrix} s+l_1 & -1 \\ l_2 & s+a \end{bmatrix} = (s+l_1)(s+a) - (-l_2) = 0. \tag{6}$$

Values $l_1 = 236$ and $l_2 = 9,0385$ are calculated by equalizing coefficients from (5) and (6).

The final form of the designed state observer can be presented as:

$$\begin{aligned} \dot{\hat{\theta}}(t) &= \hat{\omega}(t) + 236\left(\theta(t) - \hat{\theta}(t)\right), \\ \dot{\hat{\omega}} &= -10,56\hat{\omega}(t) + 2273,68u_r(t) + 9,0385\left(\theta(t) - \hat{\theta}(t)\right). \end{aligned} \tag{7}$$

The experimental results obtained from the servo system are shown in Figs. 1 and 2. The estimated time response of the angular position from observer compared to the angular position of the motor is shown in Fig. 1. The time response of the estimated angular velocity of the observer is compared to the real angular motor velocity (Fig. 2). The presence of disturbances and noises, and the inability of their filtering are disadvantages of a standard linear observer. Those disturbances are present because some of them are slowly varying parameters which are not measurable. Position estimation is not significantly sensitive to the effects of disturbances and it is therefore accurate (Fig. 1). The velocity is sensitive to these disturbances, therefore, the estimation is not completely accurate. In practice, rotor current is introducing as a disturbance signal in the observer for the purpose of reducing a velocity estimation error. In our case, the motor current is not measurable parameter. As a result, it is not possible to introduce rotor current as a disturbance signal in the observer. As a result, we have velocity estimation error, which can be seen in Fig. 2.

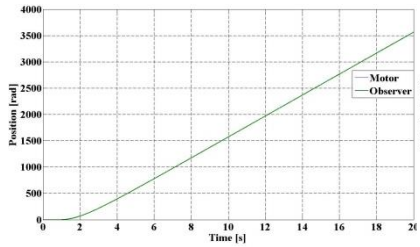



Figure 1
Time responses of estimated angular positions from observer and motor

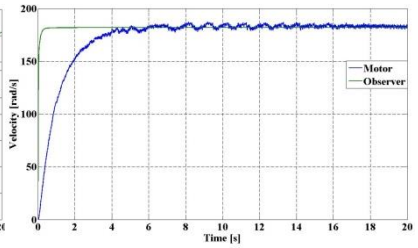



Figure 2
Time responses of estimated angular velocities from observer and motor

# 3 Variable Structure Controller

Variable Structure Control (VSC) is a control algorithm frequently used within nonlinear control systems. The main advantage of this approach is low sensitivity to parameter perturbations and disturbances, which makes it a robust control method [20, 21]. The dynamics of the second order system is represented by the following differential equations:

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= -ax_2 + bu, \quad a > 0,\ b > b_0 > 0. \end{aligned} \tag{8}$$

The design procedure of VSC logic consists of two steps. The first one is establishing a reaching motion within the system trajectory, which will move towards the sliding manifold and reach it in a finite time; the second step is to keep the motion of the trajectory on the manifold as $t$ tends to infinity.

If we choose the switching function as:

$$g = cx_1 + x_2, \tag{9}$$

the sliding manifold will be defined as $g = 0.$

Therefore, the motion of the system trajectory is governed by:

$$\dot{x}_1 = -cx_1. \tag{10}$$

From (10) it can be seen that the order of the sliding mode equation is less than the order of the original system, and the dynamics are determined by the parameter *c*. That means the dynamics during the sliding mode do not depend on the original system dynamics.

In order to provide stability for the system, the Lyapunov stability theory is used:

$$V = \frac{1}{2} g^2, \tag{11}$$

as a Lyapunov function candidate. The derivative of $V$ is:

$$\dot{V} = g\dot{g} = g(c\dot{x}_1 + \dot{x}_2) = g(cx_2 - ax_2 + bu) = g(c-a)x_2 + bgu. \tag{12}$$

If it is assumed that the derivative of the switching function $g$ satisfies the inequality:

$$\left|\frac{(c-a)}{x_2}\right| < \rho(x), \tag{13}$$

for some known $\rho(\mathrm{x})$ functions, from (12) and (13) the following inequality is obtained:

$$\dot{V} = g\dot{g} \le b|g|\rho(x) + bgu. \tag{14}$$

By choosing the control input signal $u$ as:

$$u = -\psi \operatorname{sign}(g), \quad \psi \ge \rho(\mathrm{x}) + \psi_0 > 0, \tag{15}$$

where:

$$\operatorname{sgn}(g) = \begin{cases} 1, \text{ if } g>0 \\ 0, \text{ if } g = 0 \\ -1, \text{ if } g < 0, \end{cases} \tag{16}$$

and substituting (15) in (14), the reaching and existence condition can be expressed as:

$$\dot{V} \le -b_0\psi_0. \tag{17}$$

We can calculate the parameter $c$ if condition (17) is fulfilled.

If the control input is defined as a quasi-relay sliding mode control:

$$u = \psi |x_2| \operatorname{sign}(g), \tag{18}$$

the stability, reaching, and existing conditions are calculated from:

$$\dot{V} = g\dot{g} \le 0,$$

$$g(c\dot{x}_1 + \dot{x}_2) < g(cx_2 - ax_2 + bu) < g\left[(c-a)x_2\right] + bgu < 0. \tag{19}$$

If we substitute (18) in (19), the following inequality is obtained:

$$g\left[(c-a)x_2\right] + b\psi |x_2||s| < 0, \tag{20}$$

and it is correct for all $c < a$.

The block diagram of a servo system with the observer and VSC is shown in the following Figure.

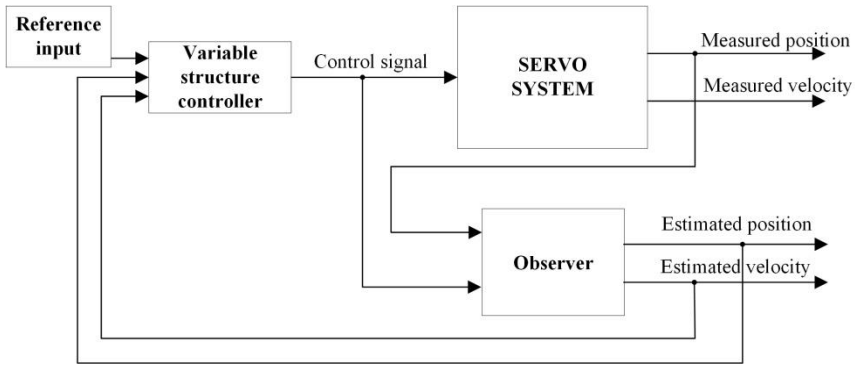



Figure 3
System block diagram with implemented state observer and variable structure controller

# 4 Servo System with Implemented VSC – Experiments and Poor Estimation Analysis

A graphic representation of the servo system used in all the experiments is shown in Fig. 4. The brass load weighing 2,030 kg, with the moment of inertia $J_{bi} = 0,001105 kgm^2$, is connected in Series to the shaft of DC motor.

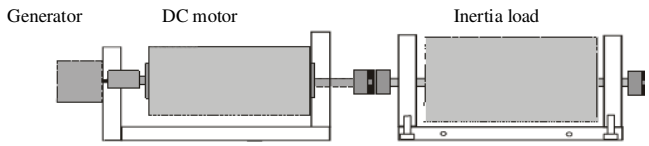



Figure 4
A servo system with brass load attached to the motor shaft

Six different step referent input values are used for experiments (Table 1). The main task is to find the system responses for different values of the proposed motor positions and to check the estimation quality. Experimental results of the servo system with implemented variable structure controller are shown in Figs. 5 to 10. From Figs. 5, 6, and 7 it can be seen that the observer precisely follows the angular position of the motor for all six input values.

Table 1
Referent inputs

| Signal type | Specified positions — Final values of a referent input signal | | | | | |
|---|---|---|---|---|---|---|
| STEP | $-25\cdot 2\pi$ | $-15\cdot 2\pi$ | $-5\cdot 2\pi$ | $5\cdot 2\pi$ | $15\cdot 2\pi$ | $25\cdot 2\pi$ |

A big deficiency of the estimated velocity from observer, is the offset appearance when the motor reaches desired positions. When the desired position is reached, Figs. 5-7, velocity should converge to zero. Analyzing Figs. 8, 9, and 10, it can be concluded that for all the input values, the observer estimates non-zero velocities after reaching the desired positions. These results are not satisfactory. The offset appearance cannot be tolerated, because it decreases observer reliability and related estimation accuracy. The second notable problem, from Figs. 8-10, is the poor estimation performance of transient processes. Requirements for optimal transient responses (13) are determined and based on [22]. The duration of transient conditions is a time period of starting, braking, and transition from one velocity to another. The time required to change the speed of the drive from $\omega_1$ to $\omega_2$, when all parameters have constant values, can be calculated as:

$$t_{1,2} = J\frac{\omega_2 - \omega_1}{T - T_L}, \tag{13}$$

where $T$ is the motor torque, and $T_L$ is load torque.

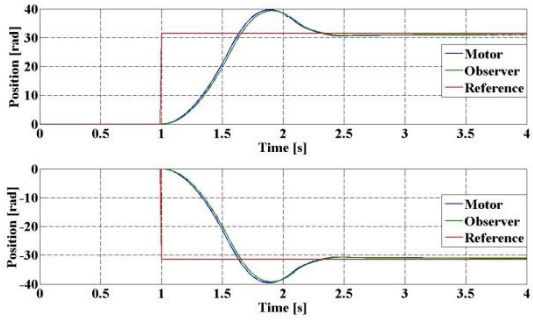



Figure 5
Estimated angular positions of observer and motor for specified motor positions: $5 \cdot 2\pi$ and $-5 \cdot 2\pi$

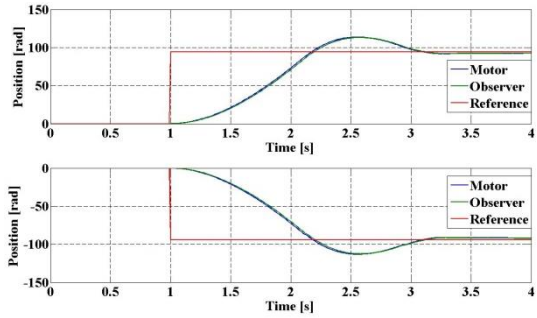



Figure 6
Estimated angular positions of observer and motor for specified motor positions: $15 \cdot 2\pi$ and $-15 \cdot 2\pi$

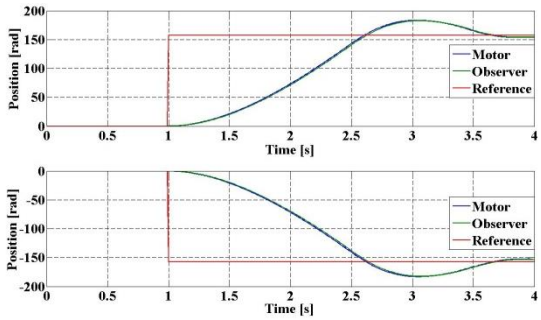



Figure 7
Estimated angular positions of observer and motor for specified motor positions: $25 \cdot 2\pi$ and $-25 \cdot 2\pi$

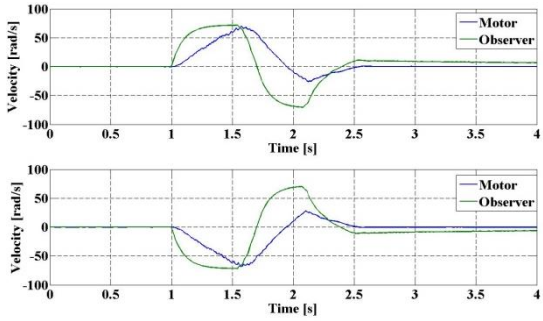



Figure 8
Estimated angular velocities for observer and motor for specified motor positions: $5 \cdot 2\pi$ and $-5 \cdot 2\pi$

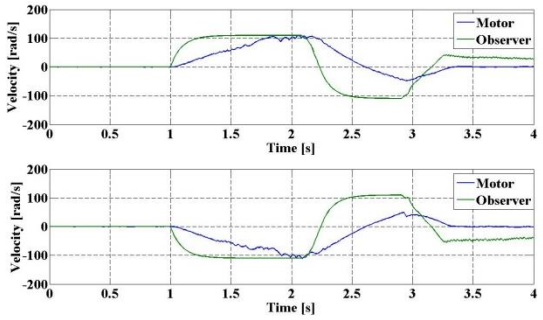



Figure 9
Estimated angular velocities for observer and motor for specified motor positions: $15 \cdot 2\pi$ and $-15 \cdot 2\pi$

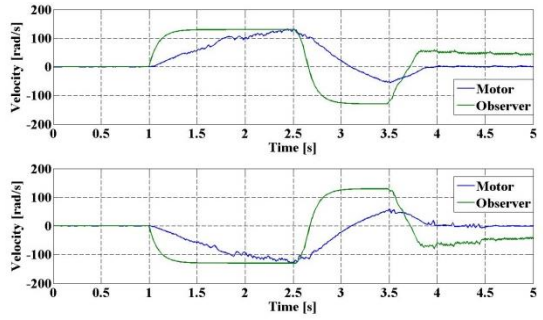



Figure 10
Estimated angular velocities for observer and motor for specified motor positions: $25 \cdot 2\pi$ and $-25 \cdot 2\pi$

Servo system time period $t_{1,2}$ varies according to the desired motor position from Table 1. It can be concluded from Figs. 8, 9, and 10 that estimated observer offset is getting larger with the increase of referent input signal. The relation between the moment of inertia of motor $J_{mot}$ and brass load inertia $J_{bi}$ can be calculated as follows:

$$\frac{J_{bi}}{J_{mot}} = \frac{0.001105 kgm^2}{0.000018 kgm^2} \approx 61. \tag{14}$$

Large load inertia compared to the motor inertia (14), and transition from one velocity to another are two main reasons for significant changes of time period $t_{1,2}$ (13). The transient process time period $t_{1,2}$ increases with the increase of the attached load on the motor shaft.

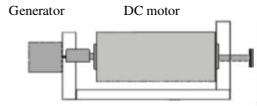



Figure 11
Servo system without load on the motor shaft

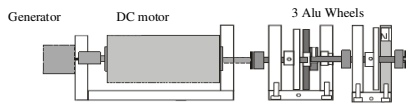



Figure 12
Servo system with aluminum wheels load

Two experiments are performed in order to show the validity of the previous analysis. In these experiments, only velocity time responses will be taken into consideration. The first experiment is based on the servo system from Fig. 11. The system is formed without any load attached to the motor shaft ( $J_l = 0$ ). VSC is included in the control logic and recorded angular velocity is presented in Fig. 13. In the second experiment, 3 aluminum wheels are attached to the motor shaft (Fig. 12). The total weight of the wheels is 0.15 kg and the moment of inertia is $\boldsymbol{J_{aw} = 0.00008\ kgm^2}$. Time response from this experiment is shown in Fig. 14. The value of referent input signal is $25 \cdot 2\pi$ in both experiments.

The relation between moment of inertia of the DC motor ( $J_{mot}$ ) and total moment of inertia of aluminum wheels ( $J_{a\omega}$ ) is:

$$\frac{J_{aw}}{J_{mot}} = \frac{0.00008 kgm^2}{0.000018 kgm^2} \approx 5. \tag{15}$$

The designed observer estimates velocity in a satisfactory manner while the system operates without any load attached. The observer ability to estimate the motor velocity decreases when the load moment of inertia is attached and increased. Velocity offset is notable between 2nd and 4th second on Fig. 14, but it converges to zero after 5 seconds. In the next chapter, a neural network will be designed with the purpose to improve observer estimation performances and overall system control.

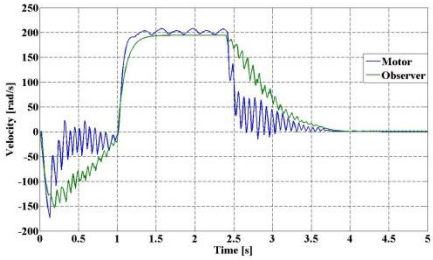



Figure 13
Experiment 1 — Estimated velocity without load torque on the motor shaft

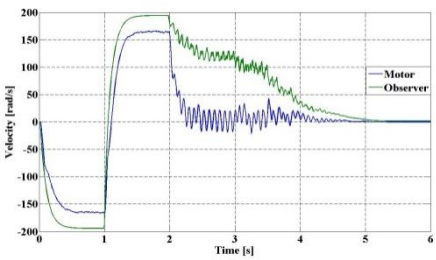



Figure 14
Experiment 2 — Estimated velocity with attached aluminum wheels to the motor shaft

# 5 Estimated Velocity Compensator Based on the Feedforward Neural Network

The block diagram of the servo system with the integrated neural network is shown in Fig. 15. A standard feedforward network is used for neural network realization. Real values from experimental model are imported for training purposes. Velocity test data from the observer and motor are used for network inputs and outputs, respectively. Four different signal types are used as referent input signals for training purposes: step, sinus, sawtooth, and square signal.

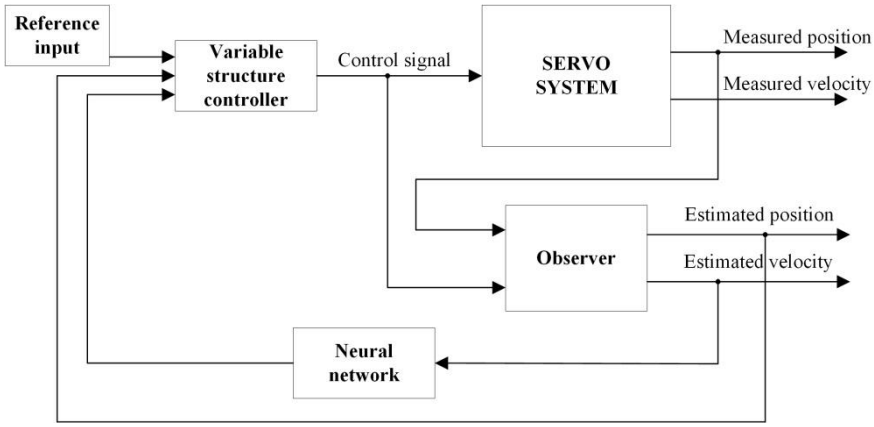



Figure 15
Modified servo system block diagram

Every input and output vector at the start of the process contained 1000 elements. Each input/output pair is obtained by performing experiments with different referent input values. Sixteen different referent input signals are used for this

purpose and they are presented in Table 2. Experiments are performed on the system with default control logic, which is presented in Fig. 3. The total number of elements of all vectors was 36000, which made the entire database large. The initial training procedures showed poor results with this database and the problem is solved with database reduction procedure.

Table 2

Referent inputs

| Signal type | $Is_{fv}$ — Final value of a referent input signal (Specified position) | | | | | |
|---|---|---|---|---|---|---|
| STEP | $1 \cdot 2\pi$ | $5 \cdot 2\pi$ | $10 \cdot 2\pi$ | $15 \cdot 2\pi$ | $20 \cdot 2\pi$ | $25 \cdot 2\pi$ |
| | $-1 \cdot 2\pi$ | $-5 \cdot 2\pi$ | $-10 \cdot 2\pi$ | $-15 \cdot 2\pi$ | $-20 \cdot 2\pi$ | $-25 \cdot 2\pi$ |
| SINUS | $25 \cdot 2\pi$ | $-25 \cdot 2\pi$ | | | | |
| SAWTOOTH | $25 \cdot 2\pi$ | $-25 \cdot 2\pi$ | | | | |
| SQUARE | $25 \cdot 2\pi$ | $-25 \cdot 2\pi$ | | | | |

A reduction of elements inside every vector was performed in order to optimize the training procedure. Each new formed vector included 100 elements instead of 1000, which made the database significantly smaller. The database reduction did not make a bad influence on neural network learning procedure. The main condition for neural network implementation into control logic is to design one-input/one-output network. The real time experimental environment has only processed vectors whose length is *mx1*. Input and output vectors are formed by merging all reduced input and output vectors respectively. The number of iterations, the type of training procedure, and the number of neurons in the hidden layer are determined experimentally. Neural network activation function is selected to be default hyperbolic tangent sigmoid transfer function (tansig). Sixteen different types of training processes are used for the initial testing phase, and the training results are shown in Table 3. Fields in Table 3 that are labeled with "x" indicate unsatisfactory training procedures, due to performance divergence, extremely poor results or too slow convergence of performance factors. Those data will not be analyzed. Standard training types integrated in Matlab software [23] are applied in this paper. The decision parameter that is used for selecting training process is total error which is made during the training. The total error is presented as "Performance" category in Tables 3 and 4.

The perfect result would be the one where a training procedure has Performance index equal to 0. Seven best training types from Table 3 are selected on the basis of performance results. Those 7 training types were analyzed further, the neuron number zones which could potentially give better results are experimentally selected, and new training procedures are performed.

Table 3
Initial training/testing phase

| Neurons: | 5 | 10 | 20 | 50 | 100 | 200 | 300 |
|---|---|---|---|---|---|---|---|
| Iterations: | 1000 | | | | | | |
| Training type | Performance | | | | | | |
| trainbfg | 2030.81 | 2121.26 | 2045.74 | 2539.57 | 3610.38 | 8231 | 13293.6 |
| trainbr | $2.35*10^{6}$ | $2.33*10^{6}$ | $2.02*10^{6}$ | $1.87*10^{6}$ | $1.71*10^{6}$ | x | x |
| trainbuwb | x | x | x | x | x | x | x |
| trainc | x | 10013.1 | 11378.3 | 8266.13 | x | x | x |
| traincgb | 2144 | 2140.42 | 2031.82 | 2054.61 | 1903.4 | 1744.45 | 3280.25 |
| traincgf | 2332.31 | 2166.54 | 1969.6 | 2009.24 | 2447.35 | 3909.72 | x |
| traincgp | 2048.78 | 2262.55 | 2020.01 | 2069.71 | 1937.8 | 2959 | 2698.41 |
| traingd | $2.7*10^{44}$ | $2.7*10^{48}$ | $2.17*10^{51}$ | $2.2*10^{57}$ | x | x | X |
| traingda | 2486.53 | 2935.52 | 4174.84 | 7837.16 | 16314.6 | x | x |
| traingdm | $3.66*10^{32}$ | $3.01*10^{36}$ | $1.17*10^{42}$ | $4.7*10^{47}$ | x | x | x |
| traingdx | 2278.56 | 2161.79 | 15606.5 | 3500.76 | 8335.5 | 12280.2 | x |
| trainlm | 2300.2 | 2114.77 | 2111.07 | 1946.85 | 1578.29 | 992.037 | 946.069 |
| trainoss | 2266.92 | 2240.2 | 1964.89 | 2368.82 | 2470.58 | 4092.24 | x |
| trainr | x | 4782.69 | 5338.76 | 6632.04 | x | x | x |
| trainrp | 2191.89 | 2179.37 | 2124.28 | 1929.01 | 2068.66 | 1678.16 | 2074.26 |
| trainscg | 2100.9 | 2102.76 | 2015.77 | 2185.61 | 2093.41 | 1748.67 | 2061.14 |

Table 4
Final training/testing phase

| Training type | trainscg | trainrp | **trainlm** | traingdx | traincgp | traincgb | trainbfg |
|---|---|---|---|---|---|---|---|
| Neurons | 30 | 40 | **600** | 7 | 30 | 15 | 3 |
| Performance | 2106.3 | 2078.4 | **617.1** | 2549.7 | 2594.4 | 2101.1 | 2342.9 |
| Neurons | 40 | 60 | **700** | 12 | 40 | 30 | 8 |
| Performance | 1995.1 | 2346.8 | **549.3** | 2151.6 | 2118.3 | 2005.2 | 2399.2 |
| Neurons | 125 | 150 | **800** | 16 | 70 | 40 | 15 |
| Performance | 2277.7 | 1914.4 | **429.5** | 2309 | 2721.1 | 2108.6 | 2064 |
| Neurons | 150 | 170 | **900** | 18 | 80 | 60 | 20 |
| Performance | 1733.7 | 1886.3 | **665.6** | 2282.6 | 2176.7 | 1884.1 | 2148.9 |
| Neurons | 170 | 190 | **1000** | * | 90 | 80 | 100 |
| Performance | 2398.7 | 1805.1 | **572** | * | 2034.6 | 1899 | 1892.9 |
| Neurons | 190 | 230 | **1100** | * | * | 120 | * |
| Performance | 2375.4 | 1715.1 | **534.3** | * | * | 2314.2 | * |
| Neurons | * | 350 | **1200** | * | * | 150 | * |
| Performance | * | 1688.2 | **339.4** | * | * | 1947.7 | * |

The results of those additional analyses are presented in Table 4. *Trainlm* is a network training function that updates weight and bias values according to the Levenberg–Marquardt optimization method. It is considered to be one of the fastest backpropagation algorithms. A deficiency of this function is the requirement of much more computer memory for its realization, in comparison to the other algorithms. The quantity of neurons in the hidden layer, the memory that is used, and the elapsed time for training processes are not considered while making the network choice. The final decision was to use 1200 neurons in the hidden layer and to use the Levenberg–Marquardt optimization training method on the basis of the results in Table 4.

# 6 Modified Servo System with Implemented VSC – Experiments

Figures 16-21 present time responses of controlled modified servo system from Fig. 15. Six step referent inputs, shown in Table 2, are used, and the system responses for each input signal are determined experimentally. The performance evaluation will be based on four parameters: a quality of the estimated angular position, an offset appearance degree, a speed of transient process, and a quality of the estimated angular velocity during the transient process.

The observer kept the estimation quality of angular positions for all input signals, as it is shown in Figs. 16, 17, 18. Figures 19, 20, and 21 show comparisons between estimated angular velocities from observer and motor. The observer velocity estimations in the steady states improved in comparison to the system responses presented in Figs. 8, 9, 10. Offset appearances problem is resolved and the errors for all the referent input signals from Table 1 were removed.

Table 5 is formed on the basis of the analysis of the transient processes before and after neural network implementation. The servo system transient process time duration is labeled as $t_{tp\_ss}$, the modified servo system (with implemented neural network) transient process time duration is labeled as $t_{tp\_ss+nn}$, $t_i$ represents increased time duration of transient process of the servo system after neural network implementation, $Is_{fv}$ is a symbol for the final value of referent input signal, and $I_t(\%)$ represents increased time duration comparing systems with and without a neural network. It can be concluded that all transient process speeds decrease in the range between 25.6% and 46.2% after neural network implementation. That implies that overall system responses get slower after artificial network implementation. The influence on speed performances can be classified as a deficiency, if there is a need for obtaining faster responses.

Table 6 represents observer estimation quality comparisons between experimental results of the system before and after neural network implementation. $Err_{ss}$ and $Err_{ss+nn}$ represent absolute errors of the systems without and with neural network, respectively. Each error is formed as an absolute value of the difference between the velocity estimated by observer and the actual motor velocity. It must be noted that only transient process parts of signals are analyzed. Further, $Erpr_{ss}$ represents the estimated error per second for the default servo system, $Erpr_{ss+nn}$ the estimated error per second in the modified servo system, ***EQI*** represents the Estimation Quality Improvement after neural network implementation, and $I_{OE}(\%)$ represents an Improvement of Observer Estimation in percentage — comparing the results before and after neural network implementation.

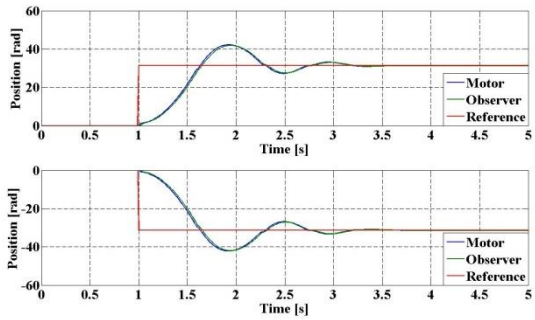



Figure 16
Estimated angular positions from observer and motor for specified motor positions: $5 \cdot 2\pi$ and $-5 \cdot 2\pi$, after neural network implementation

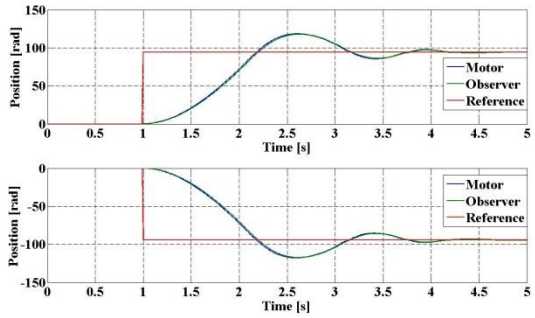



Figure 17
Estimated angular positions for observer and motor for specified motor positions: $15 \cdot 2\pi$ and $-15 \cdot 2\pi$, after neural network implementation

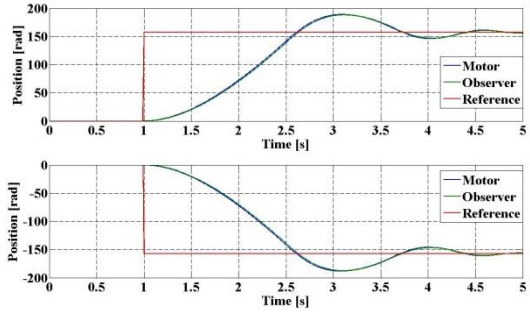



Figure 18
Estimated angular positions for observer and motor for specified motor positions: $25 \cdot 2\pi$ and $-25 \cdot 2\pi$, after neural network implementation

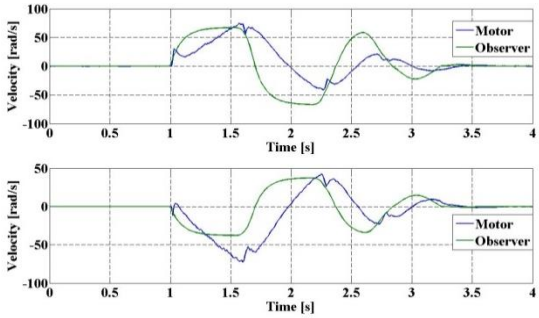



Figure 19
Estimated angular velocities for observer and motor for specified motor positions: $5 \cdot 2\pi$ and $-5 \cdot 2\pi$, after neural network implementation

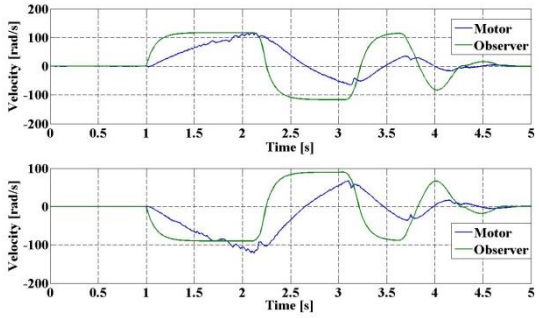



Figure 20
Estimated angular velocities for observer and motor for specified motor positions: $15 \cdot 2\pi$ and $-15 \cdot 2\pi$, after neural network implementation

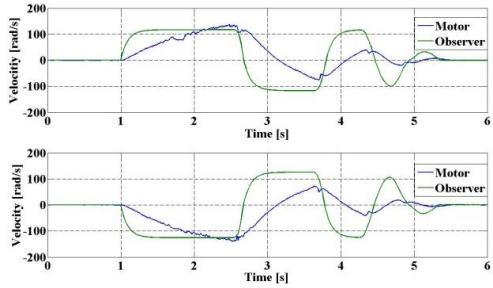



Figure 21
Estimated angular velocities for observer and motor for specified motor positions: $25 \cdot 2\pi$ and $-25 \cdot 2\pi$, after neural network implementation

Table 5
Servo system transient processes time durations

| $Is_{fv}$ | $t_{tp\_ss}$ (s) | $t_{tp\_ss+nn}$ (s) | $t_i$ (s) | $I_t(\%)$ |
|---|---|---|---|---|
| $-25 \cdot 2\pi$ | 4.45 | 5.59 | 1.14 | 25.6 |
| $-15 \cdot 2\pi$ | 3.35 | 4.90 | 1.55 | 46.2 |
| $-5 \cdot 2\pi$ | 2.55 | 3.55 | 1.00 | 39.2 |
| $5 \cdot 2\pi$ | 2.60 | 3.71 | 1.11 | 42.7 |
| $15 \cdot 2\pi$ | 3.40 | 4.91 | 1.51 | 44.4 |
| $25 \cdot 2\pi$ | 4.01 | 5.56 | 1.55 | 38.6 |

It can be concluded, from Table 6 (column $I_{OE}(\%)$), that the precision of the estimation is improved in 5 experiments. A neural network improves observer performances for all negative input signal values. For positive selection of input signals, a deficiency occurs when the final value of referent input signal is $15 \cdot 2\pi$. For all the other experiments from Table 5, the observer improvement is in the range between 2.2% and 28.2% after implementation of the neural network. The disadvantage of the modified servo system is increased duration of the time responses, which is a regular occurrence after artificial network induction. The overall conclusion is that implementation of the standard feedforward neural network can be a satisfactory solution to improve velocity estimations of the state observer. The network can compensate an error that occurs as a result of variable structure control logic and effects of large load moments of inertia.

Table 6
Observation quality comparisons before and after neural network implementation

| $Is_{fv}$ | $Err_{ss}$ | $Err_{ss+nn}$ | $Erpr_{ss}$ | $Erpr_{ss+nn}$ | ***EQI*** | $I_{OE}(\%)$ |
|---|---|---|---|---|---|---|
| $-25 \cdot 2\pi$ | $23.666 \cdot 10^3$ | $28.655 \cdot 10^3$ | $5.318 \cdot 10^3$ | $5.126 \cdot 10^3$ | ***yes*** | **3.6** |
| $-15 \cdot 2\pi$ | $13.826 \cdot 10^3$ | $16.990 \cdot 10^3$ | $4.127 \cdot 10^3$ | $3.467 \cdot 10^3$ | ***yes*** | **16.0** |
| $-5 \cdot 2\pi$ | $4.937 \cdot 10^3$ | $4.935 \cdot 10^3$ | $1.936 \cdot 10^3$ | $1.390 \cdot 10^3$ | ***yes*** | **28.2** |
| $5 \cdot 2\pi$ | $5.128 \cdot 10^3$ | $7.151 \cdot 10^3$ | $1.972 \cdot 10^3$ | $1.927 \cdot 10^3$ | ***yes*** | **2.2** |
| $15 \cdot 2\pi$ | $13.566 \cdot 10^3$ | $22.459 \cdot 10^3$ | $3.990 \cdot 10^3$ | $4.574 \cdot 10^3$ | ***no*** | **-12.8** |
| $25 \cdot 2\pi$ | $19.982 \cdot 10^3$ | $26.938 \cdot 10^3$ | $4.983 \cdot 10^3$ | $4.844 \cdot 10^3$ | ***yes*** | **2.8** |

## Conclusions

This paper presents a method of neural network use, in the servo system control logic, for the purpose of obtaining better estimation performance. The servo system that is used is based on a separately excited DC motor. System control includes Variable Structure Control (VSC) logic. Position and velocity are estimated by the Luenberger state observer. Poor observer velocity estimations are noted and it is experimentally shown what causes poor performance. The feedforward neural network is designed to properly compensate the control signal and solve those problems. The training type and number of neurons in the hidden layer are empirically determined by comprehensive simulation procedures. Modified control logic is tested by a series of experiments.

Velocity offsets in steady states are eliminated for each experiment. Further, modified control logic significantly reduced overall velocity estimation errors. The main disadvantage of the neural network implementation is the increased time duration of the servo system processes, in comparison to default control logic performance. In general, control logic should be modified in the way demonstrated in this work, if there is need for more reliable and accurate observer velocity estimations. On the other hand, it is better to avoid feedforward neural network implementation if the speed of a system time responses is of greater importance.

## Acknowledgement

This paper was realized as a part of the projects III 43007, III 44006, and TR 35005, financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia within the framework of integrated and interdisciplinary research.

## References

[1] A. Bretas, A. Phadke: Artificial Neural Networks in Power System Restoration, IEEE Transactions on Power Delivery, Vol. 18, No. 4, pp. 1181-1186, 2003

[2] T. Haque, M. Kashtiban: Application of Neural Networks in Power Systems, World Academy of Science, Engineering and Technology, Vol. 1, No. 6, pp. 917-927, 2005

[3] A. Wahab, A. Mohamed: Transient Stability Assessment of Power Systems using Probabilistic Neural Network with Enhanced Feature Selection and Extraction, International Journal on Electrical Engineering and Informatics Vol. 1, No. 2, pp. 103-114, 2009

[4] V. Vankayala, N. Rao: Artificial Neural Networks and Their Applications to Power Systems - A Bibliographical Survey, Vol. 28, No. 1, pp. 67-79, 1993

[5] A. Ishiguro, T. Furuhashi, S. Okuma: A Neural Network Compensator for Uncertainties of Robotics Manipulators, IEEE Transactions on Industrial Electronics, Vol. 39, No. 6, pp. 565-570, 1992

[6] A. Kusagur, S. F. Kodad, B. V. Sankar: Speed Control of Separately Excited DC Motor using Neuro Fuzzy Technique, International Journal of Computer Applications, Vol. 6, No. 12, pp. 29-44, 2010

[7] R. E. Precup, S. Preitl: Stability and Sensitivity Analysis of Fuzzy Control Systems, Mechatronics Applications, Acta Polytechnica Hungarica, Vol. 3, No. 1, pp. 61-76, 2006

[8] G. MadhusudhanaRao, B. SankerRam: A Neural Network Based Speed Control for DC Motor, International Journal of Recent Trends in Engineering, Vol. 2, No. 6, pp. 121-124, 2009

[9] A. Atri, M. Ilyas: Speed Control of DC Motor using Neural Network Configuration, International Journal of Advanced Research Computer Science and Software Engineering, Vol. 2, No. 5, pp. 209-212, 2012

[10] S. Mehta, J. Chiasson: Nonlinear Control of a Series DC Motor: Theory and Experiment, IEEE Transactions on Industrial Electronics, Vol. 45, No. 1, pp. 134-141, 1998

[11] P. Dzung, L. Phuong: Control System DC Notor with Speed Estimator by Neural Networks, Power Electronics and Drives Systems, Vol. 2, pp. 1030-1035, 2005

[12] N. Rai, B. Rai: Neural Network Based Closed Loop Speed Control of DC Motor using Arduino Uno, International Journal of Engineering Trends and Technology, Vol. 4, No. 2, pp. 137-140, 2013

[13] B. Mouna, S. Lassaad: Neural Network Speed Controller for Direct Vector Control of Induction Motors, International Journal of Engineering Science and Technology, Vol. 2, No. 12, pp. 7470-7480, 2010

[14] F. Farhani, A. Zaafouri, A. Chaari: Gain-scheduled Adaptive Observer for Induction Motors: An LMI Approach, Acta Polytechnica Hungarica, Vol. 11, No. 1, pp. 49-61, 2014

[15] S. Yogesh, M. Gupta, M. Garg: DC motor Speed Control using Artificial Neural Network, International Journal of Modern Communication Technologies & Research, Vol. 2, No. 2, pp. 19-24, 2014

[16] D. G. Luenberger, Introduction to Dynamic Systems: Theory, Models, and Applications, New York: John Wiley & Sons, 1979

[17] Inteco, Krakow, Poland, Modular Servo System - User's Manual, http://sharif.edu/~namvar/index_files/servo_um.pdf

[18] www.dc-motorshop.com/Motoren/DC+Motoren/artikel/1.13.044.236+-+24+V+DC+Motor.html

[19] R. E. Precup, R. C. David, E. M. Petriu, St. Preitl, M. B. Rădac: Fuzzy Control Systems with Reduced Parametric Sensitivity Based on Simulated Annealing, IEEE Transactions on Industrial Electronics, Vol. 59, No. 8, pp. 3049-3061, 2012

[20] V. I. Utkin, Sliding Mode Control, in Variable structure systems from principles to implementation, Eds. A. Sabanovic, L. M. Fridman, S. Spurgeon, pp. 3-18, 2004

[21] H. Khalil, *Nonlinear Systems*, New Jersey: Prentice Hall, 2002

[22] M. Chilikin: *Electric Drive*, MIR Publishers, Moscow, 1976

[23] H. Demuth, M. Beale, M. Hagan: Neural Network Toolbox™ 6 - User's Guide, 2013

# Strategies to Fast Evaluation of Tree Networks

**Raed Basbous**

Eastern Mediterranean University, Faculty of Arts and Sciences,
Department Mathematics,
Famagusta, North Cyprus via Mersin 10, Turkey, and
Al-Quds Open University, Department of Administrative Affairs,
Jerusalem, PO Box 58100, Palestine
E-mail: rbasbous@qou.edu

**Benedek Nagy**

Eastern Mediterranean University, Faculty of Arts and Sciences,
Department Mathematics,
Famagusta, North Cyprus via Mersin 10, Turkey, and
University of Debrecen, Faculty of Informatics, Department of Computer Science,
4010 Debrecen, PO Box 12, Hungary
E-mail: nbenedek.inf@gmail.com

*Abstract: Special tree graphs could model Cognitive Infocommunication Networks. The various modalities of the network are represented by various types of vertices, e.g., additions, multiplications. Tree graphs/networks are widely used in several other theoretical and practical fields. Expression trees are well-known tools to visualize the syntactic structure of the expressions. They are helpful also in evaluations, e.g., decision trees are widely used. Games and game theory form an important field in Artificial Intelligence and it has several connections to Optimization, Business and Economy. Game trees are used to represent games.*

*In this paper, certain types of tree networks are considered using various operations at their inner vertices, e.g., multiplication, (constrained) addition and the usual minimum and maximum (related to conjunction and disjunction of Boolean algebra). Evaluation techniques are presented, as well as, various pruning algorithms (related to short circuit evaluation in the Boolean case) that can quicken the evaluation in most cases. Based on the commutativity of the used operations, the evaluations can be more effective (faster) by reordering the branches of the tree. The presented techniques are useful to optimize (minimize) the size of the tree networks in various cases without affecting the final result/decision of the network.*

*Keywords: expression trees; game trees; short circuit evaluation; pruning of trees; decision making; reordered trees*

# 1 Introduction

Cognitive infocommunications (CogInfoCom) form an interdisciplinary field that works with various modalities [1]. These multimodal sensors usually form various networks. In this paper, we deal with special networks that form tree graphs. The multimodality is modelled by various types of nodes, let us say, operators in the tree. In our model the leaves may represent some measured data, and via the network a value (a decision, an evaluation of the situation, etc.) must be determined/computed. The edges of the tree specify the communication channels in which some data or (sub)result can be transferred from one node to another. Our model is also connected to the efficient evaluation process of various expression-trees that may represent CogInfoCom Networks.

Evaluation of various types of expressions is essential in every field connected to Mathematics, Computer Science and Engineering. Logical expressions can be evaluated in a fast way based on the following facts: if a member of a conjunctive formula is false, then the whole formula is false; if a member of a disjunctive formula is true, then the whole formula is true [10]. These facts are used in several programming languages to have a fast evaluation of logical formulae, e.g., in conditions [4]. This technique is called short circuit evaluation. There is a very similar idea that is used in game trees in Artificial Intelligence (AI). At the most investigated zero-sum two-player games, the minimax algorithm gives the best strategies for the players and it also answers the question, who has a winning strategy [8]. To compute the minimax algorithm every leaf of the tree is computed and the whole tree is evaluated. However, in most of the cases, it can be done with a much less effort, using alpha and beta pruning techniques [5, 8].

We assume that (AI) agents are starting from the leaves of the tree and they are communicating until the result is given at the root. In our CogInfoCom Network model the agents are various types of cognitive sensors. The sensors at leaves measure the environment/receive input data. The sensors/nodes communicate to each other in a bottom-up direction. When an inner cognitive sensor receives the data from its children nodes, it computes its task (based on its modality, i.e., type of operation) and sends its result to its parent node. The root node generates the final result/decision of the whole CogInfoCom Network. By using pruning techniques, there, usually a (much) less number of agents is enough to obtain the same results. We can reduce the communication cost of the network and possibly we do not need to use all the measured data, some subset of these data could be enough to infer the exact result of the network/expression tree. In this way fast evaluation techniques are important not only by saving time and space in a one-processor system when an expression is evaluated, but they also help to optimize the parallel resources in (bounded) parallel systems, e.g., CogInfoCom networks. There is also an enhancement of the previously mentioned pruning techniques: by reordering the branches of the tree network in such a way that the evaluation starts with the shortest branch.

Our algorithms can also be applied in artificial decision making systems that are closely related to systems used in cognitive-informatics and info-communications [9, 11]. Related human decisions and cognitive systems were also presented in [3].

In this paper, we consider special formula/expression trees (as models of CogInfoCom Networks) that can be considered a type of extension of the usual game trees/logical expression trees. Special extensions, as a mixture of decision and game-trees were already discussed in [6]. Here, by a further step, such extensions of game-trees are investigated in which some operations may not be directly connected to games, but with usual (mathematical or logical) expressions to give more modalities to our systems. We use values 0 and ±1 at the leaves of the trees. In some cases, it is not necessary to know the value of every descendant to evaluate a node of the tree, these cases lead to various pruning techniques that are presented here, in this way simplifying the (evaluation of the) network.

In the next section we recall some analogous techniques (mostly from [6, 8]) and we also fix our notations. We also show that reordering the branches of the tree (restructuring a Boolean network based on two modalities represented by conjunction/minimum and disjunction/maximum) may lead to an even faster evaluation. In Section 3 we present the results about pruning the tree networks considered here. In Section 4, it is presented that reordering of the branches of the tree may lead to a faster exact evaluation of the network. Finally, a short concluding section closes the paper.

# 2 Preliminaries

In this section we recall some related concepts, specially decision trees, game trees, minimax algorithm and alpha-beta pruning. We also present short circuit evaluation techniques for Boolean logic and, further, in Section 4 we show, as one of our new results, its more advanced version, based on the reordering of the branches of the expression tree.

In this paper, we use the term inner node for every node (including the root node) that is not a leaf.

## 2.1 Decision Trees

In this subsection we recall decision trees (see, e.g. in [6]); they can also be used to evaluate games against the “Nature”. Let a person be given who has some decision points, and some random events with known probabilities, also let us consider a tree with decision nodes and chance nodes. At decision nodes the person chooses a successor node. At chance nodes a random event happens: the successor node is chosen randomly with the probabilities known in advance. The

leaves of the tree represent the payoff values. The aim of the player is to maximize the payoff value and for this purpose she/he chooses the branch that leads to the highest expected value at every decision node.

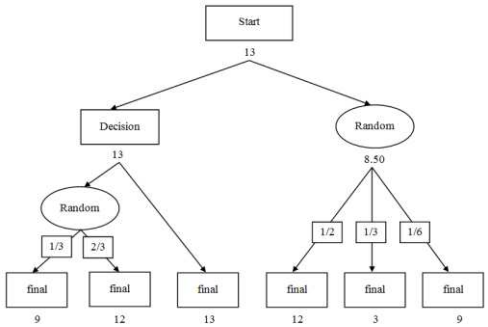



Figure 1
A decision tree

Assume that P(N) denotes the probability of the event N. Figure 1, shows an example of a decision tree. It includes decision nodes, where the person chooses among the successor, i.e. children nodes (represented by rectangles). At nodes represented by ellipses random events will determine the successor node (the probabilities are written on the edges). A numeric value, the expected outcome, is assigned to every node of the tree; it is shown under the rectangle/ellipse. These values are computed by Algorithm 1 (it is from [6]).

**Algorithm 1 (Decision)**

```
1.    function Dec(N)
2.    begin
3.    if N is leaf then
4.      return the value of this leaf
5.    else
6.     let N1, N2, …, Nm be the successors of N
7.     if N is a decision node then
8.       return max {Dec(N1), Dec(N2), …, Dec(Nm)}
9.     if N is a chance node then
10.      return P(N1)Dec(N1)+ … + P(Nm)Dec(Nm)
11.   end Dec
```

The technique presented in Algorithm 1 is called expectimax [6], since at decision points it computes the maximum of the expected values of the successor nodes, while at chance nodes the expected values of the successor nodes are computed by finding the sum of the assigned probabilities multiplied by the value of the given successors.

In this paper we deal only with problems having a bounded set of possible outcomes $\{0, \pm 1\}$. This restriction/simplification on the input/measured data allows us to develop some efficient algorithms.

## 2.2 Game Theory

An important and widely investigated field of game theory is about deterministic, strategic, two-player, zero-sum, finite games with perfect information. An instance of a game begins with the first player's choice from a set of specified alternatives, called moves. After a move, a new state of the game is obtained; the other player is to make the next move from the alternatives available to that player. In same state of the game there is no move possible, the instance of the game has been finished. In such states each player receives a payoff, such that their sum is zero, and therefore it is enough to know the payoff of the first player. In some typical zero-sum games, the value +1 assigned to the player in case of win, -1 in case of lose, and 0 in case of draw.

Game trees can be used to represent games. The nodes represent the states of the game, while the arcs represent the available moves. In the game tree there are two kinds of nodes representing the decision situations of the two players. At the root node the first player has decision. The leaves represent terminal states with their payoff values (for the first player).

In every two player, zero sum, deterministic game with perfect information there exists a perfect strategy for each player that guarantees the at least result in every instance of the game. The most fundamental result of game theory is the minimax theorem and the minimax algorithm. The theorem says: If a minimax of one player corresponds to a maximin of the other player, then that outcome is the optimal for both players.

### 2.2.1 Minimax Algorithm

Minimax theorem is a fundamental result of game theory. Players adopt strategies which maximize their gains, while minimizing their losses. Therefore, the solution is the optimal for each player that she/he can do for him/herself in the face of opposition of the other player. These optimal strategies and the optimal payoff can be determined by the minimax algorithm. It uses simple recursive functions to compute the minimax values of each successor state [8].

Algorithm 2 represents the way that minimax algorithm works (it is recalled from [6, 8]), see also, e.g., Figure 2 for examples.

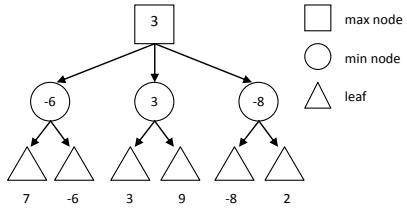



Figure 2
A minimax tree

**Algorithm 2 (MIN and MAX)**

```
1.   function MAXValue(N)
2.   begin
3.   if N is leaf then
4.      return the value of this leaf
5.   else
6.      let v = -∞
7.      for every successor N_i of N do
8.         let v= max{v, MINValue(N_i)}
9.   return v
10.  end MAXValue

1.   function MINValue(N)
2.   begin
3.   if N is a leaf then
4.      return the value of this leaf
5.   else
6.      let v = +∞
7.      for every successor N_i of N do
8.         let v= min{v, MAXValue(N_i)}
9.   return v
10.  end MINValue
```

### 2.2.2 Alpha-Beta Pruning

The minimax algorithm evaluates every possible instance of the game, and thus to compute the value of the game, i.e. its optimal payoff, takes usually exponential time on the length of the instances of the game. To overcome on this issue special cut techniques can be used. The alpha-beta pruning helps find the optimal values without looking at every node of the game tree. While using minimax, some situations may arise when searching of a particular branch can safely be terminated. So, while doing search, these techniques figure out those nodes that do not require to be expanded.

The way that this algorithm works can be described as below.

- Max-player cuts off search when she/he knows that Min-player can force a clear bad (for the first player, i.e. for Max-player) outcome.
- Min-player cuts off search when she/he knows that Max-player can force a clear good (for Max-player) outcome.
- Applying alpha-pruning (beta-pruning) means the search of a branch is stopped because a better opportunity for Max-player (Min-player) is already known elsewhere. Applying both of them is called alpha-beta pruning technique.

Figure 3 shows an example of beta-pruning, when β becomes smaller than or equal to α, we can stop expanding the children of N.

These algorithms, shown in Algorithm 3, are recalled from [6, 8].

**Algorithm 3 (MIN and MAX PRUNING)**

```
1.   function ALPHAPrune(N, α, β)
2.   begin
3.   if N is leaf then
4.      return the value of this leaf
5.   else
6.      let v = -∞
7.      for every successor Ni of N do
8.         let v= max{v, BETAPrune(Ni, α, β)}
9.         if v≥ β then return v
10.        let α = max{α,v}
11.  return v
12.  end ALPHAPrune

1.   function BETAPrune(N, α, β)
2.   begin
3.   if N is a leaf then
4.      return the value of this leaf
5.   else
6.      let v = +∞
7.      for every successor Ni of N do
8.         let v= min{v, ALPHAPrune(Ni, α, β)}
9.         if v ≤ α then return v
10.        let β = max{β,v}
11.  return v
12.  end BETAPrune
```

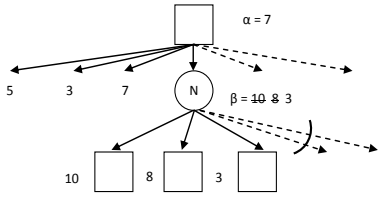



Figure 3
A beta-pruning example

## 2.3 Short Circuit Evaluation

Short circuit evaluation of a logical expression is widely used in programming languages for optimization. The short circuit evaluation technique is a type of time saving and, in some cases, it is also used for safety reasons [7]. For example, if it is known that all three conditions/variables must be true in order to proceed, it is not necessary to check the second and third conditions if the first one is already known to be false. In some programming languages the symbols && and ‖ are used for the logical operations AND and OR, respectively. These operations are working in short circuit evaluation.

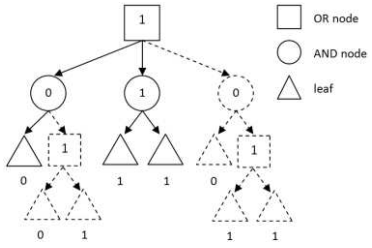



Figure 4
An example of applying the short circuit evaluation on a tree. Only three inner nodes and three leaves (total 6 nodes) out of six inner nodes and eight leaves (total 14 nodes) are explored and evaluated to get the final result in the root.

### 2.3.1 The AND Version

Let A, B and C be three Boolean expressions. Let them be in a conditional statement using && as follows.

```
if ( A && B && C )          {   statement; }
```

The only way this expression can be true is if A and B and C are all true. If A is false, we don't even need to consider B and C because we already know that the entire expression is false, etc. There is no need for further evaluation, as soon as it can be determined, through short circuit evaluation that the expression is false.

### 2.3.2 The OR Version

Similar to the previously described AND operation, the OR operation can also be terminated early in some cases. Let us consider the following statement.

```
if (A || B || C)        {  statement; }
```

In the case of OR, if it is known that the first condition is true, then there is no need to check the value of second and third conditions. If A is false, but B is true, then it is not needed to evaluate C to know the truth value of the condition.

### 2.3.3 The Mixed Version

Let us consider a more complex Boolean formula having only conjunctions and disjunctions. In this way, these special Boolean formulae can model tree networks having two types of operations/modalities. Figure 4, shows an example of a complex tree and also of applying the short circuit evaluation on the tree. As shown in this example, when evaluating the AND nodes, if a zero value (0) returned back from one of the children nodes, then we cut the next connected nodes and leaves and there is no need to evaluate them since their value will not affect the final result. The same technique is applied in evaluating the OR nodes, the idea here is to cut-off when the value one (1) returned back from a connected node or leaf. The final result of evaluating that node will be one (1) regardless the value of the remaining connected children nodes. See also Algorithm 4 below which describes how the idea of the short circuit evaluation is used. We allow not only binary conjunctions and disjunctions and thus, we may assume that they are alternating by levels.

# 3 Algorithms for Trees with Several Modalities

Now, various algorithms are proposed to quicken the evaluation of special trees in which there are specific operations are used. We are dealing with trees with a bounded set of payoff values: -1, 0, 1. Apart from the usual min and max operations (that can be seen as AND and OR by restricting the values to the Boolean set, i.e. to {0,1}) we use the operations multiplication (that can also be seen as AND on the Boolean set) and sum (that usually can be seen as binary addition, i.e. OR on the Boolean set), as well. We keep only the sign of the sum to have the result inside the domain {-1, 0, +1}. The proposed modified minimax, sum and product pruning, minimax with sum, and minimax with product algorithms are described below in detail.

**Algorithm 4 (OR and AND PRUNING)**

```
1.    function ORPrune (N)
2.    begin
3.    if N is leaf then
4.       return the value of this leaf
5.    else
6.      let v = 0
7.      for every successor N_i of N do
8.        while v ≠ 1  do
9.           let v= v + ANDPrune(N_i)
10.       return v
11.   end ORPrune

1.    function ANDPrune (N)
2.    begin
3.    if N is a leaf then
4.       return the value of this leaf
5.    else
6.      let v = 1
7.      for every successor N_i of N do
8.         while v ≠ 0  do
9.            let v = v · ORPrune(N_i)
10.   return v
11.   end ANDPrune
```

## 3.1 Modified Alpha-Beta Pruning Algorithm

We start with an obvious modification of Algorithm 3. Since the set of payoff values is bounded, we can modify the alpha-beta pruning algorithm to do a cut when the maximum (+1) or the minimum (-1) value of the set is already found in ALPHAPrune and BETAPrune, respectively, as shown in Algorithm 5 below. Line 8 of Algorithm 3 is modified in both ALPHAPrune and BETAPrune functions to test if the values +1 and -1 (the possible maximum and minimum, respectively) have been found in the evaluated node, thus there is no need to visit the remaining successors for that node: a cut can be done.

Figure 5 shows an example for the usage of the modified pruning algorithm.

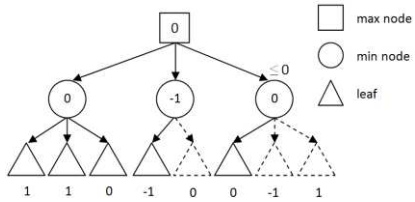



Figure 5
A modified minimax alpha-beta pruning. Only five leaves out of eight are explored to have an exact evaluation of the tree.

**Algorithm 5 (MIN and MAX PRUNING)**

```
1.    function MAXPrune(N, α, β)
2.    begin
3.    if N is leaf then
4.       return the value of this leaf
5.    else
6.      let v = -∞
7.      for every successor N_i of N do
8.        while v < +1  do
9.          let v= max{v,  MINPrune(N_i, α, β)}
10.         if v≥ β then return v
11.         let α = max{α,v}
12.   return v
13.   end MAXPrune

1.    function MINPrune(N, α, β)
2.    begin
3.    if N is a leaf then
4.       return the value of this leaf
5.    else
6.      let v = +∞
7.      for every successor N_i of N do
8.        while v > -1  do
9.          let v= min{v,  MAXPrune(N_i, α, β)}
10.         if v ≤ α then return v
11.         let β = max{β,v}
12.   return v
13.   end MINPrune
```

## 3.2 Sum and Product Pruning Algorithm

Now let us consider a new type of expression in which product operations follow the additions and vice versa, e.g., expressions of the form abc+de+fghi and also more complex expressions with these two operations appearing in the expression tree. Remember that both the possible values of the variables (leaves) and of the expressions (other nodes) are restricted to the set {-1, 0, +1} and thus both the Sum and Product functions can have these three output values.

Ideas similar to the short circuit evaluation can be used to cut during the evaluations of sum and product (multiplication) functions as they are described in Algorithm 6, see also, e.g. Figure 6 for examples.

**Algorithm 6 (MUL and SUM PRUNING)**

```
1.    function SUMPrune (N)
2.    begin
3.    if N is leaf then
4.       return the value of this leaf
5.    else
6.     let v = 0
7.     for every successor N_i of N do
8.       while |v|≤ (the number of remaining nodes to visit do)
9.          let v= v + MULPrune(N_i)
10.      if v > 0 then
11.            return 1
12.      if v< 0 then
13.            return -1
14.      else
15.            return 0
16.   end SUMPrune

1.    function MULPrune (N)
2.    begin
3.    if N is a leaf then
4.       return the value of this leaf
5.    else
6.      let v = 1
7.      for every successor N_i of N do
8.         while v ≠ 0  do
9.            let v = v * SUMPrune(N_i)
10.   return v
11.   end MULPrune
```

The possible cuts that are applied here are the following: For the SUM function, if the absolute value of the actual value is greater than the remaining successors to visit, then pruning is applied since there is no need to evaluate the remaining nodes: the result is already known: the return value is 1 if the sum value greater than zero, and -1 if the sum value is less than zero. This can be done using the while loop in line 8 in the function SUMPrune.

At the product operator a zero-cut is done, since 0 is the zero element of the multiplication, as it is written in the function MULPrune.

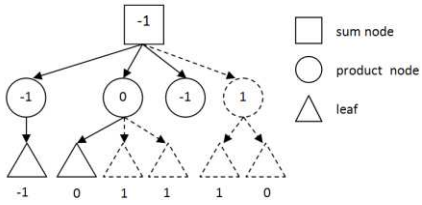



Figure 6

A sum and product pruning example. Only four inner nodes and two leaves (total 6 nodes) out of five inner nodes and six leaves (total 11 nodes) are explored and evaluated to get the final result in the root.

## 3.3 Minimax-Product Pruning Algorithm

In minimax-product case, the evaluation of the tree can be quickened by applying the zero-cut at product layer when a node with zero value is returned from one of the connected successors. In addition to this, similarly to the usual alpha-beta pruning, pruning can be applied in max and min layers. Example for these kinds of pruning are shown in Figure 7: a cut applied to the second node in min layer when a zero value (β) returned from the product layer, which is less or equal to the maximum value (α) that is found in the first node.

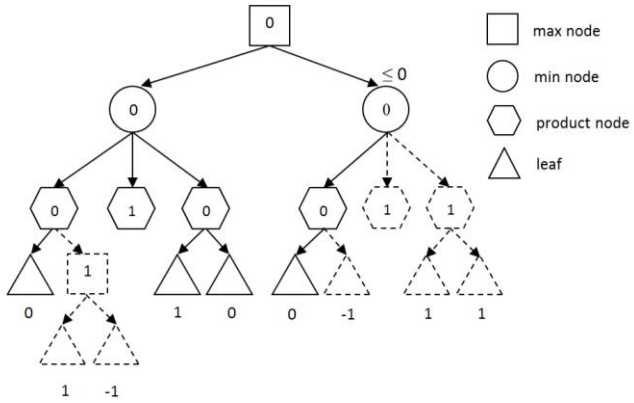



Figure 7

A minimax and product pruning example. Only seven inner nodes and three leaves (total 10) out of ten inner nodes and nine leaves (total 19) are explored and evaluated to get the final result in the root.

The proposed process for pruning is shown in Algorithm 7. The same MULPrune function that proposed previously in Algorithm 6 is used here with MAXPrune and MINPrune functions. (They can easily be modified having other order of layers in the expression tree.)

**Algorithm 7 (MUL, MIN and MAX PRUNING)**

```
1.   function MAXPrune(N, α, β)
2.   begin
3.   if N is leaf then
4.        return the value of this leaf
5.   else
6.      let v = -∞
7.      for every successor N_i of N do
8.          while v <+1  do
9.             let v=  max{v,  MINPrune(N_i, α, β)}
10.         let α = max{α,v}
11.  return v
12.  end MAXPrune
```

```
1.    function MINPrune(N, α, β)
2.    begin
3.    if N is a leaf then
4.       return the value of this leaf
5.    else
6.       let v = +∞
7.       for every successor Ni of N do
8.          while v > -1  do
9.             let v=  min{v,  MULPrune(Ni, α, β)}
10.         if v ≤ α then return v
11.         let β = max{β,v}
12.      return v
13.   end MINPrune
```

```
1.    function MULPrune(N, α, β)
2.    begin
3.    if N is a leaf then
4.       return the value of this leaf
5.    else
6.       let v = 1
7.       for every successor Ni of N do
8.          while v ≠ 0 do
9.             let v = v * MAXPRUNE(Ni, α, β)
10.      return v
11.   end MULPrune
```

## 3.4 Minimax-Sum Pruning Algorithm

In expression trees with three layers (max, min and sum), we use the idea of sum short circuit evaluation initiated in Subsection 3.2.

As mentioned before, the sum function has three output values (-1, 0, +1). When the absolute value of the sum of the visited successors is more than the number of remaining successors, then the cut can be done and if the sum is positive, then +1 is returned; if the sum is negative, then -1 is returned.

Furthermore, similar alpha and beta pruning for min and max functions can be done as in the previous algorithms to cut and quicken the evaluation of the tree (see Algorithm 8). In Figure 8, in the minimum layer a pruning is applied when the minimum value (-1) has been found, and after that in maximum layer again, when the maximum value (+1) has been found. The sum short circuit is applied too; since the absolute value of the sum function (+2) is greater than the remaining nodes to evaluate (we have one node remain to evaluate, which is less than 2).

**Algorithm 8 (ADD, MIN and MAX PRUNING)**

```
1.  function MAXPrune(N, α, β)
2.  begin
3.  if N is leaf then
4.      return the value of this leaf
5.  else
6.     let v = -∞
7.     for every successor N_i of N do
8.        while v < +1  do
9.           let v=  max{v,  MINPrune(N_i, α, β)}
10.       let α = max{α,v}
11. return v
12. end MAXPrune

1.  function MINPrune(N, α, β)
2.  begin
3.  if N is a leaf then
4.      return the value of this leaf
5.  else
6.      let v = +∞
7.      for every successor N_i of N do
8.         while v > -1  do
9.            let v=  min{v,  ADDPrune(N_i, α, β)}
10.        if v ≤ α then return v
11.        let β = max{β,v}
12. return v
13. end MINPrune

1.  function ADDPrune(N, α, β)
2.  begin
3.  if N is a leaf then
4.     return the value of this leaf
5.  else
6.    let v = 0
7.    for every successor N_i of N do
8.      while |v|≤  (the number of remaining nodes to visit do)
9.           let v= v + MAXPRUNE(N_i, α, β)
10.     if v > 0 then
11.          return 1
12.     if v< 0 then
13.          return -1
14.     else
15.          return 0
16. end ADDPrune
```

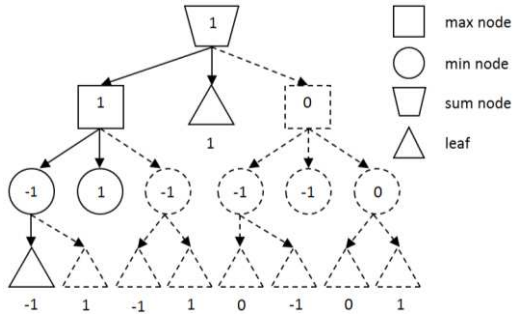



Figure 8

A minimax and sum pruning example. Only four inner nodes and four leaves (total 8) out of nine inner nodes and nine leaves (total 18) are explored and evaluated to get the final result in the root.

# 4 Reordering the Branches of the Trees

Notice that each of the used operations are commutative, and thus, the result of the operand does not depend on the order of the children branches. Therefore, in addition to the above proposed algorithms to quicken the evaluation of the tree networks with the presented logical and mathematical operations, a reordering technique can be applied on these kinds of trees before starting the evaluation.

The intuitive idea behind the reordering is that shorter branches can be computed faster. Therefore, the aim is to move the node that is the root of a subtree having least depth to the left side to be evaluated first. The process of reordering must be started from the lowest layers and goes up to the root. Then, after the reordering is done, the evaluation process can be started where the previously mentioned pruning techniques can be applied to quicken the evaluation.

## 4.1 Reordering the Branches of Boolean Expressions

In this subsection we show how the evaluation of a Boolean expression/CogInfoCom tree-network with two modalities can be done in a more efficient way. Figure 9, shows how the evaluation of the tree in Figure 4, is done after reordering its branches. After applying the reordering process, only two inner nodes (operators) out of six evaluated to get the final result in the root. While, only two of the leaves are used instead of the total eight during the evaluation process using the operators/modalities (AND and OR). The process of reordering the tree branches described in Algorithm 9. The same idea can be applied if we have different types of operators/modalities (e.g., SUM).

**Algorithm 9 (AND, and OR Tree Reordering)**

```
1.   function ORReorder(N)
2.   begin
3.   if N is leaf then
4.        move the leaf to the left  side
5.   else
6.      for every successor N_i of N do
7.           ANDReorder(N_i)
8.           Count the number of connected leaves and nodes
9.    move the node N_i with less leaves and nodes to the left side.
10.  end ORReorder

1.   function ANDReorder(N)
2.   begin
3.   if N is leaf then
4.        move the leaf to the left  side
5.   else
6.      for every successor N_i of N do
7.           ORReorder(N_i)
8.           Count the number of connected leaves and nodes
9.      move the node N_i with less leaves and nodes to the left side.
10.  end ANDReorder
```

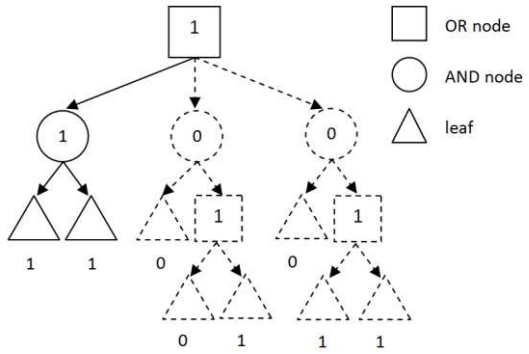



Figure 9

An example of reordering and pruning algorithm. Only 2 inner nodes and 2 leaves out of 4 inner nodes and 6 leaves were evaluated after the reordering and pruning process.

## 4.2 Reordering and Pruning Complex Trees

In this subsection we use the reordering technique for tree networks with more than two modalities and show that it can be used very efficiently in these cases as well.

In the first example of this subsection an expression tree is shown (see Figure 10 below). This tree includes four different operators (SUM, MULTIPLICATION, MAX, and MIN) in such a way that in the same level only the same operator is used. Since, in this case, the order of the operators are fixed (as at game trees), the evaluation functions call each other in a predefined order. To evaluate the tree of that example, without applying the reordering technique and pruning algorithms mentioned and detailed in this paper, 23 inner nodes and 26 leaves must be explored and evaluated. Figure 11 shows the same tree evaluated after applying the proposed pruning algorithms without reordering the branches. The same result of evaluation returned back to the root, but this was by exploring and evaluating only 19 inner nodes and 19 leaves (total 38) out of 23 inner nodes and 26 leaves (total 49).

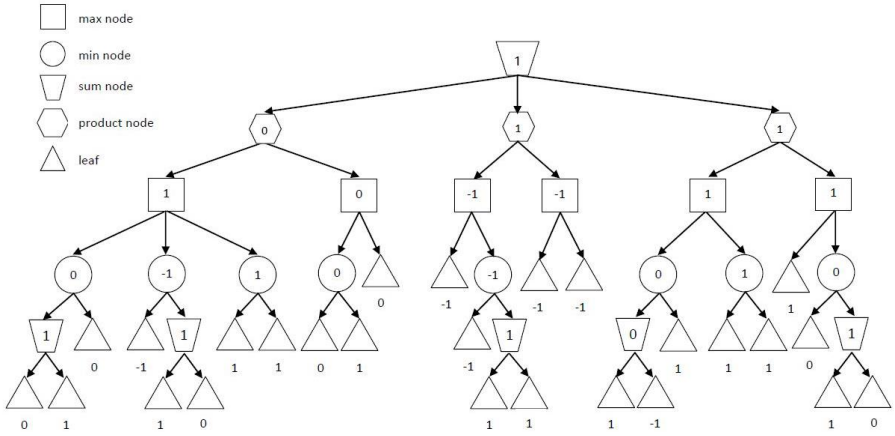



Figure 10

An example of an expression tree with SUM, MULTIPLICATION, MAX, and MIN operators. The tree contains 23 inner nodes and 26 leaves (total 49 nodes).

Figure 12 shows the same tree evaluated after reordering the branches and then applying the proposed pruning algorithms. The same result of evaluation returned back to the root, but this was by exploring and evaluating only 12 inner nodes and 7 leaves (total 19) out of 23 inner nodes and 26 leaves (total 49). This result shows how fast the evaluation process can be after applying the proposed technique to evaluate an expression trees that include various modalities (logical and mathematical operators, here).

In our other example, shown in Figure 13, the operators have no fixed order in the expression, which maybe much closer to real word applications in some cases. Also, leaves can be found in various levels (i.e., various depths) of the tree. To evaluate the tree without applying the reordering technique and pruning algorithms mentioned above, 11 inner nodes and 13 leaves must be explored and evaluated (total 24). Using various pruning strategies the number of inner nodes

that must be evaluated is 9, while the number of leaves that must be explored is 7 (total 16 nodes, see Figure 14). The proposed reordering technique together with the pruning algorithms evaluates the expression tree of this example by exploring and evaluating only 5 inner nodes and 4 leaves (total 9 nodes are visited for the evaluation) as it can be seen in Figure 15.

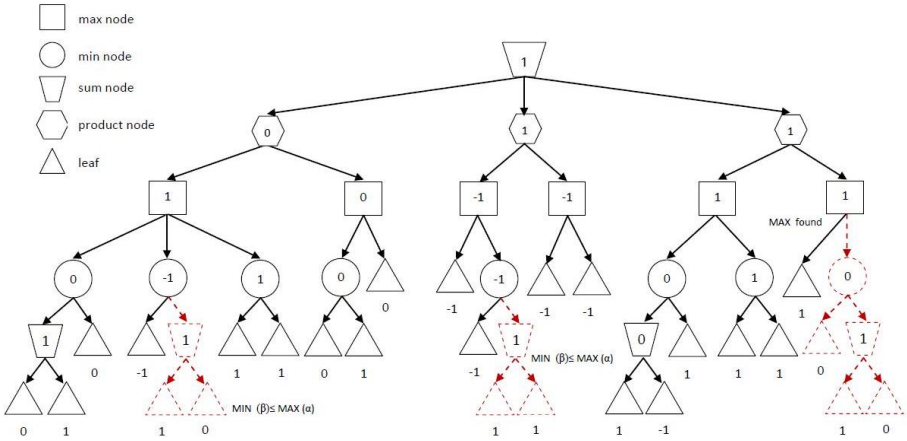



Figure 11

The expression tree of the example of Figure 10 is evaluated by pruning techniques. After applying the pruning algorithms without reordering the branches, 19 inner nodes and 19 leaves remained (total 38) out of 23 inner nodes and 26 leaves (total 49 nodes).

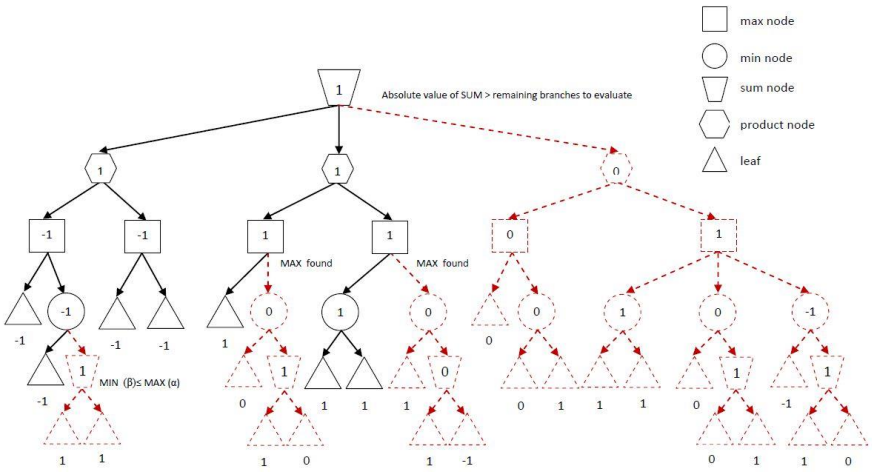



Figure 12

The expression tree of Figures 10 and 11 is evaluated by reordering and pruning. After applying the reordering and then the pruning algorithms, only 9 inner nodes and 7 leaves remained (total 16) out of 23 inner nodes and 26 leaves (total 49 nodes).

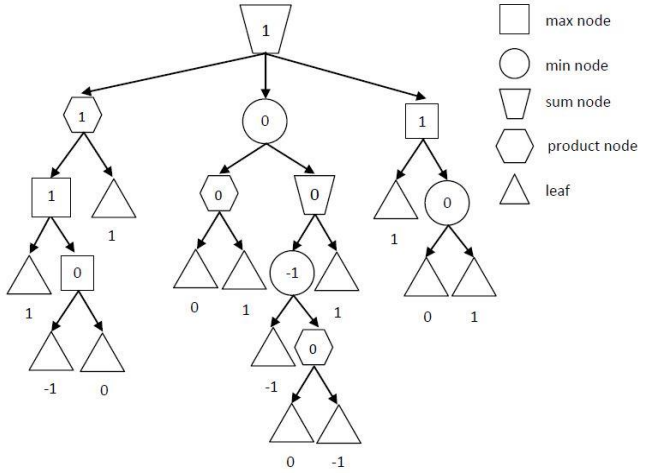



Figure 13
An example of an expression tree with SUM, MULTIPLICATION, MAX, and MIN operators in various order. The tree contains 11 inner nodes and 13 leaves (total 24 nodes).

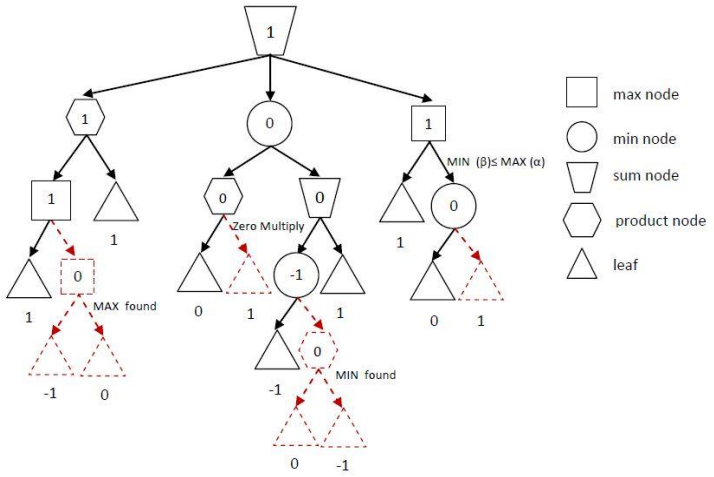



Figure 14
The example of Figure 13 is evaluated applying the pruning algorithms without reordering the branches: 9 inner nodes and 7 leaves evaluated and explored (total 16) out of 11 inner nodes and 13 leaves (total 24).

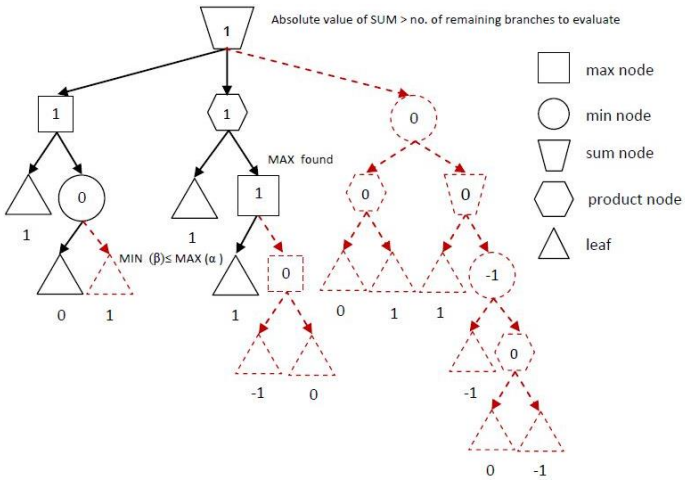



Figure 15

The expression tree of Figure 13 is evaluated by applying both the reordering and then the pruning algorithms: only 5 inner nodes and 4 leaves (total 9 nodes) are needed for the evaluation.

## Conclusions

We have considered simple network models, in which there are only finite communication channels. In our tree networks children nodes send their data/results to their parents and thus, the whole process is finite and can be modelled by evaluation techniques. There are well-known techniques to quicken the evaluation of Boolean expressions and similar techniques are used at game trees in alpha-beta pruning algorithms. In this paper, we have shown other pruning strategies: techniques for expressions where sum, product, minimum and maximum modeling various modalities in CogInfoCom networks. Similar optimization techniques could help in fast evaluations of various expressions (other types of trees and networks). Reordering the branches, by evaluating first, the shorter ones, may also help a lot to save time and energy, especially, when the evaluation costs of the leaves (or other nodes) consumes a large amount of energy or other limited source.

Considering a kind of parallel approach in which various agents can work together on the evaluation, these types of techniques can easily reduce the number of required agents and also to reduce their communication costs. Therefore, the presented algorithms can reduce algorithmic costs (time, space, number of agents/processors) in decision making systems and in various extensions of two-player zero-sum games. Artificial cognitive systems, CogInfoCom systems and other systems based on artificial/computational intelligence can be used in a more efficient way.

**Acknowledgement**

Some parts of the results of this paper were already presented in the conference CogInfoCom 2014, [2].

**References**

[1] Péter Baranyi, Ádám Csapó: Cognitive Infocommunications, CogInfoCom, Proceedings of 11th CINTI: IEEE International Symposium on Computational Intelligence and Informatics, Budapest, Hungary, pp. 141-146, 2010

[2] Raed Basbous, Benedek Nagy: Generalized Game Trees and their Evaluation, Proceedings of CogInfoCom 2014: 5th IEEE International Conference on Cognitive Infocommunications, Vietri sul Mare, Italy, pp. 55-60, 2014

[3] Peter Földesi, János Botzheim: Computational Method for Corrective Mechanism of Cognitive Decision-Making Biases, Proceedings of CogInfoCom 2012: IEEE 3rd International Conference on Cognitive Infocommunications, Kosice, Slovakia, pp. 211-215, 2012

[4] Brian Kernighan, Dennis Ritchie: The C Programming Language. Prentice Hall, Englewood Cliffs, NJ, 1988

[5] Donald Knuth, Ronald Moore: An Analysis of Alpha-Beta Pruning, Artificial Intelligence, Vol. 6, pp. 293-326, 1975

[6] Ervin Melkó, Benedek Nagy: Optimal Strategy in Games with Chance Nodes, Acta Cybernetica, Vol. 18, pp. 171-192, 2007

[7] Benedek Nagy: Many-valued Logics and the Logic of the C Programming Language, Proc. of ITI 2005: 27th International Conference on Information Technology Interfaces (IEEE) Cavtat, Croatia, pp. 657-662, 2005

[8] Stuart Russell, Peter Norvig: Artificial Intelligence, a Modern Approach. Prentice-Hall, 2003

[9] David Schum, Gheorghe Tecuci, Dorin Marcu, Mihai Boicu: Toward Cognitive Assistants for Complex Decision Making under Uncertainty, Intelligent Decision Technologies, Vol. 8, pp. 231-250, 2014

[10] Joseph Shoenfield: Mathematical Logic, A K Peters, 2001

[11] Yingxu Wang, Guenther Ruhe: The Cognitive Process of Decision Making, International Journal of Cognitive Informatics and Natural Intelligence (IJCINI) Vol. 1, pp. 73-85, 2007

# Ageing of Edges in Collaboration Networks and its Effect on Author Rankings

**Dalibor Fiala[1], Gabriel Tutoky[2], Peter Koncz[2], Ján Paralič[2]**

[1]University of West Bohemia, Univerzitní 8, 30614 Plzeň, Czech Republic
email: dalfia@kiv.zcu.cz

[2]Technical University of Košice, Letná 9, 04001 Košice, Slovakia
email: {gabriel.tutoky, peter.koncz, jan.paralic}@tuke.sk

*Abstract: In this paper we show that assigning weights to the edges in a collaboration network of authors, according to a decreasing exponential function depending on the time elapsed since the publication of a common paper, may add valuable information to the process of ranking authors based on importance. The main idea is that a recent collaboration represents a stronger tie between the co-authors than an older one and, therefore, reduces the weight of potential citations between the co-authors. We test this approach, on a well-known data set and with an established methodology of using PageRank-based ranking techniques and reference sets of awarded authors and demonstrate that edge ageing may improve the ranking of authors.*

*Keywords: collaborations; citations; PageRank; scholars; rankings*

# 1 Introduction

The ranking of authors of scholarly literature based on various bibliometric aspects has been popular in recent years for the purpose of the research assessment of individual scientists and related activities, such as prize and grant awarding, hiring, or promotion. It may indeed be considered a valuable complementary tool to manual research evaluation. In our previous research [1, 2], we were concerned with applying the recursive Google PageRank algorithm by Brin and Page [3] to bibliographic networks in order to rank researchers by prestige that is based on both the citation and the collaboration network of authors [4]. We have shown that such a ranking scheme may be more objective and fair. On the other hand, in another study [5] we have tested the effect of edge ageing in a collaboration social network [6] of teenagers. The main idea is that the strength of a collaboration tie (i.e. the edge weight) diminishes within the course of time. The goal of this present short paper is to show that we are able to combine both approaches (bibliographic PageRank and edge ageing) to produce author rankings that may

better reflect reality. There was other related work, prior to [1], for example, the pioneering studies on the usage of PageRank in bibliographic networks [7, 8, 9] and many more appeared later, e.g. [10, 11, 12, 13]. From the numerous papers on scientific collaboration networks, we discuss herein, two of the most well-known works – [14] and [15]. And, of course, the natural extension of each collaboration network analysis is a visualization process, e.g. [16] or [17], which is, however, beyond the scope of this article.

# 2 Methods and Data

The importance of nodes in a directed graph is often assessed by means of the number of in-coming edges and more advanced techniques are recursive in that they study the significance of these in-linking nodes by inspecting their in-links and so on. The importance, in this sense, is sometimes called prestige. One of these methods is the well-known PageRank algorithm by Brin and Page [3] originally conceived for the ranking of webpages, which may, nevertheless, be applied to any directed graph. Thus, let $G = (V, E)$ be a directed graph of citations between authors, with $V$ as the set of vertices (authors) and $E$ as the set of edges (with all citations between two authors in the same direction merged to one edge). So if author $v$ cites author $u$ (once or multiple times), there is exactly one edge $(v, u) \in E$. Then the PageRank score PR($u$) of author $u$ is computed recursively and the result depends on the scores of all citing authors, in the following way:

$$\mathrm{PR}(u) = \frac{1-d}{|V|} + d \sum_{(v,u)\in E} PR(v)\Omega \tag{1}$$

where $d$ is the damping factor (set to 0.85 in the original web experiments), and $\Omega$ is either $D_{out}^{-1}(v)$ (with $D_{out}$ being the out-degree of $v$) like in the standard PageRank by [3] or $\sigma_{v,u} \Big/ \sum_{(v,k)\in E} \sigma_{v,k}$ like in the bibliographic PageRank by Fiala et al. [1], where $\sigma_{v,k} = w_{v,k} \Big/ \left( \left[ (c_{v,k}+1)/(b_{v,k}+1) \right] \sum_{(v,j)\in E} w_{v,j} \right)$ and $w$, $b$, and $c$ are various coefficients that determine the weights of edges between authors. The parameters $w$, $b$, and $c$ are themselves based on the topology of both the citation and collaboration network of authors and for details we refer to [1] or to [2]. Here, let us say, that these coefficients help assign less weight to citations from colleagues and that the strength of relationship between two authors does not rely solely on the number of joint papers but also on other factors such as the number of coauthors in those papers. Based on the combination of the above parameters, Fiala [2] called the seven new PageRank (PR) variants tailored for bibliographic networks in this way: (*PR) collaboration*, *publications*, *allCoauthors*, *allDistCoauthors*, *allCollaborations*, *coauthors*, and *distCoauthors*.

As far as collaboration networks are concerned, we can additionally consider several approaches to the projection of collaborations among authors like binary weights, counting of co-occurrences, Newman's weights determination (see Eq. 2) and the *edge weights ageing principle*. We discuss all of these methods in detail in [5]. Briefly, a collaboration network of authors is a two-mode network [18] with two types of nodes – authors and their publications. In this network, actors are connected together via their common publications, but there does not exist, a direct connection between them (or between publications). This network needs to by projected onto a one-mode network with a single type of nodes – authors. See Figure 1 for an example of such a projection.

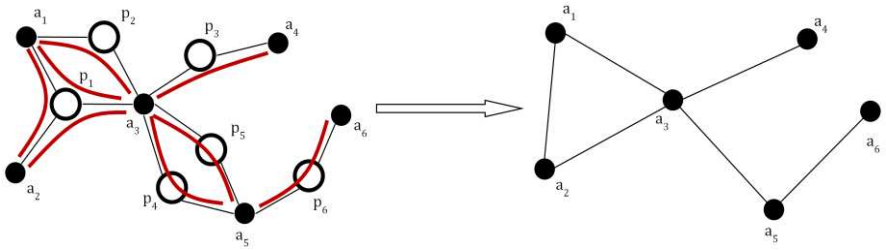



Figure 1
Projection of an author collaboration two-mode network
(a – authors, p – publications) onto a one-mode network

A projected relation (collaboration) between two authors depends on several factors, e.g. the number of co-authors of a single publication or the number of common publications of two selected authors, etc. An expression of that relationship in a binary form can be expressed as 1, if two authors have a common publication and 0 if not. A better approach is that the relation weight is expressed in the interval <0, 1> where 0 represents no relation and 1 represents the maximum strength of the relation. The first factor – number of co-authors of a single publication ($N_a$) can be expressed by the formula proposed by [19]:

$$w_{ij}^{p} = \frac{1}{N_a - 1} \tag{2}$$

or by the formulas proposed in [5], one of which, is an exponential expression with the decay parameter $\alpha_e$, which can be adjusted with respect to the particular type of collaboration network. In this way, it can influence the shape of the exponential curve:

$$w_{ij}^{p} = \alpha_e^{2-N_a} \tag{3}$$

where $w_{ij}^{p}$ is a projected weight of a single publication *p* between authors *i* and *j*.

Additionally, we can consider that collaboration strength between two authors is related to time-based factors – year of publication or more accurately to the time spent between their common publications and their frequency. We suppose that if the frequency of common publications of two authors is higher and/or their common publications are newer then their collaboration strength is larger. And vice versa, if the frequency of common publications is lower and/or their common publications are older, then the collaboration strength of those authors is smaller. In [5] we propose time dependent weights in the representation of a one-mode projected affiliation network – a kind of ageing of the ties (edges). This should be considered as a similar approach to the one presented in [20] or [21] where authors considered ageing of the nodes in the context of citation networks. They describe a node's age as influence on the probability of connecting the current node to new nodes in the network.

The ageing of the ties (collaborations) among authors can be described as an evolutionary process depicted in Figure 2 where $e_1, e_2, \ldots, e_m$ are events (common publications) on which authors $i$ and $j$ collaborate together. This ageing allows for the termination of sporadic and insignificant relations (e.g. when two authors have rarely participated in common publications) and, by contrast, if the relations are periodically repeated (we assume that these relations are significant), they are "highlighted" in the network and their weight is increased (see events $e_1$ through $e_3$ in Figure 2). After the creation of the first collaboration (based on the first event $e_1$), a tie with a value of 1 comes into being which decreases exponentially in time according to the formula

$$w_{ij}(t+\delta) = \left\{ \begin{array}{ll} w_{ij}(t)\, e^{-\theta \Delta t} & \text{when } w_{ij}(t)\, e^{-\theta \Delta t} > \varepsilon \\ 0 & \text{otherwise} \end{array} \right\} \tag{4}$$

where $w_{ij}(t)$ is the weight of the collaboration in time $t$ and $w_{ij}(t+\delta)$ is the collaboration weight after time $\delta$. The value $\varepsilon$ is the threshold value of a minimum collaboration weight and the factor $\theta$ is called the ageing factor which designates the "ageing speed" and is described by $\theta = \frac{\ln 2}{t_{1/2}}$ where $t_{1/2}$ is the time span after which the weight of ties decreases by 50% in the ageing process. For more details see [5].

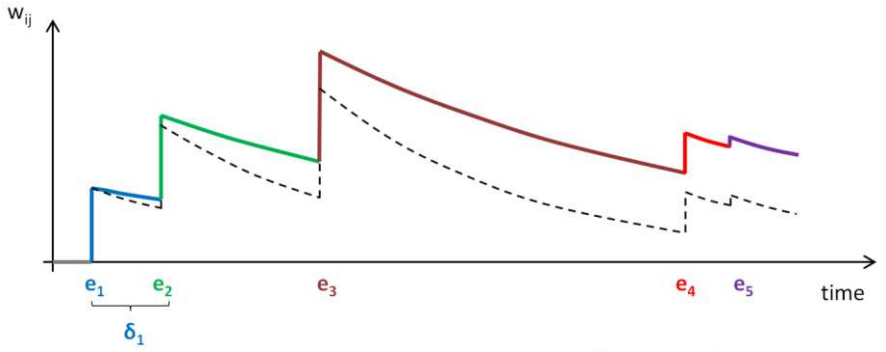



Figure 2

Evolution in time of a single collaboration tie between authors *i* and *j* for two various ageing factors (Events $e_1$, .., $e_5$ represent common publications between authors *i* and *j*. At the time when a publication is produced, the initial strength of the tie between the authors (value $w_{ij}$) is determined by the projection method and subsequently the tie strength decreases as time passes since the creation of the new publication.)

As for the data examined in our analysis, we investigated the same data set as in the study by Fiala [2], where there was a citation graph of 205,780 computer science publications from the period 1996-2005 with 276,957 citations between them. This paper citation graph was subsequently transformed to an author citation graph with 187,016 authors and 1,471,312 citations between them (with no self-citations allowed). The authors were represented by their surnames and given names initials and we did not perform any name disambiguation or unification. Self-citations of authors were removed before we started our analysis. The resulting (two-mode) collaboration network thus consisted of 392,796 nodes (publications + authors) and 492,284 edges (publication-author connections).

# 3 Results and Discussion

We applied those seven ranking techniques mentioned in the section on methods with ageing factors set on (i.e. "new methods") to the same author citation network described in [2]. The result of this application was seven author rankings which we compared to the rankings from the above study, in which edge weight ageing did not occur (i.e. "old methods"). The comparison was twofold: first, we calculated five statistics (minimum, maximum, median, mean value, and standard deviation) for the "new" and "old" rankings using the ranks of Codd[1] and Turing[2]

[1] Codd Award, http://www.sigmod.org/sigmod-awards/sigmod-awards#innovations
[2] Turing Award, http://awards.acm.org/homepage.cfm?srt=all&awd=140

awardees and plotted them as lines in Figure 3 and Figure 4 and, second, created boxplots for each pair of "new" and "old" rankings (with the first quartile, median, and third quartile of ranks forming the bar and the minimum and maximum ranks being the whiskers) displayed in Figure 5 and Figure 6.

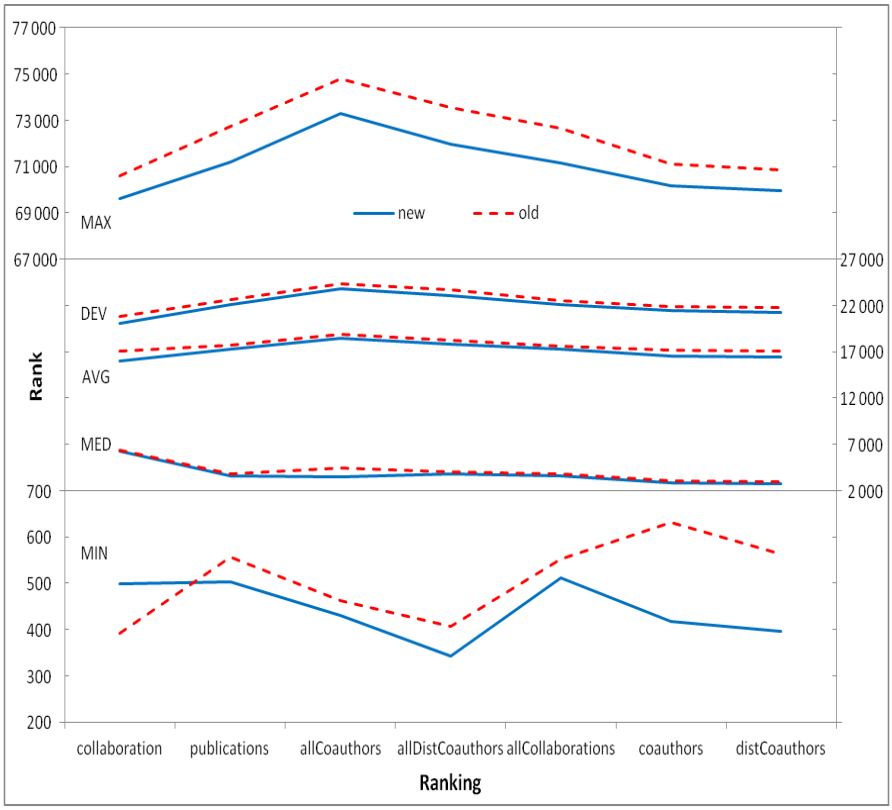



Figure 3

Codd Award and five statistics of new and old methods

(MAX – maximum rank, MIN – minimum rank, MED – median rank, AVG – mean rank, DEV – standard deviation of ranks produced by a particular ranking method. The lower the values of MAX, MIN, MED, and AVG, the better the ranking. Even for DEV, lower values mean a more compact ranking, which is a desirable property.)

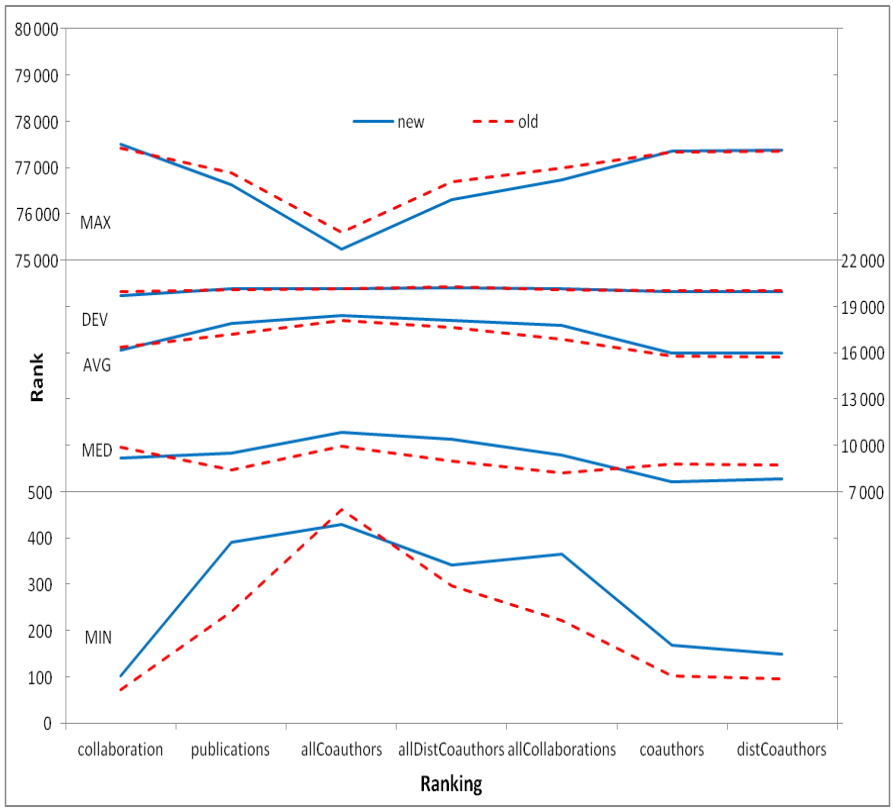



Figure 4

Turing Award and five statistics of new and old methods

(MAX – maximum rank, MIN – minimum rank, MED – median rank, AVG – mean rank, DEV – standard deviation of ranks produced by a particular ranking method. The lower the values of MAX, MIN, MED, and AVG, the better the ranking. Even for DEV, lower values mean a more compact ranking, which is a desirable property.)

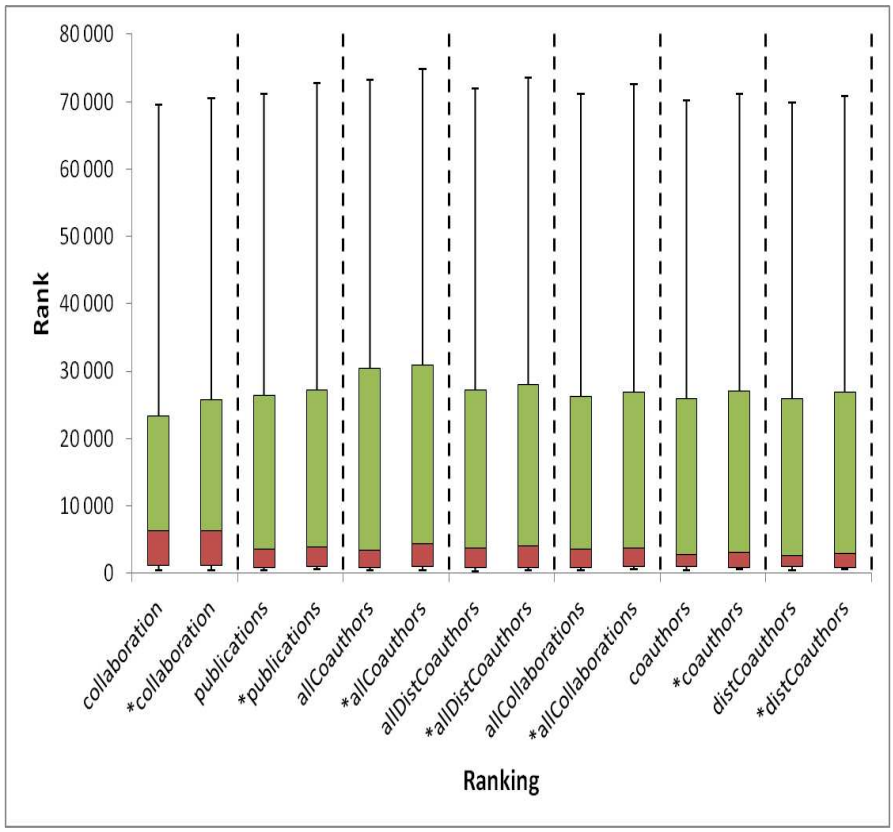



Figure 5

Codd Award boxplots (* = old method)

(The bottom of each bar represents the 25th percentile and the top is the 75th percentile. The median rank of each ranking method lies on the boundary between the red and the green rectangle. The minimum and maximum ranks achieved are marked with the whisker lines. The smaller a box, the greater the trend to produce better ranks for the awarded authors.)

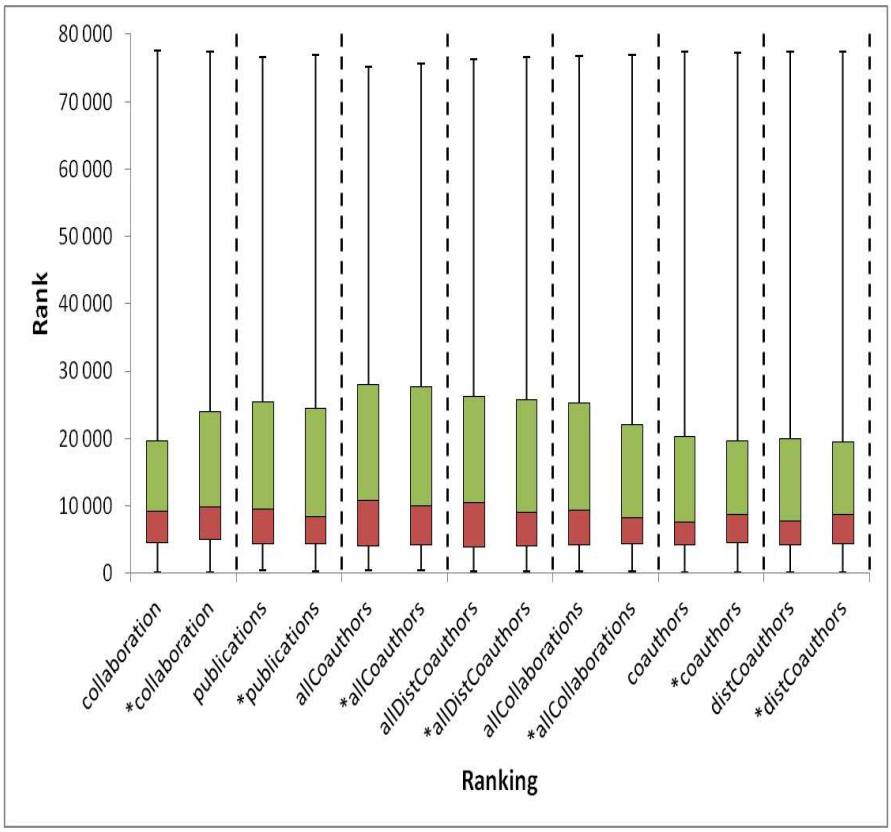



Figure 6

Turing Award boxplots (* = old method)

(The bottom of each bar represents the 25th percentile and the top is the 75th percentile. The median rank of each ranking method lies on the boundary between the red and the green rectangle. The minimum and maximum ranks achieved are marked with the whisker lines. The smaller a box, the greater the trend to produce better ranks for the awarded authors.)

The blue solid line represents the new methods (with ageing factor enabled) and the red dashed line represents the old methods from [2]. Since lower ranks mean better ranks (i.e. position 1 is better than position 100), we can see in Figure 5 that the new methods outperform the old ones, most remarkably in the terms of the maximum and minimum rank achieved (MAX and MIN) and only slightly regarding the other three statistics (median rank – MED, mean rank – AVG, and standard deviation of ranks – DEV). However, unlike Codd Award this phenomenon is much less visible in Figure 5 with Turing Award winners. Only the MAX statistics is clearly better there for the new methods. DEV is approximately the same and the remaining three are worse for the new methods.

Similar trends may be observed with the boxplots. While the bars tend to be positioned lower (towards better ranks) with the new methods for Codd Award winners in Fig. 5, it is almost the opposite for Turing Award laureates in Fig. 6. A first explanation of this observation may be the different nature of these awards and some specific features of the data set underlying our analysis. While the Codd Award, intended for database researchers, seems to be well modelled by the citation patterns in our sample of Web of Science data, the Turing Award for achievements in general computer science is less so. It appears that much more than the analysis of citation and collaboration networks is needed to identify the winners of such a broadly defined award. A further investigation into the different outcomes of the application of the ranking methods based on edge weight ageing to Codd Award and Turing Award winners will be necessary in the future.

The main weakness of the ageing of ties is the right determination of the $\theta$ factor used in Eq. 4. This factor must be estimated considering the data set and the time range of the publications and the awards considered. If we determine $\theta$ to yield a too "slow" ageing, the method does not have the required influence on the results and, conversely, when a too "fast" ageing is produced, the method suppresses nearly all of the ties in the network and we will not have enough ties sufficiently distributed across all possible strengths. As an optimum value for $\theta$ we have chosen $t_{1/2} = 730$ days, which means that the strength of all ties in the network decreases by a half of their value during 2 years. This value was determined based on previous experiments similar to the experiments described in [5] for estimating $\theta$, which are not described in this article.

**Conclusion**

In this short paper, we showed that the ageing of edges and their weights in the collaboration networks of authors, might add some useful information to the process of assessing the importance of individual authors. More precisely, by weighing the edges in a collaboration network of authors according to a decreasing exponential function, that depends on the time elapsed since a common publication was produced, we are able to differentiate between fresh and obsolete collaborations and to decide on the strength of relationship between two authors. This factor may then be input in a ranking scheme as in [1] or [2] that ranks authors in a citation network, taking into account the information from the corresponding collaboration network. We demonstrated this approach on the same data set and methodology used by [2] and showed that edge ageing improved the ranking of researchers when Codd Award winners were used as a reference set of outstanding scholars but did not make the ranking better when Turing Award winners were used. We argue that this contrast may be the result of the distinct nature of these awards (a specialized versus a general award) that are reflected differently by the underlying data set. We think examining the influence of the ageing of edges, in collaborative networks, on the ranking of authors is worth of further research.

**Acknowledgement**

This work was supported by the European Regional Development Fund (ERDF), project "NTIS - New Technologies for Information Society", European Centre of Excellence, CZ.1.05/1.1.00/02.0090 and in part by the Ministry of Education of the Czech Republic under grant MSMT MOBILITY 7AMB14SK090. This work was also supported by the Slovak Research and Development Agency under the contract No. SK-CZ-2013-0062 and by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/1147/12. Thanks are due to M. Dostal and C. Havrilová for their useful comments as well.

**References**

[1] Fiala, D., Rousselot, F., & Ježek, K. (2008) PageRank for Bibliographic Networks. *Scientometrics*, 76(1) 135-158

[2] Fiala, D. (2012) Time-Aware PageRank for Bibliographic Networks. *Journal of Informetrics, 6*(3) 370-388

[3] Brin, S., & Page, L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *Proceedings of the 7th World Wide Web Conference*, Brisbane, Australia, 107-117

[4] Newman, M. E. J. (2001) Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality. *Physical Review E*, *64*, art. no. 016132

[5] Tutoky, G. (2011) Discovery and Exploitation of Knowledge in Collaboration Social Networks. *Information Sciences and Technologies Bulletin of the ACM Slovakia*, *3*(4) 28-36

[6] Wasserman, S., & Faust, K. (1994) Social Network Analysis. Cambridge University Press, Cambridge, UK

[7] Liu, X., Bollen, J., Nelson, M. L., & Van De Sompel, H. (2005) Co-Authorship Networks in the Digital Library Research Community. *Information Processing and Management*, *41*(6) 1462-1480

[8] Bollen, J., Rodriquez, M. A., & Van De Sompel, H. (2006) Journal Status. *Scientometrics*, *69*(3) 669-687

[9] Chen, P., Xie, H., Maslov, S., & Redner, S. (2007) Finding Scientific Gems with Google's PageRank Algorithm. *Journal of Informetrics*, *1*(1) 8-15

[10] Ma, N., Guan, J., & Zhao, Y. (2008) Bringing PageRank to the Citation Analysis. *Information Processing and Management*, *44*(2) 800-810

[11] Yan, E., & Ding, Y. (2009) Applying Centrality Measures to Impact Analysis: A Coauthorship Network Analysis. *Journal of the American Society for Information Science and Technology*, *60*(10), 2107-2118

[12] Ding, Y. (2011) Applying Weighted PageRank to Author Citation Networks. *Journal of the American Society for Information Science and Technology*, *62*(2) 236-245

[13] Yan, E., & Ding, Y. (2011) Discovering Author Impact: A PageRank Perspective. *Information Processing and Management*, *47*(1) 125-134

[14] Newman, M. E. J. (2001) The Structure of Scientific Collaboration Networks. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2) 404-409

[15] Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002) Evolution of the Social Network of Scientific Collaborations. *Physica A*, *311*(3-4) 590-614

[16] Shannon, P., Markiel, A., Ozier, O., Baliga N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11) 2498-2504

[17] Horváth, A. (2013) The Cxnet Complex Network Analyser Software. *Acta Polytechnica Hungarica*, *10*(6) 43-58

[18] Latapy, M., Magnien, C., & Vecchio, N. D. (2008) Basic Notions for the Analysis of Large Two-Mode Networks. *Social Networks*, *30*(1) 31-48

[19] Newman, M. E. J. (2004) Who Is the Best Connected Scientist? A Study of Scientific Coauthorship Networks. *Lecture Notes in Physics*, *650*, 337-370

[20] Hajra, K. B., & Sen, P. (2005) Aging in Citation Networks. *Physica A*, *346*(1-2) 44-48

[21] Zhu, H., Wang, X., & Zhu, J.-Y. (2003) Effect of Aging on Network Structure. *Physical Review E*, *68*(5) art. no. 056121

# The Emergent Technological and Theoretical Paradigms in Education: The Interrelations of Cloud Computing (CC), Connectivism and Internet of Things (IoT)

**Mustafa Tuncay Sarıtaş**

Balikesir University, Necatibey Faculty of Education, Department of Computer Science and Instructional Technology, Soma cad., 10100, Balikesir, Turkey, tsaritas@balikesir.edu.tr

*Abstract: With the rapid development and increase in Internet speed in network technology, communication and interaction with others, even with objects, as well as sharing and creation of knowledge is supported and facilitated more effectively than ever before. The characteristics of society are in an incessant change towards a more social, networked, and connected one. The education sector is also in a constant state of evolution, naturally as a result of emerging technologies and theories. Based on this fact, educational institutions, inevitably, have to transform their existing frameworks to promote learning and teaching in the 21st Century. This paper introduces background and fundamentals about emerging technology paradigms – Cloud Computing (CC) and Internet of Things (IoT), and an emerging learning theory – Connectivism. The relationships between these three are investigated to provide insights and raise awareness on their rich potentials for and impacts on learning experiences.*

*Keywords: Cloud computing; connectivism; Internet of Things; education; learning*

# 1 Introduction

Dramatic developments in information technology, and more specifically, the emergence of mobile devices (e.g., tablets and smart phones) and applications have transformed the way people attain, use, and save the information. Additionally, the vast expansion of wireless networks provides new opportunities and infrastructures for individuals and institutions to be easily able to utilize new IT resources more effectively and cost efficient. The amazing speed in the expansion of Internet-based information has created a circumstance, in other words, a challenge for organizations to have enough capacities and capabilities for service-oriented computing. Including, virtual, dynamically-scalable, manageable

computing power, storage, platforms, and other services that can be delivered on demand to external customers over the Internet [15].

In recent years, "Cloud Computing (CC)" has emerged as a practical and assuring innovative solution to transforming traditional server infrastructure into a dynamic environment "for enabling convenient, on- demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [37, p590]. The concept of cloud computing has gained a significant amount of attention and has been defined with different perspectives.

Johnson, et al. [21, p36] defines cloud computing as:

> …expandable, on demand services and tools that are served to the user via the Internet from specialized data centers and consume almost no local processing or storage resources. Cloud computing resources support collaboration, file storage, virtualization, and access to computing cycles, and the number of available applications that rely on cloud technologies.

According to Sultan and Sultan [25, p3], cloud computing is:

> …a modality, that uses advances in ICTs such as virtualization and grid computing for delivering a range of ICT services through software, and virtual hardware (as opposed to physical) provisioned (by data centres owned and operated by cloud providers and/or end users) according to user demands and requirements and delivered remotely through public (e.g., Internet), private networks or a mix (i.e., hybrid) of the two delivery modes.

Using the main characteristics associated with the notion of cloud computing in the literature, Vaquero, et al. [23, p51] sums its definitions up as below:

> Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a payper-use model in which guarantees are offered by the Infrastructure Provider.

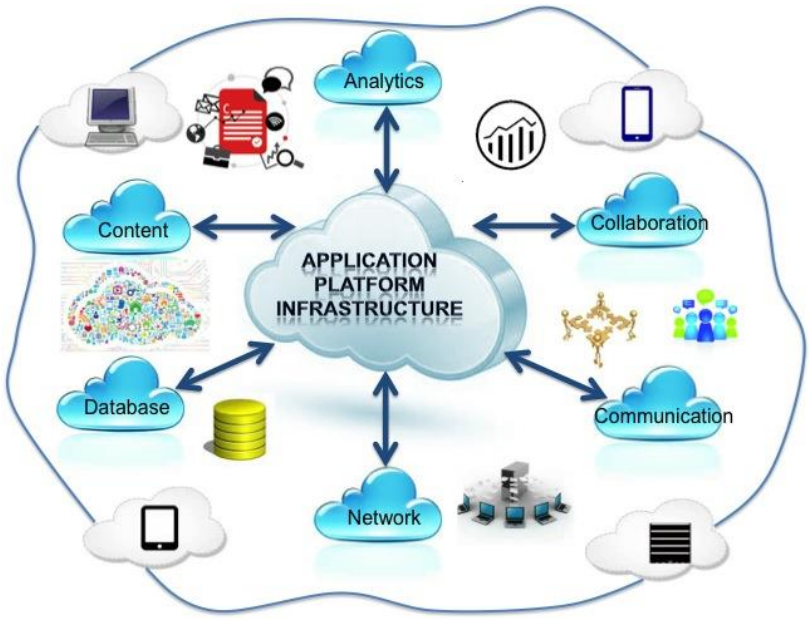



Figure 1
Cloud computing

The term "cloud" is used commonly as a metaphor of remote environments (e.g., Internet), which is illustrated with cloud-like images showing the complexity lying under full of servers providing different services through Internet connection. Understanding the cloud computing services could provide a better overview for what this new paradigm is all about. In general, CC services are categorized in three main types: Software as a Service (SaaS), Infrastructure as a Service (IaaS), and Platform as a Service (PaaS). Internet connection is the main requisite for three services. SaaS provides services (i.e., any application resides in the cloud) primarily for end-users who could access applications hosted in companies' servers. IaaS (also called utility computing) provides services (e.g., virtual machine instances, storage, and computation) generally for system administrators who could obtain general processing, storage, database management, and other applications through the network [30]. PaaS provides services (e.g., web hosting and data hosting services) mostly for developers who could design, build, and test applications within cloud infrastructure and then deliver those apps to end-users from the servers [30]. Table 1, below provides additional descriptions for three cloud services adopted from Sultan [24, p110].

Table 1
Cloud computing services

| User Levels | Service Type | Description |
|---|---|---|
| End-users (individuals, students, administrators, staff, etc.) | **Software as a Service (SaaS)** *Ex. Google Docs, YouTube, Facebook, twitter, flickr, and virtually every other Web 2.0 application* | Under this service, applications are delivered through the medium of the Internet as a service. Instead of installing and maintaining software, you simply access it via the Internet, freeing yourself from complex software and hardware management. This type of cloud service offers a complete application functionality that ranges from productivity (e.g., office-type) applications to programs such as those for Customer Relationship Management (CRM) or enterprise-resource management. |
| System administrators, researchers, end-users (individuals, students, administrators, staff, etc.) | **Infrastructure as a Service (IaaS)** *Ex. Data storage/ hosting and archiving/ preservation* | Products offered via this mode include the remote delivery (through the Internet) of a full computer infrastructure (e.g., virtual computers, servers, storage devices, etc.) |
| Developers | **Platform as a Service (PaaS)** *Ex. Amazon's S3 and EC2 offerings* | To understand this cloud-computing layer, one needs to remember the traditional computing model where each application managed locally required hardware, an operating system, a database, middleware, Web servers, and other software. One also needs to remember the team of network, database, and system management experts that are needed to keep everything up and running. With cloud computing, these services are now provided remotely by cloud providers. |

# 2 Cloud Computing in Education

With the spread of information and communication technology, a variety of new learning situations have been proposed for learners, individually or in groups, who have the opportunity for carrying out learning activities and experiences whenever and wherever they want by switching from one learning scenario to another easily and quickly [14]. Cloud computing is not just an innovative information and communication technology that promises to bring various exciting services into our daily life. It is already a reality and paradigm in terms of its implementations in many sectors. The education sector is one of those in which cloud computing has already attracted institutions, schools, educators, and researchers. Due to its rich capability in delivering virtual computing environments of almost any scale, cloud computing offers new tools that foster best computing practices for easy

sharing and mobility which improves productivity and expands collaboration in education. For instance, cloud computing provides engineering students with versatile and ubiquitous access to software commonly used in the field from any where, at any time [1].

The notion of providing computing-related services on the cloud (some free, some on a pay-as-you-go basis) offers numerous advantages and opportunities for data and software sharing, collaboration systems, and virtual learning environments. The cloud concept has imperative implications in virtual communication, collaboration, and learning contexts. This innovative communication and collaboration medium enables virtual communities of educators, researchers and practitioners to work in collaborative groups on the Internet with links to real time data, therefore, promoting a metacognitive approach in educational problems and practices via a dynamic research platform where data are dynamically stored, dynamically navigated, dynamically visualized, and so others [28]. From this perspective, CC has strong potential for fostering and facilitating social interactivity and metacognitive activities. Within a learning community, CC highly supports sharing of intellectual work, negotiation and co-creation of meanings or knowledge, research, reflection, evaluation, and application of newly constructed knowledge.

Stein, et al. [36, p241] addresses the educational benefits of Cloud computing for pedagogical aims in the following areas:

- *Diversity of software applications*: Cloud computing offers a cost-effective way of circumventing such restrictions as reaching diverse learning applications for school and student computers, and of providing richer curricula to all school systems.
- *Software license cost savings*: Schools can share software applications legally and cost-effectively using the Cloud by purchasing a limited number of "seat" licenses and managing the concurrent use of the software.
- *Security*: Security of data and transactions are concerns for every user of networked computation. Cloud computing can be configured to meet the Internet Protection Act compliance policies put in place by schools. Cloud-based instructional software can be accessed and used securely by making schools' firewalls impervious to hackers.
- *Extending machine life*: Significant cost savings can be realized through Cloud computing by decreasing the replacement rate for student computers. Student access to software is made more equitable across socioeconomic differences since older, less powerful computers can remotely operate advanced software.
- *Availability of broadband connectivity outside of school*: Learning environments based on CC can specifically be designed to provide access to learning tools from anywhere at anytime.

- *Affordability of class time delivery*: Face-to-face education employs strict time periods during which software must be available with minimal delay. Using CC platform, designed specifically for educational needs, block allocations (i.e., access to specified software and student registration for specific days and times) allow teachers to register their students and make pre-loaded software remotely available in less than a minute.

Cloud computing provides opportunities for teachers with quickly applicable educational applications and services that enrich the learning process based on individual performance data and learning style [1]. Teachers and students today are inundated with many low-cost or free cloud-computing applications offering SaaS that is accessible through diverse digital devices. Two of the most prevalent and free CC applications being adopted by educators are *Google Apps for Education* (including Gmail, Gdrive, Google Docs) and *Microsoft's cloud service – Office 365 for Education* (including Office, OneDrive, OneNote). Both of these applications provide educational institutions, teachers and students with platforms, which support dynamic learning experiences in and outside of the classroom, provide access to centralized information stores, and increase reporting consistency [1].

For instance, Lin, et al. [38], in their research study, proposed a cloud-based reflective learning environment, which was designed by using Google applications (Google Plus, Google Drive, and Google Site) as a set of teaching and learning tools. They conducted an experiment (including tests, questionnaires, and interviews) to evaluate the effectiveness of the proposed environment on the students' ability to develop and strengthen their reflection and motivation. According to the findings, students in the experimental group (i.e., those in the proposed cloud-based learning environment) showed higher learning motivation and reflection performance. The authors provided results that verified the cloud-based reflective learning environment supports instructors and students who effectively administered and conducted reflective learning activities during and after actual class sessions.

Sometimes, institutions choose a way of creating or joining a consortium where they define their position as a leader of innovation and progress in a certain field. The cloud-based consortiums particularly allow institutions to work collaboratively with other organizations and experts that could increase the availability of resources and speed of research activities. To give an example, the Open Cloud Consortium (OCC) (http://opencloudconsortium.org/) is "an open cloud computing infrastructure that facilitates community based science, in which researchers from member institutions, including the University of Chicago and Johns Hopkins University among others, can compile, analyze, and share huge data sets via the Open Science Data Cloud" [22, p10].

# 3 Connectivism

From an educational perspective, cloud-computing technology is commonly considered and practiced within the concept of virtual or online learning. CC is widely accepted to be suitable for online learning environment with its capabilities in facilitating information managing, data sharing, and collaborating, and so on. Considering today's students who are under the influence of Internet (those called *Net Generation*) and digital media and technologies (those called *Digital Natives*) have got unique habits, actions, learning styles, preferences, communication patterns within the ecology of networked content and communities. These students possess the skills of self-directing their learning progress, obtaining information from a variety of online resources, doing multitask quickly - sometimes simultaneously, and working in teams.

Therefore, along with the advancements in network technologies, educators have started to employ many different educational philosophies, strategies, and approaches for emerging needs and demands of students in the online milieu. Yet, there was a need to develop new pedagogical approaches, which take into consideration and explain new learning experiences specifically in a networked learning setting. The connectivism, introduced and advocated by George Siemens and Stephen Downes, was called "A Learning Theory for the Digital Age" [9]. George Siemens [9] defines connectivism as

> …the integration of principles explored by chaos, network, and complexity and self-organization theories. Learning is a process that occurs within nebulous environments of shifting core elements – not entirely under the control of the individual. Learning (defined as actionable knowledge) can reside outside of ourselves (within an organization or a database), is focused on connecting specialized information sets, and the connections that enable us to learn more are more important than our current state of knowing.

George Siemens [11] explores some of the main categories and aspects of learning that are experienced today. According to Siemens, learning can be categorized as formal or structured learning; experience, games and simulation learning; mentoring and apprenticing; performance support; self-learning; communities of practice or learning communities; and informal learning. These seven learning categories are all interconnected with each other, which are driven by dimensions of learning (e.g., learning what, why, where), learning concepts (data, knowledge, meaning, understanding), conduits (e.g., language, media, technology), and filters for learning (e.g., values, beliefs, perspective). Change and transformation in knowledge with learner's intention to learn takes place in connected network of learning domains (see Figure 2).

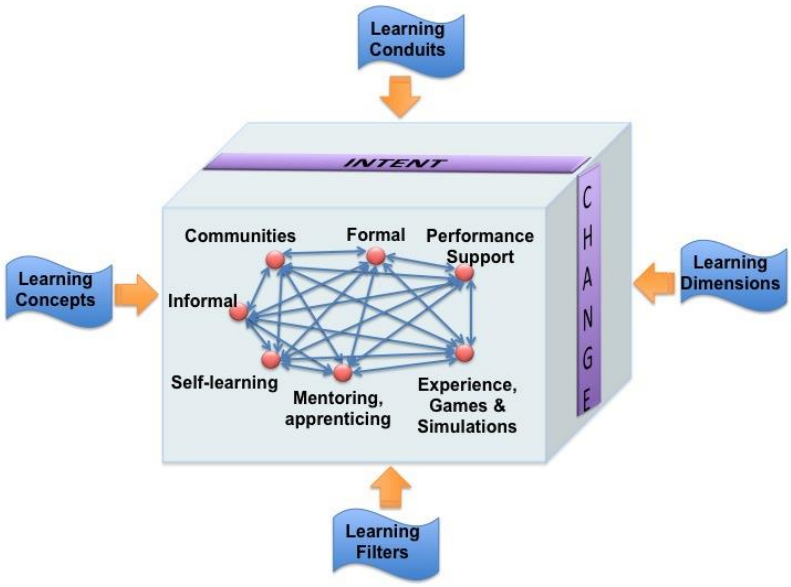



Figure 2
Learning Network

Connectivism is, simply, a learning theory grounded in the interactions within networks [34]. Stephen Downes [32] described it as, "… the thesis that knowledge is distributed across a network of connections, and therefore that learning consists of the ability to construct and traverse those networks." Siemens [9] identifies the fundamental principles of connectivism:

- Learning and knowledge rests in diversity of opinions
- Learning is a process of connecting specialized nodes or information sources
- Learning may reside in non-human appliances
- Learning is more critical than knowing
- Maintaining and nurturing connections is needed to facilitate continual learning
- Perceiving connections between fields, ideas and concepts is a core skill
- Currency (accurate, up-to-date knowledge) is the intent of learning activities
- Decision-making is itself a learning process. Choosing what to learn and the meaning of incoming information is seen through the lens of a shifting reality. While there is a right answer now, it may be wrong tomorrow due to alterations in the information climate affecting the decision.

Downes [33] states that according to the connectivism: i) learning occurs as a distributed process in a network, based on recognizing and interpreting patterns; ii) the learning process is influenced by the diversity of the network, strength of the ties; iii) memory consists of adaptive patterns of connectivity representative of current state; iv) transfer occurs through a process of connecting; and v) best for complex learning, learning in rapidly changing domains. Knowledge is viewed in connectivism as a pattern of connectedness [13] (see Figure 3). Connectivism claims "knowledge - and therefore the learning of knowledge - is distributive that is not located in any given place (and therefore not 'transferred' or 'transacted' per se). It rather consists of the network of connections formed from experience and interactions with a knowing community" [31].

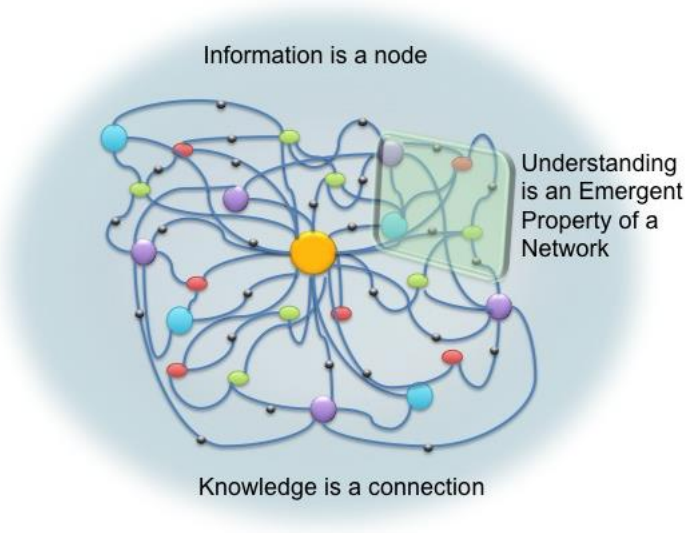



Figure 3
Knowledge in connectivism theory

Siemens [12] emphasized that, "Exponentially developing knowledge and complexification of society requires nonlinear models of learning (process) and knowing (state)." In brief, connectivism is described as, "the amplification of learning, knowledge, and understanding through the extension of a personal network" [8].

In a networked, technology-enriched arena, new instructional models and practices based on connectivism learning theory attracted a great interest of educators as well as researchers for promoting students' problem-solving skills [35], motivation in learning [7], and information literacy self-efficacy in a cloud-based virtual classroom [19]. The MOOC (which stands for massively open online courses), perhaps, best represents learning and teaching practices based on the

concept of connectivism. The essence of the MOOC concept is that thousands or millions of participants from all over the world at any time could take a course in a networked environment (e.g., web course). One of the unique features of MOOCs is that a course (open and free of charge) is designed and/or contributed by a variety of experts, educators, and instructors supporting an expansive and a diverse set of contents for which the materials are not necessarily designed to go together, but become associated with each other [20]. MOOCs are one of the most prominent trends in online education since they support active and participative involvement within a learning community, where the contributions are shared and educational resources are distributed through an online platform to high volume learners. Cloud-based services (e.g., WikiSpaces, YouTube, and Google Hangouts, etc.) and the three major tools - Coursera, edX, and Udacity - are predominantly being used for offering MOOCs, for instance, to create discussion forums, share videos, assess learner performances and engage in all other learning and teaching activities in online ecosystem.

# 4 Internet of Things (IoT)

## 4.1 Definition and Scope of the IoT

The notion of connectivism, as stated above, supports the idea of learning that rests in the network of connected and relational patterns of actions and information. Over recent past years, the concept of connectedness has gained a growing interest in its place not only between human and human; or human and objects but also between/amongst objects – omnipresent connection of "objects" that intelligently connect and communicate internally or with others – with the ability of embedded smart devices, sensors, and networks. The term, Internet of Things (IoT) has been introduced to provide a better explanation about a network of connected objects that turn the information into actions in the physical world (e.g., mobile phones, smart houses, smart sensors, and Internet-enabled appliances, etc.).

> Embedded chips, sensors, or tiny processors attached to an object allow helpful information about the object, such as cost, age, temperature, color, pressure, or humidity to be transmitted over the Internet. This simple connection allows remote management, status monitoring, tracking, and alerts… Many web tools allow objects to be annotated with descriptions, photographs, and connections to other objects, and other contextual information; the Internet of Things makes access to these data as easy as it is to use the web [21, p. 42].

The Internet of Things (IoT), also called the Internet of Everything, is considered the next evolution of Internet (see Figure 4); a new technology paradigm; and a next step in Information Industry. The evolution of Internet begins with a limited network (e.g., two-three computers connected), then moves towards to World Wide Web, then mobile devices with Internet connection appears, then social networks including people's identities and various information takes place, and finally the Internet of Things connecting every day objects to the Internet comes into our life [4].

IoT is an interdisciplinary field of which many alternative definitions have been introduced. The Internet of Things is defined as

> A global network infrastructure, linking physical and virtual objects through the exploitation of data capture and communication capabilities. This infrastructure includes existing and evolving Internet and network developments. It will offer specific object-identification, sensor and connection capability as the basis for the development of independent cooperative services and applications. These will be characterised by a high degree of autonomous data capture, event transfer, network connectivity and interoperability [5, p10].

According to this definition, three categories of hardware technology are suggested, to realize IoT: i:) identification and data capture technologies forming the physical interface layer, ii) fixed, mobile, wireless and wired communication technologies, with associated interface support, for data and voice communications, and iii) network technologies (in combination with communication technologies) to facilitate grouping of supported Objects for application and service purposes [5, p16]. In addition to this, software, middleware components and associated protocols are necessary for linking objects and driving the hardware and service to establish a fully operational system or systems [5].

Another definition by Vermesan and Friess [26, pp7, 8] states that "Internet of Things (IoT) is a concept and a paradigm that considers pervasive presence in the environment of a variety of things/objects that through wireless and wired connections and unique addressing schemes are able to interact with each other and cooperate with other things/objects to create new applications/services and reach common goals.''

Antoine de Saint-Exupery [2, p8] provides a definition, which covers a wide-ranging vision of IoT as "The Internet of Things allows people and things to be connected Anytime, Anyplace, with Anything and Anyone, ideally using Any Path/Network and Any Service. This implies addressing elements such as Convergence, Content, Collections (Repositories), Computing, Communication, and Connectivity in the context where there is seamless interconnection between people and things and/or between things and things so the A and C elements are present and addressed."

## 4.2 Impact of Internet of Things

Internet of Things has been acknowledged as one of the emerging technologies in Information Technology. Through various enabling technologies such as nanoelectronics, sensors/actuators, cloud computing, IoT includes systems which bridge the gap between cyberspace and the physical world of real things [27]. As illustrated in Figure 5 below, IoT becomes a system consisting of different systems.

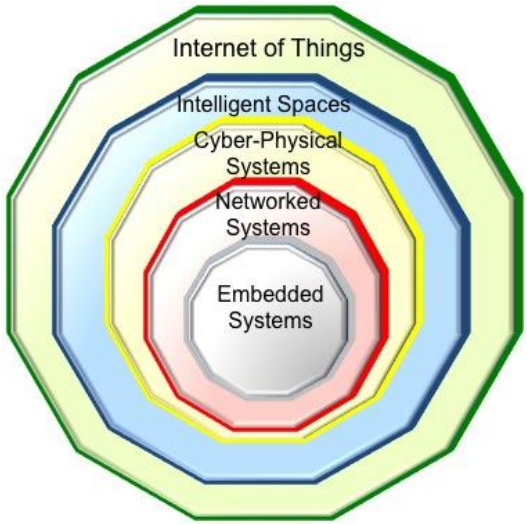



Figure 4
The relations of systems within IoT

With increasing levels of connectivity through IoT, smart life including smart cities, smart business, smart houses and so others will take place in our daily lives. It is predicted that there will be a connected society with 50 billion connected smart devices and approximately 7 connected devices per person around the world by 2020 [6]. This, therefore, shows us that new IoT products and services will grow exponentially in the following years. Due to this fact, major developed countries in the world consider and make investments on the IoT as an area of innovation and growth.

European Union is one of those where the number of IoT connections will increase from approximately 1.8 billion in 2013 (the base year) to almost 6 billion in 2020 [29]. Hence, according to the analysts, IoT revenues in the EU will be expected to increase from more than €307 billion in 2013 to more than €1,181 billion in 2020 [29]. Based on the research and predictive studies, in order to create a dynamic European IoT ecosystem, The Alliance for Internet of Things Innovation (AIOTI) (www.aioti.eu) was launched in 2015 by the European

Commission. The AIOTI's goals are to provide and support interaction for future IoT research as well as innovation and standardization of policies among IoT stakeholders.

Within the market of the EU, the IoT is developing rapidly, generating multiple new services, and moving towards a more user-oriented perspective combining with Cloud Computing and Big Data. Aguzzi, et al. [29] provides a recommendation for EU's H2020 Program about prioritizing and increasing the investment(s) in the IoT and Cloud Computing combination for the development of technologies, research and innovation areas to take-up actions and accelerate the market and relevant ecosystem development.

Actually, there is a broad-spectrum expectation in many sectors in the world that new opportunities will get into our life from the combination of IoT and Cloud computing. Cloud computing is anticipated to be a vital part of the IoT. Applications of IoT will be almost anywhere producing an enormous amount of data, which have to be stored, accessed, processed, and transferred. Intelligent objects with sensors (of which processors become smaller, cheaper and more powerful), a truly revolutionary potential of IoT, will need a platform which can handle the speed and flow of the data accessible anywhere from any device. This is where cloud-computing platforms play their role of being an area of immense data storage and analytics. In addition, the interdependency of IoT and CC provides an effective opportunity especially for IoT developers who can use the Cloud as a development environment for IoT applications primarily by connecting devices through the Cloud [18]. According to IEEE [16],

> The combination of cloud computing and IoT can enable ubiquitous sensing services and powerful processing of sensing data streams beyond the capability of individual "things", thus stimulating innovations in both fields. For example, cloud platforms allow the sensing data to be stored and used intelligently for smart monitoring and actuation with the smart devices. Novel data fusion algorithms, machine learning methods, and artificial intelligence techniques can be implemented and run centralized or distributed on the cloud to achieve automated decision-making.

## 4.3 IoT in Education

Internet-enabled appliances are coming into focus and utilization, more and more, every single day in the consumer industry. Finally, like many other emerging technologies transferred from industry sector and integrated into the world of education, IoT will find a place in various learning environments. It is forecasted that the IoT market in education and health will increase from € 22 060 in 2014 to € 66 925 in 2020 (€ Million) in the European Union context [29]. Thus, it is strongly recommended to collaborate education and training actors in the field to

develop appropriate curricula and main skills for adoption and integration of IoT into pedagogical situations [29].

Although concrete implementations of IoT, today, are difficult to find, its potentials in fostering learning attract educational institutions and educators. For instance, collaboration and engagement between teachers and students; and sharing and accessing digital content are easily encouraged and facilitated with the structure and features of IoT. "For instruction, IoT in higher education takes the form of blended learning models that integrate personalized materials and formative assessment technologies that deliver instant feedback. In this landscape, students will have the ability to monitor their own environment and collect real-time data for further study" [22, p47].

With the increase in connection speed to Internet, and reduction in equipment and software costs, IoT could offer better learning resources widely accessible, and real-time learning experiences that are immersive and interactive than many existing lectures and classes do so. For instance,

> Archaeologists from the University of Bristol are embedding historical objects from the transatlantic slave trade for "Reflector," a project that aims to share stories through authentic pieces of history that would otherwise not be available to the masses. Every artifact has a story and presents an opportunity for learning about history and culture, and the Internet of Things is making it easier — and more automatic — to communicate them" [21, p43].

An empirical research study by Uzelac, et al. [3] was conducted to examine the impact of different parameters in the physical learning environment on students' focus based on the IoT concept. The researchers claimed that this was the first study in the literature providing insights in terms of the correlation between lecturer's voice features and students' focus on the lecture. 5 different parameters, namely, CO2 (CO2 sensor used), Temperature (Temperature sensor used), Air pressure (Air pressure sensor used), Humidity (Humidity sensor), Noise level (Noise level sensor on smartphone used), Lecturer's voice (Bluetooth headset connected to a netbook used) were measured in real classroom setting. The findings indicated that five uncorrelated parameters had significant impact on students' focus. The results also suggested that this kind of study presents a practical strategy for recognizing the students' focus during the lecture.

## Conclusions

With today's ubiquitous computing technologies, an enormous amount of information has become available to an individual that can be stored, retrieved, accessed, and shared with others through networks at the point of need any time anywhere, even during mobility. Cloud computing, indeed, has entered in our life as an innovative technology that evokes a new kind of social phenomenon at different levels from posting a photo to Facebook or using the mobile Google

Maps for driving directions to using software from any device and developing apps, products and activities. In addition, cloud computing enables a new kind of learning and teaching experiences. CC can offer a powerful platform and new computing tools for "a new instructional paradigm that makes a shift possible from the traditional practice of teaching as a private affair to a peer-reviewed transparent process, and makes it known how student learning can be improved generally, not only in one's own classroom but even beyond it" [28, p222].

Therefore, students, today, have the opportunity of accessing the knowledge by sharing distributed E-learning resources (OER) in cloud systems [1]. Although there is a growing interest in cloud computing for educational use, administrators, teachers and instructional designers must be careful to discern what truly improves and/or facilitates the learning and teaching process and what does not, with regard to CC. This technology creates teaching opportunities, but on the other hand, requires new instructional techniques and strategies that are effective for and go compatible with new networked learning activities.

Connectivism, a new learning theory proposed by Siemens (2005), can be considered an alternative and effective theory to support the learning and teaching through cloud computing technology. Since connectivism relies on sharing, making connections between people and ideas, it strongly encourages teachers to utilize CC tools. The principal pedagogical implication of connectivism is the development of knowledge through rich connections in a networked environment [10]. "Learning becomes the critical recognition of connections that change the network itself, simultaneously adding new connections, potentially in the absence of an instructor or authority" [17, p688].

The pace of change in our daily activities, in particular, communication, collaboration, and interaction with individuals in the society and objects around us is so rapid. The penetration of computing systems into society at every level is changing the way we see our relations, our life style and conditions, our work habits, and our interactions with individuals and objects. Perhaps, we are unaware of the fact that our lifestyle has already been replaced with another. We will experience this change even more with the adoption of Internet of Things, a recent phenomenon, which makes objects smarter, in that, they have the ability of sensing, actuation, communication with each other, collection of information from vast amounts of data and creation of new knowledge in order for people to take action(s) when needed.

This paper provides insights about emerging technology paradigms – *Cloud Computing* and *Internet of Things* – and an emerging learning theory – *Connectivism*, which promise to provide innovative and alternative solutions in digital age. It is believed that the impact of those on many sectors will be experienced and visible and tangible in the near future. Education is, perhaps, one of the critical sectors among those because digital-born students are already actively using and becoming easily familiar with these emerging technologies in

their social life. These are still early days, and educational institutions are at the beginning of a transition period facing many challenges in the adoption of technological, theoretical, and practical paradigms. It is strongly suggested that educational institutions should develop a comprehensive strategy including curricula, professional development of teachers, educational philosophy, data security, legal and political issues, and transformation of resources and infrastructure to be able to address the many unique challenges that lie ahead.

**References**

[1] A. Bouyer and B. Arasteh, "The Necessity of Using Cloud Computing In Educational System," *Procedia - Social and Behavioral Sciences,* Vol. 143, pp. 581-585, 2014

[2] A. De Saint-Exupery, "Internet of Things: Strategic Research Roadmap. The Cluster of European Research Projects," Tech. Rep., September, 2009 [Online] Available: http://www.internet-of-things-research.eu/pdf/IoT_Cluster_Strategic_Research_Agenda_2009.pdf [Accessed: Jun. 04, 2015]

[3] A. Uzelac, N. Gligoric, S. Krco, "A Comprehensive Study of Parameters in Physical Environment that Impact Students' Focus during Lecture using Internet of Things," *Computers in Human Behavior*, Vol. 53, pp. 427-434, 2015

[4] C. Perera, A. Zaslavsky, P. Christen, D. Georgakopoulos, "Context Aware Computing for The Internet of Things: A Survey," *IEEE Communications Surveys & Tutorials*, Vol. 16, No. 1, pp. 414-454, 2014 [Online] Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6512846 [Accessed: Jun. 07, 2015]

[5] Casagras, "RFID and the Inclusive Model for the Internet of Things," European Union, Brussels, Belgium, Rep. 216803, pp. 16-23, 2011

[6] D. Evans, "The Internet of Things-How the Next Evolution of the Internet is Changing Everything," Cisco Internet Business Solutions Group White Paper, 2011 [Online] Available: http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf [Accessed: Jun. 04, 2015]

[7] E. Trnova and J. Trna, J. "Motivational Effect of Communication Technologies in Connectivist Science Education," *Online Journal of Communication and Media Technologies*, Vol. 5, No. 3, pp. 107-119, 2015

[8] G. Siemens (2004) "*Connectivism: A Learning Theory for the Digital Age*". [Online] Available: http://www.elearnspace.org/Articles/connectivism.htm [Accessed: Jun. 07, 2015]

[9] G. Siemens, "Connectivism: A Learning Theory for the Digital Age," *International Journal of Instructional Technology and Distance Learning,* Vol. 2, No. 1, 2005 [Online] Available:

http://www.itdl.org/Journal/Jan_05/article01.htm [Accessed: May. 07, 2015]

[10] G. Siemens (2006, Nov. 12) "*Connectivism: Learning Theory or Pastime for the Self-Amused*" Elearnspace [Online] Available: http://www.elearnspace.org/Articles/connectivism_self-amused.htm [Accessed: May. 07, 2015]

[11] G. Siemens (2007). "*Professional and Personal Development*" [Online] Available: http://www.elearnspace.org/media/ProfessionalDevelopment/player.html [Accessed: May. 07, 2015]

[12] G. Siemens (2009) "*elearnspace*" [Online] Available: http://www.elearnspace.org/blog/. [Accessed: May. 07, 2015]

[13] G. Siemens (2011) "*Re: Explanation Needed - Definition of Knowledge and Knowledge Models*" [Online] Available: http://scope.bccampus.ca/mod/forum/discuss.php?d=16450 [Accessed: May. 07, 2015]

[14] G. Zurita, N. Baloian, J. Frez, "Using the Cloud to Develop Applications Supporting Geo-Collaborative Situated Learning," *Future Generation Computer Systems*, Vol. 34, pp. 124-137, 2014

[15] I. Foster, Z. Yong, I. Raicu, S. Lu, "Cloud Computing and Grid Computing 360-Degree Compared," in Grid Computing Environments Workshop, Nov 2008, pp. 1-10

[16] IEEE. "*Call for papers IEEE Internet of Things Journal Special Issue on Cloud Computing for IoT,*" IEEE Internet of things journal [Online] Available: http://iot-journal.weebly.com/uploads/1/8/8/0/18809834/ieee_iot_journal_si_cloud_iot_cfp_final.pdf [Accessed: Jun. 05, 2015]

[17] J. Barnett, V. McPherson, R. M. Sandieson, "Connected Teaching and Learning: The Uses and Implications of Connectivism in an Online Class," *Australasian Journal of Educational Technology*, Vol. 29, No. 5, pp. 685-698, 2013

[18] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, "Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions," *Future Generation Computer Systems*, Vol. 29, pp. 1645-1660, 2013

[19] K. Kultawanicha, P. Koraneekija, J. Na-Songkhla, "A Proposed Model of Connectivism Learning Using Cloud-based Virtual Classroom to Enhance Information Literacy and Information Literacy Self-Efficacy for Undergraduate Students," *Procedia - Social and Behavioral Sciences*, Vol. 191, pp. 87-92, 2015

[20] L. Johnson, S. Adams Becker, M. Cummins, V. Estrada, A. Freeman, H. Ludgate, "NMC Horizon Report: 2013 Higher Education Edition," The New Media Consortium, Austin, Texas, 2013 [Online] Available:http://www.nmc.org/pdf/2013-horizon-report-HE.pdf [Accessed: Jun. 03, 2015]

[21] L. Johnson, S. Adams Becker, V. Estrada, A. Freeman, "NMC Horizon Report: 2014 K-12 Edition," The New Media Consortium, Austin, Texas, 2014 [Online] Available: http://cdn.nmc.org/media/2014-nmc-horizon-report-k12-EN.pdf [Accessed: Jun. 03, 2015]

[22] L. Johnson, S. Adams Becker, V. Estrada, A. Freeman, "NMC Horizon Report: 2015 Higher Education Edition," The New Media Consortium, Austin, Texas, 2015 [Online] Available: http://cdn.nmc.org/media/2015-nmc-horizon-report-HE-EN.pdf [Accessed: Jun. 07, 2015]

[23] L. M. Vaquero, L. M. Lindner, L. Rodero-Merino, J. Caceres, "A Break in the Clouds: towards a Cloud Definition," *Computer Communication Review*, Vol. 39, No. 1, pp. 50-55, 2009

[24] N. Sultan, "Cloud Computing for Education: A New Dawn?" *International Journal of Information Management*, Vol. 30, No. 2, pp. 109-116, 2010

[25] N. Sultan and Z. Sultan, "The Application of Utility ICT in Healthcare Management and Life Science Research: A New Market for a Disruptive Innovation?" In The European Academy of Management Conference, Rotterdam, The Netherlands, June 6-8, 2012 [Online] Available: http://www.researchgate.net/publication/277310972_The_application_of_utility_ICT_in_healthcare_management_and_life_science_research_a_new_market_for_a_disruptive_innovation [Accessed: May. 09, 2015]

[26] O. Vermesan and P. Friess, *Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems*. Aalborg, Denmark: River Publishers, 2013, pp. 7-151

[27] O. Vermesan, M. D. Nava, H. Grindvoll, "Driving Innovation through the Internet of Things - Distruptive Technology Trends," in *Building the Hyperconnected Society: IoT Research and Innovation Value Chains, Ecosystems and Markets*, O. Vermesan and P. Friess, Eds. Aalborg, Denmark: River Publishers, 2015, pp. 279-308

[28] P. Y. Thomas, "Cloud Computing," *The Electronic Library*, Vol. 29, No. 2, pp. 214-224, 2011

[29] S. Aguzzi, D. Bradshaw, M. Canning, M. Cansfield, P. Carter, G. Cattaneo, S. Gusmeroli, G. Micheletti, D. Rotondi, R. Stevens, "Definition of a Research and Innovation Policy Leveraging Cloud Computing and IoT Combination," European Commission, Directorate-General of Communications Networks, Content & Technology, Brussels, Belgium, Rep. SMART number 2013/0037, 2014 [Online] Available:

http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=9472 [Accessed: Jul. 04, 2015]

[30] S. C. Park and S. Y. Ryoo, "An Empirical Investigation of End-Users' Switching toward Cloud Computing: A Two Factor Theory Perspective," *Computers in Human Behavior,* Vol. 29, pp. 160-170, 2013

[31] S. Downes (2006) "*Learning Networks and Connective Knowledge"* [Online] Available: http://itforum.coe.uga.edu/paper92/paper92.html [Accessed: May. 13, 2015]

[32] S. Downes (2007) "*What Connectivism is*" [Online] Available: http://halfanhour.blogspot.com/2007/02/what-connectivism-is.html [Accessed: May. 21, 2015]

[33] S. Downes (2010) "*Connectivism and its Critics: What Connectivism is not*" [Online] Available: http://www.downes.ca/post/53657 [Accessed: May. 21, 2015]

[34] S. Downes, *Connectivism and Connective Knowledge: Essays on Meaning and Learning Networks*. National Research Council (Canada) 2012 [E-book] Available: http://www.downes.ca/files/books/Connective_Knowledge-19May2012.pdf [Accessed: May. 16, 2015]

[35] S. Sitti, S. Sopeerak, N. Sompong, N, "Development of Instructional Model Based on Connectivism Learning Theory to Enhance Problem-Solving Skill in ICT for Daily Life of Higher Education Students," *Procedia - Social and Behavioral Sciences*, Vol. 103, pp. 315-322, 2013

[36] S. Stein, J. Ware, J. Laboy, H. E. Schaffer, "Improving K-12 Pedagogy via a Cloud Designed for Education," *International Journal of Information Management*, Vol. 33, pp. 235-241, 2013

[37] V. H. Pardeshi, "Cloud Computing for Higher Education Institutes: Architecture, Strategy and Recommendations for Effective Adaptation," *Procedia Economics and Finance*, Vol. 11, pp. 589-599, 2014

[38] Y-T. Lin, M-L. Wen, M. Jou, D-W. Wu, "A Cloud-based Learning Environment for Developing Student Reflection Abilities," *Computers in Human Behavior,* Vol. 32, pp. 244-252, 2014

# Towards the Coevolution of Incentives in BitTorrent

**Tamás Vinkó and David Hales**

University of Szeged, Department of Computational Optimization
Árpád tér 2, H-6720 Szeged, Hungary
tvinko@inf.u-szeged.hu, daphal@inf.u-szeged.hu

*Abstract: BitTorrent is a peer-to-peer file sharing system that is open to variant behavior at the peer level through modification of the client software. A number of different variants have been released and proposed. Some are successful and become widely used whereas others remain in a small minority or are not used at all. In previous work we explored the performance of a large set of client variants over a number of dimensions by applying Axelrod's round-robin pairwise tournament approach. However, this approach does not capture the dynamics of client change over time within pairwise tournaments. In this work we extend the tournament approach to include a limited evolutionary step, within the pairwise tournaments, in which peers copy their opponents strategy (client variant) if it outperforms their own and also spontaneously change to the opponents strategy with a low mutation probability. We apply a number of different evolutionary algorithms and compare them with the previous non-evolutionary tournament results. We find that in most cases cooperative (sharing) strategies outperformed free riding strategies. These results are comparable to those previously obtained using the round-robin approach without evolution. We selected this limited form of evolution as a step towards understanding the full coevolutionary dynamics that would result from evolution between a large space of client variants in a shared population rather than just pairs of variants. We conclude with a discussion on how such future work might proceed.*

*Keywords: peer-to-peer; BitTorrent; selection rules; design space analysis*

## 1 Introduction

Many popular content sharing systems are based on peer-to-peer (P2P) technology. In P2P systems the participating users, also called peers, are the content providers and demanders at the same time. Among these systems, BitTorrent [1] is the best representative example, used by millions of Internet users every day. In a BitTorrent system, peers can use different clients, i.e. a

computer program which runs the BitTorrent protocol, or a modified version of it. Modification usually means some extension upon the original protocol in order to provide the user with better quality of service. Apart from the protocol modifications actually deployed as clients –for example uTorrent, VuZe, Tribler, etc.–, many interesting variants were proposed in the scientific literature of BitTorrent systems, see, e.g. [2, 3, 4, 5]. Most noticeably, the paper of Rahman *et al.* [6] gives a list of more than 3000 protocol variants. However, the aim of the paper was to lay down the methodology of Design Space Analysis (DSA) of distributed protocols for measuring their performance, robustness and aggressiveness.

In this paper we give an extension of the DSA concept by studying the effects of applying evolutionary approaches to the peers.

## 1.1 Motivation

BitTorrent and open peer-to-peer protocols in general require the cooperative interaction of individual peers if they are to function optimally. This is because performance is collectively produced yet actions are individually selected. Since each individual peer may run a different client variant (so long as it implements the conventions of the specified protocol) there is the possibility for strategic interaction and consequent collective action problems. In the context of BitTorrent this involves free riding (downloading data, but not uploading data) rather than sharing data.

However, this problem is quite general in any open distributed system where the client software that runs on each peer is not under the control of a central authority or designer.

Consequently, designers of open peer-to-peer protocols (such as BitTorrent) include incentive mechanisms to encourage cooperation. In its simplest form this involves client software which punishes other peers who do not operate in a cooperative way. If a sufficient number of peers execute such a client, then free riding can be controlled since it is not then in the individual interests of a single peer to change to free riding behavior.

This general problem of cooperative collective action has been addressed through game theory and computer simulations. Axelrod [7] used the pairwise Iterated Prisoner's Dilemma game and computer programs implementing game strategies to perform a "round-robin tournament" (RRT) to examine which individual strategies did well on average against all other strategies.

For a given P2P application it is possible to view client variants as strategies in a complex multiplayer game. In general, such games are too complex to be tractable within analytical game theoretic frameworks without gross simplifications and

assumptions that do not hold in the real world. Hence, computer simulation is often used.

In previous work we applied a form of Axelrod's RRT (which we called DSA) to the BitTorrent protocol identifying a number of interesting, counterintuitive and high performance variants. This approach involved taking every pair of possible client variants (or strategies) and performing simulation runs in which the population was partitioned into two fixed size subpopulations running the two different variants. Both variants then mixed freely during interactions which involved downloading and uploading file pieces following the BitTorrent wire protocol. Results for each strategy were calculated based on average performance of that strategy against all other strategies by aggregating all of the relevant simulation runs and calculating several statistics.

We modify the RRT such that rather than only examining the performance of one strategy (or client variant) against another in *fixed* subpopulations we allow clients to apply evolution such that they can copy and mutate variants from others who outperform them. Hence, this approach allows for relative sizes of the two subpopulations to change during interaction over time within a tournament.

This extension is motivated by several issues: firstly, are the original DSA results reproduced when peers are given the ability to evolve directly? Secondly, do those strategies previously identified as robust against other strategies evidence evolutionary robustness; and thirdly, how do different evolutionary algorithms effect the outcomes?

Hence, in this work we are extending the previous DSA approach by applying evolution *within* each tournament between only pairs of strategies. We are not applying general evolution to the entire space of strategies in one large population.

This allows us to test if the original DSA results are consistent with evolution applied to pairs of strategies and to examine in more detail the robustness and performance of client variants. We consider this *limited* evolutionary approach as a step towards a full coevolutionary analysis over a large strategy space while maintaining the ability to produce meaningful insights.

This approach applies a limited form of evolution between only pairs of solutions (client variants) because it extends the round-robin tournament approach. Hence, it is not possible to apply operators such as crossover since these would introduce more than two variants into the tournament. Also, for the same reason mutation is implemented as swapping to the other client variant rather than producing a new client variant. It would be of interest to compare the results from a full co-evolutionary analysis in which all variants compete and evolve within a single population. However, based on previous results, as discussed in section 1.2, we expect such an approach to produce highly variable and difficult to analyse results. This is because coevolutionary dynamics over complex strategies tend to lead to cyclical and highly contingent dynamics such that, in the worst case, outcomes

may appear indistinguishable from noise. However, we discuss ways forward in this regard in the conclusion.

### 1.2 Related Work

Several previous works have studied the application of incentive mechanisms to regulate open distributed systems. This area of research has been termed distributed computational mechanism design [8]. Approaches essentially fall into two broad categories. Firstly, those that apply analytic game theoretic formulations [9, 10, 11] and secondly, those that utilize simulation approaches [12, 13, 14, 15]. In general, analytic approaches require high levels of abstraction that limit their applicability to the design of realistic deployable protocols. Alternatively, simulation can be applied to highly realistic scenarios, but often lacks detailed sensitively analysis due to the large parameter spaces of real protocols. Previous work attempted to balance these two aspects via the use of simulation over a large yet tractable space of parameters related to realistic protocol designs [6]. We build on this latter work in the present article. Jin *et al.* [16] presented an evolutionary analysis of BitTorrent P2P protocol variants by simulating the coevolution of six existing deployed variants within a single population. Our work differs in that we investigate on the coevolution between two variants at a time in the population. However, by limiting evolution in this way, to within tournament pairs, we are able to examine all coevolutionary pairings between over 500 individual protocols. Hence, through this abstraction we aim to move towards an understanding of the coevolutionary dynamics of more extensive forms of evolution. It is notable in the work of Jin *et al.* that the coevolutionary dynamics of populations composed of six protocol variants are difficult to analyze due to what appear to be highly contingent outcomes.

## 2 Background terminology

### 2.1 Design Space Analysis

Inspired by the seminal work of Axelrod [7], a method called Design Space Analysis (DSA) was proposed by Rahman *et al.* [6] to comprehensively model incentives in distributed protocols. This method models interactions between participating users playing repeated games. It combines the specification of a large design space of protocol variants together with their analysis done by simulations. The specification has two steps: (1) parametrization, which is the determination of the design space's dimensions, and (2) actualization, in which the actual values of the individual dimensions get specified. After these two steps, a solution concept can be used, where every element of the design space gets characterized by

different measures. DSA proposed three measures: Performance, Robustness and Aggressiveness – also called PRA quantification. Given a utility function, the Performance of a protocol is the average performance of the whole system under the assumption that all peers use the same protocol variant. The utility function is always domain specific. In a content distribution system the utility function is usually the average download speed of the users, but other measures could be used. The Robustness and Aggressiveness measures are defined in a system composed by different protocol variants and they indicate the ability of outperforming other protocol variants. By these three measures, the properties of all the protocols can be characterized as a three dimensional point.

Robustness indicates the ability of a given protocol variant to outperform its opponent variants (averaged over all tournaments). Hence, it measures how "robust" a given variant is to being dominated (i.e., outperformed) by other variants in a head-to-head tournament. In this sense it is a measure of relative performance a detailed description of how Robustness is calculated can be found in Section 4.3. We do not use the Aggressiveness measure in the present paper, but its definition can be found in [6].

## 2.2 BitTorrent

Maybe the most important idea behind the BitTorrent P2P protocol is that the files to be shared are divided into *pieces*. During the download of a particular file the peers (or nodes) obtain the pieces from a (usually dynamically changing) set of different nodes, which can consist of two types of users, leechers and seeders. *Leechers* are peers who are currently obtaining the file, i.e. those who do not have a copy of the entire file. *Seeders* are uploading exclusively, they do have all the pieces of the file. What makes the BitTorrent protocol highly scalable and very efficient is that the leechers can be uploaders as well. By default leechers follow a *rarest-first* rule to obtain pieces. Due to this, leechers have good chance to have pieces which can be traded for other pieces with other leechers, also called its neighbors. Each peer has upload capacity, which is divided into slots of two types: *regular unchoke* and *optimistic unchoke slots*. Regular unchoke slots are assigned to the subset $R$ of neighbors which recently provided data with the highest speed. The assignments get re-evaluated in every unchoke interval, which is usually fixed to 10 seconds. On the other hand, optimistic unchoke slots are assigned to strangers (i.e., randomly selected neighbors), which also get re-evaluated in fixed period of time. In case an optimistically unchoked peer $p$ is found to be faster than any of the regularly unchoked peers, then peer $p$ is moved to the set $R$, replacing the slowest peer in $R$.

The set of leechers and seeders who exchange pieces of a particular file is called a *swarm*. A collection of swarms are called a community. Communities usually emerge around a central web server, called a *tracker,* which is an important part of the network. A peer upon joining a swarm it requests the IP addresses of other

peers participating in this swarm. The tracker can have other features like providing a searchable database of available content.

### 2.3 Parametrization

The protocol design space of BitTorrent can be spanned over the following dimensions:

**Peer discovery**: in order to participate and possibly interact with each other in the same swarm, peers need to find each other.

**Stranger policy**: this policy is applied when a peer is interacting with a previously unknown peer.

**Selection function**: this function decides which of the known peers should be selected for interaction.

**Resource allocation**: defines the way a peer divides its (upload) resources among the selected peers (which are given by the Selection function).

Using this design space a user of the BitTorrent network can enter with a client using either default protocol parameters or modified ones. Modification of the protocol can be motivated by aiming at improving the individual performance or even to trick the system by freeriding (i.e. being only a leecher and not uploading to any other leechers). We term a specific actualization of the parameters listed above a *strategy.*

## 3 Evolutionary approaches

In this section we give details of the evolutionary algorithm which can be used to find out the evolutionary behavior of a pair of unique protocols. The algorithm starts with $N$ peers, and two strategies, $A$ and $B$. Initially, half of the peers use strategy $A$ and the other half use strategy $B$. The *fitness* of peer $i$ is defined as $f_i = 1 - w + wU_i$, where $U_i$ is the utility of peer $i$ and $w \in [0,1]$ measures the intensity of selection (strong selection means $w = 1$ and weak selection means $w \ll 1$). By default, we measure the utility $U_i$ as the average download speed in the current time interval, i.e., in the last $R$ rounds.

The steps of the algorithm are the following:

**Step 1** Let $f_i = 0$ for all $i = 1, \ldots, N$.

**Step 2** For $R$ rounds let the peers play the 'BitTorrent game'. Within a round, peers connect to each other (using a prescribed rule) and exchange pieces of the file they want to download.

**Step 3** In each *R*th round apply a selection rule on *K* pairs of peers in order to update the composition of these peers' strategies.

**Step 4** If the total number of steps equals to $R_T$ then stop, otherwise go back to Step 1.

Note that *K* represents the *selective pressure* meaning how much selection is performed per round. A collection of possible *selection rules*, which can be applied in Step 3, will be discussed in Section 3.1. It is worth to note that with the choice $R=R_T$, $K=N$ and $w=1$we get back the original robustness test used in [6].

## 3.1 Selection rules

Now we give a list of selection rules that will be used later in the experimental part (Section 4). All of them use the parameters *K* and *m*, which must be set up at the beginning of the experiment. Note that in all rules we apply a *mutation* operator, which is as follows. With probability *m*, peer *j* switches its strategy to the opposite one. Technically, this means that we generate a random number *r* between 0 and 1 and if $r \leq m$ holds, then we switch.

**Tournament selection**

This involves repeated tournaments between pairs of randomly selected peers as applied in [17].

Repeat *K* times:

Select two peers, peer *i* and *j* uniformly at random, with replacement.

If $f_i \geq f_j$ holds, then

Set the strategy of peer *j* to be the same as the strategy of peer *i.*

Apply mutation on peer *j.*

Reset $f_j = 0$.

end if.

**Death-birth updating**

This involves selecting a peer randomly and setting its strategy based on the proportion of the strategies used in the entire population [18].

Repeat *K* times:

Select a peer *j* uniformly at random.

Change its strategy according to: $F_{A\vee B}/\left(F_A+F_B\right)$, where $F_A=\sum_{q\ uses\ A} f_q$, $F_B=\sum_{r\ uses\ B} f_r$, and $F_{A\vee B}$ equals to $F_A$ if peer $j$ uses strategy $B$, and equals to $F_B$ if peer $j$ uses strategy $A$.

Apply mutation on peer $j$.

Reset $f_j=0$.

**Birth-death updating**

This involves selection of an individual proportional to its fitness value and then its offspring is replacing another randomly chosen peer.

Repeat $K$ times:

Select a peer $i$ proportional to its fitness. This is done in the following way: generate a uniformly random number $r$ from [0,1]; sort the peers according to their fitness values; start summing up the fitness values of the (sorted) peers, and denote this (partial) sum by $s$; normalize $s$; select the peer $i$ at which $s$ is greater than $r$.

Select another peer $j$ uniformly at random and change its strategy to the strategy of peer $i$.

Apply mutation on peer $j$.

Reset $f_j=0$.

**Satisficing updating**

In this decentralized selection rule we assume that all the peers keep record on their own fitness value from the previous $R$ rounds. This value is $\hat{f}_i$, where $i=1,...,n$. The general idea of satisficing was proposed in [19]. It captures the idea that individuals may be satisfied with an internally calculated threshold rather than comparing themselves to others.

Repeat $K$ times:

Select peer $j$ uniformly at random

If $f_j\geq\hat{f}_j$ holds, then peer $j$ keeps its current strategy, otherwise it switches to the opposite one.

Apply mutation on peer $j$.

# 4 Experiments

We compare the strategies tested in the simulations using the DSA approach previously discussed. Firstly, every strategy (protocol variant) is evaluated for Performance (based on average download time) when it is in a population composed entirely of the same strategy. This involves a realistic BitTorrent simulation (BitTorrent game) in which each peer attempts to download a file.

Then every possible pair of strategies are pitted against each other in a tournament by creating a population composed of half of each and executing the BitTorrent simulation. At the end of the simulation run the best performing strategy in terms of average performance is deemed to have "won" the tournament. A robustness measure is calculated for each protocol by averaging wins over all tournaments against all other protocols. Hence, a robustness value of 1 indicates a protocol wins against all others whereas a value of 0 indicates it loses against all others.

Finally, we apply the evolutionary extension by modifying the above tournament process in the following way. During a tournament, periodically, peers may change their strategy to their opponents strategy using one of the evolutionary algorithms described above. At the end of a simulation run the best performing strategy in terms of evolutionary success is deemed to have "won" the tournament. Wins are classified as "weak" or "strong" (see below for definitions). A strong and weak evolutionary robustness measure is calculated for each protocol by averaging over all tournaments.

In all cases simulations involved 10 independent runs starting with different pseudo-random number seeds and averages were calculated. This is necessary because the BitTorrent game simulator involves several stochastic elements.

## 4.1 Actualization of BitTorrent strategies

We selected a subset of the BitTorrent strategies tested in [6] covering the main interesting behavioral variants. The dimensions and possible values for each dimension are described below.

*Stranger policy*: Three kinds of policies are used here:

- Periodic: Give resources to strangers periodically.
- When needed: Only give resources to strangers when do not have enough regular partners.
- Defect: Never give resources to strangers.

*Number of strangers* can be either 0, 1 or 2.

*Selection function*: This depends on the Candidate list, Ranking function and the number of peers selected:

- For the *Candidate list* we use the BitTorrent's default *TFT*, in which a peer only places those peers in the candidate list who reciprocated to it in the last round.
- For the *Ranking function* we use all six actualizations: *Sort Fastest*, *Sort Slowest*, *Sort Based on upload bandwidth proximity* (called '*Birds*'), *Sort Adaptive* (ranks peers in order of proximity to an aspiration level, which is adaptive and changes based on a peer's evaluation of its performance), *Sort Loyal* (ranks peers in order of those who have cooperated with the peer for the longest durations), and *Random*.
- the number of top $k$ peers selected after applying the raking function can be from the range [0,4]. Note that the case $k$=4 is in line with the default BitTorrent parameter for the number of partners in a peer's unchoking slots.

*Resource allocation*: Two allocation methods are tested:

- *Equal split* gives all selected peers equal resources (upload bandwidth).
- *Freeride* gives nothing to partners.

Altogether, the above dimensions specify a space of 540 different protocol variants.

## 4.2 Strong and weak evolutionary robustness

Assuming that $R=R_T$, $K=N$ and $w$=1 in our algorithm, the robustness value of a protocol $P$ is calculated in the following way. For each run, we compare the average fitness value of $P$ with the average fitness of the other protocol. If the average fitness of $P$ is greater than the average fitness of the other protocol, we mark it as a '*Win*' for *P,* otherwise we mark it as a '*Loss*' for $P$. The robustness value for $P$ is calculated by number of games that it wins against all opponents in all runs divided by the total number of games that it plays, which is constant for all protocols.

This single robustness measure is sufficient when no evolutionary algorithm is applied. However, when applying an evolutionary algorithm this approach does not capture the possible dynamics of strategy change over time. We therefore introduce two new robustness measures for evaluating a protocol variant within an evolutionary algorithm: *weak evolutionary robustness* and *strong evolutionary robustness*.

Having evaluated the selection rule in Step 3 in the Algorithm explained in Section 3, we denote the number of peers associated with strategy $A$ and $B$ as $N_A$ and $N_B$, respectively.

Based on this fact, we define two kinds of *'win'*:

- If $N_A > N_B$ holds in each and every $R$th round of the algorithm, then we mark this as a '*strong win*' for protocol $A$.
- If $N_A \geq N_B$ holds at some $R$th round of the algorithm, and $\sum N_A \geq \sum N_B$, then we mark this as a '*weak win*' for protocol $A$.

The *strong evolutionary robustness* (and the *weak evolutionary robustness*) value of protocol $P$ is calculated by the number of games that it strongly wins (respectively weakly wins) against all opponents in all runs divided by the total number of games that it plays. For illustration of the definitions strong and weak evolutionary robustness see Figure 1. A protocol is strongly robust against another one if it never becomes a minority in the population during the simulation run. If the protocol was in the majority for most of the time, but not in all rounds, then it is weakly robust against the competing protocol variant[1].

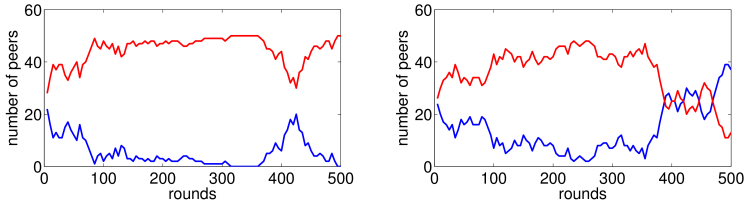



Figure 1

Illustration of strong evolutionary robustness (left) and weak evolutionary robustness (right) of a protocol denoted with red line

Weak and strong evolutionary robustness measures differ from the standard non-evolutionary robustness measure because they do not measure relative performance between protocol variants directly, but rather measure their replication success based on the outcome of an evolutionary algorithm.

# 5 Results

We have extended the cycle-based simulator used in [6] with the selection rules (listed in Section 3.1) and with the ability to measure the strong and weak evolutionary robustness of protocols (which were introduced in Section 4.2). In the following we use the terminologies introduced in Section 4.1.

[1] The situation where both protocols stayed evenly matched throughout all rounds was not found to occur in practice.

The simulation models a single swarm where peers share the same file. Time in this model consists of *discrete rounds*. For peer discovery we assume that all peers can connect to each other. In every round each peer is fired in random order and engages in connection and data transfer activities. It is assumed that all peers always have data that others are interested in. In each round, a peer:

- decides to upload to (maximum) 4 peers which are selected by the Selection function;
- uses its Resource allocation policy to decide how much to give to each of the selected partners;
- decides whether to cooperate or not with strangers, which is based on its Stranger policy.

Each peer maintains a short history of actions by others. This is implemented as a list, which has typically a length of 10-15 elements. At the same time, a peer also has some rate of requesting services from other peers that depends on specific actualizations. This means that different protocols will request services at different rates. This is how optimistic unchoking is mapped in our scenario. For example, a freeriding protocol will request a service with a partner every round and offer no bandwidth. If the chosen partner is a sucker (i.e., it accepts those protocols which provide zero bandwidth), then this partnership will be maintained further on.

Each simulation experiment was run with 50 peers and for 500 rounds in total. The peers' upload bandwidths are initialized using the bandwidth distribution provided by [2]. This was done by deriving uniformly at random samples from this dataset to assign to the peers in the simulations, hence preserving the distribution. The distribution represents real download rates that were empirically collected. Download is assumed to be infinite[2].

In the evolutionary approaches tested below we used the parameters: number of rounds $R_T$=500, number of peers $R$=50, selective pressure $K$=5, mutation rate $m$=0.01, and intensity of selection $w$=1 as defaults. Note that we also used other parameter setups and found no significant differences from the results obtained using the defaults.

Given the abstraction level of the simulator only the above parameters are required. There is no explicit representation of pieces or seeding/leeching behavior. Hence, the focus is on the effect of strategy interactions on data sharing[3].

[2] Since upload bandwidth is the main constraint in file sharing systems this assumption does not significantly effect results.

[3] A previous version of the simulator was validated against an actual BitTorrent client implementation and experiments [6], which gives us some confidence in the validity of the results.

## 5.1 Without Evolution

Firstly, we calculated performance and robustness values for all our 540 protocol variants *without* applying evolution - see Figure 2. Note that Performance is normalized over the entire protocol design space. These results serve as a baseline for comparison with the evolutionary approaches. Also, we were interested to see the effect of selecting only a subset of potential protocol variants from [6]. We categorize those protocols which never share as freeriders.

We found that the ranking of the protocols is comparable to those previously obtained. This result is non-trivial and reassuring. This is because it is the nature of the DSA approach that the performance and robustness of any given strategy is only calculated relative to the other strategies in the design space. This means that results obtained for one space of strategies can not be generalized to either a subset or superset of strategies without testing. Hence our reproduction of results for our chosen subset allows us to be more confident that the previous results obtain were not merely the result of an artifact of large the design space chosen there.

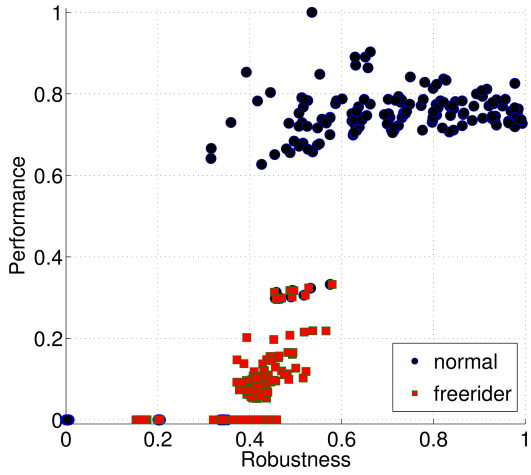



Figure 2

Robustness against Performance using 540 protocol variants without apply evolution

## 5.2 Tournament selection[4]

We performed evolutionary simulations for all protocol variants using the tournament selection approach. We then analyzed these results by comparing them to the Performance and Robustness obtained from the previous non-evolutionary

[4] Note that the term "tournament selection" should not be conflated with the "tournament" nature of the DSA approach. These refer to two different kinds of tournament occurring at two different levels.

simulations. As previously stated (in Section 4.2), each protocol is measured along two new measures - strong evolutionary robustness and weak evolutionary robustness.

Figure 3 shows the results we got for the Robustness test. We can immediately see that strategies cannot be quantified using weak evolutionary robustness, all the protocol variants have roughly the same value. On the other hand, strong evolutionary robustness correlates with the previously obtained non-evolutionary robustness. The Pearson's correlation coefficient of robustness and strong evolutionary robustness is 0.939.

This result indicates that evolution based on tournament selection effectively reproduces the non-evolutionary DSA approach so long as strong evolutionary robustness is used as a measure for winning. Or to put this another way, the non-evolutionary DSA approach is a good predictor of strong evolutionary robustness in this evolutionary setting. This is a non-trivial result because strong evolutionary robustness measures the outcome of an evolutionary processes over time, whereas non-evolutionary robustness is based on relative performance in one-shot interactions.

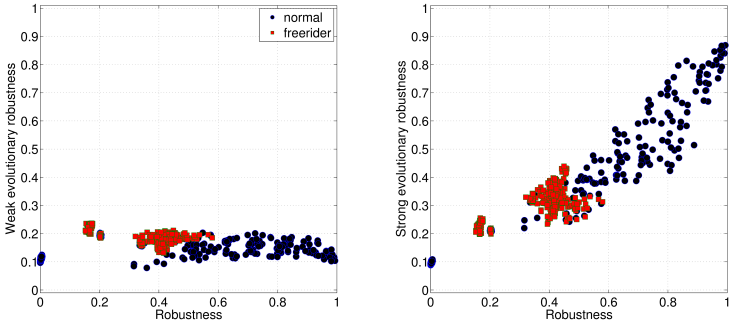



Figure 3

Tournament selection – weak evolutionary robustness and strong evolutionary robustness compared to non-evolutionary robustness

Figure 4 shows weak and strong evolutionary robustness against performance. Note that performance is calculated identically for both non-evolutionary and evolutionary approaches. Hence, performance values are identical to those given in Figure 2. As we would expect, from the results given in Figure 3, we see that weak evolutionary robustness does not distinguish between performance whereas strong evolutionary robustness reproduces the non-evolutionary results and is directly comparable with Figure 2. The Pearson's correlation coefficient of performance and strong evolutionary robustness is 0.761.

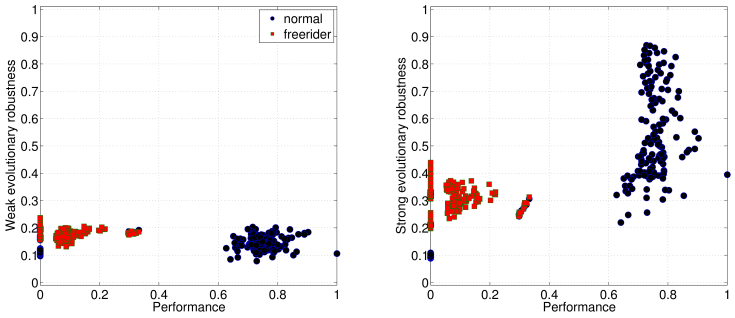



Figure 4

Tournament selection – performance

## 5.3 Death-birth selection, Birth-death selection

For both Death-birth and Birth-death selection we obtain almost identical results to those for Tournament selection. Figure 5 shows the correlation between the weak and strong evolutionary robustness measures of the different protocol variants we found using Tournament and Birth-death selection. As we can see the weak evolutionary robustness values are very strongly correlated. In the case of strong robustness we notice that some strategies can have significantly higher value under Birth-death selection. Closer inspection of these particular variants indicated they used the Periodic or When-needed stranger policy, Sort fastest ranking function, 1 or 2 regular and optimistic unchoke slots and Equal split as resource allocation.

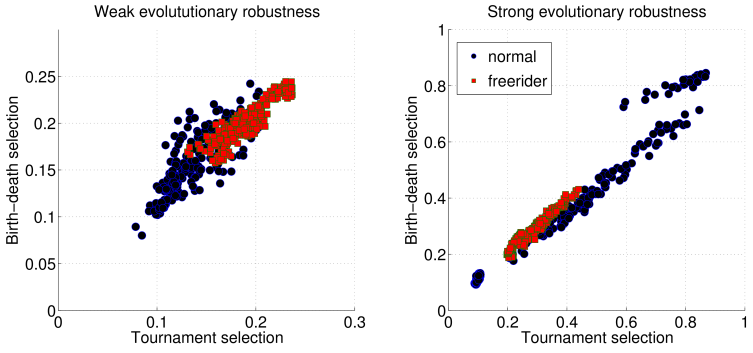



Figure 5

Comparison of Tournament and Birth-death selection.

Interestingly, the Death-birth selection mechanism produced very much the same ranking as Birth-death selection. Namely, we got 0.987 and 0.998 Pearson's correlation coefficients for weak and strong evolutionary robustness, respectively.

We can conclude that these evolutionary algorithms also reproduce the non-evolutionary results when the strong evolutionary robustness measure is used. This indicates that different evolutionary algorithms based on fitness comparisons are predicted by the non-evolutionary DSA approach.

## 5.4 Satisficing selection

This selection mechanism differs from the above tested ones as it is using only local information. This means that no fitness comparisons are made between peers but rather individual peers assess their own performance against an internal threshold or target.

As can be seen in Figure 6 and 7, neither weak nor strong evolutionary robustness make any significant distinction between the different protocol variants. In this case we cannot even distinguish between freeriders and normal strategies.

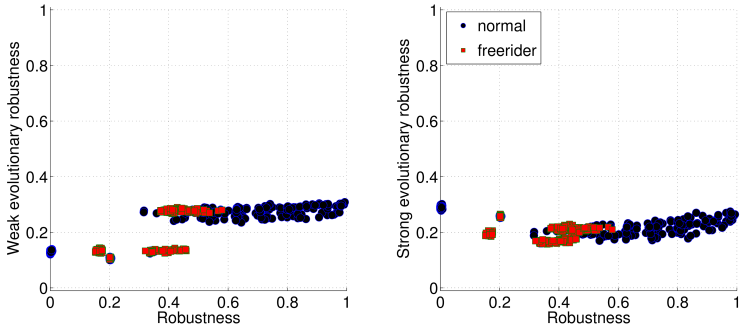



Figure 6

Satisficing selection – robustness

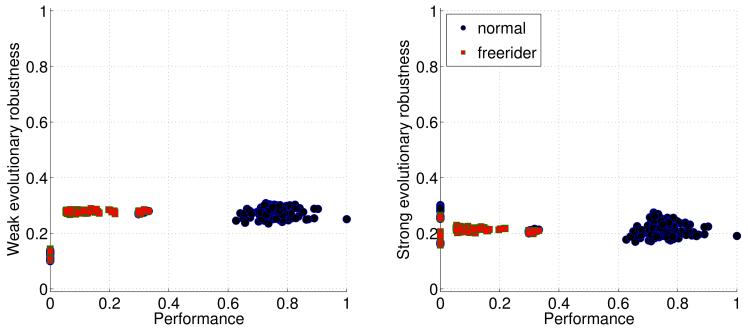



Figure 7

Satisficing selection – performance

In general outcomes appear to be little more than random noise. Since there are four possible outcomes from a tournament we would expect uniformly random outcomes to produce weak and strong evolutionary robustness values around 0.25.

Note, however, a small cluster of freeriders can be seen to have noticeably lower weak evolutionary robustness values around 0.4 on the *x*-axis (Figure 6, left hand side). These appear to be a special case of very poorly performing protocols, as can be seen in Figure 7 (left hand side) where this cluster is bunched at 0.0 on the *x*-axis.

More interestingly, a small cluster of 72 non-freerider variants can be seen in Figure 6 (right hand side) with near-zero robustness, but relatively high levels of strong evolutionary robustness (around 0.3). This means that these variants which have almost zero non-evolutionary robustness do much better in an evolutionary setting. These protocols do not communicate with strangers (thus they are unable to bootstrap in an homogeneous environment, i.e., producing zero performance), use Equal split as resource allocation and have $k>0$ regular partners. Closer inspection of the simulation results revealed that these particular variants *always* dominate in scenarios where they are paired with freeriders. This does not imply, though, that these protocols do better than freeriders. Due to the local nature of the calculation of the satisfaction threshold this only means that variants using these kind of protocols do gradually better than in the previous round, whereas the competing freerider protocol variant' performance varies up and down.

Overall we can see that the non-evolutionary DSA approach does not predict the outcome of the satisficing approach. However, there are *many* ways to implement a satisficing approach and it would be interesting to explore other implementations to identify conditions under which (if any) results would converge to the previous DSA results. For example, some satisficing approaches utilize an adaptive satisfaction threshold that may increase above over all previous performances obtained through the application of noise (a "trembling hand effect") [20]. Alternatively a minimum as well as maximum threshold could be employed.

## Conclusions

We developed a limited evolutionary extension to the Design Space Analysis approach [6]. We applied this extended approach to an exploration of BitTorrent protocol variants capturing the possibility of dynamic changes in protocol variants over time within pairwise tournaments between protocols. We found that the results obtained were broadly consistent, in most cases, with those previously obtained thus increasing our confidence in previous results as a predictor for this form of limited evolution.

The subset of protocol variants that we selected ranked similarly to the larger design space results. This is a non-trivial observation since results form a given design space do not necessarily generalize to a subspace. This is due to the co-evolutionary nature of open peer-to-peer systems - where the utility of each peer is dependent on the dynamic composition of protocol variants in the population as a whole rather than being related to a fixed and equal partition.

We found that applying a satisficing approach did *not* reproduce the previous results and in fact, as implemented here, produced almost random outcomes. However, some of the results point towards the kinds of alternative satisficing approaches that might produce comparable outcomes. It is of interest how different satisficing approaches behave because it could be argued that such a procedure may capture the kinds of user behavior that leads to a change of protocol variants over time - since users tend to have access only to local information on protocol performance.

The work here limits coevolution to all pairs of variants. To fully understand co-evolutionary dynamics it would be necessary to allow for many protocol variants from the design space to exist in the population simultaneously. However, previous work has demonstrated the difficulty in analyzing the results of evolution applied to populations with design spaces as low as six variants [16]. Our aim in limiting evolution in this way is as a step *towards* modeling and understanding co-evolutionary processes in large design spaces while linking back to previous results. This requires careful experimental design and analysis to avoid a combinatorial explosion and to filter for noise which plays a large role in protocol pairings and can be magnified by coevolution.

Future work may develop coevolutionary algorithms that can be applied to a large design space within a single population by grounding the algorithms in a theory of user and developer behavior. Developers modify and release new protocol variants and users decide if to download and use them. It is not beyond current approaches to model this process in a coevolutionary simulation, but more detailed user and developer models would need to be formulated.

**Acknowledgement**

This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013). T. Vinkó was supported by the Bolyai Scholarship of the Hungarian Academy of Sciences.

**References**

[1] B. Cohen, Incentives build robustness in BitTorrent, In Proceedings of Workshop on Economics of Peer-to-Peer Systems, 68-72, 2003

[2] M. Piatek, T. Isdal, T. Anderson, A. Krishnamurthy, and A. Venkataramani. Do incentives build robustness in BitTorrent. In Proceedings of NSDI, vol. 7., 2007

[3] T. Locher, P. Moor, S. Schmid, and R. Wattenhofer. Free riding in BitTorrent is cheap. In Proc. Workshop on Hot Topics in Networks (HotNets), 85-90. 2006

[4] U.W. Khan, and U. Saif, BitTorrent for the less privileged. In Proceedings of the 10th ACM Workshop on Hot Topics in Networks (HotNets-X), 2011

[5] N. Liogkas, R. Nelson, E. Kohler, and L. Zhang, Exploiting BitTorrent For Fun (But Not Profit), In Proc. of IPTPS, 2006

[6] R. Rahman, T. Vinkó, D. Hales, J. Pouwelse, and H. Sips, Design Space Analysis for Modeling Incentives in Distributed Systems. In Proceedings of the ACM SIGCOMM 2011 Conference (SIGCOMM'11) 182-193, 2011

[7] R. Axelrod, The Evolution of Cooperation. Basic Books, 1984

[8] R. Dash, N. Jennings, and D. Parkes. Computational-mechanism design: A call to arms. IEEE Intelligent Systems, 18:40–47, 2003

[9] J. Feigenbaum and S. Shenker. Distributed Algorithmic Mechanism Design: Recent Results and Future Directions. In ACM DIALM, 2002

[10] R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan, Experiences applying game theory to system design. In Proceedings of the ACM SIGCOMM workshop on Practice and theory of incentives in networked systems (PINS '04). ACM, New York, NY, USA, 183-190, 2004

[11] C. Buragohain, D. Agrawal, and S. Suri. A game theoretic framework for incentives in P2P systems. In Proceedings of IEEE P2P, 2003

[12] M. Feldman, K. Lai, I. Stoica, and J. Chuang. Robust incentive techniques for peer-to-peer networks. In Proceedings of ACM EC, 102–111, 2004

[13] A. Bharambe, C. Herley, and V. Padmanabhan. Analyzing and improving a BitTorrent network's performance mechanisms. In Proceedings of INFOCOM, 1-12, 2006

[14] M. Meulpolder, J. Pouwelse, D. Epema, and H. Sips. BarterCast: A practical approach to prevent lazy freeriding in P2P networks. In Proceedings of IEEE IPDPS, 2009

[15] J. Mol, J. Pouwelse, M. Meulpolder, D. Epema, and H. Sips. Give-to-get: Free-riding-resilient video-on-demand in p2p systems. In Proceedings of SPIE/ACM MMCN, 2008

[16] X. Jin, Y.-K. Kwok, J. Deng, Variegated competing peer-to-peer systems with selfish peers, Computer Networks, (24)313-330, 2014

[17] R.L. Riolo, M.D. Cohen, and R. Axelrod, Evolution of cooperation without reciprocity. Nature, 414(6862), 441-443, 2001

[18] H. Ohtsuki, C. Hauert, E. Lieberman, and M.A. Nowak, A simple rule for the evolution of cooperation on graphs and social networks. Nature, 441(7092), 502-505, 2006

[19] H. A. Simon, Models of Bounded Rationality, Vol. 1, MIT Press, Boston, 1984

[20] C.P. Roca and D. Helbing, Emergence of social cohesion in a model society of greedy, mobile individuals, PNAS 108 (28) 11370-11374, 2011

# Operational Study of a Frequency Converter with a Control Sequence, Utilizing Xilinx Software

**Corina Daniela Cunţan, Ioan Baciu, Mihaela Osaci**

"Politehnica" University of Timisoara, Revolutiei no. 5, 331128 Hunedoara, Romania, corina.cuntan@upt.ro, ioan.baciu@upt.ro, mihaela.osaci@upt.ro

*Abstract: The paper analyses the operation of a single phase – three phase frequency converter, built with insulated-gate bipolar transistors (IGBT). The bridge inverter control is made via the Basys 2 development board which generates voltages, at the output terminals, compatible with the logic levels of complementary metal–oxide–semiconductor (CMOS) circuits. The diagram for obtaining the control pulses is realized with Xilinx software that allows changing the output frequency to a value of 1 KHz. The analysis of converter operation shall be made using either a resistive load or an inductive load. The installation model presented in the paper highlights the operation modality and the study possibility, by experiments at various frequencies and load circuits. By analyzing the waveforms obtained in various points of the circuit diagram, it results the possibility to improve certain operating parameters over others, according to the user requirements and the regulations in force for a harmonic regime.*

*Keywords: frequency converter; galvanic separation circuit; pulse sequence; development board; circuit switching; DC/AC power convertors*

# 1 Introduction

Static power converters include complex techniques including circuit switching, control, regulation and conversion, using power semiconductors, diodes, transistors, etc., completed by auxiliary measurement and protection circuits. A spectacular evolution in the conversion of electrical energy was produced by the development and emergence of new power semiconductor devices which facilitated the improvement and diversification of the power converter, which is the element situated between the power supply and the power consumer. The purpose of this element is to convert the shape and parameters of electrical energy according to customer requirements, conversion characterized by high efficiency and reduced generation of high-order harmonics in power networks or to the power consumer. [8]

Most of the current technological processes of the various industrial sectors, increasingly require cheap, but robust electric motors. The DC motors with an adjustable speed have limited the scope of use due to the presence of brushes and mechanical collectors, which reduce operational reliability and is a risk factor, due to the sparks that can occur in presence of a flammable (explosive) environment. Therefore, it became imperative to develop drive systems with motors without sliding contacts, which have at least the adjustment performances in dynamic regime by the previous DC motors.

It should be noted the progress made by the emergence of digital signal processors (DSP), which, thanks to their greatly increased computing power (i.e. multiplication and addition in one cycle, clock rates of hundreds of MHz), enable the real-time implementation of complex and sophisticated algorithms.

The rapid growth of technology development and the progress in the field of power electronics, now allow for the development of extremely efficient electric drive equipment. The commensurate development of microelectronics and information technology has resulted in major changes within the field of electronic machine control, and the emergence of digital signal processors dedicated to digital motion control make possible the widespread deployment of control numerical algorithms. An important benefit for electric drives is shown in the development of intelligent power modules, incorporating either control or protection devices. Given the currently evolving direction of the high performance electric drives, it is obvious that large-scale development of numerical control algorithms plays an extremely important role.

The reconfigurable circuit board (FPGA), offers a compromise between the performances of a specific circuit and the flexibility of a software programmable circuit. These circuits are characterized by their ability to implement specialized circuits directly in the hardware. In addition, they can be easily modified when the operating conditions and/or dataset changes. [3]

The digital circuits are easier to design and test, because they can be designed from the logic gate level or functional level. The FPGA circuits have one of the shortest development and market integration timeframes. The FPGA circuits are inherently reconfigurable circuits, this feature alone allows for greater flexibility.

# 2 The Control Circuit

The control sequence requires 6 pulse trains of alternating ones and zeroes, so as to obtain simultaneously, three signals on a 1 and the others on a 0, corresponding to the transistors needing to be controlled in the three-phase output voltage system. For this, we used the clock signal of the development board, whose frequency is reduced by a division circuit made in the VHDL code (Figure 1) [1].

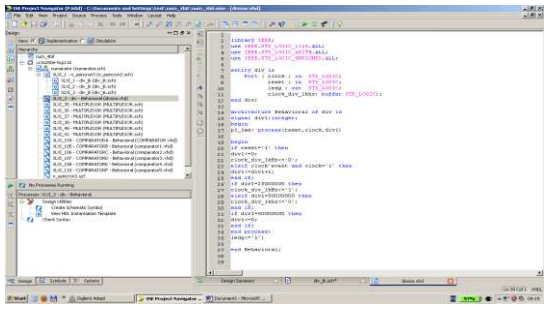


Figure 1

The VHDL code for the frequency divider

By means of the division circuit, we obtain the desired control frequency. From the division circuit we have also obtained the clock signal for the counter, comparators and multiplexers. The counter (Figure 2) is a three-bit asynchronous counter made with T-type flip-flops (Figure 3). For obtaining the 6 control sequences of the IGBT transistors, the counter was supplemented with a circuit for the identification of the 6$^{th}$ status (110), which then, initializes its status.

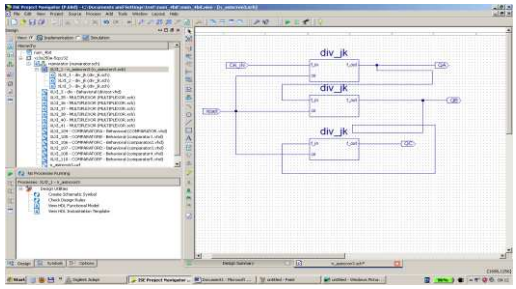



Figure 2

The 3-bit asynchronous counter realized in Xilinx

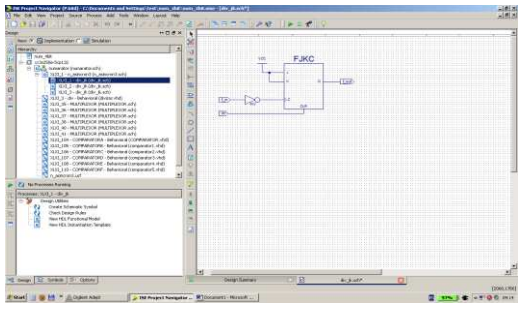



Figure 3

T-type flip-flop circuit

For each counter output sequence, the selection of each comparator is made via the control logic. We made six comparators in VHDL code, each comparator having entered in its structure the code corresponding to the IGBT transistors for 1/6 of the control sequence (Figure 4).

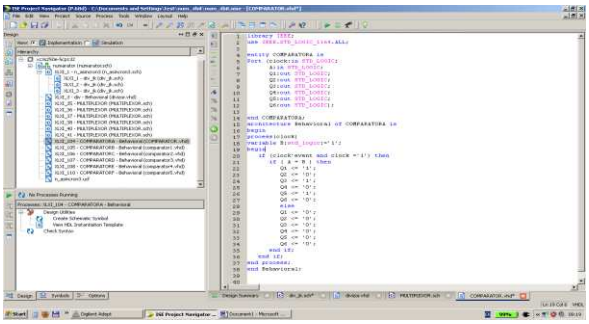


Figure 4

The VHDL code for a comparator

The comparator outputs are connected to the inputs of six multiplexers controlled by the counter output sequence (Figure 5). For each of the six control sequences, each multiplexer selects one of the counter output sequence. The multiplexer selection is made via a logical structure that receives commands from the counter output. Each of the six output states of the counter allows for the selection of a different sequence, corresponding to the control requirement of the bridge inverter transistors. The multiplexer outputs are brought via the output pins of the Basys2 board to the galvanic separation circuit for the transistor inverter controls.

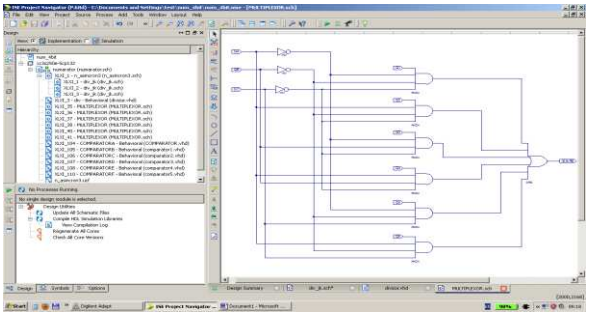


Figure 5

The multiplexer realized in Xilinx

For obtaining the control signals with "dead time" between the transistor control sequences we use a second divider, whose frequency is 10 times higher than the frequency of the first divider. This signal is used as a clock pulse for the second modulo-10 counter. The signals from the counter output go through an OR gate with 4 inputs which controls 6 AND gate circuits with two inputs and to the other

input of each AND gate arrive the signals from the multiplexer outputs (Fig. 6). Thus, at the output we obtain a control sequence for each transistor with a pause time 10 times lower than the normal clock sequence.

From the above, we can obtain the complete diagram (Figure 6) which allows obtaining the control sequence (Figure 7) at the output connectors of the board. [2]

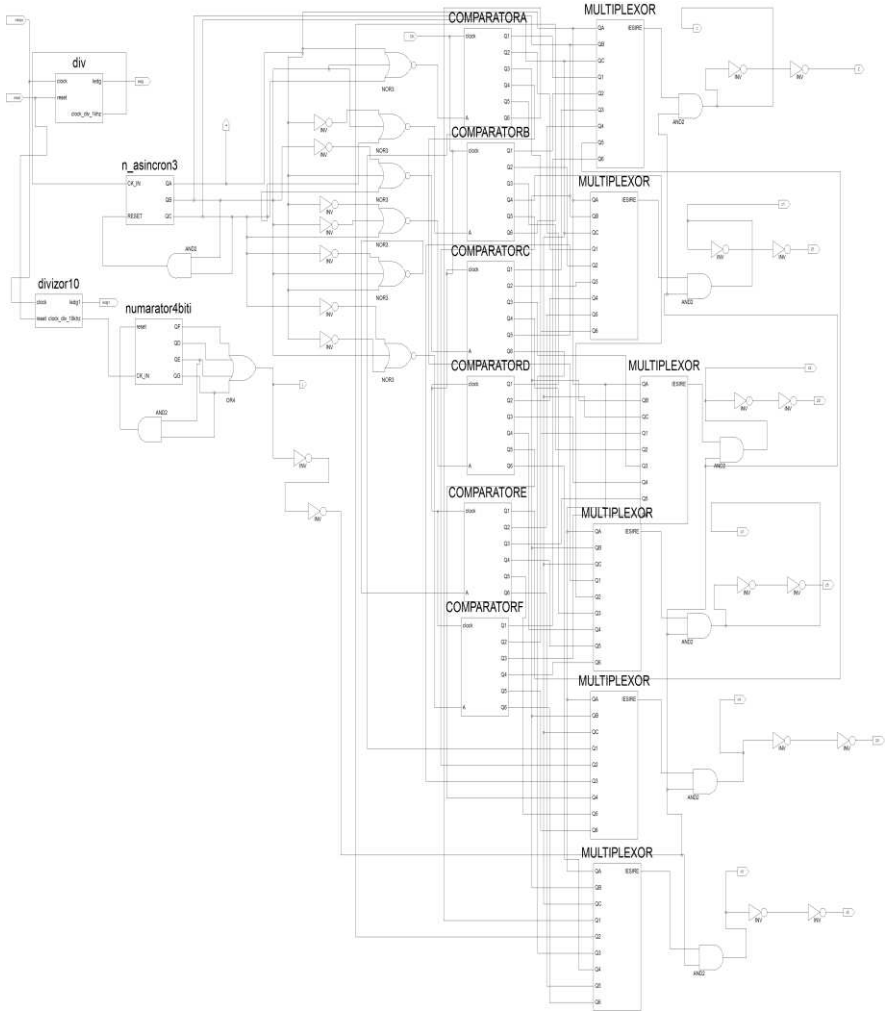



Figure 6
The control circuit diagram for obtaining the control sequences realized in Xilinx

# 3 Results and Discussions

We can see (Figure 7) the six control sequences corresponding to the transistors of the bridge inverter, either in graphic form, or as number sequence. [3]

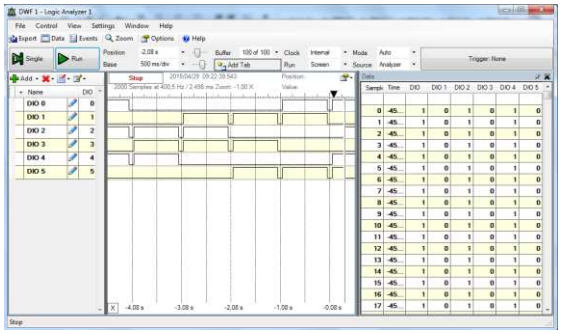


Figure 7
Inverter control sequence

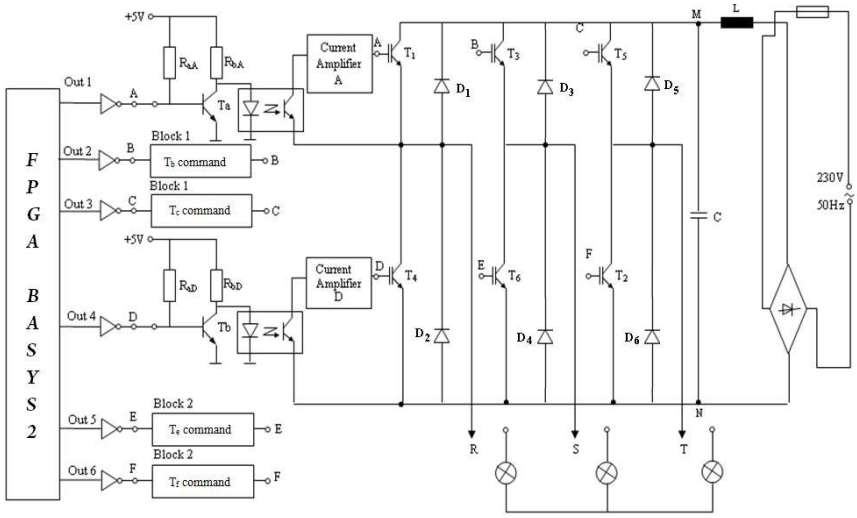


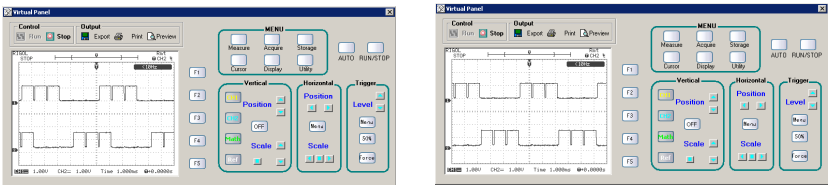

Figure 8
The diagram of the transistor voltage inverter – IGBT

The converter diagram (Figure 8) contains a galvanic separation circuit, a control pulse amplifying circuit, a single phase rectifier circuit, an intermediate DC circuit and the transistor bridge inverter IGBT. [4], [5], [6]

The control sequence of the Basys2 board is brought via MOS inverters and power amplifiers at the input of six optocouplers. They control, via the amplifier circuits, the transistor gates of the bridge inverter. [11]

The control sequence is obtained from the clock counter and is formed of six intervals, the resulting frequency being six times lower than the clock frequency.

From the pins of the Basys2 board connector, via a MOS inverter, the control signal is brought amplified in current by the transistors Ta…Tf at the inputs of the six optocouplers [1]. Examples are presented for three pairs of such sequences, i.e.: UCE voltage for Ta, UCE voltage for Tb (Figure 9a), UCE voltage for Ta, UCE voltage for Td (Figure 9b), corresponding to the frequency of 166.66 Hz, value obtained from the 1 KHz clock frequency of the counter divided by six, the circuit having a resistive load connected to the output.



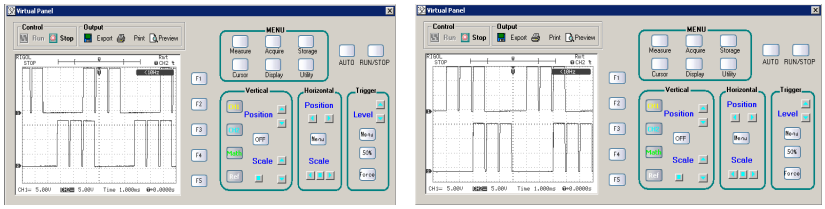
(a) (b)

Figure 9

Control voltages at optocouplers in the collector terminal of the transistors Ta and Tb (a) and Ta and Td (b)

By analyzing the waveforms obtained using the digital oscilloscope and stored on a PC, it appears to be a delay in the control sequence specific to the transistor pairs.

The control sequences in the collector terminals of the transistors Ta…Te are galvanicaly separated using optocouplers, and amplified in current with six amplifiers, each one provided with a high power switching transistor. The amplifier output is brought to the transistor gates of the bridge inverters. We present the emitter gate voltages for the transistors T4 and T6 (Figure 10a), T4 and T2 (Figure 10b).



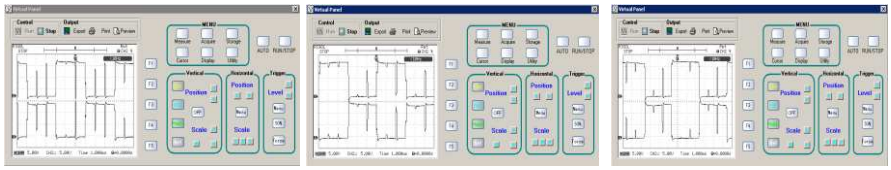
(a) (b)

Figure 10

The emitter gate voltage for the transistors T4 and T6 (a) and T4 and T2 (b)

We further present the waveforms for the emitter gate voltages for each of the three transistor pairs of the bridge inverter that has a resistive load connected as consumer. The visualization of the voltage pairs was made with a two-channel

digital oscilloscope which enables the data transfer into a computer system, as follows: the collector-emitter voltage $U_{CE}$ for $T_1$ and $T_4$ (Figure 11a), $U_{CE}$ for $T_3$ and $T_6$ (Figure 11b), $U_{CE}$ for $T_2$ and $T_5$ (Figure 11c).

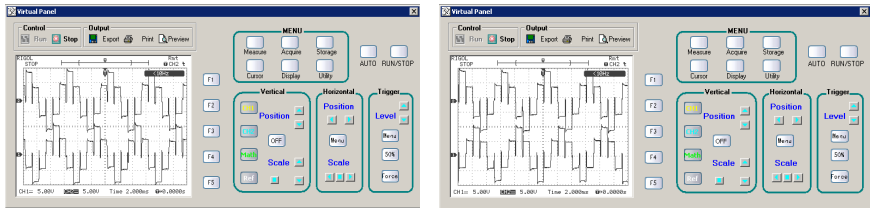


(a) (b) (c)

Figure 11

The collector-emitter voltage for the transistors T1 and T4 (a), for the transistors T3 and T6 (b) and for the transistors T5 and T2 (c)

From the analysis of the waveforms corresponding to the collector - emitter voltages of the transistor pairs of the bridge inverter at the frequency of 166 Hz, we obtain a complementary operation without switching delay or accidental switching. Further, we present the waveforms for the output voltages at the resistive load, as follows: the output voltages for the phases R and S (Figure 16), and the output voltages for the phases R and T (Figure 17). It can be seen that the output voltages have synthetic shape, with the phase angle shift of $120^o$ and without random switching.




(a) (b)

Figure 12

The output voltages: Phases R and S (a) and Phases R and T (b)

We present the emitter gate voltages and the emitter collector voltages for the transistors IGBT, at the clock frequency of 5 KHz, and the output voltages at the same frequency, the converter being connected to the same resistive load (Figure 13). [11]

In accordance with the commands at 833 Hz, we represent the emitter collector voltages for the transistors T1 and T4 (Figure 14), being the same for the other pairs of transistors T3-T6 and T5-T2, the reason why they will not, will be shown in separate figures.

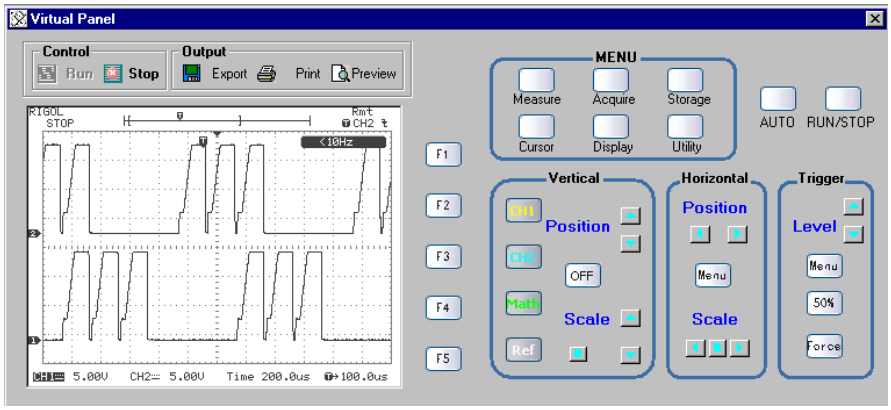



Figure 13
The GE voltages for the transistors T4 and T6 at 833 Hz

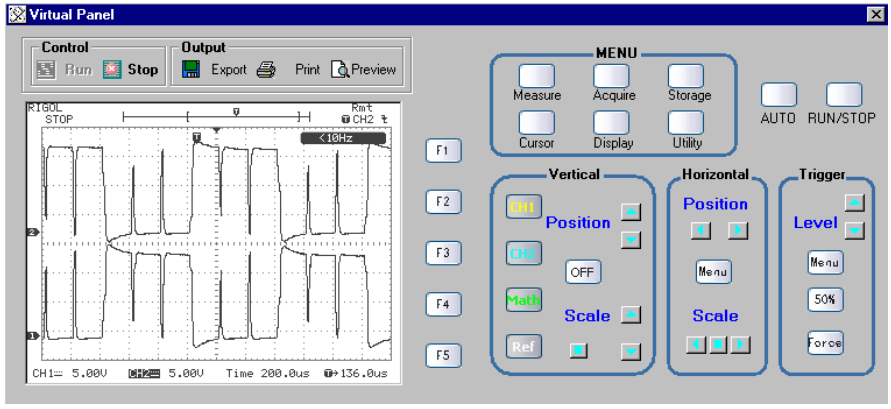
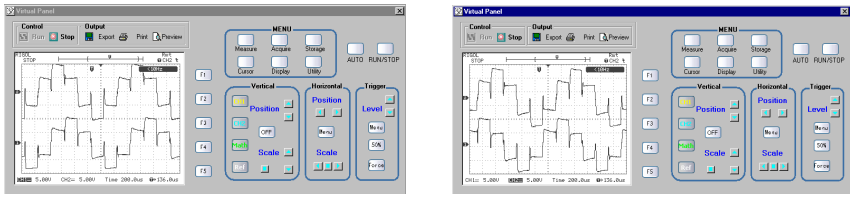



Figure 14
The UCE voltages for the transistors T1 and T4

The proper operation of the inverter at resistive load is evidenced by the waveforms of the output voltages $U_{RS}$ and $U_{RT}$.



(a) (b)

Figure 15
The output voltages: Phases R and S (a) and Phases R and T (b)

The inverter operation was also checked with inductive load; more precisely, we controlled a 0.37 kW three-phase asynchronous motor with squirrel cage rotor. [7], [9]

The waveforms for the output voltages $U_R$ and $U_S$ at inductive load are visualized on an analogue oscilloscope at the clock frequency of 1 KHz, which corresponds to the motor control frequency of 166.6 Hz (Figure 16), and to the clock frequency of 2 KHz, which corresponds to the motor control frequency of 333.3 Hz (Figure 17).

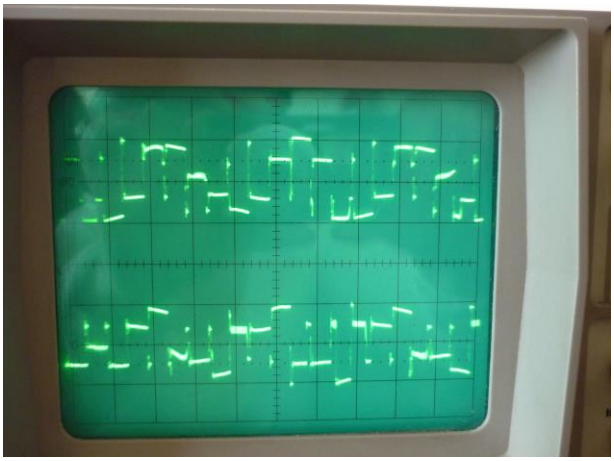


Figure 16
The $U_R$ and $U_S$ voltages

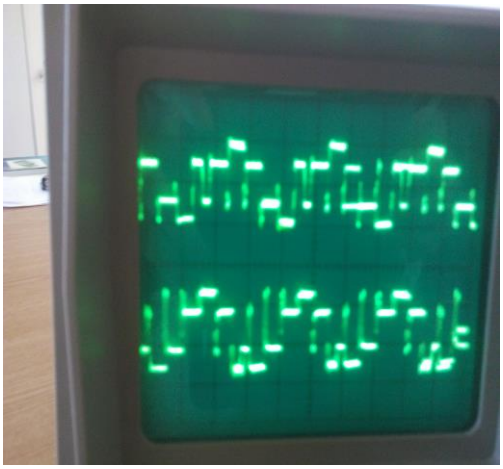


Figure 17
The output voltages - Phases $U_R$ and $U_S$

There is deterioration in the operation of the bridge inverter transistors, particularly, during their blocking, when there is a relatively small exponential variance, but which does not affect the output voltage so profoundly. This can be seen in Figure 18. [10]

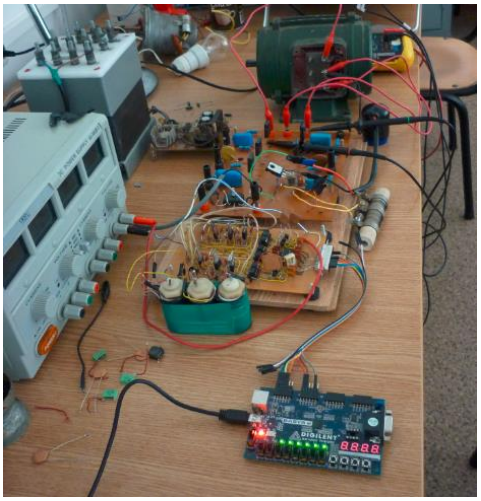


Figure 18
The model made to study the frequency converter

## Conclusions

Herein it was found that the frequency converter operates properly. The converter is designed to operate with full wave control signal and being only laboratory equipment, was operated with a low load. The correct operation of the converter components is shown through the waveforms at certain points within the circuit. The operation analysis was carried out at resistive load, up to the frequency of 833 Hz and at inductive load at a frequency of 333.3 Hz. These values resulted by dividing the clock frequencies of 5 KHz and 2 KHz by 6, a number representing the control sequences of the inverting bridge. The measurements for the resistive load were made with a digital oscilloscope and those for the inductive load were made with an analog oscilloscope. The converter output frequency can be increased to greater than 800 Hz, but verification can be made only on the resistive load, because certain problems usually occur with the inductive load at frequencies above 400 Hz (iron loss and overheating of the bearings).

The installation model presented in this work (Figure 18) highlights the operational modality and the study possibilities by experimentation at various frequencies and load circuits.

## References

[1] Mekhilef, S. Rahim, N. A., *Xilinx FPGA-based Three-Phase PWM Inverter and its Application for Utility Connected PV System*, TENCON '02 Proceedings. IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering, 2079-2082, 28-31 Oct. 2002

[2] Mekhilef, S. Masaoud, A., *Xilinx FPGA-based Multilevel PWM Single Phase Inverter*, ICIT 2006, IEEE International Conference on Industrial Technology, Mumbai, 15-17 Dec. 2006, pp. 259-264

[3] Arshak, Khalil, Jafer, E., Ibala, C., *Testing FPGA-based Digital System using XILINX ChipScope Logic Analyzer*, ISSE '06, 29th International Spring Seminar on Electronics Technology, St. Marienthal 10-14 May 2006, pp. 355-360

[4] Baciu Ioan, Cunţan Corina Daniela, *Operation Analysis of a Frequency Converter with Control Realized in LabView,* 2012 IEEE International Conference on Industrial Technology, 19-21 March 2012

[5] E. C. dos Santos, C. B. Jacobina, N. Rocha, J. A. A. Dias, M. B. R. Corea, *Single-Phase to Three-Phase Four-Leg Converter Applied to Distributed Generation System*, IET Power Electronics, Vol. 3, Issue 6, 2010, pp. 892-903

[6] Dimitrie Alexa, Octavian Hrubaru, *Applications of the Static Power Converters*, Publisher: Tehnică, Bucharest, 1989

[7] Măgureanu, R., Micu, D., *Static Frequency Converters in Drives with Induction Motors*, Publisher: Tehnică, Bucharest, 1994

[8] Ionescu F., Florincău D., Niţu S., Six J. P., Delarue Ph., Boguş C., *Power Electronics. Static Converters*, Publisher: Tehnică, Bucharest, 1998

[9] Liu W., Cheng G. W., Deng C. N., *Analysis of Electric Vehicle's Driving Motor*, Motor Technology, 2006, pp. 26-28

[10] Bitoleanu, A., Popescu, M., Linca, M., Mitran, A., *Energetical Performances Analysis of PWM Asynchronous Motor and Voltage Inverter Driving System*, Advanced Topics in Electrical Engineering (ATEE) 2011, 7th International Symposium on, Issue Date: 12-14th May 2011, pp. 1-6

[11] Hisayuki Sugimura; Ahmad Mohamad Eid; Soon-Kurl Kown; Hyun Woo Lee; Hiraki, E.; Mutsuo Nakaoka, *High Frequency Cyclo-Converter using One-Chip Reverse Blocking IGBT-based Bidirectional Power Switches*, Proceedings of the Eighth International Conference on Electrical Machines and Systems, ICEMS 2005, 29-29th Sept. 2005, pp. 1095-1100, Vol. 2

# Hand Vein-based Multimodal Biometric Recognition

**S. Bharathi[1], R. Sudhakar[1], Valentina E Balas[2]**

[1]Department of Electronics and Communication Engineering, Dr.Mahalingam College of Engineering and Technology, Pollachi -642 003, Tamilnadu, India
E-mail: sbharathi@drmcet.ac.in, hod_ece@drmcet.ac.in

[2]Department of Automatics and Applied Software, Aurel Vlaicu University of Arad, 77 B-dul Revolutiei, 310130 Arad, Romania
E-mail: valentina.balas@uav.ro

*Abstract: Multi-modal biometric recognition utilizes more than one modality for recognition of a person, when compared to the single modality with the help of enhanced security. The vascular patterns are one of the physiological biometrics and internal feature of the body, hence, it cannot be easily cracked, falsified or spoofed. This paper proposes a novel approach for biometric recognition, using finger vein, palm vein and dorsal vein of the hand. Our proposed method utilizes both the Shearlet transform and Scale-invariant feature transform to extract features from the hand vein images. The extracted features in the form of coefficients are stored in the data base. Then the matching is done between the coefficients of the input test images and the features stored in the data base using distance measure and finally the fusion is carried out using the maximum likelihood ratio technique. This approach was tested on standard data bases of finger vein, palm vein and dorsal vein images of hands. The proposed method provides a maximum accuracy of 94%, with a reduction in false acceptance and false rejection rates, illustrating the efficiency of the technique compared to other current methods.*

*Keywords: hand vein; shearlet transform; scale-invariant feature transform; multi-modal biometric recognition; maximum likelihood-based fusion*

# 1 Introduction

Biometrics has many advantages over conventional safety measures such as Non-repudiation, Accuracy, Screening and Security. Nowadays, biometric recognition is a familiar and reliable way to authenticate the identity of a person through physical measurements of unique human characteristics or behavior. A physiological characteristic is a distinctive personal traits, such as fingerprint, palm print, hand vein, iris pattern and many more [1]. Signature, gait and

keystrokes are some of the behavioral characteristics. Biometric technology provides assurance of a trouble-free, safe technique to make an exceedingly precise authentication of a person [2, 3]. Alternate symbols of distinctiveness like passwords and identity cards have the vulnerability of being easily misplaced, shared or stolen. Passwords may also be easily cracked by means of social engineering and dictionary attacks [4, 5] and offer negligible protection. As biometric techniques make it a precondition for the client to be physically present during validation, it serves as a deterrent against the possibility of clients emerging with false refutation claims at a later stage [6].

The excellence of a biometric technique is assessed by means of its inherent competence. Regardless, even the finest biometrics, of our modern age is haunted by many problems, several of them innate in the technology itself. Multi-biometrics is a comparatively novel method to conquer these obstacles by means of manifold models of a solitary biometric trait, known as, multi-sample biometrics or models of multiple biometric traits called multi-source or multi-modal biometrics [7, 8]. Multi-biometric techniques are capable of diminishing several of the constraints of uni-biometric methods, as the diverse biometric sources generally compensate for the inborn constraints of the supplementary sources [9]. The significant objective of multi-biometrics, is to scale up the accuracy of detection over a specific technique by synthesizing the outcomes of several traits, sensors or algorithms. In multi-modal biometrics, selection of true modality is a demanding task in the identification of a person. The personal recognition by means of hand vein has, nowadays, attracted the ever-increasing enthusiasm of investigators because of its superior biometric traits over other modalities [10, 11, 12].

In multimodal biometric recognition systems, the fusion [13] of various traits can be carried out at different levels, such as sensor level, feature level, [14, 15] score level [16, 17] and decision level [18]. Score level blending is the most desirable factor in multimodal biometric systems in view of the fact that matching scores encompass ample data to distinguish between true and false cases and they are comparatively easy to access. This facilitates trouble-free amalgamation of data gathered from specific modalities, by means of score level synthesis, making it realistic and reasonable. The amalgamation procedure can be carried out using the approaches like transformation-based, classifier-based and density-based at the score level. Scores from multiple matchers are treated as a feature vector and a classifier is constructed to discriminate genuine and impostor scores in a classifier based fusion. The density-based approach [19] has the advantage that it directly achieves optimal performance at any desired operating point, provided the score densities are estimated accurately.

Herein, the proposed technique, employs Shearlet transformation and Scale-invariant Feature Transformation (SIFT) to obtain feature coefficients from hand vein images and a maximum likelihood ratio-based fusion, at the score level. The rest of the paper is organized as follows: A brief review of research related to the

proposed technique is presented in Section 2. The proposed technique is presented in Section 3. The detailed experimental results and discussion are given in Section 4 and the concluding remarks are provided at the end.

# 2 Background of the Research Work

A multimodal biometric recognition based on finger images was discussed in [20].The local binary patterns algorithm was used to extract features and match for the fingerprints and finger veins, while the oriented FAST and Rotated BRIEF algorithm was applied for knuckle prints. Finally, score-level fusion was performed on the matching results from the above three finger biometrics. In [21], a multimodal biometric recognition using iris and facial images was discussed. Contourlet transform and two dimensional principal component analyses were used here to extract the iris features and the facial features respectively, and a feature vector was formed by the combination of the iris and facial features. A fixed random matrix was used here, to improve the recognition efficiency. The excellence of sum rule-based and support vector machine based score level fusion were deliberated in [22]. Three biometric characteristics were taken into consideration, for the purpose of the investigation, including fingerprints, faces, and finger veins. Authors have formulated vigorous normalization systems, such as, diminution of high-scores effect normalization which has been obtained from min-max normalization technique.

Shubhangi Sapkal [23] have smartly developed a novel level synthesis method to fine tune population coverage and scale down spoofing, that possess the quality of flexibility to error forbearance of various mono-modal biometric techniques. This method was intended as an access management system necessitating the enhanced safety in permitting access to significant data. Poh, N. [24] have proficiently proposed a technique to mechanically validate the uniqueness of an individual by way of biometrics, employing face and fingerprint. They established the fact that the superior performing fusion algorithms were those that make the utmost use of the mechanically mined biometric trait quality calculated with a view to detect the utmost possible biometric mechanism from which the query biometric data was obtained.

Recently, personal identification by means of hand veins has attained progressively more research attention. Hand vein recognition has been deliberated in [25], based on the statistical processing of the hand vein patterns. The hand vein database has been collected under practical conditions and subjected to go through different procedures. Here a combination of geometric and appearance-based techniques were used for feature extraction and distance metrics for recognition. A bank of Gabor filters was used for feature extraction [26] of finger vein images, from which finger vein codes were generated using the local and global features of

the vein. Nearest cosine classifier was used for classification and fusion was carried out at the decision level. In [27], Palm-dorsal vein recognition method based on histogram of local Gabor phase XOR Pattern has been suggested. They have used chi-square distance measure for recognition. The modified two directional two-dimensional linear discriminant analysis was proposed by Lee [28] for personal verification approach using palm vein patterns. A minimum distance classifier was used here for identification.

From these discussions, it is clear that the vein based biometrics provide improved security and it cannot be easily spoofed or falsified. Hence, in our proposed system, we have used hand vein biometrics, such as, finger vein, palm vein and dorsal vein of the hand, for multimodal biometric recognition.

# 3 Proposed Method

Biometric recognition has the advantage of being reliable and secure for authentication purposes. Multi-biometrics uses more than one trait and overcomes the draw backs of using single modality. It improves security, but choosing the right modality and techniques involved, is of utmost importance. Finger vein, palm vein and dorsal vein of the hand are used here as the biometric modalities. The proposed technique employs Shearlet transform [29] and SIFT [30] for feature extraction and the fusion is carried out using maximum likelihood ratio-based technique. The block diagram of the proposed technique is given in Fig. 1.

## 3.1 Formation of feature set

In order to generate the required feature set, the input vein images of hands are transformed using Shearlet transform and Scale-invariant feature transform.

### 3.1.1 Shearlet Transformation

The Shearlet is an affine system with a single generating mother Shearlet function parameterized by a scaling, shear, and translation parameter. The Shearlet transform thus overcomes this drawback while retaining most aspects of the mathematical framework of wavelets. Shearlet has the properties that the associated system forms an affine system and the transform can be regarded as matrix coefficients of a unitary representation of a special group.

Shearlet can be represented as:

$$\Psi_{a,s,t}(x) = a^{-3/4}\Psi((D_{a,s}{}^{-1}(x-t)), \quad where \quad D_{a,s} = [a, -a^{1/2}s; 0, a^{1/2}] \qquad (1)$$

Where, $a$ is a scaling parameter, $s$ is a shear parameter and $t$ is a translation parameter.

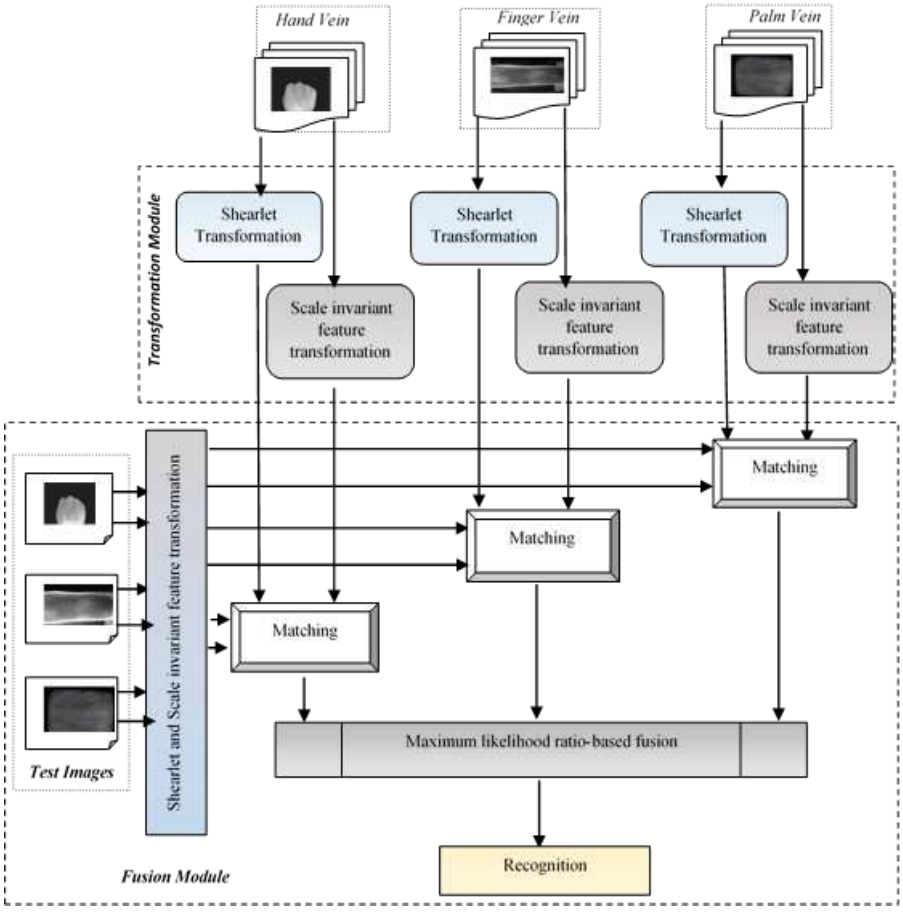



Figure 1
Block diagram of the proposed recognition technique

The mother Shearlet function $\Psi$ is defined as:

$$\Psi(\xi_1, \xi_2) = \Psi_1(\xi_1)\Psi_2\left(\frac{\xi_2}{\xi_1}\right) \tag{2}$$

Where, $\Psi_1$ is a wavelet and $\Psi_2$ is a bump function.

The associated continuous Shearlet transform depends on the scaling, shear and translation parameters and is defined by

$$ST\left[f(a,s,t)\right] = \left\langle f, \Psi_{a,s,t} \right\rangle \tag{3}$$

This transform can also be regarded as matrix coefficients of the unitary representation:

$$\sigma(a,s,t)(\Psi)(x) = \Psi_{a,s,t}(x) = a^{-3/4}\Psi\left(D_{a,s}{}^{-1}(x-t)\right) \tag{4}$$

Let the input hand vein, finger vein and palm vein images be represented as

$$H = \{h_1, h_2, h_3, ..., h_n\}, D = \{d_1, d_2, d_3, ..., d_n\} \ \ and \\ P = \{p_1, p_2, p_3, ..., p_n\} \tag{5}$$

Here, $n$ is the number of images. Then the Shearlet transformed hand vein, finger vein and palm vein images can be represented by

$$SH = \{sh_1, sh_2, sh_3, ..., sh_n\},\ SD = \{sd_1, sd_2, sd_3, ...sd_n\}\ \ and \\ SP = \{sp_1, sp_2, sp_3, ...sp_n\} \tag{6}$$

### 3.1.2 Scale-Invariant Feature Transformation

Scale-invariant feature transform (SIFT) is a step by step procedure to identify and delineate the local features in images. There are mainly four steps involved in SIFT algorithm namely scale space extrema detection, key point localization, orientation assignment, key point descriptor and key point matching.

Scale-space extrema detection is employed to detect larger corners using larger windows. Here, Laplacian of Gaussian (LoG) is found for the image with various scaling parameter values ($\theta$). LoG acts as a blob detector which detects blobs in various sizes. As LoG is a little costly, SIFT algorithm uses Difference of Gaussians which is an approximation of LoG. Difference of Gaussian (DoG) is obtained as the difference of Gaussian blurring of an image with two different $\theta$ and $k\theta$. Once this DoG is found, images are searched for local extrema over scale and space.

Once potential key points locations are found, they have to be refined to get more accurate results in the key point localization. If the intensity at this extrema is less than a threshold value, it is rejected so as to eliminate low contrast key points. Similarly, edge threshold is used to remove low edge key points. These processes would rise to retain of strong interest points.

The SIFT transformed images of hand vein, finger vein and palm vein images can be represented by

$$FH = \{fh_1, fh_2, fh_3, ..., fh_n\},\ FD = \{fd_1, fd_2, fd_3, ...fd_n\}\ \ and \\ FP = \{fp_1, fp_2, fp_3, ...fp_n\} \tag{7}$$

## 3.2 Maximum Likelihood Ratio-based Fusion and Recognition

The feature set obtained from the transformations are matched and fused for recognition. Initially, images of hand vein, finger vein and palm vein are transformed using the above transforms and stored in the database. The database (DB) would consist of both the Shearlet and the SIFT transformed images:

$$DB = \{SH, SD, SP, FH, FD, FP\} \tag{8}$$

Where,

$$SH = \{sh_1, sh_2, sh_3, ..., sh_n\}, SD = \{sd_1, sd_2, sd_3, ..., sd_n\},$$
$$SP = \{sp_1, sp_2, sp_3, ..., sp_n\}, FH = \{fh_1, fh_2, fh_3, ..., fh_n\}$$
$$FD = \{fd_1, fd_2, fd_3, ..., fd_n\} \text{ and } FP = \{fp_1, fp_2, fp_3, ..., fp_n\}$$

Subsequently, the matching score is compared to the test input images

and those in the database. The test images are represented by:

$$T = \{h_{test}, d_{test}, p_{test}\} \tag{9}$$

The transformed images are represented as:

$$T^* = \{sh_{test}, sd_{test}, sp_{test}, fh_{test}, fd_{test}, fp_{test}\} \tag{10}$$

These features are compared with those in the database using the Euclidean distance measure which gives the matching score. If the Euclidean distance between the test image and that of the data base image is less than the threshold set, then the images are said to be in a matched condition. Suppose the Euclidean distance between images $Im_1$ *and* $Im_2$ is represented by *dis* and the threshold set is represented as $d_{thr}$, then,

*if* $dis < d_{thr}$ *, then* $Im_1$ *and* $Im_2$ are in matched condition (11)

Each of the images are matched with the database image and then matching scores of all images are discovered. After the matching process, a fusion process is carried out with the use of maximum likelihood ratio based fusion. It is basically a density based score fusion which requires explicit estimation of genuine and impostor match score densities. Each comparison of the test image with that in the database would yield a matching score and for the fusion process each of the matching scores are taken into consideration. It is supposed that total of *m* matching is carried out, to get the matching score vector given by

$$SV = \{sv_1, sv_2, ..., sv_m\} \tag{12}$$

Let the conditional joint densities of the M match scores for the genuine and impostor classes be represented by *Ge(sv) and Ip(sv)*. The respective class of genuine or impostor is assigned by analyzing the score vector and Gaussian mixture model is employed for finding out the score densities.

Let the M-variant Gaussian density be represented as $g(sv;\mu,c)$, where $\mu$ is the mean vector and $c$ is the covariance matrix.

The estimates obtained can be represented as $\hat{Ge}(sv)$ *and* $\hat{Ip}(sv)$. Then the maximum likelihood ratio is given by:

$$L(sv) = \frac{\hat{Ge}(sv)}{\hat{Ip}(sv)} \tag{13}$$

The matching score vector $sv$ is assigned to genuine class, if $L(sv) >$ decision threshold or assigned to impostor class, if $L(sv) \leq$ decision threshold. The decision threshold is determined based on the specified False acceptance rate (FAR).

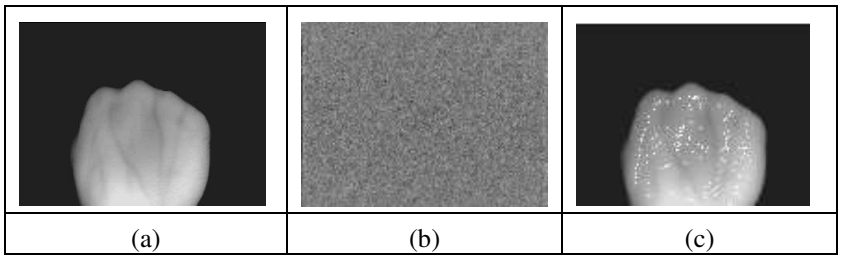
# 4 Results and Discussion

The experimental results of the proposed method for vein based biometric recognition are discussed here. The evaluation metrics employed here are accuracy, FAR (False Acceptance Rate) and FRR (False Rejection Rate).

FAR is the measure of the likelihood that a biometric security system will incorrectly accept an access attempt by an unauthorized user. FAR typically is stated as the ratio of the number of false acceptances divided by the number of identification attempts.

FRR is the measure of the likelihood that the biometric security system will incorrectly reject an access attempt by an authorized user. FRR typically is stated as the ratio of the number of false rejections divided by the number of identification attempts.

The database utilized for our experimentation is taken from the standard data bases [31, 32, 33] for hand vein, palm vein and finger vein images. The experimental results at various stages of hand vein, palm vein and finger vein images are given in Figure 2.



(a) (b) (c)

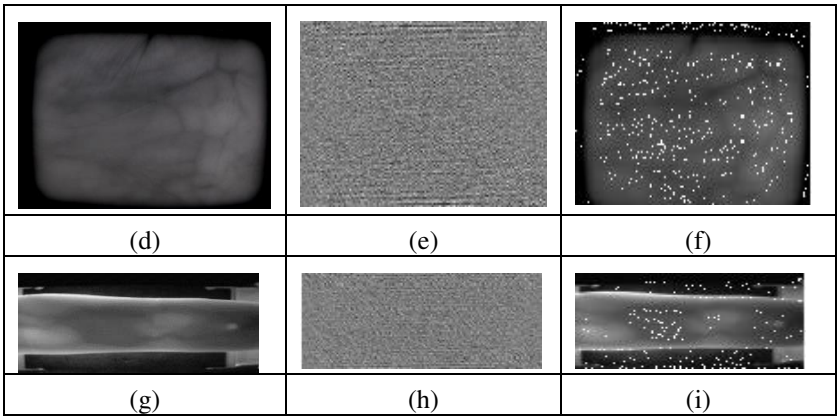



Figure 2
Images at various stages (a) Input dorsal hand vein image (b) Shearlet transformed dorsal vein image (c) Scale-invariant feature transformed dorsal vein image (d) Input palm vein image (e) Shearlet transformed palm vein image (f) Scale-invariant feature transformed palm vein image (g) Input finger vein image (h) Shearlet transformed finger vein image (i) Scale-invariant feature transformed finger vein image

## 4.1 Performance Analysis

The performance of the proposed technique is evaluated using metrics of FAR, FRR and accuracy. The values are taken for each modality of hand vein images such as dorsal vein, finger vein and palm vein. The evaluation metrics is also taken for feature one (when Shearlet transform is only performed), feature two (when SIFT transform is only performed) and product (when the fusion is carried out using product rule). The analysis is carried out and they are illustrated in the graphs separately.

The performance graph of FAR, FRR, FAR-FRR, ROC and Accuracy is plotted in Figures 3 to 7. Here, initially the experimental analysis is carried out with the hand vein features, which are extracted by using the shearlet and SIFT transforms separately. Then the analysis is performed with the vein features (extracted by using both shearlet and SIFT) fused at the score level using product rule. Finally, the above results are compared with our proposed technique. In Figure 3, we have obtained FAR with good performance for which, the threshold is above 0.3. From the graph shown in Figure 4, FRR of the proposed method attained lower value compared with the product rule based fusion at the score level. FRR-FAR curve obtained are shown in Figure 5 and the lines meet at a score threshold of 0.4 when FRR-FAR values were below 0.1 and this indicates improved performance of the proposed system.

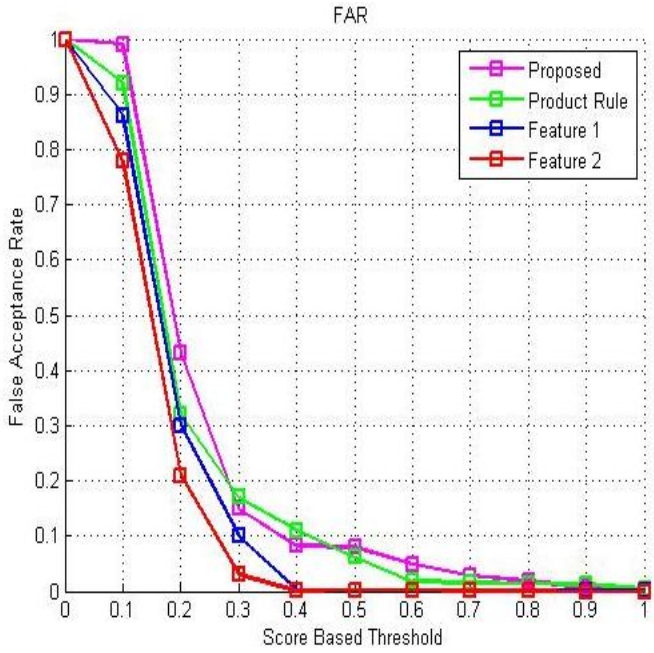



Figure 3
Plot of FAR

The accuracy graph (Figure 6) is plotted by varying score based threshold, in the range of zero to one (0-1). For each threshold value set, we can see that our proposed technique has achieved a higher accuracy. Here, we found that our proposed technique provides low FAR, FRR and high accuracy of 94%. Also, we observed that from the Receiver operating characteristics (Figure 7), the proposed technique achieved a low Equal error rate (EER) of 0.04, in comparison with other techniques.

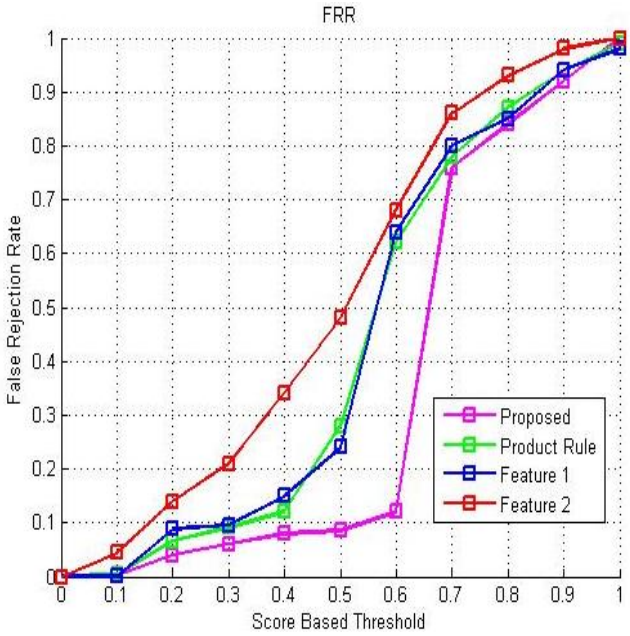



Figure 4
Plot of FRR

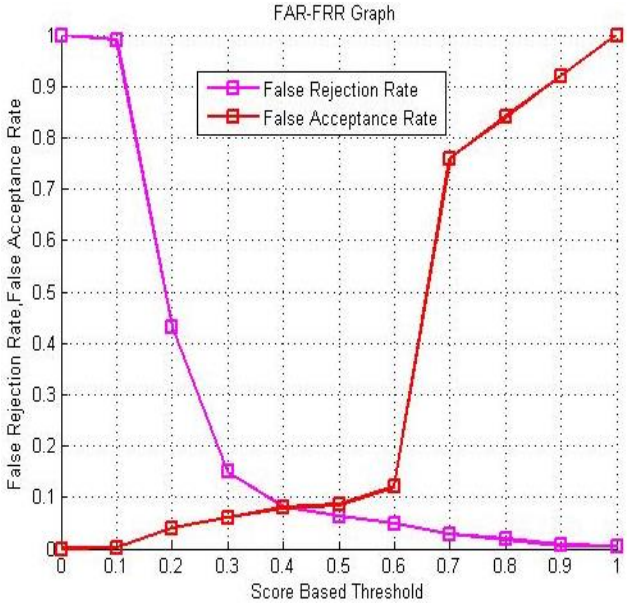



Figure 5
Plot of FAR-FRR

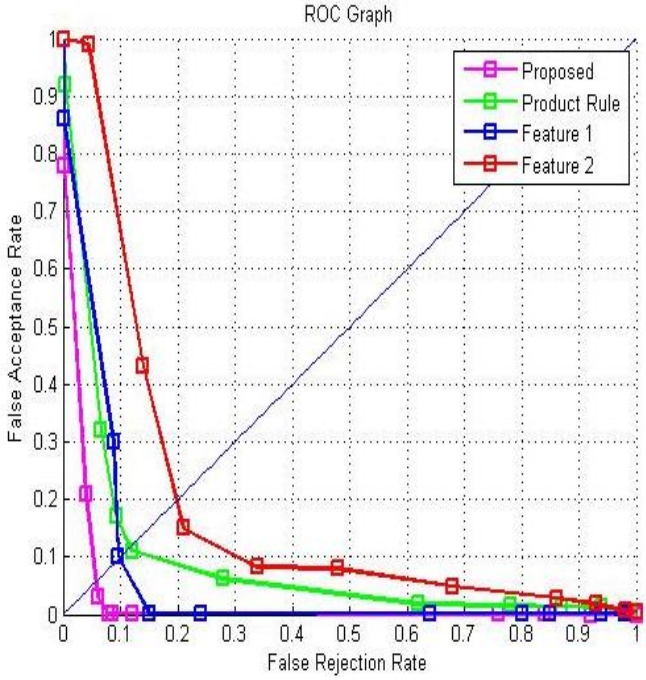



Figure 6
Plot of ROC

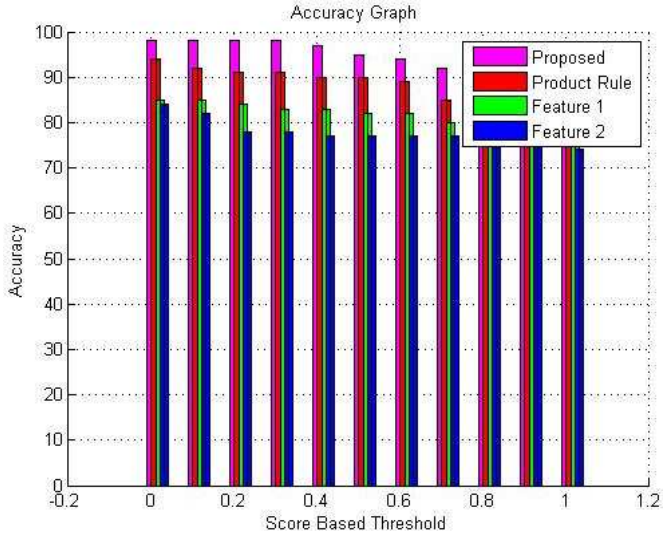



Figure 7
Plot of Accuracy

Table 1
Comparative analysis of related work on vein based biometric recognition

| Author | Biometric features | Methodology | Imaging | Data base | Performance |
|---|---|---|---|---|---|
| Lin, C.L., and Fan, K.C [34] | Palm-dorsa vein | Multi resolution analysis and combination | Thermal imaging | 32 users | FAR = 1.5<br>FRR = 3.5<br>EER = 3.75 |
| Kumar, A., and Prathyusha, K.V [35] | Hand vein and knuckle shape | Matching vein triangulation and shape features | Near IR imaging | 100 users | FAR = 1.14<br>FRR = 1.14 |
| Raghavendra, R., Imran, M., Rao, A., and Kumar, G.H [36] | Hand vein and palm print images | Log Gabor transform | Near IR imaging | 50 users | FAR = 7.4,<br>FRR = 4.8 |
| | | Non-standard mask | | | FAR = 2.8,<br>FRR = 1.4 |
| Ferrer, M.A., Morales, A., Travieso, C.M. and Alonso, J.B [37] | Hand geometry, palm and finger textures, dorsal hand vein | Simple Sum rule | Near IR imaging | 50 users | FAR = 0.01,<br>FRR = 0.01,<br>EER = 0.01 |
| Yuksel, A., and Akarun, L [38] | Hand vein | ICA 1, ICA2, LEM and NMF | Near IR imaging | 100 users | EER =5.4,<br>7.24,7.64<br>and 9.17 |
| Proposed method | Dorsal hand vein, palm vein and finger vein | Maximum likelihood | Near IR imaging | 100 users | FAR = 0.04,<br>FRR = 0.05,<br>EER = 0.04 |

## 4.2 Comparative Analysis

The comparative analysis of the related work on vein based authentication is given in Table 1. All the results presented in this table are in terms of Equal error rate (EER), FAR and FRR. EER is defined as a point, at which, FAR is equal to FRR. The lower the values of EER, the better the performance of the system, but it varies according to the imaging technique, type of biometric trait, methodologies used for feature extraction, method and type of fusion of these features and number of users in the data base.

**Conclusions**

In this work, a novel multi-modal biometric recognition was presented, using only the vein biometric traits present in the hand. Here, we used the finger vein, palm

vein and dorsal vein biometric traits of the hand. Shearlet Transform and Shift Invariant Feature Transform (SIFT) were used for feature extraction and the fusion was carried out using the maximum likelihood ratio method at the score level. All the biometric traits used here are the vascular patterns of the body so it cannot be spoofed and the trait, provides a better security, compared to other kinds of multi-modal biometric systems. The proposed system is evaluated using the parameters of accuracy, FAR and FRR. The experimental results show improved performance of the system, when compared with other current techniques, indicating the effectiveness of our proposed technique. This technique provides the highest accuracy of 94% and minimum FAR and FRR.

**References**

[1] Jain, A. K., Ross, A., and Prabhakar, S: An Introduction to Biometric Recognition, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 1, 2004, pp. 4-20

[2] Nicolaie Popescu-Bodorin, Valentina E. Balas: Learning Iris Biometric Digital Identities for Secure Authentication: A Neural-Evolutionary Perspective Pioneering Intelligent Iris Identification, Recent Advances in Intelligent Engineering Systems, Series: Studies in Computational Intelligence, Springer, Vol. 378, 2012, pp. 409-434

[3] Nicolaie Popescu-Bodorin, Valentina Emilia Balas: Fuzzy Membership, Possibility, Probability and Negation in Biometrics, Acta Polytechnica Hungarica, Vol. 11, No. 4, 2014, pp. 79-100

[4] Akhtar, Z, Fumera, G, Marcialis, G. L, Roli, F: Evaluation of Multimodal Biometric Score Fusion Rules under Spoof Attacks, Proceedings of IEEE International Conference on Biometrics, Arlington, 2012, pp. 402-407

[5] Jain, A. K., Ross, A., and Pankanti, S: Biometrics: A Tool for Information Security, IEEE Transactions on Information Forensics and Security, Vol. 1, No. 2, 2006, pp. 125-143

[6] Jain, A. K., Nandakumar, K., Lu, X., and Park,U: Integrating Faces, Fingerprints and Soft Biometric Traits for User Recognition, Proceedings of ECCV International Workshop on Biometric Authentication (BioAW) Vol. 3087, 2004, pp. 259-269

[7] Ross, A. and Jain, A. K: Multimodal Biometrics: An Overview, Proceedings of 12th European Signal Processing Conference, 2004, pp. 1221-1224

[8] Zhu Le-qing, Zhang San-yuan: Multimodal Biometric Identification System Based on Finger Geometry, Knuckle Print and Palm Print, Pattern Recognition Letters, Vol. 31, 2010, pp. 1641-1649

[9] Yadav, S. S., Gothwal, J. K., and Singh, R: Multimodal Biometric Authentication System: Challenges and Solutions, Global Journal of Computer Science and Technology, Vol. 11, No. 16, 2011, pp. 57-61

[10] Ali Mohsin Al-juboori, Wei Bu, Xingqian Wu and Qiushi Zhao: Palm Vein Verification using Gabor Filter, International Journal of Computer science issues, Vol. 10, No. 1, 2013, pp. 678-684

[11] Yi-Bo Zhang, Qin Li, Jane You and Prabir Bhattacharya: Palm Vein Extraction and Matching for Personal Authentication, Advances in Visual Information Systems, Vol. 4781, 2007, pp. 154-164

[12] Yang, J. and Shi, Y: Finger-Vein ROI Localization and Vein Ridge Enhancement, Pattern Recognition Letters, Vol. 33, No. 12, 2012, pp. 1569-1579

[13] Ross, A, and Jain, A, K: Information Fusion in Biometrics, Pattern Recognition Letters, Vol. 24, No. 24, 2003, pp. 2115-2125

[14] Chin Y. J, Ong T. S, Teoh A. B. J, Goh K. O. M: Integrated Biometrics Template Protection Technique Based on Finger Print and Palm Print Feature-Level Fusion, Journal of Information fusion, 2014, Vol. 18, pp. 161-174

[15] Yang J, Zhang X: Feature-Level Fusion of Fingerprint and Finger-Vein for Personal Identification. Pattern Recognition Letters, 2012, Vol. 33, No. 5, pp. 623-628

[16] Hanmandlu M, Grover J, Gureja A, and Gupta HM: Score Level Fusion of Multimodal Biometrics using Triangular Norms. Pattern recognition letters, 2011, Vol. 32, No. 14, pp. 1843-1850

[17] Bharathi S, Sudhakar R, Balas V. E: Biometric Recognition Using Fuzzy Score Level Fusion. International Journal of Advanced Intelligence paradigms, 2014, Vol. 6, No. 2, pp. 81-94

[18] Ibrahim A Saleh, Laheeb M Alzoubiady: Decision Level Fusion of Iris and Signature Biometrics for Personal Identification using Ant Colony Optimization, International Journal of Engineering and Innovative Technology, 2014, Vol. 3,No. 11, pp. 35-42

[19] Karthik Nandakumar, Yi Chen, Sarat C. Dass and Anil K. Jain: Likelihood Ratio-based Biometric Score Fusion, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No. 2, 2008, pp. 342-348

[20] Wenxiong Kang, Xiaopeng Chen, Qiuxia Wub: The Biometric Recognition on Contactless Multi-Spectrum Finger Images, Infrared Physics & Technology, Vol. 68, 2015, pp. 19-27

[21] Ying Xu, FeiLuo, Yi-Kui Zhai and Jun-Ying Gan: Joint Iris and Facial Recognition Based on Feature Fusion and Biomimetic Pattern Recognition,

IEEE Conference on Wavelet Analysis and Pattern Recognition (ICWAPR) 2013, pp. 202-208

[22] Shi-Jinn Horng, Yuan-Hsin Chen, Ray-Shine Run, Rong-Jian Chen, Jui-Lin Lai and Sentosal, K. O: An Improved Score Level Fusion in Multimodal Biometric Systems, Parallel and Distributed Computing, Applications and Technologies, 2009, pp. 239-246

[23] Shubhangi Sapkal: Data Level Fusion for Multi Biometric System using Face and Finger, International Journal of Advanced Research of Computer Science and Electronics Engineering, Vol. 1, No. 2, 2012

[24] Poh, N: Benchmarking Quality-Dependent and Cost-Sensitive Score-Level Multimodal Biometric Fusion Algorithms, Information Forensics and Security, Vol. 4, No. 4, 2009, pp. 849-866

[25] Yuksel, A., Akarun, L. and Sankur, B: Hand Vein Biometry Based on Geometry and Appearance Methods, IET Computer Vision, Special Issue: Future Trends in Biometric Processing, Vol. 5, No. 6, 2011, pp. 398-406

[26] Jinfeng Yang, Yihua Shi, Jinli Yang: Personal Identification Based on Finger-Vein Features, Computers in Human Behavior, Vol. 27, 2011, pp. 1565-1570

[27] Meng, Z. and Gu, X: Palm-Dorsal Vein Recognition Method Based on Histogram of Local Gabor Phase XOR Pattern with Second Identification, Journal of Signal Processing Systems, Vol. 73, No. 1, 2013, pp. 101-107

[28] Lee, Y, P: Palm Vein Recognition Based on a Modified $(2D)^2$ LDA, International journal of Signal Image and Video Processing, 2013, pp. 101-114

[29] Zhiyong Zeng, Jianqiang Hu: Face Recognition Based on Shearlets and Principle Component Analysis, IEEE International conference on Intelligent Networking and collaborative systems, Xian, 2013, pp. 697-701

[30] Alense-Femandez, F, Tome-Gonzalez P, Ruiz-Albacete, V, Ortega Gavein J: Iris Recognition Based on SIFT Features, IEEE International conference on Biometrics, Identity and Security BIDS, 2009, pp. 1-8

[31] Badawi A. M: Hand Vein Database, at Systems and Biomedical Engineering, Cairo University, Cairo, Egypt, 2005

[32] Palm Vein Data Base. Available at http://biometrics.put.poznan.pl/vein-dataset

[33] Finger Vein Database. Available at http: //mla.sdu.edu.cn/sdumla-hmt.html

[34] Lin, C. L, and Fan, K. C: Biometric Verification Using Thermal Images of Palm-Dorsa Vein Patterns, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 2, 2004, pp. 199-213

[35] Kumar, A, and Prathyusha, K. V: Personal Authentication Using Hand Vein Triangulation and Knuckle Shape, IEEE Transactions on Image processing, Vol. 18, No. 9, 2009, pp. 2127-2136

[36] Raghavendra, R, Mohammad Imran, Ashok Rao and Kumar, G. H: Multi Modal Biometrics: Analysis of Hand Vein and Palm Print Combination Used for Personal Verification, Proceedings of IEEE International conference on Emerging trends in Engineering and Technology, 2010, pp. 526-530

[37] Ferrer, M. A, Morales, A., Travieso, C. M., and Alonso, J. B: Combining Hand Biometric Traits for Personal Identification, Proceedings of IEEE International Carnahan Conference on Security Technology, 2009, pp. 155-159

[38] Yuksel, A., and Akarun, L: Biometric Identification through Hand Vein Patterns, Proceedings of International Workshop on Emerging techniques and challenges for Hand based biometrics, 2010, pp. 1-6

# Developing a Complex Decision-Making Framework for Evaluating the Energy-Efficiency of Residential Property Investments

**András Szűts**

Óbuda University
Bécsi út 96/b, H-1034 Budapest, Hungary
szuts.andras@phd.uni-obuda.hu

*Abstract: Currently, energy efficiency and energy security are among the most important energy issues that face Hungary and the European Union. Since households are responsible for one third of the primary energy consumption, we believe that significant results can be achieved in this area. According to our research, the typical design methods currently utilized are not entirely suitable to achieve the stated energy goals, since the goal of these methods is not to search and design an energetically optimal solution, but rather to comply with target values stipulated by law. Therefore, in order to increase the energy efficiency of the design of the buildings, we have developed a complex decision support model that enables in the conceptual design phase to effectively evaluate the emerging alternatives, taking into consideration both convenience and economy, as well as various subjective elements. This article presents a methodology developed through examples of four typical family homes in Hungary.*

*Keywords: analytic hierarchy process; energy; households; property investments; net present value*

# 1 Introduction

Currently, one of the main areas of research in the European Union is focused on reducing the energy consumption of households and making them more energy efficient, given that households are responsible for one third of the primary energy consumed. However, this topic requires a complex, systematic approach, whereas nowadays the general construction energy strategy is mostly limited to the technical parameters, which could lead in erroneous directions. In the course of my research, I have found that the consumption habits and lifestyles of consumers influence household energy consumption, at least as much as the technical solutions applied. In addition, existing rules in different countries can have a decisive impact on the amount of energy consumed, as well as on the energy

efficiency of the building. It is essential that we make consumers (owners and renters) interested in the implementation/construction of energy-conscious designs by delivering an optimal, or at least adequate, solution for them from the point of view of aesthetics and comfort, including finances. This objective requires designers to adopt new and unconventional attitudes and methods.

During the application of current building design practices, designers rely on their experience of energy-conscious methodologies as they try to meet legal regulations and requirements valid at the given location and time. In recent years, these technical requirements have become considerably stricter, and they are expected to become stricter still in the coming years, as starting from 2020 only those buildings can be constructed in the EU whose energy consumption is near zero. [1] As it was already mentioned, this fact alone may not be enough for the EU to achieve the energy efficiency targets it has set, so it is also necessary to examine the design phase of construction projects, in particular the energy-efficiency criteria.

Looking at the design process, we can see that the decisions made in the early, conceptual stages of the design process can significantly influence the future energy consumption of the building. Therefore, in the course of our research, we developed a decision support model that can efficiently solve the ranking of the various options arising in the conceptual design phase. The decision support model and the possibilities of its application will be described in the following paragraphs. Other decisive factors that can significantly affect the energy consumption of a building or household are the habits of the consumers and the factors associated with their living conditions. Therefore, a model was developed to help to manage the uncertainties arising from such factors, as well as to determine the future energy consumption of the building in the early design phase accurately and personalized. During the creation of the methodology, financial investment considerations and the reduction of greenhouse gas emissions were considered.

# 2 The Key Factors of Households Energy Consumption

In order to estimate the energy consumption of a building, or the change in efficiency due to a change in it or its parts, as well as to achieve the savings impact desired, it is necessary to understand the factors affecting the energy consumption of the household or building. In both Hungary and the rest of the European Union, a mix of general primary energy sources (e.g. wood, gas) and secondary energy sources (e.g. electricity) are used to satisfy the various energy needs of buildings, which can be grouped accordingly [2] [4]:

- Energy needed to ensure comfortable air space in living areas (e.g. heating, cooling, ventilation)
- Energy needed to produce domestic hot water
- Energy needed to operate lighting and electrical appliances (e.g. refrigerators, cooking appliances, etc.)

According to G. Swan et al. [2], the energy needed to ensure adequate air comfort depends on the technical characteristics and solutions of the building as well as on climatic conditions of the area; thus the relevant energy use can be calculated utilizing technical and meteorological data. However, as for the other two factors, Swan [2] shows that demographic factors, the number of household members, consumer behavior, and the energy efficiency of the appliances also influence the level of energy consumption. Therefore, in order to determine the energy use for hot water, appliances and lighting, he recommends the use of statistical data [2]. Professionals should not rely exclusively on technical and other data for determining energy use for these factors, independently from considering the user, since even if the building is nearly a zero-energy building assuming an average consumer, it will not achieve near zero energy use if the consumer does not use the building according to the principles of energy efficiency. In case of residential buildings, it is very difficult to influence consumer behavior by way of technical solutions. Residents generally have the need for what is usually expressed as "manual control", and thus intervene in the operation of the building; one need only think of such a simple thing as opening the windows to let in some air, for example. Consequently, energy efficiency cannot be guaranteed even though the best available technologies are used during the design and the construction of the building.

In order to take into consideration the issues cited above, the proposed decision support model consists of two major parts. One part deals with the uncertainties resulting from consumer behavior, tailoring the projected energy consumption of a household to a specified consumer, in order to eliminate inaccuracies resulting from projections based on the average consumer. The other part of the model considers technical factors that are independent from consumer habits, as it ranks the alternatives arising during the conceptual design and estimates the consumption of the building, or part of the building, based on the consumption rate of the specific household, which had been calculated with the first part of the model.

# 3 The Design of the Complete Decision Support System

As explained in the previous paragraph, in order to estimate accurately the energy consumption of a household, it is extremely important to consider the effects of consumer habits and living conditions. Towards this end, an artificial neural network based interference system was created that allows us to determine the average annual primary energy consumption of a household in accordance with the relevant statistical data. These factors, according to Aydinalp et al. [3] and supplemented with considerations of domestic peculiarities, are the following [4]:

- Size of the settlement
- Type of housing arrangement (owner, renter)
- Number of people living in the household
- Number of children living in the household
- Number of household members actively employed
- Highest education level of the head of the household
- Household income decile
- Size of the actual living space of the property
- Type of housing (house, flat)

To train the artificial neural network a starting database was needed, which was created based on the 2011 Census and the income and consumption statistics of the same year. [4] One benefit of the developed system is that the database can be expanded and modified, so the neural network model can be retrained with the up-to-date data in the future, thereby increasing the accuracy of the conclusions. To prepare and train the neural network based model, the Matlab software package, "Neural Network Fitting Tool" was used.

For the other part of the model an AHP (Analytic Hierarchy Process) based model similar to the models traditionally used in multi-criterion decision problems was created, which allows us to rank the alternatives emerging in the early planning stages. A complete evaluation is essential, as the owner or tenant, or in some cases the investor, will only be satisfied with the investment if it meets his/her needs on all points, since no matter how energy efficient a specific concept is, if it is unacceptable from the point of view of aesthetics or comfort, then that particular version will not be built. Accordingly, the main criteria of the AHP system are as follows [5]:

- Conceptual efficiency
- The performance of the building structures

- The energy efficiency of the equipment
- Architectural value, design
- Interior comfort
- Lifetime and reliability

The complete AHP hierarchy with the criteria and the sub-criteria along with their relationships are shown in Figure 1. During the evaluation, the classic AHP method was applied with the use of the basic Saaty scale and the eigenvector method. [5]

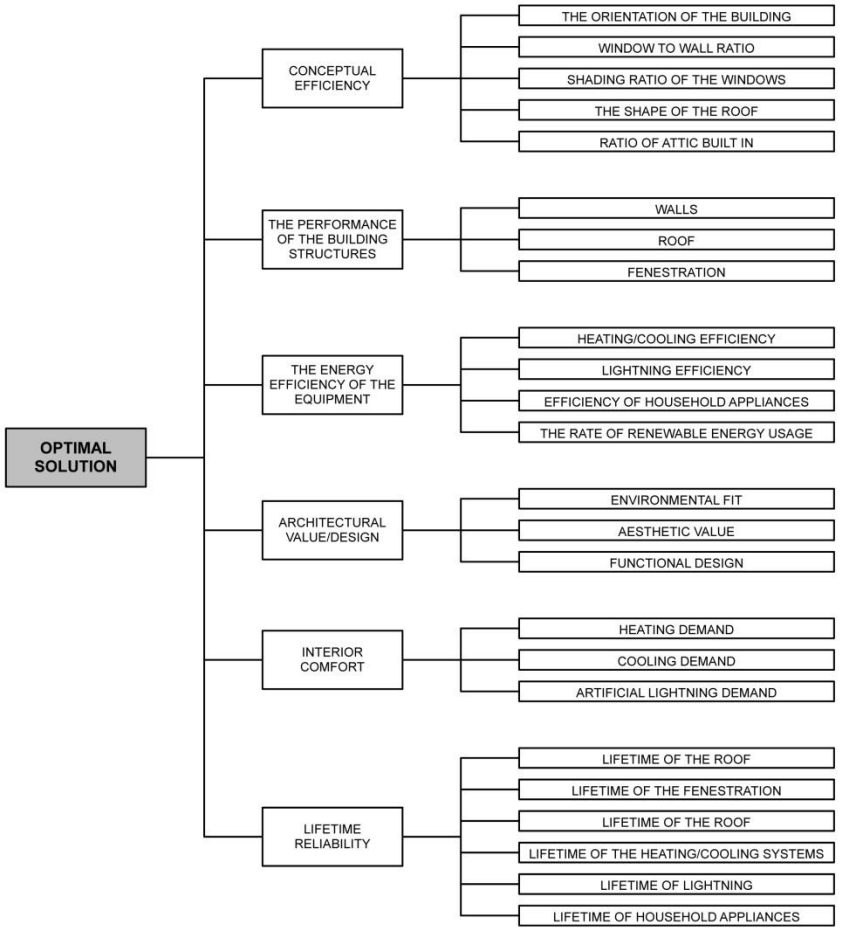



Figure 1
The AHP system with main- and sub-criteria [5]

Based on studies conducted during the research, it was observed that the use of the traditional AHP system is a relatively labor-intensive system. [5] Therefore, an improvement of the model was made as a Hybrid Fuzzy AHP system. This can significantly reduce labor-intensity and evaluation time. [5] The substance of the applied fuzzy AHP system is that instead of the paired comparison matrixes, we use fuzzy inference systems on the input side, which define the adequate SPIs (Site Performance Indexes) for specific factors and alternatives. [5] With the help of the SPIs, the evaluation continues within the established AHP structure, but by this point, the main and sub-criteria have already undergone a preliminary and generalized evaluation and received their assigned weights, which also accelerates the ranking process. [5] To prepare the model, the Matlab software package "Simulink" module was used. As discussed above, some economic calculation will be needed in order to complete the developed decision support system. The flowchart of the whole decision method along with the new modules is shown on Figure 2. It contains three major parts related to the economic investigation:

- Energy consumption estimate system
- Cost estimate system
- Return of investment calculation

The above listed systems and the results obtained with the completed method on the four fictional family houses will be described in the following paragraphs.

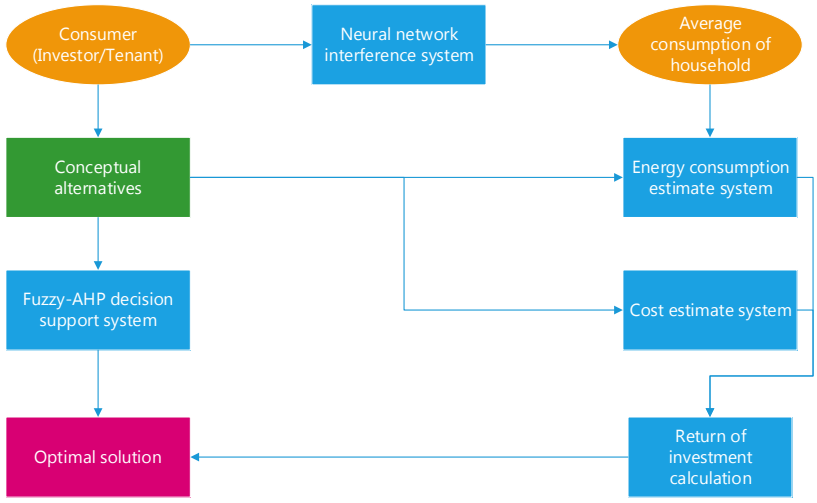



Figure 2
The flowchart of the developed decision method

# 4 Estimating Investment Costs

For a complex analysis of energy efficient investments, it is essential to examine the economics of the investments as well. However, in the studied design phase, it is difficult to determine the cost of the investment, since certain issues that can significantly affect the costs have not been definitively decided, at this point. In addition, the designer cannot be expected to write up a detailed budget for all the incurred alternatives, since this would require much more detailed plans than the conceptual design, which in turn would make the work economically unfeasible for the designer. To overcome this problem, a fuzzy inference based investment cost estimate system was developed and harmonized with the AHP based decision support models, which enables estimation for the cost of a building (or part of a building) by knowing some basic data. The reason for a system separate from the AHP model, is the extremely high price sensitivity of the investors/occupants; since in the AHP system the price has an unrealistically high weight, it would distort the results. In the future it would clearly be worth examining various financing structures, as they significantly affect the profitability of an investment, but that analysis is beyond the scope of this paper.

The backbone of the developed cost estimate system is made from the criteria and sub-criteria set in the AHP model, omitting those that do not at all or do not significantly affect the cost of the investment. During the development of the system, the key factor was matching the input data structure to the decision support models described above in order to ensure interoperability between the AHP based decision support model and the cost estimate system. The model was prepared in the "Simulink" module of the Matlab software package. The structure of the model is shown in Figure 3. Due to the design of the system and the variables used, the system can be fully integrated into the system already described and also systems made with the "Simulink" package.

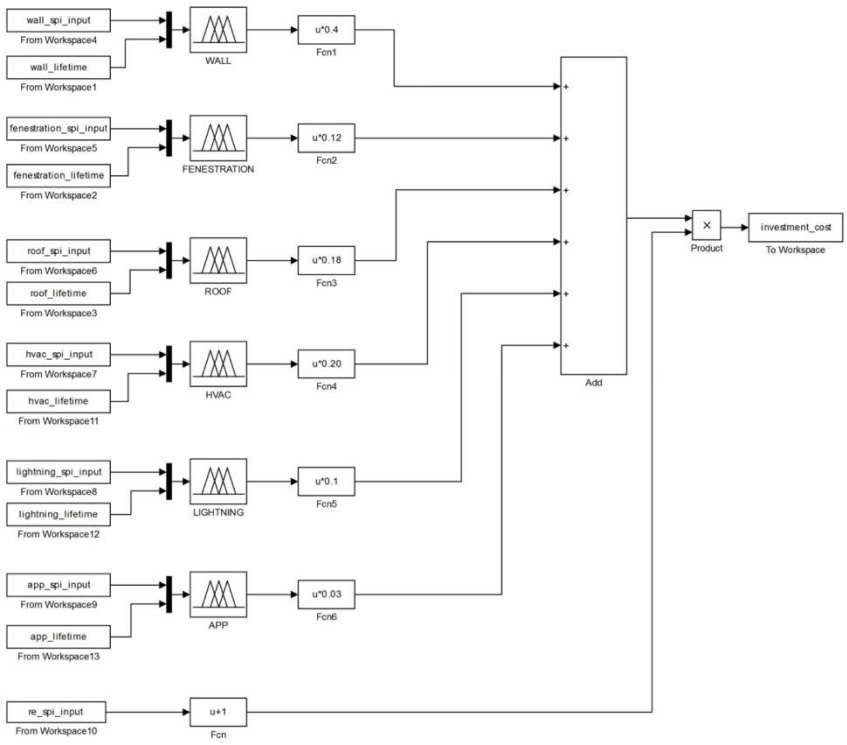



Figure 3
The structure of the cost estimate system in Matlab

To determine the cost of the structures and equipment, fuzzy controllers were used, with two input variables and one output variable. The fuzzy controller determines the relevant cost factor of the associated device or equipment by its performance and lifetime. In the event that the exact input values are unknown, with the help of the fuzzy controllers, linguistic variables [6, 7] could be used, which in the case of performance, for example, may be as follows: available worst, below average, average, above average, the best available. The membership functions of a particular fuzzy controller and its controlling surface is shown in Figure 4.

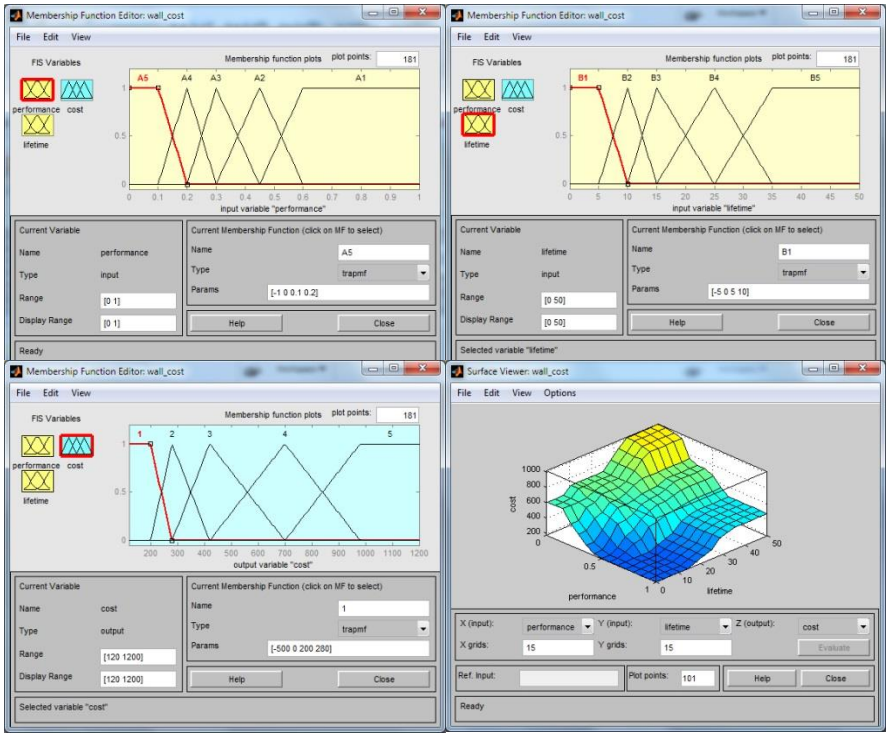



Figure 4

The membership functions of the wall-cost fuzzy controller and its controlling surface

For the calculation, the costs of different structures and equipment based on current construction statistics in Hungary were used. Obviously, the data that not yet known in the early design phase are not taken into account, so the actual investment cost may vary from the estimates; however, it is still possible to compare the alternatives that arise, particularly for the demonstration of differences in the quality of energy-conscious investments.

# 5 Determining the Costs of Operation

The determination of the operating costs is worth a more detailed investigation than the estimation of the investment cost. Based on our research, the expenses occurring over the lifetime of a building or a household can be divided into two independent groups, which would be the following:

- The cost of energy use
- The cost of replacement investments, due to the known lifetime of structures and equipment

The starting point for the calculation of the costs of energy use is the previously developed artificial neural network based inference system, based upon which we know the average annual primary energy consumption for a particular customer (household). In the case of a particular household, to evaluate the effects of different energy efficient investments, a high number of simulations with hypothetical reference buildings and households are required. Using these results, we developed a system to estimate the energy savings, which helps to compare the alternatives that arise during the conceptual design phase. It needs to be emphasized that the system does not replace the simulation methods, since a specific building has several parameters that are not known, or only approximately known; therefore it is not possible to calculate the precise energy consumption, which is required at later stages of the design. But the basic objective, the ranking of different versions by energy consumption, had been achieved. For the simulations needed to create the system, the Open Studio software system from NREL (the National Renewable Energy Laboratory) was used.

The model for estimating energy savings is also designed to be interoperable with the AHP based decision support system, as well as based on design variables already introduced. The system consists of fuzzy controllers that can help to estimate the impact of an investment on energy consumption utilizing the same factors that are used in the AHP system. This module was also prepared in the "Simulink" module of the Matlab software package, as was the above-described models. The schematic diagram of the model is shown in Figure 5.

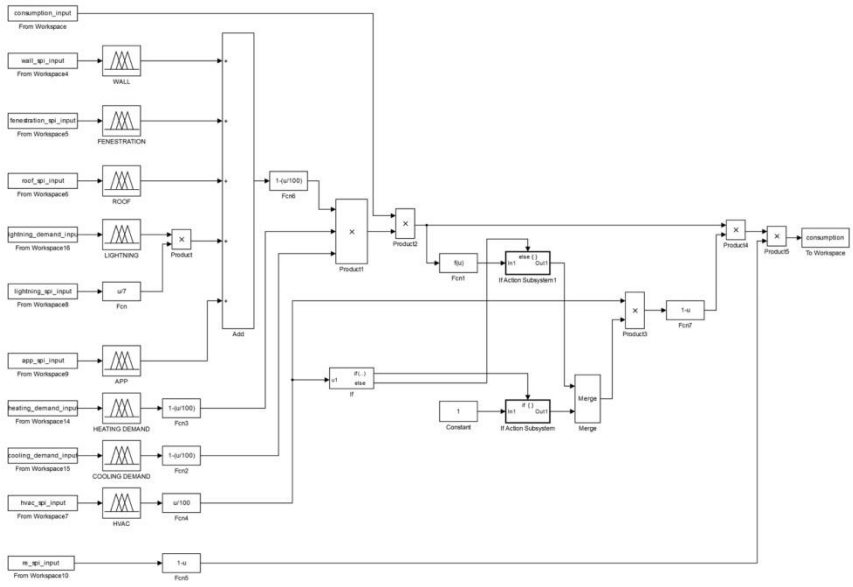



Figure 5

The schematic diagram of the energy consumption estimate system

The costs of the replacement investments occurring over the lifetime of the building or household were considered along with the useful life of the structures and equipment used in the investment. During the research, the design longevity was set at 50 years, while the replacement investments are cyclically repeating, according to the lifetime of the structure or equipment.

# 6 Results

The developed system was tested with 4 typical family houses in Hungary. In order to have an objective comparison, the four selected buildings have the same floor space and replaceable design. The reference household was a private "owned" urban household. There were four inhabitants, two of which are children. Of the four, 1 person is working and has finished secondary school, and they belong to the 6th decile income group. During the evaluation, I determined the average annual energy consumption of the reference household, ranked the alternatives with the developed fuzzy AHP hybrid model, with the estimating module I determined their investment cost and estimated their energy consumption and the reduction of resulting greenhouse gas emissions. Based on the results of all 4 versions, I made net present value calculations for the design longevity and calculated the internal rate of return for the initial investment. For verification, I made a detailed energy simulation of each alternative; however, in our previous researches we have already established that with the developed AHP and Fuzzy AHP based model, the energy ranking of the alternatives is solvable with adequate accuracy in the early design phase. [5] The design data of the selected 4 variants is in Table 1.

Table 1
The data of the alternatives

| | Alternatives | | | | |
|---|---|---|---|---|---|
| **Criteria** | **A1** | **A2** | **A3** | **A4** | |
| **Conceptual efficiency** | | | | | |
| The orientation of the building | 90 | 90 | 90 | 90 | *[°] N - 0°* |
| Window to wall ratio | 0.18 | 0.18 | 0.25 | 0.43 | |
| Shading ratio of the windows | 0.00 | 0.00 | 0.52 | 1.00 | |
| The shape of the roof | 38.83 | 38.83 | 30.00 | 5.00 | *[°]* |
| The ratio attic building in | 0.00 | 0.00 | 0.00 | 0.00 | |
| **The performance of the building structures** | **A1** | **A2** | **A3** | **A4** | |
| Performance of the walls | 1.79 | 0.39 | 0.24 | 0.14 | $[W/m^2K]$ |
| Performance of the fenestration | 2.20 | 1.60 | 1.60 | 0.80 | $[W/m^2K]$ |
| Performance of the roof | 1.69 | 1.69 | 0.74 | 0.12 | $[W/m^2K]$ |

| **The energy efficiency of the equipment** | **A1** | **A2** | **A3** | **A4** | |
|---|---|---|---|---|---|
| Heating/cooling efficiency | 0.80 | 0.80 | 1.00 | 1.50 | |
| Lighting efficiency | 10.50 | 10.50 | 7.00 | 5.25 | *[W/m$^2$]* |
| Efficiency of household appliances | 3.45 | 3.45 | 2.30 | 1.80 | *[W/m$^2$]* |
| The rate of renewable energy usage | 0.00 | 0.00 | 0.10 | 0.20 | |
| **Architectural value/design** | **A1** | **A2** | **A3** | **A4** | |
| Environmental fit | 0.20 | 0.20 | 0.50 | 1.00 | |
| Aesthetic value | 0.20 | 0.20 | 0.30 | 0.80 | |
| Functional design | 0.50 | 0.50 | 0.40 | 0.60 | |
| **Interior comfort** | **A1** | **A2** | **A3** | **A4** | |
| Heating demand | 22 | 21 | 20 | 21 | *[°]* |
| Cooling demand | - | - | 25 | 24 | *[°]* |
| Artificial lighting demand | 13 | 13 | 11 | 8 | *[hour]* |

| **Lifetime/reliability** | **A1** | **A2** | **A3** | **A4** | |
|---|---|---|---|---|---|
| Lifetime of the walls | 20 | 20 | 25 | 50 | *[year]* |
| Lifetime of the fenestration | 10 | 10 | 20 | 25 | *[year]* |
| Lifetime of the roof | 10 | 20 | 20 | 25 | *[year]* |
| Lifetime of the heating/cooling system | 5 | 10 | 10 | 20 | *[year]* |
| Lifetime of lighting | 5 | 10 | 10 | 10 | *[year]* |
| Lifetime of household appliances | 5 | 10 | 10 | 10 | *[year]* |

For the calculations of investment profitability, I set up different scenarios with varying energy prices, since considering the 50-year longevity, it is difficult to rely only on one data series. Accordingly, I relied on both domestic and international literature as well as my estimates, with regards to expected energy prices in the next 50 years. The 4 different scenarios are shown in the following charts.

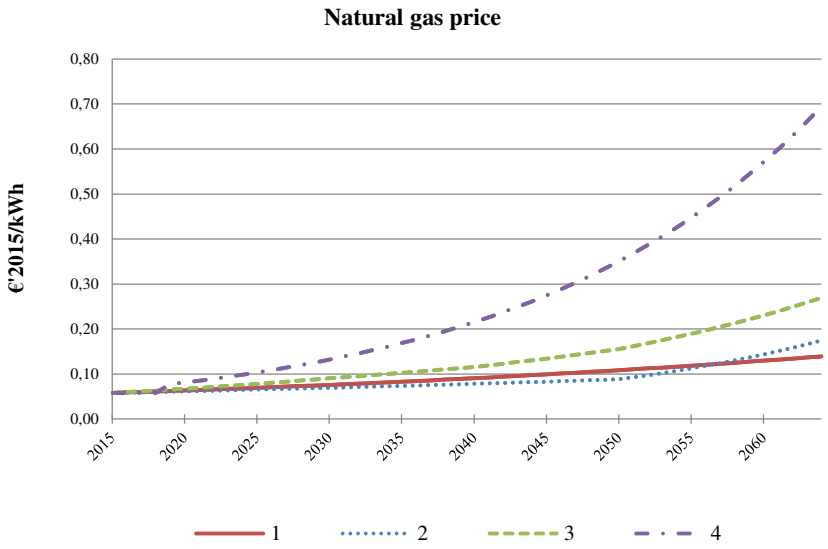



Figure 6
Price of natural gas based on the 4 scenarios

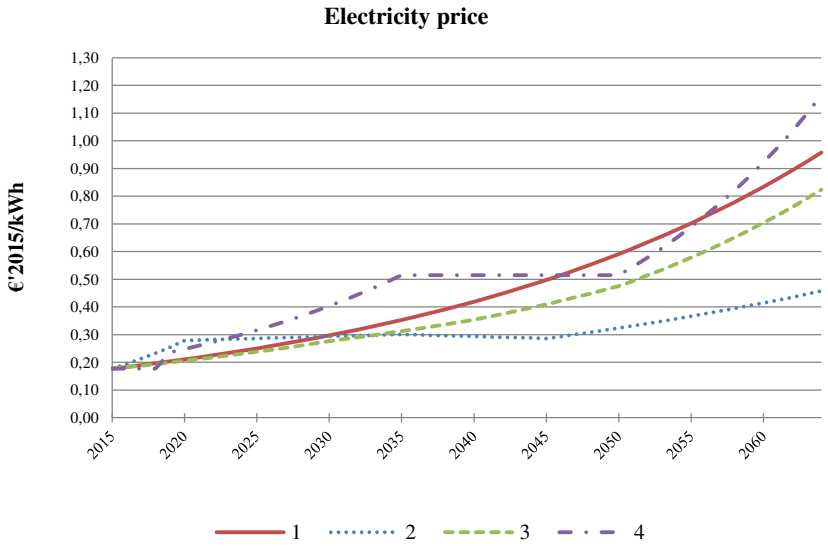



Figure 7
Price of electricity based on the 4 scenarios

The first scenario was based on research by the Hungarian Energy Club, according to which energy prices are expected to increase with relatively stability, meaning a gas price increase at an average annual rate of 4.3%, and an electricity price increase of an average of 5%. [8] In the second scenario, which is based on the European Commission's forecast, gas prices steadily rise at a slightly lower rate compared to the previous scenario until 2050, after which it will rise at a higher rate. However, electricity prices will increase considerably until 2020, and then stagnate until 2035, followed by a moderate increase. [9] The intermittent rise and stagnation in electricity prices is a consequence of the life cycle of investments made at different times in the electricity sector. In the third version, which was prepared according to the forecast of the European Renewable Energy Council, the rise in energy prices is moderate until 2030, higher until 2040, and even higher after 2040. [10] According to my own estimates, which is the fourth version, in 2018, a sharp correction in energy prices will be felt, resulting in an immediate 33% increase compared to the period of stagnation before. In the subsequent period, gas prices will steadily rise with an average annual rate of 5%, while electricity prices increase fractionally, similarly to the third scenario. To perform the economic calculations, a supposed 2.5% cost of capital, during the entire life cycle, and the funds for the investment are considered as available. It would require a separate analysis to forecast the various economic indicators, which is beyond the scope of our research; however, the model provides the opportunity to the change the capital cost as well as to consider other forms of financing. The results of the economic calculation of the 4 selected family houses are in Table 2.

Table 2
The results

| | Alternatives | | | | |
|---|---|---|---|---|---|
| **Decision support system** | **A1** | **A2** | **A3** | **A4** | |
| Fuzzy-AHP results | 0.40 | 0.52 | 0.62 | 0.83 | *[0-1 point]* |
| **Costs** | **A1** | **A2** | **A3** | **A4** | |
| Net square footage | 85 | 85 | 85 | 85 | *[$m^2$]* |
| Cost per square footage | 816 | 1603 | 2203 | 3735 | *[€/$m^2$]* |
| Investment cost | 69371 | 136274 | 187274 | 317516 | *[€]* |
| **Energy consumption** | **A1** | **A2** | **A3** | **A4** | |
| Simulation results | 509 | 368 | 294 | 170 | *[kWh/$m^2$*year]* |
| Estimated results | 494 | 332 | 263 | 159 | *[kWh/$m^2$*year]* |
| Deviation | 3.07 | 10.96 | 11.71 | 6.86 | *[%]* |
| GHG emission | 10.54 | 7.09 | 5.61 | 3.39 | *[t/year]* |
| GHG emission reduction | - | 32.79 | 46.76 | 67.81 | *[%]* |
| Not appropriate comfortable | 1845 | 2336 | 1229 | 437 | *[h]* |
| **Economical results** | **A1** | **A2** | **A3** | **A4** | |
| **Scenario 1** | | | | | |
| Lifetime net present value | 438877 | 379374 | 389999 | 427334 | *[€]* |

| | | | | | |
|---|---|---|---|---|---|
| NPV - deviation from A1 | - | 59503 | 48878 | 11543 | *[€]* |
| Internal rate of return | - | 6.07 | 4.30 | 2.72 | *[%]* |
| **Scenario 2** | | | | | |
| Lifetime net present value | 394193 | 349343 | 366210 | 412952 | *[€]* |
| NPV - deviation from A1 | - | 44849 | 27983 | -18759 | *[€]* |
| Internal rate of return | - | 5.70 | 3.84 | 2.25 | *[%]* |
| **Scenario 3** | | | | | |
| Lifetime net present value | 492655 | 415516 | 418630 | 444643 | *[€]* |
| NPV - deviation from A1 | - | 77138 | 74025 | 48012 | *[€]* |
| Internal rate of return | - | 6.45 | 4.75 | 3.19 | *[%]* |
| **Scenario 4** | | | | | |
| Lifetime net present value | 711682 | 562717 | 535237 | 515139 | *[€]* |
| NPV - deviation from A1 | - | 148965 | 176444 | 196542 | *[€]* |
| Internal rate of return | - | 7.71 | 6.16 | 4.58 | *[%]* |

*based on 310 HUF/€ exchange rate

## Conclusions

Looking at the results, it can be observed that the fuzzy inference system developed to estimate the energy savings estimated the energy consumption of a particular investment with a margin of error under 15%. This is sufficiently accurate for the examined planning stage, especially considering my main goal was to rank the alternatives and to select the best version for further development. The ranking made by the fuzzy AHP system harmonizes well with the energy results, but as to the point of comfort, it leads to disparity. In this respect, the A2 version performs even worse than the A1 alternative, even though it is the favored solution considering energy consumption, due to a simple and/or poorly executed energy refurbishment. We can see many examples of the above these days, e.g. different air condition problems such as mold, too low or too high humidity and discomfort in isolated parts of the building.

In the results of the economic calculations, it can be seen that, apart from scenario 2, in virtually every case, the A1 variant had the highest expenses, calculated for longevity. Consequently, we can conclude that, in almost all cases, the energy-conscious investments should be implemented, since if we look at the whole life cycle, in most cases they are economically worthwhile as well. This fact could impact the perspective of the consumer, because if the investor sells the property, or if for some reason it comes under another person's use, then the investor does not realize these economic benefits. Nevertheless, the social and environmental impacts of the project are not negligible. Considering the complex evaluation, it can be concluded that the fuzzy AHP system herein, established an adequate ranking via the economic criteria as well, since if we look at the evaluation of the A2 and A3 versions, it shows that the expenses for the entire lifespan are similar, but the A3 alternative performs better as regards both energy and comfort.

With regard to greenhouse gas emissions, obviously, the lowest-energy building achieves the best results. If we take this fact into consideration, the nearly 3% internal rate of return is not a bad result, since the amount of the emitted gases decreases by 2/3. However, the future social and environmental impacts of this cannot be measured objectively via economic indicators. In the view of the Author, however, taking a long-term perspective, the issue of greenhouse gas emissions is a more important factor than different economic factors, especially with regard to fulfilling the objectives set by the EU. Based on these results, the developed method is suitable for evaluating the alternatives available in the early stages of planning and for choosing the best alternatives for future plans. However, concerning economic issues, it needs to be mentioned, that a significantly better return on the investment can be achieved with adequate incentive and/or funding programs.

**Acknowledgement**

This work was supported by the Doctoral School of Applied Informatics and Applied Mathematics of Óbuda University.

**References**

[1] Directive 2010/31/EU of the European Parliament and of the Council: on the energy performance of buildings, 19.05.2010

[2] Lukas G. Swan, V. Ismet Ugursal, Ian Beausoleil-Morrison: Occupant-related Household Energy Consumption in Canada: Estimation Using a Bottom-Up Neural-Network Technique, Energy and Buildings, Vol. 43, 2011, pp. 326-337

[3] Merih Aydinalp, V. Ismet Ugursal, Alan S. Fung: Modeling of the Space and Domestic Hot-Water Heating Energy-Consumption in the Residential Sector Using Neural Networks, Applied Energy, Vol. 79, 2004, pp. 159-178

[4] András Szűts, István Krómer: Estimating Hungarian Household Energy Consumption Using Artificial Neural Networks, Acta Polytechnica Hungarica, Vol. 11, No. 4, pp. 155-168, 2014

[5] András Szűts, István Krómer: Developing a Fuzzy Analytic Hierarchy Process for Choosing the Energetically Optimal Solution at the Early Design Phase of a Building, Acta Polytechnica Hungarica, Vol. 12, No. 3, pp. 25-39, 2015

[6] Thomas L. Saaty: Decision Making with the Analytic Hierarchy Process, Int. J. Services Sciences, Vol. 1, No. 1, pp. 83-98, 2008

[7] Alessio Ishizaka, Nam Hoang Nguyen: Calibrated Fuzzy AHP for Current Bank Account Selection, Expert Systems with Applications, Vol. 40, pp. 3775-3783, 2013

[8] Épületek energetikai követelményeinek költségoptimalizált szintjének megállapítását megalapozó számítások, Energiaklub, 2013

[9] EU Energy, Transport and GHG Emissions Trends to 2050 Reference Scenario 2013, European Comission, 2013

[10] RE-thinking 2050: A 100% Renewable Energy Vision for the European Union, European Renewable Energy Council, 2010

[11] Directive 2009/28/EC of the European Parliament and of the Council: on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC, 23.04.2009

[12] Directive 2010/31/EU of the European Parliament and of the Council: on the energy performance of buildings, 19.05.2010

[13] Statute 7/2006. TMN (V.24): Regarding the Definition of Building Energy Characteristics, 2006 (in Hungarian: Az épületek energetikai jellemzőinek a meghatározásáról)

[14] Christina J. Hopfe, Godfried L. M. Augenbroe, Jan L. M. Hensen: Multi-Criteria Decision Making under Uncertainty in Building Performance Assessment, Building and Environment, Vol. 69, pp. 81-90, 2013

[15] https://www.openstudio.net/, download time: 20.05.2014

[16] http://apps1.eere.energy.gov/buildings/energyplus/, download time: 20.05.2014

[17] Johnny K. W. Wong, Heng Li: Application of the Analytic Hierarchy Process (AHP) in Multi-Criteria Analysis of the Selection of Intelligent Building Systems, Building and Environment, Vol. 43, pp. 108-125, 2008

[18] Guozhong Zheng, Youyin Jing, Hongxia Huang, Guohua Shi, Xutao Zhang: Developing a Fuzzy Analytic Hierarchical Process Model for Building Energy Conservation Assessment, Renewable Energy, Vol. 35, pp. 78-87, 2010

[19] R. Ramanathan: A Note on the Use of the Analytic Hierarchy Process for Environmental Impact Assessment, Journal of Environmental Management, Vol. 63, pp. 27-35, 2001

[20] Yang Yu-lan, Tai Hui-xin, Shi Tao: Weighting Indicators of Building Energy Efficiency Assessment Taking Account of Experts' Priority, J. Cent. South Univ., Vol. 19, pp. 803-808, 2012

[21] Yu-Ting Lai, Wei-Chih Wang, Han-Hsiang Wang: AHP- and Simulation-based Budget Determination Procedure for Public Building Construction Projects, Automation in Construction, Vol. 17, pp. 623-632, 2008

[22] R. Judkoff, D. Wortman, B. O'Doherty, and J. Burch: A Methodology for Validating Building Energy Analysis Simulations, National Renewable Energy Laboratory, Technical Report NREL/TP-550-42059, 2008

[23] Mehmet Kabaka, Erkan Köseb, O˘guzhan Kırılmaza, Serhat Burmao˘gluca: A Fuzzy Multi-Criteria Decision Making Approach to Assess Building Energy Performance, Energy and Buildings, Vol. 72, pp. 382-389, 2014

[24] Building Information Centre Ltd.: Construction Cost Estimation Guide, 2011 (in Hungarian: Építésügyi Tájékoztatási Központ Kft.: Építőipari költségbecslési segédlet, 2011)