Preface

Fuzzy theory and more generally, soft computing technologies have been widely used in different applications, as proven by their impeccable track record for the past decades. Theoretical and practical developments are increasingly gaining ground in scientific publications. One such prominent venue for these achievements is the FSTA, the International Conference on Fuzzy Set Theory and Applications, which is organized biannually and offers a comprehensive review of the significant results in this field of the previous two years. This series of conferences is also supported by EUSFLAT (European Society of Fuzzy Logic and Technology) whose aim is to promote the cooperation between the European fuzzy research centers. The Hungarian Fuzzy Association is an important partner in this venture, as it regularly supports international conferences organized in Hungary as well as outside the borders of Hungary (SISY, SAMI, SACI, and CINTI) where publications on intelligent systems and model based on soft computing are featured prominently.

The present Special Issue of selected papers is of the result the application-oriented publications presented at the 10th FSTA (Tenth Conference on Fuzzy Set Theory and Applications, Liptovský Mikuláš, 2010, http://www.math.sk/fsta2010/) and other conferences supported by Hungarian Fuzzy Association.

This selection offers a cross section of soft computing technologies in terms of their widespread application. The aim of this publication is to offer scientists and researchers an insight into current directions of development and imminent trends of research in this field.

Márta Takács

Special Issue Guest Editor

Aggregation Functions and Personal Utility Functions in General Insurance

Jana Špirková

Department of Quantitative Methods and Information Systems,
Faculty of Economics, Matej Bel University
Tajovského 10, 975 90 Banská Bystrica, Slovakia, jana.spirkova@umb.sk

Pavol Kráľ

Department of Quantitative Methods and Information Systems,
Faculty of Economics, Matej Bel University
Tajovského 10, 975 90 Banská Bystrica, Slovakia

Institute of Mathematics and Computer Science, UMB and MÚ SAV
Ďumbierska 1, 974 11 Banská Bystrica, Slovakia, pavol.kral@umb.sk

Abstract: The modeling of a utility function's forms is a very interesting part of modern decision making theory. We apply a basic concept of the personal utility theory on determination of minimal net and maximal gross annual premium in general insurance. We introduce specific values of gross annual premium on the basis of a personal utility function, which is determined empirically by a short personal interview. Moreover, we introduce a new approach to the creation of a personal utility function by a fictive game and an aggregation of specific values by mixture operators.

Keywords: Utility function; Expected utility; Mixture operator; General Insurance

1 Introduction

This paper was mainly inspired by the books Modern Actuarial Risk Theory [3] and Actuarial models – The Mathematics of Insurance [13]. The authors of the above-mentioned books assume utility functions as linear utility $u(w) = w$, quadratic utility $u(w) = -(a-w)^2$, power utility $u(w) = w^c$, etc. Lapin in [5] describes and explains an application of the utility function in decision making in a really interesting way. In this book also the generation of the utility function

using information extracted from a personal interview is explained. A modern theoretical approach to the utility function is also described by Norstad in [8].

We can find a very interesting discussion about utility functions in [4]. An alternative approach to the determination of a utility function on the basis of the aggregation of specific utility values can be found in [18].

However, in real life people do not behave according to the theoretical utility functions. It is a psychological problem rather than a mathematical one. The seriousness and also the uncertainty of a respondent's answers depend on the situation, on the form of the asked questions, on the time which the respondents have, and on a lot of other psychological and social factors. In our paper we introduce the possibility of determining a personal utility function on the basis of a personal interview with virtual money.

Moreover, we recall and apply one type of aggregation operators [2], the so-called mixture operators – M_g, the generalized mixture operators – M_{g}, and the specially ordered generalized mixture operators – M_{g}' on the aggregation of so-called risk neutral points, see [6-7], [9-11], [14-16].

This paper is organized as follows: in Section 2 we recall the basic properties of utility functions and their applications in general insurance. In Section 3 we also recall mixture operators and their properties, namely the sufficient conditions of their non-decreasing-ness. In Section 4 we describe the personal utility function of our respondent who took part in our short interview. Using this function we calculate the maximal gross premium for a general insurance policy. In this section we also describe an alternative approach, where a personal utility function is determined through the result of a fictive game and theoretical utility functions. The resulting utility function is then used for the computation of the maximal gross premium. Moreover, we evaluate the minimal net annual premium by means of the theoretical utility function for the insurer. Finally, some conclusions and indications as to our next investigation about the mentioned topic are included.

2 Utility Functions

Individuals can have very different approaches to risk. A personal utility function can be used as a basis for describing them. In general, we can identify three basic personalities with respect to risk. The risk-averse individual, who accepts favorable gambles only, a risk seeker, or in other words risk-loving individual, who pays a premium for the privilege of participation in a gamble, and the risk-neutral individual, who considers the face value of money to be its true worth. Throughout most of their life people are typically risk averse. Only gambles with high expected payoff will be attractive to them. The risk-averse individual's

marginal utility diminishes as the benefits increase, so that the risk-averse individual's utility function exhibits a decreasing positive slope as the level of monetary payoff becomes higher. Such a function is concave, see Figure 1.

The behavior of a risk-loving individual is opposite. The risk-loving individual prefers some gambles with negative expected monetary payoffs. Their marginal utility increases. Each additional euro provides a disproportionately greater sense of well-being. Thus, the slope of the risk-loving individual's utility function increases as the monetary change improves. This function is convex (see Fig. 2). The utility function for a risk-neutral individual is a straight line. The utility is equal to the utility of expected value. Risk-neutral individuals buy no casually insurance since the premium charge is greater than the expected loss. Risk-neutral behavior is typical for persons who are enormously wealthy.

Of course, a lot of people may be risk averse and risk loving at the same time, depending on the range of monetary values being considered, which can be illustrated using the behavior of the personal utility function of our respondent.

2.1 The Personal Utility Function

The fundamental proposition of the modern approach to utility is the possibility to obtain a numerical expression for individual preferences. As people usually have different approaches to risk, two persons faced with an identical decision may actually prefer different courses of action. In this section we will discuss utility as an alternative expression of payoff that reflects personal approaches.

Suppose that our respondent owns capital w, and that he values wealth by the utility function u. The next Theorem 1, or in other words Jensen`s inequality, describes the properties of the utility function and its expected value [3], (see also Figure 2). It can be written as follows.

Theorem 1 [3] (Jensen's inequality)

If $u(x)$ is a convex function and X is a random variable, then the expected utility is greater or equals to a utility value

$$E[u(X)] \geq u(E[X]) \tag{1}$$

with an equality if and only if $u(x)$ is linear with respect to X or $\mathrm{var}(X) = 0$.

From Jensen`s inequality and Figure 1 it follows that for a concave utility function it holds

$$E[u(w - X)] \leq u(E[w - X]) = u(w - E[X]). \tag{2}$$

In this case the decision maker is called risk averse. He prefers to pay a fixed amount $E[X]$ instead of a risk amount X.

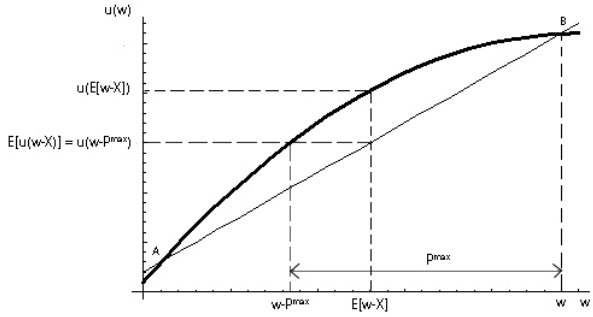

Figure 1
Concave utility function - risk averse approach

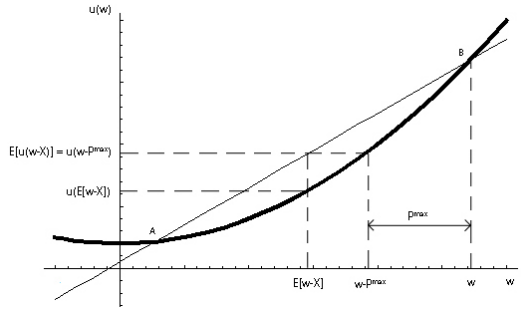

Figure 2
Convex utility function - risk loving approach

In the next part we illustrate whether to buy insurance or not by evaluating an individual's decision. Now suppose that our respondent has two alternatives, to buy insurance or not. Assume he is insured against a loss X for a premium P.

If he is insured, this means a certain alternative. This decision gives us the utility value $u(w-P)$.

If he is not insured, this means an uncertain alternative. In this case the expected utility is $E[u(w-X)]$.

From Jensen's inequality (2) we get

$$E[u(w-X)] \le u(E[w-X]) = u(w-E[X]) = u(w-P). \tag{3}$$

Since a utility function u is a non-decreasing continuous function, this is equivalent to $P \le P^{\max}$, where $P^{\max}$ denotes the maximum premium to be paid.

This so-called zero utility premium is a solution of the following utility equilibrium equation

$$E[u(w-X)] = u\left(w - P^{\max}\right). \tag{4}$$

The difference $\left(w - P^{\max}\right)$ is also called the certainty equivalent - CE. In [3] the certainty equivalent is defined as follows.

Definition 1 The certainty equivalent is that payoff amount that the decision maker would be willing to receive in exchange for undergoing the actual uncertainty, taking into account its benefits and risks.

Remark 1 We recall that the expected utility is calculated by means of the well-known formula

$$E[u(X)] = \sum_{i=1}^{n} u(x_i) \cdot p_i, \tag{5}$$

where $X = (x_1, x_2, ..., x_n)$ is a vector of the possible alternatives and p_i, for $i = 1,2,...,n$, are respective probabilities.

Expected utilities can be calculated as function values of a linear function, which is assigned uniquely by points A and B, where point A represents the worst outcome and B the best outcome.

Remark 2 [5] When possible monetary outcomes fall into the decision maker's range of risk averse, the following properties hold (see Figure 1):

1) Expected payoffs $EP = E[w-X]$ are greater than their counterpart certainty equivalent $CE = w - P^{\max}$.

2) Expected utilities $E[u(w-X)]$ will be less than the utility of the respective expected monetary payoff $u\left(w - P^{\max}\right)$.

3) Risk premiums $RP = EP - CE$ are positive.

If possible monetary outcomes fall into the decision maker's range of risk loving, the following properties hold (see Figure 2):

1) Expected payoffs $EP = E[w-X]$ are less than their counterpart certainty equivalent $CE = w - P^{\max}$.

2) Expected utilities $E[u(w-X)]$ will be greater than the utility of the respective expected monetary payoff $u\left(w - P^{\max}\right)$.

3) Risk premiums $RP = EP - CE$ are negative.

The insurer with a utility function U and capital W, with insurance of loss X for a premium P must satisfy the inequality

$$E[U(W+P-X)] \geq U(W), \tag{6}$$

and hence for the minimal accepted premium $P^{\min}$

$$U(W) = E\left[U\left(W+P^{\min}-X\right)\right]. \tag{7}$$

2.2 The Risk Aversion Coefficient

On the basis of equation (3) we can evaluate a risk aversion coefficient. Let μ and σ^2 be the mean and variance of loss X. Using the first terms in the Taylor expansion of the utility function u in $w-\mu$, we obtain

$$u(w-X) \approx u(w-\mu) + u'(w-\mu)\cdot(\mu-X) + \frac{1}{2}u''(w-\mu)\cdot(\mu-X)^2.$$

The expected utility from $u(w-X)$ is given by

$$\begin{aligned} &E[u(w-X)] \approx \\ &\approx E\left[u(w-\mu) + (\mu-X)\cdot u'(w-\mu) + \frac{1}{2}(\mu-X)^2\cdot u''(w-\mu)\right] \end{aligned}$$

After some processing we get

$$E[u(w-X)] \approx u(w-\mu) + \frac{1}{2}\sigma^2\cdot u''(w-\mu). \tag{8}$$

The Taylor expansion of the function on the right side of equation (3) is given by

$$u\left(w-P^{\max}\right) \approx u(w-\mu) + \left(\mu-P^{\max}\right)\cdot u'(w-\mu). \tag{9}$$

From the equality of equations (8) and (9) we have

$$u(w-\mu) + \frac{1}{2}\sigma^2\cdot u''(w-\mu) \approx u(w-\mu) + \left(\mu-P^{\max}\right)\cdot u'(w-\mu). \tag{10}$$

After some processing we get

$$P^{\max} \approx \mu - \frac{1}{2}\sigma^2\frac{u''(w-\mu)}{u'(w-\mu)}, \tag{11}$$

where a risk aversion coefficient $r(w)$ of the utility function u at a wealth $w-\mu$ is given by

$$r(w) = -\frac{u''(w-\mu)}{u'(w-\mu)}. \tag{12}$$

$$P^{\max} \approx \mu + \frac{1}{2} r(w-\mu) \cdot \sigma^2. \tag{13}$$

From (13) you can see that, if the insured has greater risk aversion coefficient, then he is willing to pay greater premium.

3 Mixture Operators

In this part we review some mixture operators introduced in [6], [7], [9-11]. Suppose that each alternative $\boldsymbol{x}$ is characterized by a score vector $\boldsymbol{x}=(x_1,...x_n) \in [0,1]^n$, where $n \in N - \{1\}$ is the number of applied criteria. A mixture operator can be defined as follows:

Definition 2 A mixture operator $M_g : [0,1]^n \to [0,1]$ is the arithmetic mean weighted by a continuous weighting function $g : [0,1] \to]0,\infty[$ given by

$$M_g(x_1,...,x_n) = \frac{\sum_{i=1}^{n} g(x_i) \cdot x_i}{\sum_{i=1}^{n} g(x_i)}, \tag{14}$$

where $(x_1,...,x_n)$ is an input vector.

Observe that due to the continuity of weighting function g, each mixture operator M_g is continuous. Evidently, M_g is an idempotent operator, [2], [6], [9-10]. Note that sometimes different continuous weighting functions are applied for different criteria score, which leads to a generalized mixture operator, see [6], [9-10].

Definition 3 A generalized mixture operator $M_{\mathbf{g}}$: $[0,1]^n \to [0,1]$ is given by

$$M_g(x_1,...,x_n) = \frac{\sum_{i=1}^{n} g_i(x_i) \cdot x_i}{\sum_{i=1}^{n} g_i(x_i)}, \tag{15}$$

where $(x_1,...,x_n)$ is an input vector and $\mathbf{g}=(g_1,...,g_n)$ is a vector of continuous weighting functions.

Obviously, generalized mixture operators are continuous and idempotent. A generalized mixture operator based on the ordinal approach can be defined as follows.

Definition 4 An ordered generalized mixture operator $M'_{\mathrm{g}}: \quad [0,1]^n \to [0,1]$ is given by

$$M'_{\mathrm{g}}(x_1,...,x_n)=\frac{\sum_{i=1}^{n} g_i\left(x_{(i)}\right)\cdot x_{(i)}}{\sum_{i=1}^{n} g_i\left(x_{(i)}\right)}, \tag{16}$$

where $\mathbf{g}=(g_1,...,g_n)$ is a vector of continuous weighting functions and $(x_{(1)},...,x_{(n)})$ is a non-decreasing permutation of an input vector.

An ordered generalized mixture operator is a generalization of an *OWA* operator [19], corresponding to constant weighting functions $g_i = w_i$, $w_i \in [0,1]$, $\sum_{i=1}^{n} w_i = 1$.

However, a mixture operator need not be non-decreasing. Marques-Pereira and Pasi [6] stated the the first sufficient condition for a weighting function g in order to a mixture operator (8) is to be non-decreasing. It can be written as follows:

Proposition 1 Let $g:[0,1]\to]0,\infty[$ be a non-decreasing smooth weighting function which satisfies the next condition:

$$0 \le g'(x) \le g(x) \tag{17}$$

for all $x \in [0,1]$. Then $M_g:[0,1]^n \to [0,1]$ is an aggregation operator for each $n \in N$, $n>1$.

We have generalized sufficient condition (17) in our previous work. In the next part we recall more general sufficient conditions mentioned in [7], [14-16].

From (14) we see that

$$\frac{\partial M_g}{\partial x_1} = \frac{\left(g(x_1) + g'(x_1)\cdot x_1\right)\cdot \sum_{i=1}^{n} g(x_1) - \left(\sum_{i=1}^{n} g(x_1)\cdot x_1\right)\cdot g'(x_1)}{\left(\sum_{i=1}^{n} g(x_1)\right)^2} \geq 0 \tag{18}$$

if and only if

$$g^2(x_1) + \alpha\left(g(x_1) + g'(x_1)\cdot(x_1 - \beta)\right) \geq 0, \tag{19}$$

where

$\alpha = \sum_{i=2}^{n} g(x_i)$ and $\alpha\cdot\beta = \sum_{i=2}^{n} g(x_i)\cdot x_i$, and thus necessarily $\beta \in [0,1]$ and $\alpha \in [(n-1)\cdot g(0), (n-1)\cdot g(1)]$.

Now it is easy to see that (17) implies (19). However, (19) is satisfied also whenever

$$g(x_1) + g'(x_1)\cdot(x_1 - \beta) \geq 0 \tag{20}$$

for each $x_1 \in [0,1]$ and each $\beta_1 \in [0,1]$.

Because $g'(x_1) \geq 0$, (20) is fulfiled whenever

$$0 \leq g'(x)\cdot(1-x) \leq g(x) \text{ for all } x \in [0,1].$$

We have just shown a sufficient condition more general than (17).

Proposition 2 Let $g : [0,1] \to]0,\infty[$ be a non-decreasing smooth weighting function which satisfies the condition:

$$0 \leq g'(x)\cdot(1-x) \leq g(x) \tag{21}$$

for all $x \in [0,1]$. Then $M_g : [0,1]^n \to [0,1]$ is an aggregation operator for each $n \in N$, $n > 1$.

Moreover, we have improved sufficient condition (21), but constrained by n.

Proposition 3 For a fixed $n \in N$, $n > 1$, let $g : [0,1] \to]0,\infty[$ be a non-decreasing smooth weighting function satisfying the condition:

$$\frac{g^2(x)}{(n-1)\cdot g(1)} + g(x) \geq g'(x)\cdot(1-x) \tag{22}$$

for all $x \in [0,1]$.Then $M_g : [0,1]^n \to [0,1]$ is an aggregation operator. In the next proposition we introduce a sufficient condition for the non-decreasing-ness of generalized mixture operators.

Proof. Minimal value of $g(x_1) + g'(x_1) \cdot (x_1 - \beta)$ for $\beta \in [0,1]$ is attained for $\beta = 1$, i. e., it is $g(x_1) + g'(x_1) \cdot (x_1 - 1)$. Therefore, (19) is surely satisfied whenever

$$\frac{g^2(x_1)}{\alpha} + g(x_1) \geq g'(x_1) \cdot (1 - x_1).$$

Suppose that (22) holds. Then

$$\frac{g^2(x_1)}{\alpha} + g(x_1) \geq \frac{g^2(x_1)}{(n-1) \cdot g(1)} \geq g'(x_1) \cdot (1 - x_1),$$

i. e., (19) is satisfied and thus g is a fitting weighting function.

□

In the next proposition we introduce a sufficient condition for the non-decreasing-ness of generalized mixture operators.

Proposition 4 For a fixed $n \in N$, $n > 1$, $i = 1,2,...,n$, let $g_i : [0,1] \to]0, \infty[$ be a non-decreasing smooth weighting functions, such that

$$\frac{g_i^2(x)}{\sum_{j \neq i} g_j(1)} + g_i(x) \geq g_i'(x) \cdot (1 - x) \tag{23}$$

for all $x \in [0,1]$. Then $M_{\mathbf{g}}' : [0,1]^n \to [0,1]$, where $\mathbf{g} = (g_1,..., g_n)$, is an aggregation operator.

4 Maximal Premium Determined by a Personal Utility Function

In practice, the utility function can be determined empirically by a personal interview made by a decision maker. In our opinion, there are at least two suitable ways to do this. The first one is based on an interview which provides us with probabilities estimated by an interviewed subject; the second one on a game with known probabilities where the interviewed subject gives us only information about a personal breaking point. The personal breaking point is the amount of wealth at which our individual is changed from risk averse to a risk seeker, or

vice versa. An appropriate curve for a risk averse and risk loving part is then selected from the theoretical utility functions.

4.1 A Personal Utility Function – a Probability-oriented Approach

Following this approach a personal utility function can easily be constructed from the information gleaned from a short interview using the classical regression analysis. The decision maker can use this function in any personal decision analysis in which the payoff falls between 0 and 30000 €. Now we recall the interview, which is compiled as follows [4].

Let us suppose you are owner of an investment which brings you zero payoff now or a loss of 30000 €. However, you have a possibility to step aside from this investment under the penalty in the amount of a sequence: A: 1000 €, B: 5000 €, C: 10000 €, D: 15000 €, E: 25000 €. Your portfolio manager can provide you with information expressing the probability loosing the 30000 €. Think. What would be the biggest probability of the loss, so that you retain the above mentioned investment? Only a few well-proportioned graphic points are required. From our interview we took the respective person's data points $(0,1)$, $(-1000,0.8)$, $(-5000,0.75)$, $(-10000,0.60)$, $(-15000,0,60)$, $(-25000,0.40)$, $(-30000,0.00)$, and created the appropriate utility function of our respondent as shown in Fig. 3. This curve has an interesting shape that reflects our respondent's approach to risk. The different personal utility functions for our respondent were created using the IBM SPSS 18.0 system for the purpose of comparison. The maximal premium $P^{\max}$ was calculated by u^{-1} inverse function to the utility equilibrium equation (4)

$$P^{\max} = w - u^{-1}\left(E\left[u\left(w - X\right)\right]\right) \tag{24}$$

with system Mathematica 5.

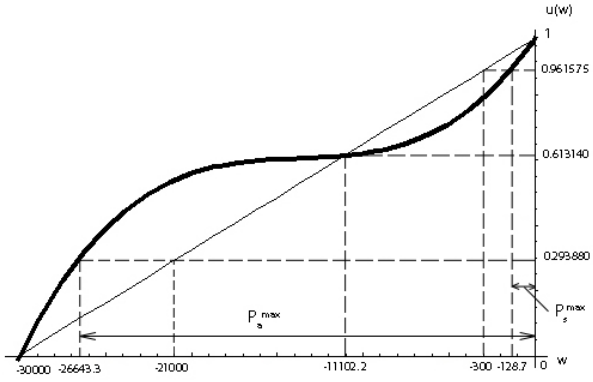

Figure 3
A utility function and the expected utility of our respondent (the function 2 from Table 1)

Utility functions are used to compare investments mutually. For this reason, we can scale a utility function by multiplying it by any positive constant and (or) transfer it by adding any other constant (positive or negative). This kind of transformation is called a positive affine transformation. All our results are the same with respect to such a transformation. Quadratic and cubic utility functions are written in Table 1. On the basis of statistical parameters (adjusted R square, p-values) we can assume that the cubic function is the best fitting function. Moreover, Table 1 also consists of appropriate expected utilities expressed by linear functions.

Remark 3 Expected utilities (for the utility functions from Table 1) can be calculated by means of a linear function which is assigned uniquely by points $(-30000, u(-30000))$ and $(0, u(0))$, or by the formula (8), alternatively. In both cases we get the same values for the expected utilities.

In Figure 3 you can see the personal utility function of our respondent, as well as three interesting points that are highlighted (also in Table 3). Maximal premium $P_a^{\max}$ represents the area where our respondent is risk averse, and $P_s^{\max}$, where he is risk seeking (loving).

Table 1
A utility function and the expected utility

	A utility function and the expected utility
1	$u(x) = -2.820 \cdot 10^{-10} x^2 + 1.856 \cdot 10^{-5} x + 0.904$ $E[u(x)] = 27.02 \cdot 10^{-6} x + 0.904$
2	$u(x) = 1.310 \cdot 10^{-13} x^3 + 5.383 \cdot 10^{-9} x^2 + 7.588 \cdot 10^{-5} x + 0.971$ $E[u(x)] = 3.23 \cdot 10^{-5} x + 0.971$

From Table 2 you can see that the insured person is willing to pay more than the expected loss to achieve his peace of mind".

Table 2
The expected utility and maximal premium with respect to a quadratic function

p **Probability**	$E[u]$ **with respect to quadratic function**	$P^{\max}$ **(€)**	$E[X]$ **(€)**	$P^{\min}$ **(€)**
0.00	0.904000	0.0	0.0	0
0.01	0.895894	433.9	300.0	301.69

0.05	0.863470	2115.7	1500.0	1508.10
0.10	0.822940	4110.7	3000.0	3015.34
0.20	0.741880	7808.5	6000.0	6027.24
0.30	0.660820	11197.3	9000.0	9035.69
0.40	0.579760	14336.3	12000.0	12040.70
0.50	0.498700	17293.4	15000.0	15042.40
0.60	0.417640	20079.0	18000.0	18040.60
0.70	0.336580	22725.4	21000.0	21035.50
0.80	0.255520	25251.4	24000.0	24027.00
0.90	0.174460	27672.4	27000.0	27015.20
1.00	0.093400	30000.0	30000.0	30000.00

Table 3
The expected utility and maximal premium with respect to a cubic function

p **Probability**	$E[u]$ **with respect to cubic function**	$P^{\max}$ **(€)**	$E[X]$ **(€)**	$P^{\min}$ **(€)**
0.00	0.971252	0.0	0.0	0,00
0.01	**0.961575**	**128.7**	**300.0**	301.69
0.05	0.922868	669.1	1500.0	1508.10
0.10	0.874485	1411.9	3000.0	3015.34
0.20	0.777717	3234.4	6000.0	6027.24
0.30	0.680950	6021.0	9000.0	9035.69
0.37	**0.613140**	**11102.2**	**11102.2**	**10986.70**
0.40	0.584182	18165.0	12000.0	12040.70
0.50	0.487415	22797.1	15000.0	15042.40
0.60	0.390647	25033.0	18000.0	18040.60
0.70	**0.293880**	**26643.3**	**21000.0**	21035.50
0.80	0.197112	27938.0	24000.0	24027.00
0.90	0.100345	29036.7	27000.0	27015.20
1.00	0.003577	30000.00	30000.0	30000.00

We determine the minimal premium by means of (7) with respect to the utility function for insurer $U(x) = \ln x$ with his basic capital $W = 2655513.51$ € and loss $X = 30000$ €.

The equation can be rewritten as follows:

$$U(W) = p \cdot U\left(W + P^{\min} - X\right) + (1-p) \cdot U\left(W + P^{\min}\right) \tag{25}$$

and hence

$$W = \left(W + P^{\min} - X\right)^{p} \cdot \left(W + P^{\min}\right)^{(1-p)}. \tag{26}$$

We determined individual minimal premiums with corresponding probability with the system Mathematica 5.

4.2 A Personal Utility Function– a Game-based Approach

Our expectation that our subject can appropriately estimate probabilities is the main drawback of the previous approach. In fact, we can doubt whether somebody without appropriate knowledge about probabilities can provide us with reliable answers. In order to avoid this problem we can assume a game with probabilities which are easy to understand, e.g. games based on coin tossing. Let us assume the following game. You have two possibilities: either to toss a coin with two possible results, head means you will get 10 €, tail means you will get nothing; or to choose 5 € without playing. What is amount of money for which you will start (stop) playing? It is easy to see that the expected value is the same in both cases and we make our decision about playing with respect to our personal utility function. The point at which we give up (stop) playing is the above mentioned breaking point. For simplicity we will assume the quadratic utility function. This approach allows us to combine easily personal utility functions to a group utility function using aggregation operators. The group utility function can represent a specific group of customers of our insurance company. Let us assume three utility functions based on different breaking, a utility function for $x = 29900$

$$u_1(x) = \begin{cases} -5.592778604 \cdot 10^{-10} \cdot (x - 29900)^2 + 0.5 & \text{for } 0 \le x \le 29900 \\ 5 \cdot 10^{-5} \cdot (x - 29900)^2 + 0.5 & \text{for } 29900 \le x \le 30000 \end{cases}, \tag{27}$$

a utility function for $x = 29800$

$$u_1(x) = \begin{cases} -5.63037701 \cdot 10^{-10} \cdot (x - 29800)^2 + 0.5 & \text{for } 0 \le x \le 29800 \\ 1.25 \cdot 10^{-5} \cdot (x - 29800)^2 + 0.5 & \text{for } 29800 \le x \le 30000 \end{cases}, \tag{28}$$

A utility function for $x = 29650$

$$u_1(x) = \begin{cases} -5.687489514 \cdot 10^{-10} \cdot (x - 29650)^2 + 0.5 & \text{for } 0 \le x \le 29650 \\ 0.4081 \cdot 10^{-5} \cdot (x - 29650)^2 + 0.5 & \text{for } 29650 \le x \le 30000 \end{cases}. \tag{29}$$

To construct the combined utility function we can use for example an ordered generalized mixture operator M'_{g} with weighting vector $\mathbf{g} = (g_1, g_2, g_3)$, where $g_1(x) = 0.2x + 0.8$, $g_2(x) = 0.5x + 0.5$ and $g_3(x) = 0.75x + 0.25$. Let us note that the selected weighting functions satisfy the conditions required for of non-decreasing aggregation operators.

Values 29900, 29800, 29650 we transform to the unit interval and aggregate them by means of M'_{g}. We obtain an aggregated value $M'_{\mathrm{g}} = 0.00575309$, and after transformation we have point of division $x = 29827.41$, where the insured is neutral to risk. On the basis of this division point we can create a new combined utility function

$$u(x) = \begin{cases} -5.620033657 \cdot 10^{-10} \cdot (x - 29827{,}41)^2 + 0.5 & \text{for } 0 \le x \le 29827{,}41 \\ 1.6785 \cdot 10^{-5} \cdot (x - 29827{,}41)^2 + 0.5 & \text{for } 29827{,}41 \le x \le 30000 \end{cases} \quad (30)$$

and appropriate expected utility

$$E(x) = \begin{cases} 1.6763 \cdot 10^{-5}\, x & \text{for } 0 \le x \le 29827{,}41 \\ 2.897039 \cdot 10^{-3} & \text{for } 29827{,}41 \le x \le 30000 \end{cases} \quad (31)$$

Table 4
The expected utility and maximal premium with respect to a function (28)

X	$w - X$	$E[u(w-X)]$	$w - P^{\max}$	$P^{\max}$	$P^{\min}$
0	30000,00	1,000000	30000,00	0,00	0
50	29950,00	0,855148	29972,90	27,10	50
100	29900,00	0,710296	29939,30	60,70	100
172,59	29827,41	0,500000	29827,40	172,60	172,59
200	29800,00	0,499541	28923,70	1076,30	200
300	29700,00	0,497864	27877,90	2122,10	300
1500	28500,00	0,477748	23535,00	6465,00	1500
3000	27000,00	0,452604	20644,10	9355,90	3000
6000	24000,00	0,402315	16643,50	13356,50	6000
9000	21000,00	0,352025	13600,90	16399,10	9000
12000	18000,00	0,301736	11044,90	18955,10	12000
15000	15000,00	0,251447	8797,37	21202,63	15000
18000	12000,00	0,201157	6767,79	23232,21	18000
21000	9000,00	0,150868	4902,97	25097,03	21000
24000	6000,00	0,100579	3168,28	26831,72	24000
27000	3000,00	0,050289	1539,73	28460,27	27000
30000	0,00	0,000000	0,00	30000,00	30000

The minimal premium we evaluated on the basis of formula (7) with the utility function for insurer $U(x) = \ln(x+1)$ with the system Mathematica. From Table 4 and also from the formula (7) you can see that the minimal premium is given by the size of the expected loss. A newly-gained utility function would be required for evaluating a decision with more extreme payoffs or if our respondent's

attitudes change because of a new job or lifestyle change. Moreover, the utility function must be revised from the viewpoint of time.

Conclusions

We have shown two approaches to creating a personal utility function and we have calculated the maximum premium against the loss of 30000 € with respect to it. We think that the personal utility function of an insured person would be very important for an insurer. On the basis of the personal utility function the insurer would know what approaches to risk the customers have and thus, how they will behave towards their own wealth. Creating a utility function for the insurer is very difficult. Moreover, in our next work we want to investigate the insurer's utility function and we want to determine the minimal premium against the loss of 30000 € with respect to a concrete real insurer's utility function.

Acknowledgement

This work was supported by grant VEGA 1/0539/08.

References

[1] Calvo, T., Mesiar, R., Yager, R. R.: Quantitative Weights and Aggregation, Fellow, IEEE Transaction on fuzzy systems, Vol. 2, No. 1, 2004, pp. 62-69

[2] Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Monograph: Aggregation Functions, Acta Polytechnica Hungarica, 6(1), 2009, pp. 79-94

[3] Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M.: Modern Actuarial Risk Theory, Kluwer Academic Publishers, Boston, 2001, ISBN 1-4020-2952-7

[4] Khurshid, A.: The End of Rationality?, Abstracts, Tenth International Conference on Fuzzy Set Theory and Applications, FSTA 2010, Slovak Republic, (2010), pp. 10-11

[5] Lapin, L. L.: Quantitative Methods for Business Decisions, 5th edition, San Jose State University, 1975, ISBN 0-15-574331-7

[6] Marques-Pereira, R. A., Pasi, G.: On Non-Monotonic Aggregation: Mixture operators. In: Proceedings of the 4th Meeting of the EURO Working Group on Fuzzy Sets (EUROFUSE 99) and 2nd International Conference on Soft and Intelligent Computing (SIC 99), Budapest, Hungary, (1999), pp. 513-517

[7] Mesiar R., Špirková, J.: Weighted Means and Weighting Functions, Kybernetika 42, 2, 2006, pp. 151-160

[8] Norstad, J.: An Introduction to Utility Theory, http://homepage.mac.com/j.norstad, 2005

[9] Ribeiro, R. A., Marques-Pereira, R. A.: Generalized Mixture Operators Using Weighting Functions: A Comparative Study with WA and OWA, European Journal of Operational Research 145, 2003, pp. 329-342

[10] Ribeiro, R. A., Marques-Pereira, R. A.: Weights as Functions of Attribute Satisfaction Values. In: Proceedings of the Workshop on Preference Modelling and Applications (EUROFUSE), Granada, Spain, 2001, pp. 131-137

[11] Ribeiro, R. A., Marques-Pereira, R. A.: Aggregation with Generalized Mixture Operators Using Weighting Functions. In: Fuzzy Sets and Systems 137, 2003, pp. 43-58

[12] Riesová, R.: Utility Function in a Theory and Practice (in Slovak), Student Scientific Activity, Banská Bystrica, 2009

[13] Rotar V. I.: Actuarial Models – The Mathematics of Insurance, Chapman & Hall/CRC, Taylor & Francis Group, 2006

[14] Špirková, J.: Weighting Functions for Aggregation Operator: 3rd International Summer School on Aggregation Operators, Universita Della Svizzera Italiana, Lugano, Switzerland, 2005, pp. 127-130

[15] Špirková, J.: Mixture and Quasi-Mixture Operators, IPMU 2006, France, 2006, pp. 603-608

[16] Špirková, J.: Dissertation thesis, Weighted Aggregation Operators and Their Applications, Bratislava, 2008

[17] Špirková, J.: Utility Function in an Insurance, Proceedings from AMSE 2009, Uherské Hradiště, 2009, pp. 417-427

[18] Špirková, J. Kráľ, P.: Aggregation Functions and Utility Functions in General Insurance, in Abstract of Proceedings of 10th International Conference on Fuzzy Set Theory and Applications, Liptovský Ján, Slovakia, 2010, p. 124

[19] Yager, R. R.: Generalized OWA Aggregation Operators, Fuzzy Optimization and Decision Making 3, 2004, pp. 93-107

Function Approximation Performance of Fuzzy Neural Networks

Rita Lovassy[1], László T. Kóczy[2,3], László Gál[2,4]

[1] Inst. of Microelectronics and Technology, Kandó Kálmán Faculty of Electrical Engineering, Óbuda University, Budapest, Hungary

[2] Inst. of Informatics, Electrical and Mechanical Eng., Faculty of Engineering Sciences, Széchenyi István University, Győr, Hungary

[3] Dept. of Telecommunication and Media Informatics, Budapest University of Technology and Economics, Budapest, Hungary

[4] Dept. of Technology, Informatics and Economy, University of West Hungary, Szombathely, Hungary

e-mail: lovassy.rita@kvk.uni-obuda.hu, koczy@sze.hu, koczy@tmit.bme.hu, laci.gal@gmail.com

Abstract: In this paper we propose a Multilayer Perceptron Neural Network (MLP NN) consisting of fuzzy flip-flop neurons based on various fuzzy operations applied in order to approximate a real-life application, two input trigonometric functions, and two and six dimensional benchmark problems. The Bacterial Memetic Algorithm with Modified Operator Execution Order algorithm (BMAM) is proposed for Fuzzy Neural Networks (FNN) training. The simulation results showed that various FNN types delivered very good function approximation results.

Keywords: fuzzy flip-flop neurons; Fuzzy Neural Networks; Bacterial Memetic Algorithm with Modified Operator Execution Order

1 Introduction

Over the last few decades many different approaches to the hybridization of neural networks and fuzzy systems have been introduced and studied [9], [11], [22] as new neuro-fuzzy structures. Based on this idea, in this paper the concept of the fuzzy flip-flop neuron is introduced. The function approximation capability of the novel Fuzzy Flip-Flop-based Neural Networks (FNN), as a new type of neural networks is studied. A comparison of the effect of applying some selected fuzzy operations in the investigation of the fuzzy flip-flop (F^3)-based neurons, and the Multilayer Perceptrons (MLPs) based on them, are presented. The proposed

network is a structure consisting of the same types of F^3. The proposed training method is a particular combination of evolutionary and gradient based (global and local search) algorithms. The effect of fuzzy operations types and fuzzy neurons numbers are studied. Illustrative examples are presented in order to demonstrate the success of this work in terms of the function approximation capability of the proposed FNNs.

The outline of this paper is as follows. In Section II we present the fuzzy J-K and D flip-flops in general. The Fuzzy Flip-Flop based Neural Network as a novel implementation possibility of multilayer perceptron neural networks is proposed in Section III. The FNNs neuron element may be any fuzzy flip-flop derived from the original F^3 with two additional modifications (feedback and fixed internal state) with more or less sigmoid transfer characteristics. After simplifications the fuzzy J-K and D flip-flop neurons block diagram for a fix Q value is given. The Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM) is carried out in Section IV. We apply this method to demonstrate that the proposed FNNs built up from fuzzy J-K and D flip-flops based on algebraic, Łukasiewicz, Yager, Dombi and Hamacher operations can be used for learning and approximating a battery cell charging characteristics, two dimensional trigonometric functions, and two and six dimensional benchmark problems. The target activation function is *tansig*, a MATLAB built-in sigmoid transfer function. Finally, the FNNs function approximation performance comparison thought simulation results are discussed in Section V, followed by a brief Conclusion and References.

2 Fuzzy J-K and D Flip-Flops

The concept of fuzzy flip-flop was introduced in the middle of 1980's by Hirota (with his students) [7]. The Hirota Lab recognized the essential importance of the concept of a fuzzy extension of a sequential circuit and the notion of fuzzy memory. From this point of view they proposed alternatives for "fuzzifying" digital flip-flops. The starting elementary digital units were the binary J-K flip-flops. Their definitive equation was used both in the minimal disjunctive and conjunctive forms. As fuzzy connectives do not satisfy all Boolean axioms, the fuzzy equivalents of these equations resulted in two non-equivalent definitions, "reset and set type" fuzzy flip-flops, using the concepts of fuzzy negation, t-norm and t-conorm operations. In [8] Hirota et al. recognized that the reset and set equations cannot be easily used as elements of memory module, because of their asymmetrical nature. In their paper [19] Ozawa, Hirota and Kóczy proposed a unified form of the fuzzy J-K flip-flop characteristic equation, involving the reset and set characteristics, based on min-max and algebraic norms. A few years later, the hardware implementation of these fuzzy flip-flop circuits in discrete and continuous mode was presented in [20]. The next state Q_{out} of a fuzzy J-K flip-

flop is characterized as a function of both the present state Q and the two present inputs J and K. The unified formula of the fuzzy J-K flip-flop was expressed as follows [19]:

$$Q_{out} = (J \vee \overline{K}) \wedge (J \vee Q) \wedge (\overline{K} \vee \overline{Q}) \tag{1}$$

The over bar denotes the standard negation (e.g. $\overline{K} = 1 - K$); furthermore, $\wedge$ and $\vee$ denote fuzzy operations (t-norm and t-conorm, labeled in the next as *i* and *u*). $J, K, Q, Q_{out} \in [0,1]$. In [12] a F^3 derived from fuzzy J-K flip-flop where $\overline{Q}$ is fed back to K ($K = 1 - Q$) is proposed. The characteristic equation of a fuzzy J-K flip-flop with feedback is

$$Q_{out} = (J \; u \; Q) \; i \; (J \; u \; Q) \; i \; (Q \; u \; (1-Q)) \tag{2}$$

The concept of a novel fuzzy D flip-flop type was introduced in [13]. Connecting the inputs of the fuzzy J-K flip-flop in a particular way, namely by applying an inverter in the connection of the input J to K, case of $K = 1 - J$, a fuzzy D flip-flop is obtained. Substitute $\overline{K} = J$ in equation (1) and let $J = D$, the fundamental equation of fuzzy D flip-flop is

$$Q_{out} = (D \; u \; D) \; i \; (D \; u \; Q) \; i \; (D \; u \; (1-Q)) \tag{3}$$

In our previous papers [13], [14] the unified fuzzy J-K flip-flop based on various norms combined with the standard negation was analyzed in order to investigate, whether and to what degree they present more or less sigmoid (s-shaped) $J \to Q_{out}$ characteristics in particular cases, when $K = 1 - Q$ (unified fuzzy J-K flip-flop with feedback), $K = 1 - J$ (new fuzzy D flip-flop derived from the unified fuzzy J-K one) with fixed value of Q. We conducted extensive investigations and found that the $J \to Q_{out}$ transfer characteristics of fuzzy J-K flip-flops with feedback based on Łukasiewicz, Yager, Dombi and Hamacher norms, as well as the $D \to Q_{out}$ characteristics of (new) fuzzy D flip-flops of Łukasiewicz, Yager and Hamacher operations, show quasi sigmoid curvature for selected Q values. The fuzzy J-K and D F^3s based on algebraic norms as well as the fuzzy D F^3s based on Dombi norms have non-sigmoid behavior.

3 Fuzzy Neural Networks

3.1 Fuzzy Flip-Flop Neurons

We proposed the construction of a neuron unit, a combinational sigmoid generator derived from arbitrary fuzzy J-K flip-flop where $\overline{Q}$ is fed back to K and (old) Q is fixed (Figure 1). In this approach, the output of fuzzy J-K flip-flop neuron depends from the value of Q_{fix} and input values of J. Substitute $\overline{K} = Q$ (1 - $K = Q$) in the unified formula of the fuzzy J-K flip-flop (equation 1), for a fix Q value, the characteristic equation of fuzzy J-K flip-flop neuron is

$$Q_{out} = (J \; u \; Q_{\text{fix}}) \; i \; (J \; u \; Q_{\text{fix}}) \; i \; \left(Q_{\text{fix}} \; u \; \left(1 - Q_{\text{fix}}\right)\right) \tag{4}$$

where i and u denotes various t-norms and t-conorms.

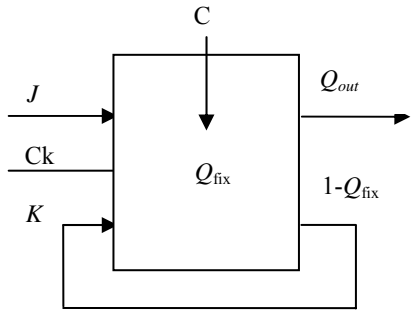

Figure 1
Fuzzy J-K flip-flop neuron

The clocked fuzzy J-K flip-flop neuron circuit can be implemented using hardware blocks (denoted by i, u and n symbols) to realize various t-norms, t-conorms and fuzzy negations [24]. Since t-norms and t-conorms are functions from the unit square into the unit interval, the fuzzy J-K flip-flop neuron block diagram differs from the binary J-K flip-flop structure. The input J is driven by a synchronized clock pulse in the sample-and-hold (S/H) circuit (Figure 2) which could be a simple D flip-flop used as register.

We proposed the construction of the fuzzy D flip-flop neuron (Figure 3), which is a combinational sigmoid generator. This unit is derived from arbitrary fuzzy J-K flip-flop by connecting the inputs of the fuzzy J-K flip-flop in a particular way, namely by applying an inverter in the connection of the input J to K.

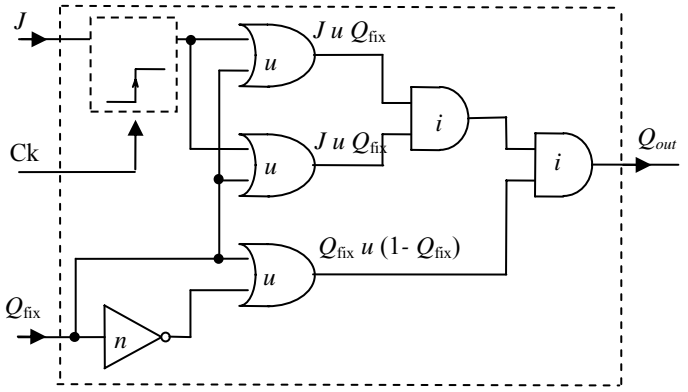

Figure 2
Fuzzy J-K flip-flop neuron block diagram

Starting from the fundamental equation of the fuzzy J-K flip-flop and substituting $\overline{K} = J$ in equation (1) and letting $D = J$ for a fix Q value, the characteristic equation of fuzzy D flip-flop neuron is

$$Q_{out} = (D\ u\ D)\ i\ (D\ u\ Q_{\text{fix}})\ i\ \left(D\ u\ \left(1 - Q_{\text{fix}}\right)\right) \tag{5}$$

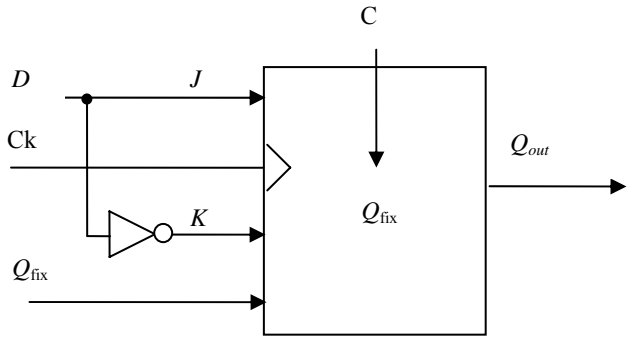

Figure 3
Fuzzy D flip-flop neuron

Interconnecting the blocks of fuzzy operations in a different way, the fuzzy D flip-flop neuron block diagram is obtained (Figure 4).

The use of four-layered (that have two sigmoid hidden layers) neural network as universal approximators of continuous functions have been investigated by Funahashi [4] and Hecht-Nielsen [6]. Kurkova [10] studied the function approximation capabilities of multilayer feedforward networks with sigmoid activation, analyzing also their computational complexity issues.

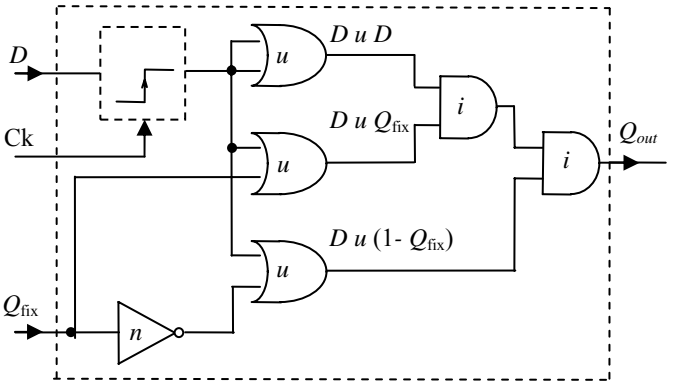

Figure 4
Fuzzy D flip-flop neuron block diagram

The neuro-fuzzy system proposed is based on two hidden layers where the approximation process becomes more manageable. The first hidden layer neurons are used to partition the input space into regions and learn the local features. A neuron in the second layer learns the global features for a particular region of the input space and outputs zero elsewhere [4]. The FNNs constituted from fuzzy flip-flop neurons are supervised feedforward network, applied in order to approximate various test functions. The functions to be approximated are represented by a set of input/output pairs.

In this approach the weighted input values are connected to input *J* of the fuzzy flip-flop neuron based on a pair of t-norm and t-conorm, having quasi sigmoid transfer characteristics. The output signal is then computed as the weighted sum of the input signals transformed by the transfer function. Based on previous hardware implementation results [19], FNNs with fixed structure can be stated as easily implemented in real hardware neural networks.

The nonlinear characteristics exhibited by fuzzy neurons are represented by quasi sigmoid transfer functions given by fuzzy J-K and D flip-flop neurons based on algebraic, Łukasiewicz, Yager, Dombi and Hamacher operations. The proposed network activation function is the same at each hidden layer, from unit to unit. The function approximation goodness is strongly dependent on the number of fuzzy neurons in the hidden layers [15], [16].

4 Bacterial Memetic Algorithm with Modified Operator Execution Order

The process of biological evolution has inspired a large amount of optimization algorithms. It has been shown that evolutionary algorithms work efficiently for solving nonlinear and constrained optimization problems. These methods do not use derivatives of the functions, such as the gradient-based training algorithms. Similarly to biological recombination, these methods are based on the search for a large number of solutions.

In particular, the Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM) [5] evolutionary approach is proposed for FNNs training. This chapter presents how the bacterial evolutionary algorithm [18] can be improved with another optimization method, the Levenberg-Marquardt (LM) technique, to achieve better results in the function approximation process, developing an effective hybrid combination.

The BMAM is a memetic algorithm which eliminates the imperfection of traditional evolutionary and LM algorithm. The evolutionary algorithm is able to find the global optimum region but miss the local optimum solution. The LM algorithm is fast and efficient for training feedforward neural networks and is able to find the local optimum, but is very sensitive to the initial position of the search space.

The learning of the FNNs is formulated as a parameter optimization problem, using the mean square error as the fitness evaluation set-up. The algorithm starts with a random population of initial solutions to the optimization problem. The solutions are coded as an array of floating point or integer values. The basic steps embrace the bacterial mutation operation and the LM method. In this application the population number was initialized according to the network size.

During simulations 30 generations of 5 individuals with 5 clones were chosen to obtain the best fitting variable values, with the lowest performance. Then the same part, or parts, of the chromosome is chosen and mutated randomly. The LM method nested into evolutionary algorithm is applied 3 times for each individual. The selection of the best clone is made and transfers its mutated parts to the other clones. The part choosing-mutation-LM method-selection-transfer cycle is repeated until all the parts are mutated, improved and tested. The best individual remains in the population and all other clones are deleted. This process is repeated until all individuals have gone through the modified bacterial mutation. Then the Levenberg-Marquardt method is applied 7 times for each individual executing several LM cycles during the bacterial mutation after each mutation step.

Gene transfer operation is done 3 times for a partial population. The number of the gene transfers in a generation is the algorithm parameter; it could be 0. The quasi optimal values can be identified at the end of the BMAM training algorithm.

5 Numerical Simulation and Results

The FNNs architecture is predefined, depending on the input function complexity. In this approach the choice of an optimal network design for a given problem is a guessing process. In particular, the application of a recently improved BMAM algorithm is applied for training various FNNs with different structures. This new, complex software is able to train all the FNNs parameters, eliminating completely the imprecision caused by training them with the LM algorithm. The simulation results published in our previous papers obtained under the same conditions could turn out to be different because the LM method is very sensitive to the initial values of the search space. The FNNs approximate a one-dimensional real-life application, two dimensional trigonometric functions, two benchmark problems whose dates were selected from the input/output test points of a gas furnace benchmark data set, and a six dimensional non-polynomial function. The test functions are arranged in the order of complexity.

A Simple Real - Life Application: Approximation of a Nickel-Metal Hydride Battery Cell Charging Characteristics

In this particular case, the FNNs approximate a Nickel-Metal Hydride (NiMH) Battery Cell charging characteristics [2], a one-input real-life application.

The nickel-metal hydride batteries can be repeatedly charged and discharged for more than 500 cycles. The charging process duration can be different, from 15 minutes to 20 hours. The charge characteristics are affected by current, time and temperature. In this experiment it was more than 1 hour. The test function is a characteristic between the battery capacity input and the cell voltage. The battery type was GP 3.6V, 300 mAH, 3x1.2V NiMH, charged for 1.5 hours with 300 mA and 25˚C.

B The Box-Jenkins' Gas Furnace Benchmark Data Set

The gas furnace data set presented by Box and Jenkins in 1970 is a frequently used benchmark data set. The set consists of 296 input-output data; a pair is sampled at every 9 seconds. The input signal represents the flow rate of the methane in a gas furnace, while the output of the model corresponds to the CO_2 concentration in the gas mixture flowing out of the furnace under a steady air supply [17]. We used as the most studies [21], [23] the inputs $y(k\text{-}1)$ and $u(k\text{-}4)$ which have the highest correlation with the output $y(k)$.

C Two - Input Trigonometric Functions

We used the next two two-dimensional polynomial input functions as test functions

$$y_1 = \left(\sin\left(c_1 \cdot x_1\right)^5 \cdot \cos\left(c_2 \cdot x_2\right)^3\right)/2 + 0.5 \tag{6}$$

$$y_2 = e^{-\frac{r^2}{100}} \cdot \cos\left(\frac{r}{2}\right); \text{ where } r = \sqrt{x_1^2 + x_2^2}, \tag{7}$$

$x_1,\ x_2 \in [-20, 20]$

D Six Dimensional Non-Polynomial Function

This widely used target function originates from paper [1] and is given by the following expression

$$y_3 = x_1 + x_2^{0.5} + x_3 x_4 + 2e^{2(x_5 - x_6)} \tag{8}$$

where $x_1,\ x_2 \in [1,\ 5],\ x_3 \in [0,\ 4],\ x_4 \in [0,\ 0.6],\ \ x_5 \in [0,\ 1],\ x_6 \in [0,\ 1.2]$.

The real-life application (denoted by 1D) is approximated with a 1-2-2-1 FNN size, described by a set of 543 input/output samples selected equidistantly from a set of 2715 test points. To approximate the gas furnace benchmark data set we proposed a 1-2-2-1 FNN (2D-gas) size given by 296 dates. A 1-20-20-1, respectively a 1-15-15-1 feedforward neural networks structure based on F^3 neurons were proposed to approximate the two two-input trigonometric functions, (equations (6) and (7), labeled as 2D-trig and 2D-hat), represented by 1600 input/output samples. To approximate the six-dimensional benchmark problem we proposed a 1-10-10-1 FNN (6D) size given by 200 samples.

The number of neurons was chosen after experimenting with different size hidden layers. Smaller neuron numbers in the hidden layer result in worse approximation properties, while increasing the neuron number in a complex FNN structure results in better performance, but longer simulation time.

In [15] we proposed a method to find the optimal *Q* and fuzzy operation parameter pairs for J-K and D type F^3 neurons based on algebraic, Yager, Dombi and Hamacher norms by training a 1-8-8-1 FNN with the Bacterial Memetic Algorithm. The optimal variable values depend on the fuzzy flip-flop neuron and fuzzy operation types. A change of the operations and parameter values in the characteristic equations of the fuzzy J-K and D flip-flops leads to the modification of the slope of the transfer function, which will affect the learning rate in the implementation of neural networks. In the next, the algebraic, Łukasiewicz, Yager, Dombi and Hamacher operations and two different fuzzy flip-flop neuron types will be compared from the point of view of the respective fuzzy-neural networks approximation capability.

Figures 5 and 6 compare the function approximation performance of J-K and D FNNs in case of various test functions. Tables I and II present the 10 runs average approximation goodness, by indicating the median MSE (mean squared error) of the various FNNs training values. During evaluation we compared the median MSE values, considering them as the most important indicators of trainability. The median is a robust estimate of the center of a data sample, since outliers have little

effect on it. According to the numerical illustrations the average of 10 runs MSE values the sequence is almost the same in every test function cases. By extensive simulation experiments it is proved that the function approximation goodness of FNNs based on fuzzy J-K flip-flop neuron with Dombi and Łukasiewicz norms are the best ones.

The error of approximation of the battery cell and Box-Jenkins' gas furnace benchmark dates characteristics obtained by traditional *tansig* function is irrelevantly less than that obtained in our simulations. As can be seen from Figure 6, and comparing the simulation results from Table II the function approximation by D FNNs may be considered sufficiently good in case of Łukasiewicz and Yager type fuzzy operations.

Conclusions

In this paper, we found that the fuzzy J-K flip-flop neurons based on Dombi and Łukasiewicz as well as the fuzzy D flip-flop neurons based on Łukasiewicz and Yager norms are the most suitable ones for constructing FNNs in order to approximate various test functions. As these FNN types produced more or less low MSE values in all simulation experiments. Thus, we proposed the construction of real hardware fuzzy neural networks constructed of the above-mentioned F^3 neuron types. The accuracy of the approximations not only depends on the network structures and parameters selected, such as the number of layers, and of the hidden units, but is strongly influenced by the fuzzy flip-flop neuron and fuzzy operation type. In the future we intend to propose new types of fuzzy flip-flop based on pliant inequality [2], and to improve the function approximation capability of FNNs.

Acknowledgement

Research supported by the National Scientific Research Fund Grant OTKA K75711, further by Széchenyi István University Main Research Direction Grant 2010, and Óbuda University Grants.

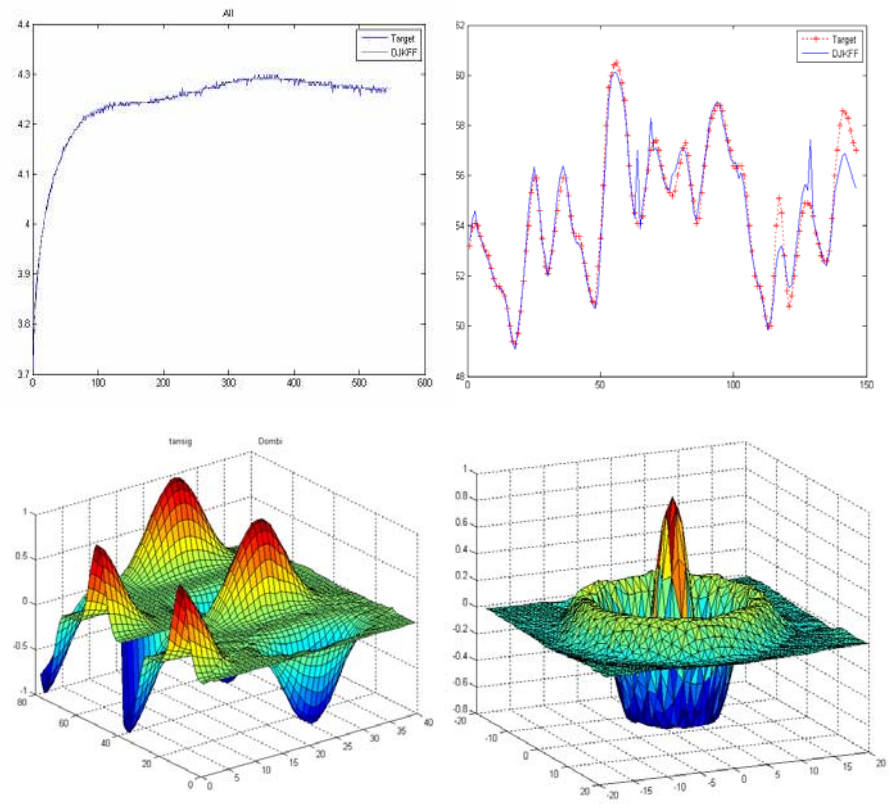

Figure 5
Function approximation capabilities of J-K FNNs

Table I
Training Median MSE Values for J-K Type FNNs

Fuzzy op.	Bat. cell 1-2-2-1 FNN	2D-gas 1-2-2-1 FNN	2D-trig 1-20-20-1 FNN	2D-hat 1-15-15-1 FNN	6D 1-10-10-1 FNN
tansig	$1.32\text{x}10^{-5}$	$7.71\text{x}10^{-2}$	$9.07\text{x}10^{-7}$	$4.26\text{x}10^{-7}$	$1.12\text{x}10^{-4}$
Algebraic	$3.32\text{x}10^{-4}$	$1.07\text{x}10^{0}$	$4.32\text{x}10^{-2}$	$3.17\text{x}10^{-2}$	$9.69\text{x}10^{-1}$
Łukasiewicz	$\mathbf{7.11\text{x}10^{-5}}$	$\mathbf{7.27\text{x}10^{-2}}$	$3.71\text{x}10^{-4}$	$9.46\text{x}10^{-4}$	$5.78\text{x}10^{-1}$
Yager	$1.47\text{x}10^{-4}$	$8.23\text{x}10^{-1}$	$1.53\text{x}10^{-2}$	$1.49\text{x}10^{-2}$	$5.92\text{x}10^{-1}$
Dombi	$\mathbf{3.52\text{x}10^{-5}}$	$\mathbf{7.30\text{x}10^{-2}}$	$\mathbf{8.75\text{x}10^{-6}}$	$\mathbf{1.76\text{x}10^{-4}}$	$2.98\text{x}10^{-1}$
Hamacher	$2.59\text{x}10^{-4}$	$8.27\text{x}10^{-1}$	$2.18\text{x}10^{-2}$	$2.86\text{x}10^{-2}$	$7.43\text{x}10^{-1}$

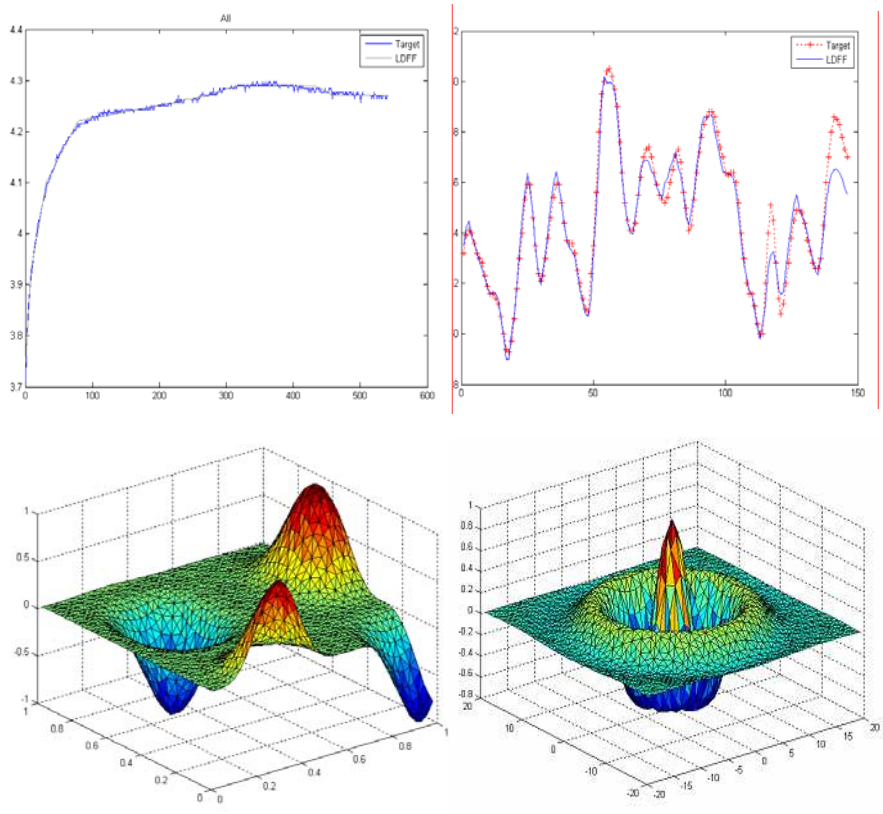

Figure 6
Function approximation capabilities of D FNNs

Table II
Training Median MSE Values for D Type FNNs

Fuzzy op.	Bat. cell 1-2-2-1 FNN	2D-gas 1-2-2-1 FNN	2D-trig 1-20-20-1 FNN	2D-hat 1-15-15-1 FNN	6D 1-10-10-1 FNN
tansig	$1.32\text{x}10^{-5}$	$7.71\text{x}10^{-2}$	$9.07\text{x}10^{-7}$	$4.26\text{x}10^{-7}$	$1.12\text{x}10^{-4}$
Algebraic	$1.38\text{x}10^{-4}$	$1.18\text{x}10^{0}$	$2.71\text{x}10^{-2}$	$2.94\text{x}10^{-2}$	$6.56\text{x}10^{-1}$
Łukasiewicz	**$4.95\text{x}10^{-5}$**	**$7.26\text{x}10^{-2}$**	**$7.48\text{x}10^{-4}$**	**$1.52\text{x}10^{-3}$**	$7.15\text{x}10^{-1}$
Yager	$1.09\text{x}10^{-4}$	**$7.69\text{x}10^{-2}$**	$8.21\text{x}10^{-3}$	$1.86\text{x}10^{-2}$	**$1.45\text{x}10^{-1}$**
Dombi	$6.41\text{x}10^{-4}$	$1.25\text{x}10^{0}$	$2.93\text{x}10^{-2}$	$3.23\text{x}10^{-2}$	$1.58\text{x}10^{0}$
Hamacher	$1.12\text{x}10^{-4}$	$7.91\text{x}10^{-2}$	$5.25\text{x}10^{-3}$	$1.48\text{x}10^{-2}$	$2.52\text{x}10^{-1}$

References

[1] J. Botzheim, B. Hámori, L. T. Kóczy, A. E. Ruano: Bacterial Algorithm Applied for Fuzzy Rule Extraction, IPMU 2002, Annecy, France, pp. 1021-1026

[2] J. Dombi: Pliant Arithmetics and Pliant Arithmetic Operations, Acta Polytechnica Hungarica, Vol. 6, No. 5, 2009, pp. 19-49

[3] R. Fan: NiMH Battery Charger Reference Design, Designer Reference Manual, 2003

[4] K. I. Funahashi: On the Approximate Realization of Continuous Mapping by Neural Networks, Neural Networks, 2, 1989, pp. 183-192

[5] L. Gál, J. Botzheim, L. T. Kóczy: Improvements to the Bacterial Memetic Algorithm used for Fuzzy Rule Base Extraction, Proc. of CIMSA 2008, Computational Intelligence for Measurement Systems and Applications, Istanbul, Turkey, 2008, pp. 38-43

[6] R. Hecht-Nielsen: Theory of the Backpropagation Neural Network, in Proc. of the Neural Networks, IJCNN, International Joint Conference on Neural Networks, 1989, pp. 593-605

[7] K. Hirota, K. Ozawa: Concept of Fuzzy Flip-Flop, Preprints of 2nd IFSA Congress, Tokyo, 1987, pp. 556-559

[8] K. Hirota, K. Ozawa: Fuzzy Flip-Flop as a Basis of Fuzzy Memory Modules, In: M. M. Gupta et al., eds., Fuzzy Computing. Theory, Hardware and Applications, North Holland, Amsterdam, 1988, pp. 173-183

[9] J.-S. R. Jang, C. T. Sun, E. Mizutani: Neuro-Fuzzy and Soft Computing. A Computational Approach to Learning and Machine Intelligence, Upper Saddle River, NJ: Prentice Hall, 1997

[10] V. Kurkova: Kolmogorov's Theorem and Multilayer Neural Networks Neural Networks, 5. 1992, pp. 501-506

[11] R. S. T. Lee: Fuzzy-Neuro Approach to Agent Applications, Springer, 2006

[12] R. Lovassy, L. T. Kóczy, L. Gál: Analyzing Fuzzy Flip-Flops Based on Various Fuzzy Operations, Acta Technica Jaurinensis Series Intelligentia Computatorica, Vol. 1, No. 3, 2008, pp. 447-465

[13] R. Lovassy, L. T. Kóczy, L. Gál: Applicability of Fuzzy Flip-Flops in the Implementation of Neural Networks; Proc. of CINTI 2008, 9th International Symposium of Hungarian Researchers on Computational Intelligence, Budapest, Hungary, 2008, pp. 333-344

[14] R. Lovassy, L. T. Kóczy, L. Gál: Function Approximation Capability of a Novel Fuzzy Flip-Flop-based Neural Network, Proc. of IJCNN 2009, International Joint Conference on Neural Networks, Atlanta, USA, 2009, pp. 1900-1907

[15] R. Lovassy, L. T. Kóczy, L. Gál: Parameter Optimization in Fuzzy Flip-Flop Based Neural Networks, in Proc. of International Symposium on Innovations in Intelligent Systems and Applications, INISTA 2009, Trabzon, Turkey, 2009, pp. 205-209

[16] R. Lovassy, L. T. Kóczy, L. Gál: Fuzzy Flip-Flops as Basic Components in Fuzzy Neural Networks-abstract, Tenth International Conference on Fuzzy Set Theory and Applications, FSTA 2010, Liptovsky Ján, the Slovak Republic, 2010, pp. 92-93

[17] S. Lukasik, P. Kowalski, M. Charytanowicz, P. Kulczycki: Fuzzy Model Identification Using Kernel-Density-based Clustering, in: Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Applications. Volume II, Krassimir Atanassov et al. (eds.), EXIT Publishing House, Warsaw (Poland), 2008, pp. 135-146

[18] N. E. Nawa, T. Furuhashi: Fuzzy System Parameters Discovery by Bacterial Evolutionary Algorithm, IEEE Trans. on Fuzzy Systems, 7(5), 1999, pp. 608-616

[19] K. Ozawa, K. Hirota, L. T. Kóczy, K. Omori: Algebraic Fuzzy Flip-Flop Circuits, Fuzzy Sets and Systems 39/2, North Holland, 1991, pp.215-226

[20] K. Ozawa, K. Hirota, L. T. Kóczy: Fuzzy Flip-Flop, In: M. J. Patyra, D. M. Mlynek, eds., Fuzzy Logic. Implementation and Applications, Wiley, Chichester, 1996, pp. 197-236

[21] W. Pedrycz: An Identification Algorithm in Fuzzy Relation Systems, Fuzzy Sets Syst. 13, 1984, pp. 153-167

[22] L. Rutkowski: Flexible Neuro-Fuzzy Systems, Kluwer 2004

[23] C. W. Xu, Y. Z. Lu: Fuzzy Model Identification and Self-Learning for Dynamic Systems, IEEE Trans. Syst. Man and Cyber., 17, 1987, pp. 683-689

[24] A. H. Zavala, O. C. Nieto, I. Batyrshin, L. V. Vargas: VLSI Implementation of a Module for Realization of Basic t-norms on Fuzzy Hardware, Proc. of FUZZ-IEEE 2009, IEEE Conference on Fuzzy Systems, Jeju Island, Korea, August 20-24, 2009, pp. 655-659

Fuzzy Interpolation According to Fuzzy and Classical Conditions

I. Perfilieva, M. Wrublová, P. Hoďáková

University of Ostrava, Inst. for Research and Applications of Fuzzy Modeling
30. dubna 22, 701 03 Ostrava 1, Czech Republic
e-mail: {Irina.Perfilieva,Michaela.Wrublova,Petra.Hodakova}@osu.cz

We will be focused on the interpolation approach to a computation with fuzzy data. A definition of interpolation of fuzzy data, which stems from the classical approach, is proposed. We investigate another approach to fuzzy interpolation (published in [5]) with relaxed interpolation condition. We prove that even if the interpolation condition is relaxed the related algorithm gives an interpolating fuzzy function which fulfils the interpolation condition in the classical sense.

Keywords: Fuzzy function; fuzzy equivalence, fuzzy space, fuzzy interpolation, fuzzy rule base interpolation

1 Introduction

In the following, we will deal with a problem of a fuzzy interpolation. We will first recall a classical approach to interpolation, because the problem of fuzzy interpolation closely relates to it.

Let f be a real function of a real argument with a domain $\mathcal{M} = \{x_i,\ i = 1,\dots,n\} \subset \mathbb{R}$, and $\{(x_i,\ f(x_i)),\ i = 1,\dots,n\} \subseteq \mathbb{R}$ be the interpolation data. Let $\mathcal{M} \subset P \subseteq \mathbb{R}$. An interpolation function $g : P \longrightarrow \mathbb{R}$ is a function that fulfils the interpolation condition:

$$f(x_i) = g(x_i),\ i = 1,\dots,n.$$

In this paper, we will be focused on the interpolation approach to a computation with fuzzy data. A precise definition of *interpolation of fuzzy data* will be given below in the subsection 2.4. Freely speaking, this is a problem of extension of a fuzzy function given on a restricted domain to a fuzzy function given on a wider domain (similar to the case considered above).

There are other approaches to the problem of fuzzy interpolation. They differ one from the other one by restrictions on interpolation functions. The following list remembers the most popular approaches : level cuts interpolation [17, 18], analogy-based interpolation [3, 4, 6], interpolation by convex completion [8, 23], interpolation by geometric transformations [1], interpolation in a family of interpolating relations [2], polar cut interpolation [15], interpolation based on closeness relations [5], flank functions interpolation [13, 14], analytic fuzzy relation-based interpolation [20], and fuzzy interpolation based on fuzzy functions [11].

The sections below are arranged as follows : basic concepts as well as definition of fuzzy interpolation will be given in Section 2. In Section 3, we will recall the approach to fuzzy interpolation introduced by Godo, Esteva, ets. in [5]. The last Section 4 is devoted to a new approach.

2 Fuzzy Interpolation and Its Analytic Representation

In this section, we will introduce the problem of fuzzy interpolation as a problem of extension of a partially given fuzzy function. Moreover, we expect that a solution will be represented analytically with the help of a structure known as residuated lattice.

First of all, we will recall the notion of residuated lattice, then we will introduce the notion of fuzzy space and fuzzy function. Finally, we will discuss a certain class of interpolating fuzzy functions and their analytical representation.

2.1 Residuated Lattice

A residuated lattice is an ordered algebraic structure with two residuated binary operation. We will recall its definition from [19].

Definition 1
A residuated lattice[*)] *is an algebra*

$$\mathcal{L} = \langle L, \vee, \wedge, *, \rightarrow, \mathbf{0}, \mathbf{1} \rangle.$$

with a support L and four binary operations and two constants such that

- $\langle L, \vee, \wedge, \mathbf{0}, \mathbf{1} \rangle$ *is a lattice where the ordering $\leq$ defined using operations $\vee, \wedge$ as usual, and $\mathbf{0}, \mathbf{1}$ are the least and the greatest elements, respectively;*
- $\langle L, *, \mathbf{1} \rangle$ *is a commutative monoid, that is, $*$ is a commutative and associative operation with the identity $a * \mathbf{1} = a$;*
- *the operation $\rightarrow$ is a residuation operation with respect to $*$, i.e.,*

$$a * b \leq c \quad \textit{iff} \quad a \leq b \rightarrow c.$$

A residuated lattice is complete if its underlying lattice is complete.

The derived operation is biresiduum:

$$a \leftrightarrow b = (a \rightarrow b) \wedge (b \rightarrow a).$$

Our investigation will be based on Łukasiewicz algebra $\mathcal{L}_{\text{Ł}}$. It is a residuated lattice with the support $L = [0, 1]$ where

$$\begin{aligned} a * b &= 0 \vee (a + b - 1), \\ a \rightarrow b &= 1 \wedge (1 - a + b), \\ a \leftrightarrow b &= 1 - \mid a - b \mid . \end{aligned}$$

The other well known examples of residuated lattice are Boolean algebra, Gödel algebra and product algebra.

2.2 L-fuzzy Space

Assume that we are given a complete residuated lattice $\mathcal{L}$ and a non-empty universal set $X \subseteq \mathbb{R}$ where $\mathbb{R}$ is the set of real numbers. An *L-valued fuzzy set* is a mapping $A : X \rightarrow L$. A *core* of a fuzzy set A is the set $\text{Core}(A) = \{x \in X \mid A(x) = 1\}$. We say that a fuzzy set is *normal* if there exists $x_A \in X : \ A(x_A) = 1$. The class of L-valued fuzzy sets of X will be denoted by L^X.

Let $A, \ B \in L^X$ be fuzzy sets. A *fuzzy equality* $(A \equiv B)$ is given by the following formula

$$(A \equiv B) = \bigwedge_{x \in X} (A(x) \leftrightarrow B(x))^{\dagger)}. \tag{1}$$

*)In this paper we assume a residuated lattice to be bounded, commutative and integral.

†)for a general definition of fuzzy equality, e.g., [7, 12, 16]

The fuzzy equality determines a degree of coincidence of two fuzzy sets expressed by an element of the residuated lattice.

It is known that

$$(A \equiv B) = 1 \quad \text{iff} \quad \forall x \in X, \quad A(x) = B(x),$$

In this case we will write $A = B$ instead of $A \equiv B$.

Definition 2
The pair $(L^X, \equiv)$ is a fuzzy space on X.

The definition of fuzzy space introduces a basic set of objects together with the basic relation of equality.

Example 1
Let $X = [a, b]$, $\mathcal{L} = \mathcal{L}_{\mathbf{L}}$. Let us show how the fuzzy equality $\equiv$ is expressed. Assume that $A,\ B \in [0, 1]^{[a,b]}$.

$$A \equiv B = \bigwedge_{x\in[a,b]} (A(x) \leftrightarrow B(x)) = \bigwedge_{x\in[a,b]} (1- \mid A(x) - B(x) \mid) = 1 - \bigvee_{x\in[a,b]} \mid A(x) - B(x) \mid .$$

Thus the pair $([0, 1]^{[a,b]},\ 1-\bigvee_{x\in[a,b]} \mid A(x)-B(x) \mid)$ is a fuzzy space on $[a, b]$ determined by Łukasiewicz algebra.

In the special case where fuzzy sets on $[a, b]$ are continuous mappings, the fuzzy equality $\equiv$ between them can be simplified to

$$1 - \bigvee_{x\in[a,b]} \mid A(x) - B(x) \mid = 1 - \max_{x\in[a,b]} \mid A(x) - B(x) \mid = 1 - d(A, B),$$

where $d(A, B)$ is the distance in the metric space of continuous functions.

2.3 Fuzzy Function

We will use the notion of fuzzy function introduced in [20]. According to [20], a fuzzy function is an ordinary mapping between two fuzzy spaces. In more details,

Definition 3
Let $\mathcal{L}$ be a complete residuated lattice and $(L^X, \equiv)$, $(L^Y, \equiv)$ fuzzy spaces on X and Y, respectively. A mapping $f : L^X \longrightarrow L^Y$ is a fuzzy function if for every $A, B \in L^X$,

$$A = B \text{ implies } f(A) = f(B). \tag{2}$$

Let us remark that there are other definitions of fuzzy function in [7, 9, 10, 16] where fuzzy function is defined as a fuzzy set of function or as a special fuzzy relation.

Below, we give an example of a fuzzy function which is reproduced from [21].

Example 2 (Fuzzy functions determined by a fuzzy relation)
Let $(L^X, \equiv)$, $(L^Y, \equiv)$ be fuzzy spaces on X and Y respectively, $R \in L^{X\times Y}$ a fuzzy relation. For every $A \in L^X$, we define the $\circ$-composition *of A and R by*

$$(A \circ R)(y) = \bigvee_{x\in X} (A(x) * R(x, y)). \tag{3}$$

Composition (3) determines the fuzzy set $A \circ R$ on Y. The corresponding mapping $f_{\circ R} : A \mapsto A \circ R$ is a fuzzy function defined on the whole fuzzy space L^X.

2.4 Fuzzy Interpolation

The problem of fuzzy interpolation includes two subproblems: a choice of a set of interpolation functions and an extension of an original fuzzy function.

In other words, let $\{(A_i, B_i),\ i = 1, \ldots, n\}$ be a set of fuzzy data and $A_i \in L^X,\ i = 1, \ldots, n$ are pairwise different fuzzy sets with respect to $=$, $B_i \in L^Y,\ i = 1, \ldots, n$. Let a fuzzy function $f : A_i \to B_i,\ i = 1, \ldots, n$ have the domain $\mathcal{M} = \{A_1, \ldots, A_n\}$, and P be a domain of an interpolation fuzzy function g where $\mathcal{M} \subset P \subseteq L^X$. Let $\mathcal{N} \subseteq \{g \mid g :\ P \longrightarrow L^Y\}$ be a chosen subset of a fuzzy function for the fuzzy interpolation. Our goal is to find an fuzzy function $g \in \mathcal{N}$ satisfying the interpolation condition

$$g(A_i) = B_i,\ i = 1, \ldots, n. \tag{4}$$

The fuzzy function g is called *an interpolation fuzzy function* for fuzzy data. Also we call the interpolation fuzzy function g an extension of f on the domain P.

We can also rewrite the interpolation condition (4) as follows:

$$A = A_i \text{ implies } g(A) = B_i,\ i = 1, \ldots, n. \tag{5}$$

2.5 Similarity and Fuzzy Point

A binary fuzzy relation E on X is called a similarity on X if for all $x, y, z \in X$, the following properties hold:

1. $E(x, x) = 1$,
2. $E(x, y) = E(y, x)$,
3. $E(x, y) * E(y, z) \leq E(x, z)$.

Let E be a similarity on X. A fuzzy set $E_t,\ t \in X$, where $E_t(x) = E(t, x)$ for all $x \in X$ is called an *E-fuzzy point* of X.

3 Fuzzy Rule Base Interpolation

In this contribution, we will investigate another approach to fuzzy interpolation, proposed in [5]. It assumes that an original function is expressed by a set of fuzzy IF-THEN rules

$$RB = \{\text{“If } x \text{ is } A_i \text{ then } y \text{ is } B_i\text{”}\}_{i=1,\ldots,n} \tag{6}$$

(A_i and B_i are respective fuzzy sets on X and Y), and the rules are sparse in the sense that $A_i \cap A_j = \emptyset$, $i \neq j$. The fuzzy interpolation is proposed to be realized in a form of an algorithm which produces a consequence B to an antecedence A (A and B are fuzzy sets too). An interpolating algorithm should respect the following requirement:

$$\begin{aligned} &\text{“The more the input } A \text{ is } \textit{close to } A_i \\ &\text{ the more the output } B \text{ must be } \textit{close to } B_i\text{.”} \end{aligned} \tag{7}$$

In [5], a general interpolating algorithm is proposed. Below, we give its essential details that characterize relations of closeness on both universes and describe the way of computing B according to the requirement (7).

The closeness relations between two fuzzy sets show how much one is similar or is included into the other one. Let $S = \{S_\lambda;\ 0 \leq \lambda \leq\ +\infty\}$, be any nested family of fuzzy similarity relations on $\mathbb{R}$ such that S_0 is the crisp equality and $S_{+\infty}$=1. Then

$$\text{close}_\lambda(E, D) = min(I_{S_\lambda}(D|E), I_{S_\lambda}(E|D)), \tag{8}$$

where

$$I_{S_\lambda}(D|E) = inf_{u\in\mathbb{R}}\left\{E(u) \to (S_\lambda \circ D)(u)\right\}. \tag{9}$$

According to [5], the value of (algorithmically defined) interpolating function at point A is fuzzy set B such that

$$B = \text{Interpol}_{RB}(A) = \cap_{R_i\in K(A)} \cap_{\lambda\geq 0} \text{close}_\lambda(A, A_i) \to (S_{f(\lambda)} \circ B_i), \tag{10}$$

where $K(A)$ is a subset of fuzzy rules related to A, $\to$ is the residuum of a left-continuous t-norm $*$, and $\circ$ is the max-* composition. Moreover, for each $0 < \lambda \leq +\infty$,

$$f(\lambda) = \inf\left\{\mu \mid \text{close}_\lambda(A_1, A_2) \leq \text{close}_\mu(B_1, B_2)\right\}, \tag{11}$$

where μ is any parameter which satisfies the inequality $\text{close}_\lambda(A_1, A_2) \leq \text{close}_\mu(B_1, B_2)$. It is proved in [5] that thus proposed algorithm fulfils the requested requirement (7).

It is seen from the description above, that a specification of the algorithm requires a choice of a parametric family of fuzzy similarity relations *S* and operations from a certain residuated lattice. One partial specification was proposed in [5] as well. It is based on an arbitrary left-continuous t-norm $*$ and the parametric family of fuzzy similarity relations on $\mathbb{R}^2$:

$$S_\lambda(x, y) = \max(1 - \frac{\mid x - y \mid}{\lambda}, 0). \tag{12}$$

4 Main Result

Our purpose it to show that the interpolation algorithm presented in [5] and based on the fuzzy prescription (7) satisfies the interpolation condition in the sense (4). It means that the interpolation function $\text{Interpol}_{RB}(A)$ (cf. (10)) fulfils the interpolation condition in the form

$$(\forall A_i) \quad \text{Interpol}_{RB}(A_i) \ =\ B_i,\ i = 1, \ldots, n.$$

4.1 Assumptions and Preliminaries

The Łukasiewicz algebra is chosen as an underlying residuated lattice. Without lost of generality, we assume that only two IF-THEN fuzzy rules specify an original function so that the subset of IF-THEN fuzzy rules connected with A is

$$K(A) = \{A_1 \to B_1, A_2 \to B_2\}.$$

Let $A_1,\ A_2$ be normal and triangular shaped fuzzy sets (inputs) defined on $X \subset \mathbb{R}$ and $B_1,\ B_2$ be normal and triangular shaped fuzzy sets (outputs) defined on $Y \subset \mathbb{R}$. Obviously, the core of a triangular shaped fuzzy set consists of one element – the *core point*. Denote core points of $A_1,\ A_2$ by x_{A_1} and x_{A_2}, and similarly, core points of $B_1,\ B_2$ by y_{B_1} and y_{B_2}, respectively. Denote (arbitrary) normal and triangular shaped fuzzy set on $X \subset \mathbb{R}$ by A, and its core point by x_A. Our aim is to prove that the fuzzy set B given by (10) fulfils (4).

Let S_λ be a similarity on X and $\lambda \geq 0$ a fixed real number. The composition between S_λ and A is given by

$$(S_\lambda \circ A)(x) = \bigvee_u (S_\lambda(x, u) * A(u)).$$

In the following proposition we will show how the similarity relation S_λ affects a fuzzy set.

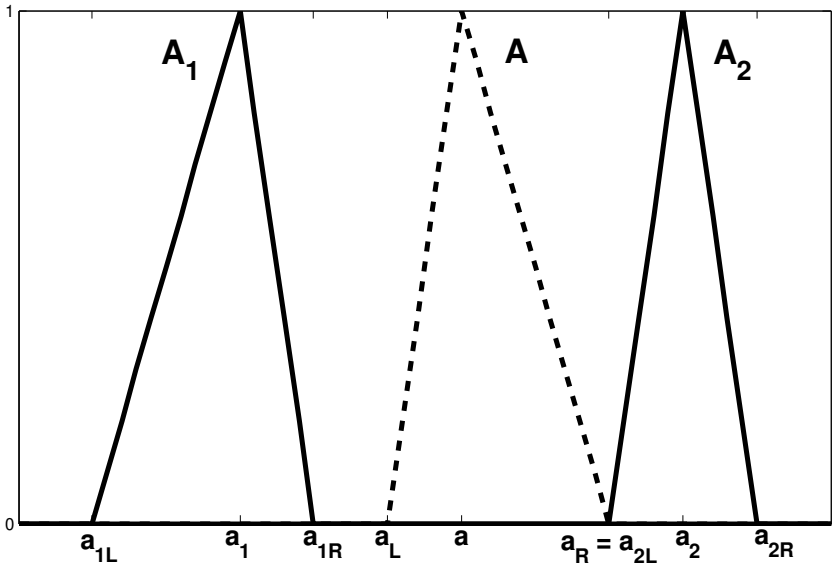

Figure 1: Fuzzy sets on X

Proposition 1
Let A be a normal fuzzy set and $x_A \in X$ be its core point. Let moreover, $S^{\lambda}_{x_A}$ be the S_λ-fuzzy point determined by x_A, i.e. $S^{\lambda}_{x_A} = S_\lambda(x_A, x)$. Then for all $x \in X$:

1. $(S_\lambda \circ A)(x) \geq A(x)$,
2. $(S_\lambda \circ A)(x) \geq S^{\lambda}_{x_A}(x)$,
3. *there exists* $\lambda^* \geq 0$ *such that* $A(x) \leq S^{\lambda^*}_{x_A}(x)$ *and then* $(S_{\lambda^*} \circ A)(x) = S^{\lambda^*}_{x_A}(x)$.

PROOF: By the assumption, $A(x_A) = 1$. We will use properties of a similarity relation and obtain:

1. $(S_\lambda \circ A)(x) = \bigvee_u (S_\lambda(x,u) * A(u)) \geq (S_\lambda(x,x) * A(x)) = A(x)$.
2. $(S_\lambda \circ A)(x) = \bigvee_u (S_\lambda(x,u) * A(u)) \geq (S_\lambda(x,x_A) * A(x_A)) = S_\lambda(x,x_A) = S_\lambda(x,x_A) = S^{\lambda}_{x_A}(x)$.
3. Assume that $A(u) \leq S^{\lambda^*}_{x_A}(u)$ so that the following holds: $(S_{\lambda^*} \circ A)(x) = \bigvee_u (S_{\lambda^*}(x,u) * A(u)) \leq \bigvee_u (S_{\lambda^*}(x,u) * S_{\lambda^*}(x_A,u)) = \bigvee_u (S_{\lambda^*}(x,u) * S_{\lambda^*}(u,x_A)) = S_{\lambda^*}(x,x_A) = S_{\lambda^*}(x_A,x)$. On the other side, $(S_{\lambda^*} \circ A)(x) \geq S^{\lambda^*}_{x_A}(x)$ for $\lambda^* \geq 0$. Therefore, $(S_{\lambda^*} \circ A)(x) = S^{\lambda^*}_{x_A}(x)$.

□

By the proposition above, we can rewrite (8) as follows:

$$\begin{aligned} \text{close}_\lambda(A_1, A_2) &= min(I_{S_{\lambda^*}}(A_2|A_1), I_{S_{\lambda^*}}(A_1|A_2)) = \\ &= \bigwedge_x (S_{\lambda^*} \circ A_1 \leftrightarrow S_{\lambda^*} \circ A_2) \end{aligned}$$

where

$$\begin{aligned} I_{S_\lambda}(A_2|A_1) &= inf_{x \in \mathbb{R}} \{A_1(u) \to (S_{\lambda^*} \circ A_2)(u)\} = \\ &= inf_{x \in \mathbb{R}} \{S_{\lambda^*}(x_{A_1}, x) \to S_{\lambda^*}(x_{A_2}, x)\}, \end{aligned}$$

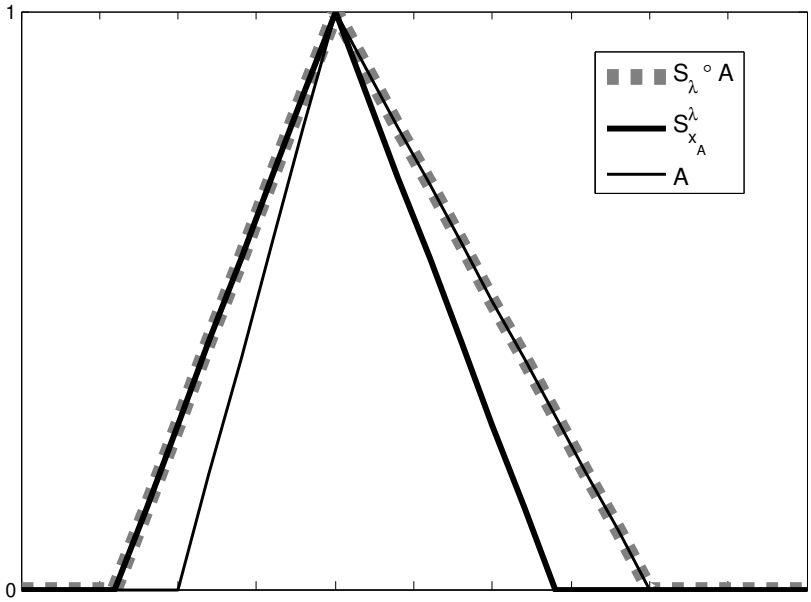

Figure 2: Properties $(S_\lambda \circ A)(x) \geq A(x)$ and $(S_\lambda \circ A)(x) \geq S^\lambda_{x_A}(x)$

and similarly $I_{S_\lambda}(A_1|A_2)$.

Let us simplify the expression (10) by applying assumptions that are accepted at the beginning of this subsection. Moreover, we replace similarities by distances (similar approach has been used in [22]). The distance between two triangular shaped fuzzy sets is considered as a distance between their core points[‡)]:

$$d(A, A_1) = \mid x_A - x_{A_1} \mid,\ d(A, A_2) = \mid x_A - x_{A_2} \mid,$$
$$d(A_1, A_2) = \mid x_{A_1} - x_{A_2} \mid$$

and

$$d(B_1, B_2) = \mid y_{B_1} - y_{B_2} \mid .$$

Proposition 2
Let $A_1,\ A_2$ be normal and triangular shaped fuzzy sets with $x_{A_1},\ x_{A_2} \in X$ as respective cores. Let S_{λ^*} be given by (12), and $\lambda^* \geq 0$. Then

$$\bigwedge_x (S_{\lambda^*} \circ A_1 \leftrightarrow S_{\lambda^*} \circ A_2) = S_{\lambda^*}(x_{A_1}, x_{A_2}).$$

PROOF: We use the following property of the absolute value: $\mid a - b \mid \geq \mid\mid a \mid - \mid b \mid\mid$ or equivalently, $-(\mid\mid a \mid - \mid b \mid\mid) \geq -(\mid a \mid - \mid b \mid)$.

[‡)]Recall that each triangular shaped fuzzy set has exactly one core point so that our definition of a distance is correct.

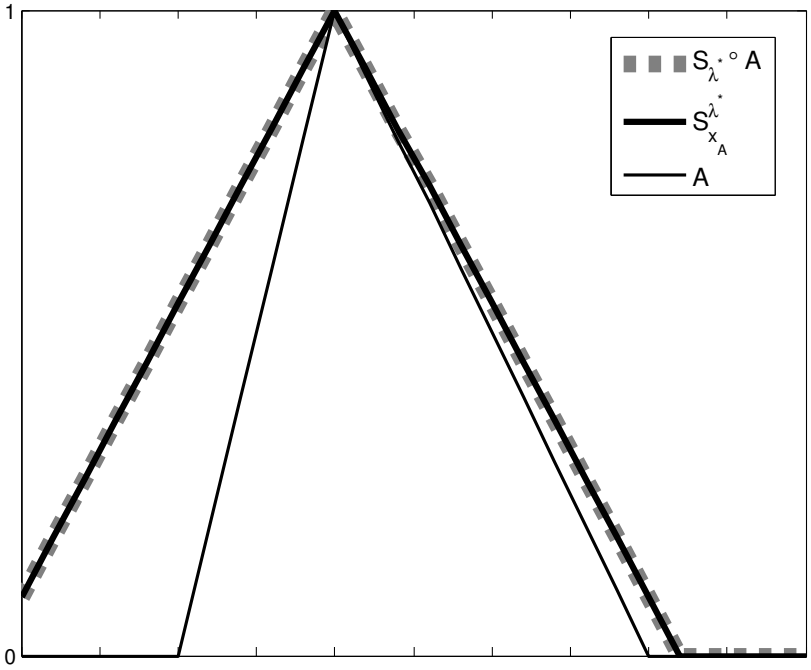

Figure 3: Property $A(x) \leq S^{\lambda^*}_{x_A}(x)$, $(S_{\lambda^*} \circ A)(x) = S^{\lambda^*}_{x_A}(x)$

$$\bigwedge_x (S_{\lambda^*} \circ A_1 \leftrightarrow S_{\lambda^*} \circ A_2) = \bigwedge_x (S^{\lambda^*}_{x_{A_1}}(x) \leftrightarrow S^{\lambda^*}_{x_{A_2}}(x)) = \bigwedge_x (1 - \mid S_{\lambda^*}(x_{A_1}, x) -$$

$$S_{\lambda^*}(x_{A_2}, x) \mid) = \bigwedge_x \left(1 - \left|1 - \frac{\mid x_{A_1} - x \mid}{\lambda^*} - 1 + \frac{\mid x_{A_2} - x \mid}{\lambda^*}\right|\right) =$$

$$\bigwedge_x \left(1 - \left|\frac{\mid x_{A_2} - x \mid}{\lambda^*} - \frac{\mid x_{A_1} - x \mid}{\lambda^*}\right|\right) = \bigwedge_x \left(1 - \frac{1}{\lambda^*} \mid\mid x_{A_2} - x \mid - \mid x_{A_1} - x \mid\mid\right)$$

$$= 1 - \bigvee_x \frac{1}{\lambda^*}(\mid\mid x_{A_2} - x \mid - \mid x_{A_1} - x \mid\mid) \geq 1 - \bigvee_x \frac{1}{\lambda^*}(|x_{A_2} - x - x_{A_1} + x|) =$$

$$1 - \bigvee_x \frac{1}{\lambda^*}(|x_{A_2} - x_{A_1}|) = 1 - \frac{1}{\lambda^*}(|x_{A_2} - x_{A_1}|) = S_{\lambda^*}(x_{A_1}, x_{A_2})$$

Assume that $x \leq x_{A_1} \leq x_{A_2}$. Three cases are possible.

1. $x \leq x_{A_1} \leq x_{A_2}$. In this case, $\mid x_{A_2} - x \mid = x_{A_2} - x$. Similarly for $\mid x_{A_1} - x \mid$.

$$1 - \bigvee_x \frac{1}{\lambda^*}(\mid\mid x_{A_2} - x \mid - \mid x_{A_1} - x \mid\mid) = 1 - \bigvee_x \frac{1}{\lambda^*}(|x_{A_2} - x - x_{A_1} + x|)$$

$$= 1 - \bigvee_x \frac{1}{\lambda^*}(|x_{A_2} - x_{A_1}|) = 1 - \frac{1}{\lambda^*}(|x_{A_2} - x_{A_1}|) = S_{\lambda^*}(x_{A_1}, x_{A_2})$$

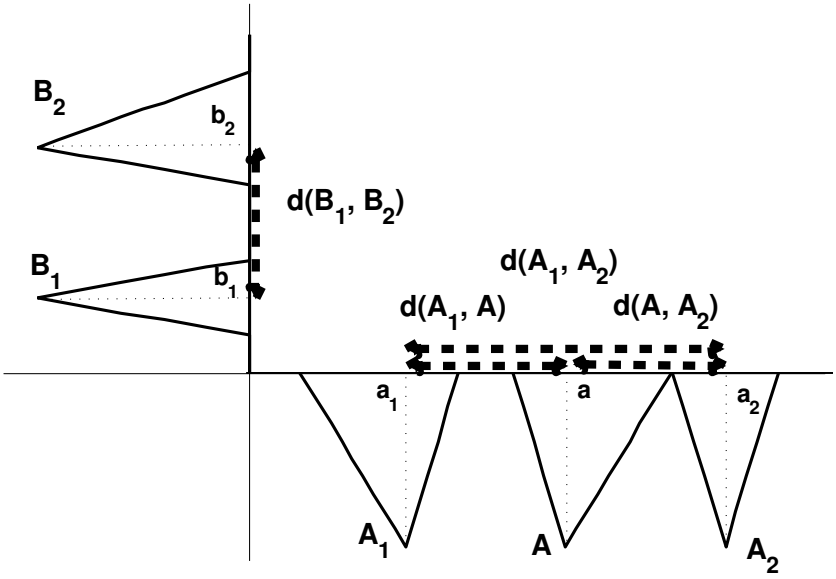

Figure 4: Distances between fuzzy sets

2. Assume that $x_{A_1} \leq x_{A_2} \leq x$ so that $\mid x_{A_2} - x \mid = x - x_{A_2}$, and similarly for $\mid x_{A_1} - x \mid$.

$$1 - \bigvee_x \frac{1}{\lambda^*}(|| \mid x_{A_2} - x \mid - \mid x_{A_1} - x \mid ||) = 1 - \bigvee_x \frac{1}{\lambda^*}(|-x_{A_2} + x + x_{A_1} - x|)$$
$$= 1 - \bigvee_x \frac{1}{\lambda^*}(|x_{A_1} - x_{A_2}|) = 1 - \frac{1}{\lambda^*}(|x_{A_2} - x_{A_1}|) = S_{\lambda^*}(x_{A_1}, x_{A_2})$$

3. Finally, let $x_{A_1} \leq x \leq x_{A_2}$ so that$\mid x_{A_2} - x \mid$ is equal to $x_{A_2} - x$. The absolute value $\mid x_{A_1} - x \mid$ is equal to $\mid x_{A_1} - x \mid = x - x_{A_1}$. Without lost of generality, let us choose $x = x_{A_1}$.

$$\bigwedge_x \left(1 - \frac{1}{\lambda^*}(|| \mid x_{A_2} - x \mid - \mid x_{A_1} - x \mid ||)\right) =$$
$$\bigwedge_x \left(1 - \frac{1}{\lambda^*}(|x_{A_2} - x + x_{A_1} - x|)\right) \leq 1 - \frac{1}{\lambda^*}(|x_{A_2} - x_{A_1} + x_{A_1} - x_{A_1}|) =$$
$$1 - \frac{1}{\lambda^*}(|x_{A_2} - x_{A_1}|) = 1 - \frac{1}{\lambda^*}(|x_{A_2} - x_{A_1}|) = S_{\lambda^*}(x_{A_1}, x_{A_2})$$

□

4.2 The Main Result

The following are assumptions of the main result:

1. Let $\lambda_1 \leq \lambda_2$ then $S_{\lambda_1} \leq S_{\lambda_2}$,
2. A_1, A_2 are normal fuzzy sets and $x_{A_1} \leq x_{A_2}$,
3. $\exists \lambda^* \geq 0 : \forall i = 1, 2 \quad A_i \leq S^{\lambda^*}_{x_{A_i}}$,

4. $S^{\lambda^*}_{x_{A_1}} \wedge S^{\lambda^*}_{x_{A_2}} = 0$.

We will describe and concretize the other parts of the expression (10) by means of distances.

Let $\lambda^* \leq \lambda \leq +\infty$. Let us remind that the same parametric family of fuzzy similarity relations on $\mathbb{R}^2$ is given by (12).

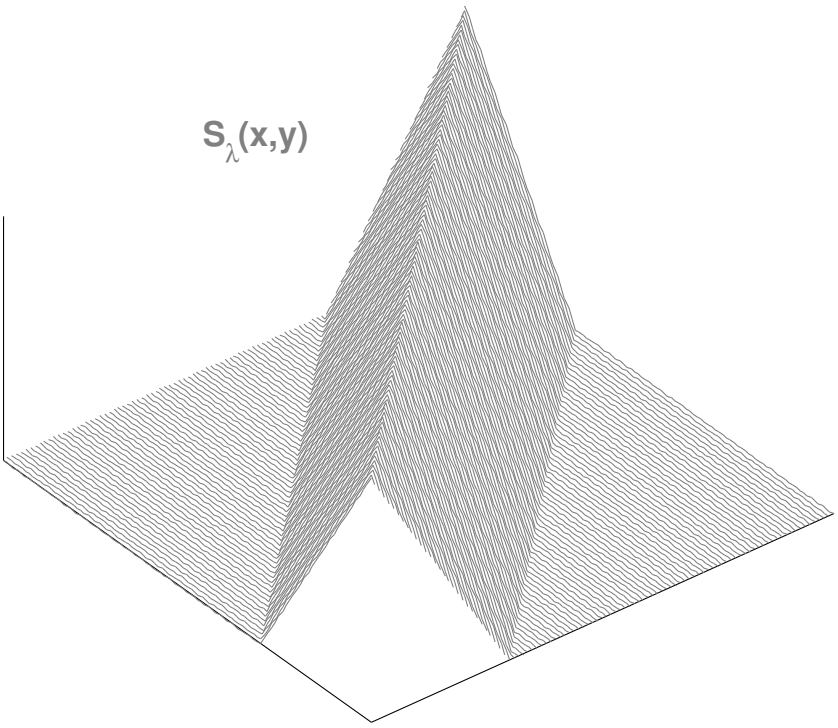

Figure 5: Similarity relation

The idea is that we extend the fuzzy set A_1 by applying to it S_λ where $\lambda^* \leq \lambda \leq +\infty$.

Now we rewrite the expression (8) (degree of closeness) using distances. For each $0 < \lambda \leq +\infty$,

$$\text{close}_\lambda(A_1, A_2) = min(I_{S_\lambda}(A_2|A_1), I_{S_\lambda}(A_1|A_2)) = \frac{\lambda - \mid x_{A_1} - x_{A_2} \mid}{\lambda}, \tag{13}$$

and respectively,

$$\text{close}_\lambda(A, A_i) = \frac{\lambda - \mid x_A - x_{A_i} \mid}{\lambda}, \quad i = 1, 2. \tag{14}$$

The respective value $f(\lambda)$ can now be expressed as

$$f(\lambda) = \frac{\lambda \mid y_{B_1} - y_{B_2} \mid}{\mid x_{A_1} - x_{A_2} \mid}. \tag{15}$$

The expression (15) is equivalent to (11). However, (15) is represented with the help of distances and by this, its meaning is clearer.

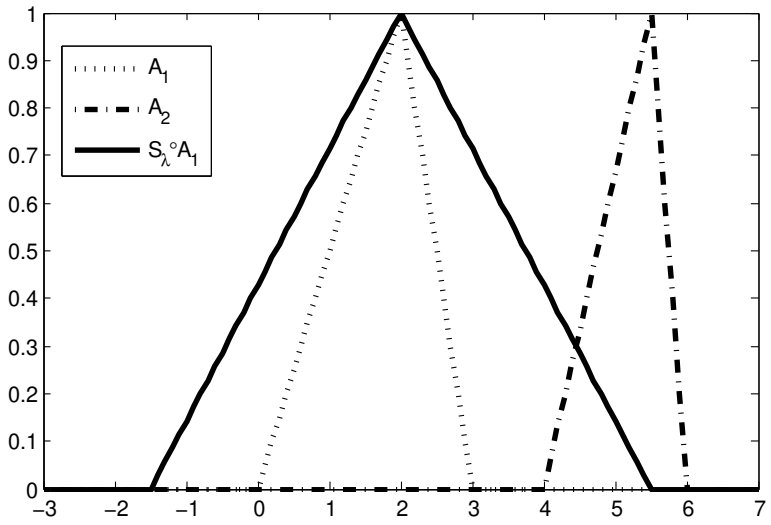

Figure 6: Extension fuzzy set A_1

Finally, we will characterize the fuzzy set B given by (10) with the help of distances too.

$$B = \bigwedge_{i=1}^{2} \bigwedge_{\lambda} (\text{close}_\lambda(A, A_i) \to (S_{f(\lambda)} \circ B_i)(y)) =$$

$$\bigwedge_{i=1}^{2} \bigwedge_{\lambda} \left(1 - 1 + \frac{|\, x_A - x_{A_i} \,|}{\lambda} + (S_{f(\lambda)} \circ B_i)(y)\right) =$$

$$\bigwedge_{i=1}^{2} \bigwedge_{\lambda} \left(\frac{|\, x_A - x_{A_i} \,|}{\lambda} + (S_{f(\lambda)} \circ B_i)(y)\right)$$

Now, we can prove that the interpolation condition (5) is fulfilled.

Theorem 1
If $A = A_i$ then $B = B_i$ for $i = 1, 2$.

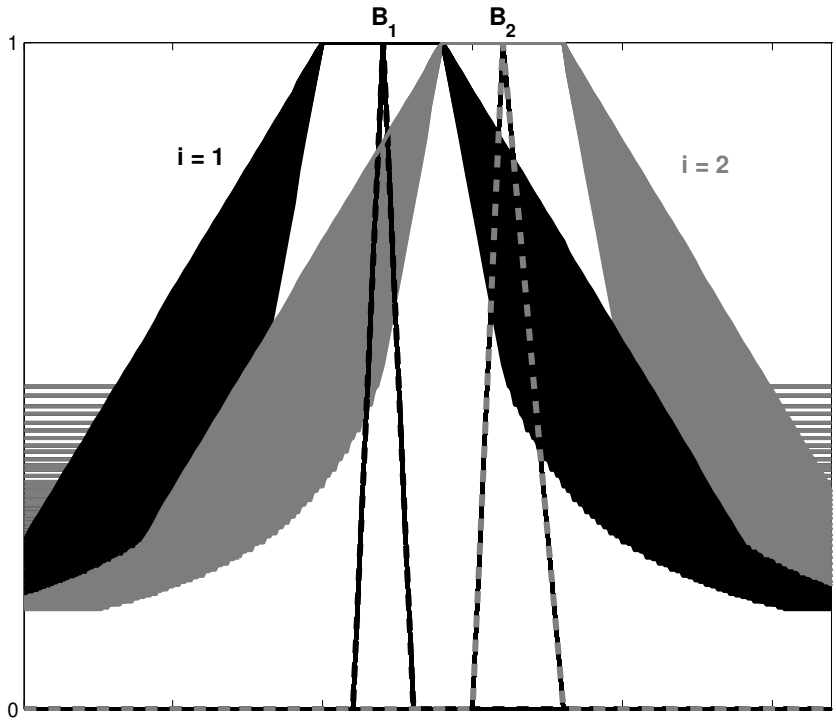

Figure 7: Construction of output fuzzy set B

PROOF: Conclusion

$$B = \bigwedge_{i=1}^{2} \bigwedge_{\lambda} (\text{close}_\lambda(A, A_i) \to (S_{f(\lambda)} \circ B_i)(y)) =$$

$$\bigwedge_{i=1}^{2} \bigwedge_{\lambda} \left(0 \vee \left(1 - 1 + \frac{\mid x_A - x_{A_i} \mid}{\lambda} + (S_{f(\lambda)} \circ B_i)(y)\right)\right) =$$

$$\bigwedge_{i=1}^{2} \bigwedge_{\lambda} \left(0 \vee \left(\frac{\mid x_A - x_{A_i} \mid}{\lambda} + (S_{f(\lambda)} \circ B_i)(y)\right)\right) =$$

$$\left[\bigwedge_{\lambda} \left(\frac{\mid x_A - x_{A_1} \mid}{\lambda} + (S_{f(\lambda)} \circ B_1)(y)\right)\right] \bigwedge$$

$$\left[\bigwedge_{\lambda} \left(\frac{\mid x_A - x_{A_2} \mid}{\lambda} + (S_{f(\lambda)} \circ B_2)(y)\right)\right] = B' \wedge B''$$

For each $i = 1, 2$, we will prove that $A = A_i \ \Rightarrow \ B = B_i$.

Let $A = A_1$.

$$B' = \bigwedge_{\lambda}(\text{close}_\lambda(A_1, A_1) \to (S_{f(\lambda)} \circ B_1)(y)) =$$

$$\bigwedge_{\lambda}\left(\frac{\mid x_{A_1} - x_{A_1} \mid}{\lambda} + (S_{f(\lambda)} \circ B_1)(y)\right) =$$

$$\bigwedge_{0<\lambda\leq\lambda'}\left(\frac{\mid x_{A_1} - x_{A_1} \mid}{\lambda} + (S_{f(\lambda)} \circ B_1)(y)\right) \wedge$$

$$\bigwedge_{\lambda>\lambda'}\left(\frac{\mid x_{A_1} - x_{A_1} \mid}{\lambda} + (S_{f(\lambda)} \circ B_1)(y)\right) =$$

$$\bigwedge_{0<\lambda\leq\lambda'}\big(0 + (S_{f(\lambda)} \circ B_1)(y)\big) \wedge \bigwedge_{\lambda>\lambda'}\big(0 + (S_{f(\lambda)} \circ B_1)(y)\big) =$$

$$B_1(y) \wedge \left[\bigwedge_{\lambda>\lambda'}\big(0 + (S_{f(\lambda)} \circ B_1)(y)\big)\right] = B_1(y) \wedge \left[\bigwedge_{\lambda>\lambda'}\big((S_{f(\lambda)}(y_{B_1}, y)\big)\right] = B_1$$

The latter equality follows from $\bigwedge_{0<\lambda\leq\lambda'}\big((S_{f(\lambda)} \circ B_1)(y)\big) = B_1$ which can be justified by the following chain of inequalities:

$$\begin{aligned}
&f(\lambda') \leq \mid y_{B_1} - y \mid,\\
&\lambda' \frac{\mid y_{B_1} - y_{B_2} \mid}{\mid x_{A_1} - x_{A_2} \mid} \leq \mid y_{B_1} - y \mid,\\
&\lambda' \frac{\mid y_{B_1} - y_{B_2} \mid}{\mid x_{A_1} - x_{A_2} \mid} \leq \mid y_{B_1} - y \mid,\\
&\lambda' \leq \mid y_{B_1} - y \mid \frac{\mid x_{A_1} - x_{A_2} \mid}{\mid y_{B_1} - y_{B_2} \mid}.
\end{aligned}$$

$$B'' = \bigwedge_\lambda (\text{close}_\lambda(A_1, A_2) \to (S_{f(\lambda)} \circ B_2)(y)) =$$

$$\bigwedge_\lambda \left(\frac{\mid x_{A_1} - x_{A_2} \mid}{\lambda} + (S_{f(\lambda)} \circ B_2)(y) \right) =$$

$$\bigwedge_\lambda \left(\frac{\mid x_{A_1} - x_{A_2} \mid}{\lambda} + (1 - \frac{\mid y_{B_2} - y \mid}{f(\lambda)} \vee 0) \right) =$$

$$\bigwedge_\lambda \left(1 + \frac{\mid x_{A_1} - x_{A_2} \mid}{\lambda} - \frac{\mid y_{B_2} - y \mid}{f(\lambda)} \right) =$$

$$\bigwedge_\lambda \left(1 - \left(\frac{\mid y_{B_2} - y \mid}{f(\lambda)} - \frac{\mid x_{A_1} - x_{A_2} \mid}{\lambda} \right) \right) =$$

$$\bigwedge_\lambda \left(1 - \left(\frac{\mid y_{B_2} - y \mid}{\frac{\lambda \mid y_{B_1} - y_{B_2} \mid}{\mid x_{A_1} - x_{A_2} \mid}} - \frac{\mid x_{A_1} - x_{A_2} \mid}{\lambda} \right) \right) =$$

$$\bigwedge_\lambda \left(1 - \left(\frac{\mid y_{B_2} - y \mid\mid x_{A_1} - x_{A_2} \mid}{\lambda \mid y_{B_1} - y_{B_2} \mid} - \frac{\mid x_{A_1} - x_{A_2} \mid}{\lambda} \right) \right) =$$

$$\bigwedge_\lambda \left(1 - \frac{1}{\lambda} \left(\frac{\mid y_{B_2} - y \mid\mid x_{A_1} - x_{A_2} \mid}{\mid y_{B_2} - y_{B_1} \mid} - \mid x_{A_1} - x_{A_2} \mid \right) \right) =$$

$$\bigwedge_\lambda \left(1 - \frac{1}{\lambda} \left(\left| \frac{(y_{B_2} - y)(x_{A_1} - x_{A_2})}{(y_{B_2} - y_{B_1})} \right| - \mid x_{A_1} - x_{A_2} \mid \right) \right) \geq$$

$$\bigwedge_\lambda \left(1 - \frac{1}{\lambda} \left(\left| \frac{(y_{B_2} - y)(x_{A_1} - x_{A_2}) - ((x_{A_1} - x_{A_2})(y_{B_2} - y_{B_1}))}{(y_{B_2} - y_{B_1})} \right| \right) \right) =$$

$$\bigwedge_\lambda \left(1 - \frac{1}{\lambda} \left(\left| \frac{-y(x_{A_1} - x_{A_2}) + x_{A_1} y_{B_1} - x_{A_2} y_{B_1}}{(y_{B_2} - y_{B_1})} \right| \right) \right) =$$

$$\bigwedge_\lambda \left(1 - \frac{1}{\lambda} \left(\left| \frac{y_{B_1}(x_{A_1} - x_{A_2}) - y(x_{A_1} - x_{A_2})}{(y_{B_2} - y_{B_1})} \right| \right) \right) =$$

$$\bigwedge_\lambda \left(1 - \frac{1}{\lambda} \left(\left| \frac{(x_{A_1} - x_{A_2})(y_{B_1} - y)}{(y_{B_2} - y_{B_1})} \right| \right) \right) =$$

$$\bigwedge_\lambda \left(1 - \frac{1}{\lambda} \left(\left| \frac{(x_{A_1} - x_{A_2})}{(y_{B_2} - y_{B_1})} \right| \mid y_{B_1} - y \mid \right) \right) =$$

$$\bigwedge_\lambda \left(1 - \frac{1}{\lambda} \left| \frac{(x_{A_1} - x_{A_2})}{(y_{B_1} - y_{B_2})} \right| (\mid y_{B_1} - y \mid) \right) =$$

$$\bigwedge_\lambda \left(\left(1 - \frac{1}{f(\lambda)} (\mid (y_{B_1} - y) \mid) \right) \vee 0 \right) =$$

$$\bigwedge_\lambda \left(S_{f(\lambda)} \circ B_1 \right)$$

Finally,

$$A = A_1 \Rightarrow B = B' \wedge B'' = B_1 \wedge B'' = B_1,$$

where $B'' \geq \bigwedge_\lambda \left(S_{f(\lambda)} \circ B_1 \right)$ and $\bigwedge_\lambda \left(S_{f(\lambda)} \circ B_1 \right) = B_1$.

Similarly for $A = A_2$.

So it holds

$$A = A_i \;\Rightarrow\; B = B_i,\; i = 1,\; 2.$$

□

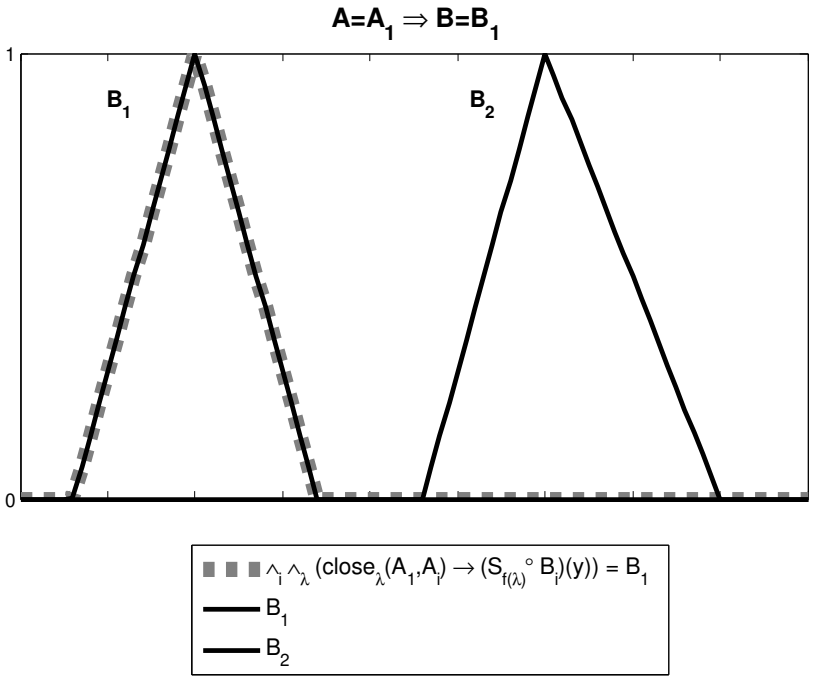

Figure 8: Interpolation condition, $A = A_1 \to B = B_1$

Acknowledgement

The paper has been partially supported by the grant "F-transform in various aspects of image processing" of the MŠMT ČR.

Conclusions

We have proposed a definition of interpolation of fuzzy data, which stems from the classical approach based on rigorous interpolation condition. We investigated another approach to fuzzy interpolation (published in [5]) with relaxed interpolation condition. We simplified and illustrated in various pictures the interpolation algorithm that is based on the proposed in [5] approach. We proved that even if the interpolation condition is relaxed the related algorithm gives an interpolating fuzzy function which fulfils the interpolation condition in the classical sense. Thus the interpolation algorithm in [5] is in the agreement with the definition of interpolation of fuzzy data which we proposed.

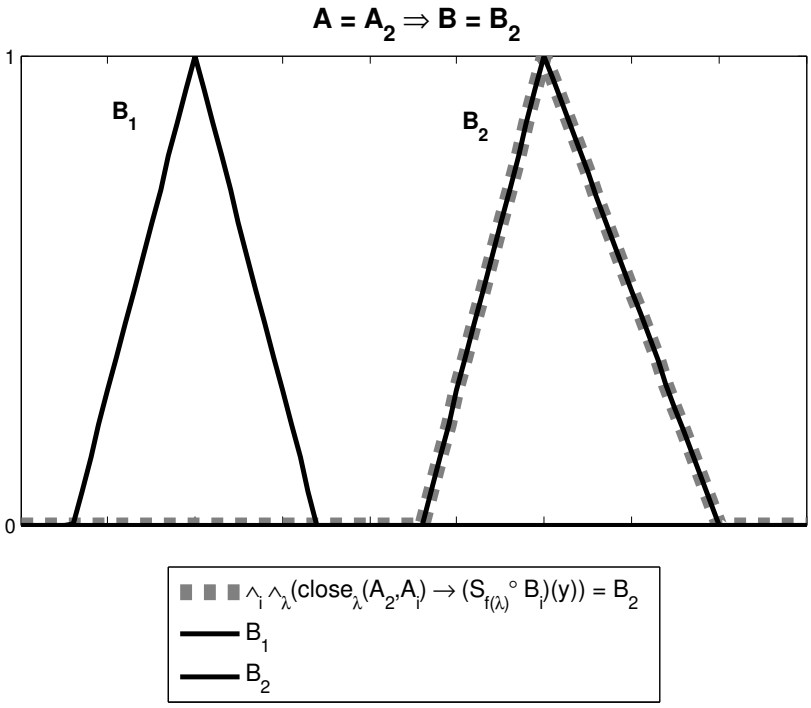

Figure 9: Interpolation condition, $A = A_2 \to B = B_2$

References

[1] P. Baranyi, T. D. Gedeon, Rule interpolation by spatial geometric representation, in: Proc. of IPMU'96 Conf., Granada, 1996.

[2] P. Baranyi, L. T. Kóczy, T. D. Gedeon, A generalized concept for fuzzy rule interpolation, IEEE Transactions on Fuzzy Systems 12 (2004) 820–837.

[3] B. Bouchon-Meunier, J. Delechamp, C. Marsala, N. Mellouli, M. Rifqi, L. Zerrouki, Analogy and fuzzy interpolation in case of sparse rules, in: Proc. of the EUROFUSE-SIC Joint Conf., 1999.

[4] B. Bouchon-Meunier, D. Dubois, C. Marsala, H. Prade, L. Ughetto, A comparative view of interpolation methods between sparse fuzzy rule, in: Proc. of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference, 2001.

[5] B. Bouchon-Meunier, F. Esteva, L. Godo, M. Rifqi, S. Sandri, A principled approach to fuzzy rule base interpolation using similarity relations, in: Proc. of the EUSFLAT-LFA Joint Conf., 2005.

[6] B. Bouchon-Meunier, C. Marsala, M. Rifqi, Interpolative reasoning based on graduality, in: Proc. of FUZZ-IEEE'2000 Int. Conf., San Antonio, 2000.

[7] M. Demirci, Fundamentals of m-vague algebra and m-vague arithmetic operations, Int. Journ. Uncertainty, Fuzziness and Knowledge-Based Systems 10 (2002) 25–75.

[8] D. Dubois, R. Martin-Clouaire, H. Prade, Practical computing in fuzzy logic, in: M. M. Gupta, T. Yamakawa (eds.), Fuzzy Computing, North-Holland, 1988.

[9] D. Dubois, H. Prade, Fuzzy Sets and Systems. Theory and Applications, Academic Press, New York, 1980.

[10] P. Hájek, Metamathematics of Fuzzy Logic, Kluwer, Dordrecht, 1998.

[11] P. Hoďáková, M. Wrublová, I. Perfilieva. Fuzzy Rule Base Interpolation and its Graphical Representation. 17th Zittau East-West Fuzzy Colloquium. Zittau/Grlitz, 2010. s. 132-137. [2010-09-15]. ISBN 978-3-9812655-4-5

[12] U. Höhle, L. N. Stout, Foundations of fuzzy sets, Fuzzy Sets and Systems 40 (1991) 257–296.

[13] S. Jenei, Interpolation and extrapolation of fuzzy quantities revisited an axiomatic approach, Soft Comput. 5 (2001) 179–193.

[14] S. Jenei, E. P. Klement, R. Konzel, Interpolation and extrapolation of fuzzy quantities revisited the multiple-dimensional case, Soft Comput. 6 (2002) 258–270.

[15] Z. C. Johanyák, S. Kovács, Fuzzy rule interpolation based on polar cuts, in: B. Reusch (ed.), Computational Intelligence, Theory and Applications, Springer, 2006.

[16] F. Klawonn, Fuzzy points, fuzzy relations and fuzzy functions, in: V. Novák, I. Perfilieva (eds.), Discovering the World with Fuzzy Logic, Springer, Berlin, 2000, pp. 431–453.

[17] L. T. Kóczy, K. Hirota, Approximate reasoning by linear rule interpolation and general approximation, Int. J. Approximate Reasoning 9 (1993) 197–225.

[18] L. T. Kóczy, K. Hirota, Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases, Inform. Sci. 71 (1993) 169–201.

[19] V. Novák, I. Perfilieva, J. Močkoř, Mathematical Principles of Fuzzy Logic, 1999, Boston, Kluwer, 320.

[20] I. Perfilieva, Fuzzy function as an approximate solution to a system of fuzzy relation equations, Fuzzy Sets and Systems 147 (2004) 363–383.

[21] I. Perfilieva, D. Dubois, H. Prade, L. Godo, F. Esteva, P. Hokov, Interpolation of fuzzy data: analytical approach and overview, Fuzzy Sets and Systems (2011) to appear.

[22] M. Takacs, The g-calculus based compositional rule of inference, Journ. of Advanced Computational Intelligence and Intelligent Informatics, 10 (2006) 534-541.

[23] L. Ughetto, D. Dubois, H. Prade, Fuzzy interpolation by convex completion of sparse rule bases, in: Proc. of FUZZ-IEEE'2000 Int. Conf., San Antonio, 2000.

Fractal Analysis of Forward Exchange Rates[1]

Mária Bohdalová, Michal Greguš

Faculty of Management Bratislava, Comenius University
Odbojárov 10, P.O.BOX 95, 820 05 Bratislava, Slovakia
maria.bohdalova@fm.uniba.sk, michal.gregus@fm.uniba.sk

Abstract: In this paper we work with nonparametric methods in modeling and analyzing the financial times series. We use the concept of fractal dimension for measuring the complexity of time series of observed financial data. The aim of this paper is to distinguish between the randomness and determinism of the financial information. We will compare the fractal analysis of the selected forward exchange rates. Fractal analysis has been introduced into financial time series by Mandelbrot and Peters. Due to the financial crisis this theory has gained new momentum. Fractal analysis indicates that conventional econometric methods are inadequate for analyzing financial time series. Adequate analysis of the financial time series allows us to predict precisely the future values and risks connected with portfolios that are influenced. We test for fractional dynamic behavior in a 1-month forward exchange rate USD into GBP and Gold Price against USD.

Keywords: fractal analysis; estimation dimension; long memory; financial time series

1 Introduction

The purpose of this paper is to show a potential presence of stochastic long memory in economic and financial time series. The long term memory property describes the high-order correlation structure of a series. The long memory existence in financial time series may be caused by investors' reactions to market information. Some investors react to information as it is received, while some investors wait for confirmation of the new information and they do not react until a trend is clearly established. Classical capital market theory assumes that the markets follow a random walk, and this means that the current prices reflect all available information and future price changes can be determined only by new information. With all prior information already reflected in prices. This means each day's price movement is unrelated to the previous day's activity. It is assumed that all investors immediately react to new information, so that the future

[1] The preliminary version of this contribution was presented at FSTA 2010 in Liptovský Ján

is unrelated to the past or the present. Thus, all investors react to new information with equal probability. This assumption has been necessary for the application of the Central Limit Theorem to capital market analysis. But investors really do not make their decisions in this manner. Although the reaction of the investors is random, they may prefer some information and therefore the probability of their decisions is not identical; their decision is biased in some direction. Therefore the market may follow a biased random walk.

Biased random walks were studied by Hurst in 1940s. Hurst was a hydrologist studying the discharge rates of the River Nile at Aswan. He detected long memory behavior on the River Nile data. Hurst attributed this result to be a consequence of the flow rate having serial correlation. The *Hurst parameter H* displays long range dependence. Hurst [5] motivated Mandelbrot and his co–workers (Mandelbrot and van Ness [9]) to introduce fractional Gaussian noise to model long memory phenomena.

Long memory or long term dependence is observed in contemporary financial time series. There exist a number of studies that have investigated the issue of persistence in financial asset returns. Using the Hurst rescaled-range (R/S) method, Greene and Fielitz (1977) reported long memory in daily stock returns series. This result has been overturned by Lo (1991) via the development and implementation of the more appropriate modified R/S method. The absence of long memory in stock returns is also reported by Aydogan and Booth (1988), Cheung, Lai, and Lai (1993), Cheung and Lai (1995), Crato (1994), and Barkoulas and Baum (1996). Booth, Kaen, and Koveos (1982) and Cheung (1993) report long-memory evidence in spot exchange rates. Helms, Kaen, and Rosenman (1984), Cheung and Lai (1993), Fang, Lai, and Lai (1994), and Barkoulas, Labys, and Onochie (1997) report that stochastic long memory may be a feature of some spot and futures foreign currency rates and commodity prices.

The presence of fractal structure in asset returns raises a number of theoretical and empirical issues. First, as long memory represents a special form of nonlinear dynamics, it calls to question linear modeling and invites the development of nonlinear pricing models at the theoretical level to account for long memory behavior.

The rest of this paper is organized as follows. Section two introduces the stochastic processes and self–similar stochastic processes. Section three briefly describes fractional Brownian motion. Section four describes fractal dimension and fuzzy sets. Data and empirical estimates are discussed in section five. The paper ends with a summary of our results.

2 Stochastic Processes and Self-Similar Stochastic Processes

Given an observed time series, a question which is of interest is whether the data were generated by a dynamical system of finite dimension or whether the system is stochastic. In many observed time series it is not clear what the fundamental underlying process is that drives the system. However, in real processes we observe certain aspects that are the evidence of an underlying, more complex process. We can observe processes that display power-law scaling and long range dependence. The main problem is, if a given time series is related to an underlying more substantial process, whether it is possible to determine whether the underlying process is driven by a deterministic set of equations or a stochastic system, or whether the process is self-similar. This opens the question of what is the difference between a deterministic and a stochastic process and whether is it possible to make this distinction based on empirical observations.

Cutler [2] defined what is meant by stochastic and deterministic time series. Lamperti [6] introduced the definition of a self-similar stochastic process. Mandelbrot [8] introduced Fractional Brownian Motion. We give a brief summary in this section.

Definition 1

The time series $\{X(t_n)\colon n=1,2,\ldots\}$ is said to be *strictly stationary* if for any finite collection $t_1, t_2, \ldots, t_n$ and for all τ,

$$Pr\{X(t_1)<x_1, \ldots, X(t_n)<x_n\} = Pr\{X(t_1+\tau)<x_1, \ldots, X(t_n+\tau)<x_n\}.$$

Definition 2

A mapping $g\colon \chi \to \psi$, between the metric spaces χ and ψ with metrics ρ_1 and ρ_2 respectively, is said to satisfy a *Lipschitz condition* if, for all $x_1, x_2 \in \chi$,

$$\rho_2(g(x_1),g(x_2)) \le k\rho_1(x_1, x_2),$$

where k is a constant. If in addition, g is one to one and g^{-1} also satisfies a Lipschitz condition on its domain, then g is bi-Lipschitz.

Definition 3

Let $\{X(t_n)\colon n=1,2,\ldots\}$ be a strictly stationary time series with values in ψ. The *predictive dimension,* denoted by ζ, is defined as the smallest $n \ge 1$ such that there exists a mapping $\Psi\colon \psi^n \to \psi$ such that

$$X(t_n) = \Psi[X(t_1),\ldots, X(t_{n-1})],$$

with probability 1. If no function Ψ exists for all $n \ge 1$, then $\zeta = \infty$.

Cutler formulated a theorem that says that a strictly stationary process with known predictor function Ψ and finite predictive dimension ζ can be predicted as a

function of the previous ζ observations. Subsequently this is used for defining a stochastic and deterministic time series.

Theorem

Let $\{X(t_n)$: n=1,2,…$\}$ be a strictly stationary time series with finite predictive dimension ζ and predictor function Ψ. Then, for all integers $m \geq 0$,

$$Y(t_{m+1+\zeta})=\Psi(Y(t_{m+1}), Y(t_{m+2}), \ldots, Y(t_{m+\zeta})),$$

with probability 1.

Definition 4

A strictly stationary time series $\{X(t_n)$: n=1,2,…$\}$ is said to be *deterministic* if $\zeta<\infty$ and *stochastic* if $\zeta=\infty$, where ζ is the predictive dimension.

The following discussion will be concerned with Lamperti's [6] idea of scaling in a process $X(t)$. Firstly we introduce the notion of *equality of finite dimensional distributions.*

Definition 5

Let $X_1(t)$ and $X_2(t)$ be two stochastic processes. We will say that these processes have the same finite dimensional distributions if, for any $n \geq 1$ and $t_1, t_2, \ldots, t_n$

$$(X_1(t_1), X_1(t_1), \ldots, X_1(t_n)) \overset{d}{=} (X_2(t_1), X_2(t_1), \ldots, X_2(t_n)) \text{ or } (X_1(t)) \overset{d}{=} (X_2(t)),$$

where $\overset{d}{=}$ denotes equality of probability distributions [4].

Definition 6

d–dimensional process $X(t)$ is a *semi–stable process,* if it obeys a simple continuity condition and, for $s>0$, the relationship

$$\{X(st)\} \overset{d}{=} \{b(s)\,(X(t)+c(s))\}$$

holds, where $b(s)$ is a positive function and $c(s) \in R^d$.

Lamperti [6] showed that if $X(t)$ is a proper semi-stable process and $X(0)=0$, then $c(s)=0$ and $b(s)=s^H$ where H is a positive constant. That is,

$$\{X(st)\} \overset{d}{=} \{s^H X(t)\}. \tag{1}$$

Definition 7

The increments of a random function $\{X(t)$: $-\infty<t<\infty\}$ are said to be self–similar with parameter H if for any $s>0$ and any τ

$$\{X(\,st+\tau\,) - X(\,\tau\,)\} \overset{d}{=} \{\,s^H\,(X(\,t+\tau\,) - X(\,\tau\,))\}.$$

If the increments of $X(t)$ are self–similar and $X(0)=0$, then $X(t)$ is also self–similar (see equation (1)). If $X(t)$ has self–similar and stationary increments and is mean square continuous, then it can be shown that $0 \leq H < 1$.

The covariance structure is derived from scaling law as follows [4]:

Let $X(t)$ be process with stationary self–similar increments. Then the *covariance function* is

$$E[(X(t+\tau+1) - X(t+\tau))(X(t+1) - X(t))] = \frac{1}{2}\sigma_H^2 \left\{ |\tau+1|^{2H} + |\tau-1|^{2H} - 2|\tau|^{2H} \right\},$$

where $\sigma_H^2 = E[(X(t+1) - X(t))^2]$ for all t.

The process $X(t)$ is said to be *isotropic* if

$$\{X(t) - X(s)\} \overset{d}{=} \{X(|t-s|)\}. \tag{2}$$

3 Fractional Brownian Motion

A *Gaussian process* is uniquely determined by its auto covariance function. *Fractional Brownian Motion* is a unique Gaussian self-similar process that we will denote as $B_H(t)$ [4]. The increments of fractional Brownian motion are referred to as *fractional Gaussian noise*. If $B_H(0)=0$, then the process $B_H(t)$ is isotropic (see equation (2)).

When $H=0.50$, $B_H(t)$ is simply *Brownian motion*. The system is independently distributed. When H differed from 0.50, the observations are not independent. Each observation carried a "memory" of all the events that preceded it. What happens today influences the future. Where we are now is a result of where we have been in the past. Time is important. The impact of the present on the future can be expressed as a correlation:

$$C = 2^{(2H-1)} - 1, \tag{3}$$

where C is the correlation measure and H is the Hurst exponent. The time series is random, and events are random and uncorrelated. The present does not influence the future. Its probability density function can be a normal curve, but it does not have to be.

When $H>0.50$ the autocorrelations are positive and have a power-law decay, hence long range dependence. If $0.50 \leq H < 1.00$, the time series have a persistent or trend–reinforcing character. If the time series was up (down) in the last period, then the chances are that it will continue to be positive (negative) in the next period. Trend is apparent. The strength of the trend-reinforcing behavior, or

persistence, increases as H approaches 1.0. The strength of the bias depends on how far H is above 0.50. The closer H is to 0.5, the noisier it will be, and the less defined its trends will be. Persistent series are called by Mandelbrot as fractional Brownian motion, or biased random walk.

When $H<0.50$ the correlation are negative and have a rapid decay. For $0 \leq H<0.50$ the time series is antipersistent, or ergodic. If the time series was up in the previous period, it is more likely to be down in the next period. Conversely, if it was down before, it is more likely to be up in the next period. The strength of this antipersistent behavior depends on how close H is to zero. The closer it is to zero, the closer C in equation (3) moves toward –0.50, or negative correlation. This time series is more volatile than a random series.

The Hurst coefficient H characterizes long-memory dependence. Self-similarity of the time series is characterized by fractal dimension. Fractal dimension expresses the regularity of series and states how similarity scales up when such a time series is observed over a longer time interval. The self-similarity could be also regarded as a measure of geometrical complexity of an object under discussion.

In principle, fractal dimension and Hurst coefficient are independent of each other: fractal dimension is a local property, and long-memory dependence is a global characteristic. Nevertheless, the two notions are closely linked. For self-affine processes, the local properties are reflected in the global ones, resulting in the relationship $D+H=2$ between fractal dimension, D, and Hurst coefficient, H.

The determination of the fractal dimension is inherently associated with set-based constructs. The generic box dimension [8] measures in which way the number of occupied boxes (those including the elements of the time series) increases when the size of the box decreases. The other common techniques of fractal determination uses a so-called correlation dimension in which a count of elements concerns a family of spheres constructed around each data point. What is common to the existing techniques (in spite of evident technical differences) is that all of them exploit sets regarded as information granules that allow us to see only a certain part of the phenomenon. The changes in the size of the information granules imply how large a part we are taking into consideration. Information granulation is an example of abstraction. There are numerous facets of the granular information processing, and there are a variety of formal frameworks in which such information granulation takes place. These include, for instance, set theory, fuzzy sets, random sets, rough sets and many others [7].

4 Fractal Dimension and Fuzzy Sets

In this section, we review the main constructs of fractal dimension and then proceed with their generalization in terms of information granules expressed in the language of fuzzy sets [7].

Consider a time series $\{X(t_n): n=1,2,\ldots, N\}$, where $X(t_n) \in R$ and t_n denote discrete time moments in which the values of this time series are recorded. In this sense, we are provided with a collection of two-dimensional elements $S=\{ X(t_n): n=1,2,\ldots, N\}$. The structural complexity of S is measured by a fractal dimension $\hat{D}$ defined by the following limit

$$\hat{D} = \lim_{\varepsilon \to 0} \frac{\log(N(\varepsilon))}{\log(\varepsilon)}, \tag{4}$$

where $N(\varepsilon)$ is a number of boxes of size ε used to cover the object (here the given time series). In essence, the above relationship relates to the power law stating that $N = \varepsilon^{\hat{D}}$. In practice, the fractal dimension has to be estimated with the use of some experimental data. A collection of ''c'' experiments concerns a determination of the number of boxes $N(\varepsilon)$ for a given value of the size of the box. Then experimental pairs $(\varepsilon_j, N(\varepsilon_j))$, $j= 1, 2, \ldots; c$ are used to determine parameters of the linear model. It can be easily shown (from (3)) that in a double logarithmic model of the form

$$\log N(\varepsilon) = D \log(\varepsilon) + C \tag{5}$$

the fractal dimension D appears as a slope of the computed regression line. The regression model itself is constructed through a minimization of the well known performance index Q treated as a sum of squared errors

$$Q = \sum_{k=1}^{c} (\log N(\varepsilon_k) - D \log(\varepsilon_k) - c)^2 . \tag{6}$$

The most intuitive approach to the determination of the fractal dimension uses the box method [8]. Another method uses a sphere of radius ε. Total number of points covered by the spheres $N(\varepsilon)$ is equal to

$$N(\varepsilon) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \Omega_{ij}(\varepsilon) \tag{7}$$

where $\Omega_{ij}(\varepsilon)$ is a sphere defined as follows

$$\Omega_{ij}(\varepsilon) = \begin{cases} 1 \text{ if } \sqrt{(t_i - t_j)^2 + (x_i - x_j)^2} \leq \varepsilon \\ 0 \text{ otherwise} \end{cases} . \tag{8}$$

Pedrycz and Bargiela [7] used fuzzy set $A_{ij}(\varepsilon)$ to compute of the fractal dimension:

$$N(\varepsilon) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \Omega_{ij}(\varepsilon) A_{ij}(\varepsilon) \text{, where} \tag{9}$$

$$A_{ij}(\varepsilon) = \exp\left(\frac{-(x_i - x_j)^2}{\varepsilon^2} \right). \tag{10}$$

For the time series we can use:

$$\Omega_i(\varepsilon) = \sum_{i=1}^{\varepsilon} (x_i - M_\varepsilon) \tag{11}$$

where $\Omega_i(\varepsilon)$ is cumulative deviation over ε period, M_ε is an average x_i over period of length of ε.

$$N(\varepsilon) = \max(\Omega_i(\varepsilon)) - \min(\Omega_i(\varepsilon)). \tag{12}$$

This approach is known as *R/S* analysis ([7, 10, 11]), where usually $N(\varepsilon)$ is denoted as R/S and ε as *n*.

The detailed computations of the fractal dimension are described, for example, in [7, 10] and they are realized on the basis of the regression model (5).

Mandelbrot used *R/S* analysis which was developed by Hurst [5]. Mandelbrot, Taqqu and Wallis demonstrated the superiority of *R/S* analysis over more conventional methods of determining long-range dependence, such as analyzing autocorrelations, variance ratios and spectral decompositions, in their several papers. In this paper our analysis will be based on the study described in Peters [10] or Robinson [11]. In this paper we compute Hurst coeficient *H* and his expected value *E*(*H*) using modified R/S analysis[2] and we will verify null hypothezis: *The time series is random walk.*

To verify this hypothesis, we calculate expected value of the adjusted range[3] $E(R/S_n)$ and its variance[4] $Var(E(R/S_n))$.

[2] The *R/S* statistics is modified so that its statistical behavior is invariant over a general class of short memory processes, but deviates for long-memory processes. ([11], p. 91)

[3] This formula was derrived by Anis and Lloyd ([10], p. 71)

[4] Variance was calculated by Feller ([10], p. 66)

$$E(R/S_n) = \frac{n-0.5}{n} \cdot \left(n \cdot \frac{\pi}{2} \right)^{-0.5} \sum_{r=1}^{n-1} \sqrt{\frac{(n-r)}{r}} \tag{13}$$

$$Var(E(R/S_n)) = \left(\frac{\pi^2}{6} - \frac{\pi}{2} \right) \cdot n \,. \tag{14}$$

Using the results of equation (13) we can generate the expected values of the Hurst exponent. The expected Hurst exponent will vary depending on the values of n we use to run the regression. Any range will be appropriate as long as the system under study and the $E(R/S_n)$ series cover to the same values of n. For financial purpose, we will begin with n=10. The final value of n will depend on the system under study.

R/S values are random variables, normally distributed and therefore we would expect that the values of H would also be normally distributed ([10], p. 72):

$$Var(H_n) = \frac{1}{T}, \tag{15}$$

where T is total number of observations in the sample. Note that the $Var(H_n)$ does not depend on n or H, but it depends on the total sample size T. Now t-statistics will be used to verify the significance of the null hypothesis.

If Hurst exponent H is approximately equal to its expected value $E(H)$, it means that the time series is independent and random during the analysed period (the Hurst exponent is insignificant). If the Hurst exponent H is greater (smaller) than its expected value $E(H)$, the time series is persistent (antipersistent) (the Hurst exponent is significant). If the series exhibits a persistent character, then the time series has long memory and the ratios R/S_n will be increasing. If the ratios R/S_n will be decreasing the time series will be antipersistent. The "breaks" may signalize a periodic or nonperiodic component in the time series with some finite frequency. We calculated the V–statistics to estimate precisely where this break occurs [10]:

$$V_n = (R/S)_n \big/ \sqrt{n} \tag{16}$$

5 Data and Empirical Results

The data set consists of daily forward 1-month exchange rate USD into GBP and Gold Price against USD from 02/01/1979 to 04/11/2010 for a total 8050 daily observations. These were obtained from Bank of England[5].

[5] http://www.bankofengland.co.uk

We begin by applying *R*/*S* analysis to the 1-month forward exchange rates USD into GBP. During the period 02/01/1979–04/11/2010, the Hurst coefficient *H* of the 1-month forward exchange rate USD into GBP is equal to 0.5702. The expected Hurst exponent is equal to $E(H)$=0.5407. The variance of $E(H)$ is $1/T$=1/8050, for Gaussian random variables. The standard deviation of $E(H)$ is 0.0111. The Hurst exponent for the daily 1-month forward exchange rates USD into GBP is 2.6513 standard deviations away from its expected value. This is highly significant result at the 95% level. The time series has persistent character. Also plotted is $E(R/S_n)$ (dashed line) as a comparison against the null hypothesis that the system is an independent process (Figure 1). There is clearly a systematic deviation from the expected values. However, breaks in R/S graph (see Figure 1) appear. To estimate precisely where this break occurs, we calculated *V*–statistics (Figure 1). V-statistics clearly stops groving at n=50, n=322, n=575 or n=805 observations. These "breaks" may be signal of a periodic or nonperiodic component in the time series. We will run regression to estimate the Hurst exponent for R/S_n values in the next subperiods: n<50, 50≤ n ≤4025, 10<n<322, 322≤ n ≤4025, 10<n<575, 575≤ n ≤4025, 10<n<805 and 805≤ n ≤4025. Table 1 and Table 2 show the regression results. During periods for n<50, 10<n<322 and 10<n<575 the time series has random character. The Hurst exponents are insignificant. During periods for 50≤ n ≤4025, 322≤ n ≤4025 and 575≤ n ≤4025 the time series has persistent character. The Hurst exponent is significant. It means that ancient history had random character and recent history has a long memory effect. During periods for 50≤ n ≤805 the time series has a persistent character, but during period 805≤ n ≤4025 the time series has an antipersistent character and the Hurst exponent is significant. We have found that 1-month forward exchange rate USD into GBP has 4 nonperiodic cycles. The longest is a 805-day cycle, or about 3 years. The shortest is a 50-day cycle, or about 10 weeks.

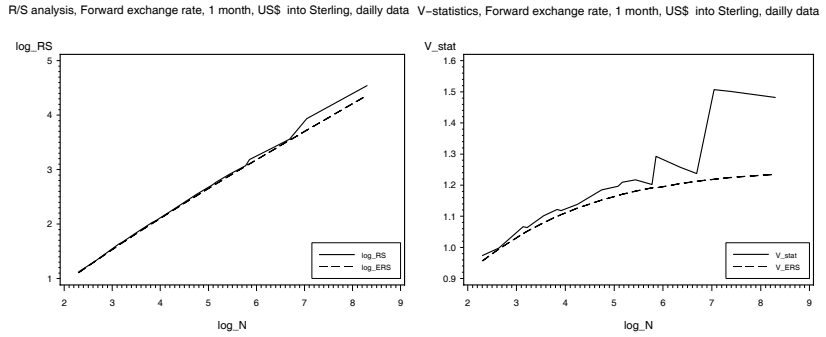

Figure 1
R/S analysis and V statistics of the daily log return of USD into GBP, (1979-2010)
H=0.5702, $E(H)$= 0.5407

Table 1
Regression results, 1-month forward FX rate USD vs GBP, estimation of the Hurst exponent, (1979-2010, daily data)

1-month forward	R/S	E(R/S)	R/S	E(R/S)	R/S	E(R/S)	R/S	E(R/S)
FX rate GBPvsUSD	10<*n*<50		50≤ *n* ≤4025		10<*n*<322		322≤ *n*≤4025	
Intercept	-0.236	-0.241	-0.200	0.041	-0.149	-0.152	-0.247	0.093
Hurst exponent	0.592	0.589	0.574	0.522	0.564	0.561	0.581	0.515
Standard Error	0.009	0.006	0.048	0.008	0.019	0.016	0.072	0.003
R squared	0.999	1.000	0.995	1.000	0.999	0.999	0.984	0.999
Number of obs.	7		12		13		6	
Significance	0.260		4.701		0.323		5.993	

Table 2
Regression results, 1-month forward FX rate USD vs GBP, estimation of the Hurst exponent, (1979-2010, daily data)

1-month forward	R/S	E(R/S)	R/S	E(R/S)	R/S	E(R/S)	R/S	E(R/S)
FX rate GBP vs USD	10<*n*<575		575≤ *n*≤4025		10<*n*<805		805≤ *n*≤4025	
Intercept	-0.146	-0.131	-0.235	0.118	0.558	-0.119	0.508	0.126
Hurst exponent	0.563	0.554	0.580	0.511	-0.126	0.551	0.486	0.510
Standard Error	0.021	0.019	0.095	0.001	0.025	0.020	0.001	0.001
R squared	0.999	0.999	0.964	1.000	1.000	0.999	1.000	1.000
Number of obs.	15		4		16		3	
Significance			6.191		-60.715		-2.153	

R/*S* analysis of Gold Price against USD from 02/01/1979 to 04/11/2010 exhibits random behavior. The Hurst coefficient *H* is equal to 0.547, $E(H) = 0.540$ and it is insignificant (see Figure 2). Table 3 summarizes the regression results. However we found 2 breaks on *R*/*S* plot (respectively in *V*-statistics plot, see Figure 2) for n=161 and n=322. During periods $10<n<161$ and $10\leq n\leq 322$ the time series has random character, but during periods $161<n<4025$ and $322\leq n \leq 4025$ the time series has persistent character. The presence of the persistent value of *H* confirms that Gold prices against USD have fractal structure in recent history. We found one periodic cycle with length 161 (or approximately 32 weeks).

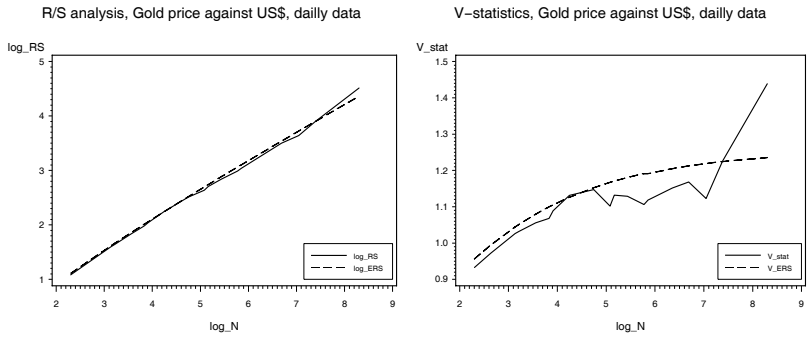

Figure 2
R/S analysis and V statistics of the daily log return of Gold prices against USD, (1979-2010)
H=0.547, $E(H)$ = 0.540

Table 3
Regression results, Gold vs GBP, estimation of the Hurst exponent, (1979-2010, daily data)

Gold versus GBP	R/S	E(R/S)	R/S	E(R/S)	R/S	E(R/S)	R/S	E(R/S)
	10<n<161		161≤ n ≤4025		10<n<322		322≤ n≤4025	
Intercept	-0.2523	-0.2040	0.2446	0.0672	-0.1633	-0.1644	-0.4465	0.0894
Hurst exponent	0.5854	0.5763	0.5631	0.5181	0.5582	0.5639	0.5910	0.5150
Standard Error	0.0113	0.0109	0.0482	0.0041	0.0268	0.0146	0.0469	0.0027
R squared	0.9994	0.9994	0.9941	0.9999	0.9979	0.9934	0.9934	0.9999
Number of observation	9		10		12		7	
Significance	0.8165		4.0375		-0.5114		6.8189	

Conclusion

In this paper, we propose a fractal analysis of the selected financial time series. In both causes, we found fractal structure. Nonperiodic cycles for forward exchange rate affirm evidence that the currency markets may be nonlinear systems. Currency markets are characterized by abrupt changes traceable to central bank intervention attempts by governments to control the value of each respective currency.

Periodic cycle in the time series Gold prices against USD may be related to the economic cycle. The cycle length measures how long it takes for a single period's influence to reduce to immeasurable amounts. In statistical terms, it is the decorrelation time of the series. In terms of nonlinear dynamics, memory effect is lost when this time expires.

Information obtained by fractal analysis can be used as the basis for momentum analysis and other forms of technical analysis. The second use is in choosing periods for model development, particularly for back testing.

Acknowledgement

The research summarized in this paper was supported by VEGA grant agency, grant number 1/4024/07 and VEGA grant agency, grant number 1/0373/08

References

[1] Cutler, C. D. *A Theory of Correlation Dimension for Stationary Time Series*. Phil. Trans. Royal Soc. London A 348 (1688). 1994; 343-355

[2] Cutler, C. D. *A General Approach to Predictive and Fractal Scaling Dimensions in Discrete-Index Time Series*. Fields Instit. Comm. 11. 1997; 29-48

[3] Davies, R. B., Harte, D. S. *Test for Hurst effect*. Biometrika 74 (1). 1987; 95-101

[4] Harte, D. *Multifractals. Theory and Applications*. Chapman&Hall/CRC. 2001; p. 248

[5] Hurst, H. E. *Long Term Storage Capacity of Reservoirs*. Trans. Amer. Soc. of Civil Engineers 116. 1951; 770-808

[6] Lamperti, J. *Semi-Stable Stochastic Processes*. Trans. Amer. Math. Soc. 104. 1962; 62-78

[7] Pedrycz, W., Bargiela, A.: *Fuzzy Fractal Dimensions and Fuzzy Modeling*. Information Sciences 153, 2003, pp. 199-216

[8] Mandelbrot, B. (Ed.), *The Fractal Geometry of Nature*, Freeman, New York, 1983, pp. 14-57

[9] Mandelbrot, B. and van Ness, J. W., *Fractional Brownian Motions, Fractional Noises and Applications.* SIAM Review 10, 1968

[10] Peters, E. E. *Chaos and Order in the Capital Markets*. New York: John Wiley & Sons, Inc.. 1996; p. 274

[11] Robinson, P. M. *Time Series with Long Memory*. Advanced Texts in Econometrics. New York: Oxford University Press. 2003; p. 375

[12] Rosenblatt, M. *Independence and Dependence*. In: Proceedings of the 4th Berkley Symposium on Mathematical Statistics and Probability. Volume II Probability theory, (Edited by J. Neyman). University of California Press, Berkley.1961; 431-443

[13] Zadeh, L. A., Kacprzyk, J. (Eds.). *Computing with Words in Information/Intelligent Systems*, Vol. I and II, Physical Verlag, Heidelberg, 1999

An Integrated Intelligent Neuro-Fuzzy Algorithm for Long-Term Electricity Consumption: Cases of Selected EU Countries

Ali Azadeh[1], Morteza Saberi[2], Vahid Nadimi[2], Mino Iman[3], Alireza Behrooznia

[1] Department of Industrial Engineering, Center of Excellence for Intelligent-based Experimental Mechanics, College of Engineering, University of Tehran, Iran

[2] Member of Young Research Club of Azad University, Tafresh Branch

[3] Islamic Azad University, Tafresh Branch, Iran

E-mail: aazadeh@ut.ac.ir, mortezza_saber@yahoo.com, vahidndimi60@yahoo.com, alirezabehroznia@gmail.com

Abstract: This paper presents an adaptive-network-based fuzzy inference system (ANFIS) for long-term natural Electric consumption prediction. Six models are proposed to forecast annual Electric demand. 104 ANFIS have been constructed and tested in order to find the best ANFIS for Electric consumption. Two parameters have been considered in the construction and examination of plausible ANFIS models. The type of membership function and the number of linguistic variables are two mentioned parameters. Six different membership functions are considered in building ANFIS, as follows: the built-in membership function composed of the difference between two sigmoidal membership functions (dsig), the Gaussian combination membership function (gauss2), the Gaussian curve built-in membership function (gauss), the generalized bell-shaped built-in membership function (gbell), the Π-shaped built-in membership function (pi), psig. Also, a number for linguistic variables has been considered between 2 and 20. The proposed models consist of input variables such as: Gross Domestic Product (GDP) and Population (POP). Six distinct models based on different inputs are defined. All of the trained ANFIS are then compared with respect to the mean absolute percentage error (MAPE). To meet the best performance of the intelligent based approaches, data are pre-processed (scaled) and finally our outputs are post-processed (returned to its original scale). The ANFIS model is capable of dealing with both complexity and uncertainty in the data set. To show the applicability and superiority of the ANFIS, the actual Electric consumption in industrialized nations including the Netherlands, Luxembourg, Ireland, and Italy from 1980 to 2007 are considered. With the aid of an autoregressive model, the GDP and population by 2015 is projected and then with yield value and best ANFIS model, Electric consumption by 2015 is predicted.

Keywords: Natural Electric Demand; Long-Term prediction; Adaptive Network-based Fuzzy Inference System (ANFIS)

1 Introduction

Energy is a vital input for the social and economic development of any nation. The growth in energy consumption is intrinsically linked to the growth in the economy. Various models have been applied to describe and forecast the evolution of energy demand. Chavez et al. (1999) used Box–Jenkins time-series analyses (ARIMA) models to formulate the forecasting model for the prediction of energy production and consumption in Asturias, Northern Spain. The trend in current and near future energy consumption from a statistical perspective by considering two factors, namely, increasing population and economic development, has been discussed by Kadoshin et al. [19].

Khotanzad et al. (2000) focused on the combination of artificial neural-network (ANN) forecasters with application to the prediction of daily natural gas consumption needed by gas utilities [20]. Saab et al. (2001) investigated different univariate-modeling methodologies for the forecasting of the monthly electric energy consumption in Lebanon [25]. Three univariate models were used, namely, autoregressive, autoregressive integrated moving average (ARIMA) and a novel configuration combining an AR(1) with a high pass filter.

An oil and gas supply model (OGSM) waa solved and the projections of oil and natural gas supply and demand to the year 2020 for Canada were presented in (Jai Persaud and Uma Kumar, 2001). Chow (2001) has discussed the sectoral energy consumption in Hong Kong for the period 1984-97 with special emphasis on the household sector [9]. Ediger and Tatlldil (2002) used a semi-statistical technique to formulate the forecasting model to predict the primary energy demand in Turkey and analysis of cyclic patterns. Reddy and Balachandra (2003) looked at various factors that influence the energy demand in India and developed the energy and environmental outlook for the year 2010. This was done by developing an integrated mathematical model incorporating various factors such as GDP and population growth.

The model presented in Siemek et al. (2003) estimates natural-gas demand, based on the average trend of economic development; the model considered the natural production/demand maxima of energy carriers. The prognosis was loaded with an error resulting from the use of average data related to yearly increases of the national gross product [27]. Gorucu et al. (2004) trained the ANNs to decide the optimum parameters to be used in forecasting gas consumption for short-term applications [13]. Gorucu and Gumrah (2004) forecasted short term gas consumption by multivariable regression analysis for the capital city of Ankara, Turkey [14].

Gutierrez et al. (2005) examined the application of a Gompertz-type innovation diffusion process for stochastic modeling and capturing the growth process of natural-gas consumption in Spain [15]. Sanchez-Ubeda and Berzosa (2007) Forecasted industrial end-use natural gas consumption in a medium-term horizon

(1-3 years) with a very high resolution (days) based on a decomposition approach [26]. The forecast was obtained by the combination of three different components: one that captures the trend of the time series, a seasonal component and a transitory component. Parikh et al. (2007) estimated demand projections of petroleum products and natural gas in India [22]. They considered GDP and population as inputs of their NG estimation model.

Azadeh et al. forecasted electrical consumption by the integration of a neural network, a time series and ANOVA [1]. They found that ANN had better estimated values for total electricity consumption. Azadeh et al. developed an integrated artificial neural network and genetic algorithm framework to predict electrical energy consumption [2]. Azadeh and Tarverdian proposed an integrated approach based on genetic algorithm, computer simulation and design of experiments for forecasting electrical energy consumption [3]. Azadeh et al. presented an integrated fuzzy system, data mining and a time series framework to estimate and predict electricity demand for seasonal and monthly changes in electricity consumption in developing countries such as China and Iran [4]. Azadeh et al. (2008b) employed an artificial neural network (ANN) approach for annual electricity consumption in high energy consumption industrial sectors [5]. Azadeh et al. developed an integrated algorithm for forecasting monthly electrical energy consumption based on an artificial neural network (ANN), computer simulation and a design of experiments using stochastic procedures [6]. Azadeh et al. (2009) proposed a new hybrid ANFIS computer simulation for improvement of electricity consumption estimation [9]. Yoo et al. estimated households' demand function for natural gas by applying a sample selection model using data from a survey of households in Seoul [30].

In this study we estimate long-term Electric demand by employing adaptive neuro fuzzy inference systems. We present six models to forecast yearly NG demand in Iran. The ANFIS is capable of dealing with uncertainty and complexity in the given data set and thus provides a better solution and estimation regarding this valuable commodity.

2 Methodology

Because of the changeable nature of electricity demand, the use of conventional methods may not give us accurate results. Thus, we employ adaptive network based fuzzy inference systems (ANFIS) to alleviate this problem. The main structure of the intelligent approach is explained in the following.

The algorithm has the following basic steps:

Step 1: Determine the inputs of the model. Considering previous studies, the most important variables that have considerable impact on gas consumption are considered.

Step 2: Collect a data set in all available previous periods for each of the input variables and output variable. In addition, all of inputs and output data are scaled and normalized using a normalization method.

Step 3: Divide the data into two sets, one for estimating the models, called the train data set, and the other one for evaluating the validity of the estimated model, called the test data set. Usually the train data set contains 70% to 90% of all data and the remaining data are used for the test data set [15].

Step 4: This step is concerned with running and estimating all of the plausible ANFIS models regarding two main parameters. The type of membership the function and number of linguistic variables are two mentioned parameters.

Step 5: The models' prediction capability is evaluated in this step through MAPE. As input data used for the model estimation have different scales, the MAPE method is the preferred method to estimate relative errors. Considering the value of MAPE, the best ANFIS is selected in each case in order to predict gas consumption.

Step 6: The value of the input variables in the coming periods are predicted by using the Auto regressive model.

Step 7: The yield values are fed to the selected ANFIS that is determined in step 5. Finally, the value of gas consumption (scaled) in each coming year is calculated by selected ANFIS and the values of the input variables that are fed to ANFIS. Finally, with post processing, the gas consumption predicted value for each year is calculated in its original scale and is suitable for analysis and usage in decision-making.

The significance of the proposed ANFIS for long term NG demand prediction is five fold. First, it uses pre-processing and post-processing approaches to eliminate possible noise. Second, it identifies the best ANFIS model based on minimum MAPE. Third, it considers standard input variables of long term Electric demand estimation. Fourth, it provides a more accurate solution than previous approaches (such as conventional regression) because it uses ANFIS, which uses adaptive neural modeling and fuzzy logic. This efficiently handles uncertainty, noise, and non-linearity in the given data set and provides the optimum solution. Fifth, it is applicable for gas consumption prediction in coming years. Figure 1 depicts the steps of the proposed ANFIS approach.

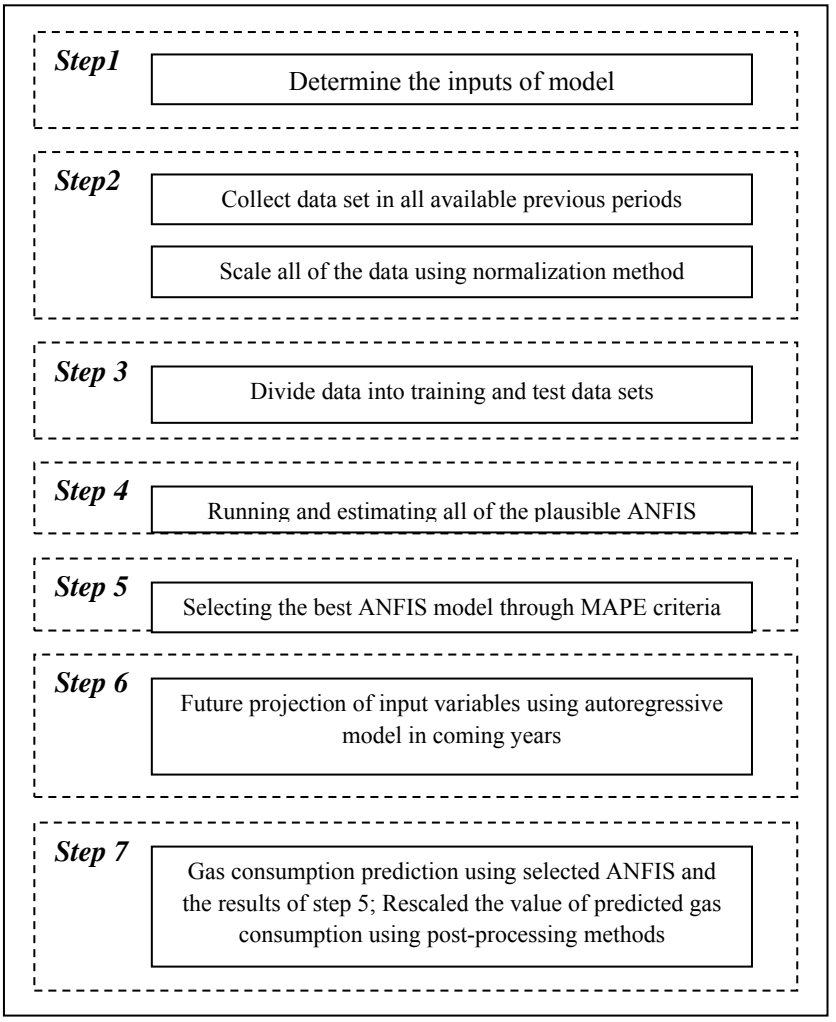

Figure 1
The proposed ANFIS approach for Long-term gas consumption forecasting

2.2 Adaptive Neuro-Fuzzy Inference System (ANFIS)

Neuro-fuzzy modeling refers to the way of applying various learning techniques developed in the neural network literature to fuzzy modeling or a fuzzy inference system (FIS) [8, 17]. A neuro-fuzzy system, which combine neural networks and fuzzy logic, have recently garnered a lot of interest in research and application. The neuro-fuzzy approach has added the advantage of reduced training time, not only due to its smaller dimensions but also because the network can be initialized with parameters relating to the problem domain. Such results emphasize the

benefits of the fusion of fuzzy and neural network technologies as it facilitates an accurate initialization of the network in terms of the parameters of the fuzzy reasoning system. Various types of FIS are reported in literature [21, 28, and 29] and each is characterized by their consequent parameters only.

A specific approach in neuro-fuzzy development is the adaptive neuro-fuzzy inference system (ANFIS), which has shown significant results in modeling nonlinear functions (Jang et al., 1997). ANFIS uses a feed forward network to search for fuzzy decision rules that perform well on a given task. Using a given input-output data set, ANFIS creates a FIS whose membership function parameters are adjusted using a back propagation algorithm alone or a combination of a back propagation algorithm with a least squares method. This allows the fuzzy systems to learn from the data being modeled.

Consider a first order Takagi-Sugeno fuzzy model with a two input (x,y), one output system having two membership functions for each input. A first-order Sugeno fuzzy model has two rules:

- Rule1: If x is A1 and y is B1, then f1 = p1x + q1y + r1
- Rule2: If x is A2 and y is B2, then f2 = p2x + q2y + r2

Layer 1: Then, the functioning of ANFIS is a five-layered feed-forward neural structure, and the functionality of the nodes in these layers can be summarized as:

$$\begin{aligned} o_{1,i} &= \mu_{A_i}(x) \qquad for\, i = 1,2 \\ o_{1,i} &= \mu_{B_{i-2}}(y) \qquad for\, i = 3,4 \end{aligned} \tag{3}$$

Layer 2: Where *x* or *y* is the input to the node, *Ai* or *Bi-2* is a fuzzy set associated with this node. At the first layer, for each input, the membership grades in the corresponding fuzzy sets are estimated. $O_{1,i}$ is the membership grade of a fuzzy set (A_1,A_2,B_1,B_2). At the second layer, all potential rules between the inputs are formulated by applying fuzzy intersection (AND). The product operation is used to estimate the firing strength of each rule.

$$o_{2,i} = w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \qquad \text{i= 1, 2} \tag{4}$$

Layer 3: The third layer is used for estimation of the ratio of the *i*th rule's firing strength to the sum of all rule's firing strengths.

$$o_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2} \qquad i = 1,2 \tag{5}$$

Layer 4:

$$o_{4,i} = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \tag{6}$$

Where $\overline{w}_i$ is the output of layer 3 and {*pi, qi, ri*} is the parameter set. Parameters in this layer will be referred to as consequent parameters.

Layer 5: The final layer computes the overall output as the summation of all incoming signals from layer 4.

$$\text{Overall output} = o_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{7}$$

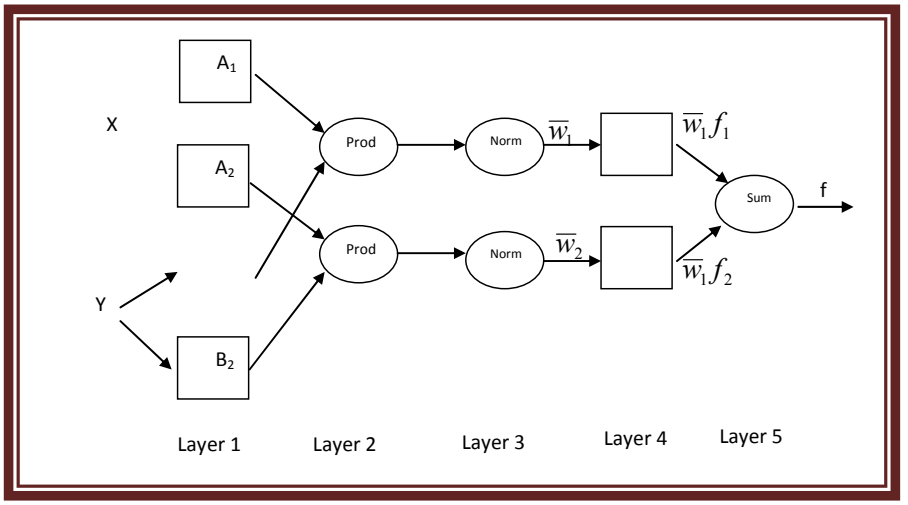

Figure 2
ANFIS structure with two inputs

Figure 2 shows the structure of the explained ANFIS. Optimizing the values of the adaptive parameters is of vital importance for the performance of the adaptive system. Jang et al. (1997) developed a hybrid learning algorithm for ANFIS which is faster than the classical back-propagation method to approximate the precise value of the model parameters. The hybrid learning algorithm of ANFIS consists of two alternating phases: (1) gradient descend, which computes error signals recursively from the output layer backward to the input nodes, and (2) the least squares method, which finds a feasible set of consequent parameters. We observe that, given fixed values of elements of premise parameters, the overall output can be expressed as a linear combination of the consequent parameters. The ANFIS architecture is not unique. Some layers can be combined and still produce the same output. In this ANFIS architecture, there are two adaptive layers (1, 4). Layer 1 has three modifiable parameters (a_i, b_i and c_i) pertaining to the input MFs. These parameters are called *premise* parameters. Layer 4 has also three modifiable

parameters (p_i, q_i and r_i) pertaining to the first order polynomial. These parameters are called *consequent* parameters. Figure 3 presents the structure of the proposed ANFIS for long term consumption. The reader should note that only two out five input layers are illustrated (population and demand of previous year) and because of redundancy the remaining three are only stated in the figure. Moreover, their operation are the same as the two shown inputs.

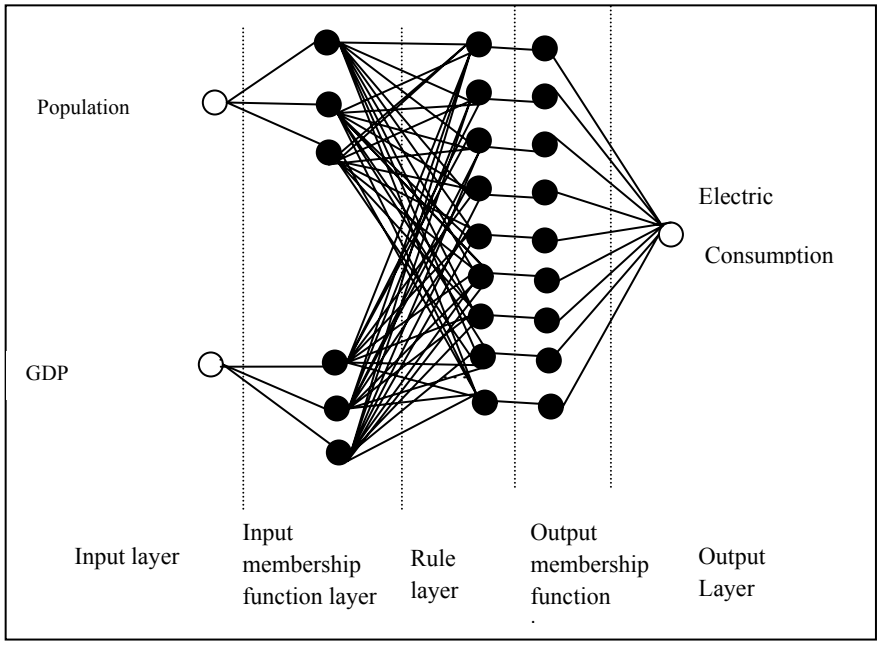

Figure 3
The structure of ANFIS model for long term Electric consumption estimation and forecasting

3 The Case Study

The proposed algorithm is applied to 27 set of data which are the annual electricity net consumption in industrialized nations which are the Netherlands, Luxembourg, Ireland, and Italy from 1980 to 2006. The proposed algorithm is applied to the data set as follows.

3.1 Step 1

Population and Gross Domestic Product (GDP) are considered as the input variables of the ANFIS in the present study.

3.2 Step 2

As regards the input variables (population & GDP) that are determined in step 1 and the output variable (electricity consumption), the related data are drawn from the World Bank Development Indicators, 2008, the World Bank. The raw data with respect to the two independent variables for the Netherlands, Luxembourg, Ireland, and Italy are shown in Tables 1 and 2.

Table 1

Raw data, Luxembourg, Italy[1]

Year	Luxembourg Population (Millions)	Luxembourg GDP (per capita)	Luxembourg Electric consumption	Italy Population (Millions)	Italy GDP (per capita)	Italy Electric consumption
1980	0.36	10.69	3.709	56,47	1152,10	165,849
1981	0.37	10.63	3.582	56,52	1161,10	164,273
1982	0.37	10.75	3.618	56,56	1168,43	166,813
1983	0.37	11.08	3.721	56,65	1182,87	165,992
1984	0.37	11.076	3.947	56,72	1215,50	175,4
1985	0.37	12.10	3.998	56,75	1251,66	179,513
1986	0.37	13.06	4.034	56,76	1283,27	184,902
1987	0.37	13.35	4.12	56,75	1321,58	194,685
1988	0.38	14,73	4.037	56,76	1373,76	187,88
1989	0.38	16.18	4.365	56,76	1413,21	213,1269
1990	0.38	16.53	4.433	56,77	1441,11	220,2536
1991	0.39	17.55	4.702	56,77	1461,14	225,4425
1992	0.39	18.33	4.477	56,86	1472,26	230,9819
1993	0.40	19.93	4.588	57,05	1459,25	230,4566
1994	0.40	20.77	4.968	57,20	1491,46	237,8341
1995	0.41	21.51	5.329	57,30	1534,02	244,9667
1996	0.42	22.23	5.47	57,39	1550,78	247,6424
1997	0.42	24.08	5.47485	57,51	1582,21	254,9867
1998	0.43	25.74	5.621	57,58	1610,59	262,7742
1999	0.43	27.75	5.772	57,63	1637,39	269,4803
2000	0.44	30.26	5.995	57,75	1686,95	281,1846
2001	0.44	30.73	5.94895	57,87	1716,71	287,0337
2002	0.45	31.49	5.87395	57,95	1723,29	293,7765
2003	0.46	32.41	6.168	58,03	1737,67	298,1345
2004	0.46	33.88	6.4095	58,09	1748,83	294,0081
2005	0.47	34.95	6.315	58,13	1761,70	309,5107
2006	0.47	37.43	6.74795	58,16	1782,93	316,2999

[1] Source: The data are drawn from World Bank Development Indicators, 2008, World Bank.

Table 2
Raw data for Netherland, Ireland

Year	Nether-land	Population (Millions)	GDP (per capita)	Electric consumption	Ire-land	Population (Millions)	GDP (per capita)	Electric consumption
1980		14.14	348.20	59.889		3.40	53.98	8.640
1981		14.25	346.42	59.317		3.44	55.77	8.527
1982		14.31	342.39	58.087		3.48	57.05	8.688
1983		14.36	348.23	59.24		3.50	56.91	8.626
1984		14.42	359.69	61.153		3.53	59.38	8.868
1985		14.49	370.77	62.111		3.54	61.22	9.691
1986		14.57	380.97	63.044		3.54	60.95	9.957
1987		14.67	386.37	65.841		3.54	63.80	10.829
1988		14.76	396.50	69.312		3.53	67.13	10.769
1989		14.85	415.07	71.512		3.51	71.03	11.523
1990		14.95	432.13	72.781		3.51	77.04	12.134
1991		15.07	441.93	75.832		3.53	78.53	12.698
1992		15.17	450.88	78.05995		3.56	81.15	13.46675
1993		15.27	454.32	79.4566		3.58	83.34	13.80225
1994		15.38	468.96	82.14895		3.60	88.14	14.39405
1995		15.46	482.02	84.2162		3.61	96.86	15.0712
1996		15.53	496.67	87.34025		3.64	106.69	16.0523
1997		15.60	515.74	89.34798		3. 67	116.30	16.87472
1998		15.70	538.17	92.6172		3.71	126.35	17.83555
1999		15.80	559.67	94.96385		3.75	140.60	18.899
2000		15.91	579.08	99.1732		3.79	155.11	20.377
2001		16.02	587.34	101.3283		3.84	164.43	21.031
2002		16.12	590.68	102.5648		3.88	174.51	21.94
2003		16.22	585.49	103.7684		3.92	180.89	22.5353
2004		16.32	593.91	106.7292		4.97	189.70	23.28025
2005		16.41	603.00	108.1612		4.02	198.53	24.089
2006		16.49	621.12	109.5509		4.06	209.92	25.674

3.3 Step 3

The 28 rows of data are divided into 24 training data (1980-2003) sets and 4 test data sets (2004-2006). According to our problem, the extrapolation and prediction ability of ANN should be calculated; therefore the data for test use chosen of the period which is closer to the last year (2004- 2006).

3.4 Step 4

Two parameters have been considered in the construction and examination of plausible ANFIS models. The type of membership function and the number of linguistic variables are two mentioned parameters. Six different membership functions are considered in building the ANFIS, as follows: a built-in membership function composed of the difference between two sigmoidal membership functions (dsig), the Gaussian combination membership function (gauss2), the Gaussian curve built-in membership function (gauss), The generalized bell-shaped built-in membership function (gbell), the Π-shaped built-in membership function (pi), and the Trapezoidal-shaped built-in membership function (trap). Also, the number of linguistic variables have been considered as between 2 and 20. MATLAB software has been used to Running ANFIS in the present study.

3.5 Step 5

The architectures which are shown below have minimum MAPE among all of the other architectures. The architectures with the minimum of MAPE are shown below. Table 3 presents the structures of long-term models. MF stands for Membership Function in the ANFIS approach.

Table 3

The results of MAPE for the best ANFIS model for each country

Countries	Number of MF's	MF Type	MAPE (%)
Italy	**2**	**gaussmf**	**0.014**
Luxembourg	**2**	**gauss2mf**	**0.02**
the Netherlands	**2**	**Trapmf**	**0.007**
Ireland	**3**	**gbell**	**0.012**

3.6 Step 6

In order to forecast with the ANFIS, there is a need to forecast the independent variables *population and GDP*. In addition, previous studies used time series to forecast the independent variables [1]. The future projection for each independent variable is performed by autoregressive model that is one of main models in the time series area.

3.6.1 Data Structure

AR (1) is considered thetime series model because of the shortage of annual data. Therefore the time series model for independent variable X_i is:

$$X_i = a\, X_{i-1} + b \tag{15}$$

Tables 4 and 5 show coefficients estimation of models for Luxembourg-Italy and the Netherlands - Ireland, respectively.

Table 4
Models information in Luxembourg, Italy

	Luxembourg	a	B	Italy	A	B
GDP		1.01387	0.646877		0.977829	0.44674
Population		0.991206	0.006972		0.99747	0.203405

Table5
Models information in Netherland, Ireland

	Netherland	A	b	Ireland	A	b
GDP		0.996963	0.20975		0,97	0,04
Population		0.989339	0.245892		0,92	0,08

3.7 Step 7: Electric Future Projection

According to the data structure, the values of the independent variables from 2008 to 2015 are fed to the selected ANFIS as inputs, and then the output values are obtained for the same period. The projections of electricity consumption in the Netherlands, Luxembourg, Ireland, and Italy during 2008-2015 are shown in Table 6 and Figures 4 to 7.

Table 6
Projections of Electric consumption during 2008-2015

	Italy	Luxembourg	Netherland	Ireland
2009	322.2903	*6.9407*	112.9254	32.04722
2010	322.9516	*6.9933*	112.3662	34.48428
2011	329.6455	*7.1541*	114.4998	37.08778
2012	330.2368	*7.1932*	114.2775	39.86907
2013	336.1318	*7.3405*	116.1477	42.84032
2014	336.5691	*7.3679*	115.9514	46.01447
2015	341.7443	*7.495*	117.5679	49.4054

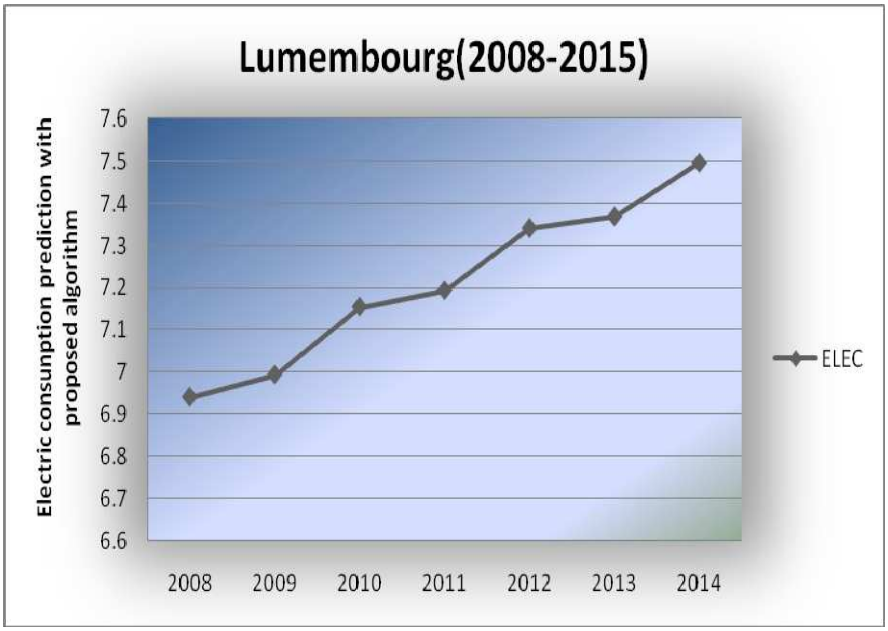

Figure 4
Luxembourg annually consumption forecasting with proposed algorithm (2008- 2015)

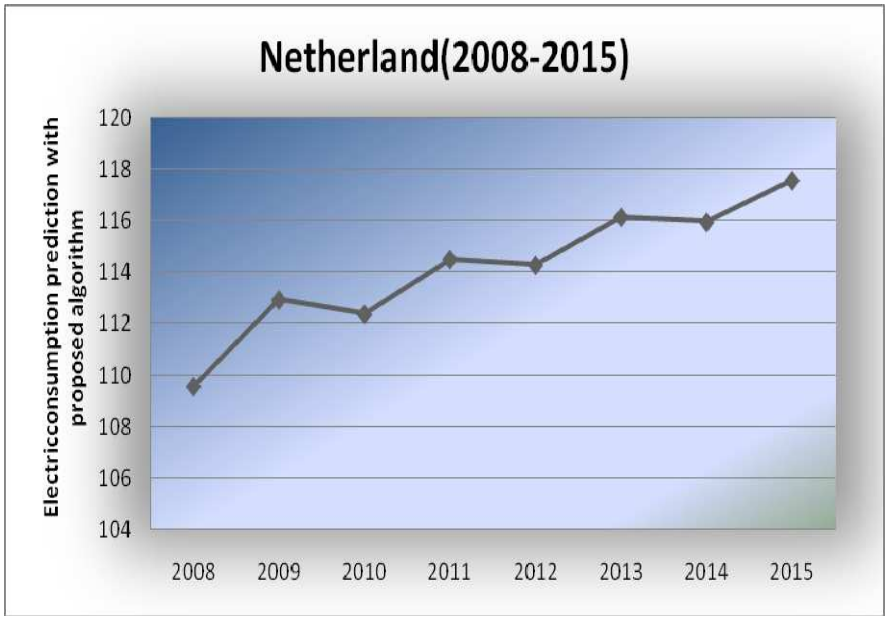

Figure 5
Netherland annually consumption forecasting with proposed algorithm (2008- 2015)

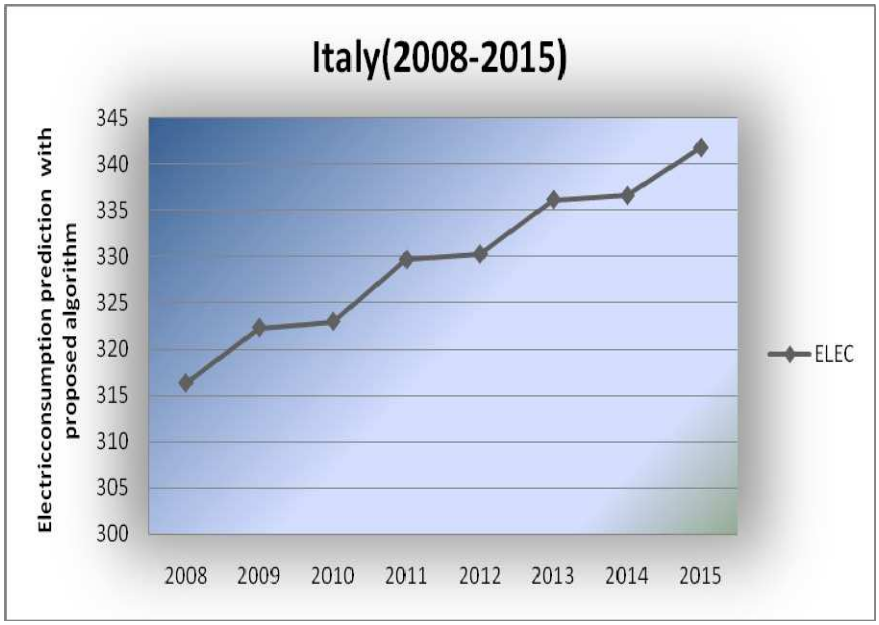

Figure 6
Italy annually consumption forecasting with proposed algorithm (2008- 2015)

Figure 7
Ireland annually consumption forecasting with proposed algorithm (2008- 2015)

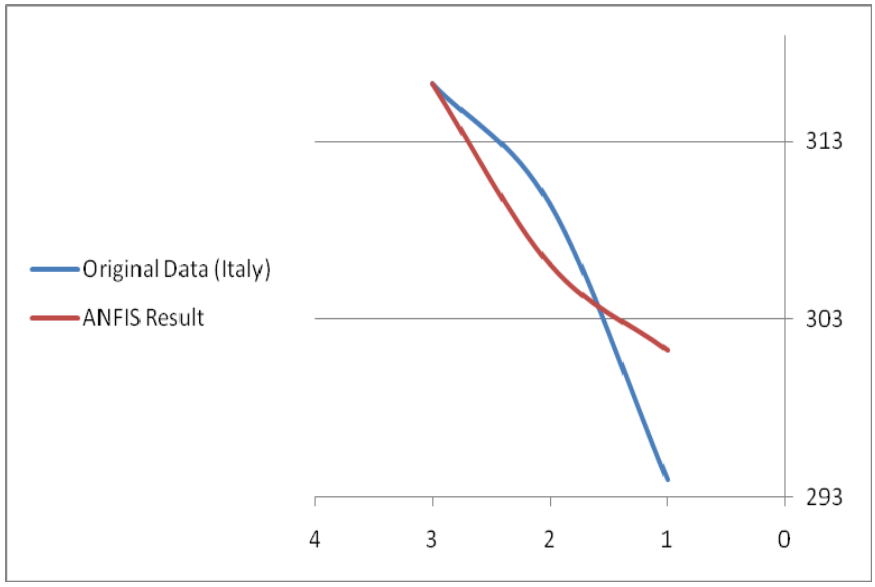

Figure 8
ANFIS results versus actual Electric consumption for 2004 to 2006 (Italy)

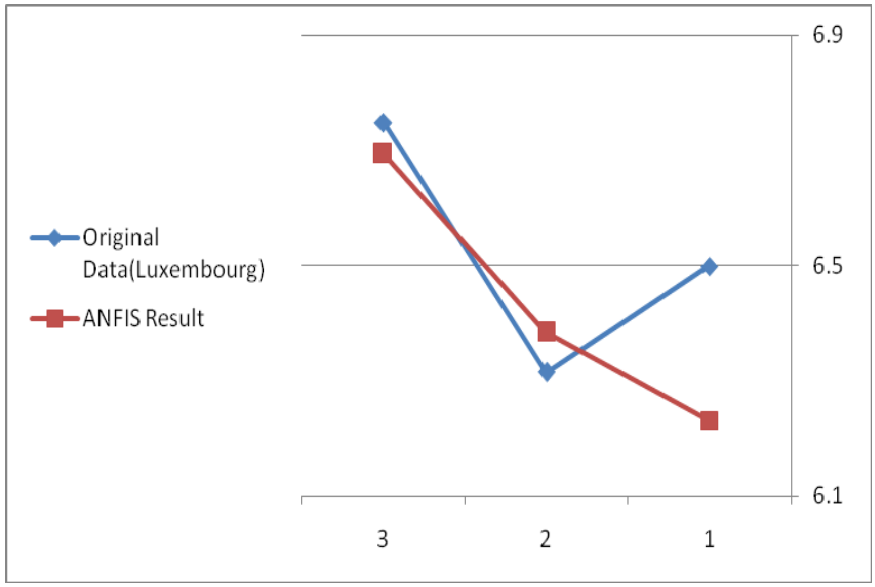

Figure 9
ANFIS results versus actual Electric consumption for 2004 to 2006 (Luxembourg)

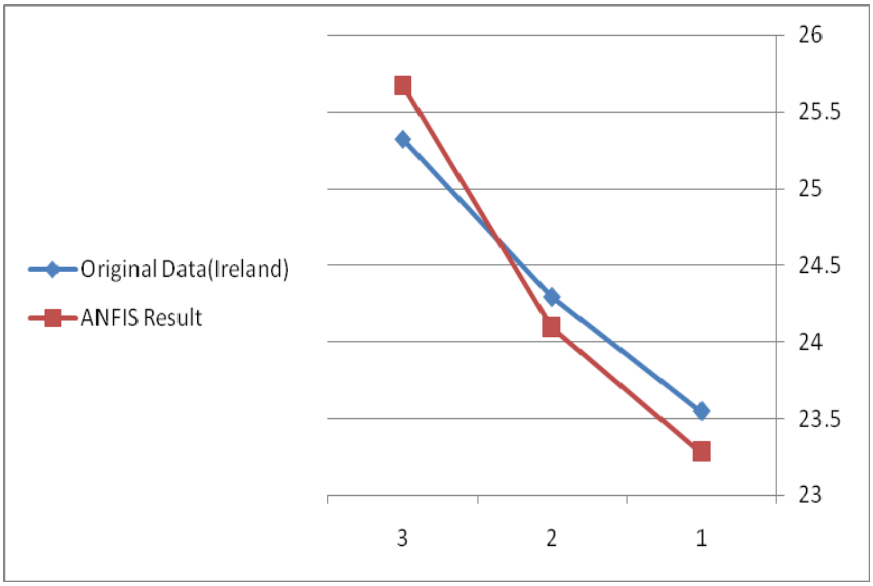

Figure 10
ANFIS results versus actual Electric consumption for 2004 to 2006 (Ireland)

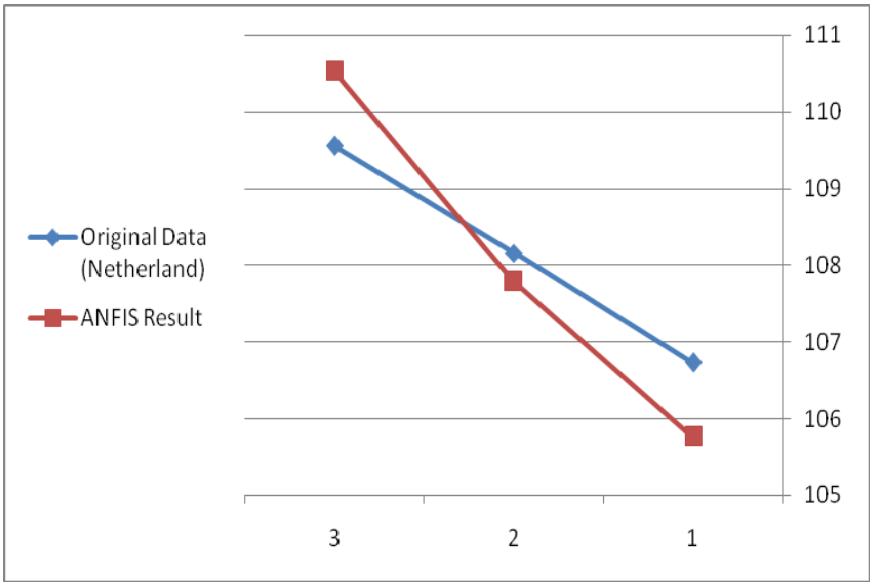

Figure 11
ANFIS results versus actual Electric consumption for 2004 to 2006 (Netherland)

Conclusions

This study presented an ANFIS approach for long term electricity consumption prediction. It presented the ANFIS and AR models for forecasting long-term natural gas demand prediction. The result of the ANFIS approach indicates that the proposed models are suitable and accurate for predicting electricity demand in industrialized nations, which in this case are the Netherlands, Luxembourg, Ireland, and Italy. In our long-term model we considered explanatory variables such as population and GDP. These variables showed a strong explanatory capability for estimating the actual electricity demand. Moreover, optimum results are obtained when all variables are used in ANFIS models. Figures 8 to 11 present the ANFIS results versus actual electricity consumption for 2004 to 2006. As can be seen, the ANFIS results are very close to the actual consumption. The approach of this study may be used for other cases to estimate and forecast optimum gas consumption.

The ANFIS algorithm is also compared with some of the current studies in the estimation of gas estimation. Its features are compared with previous models to show its advantages over previous models (Table 7). The ANFIS algorithm of this study is capable of dealing with both data complexity and ambiguity due to its mechanism. Also, it pre-processes and post-processes the given data to provide higher precision. In addition, it dominates all recent studies and is capable of dealing with complexity, ambiguity and uncertainty.

Table 7

The features of the ANFIS approach versus other methods

Feature / Method	Data Complexity and Non-Linearity	Data Uncertainty and Non-Crisp Data set	Intelligent Modeling and Forecasting	Fuzzy Data Modeling	High Precision and Reliability	Dealing Ambiguity	Data Pre-Processing and Post-Processing
The ANFIS approach	√	√	√	√	√	√	√
ANN	√		√		√		
Fuzzy Regression						√	
Linear Regression	√				√		
Nonlinear Regression	√		√		√	√	
Decision Tree	√			√	√	√	
Genetic Algorithm	√			√	√	√	

References

[1] Azadeh, A. Ghaderi, S. F. Sohrabkhani, S. (2007a) Forecasting Electrical Consumption by Integration of Neural Network, Time Series and ANOVA. Applied Mathematics and Computation, 186, 1753-1761

[2] Azadeh, A. Ghaderi, S. F. Tarverdian, S., Saberi, M. (2007b) Integration of Artificial Neural Networks and Genetic Algorithm to Predict Electrical Energy Consumption. Applied Mathematics and Computation, 186, 1731-1741

[3] Azadeh, A., Tarverdian, S. (2007c) Integration of Genetic Algorithm, Computer Simulation and Design of Experiments for Forecasting Electrical Energy Consumption. Energy Policy, 35, 5229-5241

[4] Azadeh, A. Ghaderi, S. F. Sohrabkhani, S. (2008a) A Simulated-based Neural Network Algorithm for Forecasting Electrical Energy Consumption in Iran. Energy Policy, 36(7), 2637-2644

[5] Azadeh, A. Ghaderi, S. F., Sohrabkhani, S. (2008b) Annual Electricity Consumption Forecasting by Neural Network in High Energy Consuming Industrial Sectors. Energy Conversion and Management, 49(8), 2272-2278

[6] Azadeh, A. Saberi, M. Ghaderi, S. F. Gitiforouz, A., Ebrahimipour, V. (2008c) Improved Estimation of Electricity Demand Function by Integration of Fuzzy System and Data Mining Approach. Energy Conversion and Management, 49, 2165-2177

[7] Azadeh, A. Saberi, M. Gitiforouz, A., Saberi, Z. (2009) A Hybrid Simulation Adaptive-Network-based Fuzzy Inference System for Improvement of Electricity Consumption Estimation. Expert Systems with Applications, In Press

[8] Brown, M., Harris, C. (1994) Neurofuzzy Adaptive Modeling and Control. New Jersey: Prentice Hall

[9] Chow, L. C. H. (2001) A Study of Sectoral Energy Consumption in Hong Kong (1984-97) with Special Emphasis on the Household Sector. Energy Policy, 29, 1099-110

[10] Ediger V. S., Tatlldil, H. (2002) Forecasting the Primary Energy Demand in Turkey and Analysis of Cyclic Patterns. Energy Conversion and Management, 43, 473-87

[11] Florides G. A., Kalogirou S. A., Tassou S. A., Wrobel L. C. (2000) Modelling of the Modern Houses of Cyprus and Energy Consumption Analysis. Energy, 25, 915-37

[12] Gonzales Chavez, S., Xiberta Bernat, J., Llaneza Coalla, H. (1999) Forecasting of Energy Production and Consumption in Asturias (Northern Spain). Energy, 24, 183-98

[13] Gorucu, F. B., Geris, P. U., Gumrah, F. (2004) Artificial Neural Network Modeling for Forecasting Gas Consumption. Energy Sources, 26(3), 299-307

[14] Gorucu, F. B., Gumrah, F. (2004) Evaluation and Forecasting of Gas Consumption by Statistical Analysis. Energy Sources, 26(3), 267-276

[15] Gutierrez, R. Nafidi, A., Sanchez, R. G. (2005) Forecasting Total Natural-Gas Consumption in Spain by Using the Stochastic Gompertz Innovation Diffusion Model. Applied Energy, 80(2), 115-124

[16] Jai Persaud, A., Kumar, U. (2001) An Eclectic Approach in Energy Forecasting: a Case of Natural Resources Canada's (NRCan's) Oil and Gas Outlook. Energy Policy, 29, 303-13

[17] Jang, J. S. R. (1993) ANFIS: Adaptive Network-based Fuzzy Inference System. IEEE Transactions on Systems, Man and Cybernetics, 23(3), 665-683

[18] Jang, R., Sun, C., Mizutani, E. (1997) Neuro-Fuzzy and Soft Computation, New Jersey: Prentice Hall

[19] Kadoshin S, Nishiyama T., Toshihide, I. (2000) The Trend in Current and Near Future of Energy Consumption from a Statistical Perspective. Applied Energy, 67, 407-17

[20] Khotanzad, A. Elragal, H., Lu, T. (2000) Combination of Artificial Neural-Network Forecasters for Prediction of Natural Gas Consumption. IEEE Transactions on Neural Networks, 11(2), 464-473

[21] Mamdani, E. H., Assilian, S. (1975) An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. International Journal of Man-Machine Studies, 7(1), 1-13

[22] Parikh, J., Purohit, P., Maitra, P. (2007) Demand Projections of Petroleum Products and Natural Gas in India. Energy, 32, 1825-1837

[23] Reddy, B. S., Balachandra, P. (2003) Integrated Energy–Environment-Policy Analysis: a Case Study of India. Utility Policy, 11(2), 59-73

[24] Rumelhart, D. E., McClelland J. L. (1986) Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Foundations, Cambridge, MA: MIT Press

[25] Saab, S., Badr, E., Nasr, G. (2001) Univariate Modeling and Forecasting of Energy Consumption: the Case of Electricity in Lebanon. Energy, 26, 1-14

[26] Sanchez-Ubeda, E. F., Berzosa, A. (2007) Modeling and Forecasting Industrial End-Use Natural Gas Consumption. Energy Economics, 29(4), 710-742

[27] Siemek, J., Nagy, S., Rychlicki, S. (2003) Estimation of Natural-Gas Consumption in Poland Based on the Logistic-Curve Interpretation. Applied Energy, 75(1-2), 1-7

[28] Takagi, T., Sugeno, M. (1985) Fuzzy Identification of Systems and its Application to Modeling and Control. IEEE Transactions on Systems, Man and Cybernetics 15(1), 116-132

[29] Tsukamoto, Y. (1979) An Approach to Fuzzy Reasoning Method. In M. M. Gupta, R. K. Ragade, R. R. Yager (Eds.). Advances in Fuzzy Set Theory and Application, North-Holland, Amsterdam, 137-149

[30] Yoo, S., Lim, H. J., Kwak, S. J. (2009) Estimating the Residential Demand Function for Natural Gas in Seoul with Correction for Sample Selection Bias. Applied Energy, 86(4), 460-465

[31] Werbos, P. I. (1974) Beyond Regression: New Tools for Prediction and Analysis in the Behavior Sciences. Ph.D Thesis, Harvard University, Cambridge, MA, USA

[32] Azadeh A, Saberi M, Nadimi V, Iman M, An Adaptive-Network-based Fuzzy Inference System for Long-Term Electricity Consumption Forecasting (2008-2015): A Case Study of Netherland, Luxembourg, Ireland, Italy, 10th International Conference on Fuzzy Set Theory and Applications. Liptovský Ján, Slovak Republic, February 1-5, 2010

[33] Tiberiu Tudorache1, Mihail Popescu, FEM Optimal Design of Wind Energy-based Heater, Acta Polytechnica Hungarica, Vol. 6, No. 2, 2009

Multilevel Fuzzy Approach to the Risk and Disaster Management

Márta Takács

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

Abstract: In this paper a short general review of the main characteristics of risk management applications is given, where a hierarchical, multilevel risk management method can be applied in a fuzzy decision making environment. The given case study is a travel risk-level calculation based on the presented model. In the last section an extended model and a preliminary mathematical description is presented, where the pairwise comparison matrix of the grouped risk factors expands the previous principles.

Keywords: risk management; fuzzy multilevel decision making; comparison matrix

1 Introduction

The economical crisis situations and the complex environmental and societal processes over the past years indicate the need for new mathematical model constructions to predict their effects. The health diagnostic as a multi-parameter and multi-criteria decision making system is, as well, one of the models where, as in the previous examples, a risk model should be managed.

Haimes in [1] gives an extensive overview of risk modeling, assessment, and management. The presented quantitative methods for risk analysis in [2] are based on well-known mathematical models of expert systems, quantitative optimum calculation models, statistical hypothesis and possibility theory. The case studies present applications in the fields of economics and environmental protection. It is observable that the statistical-based numerical reasoning methods need long-term experiments and that they are time- and computationally demanding. The complexity of the systems increases the runtime factor, and the system parameter representation is usually not user-friend. The numerical methods and operation research models are ready to give acceptable results for some finite dimensional problems, but without management of the uncertainties. The complexity and uncertainties in those systems raise the necessity of soft computing based models.

Nowadays the expert engineer's experiences are suited for modeling operational risks, not only in the engineering sciences, but also for a broad range of applications [13]. Wang introduces the term of risk engineering related to the risk of costs and schedules on a project in which there is the potential for doing better as well as worse than expected [3]. The presented case studies in his book are particularly based on long-term engineering experiences, for example on fuzzy applications, which offer the promised alternative measuring of operational risks and risk management globally.

The use of fuzzy sets to describe the risk factors and fuzzy-based decision techniques to help incorporate inherent imprecision, uncertainties and subjectivity of available data, as well as to propagate these attributes throughout the model, yield more realistic results. Fuzzy logic modeling techniques can also be used in risk management systems to assess risk levels in cases where the experts do not have enough reliable data to apply statistical approaches.

There are even more applications to deal with risk management and based on fuzzy environments. Fuzzy-based techniques seem to be particularly suited to modeling data which are scarce and where the cause-effect knowledge is imprecise and observations and criteria can be expressed in linguistic terms. [4]

The structural modeling of risk and disaster management is case-specific, but the hierarchical model is widely applied. The system characteristics are as follows: it is a multi-parametrical, multi-criteria decision process, where the input parameters are the measured risk factors, and the multi-criteria rules of the system behaviors are included in the decision process. The Analytical Hierarchy Process (AHP) expands this complex system with the pairwise comparison of the factors' importance and interaction [5].

In this paper, after a short general review of the main characteristics of risk management applications, a hierarchical, multilevel risk management method will be presented in a fuzzy environment. The given case study is a travel risk-level calculation based on the presented model. In the last section a preliminary mathematical description is presented based on a pairwise comparison matrix and AHP expanded principles.

2 Risk Management

Risk management is the identification, assessment, and prioritization of risks, defined as the effects of uncertainty of objectives, whether positive or negative, followed by the coordinated and economical application of resources to minimize, monitor, and control the probability and/or impact of unfortunate events [6].

The techniques used in risk management have been taken from other areas of system management. Information technology, the availability of resources, and

other facts have helped to develop the new risk management with the methods to identify, measure and manage the risks, thereby reducing the potential for unexpected loss or harm [7]. Generally, a risk management process involves the following main stages.

The first step is the identification of risks and potential risks to the system operation at all levels. Evaluation, the measure and structural systematization of the identified risks, is the next step. Measurement is defined by how serious the risks are in terms of consequences and the likelihood of occurrence. It can be a qualitative or quantitative description of their effects on the environment. Plan and control are the next stages to prepare the risk management system. This can include the development of response actions to these risks, and the applied decision or reasoning method. Monitoring and review, as the next stage, is important if we are to have a system with feedback, and the risk management system is open to improvement. This will ensure that the risk management process is dynamic and continuous, with correct verification and validity control. The review process includes the possibility of new additional risks and new forms of risk description. In the future the role of complex risk management will be to try to increase the damaging effects of risk factors.

2.1 Fuzzy Risk Management

Risk management is a complex, multi-criteria and multi-parametrical system full of uncertainties and vagueness. Generally the risk management system in its preliminary form contains the identification of the risk factors of the investigated process, the representation of the measured risks, and the decision model. The system can be enlarged by monitoring and review in order to improve the risk measure description and decision system. The models for solving are knowledge-based models, where linguistically communicated modeling is needed, and objective and subjective knowledge (definitional, causal, statistical, and heuristic knowledge) is included in the decision process. Considering all these conditions, fuzzy set theory helps manage complexity and uncertainties and gives a user-friendly visualization of the system construction and working model.

Fuzzy-based risk management models assume that the risk factors are fuzzified (because of their uncertainties or linguistic representation); furthermore the risk management and risk level calculation statements are represented in the form of *if premises then conclusion* rule forms, and the risk factor calculation or output decision (summarized output) is obtained using fuzzy approximate reasoning methods. Considering the fuzzy logic and fuzzy set theory results, there are further possibilities to extend fuzzy-based risk management models modeling risk factors with type-2 fuzzy sets, representing the level of the uncertainties of the membership values, or using special, problem-oriented types of operators in the fuzzy decision making process.

The hierarchical or multilevel construction of the decision process, the grouped structural systematization of the factors, with the possibility of gaining some subsystems, depending on their importance or other significant environment characteristics or on laying emphasis on risk management actors, is a possible way to manage the complexity of the system. Carr and Tah describe a common hierarchical-risk breakdown structure for developing knowledge-driven risk management, which is suitable for the fuzzy approach [8].

Starting with a simple definition of the risk as the adverse consequences of an event, such events and consequences are full of uncertainty, and inherent precautionary principles, such as sufficient certainty, prevention, and desired level of protection. All of these can be represented as fuzzy sets. The strategy of the risk management may be viewed as a simplified example of a precautionary decision process based on the principles of fuzzy logic decision making [9].

3 Grouped, Weighted Fuzzy Model

Based on the main ideas from [8] a risk management system can be built up as a hierarchical system of risk factors (inputs), risk management actions (decision making system) and direction or directions for the next level of risk situation solving algorithm. Actually, those directions are risk factors for the action on the next level of the risk management process. To sum this up: risk factors in a complex system are grouped to the risk event where they figure. The risk event determinates the necessary actions to calculate and/or increase the negative effects. Actions are described by 'if … then' type rules.

With the output those components frame one unit in the whole risk management system, where the items are attached on the principle of the time-scheduling, significance or other criteria (Fig. 1). Input Risk Factors (RF) grouped and assigned to the current action are described by the Fuzzy Risk Measure Sets (FRMS) such as 'low', 'normal', 'high', and so on. Some of the risk factor groups, risk factors or management actions have a different weighted role in the system operation. The system parameters are represented with fuzzy sets, and the grouped risk factors values give intermitted results [14]. Considering some system input parameters, which determine the risk factors' role in the decision making system, intermitted results can be weighted and forwarded to the next level of the reasoning process.

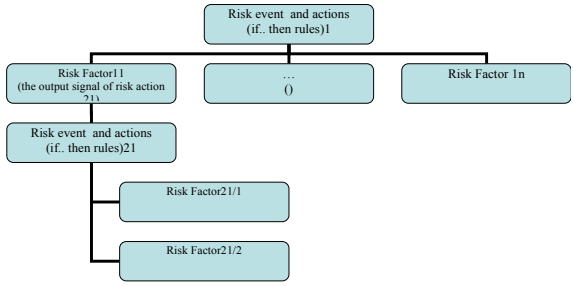

Figure 1
The hierarchical risk management construction

3.1 Disaster Management - Case Studies

Disaster event monitoring as one of the steps in risk and crisis management is a very complex system with uncertain input parameters. Fuzzified inputs, the fuzzy rule base, which is constructed using objective and subjective definitional, causal, statistical, and heuristic knowledge, is able to present the problem in a user-friendly form. The complexity of the system can be managed by the hierarchically-structured reasoning model, with a thematically-grouped, and if necessary, gained risk factor structure.

Crisis or disaster event monitoring provides basic information for many decisions in today's social life. The disaster recovery strategies of countries, the financial investments plans of investors, or the level of the tourism activities all depend on different groups of disaster or crisis factors. A disaster can be defined as an unforeseen event that causes great damage, destruction and human suffering, evolved from a natural or man-made event that negatively affects life, property, livelihood or industry. A disaster is the start of a crisis, and often results in permanent changes to human societies, ecosystems and the environment.

Based on the experts' observations [11], [12], the risk factors which prejudice disaster situation can be classified as follows:

- natural disasters;

- man-made disasters (unintended events or willful events).

Natural disasters arise without direct human involvement, but may often occur, because of human actions prior, during or after the disaster itself (for example, a hurricane may cause flooding by rain or by a storm surge).

The natural disasters can also be grouped primarily based on the root cause:

- hydro-meteorological disasters: floods, storms, and droughts;

- geophysical disasters: earthquakes, tsunamis and volcanic eruptions;

- biological disasters: epidemics and insect infestations;

or they can be structured hierarchically, based on sequential supervention.

The example, presented in this paper, is constructed based on the first principle, with fuzzified inputs and a hierarchically-constructed rule base system (Figure 2). The risk or disaster factors, as the inputs of one subsystem of the global fuzzy decision making system, give outputs for the next level of decision, where the main natural disaster classes result is the total impact of this risk category.

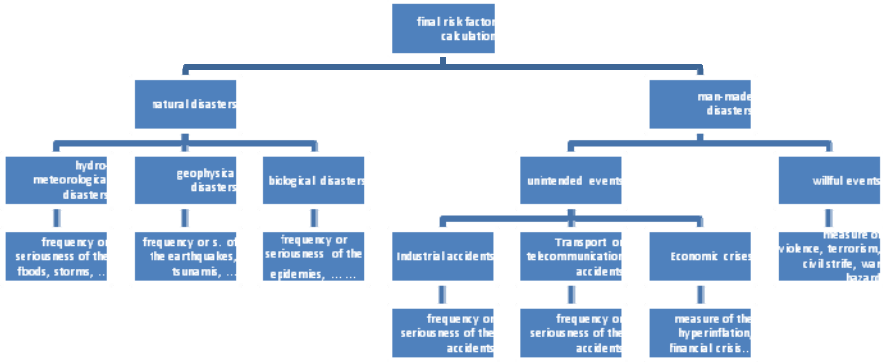

Figure 2

Hierarchically constructed rule base system

This approach allows additional possibilities to handle the set of risk factors.

It is easy to add one factor to a factors-subset; the complexity of the rule base system is changed only in the affected subsystem.

In different seasons, environmental situations etc., some of the risk groups are more important for the global conclusion than others, and this can be achieved with an importance factor (number from the [0,1]).

Man-made disasters have an element of human intent or negligence. However, some of those events can also occur as the result of a natural disaster. Man-made factors and disasters can be structured in a manner similar to the natural risks and events. One of the possible classifications of the basic man-made risk factors or disaster events (applied in our example) is as follows:

1. unintended events:

   - Industrial accidents (chemical spills, collapses of industrial infrastructures);

   - Transport or telecommunication accidents (by air, rail, road or water means of transport);

- Economic crises (growth collapse, hyperinflation, and financial crisis);

2. willful events (violence, terrorism, civil strife, riots, and war).

In the investigated example, the effects of man-made disasters as inputs in the decision making process are represented with their relative frequency, and the premises of the related fuzzy rules are very often represented with the membership functions: never, rarely, frequently, etc.[1]

The input parameters are represented on the unit universe [0,1] with triangular or trapezoidal membership functions describing the linguistic variables such as the frequency of the floods, for example: "low", "medium" or "high" (Fig. 3). The system was built in the Matlab Fuzzy Toolbox and Simulink environment.

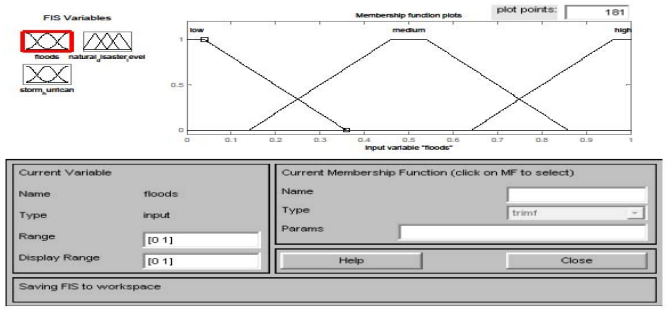

Figure 3
Membership functions of the flood frequencies

The risk and disaster factors are grouped in two main groups: human- and nature-based group. The inputs are crisp, but the rule base system is hierarchically constructed (Fig. 4), and the decision making is Mamdani type approximate reasoning with basic *min* and *max* operators.

[1] The Matlab Fuzzy Toolbox and Simulink elements were in the preliminary, partial form constructed by Attila Karnis, student of the Óbuda University as the project on the course "Fuzzy systems for engineers".

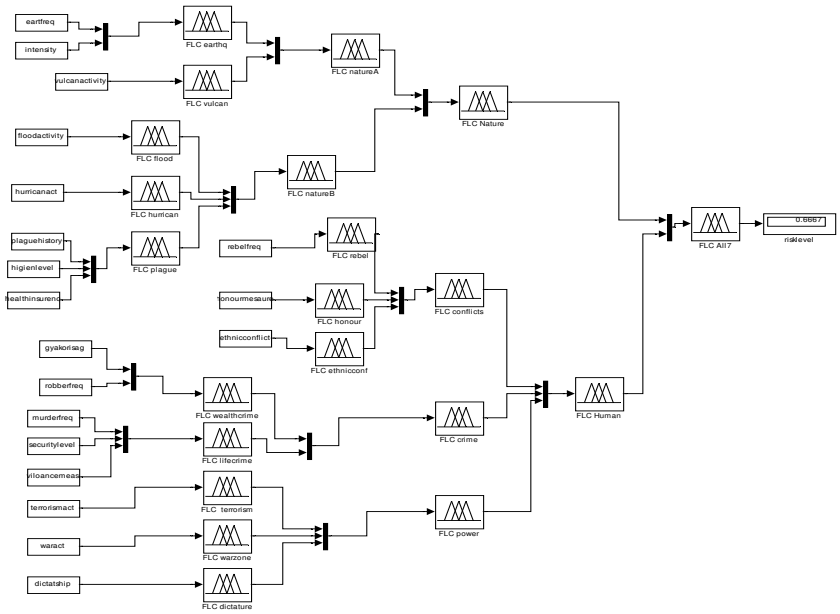

Figure 4

The system construction for the effects of disasters to calculate the travel risk level in a country

The final conclusion based on both disasters' as risk factors' groups is shown in Figure 5.

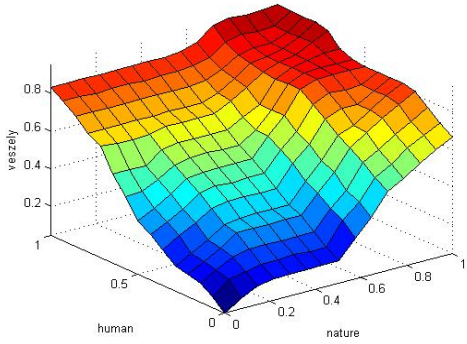

Figure 5

The final conclusion based on both disasters' as risk factors' groups

4 First Step to the Fuzzy AHP Model for Group-based Risk Management Model

Let $X_1, X_2, ..., X_n$ be the set of elements in a decision making system. It is a natural way to use the framework of a $A_{n\times n}$ square matrix to represent the pairwise comparisons of the dominance and interaction of those elements. Analytical Hierarchy Process (AHP) is a method for estimating the preference values from the pairwise comparison matrix. APH allows for the consideration of both qualitative and quantitative aspects of the decision, expanding the decision with the one-to-one comparison of the objectives, criteria, constraints or alternatives in the system model. The pairwise comparison in the AHP assume that the decision-maker can compare any two elements, for example X_i and X_j at the same level of the hierarchy in the system and provide a numerical value a_{ij} for the ratio of their importance. Saaty suggests using scale 1 to 9 to describe the preference measures [5], but in different applications there are presented other possible scales too [10].

Let $a_{ij} > 1$ if the element X_i is preferred to X_j, correspondingly, the reciprocal property $a_{ji} = 1/a_{ij}$ for *i*=1,2,...n, *j*=1,2,...n.

Each set of comparisons for a level with *n* elements requires $\frac{n \cdot (n-1)}{2}$ judgments, which are further used to construct a positive reciprocal matrix $A_{n\times n}$ of pairwise comparisons [10].

Let us interpret the comparison matrix $A_{n\times n}$ as the matrix of the dominance measures regarding the set of risk factors in a risk management system.

If the factors are grouped, and the groups are more or less independent, the comparison matrix has the block diagonal matrix form, and this allows us to pare down the computation complexity.

Example. Let $X_1, X_2, ..., X_n$ be the set of risk factors grouped in *p* groups, and let it contain the first factors group the factors X_1, X_2, X_3. The pairwise comparison of them is represented with the 3×3 dimensional sub-matrix A_{11}. The further representations are similar to this, so the next to last group contains two factors: X_{n-2}, X_{n-1}, with the 2×2 dimensional sub-matrix $A_{p-1,p-1}$, the last group holds only one factor.

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & \\ a_{21} & a_{22} & a_{23} & 0 & 0 & 0 & \\ a_{31} & a_{32} & a_{33} & & & & \\ 0 & & & \ddots & 0 & 0 & \\ 0 & & 0 & a_{n-2,n-2} & a_{n-2,n-1} & 0 & \\ & & & a_{n-1,n-2} & a_{n-1,n-1} & & \\ 0 & & 0 & 0 & & a_{nn} & \end{bmatrix}=\begin{bmatrix} A_{11} & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & A_{p-1,p-1} & 0 \\ 0 & 0 & 0 & A_{p,p} \end{bmatrix}$$

It is natural that the comparison values a_{ii} are units, $a_{ii}=1$ for all i=1,2,...n.

Let $x=(x_1,x_2,...,x_n)$ be the actual input vector of the risk factors' vector $X=(X_1,X_2,...,X_n)$. The influence of the pairwise dominance comparison of the factors on the actual input vector can be represented as a transformation described with the matrix operation $A\cdot x^T$. The goal is to forward a weighted input vector to the system, where the weight-multiplier λ holds up the information about the pairwise dominance comparison of the input factors:

$A\cdot x^T=\lambda\cdot x^T$.

The method for computing the λ multiplier can be the eigenvalue method. On a practical score only real eigenvalues can be accepted. If there are not real eigenvalues in the set of solutions, the multiplier λ is a unit one, λ=1.

If there exists more than one solution with the proposed conditions, the chosen one should be the eigenvalue which keep the input vectors in their universe, but permits the highest efficiency of the decision. The AHP should be applied before the risk level calculation or decision making process.

The open problems are:

- to find the best way to create pairwise comparison of the factors, because the values are the judgments obtained from an appropriate semantic scale. In practice the decision-makers usually give some or all pair-to-pair comparison values with an uncertainty degree rather than precise ratings;
- to adjust the scale of the comparison values to keep the weighted input vector in their universe, but permitting the highest efficiency of the decision;
- to build up a fuzzy AHP model for the preliminary comparison of the risk factors in the risk management system.

Conclusions

Risk management applications are complex, multi-criteria and usually multilevel decision systems, required to manage uncertainties. The fuzzy environment is able

to represent the ambiguous risk factors and rules in an acceptable form, where the risk factors are grouped based on their roles in the decision-making system. The given case study is a travel risk-level calculation based on the presented model.

The pairwise comparison matrix is the first step in introducing the fuzzy AHP model for the multilevel, hierarchically-structured risk management system, with further open problems and the possibility for fine tuning in the reasoning process.

Acknowledgement

This work was supported by the Óbuda University and Vojvodina Academy of Science and Art (project title: Mathematical Models for Decision Making under Uncertain Conditions and Their Applications).

References

[1] Haimes, Y. Y., *Risk Modeling, Assessment, and Management*. 3rd Edition. John Wiley & Sons (2009)

[2] Vose, D., Risk Analysis: a Quantitative Guide, 3rd Edition. John Wiley & Sons (2009)

[3] Wang, J. X., Roush, M. L., http://www.amazon.com/gp/product/0824793013/ref=pd_lpo_k2_dp_sr_2?pf_rd_p=1278548962&pf_rd_s=lpo-top-stripe-1&pf_rd_t=201&pf_rd_i=047199765X&pf_rd_m=ATVPDKIKX0DER&pf_rd_r=1NEWTG6PBE2QGGHKFGTY *What Every Engineer Should Know About Risk Engineering and Management*, Marcel Dekker Inc, (2000)[4] Kleiner, Y., Rajani, B., Sadiq, R., *Failure Risk Management of Buried Infrastructure Using Fuzzy-based Techniques*, Journal of Water Supply Research and Technology: Aqua, Vol. 55, No. 2, pp. 81-94, March (2006)

[5] Saaty, T. L., Vargas, L. G. *Models*, *Methods*, *Concepts and Applications of the Analytic Hierarchy Process*, Kluwer Academic press, (2001)

[6] Douglas, H. *The Failure of Risk Management: Why It's Broken and How to Fix It*. John Wiley & Sons. p. 46 (2009)

[7] http://www.scotland.gov.uk/Publications/2008/11/24160623/3

[8] Carr, J. H. M., Tah, V.: *A Fuzzy Approach to Construction Project Risk Assessment and Analysis: Construction Project Risk Management System,* Advances in Engineering Software, Volume 32, Number 10, pp. 847-857(11) October (2001)

[9] Cameron, E., Peloso, G. F. *Risk Management and the Precautionary Principle*: A Fuzzy Logic Model, Risk Analysis, Vol. 25, No. 4, pp. 901-911, August (2005)

[10] Mikhailov, L., *Deriving Priorities from Fuzzy Pairwise Comparison Judgements*, Fuzzy Sets and Systems Vol. 134, pp. 365-385, (2003)

[11] Yasuyuki Sawada, *The Imapct of Natural and Manmade Disasters on Household Welfare*, http://www.fasid.or.jp/kaisai/080515/sawada.pdf

[12] Didadster Definition, http://www.wordiq.com/definition/Disaster

[13] Németh-Erdődi, K., *Risk Management and Loss Optimiyation at Design Process of Products,* Acta Polytechnica Hungarica, Vol. 5, No. 3 (2008)

[14] Takács, M., *Structured Risk Management Model with Fuzzy Background*, Proc. of FSTA 2010, Tenth International Conference On Fuzzy Set Theory And Applications, Liptovský Ján, Slovak Republic, February (2010)

The DEA – FUZZY ANP Department Ranking Model Applied in Iran Amirkabir University

Babak Daneshvar Rouyendegh, Serpil Erol

Department of Industrial Engineering, Atılım University
P.O. Box 06836, İncek, Ankara, Turkey
E-mail: babekd@atilim.edu.tr

Department of Industrial Engineering, Gazi University
P.O. Box 06570, Maltepe, Ankara, Turkey

Abstract: Proposed in this study is a hybrid model for supporting the department selection process within Iran Amirkabir University. This research is a two-stage model designed to fully rank the organizational departments where each department has multiple inputs and outputs. First, the department evaluation problem is formulated by Data Envelopment Analysis (DEA) and separately formulates each pair of units. In the second stage, the pair-wise evaluation matrix generated in the first stage is utilized to fully rank-scale the units via the Fuzzy Analytical Network Process (FANP). The FANP method adopted here uses triangular fuzzy numbers. ANP equipped with fuzzy logic helps in overcoming the impreciseness in the preferences. DEA-FANP ranking does not replace the DEA classification model; rather, it furthers the analysis by providing full ranking in the DEA context for all departments, efficient and inefficient.

Keywords: Data Envelopment Analysis (DEA); Fuzzy Analytical Network Process (FANP); Performance; Efficiency; Fully Rank

1 Introduction

Multi-attribute decision-making (MADM) ranks elements based on single or multiple criteria, where each criterion contributes positively to the overall evaluations. The decision maker often carries out the evaluations subjectively. However, DEA deals with classifying the units into two categories, efficient and inefficient, based on two sets of multiple outputs contributing positively to the overall evaluation [1].

Many researchers (Belton & Vickers, 1993) highlight the relationship between DEA and MCDM: "According to them, DEA utilizes a process of allocating weights to criteria, just like other approaches to multi criteria and analysis".

Ranking is very common in MCDM literature, especially when we need to describe lists of elements or alternatives with single or multiple criteria that we wish to evaluate, and then compare or select. Various approaches have been proposed in the literature for full- ranking of the element, ranging from the utility theory approach to AHP developed by Saaty [2], [3].

Throughout the process of reviewing the literature, it appeared that limited research has been carried out regarding DEA-FANP methods, and only the DEA-AHP method in which connections among factors are not considered has been addressed. The idea of combining AHP and DEA is not new, and there have been several attempts to use them in actual situations. Some of these examples include: Bowen [4], Shang and Sueyoshi [5], Zhang and Cui [6], Zilla Sinuany-Stern et al. [3], Taho Yang, Chunwei Kuo [7], Takamura and Tone [8], Saen et al. [9], Ramanathan [10], and Wang et al. [11].

This paper is divided into four sections. In Section 1, the studied problem is introduced. Section 2 briefly describes the DEA-FANP method and the stages of the proposed model and steps are determined in detail. How the proposed model is used in an example in the real world is explained in Section 3. Finally, in Section 4, conclusions and future research areas are discussed.

2 The DEA-FUZZY ANP Method

2.1 Fuzzy Sets and Fuzzy Number

Zadeh (1965) introduced the Fuzzy Set Theory to deal with the uncertainty due to imprecision and vagueness. A major contribution of this theory is its ability to represent vague data; it also allows mathematical operators and programming to be applied to the fuzzy domain. A fuzzy set is a class of objects with a continuum of grades of membership. Such a set is characterized by a membership (characteristic) function which assigns to each object a grade of membership ranging between zero and one [12].

A tilde '~' will be placed above a symbol if the symbol represents a fuzzy set. A triangular fuzzy number (TFN), $\tilde{M}$ is shown in Fig. 1. A TFN is denoted simply as (l/m, m/u) or (l,m,u). The parameters l, m and u ($l \leq m \leq u$) denote respectively the smallest possible value, the most promising value, and the largest possible value that describe a fuzzy event. The membership function of triangular fuzzy numbers is as follows:

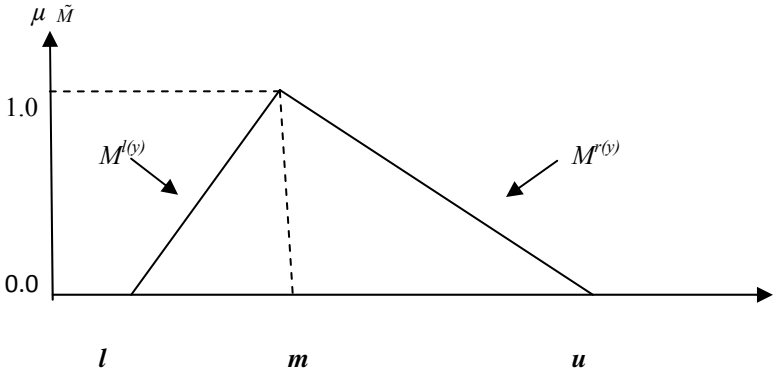

Figure 1
A triangular fuzzy number $\tilde{M}$

Each TFN has linear representations on its left and right side, such that its membership function can be defined as

$$\mu(x/M)=\begin{cases} 0, & x<l, \\ (x-l)/(m-l), & l\le x\le m, \\ (u-x)/(u-m), & m\le x\le u, \\ 0, & x>u. \end{cases} \tag{1}$$

A fuzzy number can always be given by its corresponding left and right representation of each degree of membership:

$$\tilde{M} = (M^{l(y)}, M^{r(y)} = (l+(m-l)y, u+(m-u)y), \qquad y\in[0,1] \tag{2}$$

where $l(y)$ and $r(y)$ denote the left side representation and the right side representation of a fuzzy number, respectively. Many ranking methods for fuzzy numbers have been developed in literature. These methods may provide different ranking results, and most of them are tedious in graphic manipulation, requiring complex mathematical calculation [13].

2.2 Fuzzy ANP

ANP, also introduced by Saaty, is a generalization of the analytic hierarchy process (AHP). Whereas AHP represents a framework with a uni-directional hierarchical AHP relationship, ANP allows for complex interrelationships among decision levels and attributes. The ANP feedback approach replaces hierarchies with networks in which the relationships between levels are not easily represented as higher or lower, dominant or subordinate, direct or indirect. For instance, not

only does the importance of the criteria determine the importance of the alternatives, as in a hierarchy, but also the importance of the alternatives may impact on the importance of the criteria [15].

ANP does not require this strictly hierarchical structure; it allows factors to `control' and be `controlled' by the varying levels or `clusters' of attributes. Some controlling factors are also present at the same level. This interdependency among factors and their levels is defined as a systems-with-feedback approach.

The ANP approach is capable of handing interdependent relationships among elements by obtaining composite weights through the development of a supermatrix. The supermatrix concept contains parallels to the Markov chain process [14-15], where relative importance weights are adjusted by forming a supermatrix from the eigenvectors of these relative importance weights. The weights are then adjusted by determining the products of the supermatrix.

The AHP method provides a structured framework for setting priorities on each level of the hierarchy using pair-wise comparisons that are quantified using a 1-9 scale, as demonstrated in Table 1. In contrast, the ANP method allows for more complex relationships among decision layers and their properties.

Table 1
The 1-9 scale for AHP [15]

Importance intensity	***Definition***	**Explanation**
1	Equal importance	Two activities contribute equally to the objective
3	Moderate importance of one over another	Experience and judgment slightly favor one over another
5	Strong importance of one over another	Experience and judgment strongly favor one over another
7	Very strong importance of one over another	Activity is strongly favored and its dominance is demonstrated in practice
9	Extreme importance of one over another	Importance of one over another affirmed on the highest possible order
2,4,6,8	Intermediate values	Used to represent compromise between the priorities listed above

The inability of ANP to deal with the impression and subjectiveness in the pair-wise comparison process has been improved in fuzzy ANP. Instead of a crisp value, fuzzy ANP applies a range of values to incorporate the decision maker's uncertainly [16]. In this method, the fuzzy conversion scale is as in Table 2. This scale will be used in the Mikhailov [17] fuzzy prioritization approach.

Table 2
The 1-9 Fuzzy conversion scale [17]

Importance intensity	*Triangular fuzzy scale*
1	(1,1,1)
2	(1.6,2.0,2.4)
3	(2.4,3.0,3.6)
4	(3.2,4.0,4.8)
5	(4.0,5.0,6.0)
6	(4.8,6.0,7,2)
7	(5.6,7.0,8.4)
8	(6.4,8.0,9.6)
9	(7.2,9.0,10.8)

2.3 Data Envelopment Analysis (DEA)

DEA has been successfully employed for assessing the relative performance of a set of firms, usually called decision-making units (DMU's), which use a variety of identical inputs. The concept of Frontier Analysis, suggested by Farrel (1957), forms the basis of DEA, but the recent series of discussions started with an article by Charnes et al. [18].

DEA is a method for mathematically comparing different DMUs' productivity based on multiple inputs and outputs. The ratio of weighted inputs and outputs produces a single measure of productivity called relative efficiency. The DMUs which have a ratio of 1 are referred to as 'efficient', given the required inputs and produced outputs. The units that have a ratio less than 1 are 'less efficient' relative to the most efficient units. Because the weights for the input and the output variables of DMUs are computed to maximize the ratio and are then compared to a similar ratio of the best-performing DMUs, the measured productivity is also referred to as 'relative efficiency'.

2.4 The Proposed DEA- Fuzzy ANP Method

In this study, fuzzy ANP and DEA for efficiency measurement have advantages over other fuzzy ANP approaches. The priorities obtained from the Fuzzy ANP method based on DEA are defined as a two-staged approach. In the first stage, the pair-wise comparison of the results obtained from the model is based on DEA; in the second stage, a whole hierarchy is carried out by the Fuzzy ANP method on the results obtained from the first stage. A schematic diagram of the proposed model for measurement is shown in Figure 2.

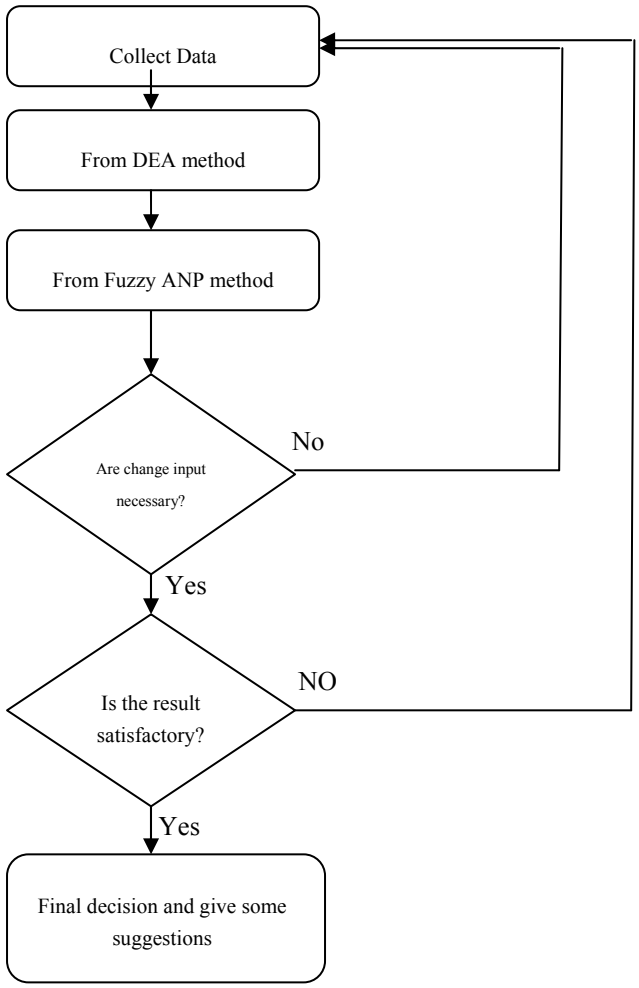

Figure 2
Schematic diagram of the proposed model for measurement

2.4.1 First Stage of the Method (DEA pair-wise comparisons)

In the Fuzzy ANP and DEA method, a pair-wise comparison in a decision-making unit is carried out. For instance, the DMUs are used for the production of x_{ij} (i=1, 2,…, m) entries and y_{rj} (r=1,2,...,s) outputs. X ($s \times n$) and Y ($m \times n$) are the amounts of the entries and outputs, respectively. In DEA, each unit is compared with all units, whereas in the DEA-Fuzzy ANP method, the DMUs are compared in a pair-wise method against each other.

Mathematical (Weighted Linear) Representation of the Problem:

$$e_{k,k'}=\max \sum_{r=1}^{s} u_r y \quad (3)$$

s.t:

$$\sum_{i=1}^{m} v_i x_k=1 \quad (4)$$

$$\sum_{r=1}^{s} u_r y_{rk}-\sum_{i=1}^{m} v_i x_{ik} \leq 0 \quad (5)$$

$$\sum_{r=1}^{s} u_r y_{rk}-\sum_{i=1}^{m} v_i x_{ik'} \leq 0 \quad (6)$$

$u_r \geq 0 \quad r=1,2,\ldots..,s \quad v_i \geq 0 \quad i=1,2,\ldots..,m$

By solving this mathematical model, $e_{k,k'}$ elements are solved and the pair-wise compared E matrix is obtained (k'=1,….,n, k=1,….,n and $k \neq k'$). In the second stage of the DEA- Fuzzy ANP method process, a two-level FANP model is given.

2.4.2 Second Stage of the Method (FANP ranking)

In the second level, based on the pair-wise comparison matrix E and after the hierarchy of FANP has been developed, the next stage creates matrices considering the interaction between pair-wise items for the factors and sub factors. We modify the selection process to a nine step method procedure, as follows:

Step 1. The calculation of $a_{k,k'}$: The components of the pair-wise comparative matrix are obtained via the following formula.

$$a_{k,k'}=e_{k,k'}/e_{k',k} \quad (7)$$

Step 2. The calculation of Triangular Fuzzy Numbers: we setup the Triangular Fuzzy Numbers. Each expert makes a pair-wise comparison of the decision criteria and gives them relative scores.

$$\hat{G}_1=(l_i, m_i, u_i) \quad (8)$$

Step 3. The calculation of $\hat{G}_1$: Establishing the Triangular Fuzzy Numbers, we setup the Triangular Fuzzy Numbers using the ANP method based on the Fuzzy numbers. Each expert makes a pair-wise comparison of the decision criteria and gives them relative scores.

$$\hat{G}_1=(l_i,m_i,u_i) \tag{9}$$

$$l_i=(l_{i1} \otimes l_{i2} \otimes \ldots \otimes l_{ik})^{1/k} \qquad i=1,2,\ldots,k \tag{10}$$

$$m_i=(m_{i1} \otimes m_{i2} \otimes \ldots.. \otimes m_{ik})^{1/k} \qquad i=1,2,\ldots,k \tag{11}$$

$$u_i=(u_{i1} \otimes u_{i2} \otimes .... \otimes u_{ik})^{1/k} \qquad i=1,2,\ldots,k \tag{12}$$

Step 4. The calculation of $\hat{G}_T$: Establishing the geometric fuzzy mean of the total row using:

$$\hat{G}_T=(\sum_{i=1}^{k} l_i,\sum_{i=1}^{k} m_i,\sum_{i=1}^{k} u_i) \tag{13}$$

~

Step 5. The calculation of *W:* Fuzzy geometric mean of the fuzzy priority value calculated with normalization priorities for factors using:

$$\tilde{W} = \hat{G}_i /\hat{G}_T=(l_i ,m_i ,u_i)/(\sum_{i=1}^{k} l_i, \sum_{i=1}^{k} m_i,\sum_{i=1}^{k} u_i)=(l_i /\sum_{i=1}^{k} u_i, m_i /\sum_{i=1}^{k} m_i , u_i /\sum_{i=1}^{k} l_i) \tag{14}$$

Step 6. The calculation of wi_{al}: Factors belonging to nine different *α-cut* values α for the calculated, fuzzy priorities are applied for lower and upper limits for each α value:

$$wi_{al}=(wil_{al} \text{ , } wiu_{\alpha}) \; i=1,2,\ldots,k \; l=1,2,\ldots,L \tag{15}$$

Step 7. The calculation of W_{il},W_{iu}: Combine the entire upper values and the lower values separately, then divide by the total sum of the α value:

$$W_{il} = \sum_{l=1}^{L} \alpha\,(w_{il})_l \; / \; \sum_{l=1}^{L} \alpha_l \qquad i= 1,2,\ldots,k \qquad l= 1,2,\ldots,L \tag{16}$$

$$W_{iu} = \sum_{l=1}^{L} \alpha\,(w_{iu})_l \; / \; \sum_{l=1}^{L} \alpha_l \qquad i= 1,2,\ldots,k \qquad l= 1,2,\ldots,L \tag{17}$$

Step 8. The calculation of W_{id}: Use the following formula in order to defuzzify by the Combined upper limit value and lower limit value using the optimism index (λ)

$$w_{id} = \lambda W_{iu} + (1-\lambda)W_{il} \qquad \lambda \in [0,1] \qquad i = 1,2,...,k \tag{18}$$

Step 9. The calculation of W_{in}: Normalization of defuzzification value priorites using

$$W_{in} = w_{id} / \sum_{i=1}^{k} w_{id} \qquad i= 1,\ldots,k \tag{19}$$

Step 10. The calculation of $w_k \times W_{in}$: The final step deals with determining the degree of relations among different units by multiplying the matrices,

$W_k =$

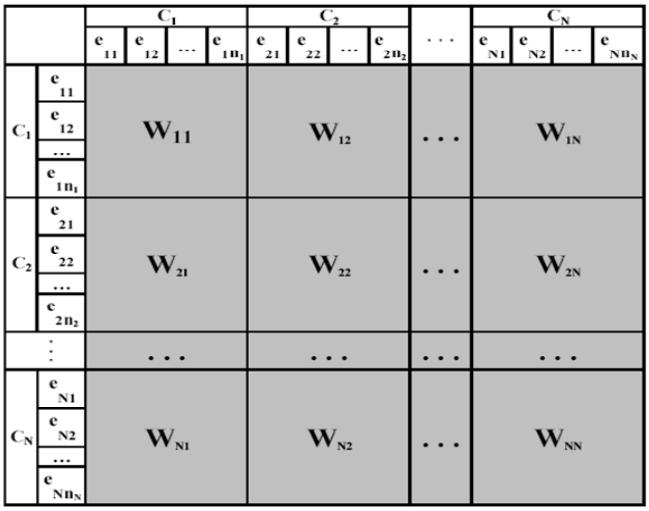

Figure 3
Relations among different units (super matrix) [15]

3 Applying the Sequential Methodology: An Illustrative Problem

The suggested hybrid model is demonstrated via an example of a selected department, supported by Iran Amirkabir University. Amirkabir University (Tehran Polytechnic) was established in 1958 as the first technical university of Iran. Through its rapid educational and research expansion, the university was able to gain a high ranking among all other universities and research centers. The achievements of this university in the area of research are evident from the many publications and the national and international prizes awarded for research activities. Thirteen departments have been considered in our evaluation. In our study, we employ a six-input evaluation criteria and four-output evaluation criteria:***Inputs:*** Number of Professor Doctors, Associated Professors, Assistant Professors, and Instructors; Budget of departments; and Number of credits.

Outputs: Number of alumni (undergraduates and graduate students), Evaluation of instructors, Number of academic congeries, and Number of academic papers (SCI-SSCI-AHCI).

Table 3
The DEA-Fuzzy ANP fully-ranking score

DMU	DEA-FANP score
D1	1.12449
D2	0.67602
D3	1.36825
D4	2.25443
D5	**2.82427**
D6	0.56335
D7	1.01403
D8	0.78684
D9	0.89915
D10	0.56435
D11	**0.55231**
D12	0.67926
D13	1.23937

The result score is always the-bigger-the-better. As visible in Table 3, department 5 has the largest score due to its highest efficiency and performance. Department 11 has the smallest score of the thirteen departments and is ranked in the last place. The relevant results can be seen in Table 3. Obviously, the best selection is candidate D5.

Conclusion

We have presented an effective model for rank scaling of the units with multiple inputs and multiple outputs using both DEA and FANP. In this paper, a two-stage hybrid methodology is provided where the binary comparison of the results obtained from the model is based on DEA. The second stage of the methodology assists in fully-ranking of the alternatives based on the results obtained from the first stage. The result of the methodology is a rank order of the alternatives, which can be used to select an individual project or a portfolio of projects.

The advantage of the DEA-FANP ranking model is that FANP pair-wise comparisons have been derived mathematically from multiple input/output data by running pair-wise DEA runs. Thus, there is no subjective evaluation.

The DEA and the FANP methods are commonly used in practice and, yet, both have limitations. The DEA-FANP method combines the best of both models by avoiding the pitfalls of each. ANP is designed for subjective evaluation of a set of alternatives based on multiple criteria organized in a hierarchical structure. In this model, we work with given tangible inputs and outputs of units, and no subjective assessment of the decision maker's evaluation is involved. The Pareto optimum limitation of DEA is resolved by the full- ranking performed here by means of the FANP. It is important to note that DEA-FANP does not replace DEA, but rather, it provides further analysis of DEA to full ranking the units.

The performance measurement model developed here structured the performance measurement problem in a hierarchical form, critical areas and performance measures. The developed performance measurement model contributes to the previous performance measurement models by including and quantifying interdependencies that exist between system components. In addition, the involvement of fuzzy theory can adequately resolve the inherent uncertainty and imprecision associated with the mapping of a decision maker's perception to exact numbers.

References

[1] J. A. Ganley, J. S. Cubbin, "Public Sector Efficiency Measurement: Applications of Data Envelopment Analysis", Elsevier Science Publishers, 1992

[2] T. L. Saaty, "The Analytic Hierarchy Process", McGraw-Hill, New York, 1980

[3] Z. Sinuany-Stern, A. Mehrez, Y.Hadad, "An AHP/DEA Methodology for Ranking Decision Making Units", International Transactions in Operational Research 7(2), pp. 109-124, 2000

[4] W. M. Bowen, "Subjective Judgments and Data Envelopment Analysis in Site Selection Computers", En. Urb. Sys. 14, pp. 133-144, 1990

[5] J. Shang, T. Sueyoshi, "A United Framework for the Selection of a Flexible Manufacturing System", European Journal of Operation Research 85(2), pp. 297-315, 1995

[6] X. S. Zhang, J. C. Cui, "A Project Evaluation System in the State Economic Information System of China: An Operations Research Practice in Public Sectors", International Transactions in Operational Research 6(5), pp. 441-452, 1999

[7] T. Yang, C. Kuo, "A Hierarchical AHP/DEA Methodology for the Facilities Layout Design Problem", European Journal of Operation Research 147(1), pp. 128-136, 2003

[8] Y. Takamura, K. Tone, "A Comparative Site Evaluation Study for Relocating Japanese Government Agencies out of Tokyo", Journal of Socio Economic Planning 37(2), pp. 85-102, 2003

[9] R. F. Saen, A. Memariani, F. H. Lot, "Determining Relative Efficiency of Slightly Non-Homogeneous Decision-Making Units by Data Envelopment Analysis: A Case Study in IROST", Journal of Applied Mathematics and. Computation 165(2), pp. 313-328, 2005

[10] R. Ramanathan, "Data Envelopment Analysis for Weight Derivation and Aggregation in the Analytic Hierarchy Process", Journal of Computer and Operation Research 33(5), pp. 1289-1307, 2006

[11] Y. M. Wang, J. Liu, T. M. S. Elhag, "An Integrated AHP-DEA Methodology for Bridge Risk Assessment", Journal of Computer and Industrial Engineering, pp. 1-13, 2007

[12] C. Kahraman, D. Ruan, İ. Doğan, "Fuzzy Group Decision-Making for Facility Location Selection", Information Sciences 157, pp. 135-15, 2003

[13] C. Kahraman, D. Ruan, E. Tolga, "Capital Budgeting Techniques using Discounted Fuzzy versus Probabilistic Cash Flows", Information Sciences 42, pp. 57-76, 2002

[14] T. L Saaty, Vargas, L. G, "Decision Making with the Analytic Network Process", Springer Science, LLC, pp. 1-23, 2006

[15] T. L Saaty, "Decision Making with Dependence and Feedback, The Analytic Network Process", RWS Publications, Pittsburgh, 1996

[16] R. Kuswandari, "Assessment of Different Methods for Measuring the Sustainability of Forest Management", International Institute for Geo-Information Science and E.O.E, Netherlands, 2004

[17] L. Mikhailov, "Deriving Priorities from Fuzzy Pair Wise Comparison Judgments", Fuzzy Sets and Systems 134, pp. 365-385, 2003

[18] A. Charnes, W. W. Cooper, E. Rhodes, "Measuring the Efficiency of Decision-Making Units", European Journal of Operational Research 2, pp. 429-44, 1978

Combining Fuzzy/Wavelet Adaptive Error Tracking Control Design

Martin Kratmüller

SIEMENS PSE sro Slovakia
Dúbravská cesta 4, 845 37 Bratislava, Slovak Republic
E-mail: martin.kratmueller@siemens.com

Abstract: A combining adaptive fuzzy-wavelet control algorithm is proposed for a class of continuous time unknown nonlinear systems. An application of wavelet networks to control problems of nonlinear systems is investigated in this work. A wavelet network is constructed as an alternative to a neural network to approximate a nonlinear system. Based on this wavelet network and fuzzy approximation, suitable adaptive control laws and appropriate parameter update algorithms for nonlinear uncertain (or unknown) systems are developed to achieve tracking performance. The stability analysis for the proposed control algorithm is provided. A nonlinear system simulation example is presented to verify the effectiveness of the proposed method.

Keywords: fuzzy control; adaptive control; wavelet approximation; feedback linearization

1 Introduction

In recent years, wavelet neural networks which combine the learning ability of feed forward neural networks and time-frequency localization properties of wavelets have become a popular tool for multiscale analysis and synthesis, time-frequency signal analysis in signal processing, function approximation, approximation in solving partial differential equations, and so on [1]-[8].

At present, there are two kinds of wavelet neural network structures. The first one is the fixed wavelet basis, where the dilation and translation parameters of wavelet basis are fixed, and the output layer weights are adjustable. The second one is the variable wavelet basis. The dilation parameters, translation parameters, and the output layer weights are adjustable in this type of wavelet neural network.

On the other hand, considerable study has been performed to integrate the excellent learning capability of neural networks with the perfect inference mechanism of fuzzy systems, which are called neuro-fuzzy systems [9], to obtain the rule-base membership function parameters from the input-output data. These

neuro-fuzzy systems have fast and accurate learning and good generalization capabilities, and both have the ability to accommodate expert knowledge about the problem under consideration.

Fuzzy logic controllers are generally considered applicable to plants that are mathematically poorly understood and where experienced human operators are available. However, fuzzy controllers have not been regarded as an exact science due to the lack of a guarantee of global stability and acceptable performance. Nonetheless, some researchers propose the stability analysis of fuzzy control systems (e.g., [10]). The mathematical model of the plant is assumed to be known in [10]. Hence, this contradicts the very fundamental premise of fuzzy control systems. In fact, if the model of plant is known, then we should give the conventional linear or nonlinear control methods high priority.

The proposed control scheme provides good transient and robust performance. In this paper, it is proved that the closed-loop system is globally stable in the Lyapunov sense and the system output asymptotically stable with modeling uncertainties and disturbances.

Fuzzy controllers are assumed to work in situations where the plant parameters and structures have some uncertainties or unknown variations. The basic objective of adaptive control is to maintain the consistent performance of a system in the presence of uncertainties. So, advanced fuzzy control or wavelet approximation might be adaptive. This work is involved by combining the characteristics of wavelet, the technique of feedback linearizations, the adaptive control scheme and the fuzzy control to solve the tracking control design problem for nonlinear systems with bounded unknown or uncertain parameters and external disturbances.

This paper is organized as follows. First, the problem formulation is presented in Section 2. A brief description of a wavelet system is included in Section 3. In Section 4, the adaptive fuzzy-wavelet control is proposed. Simulation results for the proposed control concept are shown in Section 5. Finally, the paper is concluded in Section 6.

2 Problem Formulation

Consider an nth order SISO nonlinear system with $n \geq 2$ of the following form

$$\begin{aligned} \dot{x}_1 &= x_2 \\ &\vdots \\ \dot{x}_n &= f(\underline{x}) + g(\underline{x})u \\ y &= x_1 \end{aligned} \tag{1}$$

where $\underline{x} = [x, \dot{x}, \cdots, x^{(n-1)}]^T = [x_1, x_2, \cdots, x_n]^T \in R^n$ is the state vector, u is the control input and y is the output of the system. All the elements of the state vector $\underline{x}$ are assumed to be available. At the beginning, $f(\underline{x})$ is assumed to be smooth and $g(\underline{x})$ is assumed to be smooth and bounded away from zero. Differentiating the output y with respect to time for n times we obtain the following input/output form

$$y^{(n)} = f(\underline{x}) + g(\underline{x})u \tag{2}$$

Note that the above system has a relative degree of n.

If $f(\underline{x})$ and $g(\underline{x})$ are known, a nonlinear tracking control can be obtained. Let y_r be the desired continuous differentiable uniformly bounded trajectory and let

$$\underline{e} = y - y_r = (e, \dot{e}, \cdots, e^{(n-1)})^T \in R^n \tag{3}$$

be the tracking error. Then by employing the technique of feedback linearization a suitable control law can be derived to achieve the tracking control goal as

$$u = \frac{1}{g(\underline{x})}\left[-f(\underline{x}) + u_p + \nu\right] \tag{4}$$

where u_p is an auxiliary control variable yet to be specified and

$$\nu = y_r^{(n)} + \alpha_1\left(y_r^{(n-1)} - y^{(n-1)}\right) + \cdots + \alpha_n\left(y_r - y\right) \tag{5}$$

Note that the coefficients $\alpha_1, \ldots, \alpha_n$ are positive constants to be assigned such that the polynomial $s^n + \alpha_1 s^{n-1} + \ldots + \alpha_n$ is Hurwitz. As a result, the error dynamic of the system has the following input/output form

$$e^{(n)} + \alpha_1 e^{(n-1)} + \ldots + \alpha_n e = u_p \tag{6}$$

which can be represented in state space form as

$$\dot{\underline{e}} = \mathbf{A}\underline{e} + \mathbf{B}u_p \tag{7}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\alpha_n & -\alpha_{n-1} & -\alpha_{n-2} & \cdots & -\alpha_1 \end{bmatrix} \tag{8}$$

$$\mathbf{B} = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}^T \tag{9}$$

$$\underline{e} = \begin{bmatrix} e & \cdots & e^{(n-2)} & e^{(n-1)} \end{bmatrix}^T \tag{10}$$

Note that the above design method is useful only if $f(\underline{x})$ and $g(\underline{x})$ are known exactly. If $f(\underline{x})$ and $g(\underline{x})$ are unknown then adaptive strategies must be employed. Let us now discuss a wavelet-network based adaptive algorithm.

First we employ two wavelet networks

$$\hat{f}(\underline{x}, \underline{\theta}_f) = \underline{\theta}_f^T W_f\left(\underline{c}_f^T \underline{x}\right) \tag{11}$$

$$\hat{g}(\underline{x}, \underline{\theta}_g) = \underline{\theta}_g^T W_g\left(\underline{c}_g^T \underline{x}\right) \tag{12}$$

to approximate (or model) the nonlinear functions $f(\underline{x})$ and $g(\underline{x})$ of the system, respectively.

3 A Review of Wavelet Networks

In this section a brief introduction to wavelet networks is given. Several kinds of wavelet bases have successfully been developed and widely applied in many different areas, such as in time-frequency signal analysis in signal processing, function approximation, approximation in solving partial differential equations and so on. Further development of new families of wavelet bases continues to receive considerable attention from researchers.

Consider the closed space U_i, $\forall i \in Z$ with the following properties [11]

$$U_i \cdots \subset U_{-1} \subset U_0 \subset U_1 \cdots \tag{13}$$

$$\cap_{i\in Z} U_i = \{0\} \tag{14}$$

$$U_{i+1} = U_i \oplus W_i \quad \forall i \in Z \tag{15}$$

$$f(\underline{x}) \in U_i \Leftrightarrow f(2\underline{x}) \in U_{i+1} \quad \forall i \in Z \tag{16}$$

where Z is the set of all integers, $\cap$ is the intersection operator and $\oplus$ is the direct sum. It is seen that the decomposition of the whole space S can be rewritten as follows

$$S = U_i \oplus W_i \oplus W_{i+1} \oplus \cdots \oplus W_0 \oplus W_1 \oplus \cdots \tag{17}$$

for some $i \in Z$. Let $\phi(\underline{x}) \in S$ be a basic scaling function such that $U_i = \text{span}\{\phi_{ij}(\underline{x})\}$ with $\phi_{ij}(\underline{x}) = 2^{\frac{i}{2}}\phi(2^i\underline{x} - j)$, for all $i, j \in Z$; then, there exists a basic function $\psi(\underline{x}) \in S$ such that $W_i = \text{span}\{\psi_{ij}(\underline{x})\}$ with $\psi_{ij}(\underline{x}) = 2^{\frac{i}{2}}\psi(2^i\underline{x} - j)$, for all $i, j \in Z$.

Now, consider a function $f(\underline{x})$ is S. It is obvious that $f(\underline{x})$ can be rewritten as [11], [12]

$$f(\underline{x}) = \sum_i \sum_j \theta_{ij}\psi_{ij}(\underline{x}) \tag{18}$$

where

$$\theta_{ij} = \int_{-\infty}^{\infty} f(x)\psi_{ij}(x)dx \tag{19}$$

with $\psi_{ij}(\underline{x}) = 2^{\frac{i}{2}}\psi(2^i\underline{x} - j)$, for all $i, j \in Z$. The above expression of $f(\underline{x})$ is called a wavelet series expansion of the function $f(\underline{x})$.

Based on the wavelet series expansion, a wavelet network of the form [13], [14]

$$\hat{f}(\underline{x}, \underline{\theta}) = \sum_{i=M_1}^{M_2} \sum_{j=N_1}^{N_2} \theta_{ij}\psi_{ij}(\underline{x}) = \underline{\theta}^T W(\underline{x}) \tag{20}$$

can be constructed to approximate a nonlinear function $f(\underline{x})$ in space S, for some integers M_1, M_2, N_1 and N_2 where

$$\underline{\theta} = \left[\theta_{M_1N_1} \cdots \theta_{M_1N_2} \cdots \theta_{M_2N_1} \cdots \theta_{M_2N_2} \right]^T \tag{21}$$

and

$$W(x) = \left[\psi_{M_1N_1}(x) \ldots \psi_{M_1N_2}(x) \ldots \psi_{M_2N_1}(x) \ldots \psi_{M_2N_2}(x) \right]^T \tag{22}$$

This wavelet network represents an alternative to a neural network approximation.

If $\epsilon(M_1, M_2, N_1, N_2) = f(x) - \hat{f}(\underline{x}, \underline{\theta})$ is the approximation error, then for arbitrary constant $\varepsilon \geq 0$ there exist some constants $M_1, M_2, N_1, N_2 \in Z$ such that $\left\| \epsilon(M_1, M_2, N_1, N_2) \right\|_2 \leq \varepsilon$, for all c in compact set $X \subset R$. This means that the wavelet network $\hat{f}(\underline{x}, \underline{\theta})$ can approximate $f(\underline{x})$ to any desired accuracy.

In the case of a function $f(\underline{x})$ defined on $X \subset R^n$ with $\underline{x} = [x_1, x_2, \cdots, x_n]^T$, the proposed wavelet network $\hat{f}(\underline{x}, \underline{\theta})$ cannot be applied directly because $\hat{f}(\underline{x}, \underline{\theta})$ is defined on $X \subset R$, not on $X \subset R^n$. We must first make a minor modification by replacing the wavelet bases in Eq. (20) by $\psi_{ij}(\underline{c}^T, \underline{x}) = \psi_{ij}\left(\sum_{i=1}^{n} c_i x_i \right)$ with some weighting constants c_i.

Then the modified wavelet network becomes

$$\hat{f}(\underline{x}, \underline{\theta}) = \sum_{i=M_1}^{M_2} \sum_{j=N_1}^{N_2} \theta_{ij} \psi_{ij}(\underline{c}^T \underline{x}) = \underline{\theta}^T W(\underline{c}^T \underline{x}) \tag{23}$$

Note that this modified wavelet network is composed of four layers. The first layer is the input layer with available input vector $\underline{x} = [x_1, x_2, \cdots, x_n]^T$. A weighting summer $\underline{c}^T \underline{x}$ is given in the second layer. The third layer is composed of the wavelet bases. The output layer is a weighted combination of the wavelets.

4 Adaptive Fuzzy/Wavelet Control

According to the description in Section 3, guaranteeing $\underline{x}$ in a compact region is very important when the wavelet networks $\hat{f}\left(\underline{x},\underline{\theta}_f\right)$ and $\hat{g}\left(\underline{x},\underline{\theta}_g\right)$ are used to approximate $f\left(\underline{x}\right)$ and $g\left(\underline{x}\right)$, respectively. In general there is still not an efficient way to ensure satisfaction of this requirement. In practical applications one may assign a very large compact set to avoid violation of this requirement. However, a very large wavelet basis is needed in this situation. This may result in a large computational burden. Fortunately, in many physical systems such as mechanical systems and electrical systems, an appropriate selection of the pre-assigned compact set can be obtained via knowledge of some physical limitations.

Let

$$\underline{\theta}_f^* = \arg\min_{\underline{\theta}_f}\max_{\underline{x}}\left|\hat{f}\left(\underline{x},\underline{\theta}_f\right)-f\left(\underline{x}\right)\right| \tag{24}$$

$$\underline{\theta}_g^* = \arg\min_{\underline{\theta}_g}\max_{\underline{x}}\left|\hat{g}\left(\underline{x},\underline{\theta}_g\right)-g\left(\underline{x}\right)\right| \tag{25}$$

be the best approximation parameters of $\underline{\theta}_f$ and $\underline{\theta}_g$, respectively.

System (1) can be rewritten as

$$x_1^{(n)} = f\left(x_1,\ldots,x_n\right)+g\left(x_1,\ldots,x_n\right)u \tag{26}$$

where $\underline{x}=[x,\dot{x},\cdots,x^{(n-1)}]^T=[x_1,x_2,\cdots,x_n]^T\in R^n$ is the state vector and the functions $f\left(\underline{x}\right)$ and $g\left(\underline{x}\right)$ are unknown nonlinear functions of the states and time. The objective of the adaptive wavelet error tracking control design is to update the controller parameters in such as a way that the system output can asymptotically track the desired reference model output $y_r = x_m\left(t\right)$ in spite of function uncertainties.

The reference model is a linear system in form

$$x_{m_1}^{(n)} + a_{n-1}x_{m_1}^{(n-1)} + \cdots + a_1\dot{x}_{m_1} + a_0 x_{m_1} = br \tag{27}$$

where $\underline{x}_m=[x_{m_1},\dot{x}_{m_1},\cdots,x_{m_1}^{(n-1)}]^T=[x_{m_1},x_{m_2},\cdots,x_{m_n}]^T\in R^n$ is the state vector of the reference model.

To follow the reference model, the controller must be chosen so as to cancel the nonlinearities in the nonlinear system and provide pole placement to the system, i.e. feedback linearization. For example, the controller is chosen in the form

$$u = \frac{1}{\hat{g}(\underline{x})}\left[-\hat{f}(\underline{x}) - \hat{a}_{n-1}x_1^{(n-1)} - \cdots - \hat{a}_1\dot{x}_1 - \hat{a}_0 x_1 + \hat{b}r\right] \tag{28}$$

In this article the set of fuzzy systems is used with a singleton fuzzifier, product inference, a centroid defuzzifier, a triangular antecendent membership function and a singleton consequent membership function with n inputs of $x_i \in \left[c_{x_i} - k_{x_i}, c_{x_i} + k_{x_i}\right]$ for $i = 1,\ldots,n$ and $\bar{u} \in [0,1]$ as the normalized output. The generalized expression of the class of the fuzzy controllers can be written as

$$\bar{u} = \sum_{i_1=1}^{2}\cdots\sum_{i_n=1}^{2} N_{i_1\cdots i_n} x_1^{i_1-1}\cdots x_n^{i_n-1} \tag{29}$$

$$N_{i_1\cdots i_n} = \frac{\left[\sum_{j_1=1}^{2}\cdots\sum_{j_n=1}^{2} R_{j_1\cdots j_n} K_{j_1\cdots j_n} C_{j_1\cdots j_n}\right]}{2^n\prod_{i=1}^{n} k_{x_i}} \tag{30}$$

$$C_{j_1\cdots j_n} = \left[\frac{(-1)^{j_1}}{k_{x_1} - (-1)^{j_1} c_{x_1}}\right]^{i_1-1}\cdots\left[\frac{(-1)^{j_n}}{k_{x_n} - (-1)^{j_n} c_{x_n}}\right]^{i_1-1} \tag{31}$$

$$K_{j_1\cdots j_n} = \left[k_{x_1} - (-1)^{j_1} c_{x_1}\right]\cdots\left[k_{x_n} - (-1)^{j_n} c_{x_n}\right] \tag{32}$$

On the other hand, given the coefficients of the explicit form $N_{i_1\cdots i_n}$ we can reconstruct the rule base from the generalized expression of the class of fuzzy systems [15] by using the following theorem.

Theorem 1 [15]: For a class of fuzzy logic systems (FLS) with a singleton fuzzifier, product inference, a centroid defuzzifier, a triangular antecendent membership function and a singleton consequent membership function, i.e. given the coefficients of the explicit form, i.e. $N_{i_1\cdots i_n}$, the control function can be expressed in terms of fuzzy rules as

$$R_{j_1\cdots j_n} = \sum_{i_1=1}^{2}\cdots\sum_{i_n=1}^{2} N_{i_1\cdots i_n} D_{j_1\cdots j_n} \tag{33}$$

with

$$D_{j_1\cdots j_n} = \left[c_{x_1} + (-1)^{j_1} k_{x_1}\right]^{i_1-1}\cdots\left[c_{x_n} + (-1)^{j_n} k_{x_n}\right]^{i_n-1} \tag{34}$$

Proof: The proof is found by directly expanding terms and comparing coefficients. For details, please refer to [15].

Therefore, one can express an equation in the form of generalized multilinear equations, such as polynomials, exactly as a rule base of FLS. Theorem 1 is useful in cases where the implementation of an FLS performs inference on a given fuzzy rule base but without any numerical computation capability.

We can express the fuzzy controller in the form of fuzzy IF-THEN rules.

$$\text{RULE i: IF r is } A_1^r \text{ and ... and } x_n \text{ is } A_1^{x_n}\text{, THEN } \bar{u}_p = R_i$$

The generalized expression of the class of fuzzy controller with n+1 inputs, i.e. r and $\underline{x}$ can be written as

$$\bar{u}_p = \sum_{i_0=1}^{2} \cdots \sum_{i_n=1}^{2} N_{i_0 i_1 \cdots i_n} r^{i_0-1} x_1^{i_1-1} \cdots x_n^{i_n-1} \tag{35}$$

By applying Theorem 1, one can find a set of R_i's to represent exactly the given pole-placement equation as $u_p = -\hat{a}_{n-1} x_1^{(n-1)} - \cdots - \hat{a}_1 \dot{x}_1 - \hat{a}_0 x_1 + \hat{b} r$.

The controller for pole-placement can be written as

$$u_p = \underline{\theta}_p^T \underline{\omega}_p \tag{36}$$

with $\underline{\theta}_p^T = \left(k_0, \underline{k}_b^T, \underline{k}_c^T\right)$

and $\underline{\omega}_p^T = \left(r, \underline{x}^T, \underline{x}_c^T\right)$

with

$$k_0 = 2N_{211\ldots111}$$

$$k_1 = 2N_{121\ldots111}$$

$$k_{n-1} = 2N_{111\ldots121}$$

$$k_n = 2N_{111\ldots112}$$

where $\underline{k}_b = \left[k_1, \cdots, k_n\right]^T$. The composite state vector $\underline{x}_c$ and the associated parameter vector $\underline{k}_c$ are defined as

$$\underline{x}_c^T = \left(rx_1x_2\ldots x_n, rx_1x_2\ldots x_{n-1}, \cdots, x_{n-1}x_n, 1\right) \tag{37}$$

$$\underline{k}_c^T = \left(k_{n+1}, k_{n+2}, \cdots, k_{n+n_c-1}, k_{n+n_c}\right) \tag{38}$$

with

$$k_{n+1} = 2N_{222\ldots222}$$

$$k_{n+2} = 2N_{222\ldots221}$$

$$k_{n+n_c-1} = 2N_{111\ldots122}$$

$$k_{n+n_c} = 2N_{111\ldots111}$$

where $n_c = 2^{n+1} - (n+1)$

Controller can be stated as

$$u = \frac{1}{\hat{g}(\underline{x})}\left[u_p - \hat{f}(\underline{x})\right] \tag{39}$$

From the nonlinear system (26) we have

$$\begin{aligned} x_1^{(n)} &= f(\underline{x}) + g(\underline{x})u \\ &= f(\underline{x}) + g(\underline{x})u - \hat{g}(\underline{x})u + \hat{g}(\underline{x})u \\ &= f(\underline{x}) + \hat{g}(\underline{x})u + \left(g(\underline{x}) - \hat{g}(\underline{x})\right)u \end{aligned} \tag{40}$$

By substituting (39) into the previous equation it becomes

$$x_1^{(n)} = \underline{k}_b^T \underline{x} + k_0 r + \underline{k}_c^T + \left(f(\underline{x}) - \hat{f}(\underline{x})\right) + \left(g(\underline{x}) - \hat{g}(\underline{x})\right)u \tag{41}$$

By substracting the closed-loop plant dynamic equation (above) with the reference model dynamic (27) we have the following

$$
\begin{aligned}
x_1^{(n)} - x_{m_1}^{(n)} &= \underline{k}_b^T \underline{x} + k_0 r + \underline{k}_c^T \underline{x}_c \\
&+ \left(f(\underline{x}) - \hat{f}(\underline{x}) \right) + \left(g(\underline{x}) - \hat{g}(\underline{x}) \right) u \\
&+ \sum_{j=1}^{n-1} a_j x_{m_1}^{(j)} - br \\
&= -\sum_{j=1}^{n-1} \left[a_j \left(x_1^{(j)} - x_{m_1}^{(j)} \right) \right] \\
&+ \sum_{j=1}^{n-1} \left[\left(k_j + a_j \right) x_1^{(j)} \right] \\
&+ \left(k_0 - b \right) r + \underline{k}_c^T \underline{x}_c \\
&+ \left(f(\underline{x}) - \hat{f}(\underline{x}) \right) + \left(g(\underline{x}) - \hat{g}(\underline{x}) \right) u
\end{aligned}
\tag{42}
$$

For the time derivative of the signal error vector $\underline{e} = \underline{x} - \underline{x}_m$ the following equality holds

$$
\begin{aligned}
e_1^{(n)} &= -\sum_{j=1}^{n-1} \left[a_j e_1^{(j)} \right] + \sum_{j=1}^{n-1} \left[\left(k_j + a_j \right) x_1^{(j)} \right] \\
&+ \left(k_0 - b \right) r + \underline{k}_c^T \underline{x}_c + \left(f(\underline{x}) - \hat{f}(\underline{x}) \right) \\
&+ \left(g(\underline{x}) - \hat{g}(\underline{x}) \right) u
\end{aligned}
\tag{43}
$$

We can rewrite the error (43) in matrix representation

$$
\dot{\underline{e}} = \mathbf{A}_m \underline{e} + \underline{b}_I \underline{\phi}^T \underline{\omega} \tag{44}
$$

The error vector $\underline{e}$ is defined as

$$
\underline{e} = \begin{pmatrix} e_1 \\ \dot{e}_1 \\ \vdots \\ e_1^{(n-1)} \end{pmatrix} = \begin{pmatrix} x_1 \\ \dot{x}_1 \\ \vdots \\ x_1^{(n-1)} \end{pmatrix} - \begin{pmatrix} x_{m_1} \\ \dot{x}_{m_1} \\ \vdots \\ x_{m_1}^{(n-1)} \end{pmatrix} \tag{45}
$$

The matrix $\mathbf{A}_m$ and vector $\underline{b}_I$ are defined as

$$\mathbf{A}_m = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_1 & -a_2 & -a_3 & \cdots & -a_n \end{pmatrix} \tag{46}$$

$$\underline{b}_I = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{47}$$

with the parameter error vector $\underline{\phi}$ defined as

$$\underline{\phi}^T = \begin{Bmatrix} k_0 - b & k_1 + a_1 & \cdots & k_n + a_n \\ k_{n+1} & k_{n+2} & \cdots & k_{n+n_c} \\ \left(\underline{\theta}_f^* - \underline{\theta}_f\right)^T & \left(\underline{\theta}_g^* - \underline{\theta}_g\right)^T & & \end{Bmatrix} \tag{48}$$

$$\underline{\omega}^T = \begin{Bmatrix} r & x_1 & \cdots & x_n \\ r x_1 x_2 \ldots x_n & r x_1 x_2 \ldots x_{n-1} & \cdots & 1 \\ W_f\left(\underline{c}_f^T \underline{x}\right) & W_g\left(\underline{c}_g^T \underline{x}\right) & & \end{Bmatrix} \tag{49}$$

where $\underline{\theta}_f^* W_f\left(\underline{c}_f^T \underline{x}\right) \approx f\left(\underline{x}\right)$ and $\underline{\theta}_g^* W_g\left(\underline{c}_g^T \underline{x}\right) \approx g\left(\underline{x}\right)$. The system's error (44) consists of a linear part governed by $\mathbf{A}_m$ and $\underline{b}_I$ plus a nonlinear control $\underline{\phi}^T \underline{\omega}$. In the following we show stable adaptive laws for the system.

Theorem 2: Consider the error equation given by (43) whose parameters are adjusted according to the following adaptive laws.

1) For the nonlinear-cancellation for $f\left(\underline{x}\right)$ the adaptive law is

$$\dot{\underline{\theta}}_f = -\gamma\left(\underline{p}^T \underline{e}\right) W_f\left(\underline{c}_f^T \underline{x}\right)$$

2) For the nonlinear-cancellation for $g\left(\underline{x}\right)$ the adaptive law is

$$\dot{\underline{\theta}}_g = -\gamma\left(\underline{p}^T \underline{e}\right) W_g\left(\underline{c}_g^T \underline{x}\right)$$

Then we have

1) $\underline{e}$ and $\underline{\phi}$ are uniformly bounded

2) $\lim_{t\to\infty} \underline{e} = 0$

where $\underline{p}$ is a vector consisting of the n-th column of positive definite symmetric matrix **P** (see Eq. 45).

Proof: The choice of the Lyapunov function is normally a quadratic function of both the signal error vector $\underline{e}$ and the parameter error $\underline{\phi}$

$$V = \underline{e}^T \mathbf{P} \underline{e} + \underline{\phi}^T \mathbf{\Gamma}^{-1} \underline{\phi} \tag{50}$$

with the adaptation gain matrix defined as $\mathbf{\Gamma} = \gamma \mathbf{I}_{2^{n+1}\times 2^{n+1}}$, where $\mathbf{I}_{2^{n+1}\times 2^{n+1}}$ is a $2^{n+1} \times 2^{n+1}$ identity matrix. Since $\mathbf{\Gamma}$ is positive definite, $\mathbf{\Gamma}^{-1}$ is also positive definite. Matrix **P** must be chosen as a positive definite symmetric matrix and it will follow from the adaptive law derivation shown in the following. To obtain an asymptotically stable adaptive system, $\dot{V}$ must be negative definite. Differentiating V yields with

$$\dot{V} = \underline{e}^T \left(\mathbf{A}_m^T \mathbf{P} + \mathbf{P} \mathbf{A}_m \right) \underline{e} + 2\underline{e}^T \mathbf{P} \underline{b}_I \underline{\phi}^T \underline{\omega} + 2\underline{\phi}^T \mathbf{\Gamma}^{-1} \dot{\underline{\phi}} \tag{51}$$

By applying the second method of Lyapunov, positive definite symmetric matrices **P** and **Q** can be found such that the first part of the equation satisfies

$$\underline{e}^T \left(\mathbf{A}_m^T \mathbf{P} + \mathbf{P} \mathbf{A}_m \right) \underline{e} = -\underline{e}^T \mathbf{Q} \underline{e} \tag{52}$$

By putting the last two terms of the equation to zero the adaptive laws emerges

$$\begin{aligned} 2\underline{e}^T \mathbf{P} \underline{b}_I \underline{\phi}^T \underline{\omega} + 2\underline{\phi}^T \mathbf{\Gamma}^{-1} \dot{\underline{\phi}} &= 0 \\ \dot{\underline{\phi}} &= -\mathbf{\Gamma} \underline{e}^T \mathbf{P} \underline{b}_I \underline{\omega} \\ &= -\mathbf{\Gamma} \left(\underline{p}^T \underline{e} \right) \underline{\omega} \end{aligned} \tag{53}$$

The product $\mathbf{P}\underline{b}_I$ is a vector consisting of the n-th column p of **P**, while the model and process parameters are assumed constant. From the definition of $\underline{\phi}$, it follows that

$$\dot{\underline{\theta}} = \mathbf{\Gamma}' \left(\underline{p}^T \underline{e} \right) \underline{\omega} \tag{54}$$

with $\mathbf{\Gamma}' = \frac{\mathbf{\Gamma}}{b_{pn}}$. By partitioning the parameter vectors, we can obtain the adaptive laws for the parameters of the two approximators. Since $\dot{V} < 0$ from (51) we obtain that $\underline{e}$ and $\underline{\phi}$ are uniformly bounded. Because of the boudedness of $\underline{e}$, $\underline{\phi}$ and $\underline{\omega}$ we see from (43) that $\dot{\underline{e}}$ is bounded as well. Thus $\underline{e}$ is uniformly continuous and so is $\dot{V}\left(\underline{e}, \underline{\phi}\right)$. From the fact that

$$V = \underline{e}^T \mathbf{P} \underline{e} + \underline{\phi}^T \mathbf{\Gamma}^{-1} \underline{\phi} \tag{55}$$

$$\dot{V} = -\underline{e}^T \mathbf{Q} \underline{e} \tag{56}$$

we have that

$$\lim_{t \to \infty} V = V^* \tag{57}$$

exists, with

$$V^* - V_0 = -\int_0^\infty \underline{e}^T \mathbf{Q} \underline{e} dt \tag{58}$$

Since the left-hand side is known to be finite, we know that the term on the right-hand side must be finite. We known that since $\underline{e}^T \mathbf{Q} \underline{e}$ is positive, uniformly continuous and has a finite integral that

$$\lim_{t \to \infty} \underline{e}^T \mathbf{Q} \underline{e} = 0 \tag{59}$$

and thus

$$\lim_{t \to \infty} \underline{e} = 0 \tag{60}$$

Notice that the sign of the actual adaptation gain matrix $\mathbf{\Gamma}'$ is found to depend on the sign of b_{pn} and so to be able to implement the adaptive law with a proper sign, the sign of b_{pn} must be known. This condition appears in all MRAC schemes. The equations form the adaptive laws that provide a stable adaptive system. The matrix **P** and so the vector $\underline{p}$ can be calculated with Lyapunov's equation starting with a chosen definite symetric matrix **Q**. Furthermore, the product of vectors $\left(\mathbf{P} \underline{b}_I\right)^T \underline{e}$ is called the "compensated error" in adaptive control literature. This adaptive law has the same form as the MIT adaptive laws,

which use the error $\underline{e}$ instead of the compensated error $\underline{p}^T \underline{e}$. Since it can be shown that using the compensated error in the adaptation laws preserves the system stability, the word "compensated" refers to the compensation of the error in order to preserve system stability.

5 Simulation Example

Example 1

The above described adaptive fuzzy/wavelet control algorithm will now be evaluated using the inverted pendulum system depicted in Fig. 1.

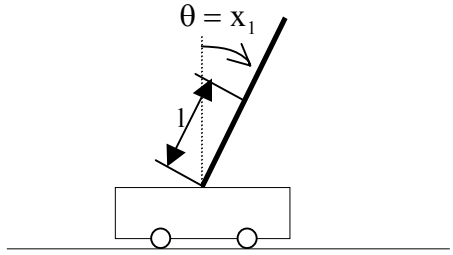

Figure 1
The inverted pendulum system

Let $x_1 = \theta$ and $x_2 = \dot{\theta}$. The dynamic equation of the inverted pendulum is given by [16]

$$\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= \frac{g \sin x_1 - \dfrac{m l x_2^2 \cos(x_1)\sin(x_1)}{m_c + m}}{l\left(\dfrac{4}{3} - \dfrac{m\cos^2(x_1)}{m_c + m}\right)} \\
&\quad + \frac{\dfrac{\cos(x_1)}{m_c + m}}{l\left(\dfrac{4}{3} - \dfrac{m\cos^2(x_1)}{m_c + m}\right)} u_c + d \\
y &= x_1
\end{aligned} \tag{61}$$

where g is the acceleration due to gravity, m_c denotes the mass of the cart, m is the mass of the pole, l is the half-length of the pole, the force u_c represents the control signal and d is the external disturbance. In simulations the following parameter values are used: $m_c = 1Kg$, $m = 0.1Kg$ and $l = 0.5m$. The reference signal is assumed to be $y_r(t) = (\pi/30)\sin(t)$ and an external disturbance $d(t) = 0.1\sin(t)$.

If we require

$$|\underline{x}| \le \frac{\pi}{6}, \; |u| \le 180 \tag{62}$$

and substitute the functions sin(.) and cos(.) by their bounds, we can determine the bounds

$$f^M(x_1, x_2) = 15.78 + 0.366x_2^2 \tag{63}$$

$$g^M(x_1, x_2) = 1.46, \; g_m(x_1, x_2) = 1.12 \tag{64}$$

$k_1 = 2$, $k_2 = 1$ and $Q = \mathrm{diag}(10,10)$ are set. Then the algebraic Riccati equation solution is $P = \begin{bmatrix} 15 & 5 \\ 5 & 5 \end{bmatrix}$ and $\lambda_{\min}(P) = 2.93$. To satisfy the constraint related to $|\underline{x}|$ we choose $M_f = 16$, $M_g = 1.6$ and $\gamma = 0.48$. Five Gaussian membership functions for both x_1 and x_2 (i=1,2) are selected to cover the whole universe of discourse

$$\mu_{F_i^1}(x_i) = \exp\left(-\left(\frac{x_i - \pi/6}{\pi/24}\right)^2\right) \tag{65}$$

$$\mu_{F_i^2}(x_i) = \exp\left(-\left(\frac{x_i - \pi/12}{\pi/24}\right)^2\right) \tag{66}$$

$$\mu_{F_i^3}(x_i) = \exp\left(-\left(\frac{x_i}{\pi/24}\right)^2\right) \tag{67}$$

$$\mu_{F_i^4}(x_i) = \exp\left(-\left(\frac{x_i + \pi/12}{\pi/24}\right)^2\right) \tag{68}$$

$$\mu_{F_i^5}(x_i) = \exp\left(-\left(\frac{x_i + \pi/6}{\pi/24}\right)^2\right) \tag{69}$$

Using the method of trial and error $\gamma_f = 50$ and $\gamma_g = 1$ are chosen. The pendulum initial position is chosen as far as possible $(\theta(0) = x_1 = \pi/20)$ to emphasize the efficiency of our algorithm.

The Haar wavelets are chosen to be the basis of the wavelet network. The vectors $\underline{c}_f$ and $\underline{c}_g$ are both chosen as $\underline{c}_f = \underline{c}_g = \underline{c} = [1 \quad 1]^T$, and the size of our network is chosen as $M_1 = -2$, $M_2 = 2$, $N_1 = -1$ and $N_2 = 1$. In this example, the wavelet bases for $f(\underline{x})$ and $g(\underline{x})$ are chosen and are the same. Therefore, $W_f(\underline{c}_f^T \underline{x}) = W_g(\underline{c}_g^T \underline{x}) = W(\underline{c}^T \underline{x})$.

Two cases have been considered in order to show the influence of the linguistic rules incorporation into the control law:

Case one: the initial values of $\underline{\theta}_f$ and $\underline{\theta}_g$ are chosen arbitrarily.

Case two: the initial values of $\underline{\theta}_f$ and $\underline{\theta}_g$ are deduced from the fuzzy rules describing the system dynamic behavior. For example, if we consider the unforced system, i.e. $u_c = 0$, the acceleration is equal to $f(x_1, x_2)$. Thus we can state intuitively:

"The bigger is x_1, the larger is $f(x_1, x_2)$".

Transforming this fuzzy information into a fuzzy rule we obtain

$$R_f^{(1)}: \text{IF } x_1 \text{ is } F_1^5 \text{ and } x_2 \text{ is } F_2^5, \text{ THEN } f(x_1, x_2) \text{ is Positive Big}$$

where "Positive Big" is a fuzzy set whose membership function is $\mu_{F_i^l}(x_i)$ given by (65)-(69). The acceleration is proportional to the gravity, i.e. $f(x_1, x_2) \cong \alpha \sin(x_1)$, where α is a constant. As $f(x_1, x_2)$ achieves its maximum at $x_1 = \pi/2$, using (63)-(64) we obtain $\alpha \cong 16$. The resulting set of 25 fuzzy rules characterizing $f(x_1, x_2)$ is given in Tab. 1.

Table 1
Linguistic rules for $f(x_1, x_2)$

$f(x_1, x_2)$			x_1				
			F_1^1	F_1^2	F_1^3	F_1^4	F_1^5
			$-\frac{\pi}{6}$	$-\frac{\pi}{12}$	0	$\frac{\pi}{12}$	$\frac{\pi}{6}$
x_2	F_2^1	$-\frac{\pi}{6}$	-8	-4	0	4	8
	F_2^2	$-\frac{\pi}{12}$	-8	-4	0	4	8
	F_2^3	0	-8	-4	0	4	8
	F_2^4	$\frac{\pi}{12}$	-8	-4	0	4	8
	F_2^5	$\frac{\pi}{6}$	-8	-4	0	4	8

Now the following observation is used to determine the fuzzy rules for $g(x_1, x_2)$:

"The smaller is x_1, the larger is $g(x_1, x_2)$".

Similarly to the case of $f(x_1, x_2)$ and based on the bounds (63)-(64) this observation can be quantified into the 25 fuzzy rules summarized in Tab. 2.

Table 2
Linguistic rules for $g(x_1, x_2)$

$g(x_1, x_2)$			x_1				
			F_1^1	F_1^2	F_1^3	F_1^4	F_1^5
			$-\frac{\pi}{6}$	$-\frac{\pi}{12}$	0	$\frac{\pi}{12}$	$\frac{\pi}{6}$
x_2	F_2^1	$-\frac{\pi}{6}$	1.26	1.36	1.46	1.36	1.26
	F_2^2	$-\frac{\pi}{12}$	1.26	1.36	1.46	1.36	1.26
	F_2^3	0	1.26	1.36	1.46	1.36	1.26
	F_2^4	$\frac{\pi}{12}$	1.26	1.36	1.46	1.36	1.26
	F_2^5	$\frac{\pi}{6}$	1.26	1.36	1.46	1.36	1.26

To obtain the same tracking performances the attenuation level ρ is equal to 0.2 in the first case and to 0.8 in the second one.

The tracking performance of both cases for a sinusoidal trajectory is illustrated in Fig. 2.

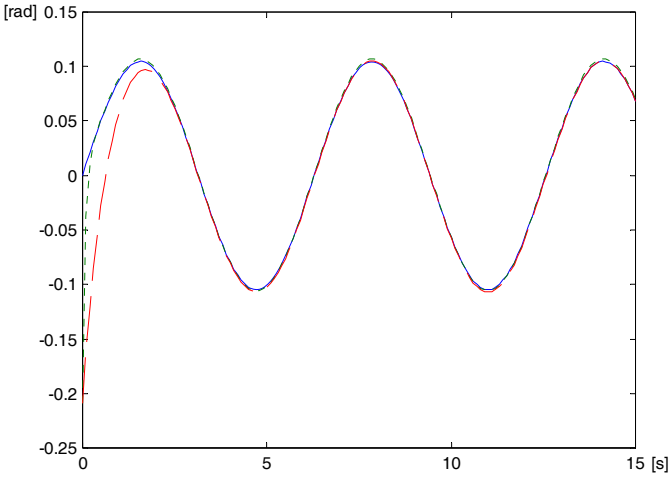

Figure 2

The state x_1 in case 1(red dashed line), in case 2 (green dotted line) and desired value $y_r(t)$ (blue solid line) for $\underline{x}(\underline{0}) = (\pi/12, 0)^T$

Example 2

In this example, we apply the adaptive fuzzy/wavelet controller to the system

$$y'' + \frac{1}{0.25 + y} y' + 1.7y - 0.5u = 0 \tag{70}$$

Define six fuzzy sets over interval <-10, 10> with labels N3, N2, N1, P1, P2, P3. The membership functions are

$$\mu_{N1}(x) = \frac{1}{e^{(x+0.5)^2}} \tag{71}$$

$$\mu_{N2}(x) = \frac{1}{e^{(x+1.5)^2}} \tag{72}$$

$$\mu_{N3}(x) = \frac{1}{1+e^{5(x+2)}} \tag{73}$$

$$\mu_{P1}(x) = \frac{1}{e^{(x-0.5)^2}} \tag{74}$$

$$\mu_{P2}(x) = \frac{1}{e^{(x-1.5)^2}} \tag{75}$$

$$\mu_{P3}(x) = \frac{1}{1+e^{-5(x-2)}} \tag{76}$$

The reference model is assumed to be

$$M(s) = \frac{1}{s^2+2s+1} \tag{77}$$

and the reference signal is the square periodic signal of magnitude 1.5 and frequency 0.01 Hz.

We choose $P = \begin{bmatrix} 50 & 30 \\ 30 & 20 \end{bmatrix}$, $k_1 = 2$, $k_2 = 1$, and $\lambda_{min}(P) = 1.52$. To satisfy the constraint related to $|\underline{x}|$ we choose $\overline{V} = 0.25$, $M_f = 20$, $M_g = 2.1$ and $\gamma = 0.25$.

At the 200th second of simulation the system (64) was switched to another system

$$y''' + 5y'' + \left[\frac{1}{(0.25+y)^2} - 1.7\right] y' + y - 5u = 0 \tag{78}$$

All initial states have been set to zero $y(0) = y'(0) = y''(0) = y'''(0) = 0$.

As can be seen from Fig. 3, the simulation results confirm the good adaptation capability of the proposed control system. The system dynamic changes are in particular manifested by changes of the control input signal (Fig. 4).

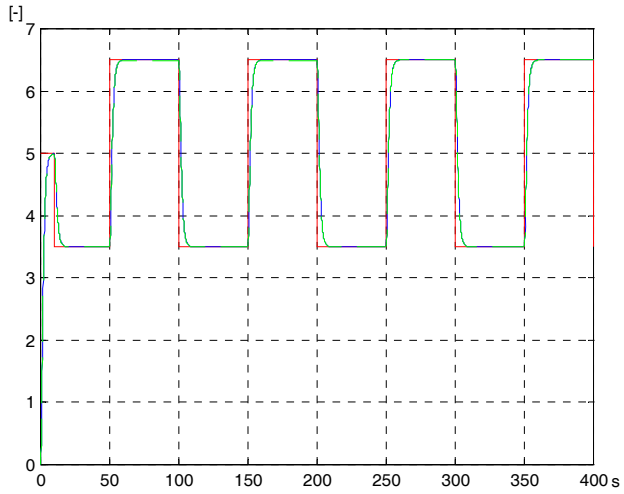

Figure 3

The state x_1 (blue dashed line), its desired reference model value $y_m(t)$ (green solid line) and reference signal (red solid line)

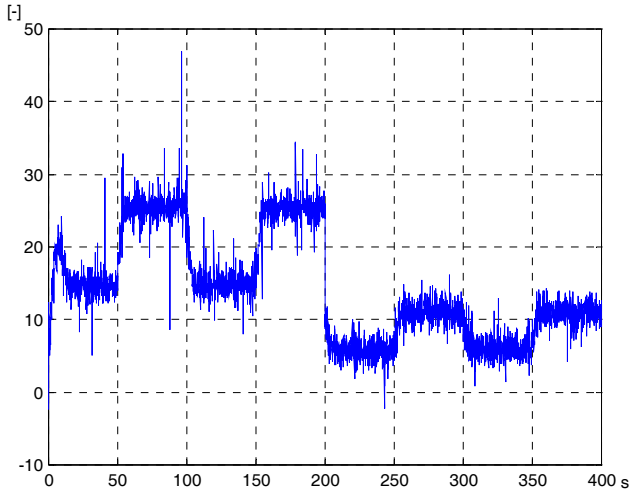

Figure 4

Control signal

Conclusions

The adaptive control technique has been combined with a wavelet network algorithm and a fuzzy approximation method in this study to achieve the desired attenuation of disturbance due to the approximation error and external noise in a class of nonlinear system under a large uncertainty or unknown variation in plant parameter and structure. The major advantage lies in that the accurate mathematical model of the system is not required to be known. The proposed method can guarantee the global stability of the resulting closed-loop system in the sense that all signals involved are uniformly bounded. In addition, the specific formula for the bounds is also given. Finally, the indirect adaptive controller has been used to control a nonlinear system to the origin.

References

[1] Daniel W. C. Ho, Ping-An Zhang, Jinhua Xu, "Fuzzy Wavelet Networks for Function Learning", *IEEE Transactions on Fuzzy Systems*, Vol. 9, No. 1, February 2001

[2] A. Shmilovici, O. Maimon, "On the Solution of Differential Equations with Fuzzy Spline Wavelets", *Fuzzy Sets and Systems*, Vol. 96, 1998, pp. 77-99

[3] T. Kim; Y. Shin; E. J. Powers; S. K. Min; S. S. Lam, Application of Wavelet Denoising to the Detection of Shared Congestion in Overlay Multimedia Networks, IEEE Workshop on Multimedia Signal Processing No. 6, Siena, Italy, pp. 474-477, 2004

[4] W. Wang, J. Ding, "Wavelet Network Model and Its Application to the Prediction of Hydrology", *Nature and Science*, 1(1), 2003

[5] A. Barinka, W. Dahmen, R. Schneider, „Fast Computation of Adaptive Wavelet Expansions", *Numerische Mathematik*, Vol. 105, No. 4, February 2007

[6] Z. Miao, H. Li, J. Wang, „An Optimal Adaptive H-Infinity Tracking Control Design via Wavelet Network", *Journal of Control Theory and Applications*, Vol. 6, No. 3, August 2008

[7] Z. Zhang, „Adaptive Controller Based on Wavelets Neural Network for a Class of Nonlinear Systems", Intelligent Computing, Springer Berlin, Heidelberg, 2006

[8] X. Li, „An Adaptive Wavelet Networks Algorithm for Prediction of Gas Delay Outburst", Advances in Neural Networks, ISNN 2009, 2009

[9] W. Sun, Y. Wang, X. Zhai, „Adaptive Control Based on Recurrent Fuzzy Wavelet Neural Network and Its Application on Robotic Tracking Control", Advances in Neural Networks, ISNN 2006, Springer Berlin, Heidelberg, 2006

[10] L.-X. Wang, "Stable Adaptive Fuzzy Control of Nonlinear System", *IEEE Trans. Fuzzy Syst.*, Vol. 1, pp. 146-155, 1993

[11] M. Thuillard, A Review of Wavelet Networks, Wavenets, Fuzzy Wavenets and their Applications, ESIT 2000, 14-15 September 2000, Aachen, Germany

[12] C. K. Chui, An Introduction to Wavelets, Academic Press, 1992

[13] B.-S. Chen, Y.-M. Cheng, Adaptive Wavelet Network Control Design for Nonlinear Systems, Proceedings of the 35th Conference on Decision and Control, Kobe, Japan, December 1996

[14] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press (San Diego), 1998

[15] T. J. Koo, Analysis of a Class of Fuzzy Controllers, in Proc. 1st Asian Fuzzy Systems Sump., Singapore, Nov. 1993

[16] L.-X. Wang, "Stable Adaptive Fuzzy Controllers with Application to Inverted Pendulum Tracking", *IEEE Trans. on Syst., Man and Cybernetics-part B*, Vol. 26, pp. 677-691, 1996

Computer Adaptive Testing of Student Knowledge

Sanja Maravić Čisar[1], Dragica Radosav[2], Branko Markoski[2], Robert Pinter[1], Petar Čisar[3]

[1]Subotica Tech, Department of Informatics
Marka Oreškovića 16, 24000 Subotica, Serbia
sanjam@vts.su.ac.rs, probi@vts.su.ac.rs

[2]Technical Faculty "Mihajlo Pupin", Department of Informatics
Đure Đakovića bb, 23000 Zrenjanin, Serbia
radosav@zpupin.tf.zr.ac.rs, markoni@uns.ac.rs

[3]Telekom Srbija, Subotica
Prvomajska 2-4, 24000 Subotica, Serbia
petarc@telekom.rs

Abstract: Technological progress, responsible for the declining costs of computers, coupled with the advancement of computer adaptive software have promoted computer adaptive testing (CAT) in higher education, offering alternatives to the conventional paper and pencil examinations. The CAT testing process, statistically conducted through Item Response Theory, is able to react to the individual examinee, keeping examinees on target with test items of an appropriate level of difficulty. The basic goal of adaptive computer tests is to ensure the examinee is supplied questions that are challenging enough for them but not too difficult, which would lead to frustration and confusion. The paper presents a CAT system realized in MATLAB along with its development steps. The application can run from a Matlab command window, or it is possible to make a stand-alone application that does not require the installation of Matlab. The questions are written in a .txt file. This allows the examiner to easily modify and extend the question database, without specific knowledge of the syntax of any programming language. The only requirement is for the examiner (but it is only required) to follow a pre-determined format of question writing. The program enables the testing of student knowledge in C++.

Keywords: computer adaptive testing; Item Response Theory; e-assessment

1 Introduction

Testing is one of the most common ways of knowledge testing. The main goal of testing is to determine the level of a student's knowledge of one or more subject areas in which knowledge is checked. Different methods of knowledge evaluations are in use, such as in-class presentations, writing essays, projects, etc. However, the most common "tool" that is used to test knowledge is the test and oral exam. Since the computer as a teaching tool has been in use more and more in recent decades, and since its use has spread to all levels of education, the computer-based test has become very popular.

Out of all testing methods available today, computer adaptive testing provides the maximal balance of accuracy and efficiency. Over the past few decades, CAT has been used extensively in the areas of education, certification, and licensure [3]. This paper presents a computer adaptive test that was realized with the software package Matlab. The application was done in Matlab based on the program code that can be found at the web address [6]. The original code presents a computer adaptive test for GRE (Graduate Record Exam) and enables questions of the following types: analogy, antonym, and fill in the blanks. It was modified to allow for testing of the basic concepts of C++ in the form of multiple choice questions.

The remainder of this paper is organized as follows: Section 2 briefly reviews the theoretical basis of computerized adaptive tests, along with its benefits and drawbacks. Some basic concepts of Item Response Theory are presented in Section 3, as this is the theoretical foundation behind CAT. Section 4 gives a description of the application. Finally, some future research topics are suggested in Section 5.

2 Theoretical Basis of Computerized Adaptive Tests

CAT (Computerized Adaptive Testing) is a type of test developed to increase the efficiency of estimating the examinee's knowledge. This is achieved by adjusting the questions to the examinee based on his previous answers (therefore often referred to as tailored testing) during the test duration. The degree of difficulty of the subsequent question is chosen in a way so that the new question is neither too hard, nor too easy for the examinee. More precisely, a question is chosen for which it is estimated, with a probability of 50% that the examinee would answer correctly. Of course, the first question cannot be selected in this way because at this point nothing is known about the examinee's capabilities (a question of medium difficulty is chosen), but the selection of the second question can be better adapted to each examinee. With every following answered question, the computer is increasingly better able to evaluate examinee's knowledge.

Some benefits of the CAT are [9] as follows: (a) Tests are given "on demand" and scores are available immediately, (b) Neither answer sheets nor trained test administrators are needed. Test administrator differences are eliminated as a factor in measurement error. (c) Tests are individually paced so that an examinee does not have to wait for others to finish before going on to the next section. Self-paced administration also offers extra time for examinees that need it, potentially reducing one source of test anxiety. (d) Test security may be increased because hard copy test booklets are never compromised. (e) Computerized testing offers a number of options for timing and formatting. Therefore it has the potential to accommodate a wider range of item types. (f) Significantly less time is needed to administer CATs than fixed-item tests since fewer items are needed to achieve acceptable accuracy. CATs can reduce testing time by more than 50% while maintaining the same level of reliability. Shorter testing times also reduce fatigue, a factor that can significantly affect an examinee's test results. (g) CATs can provide accurate scores over a wide range of abilities while traditional tests are usually most accurate for average examinees.

Despite the above advantages, computer adaptive tests have numerous limitations, and they raise several technical and procedural issues [9]: (a) CATs are not applicable for all subjects and skills. Most CATs are based on an item-response theory model, yet item response theory is not applicable to all skills and item types. (b) Hardware limitations may restrict the types of items that can be administered by computer. Items involving detailed art work and graphs or extensive reading passages, for example, may be hard to present. (c) CATs require careful item calibration. The item parameters used in a paper and pencil testing may not hold with a computer adaptive test. (d) CATs are only manageable if a facility has enough computers for a large number of examinees and the examinees are at least partially computer-literate. This can be a great limitation. (e) The test administration procedures are different. This may cause problems for some examinees. (f) With each examinee receiving a different set of questions, there can be perceived inequities. (g) Examinees are not usually permitted to go back and change answers. A clever examinee could intentionally miss initial questions. The CAT program would then assume low ability and select a series of easy questions. The examinee could then go back and change the answers, getting them all right. The result could be 100% correct answers which would result in the examinee's estimated ability being the highest ability level.

The CAT algorithm is usually an iterative process with the following steps:

1 All the items that have not yet been administered are evaluated to determine which will be the best one to administer next given the currently estimated ability level
2 The "best" next item is administered and the examinee responds
3 A new ability estimate is computed based on the responses to all of the administered items.
4 Steps 1 through 3 are repeated until a stopping criterion is met.

The flowchart below serves as an illustration of the CAT algorithm.

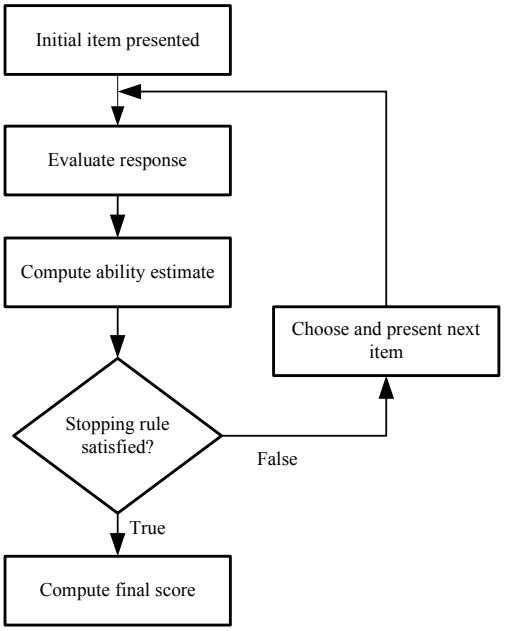

Figure 1
Illustration of the CAT algorithm

Several different methods can be used to compute the statistics needed in each of these three steps, one of them is Item Response Theory (IRT). IRT is a family of mathematical models that describe how people interact with test items [2].

According to the theory of item response, the most important aim of administering a test to an examinee is to place the given candidate on the ability scale [5]. If it is possible to measure the ability for every student who takes the test, two targets have already been met. On the one hand, evaluation of the candidate happens based on how much underlying ability they have. On the other hand, it is possible to compare examinees for purposes of assigning grades, awarding scholarships, etc.

The test that is implemented to determine the unknown hidden feature will contain N items, and they all measure some aspect of the trait. After taking the test, the person taking the test responds to all N items, with the scoring happening dichotomously. This will bring a score of either a 1 or a 0 for each item in the test. Generally, this item score of 1 or 0 is called the examinee's item response. Consequently, the list of 1's and 0's for the N items comprises the examinee's item response vector. The item response vector and the known item parameters are used to calculate an estimate of the examinee's unknown ability parameter.

According to the item response theory, maximum likelihood procedures are applied to make the calculation of the examinee's estimated ability. Similarly to item parameter estimation, the afore-mentioned procedure is iterative in nature. It sets out with some a priori value for the ability of the examinee and the known values of the item parameters. The next step is implementing these values to compute the likelihood of accurate answers to each item for the given person. This is followed by an adjustment to the ability estimate that was obtained, which will in turn improve the correspondence between the computed probabilities and the examinee's item response vector. The process is repeated until it results in an adjustment that is small enough to make the change in the estimated ability negligible. The result is an estimate of the examinee's ability parameter. This process is repeated separately for each person taking the test. Nonetheless, it must be pointed out that the basis of this process is that the approach considers each examinee separately. Thus, the basic problem is how the ability of a single examinee can be estimated.

The estimation equation used is shown below:

$$\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum_{i=1}^{N} -a_i[u_i - P_i(\hat{\theta}_s)]}{\sum_{i=1}^{N} a_i^2 P_i(\hat{\theta}_s) Q_i(\hat{\theta}_s)} \tag{1}$$

where: $\hat{\theta}_s$ is the estimated ability of the examinee within iteration s, a_i is the discrimination parameter of item i, i = 1, 2,... .N.

- u_i is the response made by the examinee to item i: u_i = 1 for a correct response, u_i = 0 for an incorrect response.
- $P_i(\hat{\theta}_s)$ is the probability of correct response to item i, under the given item characteristic curve model, at ability level $\hat{\theta}$ within iteration s.
- $Q_i(\hat{\theta}_s)=1 - P_i(\hat{\theta}_s)$ is the probability of incorrect response to item i, under the given item characteristic curve model, at ability level $\hat{\theta}$ within iteration s.

The CAT problems have been addressed before in the literature [1], [4], [5].

3 Computer Adaptive Tests Based on IRT

For computer adaptive tests which implement IRT (Item Response Theory) a relatively large base of questions for a given task is developed and their informational functions are defined. A well-formed question bank for CATs contains questions that together provide information through a whole range of

properties (θ). The examinee starts the test with an initial estimate of theta (θ), which may be identical for each examinee, or it may be used as predefined information available on the candidate (e.g. results attained in other tests, marks or information from the professor). The question is administered on the basis of the initial theta estimate and immediately evaluated by the computer that generated the test.

3.1 Question Selection

With computer adaptive tests (CAT) based on IRT the subsequent question is selected on the basis of the examinee's scored answers to all previously set questions. In the initial phase of CATs, though only the first or first two questions have been evaluated, the subsequent question is selected based on the rule of "step" – if the first question was answered correctly, the examinee's previous theta estimate will be increased by some value (e.g. 0.50); while, if the first given answer was incorrect, the original theta estimate will be decreased by the same value. As the test continues, an answer sheet is generated which consists of at least one correct and one incorrect answer to the question, thus the MLE (*Maximum Likelihood Estimation*) is used to calculate the new theta estimate, which is based on all the answers that the examinee has given up to that point in the test [11].

After each processed question, the new theta estimate is used for selecting the next question. That question is an un-administered question from the question bank that provides the most information for the currently estimated theta value. Figures 2, 3, and 4 illustrate the "maximum information" questions selected in the computer adaptive test. Figure 2 presents information functions for 10 questions, for the initial theta estimate for a fictitious examinee (indicated by a vertical line). This value is presented at 0.0, which is the mean value of the theta scale. The values of information are calculated for all questions for that theta level. Figure 2 shows that Question 6 provides the most information of the 10 questions for theta = 0.0. Thus, Questions 6 is processed and evaluated [11].

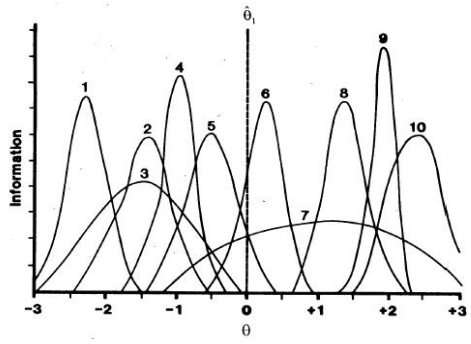

Figure 2
Information functions for the 10 questions [11]

Based on this score (incorrect answer, in this case), the new theta value is defined with a step 1.0, and thus now it is -1.0. Based on the rule of question selection with maximum information, Question 4 was selected (Figure 3) because at the given theta level it contains the most information, and it is evaluated. Given the assumption that the answer to Question 4 is correct, the MLE can be used for the new theta estimate.

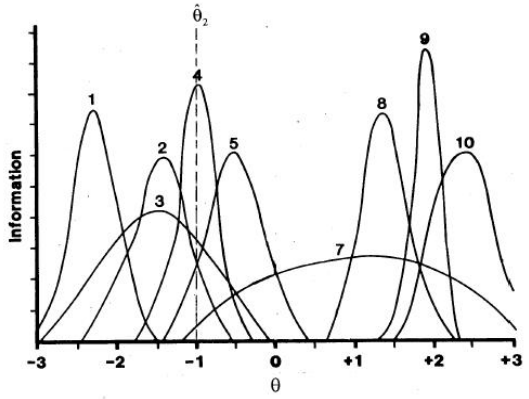

Figure 3
Information functions for 9 questions [11]

The result is theta = -0.50. Again, by selecting the question based on the (Figure 4). The evaluation, theta estimation and question selection continues until the criterion for termination is not met [11].

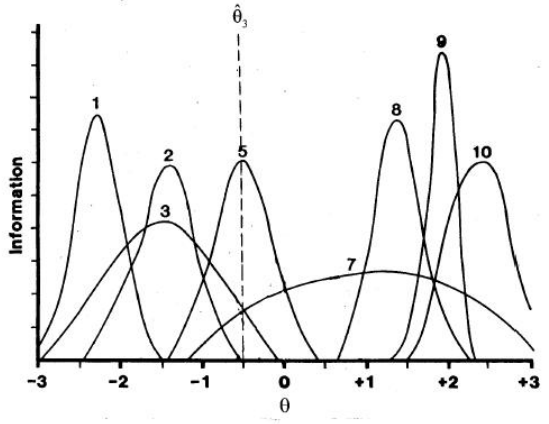

Figure 4
Information functions for 8 questions [11]

3.2 Termination of the Computer Adaptive Test

One of the most important properties of these adaptive tests is the criterion of discontinuing the test may vary depending on different goals of the test. Some tests are used for selection or classification, e.g. whether the subject has managed the acquisition of a certain unit of the learning material, which student will be admitted for secondary school or university, or who will be chosen for a job. Other tests are used for counseling or clinical purposes. The goal of such tests is determine the abilities of the subjects as well as possible. In the context of adaptive tests, these two aims are realized by the two different rules of test termination.

The aim of the classification is that the candidate's results are compared with some cutoff value. The aim is to create the most precise classification. In order for this to be implemented in the context of computer adaptive tests, the theta estimate and its standard error measurement is used. The candidate is classified as above the cutoff value (expressed on the theta scale) if the theta estimate as well as its 95% confidence interval (calculated as ± two standard error measurement) is above or beneath the cut score. As CAT can evaluate this decision after every processed question, the test can be terminated when this condition is met. The result of this test will be the sum of the classification made for the group of examinees where all will have a 5% error rate. The error rate can be controlled by the size of the SEM confidence interval around the theta estimate.

When CATs are not used for classification, a different rule is applied for the termination of the test. In that case it is advisable to evaluate every examinee to the desired level of precision, which is determined in advance by the level of standard error measurement.

This will results in the sum of "equally precise" evaluations, so that all examinees will have results which are equally precise, thus defining a new concept, "fair test". In order to implement equally precise evaluation, CAT enables the user to specify the level of the SEM desired for every examinee. Assuming that the question bank contains enough questions correctly spread along the theta scale and it is possible to continue the test long enough for the examinee, this goal will be realized if the test is terminated when the given SEM level is achieved [11].

3.3 Development of a CAT

According to [3] the final pool of items should consist of approximately five to ten times the number of items that an examinee will eventually see when taking the test. Thus, for a 30-item test, an item bank of 150-300 quality items is highly recommended. Item writing, in and of itself, is a tedious and rigorous process. Developing the initial set of items that will eventually be reduced through the analysis process is a major undertaking, as upwards of 400 items may be needed in order to get to a final pool of 150-300 items.

Once the initial item pool is established, data is collected on each item. IRT analyses typically require at least 300 data points for each item, with 500 being preferred. Since it is not advisable to attempt to get 300 people to complete all items in the initial item pool, often the items have to be split into sub-pools small enough to collect accurate data.

With a sufficient sample size of examinees, the item parameters (discrimination, difficulty, and guessing) can be estimated. These parameters are used to determine which items will be retained in the final item pool, and which items will be revised or discarded. The final pool is then entered into the CAT system, which then creates optimal item selection paths for test takers.

4 Description of the Application

The program that can be found at the web address [6] presents a computer adaptive test and was modified to enable the testing of student knowledge in C++. The application can run from a Matlab command window, or it is possible to make a stand alone application that does not require the installation of Matlab. The MATLAB and Simulink product families are fundamental computational tools at the world's educational institutions. Adopted by more than 5000 universities and colleges, MathWorks products accelerate the pace of learning, teaching, and research in engineering and science. MathWorks products also help prepare students for careers in industry, where the tools are widely used for research and development [10]. Some examples of implementing Matlab as an educational tool can be found in [7], [8].

After starting the program the main window is displayed as is the dialog box for entering basic data on the student (name, surname and index number). Pressing the Enter command button starts the test, as shown in Figure 5.

Figure 5
Startup screen

After pressing the button *Pocetak testa* (Start), the function *pocetak_testa* (test_start) is called and the visibility of objects that are no longer needed has to be set to "off" and the visibility of the edit control (for question displaying), option buttons (for showing multiple choices as answers) and patch object is set to "on". Then the function *final_test* is called, which has two output parameters: an array with correct/incorrect answers (in this case 30) and the second parameter is an array which contains the time (given in seconds) that has elapsed since the student has given the answer for each questions.

After registering for the test a new window opens with the first question. At all times the student can see on the screen which question the student is on, the total number of questions, the text of the question with multiple choice answers, as can be seen in Figure 6. At the bottom of the screen there is a progress bar which illustrates the progress of the student during the test.

Figure 6
Screenshot of a question

```
function [ans_array,time_arr]=final_test
[a b c t1]=ask_qn(1,1,grupa_pitanja,ones(4,4),1);
      ans_array = b;
       time_arr = t1;
end
```

As shown, the next function that is called is the function *ask_qn* which has five input parameters and four output parameters. In the function *ask_qn* everything is handled in one *for* loop which is repeated as many times as there are questions *(comm_arr*). The first calculation is for the determination of the question's difficulty that needs to be answered.

The questions are divided based on their difficulty into three groups, easy, medium and difficult question (parameter *question_set* could be 1, 2 or 3).

```
deciding_factor=ask_1;

question_set=normalize_qno(question_set,deciding_factor,1,3);
```

At the beginning, the parameter *question_set* is 1 and also the parameter *ask_1*. The parameter *ask_1* determines by how much to increase or decrease the parameter *question_set*. In this case, the test starts with a question of medium difficulty, which is assigned in test the results with number 2. If the student gives a correct answer to this question, the algorithm of the test passes to the first question in the group of difficult questions (assigned number 3), and if the answer given to the first question is incorrect then the group with easy questions is selected (assigned number 1). The questions are written in a txt file and they are invoked by calling the appropriate function in the program. This allows the examiner to easily modify and extend the question database, without knowledge of the syntax of any programming language; it is only required to follow a determined format of questions writing.

Also, the type of question that will be selected as the next question is determined. In this test there is only one type of questions (MCQ), but it is also possible to set some other types of questions (analogy, antonym, fill in the blanks etc.). So, in this case the array *com_arr(i)* consists of only the ones.

```
ask_1_char_type=question_type(com_arr(i));

ask_1_char_type = 'pitanja'

function question_str=question_type(number)

if(number==1)

    question_str='pitanja';

end
```

The next parameter that is necessary to obtain is *question_status*, which contains data in form of a matrix (question difficulty and the type of question). In the case of questions with a difficulty of level three and only one type of the question, it would be a 1-dimensional array initialized with the *ones (1,3)*. After that the function *ask* is called:

```
[ask_1 q_time]=

=ask(ask_1_char_txt,question_status(com_arr(i),question_set))
;
```

The first parameter of the function gives the information from which .txt file to read the questions, and the second parameter *question_status (com_arr(i),question_set))* obtains the information from which line in the .txt file to start reading. The output parameters are placed in variables *qn_time* and *ans_array*:

```
qn_time=[qn_time q_time];

ans_array=[ans_array ask_1];
```

The next step is to start measuring the time that passes before the student selects any of the five given options as answer. The elapsed time is recorded in the variable *q_time=[q_time toc]*; there is verification whether the given answer is correct (`if(taster==num_tline)`) and if it is, the related variable is set to 1.

```
q_time=[q_time toc];

 if(taster==num_tline)

        output_check=1;

      end

end
```

Once the student has given answers to all questions, the program continues to run in the function *pocetak_testa* from the part where the function was called:

```
[a b]=final_test;
```

where *a* is the array with answers and *b* is the array with the time elapsed per each question. The final result is calculated with the call of the function *totaling* with the parameter *a*.

```
total_marks=int2str(totaling(a));
```

After answering the last question, the examinee can see their results immediately on the screen. If the examinee selects the option to save the test results, the appropriate function parameters are called. From the text file can be seen the level of the question's difficulty, whether or not the answer was correct or incorrect, and the time needed for answering each question (i.e. until pressing the command button Next question/Show results).

Figure 7
Test results view

Conclusions

Computerized adaptive tests offer many advantages over conventional paper and pencil tests: efficiency in the form of reduced testing time; appropriate challenge for the individual examinee's ability level; flexibility in arranging testing time and location; the potential for the use of sophisticated dynamic graphics via the computer; immediate scoring; the potential for the synchronous collection of data during testing and so on [12].

This paper reports on the use of a computer adaptive test for examining a student's knowledge in C++. The motivation behind this work was to investigate techniques for the improvement of student assessment. Future work will involve the further analysis of the test statistics and the improvement of the classification of questions based on the student's test results.

References

[1] Baker, F. "The Basics of Item Response Theory", Chapter 5, Estimating an Examinee's Ability, http://echo.edres.org:8080/irt/baker/chapter5.pdf, 2001

[2] Embretson, S. E., Reise, S. P., "Item Response Theory for Psychologists", Mahwah NJ, Lawrence Erlbaum Associates, 2000

[3] Fetzer, M., Dainis, A., Lambert, S., Meade, A., PreVisor's PreViewTM Computer Adaptive Testing (CAT) in an Employment Context, April 2008, White Paper, www.previsor.com/pdf/WPCAT.pdf

[4] Ju, Gin-Fon N., Bork, A., "The Implementation of an Adaptive Test on the Computer", Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05), Kaohsiung, Taiwan, 2005, pp. 822-823, http://doi.ieeecomputersociety.org/10.1109/ ICALT.2005.274

[5] Kardan, S., Kardan, A., " Towards a More Accurate Knowledge Level Estimation", 2009 Sixth International Conference on Information Technology: New Generations, Las Vegas, Nevada, pp. 1134-1139, 2009, http://doi.ieeecomputersociety.org/10.1109/ITNG.2009.154

[6] Mathworks, Computer Adaptive Test Demystified, http://www.mathworks.com/matlabcentral/fileexchange/12467-computer-adaptive-test-demystified-gre-pattern

[7] Sadeghi, S., Mirsalim, M., Hassanpour Isfahani, A., "Dynamic Modeling and Simulation of a Switched Reluctance Motor in a Series Hybrid Electric Vehicle", Acta Polytechnica Hungarica, 2010, Vol. 7, No. 1, pp. 51-71

[8] Takács, M., Szakál, A., Csikós Pajor, G., "Software Supported Mathematics Teaching", Proceedings of the 3rd International Conference on Information Technology Based Higher Education and Training, ITHET Conference, July 4-6, 2002, Budapest, Hungary

[9] http://echo.edres.org:8080/scripts/cat/catdemo.htm. Rudner, L., An On-line, Interactive, Computer Adaptive Testing Tutorial

[10] http://www.mathworks.com/academia/

[11] http://www.psych.umn.edu/psylabs/catcentral/

[12] www.rpi.edu/~faheyj2/SB/diss1.doc. Bringsjord, E., "Computer-Adaptive Versus Paper-and-Pencil Testing Environments: An Experimental Analysis of Examinee Experience"

Intelligent Mobile Robot Motion Control in Unstructured Environments

Gyula Mester

Department of Informatics, Robotics Laboratory, University of Szeged
Árpád tér 2, H-6720 Szeged, Hungary
gmester@inf.u-szeged.hu

Abstract: This paper presents the intelligent wheeled mobile robot motion control in unstructured environments. The fuzzy control of a wheeled mobile robot motion in unstructured environments with obstacles and slopes is proposed. Outputs of the fuzzy controller are the angular speed difference between the left and right wheels of the mobile robot and the mobile robot velocity. The simulation results show the effectiveness and the validity of the obstacle avoidance behavior in an unstructured environment and the velocity control of a wheeled mobile robot motion of the proposed fuzzy control strategy. Wireless sensor-based remote control of mobile robots motion in unstructured environments using the Sun SPOT technology is proposed. The proposed method has been implemented on the miniature mobile robot Khepera that is equipped with sensors. Finally, the effectiveness and efficiency of the proposed sensor-based remote control strategy are demonstrated by experimental studies and good experimental results.

Keywords: intelligent wheeled mobile robot; motion control; unknown and unstructured environments; obstacles and slopes; fuzzy control strategy; wireless sensor-based remote control; Sun SPOT technology; simulation results; experimental studies; mobile robot Khepera

1 Introduction

In recent years, there has been a growing interest in mobile robot motion control. This paper presents intelligent mobile robot motion control in unstructured environments. The paper actually is a continuation of a conference paper [1]. The paper deals with the fuzzy velocity control of a mobile robot motion in an unstructured environment with slopes and obstacles and gives the wireless sensor-based remote control of mobile robots motion in an unstructured environment with obstacles using the Sun SPOT technology.

The wheeled mobile robot must be capable of sensing its environment. Conventionally, mobile robots are equipped by ultrasonic sensors and a stereo-vision system. It is supposed that the autonomous mobile robot has groups of

ultrasonic sensors to detect obstacles in the front, to the right and to the left of the vehicle, that the model of the wheeled mobile robot has two driving wheels and that the angular velocities of the two wheels are independently controlled.

When the vehicle is moving towards the target and the sensors detect an obstacle or slope, an avoiding strategy is necessary. While the mobile robot is moving it is important to compromise between avoiding the obstacles and slopes and moving towards the target position. The fuzzy control of a wheeled mobile robot motion in unstructured environments with obstacles and slopes is proposed. Outputs of the fuzzy controller are the angular speed difference between the left and right wheels of the vehicle and the vehicle velocity. The simulation results show the effectiveness and the validity of the obstacle avoidance behavior in unstructured environments and the velocity control of a wheeled mobile robot motion of the proposed fuzzy control strategy. The wireless sensor-based remote control of mobile robots motion in unstructured environments using the Sun SPOT technology is proposed.

The proposed method has been implemented on the miniature mobile robot Khepera that is equipped with sensors and the free range Spot from the Sun Spot technology.

Finally, the effectiveness and efficiency of the proposed sensor-based remote control strategy are demonstrated by experimental studies and good experimental results of the obstacle avoidance behavior in unstructured environments.

The paper is organized as follows:

- Section 1: Introduction.
- In Section 2, the strategy of autonomous wheeled mobile robot motion control in unstructured environments is proposed.
- In Section 3, the simulation results are illustrated.
- In Section 4, the wireless robot-sensor networked systems are illustrated.
- In Section 5, the Sun-SPOT-based remote control of mobile robots is proposed.
- Conclusions are given in Section 6.

Currently much research in robotics deals with different problems of the motion of wheeled mobile robots and the motion control of wheeled mobile robots in unstructured environments. Fuzzy logic approaches to mobile robot navigation and obstacle avoidance have been investigated by several researchers. Many application works of fuzzy logic in the mobile robot field have given promising results.

[2] has presented a strategy for the autonomous navigation of field mobile robots on hazardous natural terrain using a fuzzy logic approach and a novel measure of terrain traversability. The navigation strategy is comprised of three simple,

independent behaviors: seek-goal, traverse-terrain, and avoid obstacles. This navigation strategy requires no a priori information about the environment.

The sensor-based navigation of a mobile robot in an indoor environment is very well presented in [3]. The paper deals with the problem of the navigation of a mobile robot either in an unknown indoor environment or in a partially-known one. Fuzzy controllers are created for the navigation of the real robot. The good results obtained illustrate the robustness of a fuzzy logic approach with regard to sensor imperfections.

The fuzzy reactive control of a mobile robot incorporating a real/virtual target-switching strategy has been made in [4]. Real-time fuzzy reactive control is investigated for automatic navigation of an intelligent mobile robot in unknown and changing environments. The reactive rule base governing the robot behavior is synthesized corresponding to the various situations defined by instant mobile robot motion, environment and target information.

Paper [5] presents a control method for the formation on nonholomic mobile robots. Robots track desired trajectories in the environment with static convex-shaped obstacles. The algorithm includes collision-avoidance between robots and obstacles.

2 Strategy of Autonomous Wheeled Mobile Robot Motion Control in Unstructured Environments

In this section fuzzy control is applied to the navigation of the autonomous mobile robot in unstructured environments with obstacles and slopes [1], [6], [7], [8], [9], [10], [11], [12]. It is supposed that: the autonomous mobile robot has two wheels driven independently and groups of ultrasonic sensors to detect obstacles in the front, to the right and to the left of the vehicle. When the vehicle is moving towards the target and the sensors detect an obstacle, an avoiding strategy is necessary. While the mobile robot is moving it is important to compromise between:

- avoiding the obstacles and
- moving towards the target position.

With obstacles present in the unknown environment, the mobile robot reacts based on both the sensed information of the obstacles and the relative position of the target [4]. In moving towards the target and avoiding obstacles, the mobile robot changes its:

- orientation and
- velocity.

When an obstacle in an unknown environment is very close, the mobile robot slows down and rapidly changes its orientation. The navigation strategy has to come as near to the target position as possible while avoiding collision with the obstacles in an unknown environment.

The intelligent mobile robot reactive behavior is formulated in fuzzy rules. Fuzzy-logic-based control is applied to realize a mobile robot motion in an unknown environment with obstacles.

Inputs to the fuzzy controller are:

- the obstacle distance p,
- the obstacle orientation θ_1 (which is the angle between the robot moving direction and the line connecting the robot's center with the obstacle),
- the target distance l,
- the target orientation θ_2 (which is the angle between the robot moving direction and the line connecting the robot's center with the target).

Outputs of the fuzzy controller are:

- the angular speed difference between the left and right wheels (wheel angular speed correction) of the vehicle: $\Delta\omega = \omega_r - \omega_l$ and
- the vehicle velocity.

The obstacle orientation θ_1 and the target orientation θ_2 are determined by the obstacle/target position and the robot position in a world coordinate system, respectively. The obstacle orientation θ_1 and the target orientation θ_2 are defined as positive when the obstacle/target is located to the right of the robot's direction of movement; otherwise, the obstacle orientation θ_1 and the target orientation θ_2 are negative [1].

The block diagram of the fuzzy inference system is presented in Fig. 1.

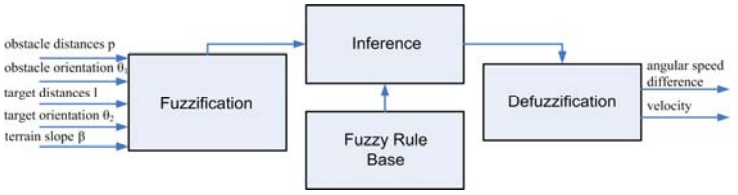

Figure 1
The block diagram of the fuzzy inference system

For the proposed fuzzy controller the input variables for the obstacle distance p are simply expressed using two linguistic labels - Gaussian membership functions - near and far ($p \in [0, 3\ m]$). The input variables for the obstacle orientation θ_1 are expressed using two linguistic labels - Gaussian membership functions - left and right ($\theta_1 \in [-\pi, \pi\ rad]$).

For the proposed fuzzy controller, the input variables for the terrain slope β is simply expressed using three linguistic labels - Gaussian membership functions – sloped left, flat and sloped right (β Є [-3.14, 3.14 rad]), β is the average slope value. The input variables for the target distance l are simply expressed using two linguistic labels - Gaussian membership functions - near and far (l Є [0, 3 m]). The input variables for the target orientation θ_2 are simply expressed using three linguistic labels - Gaussian membership functions - left, targetdirection and right (θ_2 Є [-3.14, 3.14 rad]).

The fuzzy sets for the output variables of the wheel angular speed correction $\Delta\omega = \omega_r - \omega_l$ (*turn-right, zero* and *turn-left*) of the mobile robot are shown in Fig. 2.

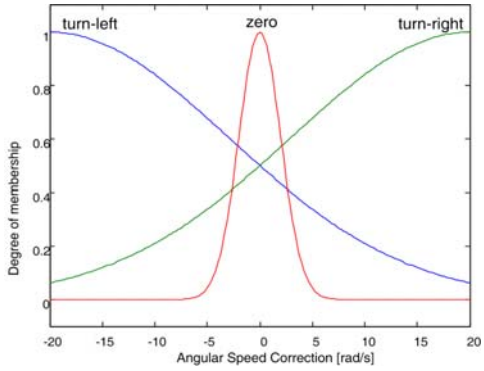

Figure 2
Membership functions of the angular speed difference Δω

The output variables are normalized between: Δω Є [-20, 20 rad/s]. The other output variable of the fuzzy controller is vehicle velocity. The output variables are normalized between: Velocity Є [-10, 20 m/s]. The fuzzy sets for the output variables - Velocity (low and high) are shown in Fig. 3.

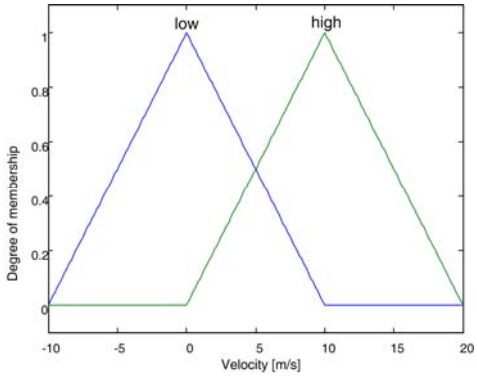

Figure 3
Membership functions of the velocity of the mobile robot

The rule-base for mobile robot fuzzy control are:

R1: If θ_2 is right and β is sloped left then $\Delta\omega$ is turn-right

R2: If θ_2 is left and β is sloped right then $\Delta\omega$ is turn-left

R3: If p is near and l is far and θ_1 is left and β is sloped left then $\Delta\omega$ is turn-right

R4: If p is near and l is far and θ_1 is right and β is sloped right then $\Delta\omega$ is turn-left

R5: If θ_2 is target direction and β is flat then $\Delta\omega$ is zero

R6: If p is far and θ_2 is target direction and β is flat then $\Delta\omega$ is zero

R7: If p is near and l is far then velocity is low

R8: If p is far and l is far then velocity is high

R9: If p is far and l is near then velocity is low.

In the present implementation of the fuzzy controller the Center of Area method of defuzzification is used.

3 Simulation Results

Simulation experiments are commonly used for the initial system analysis and control design while the experimental scalable testbed system must be used in the final phase of system evaluation and control verification. The obtained results and control architecture can afterwards be adapted to the different application of mobile robots. Based on this, the important task in system development is the accurate and valuable modeling of the observed system.

In this instance, the author applied the proposed fuzzy controller to the mobile robot moving in an unstructured environment with obstacles [14]. A simulation example of a wheeled mobile robot is presented in Fig. 4. The corresponding fuzzy control is implemented to perform tasks of obstacle and collision avoidance. The results of the simulation are shown in Fig. 4. regarding the goal seeking and the obstacle avoidance mobile robot paths.

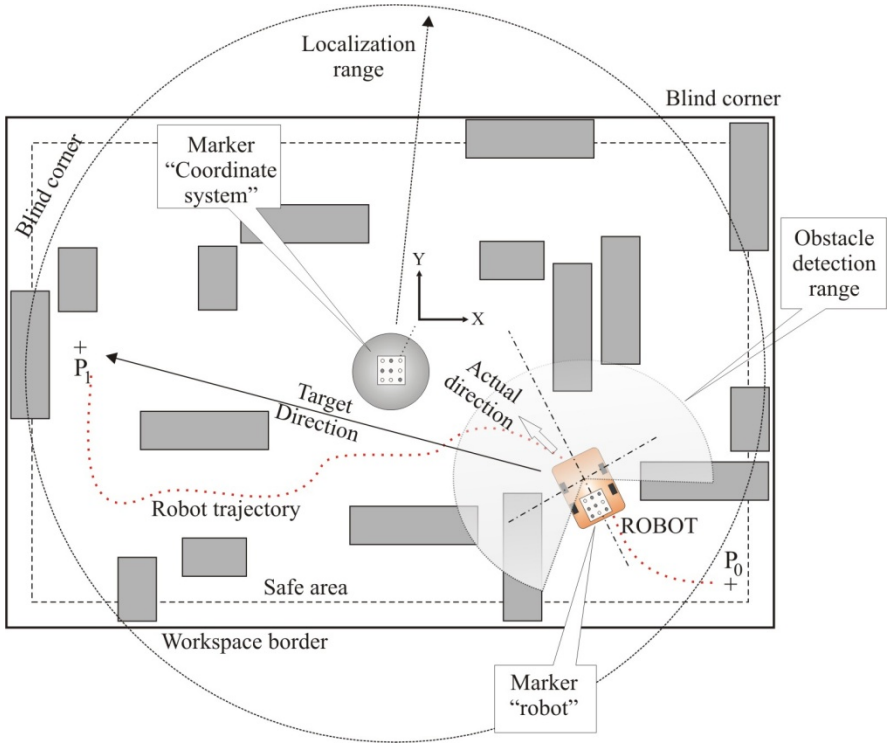

Figure 4
Example of an obstacle avoidance scenario, obstacle avoidance trajectory of mobile robot

4 Wireless Robot-Sensor Networked Systems

Wireless Robot-Sensor Networked systems refer to multiple robots operating together in coordination or cooperatively with sensors, embedded computers, and human users [13], [14]. Cooperation entails more than one entity working toward a common goal while coordination implies a relationship between entities that ensures efficiency or harmony.

Communication between entities is fundamental to both cooperation and coordination and hence the central role of the networked system. Embedded computers and sensors are now ubiquitous in homes and factories, and increasingly wireless ad-hoc networks or plug-and-play wired networks are becoming commonplace.

Robots are functioning in environments while performing tasks that require them to coordinate with other robots, cooperate with humans, and act on information derived from multiple sensors. In many cases, these human users, the robots and

sensors are not collocated, and the coordination and communication happens through a network. Networked robots allow multiple robots and auxiliary entities to perform tasks that are well beyond the abilities of a single robot [13], [14].

Robots can automatically couple to perform locomotion and manipulation tasks that either a single robot cannot perform or that would require a special-purpose larger robot to perform. They can also coordinate to perform search and reconnaissance tasks exploiting the efficiency that is inherent in parallelism. Further they can perform independent tasks that need to be coordinated.

Another advantage of networked robots is improved efficiency. Tasks like searching or mapping, in principle, are performed faster with an increase in the number of robots. A speed-up in manufacturing operations can be achieved by deploying multiple robots performing operations in parallel, but in a coordinated fashion.

Perhaps the greatest advantage of using the network to connect robots is the ability to connect and harness physically-removed assets. Mobile robots can react to information sensed by other mobile robots in the next room. Human users can use machines that are remotely located via the network.

The ability to network robots also enables fault-tolerance in design. If robots can in fact dynamically reconfigure themselves using the network, they are more tolerant to robot failures. Finally, networked robots have the potential to provide great synergy by bringing together components with complementary benefits and making the whole greater than the sum of the parts [13], [14].

5 Sun SPOT-based Remote Control of Wheeled Mobile Robots

In this paper Sun SPOT-s (Small Programmable Object Technology) have been used to creat remote control over a Khepera® mobile robot [15], [16], [17], [18].

A Sun SPOT is a small electronic device made by Sun Microsystems. The Sun SPOT is designed to be a flexible development platform, capable of hosting widely differing application modules. The Sun SPOT connection strategy [19], [20], [21], [22], [23], [24], [25], is presented in Fig. 5.

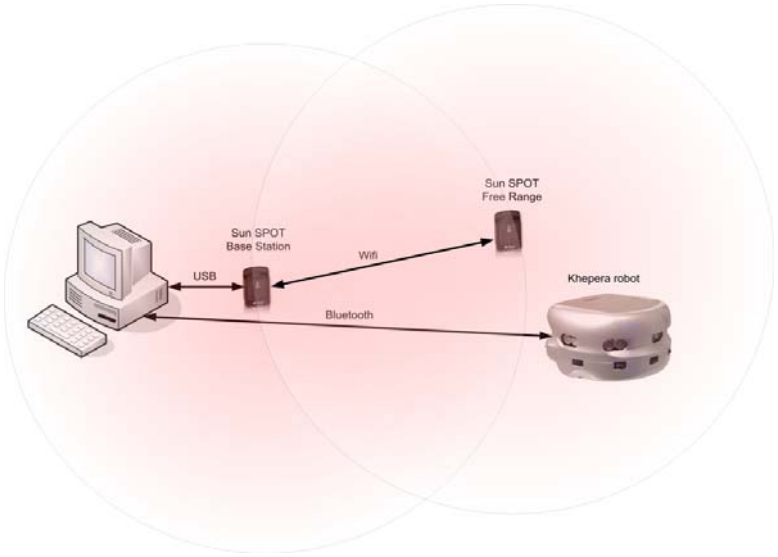

Figure 5
Remote control system

For this task 2 SunSPOT-s have been used from the development kit (Sun Microsystems, Inc. 2007). Sun SPOTs are programmed in a Java programming language, with the Java VM run on the hardware itself. It has quite a powerful main processor running the Java VM "Squawk" and which serves as an IEEE 802.15.4 wireless network node. The Sun SPOT's wireless protocol is Zigbee-based protocol [26], [27], [28], [29], [30], [31], [32], [33].

The Sun SPOT is designed to be a flexible development platform, capable of hosting widely differing application modules. The Sun SPOT base station is used to read the data from the free range SPOT and send its contents to the PC. The PC sends via Bluetooth the control signal to the mobile robot Kephera. The miniature mobile robot Khepera® is equipped with 9 infrared sensors, 5 ultrasonic sensors and an integrated Bluetooth communication module (Fig. 6).

Figure 6
Khepera mobile robot

In the Robotics Laboratory, Department of Informatics, University of Szeged it is possible to use the sensor-based remote control system [21]. The user can start control experiment of mobile robots in Sun SPOT environment (Fig. 7), [22], [23].

Figure 7
Remote control experiment

Conclusions

The paper deals with the fuzzy control of mobile robot motion in an unstructured environment with slopes and obstacles. Further, it presents the wireless sensor-based remote control of mobile robots motion in an unstructured environment with obstacles using the Sun SPOT technology. When the vehicle moves towards the target and the sensors detect an obstacle, an avoiding strategy and velocity control are necessary. With obstacles present in the unstructured environment, the mobile robot reacts based on both the sensed information of the obstacles and the relative position of the target.

The paper proposed the wireless sensor-based remote control of mobile robots motion in unstructured environments with obstacles and a fuzzy reactive navigation strategy of collision-free motion and velocity control in unstructured environments with slopes and obstacles.

The proposed method has been implemented on the miniature mobile robot Khepera® equipped with sensors. The wireless robot-sensor networked systems are illustrated.

The simulation results show the effectiveness and the validity of the obstacle avoidance behavior in unstructured environments and the velocity control of a wheeled mobile robot motion of the proposed fuzzy control strategy.

Finally, the effectiveness and efficiency of the proposed sensor-based remote control strategy are demonstrated by experimental studies and good experimental results.

Acknowledgement

The results in the paper are partially supported by the TÁMOP-4.2.2/08/1/2008-0008 program of the Hungarian National Development Agency.

References

[1] Gyula Mester, "Obstacle - Slope Avoidance and Velocity Control of Wheeled Mobile Robots using Fuzzy Reasoning", Proceedings of the IEEE 13th International Conference on Intelligent Engineering Systems, INES 2009, Barbados, pp. 245-249, ISBN: 978-1-4244-4113-6, Library of Congress: 2009901330, DOI: 10.1109/INES.2009.4924770, 2009

[2] Homayoun Seraji, Ayanna Howard, Edward Tunstel, "Terrain-based Navigation of Planetary Rovers: A Fuzzy Logic Approach", Proceedings of the 6th International Symposium on Artifical Intelligence and Robotics & Automation in Space: i-SAIRAS 2001, pp. 1-6, Quebec, Canada, 2001

[3] H. Maaref, C. Barret, "Sensor-based Navigation of a Mobile Robot in an Indoor Environment", Robotics and Autonomous Systems, Vol. 38, pp. 1-18, 2002

[4] W. L. Xu, S. K. Tso, Y. H. Fung: "Fuzzy Reactive Control of a Mobile Robot Incorporating a Real/Virtual Target Switching Strategy", Robotics and Autonomous Systems, Vol. 23, pp. 171-186, 1998

[5] Krzystof Kozlowskí, Wojciech Kowalczik, "Motion Control for Formation of Mobile Robots in Environment with Obstacles",Studies in Computational Intelligence, Towards Intelligent Engineering and Information Technology, Volume 243/2009, pp. 203-219, ISBN 978-1-642-03736-8, Library of Congress: 2009933683, DOI: 10.1007/978-3-642-03737-5_15, Springer, 2009

[6] Gyula Mester, "Intelligent Mobile Robot Controller Design", Proceedings of the Intelligent Engineering Systems", INES 2006, pp. 282-286, London, United Kingdom, 2006

[7] Gyula Mester, "Motion Control of Wheeled Mobile Robots". Proceedings of the IEEE SISY 2006, pp. 119-130, Subotica, Serbia, 2006

[8] Gyula Mester, „Improving the Mobile Robot Control in Unknown Environments", Proceedings of the YUINFO'2007, pp. 1-5, Kopaonik, Serbia, 2007

[9] Gyula Mester, „Obstacle Avoidance of Mobile Robots in Unknown Environments". Proceedings of the IEEE SISY 2007, pp. 123-128, Subotica, Serbia, 2007

[10] Gyula Mester, „Obstacle and Slope Avoidance of Mobile Robots in Unknown Environments". Proceedings of the XXV. Science in Practice, pp. 27-33, Schweinfurt, 2007, Germany

[11] Gyula Mester, „Obstacle Avoidance and Velocity Control of Mobile Robots", 6th International Symposium on Intelligent Systems and Informatics, Proceedings of the IEEE SISY 2008, pp. 1-5, IEEE Catalog

Number: CFP0884C-CDR, ISBN: 978-1-4244-2407-8 Library of Congress: 2008903275, Subotica, Serbia, 2008

[12] Gyula Mester, „Designing of the Intelligent Mobile Robot Control inthe Matlab Environment", Proceedings of the YUINFO'2008, pp. 1-5, ISBN: 978-86-85525-03-2, Kopaonik, Serbia, 2008

[13] B. Siciliano, O. Khatib, Eds., Handbook of Robotics, Springer, 2008, ISBN: 978-3-540-23957-4, pp. 943-955

[14] Aleksandar Rodic, Gyula Mester, Virtual WRSN – Modeling and Simulation of Wireless Robot-Sensor Networked Systems. Proceedings of the 8th IEEE International Symposium on Intelligent Systems and Informatics, SISY 2010, pp. 115-120, ISBN: 978-1-4244-7395-3, IEEE Catalog Number: CFP1084C-CDR, Subotica, Serbia, 2010

[15] Gyula Mester, "Web-based Remote Control of Mobile Robots Motion", Proceedings of the YUINFO'2009, pp. 1-3, ISBN: 978-86-85525-04-9, Kopaonik, Serbia, 2009

[16] Gyula Mester, "Intelligent Mobil Robot Control in Unknown Environments", Intelligent Engineering Systems and Computational Cybernetics, Part I Intelligent Robotics, pp. 15-26, ISBN 978-1-4020-8677-9, Library of Congress: 2008934137, DOI 10.1007/978-1-4020-8678-6_2, Springer, 2009

[17] Gyula Mester, Aleksandar Rodic, "Autonomous Locomotion of Humanoid Robots in Presence of Mobile and Immobile Obstacles", Studies in Computational Intelligence, Towards Intelligent Engineering and Information Technology, Volume 243/2009, pp. 279-293, ISBN 978-1-642-03736-8, Library of Congress: 2009933683, DOI 10.1007/978-3-642-03737-5-_20, Springer, 2009

[18] Gyula Mester, "Sensor-based Control of Autonomous Wheeled Mobile Robots", Ipsi Journal, TIR, Volume 6, Number 2, pp. 29-34, ISSN 1820-4503, 2010

[19] Gyula Mester, "Wireless Sensor-based Control of Mobile Robots Motion", Proceedings of the IEEE SISY 2009, pp. 81-84, IEEE Catalog Number: CFP0984C-CDR, ISBN: 978-1-4244-5349-8 Library of Congress: 2009909575, DOI 10.1109/SISY.2009.52911190, Subotica, Serbia, 2009

[20] Aleksandar Rodić, Dusko Katić, Gyula Mester, "Ambient Intelligent Robot-Sensor Networks for Environmental Surveillance and Remote Sensing", Proceedings of the IEEE SISY 2009, pp. 39-44, IEEE Catalog Number: CFP0984C-CDR, ISBN: 978-1-4244-5349-8 Library of Congress: 2009909575, DOI 10.1109/SISY.2009.5291141, Subotica, Serbia, 2009

[21] http://www.inf.u-szeged.hu/robotics

[22] Tamás Szépe, "Sensor-based Control of an Autonomous Wheeled Mobile Robot", Proceedings from PROSENSE 3rd Seminar Presentations, pp. 34-37, Institute Jožef Stefan, Ljubljana, Slovenia, January 2010

[23] Tamás Szépe, "Robotirányítás támogatása távoli érzékelőrendszerrel", VMTT konferencia, konferenciakiadvány, pp. 527-532, Újvidék, Szerbia, 2010

[24] Pozna, C., Prahoveanu, V., Precup, R-E, "A New Pattern of Knowledge Based on Experimenting the Causal Relation", Proceedings of the 14th IEEE International Conference on Intelligent Engineering Systems INES 2010, pp.61-66, Las Palmas of Gran Canaria, Spain, 2010

[25] Pozna, C., Precup, R-E., Minculete, N., Antonia, C., "Properties of Classes, Subclasses and Objects in an Abstraction Model", In Proceedings of the 19th International Workshop on Robotics in Alpe-Adria-Danube Region, RAAD, 23-25 June, 2010, pp. 291-296, Budapest, Hungary, 2010

[26] Péter Kucsera: Industrial Component-based Sample Mobile Robot System, Acta Polytechnica Hungarica, Vol. 4, No. 4, ISSN 1785-8860, 2007

[27] Péter Kucsera, "Autonomous Advertising Mobile Robot for Exhibitions, Developed at BMF", Studies in Computational Intelligence, Towards Intelligent Engineering and Information Technology, Volume 243/2009, pp. 295-303, ISBN 978-1-642-03736-8, Library of Congress: 2009933683, DOI: 10.1007/978-3-642-03737-5_21, Springer, 2009

[28] Dragan Saletic, Uros Popovic, "Fuzzy Expert System for Automatic Movement Control of a Platform on a Ground with Obstacles", Proceedings of the YuINFO 2006, pp. 1-6, Kopaonik, Serbia, 2006

[29] Matijevics István, Simon János, "Comparison of Various Wireless Sensor Networks and their Implementation", Proceedings of the Conference SIP 2009, pp 17-19, Pécs, Hungary, 2009

[30] Bojan Kuljić, János Simon, Tibor Szakáll, "PathfindiBased on Edge Detection and Infrared Distance Measuring Sensor", Acta Polytechnica Hungarica, Vol. 6, No. 1, pp 103-116, ISSN 1785-8860, 2009

[31] Matijevics Istvan, Simon János, "Control of the Greenhouse's Microclimatic Condition using Wireless Sensor Network" The IPSI BgD Transactions on Internet Research, Vol. 6, Number 2, pp. 35-38, Belgrade, Serbia, 2010

[32] Sárosi J., Gyeviki J., Véha A., Toman P., "Accurate Position Control of PAM Actuator in LabVIEW Environment", Proceedings of the IEEE SISY 2009, 7th International Symposium on Intelligent Systems and Informatics, ISBN 978-1-4244-5349-8, pp. 301-305, Subotica, Serbia, 25-26 Sept., 2009

[33] Sun™ Small Programmable Object Technology (Sun SPOT) Owner's Manual Release 3.0, Sun Microsystems, Inc. 2007

Appliance of Neural Networks in Basketball Scouting

Zdravko Ivanković[1], Miloš Racković[2], Branko Markoski[3], Dragica Radosav[4], Miodrag Ivković[5]

[1] University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia, zdravko@tfzr.uns.ac.rs

[2] University of Novi Sad, Faculty of Sciences, Institute of Mathematics, Novi Sad, Serbia, rackovic@dmi.uns.ac.rs

[3] University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia, markoni@uns.ac.rs

[4] University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia, dolores023@nadlanu.com

[5] University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia, misa.ivkovic@gmail.com

Abstract: These days, data mining is frequently used as a technology for analysing the huge amounts of data collected in sport. Basketball is one of most popular sports. Due to its dynamics, a large number of events happen during a single game. Basketball statisticians have the task of noting as many of these events as possible, in order to provide for their analysis. These data are collected by special software applications. In this paper, we used data from the First B basketball league for men in Serbia, for seasons 2005/06, 2006/07, 2007/08, 2008/09 and 2009/2010. During these five seasons, a total of 890 games were played. These data were analyzed using the feedforward technique in neural networks, which is the most often used technique in analyzing nonlinear sports data. As a final result, we concluded that the most important elements in basketball are two-point shots under the hoop and the defensive rebound, i.e. game "in the paint".

Keywords: neural networks; data mining; basketball

1 Introduction

Data mining in sport has seen an abrupt rise in recent years [10]. The developed tools and techniques have the aim to measure performances. These methods attract the attention of biggest sports associations because large amounts of money are involved in modern sports.

Before the occurrence of data mining and all its advantages, sports associations almost solely used the knowledge and expertise of people responsible for scouting. As the amount of data collected began increasing, the aim was to find more practical methods to extract knowledge from raw data. In the beginning, this led to hiring statisticians who could enable better performance measurements for a given organization and therefore more correct decision-making. The next step was to find even more practical methods; that is, to start using data mining techniques.

In order to apply data mining, data are needed. Depending on amount and richness of the available data, it is possible to extract the appropriate knowledge. Basketball is a complex game between two teams trying to score more points and win. During a game, a large number of events occur and influence the game. There are shots, offensive and defensive rebounds, turnovers, steals, blocks and assists [14], [15].

In this paper, data mining has been done using data collected in the Serbian men's First B Basketball league. Data were collected for the games in five seasons: 2005/06, 2006/07, 2007/08, 2008/09 and 2009/2010. These are official data, collected by the statistics team of the Basketball Federation of Serbia. At the end of every game, statistician send data to the Federation, and they are accumulated in a common database.

In this paper, several types of analyses were done. In analyzing the influence of shooting from different court positions, the court was divided into eleven positions: six for two-point and five for three-point score. After that, the influence of the basic basketball parameters on winning the game was analyzed. These parameters were: one-, two- and three-point shots, defensive and offensive rebounds, steals and turnovers, blocks and assists[12]. Data mining analysis is done by neural networks [1].

The paper is divided in five sections. The first section is the introduction. In section two we give some basics about data mining. Section three deals with the application of data mining in different sports. Section four contains data analysis through all six phases of data mining CRISP-DM process on concrete basketball data from First B Basketball league of Serbia. Section five contains conclusions about the data mining results.

2 Data Mining

Data mining is the analysis of an (mostly large) observed data set in order to find positive connections and to sum up data in a new way which will be both understandable and useful to the data user [2]. It can also be said that data mining is an interdisciplinary branch that encompasses techniques such as machine learning, pattern recognition, statistics, databases and visualization, all in order to provide answers to obtaining information from large databases [3].

A special intra-industrial standard was developed for data mining, independent from industry type, tool and application. The intra-industrial standard process for data mining (CRISP-DM) was developed in 1996 by analysts representing DaimlerChrysler, SPSS and NCR. The CRISP is a non-profit standard freeware, enabling the fitting of data mining to general problem-solving strategies for business and research purposes. According to CRISP-DM, a data mining project has a life span consisting of six phases. Those phases are:

- Business/Research Understanding Phase
- Data Understanding Phase
- Data Preparation Phase
- Modeling Phase
- Evaluation Phase
- Deployment Phase

The flow of the changes; i.e. the next phase often depends on output from previous phase [4].

3 Appliance of Data Mining in Sport

Sport is full of data. These data may show the individual qualities of a certain player, the events that happened during a game, and/or how a team is functioning as a whole. It is important to determine which data should be stored and a way to maximally use them [5]. By finding the right way to extract sense from these data and to transform them into practical knowledge, sports organizations provide themselves an advantage in comparison to other teams [6].

Different sports associations have different attitudes to data. This approach may be divided into five levels [10]:

- There is no connection between sports data and their use
- The experts from a given field work on predictions using their instincts and hunches
- The experts from a given field work on predictions using data collected
- The use of statistics in the decision-making process
- The use of data mining in the decision-making process.

The first type of approach is when there is no connection between sports data and their use. These sport organizations often obtain certain data about players on their games and they ignore these. This is characteristic for amateur sports organizations, since their emphasis is on fun or on introducing the sports basics.

The next type of approach is based on an expert from a given field who is predicting based on his own experience. It used to be a widely accepted notion that these experts (coaches, managers, scouts) could efficiently use their insights and experience in order to reach the correct decisions. Decisions made from this type of approach are usually based on predictions or instincts, and not on real data and information. These decisions may include taking certain types of actions or making certain player changes because such a decision "looks right".

The third type of approach is when experts start using collected data. Decisions at this level include playing certain players, for whom is has been proven that they cooperate well and are taking actions that score points more often.

The fourth type includes statistics in the decision-making process. These statistical measurements may be simple, such as the measurements of the frequency of certain events; or complex, which divides the performance of a whole team and assigning merits to each player in given game or a league. Statistics is used as a tool, helping experts in making correct decisions.

The fifth type uses data mining techniques. They do have potential to help predictions. Statistical techniques are still at the core of data mining, but statistics is used to separate a pattern or any other behavior of interest (the tendencies of opponent players) from the background noise. Statisticians do not explain relations within data, since this is a point of data mining. This type of approach has a potential to be used in order to help experts to make appropriate decision or to be used independently in order to make decisions without experts. Most sports organizations use the third or fourth type of approach between data and their use, and only a few use data mining techniques. Although data mining was introduced in sport relatively recently, the results of teams who apply these techniques are exceptional [7].

Although the use of statistics in the decision-making process is certainly a step up from decision-making based on instinct, statistics can also lead decisions in a wrong direction if there is no deep understanding of a problem. This tendency of statistics may be a consequence of imprecise performance measurements or of insistence by the sports community on certain characteristics [8]. This we may support by the fact that a certain player may have exceptional individual statistics but may still have small influence on the team as a whole.

4 Analysis of Basketball Data

When data from a basketball game are collected, the next step is to find knowledge in this information. A number of different statistical methods may be applied to basketball, which is a sport full of action and therefore rich in data.

Data analysis and predictions have been the aim for numerous individuals and organizations for a long time. This motivation leads to a number of sport-oriented developments, such as statistic simulation and machine learning techniques [9]. Neural networks are the most dominant system in machine learning used in sports [10]. Using neural networks, data sets collected from games are analyzed in order to find patterns and tendencies due to competition and financial gain. Other techniques are genetic algorithms, ID3, C4.5 and C5.0 decision trees, and a regressive variant of the Support Vector Machine (SVM) classificator called the Support Vector Regression (SVR). Data analysis, which is a subject of this paper, was done according to the CRISP-DM standard, encompassing several phases.

4.1 Business/Research Understanding Phase

Basketball is a competition between two teams with the aim to win. A win is attained by scoring more points than the other team. Some coaches like to say that the aim is to give up less points than the opponent, i.e. that the game is won by defense. In both cases, the winner is decided by the number of points scored.

Shots may be scored in several ways, and they bring a different number of points. The hardest to achieve are long-distance shots, so they bring the most points. On the floor, there is a line at 625 centimeters from a basket, and shots from outside this line bring three points (in some leagues this boundary is moved even further from the basket). Within this line, every score brings two points. During a game, it happens that a player is irregularly disrupted by opponent players, and this is called a foul. If the foul is done during an attempt to score, or if the team committing the foul have already exceeded the limit (four fouls committed during a period, or quarter), then a player has a chance to score from the free-throw line. Every shot scored from this line brings one point. Depending on whether a foul was committed while a player was trying to score two or three points, he will have opportunity to try two or three free throws, respectively.

When shooting for two points, three points and when throwing free throws, a player may make the shot or miss the shot, i.e. to score or to miss. The relation between shots and scores is called shooting percentage. In basketball statistics, there are separate percents for one-point shots, two-point shots and three-point shots.

In the Serbian First B basketball league for men, the floor is divided into six zones so that measurements of shooting percentage for two points is possible from six positions, while three-point scores measurements are possible from five positions (there are no three-point shots from a position underneath the basket). When shooting for two points, it is not the same if the shot is done underneath the hoop or from some distance. In addition, when shooting for a three-point score, it is important whether shots are tried from a position in front of a basket (position 2) or from court corners (positions 4 and 6) [13]. Figure 1 shows the division of the court into zones.

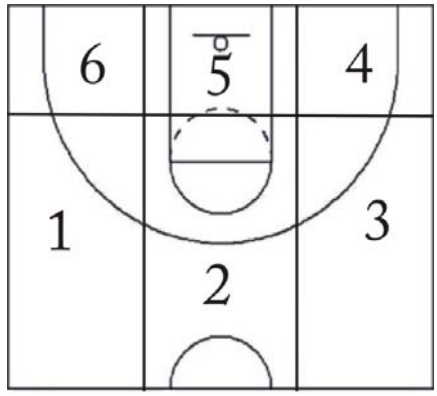

Figure 1
Basketball court

4.2 Data Understanding Phase

For keeping statistics in the Serbian First B basketball league of for men, Basketball Supervisor (BSV) software is used. This program enables entering all relevant data for a basketball game. At the end of every period, statistics recorded by this program are printed and distributed to host and guest players, commissioners of the game, and to the media. After the end of the game, collected data are sent to the Basketball Federation of Serbia where they are stored for further analysis. In this paper, statistics are analyzed for all games of the Serbian First B basketball league for men in seasons 2005/06, 2006/07, 2007/08, 2008/09 and 2009/10.

The database is organized in such a way that data regarding a shot are entered into the table *game_stat*. This table is shown in Figure 2. Here we will discuss fields P1OK and P1ALL, representing successful free throws and the total number of shots from the free-throw line, respectively. Fields P2xOK and P2xALL (x=1,2,3,4,5,6) are successful two-point shots from six different positions on the court, and the total number of two-point shots from those positions. Finally, P3yOK and P3yALL (y=1,2,3,4,6) are data regarding three-point shots from five different positions. These positions are shown in Figure 2.

The *game_stat* table encompasses all data regarding shots. It does not contain the final results games, nor who won. These data are located in a table called *game*. This table is also shown in Figure 2. The parameters of interest for use in this table are RESHOME and RESGUEST, representing the number of points scored by the host and the guest team, respectively.

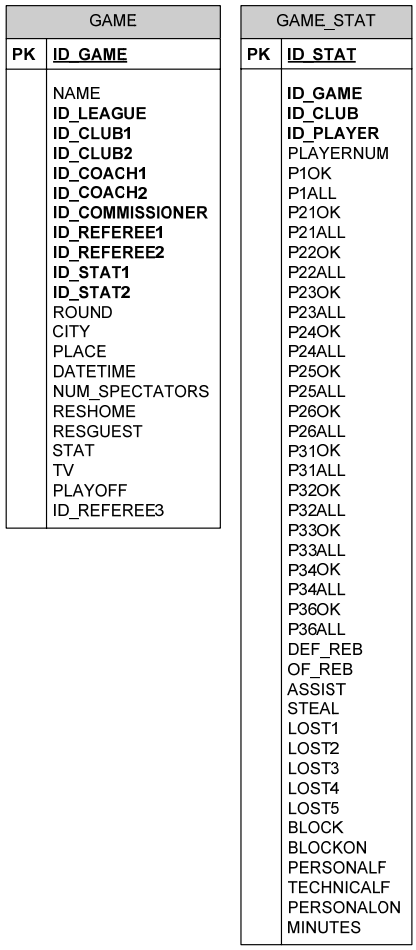

Figure 2
Data organization in the database

4.3 Data Preparation Phase

Data within the database are connected to individual players. In the analysis that is a subject of this paper, we are interested in comparing the influence of shot precision for one-, two- and three-point shots on the wins of the team observed. Therefore, we need to summarize the data regarding players in order to obtain data for a team as a whole.

Before summarizing the data, we will merge the tables *game_stat* and *game*. The merging will be done using the attribute *ID_GAME* so that in every row observed we will have shots from different positions and the result data as well.

Since we are interested in the shot percent from certain positions, we will divide the values for successful shots from a position to the values representing the total number of shots. The appearance of the SQL command that selects the appropriate data is as follows:

```
select
        if(sum(p1all) != 0, sum(p1ok)/sum(p1all), 0) p1_percent,
        if(sum(p21all) != 0, sum(p21ok)/sum(p21all), 0) p21_ percent,
        if(sum(p22all) != 0, sum(p22ok)/sum(p22all), 0) p22_ percent,
        if(sum(p23all) != 0, sum(p23ok)/sum(p23all), 0) p23_ percent,
        if(sum(p24all) != 0, sum(p24ok)/sum(p24all), 0) p24_ percent,
        if(sum(p25all) != 0, sum(p25ok)/sum(p25all), 0) p25_ percent,
        if(sum(p26all) != 0, sum(p26ok)/sum(p26all), 0) p26_ percent,
        if(sum(p31all) != 0, sum(p31ok)/sum(p31all), 0) p31_ percent,
        if(sum(p32all) != 0, sum(p32ok)/sum(p32all), 0) p32_ percent,
        if(sum(p33all) != 0, sum(p33ok)/sum(p33all), 0) p33_ percent,
        if(sum(p34all) != 0, sum(p34ok)/sum(p34all), 0) p34_ percent,
        if(sum(p36all) != 0, sum(p36ok)/sum(p36all), 0) p36_ percent,
        if(id_club1=id_club,
                if(reshome > resguest, 'win', 'loss'),
                if(reshome < resguest, 'win', 'loss')) result
from yubadata.game g, yubadata.game_stat gs
where g.id_game = gs.id_game
group by g.id_game, gs.id_club;
```

After executing the inquiry, we obtain the data as the shot percent for one point, the shot percent for two points at six positions, and the shot percent for three points from five positions, as well as the information on whether a team won or lost the game. The total number of rows obtained is 1780. This means that in five seasons, a total of 1780/2=890 games were played, and for each one we have data for the host and the guest team.

4.4 Modeling Phase

Modeling was done using neural networks. The input parameters for the network are *p1_percent, p21_percent, p22_percent, p23_percent, p24_percent, p25_percent, p26_percent, p31_percent, p32_percent, p33_percent, p34_percent* and *p36_percent*. The output parameter is *result*. Therefore, the network will have twelve input nodes and one output node. In addition to these, the network has one hidden layer. The neural network used was a feedforward neural network. Every layer within the network is fully connected to all nodes in the previous level, as well as in the next level of the network.

Network training is done by the error backward propagation method, based on the generalized delta rule. For every record brought to the network during training, information goes forward through a network in order to give a prediction in the output layer. This prediction is compared to the real output value of the

information observed, and the difference between the real and the predicted value is returned backwards through the network in order to adjust difficulty factors and improve the prediction of the next records [9], [11].

During training, the input data set was divided into 75:25 ratios, to data that will be used during network training and to data that will be used for testing. This was done in order to prevent the network from memorizing inputs and losing its robustness.

After the execution of the network training on input data, we obtain the following relation: how a certain parameter influences the final outcome of a game. Table 1 shows the influence of certain parameters on output. This relation may be shown by a histogram as in Figure 3.

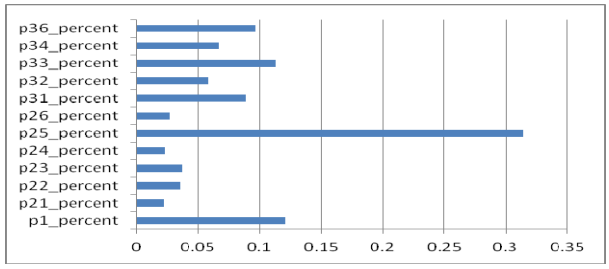

Figure 3
Influence of basketball parameters on result

In the table, it is visible that the two-point shots from position five, underneath the basket, had the highest influence on winning the game. Somewhat lower influence was the one-point shots and then the three-point shots. Two-point shots from other positions had the least influence.

Table 1
Influence of basketball parameters on result

Variable	Influence
p1_percent	0.121
p21_percent	0.022
p22_percent	0.035
p23_percent	0.037
p24_percent	0.023
p25_percent	0.314
p26_percent	0.027
p31_percent	0.089
p32_percent	0.058
p33_percent	0.113
p34_percent	0.067
p36_percent	0.096

4.5 Model Evaluation Phase

Using the data mining model, created by a neural network, we may use the testing data set in order to establish model quality in predicting the output variable. If all results of games and all predictions are reviewed, we obtain that 1182 from 1780 outcomes were correct, which is 66.4% from the total number of input data. On the other hand, the model did not correctly predict the outcome in 598 cases, or 33.6%. These results are shown in Table 2.

Table 2
Prediction results

True	1182	66.4%
False	598	33.6%
Total	1780	

From 890 wins documented (at every game there is a winner and a loser, so the number of wins and losses is identical), the algorithm correctly predicted 651, while for 239 wins it predicted losses. Regarding losses, algorithm correctly predicted 531 from 890 losses, and for 359 losses, it predicted wins. The confidence matrix is shown in Table 3.

Table 3
Confidence matrix

	won	lost
won	651	239
lost	359	531

One of the reasons for the 64.4% exactness was that the final outcome is under the influence of other parameters as well: mostly offensive and defensive rebounds, assists, steals, turnovers and blocks. In order to include these parameters in our model, we must return to data preparation stage and gather new data for modeling.

4.6 Data Preparation Phase

In the data preparation phase, we included new data in the inquiry. These data will be modeled in order to increase the model precision. These data are offensive and defensive rebounds, assists, steals, turnovers, and blocks. In addition to introducing new data, we also summed up two- and three-point shots. This means that all two-point shot percents are now located under a single field (and the same goes for three-point shots).

```
select
        if(sum(p1all) != 0, sum(p1ok)/sum(p1all), 0) p1_percent,
          (sum(p21ok)+sum(p22ok)+sum(p23ok)+
           sum(p24ok)+sum(p25ok)+sum(p26ok))/
          (sum(p21all)+sum(p22all)+sum(p23all)+
```

```
        sum(p24all)+sum(p25all)+sum(p26all)) p2_percent,
       (sum(p31ok)+sum(p32ok)+sum(p33ok)+
        sum(p34ok)+sum(p36ok))/
       (sum(p31all)+sum(p32all)+sum(p33all)+
        sum(p34all)+sum(p36all)) p3_percent,
        sum(def_reb) def_reb,
        sum(of_reb) of_reb,
        sum(assist) assist,
        sum(steal) ukradena,
        sum(lost1)+sum(lost2)+sum(lost3)+sum(lost4)+sum(lost5) lost,
        sum(block) block,
      if(id_club1=id_club,
        if(reshome > resguest, 'win', 'loss'),
        if(reshome < resguest, 'win', 'loss')) result
from yubadata.game ut, yubadata.game_stat gs
where g.id_game = gs.id_game
group by g.id_game, gs.id_club;
```

4.7 Modeling Phase

Modeling was done using neural networks. The input parameters are *p1_percent*, *p2_percent* i *p3_percent, of_reb, def_reb, assist, steal, lost* (turnover) and *block*. The output parameter is *result*. Therefore, the network has nine input nodes and one output node. In addition to these, the network has one hidden layer.

During network training, the input data set was divided into 75:25 ratios, to data that will be used in the network training and data that will be used for testing.

After the execution of the network training on input data, we obtain the following relation: how a certain parameter influences the final outcome of a game. Table 4 shows the influence of certain parameters on output. This result may be shown by the histogram in Fig. 4.

Table 4
Influence of basketball parameters on result

Variable	Influence
p1_percent	0.0796
p2_percent	0.1558
p3_percent	0.1535
def_reb	0.1588
of_reb	0.1214
assist	0.0223
steal	0.1253
lost	0.1239
block	0.0594

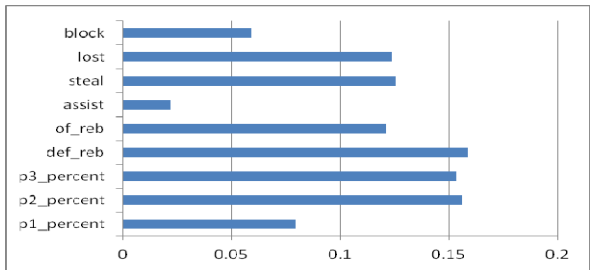

Figure 4

Influence of basketball parameters on result

From the graph it is visible that the defensive rebound has the highest influence on wining. The two-point shot has a somewhat lower influence, then the three-point shot, and the number of assists has the least influence.

4.8 Model Evaluation Phase

In this data mining model, created using a neural network, we may use again the same data set in order to see how good the model is in predicting the output variable. If all games results and all predictions are reviewed, we obtain that 1441 from 1780 outcomes were correct, which is 80.96% from the total number of input data. On the other hand, the model did not correctly predict the outcome in 339 cases, or 19.04%. These results are shown in Table 5.

Table 5

Prediction results

True	1441	80.96%
False	339	19.04%
Total	1780	

From 890 wins documented, the algorithm correctly predicted 680, while for 210 wins it predicted losses. Regarding losses, the algorithm correctly predicted 761 from 890 losses, and for 129 losses it predicted wins. The confidence matrix is shown in Table 6.

Table 6

Confidence matrix

	won	lost
won	680	210
lost	129	761

The model now including most relevant basketball parameters has relatively high prediction correctness for game outcomes based on input parameters. More than eighty percent of input data would correctly predict the game outcome. After this, the next phase is to apply this model to predict game outcomes.

Conclusions

"Get to know the enemy and get to know yourself; in a hundred battles you may never be in danger. If you don't know the enemy but get to know yourself, your chances of winning or losing are equal. If you know neither the enemy nor yourself, you are undoubtedly in danger in every battle."

Sun Tzu Wu, *The Art of War*, 450 BC

Data mining in sports provides serious advantages to its users. It makes it possible to apprehend all important elements of a basketball game and to extract knowledge from the data collected. In this way, teams get to know themselves, because they are able to see what they need in order to win, where they mostly make mistakes, which elements of the game call for improvements... Moreover, using basketball analysis and data mining as the highest level of analysis, teams may know their opponents and prepare tactics for the game. Basketball scouting, as well as the analysis of their own and the opponent's team, has become the essential part of preparation for all games in professional leagues.

Using existing data from Basketball Federation of Serbia for seasons 2005/06, 2006/07, 2007/08, 2008/09 and 2009/10, we have done the analysis applying neural networks as a data mining method, which is often used in analyzing sports events.

The general conclusion of all the analyses is that the game under the hoop is crucial for winning the game. In defense, it is important to catch the ball after the opponent's shot and preventing them from next offensive attempt, while in offense it is most important to be precise under the hoop or to score "points in the paint".

The data collected are relevant for the First Basketball League of Serbia, and the model created may be applied to other leagues of similar quality. It is to be expected that some higher-quality leagues (such as the NBA) or lower-quality ones (such as municipal leagues), as well as junior leagues, would create somewhat different models.

The flaw in keeping statistics with most of the existing programs is that a large number of basketball elements remain undocumented. With those programs, we do not know which player guarded which opponent, how much players run in offense and defense, which moves they performed and how much the ball was passed before being shot. Programs such as BSV are used in real time, during a game, so there must be some selection as to which events to include and which not. In order to obtain more complete knowledge about the game and to discover some new patterns, we need a richer data set and new software solutions so that by subsequent appraisal of a game we document all relevant events.

References

[1] Jaroslava Zilková, Jaroslav Timko, Peter Girovsky, "Nonlinear System Control Using Neural Networks", Acta Polytechnica Hungarica, Vol. 3, Issue 4, 2006

[2] D. Hand, H. Mannila, P. Smyth, "Principles of Data Mining", *MIT Press*, Cambridge, MA, 2001

[3] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees, A. Zanasi, "Discovering Data Mining: From Concept to Implementation", *Prentice Hall*, Upper Saddle River, NJ, 1998

[4] D. Larose, "Discovering Knowledge in Data - An Introduction to Data Mining", *Wiley Interscience*, 2005

[5] K. Lyons, "Data Mining and Knowledge Discovery", *Australian Sports Commission Journals* 2(4), 2005

[6] N. O'Reilly, P. Knight 2007. "Knowledge Management Best Practices in National Sport Organizations". *International Journal of Sport Management and Marketing* 2(3): 264-280, 2007

[7] R. Stefani, "A Taxonomy of Sports Rating Systems", *IEEE Transactions on Systems, Man, and Cybernetics - Part A* 29(1): 116-120, 1999

[8] L. Fieltz and D. Scott, "Prediction of Physical Performance Using Data Mining", *Research Quarterly for Exercise and Sport* 74(1): 1-25, 2003

[9] H. I. Witten, and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition". *Morgan Kaufmann*, San Francisco, 2005

[10] R. Schumaker, O. Solieman and H. Chen, "Sports Data Mining", *Springer*, 2010

[11] Rita Lovassy, László T. Kóczy, László Gál, " Function Approximation Performance of Fuzzy Neural Networks", Acta Polytechnica Hungarica, Volume 7, Issue 4, 2010

[12] Oliver, D.: Basketball on Paper - Rules and Tools for Performance Analysis. Washington DC, 2004

[13] Markoski, B., Radosav, D., Vasiljević, P., Milošević, Z.: Košarkaška statistika. International Conference Dependability and quality management ICDQM 2009, Belgrade, Serbia, 776-781, 2009

[14] FIBA - Official Basketball Rules, 2008

[15] FIBA - Basketball Statisticians' Manual, 2008

[16] Ivankovic Z., Rackovic M., Markoski B., Radosav D., Ivkovic M., "Analysis of Basketball Games Using Neural Networks", CINTI 2010, pp. 251-256, November 18-20, 2010, Budapest, Hungary