FOOTBALL PREDICTIONS BASED ON A FUZZY MODEL WITH GENETIC AND NEURAL TUNING

A. P. Rotshtein,[a] M. Posner,[a] and A. B. Rakityanskaya[b]

UDC 681.5.015:007

A model is proposed for predicting the result of a football match from the previous results of both teams. This model underlies the method of identifying nonlinear dependencies by fuzzy knowledge bases. Acceptable simulation results can be obtained by tuning fuzzy rules using tournament data. The tuning procedure implies choosing the parameters of fuzzy-term membership functions and rule weights by a combination of genetic and neural optimization techniques.

Keywords: *football match prediction, fuzzy logic, fuzzy knowledge bases, genetic algorithm, neural fuzzy network.*

INTRODUCTION

Football is a sport game attracting the greatest number of fans. Predicting the results of football matches is interesting from two points of view: demonstrating the usefulness of various mathematical methods [1, 2] and making money by betting on one or another team.

Models and PC-programs for sport predictions have long been developed (for example, *http://dmiwww.cs.tut.fi/riku*). Most of them employ stochastic methods to describe uncertainty: regressive and autoregressive analysis [3–5], the Bayesian approach combined with Markovian chains and the Monte-Carlo method [6–9]. These models are complex, use many assumptions, require large statistical samplings, and may not always be easily interpreted. Recently, neural networks have come to be used to make football predictions [10–12]. They are considered as universal approximators of nonlinear dependences, trained by experimental data. These models also require extensive statistical data and do not allow defining the physical meaning of weights between neurons after training.

Football experts and fans frequently make predictions based on simple, common-sense assumptions, such as

IF *a team* T_1 *won all previous matches*
AND *a team* T_2 *lost all previous matches*
AND *the team* T_1 *won the previous matches between* T_1 *and* T_2,
THEN *it should be expected that* T_1 would win.

Such a reasoning concentrates experts' experience and can be formalized using fuzzy logic [13]. That is why it is quite natural to use such a reasoning as a prediction model support.

A method for identiffying nonlinear dependences by fuzzy knowledge bases is proposed in [14, 15]. Its various theoretical and practical aspects are considered in [16–19]. The aim of the present paper is to analyze the possible applications of fuzzy knowledge bases and the method [14, 15] in the prediction of football matches.

A prediction model is constructed in two stages. The first stage is to determine the structure of fuzzy model that associates the result of a football match with the previous results of both teams. To this end, we use the generalized fuzzy approximator proposed in [14, 15]. At the second stage, the fuzzy model is tuned, i.e., its optimal parameters are determined from experimental data available. The tuning procedure employs a genetic algorithm and a neural network. The genetic

[a]Jerusalem College of Technology, Machon Lev, Israel, *rot@mail.jct.ac.il; posner@mail.jct.ac.il*. [b]Vinnitsa National Technical University, Vinnitsa, Ukraine, *h_rakit@hotmail.com*. Translated from Kibernetika i Sistemnyi Analiz, No. 4, pp. 171-184, July-August 2005. Original article submitted June 17, 2002.

1060-0396/05/4104-0619

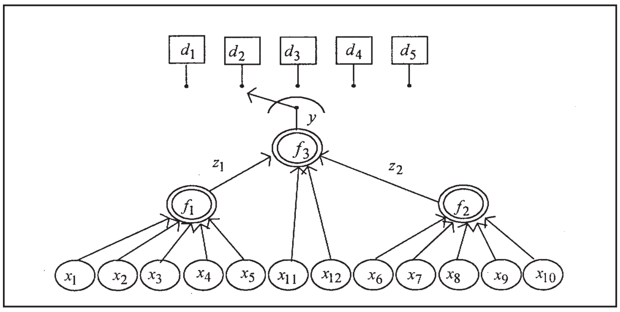


Fig. 1. Prediction model structure.

algorithm provides rough hit into the domain of global minimum of the discrepancy between the model and experimental results. The neural approach is used for fine tuning of the model parameters and their adaptive correction as new experimental data arrive.

FUZZY MODEL

Model Structure

The purpose of simulation is to predict the result of a match between teams T_1 and T_2, which is characterized by a difference y of goals scored and goals conceded. Assume that $y \in [\underline{y}, \bar{y}] = [-5, 5]$. To construct a prediction model, let us define y at the following five levels:

d_1 is a high-score loss (*big loss* – *BL*), $y = -5, -4, -3$;
d_2 is a low-score loss (*small loss* – *SL*), $y = -2, -1$;
d_3 is a drawn game (*draw* – *D*), $y = 0$;
d_4 is a low-score win (*small win* – *SW*), $y = 1, 2$;
d_5 is a high-score win (*big win* – *BW*), $y = 3, 4, 5$.

Let the following factors influence the result of a match (y):
$x_1, x_2, \ldots, x_5$ are the results of the five previous matches of the team T_1;
$x_6, x_7, \ldots, x_{10}$ are the results of the five previous matches of the team T_2;
x_{11}, x_{12} are the results of the two previous matches between the teams T_1 and T_2.

Obviously, $x_1, x_2, \ldots, x_{12}$ vary from -5 to $+5$. Hierarchical correlation between the output variable y and input variables $x_1, x_2, \ldots, x_{12}$ is shown in Fig. 1 in the form of a tree.

This tree is equivalent to the following system of relationships:

$$y = f_3(z_1, z_2, x_{11}, x_{12}), \tag{1}$$

$$z_1 = f_1(x_1, x_2, \ldots, x_5), \tag{2}$$

$$z_2 = f_2(x_6, x_7, \ldots, x_{10}), \tag{3}$$

where z_1 and z_2 are intermediate variables: $z_1(z_2)$ is a predicted result for the team T_1 (T_2) based on the previous results $x_1, x_2, \ldots, x_5$ ($x_6, x_7, \ldots, x_{10}$).

We will consider the variables $x_1, x_2, \ldots, x_{12}, z_1$, and z_2 to be linguistic variables [13], which can be estimated using the fuzzy terms introduced above: BL, SL, D, SW, and BW.

To describe relations (1)–(3), we will use expert knowledge matrices (Tables 1 and 2). These matrices correspond to IF-THEN fuzzy rules formulated based on common-sense reasoning. An example of such a rule for Table 2 is as follows:

IF $(x_{11} = BW)$ AND $(x_{12} = BW)$ AND $(z_1 = BW)$ AND $(z_2 = BL)$
OR $(x_{11} = SW)$ AND $(x_{12} = BW)$ AND $(z_1 = SW)$ AND $(z_2 = D)$
OR $(x_{11} = BW)$ AND $(x_{12} = D)$ AND $(z_1 = BW)$ AND $(z_2 = SL)$
THEN $y = d_5$.

TABLE 1. Expert Knowledge Matrices for Relations (2) and (3)

$x_1(x_6)$	$x_2(x_7)$	$x_3(x_8)$	$x_4(x_9)$	$x_5(x_{10})$	$z_1(z_2)$
BL	*BL*	*BL*	*BL*	*BL*	
BW	*SL*	*BL*	*SL*	*BW*	*BL*
SW	*BL*	*SL*	*SL*	*SW*	
SL	*SL*	*SL*	*SL*	*SL*	
D	*SL*	*SL*	*D*	*D*	*SL*
SW	*D*	*SL*	*SL*	*SW*	
D	*D*	*D*	*D*	*D*	
SL	*SW*	*SW*	*D*	*SL*	*D*
D	*D*	*SW*	*SW*	*D*	
SW	*SW*	*SW*	*SW*	*SW*	
D	*BW*	*BW*	*SW*	*D*	*SW*
SL	*SW*	*SW*	*BW*	*SL*	
BW	*BW*	*BW*	*BW*	*BW*	
SL	*BW*	*SW*	*BW*	*SL*	*BW*
BL	*SW*	*BW*	*SW*	*BL*	

TABLE 2. Expert Knowledge Matrices for Relation (1)

x_{11}	x_{12}	z_1	z_2	y
BL	*BL*	*BL*	*BW*	
BW	*D*	*BL*	*D*	d_1
SW	*BL*	*SL*	*SL*	
SW	*SL*	*D*	*SL*	
D	*SL*	*SL*	*D*	d_2
SW	*D*	*SL*	*SL*	
D	*D*	*D*	*D*	
SL	*SW*	*SW*	*D*	d_3
SL	*D*	*SW*	*SW*	
SL	*SW*	*SW*	*BW*	
D	*BW*	*BW*	*SW*	d_4
SL	*SW*	*SW*	*BW*	
BW	*BW*	*BW*	*BL*	
SW	*BW*	*SW*	*D*	d_5
BW	*D*	*BW*	*SL*	

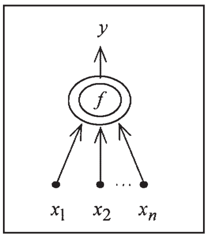


Fig. 2. Generalized fuzzy approximator.

TABLE 3. Expert Knowledge Matrix

Rule No.	**IF <inputs>**				**THEN <output>**	**Weight of the rule**
	x_1	x_2	…	x_n	y	
11	a_1^{11}	a_2^{11}	…	a_n^{11}		w_{11}
12	a_2^{1}	a_2^{1}	…	a_n^{1}	d_1	w_{12}
…	…	…	…	…		…
$1k_1$	$a_1^{1k_1}$	$a_2^{1k_1}$	…	$a_n^{1k_1}$		w_{1k_1}
…	…	…	…	…	…	…
$m1$	a_1^{m1}	a_2^{m1}	…	a_n^{m1}		w_{m1}
$m2$	a_1^{m2}	a_2^{m2}	…	a_n^{m2}	d_m	w_{m2}
…	…	…	…	…		…
mk_m	$a_1^{mk_m}$	$a_2^{mk_m}$	…	$a_n^{mk_m}$		w_{mk_m}

Fuzzy Approximator

To apply fuzzy knowledge bases (Tables 1 and 2) we will use the generalized fuzzy approximator (Fig. 2) introduced in [14, 15].

This approximator describes the dependence $y = f(x_1, x_2, \ldots, x_n)$ between the inputs $x_1, x_2, \ldots, x_n$ and the output y using an expert knowledge matrix (Table 3).

The matrix is associated with a fuzzy knowledge base:

$$\text{IF } [(x_1 = a_1^{j1}) \text{ AND } \ldots (x_i = a_j^{j1}) \text{ AND } \ldots (x_n = a_n^{j1})]$$

$$(\text{with weight } w_{j1}) \ldots$$

$$\ldots \text{ OR } [(x_1 = a_1^{jk_j}) \text{ AND } \ldots (x_i = a_i^{j\,k_j}) \text{ AND } \ldots$$

$$(x_n = a_n^{jk_j})] \text{ (with weight } w_{jk_j}), \tag{4}$$

$$\text{THEN } y = d_j, \quad j = \overline{1, m},$$

where a_i^p is a linguistic term evaluating a variable x_i in the row $p=k_j$; k_j is the number of conjunction rows corresponding to the class d_j of the output variable y; w_{jp} is a number from the interval [0,1], characterizing the subjective measure of expert's confidence as to a statement with the number $p=k_j$.

Classes d_j, $j=\overline{1,m}$, are formed by quantization of the range $[\underline{y}, \overline{y}]$ of the output variable into m levels:

$$[\underline{y}, \overline{y}] = \underbrace{[\underline{y}, y_1)}_{d_1} \cup \ldots \cup \underbrace{[y_{j-1}, y_j)}_{d_j} \cup \ldots \cup \underbrace{[y_{m-1}, \overline{y}]}_{d_m}.$$

According to [14–16], the following object approximation corresponds to fuzzy knowledge base (4):

$$y = \frac{\underline{y}\,\mu^{d_1}(y) + y_1\,\mu^{d_2}(y) + \ldots + y_{m-1}\,\mu^{d_m}(y)}{\mu^{d_1}(y) + \mu^{d_2}(y) + \ldots + \mu^{d_m}(y)}, \tag{5}$$

$$\mu^{d_j}(y) = \max_{p=\overline{1,k_j}} \left\{ w_{jp} \min_{i=\overline{1,n}} \left[\mu^{jp}(x_i) \right] \right\}, \tag{6}$$

$$\mu^{jp}(x_i) = \frac{1}{1+\left(\frac{x_i - b_i^{jp}}{c_i^{jp}} \right)^2}, \quad i=\overline{1,n}, \quad j=\overline{1,m}, \quad p=k_j, \tag{7}$$

where $\mu^{d_j}(y)$ is the membership function of the output y in the class $d_j \in [y_{j-1}, y_j]$; $\mu^{jp}(x_i)$ is the membership function of a variable x_i in a term a_i^p; b_i^{jp}, and c_i^{jp} are settings of membership functions with the following interpretation: b is the coordinate of maximum, $\mu^{jp}(b_i^{jp})=1$; and c is the concentration (contraction–extension) parameter.

Relations (5)–(7) determine the generalized model of the nonlinear function $y=f(x_1, x_2, \ldots, x_n)$ as

$$y = F(X, W, B, C), \tag{8}$$

where $X=(x_1, x_2, \ldots, x_n)$ is the vector of input variables, $W=(w_1, w_2, \ldots, w_N)$ is the vector of weights of fuzzy rules, $B=(b_1, b_2, \ldots, b_q)$ and $C=(c_1, c_2, \ldots, c_q)$ are the vectors of parameters of membership functions, N is the total number of rules, q is the total number of fuzzy terms, and F is the inputs-output operator corresponding to formulas (5)–(7).

Fuzzy Model of Prediction

Using the fuzzy approximator (8) (Fig. 2) and the derivation tree (Fig. 1), we can describe the prediction model as

$$y = F_y(x_1, x_2, \ldots, x_{12}, W_1, B_1, C_1, W_2, B_2, C_2, W_3, B_3, C_3), \tag{9}$$

where F_y is the inputs-output operator corresponding to formulas (1)–(3), $W_1 = ((w_1^{11}, \ldots, w_1^{13}), \ldots, (w_1^{51}, \ldots, w_1^{53}))$, $W_2 = ((w_2^{11}, \ldots, w_2^{13}), \ldots, (w_2^{51}, \ldots, w_2^{53}))$, and $W_3 = ((w_3^{11}, \ldots, w_3^{13}), \ldots, (w_3^{51}, \ldots, w_3^{53}))$ are the vectors of rule weights in relations (2), (3), and (1), respectively;

$$B_1 = (b_{1-5}^{BL}, b_{1-5}^{SL}, b_{1-5}^{D}, b_{1-5}^{SW}, b_{1-5}^{BW}),$$

$$B_2 = (b_{6-10}^{BL}, b_{6-10}^{SL}, b_{6-10}^{D}, b_{6-10}^{SW}, b_{6-10}^{BW}),$$

$$B_3 = (b_{11,12}^{BL}, b_{11,12}^{SL}, b_{11,12}^{D}, b_{11,12}^{SW}, b_{11,12}^{BW})$$

are the vectors of the centers of membership functions of the variables $x_1, x_2, \ldots, x_5$, $x_6, x_7, \ldots, x_{10}$ and x_{11}, x_{12} in the terms BL, SL, …, BW;

$$C_1 = (c_{1-5}^{BL}, c_{1-5}^{SL}, c_{1-5}^{D}, c_{1-5}^{SW}, c_{1-5}^{BW}),$$

$$C_2 = (c_{6-10}^{BL}, c_{6-10}^{SL}, c_{6-10}^{D}, c_{6-10}^{SW}, c_{6-10}^{BW}),$$

$$C_3 = (c_{11,12}^{BL}, c_{11,12}^{SL}, c_{11,12}^{D}, c_{11,12}^{SW}, c_{11,12}^{BW})$$

are the vectors of concentration parameters of the membership functions of variables $x_1, x_2, \ldots, x_5, x_6, x_7, \ldots, x_{10}$ and x_{11}, x_{12} in the terms $BL, SL, \ldots, BW$.

In constructing model (9), we assumed that the fuzzy terms $BL, SL, \ldots, BW$ have identical membership functions for each of the variables $x_1, x_2, \ldots, x_5$. The same assumption is made for the variables $x_6, x_7, \ldots, x_{10}$, x_{11}, and x_{12}.

FORMULATION OF THE PROBLEM OF FUZZY MODEL TUNING

Let a learning sample as M pairs of experimental data of the form

$$\langle \hat{X}_l, \hat{y}_l \rangle, \; l = \overline{1, M},$$

be composed from the tournament data, where $\hat{X}_l = \{(\hat{x}_1^l, \hat{x}_2^l, \ldots, \hat{x}_5^l), (\hat{x}_6^l, \hat{x}_7^l, \ldots, \hat{x}_{10}^l), (\hat{x}_{11}^l, \hat{x}_{12}^l)\}$ are the results of previous matches for the teams T_1 and T_2 in the lth experiment, and $\hat{y}_l$ is the result of a match between the teams T_1 and T_2 in the lth experiment.

Essentially, tuning of the prediction model is to select the parameters of the membership functions (b-, c-) and weights of fuzzy rules (w-) so as to provide minimum discrepancy between theoretical and experimental data:

$$\sum_{l=1}^{M} (F_y(\hat{x}_1^l, \hat{x}_2^l, \ldots, \hat{x}_{12}^l, W_i, B_i, C_i) - \hat{y}_l)^2 = \min_{W_i, B_i, C_i}, \quad i = 1, 2, 3. \tag{10}$$

To solve the nonlinear optimization problem (10), we combine a genetic algorithm and a neural network. The genetic algorithm provides rough off-line hit into the domain of global minimum, and the neural network is used for on-line improvement of parameter values.

GENETIC TUNING OF THE FUZZY MODEL

Structure of the Algorithm

To implement the genetic algorithm for solving the optimization problem (10), it is necessary to define the following main concepts and operations [15, 20]: a *chromosome* means a coded version of the solution; a *population* means the initial set of versions of the solution; *fitness function* means a selection criterion; *crossover* means generation of daughter chromosomes by parent chromosomes; and a *mutation* means random change of chromosome elements.

If $P(t)$ are parent chromosomes, and $C(t)$ are daughter chromosomes at the tth iteration, then the general structure of the genetic algorithm is as follows:

begin
    $t := 0$;
    Specify the initial value of $P(t)$;
    Estimate $P(t)$ using the fitness function;
    while (**not** termination conditions) **do**
        Cross $P(t)$ to produce $C(t)$;
        Mutate $C(t)$;
        Estimate $C(t)$ using the fitness function;
        Select $P(t+1)$ from $P(t)$ and $C(t)$;
        $t := t + 1$;
    end
end

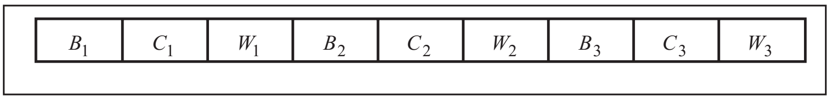


Fig. 3. Structure of a chromosome.

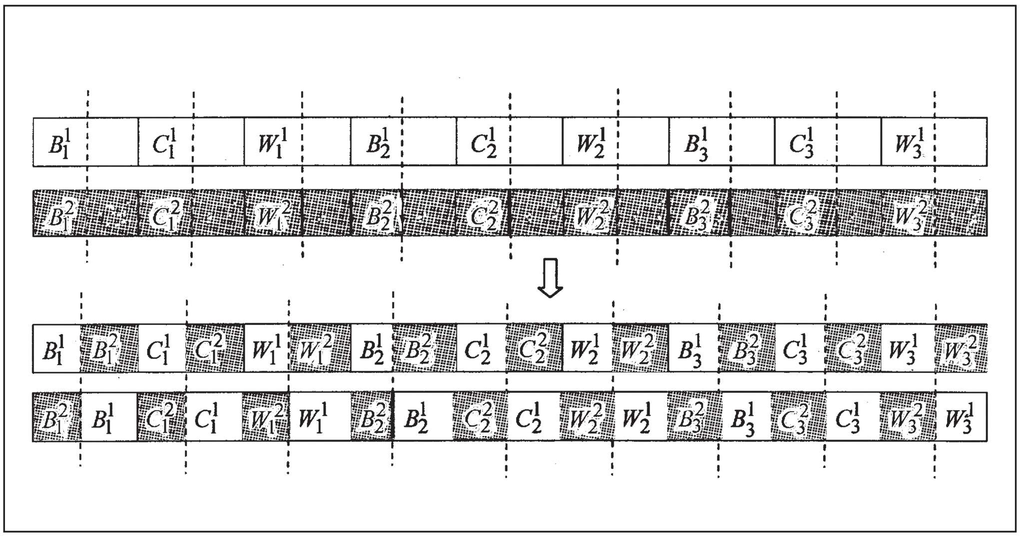


Fig. 4. Structure of the crossover operation.

Coding

Let us define a chromosome as a row vector of binary codes of the parameters of membership functions and rule weights (Fig. 3).

Crossover and Mutation

The crossover operation is defined in Fig. 4. It consists in exchange of chromosome parts in each of the vectors of membership functions ($B_1, C_1, B_2, C_2, B_3, C_3$) and each of the vectors of rule weights (W_1, W_2, W_3). The crossover points shown by dotted lines are selected arbitrarily. The superscripts (1 and 2) in the parameter vectors refer to the first and second parent chromosomes, respectively.

The mutation (Mu) provides a random change (with some probability) of chromosome elements:

$$Mu\,(w_{jp}) = RANDOM\;([0,1]),$$

$$Mu\,(b_i^{jp}) = RANDOM\;([\underline{y}, \overline{y}\,]),$$

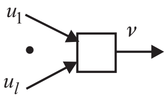
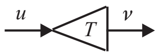
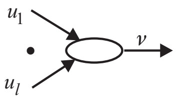
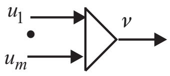
$$Mu\,(c_i^{jp}) = RANDOM\;([\,\underline{c}_i^{jp}, \overline{c}_i^{jp}\,]).$$

Here $RANDOM\;([\,\underline{x}, \overline{x}\,])$ is the operation of finding a random number uniformly distributed over the interval $[\,\underline{x}, \overline{x}\,]$.

Selection

Selection of parent chromosomes for the crossover operation should not be random. We used the selection procedure giving priority to the best solutions. The greater the fitness function of some chromosome, the higher the probability that this chromosome will generate daughter chromosomes [15, 20]. As a fitness function, we take criterion (10) with minus sign, i.e., the greater the degree of chromosome adaptability to the optimization criterion, the greater the fitness function. While the genetic algorithm is running, the population size remains constant. Therefore, after performing the crossover and mutation operations, it is necessary to delete the chromosomes with the worse values of the fitness function from the population obtained.

TABLE 4. Neural Fuzzy Network Elements

Node designation	Node name	Function	Node designation	Node name	Function
	Input	$v = u$		Class of rules	$v = \sum_{i=1}^{l} u_i$
	Fuzzy term	$v = \mu^{T}(u)$			
	Fuzzy rule	$v = \prod_{i=1}^{l} u_i$		Defuzzification ($\bar{d}_j$ is the center of class d_j)	$v = \frac{\sum_{j=1}^{m} u_j \bar{d}_j}{\sum_{j=1}^{m} u_j}$

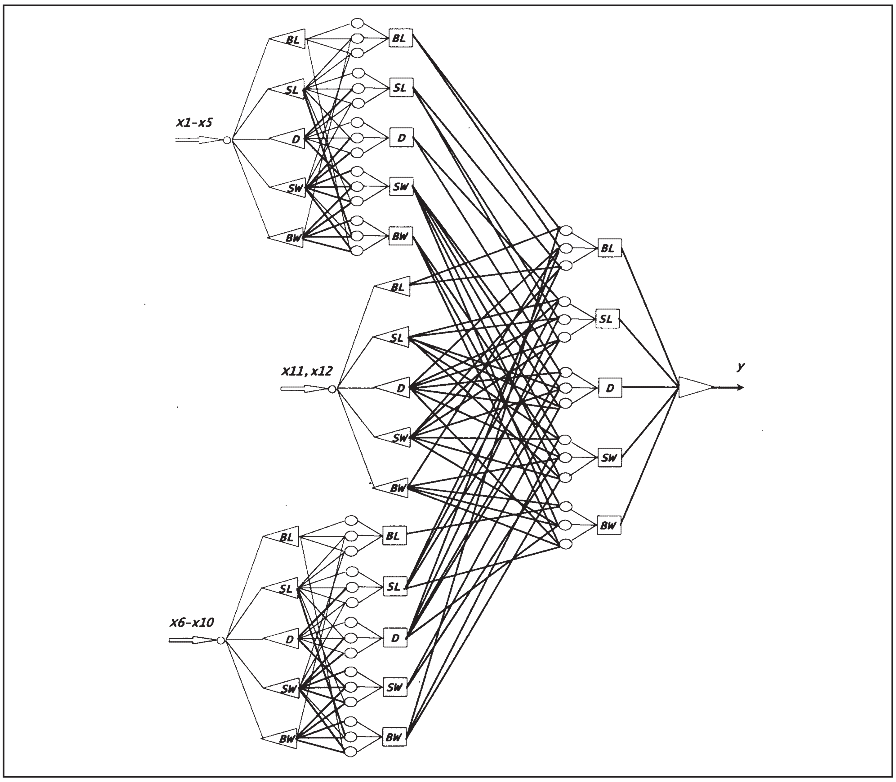


Fig. 5. Neural fuzzy prediction network.

NEURAL TUNING OF THE FUZZY MODEL

Neural Fuzzy Prediction Network

For on-line neural tuning, we implanted the IF-THEN fuzzy rules into a special neural network constructed using elements from Table 4 [18].

The neural fuzzy network obtained is shown in Fig. 5.

Relationships for the Tuning

For tuning neural fuzzy network parameters, we used the following recurrences:

$$w_{jp}(t+1) = w_{jp}(t) - \eta \frac{\partial E_t}{\partial w_{jp}(t)}, \tag{11}$$

$$c_i^{jp}(t+1) = c_i^{jp}(t) - \eta \frac{\partial E_t}{\partial c_i^{jp}(t)}, \tag{12}$$

$$b_i^{jp}(t+1) = b_i^{jp}(t) - \eta \frac{\partial E_t}{\partial b_i^{jp}(t)} \tag{13}$$

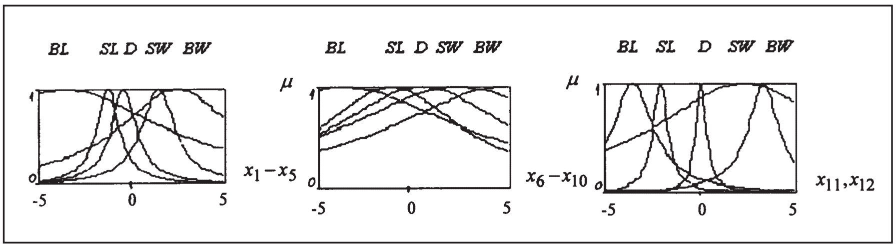


Fig. 6. Membership functions after tuning.

TABLE 5. Rule Weights in Relation (2)

Genetic tuning	Neural tuning
1.0	0.989
1.0	1.000
1.0	1.000
0.8	0.902
0.5	0.561
0.8	0.505
0.6	0.580
1.0	0.613
0.5	0.948
1.0	0.793
0.9	0.868
0.6	0.510
0.6	0.752
0.5	0.500
0.5	0.500

TABLE 6. Rule Weights in Relation (3)

Genetic tuning	Neural tuning
0.7	0.926
0.9	0.900
0.7	0.700
0.9	0.954
0.7	0.700
1.0	1.000
0.9	0.900
1.0	1.000
0.6	0.600
1.0	1.000
0.7	0.700
1.0	1.000
0.8	0.990
0.5	0.500
0.6	0.600

TABLE 7. Rule Weights in Relation (1)

Genetic tuning	Neural tuning
0.7	0.713
0.8	0.782
1.0	0.996
0.5	0.500
0.5	0.541
0.5	0.500
0.5	0.500
0.5	0.522
0.6	0.814
1.0	0.903
0.6	0.503
1.0	0.677
1.0	0.515
0.5	0.514
1.0	0.999

TABLE 8. Membership Function Parameters after Tuning

Terms	Genetic tuning						Neural tuning					
	$x_1, x_2, \ldots, x_5$		$x_6, x_7, \ldots, x_{10}$		x_{11}, x_{12}		$x_1, x_2, \ldots, x_5$		$x_6, x_7, \ldots, x_{10}$		x_{11}, x_{12}	
	b-	*c*-	*b*-	*c*-	*b*-	*c*-	*b*-	*c*-	*b*-	*c*-	*b*-	*c*-
BL	– 4.160	9	– 5.153	9	– 5.037	3	– 4.244	7.772	– 4.524	9.303	– 4.306	1.593
SL	– 2.503	1	– 2.212	5	– 3.405	1	– 1.468	0.911	– 1.450	5.467	– 2.563	0.555
D	– 0.817	1	0.487	7	0.807	1	– 0.331	0.434	0.488	7.000	0.050	0.399
SW	2.471	3	2.781	9	2.749	7	1.790	1.300	2.781	9.000	2.750	7.000
BW	4.069	5	5.749	9	5.238	3	3.000	4.511	5.750	9.000	3.992	1.234

minimizing the criterion

$$E_t = \frac{1}{2}(\hat{y}_t - y_t)^2,$$

used in neural network theory. Here y_t ($\hat{y}_t$) is the theoretical (experimental) difference between the goals scored and goals conceded at the *t*th step of learning; $w_{jp}(t)$, $c_i^{jp}(t)$, and $b_i^{jp}(t)$ are the rule weights and membership function parameters at the *t*th step of learning; and η is a learning parameter selected according to the guidelines from [21].

Models of the derivatives appearing in (11)–(13) are presented in the Appendix.

EXPERIMENTAL RESULTS

For tuning the fuzzy model, we used the tournament data for the championship of Finland, characterized by a minimum number of sensations.

TABLE 9. Fragment of Prediction Results

№	T_1	T_2	Year	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$	$x7$	$x8$	$x9$	$x10$	$x11$	$x12$	Score	$\hat{y}$	$\hat{d}$	y_G	d_G	y_N	d_N
1	Kuusysi	Reipas	1991	2	1	2	0	1	-1	0	1	-2	-3	2	1	2-0	2	d4	1	d4	1	d4
2	Ilves	PPT	1991	1	3	-1	1	0	0	2	-1	-2	0	0	0	2-1	1	d4	0	d3*	0	d3*
3	Haka	Jaro	1991	-1	2	0	-1	1	1	0	-2	-1	-2	-1	1	1-1	0	d3	0	d3	0	d3
4	MP	OTP	1991	3	1	2	0	2	-1	-2	1	-2	-3	1	3	4-0	4	d5	3	d5	3	d5
5	KuPS	HJK	1991	-1	-3	-4	1	-3	1	0	2	0	0	-2	0	1-3	-2	d2	-1	d2	-1	d2
6	TPS	RoPS	1991	3	1	2	-2	0	2	0	1	-1	1	0	-1	1-0	1	d4	0	d3*	0	d3*
7	PPT	Jaro	1991	0	-5	-1	0	1	1	2	-2	-1	1	1	-3	0-1	-1	d2	-1	d2	-1	d2
8	Haka	Reipas	1991	2	-1	3	1	4	2	-2	0	-1	0	-1	2	3-0	3	d5	2	d4*	2	d4*
9	OTP	Kuusysi	1991	-1	-2	-3	-2	0	1	3	4	-1	2	-2	-1	1-4	-3	d1	-3	d1	-3	d1
10	HJK	TPS	1991	1	1	1	0	2	0	1	-1	2	-3	0	2	2-0	2	d4	2	d4	2	d4
11	MyPa	Jaro	1992	-3	1	2	1	0	2	1	-2	-1	0	-2	0	0-0	0	d3	0	d3	0	d3
12	Jazz	Ilves	1992	2	2	1	-1	0	3	4	-1	0	1	1	-1	2-1	1	d4	0	d3*	1	d4
13	Haka	RoPS	1992	-2	-2	0	1	1	-1	1	1	1	0	1	3	1-1	0	d3	1	d4*	1	d4*
14	HJK	Oulu	1992	2	3	0	0	1	0	-5	1	-2	-1	-1	2	4-0	4	d5	2	d4*	3	d5
15	MP	Kuusysi	1992	0	1	-2	-1	-1	3	1	2	0	1	0	-2	0-3	-3	d1	-3	d1	-3	d1
16	KuPS	HJK	1992	-2	-1	-3	1	-2	4	2	1	2	1	-2	-3	0-5	-5	d1	-4	d1	-4	d1
17	Kuusysi	MP	1992	0	-1	3	2	-1	-3	2	-1	-2	0	1	0	3-1	2	d4	1	d4	1	d4
18	TPS	Haka	1992	-1	2	3	-1	-2	0	-1	0	3	1	-1	1	2-2	0	d3	0	d3	0	d3
19	RoPS	MyPa	1992	-2	-1	2	0	-1	1	-1	1	1	-2	1	-1	1-2	-1	d2	0	d3*	0	d3*
20	Jazz	Ilves	1992	-2	1	-3	5	-1	1	1	-2	0	-1	2	0	1-0	1	d4	1	d4	1	d4
21	TPS	Jaro	1992	-2	-1	2	-1	-3	1	0	2	-1	3	1	-2	0-2	-2	d2	-1	d2	-1	d2
22	Haka	MyPa	1992	1	1	-1	0	1	0	3	2	1	-1	-1	-3	0-1	-1	d2	-2	d2	-2	d2
23	HJK	RoPS	1992	1	2	0	-1	1	-1	2	2	-1	1	0	0	2-1	1	d4	0	d3*	0	d3*
24	MP	Kuusysi	1992	1	-1	-2	-3	1	1	-1	-2	2	3	-2	1	0-2	-2	d2	-1	d2	-1	d2
25	Ilves	Kups	1992	3	0	-2	2	-2	1	1	-1	0	-2	1	0	1-0	1	d4	1	d4	1	d4
26	Haka	HJK	1992	0	-2	-1	-1	0	2	3	-1	0	3	-1	-2	0-3	-3	d1	-3	d1	-3	d1
27	Jaro	MyPa	1992	-1	-1	1	2	1	-3	1	2	1	0	1	1	1-1	0	d3	1	d4*	0	d3
28	RoPS	TPS	1992	-1	1	-1	1	4	-5	-2	3	-1	-2	5	1	2-0	2	d4	2	d4	1	d4
29	MP	Ilves	1992	1	2	-1	1	0	0	1	0	0	-1	1	-2	2-3	-1	d2	-1	d2	-1	d2
30	Kuusysi	KuPS	1992	2	2	0	3	1	-1	-1	1	-3	0	2	3	4-1	3	d5	3	d5	3	d5
31	Jazz	MP	1993	2	2	2	0	3	-2	-1	0	-1	-3	4	3	5-0	5	d5	4	d5	4	d5
32	Kuusysi	TPS	1993	1	-1	0	-1	1	-2	2	0	-1	1	0	1	0-0	0	d3	0	d3	0	d3
33	MyPa	RoPS	1993	-1	-1	2	2	3	2	-1	1	2	-2	3	-1	2-0	2	d4	1	d4	1	d4
34	Haka	HJK	1993	-3	-1	-2	1	0	1	4	1	2	0	-1	-2	1-3	-2	d2	-1	d2	-1	d2
35	Jaro	Ilves	1993	2	0	-1	0	-1	-2	-1	-2	2	1	2	0	2-1	1	d4	1	d4	1	d4
36	Ilves	HJK	1993	1	-2	-1	-1	1	3	1	2	0	1	-1	-1	0-2	-2	d2	-1	d2	-1	d2
37	Jazz	Jaro	1993	2	1	0	1	5	-1	-2	-2	1	-1	2	1	3-0	3	d5	2	d4*	2	d4*
38	MyPa	MP	1993	1	3	1	-1	1	-1	0	2	-1	1	1	0	1-0	1	d4	1	d4	1	d4
39	Kuusysi	Haka	1993	-1	-2	1	1	2	-1	-3	1	-5	2	3	-1	3-1	2	d4	1	d4	1	d4
40	TPS	RoPS	1993	-1	1	-2	1	2	1	2	-1	1	-2	1	1	1-0	1	d4	1	d4	1	d4
41	MP	HJK	1993	-1	-1	0	2	-1	2	3	1	-1	1	-2	1	1-2	-1	d2	0	d3*	0	d3*
42	Kuusysi	Jaro	1993	2	2	-2	1	2	0	-1	2	-2	0	1	2	2-1	1	d4	1	d4	1	d4
43	Jazz	Haka	1993	2	3	2	-1	1	-1	-3	-4	-2	0	2	2	4-0	4	d5	3	d5	3	d5
44	FinnPa	MyPa	1993	-1	1	-2	-1	2	1	-2	-1	1	0	-1	-1	1-2	-1	d2	-1	d2	-1	d2
45	TPS	Ilves	1993	2	1	2	1	-1	2	2	-2	1	-3	0	2	2-0	2	d4	1	d4	1	d4
46	RoPS	Jazz	1993	-1	-1	2	-2	-1	4	1	5	0	2	1	-3	2-5	-3	d1	-3	d1	-3	d1
47	MyPa	Ilves	1993	5	0	2	1	1	-3	-1	-2	1	-2	3	0	5-1	4	d5	3	d5	3	d5
48	TPV	Kuusysi	1993	-2	-1	0	1	0	-1	0	2	-1	0	0	1	0-0	0	d3	0	d3	0	d3
49	RoPS	HJK	1993	-1	-1	1	-2	0	3	1	-2	1	1	-2	1	0-2	-2	d2	0	d3*	-1	d2
50	TPS	Jaro	1993	-1	-1	1	2	2	-2	-1	1	-2	1	3	1	1-0	1	d4	1	d4	1	d4

The learning sampling consists of the results of 1056 matches over eight years from 1994 to 2001. Tables 5–8 and Fig. 6 show the results of fuzzy model tuning.

For model testing, we used the results of 350 matches from 1991 to 1993. A fragment of the testing sample and prediction results is presented in Table 9, where T_1 and T_2 are team names; $\hat{y}$ and $\hat{d}$ are real and experimental results; y_G and d_G are predicted results after genetic tuning; and y_N and d_N are predicted results after neural tuning. The symbol * denotes noncoincidence of the theoretical and experimental results.

The efficiency indices of the fuzzy model tuning algorithms for the testing sample are presented in Table 10. The best prediction results are provided for extreme classes of decisions, i.e., high-score win and loss (d_1 and d_5). The worst prediction results are for low-score win and loss (d_2 and d_4).

TABLE 10. Efficiency Indices of Tuning Algorithms

Efficiency indices		Genetic tuning	Neural tuning
Tuning time (minutes)		**52**	**7**
No. of iterations		**25,000**	**5000**
Probability of correct prediction for different decisions	d_1 — a high-score loss	30/35=0.857	32/35=0.914
	d_2 — a low-score loss	64/84=0.762	70/84=0.833
	d_3 — a drawn game	38/49=0.775	43/49=0.877
	d_4 — a low-score win	97/126=0.770	106/126=0.841
	d_5 — a high-score win	49/56=0.875	53/56=0.946

CONSTRAINTS OF THE PREDICTION MODEL

The model proposed is constructed according to the time-series ideology [1, 2], which implies that the past completely determines the future. As compared with classical time series, fuzzy rules [15] have been used, which significantly decreases the number of experimental data due to expert knowledge. The input factors $(x_1, x_2, \ldots, x_{12})$ provide a convenient model tuning based on the information from the Internet. Therefore, it is expedient to apply the proposed model to football league championships. To illustrate the simulation by an example, we used the championship of Finland, which is characterized by a minimum number of sensations. However, the results obtained can hardly be applied to world championships since the model did not take into account the following important factors.

1. Number of injured players. In formalizing this factor, it is necessary to take into account the importance and performance of the injured player, who may influence the match result.

2. The number of booked and benched players.

3. Refereeing objectivity. This factor is determined by the first- and second-kind errors. A first-kind error (false alarm) is indicative of the referee's prepossession to a team (unfair warnings). The second-kind error (skip of a fault) means that the referee does not notice incorrect actions of one of the teams.

4. Weather and climatic conditions. This factor determines that a technique advantage may be lost in a game played in unfavorable conditions.

Allowing for these factors may be the subject of a special analysis and construction of a special knowledge base. However, information about such factors become available just before a match. Therefore, the proposed model can be used for a preliminary prediction, which should be specified later, accounting for current data about the injure level, referee characteristics, and climatic and psychological factors. A fuzzy knowledge base, which takes into account these factors, may be the next level of simulation; however, tuning of such a model is a challenge due to the absence of objective learning samples.

CONCLUSIONS

The model proposed makes it possible to predict the result of a football match using the previous matches of both teams. The model is based on the method of identifying a past-future nonlinear dependence by a fuzzy knowledge base.

Acceptable simulation results can be achieved by tuning fuzzy rules based on tournament data. Tuning consists in selecting the parameters of membership functions of fuzzy terms and rule weights by combining the genetic (off-line) and neural (on-line) optimization algorithms. The prediction model can be further improved by accounting for additional factors in the fuzzy rules: home/away game, number of injured players, and various psychological effects.

The model can be used for creating commercial programs of predicting the results of football matches for bookmaker offices. Moreover, the technique for constructing and tuning the fuzzy model presented in the paper can be used for design and tuning of fuzzy expert systems in other domains.

APPENDIX

The partial derivatives appearing in relations (11)–(13) characterize the error sensitivity (E_t) to the variation of parameters of a neural fuzzy network and can be calculated as follows:

$$\frac{\partial E_t}{\partial w_3^{jp}} = \varepsilon_1 \varepsilon_2 \varepsilon_3 \frac{\partial \mu^{dj}(y)}{\partial w_3^{jp}}, \quad \frac{\partial E_t}{\partial c_{11,12}^{jp}} = \varepsilon_1 \varepsilon_2 \varepsilon_3 \varepsilon_4 \frac{\partial \mu^{jp}(x_i)}{\partial c_{11,12}^{jp}}, \quad \frac{\partial E_t}{\partial b_{11,12}^{jp}} = \varepsilon_1 \varepsilon_2 \varepsilon_3 \varepsilon_4 \frac{\partial \mu^{jp}(x_i)}{\partial b_{11,12}^{jp}},$$

$$\frac{\partial E_t}{\partial w_1^{jp}} = \varepsilon_1 \varepsilon_2 \varepsilon_3 \varepsilon_5 \varepsilon_6 \frac{\partial \mu^{jp}(z_1)}{\partial w_1^{jp}}, \quad \frac{\partial E_t}{\partial c_{1-5}^{jp}} = \varepsilon_1 \varepsilon_2 \varepsilon_3 \varepsilon_5 \varepsilon_6 \varepsilon_8 \frac{\partial \mu^{jp}(x_i)}{\partial c_{1-5}^{jp}}, \quad \frac{\partial E_t}{\partial b_{1-5}^{jp}} = \varepsilon_1 \varepsilon_2 \varepsilon_3 \varepsilon_5 \varepsilon_6 \varepsilon_8 \frac{\partial \mu^{jp}(x_i)}{\partial b_{1-5}^{jp}},$$

$$\frac{\partial E_t}{\partial w_2^{jp}} = \varepsilon_1 \varepsilon_2 \varepsilon_3 \varepsilon_5 \varepsilon_7 \frac{\partial \mu^{jp}(z_2)}{\partial w_2^{jp}}, \quad \frac{\partial E_t}{\partial c_{6-10}^{jp}} = \varepsilon_1 \varepsilon_2 \varepsilon_3 \varepsilon_5 \varepsilon_7 \varepsilon_9 \frac{\partial \mu^{jp}(x_i)}{\partial c_{6-10}^{jp}}, \quad \frac{\partial E_t}{\partial b_{6-10}^{jp}} = \varepsilon_1 \varepsilon_2 \varepsilon_3 \varepsilon_5 \varepsilon_7 \varepsilon_9 \frac{\partial \mu^{jp}(x_i)}{\partial b_{6-10}^{jp}},$$

where

$$\varepsilon_1 = \frac{\partial E_t}{\partial y} = y_t - \hat{y}_t, \quad \varepsilon_2 = \frac{\partial y}{\partial \mu^{d_j}(y)} = \frac{\bar{d}_j \sum_{j=1}^{m} \mu^{d_j}(y) - \sum_{j=1}^{m} \bar{d}_j \mu^{d_j}(y)}{\left(\sum_{j=1}^{m} \mu^{d_j}(y) \right)^2},$$

$$\varepsilon_3 = \frac{\partial \mu^{d_j}(y)}{\partial(\mu^{jp}(z_1)\, \mu^{jp}(z_2)\, \mu^{jp}(x_{11})\, \mu^{jp}(x_{12}))} = w_3^{jp},$$

$$\varepsilon_4 = \frac{\partial(\mu^{jp}(z_1)\, \mu^{jp}(z_2)\, \mu^{jp}(x_{11})\, \mu^{jp}(x_{12}))}{\partial \mu^{jp}(x_i)} = \frac{1}{\mu^{jp}(x_i)} \mu^{jp}(z_1)\, \mu^{jp}(z_2)\, \mu^{jp}(x_{11})\, \mu^{jp}(x_{12}), \quad i = 11, 12,$$

$$\varepsilon_5 = \frac{\partial(\mu^{jp}(z_1)\, \mu^{jp}(z_2)\, \mu^{jp}(x_{11})\, \mu^{jp}(x_{12}))}{\partial \mu^{jp}(z_i)} = \frac{1}{\mu^{jp}(z_i)} \mu^{jp}(z_1)\, \mu^{jp}(z_2)\, \mu^{jp}(x_{11})\, \mu^{jp}(x_{12}), \quad i = 1, 2,$$

$$\varepsilon_6 = \frac{\partial \mu^{jp}(z_1)}{\partial \left(\prod_{i=1}^{5} \mu^{jp}(x_i) \right)} = w_1^{jp}, \quad \varepsilon_7 = \frac{\partial \mu^{jp}(z_2)}{\partial \left(\prod_{i=6}^{10} \mu^{jp}(x_i) \right)} = w_2^{jp},$$

$$\varepsilon_8 = \frac{\partial \left(\prod_{i=1}^{5} \mu^{jp}(x_i) \right)}{\partial \mu^{jp}(x_i)} = \frac{1}{\mu^{jp}(x_i)} \prod_{i=1}^{5} \mu^{jp}(x_i), \quad i = 1, 2, \ldots, 5,$$

$$\varepsilon_9 = \frac{\partial \left(\prod_{i=6}^{10} \mu^{jp}(x_i) \right)}{\partial \mu^{jp}(x_i)} = \frac{1}{\mu^{jp}(x_i)} \prod_{i=6}^{10} \mu^{jp}(x_i), \quad i = 6, 7, \ldots, 10,$$

$$\frac{\partial \mu^{d_j}(y)}{\partial w_3^{jp}} = \mu^{jp}(z_1)\, \mu^{jp}(z_2)\, \mu^{jp}(x_{11})\, \mu^{jp}(x_{12}), \quad \frac{\partial \mu^{jp}(z_1)}{\partial w_1^{jp}} = \prod_{i=1}^{5} \mu^{jp}(x_i), \quad \frac{\partial \mu^{jp}(z_2)}{\partial w_2^{jp}} = \prod_{i=6}^{10} \mu^{jp}(x_i),$$

$$\frac{\partial \mu^{jp}(x_i)}{\partial c_i^{jp}} = \frac{2 c_i^{jp} (x_i - b_i^{jp})^2}{\left((c_i^{jp})^2 + (x_i - b_i^{jp})^2 \right)^2}, \quad \frac{\partial \mu^{jp}(x_i)}{\partial b_i^{jp}} = \frac{2 (c_i^{jp})^2 (x_i - b_i^{jp})}{\left((c_i^{jp})^2 + (x_i - b_i^{jp})^2 \right)^2}, \quad i = 1, 2, \ldots, 12.$$

REFERENCES

1. A. G. Ivakhnenko and V.G. Lapa, Prediction of Random Processes [in Russian], Naukova Dumka, Kiev (1971).
2. S. Markidakis, S. C. Wheelwright, and R. J. Hindman, Forecasting: Methods and Applications, John Wiley and Sons, New York (1998).
3. J. Mingers, “Rule induction with statistical data — a comparison with multiple regression,” Oper. Res. Soc., **38**, 347–351 (1987).
4. K.A. Willoughby, “Determinants of success in the CFL: A logistic regression analysis,” in: Proc. Nat. Ann. Meeting to the Dec. Sci., Decision Sci. Inst., Atlanta, USA, 2 (1997), pp. 1026–1028.
5. M. E. Glickman and H. S. Stern, “A state-space model for national football league scores,” J. Amer. Statist. Assoc., **93**, 25–35 (1998).
6. H. Stern, “On probability of winning a football game,” The Amer. Statist., **45**, 179–183 (1991).
7. M. I. Dixan and S. C. Coles, “Modeling association football scores and inefficiencies in the football betting market,” Appl. Statist., **46**, No. 2, 265–280 (1997).
8. R.H. Koning, “Balance in competition in Dutch soccer,” The Statistician, **49**, 419–431 (2000).
9. H. Rue and O. Salvensen, “Prediction and retrospective analysis of soccer matches in a league,” The Statistician, **49**, 399–418 (2000).
10. S. Walczar and J. Krause, “Chaos neural networks and gaming,” in: Proc. 3rd Golden West Intern. Conf., Intelligent Systems, Las Vegas, NV, USA, Kluwer, Dordrecht (1995), pp. 457–466.
11. M. C. Purucker, “Neural network quarterbacking,” IEEE Potentials, **15**, No. 3, 9–15 (1996).
12. E. M. Condon, B. L. Colden, and E. A. Wasil, “Predicting the success of nations at the summer Olympic using neural networks,” Comput. Oper. Res., **26**, No. 13, 1243–1265 (1999).
13. L. A. Zadeh, “Outline of a new approach to the analysis of complex systems and decision process,” IEEE Trans., SMC, 3, No. 1, 28–44 (1973).
14. A. Rotshtein, “Design and tuning of fuzzy rule-based systems for medical diagnosis,” in: N.-H. Teodorescu, A. Kandel, and L. Gain (eds.), Fuzzy and Neuro-Fuzzy Systems in Medicine, CRC Press, New York (1998), pp. 243–289.
15. A. P. Rotshtein, Intelligent Identification Techniques: Fuzzy Sets, Neural Networks, and Genetic Algorithms [in Russian], Universum, Vinnitsa (1999).
16. A. P. Rotshtein and D. I. Katel’nikov, “Identification of nonlinear objects by fuzzy knowledge bases,” Cyb. Syst. Anal., Vol. 34, No. 5, 676–683 (1998).
17. A. P. Rotshtein, E. E. Loiko, and D. I. Katel’nikov, “Prediction of the number of diseases on the basis of expert-linguistic information,” Cyb. Syst. Anal., Vol. 35, No. 2, 335–342 (1999).
18. A. P. Rotshtein and Yu. I. Mityushkin, “Neurolinguistic identification of nonlinear dependences,” Cyb. Syst. Anal., Vol. 36, No. 2, 179–187 (2000).
19. A. P. Rotshtein and S. D. Shtovba, “Control of a dynamic system based on a fuzzy knowledge base,” Avtom. Vych. Tekhn., No. 2, 23–30 (2001).
20. M. Gen and R. Cheng, Genetic Algorithms and Engineering Design, John Wiley and Sons, New York (1997).
21. Ya. Z. Tsypkin, Fundamentals of the Information Theory of Identification [in Russian], Nauka, Moscow (1984).