Prognosis Of Diseases Number Using Fuzzy Logic

Alexander Rotshtein[†*], Eugene Loiko[‡], Denis Katelnikov[†]
[†] - Vinnitsa State Technical University,
Khmelnitsky Shosse 95, Vinnitsa, 286021, UKRAINE
Phone: (0432)440-157, Fax: (0432)465-772
E-mail: alex@rap.rts.vinnica.ua
[‡] - Vinnitsa State Medical University,
Pirogov str. 56, Vinnitsa286018, UKRAINE

ABSTRACT:This paper presents a new method of diseases number forecasting based on the combination of diseases number experimental data with some fuzzy expert information expressing regularities in available data.

KEYWORDS:prognosis model, fuzzy logic, membership functions, four-year cycle, appendicular peritonitis

1. INTRODUCTION

Prognosis or forecasting of the number of diseases of some types in a city or a region is the necessary element of prophylactic and treatment measures organization. From the formal point of view this task is related to some wide class of time series prognosis tasks (being combinations of values in fixed moments), and arises not only in medicine, but in physics, technology, economics, sociology and other sciences [Ivachnenko (1971)].

Nontriviality of time series prognosis is conditioned by the necessity to extrapolate data about the past into the future taking into account some unknown law about the phenomenon making the basis of the process being studied (in our case it is the number of diseases).

A lot of studies are devoted to elaboration of mathematical prognosis methods [Ivachnenko (1971)]. The most widely spread are those methods which are built on the basis of probabilistical and statistical techniques. But their use demands a great deal of experimental data which is not always successful to collect in conditions of events that took place relatively recently, for instance, the events in Chernobyl.

Artificial neural networks become widely popular in prognosis tasks lately [Mayoraz (1995)]. They are treated as some universal models very close to human brain activity which is learning to identify unknown regularities. But, as it is in the case of probabilistic-statistic methods, neural network learning needs much sampling of experimental data.

In this paper we propose a new method of prognosis which combines experimental data about the diseases number with some linguistic-expert information about the regularities identified in the data available.

The use of the linguistic-expert regularities that are formalized using fuzzy logic [Zadeh (1973)] allows to construct some prognosis model in the conditions of scant experimental samplings.

Our method can be considered as close to the proposed in [Nie (1997)] fuzzy-neural approach of nonlinear time series forecasting. But in contrast to [Nie (1997)] we do not use neural network for the prognosis model learning . Instead of neural network we use methodology of design and tuning of fuzzy if-then rules proposed in [Rotshtein (1998)].

2. EXPERIMENTAL DATA AND FOUR-YEAR CYCLES

Prevalence of appendicular peritonitis in children according to the data received at the Child Surgical Clinic in Vinnitsa (Ukraine) in 1982-1997 as summarized in Table I is located here.

Year	1982	1983	1984	1985	1986	1987	1988	1989
Number of cases	109	143	161	136	161	163	213	220
Year	1990	1991	1992	1993	1994	1995	1996	1997
Number of cases	162	194	164	196	245	252	240	225

Table I: Annual prevalence of diseases number

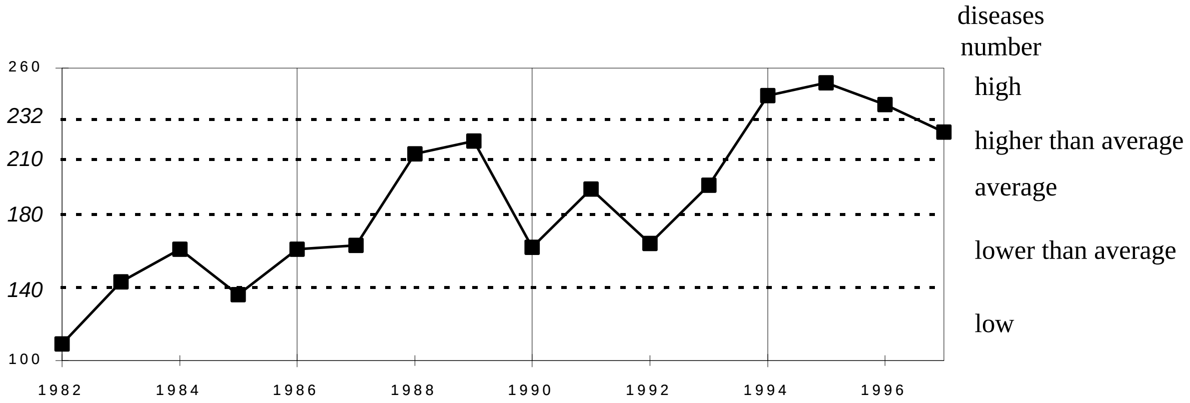


Figure 1:Dynamics of diseases number change

Studying the dynamics of diseases number change according to Fig. 1, it is easy to notice the availability of four-year cycles in which the leap-year holds the third place. Let us designate these cycles as follows:

$$\ldots\ x_4^{i-1}\}\ \{\ x_1^i\quad x_2^i\quad x_3^i\quad x_4^i\ \}\ \{\ x_1^{i+1}\ \ldots$$

the leap year

where i is the four-year cycle number, x_1^i is the number of diseases two years before the leap year, x_2^i is the number of diseases one year before the leap year, x_3^i is the number of diseases during the leap year, x_4^i is the number of diseases during the year after the leap year.

The represented designations will be used further on during regularities formations necessary for prognosis making.

3. LINGUISTIC-EXPERT REGULATIONS

The regularities being prone to study using Fig. 1, can be easily written in the form of four expert expressions in natural language. These expressions are the rules «IF-THEN», which combine diseases numbers in the i -th and in the $(i+1)$-th cycles (Fig. 2)

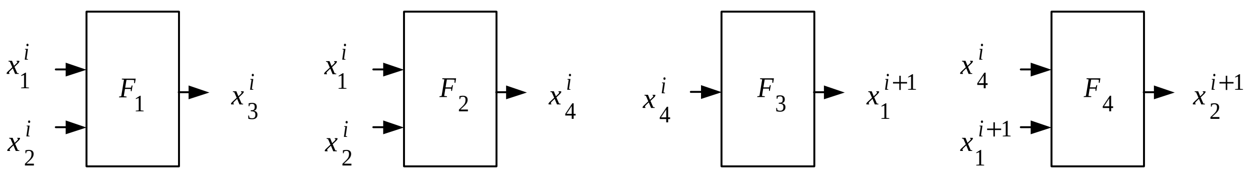


Figure 2:Functional ties among diseases numbers

F_1:

IF $x_1^i = low$
AND $x_2^i =$ *lower than average,*
THEN $x_3^i =$ *lower than average*

IF $x_1^i =$ *lower than average*
AND $x_2^i =$ *lower than average,*
THEN $x_3^i =$ *higher than average*

IF $x_1^i =$ *lower than average*
AND $x_2^i =$ *average,*
THEN $x_3^i =$ *lower than average*

IF $x_1^i = high$
AND $x_2^i = high,$
THEN $x_3^i = high$

F_2:

IF $x_1^i = low$
AND $x_2^i =$ *lower than average,*
THEN $x_4^i = low$

IF $x_1^i =$ *lower than average*
AND $x_2^i =$ *lower than average,*
THEN $x_4^i =$ *higher than average*

IF $x_1^i =$ *lower than average*
AND $x_2^i =$ *average,*
THEN $x_4^i =$ *average*

IF $x_1^i = high$
AND $x_2^i = high,$
THEN $x_4^i =$ *higher than average*

F_3:

IF $x_4^i = low$
THEN $x_1^{i+1} =$ *lower than average*

IF $x_4^i =$ *higher than average*
THEN $x_1^{i+1} =$ *lower than average*

IF $x_4^i =$ *average*
THEN $x_1^{i+1} = high$

F_4:

IF $x_4^i = low$
AND $x_1^{i+1} =$ *lower than average,*
THEN $x_2^{i+1} =$ *lower than average*

IF $x_4^i =$ *higher than average*
AND $x_1^{i+1} =$ *lower than average,*
THEN $x_2^{i+1} =$ *average*

IF $x_4^i =$ *average*
AND $x_1^{i+1} = high,$
THEN $x_2^{i+1} = high$

The dependencies network in Fig. 3 which combines the above formed rules, shows that one can make four years ahead prognosis using two first years of the i -th cycle: prognosis for two last years of the i -th cycle and for two first years of the next $(i+1)$ -th cycle.

4. PROGNOSIS MODEL

To apply linguistic-expert expressions $F_1 \div F_4$ let us make use of the fuzzy sets theory apparatus [Zadeh (1973)]. According to this theory linguistic terms «low», «lower than average», and other ones are formalized using membership functions. Let us present these functions in this form [Rotshtein (1998)]:

$$\mu^T(x) = \frac{1}{1+\left(\frac{x-b}{c}\right)^2} ,$$

where: b and c are parameters first chosen by expert way, and are then tuned by experimental data, $\mu^T(x)$ is number in the range of [0,1], which characterizes subjective measure of x value correspondence to linguistic term T.

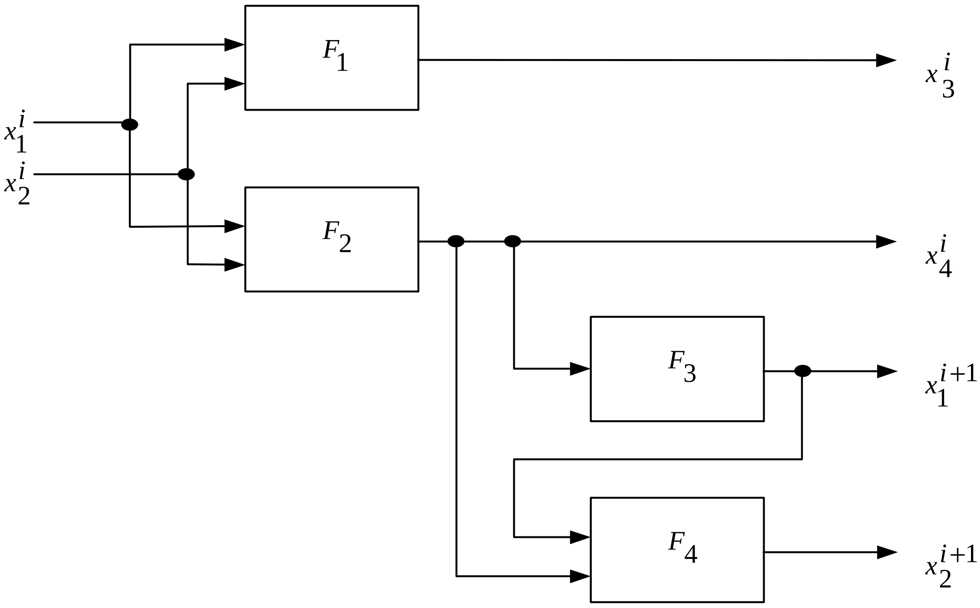


Figure 3:Dependencies network for prognosis making

Parameters b and c as chosen by expert way for various linguistic terms and, as used in rules $F_1 \div F_4$, are represented in Table II.

The membership functions received by this way are shown in Fig.4.

Linguistic terms of variables $x_1^i \div x_4^i$	Parameters before tuning		Parameters after tuning	
	b	c	b	c
low (L)	100	50	100.385	14.148
lower than average (lA)	160	30	146.602	21.046
average (A)	195	25	195.650	7.621
higher than average (hA)	222	20	234.457	19.760
high (H)	260	30	251.836	36.640

Table II:Membership functions parameters

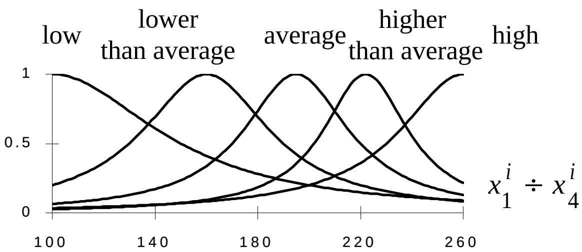


Figure 4:Membership functions of linguistic terms before tuning

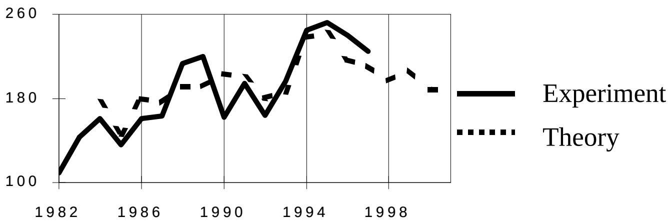
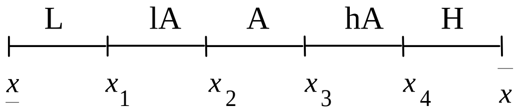


Figure 5:Comparison of experimental data and prognosis model before tuning

Let us designate the range of possible diseases numbers values by $[\underline{x}, \overline{x}]$. Let us divide this range into five parts:



,

Then, using the method of fuzzy logic equations [Rotshtein (1998)] and expressions $F_1 \div F_4$ we can write down the prognosis model in explicit form:

$$F_1: \begin{cases} x_3^i = \dfrac{x_1 \mu^{lA}(x_3^i) + x_3 \mu^{hA}(x_3^i) + x_4 \mu^{H}(x_3^i)}{\mu^{lA}(x_3^i) + \mu^{hA}(x_3^i) + \mu^{H}(x_3^i)}, \\ \mu^{lA}(x_3^i) = \max\begin{pmatrix} \min\left(\mu^{L}(x_1^i), \mu^{lA}(x_2^i)\right), \\ \min\left(\mu^{lA}(x_1^i), \mu^{A}(x_2^i)\right) \end{pmatrix}, \\ \mu^{hA}(x_3^i) = \min\left(\mu^{lA}(x_1^i), \mu^{lA}(x_2^i)\right), \\ \mu^{H}(x_3^i) = \min\left(\mu^{H}(x_1^i), \mu^{H}(x_2^i)\right). \end{cases}$$

$$F_2: \begin{cases} x_4^i = \dfrac{\underline{x} \mu^{L}(x_4^i) + x_2 \mu^{A}(x_4^i) + x_3 \mu^{hA}(x_4^i)}{\mu^{L}(x_4^i) + \mu^{A}(x_4^i) + \mu^{hA}(x_4^i)}, \\ \mu^{L}(x_4^i) = \min\left(\mu^{L}(x_1^i), \mu^{lA}(x_2^i)\right), \\ \mu^{A}(x_4^i) = \min\left(\mu^{lA}(x_1^i), \mu^{A}(x_2^i)\right), \\ \mu^{hA}(x_4^i) = \max\begin{pmatrix} \min\left(\mu^{lA}(x_1^i), \mu^{lA}(x_2^i)\right), \\ \min\left(\mu^{H}(x_1^i), \mu^{H}(x_2^i)\right) \end{pmatrix}. \end{cases}$$

$$F_3: \begin{cases} x_1^{i+1} = \dfrac{x_1 \mu^{lA}(x_1^{i+1}) + x_4 \mu^{H}(x_1^{i+1})}{\mu^{lA}(x_1^{i+1}) + \mu^{H}(x_1^{i+1})}, \\ \mu^{lA}(x_1^{i+1}) = \max\left(\mu^{L}(x_4^i), \mu^{hA}(x_4^i)\right), \\ \mu^{H}(x_1^{i+1}) = \mu^{A}(x_4^i).. \end{cases}$$

$$F_4: \begin{cases} x_2^{i+1} = \dfrac{x_1 \mu^{lA}(x_2^{i+1}) + x_2 \mu^{A}(x_2^{i+1}) + x_4 \mu^{H}(x_2^{i+1})}{\mu^{lA}(x_2^{i+1}) + \mu^{A}(x_2^{i+1}) + \mu^{H}(x_2^{i+1})} \\ \mu^{lA}(x_2^{i+1}) = \min\left(\mu^{L}(x_4^i), \mu^{lA}(x_1^{i+1})\right) \\ \mu^{A}(x_2^{i+1}) = \min\left(\mu^{hA}(x_4^i), \mu^{lA}(x_1^{i+1})\right) \\ \mu^{H}(x_2^{i+1}) = \min\left(\mu^{A}(x_4^i), \mu^{H}(x_1^{i+1})\right) \end{cases}$$

Using this derived model we can roughly make prognosis of the diseases number as it is shown in Fig. 5. To make the prognosis more precise it is necessary to pass to model tuning.

5. PROGNOSIS MODEL TUNING

The task of tuning is in the selection of such b and c parameters of linguistic terms membership functions (Fig.4) which provide for the minimum of deviation between theoretical and experimental diseases numbers. According to the method of least squares this task is formulated in this way:

To find such parameters of b and c so that:

$$\sum_{i=1}^{N}\left(x_3^i - \hat{x}_3^i\right)^2 + \sum_{i=1}^{N}\left(x_4^i - \hat{x}_4^i\right)^2 + \sum_{i=1}^{N-1}\left(x_1^{i+1} - \hat{x}_1^{i+1}\right)^2 + \sum_{i=1}^{N-1}\left(x_2^{i+1} - \hat{x}_2^{i+1}\right)^2 = \min_{b,c}$$

where x_3^i, x_4^i, x_1^{i+1}, x_2^{i+1} are theoretical diseases numbers which depend on parameters b and c of membership functions, $\hat{x}_3^i$, $\hat{x}_4^i$, $\hat{x}_1^{i+1}$, $\hat{x}_2^{i+1}$ are experimental diseases numbers, N is the number of cycles used for model tuning.

To solve the above stated optimization problem we used genetic algorithm as described in [Rotshtein (1998)].

After prognosis model tuning when $N = 3$, what corresponds to the use of data within 1982-1993, the derived membership functions parameters are shown in Table II. Membership functions after tuning are represented in Fig. 6. It was taken into account that: $\underline{x} = 100$, $x_1 = 140$, $x_2 = 180$, $x_3 = 210$, $x_4 = 232$, $\overline{x} = 260$.

The use of the tuned membership functions yields the prognosis model which is enough close to experimental data (Fig. 7).

So as experimental values of diseases numbers in 1994-1997 were not used in model tuning, enough close results of theory and experiment for these years are the evidence of the high quality of the constructed prognosis model.

Information about prognosis model error and the prognosis of the appendicular peritonitis disease number until 2001 are represented in Table III.

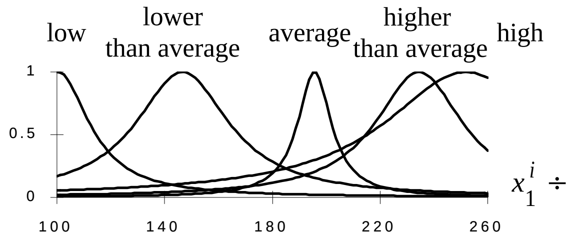


Figure 6:Membership functions of linguistic terms after tuning

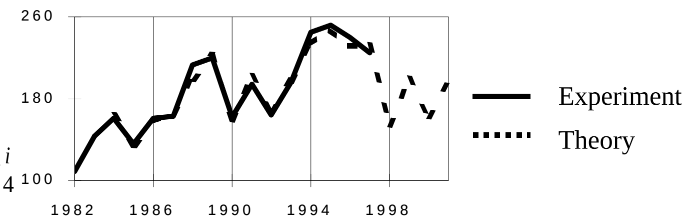


Figure 7:Comparison of experimental data and prognosis model after tuning

Year	Experiment	Theory	Error		Year	Experiment	Theory	Error
1982	109				1992	164	168	4
1983	143				1993	196	198	2
1984	161	163	2		1994	245	234	11
1985	136	135	1		1995	252	247	5
1986	161	158	3		1996	240	231	9
1987	163	165	2		1997	225	232	7
1988	213	199	14		1998		154	
1989	220	222	2		1999		199	
1990	162	159	3		2000		163	
1991	194	201	7		2001		195	

Table III:Experimental and theoretical diseases numbers

6. CONCLUSIONS

A new method of diseases number prognosis is proposed and shown in this article by way of illustrations employing appendicular peritonitis cases based on expert information about regularities identified in available experimental data.

The use of expert information in the form of natural language rules «IF-THEN» which are formalized using fuzzy logic allows to build prognosis models with some relatively small samples of experimental data provided (in contrast to statistic methods).

REFERENCES

Ivachnenko A.G., Lapa V.G., 1971, «Forecasting of random processes», Kiev, Naukova Dumka, Ukraine (In Russian).

Mayoraz F., Cornu T., Vulliet L., 1995, «Prediction of Slope Movements Using Neural Networks», in «Fuzzy Systems & A.I.», Vol. IV, 1 1, pp. 9-17.

Nie J.,1997,«Nonlinear time series forecasting:A fuzzy neural approach»,in«Neurocomputing», Vol 16,1 1,pp.63-76.

Rotshtein A. P., 1998, «Design and Tuning of Fuzzy Rule-Based Systems for Medical Diagnosis», in H.-N. Teodorescu, A. Kandel, L. Jain (ed.) «Fuzzy and Neuro-Fuzzy Systems in Medicine and Biomedical Engineering», CRC Press, USA.

Zadeh, L. 1973, «Outline of a New Approach to the Analysis of Complex Systems and Decision Processes», in «IEEE Trans. Syst. Men. Cybern.», 3, 28.