STATISTICAL METHOD OF FUZZY GRADES' ANALYSIS FOR FORECAST MODELING

I. Khutsishvili

Iv. Javakhishvili Tbilisi State University
0143 University Street 2, Tbilisi, Georgia

(Received: 11.06.05; accepted: 18.02.06)

Abstract

The present article introduces the mathematic model of forecasting through the application of the statistical method of fuzzy grades' analysis. The article focuses on the specific example of earthquakes' forecasting and takes an intensity of the electric field as the single factor-precursor. Initial data comprises the earthquakes' statistics in the Caucasus Region from 1955 to 1992.

The efficiency of the method was tested on twenty arbitrarily taken earthquakes. The method proved 70% accuracy, which is the satisfactory result taking into account the fact the intensity of the electric field is not the principal factor-precursor.

Key words and phrases: Fuzzy grades, fuzzy relative frequencies, membership functions, multi-factor linear synthesis.

AMS subject classification: 03B52, 94D05, 03E72

1 Introduction

Construction of a forecasting model requires us to classify forecasting object into a number of categories. Such a classification is to be performed in consideration of factors describing the object. For example, the forecasting object such as an earthquake can be divided into forecasting classes such as: "strong earthquake", "moderate earthquake" and "weak earthquake". It becomes obvious, that in this case, it is impossible to draw strict boundaries between classes of classification, i.e. the forecasting concept contains fuzziness. If classes of classification are the fuzzy sets, application of classical statistics methods becomes less useful, since it does not give authentic and accurate results. Thus, applying statistical method of the analysis of fuzzy grades should be deemed more expedient in such cases. This method enable one to use statistical data when membership functions are constructed and the images of fuzzy grades are obtained.

In a given article we analyze 10-year data of moderate and strong earthquakes each. Intensity of an electric field values, recorded during three days before earthquake, are considered as a single factor-precursor.

Analysis is based on a multi-factor linear synthesis of a fuzzy weight and a fuzzy frequency, while the final decision is made after applying of a principle of a maximum of opportunities to results.

2 Description of a Method

Let's describe the mathematical forecasting model constructed with application of the statistical method of fuzzy grades' analysis (fuzzy grade statistics). In mathematical model the object is presented by set of parameters, all area of these numerical values is shared into forecasting grades: $M_1, M_2, \ldots, M_n$. To each class the numerical interval is put in conformity. Corresponding membership functions are defined: $\mu_1, \mu_2, \ldots, \mu_n$. The mentioned classes are fuzzy, therefore supports of membership functions are intersected.

The forecasting value depends on the certain parameters, or of predictive (helping to make the forecast) factors: $X_1, X_2, \ldots, X_p$. Each of factors, in turn, is divided into classes (sub factors): $X_{k1}, X_{k2}, \ldots, X_{km}$, where $k = \overline{1,p}$; $X_k = \bigcup\limits_{j=1}^{m} X_{kj}$. The numbers of forecasting factors and their classes, and also range of their numerical intervals can be selected arbitrarily.

For selective frequencies of j class of X_k factor of corresponding i forecasting class we shall enter values n_{kj}^i. In mathematical model they represent the primary information (primary data) and them receive by direct supervision and measurements. By means of n_{kj}^i and μ_i numbers and under known formulas fuzzy selective frequencies and fuzzy relative frequencies also are defined: [2]

$$\widetilde{n}\,_{kj}^{m} = \sum_i \mu_i^m \cdot n_{kj}^i, \quad \widetilde{f}\,_{kj}^{m} = \frac{\widetilde{n}\,_{kj}^{m}}{\sum\limits_i \widetilde{n}\,_{kj}^{i}}\,, \tag{2.1}$$

where μ_i^m is average value of membership function when the forecasting value from i forecasting interval belongs to m forecasting grade. Fuzzy weights for each interval of the forecasting factor are calculated also:

$$w_{kj} = \frac{\sum\limits_i \widetilde{n}\,_{kj}^{i}}{\sum\limits_j \sum\limits_i \widetilde{n}\,_{kj}^{i}}\,. \tag{2.2}$$

After that for the certain sample of factors of the forecasting value (forecasting factors) it is already possible to make the forecast: it is necessary to define only fuzzy weights of each forecasting factor according to its interval and to carry out multi-factor linear synthesis of fuzzy weights and fuzzy relative frequencies, and then to compare the numerical values received for each forecasting grade.

3 Example of Application of a Method

Let's consider a concrete example of application of a fuzzy grade statistics. Primary data - statistics of earthquakes in the Caucasus region from 1955 to 1992. [3] Data are taken for a day and in day of earthquake in each of any way chosen 10 years and collected with step at 1 o'clock: $0^{00} - 1^{00}; \ldots ; 23^{00} - 24^{00}$. As the factor-precursor value of intensity of an electric field is considered. The object of forecasting, earthquake, is described by means of a linguistic variable with following values: "moderate earthquake", "strong earthquake" and is characterized by numerical value of magnitude (M). At $3 < M < 5$ "moderate earthquake" is observed; at $5 \le M \le 8$ - "strong earthquake". Let us designate the defined forecast classes as M_1 and M_2, and introduce the corresponding membership functions:

$$\mu_1(M) = \begin{cases} 0, & M \le 4, \\ \frac{1}{1+(3{,}1(M-4{,}5))^2}, & 4 < M < 6, \end{cases}$$

$$\mu_2(M) = \begin{cases} 0, & M < 5, \\ \frac{1}{1+(1{,}29(M-7))^2}, & 5 \le M \le 7, \\ 1, & M > 7. \end{cases}$$

On Figure 1 is shown the scheme of intersection of the forecasting grades and membership functions' supports.

As the forecasting grades are represented by means of the intervals, it is necessary to average membership functions on these intervals. Let μ_i^j there is an averaged value of μ_j with the consideration of an i forecasting grade's support intersection with supp μ_j:

$$\mu_1^1 = \int_{4,0}^{5,0} \frac{dM}{1+(3,1(M-4,5))^2} \approx 0,6438,$$

$$\mu_2^1 = \int_{5,0}^{5,8} \frac{dM}{1+(3,1(M-4,5))^2} \approx 0,1330.$$

$$\mu_1^2 = \int_{4,8}^{5,8} \frac{dM}{1+(1,29(M-7))^2} \approx 0,1798,$$

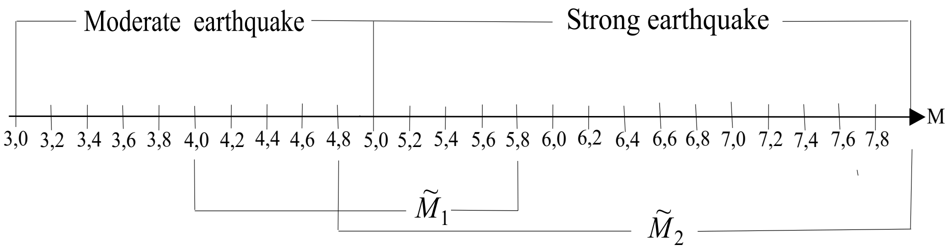

Figure 1. here $\widetilde{M}_i = supp\, \mu_i$.

$$\mu_2^2 = \int_{5,8}^{7,0} \frac{dM}{1 + (1,29(M-7))^2} \approx 0,6415.$$

Hence, the membership functions of fuzzy classification of earthquake in each interval which belongs to fuzzy sets M_1 and M_2 are represented by:

$$\mu_1(M) = \frac{0,6438}{4,0 < M \leq 5,0} + \frac{0,1330}{5,0 < M \leq 5,8},$$
$$\mu_2(M) = \frac{0,1798}{4,8 \leq M \leq 5,8} + \frac{0,6415}{5,8 < M \leq 7}.$$

Let's consider the factors describing object of forecasting $X_1, X_2, \ldots, X_{24}$, where X_i - value of intensity of an electric field in an interval of time $(i-1, i)$.

Each of factors in turn shares on sub factors $x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4}$ ($x_{i_k},\ k = \overline{1,4}$), where x_{i_1} - values of intensity of an electric field in an interval [-104; 2,8]; x_{i_2} - in an interval (2,8; 5]; x_{i_3} - in an interval (5; 10]; x_{i_4} - in an interval (10; 123,20], where $X_i = \bigcup_{j=1}^{4} x_{ij}$.

Further, let us define n_{kj}^i as frequencies of the grades, which represent the frequencies of the j sub factor of the X_k factor occurring in the i forecasting grade. The values of n_{kj}^i constitute initial data received as a result of observations (see [1]).

The $\widetilde{n}_{kj}^{\,i}$ frequencies of fuzzy grades and $\widetilde{f}_{kj}^{\,i}$ fuzzy relative frequencies is built on the basis of formula (1). Coming out from the formula (2) we will build also w_{kj} fuzzy weights of the forecasting factors. Now all the data necessary for decision-making exist.

Table 1. Frequencies of n_{kj}^i of crisp grades

Factor		M_1 n_{kj}^1	M_2 n_{kj}^2	Factor		M_1 n_{kj}^1	M_2 n_{kj}^2	Factor		M_1 n_{kj}^1	M_2 n_{kj}^2
X_1	x_{11}	6	4	X_9	x_{91}	2	1	X_{17}	x_{171}	4	2
	x_{12}	7	8		x_{92}	3	5		x_{172}	1	3
	x_{13}	5	4		x_{93}	10	6		x_{173}	7	8
	x_{14}	2	4		x_{94}	5	8		x_{174}	8	7
X_2	x_{21}	5	3	X_{10}	x_{101}	3	4	X_{18}	x_{181}	3	2
	x_{22}	7	7		x_{102}	2	2		x_{182}	2	2
	x_{23}	7	7		x_{103}	11	5		x_{183}	7	8
	x_{24}	1	3		x_{104}	4	9		x_{184}	8	8
X_3	x_{31}	5	4	X_{11}	x_{111}	2	3	X_{19}	x_{191}	3	2
	x_{32}	7	6		x_{112}	5	5		x_{192}	5	2
	x_{33}	8	6		x_{113}	10	4		x_{193}	7	10
	x_{34}	0	4		x_{114}	3	8		x_{194}	5	6
X_4	x_{41}	5	4	X_{12}	x_{121}	3	3	X_{20}	x_{201}	3	1
	x_{42}	5	8		x_{122}	3	4		x_{202}	4	3
	x_{43}	9	5		x_{123}	8	4		x_{203}	6	9
	x_{44}	1	3		x_{124}	6	9		x_{204}	7	7
X_5	x_{51}	4	2	X_{13}	x_{131}	1	4	X_{21}	x_{211}	4	2
	x_{52}	3	4		x_{132}	5	6		x_{212}	3	2
	x_{53}	8	7		x_{133}	12	3		x_{213}	8	9
	x_{54}	5	7		x_{134}	2	7		x_{214}	5	7
X_6	x_{61}	3	0	X_{14}	x_{141}	3	4	X_{22}	x_{221}	3	3
	x_{62}	4	6		x_{142}	2	5		x_{222}	5	2
	x_{63}	6	6		x_{143}	10	3		x_{223}	8	7
	x_{64}	7	8		x_{144}	5	8		x_{224}	4	8
X_7	x_{71}	0	1	X_{15}	x_{151}	3	3	X_{23}	x_{231}	4	1
	x_{72}	3	3		x_{152}	3	4		x_{232}	7	8
	x_{73}	12	7		x_{153}	10	7		x_{233}	8	4
	x_{74}	5	9		x_{154}	4	6		x_{234}	1	7
X_8	x_{81}	1	1	X_{16}	x_{161}	0	1	X_{24}	x_{241}	3	2
	x_{82}	3	4		x_{162}	4	4		x_{242}	8	6
	x_{83}	10	4		x_{163}	8	7		x_{243}	9	7
	x_{84}	6	11		x_{164}	8	8		x_{244}	0	5

Let's admit, measurements of predicting factors are presented for 2 different days, and results are following:

$X_I=$ $(x_{13}, x_{23}, x_{33}, x_{43}, x_{53}, x_{64}, x_{73}, x_{82}, x_{92}, x_{103}, x_{112}, x_{122}, x_{132}, x_{142}, x_{151},$

$$x_{164}, x_{171}, x_{181}, x_{192}, x_{201}, x_{211}, x_{221}, x_{231}, x_{241}) - I\ time\, observation$$

$$X_{II} = (x_{14}, x_{23}, x_{33}, x_{43}, x_{53}, x_{63}, x_{73}, x_{84}, x_{94}, x_{104}, x_{114}, x_{124}, x_{134}, x_{144}, x_{153}, x_{164}, x_{173}, x_{184}, x_{194}, x_{204}, x_{214}, x_{224}, x_{234}, x_{243}) - II\ time\, observation;$$

The fuzzy weights of the forecasting factors are:

$$\vec{w}_I = (0.2258,\ 0.35,\ 0.3515,\ 0.3531,\ 0.3758,\ 0.3742,\ 0.4788,\ 0.1742,\ 0.1985,\ 0.4046,\ 0.25,\ 0.1742,\ 0.2742,\ 0.3304,\ 0.15,\ 0.4,\ 0.1515,\ 0.1258,\ 0.1773,\ 0.1015,\ 0.1515,\ 0.15,\ 0.1273,\ 0.1258);$$

$$\vec{w}_{II} = (0.1485,\ 0.35,\ 0.3515,\ 0.3531,\ 0.3758,\ 0.3,\ 0.4788,\ 0.4212,\ 0.3127,\ 0.3212,\ 0.2712,\ 0.3727,\ 0.2212,\ 0.3227,\ 0.4273,\ 0.4,\ 0.3742,\ 0.4,\ 0.2742,\ 0.35,\ 0.2985,\ 0.2969,\ 0.1954,\ 0.4015);$$

Likewise, the corresponding matrixes of fuzzy relative frequencies are (matrixes are given in the transposed kind):

$$\widetilde{f}_I = \begin{pmatrix} 0.5198 & 0.4861 & 0.5294 & 0.5724 & 0.5063 & 0.4657 \\ 0.4802 & 0.5139 & 0.4706 & 0.4276 & 0.4937 & 0.5343 \end{pmatrix.$$
$$\begin{matrix} 0.5656 & 0.4424 & 0.4093 & 0.599 & 0.4861 & 0.4424 \\ 0.4344 & 0.5576 & 0.5907 & 0.401 & 0.5139 & 0.5576 \end{matrix}$$
$$\begin{matrix} 0.4583 & 0.6475 & 0.4861 & 0.4861 & 0.5866 & 0.5466 \\ 0.5417 & 0.3525 & 0.5139 & 0.5139 & 0.4134 & 0.4534 \end{matrix}$$
$$\left.\begin{matrix} 0.615 & 0.6361 & 0.5866 & 0.4861 & 0.6656 & 0.5466 \\ 0.385 & 0.3639 & 0.4134 & 0.5139 & 0.3344 & 0.4534 \end{matrix}\right);$$

$$\widetilde{f}_{II} = \left(\begin{matrix} 0.3835 & 0.4861 & 0.5294 & 0.5724 & 0.5063 & 0.4861 \\ 0.6165 & 0.5139 & 0.4706 & 0.4276 & 0.4937 & 0.5139 \end{matrix}\right.$$
$$\begin{matrix} 0.5656 & 0.3956 & 0.4153 & 0.3675 & 0.3456 & 0.4248 \\ 0.4344 & 0.6044 & 0.5847 & 0.6325 & 0.6544 & 0.5752 \end{matrix}$$
$$\begin{matrix} 0.3139 & 0.4153 & 0.5396 & 0.4861 & 0.4657 & 0.4861 \\ 0.6861 & 0.5857 & 0.4604 & 0.5139 & 0.5343 & 0.5139 \end{matrix}$$
$$\left.\begin{matrix} 0.4583 & 0.4861 & 0.435 & 0.3835 & 0.2622 & 0.524 \\ 0.5417 & 0.5139 & 0.565 & 0.6165 & 0.7478 & 0.476 \end{matrix}\right).$$

For each case we shall lead multi-factor synthesis of weights of measurements and matrixes of data. As a result we shall receive the generalized decisions (the weighed vectors of possible decisions):

$$\overrightarrow{Poss_I} = \vec{w}_I \cdot \widetilde{f}_I = (3.16384,\ 2.81227);$$
$$\overrightarrow{Poss_{II}} = \vec{w}_{II} \cdot \widetilde{f}_{II} = (3.68918,\ 4.33931).$$

Using the principle of possibility maximum, we have:
$D^{(\alpha)}_{Class} = \max\limits_i (Poss_\alpha(i))$, where $\alpha = I, II$;
$Poss_\alpha(i)$ is an i component part of $\overrightarrow{Poss_\alpha}$.
In our case for each observation we receive the following forecast:

$$D^{(I)}_{Class} = 0,316384 \ \ (\Rightarrow \ M_1 \ - \ Moderate\, earthquake),$$
$$D^{(II)}_{Class} = 4,33931 \quad (\Rightarrow \ M_2 \ - \ Strong\, earthquake).$$

The received result corresponds to statistical data: values of predicting factors of sample correspond to real data for November, 13-th, 1974 when in 2^{00} there was an earthquake with magnitude 4.7 (on our classification - "moderate earthquake"); there correspond data for December, 16-th, 1990 when in 15^{00} there was an earthquake with magnitude 5.1 (i.e. "strong earthquake").

4 Conclusion

Using an offered method it is necessary to remember, that there should be a remarkable correlation between forecasting factors and object of the forecast.

Besides, it is necessary to make sure, that the sample of primary classical frequencies does not contain much of zero values. Otherwise, it will have statistical effect.

That fact of getting satisfactory results based on relatively small amount of initial data speaks in favor of the offered method.

Table 2. Factual data and forecasting results

Classification	Data	Magnitude	Forecast	Fitting
	22.03.1972	4,5	M_1	+
	28.07.1976	4,7	M_1	+
	13.11.1974	4,7	M_1	+
M_1 (moderate	03.01.1970	4,7	M_1	-
earthquake)	02.06.1967	4,5	M_1	+
	07.04.1989	4,6	M_1	-
$3 < M < 5$	18.10.1981	4,6	M_1	-
	11.12.1980	4,3	M_1	+
	12.07.1978	4,4	M_1	+
	17.03.1978	4,4	M_1	+

Classification	Data	Magnitude	Forecast	Fitting
M_2 (strong earthquake) $5 \leq M \leq 8$	26.02.1978	5,3	M_2	-
	24.11.1976	7	M_2	+
	09.01.1975	5,2	M_2	+
	22.05.1971	6,8	M_2	+
	26.07.1967	5,8	M_2	-
	16.12.1990	5,1	M_2	+
	06.03.1986	6,1	M_2	+
	30.10.1983	6,8	M_2	+
	18.10.1981	5,4	M_2	-
	30.09.1977	5,4	M_2	+

The efficiency of the method was tested on twenty arbitrarily taken earthquakes. The method proved 70% accuracy, which is the satisfactory result taking into account the fact the intensity of the electric field is not the principal factor-precursor.

References

1. Li Zuoying, Chen Zhenpei and Li Jitao, A Model of Weather Forecast by Fuzzy Grade Statistics. *FSS* **26**(1988), No 3, 275–283.

2. Li Juzhang, Fuzzy Statistics of Classification. *Fuzzy Mathematics*, **2**(4)(1988), p. 107.

3. F. Criado, T. Gachechiladze, H. Meladze, G. Tsertsvadze, A new Approach to Analysing Fuzzy Data and Decission-making Regarding the Possibility of Earthquake Occurrence. *Intas*-9702126 (Final Report), www. cordis.lu / en / home. html, 1999.

4. Wang Peizhuang, Fuzzy Sets and its Application. *Publishing House of Science and Technology, Shanghai*, 1983.