Статистический метод анализа нечетких классов при большом числе активностей

Ирина Хуцишвили

Тбилисский Государственный Университет им. Ив. Джавахишвили
Email: khutsishvili_iri@mail.ru

Аннотация:

Статистического метод анализа нечетких классов лежит в основе многих задач принятия решений, в частности, задач прогнозирования. В статье рассматривается задача прогноза, когда прогнозируемый объект описывается большим числом активностей. Построена математическая модель прогноза, введено понятие меры возможности обобщенного решения, предложена новая модель функции совместимости прогнозируемых понятий - модификация модели Заде.

Рассмотрен конкретный пример составления прогноза землетрясений по данным о геофизических активностях атмосферы. В качестве исходных данных взята статистика землетрясений в Кавказском регионе.

Метод опробован на данных для 80 случаев «шумов» и 20 произвольным образом выбранных землетрясений. Точность прогноза (≈70%) можно считать вполне удовлетворительной, тем более, что геофизические активности атмосферы не являются главными факторами-предвестниками землетрясения.

Ключевые слова: *нечеткие множества, функции совместимости, многофакторный линейный синтез, принцип максимума возможностей, мера возможности обобщенного решения.*

1. В статье описывается один из нечетких статистических методов принятия решения и, в известной мере, дается его обобщение для случая, когда объект принятия решения характеризуется большим числом активностей. Рассматриваемый в статье прогнозируемый объект – землетрясение – относится к прогнозируемым понятиям, содержащим нечеткость в самом определении. Классы класификации, на которые можно разделить прогнозируемый объект, представляют собой нечеткие множества. Нечеткими множествами являются также и классы активностей прогнозируемого объекта. Это и предопределило применение в задаче составления прогноза статистического метода анализа нечетких классов. В известном «классическом» варианте названного метода [1, 3, 6] рассматриваемое число активностей прогнозируемого объекта невелико. Однако для конкретной задачи построения прогноза оно может быть довольно большим (в рассматриваемой конкретной задаче оно составило 9). Поэтому «классический» вариант метода в условиях большого числа активностей потребовал модификации. С этой целью в работе вводится понятие меры возможности (см. формулу (3)), которое используется для получения обобщенного решения.

Принадлежность прогноза к тому или иному классу классификации определяют с применением т.н. функций совместимости прогнозируемых понятий или, проще, функций совместимости. Функция совместимости представляет собой обобщение понятия характеристической функции обычного четкого множества. Одной из самых главных проблем статистического метода анализа нечетких классов является построение функций совместимости прогнозируемых понятий с учетом интеллектуальной активности экспертов. Этой теме посвящено огромное количество литературы. Т.к. функции совместимости строятся исходя из субъективных предпочтений экспертов, вид их может быть произвольным. «Хорошее» определение функций совместимости является основной гарантией успеха работы метода. Для конкретного случая построения прогноза в работе предложена модель функции совместимости, являющаяся новой модификацией модели Заде (см. (5)).

2. Согласно статистического метода анализа нечетких классов (впоследствии будем называть его методом статистики нечетких классов) объект прогнозирования описывается соответствующей прогнозируемой величиной. Область значений прогнозируемой величины делится на конечное число прогнозируемых классов, каждому из которых ставится в соответствие численный интервал.

Определяются соответствующие функции совместимости прогнозируемых понятий (функции принадлежности классу). Определение этих функций всегда содержит в себе субъективный момент – мнение эксперта о степени принадлежности прогнозируемого объекта прогнозируемому классу. В силу нечеткости прогнозируемых классов, суппорты функций совместимости пересекаются.

Значение прогнозируемой величины зависит от активностей (прогнозирующих факторов), каждая из которых может делиться на классы. Количество активностей и их классов, а также пределы изменения описывающих их численных интервалов могут быть выбраны произвольно.

Введем некоторые обозначения:

Прогнозируемые классы : $\boldsymbol{M_1, M_2, \ldots, M_\ell}$;

Функции совместимости прогнозируемых понятий: $\mu_1, \mu_2, \ldots, \mu_\ell$

Активности: $\boldsymbol{X_1, X_2, \ldots, X_m}$;

Классы активностей : $\boldsymbol{X_{k1}, X_{k2}, \ldots, X_{kr}}$; $\boldsymbol{k = \overline{1,m}}$; $\boldsymbol{X_k = \bigcup_{j=1}^{r} X_{kj}}$.

Введем также обозначение $\boldsymbol{n_{kj}^i}$ для выборочных частот $\boldsymbol{j}$ – того класса $\boldsymbol{X_k}$ – той активности, соответствующей $\boldsymbol{i}$ – тому прогнозируемому классу. Выборочные частоты являются результатом обработки первичной информации, которую получают непосредственным наблюдением и измерениями.

На основании величин $\boldsymbol{n_{kj}^i}$ и $\boldsymbol{\mu_i}$ по известным формулам определяются нечеткие выборочные частоты, нечеткие относительные частоты и нечеткие веса для каждого интервала активности: [1]

$$\tilde{n}_{kj}^m = \sum_i \mu_i^m \cdot n_{kj}^i \ , \qquad \tilde{f}_{kj}^m = \frac{\tilde{n}_{kj}^m}{\sum_i \tilde{n}_{kj}^i} \ , \quad w_{kj} = \frac{\sum_i \tilde{n}_{kj}^i}{\sum_j \sum_i \tilde{n}_{kj}^i} \ . \qquad (1)$$

В первой из формул $\boldsymbol{\mu_i^m}$ – это среднее значение функции совместимости, когда прогнозируемая величина из $\boldsymbol{i}$ – того прогнозируемого класса принадлежит $\boldsymbol{m}$ – тому прогнозируемому классу.

Проведя вычисления по формулам (1) и тем самым подготовив необходимую базу знаний, можно принять решение (сделать прогноз) для конкретой совокупности активностей прогнозируемой величины. Для этого нужно определить нечеткие веса каждой активности в соответствии с ее интервалом, а затем осуществить многофакторный линейный синтез нечетких весов и нечетких относительных частот.

В результате многофакторного линейного синтеза получаем соответствующее обобщенное решение (вектор возможных взвешенных решений) [3]

$$\vec{D}_\alpha = \vec{w}_\alpha \cdot \tilde{f}_\alpha . \qquad (2)$$

В условиях большого числа активностей для обобщенного решения оказывается необходимым ввести понятие меры возможности

$$\overrightarrow{Poss_\alpha} = \frac{\vec{D}_\alpha}{\max\limits_j (D_\alpha(j))}, \text{ где } D_\alpha(j) - j-\text{тая компонента вектора } \vec{D}_\alpha. \qquad (3)$$

И, наконец, для получения классического, единственного решения необходимо ввести дополнительный принцип. Например, можно воспользоваться принципом максимума возможностей. Тогда окончательное решение примет вид [4]:

$$D_{Class}^{(\alpha)} = \max_i (Poss_\alpha(i)), \text{ где } Poss_\alpha(i) \text{ есть } i - \text{тая компонента вектора } \overrightarrow{Poss_\alpha}. \qquad (4)$$

3. Рассмотрим конкретный пример применения метода статистики нечетких классов.

Для конкретной задачи принятия решения о возможности землетрясения в качестве факторов-предвестников рассматриваются некоторые геофизические данные атмосферы:

1. величина напряженности электрического поля (вольт/м);
2. температура воздуха (в градусах Цельсия);
3. температура почвы (в градусах Цельсия);
4. атмосферное давление (в мб);
5. абсолютная влажность (упругость водяного пара в мб);
6. относительная влажность (в %);
7. общая облачность (в баллах);
8. нижняя облачность (в баллах);
9. скорость ветра (в м/сек.).

Значения величин активностей замерены в течение дня с промежутком в 3 часа: $0^{\underline{00}}$, $3^{\underline{00}}$, $6^{\underline{00}}$, $9^{\underline{00}}$, $12^{\underline{00}}$, $15^{\underline{00}}$, $18^{\underline{00}}$, $21^{\underline{00}}$. Величины активностей для известных случаев землетрясения получены на основе известной статистики землетрясений в Кавказском регионе.

В качестве базовых случаев взяты данные геофизических активностей в день землетрясения по всем интервалам интенсивности. Для каждого интервала выбраны по 10 известных случаев землетрясений

Объект прогнозирования – землетрясение – делится на 3 класса классификации (M_0, M_1, M_2) со значениями лингвистических переменных «шум», «среднее землетрясение» и «сильное землетрясение» [5]. Каждый интервал интенсивности характеризуется численным значением величины магнитуды (M). При $0 \le M \le 3$ наблюдается «шум», при $3 < M < 5$ имеем «среднее землетрясение», при $5 \le M \le 8$ – «сильное землетрясение».

Модель функции совместимости, применяемая в данном методе, является модификацией модели Заде:

$$\mu_0(M) = \begin{cases} \dfrac{1}{1+(\alpha_1 M)^2}, & 0 \le M \le 3; \\ 0, & M > 3 \end{cases}; \qquad \mu_1(M) = \begin{cases} 0, & M < 4.4; \\ \dfrac{1}{1+(\alpha_2(M-4.9))^2}, & 4.4 \le M \le 5.4; \\ 0, & M > 5.4 \end{cases};$$

$$\mu_2(M) = \begin{cases} 0, & M < 4.4; \\ \dfrac{1}{1+(\alpha_3(M-8))^2}, & 4.4 \le M \le 8; \\ 1, & M > 8 \end{cases}.$$

Коэффициенты α_1, α_2 и α_3 подбираются эмпирическим путем в соответствии с имеющимися данными и рекомендациями экспертов.

В нашем случае функции совместимости имеют следующий вид:

$$\mu_0(M)=\begin{cases}\dfrac{1}{1+(0.15M)^2}, & 0\le M\le 3;\\ 0, & M>3\end{cases};\qquad \mu_1(M)=\begin{cases}0, & M<4.4;\\ \dfrac{1}{1+(4.99(M-4.9))^2}, & 4.4\le M\le 5.4;\\ 0, & M>5.4\end{cases};$$

$$\mu_2(M)=\begin{cases}0, & M<4.4;\\ \dfrac{1}{1+(0.5(M-8))^2}, & 4.4\le M\le 8;\\ 1, & M>8\end{cases}.$$

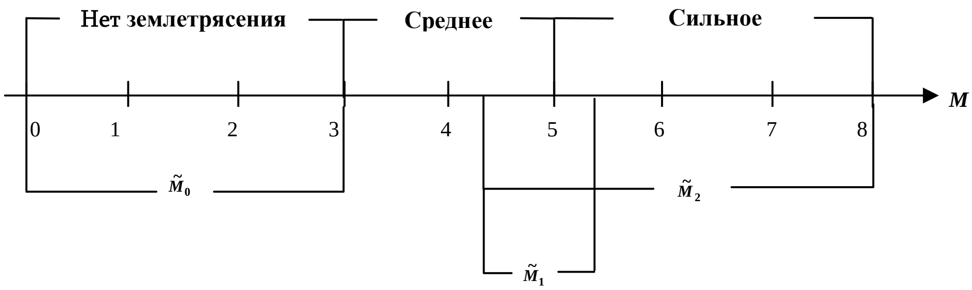
Так как прогнозируемые классы представлены в виде интервалов, необходимо усреднить функции совместимости по этим интервалам. Пусть μ_i^j есть усредненное значение μ_j с учетом пересечения суппорта i-того прогнозируемого нечеткого класса и supp μ_j. Тогда:

$$\mu_0^0=\frac{1}{3}\int_0^3\frac{dM}{1+(0.15M)^2}\approx 0.93968;\qquad \mu_1^0=0;$$

$$\mu_0^1=0;\qquad \mu_1^1=\frac{1}{5-4.4}\int_{4.4}^{5,0}\frac{dM}{1+(4.99(M-4.9))^2}\approx 0.55192;\qquad \mu_2^1=\frac{1}{0.4}\int_5^{5.4}\frac{dM}{1+(4.99(M-4.9))^2}\approx 0.3641;$$

$$\mu_1^2=\frac{1}{0.6}\int_{4.4}^{5}\frac{dM}{1+(0.5(M-8))^2}\approx 0.26968;\qquad \mu_2^2=\frac{1}{3}\int_5^8\frac{dM}{1+(0.5(M-8))^2}\approx 0.6552.$$

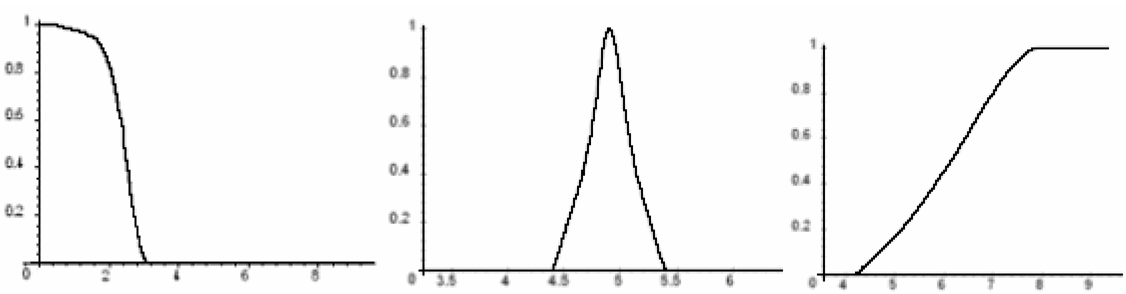
Приведем схему перекрытия интервалов прогнозируемых классов и суппортов функций совместимости:



где $\tilde{M}_i=\text{supp }\mu_i$, $i=0,1,2$

Основываясь на имеющихся исходных данных считаем, что случаи «шумов» не влияют на «среднее землетрясение», и наоборот, что и отразилось на вышеприведенной схеме.

Графики соответствующих функций совместимости имеют вид



Каждая из активностей (факторов-предвестников землетрясения) делится на три класса (подфактора). Интервалы классов активностей – нечеткие множества, границы их подбираются эмпирическим путем, согласуясь с имеющимися данными и экспертными оценками.

Введем обозначения для каждой активности и каждого ее класса

X_0 – напряженность электрического поля: x_{00} – НЭП≤ 5, x_{01} – НЭП$\in (5,9]$, x_{02} – НЭП>9;
X_1 – температура воздуха: x_{10} – ТВ<4.3, x_{11} – ТВ$\in [4.3,9)$, x_{12} – ТВ≥ 9;
X_2 – температура почвы: x_{20} – ТП≤ 2, x_{21} – ТП$\in (2,11)$, x_{22} – ТП≥ 11;
X_3 – атмосферное давление: x_{30} – Атм.д≤ 911.0, x_{31} – Атм.д$\in (911, 916.4)$, x_{32} – Атм.д ≥ 916.4;
X_4 – влажность абсолютная: x_{40} – Вл.абс≤ 5, x_{41} – Вл.абс$\in (5,9]$, x_{42} – Вл.абс>9;
X_5 – влажность относительная: x_{50} – Вл.отн<60, x_{51} – Вл.отн $\in [60,81]$, x_{52} – Вл.отн >81;
X_6 – облачность общая: x_{60} – Обл.общ.<6, x_{61} – Обл.общ.$\in [6,8]$, x_{62} – Обл.общ.>8;
X_7 – облачность нижняя: x_{70} – Обл.н≤ 3, x_{71} – Обл.н$\in (3,9)$, x_{72} – Обл.н≥ 9;
X_8 – скорость ветра: x_{80} – Ск.в≤ 1, x_{81} – Ск.в $\in (1,2]$, x_{82} – Ск.в >2.

Теперь можно подсчитать четкие выборочные частоты n^i_{kj} для каждого интервала интенсивности и каждого класса каждой активности. Например, для класса x_{00} активности X_0 из имеющихся в наличии случаев землетрясений подсчитывается количество случаев «шумов», «средних землетрясений» и «сильных землетрясений», для которых величина активности – напряженность электрического поля – имеет значение НЭП≤ 5.

Затем при помощи функций совместимости и четких выборочных частот вычисляются нечеткие выборочные частоты $\tilde{n}^i_{kj}$. Нарпример,

$$\tilde{n}^0_{00} = \mu^0_0 \cdot n^0_{00} + \mu^0_1 \cdot n^1_{00}; \quad \tilde{n}^1_{00} = \mu^1_0 \cdot n^0_{00} + \mu^1_1 \cdot n^1_{00} + \mu^1_2 \cdot n^2_{00}; \quad \tilde{n}^2_{00} = \mu^2_1 \cdot n^1_{00} + \mu^2_2 \cdot n^2_{00}.$$

Значения нечетких весов вычисляются как отношение суммы нечетких частот класса активности к сумме нечетких частот всех трех ее классов. Например,

$$w_{00} = \frac{\tilde{n}^0_{00} + \tilde{n}^1_{00} + \tilde{n}^2_{00}}{\sum_{i=0}^{2} (\tilde{n}^0_{0i} + \tilde{n}^1_{0i} + \tilde{n}^2_{0i})}.$$

Нечеткие же относительные частоты – как отношение каждой нечеткой частоты каждого подфактора к сумме всех нечетких частот этого подфактора. Например,

$$\tilde{f}^0_{00} = \frac{\tilde{n}^0_{00}}{\tilde{n}^0_{00} + \tilde{n}^1_{00} + \tilde{n}^2_{00}}; \quad \tilde{f}^1_{00} = \frac{\tilde{n}^1_{00}}{\tilde{n}^0_{00} + \tilde{n}^1_{00} + \tilde{n}^2_{00}}; \quad \tilde{f}^2_{00} = \frac{\tilde{n}^2_{00}}{\tilde{n}^0_{00} + \tilde{n}^1_{00} + \tilde{n}^2_{00}}.$$

В результате проведенных вычислений получаем следующую таблицу:

Активность		n^0_{kj}	n^1_{kj}	n^2_{kj}	$\tilde{n}^0_{kj}$	$\tilde{n}^1_{kj}$	$\tilde{n}^2_{kj}$	$\sum_i \tilde{n}^i_{kj}$	w_{kj}	$\tilde{f}^0_{kj}$	$\tilde{f}^1_{kj}$	$\tilde{f}^2_{kj}$
X_0	x_{00}	2	0	1	1.8794	0.3641	0.6552	2.8987	0.1489	0.6484	0.1256	0.2260
	x_{01}	5	5	4	4.6984	4.216	3.9692	12.884	0.6619	0.3647	0.3272	0.3081
	x_{02}	0	2	2	0	1. 832	1.8498	3.6818	0.1892	0	0.4976	0.5024
X_1	x_{10}	3	3	2	2.8190	2.3 84	2.1194	7.3224	0.3762	0.3850	0.3256	0.2894

	x_{11}	1	1	3	0.9397	1.6442	2.2353	4.8192	0.2476	0.1950	0.3412	0.4638
	x_{12}	3	3	2	2.8190	2.3840	2.1194	7.3224	0.3762	0.3850	0.3256	0.2894
X_2	x_{20}	3	2	3	2.8190	2.1961	2.5050	7.5201	0.3864	0.3749	0.2920	0.3331
	x_{21}	2	2	3	1.8794	2.1961	2.5050	6.5805	0.3381	0.2856	0.3337	0.3807
	x_{22}	2	3	1	1.8794	2.0199	1.4642	5.3635	0.2756	0.3504	0.3766	0.2730
X_3	x_{30}	1	2	2	0.9397	1.8320	1.8498	4.6215	0.2374	0.2033	0.3964	0.4003
	x_{31}	3	5	3	2.8190	3.8519	3.314	9.9849	0.5130	0.2823	0.3858	0.3319
	x_{32}	3	0	2	2.8190	0.7282	1.3104	4.8576	0.2496	0.5803	0.1499	0.2698
X_4	x_{40}	3	2	1	2.8190	1.4680	1.1946	5.4815	0.2816	0.5143	0.2678	0.2179
	x_{41}	2	4	4	1.8794	3.6641	3.6995	9.2430	0.4749	0.2033	0.3964	0.4003
	x_{42}	2	1	2	1.8794	1.2801	1.5801	4.7396	0.2435	0.3965	0.2701	0.3334
X_5	x_{50}	2	2	1	1.8794	1.4679	1.1946	4.5419	0.2333	0.4138	0.3232	0.2630
	x_{51}	4	3	5	3.7587	3.4763	4.0850	11.32	0.5816	0.3320	0.3071	0.3609
	x_{52}	1	2	1	0.9397	1.4679	1.1946	3.6022	0.1851	0.2609	0.4075	0.3316
X_6	x_{60}	4	2	3	3.7587	2.1961	2.5050	8.4598	0.4346	0.4443	0.2596	0.2961
	x_{61}	2	4	2	1.8794	2.9359	2.3891	7.2044	0.3701	0.2609	0.4075	0.3316
	x_{62}	1	1	2	0.9397	1.2801	1.5801	3.7999	0.1952	0.2473	0.3369	0.4158
X_7	x_{70}	2	3	2	1.8794	2.3840	2.1194	6.3828	0.3279	0.2944	0.3735	0.3321
	x_{71}	4	4	4	3.7587	3.6641	3.6995	11.122	0.5714	0.3379	0.3294	0.3326

	x_{72}	1	0	1	0.9397	0.3641	0.6552	1.9590	0.1006	0.4797	0.1859	0.3345
X_8	x_{80}	1	4	4	0.9397	3.6641	3.6995	8.3033	0.4266	0.1132	0.4413	0.4455
	x_{81}	5	2	2	4.6984	1.8320	1.8498	8.3802	0.4305	0.5607	0.2186	0.2207
	x_{82}	1	1	1	0.9397	0.9160	0.9249	2.7806	0.1429	0.3379	0.3294	0.3326

Теперь все величины, необходимые для принятия решения, существуют, и можно сделать прогноз.

Допустим, необходимо исследовать новый случай и значения описывающих его активностей

6.708 21.175 30.125 914.99 20.96 85.5 7.125 6 1.

Этой совокупности активностей соответствует следующий набор строк таблицы:

$$x_{01},\ x_{12},\ x_{22},\ x_{31},\ x_{42},\ x_{52},\ x_{61},\ x_{71},\ x_{80}\ .$$

Тогда вектор нечетких весов

$$\vec{w} = (0.6619,\ \ 0.3762,\ \ 0.2756,\ \ 0.5130,\ \ 0.2435,\ \ 0.1851,\ \ 0.3701,\ \ 0.5714,\ \ 0.4266)\ ,$$

а матрица нечетких относительных частот

$$\tilde{f} = \begin{pmatrix} 0.3647 & 0.3272 & 0.3081 \\ 0.3850 & 0.3256 & 0.2894 \\ 0.3504 & 0.3766 & 0.2730 \\ 0.2823 & 0.3858 & 0.3319 \\ 0.3965 & 0.2701 & 0.3334 \\ 0.2609 & 0.4075 & 0.3316 \\ 0.2609 & 0.4075 & 0.3316 \\ 0.3379 & 0.3294 & 0.3326 \\ 0.1132 & 0.4413 & 0.4455 \end{pmatrix}.$$

Проведя линейный многофакторный синтез, получим взвешенный вектор возможных решений

$$\vec{D} = \vec{w} \cdot \tilde{f} = (\ 1.11039,\ \ 1.30928,\ \ 1.20374\ )\ ,$$

которому соответствует мера возможности

$$\overrightarrow{Poss} = \frac{\vec{D}}{\max\limits_{j} D_j} = (\ 0.84809,\ \ 1,\ \ 0.91939\ ),$$ где $D_j -\ j$ -тая компонента вектора $\vec{D}$.

Воспользуемся принципом максимума возможностей и, окончательно, получим прогноз

$$D_{Class} = 1 \quad \left(\Rightarrow\ \ M_1 \equiv \textit{Среднее землетрясение} \ \right).$$

Полученный результат соответствует статистическим данным: значения прогнозирующих активностей для приведенного случая соответствуют реальным данным за 12 июля 1978 года, когда в $11^{\underline{00}}$ произошло землетрясение с магнитудой 4.4 (по нашей классификации – «среднее землетрясение»).

4. С целью тестирования метода были рассмотрены 80 случаев «шумов» (дней, когда землетрясение не наблюдалось) и 20 случаев случайным образом выбранных землетрясений как «средних», так и «сильных». Исходные данные являются результатами измерений метеорологической станции Душетского района Грузии и Гидрометцентра Грузии за период 1967-1992 гг.

В результате проверки прогноза предлагаемым методом историческая точность оказалась равной ≈70%. Это можно считать вполне удовлетворительным результатом, тем более, что геофизические активности атмосферы не являются главными факторами-предвестниками землетрясения.

Используя предложенный метод следует учитывать, что между прогнозирующими активностями и прогнозируемым объектом должна существовать заметная корреляция, а между активностями – зависимость. Кроме того, во избежание статистического эффекта, набор первичных классических частот не должен содержать большого количества нулевых значений. И наконец, в пользу предложенного метода говорит тот факт, что для получения удовлетворительного результата прогноза достаточно сравнительно небольшого количества исходных данных[2,6,7].

Литература:

1. Li Zuoying, Chen Zhenpei and Li Jitao – A Model of Weather Forecast by Fuzzy Grade Statistics – FSS 26, N 3, June (1988), pp. 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), 1999.
4. Wang Peizhuang – Fuzzy Sets and its Application – Publishing House of Science and Technology, Shanghai, 1983.
5. Л.Заде – Понятие лингвистичекой переменной и его применение в теории принятия решений, Москва, 1983.
6. I. Khutsishvili, An application of the statistical method of fuzzy grades’ analysis - Bulletin of the Georgian Academy of Sciences, Vol.173, № 2, 2006, pp. 266-268.
7. I. Khutsishvili, Statistical Method of Fuzzy Grades' Analysis for Forecast Modeling - Applied Mathematics, Informatics and Mechanics, Vol.1, 2006, pp.12-19.

Статья получена: 2008-11-29