MAKING CLASSIFICATION FORECASTING OF WEATHER BY MULTIVARIABLE MEMBERSHIP FUNCTION

Li Zuoyong* Wang Peizhuang**

*Chengdu Meteorological Institute, Chengdu, Sichuan, China

**Dept. of Math., Beijing Normal University, Beijing, China

In this paper, a method of making the classification forecating of weather by means of multivariable membership functions is presented. This method has been proved with example, and the accuracy for the method is 96%.

Keywords: Theory of Fuzzy Sets, Forecating of Weather.

1. Introduction

In the past few years, some efforts at weather by fuzzy sets have been made and have had some successes. [1]

It is known to all, multi-forecating indexes are necessary for making forecating of weather.

If regard classification forecating of prediction object U (For example, forecating of rainfall is divided into three grades ——"full of rainfall", "normal of rainfall", and "short of rainfall") as fuzzy sets $\underset{\sim}{A}$, $\underset{\sim}{B}$, $\underset{\sim}{C}$, ... in the discussion universe U. Then, only membership degree of sample $u(\in U)$ for fuzzy sets $\underset{\sim}{A}$, $\underset{\sim}{B}$, $\underset{\sim}{C}$, ...are calculated respectively, and based on principle of maximum membership degree in fuzzy model recognition.We can make rainfall forecating according to this sample. Because a prediction object is concerned with multi-forecating indexes, the key is to design multivariable membership function.

Based on the method of least squares, this paper designs a linear model of multi-indexes parameter , and determines parameter vector β, then designs multivariable membership function, and calculates the membership degree of sample to be predicted. Finally, we take a threshold values as limits of evalution grades, and make forecating of weather according to the grade of membership degree value.

2. The Theory and Method of Founding Forecating Model of Weather Applying Multivariable Membership Function

If prediction object is the rainfall in certain region, so we can take our discussion in the set U collecting of all possible vector of predictive indexes. "Full of rainfall", "normal of rainfall", and "short of rainfall" are fuzzy sets of U, they are represented $\underset{\sim}{M}$, $\underset{\sim}{N}$, and $\underset{\sim}{F}$ respectively.

Selecting n numbers of historical samples in U. They include n_1 numbers of "full of rainfall", n_2 numbers of "normal of rainfall" ; n_3 numbers of "short of rainfall", here $n_1+n_2+n_3=n$. We determine p items of predictive index for each historical sample. The determined data of p items of the number i of sample make up of p dimensional vector

$$(x_{i1}, x_{i2}, \ldots, x_{ip}) \qquad i=1, 2, \ldots, n.$$

Taking discussion universe $U=\{u \mid u(x_1, x_2, \ldots, x_p)$, x_j is index of j item$\}$ as mapping[2]

$$f(u)=\begin{cases} 1 & \text{if } u\in M \\ 0 & \text{if } u\in N \\ -1 & \text{if } u\in F \end{cases}$$

for

$$M=\{u \mid u \text{ is "full of rainfall"}\}$$
$$N=\{u \mid u \text{ is "normal of rainfall"}\}$$
$$F=\{u \mid u \text{ is "short of rainfall"}\}$$

and definiting

$$Y_{n\times 1} \triangleq \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \\ -1 \\ \vdots \\ -1 \end{pmatrix} \begin{matrix} \left.\right\} n_1 \\ \left.\right\} n_2 \\ \left.\right\} n_3 \end{matrix} ,$$

$$X_{n\times(p+1)} \triangleq \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} .$$

If rainfall is a linear model of p item indexes:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i , \quad (i=1, 2, \ldots, n) \qquad (1)$$

or it is represented as vector

$$Y = X\beta + \varepsilon , \qquad (2)$$

here

$$\beta_{(p+1)\times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} , \qquad \varepsilon_{n\times 1} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} ,$$

β is parameter to be determined , ε_i is normal random variable, and $E(\varepsilon_i) = 0$, $D(\varepsilon_i) = \sigma^2$, $(i= 1, 2, \ldots, n)$, σ —— constant.

Because of $n>p$, the question becomes to solve the minimum of residual vector $\varepsilon = Y - X\beta$. Namely, to solve the generalized solution of transcendental equation sets $Y=X\beta$. The solution are

$$\hat{\beta} = (X^T X)^{-1} X^T Y , \qquad \text{if } \operatorname{rank}(X^T X) = p+1 \qquad (3)$$

$$\hat{\beta} = (X^T X)^{+} X^T Y , \qquad \text{if } \operatorname{rank}(X^T X) < p+1 \qquad (4)$$

Here, X^T is transpose of X, $\hat{\beta}$ is estimation of least squares of β, "+" presents generalized inverse of Moore—Penrose.

To design multivariable membership function

$$\mu_{\underset{\sim}{M}}(u) = \mu_{\underset{\sim}{M}}(x_1, x_2, \ldots, x_p) = \frac{\beta_0 + \sum_{j=1}^{p} \beta_j x_j - a}{b} . \qquad (5)$$

here, a, b are constants, according to the nature of problem.

At last, selecting threshold λ_1 and λ_2, the $\underset{\sim}{M}$ is divided into three sets

$$\begin{aligned} M_1^* &= \{ u \mid \mu_{\underset{\sim}{M}}(u) \geqslant \lambda_1 \} \\ M_2^* &= \{ u \mid \lambda_1 > \mu_{\underset{\sim}{M}}(u) > \lambda_2 \} \\ M_3^* &= \{ u \mid \mu_{\underset{\sim}{M}}(u) \leqslant \lambda_2 \} \end{aligned} \qquad (6)$$

So, the rainfall is divided into three grades, and the rainfall of each sample to be determined can be forecated according to the membership function $\mu_{\underset{\sim}{M}}(u)$.

3. The Example of Forecating

In order to explain the application of this method, we take the rainfall prediction of ×× region as the example. Selecting three prediction indexes

X_1'——The mean maximum temperature of March in the past three years,

X_2'——The mean maximum temperature of first 10 days of March in the past three years,

X_3'——The mean temperature of the past three years.

Taking historical data of rainfall in May from 1957 to 1980 as samples. So n=24. If selecting

$R > 18.5$ mm	"full of rainfall"
$18.5 \geqslant R > 11.5$ mm	"normal of rainfall"
$R \leqslant 11.5$ mm	"short of rainfall"

then, $n_1 = 8$, $n_2 = 3$, $n_3 = 13$.

The original data of above 3 prediction indexes of all 24 samples are normalized according to the formula

$$x_{ij}=\frac{x'_{ij}-(x'_{ij})_{min}}{(x'_{ij})_{max}-(x'_{ij})_{min}} \qquad (7)$$

(i=1,2, ..., n, j=1, 2, 3)

The numerical value after normalization are shown in table 1.

With program calculation for formula (3), the matrix $X_{24\times1}$ are made up of normalized data of three prediction indexes in table 1, and

$$Y_{24\times1}=\begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ 0 \\ -1 \\ \vdots \\ -1 \end{pmatrix} \begin{matrix} \left.\right\} 8 \\ \left.\right\} 3 \\ \left.\right\} 13 \end{matrix} \qquad (8)$$

At last, we can get numerical value of parameters β_0 and β_i (i=1, 2, 3) as follow

$$\beta_0= -0.952211$$

$$\beta_1= 0.570905$$

$$\beta_2= -0.147114$$

$$\beta_3= 0.488126 \qquad (9)$$

Designing the multivariable membership function according to formula (5), here, a= -0.830179, and b=0.823295 are minimum value of 24 samples based on linear polynomial $\left\{ \beta_0+\sum_{j=1}^{3}\beta_j x_j \right\}$ calculation and the difference between maximum and minimum. So, there is

$$\mu_{\underset{\sim}{M}}(u)=\frac{\beta_0+\sum_{j=1}^{3}\beta_j x_j+0.830179}{0.823295} . \qquad (10)$$

Table 1.

classification	year	x_{i1}	x_{i2}	x_{i3}	$\mu_{\underset{\sim}{M}}(u)$	rainfall in May	fitting
R >18.5mm(full of rainfall)	58	0.79	0.72	0.81	0.751180	50.2	✓
	60	0.96	0.38	0.66	0.840886	20.3	✓
	63	0.95	1.00	1.00	1.000000	20.5	✓
	66	1.00	0.53	0.64	0.829963	21.7	✓
	67	0.96	0.47	0.72	0.860377	48.9	✓
	69	0.91	0.42	0.31	0.666603	18.7	✓
	72	0.89	0.21	0.53	0.745645	20.5	✓
	75	0.84	0.48	0.79	0.816880	20.1	✓
18.5≥R>11.5 mm (normal of rainfall)	61	0.89	0.30	0.42	0.664345	13.6	✓
	64	0.75	0.30	0.58	0.662168	18.5	✓
	65	0.63	0.51	0.55	0.523602	15.6	✗
R ≤11.5 mm (short of rainfall)	57	0.49	0.02	0.22	0.318424	9.1	✓
	59	0.63	0.26	0.32	0.431908	10.8	✓
	62	0.81	0.31	0.28	0.524078	0.5	✓
	68	0.54	0.04	0	0.219085	4.1	✓
	70	0.54	0.02	0.15	0.311594	8.0	✓
	71	0.65	0.54	0.51	0.508394	11.5	✓
	73	0	0	0.25	0	11.5	✓
	74	0.54	0.24	0.56	0.515369	9.6	✓
	76	0.67	0.13	0.50	0.589596	2.0	✓
	77	0.30	0.17	0.36	0.242326	5.4	✓
	78	0.54	0.33	0.64	0.546718	11.1	✓
	79	0.26	0.28	0.70	0.397062	8.2	✓
	80	0.72	0.44	0.62	0.640023	9.8	✓
	81	0.18	0.37	0.51	0.212856	1.9	✓

Here, the numerical value of β_0, β_1, β_2, β_3 are taken as formula (9). The results of calculation is shown in column in table 1.

Taking

$$\lambda_1 = 0.666603 \quad \text{and} \quad \lambda_2 = 0.640023 \ .$$

Then $\underset{\sim}{M}$ is divided into three sets

$$M_1^* = \{ u \mid \mu(u) \geq 0.666603 \} \ ;$$

$$M_2^* = \{ u \mid 0.666603 > \mu(u) > 0.640023 \} \ ;$$

$$M_3^* = \{u \mid \mu(u) \leq 0.640023\} \quad .$$

Each sample u_0 can be classificated according to membership degree $\mu(u_0)$. The results of evaluation are shown in table 2. The fitting accuracy is (8+2+13)/24=96%. The trial forecating for 1981 is "short of rainfall". It was true.

Table 2.

	M	N.	F	SUM
M_1^*	8	0	0	8
M_2^*	0	2	0	2
M_3^*	0	1	13	14
SUM	8	3	13	24

4. Conclusion

Applying this method to classification forecating of weather is rational in principle and the result corresponds with practice. Because it considers the linear fitting of method of least squares of multi-prediction indexes for sample data, and draws out multi-indexes curve, designs multivariable membership function on the basis of this, and then makes evaluation according to membership degree. Especially if the number n of historical samples and number p of prediction indexes are all big numbers, the method is comparatively simple , because of operation of computer program.

In spite of making the forecating of three grades for prediction object only, this method can also be applied for more than three grades of forecating. For example, if it is needed to make 5 grades forecating, we can make mapping

$$f(u) = \begin{cases} 2 & \text{if } u \in A \\ 1 & \text{if } u \in B \\ 0 & \text{if } u \in C \\ -1 & \text{if } u \in D \\ -2 & \text{if } u \in E \end{cases} .$$

Here, A, B, C, D, E are ordinary sets. then selecting four thresholds λ_k (i=1,2,3,4), it divides $\underset{\sim}{M}$ into five ordinary sets. In such a manner, the prediction object will be evaluated.

If sample data are few, it is not suitable to classify it too minute. Because it will make sample numbers very few in each grade, and there will be diviation between fitting curve and practice case, and it will influences accuracy of forecating.

In order to increas the accuracy of historical fitting and trial-forecating , it is not only reckon on selecting parameter and conformation of membership function properly, but also reckon on the feature of selected prediction indexes. The more the correlation coefficient between indexes and prediction object, the greater the indenpendence among the indexes, the better the results of forecating.

It will be seen that just one year of "normal of rainfall" is evaluated as "short of rainfall". It was wrong. The reason may be that the historical sample numbers of "normal rainfall" is few, and therefore it influences the accuracy of forecating of "normal rainfall".

Reference

(1) Wang Peizhuang(汪培庄), Fuzzy Sets and its Application, Publish House of Science and Technology in Shanghai, 1983, 237-273, 231-233.

(2) Sun Hongyuan(孙洪元), An Evaluation of the Human Heart Function by Multivariable Membership Function, FUZZY MATHEMATICS, Vol.2, 3,1982, 99-104.