A Numerical Method of Statistical Pattern Recognition

Xu Hong (徐宏), *Li Hongji* (李洪勣) and *Wang Ronghua* (王荣华)
Air Force Meteorological Research Institute, Beijing, 100085

Received January 4, 1989

ABSTRACT

A numerical method of statistical pattern recognition is proposed in this paper. Different from the discriminatory analysis method currently used in the mathematic statistics, it is unnecessary to assume that the predictand should be subject to a certain distribution. On the contrary, the statistical relationship between predictand and predictor has been obtained directly with computer according to actual distribution to recognize the category of patterns. Result of forecast has been improved as compared with the usual analytic discriminatory method. The influence of predictor on predictand can be seen clearly from this method and the transparency is good. Therefore, it is better to use the method in very short range forecast for which causality is more obvious.

I. INTRODUCTION

Statistical pattern recognition, also called model recognition, is to recognize and analyze totality constituted by several categories of study objects according to statistical characteristics. In weather forecast, the statistical methods frequently used in forecast operations such as Fisher discrimination, Bayes discriminatory analysis and step by step discrimination are methods of statistical pattern recognition. Such statistical methods, either assuming random variable normal distribution or using linear mathematic models, yield analytic expression of discriminatory functions and judge the category of patterns according to the magnitude of value. But in fact the distribution of random variables has many forms. Mostly, the relationship between predictand and predictor is non–linear. The discriminatory function in linear or super–plane form obtained above is unable to avoid distorting even seriously distorting statistical relationship actually existing between quantities. Thus, sometimes the result is not good in practical use though discriminatory function has been strictly established. In view of this problem, some non–linear statistical models such as Logistic (Lu Chunlian and Chen Xunhua, 1986) have been proposed. Due to the difficulties in solving coefficient equations in such models, the number of factors it includes will be limited if the order of polynomial in discriminatory function is increased. Thus just like many differential equations which are unable to solve analytic solution, the statistic relationship actually existing between predictand and multiple predictors is also difficult to express with an analytic discriminatory function.

In recent years, microprocessors and flexible display equipment have been used in weather forecast operation, providing a favourable condition for the development of pattern recognition technology. In this paper a new method called the numerical method of statistical pattern recognition, is proposed, so as to be distinguished from the analytic method using discriminatory function. It is unnecessary to assume that predictand should be subject to a certain distribution. It studies and establishes the actual statistic relationship between predictand and predictors, and performs discrimination and classification according to the actual distribution of random variables. The numerical method of statistical pattern recognition is described with Bayes method as an example, due to the extensive application of Bayes discrimination to statistical pattern recognition.

II BASIC METHOD

Assuming that a pattern ξ can be depicted with several sequential numbers $x_1, x_2, \cdots, x_n$, each ξ corresponding to a vector $\vec{x}=(x_1, x_2, \cdots, x_n)^T$ or a point in n–dimension R_n. If the set B of known pattern consists of incompatible patterns of category m, then

$$\bigcup_{1}^{m} B_i = B, \quad B_i \cap B_k = \varphi \quad (i \neq k,\ i,k=1,\ 2,\ \cdots,\ m)$$

$$B_i:\ \vec{x}_j^{(i)} = (x_{j1}^{(i)}, x_{j2}^{(i)}, \cdots, x_{jn}^{(i)})^T \quad (j=1,\ 2,\ \cdots,\ N_i)\ . \tag{1}$$

Where $N=\sum_{i=1}^{m} N_i$, total samples in B, and N_i, the number of samples in category i. When the number of samples is large enough, the occurrence of corresponding B_i pattern can be approximately regarded as climatic probability

$$P(B_i) \approx \frac{N_i}{N}, \quad \sum_{i=1}^{m} P(B_i) = 1 \tag{2}$$

If every pattern $\xi \in B_i$ is regarded as a random process, with its conditional probability density distribution being $P(\xi / B_i)$, and with its conditional density distribution being $P(\vec{x} / B_i)$ in R_n, a statistical recognition method can be established according to the set of samples and climatic probability of various categories of patterns.

1. *Loss Function*

Generally, statistical pattern recognition always makes misdiscrimination. As the emphasis of recognition and climatic probability of every recognition object are different, the results of misdiscrimination are different. For instance, in forecast of small probability event like hail, different is the degree of loss where hail is misdiscriminated as no hail and no hail, as hail. Therefore, it is necessary to establish loss function $H((C(\vec{x})), j)$ according to recognition object. For m category patterns recognized, H is m^2 constant. h_{ij} $(i,j=1,2,\cdots,m)$ is the loss caused when ξ is category j pattern and discriminatory function $C(\vec{x})=i$ is misdiscriminated as category i. As described above, if hail belongs to category B_1 and no hail belongs to B_2, $h_{12}<h_{21}$ should be taken when selecting loss function, that is to say, if the object to be forecasted in particular is misdiscriminated, larger loss function must be taken. When $i=j$ e.g. recognition is correct, $h_{ij}=0$.

Actually, loss function depends on score standard. Although the objects to be recognized are the same, loss function $\{h_{ij}\}$ is different for different score standard. If accuracy is used as score standard, it is equivalent to treat loss of various misdiscriminations equally. At this time $h_{ij}=1/m(i,j=1,\ 2,\ \cdots,\ m,\ i\neq j)$ can be taken, thus accuracy of recognition would be high. If other score standards are selected, $h_{ij} \neq h_{ji}$ should be taken when climatic probabilities of category B_i and B_j are different. Therefore, in pattern recognition research, standard for evaluating forecast quality should be established first, so as to make recognition reach optimum discriminatory results of given evaluating standard.

2. *Discriminatory Function*

Assuming that in the n-dimensional characteristic space there exists a function $C(\vec{x})$ with value domain of $\{1, 2, \cdots, m\}$, when $C(\vec{x}) = i$, $\xi \in B_i$ corresponding to $\vec{x}$ is discriminated, thus function $C(\vec{x})$ is called discriminatory function. Therefore, it is necessary to divide space R_n into m subsets r_i $(1 \leqslant r \leqslant m)$ which are not intersecting. When $\vec{x} \in r_i$ let $C(\vec{x}) = i$, judge $\xi \cdot B_i$, so as to make average loss L minimum.

$$L = \sum_{i}^{m} \int_{r_i} \sum_{j=1}^{m} h_{ij} P_j p(\vec{x} / B_j) d\vec{x} \ . \tag{3}$$

Where P_j is the climatic probability of category j image and can be approximately replaced by frequency. To satisfy equation (3), the value of multiplied function $\sum_{j=1}^{m} h_{ij} P_j p(\vec{x} / B_j)$ is smaller than other $(m-1) \sum_{j=1}^{m} h_{lj} P_j p(\vec{x} / B_j)$. Thus the discriminatory standard of Bayes method is:

$$\sum_{j=1}^{m} h_{ij} P_j p(\vec{x} / B_j) \leqslant \sum_{j=1}^{m} h_{lj} P_j p(\vec{x} / B_j) \tag{4}$$

$$(i, l = 1, 2, \cdots, m, \quad l \neq i) \ .$$

$C(\vec{x}) = i$ is discriminated, namely, ξ belongs to category i image.

3. *Conditional Probability Density Function*

It can be known from (4) that it is important to know conditional probability density function $p(\vec{x} / B_i)$ in doing pattern recognition. Here we begin with actual distribution of samples to find its corresponding conditional probability density function.

Conditional probability density function can be established from sample set (1) by using the features of potential function.

$$p(\vec{x} / B_i, h^2) = \frac{1}{N_i} \sum_{j=1}^{N_i} K(|\vec{x} - \vec{x}_j|, h^2) \ . \tag{5}$$

where $K(|\vec{x} - \vec{x}_j|, h^2)$ is potential function and it has the following features: $K(|\vec{x} - \vec{x}_j|, h^2) = K(|\vec{x}_j - \vec{x}|, h^2)$. When $\vec{x} = \vec{x}_j$, $K(|\vec{x} - \vec{x}_j|, h^2)$ reaches an extremely large value. As $(|\vec{x} - \vec{x}_j|)$ increases, K value gradually decreases until it tends to be or is equal to zero. h^2 is scale parameter. In analysis we select the following form of potential function:

$$K(|\vec{x} - \vec{x}_j|, h^2) = \frac{1}{(h^2 \pi)^{n/2}} \exp\left[-\frac{(x_1 - x_{1j})^2 + (x_2 - x_{2j})^2 + \cdots + (x_n - x_{nj})^2}{h^2} \right] ,$$

here, $x_1, x_2, \cdots, x_n$ are variables after standardization, substituting the above equation into (5), we have:

$$p(\vec{x} / B_i, h^2) = \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{(h^2 \pi)^{n/2}} \exp$$

$$\left[-\frac{(x_1 - x_{1j})^2 + (x_2 - x_{2j})^2 + \cdots + (x_n + x_{nj})^2}{h^2} \right] . \tag{6}$$

Thus $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\vec{x}/B_i, h^2)d\vec{x} = 1$. It can be shown that $p(\vec{x}/B_i, h^2)$ is the gradual non-deviative estimate of conditional probability density function $p(\vec{x}/B_i)$.

It can be known from (6), h^2 represents the influence of sample point $\vec{x}_i$ on point $\vec{x}$ in space. When value h^2 is small, the role of sample points near $\vec{x}$ will be stressed. When h^2 is large, the role of sample points farther from $\vec{x}$ can not be ignored. It is the function of sample number N_i, when sample patterns are many, smaller value can be taken as h^2. On the contrary, when samples are less, larger value should be taken.

As for subject determined, climatic probability of every category of patterns are known when loss function $|h_{ij}|$ is given, classification of subject can be made according to (4) and (6).

Generally, category m patterns in n–dimension can be classified according to (4) and (6). But when n increases, the time of computation will be largely increased. Therefore, when preparing forecast with this method, attention should be paid to selecting factors closely related to predictand and mutually independent as coordinates, so as to accomplish classification of category m patterns when dimension is relatively low. The specific procedure is to select the group of best discriminatory quality from combining two from n factors according to score standard to serve as the first and second coordinates, distinguishing between the certitude and incertitude regions in category m patterns. Then select from $n-2$ factor the third coordinate which matches the first two coordinates quite well and classify the incertitude regions, and do it by analogy. This method is called sequential discriminatory method. The quantity of computation and occupied internal storage can be reduced by using this sequential discriminatory method.

III. DISCRIPTION OF THE APPLICATION OF THIS METHOD

In considering that the two–dimensional diagram is convenient and visual, and it is the basis of commonly used methods such as sequential discriminatory methods, the recognition of two–dimension and two–category problems is taken as an example in order to illustrate the numerical method of pattern recognition. The numerical recognition of two–dimensional patterns can be regarded as a problem of objectification and automation of preparation, and correction of scatter diagram, and the improvement upon traditional scatter diagram method.

Data of non–shear type are quoted from thunderstorm forecast method in Hankou, Hubei Province as samples, of which 26 days with thunderstorms, 148 days without thunderstorms, the climatic probability of thunderstorm days is 15%. Condensation level H_c, and air pressure of the top layer releasing unstable energy P_h are used as factors to know if there are thunderstorms in the future 12 hours (Xiao Guizhen and Li Hongji, 1978).

In the two–dimensional problem, equation (6) is simplified as:

$$p(\vec{x}/B_i, h^2) = \frac{1}{N_i} \sum_{i} \frac{1}{h^2 \pi} \exp\left[-\frac{(x_1 - x_{1i})^2 + (x_2 - x_{2i})^2}{h^2} \right]. \tag{7}$$

In operational work, we took the grids 15 × 15 for data after the procedure of being standardized and tested $h^2 = 1 \sim 10$ and selection of various loss functions. Conditional probability density distribution for thunderstorms category B_1 and no thunderstorms category B_2 at recognition space grids were separately computed according to (7). Figures 1 and 2 give the forms of conditional probability density distribution for these two categories of predictands, in which broken lines are for $b^2 = 1$ and solid lines are for $b^2 = 5$. It can be seen from these two figures, when samples are constant, if different values are taken for b^2, contour lines of conditional probability density will change accordingly. When b^2 is small, it can slightly show the local characteristics of distribution; When b^2 is large, distribution of contour lines are relatively smooth, which is favourable to showing the general characteristics of distribution.

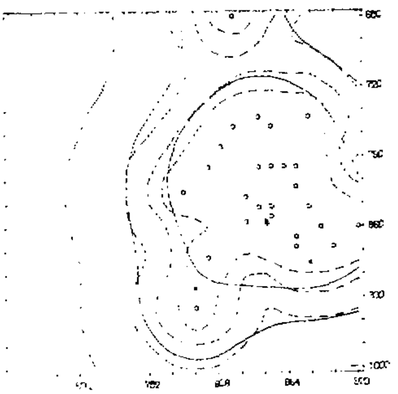

Fig. 1. Distribution of conditional probability density function $p(\vec{x}/B_1, b^2)$. Broken lines: $b^2 = 1$, solid lines: $b^2 = 5$.

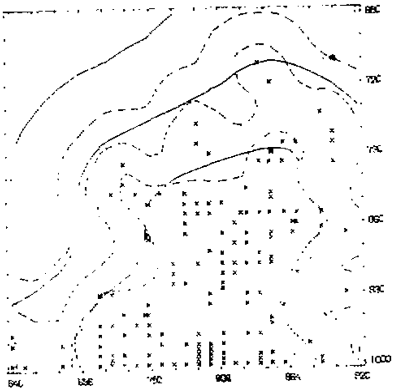

Fig. 2. Distribution of conditional probability density function $p(\vec{x}/B_2, b^2)$. Broken lines: $b^2 = 1$, solid lines: $b^2 = 5$.

For these two categories of problems m = 2, (4) can be simplified to divide two categories of regions r_1 and r_2, enabled to satisfy:

$$r_1:\ h_{12}P_2p(\vec{x}/B_2, b^2) \leq h_{21}P_1p(\vec{x}/B_1, b^2)$$

$$r_2:\ h_{21}P_1p(\vec{x}/B_1, b^2) \leq h_{12}P_2p(\vec{x}/B_2, b^2)$$

Let $$d(\vec{x}) = \ln\frac{h_{21}P_1p(\vec{x}/B_1, b^2)}{h_{12}P_2p(\vec{x}/B_2, b^2)}. \tag{8}$$

It can be known from the comparison of the above two equations.

When $d(\vec{x}) \leq 0$, discriminate $\vec{x} \in r_1$

When $d(\vec{x}) > 0$, discriminate $\vec{x} \in r_2$.

Function $d(\vec{x})$ can also be called discriminatory function and equation $d(\vec{x}) = 0$ is called judgement boundary. The curves determined by it are boundary of two categories of recognition objects. But it can be seen from (7), generally it is difficult to obtain an analytic expression of $d(\vec{x})$. In practical work, we obtained it by taking $d(\vec{x})$ values of various grid points, by interpolating to yield grid points with $d(\vec{x}) = 0$, and by connecting spline function. Fig. 3 is the discriminatory boundary output by computer based on the two categories of conditional probability density distributions in Fig. 1 and Fig. 2 when $h_{12} = 0.2$, $h_{21} = 0.8$. Similarly, broken lines in the figure correspond to $b^2 = 1$ and solid lines correspond to $b^2 = 5$. It can also be seen from the figure that different values are taken for b^2, which has a large influence on the shape of discriminatory boundary. The larger value b^2 has, the smoother the discriminatory boundary is. On the contrary, the smaller value b^2 has, the more complicated the discriminatory boundary will be.

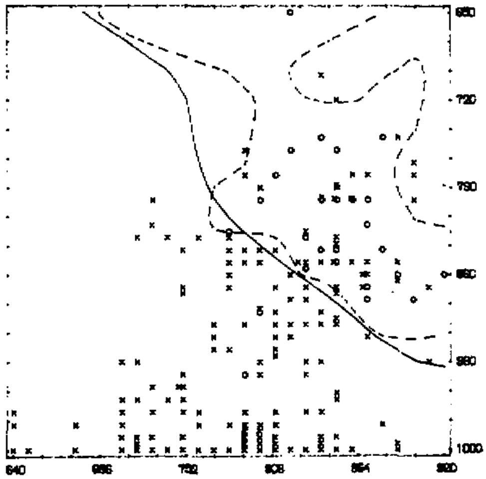

Fig. 3. Diagram of numerical discrimination when $h_{12}=0.2$ and $h_{21}=0.8$. Broken lines: $b^2=1$, solid lines: $b^2=5$

It can be known from (8) that the taking of loss function is different, so discriminatory boundary is also different. Actually, the criterion of discrimination is to make average loss of erroneous discrimination minimal and score highest under given score standard. At present score standards have been not unified, some people have used accuracy (the ratio of number of correct forecast to total number of forecast), missing rate (ratio of number of weather phenomenon not forecast to total number of weather phenomenon appeared), false alarm rate (ratio of number of weather phenomenon forecast that appeared but actually did not appear to total number of forecasting appeared), CSI (ratio of number of weather phenomenon forecast to the sum of number of weather phenomenon that appeared and number of false forecast), etc. Table 1 gives three numerical discriminatory results when $b^2=1$ with loss function $h_{12}=h_{21}=0.5$ (namely missing and false are of equal value, and correspond to the score method of accuracy); $h_{12}=0.3$, $h_{21}=0.7$; and $h_{12}=0.15$, $h_{21}=0.85$. The total number of sample is 174, number of weather that appeared is 26, figures in the table are the number of corresponding forecast results, and figures in brackets are the corresponding frequency. It can be seen obviously from the first row in the table that though its accuracy is high, yet missing is too much, that is to say, if only accuracy is used as discriminatory standard for such small probability events as thunderstorm, the result is not so good. By comparison, though the accuracy for the last two rows is lower, yet missing has been reduced largely. CSI score is obviously higher than the first row, and it seems more reasonable. This also shows that the difficulties in discrimination caused by the large difference between two categories of samples can be overcome by changing loss function.

Table 1. Numerical Discriminatory Results When $b^2=1$

h_{12}, h_{21}	Accuracy	Missing	False	CSI
0.5, 0.5	156 (0.90)	17 (0.65)	1 (0.10)	0.33
0.3, 0.7	150 (0.86)	4 (0.15)	20 (0.48)	0.48
0.15, 0.85	144 (0.83)	1 (0.04)	29 (0.54)	0.45

Shapes of boundary may be various. Some are closed curves, others may have more than two boundaries. We have worked out procedures to discriminate category which each point belongs to according to the orientation of boundaries. In the process of establishing forecast method, corresponding score results can be obtained while displaying boundaries. After giving loss function, adjust h^2, $p(\vec{x}/B_i, h^2)$ distribution changes, boundary also changes accordingly. When determining boundary, the score should be considered on the one hand and the stability of application should be considered on the other hand. Generally speaking, when h^2 value is small, score is better than that of bigger h^2 value. Generally, the smoother the better except that the boundary in area of denser sample points can change greatly. In Fig. 3 there is a relatively large difference between these two boundaries. With $h^2=1$, accuracy is 84% and missing is 8%; When $h^2=5$, accuracy is 78% and missing is 12% respectively. But from the representation of boundary and stability of operation, the latter may be slightly better.

After determining the boundary of two-dimensional problems or after dividing multi-dimensional problems r_i, recognition can be performed. In operation, the category to which it belongs can be obtained according to coordinate $\vec{x}$ of the predictand and display conditional probability density of predictand of each category near forecast point, providing reference to the forecaster. When actual observation result appears, forecast point becomes historic sample point. During the process of application, method should be self-adjusted. Each time it is used, it is equivalent that sample number of category i pattern changes from N_i to N_i+1. Distribution of $p(\vec{x}/B_i, h^2)$ for N_i+1 sample is computed from (6) on the basis of N_i. Thus, the division of boundary or region r_i is corrected. This adjustment method is called adaptive method. The accuracy of discrimination will increase as the sample increases, called "learning".

IV. COMPARISON WITH ANALYTIC DISTRIMINATORY ANALYSIS METHOD

Forecast of most weather phenomena can be summarized up as classification of predictand. In order to obtain mathematic expression of discriminatory function, some assumptions must be made in discriminatory analysis to simplify problems. An analysis is made for analytic discriminatory method and numerical discriminatory method by taking Bayers discrimination as an example.

In Bayers discrimination, usually an assumption is made that conditional probability density functions of each category of recognition objects are normal distributions of several elements.

$$p(\vec{x}/B_i)=\frac{1}{(2\pi)^{n/2}\left|\sum_i\right|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\vec{x}-\mu^{(i)})^T\sum_i{}^{-1}(\vec{x}-\mu^{(i)})\right]. \tag{9}$$

Where $\mu^{(i)}$ is the mathematical expectation of random vector $\vec{x}$ in category B_i, $\sum_i$ is covariance matrix of $\vec{x}$, $\left|\sum_i\right|$ is $\sum_i$ determinant. In practical operation, we can discriminate with discrimination function after estimating directly parameters such as mathematical expectation and covariance matrix etc. of normal distribution from sample data. But the computation is complicated when total covariance matrix of various categories is different. The linear discriminatory function used in operational work is obtained with assumption that various categories of covariance matrices are the same (Tu Qipu et al., 1984). In fact, substituting (9) into (8) will yield

$$d(\vec{x}) = -\ln\frac{p(\vec{x}/B_1)}{p(\vec{x}/B_2)} + C = \frac{1}{2}(\vec{x}-\mu^{(1)})^T \sum{}^{-1}(\vec{x}-\mu^{(1)}) - \frac{1}{2}(\vec{x}-\mu^{(2)})^T \cdot$$

$$\sum{}^{-1}(\vec{x}-\mu^{(2)}) + C = \vec{x}' \sum{}^{-1}(\mu^{(2)}-\mu^{(1)}) - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})' \sum{}^{-1}(\mu^{(2)}-\mu^{(1)}) + C \ . \quad (10)$$

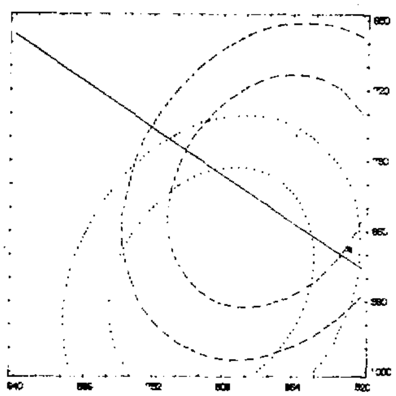

Fig.4 Conditional probability density of normal distribution and discriminatory boundary when covariances are the same. Dotted lines are isa-$p(\vec{x}/B_2)$, broken lines are isa-$p(\vec{x}/B_1)$, solid lines are discriminatory boundary (similarly here inafter).

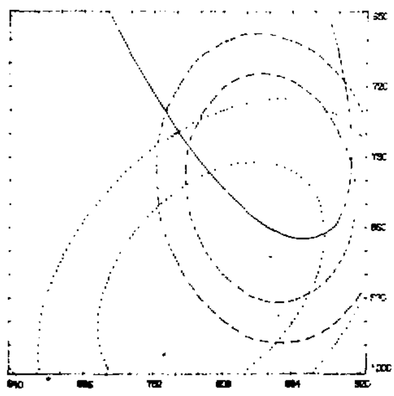

Fig.5. Conditional probability density and discriminatory boundary of normal distribution when covariances are different.

Obviously, it is a linear function of components $x_1, x_2, \cdots, x_n$ of $\vec{x}$. Accordingly, when $n>2$, judgement boundary is a superplane of n–dimension. Still taking the sample of the above thunderstorm forecast as an example, assuming that these two categories of predictands are normal distribution and covariance matrices are identical, compute its conditional probability density. We take mathematical average of covariance of two categories of corresponding predictands as common covariance. Fig. 4 shows the equivalent probability density distribution of two categories of predictands. It can be seen from this figure that two categories of curves are entirely the same in shapes on the plane. Because their mathematical expectations are different, the center of large value is different, distribution curves of one category can be also regarded as the result of translation of that of the other category. Discriminatory boundary in the figure is the straight line obtained with loss function $h_{12}=0.3$ and $h_{21}=0.7$. Compared with Figs. 1, 2 and 3, its distribution can be regarded as a simplification of actual distribution. It is worthy to note that there is a large distribution in category–one predictand.

We relax requirement with only the assumption of normal distribution remained and the covariance matrices different. Results of contour lines of conditional probability density and boundaries of these two categories of predictands are shown in Fig. 5. Compared with Fig. 4, the orientations of long axis of two groups of closed lines are inconsistent and the corresponding boundaries have also changed to quadratic curves from straight lines. The analytic expression of this boundary is

$$d(\vec{x}) = -\frac{1}{2}\ln\frac{|\sum_2|}{|\sum_1|} + \frac{1}{2}(\vec{x}-\mu^{(1)})^T\sum_1^{-1}(\vec{x}-\mu^{(1)}) - \frac{1}{2}(\vec{x}-\mu^{(2)})^T\sum_2^{-1}(\vec{x}-\mu^{(2)}) + C. \quad (11)$$

It has already been a quadratic function, usually a quadratic super surface. It can be seen that whether covariance matrices are the same or not, there is a great influence on discriminating boundary. It can be known from comparing with (8), in equations (10) and (11) that:

$$C = \ln(h_{12}P_2 / h_{21}P_1) \ .$$

Here P_1 and P_2 are known; a family of discriminatory boundaries can be obtained by changing values of h_{12} and h_{21}. This family of curves are parallel to each other. Therefore, in analytic discrimination, concept of loss function is not quoted: appropriate boundary is selected by changing constant term and translating boundary. Compared with Fig. 2 and Fig. 3, the corresponding iso–density lines and boundary in Fig. 5 are very similar to it, and is a better approximation of actual distribution. This is because the actual distribution of these two categories of predictands in this case are nearly normal distributions. Table 2 shows the results of two types of normal distribution and numerical discrimination for $b^2=1$ and $b^2=5$ when $h_{12}=0.3$ and $h_{21}=0.7$ are taken for loss function. It can be known from this table that result with different covariance ($\sum_1 \neq \sum_2$) is better than that with the same covariance ($\sum_1 = \sum_2$) but almost identical to numerical discriminatory results of $b^2=5$.

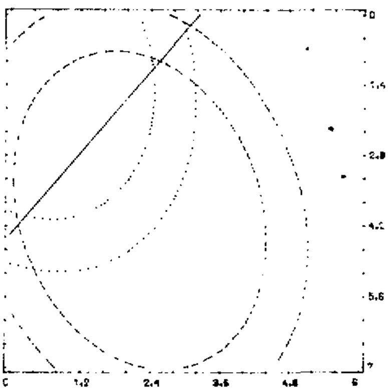

Fig. 6. Normal distribution $\sum_1 = \sum_2$.

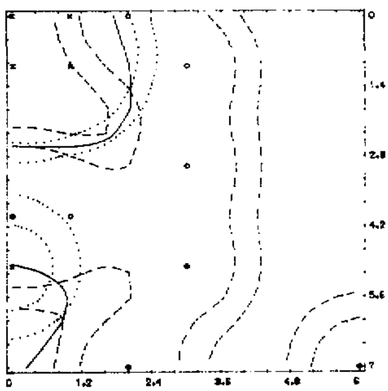

Fig. 7. Numerical method $b^2=5$.

When conditional probability density of each category of predictand is largely different from normal distribution, then assumption of normal distribution is untenable. We have done the computation with the data used in logistic discriminatory analysis in Table 1 in the reference (Lu Chunlian and Chen Xunhua, 1986). Where x_1 is the number of days with highest temperature $\geqslant 35$ ℃ from late May to early June; x_2 is the number of days with highest temperature $\geqslant 35$ ℃ in late June. Then discriminating whether the number of days with highest temperature $\geqslant 35$ ℃ is equal to or less than 2 days in late July. We have computed respectively the iso–probability density curves under three kinds of conditions in which normal distribution covariance matrices are the same (Fig. 6), normal covariance matrices are different (not shown) and numerical methods (Fig. 7) and have computed the discriminatory boundaries when loss functions are the same. Forecast accuracy of three lines are 80%, 80%

and 100% separately. When actual distributions of two types of predictands are largely different from normal distributions, if analytic method is used for discrimination, both discriminatory results are obviously not as good as the discriminatory results by numerical methods no matter whether covariance matrices are the same or not.

It can be seen from this that when actual distribution of samples is close to the assumption of analytically discriminatory analysis, results obtained by analytic discrimination method and numerical method respectively are approximate. On the contrary, when actual distribution is significantly different from assumption, their results will differ greatly. It must be pointed out that discriminatory boundary same as analytic discriminatory method can be also obtained by using the computational method described in this paper without solving equations. Namely, the boundary lines in Figs. 4, 5 and 6 can be either calculated according to analytic expression or accurately drawn with numerical method. Therefore, so long as solving practical problems, the numerical recognition of statistical patterns by microcomputer is a comparatively common and efficient discriminatory method.

Table 2. Discriminatory Results under Different Assumptions When $h_{12}=0.3$ and $h_{21}=0.7$

	Accuracy	Missing	False	CSI
$\sum_1=\sum_2$	141 (0.81)	11 (0.42)	22 (0.59)	0.31
$\sum_1\neq\sum_2$	139 (0.80)	7 (0.27)	28 (0.60)	0.35
$b^2=5$	141 (0.81)	7 (0.27)	26 (0.58)	0.37
$b^2=1$	150 (0.86)	4 (0.15)	20 (0.48)	0.48

V. CONCLUSION

A numerical method of statistical pattern recognition is proposed in this paper according to the needs for forecast operation and considering the application of computers. Without the limitation of those prerequisite in analytic discriminatory method, this method has expanded the extent to which the discriminatory analysis method is able to be used. It is relatively practical to the division of discriminate boundary starting from the actual distribution of samples. Discriminatory results are better, compared with analytic method. In analytic discriminatory method, some specific influences of various factors and sample points on discriminatory boundary are often masked by some computational processes. But the influence of numerical method on this can be usually vividly analysed; it is of good transparency. And it is convenient to check faults and correct the method. Therefore, this method is of practical importance to the forecast of weather phenomena, especially to very short range forecast in which relationship between cause and effect is obvious. Of course, the analytic discriminatory method is also developing and improvement in the existing problems can be also expected.

REFERENCES

Cheng Minde et al. (1983), Introduction to pattern recognition (in Chinese), Shanghai Science and Technology Publishing House.

Lu Chunlian and Chen Xunhua (1986), Quadratic logistic discriminatory analysis and its application to meteorology (in Chinese with English abstract), Meteorological Bulletin, Vol. 44, No. 3.

Tu Qipu et al. (1984), Meteorological applied probability statistics (in Chinese), Meteorological Publishing House.

Xiao Guizhen and Li Hongji (1978), Study on method of forecasting thunderstorms in July and August in Wangjiadun, Hankou (in Chinese), Aviation Meteorology, Vol. 5, No. 1.