

# The Multi-dimensional Customer Behavior Analysis Based On Biayes and Fisher Classification Algorithm

Guanghui Wang<sup>\*a</sup>, Fengliang Jiao<sup>b</sup>

<sup>a</sup> Heilongjiang University of Science and Technology, Computer and information Engineering College, Harbin 150022, China;

<sup>b</sup> Information Engineering Department of Weifang Business Vocational College, Shandong Province, Zhucheng 262234, China.

[guanghuiwang8668@163.com](mailto:guanghuiwang8668@163.com)

The biayes classification algorithm was based on the mutual independence between one class and another class, and the reality of multidimensional existence dependency relationship with the customer class. So this paper first used the Fisher method to classify the class, then used the bias classification algorithm for customer segmentation, identified potential customers, and effectively reduced marketing costs. Finally, multi-group data were used to test by matlab. Compared to the bias algorithm, the results show that hybrid algorithm was an effective method.

## 1. Introduction

With the development of science and technology in today's society, the various aspects and amount of social information is increasingly large, and the dimensions of the data is also increasing. Classification is becoming increasingly difficult and at the same time more important (Abdelfattah et al, 2013; Bouchaala et al, 2010). According to their different social fields different classification methods can be chosen, such as vector machine classification, clustering analysis and BP neural network, etc. In numerous classification methods, Bayes classification algorithm and Fisher discriminant method have been of great value. The Bayesian classifier has been widely used because of its simple structure learning and parameter learning, classification speed when the on class and another class are independent of each other, and its lower calculation complexity. But in real life, when classes cannot do the math background independent each other, there is a mutual dependence between the dimensions of the classes. In the Fisher discriminant method of projection, the basic idea is to group projection data to a certain direction, making them a projection of classes being as separate from each other as much as possible (Aquaro et al, 2009; Cook et al, 2000).

In order to improve on this problem, this paper proposes a kind of method based on Fisher linear discriminant analysis and an improved algorithm of Bayesian classifier. The main idea of this algorithm is that by using the transformation matrix, the original training samples are transformed, and use the classifier in the projection to the new sample space for learning classification. The original sample property is concentrated, with any two may having certain dependencies between attributes, and the new samples are assumed to be independent of each other after the projection in the new sample space (Duan et al, 2009; Hao, 2010). Through the transformation, the model which can be expressed in the measurement space with high dimension is changed into the mode of representation in the feature space with lower dimension. In this way, it can effectively realize the classification and recognition, which can more accurately reflect the nature of the classification (Guo et al, 2012; Li et al, 2010).

This paper gives a kind of method based on Bayesian and Fisher algorithm of multi-dimensional customer behavior analysis on the basis of analyzing the characteristics of the Bayesian model, combined with the Fisher linear discriminant analysis..

## 2. The Bayesian classification algorithm and Fisher discriminant analysis algorithm

The Bayesian classification algorithm, using probability and statistics to classification samples, is one of the earliest methods for dealing with uncertainty by classification. It is based on maximum a posteriori probability criterion, namely the use of a certain object by probabilistic prior probability calculation, and it selects the classes with maximum a posteriori probability as the object's class (Liu, 2014; Zhang et al, 2010).

Definition 1 Bayes formula:

Set state of A and B to meet the following conditions:

(1) Any two states are incompatible, When  $i \neq j$ , there is  $A_i \cap A_j = \phi$

(2)  $P(A_i) > 0$

(3) Sample space  $D$  is a collection of  $D = \bigcup_{i=1}^n A_i$ ,

$$\text{There are } P(x) = \sum_{i=1}^n P(A_i) \times P(x|A_i) \quad (1)$$

$$P(A_i|x) = \frac{P(A_i) \times P(x|A_i)}{P(B)} \quad (2)$$

$$P(A_i|x) = \frac{P(A_i) \times P(x|A_i)}{\sum_{j=1}^n P(A_j) \times P(x|A_j)} \quad (3)$$

Among them,  $P(A_i)$  is  $A_i$  prior probability,  $P(x|A_i)$  is the conditional probability of sample  $x$  under the condition  $A_i$ .  $P(A_i|x)$  is the conditional probability of the condition  $A_i$  in the condition of sample  $x$ , also known as the posteriori probability (When the sample  $x$  is known, it belongs to the probability of the state  $A_i$ ). So sample A is assigned to the posterior probability of the largest in the class (Li L, Ma S, Zhang Y, 2014; Sun Y, Tang Y, Ding SI, 2011).

Supposing that there are  $k$   $p$ -dimensional overall  $G_1, G_2, \dots, G_k$ , The probability density function are  $f_1(x), f_2(x), \dots, f_k(x)$ . Assume that the sample  $x$  prior probability from general  $G_i$  to  $p_i (i=1,2,\dots,k)$ , then  $p_1 + p_2 + \dots + p_k = 1$ . According to the Bayes theory, the posterior probability of sample A from total B is

$$P(G_i|x) = \frac{p_i f_i(x)}{\sum_{j=1}^k p_j f_j(x)}, i=1,2,\dots,k \quad (4)$$

The following discriminant rule applies under the condition of not considering miscalculation cost:

$$x \in G_i \text{ if } P(G_i|x) = \max_{1 \leq j \leq k} P(G_j|x).$$

If misjudgment is considered,  $R_i$  means that according to certain discriminant rule it may be sentenced to a collection of all samples  $G_i (i=1,2,\dots,k)$ ,  $c(j|i) (i, j=1,2,\dots,k)$  meaning that it will from the price of the sample  $x$  of the  $G_i$  be mistaken for  $G_j$ , then  $c(i|i) = 0$ . The conditional probability of sample  $x$  comes from  $G_i$  misjudgment of  $G_j$  is  $P(j|i) = P(x \in R_j | x \in G_i) = \int_{R_j} f_i(x) dx$ . The average miscalculation cost for any discriminant rule is:

$$\begin{aligned} ECM(R_1, R_2, \dots, R_k) &= E(c(j|i)) \\ &= \sum_{i=1}^k p_i \sum_{j=1}^k c((j|i)) P(j|i) \end{aligned} \quad (5)$$

The average misclassification cost  $ECM$  to minimum is discriminant rules  $x \in G_i$  if

$$\sum_{j=1}^k p_j f_j(x) c(i|j) = \min_{1 \leq h \leq k} \sum_{j=1}^k p_j f_j(x) c(h|j).$$

The average misclassification cost if the sample was adjusted to the  $G_i$  than the average misclassification cost attributed to other general is small, and will be assigned to group  $G_i$  samples.

Definition 2 Fisher criterion function:

Supposing there are  $n$   $k$ -status  $A_1, A_2, \dots, A_n$ , the sample taken from the overall  $A_i$  is denoted as  $x_{i1}, x_{i2}, \dots, x_{in_i}$  ( $i=1, 2, \dots, n$ ), and the sample observation data matrix and the sample mean value are:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n_1} \\ x_{21} & x_{22} & \cdots & x_{2n_2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{k1} & x_{k2} & \cdots & x_{kn_k} \end{bmatrix} \quad (6)$$

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad (i=1, 2, \dots, k) \quad (7)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij} \quad (8)$$

Assuming two classification problems  $w_1/w_2$ , there are  $n$  training samples  $x_k$  ( $k=1, 2, \dots, n$ ), where  $n_1$  samples from type  $w_1$ ,  $n_2$  samples from type  $w_2$ . Two types of training samples were used to construct the subset of the training samples  $X_1$  and  $X_2$ .

Make  $y_k = w^T x_k$ ,  $k=1, 2, \dots, n$ ,  $y_k$  as the projection of  $x_k$ . projection within class alienation sum of squared residuals and class from the sum of squared residuals are respectively:

$$S_G = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2 = w^T \left[ \sum_{i=1}^k (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \right] w \quad (9)$$

$$S_E = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (w^T x_{ij} - w^T \bar{y}_i)^2 \quad (10)$$

Obviously, the sum of squared residuals in a class reflects test results caused by many kinds of various random factors in the process of the test error, such an alienation sum of squared residuals reflects the degree of difference between various kinds of samples, and the system error caused by different levels of variation factors. If it can make the space behind the projection, classes within the sample concentration and sample separation between classes, it can achieve the purpose. If the effect of the projection is good, then  $F = \frac{S_G}{S_E}$  should be large (Sun et al, 2011; Tseng et al, 2012).

The main idea of this algorithm is to use the transformation matrix, to transform the original training sample, project to a new sample space, classify in the projection of the new sample space for learning classification. Among the original sample properties, there may be some dependence between any two attributes, but in the projection in the new sample space, the properties of the new sample are assumed to be independent of each other. Through the transformation it can indicate in the measurement of a high dimension space model into indicating the characteristics of a low dimension space model In this way it can effectively realize classification recognition, and can more accurately reflect the characteristics of the nature of classification (Thalayasingam, 2012; Yang and Wu, 2012).

### 3. Case analysis

A value-added service operator is on the basis of the basic telecom voice business, according to different user groups and the opening of the market demand for the user to choose to use the business's service. Value-added services are the result of market segmentation, and value-added service operators provide customers a higher level of information demand (Tseng P C, Woung L C, Tseng G L, 2012; Yoon I P B, 2014). Therefore, it must provide better, more thoughtful and more diverse services, in order to meet different customers personalized requirements. As the communication network of broadband, intelligent, integrated, personal direction, the boundaries of value-added service and the basic business are becoming increasingly blurred, development of more effective, more value creating business will always be the result of the telecom industry looking for new economic growth points, this is the focus of the competition between them (Yu et al, 2014).

This paper is based on using the Bayesian and Fisher algorithms for analysis of customer behavior data, identifying the customer's behavior characteristics, through the customer providing information about his own hobbies and value-added services. This increases customer loyalty, reduces marketing costs, and enhances the competitiveness of the enterprises.

The original data is from some operators in Harbin in Heilongjiang province, and from sample data for the data flow gas package users in March.

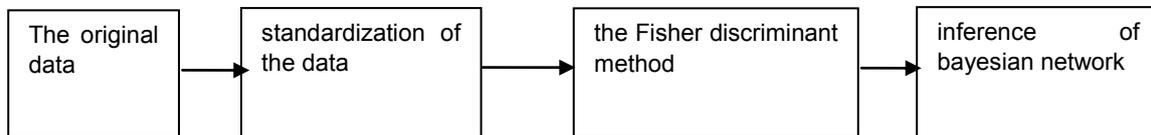


Figure 1: Algorithm flow chart

#### 3.1 The standardization of the data processing

Because the input samples belong to different dimensions, all the input samples, such as TALK\_FEE, CITY\_PHONE\_FEE, ARPU\_FEE are normalized and transformed into 0~1. Using the method of proportional compression, the specific formula is:

$$T = T_{\min} + \frac{T_{\max} - T_{\min}}{X_{\max} - X_{\min}} (X - X_{\min}) \quad (11)$$

Here, X is the original data, and the maximum and minimum for each dimension of the original data. A is transformed data, also known as the target data. For each dimension of the target data's maximum and minimum, take =0.9, =0.1.

Next, through the Fisher discriminant method is used to do the pre-treatment between attributes. The data is read (data=xlsread), data1 and data2 respectively for class 1 and class 2 test sample data. The sample number of class 1 and class 2 is calculated (r1=size(data1,1); r2=size(data2,1);). The mean of class 1 and class 2 (matrix) is calculated (m1=mean(w1); m2=mean(w2)). Various kinds of classes in discrete degree matrix (covariance matrix) are calculated (s1=cov(data1)\*(r1-1); s2=cov(data2)\*(r2-1);). The total class scatter matrix is calculated (sw=s1+s2;). The formula of the projection vector is (w=inv(sw)\*(m1-m2)'). Various kinds of mean of post-projection of a space is calculated (y1=w\*m1'; y2=w\*m2';) The calculated threshold is (w0=1/2\*(y1+y2)). The data to be measured and the class of the same symbol are classified as similar.

Next, network building can be used in the inference of Bayesian network, which is divided into the following three steps.

1. Determine the network nodes and the distribution parameters. By means of expert knowledge this will determine the predictive factors of the problems involved in these factors as network nodes, and further determine the possible values of each of the node variables.
2. Determine the network structure of G; that is, the identification of the causal relationship between the prediction of various factors, and graphical representation.
3. Determine the variable probability distribution  $\theta$ , and on the premise of known network structure, determine the conditional probability of each network node.

Because the paper uses the Bayesian network toolbox (BNT) of Matlab software to build the prediction model, through discretization processing, 500 records from the sample data were randomly selected to form a training set. The other records are included in the test set. Through machine learning applied to the training set data, the network structure can be determined. This article selects K2 algorithm for network structure learning, which is one of the earliest batches of Bayesian network structure learning algorithms. After the

network structure is determined, the conditional probability is calculated by using the maximum likelihood estimation (MLE) algorithm.

After the operation of the network, the data reduction process can eventually be applied.

### 3.2 The results analysis

In this paper, we took 10,000 samples to test, the samples includes 24 selected dimensions of TALK\_FEE, CITY\_PHONE\_FEE, ARPU\_FEE etc., 1 class attributes.

Classes were divided into two classes, recorded as 1 (set customers) and 2(customers said to have unsubscribed due to dissatisfaction). The first 24 dimension properties by Fisher criterion function classes alienation sum of squared residuals and were calculated from the sum of squared residuals, thus obtaining a new sample set. Then, the Bayesian classification algorithm was used to classify the new sample set. In order to analyze the forecast accuracy, a five-fold cross-validation was used to test classification results. Initial data sets were randomly divided into five mutually disjointed subsets  $D_1, D_2, D_3, D_4, D_5$ , and the sizes of each subset are basically the same. Each was studied and tested five times. In the  $i$  iteration,  $S_i$  was used as a test set, and the rest of the subsets of the classifier were used for training. Five iterations were taken with correct classification numbers and divided by the total number of samples in the initial data of average accuracy. The final assessment results showed 89% accuracy.

This shows that the method of this experiment is quite successful in the practical application of data sets, which mainly has the following three aspects:

1. Data pre-processing stage obtains a high quality of the sample data.
2. The dimension independence is higher after Fisher discriminant analysis.
3. Accumulation of a priori knowledge is more comprehensive and accurate, so the selection of network nodes is more reasonable.

The BP artificial neural network has a strong capability to handle and deal with nonlinear problems. The upper and lower layers of each neuron are completely connected, and the algorithm uses forward transfer of information and error return propagation of the two parts. After a pair of samples provides network learning mode, the input information from the input goes layer by layer to the output layer, and neurons in the output layer of the network input respond. Then, in the tending toward reducing the expected output and the actual output error, the connection weights are corrected by each middle layer, layer by layer, from the output layer and finally to the input layer. With the error return propagation modified continuously, the network of correct input mode response rate is rising. The shortcoming of BP model is slow convergence speed of learning algorithm, having been limited to local minimum rather than the state of the global convergence.

With the same data being calculated by Matlab programming, the prediction effect of the hybrid algorithm is better than that of both the pure Bias classification algorithm and Neural network algorithm.

Table 1: Comparison of classification accuracy of different algorithms

| Algorithm                            | Number of samples        | Classification accuracy |
|--------------------------------------|--------------------------|-------------------------|
| Bayesian and Fisher hybrid algorithm | 2000 training, 2000 test | 89                      |
| Pure Bayesian algorithm              | 2000 training, 2000 test | 73                      |
| Neural network algorithm             | 2000 training, 2000 test | 70                      |

## 4. Conclusion

It can be seen from the experimental results obtained that the combination of Fisher based on Bayesian classification algorithms performs better than only the Bayesian algorithm alone, with an increased accuracy of approximately 15%. The traditional Bayes algorithm considers the connection between the properties of each dimension, and by adding the Fisher classification algorithm and projection of each attribute correlation between is greatly reduced, thus fully improving the classification accuracy.

The classical Bayesian classifier is a simple and effective classification algorithm, but its independence assumption makes it unable to express any attribute dependency relationship between the actual data. Without the use of class information, it is only an approximate expression of the parameters of the distribution of the training sample set for each class. In this paper, an improved classifier is proposed, from another point of view, to solve the problem of the classical Bias classifier not being able to extract class information. This is accomplished through the use of Fisher's discriminant analysis method for the separation of class and class largest projection space. Then the original samples are projected to the maximum separable space, and the new samples are obtained. Using the discriminant for the new attributes, then the classical Bias classification algorithm is used to classify the samples. Experiments show that integrating the classical Bias classifier and Fisher linear discriminant analysis method, can achieve a better classification effect.

This paper, on the basis of analyzing the characteristics of the Bayesian model, combined with the Fisher linear discriminant analysis, gives a combined Bayesian and Fisher algorithm of a multi-dimensional customer behavior analysis model. Results show that the model can effectively improve classification accuracy and reduce labor cost. At the same time to it can provide customers with good quality products, which is the cornerstone of the market, and is the key to win over the trust of users.

### Acknowledgments

Heilongjiang Province Natural Science Foundation of China (F201436).

### References

- Abdefattah I.M., Khedr W.I., Sallam K.M., 2013, A TOPSIS based Method for Gene Selection for Cancer Classification J. *International Journal of Computer Applications*, 67(17): 39-44. DOI:10.5120/11490-7195.
- Aquaro V., Bardoscia M., Bellotti R., 2009, A Bayesian Networks approach to Operational Risk[J]. *Physical A Statistical Mechanics & Its Applications*, 389(906.3968): 1721–1728. DOI:10.1016/j.physa.2009.12.043.
- Bouchaala L., Masmoudi A., Gargouri F., 2010, Improving algorithms for structure learning in Bayesian Networks using a new implicit score[J]. *Expert Systems with Applications*, 37(7):5470-5475. DOI:10.1016/j.eswa.2010.02.065.
- Cook D.F., Ragsdale C.T., Major R.L., 2000, Combining a neural network with a genetic algorithm for process parameter optimization[J]. *Engineering Applications of Artificial Intelligence*, 13(4):391-396. DOI:10.1016/S0952-1976(00)00021-X
- Duan J., Wang W., Zeng J., 2009, A prediction algorithm for time series based on adaptive model selection[J]. *Expert Systems with Applications*, 36(2):1308-1314. DOI:10.1016/j.eswa.2007.11.021.
- Guo X., Li D.C., Zhang A., 2012, Improved Support Vector Machine Oil Price Forecast Model Based on Genetic Algorithm Optimization Parameters [J]. *Aasri Procedia*, 1(4):525-530. DOI:10.1016/j.aasri.2012.06.082.
- Hao H.N., 2010, Notice of Retraction Short-term forecasting of stock price based on genetic-neural network[C]// *Natural Computation (ICNC)*, 2010 Sixth International Conference on. IEEE, 1838-1841. DOI:10.1109/ICNC.2010.5584528.
- Liu C., 2014, Network Intrusion Detection Model Based on Genetic Algorithm Optimizing Parameters of Support Vector Machine [J]. *Advanced Materials Research*, 989-994:2012-2015. DOI:10.4028/www.scientific.net/AMR.989-994.2012.
- Li L., Ma S., Zhang Y., 2014, Optimization Algorithm Based on Genetic Support Vector Machine Model[C]// *Seventh International Symposium on Computational Intelligence and Design*. IEEE, 307 - 310. DOI:10.1109/ISCID.2014.99.
- Qu H.N., Li G.Z., Xu W.S., 2010, An asymmetric classifier based on partial least squares[J]. *Pattern Recognition*, 43(10):3448-3457. DOI:10.1016/j.patcog.2010.05.002.
- Sun Y., Tang Y., Ding S.I., 2011, Diagnose the mild cognitive impairment by constructing Bayesian network with missing data[J]. *Expert Systems with Applications*, 38(1):442-449. DOI:10.1016/j.eswa.2010.06.084.
- Thalayasingam M., Veerakumarasivam A., Kulanthayan S., 2012, Clinical clues for head injuries amongst Malaysian infants: Accidental or non-accidental?[J]. *Injury-international Journal of the Care of the Injured*, 43(12):2083-2087. DOI:10.1016/j.injury.2012.02.010.
- Tseng P.C., Woung L.C., Tseng G.L., 2012, Refractive change after pars plana vitrectomy[J]. *Taiwan Journal of Ophthalmology*, 2(1):18–21. DOI: 10.1016/j.tjo.2011.11.003.
- Yang Y., Wu Y., 2012, On the properties of concept classes induced by multivalued Bayesian networks [J]. *Information Sciences*, 184(1):155-165. DOI:10.1016/j.ins.2011.08.031
- Yoon I.P.B., 2014, A semantic analysis approach for identifying patent infringement based on a product–patent map[J]. *Technology Analysis & Strategic Management*, 26(8):855-874. DOI:10.1080/09537325.2014.909926.
- Yu F., Wang Z.Q., Xu X.Z., 2014, Short-Term Gas Load Forecasting Based on Wavelet BP Neural Network Optimized by Genetic Algorithm[J]. *Applied Mechanics & Materials*, 631-632(631-632):79-85. DOI:10.4028/www.scientific.net/AMM.631-632.79.