

行政院國家科學委員會專題研究計畫 期中進度報告

類神經網路特徵選取及其應用之研究(2/3)

計畫類別：個別型計畫

計畫編號：NSC92-2213-E-009-028-

執行期間：92年08月01日至93年07月31日

執行單位：國立交通大學工業工程與管理學系

計畫主持人：蘇朝墩

計畫參與人員：李得盛、許志華、薛友仁

報告類型：精簡報告

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 93 年 5 月 25 日

行政院國家科學委員會補助專題研究計畫成果報告

類神經網路特徵選取及其應用之研究 (2/3)

計畫類別：個別型計畫

計畫編號：NSC 92-2213-E-009-028

執行期間：92 年 8 月 1 日至 93 年 7 月 31 日

計畫主持人：蘇朝墩

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

執行單位：國立交通大學工業工程與管理學系

中華民國九十三年五月二十五日

行政院國家科學委員會專題研究計畫成果報告

類神經網路特徵選取及其應用之研究 (2/3)

Feature Selection for Neural Classifiers with Application to Operation Management

計畫編號：NSC 92-2213-E-009-028

執行期限：92 年 8 月 1 日至 93 年 7 月 31 日

主持人：蘇朝墩 國立交通大學工業工程與管理學系

計畫參與人員：李得盛、許志華、薛友仁

國立交通大學工業工程與管理學系

一、中文摘要

特徵選取是分類問題的主要工作之一，特徵選取的主要目的乃是要選取重要的特徵並獲得一可接受的分類精確度。類神經網路是處理分類問題的一個很受歡迎的方法，類神經網路的結構愈是簡化，其愈可改善網路之解釋和預測的能力。也就是說，降低特徵個數可以減少計算上的複雜度，並有可能可以提昇分類的精確度。

本計劃第二年首先提出二個方法：(i) 倒傳遞類神經網路的特徵選取方法，(ii) 馬氏/田口系統（本計劃發展一決定 threshold 之方法供其使用）。本計劃利用二個例子，對這二個方法進行比較，最後並給予扼要討論。

關鍵詞：特徵選取；類神經網路；倒傳遞；馬氏距離；自動門檻；馬氏/田口系統

Abstract

This project first discusses two classification approaches using back-propagation (BP) neural network and Mahalanobis distance (MD) classifier, and then proposes two classification approaches for multi-dimensional feature selection. The first one proposed is a feature selection procedure from the trained back-propagation (BP) neural network. The basic idea of this procedure is to compare the multiplication weights between input and hidden layer and hidden and output layer. In order to simplify

the structure, only the multiplication weights of large absolute values are used. The second approach is Mahalanobis-Taguchi system (MTS) originally suggested by Dr. Taguchi. The MTS performs Taguchi's fractional factorial design based on the Mahalanobis distance as a performance metric. We combine the automatic thresholding with MD; it can deal with a reduced model, which is the focus of this study. In this project, two case studies will be used as examples to compare and discuss the complete and reduced models employing BP neural network and MD classifier. The implementation results show that proposed approaches are effective and powerful for the classification.

Keywords: Feature selection, Artificial neural networks; Backpropagation; Mahalanobis distance; Automatic thresholding; Mahalanobis-Taguchi system

二、緣由與目的

In most cases, many data (such as medical examination data) are characterized by multi-dimensional information with ambiguity and variation, which make it difficult to explore the relationships among them. The traditional approach to building an expert system requires the formulation of rules by which the input data can be analyzed. The formulation of such rules is quite difficult with large sets of input data. To resolve the difficulty, artificial neural network (ANN) has been applied as an

alternative to traditional rule-based expert system. ANNs can be well trained by input data (e.g. examination results) and output response (e.g. signs or symptoms). Moreover, ANN has been applied to various pattern classifications in many fields. Giacinto et al. (2000) combined the neural and statistical algorithms for supervised classification of remote-sensing images. Sutter and Jrus (1997) used ANN to classify and quantify organic vapors. A neural network was trained by Pfurtscheller et al. (1996) to classify electroencephalogram (EEG) patterns in a real-time fashion. Kwak and Lee (1997) illustrated the capacity of ANN to classify and predict the health status of HIV/AIDS patients. In short, ANN has demonstrated its capability of pattern classification including diagnosis of diseases. Hence, ANN has been found to be more helpful than a traditional approach in dealing with the multi-dimensional data.

Even though it can basically approximate any function, an ANN still has a few problems such as time-consuming convergence, overfitted training, high complexity in computation and trained NNs are black boxes from the designer's point of view (Tsukimoto, 2000). The advanced computer hardware has contributed to the substantial improvement in the speed and ease of computation. However, the other problems are closely related to the neural network structures and training algorithms. Several algorithms (Tsukimoto, 2000) have been developed by researchers trying to understand the neural network structure. Through knowing the structures, deleting the redundant connections and extracting the rules, neural network users can learn in advance what the neural networks have discovered and how the neural networks predict. Therefore, users can apply the neural networks to some critical problems. In this project, we analyze and evaluate the complete and reduced neural network models applicable to the multi-dimensional data. The reduced neural network will be obtained from an feature selection procedure. The basic idea of this procedure is to compare the multiplication weights between input and

hidden layer and hidden and output layer. After eliminating the unimportant input nodes, the neural network still possesses the robust potential for classification.

On the other hand, the Mahalanobis distance (MD) is one of the minimum distance classifiers. In contrast to the Euclidean distance classifier, MD also considers the correlation among the multi-dimensional variables. MD is a very sensitive and useful way to determine the similarities among a group of data and detect any unknown data or outlier from a large data set. As MD has been known for some time, in fact, MD was successfully applied to spectral discrimination in analytical chemistry and pattern recognition in computer vision. Brown et al. (1998) used Mahalanobis distance metric based on multi-dimensional vector to evaluate the performance of three 100-compound spectra classifications. Shah and Gemperline (1990) qualitatively identified raw materials by near infrared (NIR) spectroscopy using a Mahalanobis distance classification method. Kato et al. (1999) proposed the asymmetric Mahalanobis distance as a fine classification technique for pattern recognition of handwritten Chinese and Japanese characters.

The Mahalanobis-Taguchi system (MTS) suggested by Dr. Taguchi, combining the Mahalanobis distance and Taguchi method, was used in the area of quality engineering (Taguchi, 1998). MTS can deal with a reduced model which determines the significant factors in the experiments by comparing the signal to noise (S/N) ratio between different levels. It is shown that the MTS is a robust approach by giving the noise to the training multi-dimensional data.

This project first discusses two classification approaches using back-propagation (BP) neural network and Mahalanobis distance (MD) classifier, and then proposes two classification approaches for multi-dimensional feature selection. The first approach proposed is an feature selection procedure from the trained back-propagation (BP) neural network. The second approach is Mahalanobis-Taguchi system (MTS) which combines the automatic

thresholding approach with MD as a performance metric. We will illustrate the effectiveness of the proposed approaches in complete and reduced models by using the real-world medical exam data and industrial product data.

三、結果與討論

Back-propagation neural networks

Procedure 1: Induction of a BP classifier

Phase I: Training process

Step 1: Collect a set of observed data.

Step 2: Divide the data into training and testing data sets.

Step 3: Set the training parameters (e.g., learning rate and momentum).

Step 4: Train the different neural network structures.

Step 5: Select a trained network with the highest classification accuracy.

Phase II: Classification process

Step 1: Obtain the unknown input data.

Step 2: Present the data to the trained network that is selected from step 5 in phase I.

Step 3: Obtain the classification results.

Feature selection from the trained BP neural network

Procedure 2: Feature selection for a neural network

Step 1: Calculate the sum of the absolute multiplication values of weights between input and hidden layers and hidden and output layers for each input node.

Step 2: Sort the values obtained from Step 1 in a descending sequence and select a cutoff value.

Step 3: Find the corresponding input features which are larger than cutoff value selected from Step 2

Step 4: Train the neural network by the selected input features and compare the classification results with that of all the original input features. If the classification result of selected input feature is satisfactory, then stop; otherwise back to Step 2 to select a new cutoff value.

Mahalanobis distance classifier

Procedure 3: Induction of a MD classifier

Phase I: Training process

Step 1: Collect a set of data obtained from multiple items (including normal and abnormal conditions).

Step 2: Normalize the individual data under normal conditions.

Step 3: Calculate the variance-covariance matrix of the normalized data.

Step 4: Calculate the MD space.

Step 5: Plot the distribution of MD space.

Step 6: Determine the threshold of the MD space, t^* .

Phase II: Classification process

Step 1: Obtain the unknown input data.

Step 2: Normalize the data based on the means and variance under normal conditions.

Step 3: Calculate the Mahalanobis Distance D^2 .

Step 4: Obtain the classification results, i.e. if $D^2 > t^*$, then this pattern belongs to an abnormal set. Otherwise, the pattern belongs to a normal set with similar properties.

Mahalanobis-Taguchi System

Procedure 4: Induction of a MTS classifier

Step 1: Collect n normal data, which are characterized by K -dimensional items.

Step 2: Calculate the D^2 for each data.

Step 3: Let M_i be the signal in Taguchi's dynamic system, i.e. $M_i = \sqrt{D_i^2}$, $i = 1 \wedge n$.

Step 4: Divide K items into L items and $(K-L)$ items; L items need to be further studied in Orthogonal Array and $(K-L)$ items represent the absolutely necessary items due to theoretic consideration or learned from previous experience.

Step 5: Select an appropriate OA and assign the L items into the column of OA. In the OA table, use two levels for each factor; 1 means not using this factor and 2 means using this factor in the experiment.

Step 6: Calculate the MD space for each row of OA. In case of all 1's in the row, it means that all the factors are not used in the experiments and we will calculate the MD space characterized by the other $(K-L)$ items. In contrary, if both 1's and 2's exist in the

same row, we will use the factors corresponding to 2's column plus ($K-L$) items to create MD space.

Step 7. Based on the MD space and the responses, calculate the S/N ratio for each row in the OA.

Step 8. Plot the factor effects and determine the important items in the experiment.

Step 9: Use Procedure 3 to obtain the MD space, determine the threshold and reach the final classification results.

四、計劃成果自評

By adopting BP (complete /reduced) and MD (complete /reduced) approaches, this project classifies the multi-dimensional examination data for diagnosis of a liver disease and glass classification. In the first example, the results show that the reduced BP network (15 items) is better than the complete BP network. The best way to elucidate the above results is the feature selection procedure that can actually classify the items into the important and unimportant classes. In contrast to the results of BP network, the complete MD classifier provides slightly more information than the reduced MD model (16 items) because MTS has lost some information during the procedure. In the second example, the results show that the reduced BP network (5 items) is better than all the other classifier. Correspondingly, MTS classifier also outperformed than the MD even MTS has reduced some features during the procedure. The analytical results indicate that these four classifiers are all robust and effective methods to classify the medical data and industrial product in this project. However, how many variables can be reduced in a MD model without serious impact on the classification accuracy is a subject for future research.

The above research results have been accepted for publication in *The Asian Journal on Quality*.

五、參考文獻

1. Andrews, R. and Diederich, J., 1996. Rules and Networks. Proceedings of Rule Extraction Trained Artificial Neural Networks Workshop, AISB.
2. Andrews, R., Diederich, J. and Tickle, A. B., 1995. Survey and critique of techniques for extracting rules from trained artificial neural networks. Knowledge-Based System 8, 373-389.
3. Antony, J., 2000. Multi-response optimization in industrial experiments using Taguchi's quality loss function and principal component analysis. Quality Reliability Engineering. International 16, 3-8.
4. Brown, C. W. and Lo, S. C., 1998. Chemical information based on neural network processing of Near-IR Spectra. Analytic Chemistry 70, 2983-2990.
5. Giacinto, G., Roli, F. and Bruzzone, L., 2000. Combination of neural and statistical algorithms for supervised classification of remote-sensing images. Pattern recognition letters 21, 385-397.
6. Guo, R. and Pandit, S. M., 1998. Automatic threshold selection based on histogram lodes and a discriminant criterion. Machine vision and applications 10, 331-338.
7. Kato, N., Suzuki, M., Omachi, S., Aso, H. and Nemoto, Y., 1999. A handwritten character recognition system using directional element feature and asymmetric Mahalanobis distance. IEEE Transaction on Pattern Analysis and Machine Intelligence. 21, 258-263.
8. Kwak, N. K. and Lee, C., 1997. A neural network application to classification of health status of HIV/AIDS patients. Journal of Medical System 21. 87-97.
9. Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Transaction on Systems Man and Cybernetics. SMC 9, 62-66.
10. Pfurtscheller, G., Kalcher, J. and Nerper, C., 1996. On-line EEG classification during externally paced hand movements using a neural network-based classifier. Electroncephalogr. Clinical Neurophysiology 99, 416-425.
11. Sarle, W. S., 2000. How to measure importance of inputs?. SAS Institute Inc., Cary, NC, USA.

<ftp://ftp.sas.com/put/neural/importance.html>.

12. Sette, S., Boullart, L., Langenhove, L. V. and Kiekens, P., 1997. Optimizing the fiber-to-yarn production process with a combined neural network/genetic algorithm approach. *Textile Research Journal* 67, 84-92. Shah, N. K. and Gemperline, P. J., 1990. Combination of the Mahalanobis distance and residual variance pattern recognition techniques for classification of Near-Infrared reflectance spectra. *Analytic Chemistry* 62, 465-470.
13. Sutter, J. M. and Jurs, P. C., 1997. Neural network classification and quantification of organic vapors based on fluorescence data from a fiber-optic sensor array. *Analytic Chemistry* 69, 856-862.
14. Taguchi, Genichi, 1998. Mathematics for Quality Engineering. *Journal of Quality Engineering Forum* 6, 5-10.
15. Tsukimoto, H., 2000. Extraction Rules from Trained Neural Networks. *IEEE Transactions on Neural Networks* 11, 377-389.
16. Younis, K. S., Rogers, T. K. and DeSimio, M. P., 1996. Vector quantization based on dynamic adjustment of Mahalanobis distance. *Aerospace and Electronics Conference Proceedings of IEEE of 1996* 1.2, 858-862.

