Differentiation of Syndromes with SVM

Zhanquan Sun, Guangcheng Xi, and Jianqiang Yi

Key Laboratory of Complex Systems and Intelligence Science,
Institute of Automation, Academy of Sciences, 100080, Beijing, China
{zhanquan.sun, guangcheng.xi, jianqiang.yi}@mail.ia.ac.cn

Abstract. Differentiation of syndromes is the kernel theory of Traditional Chinese Medicine (TCM). How to diagnose syndromes correctly with scientific means according to symptoms is the first problem in TCM. Several modern approaches have been applied, but no satisfied results have been obtained because of the complexity of diagnosis procedure. Support Vector Machine (SVM) is a new classification technique and has drawn much attention on this topic in recent years. In this paper, we combine non-linear Principle Component Analysis (PCA) neural network with multi-class SVM to realize differentiation of syndromes. Non-linear PCA is used to preprocess clinical data to save computational cost and reduce noise. The multi-class SVM takes the non-linear principle components as its inputs and determines a corresponding syndrome. Analyzing of a TCM example shows its effectiveness.

1 Introduction

Traditional Chinese Medicine (TCM) is the treasure of China with about 3000 years' history. It is the only existing systematic classic medicine all over the world now. But most people prefer western medicine and take TCM as experience science. It is unassailable that TCM has better curative effect on many stubborn and chronic diseases than western medicine. So it attracts more and more researchers to study TCM with modern techniques. Differentiation of syndromes is a method of understanding and diagnosing diseases by the theories of TCM. A syndrome is a TCM diagnosis conclusion, which comes from an assessment of all information gathered from four diagnostic techniques: inquiry, inspection, smelling and palpation. In this diagnosis process, factors of age, gender, constitution, season, weather, geographical area, as well as emotion, stress, diet, lifestyle and all aspects of symptoms and signs (tongue picture and pulse) of the body are taken into consideration to form a syndrome for a patient. The treatment strategy is given according to the syndrome. So correct diagnosis of a syndrome is very important in TCM. In the past, differentiation of syndromes is diagnosed according to the doctor's experience. Different doctor maybe diagnose different syndromes to the same patient because of complicated relation between symptoms and a syndrome. How to find the just syndrome with scientific means is a crucial problem in TCM field. Many techniques have been developed, such as cluster, fuzzy cluster, BP network, RBF network and so on[1-2]. But the classification results of all these methods can't meet with practice requirement.

Support Vector Machine (SVM) developed by Vapnik[3] is a new classification technique and has gained popularity due to many attractive features and promising

J. Wang et al. (Eds.): ISNN 2006, LNCS 3973, pp. 786 – 791, 2006.

empirical performance. SVM is a powerful supervised learning algorithm belonging to the statistical learning theory. It minimizes the structural risk performance in various classification problems. In many applications, SVM has been shown to provide higher performance than traditional learning machines and has been introduced as powerful tools for solving classification problems [4]. Furthermore, SVM is suitable to analyze small sample problems. In SVM, it will cost more computational time if there are too much feature variables. In view of practical requirement, the source data should be preprocessed. Principle Component Analysis (PCA) is a multivariate statistical method. PCA extracts the modes in a set of variables. Its main purpose is to reduce the dimensionality of the dataset by retaining only the first few modes. But PCA is a linear method and implies a potential oversimplification of the datasets being analyzed. The advent of non-linear PCA can overcome the shortcomings. Nonlinear PCA has been applied in many kinds of field, which is usually approximated by a neural network [5-6].

2 Nonlinear PCA Neural Network

Non-linear PCA can be realized with neural network. Let $\mathbf{x}$ denote an L-dimensional random vector. A non-linear neuron model is configured with weight vector $\mathbf{w}$, input sample $\mathbf{x}$ and non-linear output neuron $z = f(\mathbf{w}^T\mathbf{x})$. A simple reconstruction of input vector $\mathbf{x}$ is represented as $\hat{\mathbf{x}} = z\mathbf{w}$. The largest non-linear principal component can be achieved by optimizing the reconstruction mean-square error, i.e., the following objective:

$$\min \quad J = E\{\|\mathbf{x}-\hat{\mathbf{x}}\|^2\} = E\{\|\mathbf{x}-z\mathbf{w}\|^2\} \tag{1}$$

The optimum solution $z = f(\mathbf{w}^T\mathbf{x})$ is the largest nonlinear principle component, f is a non-linear mapping function. Here we choose f as a sigmoidal function bounded between 0 and 1.

A stochastic approximation approach will lead to the following learning rule

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \mu_t(e_t^T\mathbf{w}_t\mathbf{x}g + z\mathbf{e}_t) \tag{2}$$

where $e_t = \mathbf{x} - z\mathbf{w}_t$ is the reconstruction error, g is the derivative of z, and $g = f'(\mathbf{w}_t^T\mathbf{x})$. The learning rule Eq. (2) can be further extended to a network with M neurons. In this case, the network can be treated as autoencoder, which has equal number of L input and output nodes and M hidden nodes, $M < L$. During learning, the inputs are duplicated at the output nodes to perform identity mapping. $L \times M$ matrix w, $M \times L$ matrix $\hat{\mathrm{w}}$ denote the connection weight from input layer to hidden layer and from hidden layer to output layer respectively. In our work, we only consider the symmetrical case, i.e., $\hat{\mathrm{w}} = \mathrm{w}^T$. The structure of non-linear PCA neural network is shown as in Fig. 1. In this structure, the nonlinear principle components are a vector $\boldsymbol{z} = (z_1, z_2, \cdots, z_M)$ with M elements. The larger is the number M, the less will be the reconstruction error, and meanwhile the more computational time will be cost. So it should be determined according to practical requirement.

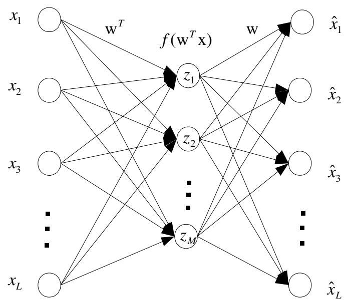

Fig. 1. Structure of nonlinear PCA neural network

3 Support Vector Machine

SVM first maps the input points into a high-dimensional feature space and finds a separating hyperplane that maximizes the margin between two classes in this space. Maximizing the margin is a quadratic programming problem and can be solved from its dual problem by introducing Lagrangian multipliers. Without any knowledge of the mapping, the SVM finds the optimal hyperplane by using the dot product functions in feature space that are called kernels. Many modern optimization methods have been developed and can be used now [7]. The solution of the optimal hyperplane can be written as a combination of a few input points that are called support vectors.

Given a training dataset (x_i, y_i), for $i = 1,2,\cdots,n$ and $x_i \in R^d$ and $y_i \in \{1,-1\}$ and a non-linear transformation to a higher dimensional space (the feature space) $(\Phi(\cdot), R^d \xrightarrow{\Phi(\cdot)} R^H)$, the SVM solves:

$$\begin{aligned} &\min_{w,\xi_i,b} \quad \left\{ \frac{1}{2}\|\mathrm{w}\|^2 + C\sum_i \xi_i \right\} \\ &\text{s.t.} \quad y^i(\Phi^T(\mathrm{x}_i)\mathrm{w} + b) \ge 1 - \xi_i, \qquad \xi_i \ge 0 \quad \forall i = 1,\cdots,n \end{aligned} \tag{3}$$

where $\mathbf{w}$ and b define a linear classifier in the feature space, and $\xi_i, i = 1,2,\cdots,n$ are relaxation variables and $\Phi(\cdot)$ is mapping function. The optimization of the problem ensures that the solution is maximum margin. By introducing Lagrangian multipliers, the optimization problem can be transformed into its dual problem.

$$\begin{aligned} &\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_i K(x_i, x_j) - \sum_{j=1}^{n} \alpha_j \\ &\text{s.t.} \quad \sum_{i=1}^{l} y_i \alpha_i = 0 \qquad 0 \le \alpha_i \le C, i = 1,2,\cdots,n \end{aligned} \tag{4}$$

where $\alpha_i, i = 1,2,\cdots,n$ are Lagrangian coefficients, $K(x_i, x_j) = (\Phi(x_i)\cdot\Phi(x_j))$ is kernel function and C is penalty coefficient prescribed according to practical requirement. The solution of the dual problem is much easier than the original problem.

A classification task usually involves training and testing data that consist of some data instances. Each instance in the training set contains one “target value” (class label) and several “attributes” (features). The goal of SVM is to produce a model which predicts target values of data instances in the test set, for which only the attributes are given. After obtaining optimum solution α^*, b^*, the following decision function is used to determine which class the sample belongs to.

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i^* K(x_i, x) + b^* \right) \tag{5}$$

It is important to choose an appropriate kernel function to SVM. Kernel function must satisfy the Mercer theorem. Many kinds of kernel functions have been developed. In this paper, we choose the most commonly used radial basis function

$$K(x_i, x_j) = \gamma \exp(\left| x_i - x_j \right|)^2 \tag{6}$$

where γ, r are kernel parameters.

SVM has been originally designed for two-class classification problems. But many practical problems are multi-class. Several methods have been proposed for multi-class SVM in recent years. In this paper, we adopt a commonly used one. The approach to solve *k*-class classification problem is to construct *k* binary SVM classifiers. Each of them is for one class. The *i*th SVM classifier will be trained with all the examples in the *i*th class with positive labels and all other examples with negative labels. The *i*th SVM classifier constructs a separating hyperplane between class i and the other classes. The multi-class SVMs trained in this way are called 1-*v*-*r*. The conclusion is given with the class that corresponds to the SVM with the highest output value

$$g^j(x) = \sum_{i=1}^{l} y_i \alpha_i^j K(x_i, x) + b^j, \; j = 1, 2, \cdots, k \tag{7}$$

In this paper, we combine non-linear PCA neural network and multi-class SVM mentioned above together. The source data are preprocessed by the non-linear PCA neural network at first. Both the dimensionality and noise of source data are reduced. The non-linear principle components are used as the inputs of multi-class SVM for classification. Through the combination, it can not only save classification time but also improve the correct rate of classification.

4 TCM Example

We have 1022 patients’ clinic data collected by Chinese Academy of Traditional Chinese Medicine. In these clinic data, 71 symptom variables are recorded. All these symptoms have been quantified. Some symptoms have two values denoted by 0 and 1, and some have four values denoted by 0, 1, 2 and 3. For consistence, some real number variables, such as age, height, are divided into several sects and denoted by 0, 1, 2 and 3. All symptoms are listed in Table 1. There are five different syndromes among these clinic data. They are Syndrome of blood stasis due to deficiency of *qi*, stagnation of *qi* and blood stasis, blood stasis due to deficiency of *yang*, mutual obstruction of phlegm and stasis, and obstruction of collaterals by stagnant blood. The syndrome is denoted by the class variable whose five different values correspond to the five different syndromes respectively.

Table 1. Symptom variables

Serial number	Symptoms	Serial number	Symptoms	Serial number	Symptoms
1	Gender	25	Tinnitus	49	Walk unsteadiness
2	Age	26	Loss of memory	50	Hemianesthesia
3	Nation	27	Insomnia	51	Acroanesthesia
4	Height	28	Drowsiness	52	Oral lip Anaesth
5	Weight	29	Painful spasm of nape	53	Thirst
6	Occupation	30	Facial distortion	54	Thin
7	Life style	31	Aphasia	55	Five feverish centres
8	Appetite hot	32	Angina pectoris	56	Polyorexia
9	Smoking	33	Chest complaint	57	Anorexia
10	Drinking	34	Chest distress	58	Diuresis
11	Psyche status	35	Cardiopalmus	59	Incontinence of urine
12	Irritable tantrum	36	Hypochondrium distending pain	60	Swell lymph
13	Fatigued and weak	37	Abdominal tenderness	61	Joint pain
14	Chilly	38	Vena epigastrica	62	Vein abnormity
15	Developmental condition	39	Rebound tenderness	63	Cacomelia
16	Operation	40	Abdominal mass	64	Swelling of limb
17	External injury	41	Nausea and vomitting	65	Red tongue
18	Palatine mucosa syndrome	42	Hemiplegia	66	Tongue ecchymosis
19	Dark gum	43	Animal force	67	Varicose lingual vessels
20	Dark labia oris	44	Muscular tension	68	Clavus
21	Scaly skin	45	Dark circles under the eyes	69	Pulses unsmooth
22	Headache	46	Dark face	70	Wiry
23	Dizziness	47	Skin ecchymosis	71	Acrotism
24	Vertigo	48	Patho-sign		

In this example, 855 clinic data are chosen to train a SVM, and the remaining are used to test. The 71 symptom variables are used as input of nonlinear PCA neural network. The kernel function is chosen as sigmoid function. The neural unit number of output is 25, which can be determined according to practical requirement. The 25 nonlinear principle components are used to diagnose a syndrome with the SVM. The setting of the SVM is: penalty factor $C = 20$, kernel function is chosen as RBF function and the width is set as $\sigma^2 = 0.1$.

After training with the method developed in this paper, the test result is: 127 clinic data out of 167 test samples can be classified correctly. The correct rate is 76.05%, which is much higher than RBF neural network used in [8]. We find that it is suitable to analyze TCM problems with SVM classifier. It can provide an objective reference for differentiation of syndromes in clinic.

5 Conclusion

In this paper, we combine non-linear PCA neural network with multi-class SVM for differentiation of syndromes. Non-linear PCA can reduce dimension and get rid of some noise of source data. Through the application on TCM example, we find the results obtained in this paper are better than other methods. Particularly, SVM is superior to other methods when sample size is small. It is suitable to solve the differentiation of syndromes with multi-class SVM. In TCM, many other non-linear classification problems can also be solved with SVM, which are our further study.

Acknowledgement

The research was supported by National Basic Research Program of China (973 program) under grant No. (2003CB517106), and MOST Projects (No. 2004DFB02100).

References

1. Huang, H.Z., Wu, W.D., Liu, C.S.: A Coordination Method for Fuzzy Multi-objective Optimization of System Reliability. Journal of Intelligent and Fuzzy Systems 16 (2005) 213-220
2. Huang, H.Z., Li, Y.H., Xue, L.H.: A Comprehensive Evaluation Model for Assessments of Grinding Machining Quality. Key Engineering Materials 291-292 (2005) 157-162
3. Cortes, C., Vapnik, V.: Support Vector Networks. Machine Learning, 20 (1995) 273–297
4. Burges, J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2 (1998) 121–167
5. Xu, L.: Least Mean Square Error Reconstruction Principle for Self-organization. Neural Networks 6 (1993) 627-648
6. Karhunen, J., Joutsensalo J.: Representation and Separation of Signals Using Nonlinear PCA type learning. Neural Networks 7 (1994) 113-127
7. Huang, H. Z., Tian, Z. G., Zuo, M. J.: Intelligent Interactive Multiobjective Optimization Method and Its application to Reliability Optimization. IIE Transactions 37 (2005) 983–993
8. Bian, Q., He, Y. M., Shi, X. C., Wang, H. W.: Neural Network Model of TCM Syndromes Based on MFBP Learning Algorithm. Chinese Journal of Basic Medicine in Traditional Chinese Medicine 7 (2001) 66-69