Research on the Classification of Audio Doppler Signals Based on SVM

Donghua Luo

Shandong Communication & Media College, Jinan, Shandong Province, China

Keywords: ultrasonic Doppler signal; maximum frequency curve; classification of signals; SVM; modulus maximum

Abstract: In order to analyze Doppler signals effectively, a SVM based classification method is proposed in the paper to classify audio signals output by supersonic Doppler detectors. At first, the percentage method is employed to get the maximum frequency curve of audio signals. Then the characteristics of signals are acquired through the modulus maximum of wavelet transform. These features can reflect the characteristics of the whole signal effectively. At last, according to these features, classification of Doppler signals is realized through SVM. The classification results can display the situation of blood vessels of patients. The simulation results show that, compared with neural network classification method, method introduced in this paper have advantages of fast classification speed and high identification rate. The classification of Doppler signals has important clinical values, since it can provide doctors with recommendations and help them treat patients more effectively.

1. Introduction

SVM, proposed by Vapnik in 1992, is a relatively new method of machine learning. It is built on the basis of statistical learning theory and Structural Risk Minimization (SRM). Based on limited sample information, it can find the best compromise between complexity of model and learning ability. With complete theoretical foundation on mathematics and unique merits, it is the best categorization method so far. In order to analyze Doppler signals effectively, we need to classify Doppler signals. In this paper, the percentage method is employed to get the maximum frequency curve. On the basis of maximum frequency curve, the characteristics of audio signals are acquired through modulus maximum of wavelet transform. Information about these features is normalized afterwards. Training sets and testing sets are established based on clinical data about the transcranial Doppler detector. A classifier based on SVM sorting algorithms is built. Finally, the classification and recognition results are analyzed

2. Principles and Methods

2.1 Support vector machine theory.

SVM is extracted from the Optimal Hyperplane of linearly separable pattern class. Its basic guiding idea goes as following. First, the input vector is mapped to a high dimensional feature space through some pre-selected nonlinear mapping. Then, an Optimal Hyperplane is constructed in this space, thus achieving the maximum separation limit between positive samples and counter-examples. Conceptually, support vectors are nearest data points from the decision plane; they determine the location of Optimal Hyperplane.

As a classification tool, SVM has good generalization ability under small sample conditions. It has many advantages in solving small sample problems and recognizing nonlinear and high dimensional patterns; it is suitable for diversified data types. Through the replacement of kernel function, it can get various kinds of surfaces. [3] It has become an alternative method in training multilayer sensors, RBF neural networks and polynomial neural networks. It has attracted widely attention, and been successfully applied in many fields, such as face recognition, text classification, gene analysis, handwriting recognition, speech recognition, function approximation and time sequence prediction. [4,5,6,7,8]

Copyright © (2018) Francis Academic Press, UK

Suppose a linear separable sample set $T=\{(x_1,y_1),...,(x_l,y_l)\}$, in which $x_i \in R^n$, $y_i \in \{1,-1\}$, $i=1,2,\cdots,l$ (l is the number of samples, $y_i \in \{1,-1\}$ is the class label); The general form of linear discrimination function in n dimension space is:

$$g(x) = w \bullet x + b \tag{1}$$

So, the equation of hyperplane is:

$$w \bullet x + b = 0 \tag{2}$$

The normalization of the discrimination function makes all two classes of samples satisfy $|g(x)| \geq 1$. The sample nearest to hyperplane is $|g(x)|=1$ with the classification interval $margin = 2/\|w\|$. So the Optimal Hyperplane needs to satisfy following conditions:

Correct classification of all samples, i. e. $y_i[(w \bullet x_i)+b]-1 \geq 0$, $i=1,2,\cdots,l$; has the largest class interval $margin = 2/\|w\|$, even if $\|w\|$ or $\|w\|^2$ is minimum.

Training samples satisfy the condition of $y_i[(w \bullet x_i)+b]-1=0$ are called Support Vectors.

The problem of finding Optimal Hyperplane is transformed into the problem of solving the minimum value of Lagrange function.

$$L(w,\alpha,b) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l}\alpha_i y_i (w^T \cdot x_i + b) + \sum_{i=1}^{l}\alpha_i,$$

$$\alpha_i \geq 0, i=1,2,\cdots,l \tag{3}$$

In which $\alpha_i > 0$ is the coefficient of Lagrange. Based on the Kuhn-Tucker condition, the solution of problem should satisfy following conditions:

$$\frac{\partial}{\partial \mathrm{w}} \mathrm{L}(\mathbf{w},\boldsymbol{\alpha},b) = \mathbf{w} - \sum_{i}^{l}\alpha_i y_i \mathbf{x}_i = 0 \tag{4}$$

$$\frac{\partial}{\partial b} \mathrm{L}(\mathbf{w},\boldsymbol{\alpha},b) = \sum_{i}^{l}\alpha_i y_i = 0 \tag{5}$$

The Optimal Hyperplane function can be obtained by substituting (4) (5) into (3).

$$f(x) = sign\left[\sum_{i\in\Omega sv} y_i \alpha_i^* (x_i, x) + b^*\right] \tag{6}$$

In which, Ωsv represents support vectors which construct a vector space, $sign()$ is a symbol function, α_i^* is the Lagrange multiplier corresponding to support vectors, $b^* = \frac{1}{2}[(w\cdot x^*(1)) + (w \cdot x^*(-1))]$, $x^*(1)$ is any support vector in the first class, $x^*(-1)$ is any support vector in the second class.

For the nonlinear separable sample set, the nonlinear space can be converted into a linear space though transformation. Then the Optimal Hyperplane can be obtained under the linear separable condition. The product of points x_i and x_j from the introduced kernel function $K(x_i,x_j)$ equals to the product of mapping points in the high dimensional feature space; the original feature space is then transformed into a new feature space. $K(x_i,x_j)$ replaces the product of point x_i and x_j. The corresponding discriminant function is:

$$f(x) = sign\left[\sum_{i\in\Omega sv} y_i \alpha_i^* K(x_i, x) + b^*\right] \tag{7}$$

The amount of calculation will be greatly reduced. It only requires calculation of the inner products in input space. Researchers do not need to know the form of nonlinear mapping, or calculate in the high dimensional feature space. The “dimension disaster” is overcome.

For linear inseparable set samples, Cortes and Vapnik introduced the concept of the soft edge optimal hyperplane in 1995, [10] in order to make SVM algorithm suitable for inseparable cases. The constraint condition of the introduced non negative variable ξ_i is loosen:

$$y_i(w^T \cdot x_i - b) - 1 \geq \xi_i, \qquad i = 1, 2, \cdots, l \tag{8}$$

At the same time, the penalty function is introduced into the objective function; the constraint condition of above formula can be written as:

$$\Phi(w, \xi) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{l} \xi_i\right) \tag{9}$$

$$y_i(w^T \cdot x_i - b) - 1 \geq \xi_i, \qquad i = 1, 2, \cdots, l \tag{10}$$

To solve this second programming problem, the Wolfe duality problem is derived as following:

$$\begin{aligned} &Maxmize \qquad && W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j}^{l} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \\ &s.t. && \sum_{i=1}^{l} \alpha_i y_i = 0 \\ & && 0 \leq \alpha_i \leq C \qquad i = 1, \cdots, l \end{aligned} \tag{11}$$

The only difference between the separable cases is the addition of an upper limit in α_i.

According to KKT (Karush-Kuhn-Tucker) condition, there are three conclusions:

$$\alpha_i = 0 \Rightarrow y_i(w^T \cdot x_i - b) > 1$$

$$\alpha_i = C \Rightarrow y_i(w^T \cdot x_i - b) < 1$$

$$0 < \alpha_i < C \Rightarrow y_i(w^T \cdot x_i - b) = 1$$

The geometric meaning goes as following. If the coefficient $\alpha_i = 0$, that is, if the corresponding sample point is not a support vector, these sample points fall outside the boundary of hyperplane, namely the area which does not have classification errors. If $0 < \alpha_i < C$, the corresponding sample point is a support vector, and falls on the boundary of hyperplane $y_i(w^T \cdot x_i + b) = 1$. If $\alpha_i = C$, these samples fall inside the boundary of hyperplane, namely the area which has classification errors.

2.2 Design of classifiers.

The most important step in classifier designing is the selection of kernel function (kernel parameter). The research of Vapnik and his colleagues shows that, the performance of SVM is not related to the type of selected kernel function (US Post handwritten digital library experiment also illustrates this); kernel parameter and error penalty factor are main factors which affect the performance of SVM. In other words, for a given sample, the performance of SVM is mainly influenced by kernel parameters and error penalty parameters. The selection of kernel function determines the structure of characteristic space; the selection of the penalty parameters determines the impact degree of empirical risk.

The LIBSVM[11] used in this paper is a simple and effective software package with SVM recognition and regression pattern; the model is established by Chih-Jen Lin, an associate professor of National Taiwan University. It can solve binary classification and multi classification problems. LIBSVM contains execution files and source codes provided by Windows system, which make it easy to improve and modify. There are a few number of parameter adjustment involved. Many default parameters are provided, which can be used to solve many problems. The function of Cross Validation is adopted in the adjustment of parameters. The software has advantages of less training samples and fast learning speed, so the author decided to adopt this software as the application software. In this paper, the correct rate of training samples and test samples is realized through svm-predict.c. Samples are trained by svm-train.c; the normalization of training samples and test samples, as well as the cross test of parameters are realized through svm-scale.c. The appropriate kernel function (kernel parameter) and penalty factor are determined by cross_validation.py Cross Validation method according to their generalization abilities. In order to achieve a high performance classifier, this paper analyzes SVM classifier design and factors related to classification performance, and refer to the classification results of simulation signals. The SVM classifier used in this paper is built based on the training of all training samples; 5 levels cross validation method is used to obtain the optimized parameters of the classifier.

According to the criterion of classification, the dimension of SVM input vector $X(x_1, x_2, \cdots x_n)$ is determined. The training sample $(x_1, y_1), \cdots, (x_n, y_n),\ x \in R^n, y \in \{+1, -1\}$ satisfies an unknown probability distribution. Through the sample normalization, formula (9), (10) about the optimal classification hyperplane is solved. The maximum classification interval $1/\|W\|$ is achieved.

This is a quadratic programming problem with linear constraints. By using the Lagrange function formula (3), the support vector α is obtained. The optimal hyperplane is determined by (11); the two classes are correctly separated. The SVM classifier has advantages in solving binary classification problems, as well as fast speed and high correct rate. In this paper, this algorithm is fully utilized in accordance with the binary classification thought.

3. Steps of Algorithm Implementation

3.1 Algorithm design.

Audio signals output by the Doppler detector are original data. After obtaining, selecting, filter denoising, as well as feature extraction and selection, the training samples with new data format are stored. Then, the design of classifier and the prediction of unknown samples are carried out in a Libsvm-2.84 environment.

3.2 Signal preprocessing.

From the quasi stationary phase, audio signals with 8000Hz frequency and 10ms time horizon are selected as samples. [12, 13] Noises are removed through Butterworth filter. Then the spectrum of signals is obtained. The maximum frequency curve of signal is extracted by using the percentage method. The result is shown in Figure 4.

The characteristics of signals are acquired through the modulus maximum of wavelet transform. 8 points are sampled in one cardiac cycle as feature points. The results of the fifth layer wavelet modulus maximum analysis are used. Data are normalized as the input of the whole classification system.

The Libsvm-2.84 environment has a specified storage requirement for training and testing data. Data is stored according to categories, namely label 1: attribute value 1, label 2: attribute value 2, label 3: property value 4... Attribute values vary in the range of [-1, 1]. 270 clinical diagnostic data are obtained. 170 are taken as training samples and 100 are taken as testing samples. Data are classified and tested through SVM. Data features are extracted and stored in a specified format.

4. Simulation and Experimental Results

Since the testing samples belong to unknown samples, all the categories of testing samples are set as 1. Because of the variety of kernel functions, SVM algorithm industry is relatively flexible. There are three kinds of kernel functions which are commonly used in SVM. For experimental data, three kinds of inner product functions are respectively used in training and classification experiments.

(1) Polynomial kernel: $K(x_i, x) = [(x \bullet x_i) + 1]^q$

(2) Radial Basis Function

$$K(x, x_i) = \exp\left\{ -\frac{|x - x_i|^2}{\sigma^2} \right\}$$

(3) Sigmoid function kernel: $K(x, x_i) = \tanh(v(x \cdot x_i) + c)$

Libsvm-2.84 is used to train and classify the classifier. Finally, the classification results of testing samples are shown in Table 1.

As can be seen from table 1, the classification accuracy of support vector machines depends on the selection of kernel parameters and error penalty factors. Optimization of kernel parameters and error penalty factors is the key to the performance of classifiers.

Table 1. Classification Results

Kernel function	kernel parameter	C	number of SVM	training time (s)	number of training samples	Classification accuracy rate (%)
Radial basis function (nuclear parameter σ)	10	50	86	105	170	95.4
	5	50	75	78	170	95.6
	2	50	60	65	170	94.3
	1	50	50	60	170	93.5
	2	100	92	64	170	92.5
	2	50	95	62	170	93.6
	2	20	87	54	170	95.8
	2	10	67	46	170	94.7
Polynomial function (Nuclear parameters q)	3	50	73	56	170	96.7
	2	50	73	48	170	96.9
	1	50	69	38	170	95.9
Sigmoid function (kernel parameter v)	1	1	34	45	170	78.6

If different kernel functions and classification accuracy rates are selected, the correct rates of classification will be slightly different. If kernel parameters are different, even with the same kernel function, the accuracy rates of classification will be slightly different. But the correct rates of classification remain above 90%. Therefore, it is feasible and has certain research values to apply SVM in classification. n order to achieve a high performance classifier, this paper analyzes SVM classifier design and factors related to classification performance, and refer to the classification results of simulation signals. The SVM classifier used in this paper is built based on the training of all training samples; 5 levels cross validation method is used to obtain the optimized parameters of the classifier.

By classifying SVM, doctors can save time and improve efficiency. This method has intelligent characteristics, and can provide information for the analysis and research of portable Doppler signals in the future. Doctors can find out the illness of patients on the basis of clinical diagnosis results and abnormal signal classification results. For example, the speed of blood flow can provide information about cerebral vasospasm, cerebral arteriosclerosis and insufficient blood supply; acoustic parameters like PI and RI indexes can show the type of diseases.

5. Conclusion

The SVM method is built on the basis of rigorous mathematical theory. It can find the Optimal Hyperplane through SRM principle, and plays an indispensable role in improving the efficiency of signal classification algorithm. Through programming, training samples are equipped with relatively good generalization abilities. Compared with neural network method, the algorithm proposed in this paper shows superiority in both theoretical basis and model selection flexibility.

References

[1] C. Cortes, V. Vapnik, Support vector networks, J. Machine Learning. 20 (1995).

[2] H. Gish, M. Schimdt, Text-independent speaker identification, J. IEEE Trans on Signal Processing Magazine. 42 (1994).

[3] S. Gao, Support Vector Machine and Personal Credit Evaluation, Xidian University Press, Xi'an, 2013.

[4] N. Cristianini, Support Vector Machines, China Machine Press, Beijing,2005.

[5] J.C. Platt, N. Cristianini, J.S. Taylor, Large margin DAGs for multi-class classification, in: Advances in Neural Information Processing Systems, MIT Press, Cambridge, 2000, pp. 547-553.

[6] J. Kindermann, E. Leopold, G. Paass, Multi-class classification with error correcting codes, In: E. Leopold, M. Kirsten (Eds.), Trefen der GI-Fachgruppe 1.1.3, Maachinelles Lemen, 2000 GMD Report 114.

[7] X.Y. Zhang, Medical Image Analysis based on Feature Extraction and Machine Learning, Nanjing University of Posts and Telecommunications, 2011.

[8] M.R. Azimi-sadjadi, S. Zekavat, Cloud Classification Using Support Vector Machine, In: Prec of the 2000 IEEE Geoscience and Remote Sensing Symposium (IGRASS 2000), Honolulu, Hawaii, 07 (2000).

[9] X.W. Yang, Support Vector Machine Algorithm Design and Analysis, Science Press, Beijing, 2013.

[10] T. Fumitake, A. Shigeo, Decision-Tree-Based Multi-class Support Vector Machines. Information on http://frenchblue.scitec.

[11] C.J. Lin, Libsvm, Information on http://www.csie.ntu.edu.tw/~cjlin/.

[12] T.R. Jin, Applying MATLAB in biomedical signals acquisition and processing, J.Computer Applications and Software. 27 (2010).

[13] D.H. Luo, Feature Extraction and Classification of Transcranial Doppler Signals, Shandong University, 2008.