An Optimization Method for Support Vector Machine Applied to Speech Emotion Recognition

Wanli Zhang [1,a], Guoxin Li[1], Wei Gao [2]

[1] Department of Electrical Information and Engineering, Changchun University, China

[2] Department of Information Engineering, Changchun University of Finance and Economics, China

Abstract. On the basis of the analysis of support vector machine model, an improved MFCC feature parameters have been adopted. Support vector machine has been used as identification model of speech emotion recognition system. For classification problems of support vector machine, an algorithm of optimization parameters has been presented. The algorithm improves the speed of solving and accuracy, and has a good classification results. Experimental results show that, in different environment, the parameter optimization method increases the recognition rate, reduces training time and has good robustness compared to traditional methods.

1 Introduction

Speech is the most accurate and diversified way of communicating emotion among human beings. Speech recognition technology is one of the cutting-edge technologies applied to man-machine interface. In many methods of human emotional state identification using by computer, emotion recognition based on speech signals is a very effective way. Voice information contains in the two aspects, one is the semantic information, on the other hand is a non-semantic information used to express the speaker's emotional tendency. Speech emotion recognition gets the speaker emotional information from the audio files by speech signal processing [1]. Differed from traditional technology such as speech recognition, emotion recognition need only on the acoustic characteristics of speech signal extraction, analysis, and summarized its characteristic parameters and summarizing the laws between the emotional types, so as to identify the emotions of audio files of unknown mood.

Support vector machine is a kind of learning algorithm based on statistical learning theory. The low dimensional linear inseparable problem is mapped to high-dimensional space by using the kernel function, so this converted into linear separable problem and has extensive applied to pattern recognition, signal processing, and other fields [2]. As one of new method of pattern recognition, support vector function better solve the small sample, nonlinear, high dimension and local minimum point and so on practical problems. Compared with the hidden Markov model based on empirical risk minimization, Gaussian mixture model and neural network, method has better classification accuracy, and is more suitable for speech emotion recognition [3].

[a] Corresponding author: wanlizhang@126.com

2 Section Heading

Today, support vector machine method is widely used in the speech emotion recognition. Support vector machine not only can be used in classification, at the same time also can reflect the contributions to the various basic emotions to identify the emotion. This is more applicable for modeling of emotion.

2.1 Linear classification

So-called linear support vector machine is developed based on the concept of the optimal separating hyperplane. In the case of separable, the optimal classification of straight is shown in Figure 1 under the linear separable cases. In Fig. 1, solid and hollow points represent samples of the two kinds, and H is partition line.
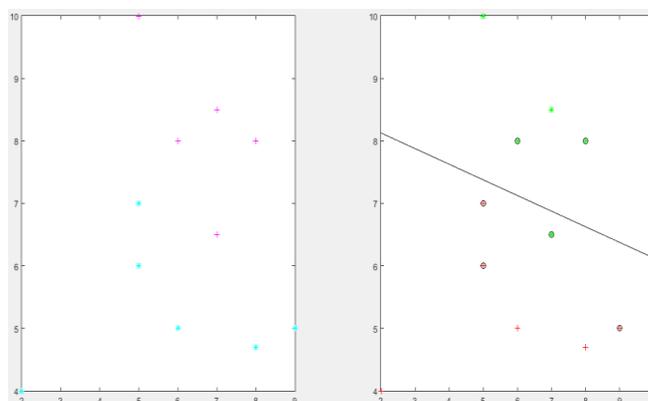

Figure 1. Two dimension linear support vector machine

Suppose two separable sample sets are $(x_i, y_i), i = 1, \cdots, n, \ x \in R^d$. $y \in \{+1, -1\}$ is category label. The general form of the linear discriminated function of d dimension is:

$$g(x) = w \cdot x + b \qquad (1)$$

Classification of surface equation is:

$$w \cdot x + b = 0 \qquad (2)$$

Two classes of samples correctly separate, thus we can get:

$$(w \cdot x_i) + b \geq 1, \ y_i = 1 \qquad (3)$$

$$(w \cdot x_i) + b \leq -1, \ y_i = -1 \qquad (4)$$

The above two inequalities are combined for the normalization conditions:

$$y_i[(w \cdot x_i) + b] \geq 1, \ i = 1, 2, \cdots, n \qquad (5)$$

A sample point to the hyperplane for distance is:

$$d = \frac{|w \cdot x + b|}{\|w\|} \qquad (6)$$

The margin is $2/\|w\|$ .Therefore, hyperplane meet the above conditions and make the $\|w\|^2$ minimum is the optimal hyperplane [4]. Based on the above discussion, constructing the optimal hyperplane is always converted into conditional extreme value problem, and strives for the minimum value of function $\phi(w) = \frac{1}{2}\|w\|^2$. Therefore, Lagrange function is defined as follows:

$$L(w, b, a) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} a_i \{y_i[(w \cdot x_i) + b] - 1\} \qquad (7)$$

In the type, a_i is the nonnegative Lagrange coefficient. The minimization problem under the original constrained condition is converted into the quadratic programming problem which minimizes the function $L(w, b, a)$. So that is:

$$\frac{\partial L(w, b, a)}{\partial w} = w - \sum_{i=1}^{n} y_i a_i x_i = 0 \qquad (8)$$

$$\frac{\partial L(w, b, a)}{\partial b} = \sum_{i=1}^{n} y_i a_i = 0 \qquad (9)$$

Solving the type and substitution to the original Lagrangian function, we get:

$$L(a) = \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j (x_i \cdot x_j) \qquad (10)$$

If a_i^* is the optimal solution, then:

$$w^* = \sum_{i=1}^{n} a^* y_i x_i \qquad (11)$$

Namely the optimal classification the weight vector is the linear combination of training sample vector. This optimization problem solution must satisfy:

$$a_i(y_i(w \cdot x_i + b) - 1) = 0, \ i = 1, \cdots, n \qquad (12)$$

To solve the above problem after get the optimal classification function is:

$$f(x) = \text{sgn}\{(w^* \cdot x) + b^*\} = \text{sgn}\left\{\sum_{i=1}^{n} a_i^* y_i(x_i \cdot x) + b^*\right\} \qquad (13)$$

2.2 Nonlinear classification

If the classification is not linearly separable, the new feature vector is constructed, which transforms the problem into a new space. Compared with the original space dimensionality the new space increases, but can use linear discriminant function to realize the nonlinear discriminant function in the original space. The optimal hyperplane in the high dimensional feature space is shown in Figure 2. Considering the nature of the optimal hyperplane algorithm, in the transform space, we just need calculate the inner product.
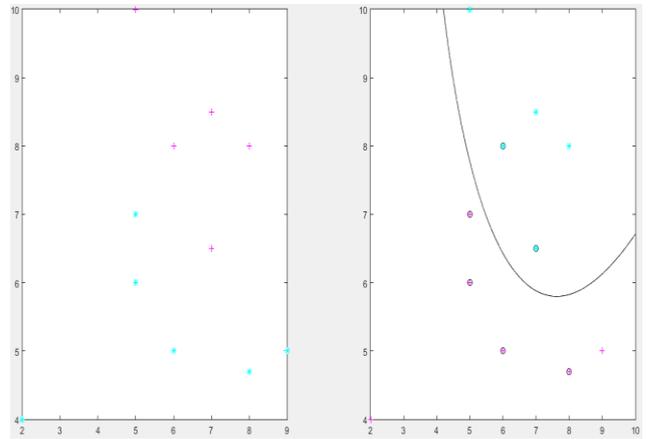

Figure 2. The linear kernel function

For any symmetric function $K(x_i \cdot x_j)$, its inner product operation in a feature space of the sufficient and necessary condition is that for any $\varphi(x) \neq 0$ and $\int \varphi^2(x) dx < \infty$ we get:

$$\iint K(x_i, x_j) \varphi(x) \varphi(x') dx dx' > 0 \qquad (14)$$

The optimization function is:

$$\max \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j K(x_i, y_j) \qquad (15)$$

The corresponding discriminant function is:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{n} a_i^* y_i K(x_i, x) + b^*\right) \quad (16)$$

In essence, kernel function determines the structure of the feature space, so it is very important in the choice [5]. Although the choice of kernel function directly affects the algorithm implementation and effects, the research on this aspect is still a lack of theoretical substantive progress. Generally we can only rely on experience, and RBF kernel function is effective [6]. The recognition effect of RBF kernel function is shown in Figure 3.
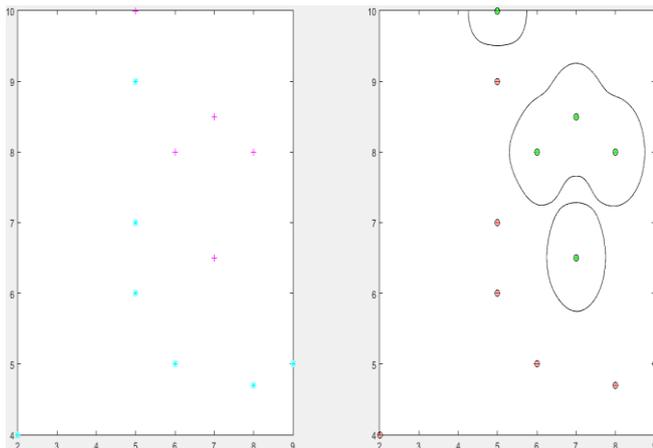

Figure 3. The Gaussian kernel function

2.3 Parameter optimization algorithm

Support vector machine is based on the structural risk minimization principle, and its core idea is the introduction of kernel function. Different types of kernel function will lead to different classification performance of support vector machine. So how to choose appropriate kernel function is the current urgent problems [7]. Therefore，the mixed kernel function is:

$$(w \cdot x_i) + b \geq 1, \ y_i = 1 \quad (17)$$

3 Results

Voice database are recorded in soundproof room. There are 30 speakers, 20 times recorded for each emotion and emotional corpus of 2400 words. All input emotional speech requires processing by Matlab 2015 with libsvm toolbox. In this paper, sampling frequency is selected for 16 KHZ, frame length of 256 points, frame at 128 points and window function using hamming window. The median filter is to smoothly process emotional speech after denoising, and the feature parameters are MFCC [8]. Table 1 is recognition results of the traditional algorithm and optimization algorithm of speech emotion, it can be seen that the optimized algorithm has different levels of the recognition rate. This indicates the optimization algorithm of support vector machine model is superior to the traditional support vector machine model.

Table 1. Recognition rate of speech emotion

Method	Emotion type			
	Calm	Joy	Anger	Sad
Tradition	75	76	72	73
Optimization	78	79	74	75

4 Conclusions

In the process of speech emotion recognition, the parameters optimization algorithm of support vector machine is proposed. Through a lot of voice and data experiment, the results show that support vector machine model for parameter optimization algorithm is reliable and effective. Compared with the traditional support vector machine, speech emotion recognition rate is obviously improved and has good robustness, widely applicable to various characteristics of each sample. From the experimental results, speech emotion recognition rate is about 80%, and different emotion recognition ability has certain disparity [9]. The reason may be that the pattern recognition method has yet to be further optimized and improved, and this is also our efforts in the direction of the future.

REFERENCES

1. Peng Song,Yun Jin,Cheng Zha, and Li Zhao.Speech emotion recognition method based on hidden factor analysis.Electronics Letters, Vol. 51, Issue 1, pp112-113, （2015）
2. IIHAN AYDIN, MEHMET KARAKOSE, ERHAN AKIN. A multi-objective artificial immune algorithm for parameter optimization in support vector machine. Applied Soft Computing, (2011)
3. Zhang wanli,Li guoxin and GAO Wei.The Research of Speech Emotion Recognition Based on Gaussian Mixture Model.Mechanical Components and Control Engineering III, Vol1, pp. 1126-1129, November (2014)
4. Arun K M, Gopal M. Reduced one-against-all method for multiclass SVM classification. Expert Systems with Application, (2010)
5. Domenico C, Rostita G. Kernel based support vector machine via semidefinite programming: Application to medical diagnosis. Computer&Operations Research, (2010)
6. Maldonado S, Weber R, Basak J. Simultaneous feature selection and classification using kernel-penalized support vector machines. Information Sciences, (2011)
7. TANG Tao, GUO Qing, YANG Mingchuan. Support vector machine based particle swarm optimization localization algorithm in WSN. Journal of Convergence Technology, (2012)

8. Wanli zhang, Guoxin Li. The Research of Feature Extraction Based on MFCC for Speaker Recognition .3rd International Conference on Computer Science and Network Technology, October(2013)

9. Zhang wanli,Li guoxin and Wang Lirong. Application of Improved Spectral Subtraction Algorithm for Speech Emotion Recognition.The 2015 International Workshop on Assistive Engineeringand Information Technologies，(2015)