Journal of System Simulation

Volume 29 | Issue 5 Article 11

6-3-2020

Improved SVDD for Speech Recognition and Simulation

Hao Rui
1. College of Information Management, Shanxi University of Finance & Economics, Taiyuan 030006, China; ;

Xiaofeng Liu
2. Taiyuan University of Technology, Taiyuan 030024, China;

Yanbo Niu
2. Taiyuan University of Technology, Taiyuan 030024, China;

Xiu Lei
1. College of Information Management, Shanxi University of Finance & Economics, Taiyuan 030006, China; ;

Improved SVDD for Speech Recognition and Simulation

Abstract

Abstract: Support vector data description (SVDD) defines multi-class data by their respective hyper-spheres. The computational complexity of the quadratic programming problem is reduced significantly and it is easier to solve multi-class classification problems. Thus, SVDD has attracted more and more attention in the field of speech recognition research. For the problems of the feature vectors of speech samples overlapping and updating, the conventional SVDD for multi-class classification was improved. *On the one hand, the spatial position of the samples was fully used to construct the decision function in overlapping domain of hyper-spheres; On the other hand, based on class incremental learning the dynamic change of support vectors was implemented.* Simulation experimental results indicate that the proposed method reduces modeling time obviously and has better recognition performance.

Keywords

support vector data description, multi-class classification, decision function, incremental learning, speech recognition system simulation

Recommended Citation

Hao Rui, Liu Xiaofeng, Niu Yanbo, Xiu Lei. Improved SVDD for Speech Recognition and Simulation[J]. Journal of System Simulation, 2017, 29(5): 1014-1020.

Improved SVDD for Speech Recognition and Simulation

Hao Rui[1], *Liu Xiaofeng*[2], *Niu Yanbo*[2], *Xiu Lei*[1]

(1. College of Information Management, Shanxi University of Finance & Economics, Taiyuan 030006, China;
2. Taiyuan University of Technology, Taiyuan 030024, China)

Abstract: Support vector data description (SVDD) defines multi-class data by their respective hyper-spheres. The computational complexity of the quadratic programming problem is reduced significantly and it is easier to solve multi-class classification problems. Thus, SVDD has attracted more and more attention in the field of speech recognition research. For the problems of the feature vectors of speech samples overlapping and updating, the conventional SVDD for multi-class classification was improved. *On the one hand, the spatial position of the samples was fully used to construct the decision function in overlapping domain of hyper-spheres; On the other hand, based on class incremental learning the dynamic change of support vectors was implemented.* Simulation experimental results indicate that the proposed method reduces modeling time obviously and has better recognition performance.

Keywords: support vector data description; multi-class classification; decision function; incremental learning; speech recognition system simulation

面向语音识别的 SVDD 改进算法及仿真研究

郝瑞[1]，刘晓峰[2]，牛砚波[2]，修磊[1]

(1. 山西财经大学 信息管理学院，太原 030006；2. 太原理工大学，太原 030024)

摘要：支持向量数据描述(SVDD)将多类样本数据每一类用各自的超球来界定，显著降低了二次规划计算复杂度，更易于解决多类分类问题，因此在语音识别研究领域越来越受到广泛关注，本文针对语音样本分类中特征向量重叠和更新等问题，对现有的 SVDD 多类分类算法进行了改进，*一方面，根据样本所在空间位置，构造超球重叠域决策函数；另一方面，基于类增量学习，实现超球类支持向量的动态改变。*仿真实验结果表明，本文所提方法明显缩短了建模时间并且具有更好的识别性能。

关键词：支持向量数据描述；多类分类；决策函数；增量学习；语音识别系统仿真

中图分类号：TP391.9　　文献标识码：A　　文章编号：1004-731X (2017) 05-1014-08

DOI: 10.16182/j.issn1004731x.joss.201705011

Introduction

Speech recognition is the key technology of the intelligent man-machine interaction. Currently, the speech recognition technology has been applied widely in many relevant areas and resulted in excellent performances.

The Support Vector Machine (SVM) proposed by Vapnik et al is a very promising classifier which is used in speech recognition applications[1-7]. Its

Received:2016-11-10　　Revised:2017-02-05;
Foundation Project: National Natural Science Foundation of China (61072087), Shanxi University of Finance & Economics School youth fund(QN2015014);
Biography: HAO Rui (1978-), W, PhD, Lecturer, Research interests: Artificial Intelligence.

theoretical fundamentals are based on the theory of VC-dimension and the principal of structural risk minimization. SVM performs well in solving the small sample, high dimensions, non-linearity, and local minimum point problems. SVM is originally designed for binary classification. But speech recognition is a typical multi-class classification problem. In order to extend it for multi-class classification, there are mainly two types of multi-class SVMs currently. One is to consider directly all classes in one optimization formulation, but since it is too complicated to solve the optimization problem, the presented method is quite difficult to be implemented in the practical applications; while the other is to construct and combine several binary classifiers[8-10] such as one-versus-one (1-V-1), one-versus-all (1-V-R), and directed acyclic graph SVM (DAG-SVM). All these methods work by building a number of optimal separating hyper-planes with maximal-margin to divide the sample space. Each class of samples is confined into a certain area by several hyper-planes. Because of their high complexity of quadratic programming, the computational time of these approaches is too long, especially for large-scale multi-class classification problems. Instead of hyper-plane SVM, Support Vector Data Description (SVDD)[11-15] is presented by Tax et al which is inspired by the SVM classifier. The objective of SVDD is to obtain a spherically shaped boundary around the target data by using kernel functions. This hyper-sphere is defined as the smallest volume that covers as possible as all the target data in a high-dimensional feature space. Because this algorithm can shift SVM convex quadratic programming to calculate a minimum ball-shaped boundary, its application in the multi-class classification problem can significantly reduce the computational complexity and effectively enhance classification speed and generalization ability. Thus, SVDD model is more suitable to deal with the large-scale speech recognition.

However natural speech carries information not only about what is said, but also about tone of voice, emotional state, language species and so on. It makes the distribution of speech sample points overlap in feature space. Actually, there is no sharp distinction between two different utterances. Thus, each class-specific hyper-sphere of speech samples could not be completely isolated. In addition, when the number of speech training samples increases retraining is constantly carried out to adjust the SVDD model which wastes a lot of training time. Thus, the conventional SVDD needs to be improved. In this paper, a new decision function in overlapping domain of hyper-spheres is designed to determine the category for the input speech samples and dynamic change of the hyper-sphere class is realized via incremental learning. Finally, the experiment verifies the proposed solution is effective for speech recognition.

1 SVDD for multi-class classification and its improvement

For a $k(k>2)$ class problem, given training date sets composed of k class examples: $T^i=(x_1^i,y^i),\cdots,(x_{m_i}^i,y^i)$, $(i=1,\cdots,k)$. Each T^i includes m_i points that belongs to the same class i and each point x_j^i , $(j=1,\cdots,m_i)$ is an n-dimensional vector. The objective of SVDD is to find the best sphere-structured decision boundary as the smallest volume that contains all possible target data in feature space. This minimum boundary hyper-sphere S^i for each class is described by

https://dc-china-simulation.researchcommons.org/journal/vol29/iss5/11
DOI: 10.16182/j.issn1004731x.joss.201705011

center a^i and minimum radius R^i which can be found by solving the following constrained quadratic optimization problem:

$$\min(R^i)^2 + \frac{C}{m_i}\sum_{j=1}^{m_i}\xi_j^i$$
$$\text{s.t.}\left\|\phi(x_j^i) - a^i\right\|^2 \leqslant (R^i)^2 + \xi_j^i$$
$$\xi_j^i \geqslant 0\text{，}\ j = 1,2,\cdots,m_i \tag{1}$$

where the slack variable ξ_j^i is used to indicate the effect of target sample points not included in the spherical description. The constant C>0 controls the tradeoff between the volume of the hyper-sphere and the number of target sample points rejected. m_i is the number of target sample points within class i. A nonlinear mapping ϕ can transform the training data into a high-dimensional feature space and compute the minimum boundary hyper-sphere in this feature space. Introducing Lagrangian multipliers to account for the constrained quadratic optimization problem, we obtain the following dual problem:

$$\max\sum_{j=1}^{m_i}\alpha_j^i K(x_j^i, x_j^i) - \sum_{j,l=1}^{m_i}\alpha_j^i\alpha_l^i K(x_j^i, x_l^i)$$
$$\text{s.t.}\sum_{j=1}^{m_i}\alpha_j^i = 1$$
$$0 \leqslant \alpha_j^i \leqslant \frac{C}{m_i}\text{，}\ j = 1,2,\cdots,m_i \tag{2}$$

where α_j^i is the Lagrange multiplier and the kernel function $K(x,y) \equiv \langle\phi(x)\cdot\phi(y)\rangle$. Solving the dual problem is to produce the k hyper-spheres with minimum volume and containing most of the target date. The points lying on the boundary surface are called support vectors which are needed for the description of the sphere. According to the above results, the decision function (3) is constructed where I denotes an indicator function. Suppose x for the test point, when meet (3), x is belongs to the target class; otherwise, x is considered as an outlier.

$$f^i(x, a^i, R^i) = I(\|\phi(x) - a^i\|^2 \leqslant (R^i)^2) =$$
$$I(K(x,x) - 2\sum_{j=1}^{m_i}\alpha_j^i K(x, x_j^i) +$$
$$\sum_{j,l=1}^{m_i}\alpha_j^i\alpha_l^i K(x_j^i, x_l^i) \leqslant (R^i)^2)$$
$$i = 1,2,\cdots,k \tag{3}$$

1.1 The decision rule in overlapping domain

The traditional SVDD speech classification model assumes that any two hyper-spheres are independent of each other and all speech samples can be classified correctly. But in fact since the speech sample points are correlated, it makes the distribution of speech sample points overlap. Thus, how to classify the speech samples in overlapping domain accurately is an important issue in current research.The decision function which is used to assess whether the test sample are accepted is an important basis for speech sample classification. If the point is excluded from all the spheres, it does not belong to any class-specific hyper-sphere. If the point is included in only one hyper-sphere, it belongs to the corresponding class. If the point is located in the area of two or more spheres, the test point needs to be re-judged in the overlapping domain. Imbalance is the most common phenomenon in the classification problem. Fig. 1 depicts that within the overlapping domain only using the distance from the test point to the center or surface of the hyper-sphere as classification decision is not reliable. Because of the different size of the sphere or the different density of the sample distribution contained in the sphere the prediction results of the classifier have a certain tendency that decreases classification performance. Thus, the decision rule in overlapping domain takes into account not only the distance between the spherical boundary and the data points but also the distribution of the data. Without considering the density distribution of the data, it is possible that the new test point in the highest density area is not classified as that class.

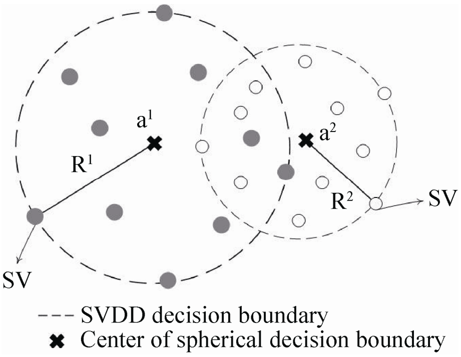

a1
a2
R1
R2
SV
SV
--- SVDD decision boundary
✖ Center of spherical decision boundary

Fig. 1 Support vector data description (SVDD) in feature space

Inspired by the nearest neighbor method and comprehensively considering the size relation of overlapping hyper-spheres, we proposed the following decision algorithm. Assume the test point *x* in the overlapping domain of the m hyper-spheres.

Step 1: For the neighbor distributing of speech samples, compute the distance from the test point *x* to each speech sample point in the m overlapping hyper-spheres.

Step 2: Select the k-nearest neighbor samples according to the kNN algorithm. Calculate the average distance between the test point and each class of the k-neighbor samples using the following formula:

$$\bar{D}(x,c_j)=\frac{1}{\mathrm{num}(c_j)}\sum_{i=1}^{k}\mathrm{dist}(x_i,x)\delta(x_i,x) \quad (4)$$

where c_j , $(j=1,2,\cdots,m)$ denotes category and $\mathrm{num}(c_j)$ is the number of samples belonging to class c_j ; $\mathrm{dist}(x_i,x)$ is the distance between x and the k-nearest neighbor of x_i .

$$\delta(x_i,x)=\begin{cases}1, x_i\in c_j\\ 0, x_i\notin c_j\end{cases}$$

is two valued class function.

Step 3: For the radii of the hyper-spheres, calculate the relative distance between *x* and the center of each overlapping hyper-sphere.

$$\rho(x,c_j)=\frac{\mathrm{dist}(x,a^j)}{R^j} \quad (5)$$

Where a^j and R^j are the center and radius of class c_j hyper-sphere respectively.

Step 4: The corresponding decision function in overlapping domain for the test sample point *x* is given by:

$$C_{\text{x fall into}}=\arg\min\bar{D}(x,c_j)\rho(x,c_j)$$
$$j=1,2,\cdots,m \quad (6)$$

It can be seen that the classification algorithm not only considers the average distance of the k-neighbor sample points in different categories, moreover also takes into account the effect of different sphere radii. Thus the proposed algorithm has better anti-noise ability and higher accuracy than the traditional KNN algorithm.

1.2 SVDD for class incremental learning

The purpose of class incremental learning [16-17] is to determine how to deal with newly added categories of training samples and use the results of the previous training to get a better classification results and avoid training the same training set repeatedly. Traditional multi-class SVM classification system will be broken when a new class is added. It is necessary to train the new class and all previous classes repeatedly to get new support vectors. This will lead to increase greatly computational time and memory cost. In the SVDD classification method, no such repetition is needed since the class-specific hyper-spheres are constructed for the data of each class separately. In the process of class incremental learning, only the samples that belong to the new incremental class are trained. The new class will not influence the previous classes. Thus the class incremental learning can be realized in a small sample set and a small memory space and the training time is reduced largely and the algorithm makes it easy for data to expand. Based on class incremental learning the multi-class SVDD for speech recognition is described below:

Let X^m be a subset of speech samples which is the mth class, $m \in \{1, 2, \cdots, N\}$. For each class of speech samples X^m, the SVDD classifier is used to construct the smallest hyper-sphere (a^m, R^m) in feature space that contains most speech samples of the class, where a^m is the center of the *mth* class hyper-sphere and R^m is the radius of the *mth* class hyper-sphere. If a new speech sample set X^{m+1} is added, the corresponding hyper-sphere (a^{m+1}, R^{m+1}) is determined by the SVDD classifier in the feature space to implement speech sample class incremental learning.

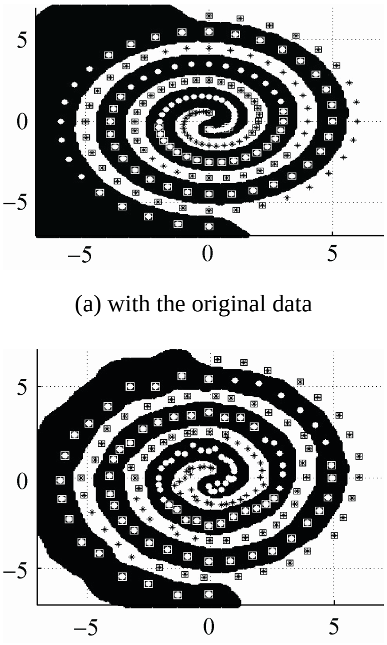
2 Simulation experiments and analysis

2.1 Bi-spiral test

The bi-spiral problem is a classic test for pattern recognition algorithm and also one of the most difficult pattern classification problems. Its purpose is to separate two highly inter-related spirals. Since the two spiral lines are intertwined with each other, that is to say, they are overlapping. It will increase the difficulty of pattern classification.

The bi-spiral data set has 97 data for each class in two dimensions. The output result is +1 which represents the positive class or -1 which represents the negative class. In order to test the anti-noise performance of bi-spiral problem, we add the Gaussian noise with the mean value of 0 and the standard deviation of 0.3 to the original data. The MATLAB and The SVDD toolbox dd_tools is use.

Fig. 2 (a) plots the classification boundary by the improved SVDD which has perfectly separated the spirals of the original data. The boxes represent support vectors, the white region represents (+1) class, and the black region represents (-1) class. The improved SVDD shows the smallest number of SVs while keeping the generalization ability for the dataset. Fig.2 (b) plots the classification boundary by the improved SVDD with noises. In this condition, the improved SVDD is also able to separate spirals correctly. The result shows that the improved SVDD has strong anti-noise ability.

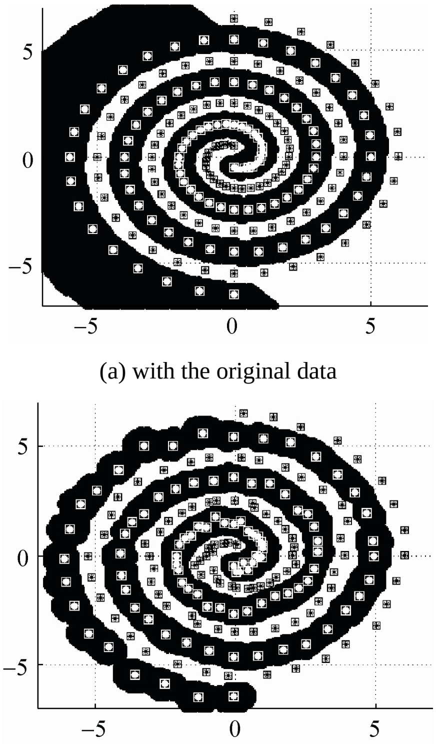


(a) with the original data

(b) with the noisy data

Fig. 2 The boundary with improved SVDD

For comparison, Fig.3 plots the classification boundary by the conventional SVDD. Fig.3 (a) shows that the conventional SVDD also has separated spirals correctly, but it has a larger number of SVs. Fig.3 (b) shows that nearly all training sets are SVs and the two regions separated are uneven and the boundary of region is not smooth. The result shows that the anti-noise ability of the conventional SVDD is limited.



(a) with the original data

(b) with the noisy data

Fig. 3 The boundary with the SVDD

2.2 Speech recognition

In the simulation experiment, we employ the embedded speech recognition system based on the TI company DM6446 processor. The SVDD optimization algorithm is implanted to an embedded chip. Through repeated experiment and test the algorithm, the system stability is available. The speech data are from Korean isolated words of small vocabularies spoken by non-specific persons obtained by the sampling system. The noise is the white Gaussian noise. The sampling rate of the speech signals is 11.025 kHz. In the process of pretreatment, the frame length N is 256 bits and the frame shift M is 128 bits. The speech feature of this speech data is Mel Frequency Cepstrum Coefficient (MFCC). The feature parameters extracted are normalized through the Dynamic Time Warping (DTW) that means every pronunciation of each single word are arranged in the speech vector sequence of 1024 dimensions.

The vocabularies of the speech data in this experiment are 10 words, 20 words, 30 words, 40 words and 50 words respectively. The training set is the speech of nine speakers in different condition of 15 dB, 20 dB, 25 dB, 30 dB and clear. Each word was uttered 3 times by each of the nine speakers. The testing set consists of the pronunciation of seven other speakers in the corresponding SNR and vocabularies. The experiments use Gaussian kernel throughout and the optimal values of the parameters C and σ are found through the method of grid search. Table 1 shows the recognition results.

From Table 1 it can be seen that under the condition of different SNR and different vocabulary, recognition rates are improved to some extent.

Tab. 1 Comparison of speech recognition rates based on SVDD and improved SVDD

Vocabulary	Algorithm	SNR (15 dB)	SNR (20 dB)	SNR (25 dB)	SNR (30 dB)	SNR (CLEAN)
10	SVDD	95.22	96.15	96.58	95.69	96.55
	Improved SVDD	95.32	97.63	97.89	98.12	98.64
20	SVDD	94.05	96.43	96.91	96.43	96.57
	Improved SVDD	95.10	97.32	97.85	98.06	98.49
30	SVDD	89.76	91.60	92.92	93.18	96.03
	Improved SVDD	93.14	95.27	96.04	96.71	97.59
40	SVDD	89.87	91.67	92.14	92.81	95.83
	Improved SVDD	93.06	94.46	94.90	96.53	97.38
50	SVDD	88.11	90.89	92.21	92.68	93.53
	Improved SVDD	92.42	93.65	93.99	95.87	96.04

As shown in Fig. 4 the training time of the conventional SVDD increases obviously as the new speech samples are added constantly, while the training time of improved SVDD based on incremental learning increased slowly. In addition it can be seen that when dealing with a large number of categories and data volumes, its advantage in training speed is more obvious.

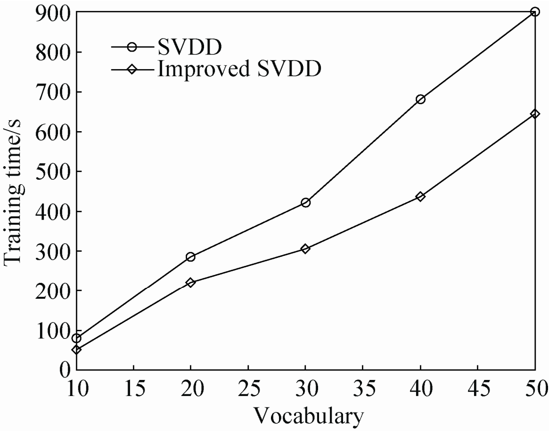


Fig. 4 Training time comparison

According to the above experimental results, the proposed approach not only reduces the modeling time, but also has better anti-noise ability and recognition performance than the conventional SVDD. It is more suitable for speech recognition system.

3 Conclusion

The speech recognition modeling of improved SVDD is proposed in this paper. By constructing the decision function in hyper-sphere overlapping domain higher, speech recognition rate is achieved and using incremental learning new category samples can be added to the original classification modeling. Thus the training time is obviously reduced. Experimental results show that the proposed method performs well on both bi-spiral problem and speech recognition. It is suitable for speech recognition system. Next work, we consider further optimization algorithm on large-scale data sets.

References:

[1] Chen Yanxiang, Xie Jian. Emotional speech recognition based on SVM with GMM [J]. Journal of Electronics (China) (S0372-2112), 2012, 29(3): 339-344.

[2] Zhang S X, Gales M J F. Structured SVMs for Automatic Speech Recognition [J]. IEEE Transactions on Audio Speech & Language Processing (S1558-7916), 2013, 21(3): 544-555.

[3] Elyes Zarrouk, Yassine Ben Ayed, Faiez Gargouri. Hybrid continuous speech recognition systems by HMM, MLP and SVM: a comparative study [J]. International Journal of Speech Technology (S1381-2416), 2014, 17(3): 223-233.

[4] Sangeetha J, Jothilakshmi S. Speech translation system for english to dravidian languages [J]. Applied Intelligence (S0924-669X), 2016, 46(3): 534-550.

[5] Kelly D, Caulfield B. Pervasive Sound Sensing: A Weakly Supervised Training Approach [J]. IEEE Transactions on Cybernetics (S2567-5471), 2016, 46(1): 123-135.

[6] Yang N, Yuan J, Zhou Y, et al. Enhanced multiclass SVM with thresholding fusion for speech-based emotion classification [J]. International Journal of Speech Technology(S1381-2416), 2016, 20(1): 27-41.

[7] Hao R, Niu Y B, Xiu L. Improved Support Vector Pre-Extracting Algorithm in Speech Recongnition Application [J]. Journal of System Simulation (S1004-731X), 2015, 27(11): 2714-2721.

[8] Ren X Y, Qi Y Z. Hadoop-based Multi-classification Fusion for Intrusion Detection [J]. Journal of Applied Sciences (S1812-5654), 2013, 13(12): 2178-2181.

[9] Wang T, Zhao D, Feng Y. Two-stage multiple kernel learning with multiclass kernel polarization [J]. Knowledge-Based Systems (S0950-7051), 2013, 48(2): 10-16.

[10] Yi H, Song X, Jiang B. Structure selection for DAG-SVM based on misclassification cost minimization [J]. International Journal of Innovative Computing Information & Control (S1349-4198), 2011, 7(9): 5133-5143.

[11] Tax D M J,Duin R P W. Support Vector Data Description [J]. Machine Learning (S0885-6125), 2004, 54(1): 45-66.

（下转第 1027 页）