Yun Jin[1,2], Wenming Zheng[1], Li Zhao[1], Jingjie Yan[1]

Southeast University (1), Jiangsu Normal University(2)

Generalized Maximal Margin Discriminant Analysis for Speech Emotion Recognition

Abstract. *A novel speech emotion recognition method based on the generalized maximum margin discriminant analysis (GMMDA) method is proposed in this paper. GMMDA is a multi-class extension of our proposed two-class dimensionality reduction method based on maximum margin discriminant analysis (MMDA), which utilizes the normal direction of optimal hyperplane of linear support vector machine (SVM) as the projection vector for feature extraction. To generate an optimal set of projection vectors from MMDA-based dimensionality reduction method, we impose orthogonal restrictions on the projection vectors and then recursively solve the problem. Moreover, to deal with the multi-class speech emotion recognition problem, we present two recognition schemes based on our proposed dimensionality reduction approach. One is using "one-versus-one" strategy for multi-class classification, and the other one is to compose the projection vectors of each pair of classes to obtain a transformation matrix for the multi-class dimensionality reduction.*

Streszczenie. *Wartykule przedstawiono metodę analizy emisji głosu pod kątem rozpoznawania emocji. Rozwiązanie bazuje na analizie dyskryminacyjnej maksymalnego marginesu GMMDA. (A**naliza dyskryminacyjna maksymalnego marginesu w rozpoznawaniu emocji w mowie**)*

Keywords: Dimensionality Reduction, Speech Emotion Recognition, Generalized Maximal Margin Discriminant Analysis, Support Vector Machine
Słowa kluczowe: redukcja wymiarowości, rozpoznawanie emocji w głosie, analiza dyskryminacyjna.

Introduction

It is well known that the same sentence carrying different emotion presents the totally different meaning. So in human-machine interaction applications, it is important for computer to recognize not only the linguistic content but also the emotional states in human speech [1].

Overall, speech emotion recognition procedures can be divided into three steps, i.e., the speech emotion features extraction step, the dimensionality reduction step of the feature vector, and the classification step. All these three steps play important roles to a successful emotion recognition system. In speech emotion features extraction, one may obtain more than thousands of emotion features by using the analytical feature generation approach [2][3] and statistical approach. Some of the famous features are the mean, standard detion, range and skewness of pitch, energy and MFCC [4]. Consequently, the extraction of a reasonably limited, meaningful, and informative set of features is always desirable for an automated speech recognition recognition [5]. In this paper, we focus our attention on the second step, i.e., devising an efficient algorithm for the dimensionality reduction of emotion feature vectors. Dimensionality reduction is an active research topic in pattern recognition and machine learning area [8]. Two of the most popular used aproaches in emotion speech recognition are principle component analysis (PCA) [9] and linear discriminate analysis (LDA) [10].

PCA is one of the most widely-known linear dimensionality reduction methods in the mean-square error sense[11] and is an unsupervised feature reduction method. Being based on the covariance matrix of the variables, it is a second-order method. In various fields, it is also known as the singular value decomposition (SVD), the Karhunen-Loeve transform, the Hotelling transform, and empirical orthogonal function(EOF) method[6]. In essence, PCA seeks to reduce the dimension of the data by finding a few orthogonal linear combinations (the PCs) of the original variables with the largest variances. The first PC is the linear combination with the largest variance. The second PC is the linear combination with the second largest variance and orthogonal to the first PC, and so on. There are as many PCs as the number of the original variables. For many data sets, the first several PCs explain most of the variance, so that the rest can be disregarded with minimal loss of information. But PCA requires the guess of the dimensionality of the target space while it is not always a easy thing.

LDA is a well-known scheme for feature extraction and dimension reduction. It has been used widely in many applications such as face recognition, image retrieval, microarray data classification, etc. Classical LDA projects the data onto a lower-dimensional vector space such that the ratio of the between-class distance to the within-class distance is maximized, thus achieving maximum discrimination. The optimal projection (transformation) can be readily computed by applying the Eigen-decomposition on the scatter matrices. An intrinsic limitation of classical LDA is that its objective function requires the nonsingularity of one of the scatter matrices. For many applications, such as face recognition, all scatter matrices in question can be singular since the data is from a very high-dimensional space, and in general, the dimension exceeds the number of data points.

Besides, it is worth to mention Margin Maximizing Discriminant Analysis (MMDA) [11]. MMDA projects input patterns onto the subspace spanned by the normals of a set of pairwise orthogonal margin maximizing hyperplanes[11]. It is a non-parametric method which is regarded as an extension of LDA without normality assumptions on the data. MMDA extract first feature component ω_1 by find the solution of a quadratic programming problem. Then transform the data by projecting it onto a space orthogonal to ω_1. With the projected data, MMDA find the second feature component ω_2 and projected onto a space orthogonal to it again. Each projection, the dimension is reduced by one. After iterations, we can obtain the data with low dimension. MMDA is proved to indeed compete with other alternative feature extraction method. So MMDA have been used in many application such as face recognition field[12]. Besides, combined with the Core Vector Machines, modified MMDA have also been used in large scale data and obtained good recognition rates with faster computation speed[13][14][15]. Some other algorithms also utilize the idea of Maximum Margin[16][17]. But one of the problems of MMDA is that once the data can not be directly separated after projection with certain time iterations, the cumulative error will greatly enhanced which will badly influence the recognition effect. Furthermore MMDA is only for two class analysis. For multi-class problems, "one-versus-all" method is used[11][13]. So, in this paper we present our proposed dimensionality reduction method based on maximum margin, use "one-versus-one" method to extend our method to multi-class and then give out a new multi-class dimensionality reduction method.

This paper is organized as follows. Section 2 introduce the principle of two-class dimensionality reduction method based on maximal margin discriminant analysis and derive the formulation for program. Section 3 firstly use “one-versus-one" method based on two class dimensionality reduction method to separate multi-class speech emotion and then introduce our proposed multi-class dimensionality reduction method which can reduce the dimensionality of multi-class at once. Which and how many emotion features chosen will be listed in the Section 4. The database used in this paper will be simply introduced in Section 5 and some experiments are carried out to prove the effective performance of our method. Discussion and conclusion will be given out in Section 6.

Proposed Two-Class Dimensionality Reduction Method Based On Maximal Margin Discriminant Analysis

The basic idea of SVM is to construct a hyperplane that separates two classes of data samples (labeled $y \in \{-1, +1\}$), such that the margin (the distance between the hyperplane and the nearest point) is maximal[17]. This gives the following optimization problem:

$$\omega_1 = \arg\min_{\omega} \phi(\omega, \varepsilon) = \frac{1}{2}(\omega \cdot \omega) + C\sum_{i=1}^{l} \varepsilon_n, \tag{1}$$

$$\text{s.t.} \begin{cases} y_i((\omega \cdot x_i) + b) \geq 1 - \varepsilon_i, & i = 1, \cdots, l \\ \varepsilon_i \geq 0, & i = 1, \cdots, l \end{cases}$$

The dual of the optimization problem in (1) is

$$W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j x_i x_j \tag{2}$$

$$\text{s.t.} \begin{cases} 0 \leq \alpha_i \leq C, & i = 1, \cdots, l \\ \sum_{i=1}^{l} \alpha_i y_i = 0, & \\ \omega = \sum_{i=1}^{l} \alpha_i y_i x_i. & \end{cases}$$

Our proposed dimensionality reduction method based on the idea of MMDA makes use of the principal idea underlying SVM and PCA: SVM searches for the maximal margin between the hyperplanes, and PCA gives out a serial of PCs of the original variables with the largest variances which are orthogonal to each others. The proposed dimensionality reduction Method combines these two characteristics to search for the maximal margin in each dimension and then we can get norm vectors which are orthogonal to each others. These vectors are sorted by the classification margin. At last we can get N norm vectors. If we want to reduce N-dimension to D -dimension, we can choose the first D norm vectors as the transform matrix.

For the binary dimensionality reduction problem, the first step is to solve the optimal solution ω_1 in (2). Suppose that we have obtained the first $i-1$ optimal discriminant vectors $\omega_1, \cdots, \omega_{i-1}$. The i-th optimal vector ω_i is defined as the optimal vector of the following optimization problem

$$\omega_i = \arg\min_{\omega} \phi(\omega, \varepsilon) = \frac{1}{2}(\omega \cdot \omega) + C\sum_{i=1}^{l} \varepsilon_n, \tag{3}$$

$$\text{s.t.} \begin{cases} \omega_i^T \cdot \omega_j = 0, & i \neq j \\ y_i((\omega \cdot x_i) + b) \geq 1 - \varepsilon_i, & i = 1, \cdots, l \\ \varepsilon_i \geq 0, & i = 1, \cdots, l \end{cases}$$

Denote $U_n = [\omega_n, \cdots, \omega_1]$ which is the last transformation matrix.

Algorithm 1: Two-Class Dimensionality Reduction Based on MMDA

Input:

- Suppose that we have M sample data with N dimension $X \in R^{M \times N}$ with the given labels $Y \in R^M$. D is the final dimension to which we want to reduce. U is a transform matrix which is an empty matrix at first.

Do the following procedures

1. Using (X, Y) , we train a SVM, and get the ω_1 which is orthogonal to the maximum margin direction.
2. the kth iteration $(k = 2, \cdots)$ denote
$$U = [U, \omega_k]$$
$$A = (I - U * U^T)$$
$$X = (A * X)^T$$
3. (Termination step): if the iteration number k reach D, the algorithm stops after the kth iteration. Otherwise, perform the $k+1$th iteration.

Output: We can obtain the transformation matrix U.

Introducing the notation

$$A_{n-1} = (I - U_{n-1}U_{n-1}^T) \tag{4}$$

The dual of the optimization problem in (2) is

$$W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j (A_{n-1}x_i)^T (A_{n-1}x_j) \tag{5}$$

$$\text{s.t.} \begin{cases} y_i(\omega^T x_i + b) \geq 1 - \varepsilon_i, & i = 1, \ldots, l \\ \varepsilon_i \geq 0, & i = 1, \ldots, l \end{cases}$$

From the equation (2) and equation (5), we can see that the iteration procedure becomes very easy. In order to calculate ω, we should solve the optimization problem (5) to get α_i. Comparing the equation (2) with (5), the only difference is x_i in (2) and $(A_{n-1}x_i)$ in (5). So in each iteration, we can substitute x_i with $(A_{n-1}x_i)$, then we use the Matlab SVM toolbox to solve the problem. If we want to reduce N-dimension to D-dimension, we choose the first D column of ω_n to obtain the U. So the transformation matrix is

$$U = [\omega_D, \cdots, \omega_1] \tag{6}$$

And the original data is

$$X = [x_1, \cdots x_l]^T \tag{7}$$

Then

$$Y = U^T X \tag{8}$$

where Y is a matrix with $D \times l$ dimension. So the N-dimension of the original data is reduced to D-dimension.

Algorithm 1 summarizes the solution procedures of solving MMDA problem to obtain the projection vector for two class problems.

Generalized Maximal Margin Discriminant Analysis(GMMDA)

MMDA-based dimensionality reduction method is only for two class problems. It is necessary to extend it for multi-class problems. So we first use “one-versus-one" method for multi-class classification(we call it GMMDA1) and then present another multi-class dimensionality reduction method based on MMDA which can reduce the multi-class dimensionality at once(we call it GMMDA2).

Given the train data set

$$(X_1, L_1), \cdots, (X_c, L_c)$$

Where $X_i,\ \ i = 1, \cdots, c$ is train data set which belongs to i class and c is the number of class. L is the label of the data.

GMMDA1 for multi-class pattern classification problem

The proposed method is only for two class dimensionality reduction problems. For multi-class classification, we can consider the problem as a collection of binary classification problems. Here we use “one-versus-one" method in which $c(c-1)/2$ transformation matrixes are constructed and each one trains data from two different classes after dimensionality reduction.

For each two class X_i, X_j, we can obtain the transformation matrix U_{ij} using our method. Then reduce the dimension of X_i, X_j from N dimension to D dimension, obtaining Y_i, Y_j which are trained by classifiers. Given a test point x_i ,we first reduce its dimension from N to D using U_{ij}, and then we can get the predicted label of x_i after the classification. Here, a majority vote across the classifiers is applied. After $c(c-1)/2$ times votes, the label of x_i is determined to the class with maximum votes.

GMMDA2 for multi-class feature extraction problem

From the above section, we can see that it is boring and complex with “one-versus-one" method. For the same X_i with different X_j, the transformation matrix U_{ij} is different by which we would obtain different Y_i for X_i. So it is difficult to analyze which feature is important for classification. Besides, for c classes, it is necessary to construct $c(c-1)/2$ dimensionality reduction matrix and train $c(c-1)/2$ classifiers which will bring much computation burden. So we proposed a multi-class dimensionality reduction method which can reduce the dimension of multi-class data at once and trained and separated by only one multi-class classifier.

Let X_i and X_j denote the data set of class i and j, respectively. Let $U_{ij} = [\omega_{ij}^1, \cdots, \omega_{ij}^D]$ denote a $N \times D$ matrix whose columns consist of the D optimal projection vectors of MMDA defined on X_i and X_j. Let

$$U = \left[U_{12}, \cdots, U_{1c}, U_{23}, \cdots, U_{(c-1)c}\right]. \tag{9}$$

Then, the matrix U represent the transform matrix of the c-class MMDA method.

$$Y_i = U^T X_i. \tag{10}$$

Speech Emotion Feature Extraction

In the paper, we extract two types of speech features for the emotion recognition, i.e., the traditional prosodic features vs. the spectral features [20]. Prosodic features consist of pitch, intensity and the first four formant frequency profiles as well as their derivatives. The spectral features are comprised of the statistics of Mel-Frequency Cepstral Coefficients (MFCC) and its derivatives.

To extract the emotional speech features, we adopt the Praat software[21] to estimate the fundamental frequency (F0) features, the formant frequencies and the voice intensity profiles as well as MFCC. Then, we compute the statistical features over the entire utterance, including the mean value, the standard deviation, the minimum, maximum and range of

Table 1. The Influence Of Parameter C On The Average Recognition Rate using GMMDA1 (The First Cross-validation)

Cvalue	0.1	0.2	0.5	1.0	2.0
Rec Rate	0.6999	**0.7044**	0.6945	0.6908	0.6866
Cvalue	5.0	10	20	50	100
Rec Rate	0.6952	0.6952	0.6908	0.6866	0.6910

F0 and its first derivative, the voice intensity and its derivative, as well as of the first four formant frequency (F1, F2, F3, F4) and their derivatives. In total, the set of utterance-level prosodic features consist of the following 60 features:

- mean, std, min, max ,range of F0 and F0 derivative.
- mean, std, min, max range of F1, F2, F3, F4 and their derivative.
- mean, std, min, max range of voice intensity and its derivative.

Utterance-level spectral features are statistics of the MFCC computed over the entire utterances. For each utterance, we computed 13 MFCC (including log-energy) using a 25 ms Hamming window at intervals of 10 ms and their derivatives. Then we computed the mean value, standard deviation, minimum, maximum and range over the entire utterances. In that case, the total number of utterance-level spectral features is 130, i.e.,

- mean, std, min, max ,range of MFCC and its derivative.

Then, we concatenate the 60 traditional prosodic features and 130 spectral features into a 190-dimensional feature vector to represent an utterance.

Experiments

In this section, we will conduct extensive experiments on the Berlin database to evaluate the performance of the proposed method. The Berlin dataset[19] is one of the most popular databases used by researchers for emotion recognition. This database contains the emotional utterances recorded by 10 German actors reading one of 10 pre-selected sentences typical of everyday communication. The utterances of the database cover the following seven emotions: anger, boredom, fear, disgust, joy, sadness, neutral emotion, in which the utterances corresponding to boredom and disgust emotions were removed in experiments. In this case, the numbers of speech files for these emotion categories in the Berlin database are: anger (127), fear (69), joy (71), neutral (79) and sadness (62).

Emotion Recognition Experiments Based On GMMDA1

In this experiments, we focuses on the emotion recognition based on the GMMDA1 method. Specifically, we adopt the “one-versus-one" pairwise classification strategy to address the five classes of emotion recognition problems, such that the our proposed two-class MMDA-based method be applicable. Since the number of training samples is less than the dimension of the feature vector, we use PCA to reduce the dimensionality of feature vector from 190 to 80. Then, we apply the GMMDA1, LDA and PCA methods, respectively, to further extract the discriminative emotion features. After obtaining the dimensionality reduction transformation matrix, we use the nearest neighbor classifier(NN), the support vector machine (SVM) classifier [18] and the Gaussian mixture models (GMM) classifier, respectively, for the classification task. By using the majority vote approach for the pairwise classification strategy, we can finally recognize the class label of each test data.

Table 4. Comparison of three dimensionality reduction method by three kind of classifiers

Method Dimension	Our Method +KNN	Our Method +SVM	Our Method +GMM	LDA +KNN	LDA +SVM	LDA +GMM	PCA+ KNN	PCA+ SVM	PCA+ GMM
1	0.8362	0.8193	0.796	0.7767	0.7723	0.7305	0.2741	0.3776	0.294
2	0.8318	0.8193	0.8153	0.7907	0.7907	0.7624	0.4447	0.5357	0.4476
3	0.8223	0.8193	0.8092	0.799	0.807	0.7663	0.5327	0.626	0.5281
4	0.8421	0.8193	0.7974	0.8063	0.807	0.7931	0.6031	0.7227	0.5829
5	0.8462	0.8193	0.8016	0.7962	0.7976	0.7976	0.6766	0.7314	0.6442
6	0.8546	0.8193	0.8021	0.8076	0.8061	0.8061	0.7084	0.7359	0.6726
7	0.8462	0.8193	0.7975	0.8041	0.8053	0.8061	0.7393	0.7192	0.7917
8	0.8462	0.8193	0.8012	0.7999	0.8053	0.8148	0.7386	0.7279	0.793
9	0.8502	0.8193	0.7969	0.7999	0.8012	0.8024	0.7492	0.7277	0.7887
10	0.8467	0.8193	0.7969	0.7999	0.8012	0.8018	0.7621	0.7346	0.7843
15	0.8294	0.8193	0.7938	0.8029	0.8012	0.7981	0.7462	0.7311	0.7843
20	0.8294	0.8193	0.7862	0.8058	0.8041	0.8018	0.742	0.7396	0.7843

Table 2. The Influence Of Parameter C On The Average Recognition Rate using GMMDA1 (The Second Cross-validation)

Cvalue	0.1	0.2	0.5	1.0	2.0
Rec Rate	0.8839	**0.8928**	0.8795	0.8881	0.8589
Cvalue	5.0	10	20	50	100
Rec Rate	0.8881	0.8589	0.8727	0.8557	0.8557

Table 3. The Influence Of Parameter C On The Average Recognition Rate using GMMDA1 (The Third Cross-validation)

Cvalue	0.1	0.2	0.5	1.0	2.0
Rec Rate	0.9306	**0.9350**	0.9261	0.9226	0.9101
Cvalue	5.0	10	20	50	100
Rec Rate	0.9268	0.9101	0.9226	0.9268	0.9268

To evaluate the recognition performance of the various methods, we adopt the 3-fold cross-validation strategy to conduct the experiments. Specifically, we randomly divide the training data set into 3 subsets with almost equal size. Then, we choose one subset as the testing data and the other two as the training data. This procedure is repeated until each subsect has been used once as the test dataset. The overall recognition result is obtained by averaging all the results.

It is notable that the penalty parameter C plays a crucial role for SVM. Since the GMMDA method is based on the SVM approach, we have to find an optimal C for our proposed method. In our experiments, we try various values of C and choose the optimal one by using the NN classifier with three-fold cross-validation strategy. The different C values vs. the classification results are listed in Table 1,Table 2,Table 3, from which we can see that the parameter C will influence the classification accuracy rate. It can be clearly seen from Table 1,Table 2,Table 3 that, when the C value is fixed at 0.2, we obtain the optimal recognition rate at each cross-validation. So in the following experiment, we choose 0.2 as the C value which can obtain the best recognition result.

Moreover, we also compare the recognition performance of the three methods with different choices of the projection vector numbers. The comparison results are listed in the Table 4 and Figure 1, where the x axis denotes the number of the projection vectors while the y axis denotes the recognition accuracy (%). The red line with "*" is the result of GMMDA1 method, the blue line with "o" is the result of LDA method, and the green line with "+" uses PCA method. The top sub-figure denotes the results of using NN classifier. The middle one denotes the results of using SVM classifier, and the bottom one denotes the results of using GMM classifier.

From the Figure 1, we can see that the GMMDA1 method obtain the best average recognition rate. Especially, when the dimension is fixed at 6, using the NN classifier the GMMDA1 method can achieve the recognition accuracy as high as 85.46%, while the best recognition accuracy of LDA and PCA are 80.76% and 76.21%, respectively. For SVM classifier,the highest recognition accuracy of our method is 81.93% when the dimension is 1, while the best result of LDA and PCA are 80.7% and 73.96%, respectively. For GMM classifier, the recognition rate of our method is 81.53% when the dimension is 2, while the best results of using LDA and PCA are is 81.48% and 79.3%, respectively. Moreover, we can also see that the best results of our method are achieved with small number of projection vectors. Specifically, for NN classifier, the highest recognition rate is obtained when the dimension is 6, while for SVM the dimension is 1 and for GMM the dimension is 2. The less the dimension is, the less the computation cost is. So using GMMDA1 method, we can obtain a good recognition rate with low time cost.

Table 5. The Influence of Parameter C On The Average Recognition Rate using GMMDA2 (The First Cross-validation)

Cvalue	0.1	0.2	0.5	1.0	2.0
Rec Rate	0.6774	0.6735	0.6900	0.6833	0.6912
Cvalue	5.0	10	20	50	100
Rec Rate	0.6817	0.6905	**0.7157**	0.7089	0.7150

Table 6. The Influence of Parameter C On The Average Recognition Rate using GMMDA2 (The Second Cross-validation)

Cvalue	0.1	0.2	0.5	1.0	2.0
Rec Rate	0.8274	0.8307	0.8476	0.8479	0.8395
Cvalue	5.0	10	20	50	100
Rec Rate	0.8354	0.8356	**0.8522**	0.8487	0.8487

Multi-class Emotion Recognition Based On GMMDA2

In this section, we used our proposed multi-class dimensionality reduction method(GMMDA2) in multi-class emotion recognition. In order to demonstrate the effectiveness of our method, we compare our method with Multi-class LDA dimensionality reduction using three kinds of classifiers NN,

Table 7. The Influence of Parameter C On The Average Recognition Rate using GMMDA2 (The Third Cross-validation)

Cvalue	0.1	0.2	0.5	1.0	2.0
Rec Rate	0.8829	0.8826	0.8736	0.8689	0.8900
Cvalue	5.0	10	20	50	100
Rec Rate	0.9025	0.9179	0.9258	0.9258	**0.9261**

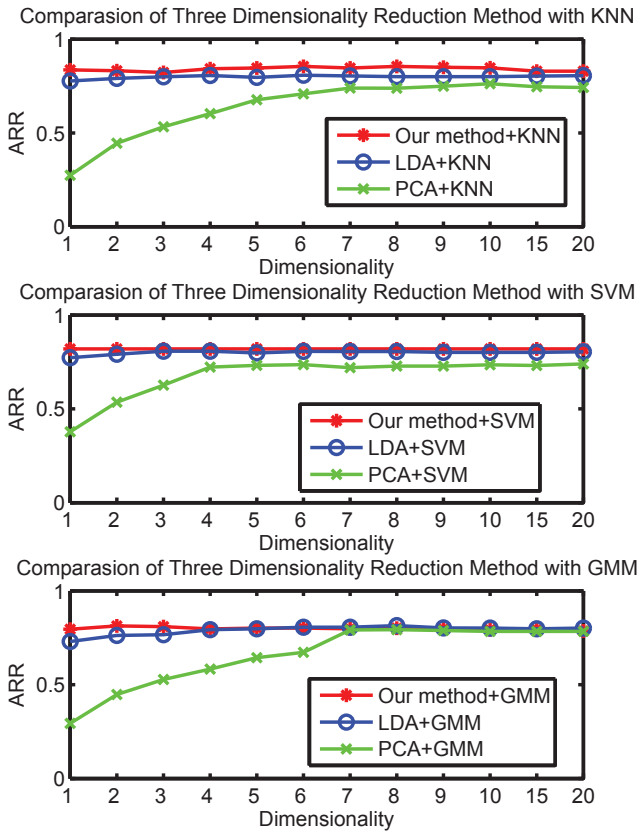

Fig. 1. Comparison of three dimensionality reduction method by different classifiers

SVM and GMM. We also adopt 3-fold cross validation method in this experiment.

Firstly, experiment with different C value is carried out to see its influence on the recognition rate and select the best parameter with 3-fold cross-validation. The result is shown in the Table 5, Table 6, Table 7.

From the Table 5, Table 6, Table 7, we can see that for the first two cross-validation, we can obtain the optimal recognition rate with multi-class dimensionality reduction method when C value is 20. For the third cross-validation, we can get the best recognition rate when the C value is 100. So in the following experiments, for the first two cross-validation, the C value is fixed at 20 and for the third cross-validation ,the C value is fixed at 100,through which we can obtain the optimal recognition result.

Table 8. the Influence of different k value with three kind of classifiers using GMMDA2

Method k value	Our Method+ KNN	Our Method+ SVM	Our Method+ GMM
1	0.8030	<u>0.8362</u>	<u>0.8268</u>
2	0.8062	0.8309	0.8017
3	0.8152	0.8237	0.3412
4	<u>0.8238</u>	0.8216	0.3672
5	0.8158	0.8242	0.3114
6	0.8184	0.8270	0.2186
7	0.8213	0.8241	0.2000
8	0.8154	0.8154	0.2000
9	0.8154	0.8154	0.2000
10	0.8126	0.8126	0.2000

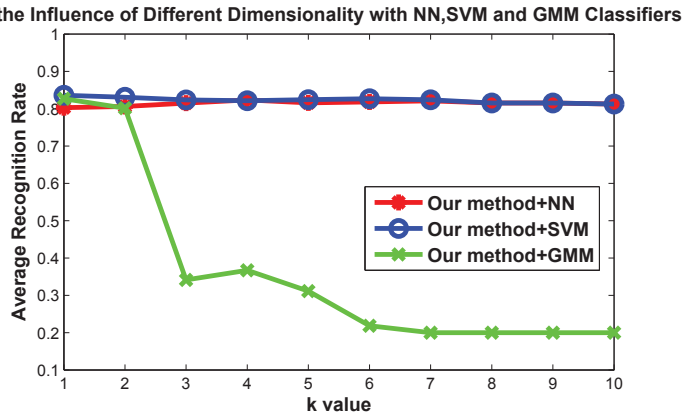

Fig. 2. the Influence of dimensionality with KNN,SVM and GMM Classifiers

Table 9. Comparison of different dimensionality reduction method with KNN, SVM,and GMM

classifier	NN	SVM	GMM
Without dimensionality reduction	0.6995	0.8073	0.2000
GMMDA2	<u>0.8030</u>	<u>0.8362</u>	<u>0.8268</u>
Multiclass LDA	0.7928	0.8022	0.8019

Then we carry out the experiments to see the influence of dimensionality (projection vector) on the recognition rate. As stated above, when $k = 1$, we will get 1 projection vector for each two classes and for c classes classification problem, there are totally $c(c-1)/2$ projection vectors. In our experiment, there are five kind of emotions, so the parameter c is 5. Then, we can obtain 10 projection vectors with $k = 1$. If $k = 2$, we can obtain 20 projection vectors and so on. The recognition rate using NN, SVM and GMM classifiers with different k value are shown in the Table 8 and Figure 2. From the Figure 2, we can see that when $k = 4$, we obtain the best recognition rate 82.38% using NN. When $k = 1$, we can obtain the best recognition rate using SVM and GMM. The result are separately 83.62%and 82.68%. When k is greater than 3, the recognition rate of GMM is dramatically declined. Because the limited train data is not enough to train a good model for GMM. So we choose $k = 1$ to continue our experiment.

In order to demonstrate the effectiveness of our multi-class dimensionality reduction method, we compare it with multi-class LDA dimensionality reduction method using NN, SVM and GMM. Firstly, we carry out the experiment without dimensionality reduction using NN, SVM and GMM, obtaining the recognition rate as benchmark to see the dimensionality reduction effect which are shown in the first row of Table 9.Then GMMDA2 and LDA are used with NN, SVM and GMM classifiers. The final recognition rate are shown in the Table 9. From the Table 9, we can see that the recognition rate with GMMDA2 using NN is 80.3%, which is much higher than that of without dimensionality reduction and higher than that of Multi-class LDA by about 1%. For SVM classifier, our method obtain 83.62% recognition rate which is better than that of without dimensionality reduction 80.73% and that of Multi-class LDA 80.22%. For GMM classifier, also for the sake of dimension, the recognition rate without dimensionality reduction is very bad. The recognition rate of 20% is only for a reference. Our proposed method obtain 82.68% and the multi-class LDA obtain 80.19%. In sum, our proposed method obtain the best performance with different classifiers which prove the effectiveness of our method.

Discussion and Conclusion

We present a multiclass speech emotion recognition method called GMMDA with our proposed dimensionality reduction approach based on the MMDA method. Two recognition schemes are adopted in this paper. One is using "one-versus-one" method for multi-class classification and the other is to combine the projection vectors of each pair of classes to get a transformation matrix which can reduce the dimensionality of multi-class speech emotion projection vector at once. For GMMDA method, the parameter C will greatly influence the recognition rate. We have to choose the C value first. For GMMDA1 method, we compare it with the other two often used classical dimensionality reduction methods LDA and PCA using three different classifiers NN, SVM and GMM. From the experiment results, we can see that our methods obtain the best recognition rate. For GMMDA2 method, we compare it with multi-class LDA method. The speech emotion recognition rate of our method with NN,SVM and GMM classifiers are all better than those of multi-class LDA method. The average performance is improved by about 2%. Moreover, GMMDA method reaches the best performance with low dimensionality which will lesson the computation cost. In sum, GMMDA method can be well applied in speech emotion recognition.

Acknowledgement

This paper is supported by the National Natural Science Funding of China (No: 61231002, No: 61273266) and funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

The authors are grateful for the insightful and constructive suggestions from the anonymous reviewers.

REFERENCES

[1] R. Cowie, et al.,"Emotion recognition in human-computer interaction," IEEE Signal Process Magazine, Vol.18, pp.32-80, 2001.
[2] T. Vogt and E.Andre, "Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition.," presented at the Proc. Multimedia and Expo(ICME05), Amsterdam, Netherlands, 2005.
[3] B. Schuller, et al., "Brute-forcing hierarchical functionals for paralinguistics: a waste of feature space?," presented at the Proc. ICASSP, Las Vegas, NV, 2008.
[4] B. Schuller, et al., "Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles," in Proc. Interspeech 2006, pp. 1818-1821.
[5] B. Schuller, et al., "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from te first challenge," speech communication, 2011.
[6] I. K.Fodor, "A survey of dimension reduction techniques," 2002.
[7] P. Pudil, et al., "Floating search methods in feature selection," Pattern Recognition Letter, vol. 15, pp. 1119-1125, 1994.
[8] C.Bishop. Pattern recognition and machine learning. Springer,2006
[9] I. T. Jolliffe, Principle Component Analysis. Berlin, Germany: Springer, 2002.
[10] K. Fukunaga, Introduction to Statistical Pattern Recogniton: Academic Press, 1990.
[11] A. Kocsor, et al., "Margin Maximizing Discriminant Analysis," ECML, pp. 227-238, 2004.
[12] K. Kovacs, et al., "Maximum Margin Discriminant Analysis based Face Recognition," presented at the Proc. Joint Hungarian-Austrian Conf. Image Process Pattern Recognition, 2005.
[13] I. W.-H. Tsang, et al., "Large-Scale Maximum Margin Discriminant Analysis Using Core Vector Machines " IEEE Transactions On Neural Network, vol. 19, pp. 610-624, 2008.
[14] I. W.Tsang, et al., "Efficient kernel feature extraction for massive data sets," presented at the Proceedings of the 12th ACM SIGKDD international conference on Knowledge Discovery and Data Mining, NY, USA, 2006.
[15] I. W. Tsang, et al., "Diversified SVM Ensembles for Large Data Sets," presented at the Machine Learning: ECML, 2006.
[16] H. Li, et al., "Efficient and Robust Feature Extraction by Maximum Margin Criterion," IEEE Transactions On Neural Networks, vol. 17, pp. 157-165, 2006.
[17] S. Gu, et al., "Discriminant analysis via support vectors," Neurocomputing, vol. 73, pp. 1669-1675, 2010.
[18] V. Vapnik, Statistical Learning Theory. New York: Wiley, 1998.
[19] F.Burkhardt, et al., "A database of german emotional speech," presented at the Interspeech, 2005.
[20] D.Bitouk, R.Verma, A.nenkova, Class-level spectral features for emotion recognition. Speech Communication.(2010), doi:10.1016/j.specom.2010.02.010
[21] P. Boersma. Praat, a system for doing phonetics by computer. Glot International, 5(9/10):341–345, 2001.

***Authors**: Yun Jin, Li Zhao, the Key Laboratory of Underwater Acoustic signal Processing of Ministry of Education, School of Information Science and Engineering, Southeast University, Nanjing 210096, P.R.China. email:jinyun9999@gmail.com.Yun Jin, School of Physics and Electronics Engineering, Jiangsu Normal University, Xuzhou, 221116, P.R.China. Wenming Zheng,Jingjie Yan, the Key Laboratory of Child Development and Learning Science, Ministry of Education, Research Center for Learning Science, Southeast University, Nanjing, Jiangsu 210096, China wenming_zheng@seu.edu.cn*.