第41卷第6期 上海师范大学学报(自然科学版) Vol. 41, No. 6
2012年12月 Journal of Shanghai Normal University(Natural Sciences) Dec., 2012

An audio signal classification algorithm based on an ELM and LDA

MAO Xianyuan, LIN Jun, KANG Qi

(College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China)

Abstract: In this paper, a new classification method is designed based on linear discriminant analysis (LDA) and extreme learning machine (ELM) for audio signals of Multimedia information. We first apply methods such as Fourier transform to extract features from each section of audio signal, which would be proportionally organized to form a high-dimensional vector. And then LDA method is applied to reduce the dimensionality of the feature vector for making it the best feature vector for classification, which can be the training and testing sample of ELM. Finally, use ELM, SVM, and BP classifier are used to do experiments on four kinds of audio signal respectively, and their performances are contrasted. The result shows that, the method promoted still exhibits better classification performance for the classes which are hard to classified, with average accuracy rate of 91%, and its computing speed is more than one thousand times faster than SVM.

Key words: ELM; LDA; feature extraction; SVM

CLC number: TP 37 **Document code:** A **Article ID:** 1000-5137(2012)06-0574-06

1 Introduction

In general, the realization of audio classification algorithm mainly consists of two parts: feature extraction and classification. Features for classification include: short-term energy in the time domain, short-term zero rate, bandwidth of the frequency domain, mass spectrum and Mel Frequency Cepstrum Coefficient (MFCC), etc.

Many classification have been promoted and improved such as: Support Vector Machine (SVM), Gaussian-Mixture Model (GMM), Neural Network (NN), Hidden Markov Model (HMM), etc.

Many results have been achieved. Lin [1] selected MFCC and sound frequency as features, and used wavelet packet and SVM as the classifier, obtained a sound results. Lavner [2] developed decision-tree algorithm to classify the speech signal. Yao [3] applied SVM to classify the feature signals which had been linearly reduced, and achieved average accuracy of 83.2%. Murugappan [4] used DWT and MFCC for the emotional speech feature extraction, and applied LDA for classification, and obtained fairly high accuracy. Liu [5] used the improved BP neural network to classify six kinds of audio signals and get the high accuracy with 90%.

This paper introduces a new classification algorithm based on LDA and ELM, and is divided into three

Received date: 2012-11-30

Biography: MAO Xianyuan (1990 -), male, graduate student, College of Information, Mechanical and Electrical Engineering, Shanghai Normal University; LIN Jun (1956 -), female, professor, College of Information, Mechanical and Electrical Engineering, Shanghai Normal University; KANG Qi (1987 -), female, graduate student, College of Information, Mechanical and Electrical Engineering, Shanghai Normal University.

parts: feature extraction, ELM Classifier, and experiments.

2 Feature extraction

Feature extraction has a decisive influence on classification performance, This paper extracts feature vectors from a section of lasting audio with length of 24, including perception features and Mel cep strum coefficients. Perception features includes frame energy, spectrum, sub-band mass energy, and bandwidth, etc. Each frame of signals is used to extract eight perception features and four MFCC coefficients, and then calculate the mean and standard deviation of these twelve coefficients are calculated to form a 24 dimensional vector, which represents the unique features of audio signals.

To obtain more conducive classification features, LDA is applied to reduce the dimensionality of high dimensional features. The basic idea is to project the high dimensional pattern sample to the optimum identified vector space to extract the classification information and compress the dimensionality of the feature space, and ensure that the the maximum between-class distance and minimum within-class distance of the pattern sample exist in the new subspace, that is, the pattern has best separability in this space. LDA is used to find a projection direction, which makes the between-class separabilities of different samples in new space as large as possible, and the within-class separabilities as small as possible.

Given an N dimensional original feature vector $X_K = (K = 1, 2, \cdots, N)$, and $y_k = (k = 1, 2, \cdots, N)$ the M dimensional feature vectors after dimensional reduction, $M < N$, linear transformation matrix $W \in R^{N \times M}$, then $y_k = W_k^{\mathrm{T}} \cdot x_k$. Choose W which will make the ratio of the between-class scatter and within-class scatter of signals maximum, and finally achieve an optimal classification effect .

The between-class scatter matrix S_B and within-class scatter matrix S_W are shown as follows

$$S_B = \sum_{i}^{C} N_i (\mu_i - \mu)(\mu_i - \mu)^{\mathrm{T}} \quad , \tag{1}$$

$$S_W = \sum_{i=1}^{C} \sum_{X_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^{\mathrm{T}} \quad . \tag{2}$$

Where μ_i is the mean of feature vector X_i, N_i is feature number, C is the category number. If S_W is the non-singular, then the best linear transformation matrix is obtained:

$$W_{opt} = \arg\max_{W} \frac{|W^{\mathrm{T}} S_B W|}{|W^{\mathrm{T}} S_W W|} = [W1, W2, \cdots, W_M] \quad . \tag{3}$$

Where $\{W_i | i = 1, 2, \cdots, M\}$ is the generalized feature vector of S_B and S_W, corresponding to the M maximums of generalized eigenvalue $\{\lambda_i | i = 1, 2, \cdots, M\}$ respectively, where $S_B W_i = \lambda_i S_W W_i$, $i = 1, 2, \cdots, M$. There exist at most C − 1 nonzero generalized eigenvalues in this equation, thus the reduced dimensionality is less than or equal to C − 1. This paper extracts 24 dimensional feature vectors from four kinds of audio samples, then achieves new three dimensional feature vectors, and takes these feature vectors as the training samples and testing samples of ELM.

3 Elm classifier

As is known to all, the training method of traditional BP neural network model adjusts the network weights on the basis on which the the principle on which the backward error propagation makes the sum of squares errors between actual and desired outputs minimum or less than a certain threshold, usually gradient descent iteration method is adopted. The BP algorithm based on gradient descent has the following defects:

(1) Requiring multiple iteration, long training time;

(2) difficult to set the original values of parameter η and W;

(3) possibly to get the local minimum solution, not the global optimum;

(4) easy to cause over-fitting.

According to above defects of the typical SLFN, Huang put forward a new learning algorithm ELM[6].

The main idea is that for any given N samples (x_i, t_i), $x_i = [x_{i1}, x_{i1}, \cdots, x_{in}]^{\mathrm{T}} \in R^n$, $t_i = [t_{i1}, t_{i1}, \cdots, t_{im}]^{\mathrm{T}} \in R^m$, there exists an activation function $g(x)$ which is infinite differentiable in any interval. If the number of neurons in hidden layer K equals to the training sample number N, for any link weights of the input and hidden layer w, and the neurons threshold of the hidden layer b, SLFN can approximate the training samples without errors; When the training sample number N is large, Choose a suitable neuron number of the hidden layer K which is less than N, the training error of SLFN can approach any positive integer ε, that is

$$\begin{cases} \sum_{j=1}^{Q} \|t_j - y_j\| = 0, K = N \\ \sum_{j=1}^{Q} \|t_j - y_j\| < \varepsilon, K < N, \varepsilon > 0 \end{cases} \quad y_i = [y_{1j}, y_{2j}, \cdots, y_{mj}]^{\mathrm{T}}, j = 1, 2, \cdots, Q \quad . \tag{4}$$

Therefore, w and b can be selected randomly before training, without adjustment during the training process. And then the link weights β between hidden and output layer can obtain the least-square solutions with formula (5):

$$\text{Minimize} \begin{cases} \|H\beta - T\|^2 \\ \|\beta\| \end{cases} \quad . \tag{5}$$

Where the output matrix of hidden layer H, and neural network T, and β can be expressed in the following formulas respectively.

$$H = \begin{bmatrix} h(x_1) \\ h(x_2) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} g(\omega_1 x_1 + b_1) & \cdots & g(\omega_K x_1 + b_K) \\ g(\omega_1 x_2 + b_1) & \cdots & g(\omega_K x_2 + b_K) \\ \vdots & \vdots & \vdots \\ g(\omega_1 x_N + b_1) & \cdots & g(\omega_K x_N + b_K) \end{bmatrix}_{N \times K} ,$$

$$T = \begin{bmatrix} t_1^{\mathrm{T}} \\ t_2^{\mathrm{T}} \\ \vdots \\ t_N^{\mathrm{T}} \end{bmatrix}_{N \times m} ,$$

$$\beta = \begin{bmatrix} \beta_1^{\mathrm{T}} \\ \beta_2^{\mathrm{T}} \\ \vdots \\ \beta_N^{\mathrm{T}} \end{bmatrix} .$$

With the application of the least square method, The Moore-Penrose generalized inverse H^+ of the output matrix H of hidden layer can be obtained, and the only solution $\beta = H^+ T$ can be calculated, if $H^{\mathrm{T}}H$ is non-singular, then $H^+ = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}$, when HH^{T} is non-singular, then $H^+ = H^{\mathrm{T}}(H^{\mathrm{T}}H)^{-1}$.

Now, the decision function ELM classifier[7] can be constructed in the form of

$$f_N(x) = \mathrm{sign}\left(\sum_{i=1}^{N} \beta_i h_{ix}\right) = \mathrm{sign}(h(x)\beta) \quad . \tag{6}$$

4 Experiments

This paper first uses Fourier transform method to extractal features from each segmental audio, and form a high dimensional vector in proportion,then applies LDA to reduce the dimensionality of feature vectors,at last applies ELM,SVM,BP classifier for 4 kinds of audio signal classifications,and compares and evaluates the performance of each classifier.

The size of sample matrix is 2000×24,after LDA transform the feature matrix is compressed to 2000×3. 1600 samples are selectad randomly as the training sets,and 400 samples the test sets. Running environment is Win7 of 64 bits,2.4GHz Intel(R) Core(TM) i3,4G internal storage,Matlab2010b.

4.1 LDA testing

Sample Matrix is of 24 dimensional,and 3 dimensional feature vector matrix can be obtained after LDA transform. An example of feature vectors after dimension reduction is shown in table 1. randomly chooses 1600 samples from the matrix as training samples,and 400 samples as testing samples.

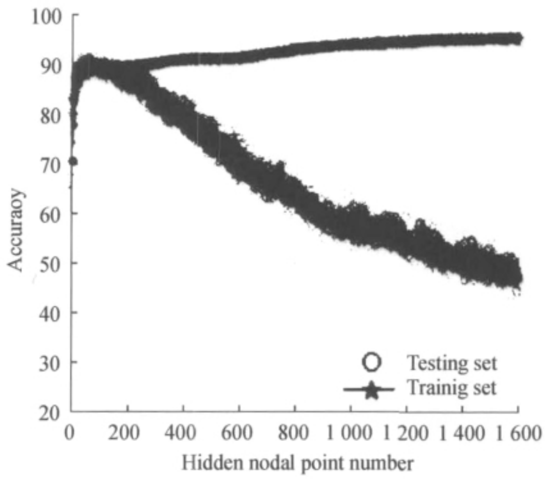
Table 1　The feature vector after LDA dimension reduction

NO.	feature vector			classes	NO.	feature vector			classes
1	9.7108	−1.4538	−0.7393	A	501	3.8828	−3.6374	1.07310	B
2	8.4983	1.1389	0.058511	A	502	2.3265	−2.3995	0.92634	B
3	8.9144	0.15093	0.286	A	503	2.3494	−2.1115	0.83630	B
4	9.4103	0.45812	0.17735	A	504	3.3445	−2.6406	0.08507	B
5	9.7547	1.2978	0.26277	A	505	3.7596	−2.0754	0.15066	B
6	9.4039	1.5186	0.49411	A	506	4.0975	−1.5155	0.08306	B
7	8.3673	0.64822	0.1806	A	507	3.3167	−2.7174	0.11200	B
8	8.3061	0.26613	0.57942	A	508	2.1208	−2.0377	0.43324	B
9	9.6992	0.44673	0.15319	A	509	2.3814	−0.2736	0.50775	B
10	10.068	0.38003	−0.43120	A	510	3.2143	−0.8074	0.31438	B
1001	8.8041	−1.0395	0.091787	C	1501	0.2151	1.3565	0.31782	D
1002	9.2844	−0.53863	0.053854	C	1502	1.2517	1.5180	1.0962	D
1003	10.027	−0.4834	−0.35657	C	1503	2.1165	1.8888	1.1500	D
1004	8.4791	0.17324	−0.10550	C	1504	1.3443	4.4898	0.82859	D
1005	7.5962	−0.77625	−0.11192	C	1505	3.7674	4.2474	0.78527	D
1006	8.1652	−0.33583	−0.10087	C	1506	4.899	3.7376	1.1021	D
1007	8.2708	−0.73155	0.002663	C	1507	3.8032	2.4274	0.99386	D
1008	7.8948	0.06065	−0.40626	C	1508	1.5700	3.0701	0.97436	D
1009	9.2168	0.00330	−0.04918	C	1509	2.6120	2.0908	1.5053	D
1010	7.8612	0.065304	0.46141	C	1510	1.6034	1.4985	0.63901	D

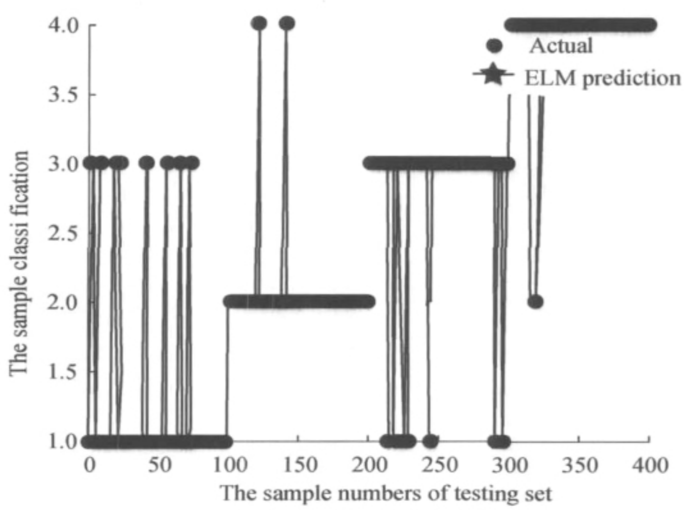
4.2 ELM algorithm testing

ELM algorithm is applied for training and classification testing on feature sample matrix with and without LDA transformed respectively. The results show that,the ELM algorithm with LDA can achieve the best classification results when the nodal point of hidden layer is 81 ,and the accuracy reaches 91% as shown in figure 1. On the contrary,the best classification result appears when the nodal point of hidden layer reaches 238 ,and the accuracy is 90% as shown in figure 2.

During iterations,classification effects calculated with different hidden nodal number tends to be different, since w and b are chosen randomly. Thus,make experiments 10 times ,sort the experiment results,and then take the mean values as the final result. Use BP neural network and SVM for training and testing on the original signal feature vectors ,Compares and analyze the performance of each algorithm.



Figuer 1 The relationships between the nodal point of hidden layer and the accuracy of ELM algorithm under LDA



Figuer 2 The result of the ELM algorithm with LDA

Table 2 Ever algorithm performance comparison results

algorithm	accuracy						time	Hidden layer Node number
	train	test	Class A	Class B	Class C	Class D		
ELM with LDA	0.905	0.910	0.832	0.981	0.915	0.932	0.5148	81
ELM without LDA	0.872	0.89	0.74	0.98	0.82	0.93	0.5928	238
BP	0.865	0.857	0.597	1.00	0.953	0.86	50.972	25
SVM	0.95	0.91	0.77	1.00	0.89	100	846.45	—

As shown in table 2,ELM algorithm has incomparable advantages on computing speed. Although BP algorithm has the same training and testing accuracy as ELM algorithm without LDA,the classification performance for the signal Class A is very poor. Compared with SVM, the ELM algorithm with LDA has a higher accuracy and computing speed, because of the less nodal number of hidden layer when getting best solution, and the speed can reach one thousand times faster.

5 Conclution

In this paper, an audio classification method has been proposed based on LDA and ELM. First LDA is applied to get lower dimensional feature vector which can be the most beneficial feature matrix for classifications. And then ELM, SVM and BP do training and testing experiments. Experiment results show that the promoted algorithm still shows good classification performance even for the classes which are hard to classify, the average accuracy of classification reaches 91%, and its computing speed is more than one thousand times faster than the SVM.

References:

[1] LIN C C, CHEN S H, TRUONG T K, et al. Audio classification and categorization based on wavelets and support vector machine [J]. IEEE Trans on Speech Audio Processing, 2005, 13: 644 – 651.

[2] LAVNER Y, RUINSKIY D. A decision-tree-based algorithm for speech/music classification and segmentation [J]. EURASIP Journal on Audio Speech and Music Processing, 2009(1): 1 – 14.

[3] YAO S Q, HU J L. Music classification based on linear discriminative analysis and support vector machine [J]. Elementary Electroacoustics, 2006, 12: 6 – 11.

[4] MURUGAPPAN M, NURUL B N Q, JERRITTA S. DWT and MFCC based Human Emotional Speech Classification Using LDA [C]. IEEE, International Conference on Biomedical Engineering (ICoBE), Penang, 2012

[5] LIU J W, YU X Q, WAN W. Multiple classification of audio based on improved BP neural network [J]. Journal of Shanghai University (Natural Science), 2012, 18(2): 127 – 132.

[6] HUANG G B, ZHOU H M, DING X J, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Transactions On System, Man, and Cybernetics—Part B: Cybernetics, 2012, 42(2): 513 – 529.

[7] HUANG G B, DING X, ZHOU H. Optimization method based extreme learning machine for classification [J]. Neurocomputing, 2010, 74(1): 155 – 163.

一种基于 ELM 与 LDA 的音频信号分类算法

毛显源，林　军，康　琪

（上海师范大学 信息与机电工程学院，上海 200234）

摘　要： 针对多媒体信息中的音频信号，提出一种基于线性判别分析（LDA）与极限学习机（ELM）的分类方法. 首先，使用傅里叶变换等方法从每一段音频中提取特征，并将它们按比例组成一个高维向量；其次，应用 LDA 对高维向量进行降维，使其成为用于分类的最优特征，作为 ELM 的训练和测试样本；最后，分别采用 ELM，SVM，BP 分类器对 4 种音频信号进行分类，并进行性能对比与分析. 实验表明：提出的算法对于较难分的类也具有较好的分类效果，平均正确率为 90%，同时运算速度比 SVM 快一千多倍.

关键词： ELM；LDA；特征提取；SVM

（责任编辑：包震宇）