2012 International Conference on System Science and Engineering
June 30-July 2, 2012, Dalian, China

The Research of speaker recognition based on GMM and SVM

Huo Chun bao Zhang Cai juan

Department of Electronics and Information Engineering
Liaoning University of technology
Jin Zhou , China
huochunbao@sohu.com

***Abstract*—GMM and SVM models with different advantages and disadvantages were often used in speaker recognition system. The Principle and the characteristics of SVM and GMM were analyzed in paper. The method of combining the advantages of GMM and SVM as the speaker recognition model, mean parameters of GMM as SVM input, which not only can realize an accurate description of the speaker voice characteristics with fewer parameters, but also can achieve the purpose of compressed data for SVM. Experimental results shown the speaker recognition system based on GMM and SVM can effectively improve recognition rate.**

Keywords- GMM; SVM; Speaker recognition

Ⅰ. INTRODUCTION

In recent years, with the development of information processing and artificial intelligence technology, more and more people pay attention to speaker recognition which was widely used in economic, judiciary and military fields.

First, the collected speech signal was pre-processed in speaker recognition. Then, the corresponding characteristic parameters were extracted to model. Last, the identification was made according to some criterion by comparing the speaker voice with ready-made speaker models in the recognition stage. The algorithm of combining GMM and SVM was proposed in paper that solved the problem of the insufficient training data can not reflect the speaker voice characteristics and the system recognition rate decline when the speech data increases in SVM.

Ⅱ. GMM AND SVM

A GMM

The principle of speaker recognition based on GMM was according to the speech parameters were extracted from speakers set to build GMM, model parameters were determined by spatial distribution of the speech feature parameters[1]. It can recognize the speaker by comparing the GMM because of different speaker has different voice model.

GMM was essentially a linear weighted combination of a multi-dimensional probability density function, it was expressed as follows

$$P(X/\lambda)=\sum_{i=1}^{M} w_i f_i(X) \tag{1}$$

Here, w_i was weight, it shown the magnitude of Gaussian distribution, and $\sum_{i=1}^{M} w_i = 1$. $f_i(X)$ was joint Gaussian probability distribution with D-dimensional, expressed as

$$f_i(X)=\frac{1}{2\pi^{D/2}\,|\sum_i|^{1/2}}\exp[-\frac{1}{2}(X-u_i)^t\sum\nolimits_i^{-1}(X-u_i)] \tag{2}$$

Here, u_i was mean, $\sum_i$ was covariance matrix. GMM was expressed by the mean, weights and covariance matrix $\lambda=\{w_i,u_i,\sum_i\}$.

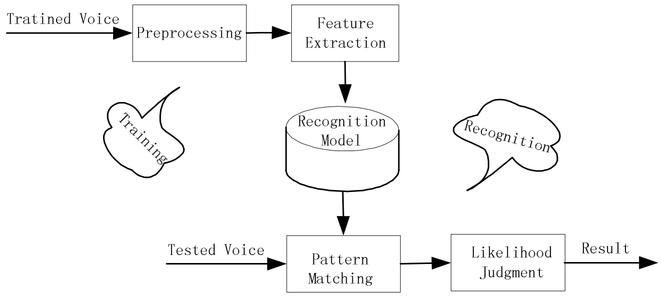

Fig. 1 Gaussian mixture model

There were four parts in the speaker recognition system based on GMM—Speech Signal Preprocessing, feature parameters extraction, speech models training, pattern matching and judgment, as shown in figure 1.

B SVM

SVM was built based on the VC dimension theory and structural risk minimization principle [2] whose idea was: firstly, the input space was transformed into a high dimensional space through the nonlinear transformation, then, the optimal linear classification surface was gained in the new space. The nonlinear transformation was realized by the inner product function. The research for SVM initially about linearly separable problem, as shown in Fig.2, two class can completely be separated by separating line H

Plan project of science and technology of Liaoning province, No 2007217003

978-1-4673-0945-5/12/$31.00 ©2012 IEEE

which was a classification function. The distance between H1 and H2 was called margin optimal separating line H can accurately separate two –class samples and margin was maximum.

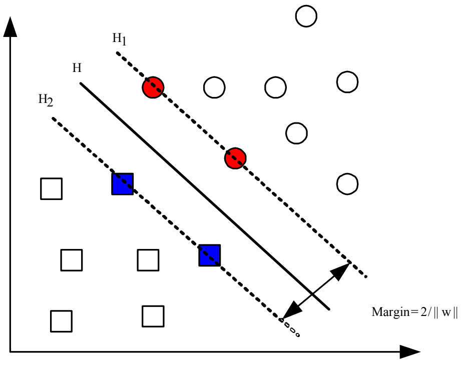

Fig.2 SVM linearly Hyperplane

For nonlinear problem, it can be transformed into a high dimensional space to solve. Optimization objective function is

$$H(w)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}y_iy_j\alpha_i\alpha_jK(x_i\cdot x_j)$$
$$s.t\quad\sum_{i=1}^{n}y_i\alpha_i=0,\alpha_i\geq 0 \qquad (3)$$

Which solution was decision function of judgment SVM, expression as follows,

$$f(x)=\mathrm{sgn}[\sum_{i=1}^{n}\sum_{j}^{n}a_i^*y_iK(x_i,x_j)+b^*] \qquad (4)$$

Where, $K(x_i,x_j)$ was Kernel function which met the Mercer condition, RBF kernel function was used in paper.

The speaker recognition system based on SVM was shown in figure 3. The SVM was built by the method of one-to-all in training stage, when recognizing the speaker, the tested speaker speech was input to the trained models, calculate the decision function value according to (4).

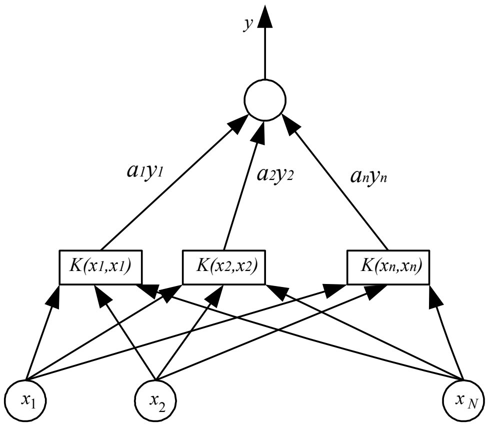

Fig.3 The structure of speaker recognition system based on SVM

III. THE GMM/SVM SPEAKER RECOGNITION SYSTEM

GMM was one of the commonly used models for the speaker recognition, recognition rate was suboptimal when the trained or tested speech data was insufficient, SVM as a learning machine can solve the small sample, nonlinear question, but it would face data aliasing problems if it was used in text-independent speaker recognition with a large amount of data. So, the advantages of GMM and SVM were combined in speaker recognition system. The GMM was obtained by the EM algorithm to train large speaker speech data, the mean parameters of GMM not only shown the distribution of the speaker voice parameters in the feature space ,but also well reflected the speaker voice personalized information. It can realize accurate description for the speaker voice characteristics and gain the goal of the compressed voice data when GMM as SVM front-end input. So, it reduced the size of the training samples because of the number of training samples for SVM was the mixture model M not speaker feature vectors L (M<<L).The speaker recognition system based on GMM and SVM was shown in Fig. 4.

Algorithm steps are as follows:

a) The feature parameters were extracted to built GMM by EM algorithm after training voice was preprocessed, then N speaker mean parameters were got from GMM, that were $u_1,u_2,\cdots,u_N$.

b) SVM was trained by the way of one-to all, that is the distance $d_{nm}^1,d_{nm}^2,\cdots,d_{nm}^N$ was calculated between the m-th mean parameter of n-th speaker and other speaker mean parameter, then training samples with N-dimensional were got.

c) Repeat step 2, all SVM training samples were obtained. Finally, N-SVM model was gained.

d) In test stage, the ultimate test voice samples were got by the method similar with step 1.

e) Test voice samples and SVM were compared, then, the speaker was recognized according to classification criteria.

□. EXPERIMENTAL RESULTS AND ANALYSIS

The experiment was made about the improved speaker recognition system in order to test the effectiveness of the algorithm. Voice library were recorded by researchers in the experiments, there were 23 people in voice samples set, the length of tested voice was 15s, the GMM order was 32, Pre-emphasis coefficient was 0.97, window function was hamming window, short-time energy and short-time zero crossing rate was combined as endpoint detection, feature parameters were 12 order MFCC and 12 order □MFCC, the experimental data for the speaker recognition system were measured in VC + + 6.0 environment.

Table □ was the recognition rate under different training voice duration. From table□, GMM, GMM and SVM system rate would increase with the length increase of training voice, and recognition rate of GMM and SVM system was always higher than the GMM system. However, the recognition rate

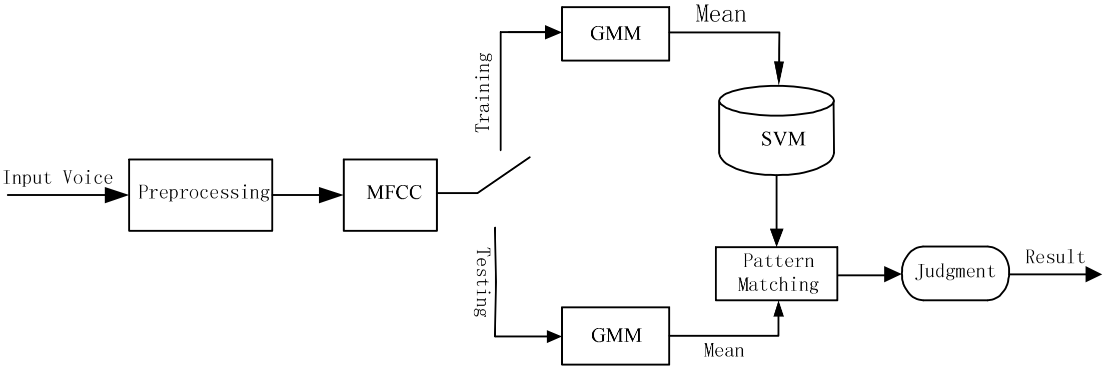

Fig. 4 The speaker recognition diagram based on GMM and SVM

of SVM system would decline with increasing length of training voice after 8s. So, the length of training voice played an important role for the recognition rate in the speaker recognition system. Model training was inadequate and can not contain most of the information for the speaker voice if training voice is too short, otherwise, it would Waste time and space , data aliasing. So, the training voice duration should be reasonable choice in the experiment.

TABLE □. RECOGNITION RATE UNDER DIFFERENT TRAINING DURATION

Training Duration	**Recognition Models(%)**		
	GMM	SVM	GMM+SVM
5	64.1	70.3	73.2
8	79.3	87.3	88.9
12	91.4	90.2	93.9
15	92.3	89.2	95.8
18	93.2	86.8	96.7

Table II was the influence of the different mixing degree for recognition rate in speaker recognition system.

TABLE II RECOGNITION RATE UNDER DIFFERENT MIXING DEGREE

Mixing Degree	**Recognition Models(%)**	
	GMM	GMM/SVM
8	63.2	72.3
16	78.4	89.7
32	91.7	94.6
64	92.3	95.9
128	91.6	96.3

From the table II, the recognition rate of the GMM system and the GMM/SVM system would increase with the mixing degree increase. The GMM/SVM system recognition rate was always higher than the GMM system recognition rate when they had the same mixing degree, both the system recognition rate increased rapidly when mixing degree were 8, 16, 32. The GMM system recognition rate was almost not growth, another still increase when mixing degree was 32.So the mixing degree had important role for recognition rate in the speaker recognition. GMM had few parameters so that the SVM classification surface description was not accurate enough When mixing degree too low, conversely, GMM were trained cost more time and SVM were trained would face the problem of large samples.

□. CONCLUSION

GMM as the speaker recognition model, the recognition rate will decline when training voice data were insufficient, and SVM had the ability to process the small-scale data, so the algorithm was proposed which combined the advantages of GMM and SVM, not only most the features of the speaker voice were covered, but also the defect of the SVM to process large-scale data was avoided. Experimental results show that, the speaker recognition system had better performance than the separate GMM and SVM.

REFERENCES

[1] He Zhiyang, Zhang Linghua. Text-Independent Speaker Identification Based on GMM Statistical Parameters and SVM. Journal of nan jing university of posts and telecommunications. Vol 26, pp78-82, Jun, 2006

[2] Yang Haiyan, Jing Xinxing, Wang Ying, Cao Yu. Application of SVM to text-independent speaker recognition. Technical Acoustics. Vol 27, pp 360-361, Oct, 2008.

[3] Cui Xuan,Sun Hua,Liu Liu. Research of the speaker verification based on the SVM-GMM mixed model. Xi hua university journal. Vol 29, pp 58-61,Jan,2010.

[4] Daniel Garcia-Romero, Julian Fierrez-Aguilar. Using Quality Measures for Multilevel Speaker recognition. Computer Speech and Language, Vol 20, pp 192-2-9, 2006

[5] He Xin, Liu Chong Qing,Li Jiegu. Text-independent Speaker identification based on support vector machines. Computer Engineering. Vol 26, pp 61-63, June , 2000.