The application research of speech feature extraction based on the manifold learning

Penghao Zhang
College of Computer Science & Information
Guizhou University
Guiyang, China
zphlwy@126.com

Li Wang
College of Computer Science & Information
Guizhou University
Guiyang, China
Wangl_@tom.com

***Abstract*—Traditional MFCC phonetic feature will lead a slower learning speed on account of it has high dimension and is large in data quantities. In order to solve this problem, we introduce a manifold learning, putting forward a new extraction method of MFCC-Manifold phonetic feature. We can reduce dimensions by making use of ISOMAP algorithm which bases on the classical MDS (Multidimensional scaling). Introducing geodesic distance to replace the original European distance data will make twenty-four dimensional data, which using the traditional MFCC feature extraction down to two dimensional data. Experiments prove that MFCC - Manifold feature extraction methods has achieved a satisfactory effect in data volume reduction**

Keywords-manifold learning; MFCC-Manifold; geodesic distance; feature extraction

I. INTRODUCTION

In the study of speech recognition, speech signal Feature Extraction technology is one of the important part of the signal Feature Extraction is usually refers to a model of group measured value transformation, to highlight the model representative characteristics of a kind of method, for the subsequent signal processing to provide necessary support, whether can extract the optimal characteristics of speech recognition system operation efficiency and success or failure has had a significant impact。

Manifold Learning (Manifold Learning) by Bregler and Omohundro published in 1995, Visual Speech Recognition for the first time in this paper carry [1] in 2000 on SCIENCE published three articles ISOMAP [2][3] LLE and perception of Manifold hypothesis [4], is considered Manifold Learning research upsurge start mark

Manifold learning as there is the frontier of machine learning, and its theory system is not mature, has been successfully applied in face recognition and obtain good effect at present, manifold learning in speech recognition aspects of the application research is less, and HE put forward the famous idea of Laplacianfaces methods face feature extraction [5], this method into the local geometric structure, from the manifold learning view feature extraction and geometry study, combining become at the beginning of this century algebra feature extraction especially projection learning classical case; M.LI and B.Yuan introducing manifold learning two-dimensional local keep projection (2 DLPP) [6] are also face feature extraction has made a certain effect, The Yale face database, AR face image library, FERET face database recognition rate were 92.11%, 58.67%, 50.50% Manifold learning whether the use of speech recognition, whether it can achieve better results, few studies. This paper put the Manifold learning used in speech recognition of the feature extraction, proposes a new MFCC - Manifold feature extraction method, to the low dimensionality reduction is effect, and the characteristic value of the data quantity have done a lot about minus.

II. SPEECH SIGNAL FEATURE EXTRACTION

Feature extraction is usually in the speech signal endpoint detection process, signal feature extraction in speech recognition of very important position, the main reason for the original speech signal carries a huge amount of data, for hardware system operation burden in addition, the original speech signal contains a large number of random factors such as noise, etc. These factors on the final recognition rate has great influence on extracting the right speech signal characteristic parameters, can make the template training and pattern matching data features more obvious, at the same time compression data quantity, reduce the computation of the system, improve the operation efficiency .In speech recognition, the characteristic parameters in the time-domain are commonly used have amplitude (or energy), zero rate, etc.; Frequency domain feature parameters have linear prediction coefficient (LPC), LPC cepstrum coefficient (LPCC), pattern of parameters (LSP) has been created, the resonant frequency (the first resonance peak F1 and second resonance peak F2, third resonance peak F3), short time spectrum, Mel frequency cepstrum coefficient (MFCC), etc. The MFCC to report of the auditory feature, its performance and robustness is the best of all parameters.

III. MEL FREQUENCY CEPSTRUM COEFFICIENT(MFCC)

The human ear for different frequency speech have different perception ability, and the experimental found in 1000 Hz the following, perception ability and the frequency of a linear relationship between, and in 1000 Hz above, perception, and frequency paired number relationship. In order to simulate the human ear for different frequency speech perception characteristics, people put forward the concept of Mel frequency, its significance for: 1 Mel for 1000 Hz tone perception level of 1/1000.

Mel frequency cepstrum coefficient (MFCC) is based on the concept of Mel frequency brought out, its extraction and calculating process as show in figure 1.

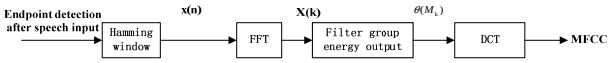


Figure1. MFCC Calculation process schematic diagram

MFCC parameter has the following advantages:

Speech information mostly concentrated in the low frequency part and high frequency part vulnerable to environmental noise interference. MFCC parameters will linear frequency converted to Mel frequency, emphasizing the voice low frequency information, so as to highlight to identification information, shielding the noise interference. MFCC parameters without any hypothesis, in a variety of situations all can use.

A. *Dynamic difference parameter*

Speech signal dynamic parameter can well reflect the signal of the time-varying characteristics; usually we use difference parameter to represent the dynamic characteristics. Difference parameter is the current speech frame, before and after the frame speech signal corresponding to the characteristic parameters of the linear combination. Assume that the gain characteristics cepstrum parameter is p dimension, the first order difference calculation formula is as follows:

$$d_{cep}(i,t)=\alpha\times\sum_{i=1}^{K}k\times[C_{cep}(i,t-k)-C_{cep}(i,t+k)] \qquad (1)$$

d_{cep} is dynamic characteristics, C_{cep} is cepstrum, K is for difference of the scope of the frame, α is for conversion characteristics of the factor.

Similarly, the second order difference features:

$$d_{cep}^{2}(i,t)=\alpha\times\sum_{i=1}^{K}k\times[d_{cep}(i,t-k)-d_{cep}(i,t+k)]=\alpha\times\sum_{k=-K}^{K}k\times d_{cep}(i,t-k) \qquad (2)$$

In the experiment of endpoint detection usually after each frame of the speech signal extraction 12 order Mel cepstrum parameter, and then calculation MFCC parameters of the first order difference parameters, so that each frame speech signal 24 characteristic parameter values. This 24 characteristic parameter value representation a frame the characteristics of the speech signal. Because the first-order difference parameter of the extraction process, the k value is 2, also is the difference parameter by its two frames before the speech signal and in it after the two frame speech signal corresponding MFCC to find the parameters, so that the beginning of the speech and the last two frame signal can't find out the corresponding difference parameter. The paper will be effective speech two frame before and after two frame characteristic parameters to remove, the rest of the characteristic vector integral voice effective section of the feature space.

IV. BASED MFCC-MANIFOLD FEATURE EXTRACTION

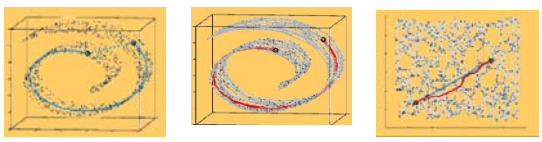
Manifold learning as through the acquisition data set intrinsic geometric for data analysis .May be defined as: by finite sample point set to calculate in high dimensional space into Euclidean space of d dimension manifold M model problem .Briefly speaking, manifold learning goal from high dimension observation data to recover low dimensional manifold structure, to find a high dimensional space the low dimensional manifold. Realize the data dimension reduction or about data visualization and obtain appropriate mapping relation.

A. *In the study of manifold ISOMAP algorithm*

Classic ISOMAP algorithm has three steps:

1) Establish neighborhood relationship diagram G (V，E): For each x_i （i=1，…，N)，according to certain criteria to determine K phase velocity point and calculate its K neighbor. To point x_i for vertex. Euclidean distance $d_o\left(x_i,x_j\right)$ construction diagram G (V，E).
2) According to the graph set good path to calculate the shortest path between any two points use them as myopia geodesic distance $d_l\left(x_i,x_j\right)$ to construct the geodesic distance matrix $D\left[d_l\left(x_i,x_j\right)\right]_{N_xN}$.
3) For geodesic distance matrix $D\left[d_l\left(x_i,x_j\right)\right]_{N_xN}$ use traditional MDS algorithm, looking for a low dimensional data $Y=\left(y_1,y_2\ldots.y_n\right)$.

ISOMAP algorithm the main thought is the original calculation of European distance between two points, with the side ground distance to replace, so more can describe the data between structures. Figure 2 (a) give the Euclidian distance between two points, (b) give the geodesic distance between two points, (c) for data in 2 d tile of Euclidean distance and side ground distance contrast.



(a) (b) (c)

Figure2 Euclidean distance and geodesic distance

B. *Mfcc-Manifold feature extraction*

Based on the Manifold learning of ISOPMAP algorithm, this paper proposed a new feature extraction MFCC - Manifold method. Processes as shown in figure 3

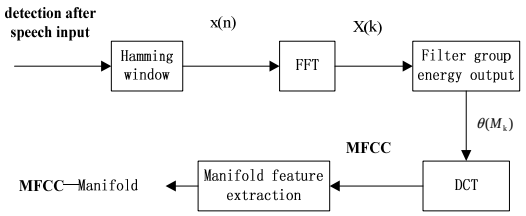


Figure3 MFCC-Manifold Feature extraction flow chart

1) After endpoint detection of speech signal period add window, get the speech signal short time domain signal $x(n)$, then use FFT will these time domain short time signal x (n) into frequency domain signal x (k).
2) Then get X(k) through a Mel filter bank $H_m(k)$, then output the logarithmic energy, computation formula is as follows type.

$$\theta(M_k) = \ln[\sum_{k=1}^{K} |X(k)|^2] H_m(k) \quad k = 1, 2, \ldots, K \qquad (3)$$

k is the first k a filter, K is total filter.

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \le k \le f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)} & f(m) < k \le f(m+1) \\ 0 & k > f(m+1) \end{cases}, \quad 0 \le m \le K \qquad (4)$$

3) On the above (4) output energy do discrete cosine transform get beauty's frequency cepstrum $c_{mel}(n)$:

$$c_{mel}(n) = \sum_{k=1}^{K} \theta(M_k) \cos(n(k - 0.5)\frac{\pi}{K}) \quad n = 1, 2, \ldots, p \qquad (5)$$

4) The discrete cosine change later got MFCC value with ISOMAP algorithm for dimension reduction. Finally get MFCC - Manifold characteristic value. In practical operation, in MFCC get after data, a data for N dimension, see it as a N a point, to calculate the distance between the two both get distance matrix, namely a diagonal element 0 and N order symmetric matrix, and then is the distance matrix as ISOMAP algorithm processing input data.

V. THE SIMULATION EXPERIMENT AND RESULT ANALYSIS

The experiment USES speech signal is in the laboratory environment quiet recording of "0" to "9" ten digital male voice section, every segment speech sampling frequency for 8 KHZ, 8 bit quantification, PCM coding. Each voice with 20 samples training, and then in another 30 a speech training samples. For convenience of shows that the characteristic value data quantity about loss of direct effect to MFCC feature extraction of characteristic value data quantity as the standard, the type definition:

Data ratio = (MFCC - Manifold characteristic value data quantity) / (Mel characteristic value data quantity)

TABLE I. IN THE ONLY "0" AND "1" TWO FOR RECOGNITION OF SPEECH RECOGNITION RATE AND DATA QUANTITY STATISTICS.

	MFCC	MFCC-Manifold
Characteristic value of the dimension	24	2
ISOMAP algorithm of K value	无	15
Data rate	1	8%
The final recognition rate	96%	90%

TABLE II. 0" TO "9" 10 TO IDENTIFY SPEECH FEATURE VALUE IN DIFFERENT DIMENSION.

	MFCC	MFCC-Manifold			
Characteristic value of the dimension	24	10	7	4	2
ISOMAP algorithm of K value	无	15	15	15	15
Data rate	1	41%	29%	17%	8%
The final recognition rate	88%	44%	45%	43%	45%

TABLE III. 0" TO "9" 10 TO IDENTIFY SPEECH FEATURE VALUE IN DIFFERENT K VALUE

	MFCC	MFCC-Manifold			
ISOMAP algorithm of K value	无	5	10	15	20
Characteristic value of the dimension	24	2	2	2	2
Data rate	1	8%	8%	8%	8%
The final recognition rate	88%	44%	46%	45%	44%

In front of the table 1 only "0" and "1" two for identification of the speech, and table 2 and table 3 for identification of the speech is "0" to "9" ten digital speech. Can be seen from table 1, and the traditional MFCC compared with its final recognition rate has little difference, while the data about reduce actually achieve 90% above. The experimental results show that in the final recognition rate under the condition of close, MFCC - Manifold of feature extraction, the data quantity are of great about minus.

Table 2 and table 3, can see MFCC - Manifold the final recognition rate is only 45% or so, and the traditional 88% compared with, is not very ideal, but the data about reducing the highest is above 90%, and in the operation efficiency, remote identification, involving data transmission, etc, it has certain practical significance. In addition, in table 1 and table 2, the data show that the introducing manifold learning feature extraction, for reducing to several dimensional characteristic value peacekeeping ISOMAP algorithm selection of value of K in the final recognition rate the impact is not big. To add, manifold learning to data dependence is very big, the space structure of data is decisive influence for manifold learning, therefore, the test can only prove the experiment data, to reduce to a few dimensional characteristic values and ISOMAP algorithm selection of value of K in the final recognition rate the impact is little.

VI. CONCLUSION

Based on the speech recognition puts forward a MFCC - Manifold new feature extraction algorithm, and the results showed that the only "0" and "1" two for identification of the data, the recognition rate and the traditional Mel feature extraction are only 0.06, the data quantity about the reduction achieved satisfactory effect. But because the manifold learning is still in the theoretical level, especially in speech recognition application study is less, in a "0" to "9" ten to identify data, low recognition rate. In addition, in this paper to reduce the dimension of selection, ISOMAP algorithm near point K value selection to do some research, the experiment shows that the recognition rate of the impact is little, therefore, to improve the recognition rate, it is necessary to start from manifold learning theory itself to do further research, in addition, if you can choose such as Bayes decision and neural networks method instead of HMM to do training and recognition in order to improve recognition rate, whether can use the manifold learning direct substitution MFCC feature extraction to do will be this paper the next step of work. No matter how, Manifold learning as the latest machine learning, has set up a file in the face recognition obtained the certain achievement in China, in the future will be in the speech recognition as an important research direction to promote the breakthrough of speech recognition.

REFERENCES

[1] Bregler C.,Omohundro S.M.，“Nonlinear Manifold Learning for Visual Speech Recognition”,Int.Conf.Computer Vision，1995.

[2] Tenenbaum J.B.,Silva V.，Langford J.C.,“A Global Geometric Framework for Nonlinear.Dimensionality Reductiaon”，Science，vol.290，no.5500，2000，2319~2323.

[3] Roweis S.T.，Saul L.K.，Nonlinear Dimensionality Reduction by Locally Linear Embedding Science，vol.290，no.5500，Dec.2000，2323~2326

[4] Seung H. S. Lee D. D.，“The manifold ways of perception”，Science，Vol.290，NO.5500.2000，2268~2269.

[5] X. He and P. Niyogi，Locality Preserving Projections. Proc. 16th Conf. Neural Information Processing Systems，2003.

[6]]M.Li，B.Yuan，2D-LDA: a statistical linear discriminant analysis for image matrix.Pattern Recognition Letters,2005,26 (5):527-532.