음성 특징의 효율성

황규웅, 이영직
한국전자통신연구소
음성언어 연구실

EFFICIENCY OF SPEECH FEATURES

Kyuwoong Hwang Youngjik Lee
Spoken Language Processing Section
Electronics and Telecommunications Research Institute(ETRI)
E-mail: hkw@zenith.etri.re.kr

Abstract

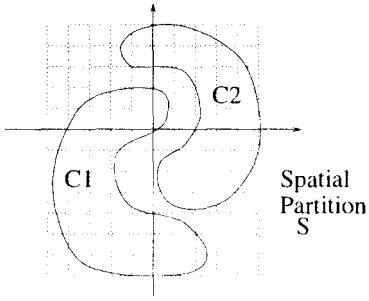
This paper compared waveform, cepstrum, and spline wavelet features with nonlinear discriminant analysis[1]. This measure shows efficiency of speech parametrization better than old linear separability criteria and can be used to measure the efficiency of each layer of certain system. Spline wavelet transform has larger gap among classes and cepstrum is clustered better than the spline wavelet feature. Both features do not have good property for classification and we will compare Gabor wavelet transform, Mel cepstrum, delta cepstrum, etc.

1 INTRODUCTION

Good features are important in pattern classification problem including speech recognition. To compare feature efficiency, people usually used linear discriminant analysis or recognition rate. But, the former is inadequate to speech because speech is not linear and the latter is recognition system or algorithm dependent method and not true efficiency comparison. We used nonlinear discriminant analysis method suggested in [1] to compare speech features. We compared waveform, wavelet coefficients, and cepstrum.

2 NONLINEAR DISCRIMINANT ANALYSIS

For complex data such as speech, even the classes are nonlinearly separable, the result looks inseparable with linear discriminant analysis. nonlinear discriminant analysis suggested in [1] is based on information theory and applicable to finite-support, multi-class distributions and suitable to analyze real data such as speech and handwritten characters.



Class partition $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_M\}$
Inter-class separability(Overlapping factor)

$$s \triangleq \frac{I(\mathcal{S};\mathcal{C})}{H(\mathcal{C})}$$

$$I(\mathcal{S};\mathcal{C}) = H(\mathcal{S}) - H(\mathcal{S}|\mathcal{C})$$

Intra-class variation(Average of scatteredness in each class)

$$v \triangleq \frac{H(\mathcal{S}|\mathcal{C})}{H_{max}}$$

If $s < 1$, the classes are overlapped. If $s = 1$ and $v < 1$, the classes are separated and clustered. If $s = 1$ and $v = 1$, each sample falls into distinct bin of the spatial partition. Let K be the smallest number of bins K that achieves $s = 1$. If K is small, the gap among the classes is large. If K is large, the gap is small.

3 EXPERIMENTS

We used part of TIMIT speech database for experiments. That is consisted of 380 utterances of 38 persons. It has 110,000 samples and the number of phoneme classes is 61. We used 16kHz, 16bit AD with preemphasis. Feature extraction conditions for compared features are as follows. We examine eigenvalue distribution and used nonlinear discriminant analysis. With eigenvectors, we mapped sample space to diagonalized one and counted the number of samples for each bin for each class. With that sample distribution, we calculated entropy and inter-calss separability and intra-class variation. To reduce calculation time, sorting is used. By parallel computing of 10 workstations, one day's work reduced to 2 hours' work. It is important to select a appropriate number of bins and the size of a bin. We assumed the volume of a feature space as the multiplication of each dimension's length. The length of a dimension is

defined as the difference of maximum value and minimum value of that dimension of samples. With this volume and determined number of bins, we can get the length of a bin. If a dimension's length is smaller than the length of a bin, that dimension is ignored. We used spline wavelet and get 55 dimensional features for each frame. Spline wavelet transform has larger gap among classes and cepstrum is clustered better than the spline wavelet feature. This is due to that wavelet feature has a lot more dimensions than cepstrum. That means wavelet feature samples are scattered sparsely in feature space and so they shows good inter-calss separability but shows worse intra-class variation. Waveform is much inferior to others in both criteria. This is the reason why people do not pattern classification directly on speech waveform.

4 CONCLUSION

We compared cepstrum and spline wavelet features with nonlinear discriminant analysis using information measures. Spline wavelet transform has larger gap among classes and cepstrum is clustered better than the spline wavelet feature. Both features do not have good property for classification and we will compare Gabor wavelet transform, Mel cepstrum, delta cepstrum, etc.

5 FURTHER WORK

We will compare other type of wavelet coefficients and use this nonlinear discriminant analysis as a probing tool to get wavelet feature. We will do speech recognition experiments for these compared features and examine the relation between feature efficiency with this method and the recognition rate.

6 REFERENCES

[1] Y. Lee and H. K. Song, 'Analysis on the efficiency of pattern recognition layers using information measures', *International Joint Conference on Neural Networks*, pp. 2129-2132, Oct. 1993.

feature	window size(msec)	overlapping(msec)	dimensionality
speech signal	10	no	160
cepstrum + energy	20	10	13
spline wavelet transform	20	10	55