A New Feature Normalization Scheme Based on Eigenspace for Noisy Speech Recognition

Yoonjae Lee and Hanseok Ko

Dept. of Electronics and Computer Engineering, Korea University, Seoul, Korea
yjlee@ispl.korea.ac.kr, hsko@korea.ac.kr

Abstract. We propose a new feature normalization scheme based on eigenspace, for achieving robust speech recognition. In particular, we employ the Mean and Variance Normalization (MVN) in eigenspace using unique and in–dependent eigenspaces to cepstra, delta and delta-delta cepstra respectively. We also normalize training data in eigenspace and get the model from the normalized training data. In addition, a feature space rotation procedure is introduced to reduce the mismatch of training and test data distribution in noisy condition. As a result, we obtain a substantial recognition improvement over the basic eigenspace normalization.

1 Proposed Scheme

We separated the feature vector into three classes as cepstra, delta and delta-delta cepstra because each class has its own definition and characteristics. Then we implemented a separated-eigenspace normalization (SEN) scheme.

When cepstral features are distorted by noisy conditions, their distribution can be moved as well as rotated by some amount from their original distribution.[2] When we rotate only the dominant eigenvector that has the largest variance or eigenvalue, the first eigenvectors of training and test features become the same and the mismatch between the training and test data distribution can be reduced. Only the first eigenvector rotation procedure is presented here simply as follows. First, we need to obtain the eigenvalue and eigenvector of full training corpus. $\tilde{v}$ denotes the first dominant eigenvector of the training distribution and v denotes the first dominant eigenvector of one test utterance. Then the rotation angle α, between the two eigenvectors, is computed from their dot product, $\alpha = \arccos(\tilde{v} \cdot v)$ and $R = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix}$ where R denotes a rotation matrix. Since the two eigenvectors are not orthogonal, the Gram-Schmidt is applied to $\tilde{v}$ in order to obtain the orthonormal basis vector $\hat{v}$ lying in the same plane of rotation, $\hat{v} = \dfrac{v - (\tilde{v} \cdot v) \cdot \tilde{v}}{\| v - (\tilde{v} \cdot v) \cdot \tilde{v} \|}$. Then we project the test features onto the plane spanned by $\tilde{v}$ and $\hat{v}$. The projection

A. Apostolico and M. Melucci (Eds.): SPIRE 2004, LNCS 3246, pp. 76–78, 2004.

matrix consists of $\tilde{v}$ and $\hat{v}$, thus $J = (\tilde{v}, \hat{v})$. Finally, a correction matrix $I - JJ^T$ with the identity matrix I has to be applied in order to restore the dimensions lost in the projection procedure. Then the full rotation matrix Q is derived as: $Q = JRJ^T + I - JJ^T$. Finally, the rotated feature is obtained by: $\hat{X}_t^c = QX_t^c$.

2 Experiments and Results

Recognition Task: The feature normalization method has been tested with the Aurora2.0 database that contains English connected digits recorded in clean environments. Three sets of sentences under several conditions (e.g. SetA: subway, car noise, SetB: restaurant, street and train station noise, SetC: subway and street noise) were prepared by contaminating them with SNRs ranging from -5dB to 20dB and clean condition. A total of 1001 sentences are included in each noise condition.

Experiments Procedure and Results: We followed the Aurora2.0 evaluation procedure for performance verification along with identical conditions suggested in the Aurora2.0 procedure. Note that we use a c0 coefficient instead of log-energy to induce improved performance, because eigenspace is defined consistently when some of elements have large variance. First we examine the baseline performance (clean condition training). We then apply MVN [3] and the eigenspace MVN to only the test data and to both training and test data together. Next, we experimented on separated-eigenspace normalization (SEN). The feature space rotation with SEN was examined also. The experiment notations of Tables are as follows: 1) MVN : mean and variance normalization in cepstral domain, 2) EIG : mean and variance in eigenspace.[1] (eigenspace normalization), 3) SEN : separated-eigenspace normalization, 4) SEN_Ro_20 : separated-eigenspace normalization +feature space rotation. The first eigenvector of the test is obtained from training noisy set's 20dB data of each.

From Table 1, we can see that SEN with feature rotation and training data normalization is more effective than basic eigenspace normalization.

We initially expected the best performance when each dominant eigenvector obtained from each SNR was applied to the corresponding SNR test set. However, it turns out that such method does not guarantee the improvement. At low SNR, the performance becomes slightly degraded. We achieved the best performance when applying an eigenvector of 20dB set to all SNR data of same test set.

Table 1. Average word accuracy for the proposed scheme of all data set in Aurora2.0(%) (_T denotes the normalization of training data)

	Baseline	MVN	EIG	EIG_T	SEN	SEN_T	SEN_Ro_20_T
SetA	59.58	77.90	79.81	80.43	80.27	80.51	81.08
SetB	57.18	79.49	81.21	82.87	81.77	82.49	-
SetC	66.81	77.90	78.96	79.23	79.32	79.10	-

At lower SNR, the data distribution in cepstral domain becomes more compressed. Consequently, their discriminative shapes (e.g. large variance) is diminished as the SNR becomes lower. That's the reason why 20dB statistics yielded the best performance. From a 20dB noisy training database, we estimated the characteristics of corresponding noise and compensated for the feature reliably. Through the proposed methods, we obtained average word accuracy up to 81.08% on the setA of Aurora2.0.

Acknowledgement

Research supported by No. 10011362 MCIE of Korea.

References

1. K. Yao, E. Visser, O. Kwon, and T. Lee, :A Speech Processing Front-End with Eigenspace Normalization for Robust Speech Recognition in Noisy Automobile Environments, Eurospeech2003 (2003) 9-12
2. S. Molau, D. Keysers and H. Ney, : Matching training and test data distributions for robust speech recognition, Speech Communication, Vol.41, No4 (2003) 579-601
3. P.Jain and H.Hermansky, : Improved Mean and Variance Normalization for Robust Speech Recognition, Proc.of ICASSP (2001)