다양한 변별분석을 통한 한국어 연결숫자 인식 성능향상에 관한 연구

송화전(부산대), 김형순(부산대)

<차 례>

<Abstract>

Performance Improvement of Korean Connected Digit Recognition Using Various Discriminant Analyses

Hwa Jeon Song, Hyung Soon Kim

In Korean, each digit is monosyllable and some pairs are known to have high confusability, causing performance degradation of connected digit recognition systems. To improve the performance, in this paper, we employ various discriminant analyses (DA) including Linear DA (LDA), Weighted Pairwise Scatter LDA (WPS-LDA), Heteroscedastic Discriminant Analysis (HDA), and Maximum Likelihood Linear Transformation (MLLT). We also examine several combinations of various DA for additional performance improvement. Experimental results show that applying any DA mentioned above improves the string accuracy, but the amount of improvement of each DA method varies according to the model complexity or number of mixtures per state. Especially, more than 20% of string error reduction is achieved by applying MLLT after WPS-LDA, compared with the baseline system, when class level of DA is defined as a tied state and 1 mixture per state is used.

* 주제어: 변별분석(discriminant analysis), 연결숫자인식(connected digit recognition)

1. 서 론

변별적 분석 (Discriminant Analysis, DA) 방법은 패턴 인식에서 클래스 사이의 변별력을 향상시키기 위해 널리 사용되는 방법이다. 음성인식에서도 음소 또는 그 이하 단위(예를 들면 HMM의 개별 상태)로 정의된 클래스 간의 변별력을 증가시키기 위해, 또한 클래스 간의 변별력에 대한 정보손실을 최소화시키면서 차원을 감소시키기 위해 DA를 사용한다.

여러 가지 변별적 방법 중에서 가장 대표적인 것이 Linear DA (LDA)[1]로서, 소용량 어휘로부터 대용량 어휘까지 다양한 음성인식기의 성능 향상을 위해 적용되어 왔다[2]. 그러나, Fisher criterion을 사용하는 LDA에서는 실제 환경에 잘 부합되지 않은 몇몇 가정에 의해 성능향상에 제약을 가져옴이 알려졌으며, 이 문제를 해결하기 위해 Weighted Pairwise Scatter LDA (WPS-LDA)[3], Heteroscedastic Discriminant Analysis (HDA)[4][5], Mutual Information Discriminant Analysis (MIDA)[6], Minimum Classification Error (MCE)에 기반한 선형변환 방법[7]과 minimum Bayes' error에 기반한 선형변환 방법[8] 등이 제안되었다.

그 중에서도 Li 등은 LDA에서 모든 클래스가 동일한 혼동가능성을 가진다는 가정의 단점을 제거하기 위해 between class scatter 행렬을 구할 때 클래스 쌍 사이의 거리에 따라 다른 가중치를 부여하는 WPS-LDA 방법을 제안하였고[3], Kumar 등은 모든 클래스는 동일한 공분산 행렬을 가진다는 LDA의 가정에 대해 within class scatter 행렬을 구할 때 각각의 클래스마다 다른 가중치를 부여함으로써 LDA를 보다 일반화시킨 HDA를 도입하였다[4].

그 외에도 연속확률분포 HMM에서 훈련 데이터 부족 및 계산량 감축을 고려하여 모델의 공분산 행렬(full covariance matrix) 대신 대각 행렬(diagonal covariance matrix)을 사용함으로 인한 분포 특성의 왜곡에 다른 성능저하 문제를 해결하기 위한 방법으로, 공분산 행렬 및 대각행렬을 사용한 경우 사이의 확률값의 차이가 최소가 되는 변환행렬을 구하는 Maximum Likelihood Linear Transformation (MLLT) [6] 등의 방법들이 개발되었다.

본 논문에서는 다양한 변별적 분석방법들을 한국어 연결숫자 인식성능을 향상시키기 위한 수단으로 사용하였다. 한국어 숫자음은 모두 단음절이며, 상당수의 혼동가능성이 높은 숫자쌍들이 존재하여 연결숫자인식의 성능을 저하시킨다. 변별적 분석방법은 이와 같이 혼동가능성이 높은 클래스 사이의 변별력을 증가시키는 데에 유용하다. 본 논문에서는 LDA를 비롯하여 HDA, WPS-LDA, 그리고 MLLT 방법에 따른 성능을 각각 평가하였으며, 또한 이들의 조합에 따른 성능향상을 도모하였다.

본 논문의 구성은 다음과 같다. 2장에서는 다양한 DA에 대해 간략히 살펴본 후, 3장에서 baseline 시스템과 다양한 DA를 적용한 경우의 성능을 비교하였다. 4장에서는 여러 가지 DA 방법들의 조합에 따른 인식성능을 검토하였으며, 마지

막으로 5장에서 결론을 맺는다.

2. 다양한 변별적 분석법

2.1. LDA[1]

Fisher의 criterion을 사용한 LDA는 다음과 같이 기술되어 진다. 먼저 N개 샘플 벡터 집합 $\{x_i\}_{1\le i\le N}$ 을 고려하자. 여기서, $x_i \in \overrightarrow{r^n}$을 n 차원 특징 벡터이고 각각은 서로 독립이다. 그리고, 각각의 벡터는 L개의 클래스중 하나에 속해 있다고 가정한다. LDA의 목적은 식 (1)과 같이 클래스 사이의 변별력에 대한 정보 손실을 최소화하는 선형 변환 행렬 (A)을 구하는 것이다.

(1) $$y = A^T x \quad A: \overrightarrow{r^n} \rightarrow \overrightarrow{r^p}$$

여기서 p≤n이다.

A를 구하기 위해서 먼저 within-class scatter matrix(S_w)와 between-class scatter matrix(S_b)를 구해야 한다. Within-class scatter matrix는 식 (2)와 같이 나타낼 수 있다.

(2) $$S_w = \sum_{i=1}^{L} N_i \Sigma_i$$

여기서 L은 클래스 수, N_i는 i번째 클래스에 속한 샘플 수이다. 또한, Σ_i는 i번째 클래스의 공분산 행렬이다. Between-class scatter matrix는 식 (3)과 같이 나타낸다.

(3) $$S_b = \sum_{i=1}^{L} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

여기서 μ_i는 i번째 클래스의 샘플 평균이고 μ는 전체 샘플의 평균이다.

LDA에서 변환 행렬을 구하기 위해 여러 가지 criterion이 사용되어 질 수 있으며, 본 논문에서는 식 (4)와 같은 형태를 사용하였다.

(4) $$J = tr(S_w^{-1} S_b)$$

여기서 tr(M)은 행렬 M의 trace를 뜻한다. J를 최대로 하기 위해서는 식 (5)와 같이 $S_w^{-1}S_b$를 고유치와 고유벡터로 분해하면 된다.

(5) $$S_w^{-1} S_b \Phi = \Phi \Lambda$$

여기서 Φ와 Λ는 $S_w^{-1}S_b$의 고유벡터와 고유치이다. 그리고, 식 (1)의 A는 Φ의 열벡터 중 고유치가 큰 순서로 p개를 선택하여 구성하면 된다.

2.2. Weighted Pairwise Scatter LDA (WPS-LDA)[3]

클래스간의 거리는 클래스간의 변별력에 대한 정보를 제공한다. 즉 클래스간의 거리가 가까울수록 변별하기가 어려워진다. 그러나, LDA에서는 각각의 클래스는 다른 모든 클래스에 대해 동일한 혼동가능성을 가진다고 가정한다. 이는 클래스들 사이의 거리에 대한 정보를 무시하는 것이다.

Li 등은 클래스 평균간의 거리가 아주 가까울 때 발생되는 문제에 대해 구체적인 예를 들어 설명하였고, 이를 해결하기 위해 between-class scatter matrix를 구할 때 pairwise 클래스간의 거리에 따라서 가중치를 부가하는 WPS-LDA 방법을 제안하였다[3]. WPS-LDA에서는 식 (3)이 다음과 같이 변한다.

(6) $$S_{b,WPS} = \frac{1}{2N} \sum_{i=1}^{L} \sum_{j=1}^{L} N_i N_j w_{ij} (\mu_i - \mu_j)(\mu_i - \mu_j)^T$$

여기서, N은 전체 샘플 수이고, w_{ij}는 클래스 쌍 (i, j)사이의 가중치이다.

Li 등은 클래스 평균간의 유클리드 거리 및 Kullback-Leibler 거리에 기반하여 클래스간 거리에 역비례하도록 가중치를 부여하였다. 그 결과, WPS-LDA에서 클래스쌍 사이의 거리가 가까울수록 많은 가중치를 부여하므로 기존의 LDA보다 상대적으로 거리가 가까운 클래스들 사이의 변별력을 향상시키게 된다. 실험결과에 따르면, 유클리드 거리에 기반한 방식이 가장 우수한 인식성능을 얻었으며, 본 논문에서도 이를 사용하였다.

2.3. HDA[4][5]

LDA에서는 모든 클래스의 공분산 행렬(full covariance matrix)이 동일하다고

가정한다. 하지만 이러한 가정은 실제 환경과 잘 부합되지 않는다. 이와 같은 LDA의 단점을 보완하는 한가지 방법으로 Kumar 등이 HDA[4]를 제안하였으며 Saon 등이 변형된 형태로 적용하였다[5]. 본 논문에서는 Saon의 방법을 사용하였다. HDA 방법은 변환행렬을 구하는데 각각의 클래스가 어느 정도 기여하는지를 고려한다. 이를 고려한 HDA에서 사용하는 criterion은 식 (7)과 같다.

$$H(\Phi) = \sum_{j=1}^{J} -N_j \log|\Phi^T \Sigma_j \Phi| + N \log|\Phi^T S_b \Phi| \tag{7}$$

HDA는 식 (7)을 최대로 하는 Φ를 구한다. 여기서 클래스는 가우시안 분포를 따른다고 가정한다.

그러나, 식 (7)에서 HDA의 변환행렬 Φ의 해를 직접 구할 수 있는 방법이 없으므로 gradient descent 방법으로 변환행렬을 구한다. 본 논문에서는 gradient descent 방법으로 GNU scientific library[11]의 quasi-Newton conjugate gradient routine을 사용하였다. 그리고, gradient descent 방법을 사용하기 위해 변환행렬 초기화는 LDA를 사용하여 구한 것을 이용한다.

2.2절의 WPS-LDA 방법은 S_b에 대한 LDA의 단점을 보완하는 방법이고, HDA는 S_w에 대한 LDA의 단점을 보완하는 방법이라고 정리할 수 있다.

2.4. MLLT[9]

연속확률분포 HMM을 이용한 인식기에서 각각의 mixture의 공분산 행렬(full covariance matrix)을 대각 행렬(diagonal matrix) 형태로 단순화시켜 사용하는데, 이는 공분산 행렬 전체를 잘 표현하기 위해서는 많은 양의 데이터가 필요하게 되며, 계산량 측면에서도 공분산 행렬 전체를 사용하는 것이 불리하기 때문이다. 그러나, 실제로는 공분산 항들은 무시하지 못하므로 단순화로 인한 인식성능 저하가 발생하게 된다.

MLLT는 이를 해결하기 위한 한 가지 방법이며, 클래스 분포를 가우시안으로 가정하고, 클래스들이 공분산 행렬을 가지는 분포와 분산으로 이루어진 대각 행렬 형태를 가지는 분포사이의 maximum likelihood (ML)의 차이가 최소가 되는 변환행렬을 구하는 방법이다[18]. Criterion은 식 (8)과 같이 주어지며, 이를 최대로 하는 Φ를 구한다.

$$M(\Phi) = \sum_{j=1}^{J} -\frac{N_j}{2} \log|diag(\Phi^T \Sigma_j \Phi)| + N \log|\Phi| \tag{8}$$

MLLT에서도 HDA와 마찬가지로 Φ를 직접 구할 수 없으므로 quasi-Newton conjugate gradient routine을 사용하였다. 식 (8)에서 보듯이 MLLT도 HDA처럼 S_w에 대한 LDA의 단점을 보완하는 방법 중의 하나이다.

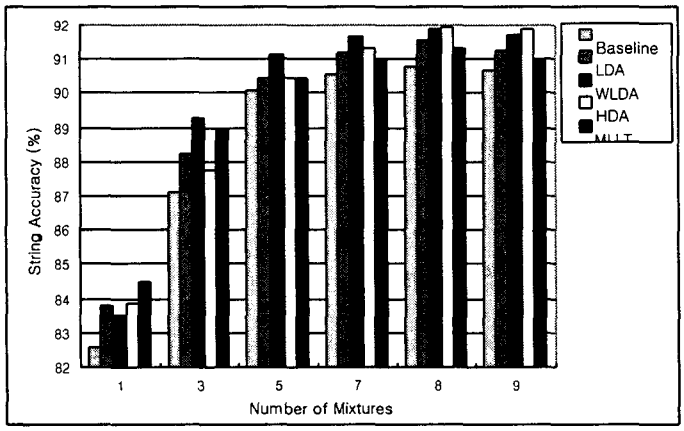
3. Baseline 시스템 및 다양한 DA 적용결과

본 논문에서 실험에 사용된 음성데이터는 원광대학교에서 구축한 전화음성 인식엔진 평가용 연속음성 DB[10]의 일부로서 8kHz로 샘플링 되었으며, 252명의 남성 화자를 50set으로 나누어 864종의 총 8000여 개 4연 숫자조합을 발성한 것이다. 각각의 화자는 집 또는 사무실에서 유/무선 전화기를 이용하여 4연 숫자를 발성하였다. 총 8000여 개의 데이터 중에서 70%를 모델훈련에, 30%를 인식실험에 사용하였다.

음성 특징 벡터는 20ms Hamming 창을 10ms씩 이동시키면서 12차 MFCC와 log 에너지, 그리고 각각의 delta 및 delta-delta를 구한 후 log 에너지는 제외시켜 총 38차의 벡터를 사용하였다. 그리고 전화망의 채널왜곡을 보상하기 위해 Cepstrum Mean Subtraction (CMS)을 적용하였다.

Triphone 기반의 연속 확률 분포 HMM을 사용하였고 모델 당 상태 수는 5개이며 mixture수를 변화시키면서 각각의 방법에 대한 성능을 평가하였다. 그리고, 모든 훈련 시 발생되지 않은 triphone을 보상하기 위해 Tree-Based Clustering (TBC) 방법을 사용하여 HMM 상태들을 tying시켰다.

2절에서 기술된 다양한 DA 방법을 적용하기 위해 각각의 방법의 criterion을 사용하여 변환행렬을 구한 후 식 (1)에 의해 음성특징벡터를 변환한 후 모델을 훈련하여 인식실험을 수행하였으며, DA 적용전인 화자독립 모델을 사용한 baseline 시스템과 성능을 비교하여 그림 1에 나타내었다.



<그림 1> Baseline과 여러 가지 DA 방법들의 성능 비교

본 논문에서는 DA을 위해 446개의 클래스를 사용하였으며, 이는 baseline에서 사용한 tied state 개수이다. 그리고, 식 (2)에서 Σ_i를 구할 때 음성특징벡터에서 MFCC와 에너지와 delta 및 delta-delta 사이의 공분산을 구하는 것은 실질적인 유용성이 없으므로 block-diagonal 행렬을 사용하였다. 마지막으로 식 (5)에서 고유벡터 및 고유치를 구할 때 먼저 whitening 변환[1]을 적용하였다. 그리고, HDA의 변환행렬을 구하기 위한 식 (7)에서 사용한 초기 변환행렬은 LDA의 변환행렬을 사용하였고, 식 (8)의 MLLT의 초기 변환행렬은 Principal Component Analysis (PCA)[1]의 변환행렬을 사용하였다.

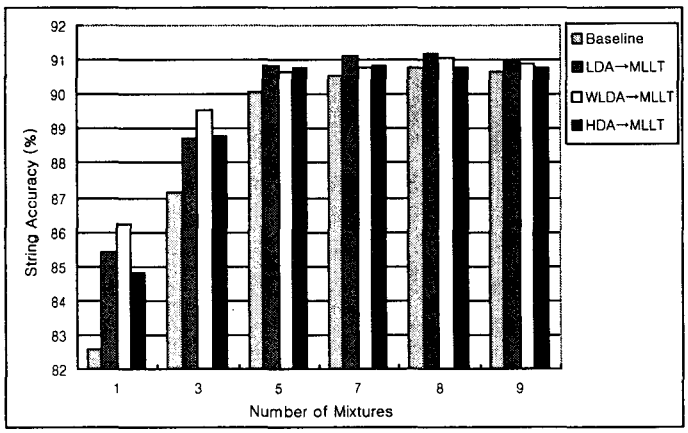
그림 1의 결과에서 보는 바와 같이, mixture가 1개인 경우에 MLLT가 좋은 성능을 나타내며 mixture가 증가할수록 WPS-LDA나 HDA의 성능이 좋음을 알 수 있다. 이는 DA에서 사용한 클래스가 mixture 1개를 가지는 tied state이고, 또한 각각의 클래스를 공분산 행렬 중에서 대각성분만 사용하는 HMM으로 표현해야 하므로 클래스 분포 특성을 HMM과 동일한 형태로 만드는 MLLT가 LDA나 HDA보다 인식 성능 향상에 효과적일 것이다. 그러나, HMM의 mixture가 증가함에 따라 대각행렬형태의 공분산을 가지는 여러 개의 가우시안 mixture들이 DA에 의해 변별력이 향상된 클래스 내부의 음향공간을 잘 표현하게 되므로 각 클래스의 분포 특성을 대각화 시키는 것에 중점을 두는 MLLT 방법보다는 각 클래스간 변별력을 증가에 중점을 두는 WPS-LDA나 HDA가 보다 좋은 성능을 나타내는 것으로 해석된다. 만약 훈련데이터가 충분한 경우에 클래스를 mixture 수준으로 정의한다면 MLLT의 성능도 증가할 것으로 사료된다. 그러나, 최적의 클래스를 정의하는 일 자체가 매우 어려운 일이기 때문에[2], 특정 클래스에 대한 성능을 미리 예측하기는 곤란하다.

4. 다양한 조합을 통한 인식성능 향상

3절에서 mixture가 1개인 경우 음성특징벡터의 분포가 HMM의 상태와 동일한 분포 형태를 가지도록 하는 MLLT가 인식성능 향상에 효과적임을 보여주었다. 이에 따라, 본 논문에서는 먼저 LDA, WPS-LDA와 HDA 등을 수행하여 각 클래스 사이의 변별력을 향상시킨 후 MLLT를 적용하여 음성특징벡터가 대각공분산행렬을 가지는 HMM의 분포특성을 가지게 하여 인식성능을 향상시키고자 하였다. 그림 2에 결과가 나타나 있다. 그림 2에서 예를 들어 LDA→MLLT는 LDA를 수행한 후 MLLT를 적용하는 과정을 의미한다.

Mixture가 1개인 경우 WPS-LDA→MLLT 방법이 가장 좋은 성능을 나타내었다. 단일 mixture를 가지는 상태는 DA 과정에서 사용한 클래스 정의와 동일한 경우이며, WPS-LDA에 의해 LDA에서의 S_b의 단점을, 그리고 MLLT에 의해 S_w의 단점을 개선함으로 성능이 향상되었음을 의미한다. Mixture가 1개인 경우

baseline과 비교하였을 때 숫자열 오인식률이 20.6% 감소하였다. HDA의 경우는 MLLT와 유사한 특성을 지니므로 이들의 조합에 의한 성능 향상 폭이 다른 조합에 비해 크지 않음을 알 수 있다. 그림 2의 경우에서도 3장에서 설명한 바와 같이 mixture수가 적은 경우에는 MLLT에 의한 영향이 크고, mixture가 증가하는 경우에서 WPS-LDA나 HDA에 의한 영향이 커짐을 알 수 있다.



<그림 2> Baseline과 여러 가지 DA의 조합에 의한 성능 비교

5. 결 론

본 논문에서는 한국어 연결숫자 인식시스템의 성능을 향상시키기 위해 다양한 DA를 적용하였다. 이 중 LDA를 사용함으로써 사용하지 않은 경우에 비해서 숫자열 오인식률이 8% 감소하였다. 또한 본 논문에서는 선형적인 방법으로 분리가 불가능한 클래스들의 변별력을 증가시키기 위해 between-class scatter matrix를 구할 때 클래스 사이에 혼동가능성 정도를 나타내는 weighting factor를 적용하였고, 클래스마다 분포 특성을 반영하는 공분산 행렬에 대해 가중치를 달리 부가하는 HDA, MLLT 등을 적용하였다. 또한 DA 방법들의 다양한 조합 중에서 WPS-LDA와 MLLT를 연속해서 사용하여 mixture가 1개인 경우 baseline보다 숫자열 오인식률이 20.6% 감소하였다. 이는 WPS-LDA와 MLLT가 LDA가 가지는 두 가지 단점을 동시에 개선함으로 향상된 결과이다.

앞으로 WPS-LDA 및 MLLT의 변환이 동시에 이루어질 수 있도록 두 가지 방법을 hybrid시키는 방법에 대한 연구가 필요하며, 클래스간의 변별력을 보다 증가시켜 인식성능을 높이기 위해서는 전처리 단계에서의 DA 적용뿐만 아니라 HMM을 구성 시에도 변별적 훈련 방법을 함께 도입해야 할 것으로 판단된다.

참 고 문 헌

[1] K. Fukunaga (1990), *Introduction to Statistical Pattern Recognition*, Second Ed., Academic Press.

[2] Haeb-Umbach, R. & H.Ney (1992), Linear discriminant analysis for improved large vocabulary continuous speech recognition, in *Proc. ICASSP I*, pp.13~16.

[3] Li, Y., Y. Gao and H. Erdogan (2000), Weighted pairwise scatter to improve linear discriminant analysis, in *Proc. ICSLP 4*, pp.608~611.

[4] Kumar, N. & A. G. Andreou (1998), Heteroscedastic discriminant analysis and reduced rank HMMs for improved speech recognition, *Speech Communications 26*, pp.283~297.

[5] Saon, G., M. Padmanabhan, R. Gopinath and S. Chen (2000), Maximum likelihood discriminant feature spaces, in *Proc. ICASSP*, pp.1129~1132.

[6] Duchateau, J., K. Demuynck, D. V. Compernolle and P. Wambacq (2001), Class definition in discriminant feature analysis, in *Proc. EUROSPEECH*, pp.1621~1624.

[7] Saon, G. & M. Padmanabhan (2000), Minimum Bayes error feature selection, in *Proc. ICSLP III*, pp.75~78.

[8] Thomae, M., G. Ruske and T. Pfau (2000), A new approach to discriminative feature extraction using model transformation, in *Proc. ICASSP*, pp.1615~1618.

[9] R. A. Gopinath (1998), Maximum likelihood modeling with Gaussian distributions for classification, *in Proc. ICASSP II*, pp.661~664.

[10] 원광대학교 음성언어과학공동연구소 음성언어자원지원센터(2001).

[11] GSL Team (2001), *GNU Scientific Library Ver.1.0.*

접수일자: 2002년 10월 27일
게재결정: 2002년 12월 12일

▶ 송화전(Hwa Jeon Song)
주소: 609-735 부산시 금정구 장전동 산30번지 부산대학교 공과대학 전자공학과
소속: 부산대학교 전자공학과 음성통신연구실
전화: 051) 510-1704
Fax: 051) 515-5190
E-mail: hwajeon@pusan.ac.kr

▶ 김형순(Hyung Soon Kim)
주소: 609-735 부산시 금정구 장전동 산30번지 부산대학교 공과대학 전자공학과
소속: 부산대학교 전자공학과 음성통신연구실
전화: 051) 510-2452
Fax: 051) 515-5190
E-mail: kimhs@pusan.ac.kr