

# LDA 를 이용한 한국어 연결숫자 인식기 성능향상에 관한 연구

송화전, 김형순  
부산대학교 전자공학과

## Study on Performance Improvement of Korean Connected Digit Recognition using LDA

Hwa Jeon Song, Hyung Soon Kim  
Dept. of Electronics Eng., Pusan National University  
E-mail : {hwajeon,kimhs}@hyowon.cc.pusan.ac.kr

### 요 약

본 논문에서는 class간의 변별력을 증가시키기 위한 유용한 방법인 LDA(Linear Discriminant Analysis)를 사용하여 한국어 숫자음간의 변별력을 높여 연결숫자 인식 성능을 높이고자 하였다. 한국어 숫자음은 모두 단음절이며 또한 혼동가능성이 높은 숫자쌍이 존재하여 이것이 전체 인식률을 저하시킨다. LDA를 사용한 경우 숫자열 오인식률이 8% 감소하였다. 그리고, 음성특징 벡터의 차수를 감소시키도 LDA 사용전 보다 약간의 인식을 증가를 보였다. 그러나, 선형적인 방법으로 분리가 불가능한 class들의 분포가 존재할 때는 LDA를 사용하여도 변별력 향상은 기대하기 어렵다. 이와 같은 분포의 class사이의 변별력을 증가시키기 위해 between-scatter covariance matrix를 구할 때 class 사이에 혼동가능성 정도를 나타내는 weighting factor를 적용하였으며, 그 결과 숫자열 오인식률이 LDA 사용전보다 9.7% 감소하였다.

### 1. 서 론

LDA(Linear Discriminant Analysis)는 패턴 인식 또는 분류 영역에서 class사이의 변별력을 향상시키기 위해 널리 사용되는 방법 중의 하나이다. 음성인식에서도 음소 또는 정의된 class 간의 변별력을 증가시키기 위해, 또

한 class간의 변별력에 대한 정보손실을 최소화 하면서 차원을 감소시키기 위해 LDA가 사용된다[2]-[5]. 보통 LDA는 음성인식기의 front-end에 사용되며, 소용량 어휘로부터 대용량 어휘까지 인식기의 성능 향상을 위해 다양하게 적용되어 왔다. 또한, 기존의 Fisher criterion을 사용한 경우의 LDA[1]에서의 문제점들의 해결하기 위해 WPS-LDA(Weighted Pairwise Scatter LDA)[2], Nonlinear PCA기반 LDA[3], HDA(Heteroscedastic Discriminant Analysis)[4], MIDA (Mutual Information Discriminant Analysis) [5] 등의 방법들이 개발되었다.

본 논문에서는 LDA(Linear Discriminant Analysis)를 사용하여 한국어 숫자음간의 변별력을 높여 연결숫자 인식 성능을 높이고자 하였다. 한국어 숫자음의 경우는 모두 단음절이며 또한 혼동가능성이 높은 숫자쌍이 존재하여 이것이 전체 인식 성능을 저하시킨다. 이와 같이 LDA는 혼동가능성 있는 class간의 변별력을 증가시키는 데에 유용한 방법이다.

본 논문의 구성은 다음과 같다. 2절에서는 기존의 LDA에 대해 간략히 살펴본 후, 3절에서 baseline 시스템과의 성능을 비교하였다. 4절에서는 LDA의 문제점 및 이를 해결하기 위한 방법을 기술하고 기존의 방법과 실험결과를 비교한 후, 마지막으로 5절에서 결론을 맺는다.

## 2. LDA [1]

Fisher criterion을 사용한 LDA는 다음과 같이 기술되어 있다.  $X \in R^n$ 을  $n$  차원 feature 벡터라 하자. LDA의 목적은 다음 식과 같이 변별력에 대한 정보 손실을 최소화하는 변환 행렬( $A$ )을 구하는 것이다.

$$Y = A^T X, A: R^n \rightarrow R^p \quad (1)$$

여기서  $p < n$ 이다. 이를 구하기 위해서 먼저 within-class, between-class 그리고 mixture scatter matrix를 구해야 한다. Within-class scatter matrix는 다음과 같이 나타낼 수 있다.

$$S_w = \sum_{i=1}^L P_i E\{(X - M_i)(X - M_i)^T | \omega_i\} = \sum_{i=1}^L P_i \Sigma_i \quad (2)$$

여기서  $L$ 는 class 수,  $\omega_i$ 는 class  $i$ 를 뜻하며,  $P_i$ 는 class의 사전확률이다. 또한  $M_i$ 는  $i$ 번째 class의 sample 평균이고,  $\Sigma_i$ 는  $i$ 번째 class의 covariance matrix이다.

Between-class scatter matrix는 다음 식과 같다.

$$S_b = \sum_{i=1}^L P_i (M_i - M_0)(M_i - M_0)^T \quad (3)$$

여기서  $M_0$ 는 전체 sample의 평균이다. 마지막으로, mixture scatter matrix는 다음과 같이 표현된다.

$$S_m = E\{(X - M_0)(X - M_0)^T\} \quad (4)$$

LDA에서 변환 행렬을 구하기 위해 여러 가지 criterion이 사용되어 질 수 있으며, 본 논문에서는 다음과 같은 형태를 사용하였다.

$$J = \text{tr}(S_w^{-1} S_b) \quad (5)$$

구하고자 하는  $A$ 는  $J$ 를 최대화 하는 변환이며 이는

$$S_w^{-1} S_b \Phi = \Phi \Lambda \quad (6)$$

로 주어진다. 여기서  $\Phi$ 와  $\Lambda$ 는  $S_w^{-1} S_b$ 의 고유벡터와 고유치이다.  $\Lambda$ 는  $\Phi$ 의 열벡터 중 고유치가 큰 순서로  $p$ 개를 선택하여 구성하면 된다.

## 3. Baseline 시스템 및 LDA 적용결과

Baseline 시스템으로는 이전의 논문[6]에서 구성된 것을 사용하였다. 기본 모델은 triphone을 사용하였다. 튜닝작업을 통해 baseline 시스템의 인식률을 향상시켰다. DB도 동일한 것을 사용하였다. 음성 특징 파라미터도 기존의 것과 동일하게 12차 MFCC, log 에너지, 각각의

delta 및 delta-delta를 구한 후 log 에너지는 제외시켜 총 38차의 파라미터를 사용하였다. 그리고 전화망의 채널 왜곡을 보상하기 위해 CMS를 적용하였다.

2절에서 기술된 LDA 방법을 사용하여 음성특징 파라미터를 식 (1)에 의해 변환한 후 모델을 훈련하여 인식 실험을 하였으며, LDA 적용전인 baseline과 성능을 비교하였다. 그 결과가 표 1과 2에 나타나 있다.

표 1. Baseline 시스템 인식률(%)

Mix- ture 수	상태수					
	3		4		5	
	숫자열	개별숫자	숫자열	개별숫자	숫자열	개별숫자
1	78.14	93.60	81.45	94.53	82.64	94.84
3	84.47	95.53	86.15	96.06	87.70	96.51
5	87.30	96.50	88.38	96.72	89.89	97.20
7	88.61	96.83	89.45	97.07	89.81	97.18
8	88.50	96.86	89.49	97.09	90.49	97.40
9	88.65	96.94	89.73	97.14	90.57	97.43

표 2. LDA 적용 후 인식률(%)

Mix- ture 수	상태수					
	3		4		5	
	숫자열	개별숫자	숫자열	개별숫자	숫자열	개별숫자
1	81.97	94.75	83.48	95.16	83.32	95.13
3	86.46	96.18	87.94	96.50	88.02	96.62
5	88.34	96.78	89.33	96.97	90.29	97.28
7	89.21	97.01	90.13	97.30	90.96	97.49
8	89.25	97.03	90.13	97.33	91.00	97.52
9	89.29	97.05	90.53	91.32	91.32	97.61

본 논문에서는 LDA를 위해 51개의 class를 사용하였다. 이는 숫자음을 표현하기 위한 17개의 monophone 당 state 3개를 사용한 경우이다. 물론 class를 정의하는 방법에는 이 이외에도 여러 가지가 있을 수 있으며, 본 논문에서는 실험을 통해 이와 같이 정하였다. 그리고, 식 (2)에서  $\Sigma_i$ 를 구할 때 음성특징 벡터에서 MFCC와

에너지와 delta 및 delta-delta 사이의 covariance를 구하는 것은 의미가 없으므로 block-diagonal matrix을 사용하였다. 마지막으로 식 (6)에서 고유벡터 및 고유치를 구할 때 먼저 whitening transformation[1]을 적용하였다.

또한 본 논문에서는 차원 감소시 성능의 변화를 살펴 보기 위해 38차 차원을 25차 및 30차로 감소시켜 인식 시험을 실시하였다. 그 결과가 표 3과 4에 나타나 있다.

표 3. 차원 감소 시 인식률(%)-(38차 → 25차)

Mix- ture 수	상태수					
	3		4		5	
	숫자열	개별숫자	숫자열	개별숫자	숫자열	개별숫자
1	79.74	94.10	81.17	94.51	81.61	94.66
3	85.43	95.85	85.71	95.95	87.06	96.37
5	87.74	96.62	89.17	96.98	89.13	96.90
7	88.54	96.87	89.89	97.25	90.21	97.26
8	88.65	96.88	90.05	97.31	90.49	97.39
9	88.81	96.93	89.97	97.33	90.61	97.43

표 4. 차원 감소 시 인식률(%)-(38차 → 30차)

Mix- ture 수	상태수					
	3		4		5	
	숫자열	개별숫자	숫자열	개별숫자	숫자열	개별숫자
1	80.61	94.36	82.13	94.84	82.01	94.79
3	85.27	95.84	87.26	96.35	87.34	96.46
5	87.90	96.62	88.85	96.88	89.97	97.21
7	88.02	96.67	89.69	97.23	90.41	97.34
8	88.26	96.73	89.89	97.24	90.64	97.42
9	88.65	96.88	89.69	97.18	90.84	97.53

표에서 보는 바와 같이 차원 감소를 적용하지 않을 경우에 비해서는 약간의 인식률 감소가 나타났지만, baseline line 시스템 보다는 여전히 인식결과가 높은 것을 알 수 있다. 결과적으로 LDA에 의한 차원 감소를 통해 성능저하 없이 인식소요 시간을 상당히 감소시킬 수 있다.

#### 4. Weighted LDA[2][3]

LDA가 해결하지 못하는 여러 가지 class의 분포가 있다. 이를 해결하기 위한 여러 가지 방법들이 연구되어 졌으며[2]-[5], 그 중 Li 등은 class 평균간의 거리가 아주 가까울 때 발생하는 문제에 대해 예를 들어 설명하고 있다[2]. 한국어 숫자음의 경우도 거리가 가까운 음소들이 존재하여 이상과 같은 문제가 발생할 수 있다.

본 논문에서도 between-class scatter matrix를 구할 때 상대적으로 거리가 가까운 class간에 많은 weighting을 부가하여 인식시스템의 성능향상을 도모하였다. 이를 위하여 본 논문에는 weighted pairwise scatter LDA(WPS-LDA)[2]와 nonlinear PCA에 기반한 eigenvector를 구하는 방법[3], 그리고 Nearest Neighborhood(NN) 방법을 사용한 weighting 방법을 적용하였다. NN 방법은 class들의 평균값을 기반으로 하여 오인식률을 NN에 의해 구한 다음, 오인식률에 비례하여 weighting을 주었다.

Li 등은 Euclidean distance 및 Kullback-Leibler(KL) distance에 기반한 두 가지 weighting 방법을 사용하였으며, 이중 Euclidean distance를 사용한 경우가 나은 성능을 나타내었다[2]. 본 실험에서도 동일한 결과를 얻었다. 이는 각각의 class를 unimodal 가우시안 모델로 가정하는 LDA의 모델링 오차에 의한 것으로 사료되며 정확한 모델링이 되어진다면 KL distance에 기반한 weighting 방법을 적용한 LDA의 성능도 증가할 것으로 사료된다.

표 5, 6, 7에 각각의 weighting 방법에 대한 결과가 나타나 있다.

표 5. WPS 방법 적용 후 인식률(%)

Mix- ture 수	상태수					
	3		4		5	
	숫자열	개별숫자	숫자열	개별숫자	숫자열	개별숫자
1	82.09	94.75	83.08	94.97	83.64	95.24
3	85.95	95.98	87.54	96.37	87.90	96.54
5	88.65	96.83	89.53	97.05	90.41	97.28
7	89.01	96.95	89.81	97.24	91.12	97.51
8	89.09	96.96	89.89	97.23	91.40	97.62
9	89.01	96.95	90.25	97.36	91.48	97.65

표 6. Nonlinear PCA 방법 적용 후 인식률(%)

Mix- ture 수	상태수					
	3		4		5	
	숫자열	개별숫자	숫자열	개별숫자	숫자열	개별숫자
1	82.01	94.71	82.96	94.99	83.84	95.28
3	86.86	96.21	87.66	96.45	88.30	96.68
5	87.94	96.63	89.93	97.15	90.09	97.17
7	88.89	96.90	89.93	97.18	90.88	97.43
8	88.89	96.92	90.37	97.33	91.20	97.54
9	88.97	96.95	90.37	97.35	91.40	97.56

표 7. NN 방법 적용 후 인식률(%)

Mix- ture 수	상태수					
	3		4		5	
	숫자열	개별숫자	숫자열	개별숫자	숫자열	개별숫자
1	81.77	94.65	83.16	95.05	83.36	95.05
3	86.27	96.05	87.62	96.51	88.93	96.87
5	88.38	96.76	89.49	97.05	90.41	97.35
7	88.73	96.87	90.33	97.33	91.12	97.47
8	88.57	96.86	90.37	97.34	91.12	97.54
9	88.57	96.85	90.29	97.31	91.24	97.59

표에서 보는 것과 같이 NN을 이용한 경우는 약간의 인식률 저하가 발생하였으나, WPS-LDA와 nonlinear PCA에 기반한 eigenvector를 구하는 방법의 경우는 기존의 LDA에 비해 미약하나마 성능의 향상등 보였다. 이는 상대적으로 거리가 가까운 class간에 변별력이 증가했음을 나타낸다. NN의 경우는 weighting 값으로 사용된 오인식률을 class의 확률 분포의 pdf 값에 의해 구하지 않고 단지 NN에 의해 구한 것에 대한 오차일 것으로 사료되며 앞으로 이에 대한 실험이 수행될 것이다.

### 5. 결 론

본 논문에서는 한국어 연결숫자 인식시스템의 성능을 향상시키기 위해 LDA를 적용하였다. LDA를 사용함

으로써 사용하지 않은 경우에 비해서 숫자열 오인식률이 8% 감소하였다. 그리고, 음성특징 벡터의 차수를 줄이더라도 인식률의 증가를 보였다 즉, LDA가 변별력에 대한 정보손실을 최소로 하여 차원 감소에 효과적인 방법임을 알 수 있었다. 또한 본 논문에서는 선형적인 방법으로 분리가 불가능한 class들의 변별력을 증가시키기 위해 between-scatter covariance matrix를 구할 때 class 사이에 혼동가능성 정도를 나타내는 weighting factor를 적용함으로써 숫자열 오인식률이 LDA 사용전보다 9.7% 감소하였다.

앞으로 class간의 변별력을 보다 증가시켜 인식성능을 높이기 위해서는 front-end에서의 LDA 적용 뿐만 아니라 HMM을 구성할 때도 변별적 훈련 방법을 도입해야 할 것으로 판단된다.

### 참고문헌

- [1] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second ed., Academic Press, 1990.
- [2] Y. Li, Y. Gao and H. Erdogan, "Weighted pairwise scatter to improve linear discriminant analysis," in Proc. ICSLP, vol.4, pp.608-611, 2000.
- [3] R. P. W. Duin, M. Loog and R. Haeb-Umbach, "Multi-class linear feature extraction by nonlinear PCA," in proc. Int'l Conf. Pattern Recognition, pp.398-401, 2000.
- [4] G. Saon, M. Padmanabhan, R. Gopinath and S. Chen, "Maximum likelihood discriminant feature spaces," in Proc. ICASSP, pp.1129-1132, 2000.
- [5] J. Duchateau, K. Demuynck, D. V. Compermolle and P. Wambacq, "Class definition in discriminant feature analysis," in Proc. EUROSPEECH, pp.1621-1624, 2001.
- [6] H. W. Song and H. S. Kim, "A study on performance of Korean connected digit recognizer by various modeling methods", in Proc. ICSP, pp.149-152, 2001.