

화자적응에서 PCA 또는 ICA를 이용한 MLLR알고리즘 연산량 감소

The Reduction of Computation in MLLR Framework Using PCA or ICA for Speaker Adaptation

김 지 운*, 정 재 호*
(Ji-Un Kim*, Jae-Ho Chung*)

*인하대학교 전자공학과 DSP Lab

(접수일자: 2003년 4월 4일; 수정일자: 2003년 7월 19일; 채택일자: 2003년 7월 22일)

본 논문은 화자 적응시 화자 독립 모델의 차수를 줄이고 MLLR (Maximum Likelihood Linear Regression) 알고리즘에서 요구되는 역행렬 횟수를 줄이는 방법을 제안한다. 주성분분석 (PCA: principal components analysis)과 독립성분분석 (ICA: independent components analysis)을 통해 모델 혼합성분 (mixture component)들간의 상관관계를 줄임으로서 모델의 차수를 감소하였다. 주성분분석 및 독립성분분석에 요구되는 추가 연산량은 화자 독립 모델을 훈련할 때 추가함으로써 화자 적응시에 추가되는 연산량은 극히 미소하다. 36차의 HMM 파라메타 차수를 PCA는 12차, ICA는 10차로 감소하였을 때 기존의 MLLR 적용방법과 유사한 단어 인식을 나타내었다. 즉, 모델 파라미터의 차수를 n 이라고 할 때 기존의 MLLR 알고리즘에서 역행렬 연산에서 요구되는 연산량은 $O(n^4)$ 에 비례하므로 PCA는 1/81, ICA는 1/167만큼 연산량을 감소하였다.

핵심용어: 화자적응, MLLR, 주성분분석 (PCA), 독립성분분석 (ICA)

투고분야: 음성처리 분야 (2.6)

We discuss how to reduce the number of inverse matrix and its dimensions requested in MLLR framework for speaker adaptation. To find a smaller set of variables with less redundancy, we adapt PCA (principal component analysis) and ICA (independent component analysis) that would give as good a representation as possible. The amount of additional computation when PCA or ICA is applied is as small as it can be disregarded. 10 components for ICA and 12 components for PCA represent similar performance with 36 components for ordinary MLLR framework. If dimension of SI model parameter is n , the amount of computation of inverse matrix in MLLR is proportioned to $O(n^4)$. So, compared with ordinary MLLR, the amount of total computation requested in speaker adaptation is reduced by about 1/81 in MLLR with PCA and 1/167 in MLLR with ICA.

Keywords: Speaker adaptation, MLLR, PCA, ICA

ASK subject classification: Speech signal processing (2.6)

1. 서론

음성인식 시스템은 화자의 특성의 변화, 음향환경의 변화, 혹은 잘못된 모델링과 같은 훈련환경과 인식환경의 차이에 매우 민감하다. 따라서 실제 환경에서 음성인식 시스템은 이러한 환경의 불일치를 보상하는 것이 매우 중요하다. 이는 강인하고 환경에 따른 변화가 적은

음성 특징 추출, 모델링 방법의 향상, 적응이나 보상 방법을 이용한 인식 파라미터나 특징 벡터의 수정, 강인한 결정 방법 등 다양한 방법으로 행해진다. 본 논문은 주어진 적응 데이터를 이용하여 기존의 모델로부터 새로운 음향학적 모델을 유도해 내기 위한 적응 방법에 대해 기술한다.

HMM에 기반한 음성 인식 시스템에 적용되는 적응 방법은 크게 2가지로 분류된다. 그 중 하나는 변환에 기초한 적응방법으로서 HMM 파라미터를 변환 함수에 의해 변환시키는 방법이다. MLLR, SM (Stochastic Matching)

책임저자: 김지운 (p2001118@invision.inha.ac.kr)
402-751 인천광역시 남구 용현동 253
인하대학교 전자공학과 디지털신호처리실
(전화: 032-860-7420; 팩스: 032-868-3654)

혹은 강제 변환 (constrained transform)이 이 부류에 해당한다[1, 8, 9]. 이런 방법들은 ML (Maximum Likelihood) 을 이용하여 군집 (cluster)에 의존적인 변환 파라미터를 구한다. 또 다른 분류는 MAP (Maximum A Posteriori) 적용 방법에 기인한 여러 가지 방법이다. 이는 사전 통계로서 SI (Speaker Independent) HMM 파라미터를 제공함으로써 평균, 분산, 혼합 가중치 등이 MAP방법에 의해 작용된다[2]. 일반적으로 적용 데이터가 제한된다면 변환에 기반한 적용이 군집에 의존적인 변환 함수에 의해 모든 HMM 파라미터를 효과적으로 변환시킬 수 있다. 반면, 만일 충분한 데이터가 제공되면 MAP방법이 변환 함수에 기반한 적용방법보다 SI HMM 파라미터를 효과적으로 적용시킬 수 있다[3, 4].

본 논문은 적은 양의 적용 데이터가 제공되는 적용 환경을 목표로 MLLR방법을 이용한다. MLLR방법은 변환에 기반한 방법으로써 변환 행렬을 구할 때 많은 양의 역행렬 연산을 수행해야 한다. 특히 음성 특징 파라미터의 차수가 커질수록, HMM 모델의 개수가 많아질수록 기하급수적으로 역행렬 연산횟수가 증가한다. 예를 들어, n 차의 특징 파라미터를 사용하고 MLLR방법에서 회귀 나무 (regression tree)의 기저 클래스 (base class)가 m 개라고 가정하면, $(n+1) \times (n+1)$ 차원의 행렬의 역행렬 연산을 $m \times n$ 회 행해야 한다. 따라서, MLLR 방법에서 변환 행렬을 구하는데 요구되는 역행렬의 연산 횟수를 줄이고 효율적으로 HMM 파라미터를 변환하는 방법이 요구된다. 잘 훈련된 SI모델의 각 혼합들을 몇 개의 클래스 (class)로 분류하고 -이때 요구되는 클래스는 MLLR방법을 적용 시에 사용되는 회귀 클래스 (regression class)를 사용한다. -이 클래스의 공간에 대한 사전정보를 이용하여 모델 및 음성 특징 파라미터의 차수를 줄이는 데에 사용한다. 즉, 각 클래스의 슈퍼 벡터 (주어진 클래스의 가우시안 평균을 연결하였다)에 주성분 분석이나 독립성분 분석을 이용하여 낮은 차수의 모델 파라미터를 얻을 수 있다. 주성분 분석이나 독립성분 분석시에 추가 되는 연산량은 SI모델을 훈련하는 과정에 추가되므로 화자 적용시에는 변환하고자 하는 혼합 성분의 평균벡터와 주성분 분석이나 독립 성분 분석을 위한 변환행렬의 곱 연산만이 추가 된다.

본논문의 구성은 2절에서 모델 파라미터의 차수 감소를 추가한 MLLR알고리즘을 유도하고, 3절에서는 독립성분분석과 주성분분석을 이용해 모델 파라미터의 차수 감소를 설명한다. 4절에서 실험방법 및 결과를 설명하고, 5절에서 결론을 맺는다.

II. 파라미터 차수 감소를 적용한 MLLR

MLLR알고리즘의 목적은 각 혼합 성분의 평균 벡터의 변환을 추정하여, SI모델을 새로운 화자의 SD모델로 바꾸는 것이다. 평균 μ_s 를 갖는 혼합 성분 s 에 대해 적용된 평균 $\hat{\mu}_s$ 는 다음과 같이 나타낸다.

$$\hat{\mu}_s = P^{-1}W_s P \xi_s \quad (1)$$

여기서 W_s 는 차수 $m \times (m+1)$ 인 변환 행렬이고 P 는 차수 $(m+1) \times (n+1)$ 인 차수감소 행렬이다. (단, $m < n$) ξ_s 는 평균벡터를 확장한 것으로

$$\xi_s = \{w, \mu_{s1}, \mu_{s2}, \dots, \mu_{sm}\} \quad (2)$$

여기서 w 는 잔류편차 (offset) 항목이다.

적용데이터를 $O = \{o_1, o_2, \dots, o_T\}$ 라고 가정하자. 여기서 T 는 관측 벡터의 개수이다. 또한, 모델 파라미터를 재추정하기 위한 보조 함수는 다음과 같이 나타낼 수 있다.

$$Q(\lambda, \bar{\lambda}) = a - \frac{1}{2} F(O|\lambda) \sum_{j=1}^m \sum_{t=1}^T \gamma_{j,t} [n \log(2\pi) + \log |\Sigma_j| + h(o_t, j)] \quad (3)$$

$$h(o_t, j) = (o_t - P^{-1}W_j P \xi_j)' \Sigma_j^{-1} (o_t - P^{-1}W_j P \xi_j) \quad (4)$$

$Q(\lambda, \bar{\lambda})$ 를 $P^{-1}W_j$ 에 대해 일차미분을 행하면 (5)와 같고,

$$\frac{d}{dW_j} Q(\lambda, \bar{\lambda}) = F(O|\lambda) \sum_{t=1}^T \gamma_{j,t} \Sigma_j^{-1} [o_t - P^{-1}W_j P \xi_j] \xi_j' P' \quad (5)$$

(5)의 최대값을 구하면 (6)과 같다

$$\sum_{t=1}^T \gamma_{j,t} \Sigma_j^{-1} o_t \xi_j' P' = \sum_{t=1}^T \gamma_{j,t} \Sigma_j^{-1} A_j P \xi_j \xi_j' P' \quad (6)$$

여기서 $A_j = P^{-1}W_j$ 이다. 식 (6)의 해를 구하는 자세한 방법은 C. J. Leggetter의 논문을 참고하라[1]. 따라서 화자 적용의 성능과 연산량 감소는 차수감소행렬 P 에 의존한다. 본 논문에서는 차수감소행렬 P 를 구하기 위해 주성분분석이나 독립성분분석을 이용하였다. 자세한 차수감소행렬 P 를 구하는 방법은 다음 장에서 설명한다.

III. 파라미터 차수 감소

본 논문에서는 파라미터의 차수 감소를 위해 주성분 분석 및 독립성분 분석을 적용한다. 주성분 분석이나 독립성분 분석을 이용해 기존의 MLLR 화자 적응 방법과 유사한 성능을 나타내도록 HMM 파라미터의 차수를 감소한다. 주성분 분석과 독립성분 분석을 적용하기 전에 먼저 SI 모델의 평균값을 몇 개의 클래스로 그룹화하는 작업이 필요한데 본 논문에서는 MLLR 방법을 수행하기 전에 만들어진 회귀 나무를 이용한다.

즉, 그림 1과 같이 회귀 나무의 기저 클래스의 각 평균 벡터들을 연결하여 하나의 슈퍼벡터, s_i 로 만든 후 이들을 하나의 열 (column) 벡터로 하는 행렬 $S = [s_1, s_2, \dots, s_k]$ 를 형성한다. 여기서 k 는 회귀 나무의 기저 클래스의 개수이다. 이와 같이 형성된 행렬 S 의 공분산 행렬을 구하여 주성분 분석 또는 독립성분 분석을 행한다.

주성분 분석의 기본 목표는 서로 상관성을 가지는 데이터들을 상관성이 없게 하는데 있다. 또한 데이터를 새로운 하위영역으로 사상하는 것이기 때문에 각 데이터의 차원을 줄일 수 있다. 수학적으로 말하면 데이터의 분산을 최대한으로 하는 축을 찾아서 축 변환에 의해서 각 데이터

들의 상관성을 최대한으로 줄이는 것이다. 주성분 분석을 통해 얻어진 고유 벡터 행렬 $V = [v_1, v_2, \dots, v_k]$ 을 이용하여 각 혼합 성분의 평균 벡터를 고유 벡터 영역으로 사상시킬 수 있다. 이때 고유값에 의거하여 사용하는 고유 벡터의 수를 조절할 수 있다. 즉, 데이터의 분포 정도가 상대적으로 작은 고유 벡터는 사용하지 않으므로써 혼합 성분의 평균 벡터의 차수를 줄일 수가 있다. 따라서 m 개의 고유 벡터를 사용한다면 차수감소 행렬 P 는 다음과 같이 나타낼 수 있다.

$$P = \begin{bmatrix} 1 & 0 \cdots 0 \\ 0 & \\ \vdots & V' \\ 0 & \end{bmatrix} \quad (7)$$

여기서 $V' = [v_1, v_2, \dots, v_m]$ 이다.

독립성분 분석에 의해 찾은 성분은 데이터의 상관성을 줄일 뿐만 아니라 서로 독립인 성분을 찾는다. 즉, 독립성분 분석이 주성분 분석의 확장된 이론이라는 것을 알 수 있다[5]. 독립성분 분석은 독립성분 분석에 의한 결과 신호가 통계적으로 독립인 선행축을 찾는 방법이다. 주성분 분석과 같은 상관관계에 의거한 변환 방법과 비교하였을 때, 독립성분 분석은 신호의 상관관계-통계학적으로 2차원적인 분석 방법이다. -를 줄일 뿐 아니라, 더 높은 차원의 통계적 의존관계를 줄일 수 있다. 따라서 일반적으로 신호의 차원을 줄이는 데에는 독립성분 분석이 주성분 분석보다 좋은 성능을 나타낸다. 그러나 입력 데이터에 독립성분이 충분히 포함되어 있지 않을 경우, PCA가 ICA보다 좋은 성능을 나타낸다[7]. 독립성분 분석에 의해 얻어진 가중행렬 A 를 이용하여 주성분 분석에서와 마찬가지로 차수감소 행렬 P 를 구할 수 있다.

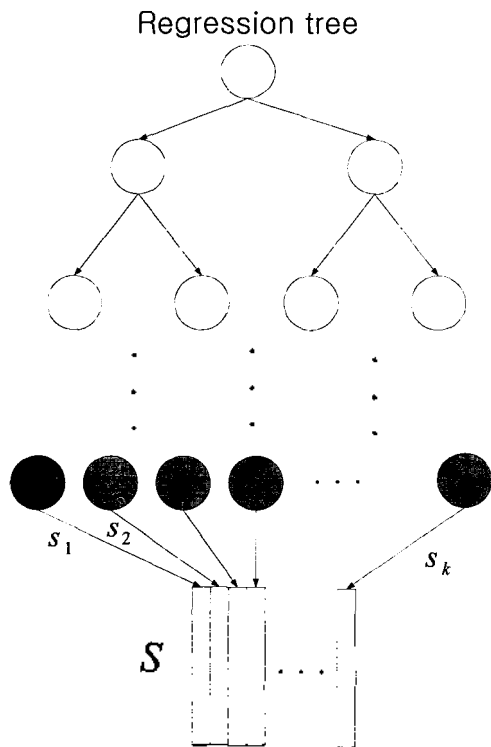


그림 1. 슈퍼벡터 행렬 형성
Fig. 1. Constructing a supervector matrix S .

IV. 실험 및 실험 결과

본 논문에서 사용한 데이터 베이스는 DARPA Resource Management Continuous Speech Database (RM)이다. 80명의 화자가 발성한 42개의 문장을 훈련 데이터로 사용하였다. 화자 적응용으로 12명의 화자로부터 각 화자마다 612개의 문장을 사용하였다. 화자 적응 성능 테스트용으로는 100개의 문장이 사용되었다. 사용한 음성의 특징 파라미터는 18차의 LPOC (Linear Predictive Cepstral Coefficient)와 델타계수를 사용하였고, 잡음 제거를 위

표 1. 최대 차수일 때 PCA와 ICA 및 MLLR의 성능 비교
Table 1. The comparison of word recognition rate when PCA, ICA, and ordinary MLLR is constrained to have maximum components.

# Utterances	1	2	3	4
CA (order: 16)	92.9	95.1	95.9	96.5
PCA (order: 36)	93.4	95.5	96.2	96.7
MLLR (order: 36)	90.8	92.8	94.3	96
SD	92.1	92.1	92.1	92.1

표 2. 차수에 따른 단어 인식률
Table 2. Word recognition rate according to the dimension of parameters.

	1	2	3	4	연산량 감소율
PCA (order: 10)	90.7	92.2	93.8	95.1	1/167
PCA (order: 12)	91.2	92.9	94.5	95.6	1/81
PCA (order: 14)	91.5	93.7	94.9	96	1/43
PCA (order: 16)	91.8	94.5	95.4	96	1/25
ICA (order: 10)	91.2	92.6	94.1	95.9	1/167
ICA (order: 12)	91.8	93.4	94.9	96.1	1/81
ICA (order: 14)	92.4	93.9	95.4	96.2	1/43
ICA (order: 16)	92.9	95.1	95.9	96.5	1/25
MLLR (order: 36)	90.8	92.8	94.8	96	1
SD	92.1	92.1	92.1	92.1	

해 CMS (Cepstral Mean Subtraction)을 행하였다. 음성 인식은 49개의 단소 음소 (monophone)으로 훈련된 모델을 이용하여 행하였다. 각 단소 음소모델은 3개의 상태 (state)를 가지고 있고 각 상태는 8개의 혼합 성분을 포함한다. MLLR방법을 이용한 실험에서는 블록 (block)의 개수가 2개인 블록 대각 (block diagonal) 변환 행렬을 사용하였고, 제안한 알고리즘에서는 블록의 수가 1개인 적응 변환 행렬을 사용하였다.

표 1은 PCA와 ICA의 차수를 최대로 사용했을 때의 MLLR과의 단어 인식률 성능을 비교하였다. MLLR과 PCA는 36차를 사용하였으며, ICA는 16차를 사용하였다. PCA와 ICA 모두 MLLR에 비해 적응 속도가 빠른 것으로 나타났다. 특히 ICA방법은 MLLR과 비교하였을 때, MLLR의 절반의 차수보다 작은 차수로 MLLR보다 높은 성능을 나타내었다.

표 2는 차수에 따른 단어 인식률과 기존의 MLLR알고리즘에 대한 상대적인 연산량 감소율을 보이고 있다. ICA에서는 10차의 독립성분이, PCA에서는 12차의 주성분이 기존의 MLLR알고리즘의 성능과 유사한 성능을 나타내었다. 더 높은 차수의 주성분이나 독립성분이 사용될 때 MLLR 알고리즘보다 높은 인식률을 나타내었다. PCA

방법의 결과와 ICA방법의 결과를 비교해 보면 ICA방법이 PCA방법보다 성능이 좋게 나타났다. 이는 독립성분분석이 통계학적으로 2차원적인 분석방법인 신호의 상관관계를 줄였을 뿐 아니라, 더 높은 차원의 통계적 의존관계를 줄였기 때문으로 사료된다. 또한 PCA방법은 분산이 상대적으로 적은 축의 고유 벡터는 사용하지 않기 때문에 차원의 감소에 의해 데이터의 손실을 초래하고, ICA는 차원의 감소에 의해 서로 비슷한 분포를 갖는 데이터는 같은 독립 성분으로 가정하므로 PCA방법에 비해 데이터의 손실을 적게 초래한 것으로 사료된다.

연산량에 대해 고려하면, 역행렬 연산은 $O(n^3)$ 에 비례하므로 n 회의 역행렬 연산이 필요한 MLLR 화자적응을 위한 연산량은 $O(n^4)$ 에 비례한다. 본 실험에서는 하나의 변환 행렬을 구하기 위해 MLLR에서 역행렬 연산은 $(36+1) * (36+1)$ 차원의 행렬의 역행렬 연산을 36회 요구하는 반면, 주성분분석을 이용한 화자적응에서의 역행렬 연산은 $(12+1) * (12+1)$ 차원의 행렬의 역행렬 연산을 12회 요구하였고, ICA에서는 $(10+1) * (10+1)$ 차원의 행렬의 역행렬 연산을 10회 요구하였다. 즉, HMM 파라메타 차수를 PCA는 12차, ICA는 10차로 감소하였을 때, 기존의 MLLR알고리즘과 유사한 성능을 나타내었으므로 연산량은 기존의 MLLR알고리즘에 비해 PCA는 1/81만큼, ICA는 1/167만큼 감소하였다. 제안한 알고리즘에서는 차원을 감소하기 위해 각 평균 벡터 μ_s 와 차원감소행렬 P 의 행렬의 곱 연산이 추가로 요구되지만 이는 $O(n^2)$ 에 비례하므로 전체 연산량에 큰 영향을 미치지 않는다. 또한, 차원 감소 행렬을 구하기 위한 주성분 분석과 독립성분 분석에 요구되는 연산량은 SI모델을 훈련하는 과정에 추가되므로 화자 적응시 요구되는 연산량은 없다.

V. 결론 및 향후 과제

본 논문은 MLLR의 연산량 감소를 위해 주성분분석과 독립성분분석을 이용하여 모델의 차수를 줄임으로서 MLLR에서 요구하는 역행렬 연산의 횟수를 줄였다. 주성분분석이나 독립성분분석에서 요구되는 연산량은 SI모델을 훈련할 때 요구되므로 화자적응시 추가되지 않고 연산량 감소 행렬을 곱하는데에 요구되는 연산량은 역행렬 연산에 비해 상대적으로 매우 적은 양이므로 영향을 미치지 않는다. 주성분분석과 독립성분분석을 행하면 차수의 감소 뿐만 아니라 각 성분끼리의 변별성이 높아진다. 즉, 회귀 나무를 형성할 때 더 깊은 나무 (tree)를 형성

할 수 있으므로 화자적응 성능을 더욱 높일 수 있다. 따라서, 주성분분석과 독립성분분석에 맞는 희귀 나무를 형성하는 연구가 요구된다.

참고 문헌

1. C. J. Leggetter, *Improved acoustic modelling for HMMs using linear transforms*, PhD Thesis, Univ. of Cambridge, 1995.
2. C.-H. Lee, C.-H. Lin, and B.-H. Juang, "A study on speaker adaptation of the parameters of continuous density hidden Markov models," *IEEE Trans. on Signal Processing*, 39 (4), 806-814, April 1991.
3. O. Siohan, C. Chesta, and C.-H. Lee, "Joint maximum a posteriori adaptation of transformation and HMM parameters," *IEEE Trans. on Speech and Audio Processing*, 9, (4), 417-428, May 2001.
4. J. T. Chien, "Online hierarchical transformation of hidden Markov models for speech recognition," *IEEE Trans. on Speech and Audio Processing*, 7 (6), 656-667, 1999.
5. A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley-Interscience, 2001.
6. M. E. Tipping, and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society, Series B*, 61, Part 3, 611-612, 1999.
7. D. Ridder, J. Kittler, and R. P. W. Duin, "Probabilistic PCS and ICA subspace mixture models for image segmentation,"

The Eleventh British Machine Vision Conference, 112-121, 2000.

8. A. Sankar, and C.-H. Lee, "A maximum-likelihood approach to stochastic matching for robust speech recognition," *IEEE Trans. on Speech Audio Processing*, 4, 190-202, 1996.
9. V. Digalakis, "On-line adaptation of hidden Markov models using incremental estimation algorithms," *Proc. 5th Eur. Conf. Speech Communication and Technology*, 4, 1859-1862, 1997.

저자 약력

● 김 지 훈 (Ji-Un Kim)



1998년2월: 인하대학교 전자공학과 공학사
2000년2월: 인하대학교 전자공학과 공학석사
2000년3월~현재: 인하대학교 전자공학과 박사과정
※ 주관심분야: 음성인식, 발화인증, 화자인식, 화자 적응

● 정 재 호 (Jae-Ho Chung)

1982년: University of Maryland (BSEE)
1984년: University of Maryland (MSEE)
1990년: Georgia Institute of Technology (Ph. D.)
1984년~1985년: 미국 국방성 산하 해군 연구소, 신호처리실, 연구원
1991년~1992년: AT&T Bell Laboratories, 음성신호처리연구실, 연구원 (MTS)
1992년~현재: 인하대학교 공과대학 전자공학과, 정교수