

FIR 필터링에 대한 HMM 파라미터 적응기법

김 남 수, 김 동 국

서울대학교 전기공학부

HMM Parameter Adaptation to FIR Filtering

Nam Soo Kim and Dong Kook Kim

School of Electrical Engineering, Seoul National University

Kwanak, P.O.Box 34, Seoul 151-742 Korea

Tel: +82-2-884-1824 Fax: +82-2-878-1452

e-mail : {nkim,dkkim11}@snu.ac.kr

요 약

본 연구에서는 finite impulse response (FIR) 필터에 의해 인식기의 입력 특징벡터가 필터링되는 경우에 hidden Markov model (HMM) 파라미터를 적응시키는 새로운 기법을 제안한다. 제안한 적응 기법은 필터링에 의해 변환된 특징벡터에 대해 HMM 파라미터를 다시 학습시킬 필요가 없으며 주어진 FIR 필터 계수만을 사용하여 HMM 파라미터를 적응시킬 수 있다. 개발된 FIR 필터링에 대한 HMM 파라미터 적응 기법은 연속 숫자음 인식 실험에서 재학습 방법과 비교 실험한 결과 low-pass 필터의 경우에 재학습 방법과 비슷한 인식 성능을 나타내었다.

1 서론

음성인식 분야에서 인식률 향상을 위해 특징벡터 선택은 중요한 문제로 여겨져 왔다. 인식기의 성능은 사용된 특징벡터에 의해 많이 좌우되는데 일반적으로 cepstrum에 기반한 특징벡터를 가장 많이 사용하고 있다. 좋은 특징벡터는 여러가지 다른 음운학적 단위들을 구별할 수 있어야 하며 화자차이, 채널왜곡 그리고 가산 잡음 등에 의한 학습과 테스트의 불일치 (mismatch)에 대해서도 robust 특성을 가져야 한다. 특징벡터는 음성 샘플로부터 추출된 특징벡터를 변환함으로써 얻어진다. 예로서 cepstrum에 대한 일차 미분 특성을 나타내는 delta-cepstrum은 cepstrum에 특정한 필터를 적용함으로써 얻어진다. 이런 delta-cepstrum을 cepstrum과 함께 사용함으로써 인식기의 변별력과 성능을 향상시킬 수 있다 [1].

특징벡터를 변환하는 기법은 배경 잡음과 같은 환경적인 불일치를 극복하기 위한 robust 음성인식 분야에서 많이 사용되었다. 잘 알려진 기법중에 하나는 천천히 변하는 잡음 성분을 제거하는데 사용되는 cep-

stral mean subtraction (CMS) 방법이다 [2]. CMS 방법은 저주파 성분을 제거하기 위해 cepstrum 열 (sequence)에 high-pass 필터를 적용함으로써 구현된다. 그리고 여러가지 잡음이나 채널왜곡 등의 환경에서 성능 향상을 위해 여러가지 필터를 사용하여 특징벡터를 변환하는 다양한 방법들이 제안되었다 [3,4,5,6].

본 논문에서는 특징벡터가 causal한 finite impulse response (FIR) 필터에 의해 변환될 때 hidden Markov model (HMM) 파라미터를 적응하는 기법을 제안한다. 기존에는 일반적으로 입력 특징벡터가 변환될 때 변환된 특징벡터에 기반하여 HMM 파라미터를 재학습시켜야 한다. 그러나 제안된 방법에서는 주어진 필터 계수만을 사용하여 HMM 파라미터를 변환하는 기법을 소개한다. 개발된 FIR 필터링에 대한 HMM 파라미터 적응 기법의 성능은 연속 숫자음 인식 실험에 의해 재학습 방법과 비교 평가되었다. 개발된 파라미터 적응 방법의 성능은 변환된 특징벡터를 재학습하는 경우와 비교하여 low-pass 필터의 경우에 비슷한 성능을 나타내었다. 그러므로 제안된 적응 방법은 재학습을 위한 시간을 줄일 수 있는 장점을 갖는다.

2 HMM 파라미터

HMM에 의해 발생되는 특징벡터 열을 x_1, x_2, \dots, x_T 로 나타낸다. 여기서 HMM은 continuous density Gaussian mixture 모델 형태이며, 파라미터는, $(\{\pi_i\}, \{A_{ij}\}, \{\theta_i\})$, $1 \leq i, j \leq N$ 이다. 이 때 N 는 전체 상태 (state) 수, $\{\pi_i\}$ 는 초기상태분포 (initial state distribution) 그리고 $\{A_{ij}\}$ 는 전이확률 (transition probability)

$$A_{ij} = P(q_t = i | q_{t-1} = j) \quad (1)$$

를 나타내며 q_t 는 시간 t 에서 상태이다. Gaussian mixture 모델에서 파라미터는

$$\theta_i = (\{\omega_{ik}, \mu_{ik}, \Sigma_{ik}\}), 1 \leq k \leq K(i) \quad (2)$$

이며, 위에서 $K(i)$ 는 상태 i 에서 mixture 성분 수이며 그리고 $\{\omega_{ik}\}$ 는

$$\omega_{ik} = P(m_t = k | q_t = i) \quad (3)$$

로 주어진 mixture 가중치 (weight)이다. 그리고 m_t 는 시간 t 에서 mixture 성분을 나타낸다. μ_{ik} 와 Σ_{ik} 는 상태 i 에서 k 번째 mixture 성분과 관련된 평균 (mean)과 공분산 (covariance)이며 다음과 같이 쓸 수 있다.

$$\begin{aligned} \mu_{ik} &= E[x_t | q_t = i, m_t = k] \\ \Sigma_{ik} &= Cov[x_t | q_t = i, m_t = k] \end{aligned} \quad (4)$$

위에서 $E[\cdot]$ 와 $Cov[\cdot]$ 는 주어진 랜덤변수 (random variable)에 대한 기대치 (expectation)와 공분산을 각각 나타낸다.

3 FIR 필터링에 대한 HMM 파라미터 적용

주어진 특징벡터 열에 필터링을 적용하기 위해 다음과 같이 causal한 FIR 필터를 사용한다.

$$z_t = \sum_{l=0}^M a_l x_{t-l} \quad (5)$$

위에서 z_t 는 시간 t 에서 필터 출력을 나타내며, M 은 필터 차수 그리고 $\{a_l\}$, $l = 1, 2, \dots, M$ 는 필터 계수이다. 제안된 방법은 필터 출력 열, z_1, z_2, \dots, z_T 에 대한 HMM 파라미터를 FIR 필터 계수와 원 HMM 파라미터만을 사용하는 적용 기법이다. 즉, 필터 계수, $\{a_l\}$ 가 주어지는 경우 필터 출력에 대한 모든 HMM 파라미터를 재학습하는 대신, 입력 특징벡터 열, x_1, x_2, \dots, x_T 에서 학습된 HMM 파라미터를 변환시켜 필터 출력에 대한 HMM 파라미터를 구하는 기법이다. 일반적으로 다른 모델 적용 기법에서와 같이 여기서도 $\{\mu_{ik}, \Sigma_{ik}\}$ 값만을 변환하고 나머지는 원래값을 사용한다.

3.1 Mean 적용화

벡터 $\psi_{ik}(t-l)$ 를 각 상태와 mixture, (i, k) 와 $0 \leq l \leq M$ 에 대해 다음과 같이 정의한다.

$$\begin{aligned} \psi_{ik}(t-l) &\equiv \\ &E[a_l x_{t-l} + \dots + a_M x_{t-M} | q_{t-l} = i, m_{t-l} = k]. \end{aligned} \quad (6)$$

(6)은 $l < M$ 에 대해 (7)과 같이 표현할 수 있다.

$$\begin{aligned} \psi_{ik}(t-l) &= a_l E[x_{t-l} | q_{t-l} = i, m_{t-l} = k] \\ &+ \sum_{i'=1}^N \sum_{k'=1}^{K(i')} E[a_{l+1} x_{t-l-1} + \dots + a_M x_{t-M} \\ &| q_{t-l} = i, m_{t-l} = k, q_{t-l-1} = i', m_{t-l-1} = k'] \\ &\cdot P(q_{t-l-1} = i', m_{t-l-1} = k' | q_{t-l} = i, m_{t-l} = k). \end{aligned} \quad (7)$$

x_{t-l} 는 $l > 0$ 동안에 q_t 나 m_t 에 대해 통계적으로 독립이고 주어진 HMM은 시불변하기 때문에 (7)은 (8)과 같이 표현된다.

$$\begin{aligned} \psi_{ik}(t-l) &= \\ &a_l \mu_{ik} + \sum_{i'=1}^N \sum_{k'=1}^{K(i')} \psi_{i'k'}(t-l-1) \Lambda(i', k' | i, k) \end{aligned} \quad (8)$$

위에서

$$\begin{aligned} \Lambda(i', k' | i, k) &= P(q_{t-l-1} = i', m_{t-l-1} = k' | q_{t-l} = i, m_{t-l} = k) \\ &= \frac{P(q_{t-l} = i | q_{t-l-1} = i') P(m_{t-l-1} = k' | q_{t-l-1} = i')}{P(q_{t-l} = i)} \\ &\cdot P(q_{t-l-1} = i') \\ &= \frac{A_{i' i'} \cdot \omega_{i' k'} \cdot P(q_{t-l-1} = i')}{P(q_{t-l} = i)}. \end{aligned} \quad (9)$$

(9)에서 확률 $P(q_{t-l} = i)$ 은 다음과 같이 반복적으로 갱신 된다.

$$\begin{aligned} P(q_{t-l} = i) &= \sum_{i'=1}^N P(q_{t-l} = i | q_{t-l-1} = i') P(q_{t-l-1} = i') \\ &= \sum_{i'=1}^N A_{i' i'} \cdot P(q_{t-l-1} = i'). \end{aligned} \quad (10)$$

$\hat{\mu}_{ik}$ 가 i 번째 상태에서 k 번째 mixture 성분의 적용된 평균 벡터를 나타낸다면

$$\begin{aligned} \hat{\mu}_{ik} &\equiv E[z_t | q_t = i, m_t = k] \\ &= \psi_{ik}(t). \end{aligned} \quad (11)$$

3.2 Covariance 적용화

공분산 행렬, $\{\Sigma_{ik}\}$ 을 변환하기 위해 한 시간에서의 특징벡터는 다른 시간에서의 특징벡터와 uncorrelated 되었다고 가정하자. 위 가정하에 $l > 0$ 동안에

$$\begin{aligned} E[x_t x'_{t-l} | q_t = i, m_t = k] &= \\ E[x_t | q_t = i, m_t = k] E[x'_{t-l} | q_t = i, m_t = k] \end{aligned} \quad (12)$$

이며, 위에서 '은 행렬의 transpose 이다. $\{\psi_{ik}(t)\}$ 가 평균 벡터를 적응하기 위해 쓰인 것같이 각각의 (i, j) 와 $0 \leq l \leq M$ 에 대해 행렬 $\Pi_{ik}(t-l)$ 을 (13)과 같이 정의한다.

$$\Pi_{ik}(t-l) \equiv \text{Cov}[a_1 x_{t-l} + \dots + a_M x_{t-l-M} | q_{t-l} = i, m_{t-l} = k]. \quad (13)$$

(12)에 주어진 가정하에서 $\{\Pi_{ik}(t)\}$ 을 (8)과 비슷하게 (14)와 같이 표현할 수 있다.

$$\begin{aligned} \Pi_{ik}(t-l) &= a_i^2 \text{Cov}[x_{t-l} | q_{t-l} = i, m_{t-l} = k] \\ &+ \sum_{i'=1}^N \sum_{k'=1}^{K(i')} \text{Cov}[a_{i+1} x_{t-l-1} + \dots + a_M x_{t-l-M} \\ &| q_{t-l} = i, m_{t-l} = k, q_{t-l-1} = i', m_{t-l-1} = k'] \\ &\cdot P(q_{t-l-1} = i', m_{t-l-1} = k' | q_{t-l} = i, m_{t-l} = k) \\ &= a_i^2 \Sigma_{ik} + \sum_{i'=1}^N \sum_{k'=1}^{K(i')} \Pi_{i'k'}(t-l-1) \cdot \Lambda(i', k' | i, k). \end{aligned} \quad (14)$$

적용된 공분산 행렬, $\{\hat{\Sigma}_{ik}\}$ 은 위에서 다음과 같이 구해진다.

$$\hat{\Sigma}_{ik} = \Pi_{ik}(t). \quad (15)$$

3.3 적응화 알고리즘

HMM 파라미터 적응화를 위해 사용된 FIR 필터에 대해 평균 벡터와 공분산 행렬을 (8), (10) 그리고 (14)를 이용하여 반복적으로 적응시킨다. 먼저 초기확률값, $\{P(q_{t-M} = i)\}$ 을 결정하는데 이는 모든 상태에서 같은 확률값을 갖는다고 가정한다. 즉, $i = 1, 2, \dots, N$ 에 대해

$$P(q_{t-M} = i) = \frac{1}{N}. \quad (16)$$

HMM 파라미터 적응을 위한 알고리즘은 다음과 같다.

HMM 파라미터 적응 알고리즘

i) 초기화 :

$i = 1, 2, \dots, N$ 과 $k = 1, 2, \dots, K(i)$ 에 대해,

$$\begin{aligned} P(q_{t-M} = i) &= \frac{1}{N}, \\ \psi_{ik}(t-M) &= a_M \mu_{ik}, \\ \Pi_{ik}(t-M) &= a_M^2 \Sigma_{ik}. \end{aligned}$$

ii) Recursion:

$l = M-1, M-2, \dots, 0$ 에 대해 (10), (8) 그리고 (14)을 이용하여 $\{P(q_{t-l} = i)\}$, $\{\psi_{ik}(t-l)\}$ 그리고 $\{\Pi_{ik}(t-l)\}$ 을 구한다.

iii) 파라미터 적응화:

$i = 1, 2, \dots, N$ 과 $k = 1, 2, \dots, K(i)$ 에 대해,

$$\begin{aligned} \hat{\mu}_{ik} &= \psi_{ik}(t) \\ \hat{\Sigma}_{ik} &= \Pi_{ik}(t). \end{aligned}$$

4 실험 및 결과

제안된 HMM 파라미터 적응 알고리즘의 성능을 평가하기 위해 화자독립 연속 숫자음 인식 실험을 수행하였다. 학습과 테스트를 위해 93명과 47명의 각각 다른 화자를 사용하였고, 화자당 3-7자리 숫자로 구성된 문장을 30-40번 발성하였다. 8kHz로 sampling된 음성샘플을 10 ms 프레임 rate로 30 ms 구간을 분석하여 얻어진 12차 mel-frequency cepstral coefficient (MFCC)를 특징벡터로 사용하였다. 인식에 사용된 기본단위는 묵음과 11개의 숫자음 단어를 사용하였다. HMM는 left-to-right 모델 형태이며, 각 모델당 7개의 state와 각 state당 4개의 mixture를 사용하였다.

위에서 얻어진 입력 cepstrum 벡터는 다음과 같은 전달함수를 갖는 causal한 FIR 필터에 의해 필터링하였다.

$$H(z) = \sum_{k=0}^M \alpha^k z^{-k} \quad (17)$$

위에서 $|\alpha| < 1$ 이고, M 은 필터 차수를 나타낸다. (17)에서 $\alpha > 0$ 인 경우는 low-pass 형태이고 $\alpha < 0$ 인 경우는 high-pass 형태의 FIR 필터이다. 본 실험에서 필터 차수는 4차를 사용하였고, α 값은 $\{-0.40, -0.25, -0.10, 0.10, 0.25, 0.40\}$ 을 사용하였다.

원래의 cepstrum 벡터가 사용되는 경우에 기본 인식 시스템은 80.6%의 단어 인식률을 나타내었다. α 값에 따라서 제안된 HMM 적응 방법에 대한 인식 성능은 다음 표 1과 같다.

α	적용된 파라미터	재학습된 파라미터
-0.40	66.0	79.1
-0.25	74.6	79.9
-0.10	79.1	80.5
0.10	81.2	80.6
0.25	80.9	80.9
0.40	78.7	79.5

표 1: 적용된 파라미터와 재학습된 파라미터의 단어 인식률 (%).

표 1에서 필터링된 특징벡터를 재학습해 얻은 HMM 파라미터와 인식 성능을 비교하였다. 결과로부터 본 방법에 의해 적용된 HMM 파라미터 성능은 $\alpha > 0$ 인 경우에 재학습된 경우에 비해 약간 저하되었다. 그리고 $\alpha = 0.10$ 인 경우는 적용된 파라미터의 인식 성능이 재학습된 경우보다 향상되었다. 한편 $\alpha < 0$ 인 high-pass

필터인 경우는 재학습된 파라미터가 적용된 파라미터에 비해 우수하였다. α 값이 작아짐에 따라 두 방법 사이의 성능은 더 커짐을 알 수 있다.

5 결론

본 연구에서는 입력 특징벡터가 FIR 필터에 의해 변환될 때, HMM 파라미터를 적용시키는 기법을 제안하였다. 제안된 적용 기법의 장점은 변환된 입력 벡터에 대해 HMM 파라미터를 재학습시키지 않고 필터계수와 원 HMM 파라미터로 적용시키는 것이다. 연속 숫자 음 인식 실험 결과를 통해 제안된 방법이 low-pass 필터 형태가 적용될 때 잘 동작함을 알 수 있었다.

6 참고문헌

- [1] S. Furui, "Speaker independent isolated word recognition based on dynamics emphasized cepstrum," *Trans. IECE Jpn.*, Vol. 69, No. 12, pp. 1310-1317, 1986.
- [2] C. Mokbel, J. Monne, and D. Juvet, "On-line adaptation of a speech recognizer to variations in telephone line conditions," *Proc. of EUROSPEECH*, Berlin, Germany, pp. 1247-1250, 1993.
- [3] H. Hermansky and N. Morgan, "RASTA processing of speech," *IEEE Trans. Speech and Audio Processing*, Vol. 2, No. 4, pp. 578-589, Oct. 1994.
- [4] B. A. Hanson and T. H. Applebaum, "Subband or cepstral domain filtering for recognition of Lombard and channel-distorted speech," *Proc. of Int. Conf. Acoust., Speech, Signal Processing*, Minneapolis, MN, Vol. II, pp. 79-82, 1993.
- [5] C. Nadeu, P. Paches-Leal, and B. -H. Juang, "Filtering the time sequence of spectral parameters for speaker-independent CDHMM word recognition," *Proc. of EUROSPEECH*, Madrid, Spain, pp. 923-926, 1995.
- [6] H. G. Hirsch, P. Meyer, and H. W. Ruehl, "Improved speech recognition using high-pass filtering of subband envelopes," *Proc. of EUROSPEECH*, Genova, Italy, pp. 413-416, 1991.