

모바일 환경을 위한 SDCHMM의 효율적인 코드북 설계

김수현, 정규준, 오영환
한국과학기술원

An efficient codebook design in SDCHMM for mobile environment

Su-hyeon Kim, Gue-Jun Hung, Yung-Hwan Oh

Department of Electrical Engineering & Computer Science Division of Computer Science

Korea Advanced Institute of Science and Technology

{ksh, sylph, yhoh}@speech.kaist.ac.kr

요약

모바일 환경에 음성 인식을 적용하기 위해서는 메모리를 효율적으로 사용할 수 있도록 모델을 구성할 필요가 있다. 이러한 필요성을 충족시킬 수 있는 SDCHMM은 가우시안 분포를 부분공간으로 나누고 각 부분 공간에 대한 코드북을 구성해 사용하여 메모리 사용면에서 CHMM의 50% 정도를 이용하면서도 인식률은 유지하는 장점을 가지고 있다. 본 논문에서는 코드북 생성 단계에서 양자화 오류를 줄일 수 있도록 부분 공간을 구성해주는 방법을 제안한다. SDCHMM에서는 코드북을 구성할 때 가우시안 분포의 평균값들을 양자화를 하기 때문에, 가우시안 분포들의 각 차원에 대한 분산을 구하고 그 분산의 크기가 근접한 차원들로 부분공간을 구성할 경우 양자화 오류를 줄일 수 있게 된다. 또한 코드북 저장에 필요한 메모리 크기를 줄이기 위해 다단계 (Multistage) 벡터 양자화를 적용하여 코드북을 구성하였다. 제안한 방법의 유효성을 검증하기 위해 RM DB를 사용하여 실험한 결과 CHMM에 비해 모델 표현을 위해 필요한 메모리를 86% 정도 줄이면서도 CHMM의 표현 능력을 유지할 수 있었다.

1. 서론

모바일 환경에서의 음성 인식은 크게 서버에서 인식

을 하는 경우와 단말기에서 인식을 하는 경우로 나누어진다. 그 중에서 단말기에서 직접 인식을 하는 경우 메모리와 컴퓨팅 파워의 제한으로 인한 성능 제약을 가지게 된다[1]. 따라서 단말기에서 음성인식을 하기 위해서는 모델을 표현하기 위해 쓰이는 메모리를 최소로 하면서 인식 성능을 크게 떨어뜨리지 않는 모델 표현 방법이 필요하다[2]. 이러한 측면에서 진행되어온 연구중 SDCHMM은 모델 표현을 위한 메모리를 크게 줄여주면서도 인식 성능을 CHMM과 비슷하게 유지해 준다[3]. SDCHMM은 가우시안 분포를 부분공간으로 나누고 각 부분공간에 대해서 코드북을 구성함으로써 모델을 표현하는 메모리를 감소시킨다.

이때 부분 공간을 어떻게 구성하느냐가 중요한 문제가 된다. 기존의 SDCHMM에서는 개념적인 정의에 의하여 부분공간을 나누어 주거나 파라미터들 사이의 상관관계를 가지고 부분공간을 구성하였다. 하지만 좀 더 정확한 인식을 위해서는 부분공간을 구성할 때 양자화 오류를 고려할 필요가 있다[4]. 따라서 본 논문에서는 양자화 오류를 고려하여 가우시안 분포의 평균 값들 중 분포가 비슷한 것끼리 같은 부분 공간으로 구성하는 방법을 제안한다.

2장에서는 SDCHMM과 기존의 부분 공간을 구성하는 방법 및 기존 방식의 문제점을 살펴보고, 3장에서 양자

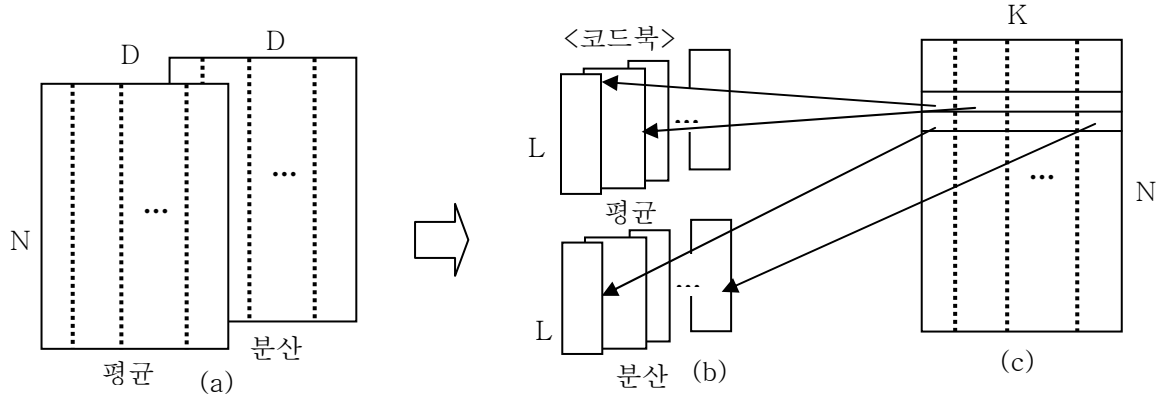


그림 1. 코드북 구성 과정 (a) 초기 CHMM (b) 코드북 생성 (c) 평균과 분산을 코드북 인덱스로 표현.
(N: 전체 가우시안분포수, D: 가우시안 분포의 차원, L: 코드북 사이즈, K: 부분 공간의 수)

화 오류를 고려한 부분공간 구성 방법 제안한다. 4장에서 실험 및 결과를 살펴보고 5장에서 결론을 맺도록 한다.

2. Subspace Distribution Clustering HMM

2.1 SDCHMM의 정의

SDCHMM은 CHMM을 이루는 가우시안 분포를 직교하는 부분공간으로 나누고 각 부분공간 별로 군집화 한 음성 모델이다[3]. 군집화 결과인 코드북으로 된 가우시안 분포를 표현할 수 있기 때문에, 적은 수의 모델 파라미터를 필요로 하여 메모리 감소의 측면에서 큰 효과를 얻을 수 있다.

$$P_s^{SDCHMM}(O) = \sum_{m=1}^{M_s} c_{sm} \left(\prod_{k=1}^K N^{tied}(O_k; \mu_{smk}, \sigma_{smk}^2) \right) \quad (1)$$

각 부분공간에 대해서 모든 CHMM의 모델들을 이루는 모든 상태(state)에 대해서 클러스터링을 하면 각 상태에서의 관측확률은 위의 식 (1)과 같이 계산된다.

2.2 상관관계에 의한 부분공간 구성법

기존의 SDCHMM에서 부분공간을 나누는 방법으로는 개념적 정의에 따른 부분공간 구성(Common Streams)과 상관관계에 의한 구성 방법(Correlated-Feature Streams)이 있다[5].

개념적 정의에 의해서 부분 공간을 구성하는 방법은 개념적으로 유사한 파라미터들이 하나의 부분공간을 이루게 된다.

상관관계에 의한 부분 공간 구성방법은 상관관계가 큰 파라미터들을 같은 부분 공간으로 구성해 주는 방법이다. 이때 상관관계를 측정하는 척도인 다중 상관관계 계수(Multiple correlation measure)는 식 (2)와 같이 정의된다.

$$R = 1 - \begin{vmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2n} \\ \rho_{31} & \rho_{32} & 1 & \cdots & \rho_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \rho_{n3} & \cdots & 1 \end{vmatrix} \quad (2)$$

이 방식은 전체 D차의 파라미터들 중에서 임의의 n 차 파라미터에 대한 다중 상관관계 계수들을 모두 계산하고, 그 중 가장 큰 값을 갖는 파라미터들을 선택하여 하나의 부분공간으로 구성한다[5].

2.3 기존 부분 공간 구성 방법의 문제점

기존의 상관관계를 이용하여 부분 공간을 구성하는 방법에서는 전역 파라미터(global parameter)들 사이의 상관관계를 계산하게 된다. 하지만 MFCC 파라미터는 상관관계가 없다는 가정하에 구해진 값이기 때문에 이 값들 사이의 상관관계를 따진다는 것은 모순이 될 뿐 아니라 이 값은 실제로 의미를 갖지 못한다.

또한 기존의 부분 공간을 구성하는 방법에서는 부분 공간들의 차원을 임의로 고정시키게 되는데, 파라미터의 분포나 특성에 따라서 부분 공간의 차원이 유동적으로 결정되는 것이 임의로 고정된 경우보다 양자화 오류가 더 작게된다. 따라서 부분 공간의 차원을 모두 동일하게 하거나 임의로 정하지 않고 양자화 오류가 작아지

도록 각 부분공간의 차원을 결정해 줄 필요가 있다.

3. 양자화 오류를 고려한 부분공간 구성

3.1 부분공간과 벡터 양자화

벡터 양자화에 있어서 벡터를 부벡터(subvector)로 나누는 다음, 각 부벡터 별로 양자화 하는 경우가 전체 벡터를 양자화 하는 경우보다 양자화 오류가 더 작다 [6][7]. 이 때, 부벡터가 나누어진 결과에 따라 양자화 오류도 달라지게 되므로 부분 공간을 나누어 주는 기준이 중요한 요소가 된다. 먼저 가우시안 분포를 부분공간으로 나누어 준 다음 양자화 되는 가정을 살펴보면 그림 1. 과 같다. 이 과정에서 벡터 양자화 오류가 최소가 되기 위해서는 벡터의 각 차의 값들이 비슷한 분포를 가져야 한다[8]. 그림 2의 경우를 보면 (a)의 경우보다 (b)의 경우에 더 양자화 오류가 적음을 알 수 있다.

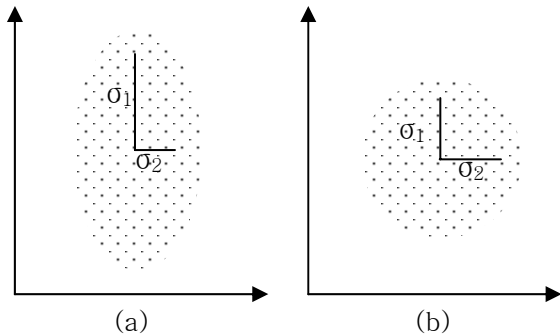


그림 2. 분산에 따른 양자화 오류

따라서 전체 가우시안 분포를 부벡터로 나눌 때 각 차의 평균값들의 분산이 비슷한 차원들을 하나의 부벡터로 만들어 주어야 한다. 즉 부분 공간을 나눌 때 가우시안 분포의 각 차원의 평균값들에 대한 분산의 비가 비슷한 차원들을 같은 부분공간으로 해 주도록 한다.

3.2 평균값의 분산에 따른 부분공간 구성

앞에서 본 것과 같이 평균값들의 분포를 고려하여 부분공간을 나누어 주기 위해서 agglomerative clustering algorithm[9]을 이용하였다. 각 부분공간의 차원은 군집화의 결과에 따라 자동으로 결정된다. 군집화 알고리즘에서 사용되는 거리 척도는 식 (3)과 같이 가우시안 분포의 각 차원의 평균값의 분산의 비로 정의해 주었다.

$$\text{Distance} = \left| \log \frac{\sigma_i}{\sigma_j} \right| \quad (3)$$

분산의 차이를 비교하는데 있어서 전체 파라미터들이 퍼져 있는 영역의 차이를 고려하기 위하여 두 분산의 비를 사용하였으며, 이 값이 0에 가까울수록 두 분포들은 비슷한 분산 값을 갖는다. 제안한 거리 척도를 이용하여 부분 공간을 자동적으로 구성 하는 과정은 다음과 같다.

- 각 차에 대해서 평균값들의 분산의 log값을 계산
- $\log \sigma_1, \dots, \log \sigma_D$ 를 agglomerative clustering algorithm을 적용하여 k개의 군집으로 구성
- 군집화의 결과를 이용하여 부분 공간 구성

이와 같은 과정을 통하여 D차의 전체 공간으로부터 k개의 부분 공간이 생성된다. 또한 각 부분 공간의 차원도 군집화 알고리즘의 결과에 따라 자동적으로 결정된다.

3.2 코드북 사이즈의 감소

SDCHMM은 기존의 CHMM에 비하여 적은 메모리를 사용하여서 모델을 표현 할수 있다. 그러나 SDCHMM에서 코드북이 차지하는 메모리의 비중이 매우 크기 때문에 코드북의 사이즈를 줄여줌으로써 더 많은 메모리의 이득을 얻을 수 있다. 본 논문에서는 코드북 사이즈의 감소를 위하여 다단계 벡터 양자화(Multistage Vector Quantization: MSVQ) 를 부분 공간의 코드북에 적용하였다[10]. MSVQ는 코드북을 두 단계로 나누어 첫 단계 코드북에서의 에러를 두번째 단계에서 양자화 해 줌으로써 코드북의 메모리를 효율적으로 감소시켜준다.

4. 실험 및 결과

4.1 실험환경

본 논문에서 제안한 부분 공간 구성 방법의 유효성을 검증하기 위해서 기존의 부분 공간 구성 방법과 제안한 방법을 비교 실험 하였다. 특징 벡터로는 12차 MFCC 계수와 에너지, 각 파라미터의 차분과 가속 파라미터등 39차 계수를 사용하였다. RM (Resource Management) DB 를 사용하여 인식 실험을 수행하였으며, 초기의 CHMM 모델을 구성하기 위해 HTK 3.1을 사용하였다.

초기 CHMM 모델은 ‘State Clustered Tri-phones’을 사용하였으며, 총 state의 수는 1578개이며 가우시안 분포 (Gaussian mixture)는 state당 3개를 사용하였다.

부분공간의 수는 실험을 통해서 메모리 감소와 인식률의 성능비로 가장 적합한 13개로 정해주었으며 SDCHMM에서의 코드북 사이즈는 부백터의 차원이 3차인 경우 최적인 256으로 정해 주었다. 또한 MSVQ를 적용한 실험에서는 첫 단계에 6bit, 두번째 단계에 2bit 씩 각각 할당해 주었다.

4.2 실험결과

실험 결과는 표1과 같다. 기존의 상관관계를 이용한 방법에 비해 제안한 평균의 분산을 이용해 부분 공간을 구성한 방법이 좀 더 양자화 오류를 줄여서 단어 오류율이 감소함을 볼 수 있다.

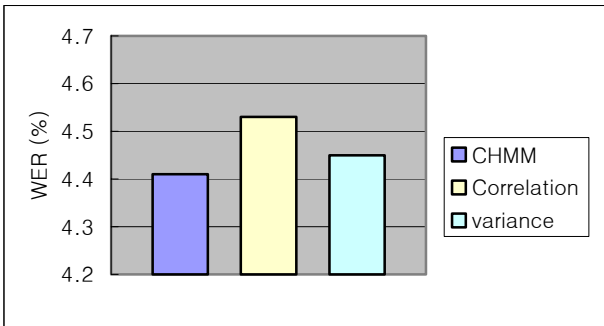


표1. 부분공간 구성법에 따른 메모리 사용량과 WER

또한 표2에서 볼 수 있듯이 MSVQ를 적용하여 코드북을 구성한 결과 기존의 SDCHMM을 사용하여 코드북을 구성한 경우에 비해서 약46%의 메모리 감소를 얻을 수 있었으나, 코드북의 사이즈가 줄어든 영향으로 약간의 단어오류율을 증가가 생겼다.

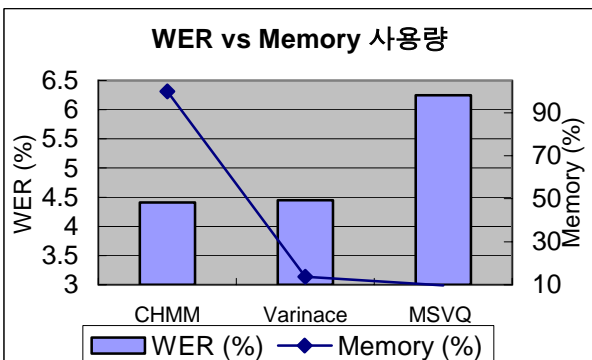


표2. 코드북 구성 방법에 따른 메모리 감소량과 WER

5. 결론

본 논문에서는 SDCHMM에서 부분 공간을 효율적으로 구성하기 위해서 가우시안 분포들의 평균 값에 대한 분산을 이용하여 부분 공간구성하고 코드북 저장을 효율적으로 할 수 있는 방법을 제안하였다. 제안한 방법으로 부분 공간을 구성한 경우 기존의 상관관계를 이용한 방법에 비해서 1.8%의 단어오류율(WER)의 감소를 얻을 수 있었다. 또한 MSVQ를 적용하여 코드북을 구성한 결과 기존의 SDCHMM에 비해 46%, CHMM에 비해 86%의 메모리 감소를 얻을 수 있었다.

현재는 부분공간의 차원에 상관없이 모두 동일한 코드북 사이즈를 사용하고 있지만 부백터의 차원에 따라서 최적의 코드북 사이즈가 달라지므로, 제안한 분산에 따른 부분공간 구성법을 적용하는데 있어서 각 부분공간별로 코드북 사이즈를 달리 하여줄 필요가 있고 이때 부분 공간의 차원에 따른 최적의 사이즈를 앞으로 결정해 주고자 한다.

참고문헌

1. Juan M. Huerta, "Speech Recognition in Mobile Environments", Ph.D Thesis, CMU, 2000.
2. Olli Viikki, "Speaker-and Language-Independent Speech Recognition in Mobile Communication Systems", Proc. of ICASSP, 2001
3. Brian Mak, "Stream derivation and clustering scheme for subspace distribution clustering hidden markov model", Proc. of the IEEE Automatic Speech Recognition and Understanding Workshop, 1997
4. Tommi Lahti, "Low Memory Acoustic Models for HMM Based Speech Recognition", Eurospeech, 2003
5. E. Bocchieri, "Subspace Distribution clustering Hidden Markov Model", IEEE Trans. Speech and Audio Processing, vol. 9, no. 3, March 2001
6. M. Ravishankar, "Sub-Vector Clustering To Improve Memory and Speed Performance of Acoustic Likelihood Computation", Eurospeech, Vol. 1, 1997
7. J. Kim, "Hidden Markov Models with Divergence Based Vector Quantized Variances", Proc. of ICASSP, 1999
8. B.H.Juang,, "Distortion performance of vector quantization for LPC voice coding, IEEE Trans.Acoust., Speech, Signal Processing, vol.ASSP-30, April 1982
9. E. M. Voorhees, "Implementing agglomerative hierarchical clustering algorithms for use in document retrieval", Information Processing & Management, no. 22, 1986
10. W.-Y. Chan, "Enhanced multistage vector quantization by joint codebook design", IEEE Trans. Communications, vol. 40, Nov. 1992