

분산의 군집화를 이용한 SDCHMM의 부분 공간 구성

김수현, 정규준, 오영환

한국과학기술원

A subspace decision method using variance clustering in SDCHMM

Su-hyeon Kim, Gue-Jun Jung, 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

음성인식에 주로 사용되는 HMM은 모델을 표현하기 위하여 많은 양의 메모리를 필요로 한다. 그러나 학습된 모델의 분포를 부분 공간으로 구성한 후 군집화하여 사용하는 SDCHMM (Subspace Distribution Clustering HMM)의 경우 모델 표현을 위한 메모리를 30%정도로 줄이면서도 CHMM과 유사한 성능을 낼 수 있다. 이런 SDCHMM에서는 전체 공간의 가우시안 분포를 어떻게 부분 공간으로 나눌 것인가가 중요한 문제가 된다. 기존의 연구에서는 파라미터들을 개념적으로 유사한 것으로 나누어 주는 방법과, 파라미터들 간의 상관관계가 높은 것들을 하나의 부분 공간으로 구성하는 방법이 있었지만, 이 두 방법 모두 코드북을 만드는데 있어서 양자화 오류를 크게 줄여주지 못한다. 본 논문에서는 코드북을 구성하는 가우시안 분포들의 평균값들에 대한 표준 편차를 이용하여 양자화 오류가 최소가 되도록 부분 공간을 나누는 방법을 제안한다. 제안된 방식의 유효성을 검증하기 위하여 고립단어 인식 실험을 한 결과 기존 방법에 비해 14.7%의 단어오류율(WER)의 감소를 보였다.

1. 서론

휴대폰이나 PDA 등과 같은 이동통신 기기의 사용량이 늘어나면서 제한된 환경에서의 음성 인식에 대한 필요성이 점차 증가되고 있다. 음성인식기술은 입력신호

에서 추출한 특징벡터와 이 특징벡터의 시간에 따른 변화추이를 확률 모델로 표현하고 활용한다. Hidden Markov Model (HMM)은 이러한 두 가지 확률 변수를 동시에 성공적으로 표현한 모델로써 음성인식기술의 상용화에 큰 기여를 하였다. 하지만 HMM을 이용한 음성인식은 방대한 연산과 메모리를 필요로 하기 때문에 휴대폰, PDA와 같은 자원이 제한적인 환경에 활용하기 어렵다. 이러한 제약을 극복하기 위해 제안된 Subspace Distribution Clustering Hidden Markov Models (SDCHMM)은 HMM을 구성하고 있는 확률 분포들이 표현하고 있는 공간을 부분공간(subspace)으로 분할하고 해당 부분 공간에 포함된 부분확률 분포들 중 비슷한 분포들을 tying하여 양자화 한다[1]. SDCHMM은 기존 HMM에 비해 모델 표현에 필요한 메모리 공간을 최소화 하면서도 모델의 표현 능력을 유지시킬 수 있기 때문에 음성인식 모델을 이동통신 기기에 적합하도록 구성할 수 있다[2].

이러한 장점들을 가진 SDCHMM은 확률 분포에서 부분공간을 어떻게 구성하는가 하는 것이 중요한 문제가 된다. 기존 연구에서 특징 파라미터로부터 부분공간을 구성하는 방법은 크게 개념적으로 유사한 파라미터들을 부분 공간으로 구성해 주는 방법과, 전체 파라미터들의 상관관계를 이용하여 부분 공간을 구성하는 방법이 있었다[3][4]. 하지만 이 두 가지 방법 모두 임의의 고정된 값을 부분공간의 차원으로 사용하므로 최적의 성능

을 나타내는데 한계가 있다. 따라서 본 논문에서는 양자화 오류를 줄이면서, 각 부분공간의 차원을 자동으로 결정하는 방법을 제안한다. 2장에서는 SDCHMM의 개요 및 기존 부분 공간 구성방법에 관해서 살펴보고, 3장에서는 본 논문에서 제안한 방법에 대해 설명한다. 4장에서는 기존 방법과 제안한 방법을 적용한 실험 결과를 제시하고, 5장에서 결론을 맺는다.

2. SDCHMM의 개요 및 부분 공간 구성법

2.1 SDCHMM의 개요

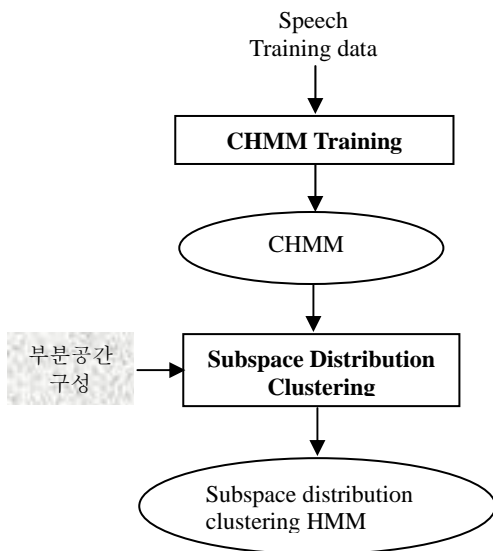


그림 1. CHMM의 SDCHMM으로의 변환

SDCHMM은 그림 1과 같이 CHMM을 재학습 시키지 않고 CHMM에 직교하는 부분 공간(orthogonal-subspace)을 이용하여 구성할 수 있다[1]. 이 때 비슷한 형태의 부분 공간 분포들을 분포의 정형(prototype)으로 tying한다. 따라서 원래의 전체 공간 분포는 식(1)과 같이 부분공간 분포의 정형들의 조합으로써 원래 분포 값에 가깝게 추정된다.

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

식에서 m은 가우시안 분포, s는 상태의 인덱스를 나타내고, K는 부분공간의 수를 나타낸다.

2.2 SDCHMM의 부분 공간 구성

2.2.1 개념적 유사성을 이용한 방법

특징 파라미터 중 개념적으로 유사한 요소들을 하나의 부분공간으로 만들어 주는 방법이다. 예를 들어 39차 MFCC(Mel Frequency Cepstral Coefficient) 파라미터의 경우, 전체를 하나의 스트림으로 나누어 줄 경우 전체 공간을 하나의 부분 공간으로 구성해 주고, 13개의 스트림으로 나누는 경우는 12개의 MFCC 각 차수 또는 에너지와 그것의 차분(delta), 가속(acceleration) 파라미터들로 이루어진다[3]. 하지만 이 방법은 파라미터의 실제 값과 관련된 특성을 반영하지 못하므로 양자화 오류를 최소화 하지 못한다.

2.2.2 상관관계를 이용한 방법

이 방법은 전체적인 특징 파라미터들 사이의 상관관계(correlation)를 이용한다. 이때 2개 이상의 파라미터들 사이의 상관관계를 정의하기 위하여 다중 상관관계 계수(Multiple-correlation coefficient) R을 식(2)와 같이 정의한다. 전체 파라미터의 차원이 D이고 n개부분공간으로 나눌 때, 각 부분 공간은 k차원의 부분 공간으로 구성된다.

$$R = 1 - \begin{vmatrix} 1 & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{21} & 1 & \cdots & \rho_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{k1} & \rho_{k2} & \cdots & 1 \end{vmatrix} \quad (2)$$

전체 D개의 파라미터 중에서 k개를 선택하는 모든 경우에 대해서 다중 상관관계 계수를 구한 다음 이 값이 최대가 되는 순서대로 파라미터 집합을 하나의 부분공간으로 구성한다[4]. 이렇게 하여 다중 상관관계 계수가 가장 크게 되는 값들끼리 부분 공간을 구성한다.

하지만 MFCC와 같은 캡스트림 영역의 파라미터는 상관관계가 거의 없다는 가정하에서 구해진 값이므로, 부분 공간을 구성하는 데 있어서 파라미터들 사이의 상관관계는 무시할 수 있다. 또한 이 방법은 부분공간 구성이 가능한 모든 경우에 대해서 상관관계 값을 구해야 하므로 계산 양도 많다는 단점이 있다.

3. 표준 편차 군집화와 부분공간 구성

3.1 평균의 표준편차와 전역 공분산

스칼라 양자화의 경우 스칼라 값들이 단일 분포(Uniform distribution)를 따를 경우 양자화 오류가 최소화

된다[5][6]. 이러한 사실을 벡터 양자화에도 적용하였다. 하지만 HMM은 파라미터 값들이 가우시안 분포를 따른다는 가정을 갖고 있기 때문에, 가우시안 분포들이 단일분포에 가장 가깝게 분포할 때 양자화 오류가 가장 적다고 할 수 있다. 따라서 파라미터들 사이에 상관 관계가 없고 각 축에 대해서 퍼져있는 정도가 비슷한 경우에 양자화 오류가 최소가 된다. 예를 들어 그림 2의 경우 (b)가 (a)의 경우보다 양자화 오류가 더 작다.

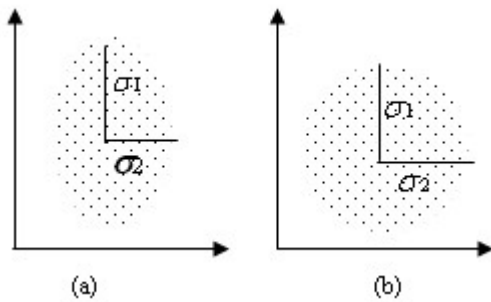


그림 2. 각 축에서 값이 퍼진 정도

따라서 본 논문에서는 각 모델을 이루고 있는 첵스트럼(Cepstrum) 파라미터들의 각 차수에서의 평균 값들 사이의 표준편차의 차이가 가장 적은 값들을 하나의 부분공간으로 구성한다. 본 논문에서 제안한 평균값의 표준 편차의 군집화에는 전체 HMM 모델의 상태들이 이루고 있는 확률 분포의 평균값들에 대한 표준편차를 사용한다. 또한 인식단계에서는 각 상태를 구성하는 가우시안 분포들의 대각 공분산(diagonal covariance)을 사용하는 대신에 전체 첵스트럼 파라미터에 대한 전역 대각 공분산(global diagonal covariance)을 사용하므로 코드북은 각 차원의 평균값을 양자화 한 벡터들로만 이루어진다.

3.2 군집화를 통한 부분공간 분포구성

파라미터의 각 차원 내에서 확률 분포의 평균 값들의 분포 범위가 비슷하면, 파라미터의 값들이 각 차원의 축에서의 퍼진 정도가 비슷해진다. 따라서 값들의 퍼진 정도, 즉 표준 편차의 차이를 이용하여 자동적으로 주어진 수의 스트림을 결정한다. 벡터 양자화에 사용되는 코드북은 각 차수에 대한 가우시안 믹스처들의 평균 값으로 구성되기 때문에 전체 데이터의 표준편차를 사용하지 않고 평균값에 대한 표준편차를 사용한다.

각 차원의 가우시안 분포들 사이의 거리를 측정하기 위하여 식(3)과 같은 거리 척도를 사용하였다.

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

식 (3)을 이용하여 계산한 값이 0에 가까울수록 두 분포들은 비슷한 표준 편차 값을 갖는다. 표준 편차의 차이를 비교하는데 있어서 전체 파라미터들이 퍼져 있는 영역의 차이를 고려하기 위하여 두 표준 편차 값의 비를 사용하였으며, 표준 편차 비의 log 값을 구함으로써 유클리드 거리에서의 차로 바꿀 수 있다. 따라서 각 파라미터의 표준편차의 log값에 대해서 절대값 차이가 가장 작은 값들을 하나의 부분 공간으로 구성해 줄 수 있다. 제안한 거리 척도를 이용하여 부분 공간을 자동적으로 구성 하는 과정은 다음과 같다.

- 각 차원에 대해서 표준편차의 log값을 계산
- $\log \sigma_1, \dots, \log \sigma_D$ 를 agglomerative clustering algorithm[7]을 적용하여 n개의 군집으로 구성
- 군집화의 결과를 이용하여 부분 공간 구성

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

4. 실험 및 결과

4.1 실험 환경

제안한 부분 공간 구성 방법의 효율성을 검증하기 위하여 제안한 방법과 기존의 부분 공간 구성 방법을 비교 실험하였다. 특징 벡터로는 0차를 포함한 13차 MFCC 계수와 차분, 가속 파라미터 등 총 39차 계수를 사용하였으며, IWR 75DB (phonetically-balanced 75Korean isolated word database)에서 7500개의 발화를 학습에 사용하였고, 2218개의 발화를 인식에 사용하였다. 초기 CHMM 구성을 위해 HTK3.1을 사용하였다.

4.2 실험 결과

적절한 부분 공간의 수를 결정하기 위하여 제안한 부분 공간 구성 방법을 이용하여 부분 공간의 수(메모리 감소량)에 따른 인식률 변화에 관한 실험을 수행하였다.

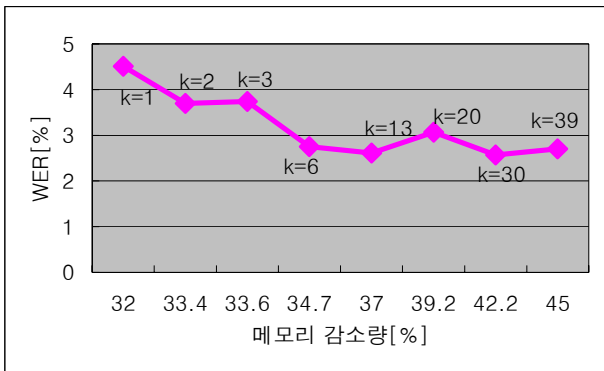


그림 3. 메모리 감소 따른 인식률 (k: 부분 공간의 수)

그림 3에서와 같이 하나의 전체 공간을 사용한 것보다 부분공간의 수가 많아질수록 단어오류율(WER)이 감소하였다. 하지만 일정 수준 이상에서는 그다지 단어오류율의 감소가 없다. 따라서 본 실험에서 성능 비교를 위해 적절한 부분 공간의 수는 13으로 정해 주었다.

다음으로 기존의 부분 공간 구성 방법과 제안한 표준편차를 이용한 부분 공간 구성 방법의 비교 실험을 수행하였으며, 결과는 그림 4과 같다. 상관관계가 양자화 오류에 주는 영향을 확인하기 위하여 상관관계가 높은 값들끼리 부분 공간을 구성하는 기존의 방법 외에 상관관계가 낮은 값끼리 부분 공간을 구성한 실험도 수행하였다. 그 결과 상관관계는 양자화 오류에 큰 영향을 주지 못함을 확인하였다. 또한 기존의 개념적 유사성을 이용한 방법과 상관 관계를 이용한 방법에 비하여 모두 제안한 방법이 양자화에 따른 오류를 최소화 하여 단어오류율이 감소하였다.

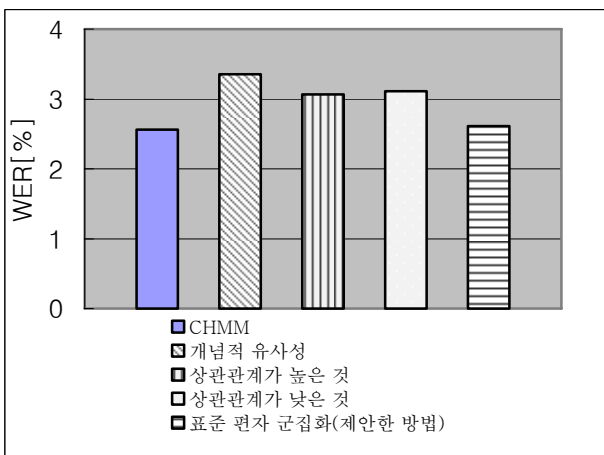


그림 4. 부분 공간 구성방법에 따른 인식률

5. 결론

본 논문에서는 SDCHMM에서 부분 공간을 효율적으로 구성하기 위하여 전체 가우시안 분포들의 평균값에 대한 표준편차를 이용하여 부분 공간을 나누는 방법을 제안하였다. 부분공간을 군집화 할 때 양자화 오류가 최소가 되는 경우에 최적의 부분 공간이 구성이 된다. 제안한 방법으로 부분 공간을 구성한 경우 기존의 상관관계를 이용한 방법보다 14.7%의 단어오류율(WER)의 감소를 보였다. SDCHMM이 CDHMM에 비하여 메모리를 30~40%가까이 사용하고도 크게 성능이 떨어지지 않지만, 실제로 휴대용 기기에 사용되기 위해서는 메모리 사용량을 좀 더 줄일 필요가 있으므로 현재 고정되어 있는 코드북 사이즈를 좀 더 줄일 수 있는 방법에 대한 향후 연구가 필요하다. 또한 현재 제안한 내용을 검증하기 위해 사용한 task가 제한적이므로, 다양한 실험 자료를 통해 성능평가가 필요하다.

참고문헌

1. Enrico bocchieri, Brian Kan-Wing Mak,, "Subspace Distribution Clustering Hidden Markov Model" , IEEE Trans. Speech Audio Processing, vol.9, No3, March 2001
2. Imre Varga, Stefanie Aalburg, "ASR in Mobile Phones--An Industrial Approach", IEEE Trans. Speech and Audio Processing, vol.10, No8, November 2002.
3. E.Bocchieri, B.Mak, "Subspace Distribution Clustering for Continuous Observation Density Hidden Markov Models", In Proc.ofEurospeech, volume 1, 1997.
4. Brian Mak, Enrico Bocchieri "Stream derivation and clustering scheme for subspace distribution clustering hidden markov model", in Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop, 1997
5. Antonio Ortega, Martin Vetterli, "Adaptive Scalar Quantization Without Side Information", IEEE Trans. On Image Processing, vol.6, no 5, May 1997
6. B.H.Juang,, D.Y.Gray, A.H.Gray, Jr, "Distortion performance of vector quantization for LPC voice coding, IEEE Trans.Acoust., Speech, Signal Processing, vol.ASSP-30, April 1982
7. E. M. Voorhees, "Implementing agglomerative hierarchical clustering algorithms for use in document retrieval", Information Processing & Management, no. 22, 1986.