

# 발산거리 기반의 신경망에 의한 가우시안 확률 밀도 함수의 군집화

정회원 박 동 철\*, 권 오 현\*\*

## Guassian pdfs Clustering Using a Divergence Measure-based Neural Network

Dong-Chul Park\*, Oh-Hyun Kwon\*\* *Regular Members*

### 요 약

음성인식 모델상의 GPDFs(Gaussian Probability Density Functions)을 효율적으로 군집화할 수 있는 알고리즘이 제안되었다. 제안된 알고리즘은 데이터 사이의 거리 척도로 발산 거리를 사용하는 새로운 형태의 CNN(Centroid Neural Network)으로, 제한된 자원을 가지는 H/W환경의 음성인식에서 메모리 사용량을 축소하는 응용에 대한 실험 결과, 음성인식 모델인 CDHMM(Continuous Density Hidden Markov Model)에서 기존의 Dk-means(Divergence-based k-means)알고리즘을 이용한 방법과 비교하여 인식 성능의 유지와 함께 약 31.3%의 GPDFs를 더 축소할 수 있었고, 군집화 알고리즘을 적용하지 않은 전체 GPDFs를 사용한 경우와 비교해서 인식 성능의 유지와 함께 약 61.8%의 GPDFs를 압축할 수 있었으며, SNR 10dB 잡음 데이터에 대한 성능평가에서도 인식 성능이 유지될 수 있었다.

Key Words : CNN, Dk-means, GPDF, memory size

### ABSTRACT

An efficient algorithm for clustering of GPDFs(Gaussian Probability Density Functions) in a speech recognition model is proposed in this paper. The proposed algorithm is based on CNN with the divergence as its distance measure and is applied to a speech recognition. The algorithm is compared with conventional Dk-means(Divergence-based k-means) algorithm in CDHMM(Continuous Density Hidden Markov Model). The results show that it can reduce about 31.3% of GPDFs over Dk-means algorithm without suffering any recognition performance. When compared with the case that no clustering is employed and full GPDFs are used, the proposed algorithm can save about 61.8% of GPDFs while preserving the recognition performance.

### I. 서 론

가전 제품, 통신용 단말기, 자동차, 원구 등에 쉽게 음성인식 기술을 적용할 수 있을 만큼 임베디드 음성인식 기술이 점차 개발되면서, 활발한 움직임을 보이고 있는 음성인식 시장에서는 성능에 대한 영

향을 최소화하며, 메모리 사용량을 최소화시키는 기술에 대한 연구가 진행되어 왔다.

Shinoda등은 MDL(Minimum Description Length)기준을 기본으로 한 가우시안 트리를 이용하여 가우시안 요소(Gaussian Component)의 하위 집합들을 선택함으로써 CDHMM상의 가우시안 확률

\*, \*\* 명지대학교 정보공학과 지능컴퓨터연구실

논문번호 : 030550-1211, 접수일자 : 2003년 12월 11일

※본 연구는 한국과학재단 지역대학우수과목자 지원연구(R05-2003-000-10992-0)의 지원으로 수행되었음.

밀도 함수(GPDFs)를 군집화 하였으며, 그 결과, 축소된 형태의 모델을 획득 하였다<sup>[1]</sup>. Watanabe등은 기존의 k-means 알고리즘과 분포사이의 거리 척도로 발산 거리 척도(Divergence Distance Measure)를 이용하여 GPDFs를 축소 하였으며<sup>[2]</sup>, L.Rigazio등은 분포사이의 거리 척도로 Bhattacharyya Measure를 이용한 최적 Bhattacharyya 중심값 알고리즘에 의해 GPDF를 축소하였다<sup>[3]</sup>. 하지만, GPDFs를 축소하고자 하는 이러한 노력에도 불구하고, 휴대 전화기와 같이 자원이 제한된 H/W 환경에서 일정 비율 이상의 GPDFs 압축은 인식 성능 저하를 피할 수 없어 그 실용성에 다소 제한적일 수밖에 없었다.

본 논문에서는 제한된 자원을 가지는 H/W 환경에서 인식 성능의 유지와 함께 메모리 사용량을 최소화할 목적으로 기존의 무지도 경쟁학습 알고리즘 중 다차원 공간상의 점들에 대한 군집화에 좋은 성능을 보였던 CNN(Centroid Neural Network)을 기반으로 다차원 공간상의 분포들에 대한 군집화 알고리즘이 제안되었다<sup>[4]</sup>. 제안되는 알고리즘은 초기 연결강도 설정 문제, 초기 학습 계수 설정 문제, 전체 Epoch 수 설정 문제들을 효과적으로 해결함으로써 자동적으로 군집화 시키는 CNN의 장점을 이용하면서, 동시에 음성 데이터들의 중요한 특징인 GPDFs로 구성된 데이터들 중 인식에 많은 영향을 미치는 밀집되어 분포하는 데이터들의 영역에 더 많은 연결강도를 배정하여 효율적으로 군집화 함으로써 메모리 사용량을 최소화함과 함께 인식 성능 유지에 적합한 알고리즘이다.

본 논문의 2장에서는 군집화 하려는 대상 데이터인 GPDFs의 구조와 GPDFs사이의 거리 척도로 사용된 Divergence Measure에 대해 살펴보았고, 3장에서는 알고리즘에 대한 설명이, 그리고 4장에서는 실험 과정과 결과가 주어지며, 본 논문의 결론은 5장에서 내려진다.

## II. GPDFs와 발산 거리 척도

본 논문에서 군집화 하려는 데이터의 형태는 일반적인 다차원 공간상의 점으로 표현되어지는 정보가 아니라, CDHMM상의 평균( $\mu$ )과 분산( $\sigma^2$ ) 정보로 구성된 다차원 공간상의 GPDFs이다. 그림 1에 음성인식에서 사용하는 단일 음소 모델(Mono Phoneme Model)의 한 예를 표현했다. 여기서  $w$ 는 Mixture Weight를 나타내고,  $\mu, \sigma^2$ 가 군집화 대상

인 GPDFs의 정보이다. 군집화 알고리즘에서 군집화 하려는 데이터 사이의 유효한 거리 측정은 군집화 성능을 좌우할 정도로 매우 중요한 부분이므로 보다 더 유효한 거리 척도가 필요하다<sup>[5,6]</sup>. 분포 사이의 거리 측정 방법으로 다음의 식 (1)에 정의되는 발산 거리 척도가 좋은 성능의 결과로 많이 사용된다<sup>[2,5,6]</sup>.

$$D(I_i, W_j) = \sum_{k=1}^d \left( \frac{\sigma_{I_i}^2(k) + (\mu_{I_i}(k) - \mu_{W_j}(k))^2}{\sigma_{W_j}^2(k)} + \frac{\sigma_{W_j}^2(k) + (\mu_{I_i}(k) - \mu_{W_j}(k))^2}{\sigma_{I_i}^2(k)} \right) \quad (1)$$

위에서  $\mu_{I_i}(k), \sigma_{I_i}^2(k)$ 는 각각  $n$ 번째 입력 GPDFs의  $k$ 번째 차원의  $\mu, \sigma^2$ 를 나타내며,  $\mu_{W_j}(k), \sigma_{W_j}^2(k)$ 는 각각  $j$ 번째 연결강도 GPDFs의  $k$ 번째 차원의  $\mu, \sigma^2$ 를 나타내며,  $d$ 는 데이터의 차원을 나타낸다. 식 (1)에서도 알 수 있듯이 발산 거리 척도는 분포의 정보인  $\mu, \sigma^2$ 를 모두 사용하여 거리를 측정한다. 만일 기존의 Euclidean 거리를 GPDFs사이의 거리 척도로 사용한다면, 각 GPDFs의  $\mu$ 와  $\mu$ 사이의 거리를 측정할 수 밖에 없으며, 이는 유효한 GPDFs사이의 거리 측정방법으로 사용되기에는 문제가 있기 때문에 GPDFs사이의 거리 측정 방법으로 발산 거리 척도를 널리 사용한다<sup>[2,5,6]</sup>.

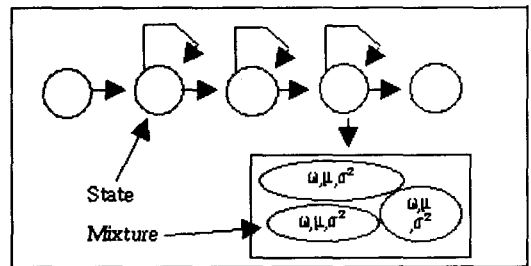


그림 1. CDHMM 단일 음소 모델 구조의 예 (5개의 State와 각 State당 3개 Mixture.)

### III. 발산척도기반의 CNN

군집화 알고리즘은 주어진 데이터 사이의 유사성을 검출하여 군집화 시키는 경쟁학습 알고리즘으로 유사성이 높은 데이터들끼리 군집화 시키는데 목적이 있다. 본 논문에서 제안되는 알고리즘은 CNN을 기본으로 구성되었다. 기존의 CNN은 승자, 패자의 연결강도의 갱신 방법에 의해 지역적으로 최적의 연결 강도를 설정함으로 주어진 데이터를 최적의 대체할 수 있는 대체 값(Weight)으로 설정하는데 탁월하며, 사전에 학습 계수나 전체 epoch수를 설정할 필요가 없으며, 더불어 기존의 군집화 알고리즘에서 문제점으로 대두 되었던 초기 연결강도 설정의 문제까지 효과적으로 해결함으로 자동적으로 최적의 군집화를 이끄는 장점이 있다<sup>[5]</sup>.

제안되는 알고리즘인 CNN-DM(CNN with Divergence Measure)은 다차원 공간상의 분포로 이뤄진 데이터를 군집화 하는데 사용되는 알고리즘이므로 기존의 Euclidean 거리대신 위의 식 (1)에서 주어진 발산 거리 척도를 사용하여 다차원 공간상의 분포들간의 거리를 최소화 시키는 연결 강도를 승자 연결 강도로 선택한다<sup>[5]</sup>. 다차원 공간상의 GPDFs을 군집화 하기 위해서는  $\mu, \sigma^2$ 에 대해 각각의 연결 강도를 갱신 해야 한다. 식(2),(3)은 승자와 패자에 대해 각각  $\mu$ 에 대한 연결강도 갱신 방법이다. 이는 CNN에서와 같이, 승자에 대한 연결강도의  $\mu$ 는 입력 데이터의  $\mu$ 방향으로 갱신되며, 패자에 대한 연결강도의  $\mu$ 는 입력데이터의  $\mu$ 의 반대 방향으로 갱신하여 최적의 지역 군집화 성능을 창출한다.

$$\mu_{W_j}(k+1) = \mu_{W_j}(k) + \frac{\mu_{I_n}(k) - \mu_{W_j}(k)}{N_j + 1} \quad (2)$$

$$\mu_{W_i}(k+1) = \mu_{W_i}(k) - \frac{\mu_{I_n}(k) - \mu_{W_i}(k)}{N_i - 1} \quad (3)$$

식 (4)에 표현된  $\sigma^2$ 에 대한 연결강도 갱신 방법은  $\sigma^2$ 의 수학적 의미로 갱신된다<sup>[2,3]</sup>.

$$\sigma_{W_j}^2(k+1) = \frac{\sum_{n=1}^{N_j} \sigma_{I_n}^2(k) + (\mu_{I_n}(k) - \mu_{W_j}(k))^2}{N_j} \quad (4)$$

이때  $\mu_{I_n}(k), \sigma_{I_n}^2(k)$ 는 각각  $j$ 군집에 속한  $n$ 번째 입력 GPDFs의  $k$ 번째 차원의  $\mu, \sigma^2$ 값을 나타내고,  $N_j$ 는  $j$ 군집에 속한 입력 GPDFs수를 나타낸다.

표 1은 제안된 알고리즘의 pseudo-code 이다.

### IV. 실험과 결과

본 논문에서 제안되는 CNN-DM은 CDHMM에서의 26차원 GPDFs 데이터의 군집화에 응용되었다. 사용되는 데이터와 실험 내용은 표2에 요약되었다. 실험은 CNN-DM을 사용하여 어느 정도까지 CDHMM에서의 GPDFs 수를 인식 성능에 영향을 미치지 않으면서 줄일 수 있는지, 기존의 방법인 Dk-means 알고리즘을 사용하는 것과는 어떠한 성능의 차이를 보이는가에 대한 검증의 과정이 수행되었다. 즉, 실험의 결과는 DCNN과 같은 발산 거리 척도를 이용하는 기존의 k-means 알고리즘인 Dk-means 알고리즘과 비교 하였으며, 1,440개의 전체 GPDFs를 쓰는 인식의 결과와 Dk-means 와 CNN-DM에 의해 재구성된 여러 가지 크기의 코드북을 채용한 인식성능을 비교 검토 하였다.

표 1. CNN-DM 의 수행코드

```

Procedure main()
  Read M, N (M:cluster 수, N: data 수)
  Initialize cluster weights for mean and variance
  k:=2, epoch := 0
  for (k <= M)
    while (input file is not empty)
      Apply a data to the network
      Find winner neuron or loser neuron
      Update weights of winner/loser for
        mean and variance by Eqs.
        (2)-(4)
    endwhile
  
```

```

if  $k \neq M$ 
    split the most erroneous group by
    adding a small vector to the weight for
    the group
endif
 $k := k+1$ 
endifor
output cluster weights for mean and variance
end main()
    
```

표 2. 실험 정보

무잡음 DB	남자11명/약 200개의 한국인명/ 무잡음환경에서 발음/ 2188개의 음성 데이터.
잡음 DB	남자12명/약 200개의 한국인명/ SNR 10dB 환경에서 발음/2388개의 음성 데이터.
군집화 알고리즘	Dk-means, CNN-DM
군집의 개수	400, 500, 550, 600
HMM model	남자 300명/약 200개의 한국인명/ Left to Right HMM 모델링 한국어 41개 단일 음소 모델 (전체 1440개 GPDFs)

본 논문에서의 실험에는 두 가지의 인식기가 사용되었는데, 1) 1,440개의 전체 GPDFs를 이용하고, 무잡음 데이터에 의해 테스트된 CDHMM을 채용하는 인식기에서, Dk-means와 CNN-DM에 의해 재구성된 코드북을 이용하여 인식의 기능을 수행하는 부분만을 대체하는 Coded-인식기와 2) CNN-DM, Dk-means에 의해 생성된 코드북과 학습 데이터들을 이용하여 최적화 시키는 AP (Adaptation Process)를 거쳐 구성된 최적화된 코드북을 테스트할 Adapted-인식기이다. 이때, 사용된 Model Adaptation 방법은 ML (Maximum Likelihood)기준이다. 인식성능의 비교 검토를 위해, 무잡음 DB는 Coded-인식기와 Adapted-인식기에서 모두 동일하게 테스트 되었고, SNR 10dB 잡음 DB는 Adapted-인식기에서만 테스트되었다.

1,440개의 전체 GPDFs를 이용하여 구성된 CDHMM을 채용하고, 1,440개의 전체 GPDFs를 사용하여 인식을 수행하는 기본적 형태의 인식성능(기본 인식성능)에 대한 무잡음 DB와 SNR 10dB 잡음 DB에 대한 인식성능이 표 3에서 주어진다.

표 3. 기본 GPDFs의 인식성능

사용된 baseline GPDFs의 수	1,440
무잡음 DB	90.06 %
잡음 DB	81.36 %

표 4에서 볼 수 있듯이, CNN-DM은 3가지 종류 (코드북, Adaptation, SNR 10dB 잡음 환경)의 인식 성능의 결과에서 모두 기존의 Dk-means보다 더 좋은 인식 성능을 낸다는 것을 알 수 있었고, 기본 GPDFs에서 무잡음 DB 인식률인 90.06%를 달성하기 위한 최소 GPDFs의 수는 Dk-means의 경우는 약 800개, CNN-DM의 경우는 약 550개였다. 즉, 550개로 군집화된 GPDF 만으로도 기존의 방법에 이용되었던 Dk-means (800개)나 전체 GPDFs (1,440개)을 적용한 경우와 대등한 인식성능을 유지 하면서, 각각 31.3%, 61.8%의 군집 수를 줄이게 된다. 이는 실제 시스템에 적용시 GPDFs를 저장하는데 필요한 메모리의 크기를 획기적으로 줄이며, 또한 탐색에 필요한 응답시간의 대폭적 감소를 가져와 실시간 시스템에의 응용을 가능케 한다. 한편, 10dB 잡음 환경에서도 유사한 결과를 얻을 수 있어, 잡음에도 강한 것을 확인할 수 있었다.

표 4. Dk-means와 CNN-DM의 인식 성능(%) 비교

	코드북 size	400	500	550	600	800
Dk-means	Coded-인식기	84.8	86.5	87.3	88.20	89.0
	Adapted-인식기	87.4	88.2	88.5	89.1	90.0
	10 dB 잡음 환경	76.4	77.7	78.7	79.3	80.1
CNN-DM	Coded-인식기	87.6	88.8	89.1	89.3	89.4
	Adapted-인식기	89.1	89.9	90.0	90.1	90.1
	10 dB 잡음 환경	78.9	79.5	81.2	81.3	81.3

### V. 결론

음성인식 모델상의 GPDFs를 효율적으로 군집화할 수 있는 무감독 경쟁 학습의 신경망 알고리즘인 CNN-DM이 제안되었다. 제안된 CNN-DM은 CNN에 기반을 두며, CNN의 장점을 가지면서, 가우시안 분포를 따르는 데이터를 군집화하는 알고리즘이다. 제한된 자원을 가지는 임베디드용 음성인식에서 인식 성능 유지와 더불어 CDHMM모델 압축에 의한 사용 메모리량 축소에 초점을 두었던 CNN-DM은 여러가지의 군집 수에 따른 인식 성능 비교에서 Dk-means 알고리즘보다도 동일한 군집에서 우수한 인식 성능을 보였으며, 특히, 기존의 방법에 사용되었던 Dk-means(800개)에 비해서는 약 31.3%, 전체 GPDF(1,440개)에 대해서는 약 61.8% 더 압축된 모델로 대등한 인식 성능을 유지할 수 있었다. 또한, 압축된 모델을 이용하여 SNR 10dB 잡음 환경에서의 인식 성능 실험에서도 유사한 결과를 얻을 수 있어서, CNN-DM이 잡음환경에서도 강인함을 알 수 있었다. 따라서 제안되는 CNN-DM은 기존의 군집화 알고리즘에 비해 더 효율적으로 군집화 시킴으로 음성인식상 CDHMM모델의 압축에 의해 임베디드용 음성인식에서의 필요 메모리량을 축소시키며, 그로 인한 인식 시간을 대폭 줄일 수 있어 실시간 음성인식기의 구현에 매우 유용한 도구로 사용될 수 있다.

### 참고 문헌

[1] K. Shinoda. and K. Iso, "Efficient reduction of gaussian components using MDL criterion for HMM-based speech recognition", *Proc. ICASSP-02*, vol.1, pp 869-872, 2002.

[2] T. Watanabe, K. Shinoda and K. Takagi and K. Iso, "High speed speech recognition using tree structured probability density function", *Proc. ICASSP-95*, vol.1, pp 556-559, 1995.

[3] L. Rigazio, B. Tsakam and J. Junqua, "An optimal Bhattacharyya centroid algorithm for Gaussian clustering with applications in automatic speech recognition", *Proc. ICASSP-00*, vol.3, pp 1599-1602, 2000.

[4] Dong-Chul Park, "Centroid Neural Network for Unsupervised Competitive Learning", *IEEE*

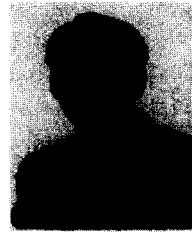
*Trans. on Neural Network*, vol. 11, no.2, pp 520-528, March 2000.

[5] E. Gokcay and J. Principe, "Information Theoretic Clustering", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, pp 158-171 February 2002.

[6] K. Fukunaga. *Introduction to Statistical Pattern Recognition*, Academic Press Inc, 2nd edition, 1990.

박 동 철(Dong-Chul Park)

정희원



1980년 2월 : 서강대학교  
전자공학과(공학사)

1982년 2월; 한국과학기술원  
전기 및 전자공학과  
(공학석사)

1990년 6월: Ph.D. in

Electrical Engineering

Univ. of Washington (Seattle)

1990년 8월 - 1994년 2월: 조교수, Florida Int'l

Univ. Dept. of Eelct. and Comp. Eng.

1994년 3월 - 현재: 명지대학교 정보공학과 교수

1997년-2000년: *IEEE Tr. on Neural Networks*,  
Associate Editor

1999년 - 현재: *IEEE Senior Member*

<주관심분야> 신경망 알고리즘 개발, 음성인식,  
신경망의 금융공학에의 응용

권 오 현(Oh-Hyun Kwon)

정희원



2001년 6월: 명지대학교  
제어계측공학과(학사)

2003년 8월: 명지대학교  
정보제어공학과(석사)

2003년 9월~ 현재 :  
(주) 정소프트

<주관심분야> 음성인식, 음성신호처리, 신경망,  
패턴인식