김일성종합대학학보
(자연과학)
주체103(2014)년 제60권 제10호

JOURNAL OF **KIM IL SUNG** UNIVERSITY
(NATURAL SCIENCE)
Vol. 60 No. 10 JUCHE103(2014).

계층구조GMM에 의한 발성자인증에 대한 연구

옥현철, 리은철

GMM-UBM은 발성자인증에서 가장 널리 리용된 모형구축방법중의 하나이다.[1, 2]

일반적으로 통용배경모형(UBM:Universal Background Model)은 수백명의 발성자들이 수시간씩 발성한 음성을 리용하여 구축되며 발성자독립의 발성특징을 표현한다.

GMM-UBM에 기초한 발성자인증에서 발성자의 가우스혼합모형(GMM: Gaussian Mixture Mode)은 학습자료를 UBM에 적응하는 방법으로 얻는다. 현재 음성인식이나 발성자인증에서 일반적으로 리용되는 적응방법들로서는 MLLR(Maximum Likelihood Linear Regression), MAP(Maximum A Posterior)와 같은것들이 있다.

그런데 UBM은 방대한 량의 학습자료를 리용하여 구축되는것으로 하여 많은(보통 1 024～2 048개) 가우스분포성분들로 이루어지며 따라서 학습자료의 량이 제한된 경우 적응을 위한 통계량추정은 충분하게 진행되지 못한다. 따라서 UBM에 대한 적응효과가 떨어지게 되며 발성자의 모형도 정확히 구축할수 없게 된다. 실례로 발성자의 학습자료가 20s(즉 2 000개 미만의 특징벡토르)인 음성인 경우 이를 리용하여 1 024～2 048개의 가우스분포성분들을 갱신하기 위한 통계량추정은 매우 불충분하며 따라서 정확한 발성자모형도 기대하기 어렵게 된다.

론문에서는 우와 같은 결함을 극복하기 위하여 학습자료량이 제한된 경우에도 UBM에 대한 MAP적응의 효과를 높여 발성자의 모형을 비교적 정확하게 구축하도록 하기 위하여 계층구조로 된 GMM형식을 제안하였다.

1. 계층구조GMM에 의한 발성자인증

론문에서 제안한 계층구조GMM(Hierachical GMM: HGMM)은 그림 1과 같다.

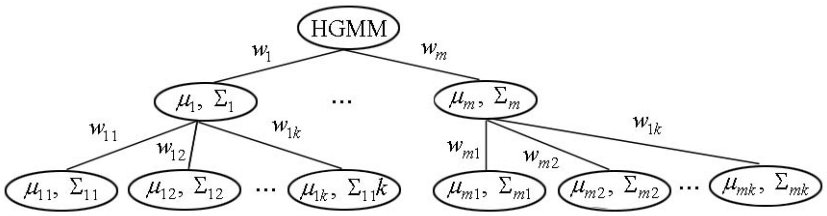

그림. HGMM의 구조

그림에서 보는바와 같이 HGMM의 웃계층은 여러 어미마디들로 이루어져있으며 아래계층의 새끼마디들은 해당 어미마디들에 종속되여있다.

HUBM은 UBM의 가우스분포성분들을 무리짓기하여 만든다. 즉 UBM의 가우스분포성

분들을 새끼마디로, 그것들을 무리짓기하여 얻어지는 클라스들을 어미마디로 한다.

HUBM에 대한 적응과정에서는 어미마디들에 대한 적응효과가 그것에 종속되는 새끼마디들의 적응에 직접 영향을 주도록 한다. 다시말하여 통계량추정이 충분한 어미마디들에 종속되는 새끼마디들에 대해서는 그 적응효과를 강조해주도록 한다.

따라서 학습자료가 제한된 경우에도 GMM-UBM에 비하여 발성자의 모형을 보다 정확히 구축할수 있게 한다.

HGMM에 대한 우도계산에서도 어미마디들은 새끼마디들에 직접 영향을 준다.

론문에서 제안한 발성자의 HGMM은 계층구조UBM(Hierachical UBM: HUBM)에 대한 MAP적응을 진행하여 얻는다.

HUBM과 적응자료렬 $X=\{x_1, x_2, \cdots, x_T\}$가 주어졌을 때 MAP적응의 첫번째 단계는 HUBM의 파라메터들의 갱신을 위한 통계량들을 추정하는것이다.

일반적으로 UBM의 파라메터들에 대한 갱신을 진행할 때 분산파라메터는 갱신하지 않아도 충분한 효과를 얻을수 있다.[1]

따라서 론문에서는 무게와 평균파라메터의 갱신에 대해서만 고찰한다.

이를 위해 그림 1에서 어미마디의 개수를 M이라고 할 때 웃계층에서 i번째 어미마디 가우스분포성분의 차지우도는 다음과 같다.

$$p(i\,|\,x_t)=w_i\,p_i(x_t)\Big/\sum_{l=1}^{M}w_l\,p_l(x_t) \tag{1}$$

여기서 w_i는 i번째 어미마디 가우스분포성분의 무게이며 $p_i(x_t)$는 가우스확률밀도함수로서 다음과 같다.

$$p_i(x_t)=\frac{1}{(2\pi)^{D/2}\,|\Sigma_i|^{1/2}}\exp\left\{-\frac{1}{2}(x_t-\mu_i)^T\Sigma_i^{-1}(x_t-\mu_i)\right\} \tag{2}$$

여기서 D는 평균과 분산의 차원수이다.

한편 아래계층에서 새끼마디들의 차지우도는 다음과 같다.

$$p_i(j\,|\,x_t)=\frac{w_{ij}\,p_{ij}(x_t)}{\sum_{k=1}^{K}w_{ik}\,p_{ik}(x_t)} \tag{3}$$

여기서 $p_{ij}(x_t)$는 i번째 어미마디에 종속된 j번째 새끼마디의 가우스확률밀도함수이다.

또한 i번째 어미마디에 종속된 j번째 새끼마디의 전체 모형에 대한 차지우도는 다음과 같다.

$$p(i,\ j\,|\,x_t)=p(i\,|\,x_t)\,p_i(j\,|\,x_t) \tag{4}$$

우의 차지우도들을 리용하여 매 가우스분포성분들의 무게, 평균파라메터를 갱신하기 위한 통계량들은 다음의 식에 의하여 계산할수 있다.

$$n_{ij}=\sum_{t=1}^{T}p(i,\ j\,|\,x_t) \tag{5}$$

$$n_i=\sum_{t=1}^{T}p(i\,|\,x_t) \tag{6}$$

$$E_{ij}(x)=\frac{1}{n_{ij}}\sum_{t=1}^{T}p(i,\ j\mid x_t)x_t \tag{7}$$

$$E_i(x)=\frac{1}{n_i}\sum_{t=1}^{T}P(i\mid x_t)x_t \tag{8}$$

모형갱신을 위한 통계량들이 다 계산되면 매 마디의 가우스분포성분들의 무게와 평균은 다음의 식들에 의하여 갱신된다.

$$\hat{w}_i=[\alpha_i^w n_i/T+(1-\alpha_i^w)w_i]\gamma \tag{9}$$

$$\hat{w}_{ij}=[\alpha_{ij}^w n_{ij}/n_i+(1-\alpha_{ij}^w)w_{ij}]\beta_i \tag{10}$$

$$\hat{\mu}_i=\alpha_i^m E_i(x)+(1-\alpha_i^m)\mu_i \tag{11}$$

$$\hat{\mu}_{ij}=\alpha_{ij}^m E_{ij}(x)+(1-\alpha_{ij}^m)\mu_{ij} \tag{12}$$

식 (9), (10)에서 γ, β_i는 어미마디와 새끼마디의 가우스분포성분들의 무게의 합이 1이 되도록 제약하는 인자이다.

적응관계곁수 α_i^ρ, α_{ij}^ρ (ρ는 w 또는 m)는 다음과 같이 정의된다.

$$\alpha_i^\rho=\frac{n_i}{n_i+r^\rho} \tag{13}$$

$$\alpha_{ij}^\rho=\frac{n_{ij}}{n_{ij}+r^\rho} \tag{14}$$

여기서 r^ρ는 적응과정에서 초기모형의 파라메터들과의 관계를 조절하는 인자로서 일반적으로 8～20사이의 값을 취한다.

2. 계층구조GMM에 대한 우도계산

그림에서 매 어미마디들은 새끼마디들로 이루어진 부분적인 GMM이라고도 볼수 있다. 따라서 i번째 어미마디에 대한 어떤 관측벡토르 x_t의 우도는 GMM에서처럼 다음과 같이 계산한다.

$$P(x_t\mid i)=\sum_{k=1}^{K}w_{ik}\,p_{ik}(x_t) \tag{15}$$

여기서 K는 i번째 어미마디에 종속된 새끼마디의 개수이며 w_{ik}는 $\sum_{k=1}^{K}w_{ik}=1$을 만족시키는 새끼마디 가우스분포성분들의 무게이다.

한편 웃계층에서 어미마디들의 차지우도는 계층구조GMM에 대한 우도계산에서 우의 우도에 대한 무게와 같이 작용한다. 즉 웃계층에서 i번째 어미마디의 차지우도가 식 (1)과 같이 주어졌을 때 관측벡토르 x_t의 모형 λ에 대한 우도는 다음과 같이 계산한다.

$$P(x_t\mid\lambda)=\sum_{i=1}^{M}\Pr(i\mid x_t)P(x_t\mid i) \tag{16}$$

3. 평 가 실 험

실험에서 리용된 발성자인증체계의 음성신호분석조건은 표 1과 같다.

표 1. 음성신호의 분석조건

지표	값	지표	값
표본화주파수	16.00kHz	고역강조결수	0.97
량자화비트수	16bit	분석창	하밍창
프레임길이	25.6ms	대역려파기수	46
프레임주기	10ms		

특징량파라메터로는 26차원 MFCC가 리용되였으며 UBM의 가우스분포성분의 수는 1 024개이다.

성능평가실험에서 리용된 코퍼스는 남자 60명, 녀자 20명의 발성자료로 이루어졌다. 이 코퍼스에서 매 사람의 등록발성자료는 10s정도의 음성파형 2개이며 인증발성자료는 5s미만의 음성파형 10개이다. 코퍼스는 한주일에 3～4개씩 3주일에 걸쳐 록음하는 원칙에서 구축되였다.

론문에서는 계층구조GMM의 어미마디의 개수를 변경시키면서 실험을 진행하여 종전의 GMM-UBM에 기초한 발성자인증체계와 성능을 비교하였다. 종전의 GMM-UBM에 기초한 발성자인증체계에서도 우와 같은 평가코퍼스를 리용하였다. 실험결과는 표 2와 같다.

표 2. GMM-UBM과 어미마디의 개수에 따르는 HGMM-HUBM의 성능비교

모형형태	어미마디의 수/개	모형크기	등오유률/%
GMM-UBM	1	1 024	4.68
HGMM-HUBM	256	〃	4.56
	128	〃	4.43
	64	〃	4.12
	32	〃	3.76
	16	〃	4.07

표 2에서 알수 있는바와 같이 계층구조GMM에 기초한 발성자인증체계는 종전의 GMM-UBM보다 성능이 좋으며 특히 뿌리마디의 수를 32개로 할 때 인증성능을 상대적으로 19%, 절대적으로 0.9%정도 개선하였다.

참 고 문 헌

[1] D. A. Reynolds et al.; Digital Signal Processing, 10, 19, 2000.

[2] A. Martin et al.; Digital Signal Processing, 10, 1, 2000.

주체103(2014)년 6월 5일 원고접수

A Study of Speaker Verification using the Hierarchical GMM

Ok Hyon Chol, Ri Un Chol

We proposed the hierarchical GMM in order to construct the speaker model more correctly by improving the MAP adapting effect on the UBM(Universal Background Model) even in the case of the limited amount of training data in the speaker verification.

We decreased the EER(Equal Error Rate) of speaker verification by 19 percent relatively or 0.9 percent absolutely using the hierarchical GMM proposed in this paper.

Key words: speaker verification, GMM, UBM, model structure, model adaptation, MAP