

HMM에 기초한 음성합성체계에서 대역분산을 고려한 자리길HMM구성방법

한철진, 김덕호

HMM에 기초한 음성합성체계는 용량이 작고 단순한 음성파라미터조작으로 발성자의 특징을 쉽게 변화시킬수 있다는 우점이 있는 반면에 합성음의 음질이 낮은 결함이 있다. 합성음질이 낮은 한가지 원인은 HMM학습에서 모형파라미터들이 지나치게 평활화되는 것이며 이 문제를 해결하기 위하여 대역분산을 고려한 파라미터생성방법이 제안되었다.[1]

한편 HMM에서의 가정에 의한 제한성을 극복하기 위한 방법으로서 동적특징량(1계 및 2계계차)의 리용방법이 제안되었으나 이 방법에서 정적 및 동적특징량들사이의 명백한 관계가 있음에도 불구하고 학습때에 그것을 무시하는 모순이 있다. 이러한 문제를 해결하기 위하여 지난 시기 자리길HMM(THMM)구성방법이 제안되었다.[2]

이 방법에서는 HMM모형으로부터 일반적인 파라미터생성방법을 리용하여 파라미터 자리길을 생성하는데 여기에 선행연구[1]에서 제안된 대역분산을 고려한 파라미터생성방법을 결합하면 모형성능을 보다 개선할수 있다.

논문에서는 HMM에 기초한 음성합성체계의 성능개선을 위하여 자리길HMM구성방법과 대역분산을 고려한 파라미터생성방법을 결합한 새로운 자리길HMM구성방법을 제안한다.

1. 목 적 함 수

대역분산을 고려한 자리길HMM(GV-THMM)구성은 주어진 HMM상태렬 $q=(q_1, q_2, \dots, q_T)$ 에 대하여 다음과 같은 목적함수 L_q 가 최대되도록 모형 λ 를 최량화하는 방법으로 진행한다.

$$L_q = P(c|q, \lambda)P(v(c)|q, \lambda, \lambda_v)^\omega \quad (1)$$

여기서 $c=[c_1^T, c_2^T, \dots, c_T^T]^T$ 는 정적특징량 $c_i=[c_i(1), \dots, c_i(M)]$ 들의 렬,

$$v(c)=[v_c(1), \dots, v_c(M)]^T \quad (2)$$

는 정적특징렬 c 의 대역분산벡터이고 $P(c|q, \lambda)$ 는 자리길HMM의 확률밀도함수[2], $P(v(c)|q, \lambda, \lambda_v)$ 는 대역분산확률밀도함수, λ_v 는 대역분산모형이며 ω 는 두 확률밀도함수의 우도균형을 조종하는 무계결수이다.

식 (2)에서 $v(c)$ 의 요소들은 다음과 같이 계산된다.

$$v_c(d) = \frac{1}{T} \sum_{i=1}^T (c_i(d) - \overline{c(d)})^2 \quad (3)$$

여기서 $\overline{c(d)}$ 는 $c_i(d)$ 의 평균값으로서 다음과 같다.

$$\overline{c(d)} = \frac{1}{T} \sum_{t=1}^T c_t(d) \quad (4)$$

한편 대역분산확률밀도함수 $P(v(c)|q, \lambda, \lambda_v)$ 는 다음과 같이 정의한다.

$$P(v(c)|q, \lambda, \lambda_v) = N(v(c)|v(\bar{c}_q), \Sigma_v) \quad (5)$$

여기서 $v(\bar{c}_q)$ 는 자리길HMM의 평균벡터의 대역분산이며 Σ_v 는 대역분산모형 λ_v 의 공분산행렬이다.

2. 모형파라미터추정

HMM의 매 상태의 평균과 대각공분산행렬 $m = [\mu_1^T, \dots, \mu_S^T]^T$, $\Sigma^{-1} = [U_1^{-1}, \dots, U_S^{-1}]^T$ 들은 위에서 설정한 목적함수 L_q 가 최대가 되도록 갱신된다.

모형파라미터의 갱신은 다음의 그라디언트들을 리용하여 반복적으로 진행한다.

$$\frac{\partial L_q}{\partial m} = \Phi_q^T U_q^{-1} W (c - \bar{c}_q + \omega P_q \bar{x}_q) \quad (6)$$

$$\begin{aligned} \frac{\partial L_q}{\partial \Sigma^{-1}} = & \frac{1}{2} \Phi_q^T \cdot \text{diag}[W(P_q + \bar{c}_q \bar{c}_q^T - c c^T) W^T - \\ & - 2\mu_q(\bar{c}_q - c)^T W^T + 2\omega W P_q \bar{x}_q(\mu_q - W \bar{c}_q)^T] \end{aligned} \quad (7)$$

여기서

$$\bar{x}_q = [\bar{x}_{q,1}^T, \dots, \bar{x}_{q,T}^T]^T \quad (8)$$

$$\bar{x}_{q,i} = [\bar{x}_{q,i}(1), \dots, \bar{x}_{q,i}(M)]^T \quad (9)$$

$$\bar{x}_{q,i}(d) = -2(\bar{c}_{q,i}(d) - \bar{c}_q(d)) \cdot [v(\bar{c}_q) - v(c)]^T p_v(d) \quad (10)$$

이며 $p_v(d)$ 는 Σ_v^{-1} 의 d 번째 열이며 행렬 Φ_q 는 $\mu_q = \Phi_q m$ 이 되도록 설정한 $3MT \times 3MS$ 차행렬이다.

3. 평가 실험

학습자료로는 22.05kHz, 16bit, Mono방식의 파형자료(2 000개 문장, 약 3h분)를, 스펙트럼파라미터로는 STRAIGHT스펙트럼으로부터 계산된 39차원 멜-케프스트라계수를 리용하였으며 F_0 은 평균진폭차함수(AMDF)와 자체상관법(ACF)을 결합한 방법을 리용하여 추출하였다.

실험을 위하여 일반적인 HMM으로 학습자료를 학습시키고 매 학습자료에 대하여 Viterbi알고리즘을 리용하여 국부최량상태열을 결정하였다.

결정된 상태열에 기초하여 자리길HMM학습을 진행하였다. 먼저 대역분산을 고려하지 않고 자리길우도가 최대가 되도록 ($\omega=0$ 으로 설정) 모형파라미터들을 갱신하고 그다음 $\omega=0.5$ 로 설정하여 대역분산을 고려한 자리길HMM학습을 진행하여 합성음질의 비교평가를 진행하였다.

합성음질의 평가는 입력문장 10개를 주고 5명의 전문가가 합성음들에 0~5사이의 점수를 주고 평균점수를 계산하여 평가하는 방식으로 진행하였다.

점수평가에서 5점은 자연음과 같은 수준의 합성음이 생성되었음을 의미하며 0점은 합성음이 전혀 알아들을수 없는 상태임을 의미한다.(표)

표. 자리길HMM의 성능평가

체계	점수1	점수2	점수3	점수4	점수5	총평
HMM	3.7	3.6	3.7	3.75	3.65	3.68
THMM	3.7	3.7	3.7	3.75	3.75	3.72
GV-THMM	3.8	3.85	3.9	3.9	3.8	3.85

표에서 보는바와 같이 자리길HMM의 성능이 일반HMM에 비하여 좋으며 대역분산을 고려한 자리길HMM을 리용하는 경우 자리길HMM을 리용하는 경우에 비하여 성능이 더 높다는것을 알수 있다.

맺 는 말

자리길HMM구성방법과 대역분산을 고려한 파라미터생성방법을 결합하여 대역분산을 고려한 자리길HMM구성방법을 제안하고 실험을 통하여 그 효과성을 검증하였다.

참 고 문 헌

[1] 김일성종합대학학보(자연과학), 57, 5, 18, 주체100(2011).

[2] K. Tokuda et al.; Computer Speech and Language, 21, 153, 2007.

주체106(2017)년 6월 5일 원고접수

Formulating Method of Trajectory HMM considering Global Variance for HMM-based Speech Synthesis System

Han Chol Jin, Kim Tok Ho

For speech quality improvement in HMM-based speech synthesis system, we proposed a formulating method of trajectory HMM considering global variance by using the formulating method of trajectory HMM and the parameter generation method considering global variance.

We proposed the objective function for this, derived model parameter estimation formulas, and verified the effectiveness of the proposed method experimentally.

Key words: trajectory HMM, global variance, speech synthesis