

# 퍼지 스무딩을 이용한 향상된 음성인식

\*김대수, \*김종교  
전북대학교 전자공학과

## A Study on Improvement of Speech Recognition by Fuzzy Smoothing

\*Dae-Su Kim, \*Chong-Kyo Kim

\*Dept. of Electronic Engineering, Chonbuk National University

### 요 약

이산 HMM을 이용한 음성인식을 할 때, 관측심볼은 훈련 데이터의 양자화과정을 수행하여 얻게 된다. 훈련 데이터는 선정된 몇몇 화자에 의해서 얻어지게 되는데, 이러한 이유로 인하여 충분하지 못한 훈련 데이터가 얻어지므로, 관측 심볼에 따라 출력확률분포값이 영(zero)이나, 거의 영에 가까운 값을 가지게 된다. 이러한 요인은 인식률의 저하를 초래하므로, 본 논문에서는 fuzzy smoothing 기법을 채택하여, 출력확률분포값이 영(zero)의 값을 가지는 것을 방지하여, 새로 구해진 파라미터로 인식실험을 하였다. Smoothing과정을 수행한 후의 인식률이 smoothing을 하지 않은 인식율에 비해 평균 1.46% 향상되었다.

### I. 서 론

최근의 음성인식은 은닉 마르코프 모델(HMM)을 이용한 음성인식이 주류를 이루고 있다. HMM은 통계적 모델링(stochastic modelling)을 이용하여 음성인식을 하게 되므로, HMM의 model parameter의 추정 여부에 따라서 인식률에 영향을 미치게 된다. 즉, 최적화된 model parameter를 추출해내는 것이 HMM을 이용한 음성인식에서는 관건이라 할 수 있다.

Discrete HMM을 이용한 음성인식에서는 관측 심볼을 얻기 위해서 훈련 데이터의 양자화를 수행하게 된

다. 훈련 데이터는 선정된 몇몇 화자에 의해서 얻어지게 된다. 이러한 양자화 에러와 선정된 화자에 의한 훈련 데이터 수집으로 인하여 최적화된 model parameter, 그 중에서도 특히 출력확률분포(output probability distribution)를 추정하기에는 충분치 못한 훈련 데이터가 얻어지게 된다. 즉, 관측 심볼에 따라서 영(zero)이나 거의 영에 가까운 출력확률분포값을 가지게 되는데, 이렇게 작은 값은 인식률의 저하를 초래한다.

충분하지 못한 훈련 데이터로 인한 인식률의 저하를 막기 위해서 본 논문에서는 smoothing을 이용하여 출력확률분포값이 영(zero)의 값을 갖는 것을 방지하여 인식실험을 수행하였다. Smoothing 기법으로 fuzzy smoothing기법을 사용하였다. 본 논문의 차례는 다음과 같다. 2, 3장에서 이산 HMM과 본 논문에서 smoothing 역할을 하는 fuzzy 양자화에 대해서 설명하고 4장에서 실험 방법과 결과를 설명하고, 5장에서 결론을 보인다.

### II. 이산 HMM

이산 HMM은 연속 신호를 양자화를 한 후 이산화된 데이터를 관측심볼로 하여 출력확률분포를 구하기 때문에 이산 HMM이라 한다. HMM은 주어진 관측열  $O$ 에 대해  $\Pr(O|\lambda)$ 가 최대가 되도록 하는 추정 과정으로 최적화된 모델 파라미터  $\lambda(A, B, \pi)$ 를 찾아내는 것이

중요하다.  $\lambda$ 에 따른 HMM에서 모델 파라미터를 구하기 위해서는 다음의 알고리즘이 필요하다.

전향 알고리즘(forward procedure)은 부분 관측열에 대한 확률값을 나타내는 것으로 다음과 같은 과정에 의해 전향 변수  $\alpha_t(i)$ 를 구한다.

1. 초기화

$$\alpha_1(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N \quad (1)$$

2. 유도과정

$$\alpha_{t+1}(j) = \left[ \sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) \quad 1 \leq t \leq T-1, 1 \leq j \leq N \quad (2)$$

3. 종료

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (3)$$

후향 알고리즘은 계산하는 관측열이  $t+1$ 부터 끝까지 계산하는 것으로 전향 알고리즘과 유사하게 구해진다.

1. 초기화

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (4)$$

2. 유도과정

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \dots, 1 \quad 1 \leq i \leq N \quad (5)$$

3. 종료

$$P(O|\lambda) = \sum_{i=1}^N \pi_i b_i(o_1) \beta_1(i) \quad (6)$$

전향-후향 변수  $\alpha_t(i)$ ,  $\beta_t(i)$ 는 Baum-Welch 알고리즘에서 함께 사용되어 파라미터 재추정시에 사용된다.

Viterbi 알고리즘은 주어진 관측열  $O$ 에 대한 최적의 상태열  $q = (q_1 q_2 \dots q_T)$ 를 찾는 알고리즘이다. Viterbi 알고리즘은 상태열을 찾아내는 이 외에 HMM에서 인식 알고리즘으로 사용된다. Viterbi 알고리즘은 다음의 과정에 의해 수행된다. 구해진 상태열을 저장하기 위해  $\psi_t(j)$  변수가 사용된다.

1. 초기화

$$\begin{aligned} \delta_1(i) &= \pi_i b_i(o_1), \quad 1 \leq i \leq N \\ \psi_1(i) &= 0 \end{aligned} \quad (7)$$

2. 재귀

$$\delta_t(i) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(o_t) \quad 2 \leq t \leq T, \quad 1 \leq j \leq N \quad (8)$$

$$\psi_t(i) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] \quad 2 \leq t \leq T, \quad 1 \leq j \leq N \quad (9)$$

3. 종료

$$\begin{aligned} P^* &= \max_{1 \leq i \leq N} [\delta_T(i)] \\ q^*_T &= \arg \max_{1 \leq i \leq N} [\delta_T(i)] \end{aligned} \quad (10)$$

4. 경로 역추적

$$q^*_t = \psi_{t+1}(q^*_{t+1}) \quad t = T-1, T-2, \dots, 1 \quad (11)$$

Baum-Welch 알고리즘은 파라미터 재추정 알고리즘으로,  $\Pr(O|\lambda)$ 가 최대가 되도록  $\lambda$ 를 갱신하는 과정이다. Baum-Welch 알고리즘에 의해서 재추정된 파라미터는 식  $\bar{\pi}_j, \bar{a}_{ij}, \bar{b}_j(k)$ 와 같다.

$$\sum_{i=1}^N \gamma_t(i) = \text{관측열 } O \text{에서 상태 } i \text{에서 천이되어지는 수}$$

$$\sum_{i=1}^N \xi_t(i, j) = \text{관측열 } O \text{에서 상태 } i \text{에서 상태 } j \text{로 천이되는 수}$$

$$\bar{\pi}_j = \gamma_1(i) \quad (12)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (13)$$

$$\bar{b}_j(k) = \frac{\sum_{t=1}^T \gamma_t(i)}{\sum_{i=1}^N \sum_{t=1}^T \gamma_t(i)} \quad (14)$$

위에서 보듯 HMM에서 전향-후향, Viterbi, Baum-Welch 알고리즘에 의해서 최적화된 모델 파라미터  $\lambda$ 를 찾아내고, 그 모델 파라미터에 의해서 Viterbi 알고리즘에 의해  $\Pr(O|\lambda)$ 를 계산하게 된다. HMM은 임의의 관측열에 대한 통계적 모델링이므로 불충분한 훈련 데이터로 인한 출력분포확률  $b_j(k)$ 의 잘못된 추정은 인식률의 저하에 영향을 주게 된다. 이러한 불충분한 훈련 데이터로 인한 파라미터 추정 에러로 인한 인식을 감소시키기 위하여  $b_j(k)$ 에 대한 smoothing을 적용한다.

### III. 퍼지 Smoothing

Smoothing technique은 불충분한 훈련 데이터로 인한 파라미터 추정 에러를 감소시키기 위한 목적으로 적용된다. 본 논문에서는 smoothing 방법으로 퍼지 (fuzzy) 양자화 알고리즘에서 사용되는 기여 (contribution) 인자를 사용하여 smoothing을 수행하였다.

퍼지 양자화 알고리즘은 아래의 목적함수  $D$ 를 최소화하는 기여인자  $m_{ki}^F$ 를 구하는 과정으로 진행된다. 아래의 수식은 일반적인 양자화 알고리즘에서 평균 왜율과 같은 의미를 지닌다.

$$D = \sum_{k=1}^T \sum_{i=1}^L m_{ki}^F d(x_k, \mu_i) \quad (15)$$

기여인자  $m_{ki}^F$ 를 구하는 공식은 다음과 같다.

$$m_{ki}^F = \left[ \sum_{j=1}^L [d(x_k, \mu_j)/d(x_k, \mu_i)]^{1/(F-1)} \right]^{-1} \quad (16)$$

그리고, 다음의 조건을 만족한다.

$$\sum_{i=1}^L m_{ki}^F = 1 \quad (17)$$

위에서 구한 기여인자와 훈련 데이터  $x_k$ 에 의해 중심값  $\mu_i$ 는 다음과 같이 구해진다.

$$\mu_i = \frac{\sum_{k=1}^T m_{ki}^F x_k}{\sum_{k=1}^T m_{ki}^F} \quad (18)$$

$T$ : 훈련 데이터의 길이

$L$ : codebook의 크기

$F$ : degree of fuzziness,  $> 1$

$d(x, y)$ : Euclidean distance

Fuzzy 양자화에서 기여 인자  $m_{ki}^F$ 는 훈련 데이터  $x_k$ 가 중심값  $\mu_i$ 에 속하는 정도를 나타낸다. Fuzzy 이론의 정의에 의해서 기여인자  $m_{ki}^F$ 를 확률의 형태로 표현하면 다음과 같다.

$$m_{ki}^F \equiv \frac{P(x_k \cap \mu_i)}{P(x_k)} \quad (19)$$

위 식은, 확률의 형태로 표현되었지만 일반적인 확률과는 다른 의미를 가지는 것으로  $F$ 에 의해서 값이 달라진다.

따라서,  $x_k$ 의 길이가  $T$ 인 데이터의  $\mu_i$ 에 대한 기여도는 다음과 같다.

$$m_{ki}^F = \sum_{k=0}^T m_{ki}^F \quad (20)$$

$F$ 는 위에서 보인 바와 같이 1보다 큰 상수로 목적에 맞는 가장 적당한 값이 어느 정도인지에 대한 관계는 아직 명확하게 정립되지 않았다. 일반적으로 1에 가까울수록 퍼지정도(fuzziness)가 약화된다.

### IV. 실험방법 및 결과

인식실험에 쓰인 데이터는 10개의 단어를 20명의 화자에 의해서 얻어졌으며, HMM 트레이닝에는 6명의 화자가 발성한 데이터에 의해 수행됐다. 데이터는 8kHz sampling rate, 16bit PCM 방식에 의해 수집됐다. 특징 파라미터로는 15차 가중 켈스트럼을 사용하였다. 트레이닝 데이터를 양자화하기 위하여 modified k-mean (MKM)을 사용하였다. HMM의 상태수는 각 단어의 음절수를 평균하여 상태수를 정하였다. 본 실험에서는 상태수를 8로 하였다.

Smoothing은 Baum-Welch 재추정 알고리즘의 출력 분포 재추정식 (14)와 fuzzy smoothing에서 보인 기여인자 식 (19), (20)에 의해 다음과 같이 구해진다.

$$b_j(k) = \frac{\sum_{all x_n} m_{jk}}{\sum_{i=1}^L \sum_{all x_n} m_{ik}} \quad (21)$$

위 식에서  $all x_n$ 는 상태  $j$ 에 포함된 모든 데이터를 나타낸다.

본 논문에서 출력확률분포  $b_j(k)$ 에 대한 smoothing 과정은 다음의 과정으로 이루어졌다. 초기 모델 파라미터  $\lambda$ 를 사용하여, 이산 HMM에서 보인 전향-후향, Viterbi, Baum-Welch 알고리즘에 의해서 재추정된 모델 파라미터  $\lambda(A, B, \pi)$ 를 구한다. 구해진 모델 파라미터를 위 식에 의해  $b_j(k)$ 에 대한 smoothing을 수행한다. 초기 파라미터  $\lambda$ 는 균등하게 하여 정했으며, HMM모델은 단순 좌-우 모델을 사용하였다. 그림 1에 단순 좌-우 모델을 나타냈다. 관측 심볼을 구하기 위해서 양자화시에 Euclidean distance를 사용하였으므로,  $F$ 값은 smoothing되어지는 데이터와 훈련되어지는 데이터간의 Euclidean distance 자체에 의해서만 기여인자가 정해지도록 2를 사용하였다.

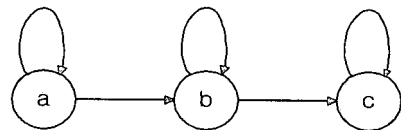


그림 1. 단순 좌-우 모델.

위 식에 의해 출력확률분포  $b_j(k)$ 를 smoothing한 모델과 smoothing하지 않은 모델의 인식율은 다음의 그림 2, 3과 같다. Fuzzy smoothing을 수행한 모델이 전체적으로 인식율이 향상된 것을 볼 수 있다. No smoothing 모델은 최저 출력확률분포값으로  $10^{-4}$ 을 적용하였다.

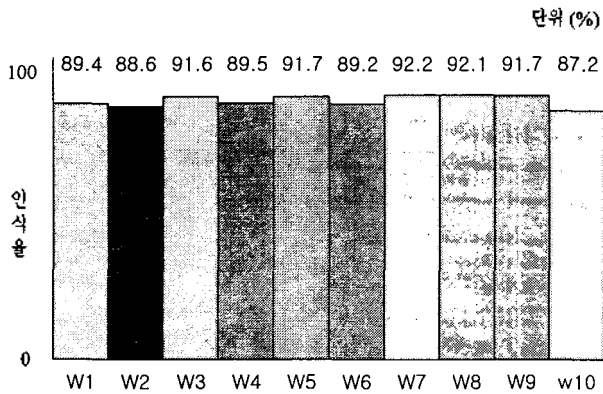


그림 2. No smoothing model의 인식율.

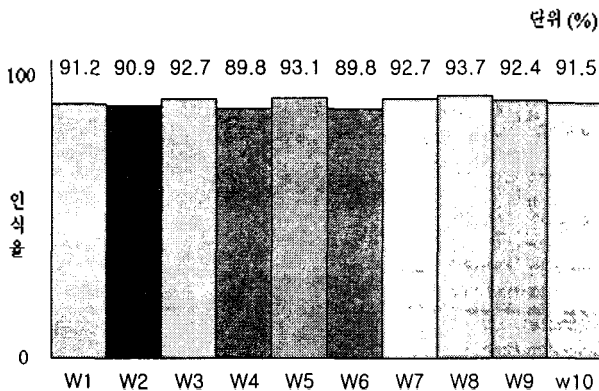


그림 3. Smoothing model의 인식율.

## V. 결론

HMM은 통계적 모델링에 의하여 인식율 하므로 불충분한 데이터로 인한 모델 파라미터의 잘못된 추정 은 인식율의 저하를 가져온다. 4장에서 보인 실험 결과에 의해 smoothing 방법으로 많은 향상은 이루어지지 않았지만, 불충분한 데이터로 인한 에러를 어느 정도 보완 할 수 있었다. 그리고,  $F$ 에 따라서 smoothing의 정도가 달라지므로, 적당한  $F$ 값을 찾아내면 더 효과적인 smoothing이 되리라 생각된다.

## 참고문헌

- [1] X.D. HUANG, Y. ARIKI, M.A. JACK, Hidden markov models for speech recognition, Edinburgh university press, 1990.
- [2] Lawrence Rabiner, Bing-Hwang Juang, Fundamentals of speech recognition. Prentice Hall, 1993.
- [3] Wilbur B. Davenport, JR, Probability and random process.
- [4] Bart Kosko, Neural networks and fuzzy systems, Prentice Hall, 1992.
- [5] Jianing Dai, "Robust estimation of HMM parameters using fuzzy vector quantization and parzen's window," Pattern Recognition, vol 28, no. 1, pp. 53-57, 1995.
- [6] M. Taylor, F. Bimbot, "Temporal decomposition for the initialization of a HMM isolated word-recognizer," ICASSP, vol 1, pp 369 - 372, 1992.
- [7] Jean-Luc Gauvain, Chin-Hui Lee, "Improved acoustic mo-delling with bayesian learning," ICASSP, vol. 1, pp. 481-484, 1992.
- [8] Paul Bao, Albert Sim, "A Hybrid Speech Recognition Model based on HMM and Fuzzy PPM," ICASSP, vol. 6, pp. 4148-4153, 1998.
- [9] A. Cavallaro, F. Beritelli, S. Casale, "A fuzzy logic-based speech algorithm for communications in noisy environments," ICASSP, vol. 1, pp. 565-568, 1998.