

한국어 음성/문자 공용인식기의 성능향상을 위한 가변 상태수 CHMM모델의 구성

°석수영, 김민정, 김광수, 정호열, 정현열

영남대학교 정보통신공학과

Difference State Number of CHMM Model to Improve the Performance of SCCRS

°Soo-Young Suk, Min-Jung Kim, Kwang-Soo Kim, Ho-Youl Jung, Hyun-Yeol Chung

Department of Information and Communication Eng., Yeungnam University

lover@yumail.ac.kr, manjuk@orgio.net, ks.kim@yumail.ac.kr, hoyoul@yu.ac.kr, hychung@yu.ac.kr

요 약

문자인식 또는 음성인식을 위해 사용되어지는 CHMM(Continuous Hidden Markov Model)모델은 일반적으로 모델의 상태수를 일정한 수로 고정하는 고정 상태수 모델 구조를 가지고 있으나, 이는 개별적인 인식 단위의 특성을 고려하지 않은 경우로써 이를 고려한 가변 상태수 모델을 사용할 경우 인식을 향상시킬 수 있다. 개별적인 인식 단위에 적합한 모델 상태수를 결정하는 방법으로 파라미터 히스토그램 방법과, BIC(Bayesian Information Criterion)방법을 사용하는 것이 대표적이다. 이들 방법들은 개별적인 인식단위의 우도값만을 향상시키기 위한 방법으로 전체인식률과 직접적으로 비례하지는 않는다. 따라서, 본 논문에서는 고정 상태수를 갖는 모델 적용 방법과 인식단위별 상태수 변화에 따른 인식률을 비교하였으며, 이를 바탕으로 각 모델별 상태수를 달리하는 가변 상태수 CHMM모델 구성 방법을 제안한다. 제안된 가변상태수 모델의 유효성을 확인하기 위해 음성/문자 공용인식기 중 필기체 문자 인식에 적용한 결과 제안한 LM(Local Maximum)으로 구성된 가변 상태수 모델이 MLE와 BIC로 구성된 모델과 인식을 면에서는 거의 동일한 성능을 유지하면서 전체 상태수는 MLE 모델에 비해 31%, BIC로 구성된 모델에 비해 22% 감소를 나타내어 제안한 모델의 유효성을 확인할 수 있었다.

1. 서 론

PDA(Personal Digital Assistants)와 같은 휴대용 단말기의 경우 키보드와 마우스 대신 문자인식, 음성인식, 펜제스처, 등을 이용하여 사용상의 편의성을 증가시키고 있

다. 특히 온라인 문자인식과 음성인식은 기본적인 멀티모달 인터페이스로 상호보완적인 동작이 필요하다. 이를 위해 개별적인 인식과정을 수행하는 음성인식과 문자인식을 결합하여 음소와 자소 단위의 CHMM(Continuous Hidden Markov Model) 모델로 구성한 후 동일한 처리과정으로 인식을 수행하는 SCCRS(Speech/Character Combined Recognition System)을 구성하였다[1]. SCCRS 엔진은 휴대용 머신에서 항상 대기상태로 존재함으로써 메모리 사용량의 효율성을 증대시킬 필요가 있으며, 실시간으로 처리가 가능한 인식속도의 개선이 필수적이다. 이를 위해 SCCRS에서 적용한 CHMM의 구조를 고려함으로써 시스템 전체의 성능을 향상시킬 필요가 있다.

일반적인 CHMM모델은 인식단위 마다 동일한 상태수와 혼합수를 가지는 고정 상태수 모델 구조를 가지고 있다. 그러나, 이것은 인식단위간의 다양한 복잡도를 고려하지 않은 문제점을 가지고 있다. 예를 들어 온라인 문자인식의 한국어 자소 “ㄱ”의 경우 “ㄷ”에 비해 상대적으로 작은 상태수로 모델링하는 것이 적합하다. 개별적인 인식 단위간의 적합한 상태수를 결정하는 방법으로 특징벡터가 출연하는 빈도로부터 만들어낸 히스토그램을 통해 결정하는 방법[2]과 확률과 정보이론으로 접근하는 BIC(Bayesian Information Criterion) 방법[3]이 대표적으로 이를 이용하여 가변 상태수 모델을 구성한 경우 고정 상태수 모델에 비해 성능의 향상을 나타내고 있다 [4]. 그러나, 개별 인식단위에 적합한 모델을 조합한 가변 상태수 모델의 경우에도 다른 인식단위 사이의 상관관계를 고려하지 못한 단점을 가지고 있다. 따라서 본 논문에서는 주어진 데이터 집합에서 다른 모델과의 상관관계를 고려하여 국부적으로 인식률이 최대가 되는 최소의 모델의 상태수를 선택하는 방법을 제안한다.

다음 2장에서는 일반적인 CHMM의 구조에 대해 살펴

보고, 3장에서는 모델 구조를 결정하는 방법 중 기존의 BIC를 이용한 방법과 인식률을 통해 구조를 결정하는 방법을 비교한다. 4장에서는 필기체 문자인식 실험을 통해 결과를 분석한 후, 5장에서는 결론과 향후 연구방향에 대해 기술한다.

2. CHMM topology

시간열을 모델링하는 음성인식과 온라인 문자인식에서는 CHMM을 통계적 모델로 널리 사용되고 있다[5][6]. CHMM을 사용하는 SCCRS에서는 음성인식의 기본단위로 48 유사음소와, 문자인식을 위해서는 67개의 자소를 사용한다. CHMM 모델 구조의 결정은 그림 1에서와 같이 몇 개의 상태수로 구성할 것인가, 각 상태는 몇 개의 가우시안 혼합수로 구성할 것인가, 그리고 상태간의 전이는 어떻게 구성할 것인가를 정하는 것이다.

이 가운데 본 논문에서는 계산량을 최소화한 휴대용 기기에 사용하기 위해 각 상태의 혼합수는 1개로 고정하고, 상태간의 전이는 1 상태 건너뛴을 허용하는 Left-to-Right 모델로 고정한 후 상태수를 가변하면서 최적의 상태수를 찾는데 초점을 두었다.

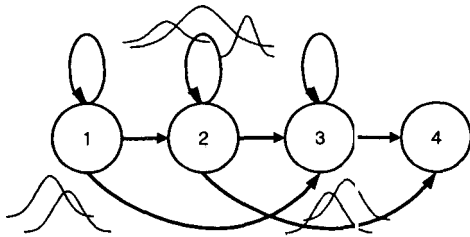


그림 1. 3상태 CHMM모델

2.1 HMM의 상태수

HMM의 상태수의 증가는 시스템의 인식시간과 소요되는 메모리에 직접적인 상관관계를 가지고 있다. 다음 표 1은 SCCRS중 음성단어 인식부분의 상태수에 따른 인식시간과 메모리 사용량을 나타내고 있다. 인식단위마다 동일한 3상태를 사용한 경우에 비해 6상태를 사용하여 모델을 구성한 경우 시간은 75%가 늘어나고, 메모리는 1.58MB가 더 필요함을 나타내고 있다. 따라서 인식률의 저하가 없는 최소 상태수의 사용이 필요하다.

표 1. 상태수에 따른 인식시간과 메모리 사용(음성단어 인식 부분, 48유사음소 모델)

상태수	시간(S)	메모리	총 상태수
3	1.00	4.5MB	144
4	1.33	5.06MB	192
5	1.57	5.55MB	240
6	1.75	6.08MB	288

2.2 고정 상태수 CHMM 모델

일반적인 CHMM모델은 인식단위 마다 동일한 상태수와 혼합수를 가지는 고정 상태수 모델 구조를 가지고 있다. 이는 음성 유사음소단위에서 평균 36 프레임의 길이를 가지는 "ja"와 평균 10 프레임의 "중성b"를 같은 상태수로 모델링 하고, 문자 자소단위에서 평균 16 프레임을 나타내는 "초성 ㄱ"모델과 평균 33 프레임을 나타내는 "중성 ㅁ"모델을 같은 상태수로 모델링 하는 것이다.

CHMM에서 각 모델의 상태수를 동일하게 증가시키는 경우 표 2에서와 같이 인식률의 증가가 나타나나, 일정수 이상 증가한 경우 프레임의 길이가 짧아 훈련이 되지 않는 인식단위가 나타나며, Overtraining 되는 모델이 생성되어 이로 인해 인식률의 저하를 보일 수 있다. 표 2만을 고려한 고정 상태수 CHMM모델을 구성하는 경우 모든 자소가 상태수 6을 가지도록 구성하는 것이 최적일 것이다.

표 2. 상태수에 따른 필기체 자소인식률의 변화

상태수	2	3	4	5	6	7	8	9
초성	58.27	81.36	84.91	90.64	91.64	92.36	95.36	95.64
중성	78.73	79.09	79.36	86.00	92.09	94.73	95.18	95.09
종성	89.18	87.76	80.41	88.16	88.78	58.78	49.59	49.8
합계	72.27	81.6	81.82	88.29	91.3	87.21	86.95	87.06
총상태수	92	138	184	230	276	322	368	414

2.2 가변 상태수 CHMM 모델

고정 상태수 모델은 인식 단위 사이의 다양한 복잡도를 고려하지 않은 문제점이 있다. 즉, 복잡도가 높은 단위의 경우 복잡도가 낮은 단위에 비해 상대적으로 많은 상태수 혹은 가우시안의 조합이 필요하다. 예를 들어 온라인 필기체 인식에서 "ㅁ"자소의 경우 "ㄱ"자소에 비해 위치 파라미터의 변화량이 크다고 생각할 수 있다. 가변 상태수 모델은 초성, 중성, 종성으로 나누어진 표 2만을 고려할 경우 초성은 9상태를 이용하고, 중성은 8상태, 종성은 6상태를 사용하는 것이 인식률의 향상을 가져올 수 있다. 본 연구에서는 가변 상태수 CHMM 모델을 구성하기 위해 각 인식단위마다 최고의 인식률을 나타내는 상태를 계산하고, 이를 조합한 가변 상태수 모델을 사용한다. 특히 온라인 문자인식에서 자소의 길이 정보를 잘 표현한 Duration Density CHMM을 사용할 경우 가변 상태수모델은 더욱 효과적이다.

3. 모델 구조 결정

모델 구조 결정의 문제는 주어진 데이터 집합을 표현하는 후보모델집합 중 하나를 선택하는 것으로 생각할 수 있다. 각 후보 모델은 다른 상태수와 다른 가우시안 혼합수를 조합한 것이다. 각 모델의 구조를 결정하는 방

법은 파라미터 변화량 히스토그램으로부터 결정하는 휴리스틱한 접근 방법과 BIC를 이용한 방법이 있다. 이중 BIC를 이용한 방법에 대해 살펴보고, 제안한 인식을 측면에서 LM된 모델구조 결정 방법에 대해 기술한다.

3.1 Bayesian Information Criterion

BIC 방법은 모델 구조를 결정하는 척도로 필터 계수를 추정하거나 HMM의 상태수를 결정하기 위해 널리 사용되어져 왔다[3]. 상태수에 따른 로그 우도값은 그림 2에서와 같이 상태의 수가 증가할수록 훈련데이터의 우도값 역시 증가한다.

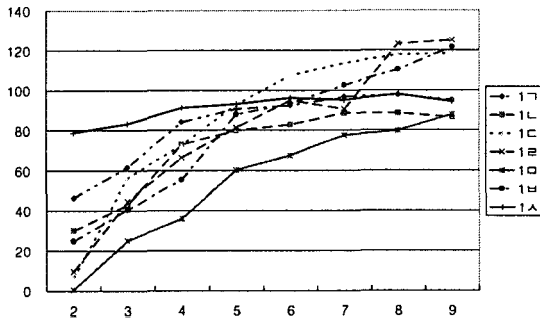


그림 2. 상태수에 따른 로그 우도값의 변화

데이터의 우도는 임의의 d 개의 파라미터 벡터 $\theta = (\theta_1, \dots, \theta_d)$ 로 구성된 데이터 X_i 의 집합 $D (X_1, \dots, X_N)$ 로부터 식(1)과 같이 나타난다.

$$p(D) = \int p(D|\theta)p(\theta)d\theta \quad (1)$$

데이터의 집합 D 가 모델 구성시 사용된 데이터의 수에 독립적이고 벡터 공간상에서 완전 분산되어 있다고 가정한다면 테일러 근사법을 이용하여 로그 우도값은 다음과 같이 나타낼 수 있다.

$$\log p(D) = \log p(D|\hat{\theta}) - \frac{d_i}{2} \log n + O(n^{-\frac{1}{2}}) \quad (2)$$

여기서, $\hat{\theta}$ 은 θ 의 MLE(Maximum likelihood Estimation) 알고리즘으로 훈련된 모델을 나타낸다. 후보 모델 집합으로부터 선택하는 최적의 모델 θ^* 은 데이터 집합의 로그우도의 합의 부분과 모델 구성시 사용된 상태수와 데이터 수에 의존하는 부분으로 나눌 수 있다.

$$\theta^* = \max_{\theta_i} \left\{ \sum_{n=1}^N \log P(X_n | \theta_i) - \frac{d_i}{2} \log N \right\} \quad (3)$$

여기서, X_n 은 n 번째 목적열, d_i 는 i 번째 모델의 상태수, N 은 목적열의 수를 나타낸다. 식 (3)에서 상태수가 증가할수록 모델의 로그 우도값은 증가하나, 상태수와 데이터 수에 의존하는 감쇄항 역시 증가한다. 그림 3, 4

에 나타난 한국어 자소 “초성 ㄱ”과 “초성 ㄴ”의 경우 로그 우도값만으로 상태수를 선택하는 경우 8상태와 9상태가 최적이거나 감쇄항의 영향으로 5상태와 8상태를 선택하는 것이 유리함을 나타내고 있다.

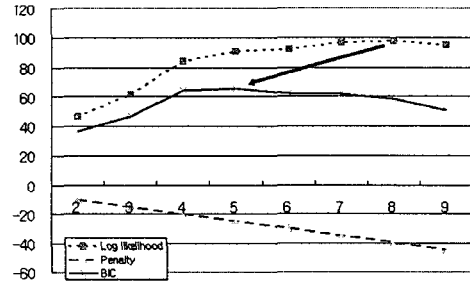


그림 3. 한국어 “초성 ㄱ” 자소의 상태수에 따른 로그 우도값, Penalty, BIC. (1 Gaussian Mixture)

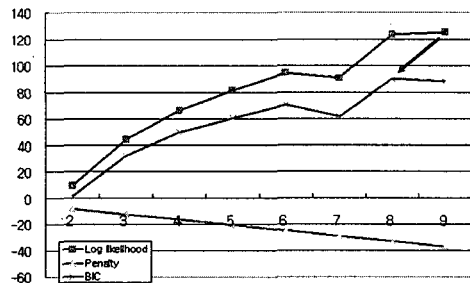


그림 4. 한국어 “초성 ㄴ” 자소의 상태수에 따른 로그 우도값, Penalty, BIC. (1 Gaussian Mixture)

BIC를 이용한 모델 구조 결정 방법에서는 개별적인 모델의 우도값만을 고려한 경우로 전체 인식률은 고려하지 않은 단점을 가지고 있다.

3.2 Local Maximize된 모델 구조 결정

각 인식단위는 자신에게 적합한 모델 구조를 가지고 있으나, MLE(Maximum Likelihood Estimation)으로 모델의 상태수를 선택하는 경우 상태수를 증가시키면서 우도값이 최대가 되는 상태수를 선택하게 된다. 그림 5의 경우 우도값이 최대가 되는 8상태를 선택하게 된다.

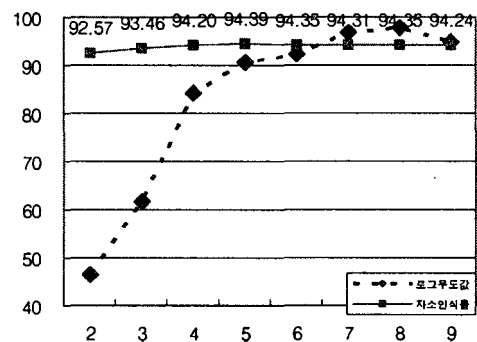


그림 5. 초성 “ㄱ” 모델의 상태수에 따른 로그 우도값과 전체 자소 인식률의 변화

그림 5에서와 같이 개별적인 인식단위의 우도값과 인식률은 항상 비례관계를 나타내지는 않는다. 즉, 인식 단위별 최대 우도값으로 구성된 모델이 전체 인식률이 최대가 되는 것은 아니며, 동일한 인식률에서 최소의 상태수를 가지는 것은 아니다. 따라서, MLE로 구성된 모델로부터 인식률을 유지하면서 최소의 상태수를 결정하기 위해서 전체 모델 집합중 전체 인식률이 최대가 되는 모델을 선택하는 것이 유리하다. 그림 5의 경우 “1” 모델이 8상태일 경우 전체 자소인식률이 94.35%를 나타내고, 5상태의 경우 94.39%를 나타내어 5상태 “1” 모델을 선택하는 것이 유리함을 나타내고 있다.

4. 인식실험 및 고찰

본 논문에서 사용된 데이터베이스는 KAIST 필기체 한글 데이터베이스를 사용하였다. 데이터의 분석 조건은 표 3과 같다.

표 3. 데이터 분석조건

전처리	100 samples/sec distance resampling
특징벡터	절대 y 좌표 1차, 국부적 각도 2차, 국부적 반곡 2차, 비트맵 9차
모델	1 Mixture Duration Density CHMM

제안된 시스템의 유효성을 확인하기 위해 SCCRS중 온 라인 문자 인식 실험을 수행하였다. 필자 독립 모델을 구성하기 위한 문자 데이터는 KAIST에서 작성된 필기체 한글 DB중 10인의 1회 필기를 이용하였다.

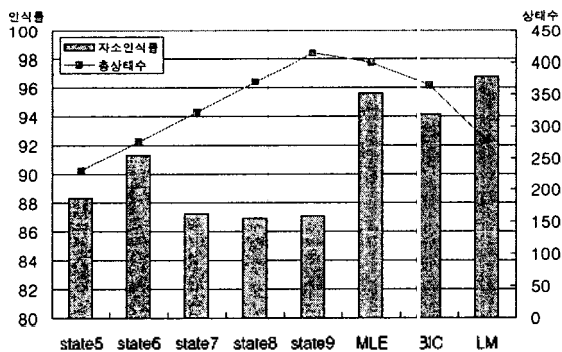


그림 6. 고정 상태수, MLE, BIC, LM 모델간의 자소 인식률과 총 상태수의 비교

자소 인식 실험 결과 그림 6에서와 같이 LM 가변모델이 MLE로 구성된 모델과 BIC로 구성된 모델에 비해 1.19%, 2.64%의 작은 인식률 차이를 나타내나, 총 상태수는 MLE으로 구성된 모델에 비해 31%, BIC로 구성된 모델에 비해 22% 감소를 나타내고 있다.

표 4. 고정 상태수, MLE, BIC, LM 모델의 단어인식률

state6	MLE	BIC	LM
84.64	91.64	90.73	91.91

글자 단위의 필기체 인식 실험을 위한 평가용 데이터로 11명의 각 100글자를 대상으로 인식실험을 수행하였다. 인식실험 결과 LM으로 구성된 가변 상태수 모델이 MLE으로 구성된 모델에 비해 0.27%, BIC로 구성된 모델에 비해 1.18%의 인식률 차이를 나타내어 글자 단위에서도 인식률 저하 없이 상태수 측면에서 효과적임을 확인할 수 있다.

5. 결 론

일반적인 CHMM 모델은 고정 상태수 모델 구조를 가지고 있으나, 이는 인식 단위 사이의 다양한 복잡도를 고려하지 않은 경우이다. 이를 개선하는 방법으로 각각의 인식단위에 적합한 CHMM 모델의 상태수를 결정하는 가변 상태수 모델 구조를 가지는 것이 유리하다. 개별적인 인식 단위에 대한 HMM 구조를 결정하는 방법으로 파라미터 히스토그램과 BIC를 이용하는 방법이 대표적이나, 이를 이용한 가변 상태수 모델이 전체 인식률을 유지하는 최소 상태수를 가지는 것은 아니다. 따라서, 본 논문에서는 자소별로 전체 자소 인식률이 최대가 되는 상태수를 찾아 조합하는 LM 방법을 제안하였다. SCCRS 중 필기체 자소 인식 실험결과 LM으로 구성된 가변 상태수 모델이 MLE과 BIC 모델에 비해 1.19%, 2.64%의 비슷한 인식률을 유지하면서, 상태수는 MLE 모델에 비해 31%, BIC로 구성된 모델에 비해 22% 감소를 나타내 제안한 가변 상태수 모델의 유효성을 확인할 수 있었다.

참 고 문 헌

- [1] S.Y.Suk, M.J.Kim, H.Y.Chung, "An on-line speech and character combined recognition system for multimodal interfaces," EALPIIT, pp89-92, January 2002.
- [2] K.S.Nathan, H.S.M.Beigi, J.Subramonia, G.J Clary, and H.Maruyama, "Real-time on-line unconstrained handwriting recognition using statistical methods," ICASSP, vol.4, pp.2619-2623,1995.
- [3] D.Li, A.Biem and J.Subrahmonia, "Hmm topology optimization for handwriting recognition," IEEE, 2001.
- [4] S.S.Chen, P.S.Gopalakrishana, "Clustering via the bayesian information criterion with applications in speech recognition," IEEE, 1998.
- [5] X.D.Huang, Y.Ariki, M.A.Jack, Hidden Markov Models for Speech Recognition, Edinburgh Univ. Press, 1990.
- [6] 조성정, 김진형, "HMM기반 온라인 한글 인식에서의 구조적, 전역적 지식의 적용," 한국정보과학회, 1998.