

온라인 필기 숫자 인식을 위한 혼동 모델 선택 기준

(Confusion Model Selection Criterion for On-Line
Handwritten Numeral Recognition)

박 미 나^{*} 하 진 영^{**}

(Mi-Na Park) (Jin-Young Ha)

요약 HMM은 파라미터의 수가 많을수록 모델링 성능이 향상되어 해당 클래스 데이터는 뿐만 아니라 혼동되는 다른 클래스 데이터에 대해서도 높은 확률을 출력하는 경향이 있다. 그러므로 단순히 파라미터 수를 증가 시키는 것은 변별력 향상에 도움이 되지 않는다. 본 논문에서는 혼동되는 클래스 데이터의 확률을 이용한 혼동 확률 선택 기준 CMC(Confusion Model Selection Criterion)과 혼동 클래스 데이터를 구성하여 혼동 모델을 만들고 이것을 이용한 새로운 인식 방법인 RCM(Recognition using Confusion Models)을 제안하였다. 제안한 인식 방법은 혼동되는 클래스 데이터의 구성으로 혼동 데이터 집합을 만들고 이것을 이용하여 별도의 혼동 모델을 훈련한 후, 혼동 모델의 확률을 해당 표준 모델의 확률에서 차감하여 해당 클래스 데이터의 오인식 가능성을 억제한다.

모델 선택 기준 CMC를 온라인 필기 숫자 데이터를 대상으로 실험하여 기존 모델 선택 기준인 ML, ALC2, BIC와 비교 분석한 결과, 제안한 방법인 CMC가 적은 파라미터로 좋은 결과를 보였으며, 제안한 혼동 모델 인식 방법인 RCM은 93.08%의 인식률을 보여 표준 모델만을 사용한 인식한 방법보다 정인식률이 약 1.5% 향상되었고, 이는 오류의 17.4%가 감소된 결과이다.

키워드 : 은닉마르코프모델, 혼동모델선택기준, 구조최적화, 혼동모델

Abstract HMM tends to output high probability for not only the proper class data but confusable class data, since the modeling power increases as the number of parameters increases. Thus it may not be helpful for discrimination to simply increase the number of parameters of HMM. We proposed two methods in this paper. One is a CMC(Confusion Likelihood Model Selection Criterion) using confusion class data probability, the other is a new recognition method, RCM(Recognition Using Confusion Models). In the proposed recognition method, confusion models are constructed using confusable class data, then confusion models are used to depress misrecognition by confusion likelihood is subtracted from the corresponding standard model probability.

We found that CMC showed better results using fewer number of parameters compared with ML, ALC2, and BIC. RCM recorded 93.08% recognition rate, which is 1.5% higher result by reducing 17.4% of errors than using standard model only.

Key words : HMM, BIC, CMC, ALC, Confusion, Topology optimization

1. 서론

패턴인식을 위해 시도 되어온 다양한 방법론 중에서 1980년대 말부터 현재까지 은닉 마르코프 모델(Hidden Markov Model: 이하 HMM)은, 시간에 따라 변해 가

는 특성을 지닌 데이터와 변형이 심한 데이터에 대한 높은 모델링 능력으로 인해 음성 인식과 온라인 필기 인식에서 우수한 성능을 보여 왔다[1,2]. 최근 보급이 확산되고 있는 PDA와 휴대용 무선 장치의 필기 인식 시스템을 위해 시스템의 메모리 용량과 인식 성능의 향상에 직접적인 영향을 미치는 HMM 구조의 최적화에 대한 연구는 절실히 필요하다. 온라인 필기인식에서 많이 사용되는 HMM은 left-to-right HMM으로 이 모델 구조는 상태 수와 상태 당 가우시안 믹스처(Gaussian

^{*} 학생회원 : 강원대학교 컴퓨터정보통신공학과
mnpark@kangwon.ac.kr

^{**} 정회원 : 강원대학교 컴퓨터학부 교수
jyha@kangwon.ac.kr

논문접수 : 2006년 7월 20일

심사완료 : 2007년 7월 25일

Mixtures) 수, 그리고 상태 사이의 전이 확률에 의해 결정된다. 이러한 HMM의 구조는 휴리스틱 방법에 의해 결정되는 것이 일반적이기 때문에 최적의 모델을 선택하는데 어려움이 있었다. 여기서 최적화란 최소한의 모델 파라미터로서 최소의 오류를 허용하는 것을 의미한다[1,2].

HMM 구조의 최적화를 위한 기존 연구는 다양한 방법으로 진행되어 왔다[3-10]. 높은 점유를 갖는 상태부터 순차적으로 분할해서 점차 상태 수를 증가시키는 방법[3]과 Dirichlet 사전 확률에 기반 한 사후 확률을 사용하여 복잡한 구조로부터 점차 구조를 감소시켜 나가는 방법이 있었다[4]. 또한 최대 확률 기준을 이용하거나 베이저안 정보기준(Bayesian Information Criterion: 이하 BIC)을 이용한 연구가 있었다[5,6]. BIC는 동질의 파라미터를 갖는 통계적으로 잘 행동하는 모델을 가정하고 있으므로 여러 종류의 파라미터로 구성된 HMM에는 적합하지 않다. BIC는 파라미터의 수는 줄였으나 인식률의 향상에는 큰 기여를 하지 못했다[2].

HMM은 파라미터의 수가 많을수록 해당 클래스 데이터는 물론 다른 클래스 데이터에 대해서도 높은 확률을 출력하는 특징이 있다. 해당 클래스 데이터의 다른 클래스의 데이터에 대한 확률의 증가를 억제하는 방법의 ALC2(Anti-Likelihood Criterion2)연구가 있었다[10]. 이와 유사한 방법으로 본 논문에서는 해당 클래스 데이터로 자주 오인식되는 클래스의 데이터들의 확률을 억제하는 방법인 CMC(Confusion Model Selection Criterion)를 제안하였다. 이 연구는 BIC보다 파라미터의 수가 감소하였고 인식률의 향상 면에서 효과적이었다. 그리고 ALC2보다 적은양의 파라미터로 인식률의 향상을 보였다. 그러나 ALC2와 마찬가지로 데이터의 크기와 종류가 다양한 데이터에 대해서는 불안정한 성능을 보였다[11,12].

HMM은 파라미터의 수가 많을수록 해당 클래스 데이터는 물론 다른 클래스 데이터에 대해서도 높은 확률을 출력하는 경향이 있다. 따라서 다른 클래스 데이터가 해당 클래스 데이터로 혼동하는 경우가 자주 발생한다. 본 논문에서는 해당 클래스 데이터로 혼동되기 쉬운 클래스에 속한 데이터로 혼동 데이터 집합을 만들고 이것으로 별도의 모델을 훈련한다. 이 모델을 혼동 모델이라고 정의한다. 혼동 모델의 확률을 표준 모델의 확률에서 차감하여 해당 데이터가 오인식될 가능성을 억제하도록 한다. 이 방법이 본 논문에서 제안하는 새로운 인식방법 RCM(Recognition Using Confusion Models)이다.

본 논문에서는 혼동되는 데이터간의 변별력 특성을 가진 CMC와 새로운 인식방법 RCM을 온라인 필기 숫자 데이터를 대상으로 실험하였다. CMC를 기존의 모델

선택 기준인 ML, ALC2, BIC와 비교 분석한 결과, 제안한 방법인 CMC가 적은 파라미터로 좋은 결과를 보였으며, 제안한 인식 방법인 RCM은 93.08%의 인식률은 표준 모델만을 이용하여 인식한 방법보다 정인식률이 약 1.5%향상되었고 17.4%의 오류 감소율을 보이고 있다.

2. 관련연구

입력 값과 출력 값으로 이루어진 데이터가 주어졌을 때 입력 값으로부터 예측되는 출력 값을 얻기 위한 가장 좋은 파라미터는 어떤 모델을 선택하느냐에 따라 달라진다.

이러한 문제들은 패턴인식 분야에서 가장 두드러진 연구가 진행되고 있으며, 모델 선택의 문제는 통계학, 기계 학습, 데이터 마이닝 등에서도 제기되고 있다. 최근 모델 선택에 대한 워크샵과 주어진 과제를 가장 효과적으로 찾아내는 대회가 캐나다에서 열렸다[13]. 잘 알려진 모델 선택기준에는 Akaike information criterion, bootstrap criteria, cross-validation criteria 그리고 Bayesian Information criterion 등의 방법이 있다[14]. LDA(Linear discriminant analysis)는 음성인식에서 파라미터 값을 변형하여 인식시스템의 성능을 향상시키는 방법이다[15]. 또 모델 자체를 변화시키는 MMI(Maximum Mutual Information)은 학습된 주어진 데이터에 관련된 모델의 사후 확률 값을 최대화하는 방법이다. 이는 상관된 두 개의 변수, 즉 모델 자체를 변화하여 모델간의 평균을 최대화 시키는 방법이다[16]. 최근 Support Vector Machine을 이용한 모델 선택 연구가 진행되고 있다[17].

음성인식이나 온라인 필기인식에서 많이 사용되는 left-to-right HMM은 일반적으로 시간에 따라 입력되는 데이터를 잘 모델링하는 특징이 있다. left-to-right HMM의 구조(topology)는 상태 수와 상태 당 믹스처 수로 결정된다. 이 논문에서 세 가지의 모델 선택 기준 연구를 소개한다. 주어진 파라미터의 최대 확률을 나타내는 구조를 선택하는 Maximum Likelihood criterion 과 모델 구조의 사전 확률을 이용한 Bayesian Model Selection, 실험 대상 클래스 데이터를 빼 나머지 데이터를 이용한 Anti-likelihood Criterion이 있다.

2.1 ML(Maximum Likelihood) Criterion

Maximum Likelihood criterion은 주어진 입력 데이터에 대해 최대우도를 나타내는 HMM 구조를 선택한다.

$$ML(M) = \frac{1}{N} \sum_{X \in C} \log P(X|\theta_{ML}) \quad (1)$$

여기서 X 는 데이터 집합이고, N 은 데이터의 수이다. θ_{ML} 은 파라미터의 최대우도 추정 값이다. HMM의 우

도는 일반적으로 모델 파라미터의 수에 따라 증가하며 ML 기준은 많은 파라미터를 갖는 HMM 구조를 선호한다. 이 방법은 계산이 단순하고 별도의 모델이나 데이터를 요구하지 않기 때문에 많이 사용한다. 학습 데이터의 양이 충분하다는 조건 하에서 아주 좋은 성능을 보이지만 ML은 모델 파라미터 수의 증가를 막는 패널티 항이 없는 단점을 갖고 있다.

2.2 베이지안 모델 선택(Bayesian Model Selection)

베이지안 구조(Bayesian Framework)에서 모델 선택은 모든 파라미터가 동일한 확률 분포를 갖는다는 가정 하에 도출된 다음 식을 사용하여 이루어진다[18-21].

$$BIC(M) = \frac{1}{N} \sum_{X \in C} \log p(X|\theta_{ML}) - \alpha \frac{k}{2} \log N \quad (2)$$

위 식에서 BIC는 우도와 $-\alpha \frac{k}{2} \log N$ 의 합인데, 후자는 모델내의 파라미터 개수에 대한 패널티(penalty) 항으로 볼 수 있다. α 는 패널티 항의 가중치 반영 비율이다. N 은 전체 훈련 데이터 수이다. 사전 확률은 자유 파라미터 개수에만 제한되고, 모델을 정의하는 각각의 파라미터 유형에 따라 별도의 고려를 하지 않는다. HMM에는 동질적이지 못한 파라미터 집합이 존재하기 때문에 이러한 가정은 부적절하다[1,5,22].

2.3 ALC(Anti-Likelihood Model selection Criterion)

HMM은 파라미터의 수가 많을수록 해당 클래스 데이터뿐만 아니라 다른 혼동 가능한 클래스 데이터에 대해서도 높은 확률 값을 보인다[2]. 이는 HMM이 유사한 패턴 클래스에 대한 변별력이 부족하기 때문이다. ALC는 해당 클래스 데이터를 뺀 나머지 클래스 데이터의 확률 값을 이용하여 반우도(Anti-Likelihood)를 생성한다. 반우도를 파라미터 수가 증가하는 것을 억제하는데 사용한다. 따라서 해당 클래스 데이터에 대해서는 잘 모델링 되고, 다른 클래스 데이터에 대해서는 잘 모델링 되지 않는 변별력을 갖게 된다.

그 첫 번째 방법으로 ALC1(Anti-Likelihood Model selection Criterion1)은 우도(Likelihood)와 반우도(Anti-Likelihood)의 차를 모델 선택 기준으로 사용하였다. 우도는 해당 클래스 데이터 확률 값이며 반우도는 해당 클래스를 제외한 나머지 클래스들의 데이터에 대한 확률 값이다.

$$ALC1 = \frac{1}{N} \sum_{X \in C} \log p(X|\theta_{ML}) - \alpha \frac{1}{N'} \sum_{X' \notin C} \log p(X'|\theta_{ML}) \quad (3)$$

(단, C는 해당 클래스)

위 식에서 반우도($p(X'|\theta_{ML})_{X' \notin C}$)는 해당 클래스 데이터를 제외한 나머지 데이터에 대한 평균 확률을 구하

는 것이다. N' 은 해당 클래스 데이터를 제외한 나머지 훈련 데이터 수이다.

반우도를 이용한 또 하나의 모델 선택 방법으로 BIC와 ALC1를 조합한 ALC2(BIC-Anti likelihood Model Selection Criterion2)가 있다.

$$ALC2 = \frac{1}{N} \sum_{X \in C} \log p(X|\theta_{ML}) - \alpha (\beta \frac{1}{N'} \sum_{X' \notin C} \log p(X'|\theta_{ML}) + (1-\beta) \frac{k}{2} \log N) \quad (4)$$

이 모델 선택 기준은 임의의 penalty 항의 가중치 α 의 적용 비율을 줄이면서 파라미터 수를 감소한다. α 는 패널티 항의 가중치를 나타내고, 임의의 β 는 반우도의 값의 반영 비율을 나타낸다. β 가 커질수록 ALC의 값이 크게 적용된다. 그러나 이 모델 선택 기준은 데이터의 크기와 종류가 다양한 UNIPEN Data에 대해서는 불안정한 향상을 보이는 반면 크기와 종류가 정규화 된 데이터에 대해서는 안정적인 향상을 보이고 있다[11,12].

2.4 MMI(Maximum Mutual Information)

MMI 학습알고리즘은 1986년 IBM 연구소의 Bahn과 그 동료들에 의해 처음 제안되었는데, 관찰열과 그에 대응하는 모델 사이의 상호 정보에 대한 음의 평균치를 비용 함수로 정의하여 이를 최대화시키는 파라미터를 구하는 알고리즘이다[23]. HMM의 파라미터 추정에 많이 사용되는 MLE(Maximum Likelihood Estimation)에 비해 MMI는 모델의 변별력을 높임으로써 인식률을 높일 수 있다. 데이터가 주어졌을 때, 클래스의 집합 $C = \{C_1, C_2, \dots, C_k\}$, m차원의 유한 특징 집합 $F_m = V_1 \times V_2 \times \dots \times V_m$, V_i 는 유한집합이다. 그리고 $\langle X, c \rangle$ $X \in F, c \in C$ 인 n개의 데이터 집합이 주어졌을 때, 서로 다른 클래스를 구별하는 가장 적합한 값의 X의 구성요소를 정하고자 한다. 특별한 구성요소를 결정짓는 적당한 값은 구성요소와 클래스와 상호의존적으로 나타난 정보로 결정한다. 이 값을 계산하기 위해 주어진 데이터의 X의 i 번째 구성요소 X_i 와 주어진 데이터의 클래스 C 두 개의 변수를 정의한다.

두 변수사이의 상호의존 정보는 $I(X_i, C) = H(C) - H(C|X_i)$ 로 구할 수 있다. $H(C)$ 의 값은 모두 같기 때문에 $H(C|X_i)$ 을 계산하는 것으로도 충분하다.

$$H(C|X_i) = - \sum_{c \in C} \sum_{x_i \in V_i} p(c, x_i) \log p(c|x_i) \quad (5)$$

가장 적합한 판별 요소는 이 값이 최대가 되는 값이다. 이 방법은 대상 데이터가 많은 경우 이들 간의 상호 정보를 추출 하려면 많은 시간이 필요하며 어느 정도에서 수렴하게 되는지 명확하지 않은 단점이 있다. MMI 방법은 앞서 언급한 방법과는 달리 HMM 구조 중 최적의 구조를 선택하는 모델 선택 방법이 아니라 정해진

HMM 구조에서 변별력 높은 파라미터를 훈련하는 방법이다[24].

3. 혼동 모델 기준과 혼동 모델을 이용한 인식 방법

대부분의 모델 구조는 대부분 최대 확률 (Maximum Likelihood) 개념을 기반으로 하기 때문에 모델간의 변별력이 없다. 따라서 혼동 패턴 간의 정확성이 떨어지는 원인이 된다. 혼동 패턴 간의 정확성을 높이기 위해서 모델간의 변별력을 향상시키기 위한 방법을 제안하였다.

데이터 간에는 서로 구별이 어려운, 즉 서로 혼동되는 데이터가 존재한다. 숫자에서는 '0'과 '9', '6' 등이, 영숫자에서는 숫자 '0'과 영문자 'O', 'o' 등이 바로 이러한 경우에 해당된다. 따라서 혼동 클래스 데이터가 해당 클래스로 오인식되는 경우가 자주 발생한다. 이런 성질을 이용하여 혼동우도(Confusion Likelihood)와 혼동모델(Confusion Model)을 각각 정의하고, 혼동우도를 이용한 모델 선택 기준인 CMC와 별도로 생성한 혼동 모델을 이용한 인식방법을 제안한다.

3.1 CMC(Confusion Model selection Criterion)

CMC는 해당 클래스 데이터와 혼동되는 클래스의 데이터의 확률 값을 이용하여 혼동우도(Confusion Likelihood)를 생성한다. 이 혼동우도는 파라미터의 수가 증가하는 것을 억제한다. 본 논문에서는 혼동우도를 이용한 모델 선택 방법으로 BIC에 혼동우도를 조합한 CMC(Confusion Model Selection Criterion)를 제안한다. 변별력 있는 모델 선택을 위한 방법으로 혼동모델 선택 기준을 제안한다. 이 방법은 베이저안 모델 선택 방법과 비슷한 과정을 가진다. M 개의 클래스집합과 각 클래스 C 에는 L 개의 모델이 있고, 각 모델 M_{ML} 에는 모델구조 T_{ML} 과 파라미터 θ_{ML} 이 있을 때, 클래스 조건 확률 $P(x|C)$ 는 $P(x|T_{ML}, \theta_{ML})$ 로 실행 할 수 있다. 혼동클래스

C 의 모델 T_{ML} 의 변별력 요소를 $D_{LC}^M = \frac{P(X|T_{ML})}{P(X_C|T_{ML})}$, $X_C = X''$ 정의한다. 여기서 $P(X_C|T_{ML})$ 는 혼동 클래스 C 에 속한 데이터 집합 X_C 의 혼동-증거 항으로 한다. 이 항은 혼동클래스 내의 데이터를 생성하기 위해 관련된 모델을 선택하는데 사용한다. 모델 M_i 의 변별력요소는 기하평균 $D_L^M = \left\{ \prod_{c \in C} D_{LC}^M \right\}^{\frac{1}{M-1}}$ 으로 나타낼 수 있다. 변별력요소에

로그를 취하면 $D(M_{ML}) = \log P(X|\theta_{ML}) - \frac{\sum_{X'' \in C} \log P(X''|\theta_{ML})}{M-1}$

으로 정의된다. 혼동모델을 이용한 변별력요소는 표준 모델의 증거와 혼동모델의 증거의 차로 정의 할 수 있다. 이 변별력요소를 사용하여 모델을 선택할 때 표준

모델의 확률은 크게 혼동모델의 확률이 최소가 되도록 모델을 선택하면, 표준 모델에서는 혼동 가능한 모델을 선택할 확률을 억제하여 최적의 모델이 선택된다.

위의 변별력 요소를 BIC의 패널티 항에 대입하여 다음과 같은 식을 얻을 수 있다.

$$CMC = \frac{1}{N} \sum_{X \in C} \log p(X|\theta_{ML}) - \alpha \left(\beta \frac{1}{N''} \sum_{X'' \in C_c} \log p(X''|\theta_{ML}) + (1-\beta) \frac{k}{2} \log N \right) \quad (6)$$

C : Standard class, C_c : Confusion classes

위 식에서 혼동우도, $p(X''|\theta_{ML})$, $X'' \in C_c$, 는 해당 클래스로 자주 오인식되는 데이터에 대한 확률이고, N'' 은 해당 클래스로 자주 오인식되는 클래스의 훈련 데이터 수이다.

이 모델 선택 기준은 임의의 penalty 항의 가중치 α 의 적용 비율을 줄이면서 파라미터 수를 감소한다. α 는 패널티 항의 가중치를 나타내고, 임의의 β 는 혼동우도의 값의 반영 비율을 나타낸다. β 가 커질수록 혼동우도의 값이 크게 적용된다. 이 모델 선택 기준은 해당 클래스 데이터에 대해 높은 확률을 출력하면서 동시에 혼동 클래스 데이터에 대해서는 높지 않은 확률을 출력하는 모델을 선호함으로써 인식할 때의 변별력을 높여주게 된다.

혼동 모델 선택 기준(CMC)과 2.3절에서 소개한 반우도 모델 선택 기준(ALC2)은 방법적인 면에서 비슷하다. 반우도 모델 선택 기준은 해당 클래스 데이터 이외의 다른 클래스 데이터를 모아 반우도를 계산하는데 사용하며 혼동 모델 선택 기준은 해당 클래스 데이터로 자주 오인식되는 데이터들을 모아 혼동우도를 계산하는 차이가 있다. 이 두 방법은 해당 클래스 데이터에 대해서는 잘 모델링 되고, 다른 클래스의 데이터나 혼동되는 클래스의 데이터에 대해서는 잘 모델링 되지 않게 하는 특징이 있다.

3.2 혼동 데이터 생성

혼동 데이터 집합을 구성하기 위해 먼저 표준 모델을 훈련하고 여기에 데이터를 인식한 후 그 결과로부터 데이터가 인식된 결과와 오인식된 결과를 확인한다. 여기서 이를 사용하여 혼동 데이터 집합을 구성한다. 그림 1에서 그 과정을 보이고 있다.

표 1과 그림 2에서 각각의 해당 클래스 데이터가 오인식된 결과와 이를 바탕으로 각각의 혼동 데이터 집합

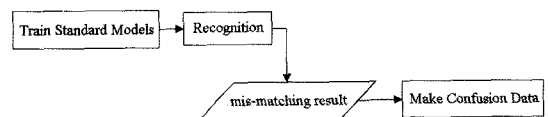


그림 1 혼동 데이터 구성 방법

표 1 혼동 데이터 구성

| 입력 데이터 | 인식 결과 |
|----------|-------|
| '6', '9' | '0' |
| '1', '9' | '7' |

| 혼동 데이터 집합 | 구성 원소 |
|-----------|----------|
| '0' | '6', '9' |
| '7' | '1', '9' |

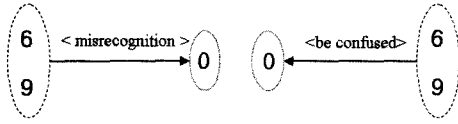


그림 2 혼동 데이터 구성

을 구성한 예를 보였다. 표준 모델을 훈련하고 데이터를 인식 한 후 그 결과를 살펴보면 임의의 데이터가 특정 데이터로 자주 오인식되는 결과를 얻게 된다. 그림 2는 데이터 '6'과 '9'가 데이터 '0'으로 오인식되는 경우를 보였다(왼쪽). 이 결과로부터 혼동 데이터 집합 '0'은 데이터 '6'과 '9'로 구성하였다. 표 1의 데이터 '1', '9'는 데이터 '7'로 자주 오인식 되며, 이 결과로부터 혼동 데이터 집합 '7'은 데이터 '1'과 '9'로 구성되는 것을 보이고 있다.

3.3 혼동 모델을 이용한 인식 방법: RCM(Recognition Using Confusion Model)

CMC는 다양한 구조를 갖는 HMM 중 변별력이 높은 모델을 선택하기 위해 혼동우도를 활용하는 모델선택기준인 반면, RCM은 혼동 모델을 별도로 훈련하여 인식 과정에서 사용하는 방법이다. 해당 클래스 데이터로 자주 오인식되는 데이터의 결과를 분석하여 혼동 데이터 집합을 구성하고 이를 훈련하여 혼동 모델을 만든다. 이 혼동 모델의 확률 값과 표준 모델 확률 값과의 연산을 통해 해당 클래스 데이터가 오인식될 가능성을 억제하고, 이를 통해 유사 클래스 사이의 변별력을 높일 수 있다.

이 인식 방법은 표준모델을 이용하여 입력데이터를 인식한 확률과 혼동모델을 이용한 입력데이터의 입력 확률 값을 비교한다. 표준모델을 이용한 결과에서 오인식된 결과와 혼동모델을 이용한 입력 데이터의 입력 확률 값의 결과와의 차를 이용하여 그 값이 큰 순위의 후보를 최종 인식된 결과로 조정하는 인식방법이다.

식 (7)이 표준모델과 혼동 모델의 차를 이용한 인식 방법(RCM)의 식이다.

$$RCM_i = \frac{1}{N} \sum_{X_i \in C} \log p(X_i | \theta_{ML}) - \alpha \frac{1}{N''} \sum_{X'_i \in C_c} \log p(X'_i | \theta''_{ML}) \quad (7)$$

C: standard class, *C_c*: confusion classes

위 식에서 혼동우도 $p(X'_i | \theta''_{ML})$ ($X'_i \in C_c$)는 해당 클래스로 자주 오인식되는 데이터에 대한 확률이고, N'' 은 해당 클래스로 자주 오인식되는 클래스의 훈련 데이터 수이다. RCM을 적용하기 위해 복수 개의 인식 후보를 생성한 후 식 (7)의 값을 적용하여 표준 모델에 의한 인식 후보를 재정렬한다. 표준 모델에 대한 확률이 높으면

서 동시에 혼동우도의 값이 작은 후보가 최종적으로 선택된다.

혼동 모델은 입력 데이터가 해당 클래스로 오인식될 가능성을 낮추며 해당 모델이 인식 될 확률 값을 향상시켜 해당 모델이 인식 될 확률을 향상 시킨다. 그림 3에서는 그 과정을 보이고 있다.

표 2와 3은 데이터 '0'과 '6'을 각각 인식한 결과 예이다. 표의 왼쪽은 데이터가 잘 인식된 결과이며, 오른쪽은 데이터가 오인식된 결과를 나타내고 있다. 인식 결과를 살펴보면 해당 데이터로 오인식되는 데이터가 자주 발생하는데 그 경우가 표의 오른쪽에 해당한다. 이 오인식된 데이터는 해당 데이터로 자주 오인식되며 확률이 높은 순서로 후보로 발생한다. 각 표의 확률 값이 음의 값을 갖는 것은 로그 확률이기 때문이다.

표 2, 3의 결과처럼 오인식된 경우를 살펴보면 확률이

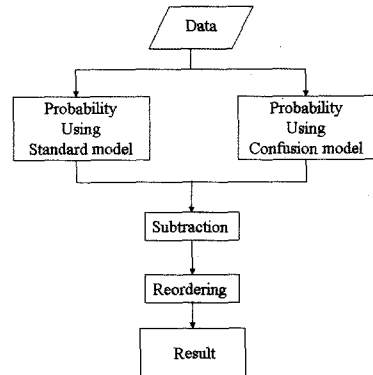


그림 3 혼동모델을 이용한 인식방법의 다이어그램

표 2 표준 모델을 이용한 데이터 '0'의 정인식(왼쪽)과 오인식(오른쪽)

| | '0' | '0' |
|-----|-----|---------|
| '0' | -17 | '8' -72 |
| '8' | -39 | '0' -86 |

※ 0 -17은 class 0과 그 로그 확률 값 -17을 의미한다. 소수점이하는 생략하였다.

표 3 표준 모델을 이용한 데이터 '6'의 정인식(왼쪽)과 오인식(오른쪽)

| | '6' | '6' |
|-----|-----|---------|
| '6' | -24 | '0' -75 |
| '0' | -73 | '6' -90 |

표 4 혼동 모델을 이용한 데이터 '0'의 '6'의 확률

| | | | |
|-----|-----|-----|-----|
| '0' | | '6' | |
| '8' | -36 | '0' | -39 |
| '0' | -56 | '6' | -65 |

가장 높은 1순위 후보에 혼동 데이터가 인식 된 경우가 많았다. 표 4는 훈련된 혼동 모델의 해당 데이터에 대한 혼란한 확률 값이다. 혼동 모델로 해당 데이터를 인식하였으므로 그 결과에는 혼동 데이터들이 확률 값이 1순위로 나타난 경우가 많다. 이렇게 각각 표준 모델과 혼동 모델을 이용하여 1순위와 1개 이상의 후보 인식 결과를 얻는다. 본 논문에서는 각각의 알고그래프(allograph)에 대해 후보를 내기 위해 25개의 후보결과를 얻었다. 표준 모델과 혼동 모델로 나온 결과에서 서로 같은 데이터의 확률 값의 차를 이용하여 그 값이 큰 순위의 후보를 최종 인식된 결과로 조정한다.

그림 4와 그림 5는 인식방법 RCM을 적용하는 예를 보여주고 있다.

정인식의 결과와 오인식 결과에 공통으로 사용하게 될 혼동모델을 이용한 1순위 데이터 '8'의 확률 값은 -36, 2순위 데이터 '0'의 확률 값은 -56이다. 그림 4는 표준 모델을 이용한 정인식 후보를 혼동 모델 확률과 연산하여 나온 결과를 보이고 있는 것이다. 표준모델을 이용한 정인식 데이터 '0'의 1순위 확률 값은 -17이고, 2 순위 데이터 '8'의 확률 값은 -39이다. 이를 혼동 모델을 이용한 결과와 연산하면 그 값은 데이터 '0'에 대한 결과는 39, '8'에 대한 결과는 17이다. 0에 대한 값이

더 크므로 이는 교환하지 않아도 된다.

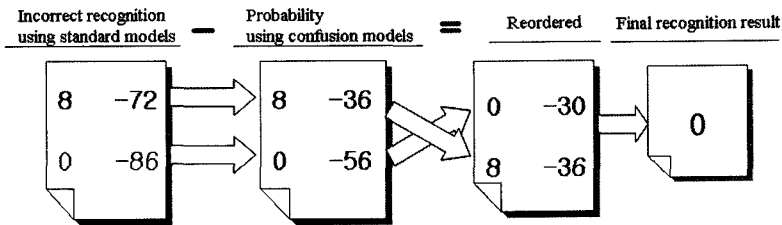
그러나 오인식에 대한 결과 그림 4에 대해서는 오인식 1순위 데이터 '8'은 -72, 2순위 데이터 '0'은 -86이며 이를 혼동 모델을 이용한 확률 값과 연산을 하면 그 값은 '0'은 -30, '8'은 -36이므로 표준 모델을 이용한 오인식 결과의 순위를 바꿔준다.

혼동될 데이터의 인식 확률 값을 정인식된 인식 확률 값의 패널티로 적용하여 혼동 데이터로 인식될 가능성을 낮춘다.

4. 실험 및 결과 분석

4.1 모델 선택 처리 절차 및 인식 실험

실험에서 사용할 데이터 집합을 HMM 훈련을 위해 훈련 데이터 집합과 교차검증을 위한 held-out 집합, 테스트를 위한 테스트 집합으로 분할하였다. 첫 번째로 훈련 데이터 집합으로 상태 수와 상태 당 믹스처 수를 변화시키면서 다양한 구조를 갖는 HMM을 훈련시킨다. HMM의 훈련이 끝나면 held-out 데이터 집합을 사용하여 훈련 된 표준 모델의 확률을 계산한다. 기존의 모델 선택 기준을 이용하여 좋은 HMM 구조를 선택 한 후 이를 이용하여 인식 결과를 얻는다. 이 인식 결과에서 오인식된 결과를 살펴본 후 이를 이용하여 혼동 데이터 집합을 구성한다. 이렇게 구성된 혼동 데이터 집합으로 다시 다양한 HMM 구조를 상태 수와 상태 당 믹스처 수를 변화시키면서 훈련시킨다. 그리고 준비된 held-out 데이터 집합을 사용하여 훈련 된 혼동 모델의 확률을 계산한다. 이렇게 계산된 표준 모델의 확률과 혼



* 0 -17은 클래스 0과 그 로그 확률 값 -17을 의미한다. 소수점이하는 생략하였다

그림 4 오인식 데이터에 대한 혼동 모델 확률을 이용한 인식결과

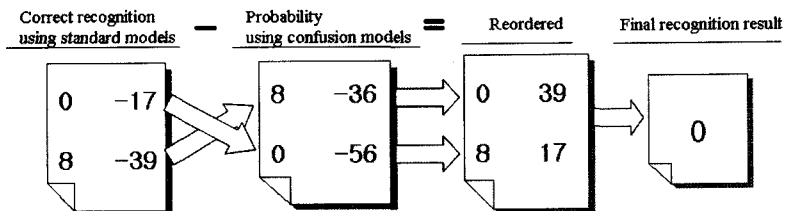


그림 5 정인식 데이터에 대한 혼동 모델 확률을 이용한 인식결과

동 모델의 확률로 각각 데이터에 대한 확률 값을 계산한다. 각각의 확률 값은 여러 개의 후보를 계산하고 이 값들의 연산을 통해 최종 인식결과를 얻는다.

그림 6은 모델 선택 처리와 선택된 모델로 인식 실험의 절차를 보이고 있다. Construction of confusion data set 블록에서는 모델 선택 기준 BIC 등을 이용하여 선택된 모델로 혼동데이터를 구성한다. 구성된 데이터를 이용하여 혼동 데이터를 다양한 구조를 갖는 HMM을 훈련하고 이렇게 선택된 훈련 모델 확률 값과 미리 계산한 표준 모델의 확률 값과의 차를 이용하여 그 값이 큰 값을 최종 인식 결과 값으로 선택하는 과정을 보이고 있다.

4.2 실험 데이터베이스

본 논문에서는 UNIPEN 데이터를 실험 대상으로 삼았다. UNIPEN 데이터 온라인 문자 인식에 관계된 세계 각국의 대학, 연구소, 기업 등 다양한 기관들이 공통의 파일 표준을 만들어 필기 데이터를 모아 놓은 것인데, 그 중 train_r01_v07을 사용하였다. UNIPEN 데이터에는 숫자, 영대소문자, 부호 등이 있지만, 본 논문에서는 숫자만을 실험 대상으로 하였다. 표준모델에 대해 실험한 데이터는 382명의 필기자로부터 수집한 9436개의 샘플을 훈련 집합으로 사용했고, 131명의 필기자로부터 얻은 3296개의 샘플을 교차 검증 집합으로 사용했다.

숫자만을 실험 대상으로 했기 때문에 모두 10개의 클래스가 있다. 필기 형태가 서로 상이한 것을 기준으로 별도의 알고그래프를 만들어 총 25개의 알고그래프를 생성하였다. 필기 데이터를 먼저 크기 정규화한 후 각 획에서 특징점을 찾고 특징점 사이의 데이터에 대해 특징을 추출하여 PCA(Principal Component Analysis) 과정을 거쳐 총 9차원 벡터를 생성하였다.

혼동 모델에 대해서는 실험한 데이터는 5050개의 샘플을 훈련 집합으로 사용했고, 2659개의 샘플은 교차 검증 집합으로 사용했다. 실험을 위한 테스트 데이터로는 121명의 필기자로부터 획득한 3221개의 샘플을 사용하였다.

4.3 혼동 데이터

실험에 사용할 데이터는 필기 형태가 서로 다른 것을 기준으로 해서 별도의 알고그래프를 만들어, 클래스 '0'의 경우 'D_00', 'D_01', 데이터 '1'의 경우 'D_10', 'D_11', 'D_12', 클래스 '9'의 경우에는 'D_90', 'D_91'와 같이 총 25개의 알고그래프로 구성되어 있다.

혼동 모델 선택 기준은 정확한 혼동 데이터 집합을 만들기 위해 한 번의 인식 실험을 통해 오인식 결과를 얻는 과정이 필요하다. 실험을 위한 데이터 훈련과 인식률 추정은 HTK을 이용하였다[25]. 그림 7은 한 번의 인식 과정을 통한 해당 클래스 데이터가 인식된 결과를 보이고 있다. 그림 7에서 보면 'D_13'가 'D_70', 'D_90'가 'D_01'으로 오인식된 것을 볼 수 있다. 이처럼 해당 클래스 데이터로 특정 데이터가 오인식되는 빈도 수가 많다는 것을 알 수 있다.

표 5 실험에서 사용 할 표준 모델

| Class | allograph |
|-------|---|
| '0' | 'D_00','D_01' |
| '1' | 'D_10','D_11','D_12','D_13','D_14','D_15' |
| '2' | 'D_20','D_21' |
| '3' | 'D_30' |
| '4' | 'D_40','D_41','D_42' |
| '5' | 'D_50','D_51','D_52' |
| '6' | 'D_60' |
| '7' | 'D_70','D_71','D_72','D_73' |
| '8' | 'D_80','D_81' |
| '9' | 'D_90' |

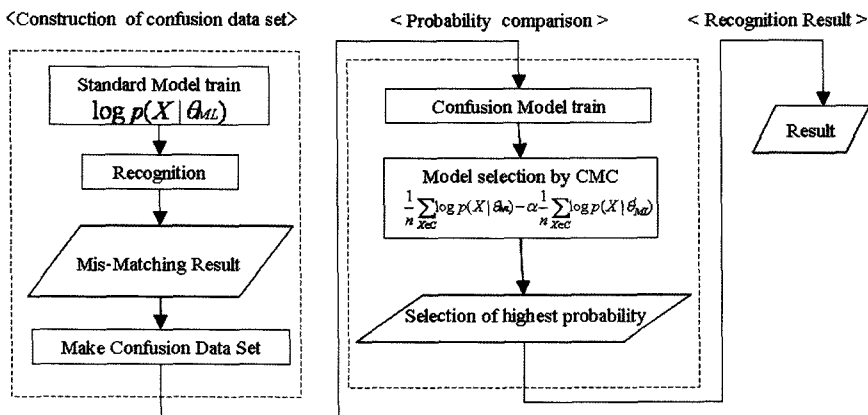


그림 6 모델 선택 및 실험 방법

해당 클래스 데이터로 자주 오인식되는 데이터를 대상으로 혼동 데이터 집합을 구성한다. 같은 클래스이지만 필기 대상에 따라 일정한 특징을 갖는 데이터의 모임으로 allograph를 만든 것은 이들만이 갖고 있는 일정한 특징이 오인식 요인이 될 수 있기 때문이다. 해당 클래스 데이터를 allograph 각각에 대해 인식하여 오인식 결과를 얻는다. 표 3은 해당 클래스 데이터의 오인식 결과를 보여주고 있다. 이를 살펴보면 데이터 'D_60', 'D_80', 'D_81', 'D_90' 등이 'D_00'으로 오인식된다. 따라서 'D_60', 'D_80', 'D_81', 'D_90'은 혼동 데이터는 'D_00'의 구성 요소가 된다. 표 7에서는 표 6의 오인식 결과로부터 각각의 혼동 데이터 집합이 구성된 결과를 보여 주고 있다.

| | D_00 | D_01 | D_10 | D_11 | D_12 | D_13 | D_72 | D_73 | D_80 | D_81 | D_90 | Ins |
|------|------|------|------|------|------|------|------|------|------|------|------|-----|
| D_00 | 193 | 24 | 0 | 0 | 0 | 1 | 1 | 5 | 0 | 0 | 0 | 4 |
| D_01 | 2 | 69 | 2 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| D_10 | 0 | 0 | 1 | 229 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 |
| D_11 | 0 | 0 | 0 | 20 | 0 | 7 | 0 | 0 | 0 | 0 | 1 | 0 |
| D_12 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| D_13 | 0 | 0 | 0 | 9 | 4 | 0 | 15 | 0 | 7 | 0 | 0 | 1 |
| D_72 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 0 |
| D_73 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 39 |
| D_80 | 4 | 5 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 292 |
| D_81 | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 34 |
| D_90 | 2 | 15 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 2 |
| Ins | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

그림 7 인식·오인식 결과

표 6 오인식 결과(일부)

| 데이터 | 오인식 결과 |
|---|--------|
| 'D_60','D_80','D_81','D_90' | 'D_00' |
| 'D_41','D_60','D_80','D_81','D_90' | 'D_01' |
| 'D_00' | 'D_60' |
| 'D_00','D_01','D_41','D_42','D_50','D_51','D_52','D_81' | 'D_90' |

표 7 혼동 모델 구성(일부)

| 혼동 데이터 | 구성 데이터 |
|--------|---|
| 'D_00' | 'D_60','D_80','D_81','D_90' |
| 'D_01' | 'D_41','D_60','D_80','D_81','D_90' |
| 'D_60' | 'D_00' |
| 'D_90' | 'D_00','D_01','D_41','D_42','D_50','D_51','D_52','D_81' |

4.4 실험 결과

그림 8은 기존 모델 선택 기준 BIC, ML, ALC2와 제안한 모델 선택 기준인 CMC를 숫자 필기 데이터에 대해 실험한 파라미터 수를 보여주고 있다.

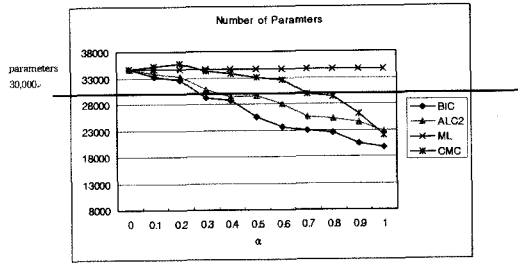


그림 8 각 모델 선택 기준의 Parameter 수

그림 8의 x축은 각 모델 선택 기준의 패널티 가중치 α 의 증가를 보이고 있다. y축은 그에 따라 변하는 파라미터의 수를 보이고 있다. α 가 증가함에 따라 ML을 제외한 모든 모델 선택 기준의 파라미터 수가 감소하는 경향을 보이고 있다. BIC는 α 값에 따라 패널티 항의 가중치가 커지면서 파라미터의 수가 감소하고 있고, ALC2도 α 값에 따라 반우도의 가중치가 점점 커지면서 파라미터의 수가 감소하고 있음을 알 수 있다. CMC의 경우에도 α 값에 따라 파라미터의 수가 감소하고 있는데 이는 패널티 항으로 사용되는 혼동우도의 비율이 점점 높아지기 때문이다. 파라미터의 수가 약 30,000로 같을 때 그림 9의 인식률을 살펴보면 BIC는 약 91.4%, ALC2는 약 91.3%, 제안한 모델 선택 기준 CMC는 약 92.4%로 좋은 성능을 보이고 있다. 반우도(ALC2)나 혼동우도(CMC) 값을 반영하면 기존의 모델 선택 기준만을 이용한 실험결과보다 좋은 결과를 얻을 수 있다.

그림 9는 기존 모델 선택 기준 BIC, ML, ALC1, ALC2와 제안한 모델 선택 기준인 CMC를 숫자 필기 데이터에 대해 실험한 인식률 결과를 보여주고 있다. 그림 9에서 α 가 0.0으로 적용될 때 인식률은 모든 모델 선택 기준에 대해 91.8%의 성능을 보이며, α 가 0.5로 적용될 때의 인식률은 모델 선택 기준 CMC는 92.64%, BIC는 92.52%의 성능을 보이며 α 가 0.6으로 적용되면 ALC2는 92.24%의 성능을 보인다. 이와 같은 결과에서 모델을 선택 할 때 모델내의 파라미터 개수에 대한 패

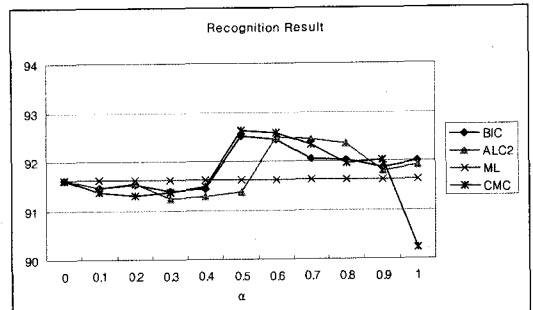


그림 9 각 모델 선택 기준의 인식률

널티가 주어지면 그렇지 않은 경우보다 그 인식률의 성능이 더 좋아지는 것을 알 수가 있다. 하지만 패널티의 비중을 너무 크게 주면 파라미터의 수가 급격히 감소하게 되고 이는 HMM의 모델링 능력을 떨어뜨리는 결과를 가져온다.

표 8은 제안한 모델 선택 기준 CMC와 기존 모델 선택기준의 비교 결과이다. 표 8은 최대의 인식률이 나왔을 때의 가중치 α 와 파라미터의 수를 보여주고 있다. 최고의 인식률을 얻었을 때 파라미터 수를 살펴보면 제안한 모델 선택 기준 CMC는 파라미터 수 22,230, 인식률 92.24%이고, BIC는 파라미터 수 25,593, 인식률 92.52%이다. CMC가 BIC 보다 인식률면에서는 크게 향상되지는 못하였으나 파라미터수를 줄이고 비슷한 인식률을 구하는데 성공하였다.

표 8 최대의 인식률 일 때 가중치 α 와 파라미터의 수

| Model Selection Criteria | Parameter | Recognition rate(%) |
|--------------------------|-----------|---------------------|
| ML | 34,697 | 91.62 |
| BIC($\alpha = 0.5$) | 25,593 | 92.52 |
| ALC2($\alpha = 0.6$) | 27,994 | 92.49 |
| CMC($\alpha = 0.5$) | 33,000 | 92.64 |

그림 10은 표준 모델과 혼동 모델의 확률을 이용한 최종 인식 결과를 보이고 있다. 표준 모델과 혼동 모델을 이용하여 인식하여 후보를 각각 22개를 구한다. 표준 모델로 인식한 결과 혼동모델을 이용한 확률 값과의 차를 이용한 인식 결과이다. x 축의 α 는 혼동모델의 값에 패널티를 준 항이다.

제안한 혼동 모델 인식 방법인 RCM은 α 가 0.1 일 때 93.08%의 인식률을 보인다. 표준 모델만을 이용하여 인식한 방법보다 약 1.5%향상되었고 17.4%의 오류 감소율을 보이고 있다. ML보다 우수한 성능을 보이는 BIC 보다 인식률이 약 1% 향상되었으며, 약 12.6%의 오류 감소율을 보인다.

혼동모델을 이용한 확률 값은 해당 클래스의 데이터와 혼동되는 데이터의 집합을 이용한 것이다. 이 값이

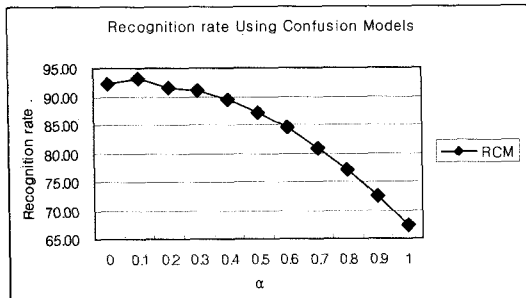


그림 10 혼동모델 인식방법으로 새로 인식한 인식결과

많이 적용이 되면 조금이라도 혼동되는 데이터의 확률 값이 많이 떨어지게 되므로 올바르게 인식될 가능성 또한 떨어지게 된다. 그러므로 혼동 모델을 이용한 확률 값은 적절히 적용이 되면 좋은 인식률을 얻을 수 있을을 위 그림 10에서 알 수 있다.

5. 결론 및 향후과제

본 논문에서는 혼동 모델의 확률을 이용한 혼동 확률 선택 기준인 CMC(Confusion Model Selection Criterion)와 혼동 클래스 데이터를 구성하여 혼동 모델을 만들고 이것을 이용한 새로운 인식 방법 RCM(Recognition using Confusion Models)을 제안하였다. 다른 클래스 데이터가 해당 클래스 데이터로 자주 오인식되는 특징을 이용하여 해당 클래스로 자주 오인식되는 데이터들의 확률을 이용하여 CMC 모델 선택 기준을 제안하였다. 이 선택 기준은 반우도를 이용한 선택 기준 ALC2 보다 적은 파라미터로 더 좋은 결과를 얻었다. ALC2가 해당 클래스를 제외한 나머지 모든 클래스들의 데이터를 사용한 반면 CMC에서는 자주 혼동되는 클래스들의 데이터만 사용함으로써 더 좋은 결과를 얻을 수 있었다.

혼동 모델을 이용한 새로운 인식 방법은 다른 클래스 데이터가 해당 클래스 데이터로 혼동하는 경우를 이용하여 혼동 데이터 집합을 만들고 이것으로 별도의 모델을 훈련하였다. 이 모델을 이용한 모델 선택기준으로 해당데이터의 혼동 모델을 이용한 확률을 구하고 표준 모델을 이용한 해당 클래스 데이터 확률과의 차를 이용하여 그 값을 재정렬하여 최종 인식 결과로 하는 방법을 사용하였다. 혼동모델을 이용한 인식 방법은 해당 클래스 데이터가 오인식 될 가능성을 억제하여 혼동되는 데이터간의 변별력을 높인다. 제안한 혼동 모델 인식 방법인 RCM은 93.08%의 인식률은 표준 모델만을 이용하여 인식한 방법보다 인식률은 약 1.5% 향상되었고, 이는 오류의 17.4%를 감소시킨 결과이다.

본 논문에서는 UNIPEN 데이터 중 필기 숫자 데이터에 대해 실험하여 제안한 방법의 유용성을 보였다. 향후 연구 과제로는 필기 숫자 이외의 여러 문자 클래스에 대한 성능 평가와, 필기 인식뿐만 아니라 음성 인식 등 HMM이 사용되는 다른 여러 분야에의 적용 등이 있다.

참고 문헌

[1] 박미나, 하진영, "HMM 모델링을 위한 HMM의 State 수와 Mixture 수 분석", 한국정보과학회 춘계 학술발표논문집, 제29권 제1호, pp. 658-660, 2002년 4월.
 [2] 하진영, A. Biem, J. Subrahmonia, 박미나, "모델의 사전 확률 추정을 통한 HMM 구조의 최적화", 한국정

- 보과학회 추계 학술발표논문집, 제28권 제2호, pp. 325-327, 2001년 10월.
- [3] H.Singer and M. ostendorf, "Maximum likelihood successive state splitting," ICASSP, pp. 601-604, 1996.
- [4] Andress Stolcke and stephen Omohundro, "Hidden Markov Model induction by Bayesian model merging," in Advances in NIPS, Vol.5. pp. 11-18, 1993.
- [5] 하진영, 신봉기, "온라인 한글 인식을 위한 HMM 상태 수의 최적화", 한국정보과학회 추계 학술발표논문집, pp. 372-374, 1998.
- [6] D. Li, A. Biem and J. Subrahmonia, "HMM Topology Optimization for Handwriting Recognition," ICASSP 2001.
- [7] S. Chen and P. S. Gopalakrishnan. "Clustering via the Bayesian Information criterion with applications in speech recognition," ICASSP, 2, pp. 645-649, 1998.
- [8] Raymond C.Vasko, Jr., Amro EL-Jaroudi J.Robert, Boston "An Algorithm To Determine Hidden Markov Model Topology," IEEE, pp. 3577-3579, 1996.
- [9] Raymond C.Vasko, Jr., Amro EL-Jaroudi J.Robert, Boston "Application of Hidden Markov Model Topology Estimation To Repetitive Lifting Data," IEEE, pp. 4073-4076, 1997.
- [10] Schwarz, "Estimating the dimension of a model," Ann, Statist, Vol.6, No.2, pp. 461-464, 1978.
- [11] Mi-Na Park, Jin-Young Ha, "On-Line Handwritten Numeral Recognition using Topology Optimized HMMs," The 5th Asia Pacific International Symposium on Information Technology, ISSN 1738-9062, pp. 87-290, 2006.
- [12] Mi-Na Park, Jin-young Ha, "HMM Topology Optimization for Handwriting and Speech Recognition," The 20th Commemorative International Technical Conference on Circutes, Systems, Computer and Communications, Proceedings Volumel, pp. 261-262, July 2005.
- [13] <http://clopinet.com/isabelle/Projects/modelselect/>
- [14] Walter Zucchini, "An Introduction to Model Selection," Journal of Mathematical Psychology 44, 41-61, 2000.
- [15] R. A. Fisher, "The use of multiple measure in taxonomic problems," Cont. Math. Stat., pp. 32. 179-32. 188, 1950.
- [16] L.R. Bahl, P. F. Brown, P. V. de Souza, and R. L. Mercer, "Maximum mutual information estimation of hidden Markov model parameters for speech recognition," in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, Tokyo, Japan, 1986, pp. 49-52.
- [17] Belén Martín Barragán, "Binarized Support Vector Machines: detecting relevant variables and interactions," International Workshop on Feature Selection for Data Mining:Interfacing Machine Learning and Statistics(FSDM), 2006.
- [18] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," IEEE Trans. Inform. Theory, Vol.44, pp. 2743-2760, 1998.
- [19] MacKay, "Bayesian interpolation," Neural computation, Vol.4, No.3, pp. 415-447, 1992.
- [20] E. Kass and A. Raftery. "Bayes factors," Technical Report 254, University of Washington, Department of Statistics. 1994.
- [21] A. Biem, J.-Y. Ha and J. Subrahmonia, "A Bayesian Model Selection Criterion for HMM Topology Optimization," ICASSP 2002. Orlando, Florida, U.S.A., pp. I-989 - I-992, May, 2002.
- [22] Jin-Young Ha, Alain Biem and Jayashere Subrahmonia "Use of Model Prior for HMM Topology Optimization," The 4th Korea-China Joint Symposium on Information Technology for Oriental Language Processing and Pattern Recognition, 2001.
- [23] L. R. Bahl, P. F. Brown, P. V. de Souza, and R. L. Mercer, "Maximum mutual information estimation of hidden Markov model parameters for speech recognition," Proc. ICASSP-86, pp. 49-52, 1986.
- [24] Nopsuwanchai Roongroj and D. Povey, "Discriminative training for HMM-based offline handwritten character recognition," Proc. of 7th International Conference on Document Analysis and Recognition, Vol.1, pp. 114-118, 2003.
- [25] Steve Young, Dan KerShaw, Julian Odell, Dave Ollason, Valtcho Valtchev, Phil Woodland, The HTK Book(for HTK version 3.0), Microsoft Corporation, 1999.



박미나

1999년 8월 강원대학교 컴퓨터공학과(학사). 2002년 8월 강원대학교 컴퓨터정보통신공학과(석사). 현재 동 대학원 박사과정, 관심분야는 문자인식, 음성인식, HCI



하진영

1987년 2월 서울대학교 전자계산기공학과(학사). 1989년 2월 한국과학기술원 전산학과(석사). 1994년 2월 한국과학기술원 전산학과(박사). 1994년 3월~1997년 2월(주)핸디소프트 기술연구소. 1997년 3월~현재 강원대학교 컴퓨터학부 교수
2007년 7월~2001년 7월 IBM T.J. Watson Research Center 방문연구원. 관심분야는 패턴인식, HCI 등