베이스 에러율의 상위 경계 최소화에 기반한 고차 곱 근사 방법과 숫자 인식기 결합에의 적용

(A High Order Product Approximation Method based on the Minimization of Upper Bound of a Bayes Error Rate and Its Application to the Combination of Numeral Recognizers)

강 희 중 †
(Hee-Joong Kang)

요 약 다수의 인식기를 결합하여 베이지안 결정 이론 하에서 클래스 분별력을 높이려면, 훈련 데이터 샘플로부터 얻은 클래스 변수와 결정 변수들로 구성된 조건부 엔트로피에 의해서 한정되는 베이스 에러율의 상위 경계를 최소화해야 한다. Wang과 Wong은 베이스 에러율의 상위 경계를 최소화하기 위하여 클래스 변수와 다수의 특징 패턴 변수들로 구성된 고차 확률 분포를 트리 의존관계로 근사하는 1차 근사 방법을 제안하였다. 본 논문에서는 이러한 베이스 에러율의 상위 경계 최소화에 기반한 기존의 1차 트리 의존관계 근사 방법을 확장하여 고차 의존관계까지 고려할 수 있는 확장된 곱 고차 근사 방법을 제안한다. 제안된 근사 방법을 CENPARMI의 무제약 필기 숫자를 인식하는 다수의 숫자 인식기 결합 방법에 적용하여 인식 실험을 하였으며, 이 방법에 의해서 보다 높은 인식률을 얻게 되었다.

Abstract In order to raise a class discrimination power by combining multiple classifiers under the Bayesian decision theory, the upper bound of a Bayes error rate bounded by the conditional entropy of a class variable and decision variables obtained from training data samples should be minimized. Wang and Wong proposed a tree dependence first-order approximation scheme of a high order probability distribution composed of the class and multiple feature pattern variables for minimizing the upper bound of the Bayes error rate. This paper presents an extended high order product approximation scheme dealing with higher order dependency more than the first-order tree dependence, based on the minimization of the upper bound of the Bayes error rate. Multiple recognizers for unconstrained handwritten numerals from CENPARMI were combined by the proposed approximation scheme using the Bayesian formalism, and the high recognition rates were obtained by them.

1. 서 론

인식기가 단일 선택을 하는 결정 수준[1]에서 다수 인식기를 결합하는 기존의 방법으로는 투표 방법[1], BKS (Behavior-Knowledge Space) 방법[2], Dempster-Shafer 방법[3,4] 그리고 독립 가정에 의한 베이지안 방법[1,5] 또는 의존관계에 기반한 곱 근사에 의한 베이지안 방법 [6, 7] 등이 있다. 특히, [7]에서와 같이 확률 이론과 베이지안 공식을 이용하여 단일 선택 결정 수준에서 다수 인식기를 결합하는 것은 다음과 같은 수식으로 정의할 수 있다. 입력 x가 K개 인식기(예를 들면, C_1, C_2, C_K)에 동시에 제공되면, K차원의 결정 벡터인 $< C_1(x)=M_1, C_2(x)=M_2, \cdots, C_K(x)=M_K>$가 관찰된다. L개의 인식 대상 클래스 집합은 $M=\{m_1, m_2, \cdots, m_L\}$으로 정의되고, 결정 변수 M_j는 인식기 C_j의 결정 클래스를 나타낸다. 이렇게 관찰된 K차원의 결정 벡터로부터 베이지안 결정 이론 하에서 다수 인식기를 결합하는 주요한 작업은

$$\max_m P(m\in M \mid C_1(x)=M_1, C_2(x)=M_2, \cdots, C_K(x)=M_K)$$

인 사후 확률 P^*를 최대화하는 가설 결정 m을 정하는 것이다. 이를 위하여, $(K+1)$차 확률 분포는 훈련 데이타

· 본 연구는 2001년도 한성대학교 교내연구비 지원과제임.
† 종신회원 : 한성대학교 정보전산학부 교수
hjkang@hansung.ac.kr
논문접수 : 2001년 1월 19일
심사완료 : 2001년 7월 20일

샘플로부터 생성되고 평가되어야 하는데, 수식으로 표현하면 $P(m, C_1(x)=M_1, C_2(x)=M_2, \cdots, C_K(x)=M_K)$와 같다.

이와 같은 고차 확률 분포를 평가하기 위한 기존의 연구는 주로 각 인식기의 결정이 주어진 가설 결정 m에 대하여 조건부 독립이라는 가정을 취하거나[1,5], Huang과 Suen의 BKS 방법에서와 같이 인식기의 독립 가정을 취하지 않고[2], 훈련 데이타에 의해 얻어진 다수 결정 결과로부터 누적된 고차 빈도수 테이블 정보로부터 직접 $(K+1)$차 확률을 평가하려 시도하였다[7]. 최근에 강희중 등은 [6,7]에서 이러한 가정을 완화시키고, Chow와 Liu가 [9]에서 제안한 1차 의존관계에 의한 곱 근사로 고차 확률 분포를 평가하는 근사 방법을 고차 의존관계에까지 확장하기 위하여, Lewis가 [8]에서 제안한 유사도를 이용하여 2차 의존관계에까지 확장시킨 근사 방법을 [6]에서 제안하였다. 그 후, 이를 기반으로 3차 이상의 고차 의존관계에까지 일반화시켜 확장시킨 근사 방법은 [7]에서 제안되었다. 제안된 근사 방법들을 이용하여 다수 인식기를 결합하기 위한 방법도 역시 제안되었다.

고차 확률 분포의 곱 근사를 결정하는 방법에 있어서, Chow와 Liu가 [9]에서 유사도를 이용하여 제안한 트리 의존관계 방법과는 달리, Wang과 Wong은 [10]에서 클래스가 잘 구별되지 않는 패턴인식 문제에 베이스 에러율의 상위 경계 최소화에 기반한 트리 근사 방법이 클래스의 구별에 더 잘 적합하다고 주장하였다. 또한, 클래스 변수와 다수의 특징 패턴 변수들을 다루기 위하여 CP(Classes-Patterns) 상호 정보를 정의하였고, 패턴 변수들의 최적 트리 구조를 찾기 위한 알고리즘을 제안하였다. 그런데, 베이스 에러율은 [11]에서 소개된 바와 같이 조건부 엔트로피와 이로부터 유도되는 상호 정보의 정의와 관계가 있다. 따라서, 상호 정보에서 사용되는 고차 확률 분포에 대한 최적의 곱 근사는 이러한 베이스 에러율의 상위 경계의 최소화에 기반한다고 볼 수 있다.

본 논문에서는 클래스 분별력을 높이는 방법으로 다수 인식기를 결합하기 위하여 Wang과 Wong의 접근 방법에서처럼, 새로이 CD(Classes-Decisions) 상호 정보를 정의하여 사용한다. 정의된 CD 상호 정보는 클래스 변수와 인식기들의 결정 변수 벡터 사이의 의존관계를 측정하는데, [10]에서와 같이 베이스 에러율의 상위 경계 최소화에 이용된다. 또한, 본 논문에서는 베이스 에러율의 상위 경계 최소화에 기반한 Wang과 Wong의 트리 의존관계 근사 방법을 2차 의존관계에 까지 고려하도록 확장된 고차 근사 방법을 제안하였으며, 이를 다수의 숫자 인식기의 결합에 적용한 새로운 베이지안 결합 방법을 제안하였다. 이와 같이 제안된 새로운 베이지안 결합 방법은 CENPARMI(Center for Pattern Recognition and Machine Intelligence, Concordia University) 무제약 필기 숫자 데이타베이스를 사용하여 실험되었으며, 기존의 다양한 베이지안 결합 방법 보다 높은 인식률을 얻었다. KAIST와 전북대학교에서 개발된 역전파 신경망 기반으로 구현된 4개의 인식기 *C1, C2, C3, C4*와 규칙 기반으로 구현된 2개의 인식기 *C5, C6* 등, 총 6개의 인식기를 실험에 사용하였다.

본 논문의 구성은 다음과 같다. 2.장에서는 베이스 에러율의 상위 경계 최소화에 기반한 곱 근사 방법과 관련된 이론적 배경 및 설명, 2차 의존관계에 의한 최적의 곱 근사를 구하는 알고리즘 등을 기술하고, 3.장은 제안된 고차 근사 방법을 이용하여 베이지안 방법으로 다수 인식기를 결합하는 방법을 소개하며, CENPARMI 데이타베이스로 제안된 방법을 실험한 결과 및 분석은 4.장에서 기술하고, 5.장에서 결론을 맺는 순서로 되어 있다.

2. 베이스 에러율의 상위 경계 최소화에 기반한 곱 근사 방법

앞서 기술한 $(K+1)$차 확률 분포를 베이스 에러율의 상위 경계 최소화에 기반하여 근사하는 방법을 기술하기 위하여 사용되는 심볼 또는 변수를 먼저 정의한다. 변수 C는 K개 인식기 결정들에 의한 K차원 결정 벡터 변수를 나타내며, 이 벡터의 원소인 C_j는 주어진 입력에 대한 j번째 인식기의 결정을 나타낸다. 또한, 변수 m은 클래스 집합 M의 원소를 가리키는 클래스 변수를 나타낸다. $(K+1)$차 확률 분포를 곱 근사 분포로 평가하기 위해서는 고려할 의존관계의 차수에서 최적의 곱 근사를 구하도록 곱 근사를 결정해야 한다. 이러한 곱 근사 결정 단계에서는 근사 분포가 베이스 에러율 P_e의 상위 경계 최소화에 얼마나 기여했는가를 평가하는 것이 필요하다. 베이스 에러율 P_e의 정의는 아래 식(1)에서와 같이 조건부 엔트로피를 구성하는 CD 상호 정보인 $D(m;C)$에 의존한다. 왜냐하면, $H(m)$은 변수 m의 엔트로피로서 상수이기 때문이다.

$$P_e \leq H(m \mid C) = \frac{1}{2}(H(m) - D(m;C)) \tag{1}$$

$$D(m;C) = \sum_{M}\sum_{C} P(m, C) \log \frac{P(m, C)}{P(m)P(C)} \tag{2}$$

위 식(2)에서와 같이 정의된 CD 상호 정보 $D(m;C)$는 결정 벡터의 발생이 클래스 원소의 발생 가능성에

대해서 평균적으로 알려주는 정량적인 기준으로서 정의된다. 식 (1)로부터, 베이스 에러율 P_e의 상위 경계를 최소화하는 것은 CD 상호 정보인 $D(m;C)$를 최대화하는 것과 같음을 명백히 알 수 있다.

Wang과 Wong은 p차 변수의 확률 분포를 $(p-1)$개의 2차 구성 분포의 곱으로 최적으로 근사하는 방법을 [10]에서 제안하였다. 그렇지만, 이들의 방법은 1차 의존관계만을 고려했기 때문에, 고차 의존관계를 고려하기에는 적합하지 않았다. 본 논문에서는 훈련 데이타 샘플로부터 얻은 $(K+1)$차 확률 분포를 2차 의존관계에 의한 최적의 곱 근사 분포로 근사하기 위하여 베이스 에러율의 상위 경계 최소화에 기반한 고차 곱 근사 방법을 제안한다. 제안된 곱 근사 방법은 Wang과 Wong에 의한 1차 의존관계 트리 방법을 2차 의존관계로까지 확장한 것이다. 확률 분포 항의 수식 표현을 간단히 하기 위하여 $C_j(x)=M_j$를 C_j로, $m\in M$를 m으로 각각 간략하게 표현한다.

결정 벡터 C의 K차 고차 확률 분포에 대하여 1차 의존관계를 고려하면, 아래 식(3)과 같은 2차 구성 분포의 곱으로 정의된다.

$$P_a(C_1,\cdots,C_K)=\prod_{j=1}^{K}P(C_{n_j}\mid C_{n_{i(j)}}),\ (0\le i(j)<j) \tag{3}$$

위 식(3)에서 C_{n_j}는 $C_{n_{i(j)}}$에 조건부로 정의되고, 역시 $(n_1,\cdots,n_K)$는 1부터 K까지의 정수에 대한 미지 순열을 의미한다. 그리하여, 변수 C와 m으로 구성된 $(K+1)$차 고차 확률 분포는 아래 식(4)와 (5)에서와 같은 3차 구성 분포의 곱으로 정의된다.

$$P_a(m,C_1,\cdots,C_K)=\prod_{i=1}^{K}P(C_{n_j}\mid C_{n_{i(j)}},m),\ (0\le i(j)<j) \tag{4}$$

$$P_a(C_1,\cdots,C_K\mid m)=\frac{1}{P(m)}\prod_{j=1}^{K}P(C_{n_j}\mid C_{n_{i(j)}},m),\ (0\le i(j)<j) \tag{5}$$

위 식(4)와 (5)에서 C_{n_j}는 $C_{n_{i(j)}}$과 m에 조건부로 정의되고, $(n_1,\cdots,n_K)$는 1부터 K까지의 정수에 대한 미지 순열을 의미한다. 정의에 의하여 $P(C_{n_j}\mid C_0,m)$는 $P(C_{n_j},m)$이라 정한다. 여기서 C_0은 정의된 근사 확률 분포를 전개했을 때, 체인 규칙과 같은 확률의 성질을 만족할 수 있도록 하고, 일반화된 구성 분포의 표준 수식으로 표현하기 위하여 도입하여 정의된 널(null) 변수항이다. 즉, 위와 같은 3차 구성 분포로 일반화시키기 위하여 맨 처음에 존재하는 2차 구성 분포 $P(C_{n_j},m)$을 3차원 구성 분포인 $P(C_{n_j}\mid C_0,m)$으로 표현한 것이다. Wang과 Wong이 제안한 1차 의존관계 트리 근사 방법을 적용하면, 선정된 최적의 의존관계 트리로부터 미지 순열 $(n_1,\cdots,n_K)$과 이에 대한 미지 조건부 순열 $(n_{i(1)},\cdots,n_{i(K)})$을 정하게 되어 최적의 곱 근사 분포를 결정하게 된다. 이 의존관계 트리에서의 가지 가중치로서는 아래 식(6)과 같은 $\varDelta$1차 평균 상호 정보가 정의되어 사용된다[10].

$$\varDelta D(C_{n_j};C_{n_{i(j)}})=D(C_{n_j};C_{n_{i(j)}},m)-D(C_{n_j};C_{n_{i(j)}}) \tag{6}$$

따라서, 최적의 의존관계 트리는 이러한 $\varDelta$1차 평균 상호 정보의 전체 합(즉, $\sum_{j=1}^{K}\varDelta D(C_{n_j};C_{n_{i(j)}})$)이 최대가 되는 최대 가중화 트리를 의미한다.

이번에는 결정 벡터 C의 K차 확률 분포에 대하여 2차 의존관계를 고려하면, 아래 식(7)과 같이 3차 구성 분포의 곱으로 근사 분포가 정의된다.

$$P_a(C_1,\cdots,C_K)=\prod_{j=1}^{K}P(C_{n_j}\mid C_{n_{i2(j)}},C_{n_{i1(j)}}),\ (0\le i2(j),i1(j)<j) \tag{7}$$

위 식(7)에서 C_{n_j}는 $C_{n_{i2(j)}}$와 $C_{n_{i1(j)}}$에 조건부로 정의되고, $(n_1,\cdots,n_K)$는 1부터 K까지의 정수의 미지 순열을 의미한다. 그리하여, 변수 C와 m으로 구성된 $(K+1)$차 고차 확률 분포는 아래 식(8)과 (9)에서와 같은 4차 구성 분포의 곱으로 정의된다.

$$P_a(m,C_1,\cdots,C_K)=\prod_{j=1}^{K}P(C_{n_j}\mid C_{n_{i2(j)}},C_{n_{i1(j)}},m),\ (0\le i2(j),i1(j)<j) \tag{8}$$

$$P_a(C_1,\cdots,C_K\mid m)=\frac{1}{P(m)}\prod_{j=1}^{K}P(C_{n_j}\mid C_{n_{i2(j)}},C_{n_{i1(j)}},m),\ (0\le i2(j),i1(j)<j) \tag{9}$$

위 식(8)과 (9)에서 C_{n_j}는 $C_{n_{i2(j)}}$과 $C_{n_{i1(j)}}$과 m에 모두 조건부로 정의되고, $(n_1,\cdots,n_K)$는 1부터 K까지의 정수에 대한 미지 순열을 의미한다. 정의에 의하여 $P(C_{n_j}\mid C_0,C_0,m)$는 $P(C_{n_j},m)$으로 정하고, $P(C_{n_j}\mid C_0,C_{n_{i1(j)}},m)$는 $P(C_{n_j}\mid C_{n_{i1(j)}},m)$으로 정한다. CD 상호 정보의 정의로부터 유도되는 수식과 평균 상호 정보를 간단한 수식으로 표현하기 위하여, 이제부터는 인식기의 결정을 표현하는 항 C_{n_j}에서 아래 첨자 n을 생략하여 C_j로 표현한다.

식(1)에서와 같이 정의된 베이스 에러율의 상위 경계를 최소화함으로써 2차 의존관계에 의한 최적의 곱 근사 분포를 얻기 위하여, 식(9)와 같은 $(K+1)$차 확률 분포와 식(7)과 같은 K차 확률 분포를 식(2)에서와 같이 정의된 CD 상호 정보의 정의에 각각 적용하면, 아래와 같은 식을 얻게 된다.

$$D(m;C)=\sum_{C}\sum_{M}P(C,m)\log\frac{P(C\mid m)}{P(C)}$$
$$=\sum_{M,C}P(C,m)\log\left[\frac{1}{P(m)}\prod_{j=1}^{K}P(C_j\mid C_{i2(j)},C_{i1(j)},m)\right]-\sum_{C}P(C)\log\prod_{j=1}^{K}P(C_j\mid C_{i2(j)},C_{i1(j)})$$
$$=-\sum_{M}P(m)\log P(m)+\sum_{j=1}^{K}\sum_{M,C}P(C,m)\log P(C_j\mid C_{i2(j)},C_{i1(j)},m)-\sum_{j=1}^{K}\sum_{C}P(C)\log P(C_j\mid C_{i2(j)},C_{i1(j)})$$
$$=H(m)+\sum_{j=1}^{K}\left[D(C_j;C_{i2(j)},C_{i1(j)},m)-D(C_j\mid C_{i2(j)},C_{i1(j)})\right] \tag{10}$$
$$H(m)=-\sum_{M}P(m)\log P(m)$$
$$\Delta D(C_j;C_{i2(j)},C_{i1(j)})=D(C_j;C_{i2(j)},C_{i1(j)},m)-D(C_j\mid C_{i2(j)},C_{i1(j)}) \tag{11}$$

위 식(10)으로부터 $D(m;C)$를 최대화하는 것은 $\Delta 2$차 평균 상호 정보의 전체 합인 $\sum_{j=1}^{K}\Delta D(C_j;C_{i2(j)},C_{i1(j)})$를 최대화하는 것과 같음을 알 수 있다. 왜냐하면, 나머지 항인 $H(m)$은 상수이기 때문이다.

이러한 이론적 배경을 바탕으로, 다음 단계는 가능한 모든 곱 근사 집합으로부터 최적의 곱 근사를 어떻게 결정하는가 이다. 2차 의존관계에 의한 최적의 곱 근사 분포를 결정하는 알고리즘은 다음과 같이 의사코드 형태로 기술된다. 이 알고리즘을 통하여 미지의 순열인 $(n_1,\cdots,n_K)$와 이의 조건부 미지 순열인 $(n_{i2(1)},\cdots,n_{i2(K)})$와 $(n_{i1(1)},\cdots,n_{i1(K)})$을 1부터 K까지의 정수에서 결정하게 된다.

입력 : w개의 샘플 데이타 $S^1,S^2,\cdots,S^w$

출력 : 2차 의존관계에 의한 최적의 곱 근사 분포

방법 :

1. 샘플로부터 2차, 3차, 4차 구성 분포를 계산, 평가한다.
2. 평가된 구성 분포로부터 $\Delta 1$차와 $\Delta 2$차 평균 상호 정보를 계산한다.
3. 계산된 $\Delta 1$차와 $\Delta 2$차 평균 상호 정보로부터 1차, 2차 의존관계에 의한 구성 분포의 평균 상호 정보의 최대 합을 구하고, 관련 곱 근사 분포를 결정한다.

```
maxTweight=0;     /* 최대 합을 지니는 변수 */
for n=1 to 1차 의존관계 구성 분포의 수 do
Twgt = 0;          /* 가능한 곱 근사 집합의 합을
                      지니는 변수 */
1차 의존관계 구성 분포 중의 하나를 선정;
Twgt = 선정된 1차 의존관계 구성 분포의 Δ1차
        평균 상호 정보;
while (탐색되지 않은 인식기의 수 > 0) do
  탐색되지 않은 인식기중의 하나를 선정;
  선정된 인식기를 지닌 2차 의존관계에 의한 최적
  구성 분포를 선정;
  Twgt += 선정된 2차 의존관계 구성 분포의 Δ2차
           평균 상호 정보;
  end
  maxTweight = MAX(maxTweight,Twgt);
  maxTweight와 관련된 1차, 2차 의존관계에 의한
  구성 분포를 저장;
end
최대인 maxTweight와 관련되어 저장된 1차, 2차
의존관계의 구성 분포 얻기;
```

4. 위에서 구한 1차, 2차 의존관계에 의한 구성 분포를 최적의 곱 근사 분포로 출력한다.

제안된 위 알고리즘은 인공지능의 휴리스틱과 같은 기법에 의해서 최적화되지 않았다. 따라서, 이 알고리즘의 계산 복잡도가 고차 의존관계를 고려함에 따라 커질 수는 있으나, 알고리즘이 최적 곱 근사 분포와 관련된 인식기들 간의 의존관계를 결정하기 위한 훈련 단계에서만 사용되고, 실제 다수 인식기의 결과를 결합하는 과정에서는 불필요하여 실행 시간에 지장을 주지 않기 때문에 문제가 없다고 여긴다. 또한, 2차 이상의 고차 의존관계를 고려한 근사 분포 식을 정의함에 따라서, 위 알고리즘도 [7]에서와 같이 일반화된 고차 의존관계에 의한 최적 곱 근사 분포를 결정하도록 쉽게 확장될 수 있다.

인식기 결정 변수와 클래스 변수를 포함한 고차 확률 분포에 대하여, [7]에서는 이들 변수들 간의 의존관계와 평균 상호 정보를 정의하고, 이에 대한 근사 분포를 결정하기 위하여 유사도 기준을 이용하여 최적의 곱 근사를 구하는 근사 분포 정의 식을 비롯한 이론적 근거와 관련 알고리즘을 제안하였다. 그러나, 본 논문에서는 인

식기 변수들만의 의존관계를 정의된 CD 상호 정보를 통하여 Δ평균 상호 정보로 정의하였으며, 베이스 에러율의 상위 경계 최소화 기준을 이용하여 최적 곱 근사를 구하는 다른 근사 분포 정의 식과 관련 알고리즘을 기술하였다. 이들 두 알고리즘은 모두 의존관계에 따른 최적의 곱 근사를 구하는 것이므로, 사용된 기준에 따라서 계산 방법은 다르지만, 전체적인 구성 형태는 매우 유사하다.

3. 곱 근사를 이용한 다수 인식기의 베이지안 결합 방법

이 장에서는 2.장에서 제안된 알고리즘을 통하여 구해진 최적의 곱 근사 분포를 이용하여 K개의 인식기를 베이지안 방법으로 결합하는 방법을 소개한다. 주어진 입력에 대하여 다수 인식기의 결정을 얻었을 때, 개개의 가설 결정 m_1에 대해서 아래 식(12)와 같은 신뢰도 함수 $Bel(m_i)$가 정의된다.

$$Bel(m_i) = P(m_i \in M \mid C_1, \cdots, C_K) \tag{12}$$

위 식(12)의 고차 사후 확률을 평가하기 위하여, 베이지안 정리와 $(K+1)$차 확률 분포 대신에 2.장에서 구한 1차 의존관계에 의한 최적의 곱 근사 분포인 식 (4)를 적용하면, 가설 결정 m_i의 사후 확률 P를 식 (13)과 같이 계산할 수 있다.

$$Bel(m_i) = P(m_i \in M \mid C_1, \cdots, C_K) = \frac{P(C_1, \cdots, C_K, m)}{P(C_1, \cdots, C_K)}$$
$$= \frac{\prod_{j=1}^{K} P(C_{n_j} \mid C_{n_{i(j)}}, m)}{P(C_1, \cdots, C_K)} \approx \eta \cdot \prod_{j=1}^{K} P(C_{n_j} \mid C_{n_{i(j)}}, m) \tag{13}$$

위 식(13)에서 η는 $\sum_{i=1}^{L} Bel(m_i) = 1$ 조건을 만족시키기 위한 상수이고, $(n_1, \cdots, n_K)$와 이의 조건부 미지 순열인 $(n_{i(1)}, \cdots, n_{i(K)})$은 1부터 K까지의 정수로 최적의 의존관계 트리로부터 결정된다. 다수 인식기를 베이지안 방법으로 결합하려면, 식(13)으로부터 계산된 개개의 $Bel(m_i)$ 값을 기준으로 가설 결정 m_i의 최대 사후 확률 P^*를 결정하고, 식 (14)와 같은 $R(C)$ 규칙에 따라서 결합된 결정을 확정하거나 기각해야 한다.

$$R(C) = \begin{cases} m_i, & \text{if } Bel(m_i) = \max_{m_j \in M} Bel(m_j) \\ L+1, & \text{otherwise} \end{cases} \tag{14}$$

다음은 고차 의존관계를 고려한 2차 의존관계에 의한 결합 방법을 살펴본다. 앞서 제안된 알고리즘을 통하여 구해진 2차 의존관계에 의한 최적의 곱 근사 분포인 식 (8)을 식(12)에서 $(K+1)$차 확률 분포 대신에 적용하면, 식(15)에서와 같이 가설 결정 m_i의 사후 확률 P를 계산할 수 있다.

$$Bel(m_i) = P(m_i \in M \mid C_1, \cdots, C_K) = \frac{P(C_1, \cdots, C_K, m)}{P(C_1, \cdots, C_K)}$$
$$= \frac{\prod_{j=1}^{K} P(C_{n_j} \mid C_{n_{i2(j)}}, C_{n_{i1(j)}}, m)}{P(C_1, \cdots, C_K)} \approx \eta \cdot \prod_{j=1}^{K} P(C_{n_j} \mid C_{n_{i2(j)}}, C_{n_{i1(j)}}, m) \tag{15}$$

위 식(15)에서도 η는 역시 $\sum_{i=1}^{L} Bel(m_i) = 1$ 조건을 만족시키기 위한 상수이고, $(n_1, \cdots, n_K)$와 이의 조건부 미지 순열인 $(n_{i2(1)}, \cdots, n_{i2(K)})$와 $(n_{i(1)}, \cdots, n_{i(K)})$은 1부터 K까지의 정수로, 제안된 알고리즘을 통하여 결정하게 된다. 2차 의존관계를 기반으로 다수 인식기를 베이지안 방법으로 결합하려면, 마찬가지 방법으로 개별 가설 결정에 대한 사후 확률을 계산하고, 식(14)와 같은 $R(C)$ 규칙을 적용하면 된다.

4. 실험 결과 및 분석

이 장에서는 CENPARMI 데이타베이스의 필기 숫자 인식에 사용되는 개별 인식기 *C1, C2, C3, C4, C5, C6*의 테스트 데이타 세트 T에 대한 인식 결과와 함께, 이들 인식기를 결합하여 수행한 실험 결과를 소개한다. CENPARMI 데이타베이스는 각 숫자 클래스 당 200개의 샘플 데이타로 구성된 3개의 데이타 세트로 총 6000자가 존재하는데, 인식기의 훈련용으로는 4000자로 구성된 훈련 데이타 세트인 A와 B를 사용하고, 인식기의 성능 테스트용으로는 2000자로 구성된 테스트 데이타 세트인 T를 사용하도록 권고한다. 이러한 맥락에서, 베이스 에러율의 상위 경계의 최소화를 기반으로 제안된 곱 근사 방법으로 n차 의존관계에 의한 최적의 곱 근사 분포를 결정하는데 있어서, 훈련 데이타 세트인 A와 B를 사용하였으며 인식기의 기각 결과는 무시하였다. 테스트 데이타 세트 T에 대한 개별 인식기의 인식 성능은 표 1에 보여진다.

표 1 테스트 데이타 세트 T에 대한 개별 인식기의 성능

인식기	인식률	기각률
C1	96.00	0.00
C3	94.15	0.00
C3	84.45	12.25
C4	78.75	0.00
C5	88.15	10.40
C6	90.95	8.15

실험에 사용되는 결합 방법으로는 투표 방법[1,5],

BKS 방법[5], 조건부 독립 가정에 의한 베이지안 방법[1,5], 유사도를 기준으로 곱 근사를 수행한 1차 및 2차 의존관계에 의한 베이지안 방법[6,7], 제안된 베이스 에러율의 상위 경계 최소화에 기반하여 곱 근사를 수행한 1차 및 2차 의존관계에 의한 베이지안 방법 등이 있으며, 이들의 인식 성능 비교는 테스트 데이타 세트 T에 대해서 수행되었다.

첫 번째 실험은 6개의 인식기로부터 5개의 인식기를 선택하여 이들을 각 결합 방법에 따라 결합하고, 그 인식 성능을 비교한 실험이다. 모두 6가지 경우의 인식기 집합이 생성되어 실험되었고, 그 결과는 표 2와 같다. 표 2로부터 대부분의 인식기 집합에서 가장 좋은 인식률을 보인 결합 방법은 제안된 베이스 에러율 기준을 이용하여 곱 근사를 수행한 1차 및 2차 의존관계에 의한 베이지안 방법이다.

표 2 5개의 인식기를 결합한 다수 인식기의 성능

결합 방법 \ 인식기 집합	C1,C2,C3,C4,C5		C1,C2,C3,C4,C6		C1,C2,C3,C5,C6	
	인식률	기각률	인식률	기각률	인식률	기각률
투표	97.20	1.35	97.50	1.05	97.40	1.60
BKS	91.15	8.30	91.75	8.00	92.90	6.75
조건부 독립 가정에 의한 베이지안	96.55	0.00	97.10	0.00	97.35	0.00
1차 의존관계에 의한 베이지안	96.55	0.00	97.10	0.00	97.00	0.00
조건부 1차 의존관계에 의한 베이지안	97.45	0.00	97.55	0.00	98.00	0.00
2차 의존관계에 의한 베이지안	97.70	0.00	97.55	0.00	97.80	0.00
제안된 1차 의존관계에 의한 베이지안	97.65	0.00	97.70	0.00	98.25	0.00
제안된 2차 의존관계에 의한 베이지안	97.75	0.00	97.90	0.00	97.90	0.00
결합 방법 \ 인식기 집합	C1,C2,C4,C5,C6		C1,C3,C4,C5,C6		C2,C3,C4,C5,C6	
	인식률	기각률	인식률	기각률	인식률	기각률
투표	97.35	1.50	96.55	2.25	96.55	2.40
BKS	92.20	7.40	92.00	7.35	90.90	8.35
조건부 독립 가정에 의한 베이지안	97.35	0.00	97.55	0.00	97.20	0.00
1차 의존관계에 의한 베이지안	97.10	0.00	97.20	0.00	96.90	0.00
조건부 1차 의존관계에 의한 베이지안	97.85	0.00	97.45	0.00	97.30	0.00
2차 의존관계에 의한 베이지안	97.70	0.00	97.40	0.00	97.70	0.00
제안된 1차 의존관계에 의한 베이지안	97.80	0.00	97.65	0.00	97.65	0.00
제안된 2차 의존관계에 의한 베이지안	98.05	0.00	97.60	0.00	97.90	0.00

두 번째 실험은 6개의 인식기를 모두 결합하고, 그 인식 성능을 비교한 실험으로, 그 결과는 표 3과 같다. 표 3에서 가장 좋은 인식률은 제안된 베이스 에러율 기준을 이용하여 곱 근사를 수행한 2차 의존관계에 의한 베이지안 방법이다. BKS 방법은 결합되는 인식기의 개수가 증가하면서 관찰되지 않은 결정 벡터의 수가 증가하게 되어 기각률이 높아지면서 인식 성능이 저하되는 현상을 보였다.

표 3 6개의 인식기를 결합한 다수 인식기의 성능

결합 방법 \ 인식기 집합	C1,C2,C3,C4,C5,C6	
	인식률	기각률
투표	97.60	1.10
BKS	89.55	10.25
조건부 독립 가정에 의한 베이지안	97.60	0.00
1차 의존관계에 의한 베이지안	97.10	0.00
조건부 1차 의존관계에 의한 베이지안	97.85	0.00
2차 의존관계에 의한 베이지안	97.95	0.00
제안된 1차 의존관계에 의한 베이지안	98.00	0.00
제안된 2차 의존관계에 의한 베이지안	98.05	0.00

이상과 같은 실험 결과를 통하여, 제안된 베이스 에러율 기준을 이용하여 곱 근사를 수행한 1차 및 2차 의존관계에 의한 베이지안 방법이 기존의 유사도를 기준으로 곱 근사를 수행한 1차 및 2차 의존관계에 의한 베이지안 방법이나 조건부 독립 가정에 의한 베이지안 방법, BKS 방법 등 보다 더욱 인식 성능이 우수함을 알 수 있었다. 대부분의 다수 인식기 집합에 있어서 2차 의존관계에 의한 베이지안 방법이 1차 의존관계에 의한 베이지안 방법 보다 높은 인식률을 보여 주어, 가능한 고차 정보를 사용하는 것이 유익함을 알 수 있었다. 결합되는 인식기의 개수가 증가하면서 인식률이 하락하는 BKS 방법은 많고 충분한 훈련 데이타 세트로부터 누적 빈도수 테이블이 생성되지 않은 이유라고 여기며, 그러한 데이타가 보장되는 경우에 사용하는 것이 바람직하다고 볼 수 있다. 실험 결과를 요약하면, 제안된 베이스 에러율 기준을 이용하여 다수 인식기를 결합하는 방법이 기존의 유사도 기준을 이용하여 결합하는 베이지안 방법 보다 Δn차 평균 상호 정보를 계산하기 위하여 더 많은 저장량과 계산량을 필요로 한다 하더라도, 다수 인식기의 결합에 의한 클래스 분별력의 향상에 더 기여함으로써 인식 성능이 향상되었음을 실험 결과로부터 얻었다고 볼 수 있다.

6. 결 론

본 논문에서는 클래스의 분별력을 높임으로써 인식 성능을 향상시키기 위하여 베이스 에러율의 상위 경계를 기반으로 확률 분포를 고차 의존관계에까지 근사하는 방법을 제안하였으며, 이를 이용하여 다수 인식기를 결합하기 위한 방법론을 제안하였다. 베이스 에러율의 상위 경계를 최소화함으로써 CD 상호 정보에 사용되는 고차 확률 분포에 대한 최적의 곱 근사를 얻을 수 있었고, 이를 기반으로 한 결합 방법이 다수 인식기의 결합 성능 실험에서도 가장 높은 인식 성능을 보였다. 이 방법론은 Wang과 Wong의 연구 결과를 2차 이상의 의존관계까지 고려하여 최적의 곱 근사를 구하는 방법으로 확장하고, 이를 다수 인식기의 결합에 적용한 것이다. 즉, 의존관계를 다루기 위한 CD 상호 정보를 정의하였으며, $\Delta 2$차 평균 상호 정보를 이용한 2차 의존관계에 의한 최적의 곱 근사 분포를 결정하는 새로운 알고리즘을 제안하였고, 이러한 곱 근사 분포를 기반으로 베이지안 방법을 사용하여 다수 인식기를 결합하였다. CENPARMI 필기 숫자 데이타베이스를 이용한 실험 결과를 통하여 베이스 에러율의 상위 경계 최소화에 기반하여 제안된 다수 인식기의 결합 방법이 다른 방법 보다 우수한 인식 성능을 보였음을 알 수 있었다.

참 고 문 헌

[1] Xu, L., Krzyzak, A., and Suen, C. Y., "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition," *IEEE Trans. on Systems, Man, and Cybernetics*, 22(3):418-435, 1992.

[2] Huang, Y. S. and Suen, C. Y., "A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(1):90-94, 1995.

[3] Mandler, E. and Schuermann, J., "Combining the classification results of independent classifiers based on the dempster/shafer theory of evidence," In E. S. Gelsema and L. N. Kanal, editors, Pattern Recognition and Artificial Intelligence, pp.381-393. 1988.

[4] Franke, J. and Mandler, E., "A comparison of two approaches for combining the votes of cooperating classifiers," In *Proceedings of the 11th IAPR Int. Conference on Pattern Recognition*, Vol.2, pp. 611-614, 1992.

[5] Lee, D.-S. and Srihari, S. N., "Handprinted Digit Recognition : A Comparison of Algorithms," In *Proceedings of the 3rd Int. Workshop on Frontiers in Handwriting Recognition*, pp.153-162, 1993.

[6] Kang, H.-J., Kim, K., and Kim, Jin H., "Optimal Approximation of Discrete Probability Distribution with *k*th-order Dependency and Its Applications to Combining Multiple Classifiers," *Pattern Recognition Letters*, 18(6), pp.515-523, 1997.

[7] 강희중, 이성환, "무제약 필기 숫자를 인식하기 위한 다수 인식기를 결합하는 의존관계 기반의 프레임워크", 정보과학회논문지 : 소프트웨어 및 응용, 제27권, 제8호, pp.855-863, 2000.

[8] Lewis, P. M., "Approximating Probability Distributions to Reduce Storage Requirement," *Information and Control*, 2:214-225, 1959.

[9] Chow, C. K. and Liu, C. N., "Approximating Discrete Probability Distributions with Dependence Trees," *IEEE Trans. on Information Theory*, 14(3): 462-467, 1968.

[10] Wang, D. C. C. and Wong, A. K. C., "Classification of Discrete Data with Feature Space Transform," *IEEE Trans. on Automatic Control*, AC-24(3):434-437, 1979.

[11] Hellman, M. E. and Raviv, J., "Probability of error, equivocation, and the Chernoff bound," *IEEE Trans. on Information Theory*, IT-16:368-372, 1970.

강 희 중

1986년 서울대학교 전자계산기공학과 졸업(학사). 1988년 한국과학기술원 전산학과 졸업(석사). 1997년 한국과학기술원 전산학과 졸업(박사). 1986년 ~ 1998년 삼성전자주식회사 기업통신개발그룹 선임연구원. 1998년 ~ 2000년 고려대학교 인공시각연구센터 연구조교수. 2000년 ~ 현재 한성대학교 정보전산학부 교수. 관심분야는 인공지능, 패턴인식, 그룹웨어