# 무제약 필기 숫자를 인식하기 위한 다수 인식기를 결합하는 의존관계 기반의 프레임워크
# (Dependency-based Framework of Combining Multiple Experts for Recognizing Unconstrained Handwritten Numerals)

강 희 중 † 이 성 환 ††
(Hee-Joong Kang)(Seong-Whan Lee)

**요 약** $K$개의 인식기로부터 관찰된 $K$개 결정을 결합하는 결합 방법론 중의 하나인 BKS (Behavior-Knowledge Space) 방법은 아무런 가정 없이 이들 결정을 결합하지만, 관찰된 $K$개 결정을 저장하고 관리하려면 이론적으로 기하학적인 저장 공간을 만들어야 한다. 즉, $K$개의 인식기 결정을 결합하기 위하여 ($K$+1)차 확률 분포를 필요로 하는데, 작은 $K$라 할지라도 그 확률 분포를 저장하거나 평가하는 것이 어렵다는 것은 이미 잘 알려져 있다. 그러한 문제점을 극복하기 위해서는 고차 확률 분포를 몇 개의 구성 분포로 나누고, 이들 구성 분포의 곱(product)으로 고차 확률 분포를 근사시켜야 한다. 그러한 이전 방법 중의 하나는 그 확률 분포에 조건부 독립 가정을 적용하는 것이고, 다른 방법으로는 [1]에서와 같이 그 확률 분포를 단지 트리 의존관계 또는 2차 구성 분포의 곱으로 근사하는 것이다. 본 논문에서는, 구성 분포의 곱으로 근사하는 방법에서, 2차 이상의 고차 구성 분포까지 고려하여 ($K$+1)차 확률 분포를 $d$차 ($1 \leq d \leq K$) 의존관계에 의한 최적의 곱으로 근사하고, 베이지안 방법과 그 곱을 기반으로 다수 인식기의 결정을 결합하는 의존관계 기반의 프레임워크를 제안한다. 이 프레임워크는 표준 CENPARMI 데이타베이스로 실험되어 평가되었다.

***Abstract*** Although Behavior-Knowledge Space (BKS) method, one of well known decision combination methods, does not need any assumptions in combining the multiple experts, it should theoretically build exponential storage spaces for storing and managing jointly observed $K$ decisions from $K$ experts. That is, combining $K$ experts needs a ($K$+1)st-order probability distribution. However, it is well known that the distribution becomes unmanageable in storing and estimating, even for a small $K$. In order to overcome such weakness, it has been studied to decompose a probability distribution into a number of component distributions and to approximate the distribution with a product of the component distributions. One of such previous works is to apply a conditional independence assumption to the distribution. Another work is to approximate the distribution with a product of only first-order tree dependencies or second-order distributions as shown in [1]. In this paper, higher order dependency than the first-order is considered in approximating the distribution and a dependency-based framework is proposed to optimally approximate the ($K$+1)st-order probability distribution with a product set of dth-order dependencies where ($1 \leq d \leq K$), and to combine multiple experts based on the product set using the Bayesian formalism. This framework was experimented and evaluated with a standardized CENPARMI data base.

· 본 연구는 과학기술부 창의적연구진흥사업의 연구비 지원을 받았음.
† 비 회 원 : 한성대학교 정보전산학부 교수
hjkang@hansung.ac.kr
†† 종신회원 : 고려대학교 컴퓨터학부 교수
swlee@image.korea.ac.kr
논문접수 : 2000년 2월 2일
심사완료 : 2000년 6월 9일

## 1. 서 론

문자 인식을 비롯한 패턴인식 분야에 있어서 최적 성능을 이루기 위하여 한 인식기의 성능을 향상시키려는 연구 방법에 대한 가능한 대안으로서, 다수 인식기의 결정을 결합하려는 연구가 지난 10여 년 동안 진행되어

왔다[2,3]. 다수 인식기를 결합하기 위해 제안된 방법은 인식기가 내리는 결정의 형태에 따라 다르게 개발되었는데, 그 결정의 형태는 측정 점수 리스트, 순위 리스트, 단일 선택 중의 하나에 속한다고 본다[4]. 다수 인식기를 결합하는 방법은 가급적 단일 선택 형태의 결정 수준에서 개발되는 것이 바람직하다. 그 이유는 개발되는 결합 방법이 임의 인식기에 독립이고, 또한 인식기를 그 결합 방법으로 쉽게 결합할 수 있기 때문이다. 이러한 단일 선택 수준에서 제안된 기존의 결합 방법으로는 투표 방법[2,3,4], BKS (Behavior-Knowledge Space) 방법[5], Dempster-Shafer 방법[6,7] 그리고 독립 가정에 의한 베이지안 방법[4,8] 등이 있다.

단일 선택 수준에서 다수 인식기를 결합하는 것은 확률 이론과 베이지안 공식을 이용하여 다음과 같은 수식으로 표현된다. 입력 $x$가 $K$개 인식기(예를 들면, $E_1, E_2, \cdots, E_K$)에 동시에 제공되면, $K$차원의 결정 벡터인 $V = < E_1(x) \cdots C_1, E_2(x) = C_2, \cdots, E_K(x) = C_K >$가 관찰된다. 여기서, 인식에 사용되는 $L$개의 클래스는 $C = \{C_1, C_2, \cdots, C_L\}$으로 표현된다. 이렇게 관찰된 $K$차원의 결정 벡터는 실제 클래스 정보와 함께 BKS 방법에서 고차 빈도수 테이블을 생성하여 다수 인식기를 결합하는데 직접 사용되고, 베이지안 기법에서는 고차 확률 분포를 생성하여 평가하므로써 다수 인식기를 결합하는데 사용된다. 베이지안 기법에서 다수 인식기를 결합하는 주요 작업은 $\max_m P(m \in C \mid E_1(x) = C_1, E_2(x) = C_2, \cdots, E_K(x) = C_K)$ 인 포스테리어 확률 $P^*$를 최대화하는 가설 결정 $m$을 결정하는 것이다. 즉, $(K+1)$차 확률 분포가 훈련 단계에서 샘플로부터 생성되고 평가되어야 한다. $(K+1)$차 확률은 $P(m, E_1(x) = C_1, E_2(x) = C_2, \cdots, E_K(x) = C_K)$ 와 같은 수식으로 표현된다.

다수 인식기를 결합하는데 있어서, 이러한 고차 확률 분포를 평가하기 위하여 기존의 연구는 주로 각 인식기의 결정이 주어진 가설 결정에 대하여 조건부 독립이라는 가정을 취하였다[4,8]. 즉, $(K+1)$차 확률은 다음과 같은 수식으로 평가된다고 여겼다.

$$P(m, E_1(x)=C_1, E_2(x)=C_2, \ldots, E_K(x)=C_K) = P(m)P(E_1(x)=C_1, E_2(x)=C_2, \ldots, E_K(x)=C_K \mid m)$$
$$\approx P(m)\prod_{i=1}^{K} P(E_i(x)=C_i \mid m)$$

그러나, Huang과 Suen은 BKS 방법에서 인식기의 독립 가정을 취하지 않고[5], 누적된 고차 빈도수 테이블 정보로부터 다음 수식과 같은 방식으로 직접 $(K+1)$차 확률을 평가하였다.

$$P(m, E_1(x)=C_1, E_2(x)=C_2, \cdots E_K(x)=C_K)$$
$$= P(m \mid E_1(x)=C_1, E_2(x)=C_2 \cdots E_K(x)=C_K) P(E_1(x)=C_1, E_2(x)=C_2 \cdots E_K(x)=C_K)$$

독립 가정에 의한 베이지안 방법이 BKS 방법 보다 계산량과 저장 공간량에 있어서 우위를 지니지만, 그러한 독립 가정이 실제 현실에서는 잘 유지되지 않는 경우가 있다. 그러나, BKS 방법은 독립 가정 없이 다수 인식기를 결합하지만, 기하학적인 계산량과 저장 공간량을 필요로 하고, 관찰되지 않은 다수 결정에 대해서는 결합을 기각하게 되어 잠재적으로 높은 기각률을 보일 수 있는 단점이 있다. 한편, 고차 확률 분포를 저차 확률 분포의 곱(product)으로 근사하는 아이디어를 기반으로 인식기 간의 의존관계를 고려하여 다수 인식기를 결합하려는 연구가 있다[9]. 인식률 향상을 위하여, 한 인식기는 다른 인식기에 통계적으로 의존적인 성향을 지니게 되므로, 인식기가 독립이라는 가정은 적절하지 않은 경우가 많기 때문이다.

본 논문에서는, 기존의 독립 가정을 기반으로 한 베이지안 방법과 그렇지 않은 BKS 방법을 모두 일반화할 수 있고, 보다 체계적으로 2차 이상의 고차 의존관계까지 고려하여 다수 인식기를 결합할 수 있는 의존관계 기반의 프레임워크를 제안한다. 이 프레임워크는 고차 확률 분포를 $d$차 ($1 \leq d \leq K$) 의존관계를 고려한 최대 $(d+1)$차 구성 분포의 곱으로 최적으로 근사하는 알고리즘과, 이러한 근사 분포와 베이지안 방법을 이용하여 독립 가정 없이 다수 인식기를 확률적으로 결합하는 결정 결합기로 구성된다. 이러한 프레임워크에 기반한 결합 방법론은 BKS 방법 보다 더 적은 공간 저장량(즉, $(K+1-d) \cdot L^{d+1}$)을 요구하고, 저차 구성 분포의 곱 근사에 의해서 높은 기각률도 보다 낮출 수 있다.

다수 인식기를 결합하는 의존관계 기반의 프레임워크는 CENPARMI (Center for Pattern Recognition and Machine Intelligence, Concordia University) 무제약 필기 숫자 데이타베이스를 사용한 실험에서 타 결합 방법과 비교되어 평가되었다. 이 필기 숫자 데이타베이스는 훈련 데이타 세트 A, B와 테스트 데이타 세트 T로 나뉘어 구성되며, 각 숫자 클래스에 200개의 필기 데이타를 지니고 있다. KAIST와 전북대학교에서 개발된 역전파 신경망 기반으로 구현된 3개의 인식기 *E1, E2, E3*와 규칙 기반으로 구현된 2개의 인식기 *E4, E5*로 총 5개의 인식기가 사용되었다.

본 논문의 구성은 다음과 같다. 2.장에서는 관련 연구를 기술하고, 의존관계 기반의 프레임워크와 관련된 이론적 배경 및 설명, $d$차 의존관계에 의한 최적의 곱 근

사를 구하는 알고리즘과 관련 예제 등은 3.장에서 기술하며, 4장은 베이지안 방법에 의한 다수 결정의 결합 방법을 소개하고, CENPARMI 데이타베이스로 제안된 프레임워크를 실험한 결과 및 분석은 5.장에서 기술되며, 6.장에서 결론을 맺는 순서로 되어 있다.

## 2. 관련 연구

BKS 방법은 훈련 단계에서 훈련 데이타 샘플로부터 $K$개 인식기의 관찰된 결정에 대해서 개별 클래스의 빈도수 테이블을 생성한 다음, 실행 단계에서 인식기의 독립 가정 없이 주어진 관찰된 $K$개 결정과 상응하는 빈도수 테이블 개체에서 최대 빈도수를 지닌 대표 클래스를 결정하여 결합된 결과로 결정하는 방법이다[5]. 그러나, 관찰되지 않은 결정에 대해서는 결합될 결정을 기각해야 하는 문제점을 지닌다. 게다가, $K$가 커지면 커질수록 이론적으로 더 많은 기하학적인 공간 저장량을 요구하게 된다. 즉, $K$개 인식기와 $L$개 클래스에 대한 공간 저장량 복잡도는 $L^{K+1}$ 이다.

Xu 등에 의해서 제안된 베이지안 방법은 인식기가 서로 독립적으로 수행한다고 가정한다[4]. 이러한 베이지안 방법은 BKS 방법 보다 더 적은 공간 저장량의 복잡도(즉, $K \cdot L^2$)를 요구한다. 이러한 조건부 독립 가정은 Lewis가 제안한 1차 의존관계에 의한 곱 근사 방법의 특수한 경우이다.

Lewis는 $n$차 이진 확률 분포를 몇 개의 구성 분포의 곱으로 근사하는 문제를 다루기 위하여 유사도(Measure of closeness)를 제안하여 연구하였으며, 적절히 제한된 조건하에서 곱 근사 분포가 최소 정보의 성질을 지닌다고 보였다[10]. Chow와 Liu는 Lewis의 연구 결과에 이어서, $n$차 분포를 1차 의존관계에 의한 $(n-1)$개 2차 구성 분포로 이루어진 최적의 곱으로 근사하는 문제를 연구하였다[1]. 그 결과, $n$개 변수 가운데 $(n-1)$개 1차 의존관계에 의한 최적의 곱 집합을 구하기 위하여 최소 정보 차이에 의해 근사 분포를 구하는 프로시듀어를 제안하였다.

강희중 등은 다수 인식기를 결합하는 문제에 의존관계 개념을 고려하여 독립 가정을 피하였다. 1차 의존관계에 의한 최적의 곱 근사를 구하는 기존의 방법에 덧붙여서, 유사도에 의한 최소 정보를 기준으로 2차 의존관계에 의한 최적의 곱 근사 분포를 구하여 다수 인식기를 결합하는 방법을 제안하였다[9]. 그러나, 보다 일반적인 2차 이상의 고차 의존관계에 의한 최적의 곱 근사 분포를 구하는 구체적인 방법론은 최근에 시도되었다[11].

Ho 등은 인식기의 결정이 순위 형태로 표현되는 것이 바람직하며, 이러한 경우에 있어서 로지스틱 regression 등을 이용한 다수 인식기를 결합하는 방법을 제안하였다[12]. 개별 인식기의 국부적인 인식 성능을 기반으로 다수 인식기의 결합을 수행하는 것이 보다 나은 결합 성능을 보일 수 있다고 Woods 등이 주장하였으며[13], 다양한 다수 인식기의 결합 방법을 sum 규칙의 형태로 일반화하려는 시도의 연구가 Kittler 등에 의해서 수행되었다[14].

## 3. 의존관계 기반의 프레임워크

$K$개의 다수 인식기를 확률적인 방법으로 결합하기 위하여, 첫 단계에서는 $d$차 의존관계에 의한 최적의 곱 근사 분포를 결정해야 한다. 고차 확률 분포를 근사하는 데 있어서, 정보 이론 모델에 의한 기준으로 유사도를 사용하여 최적의 곱 근사 분포를 결정하였다. 유사도는 실제 확률 분포에 내재된 정보와 근사 분포에 내재된 정보 사이의 차이로서 정의된다. 그러므로, 최적의 곱 근사 분포는 이러한 정보의 차이를 최소화하므로써 구해진다.

본 논문에서는 데이타 샘플로부터 얻은 $(K+1)$차 확률 분포를 $d$차 의존관계에 의한 최적의 곱 근사 분포로 근사하고, 이러한 최적의 곱 근사를 기반으로 다수 인식기를 결합하기 위한 의존관계 기반의 프레임워크를 제안한다. 이러한 프레임워크는 의존관계 트리에 의한 방법에서부터 BKS 방법까지의 확률적인 결합 방법을 포함한다. 의존관계의 차수인 $d$는 더 나은 근사 분포 또는 성능을 위하여 가능한 자원하에서 조정될 수 있다. 앞으로 확률 항에서의 수식 표현을 간단히 하기 위하여 $E_j(x)=C_j$를 $E_j$로 표현한다.

1차 의존관계(즉, $d$=1)를 고려하여 고차 확률 분포를 2차 구성 분포의 곱으로 근사하면, 아래 식(1)과 같이 정의된다.

$$P_a(E_1,\dots,E_{K+1})=\prod_{j=1}^{K+1}P(E_{n_j}\mid E_{n_{i(j)}}),\ (0\le i(j)<j). \qquad (1)$$

위 식(1)에서 $E_{n_j}$는 $E_{n_{i(j)}}$에 조건부로 정의되고, $(n_1,\dots,n_K,n_{K+1})$는 1부터 $(K+1)$까지의 정수의 미지 순열을 의미한다. 또한, $E_{K+1}$는 가설 결정 $m$을 의미한다. 정의에 의하여 $P(E_{n_j}\mid E_0)$는 $P(E_{n_j})$ 이다. 가설 결정 $m$과 $K$개 인식기의 결정 $E_1,\dots,E_K$로 구성된 $(K+1)$차 확률 분포에 1차 의존관계 트리 근사 방법을 적용하면, 선정된 최적 의존관계 트리로부터 미지 순열

$(n_1, \dots, n_K, n_{K+1})$과 이에 대한 미지 조건부 순열 $(n_{i(1)}, \dots, n_{i(K)}, n_{i(K+1)})$을 결정하여 최적의 곱 근사 분포를 결정하게 된다.

한편, 위 식(1)에서 $E_{n_{i(j)}}$가 모든 $E_{n_j}$에 대해서 동일하면, 즉, $E_{n_j}$가 주어진 가설 결정인 $E_{K+1}$에 대하여 조건부로 독립이라고 가정하면, 식(1)은 아래와 같이 간단하게 정의된다.

$$P_a(E_1, \cdots, E_{K+1}) = \prod_{j=1}^{K} P(E_j \mid E_{K+1}).$$

이러한 곱 근사 분포는 1차 의존관계에 의한 곱 근사의 특수한 경우로 조건부 독립 가정이라고 알려져 있다.

2차 의존관계(즉, $d$=2)를 고려하여 고차 확률 분포를 3차 구성 분포의 곱으로 근사하면, 아래 식(2)와 같이 정의된다.

$$P_a(E_1 \cdots, E_{K+1}) = \prod_{j=1}^{K+1} P(E_{n_j} \mid E_{n_{i2(j)}}, E_{n_{i1(j)}}),\ (0 \le i2(j), i1(j) < j). \quad (2)$$

위 식(2)에서 $E_{n_j}$는 $E_{n_{i2(j)}}$와 $E_{n_{i1(j)}}$ 모두에 조건부로 정의되고, $(n_1, \cdots, n_K, n_{K+1})$는 1부터 ($K$+1)까지의 정수의 미지 순열을 의미한다. 정의에 의하여 $P(E_{n_j} \mid E_0, E_{n_{i(j)}})$는 $P(E_{n_j}, E_{n_{i(j)}})$ 이며, $E_{n_{i(j)}}$은 $E_{n_{i2(j)}}$ 또는 $E_{n_{i1(j)}}$을 의미한다. 유사도로부터 유도되는 수식과 평균 상호 정보를 표현하는 수식을 간단히 표현하기 위하여, 이제부터는 인식기를 표현하는 항 $E_{n_j}$에서 아래 첨자 $n$을 생략하여 $E_j$로 표현한다. 또한 변수 $E$는 ($K$+1)개 변수 $E_1, \dots, E_K, E_{K+1}$의 벡터를 의미한다. 2차 의존관계에 의한 최적의 곱 근사 분포를 얻기 위하여 식(2)를 유사도 정의에 적용하면, 아래와 같은 식을 얻게 된다.

$$\begin{aligned} I(P(E), P_a(E)) &= \sum_E P(E) \log \frac{P(E)}{P_a(E)} \\ &= \sum_E P(E) \log P(E) - \sum_{j=1}^{K+1} \sum_E P(E) \log P(E_j \mid E_{i2(j)}, E_{i1(j)}) \\ &= -\sum_{j=1}^{K+1} M(E_j; E_{i2(j)}, E_{i1(j)}) + \sum_{j=1}^{K+1} H(E_j) - H(E) \qquad (3) \end{aligned}$$

$$H(E) = -\sum_E P(E) \log P(E)$$

$$M(E_j; E_{i2(j)}, E_{i1(j)}) = \sum_E P(E_j, E_{i2(j)}, E_{i1(j)}) \log \frac{P(E_j \mid E_{i2(j)}, E_{i1(j)})}{P(E_j)}$$

위 식(3)으로부터 $I(P(E), P_a(E))$를 최소화하는 것은 $\sum_{j=1}^{K+1} M(E_j; E_{i2(j)}, E_{i1(j)})$ 를 최대화하는 것과 같음을 알 수 있다. 여기서, $\sum_{j=1}^{K+1} M(E_j; E_{i2(j)}, E_{i1(j)})$는 2차 평균 상호 정보의 전체 합을 의미하고[9], 나머지 항인 $\sum_{j=1}^{K+1} H(E_j)$과 $H(E)$은 상수이기 때문이다. $\sum_{j=1}^{K+1} H(E_j)$와 $H(E)$은 각각 주어진 변수의 엔트로피를 의미한다.

한편, 식(2)에서 $E_{n_{i1(j)}}$가 모든 $E_{n_j}$에 대해서 동일하면, 즉, $E_{n_j}$가 주어진 가설 결정 $E_{K+1}$과 $E_{n_{i2(j)}}$에 대하여 조건부로 의존한다고 가정하면, 식(2)는 아래와 같이 비교적 간단하게 정의된다.

$$P_a(E_1 \cdots, E_{K+1}) = \prod_{j=1}^{K} P(E_{n_j} \mid E_{n_{i2(j)}}, E_{K+1}),\ (0 \le i2(j) < j). \quad (4)$$

위 식(4)에서 $E_{n_j}$는 $E_{n_{i2(j)}}$와 $E_{K+1}$ 모두에 조건부로 정의되고, $(n_1, \cdots, n_K)$는 1부터 $K$까지의 정수의 미지 순열을 의미한다. 정의에 의하여 $P(E_{n_j} \mid E_0, E_{K+1})$는 $P(E_{n_j}, E_{K+1})$ 이다. 이러한 근사 분포는 2차 의존관계에 의한 곱 근사의 특수한 경우로 조건부 1차 의존관계에 의한 근사 방법이라고 정의한다.

다음 단계는 가능한 모든 곱 근사 집합으로부터 최적의 곱 근사를 어떻게 결정하는가 이다. 2차 의존관계에 의한 최적의 곱 근사 분포를 결정하는 알고리즘은 참고문헌 [9]에 기술되어 있다. 본 논문에서는 보다 일반화된 $d$차 의존관계에 의한 최적의 곱 근사 분포를 결정하기 위한 이론적 배경과 알고리즘을 3.2절에서 기술하고자 한다.

### 3.1 최적의 곱 근사 분포에 대한 예

1차, 2차 의존관계에 의한 최적의 곱 근사 분포를 구하는 예제를 이 절에서 기술하고자 한다. 서론에서 소개한 3개의 인식기 *E1, E2, E5*로부터 얻어진 결정 변수를 각각 $E_1, E_2, E_3$이라 하고, 가설 결정 변수를 $E_4$이라 가정할 때, 이들로 구성된 4차 확률 분포로부터 2차 및 3차 구성 분포에 의해서 계산된 평균 상호 정보를 아래와 같이 얻었다고 가정한다.

$$\begin{array}{lll} M(E_4;E_1) = 2.179802 & M(E_4;E_2) = 2.142977 & M(E_4;E_3) = 2.102546 \\ M(E_1;E_2) = 2.070755 & M(E_1;E_3) = 2.046490 & M(E_2;E_3) = 1.990362 \\ M(E_4;E_1,E_2) = 2.271831 & M(E_1;E_4,E_2) = 2.199609 & M(E_2;E_4,E_1) = 2.162786 \\ M(E_4;E_1,E_3) = 2.258727 & M(E_1;E_4,E_3) = 2.202671 & M(E_3;E_4,E_1) = 2.125416 \\ M(E_4;E_2,E_3) = 2.273394 & M(E_2;E_4,E_3) = 2.161210 & M(E_3;E_4,E_2) = 2.120799 \\ M(E_1;E_2,E_3) = 2.199589 & M(E_2;E_1,E_3) = 2.143459 & M(E_3;E_1,E_2) = 2.119195 \end{array}$$

4차 확률 분포에 조건부 독립 가정을 적용하면 아래에 있는 첫번째 줄의 근사 분포를 얻게 되며, 의존관계 트리 방법을 적용하면 두번째 줄의 근사 분포를 얻게 된다. 이 경우에는 두 근사 분포가 동일하게 결정되었다.

$P_a(E_1,E_2,E_3,E_4)=P(E_1,E_4)P(E_2|E_4)P(E_3|E_4):\sum M$ =6.425325

$P_a(E_1,E_2,E_3,E_4)=P(E_1,E_4)P(E_2|E_4)P(E_3|E_4):\sum M$ =6.425325*

이번에는 가용한 많은 정보를 고려하기 위하여 동일한 4차 확률 분포에 2차 의존관계를 고려하여 가능한 곱 근사 분포를 구하면 아래와 같이 표현된다.

$P_a(E_1,E_2,E_3,E_4)=P(E_4,E_1)P(E_2|E_4,E_1)P(E_3|E_4,E_2):\sum M$ =6.463367*

$P_a(E_1,E_2,E_3,E_4)=P(E_4,E_1)P(E_3|E_4,E_1)P(E_2|E_1,E_3):\sum M$ =6.448677

$P_a(E_1,E_2,E_3,E_4)=P(E_4,E_2)P(E_1|E_4,E_2)P(E_3|E_4,E_2):\sum M$ =6.463365

$P_a(E_1,E_2,E_3,E_4)=P(E_4,E_2)P(E_3|E_4,E_2)P(E_1|E_4,E_2):\sum M$ =6.463365

$P_a(E_1,E_2,E_3,E_4)=P(E_4,E_3)P(E_1|E_4,E_3)P(E_2|E_1,E_3):\sum M$ =6.448676

$P_a(E_1,E_2,E_3,E_4)=P(E_4,E_3)P(E_2|E_4,E_3)P(E_1|E_4,E_2):\sum M$ =6.463365

$P_a(E_1,E_2,E_3,E_4)=P(E_1,E_2)P(E_3|E_1,E_2)P(E_4|E_1,E_3):\sum M$ =6.463365

$P_a(E_1,E_2,E_3,E_4)=P(E_1,E_2)P(E_3|E_4,E_1)P(E_2|E_1,E_3):\sum M$ =6.448677

$P_a(E_1,E_2,E_3,E_4)=P(E_1,E_3)P(E_4|E_1,E_3)P(E_2|E_1,E_3):\sum M$ =6.448676

$P_a(E_1,E_2,E_3,E_4)=P(E_1,E_3)P(E_2|E_1,E_3)P(E_4|E_1,E_3):\sum M$ =6.448676

$P_a(E_1,E_2,E_3,E_4)=P(E_2,E_3)P(E_4|E_2,E_3)P(E_1|E_4,E_2):\sum M$ =6.463365

$P_a(E_1,E_2,E_3,E_4)=P(E_2,E_3)P(E_1|E_2,E_3)P(E_4|E_1,E_3):\sum M$ =6.448678

위 결과로부터 2차 의존관계에 의한 최적의 곱 근사 분포를 $P(E_4,E_1)P(E_2|E_4,E_1)P(E_3|E_4,E_2)$으로 결정할 수 있다. 그 이유는 $P(E_4,E_1)P(E_2|E_4,E_1)P(E_3|E_4,E_2)$의 2차 평균 상호 정보의 합이 최대이기 때문이다. 또한, 조건부 1차 의존관계에 의한 최적의 곱 근사 분포는 $P(E_4,E_1)P(E_2|E_4,E_1)P(E_3|E_4,E_2)$으로 2차 의존관계에 의한 최적의 곱 근사 분포와 동일하다.

### 3.2 고차 의존관계에 의한 근사 분포

앞서 정의한 의존관계 지향의 곱 근사 분포를 구하는 방법을 보다 체계화하므로써, 고려할 의존관계의 차수는 가용한 자원하에서 쉽게 $d$차까지 확장될 수 있다. $d$차 의존관계에 의한 최적의 곱 근사 분포는 아래와 같이 정의된 식(5)를 기반으로 확률 이론에 맞추기 위하여 1개의 1차 의존관계에 의한 구성 분포, 1개의 2차 의존관계에 의한 구성 분포, …, 1개의 $(d-1)$차 의존관계에 의한 구성 분포, 그리고 $(K-d)$개의 $d$차 의존관계에 의한 구성 분포로 이루어지게 된다.

$d$차 의존관계를 고려하여 고차 확률 분포를 $(d+1)$차 구성 분포의 곱으로 근사하면, 아래 식(5)와 같이 정의된다.

$$P_a(E_1\cdots,E_{K+1})=\prod_{j=1}^{K+1}P(E_{n_j}|E_{n_{id(j)}}\cdots,E_{n_{i1(j)}}),\ (0\le id(j,\cdots,i1(j)<j). \quad (5)$$

위 식(5)에서 $E_{n_j}$는 $E_{n_{id(j)}}$부터 $E_{n_{i1(j)}}$까지의 $d$개의 모든 변수에 조건부로 정의되고, $(n_1,\cdots,n_K,n_{K+1})$는 1부터 $(K+1)$까지의 정수의 미지 순열을 의미한다. 정의에 의하여 $P(E_{n_j}|E_0,E_{n_{i(j)}})$는 $P(E_{n_j},E_{n_{i(j)}})$ 이다. $d$차 의존관계에 의한 최적의 곱 근사 분포를 얻기 위하여 식(5)를 유사도 정의에 적용하면, 아래와 같은 식을 얻게 된다.

$$\begin{aligned}
I(P(E),P_a(E)) &= \sum_E P(E)\log\frac{P(E)}{P_a(E)} \\
&= \sum_E P(E)\log P(E)-\sum_{j=1}^{K+1}\sum_E P(E)\log P(E_j|E_{id(j)},\cdots E_{i1(j)}) \\
&= -\sum_{j=1}^{K+1}M(E_j;E_{id(j)},\cdots E_{i1(j)})+\sum_{j=1}^{K+1}H(E_j)-H(E) \qquad (6)
\end{aligned}$$

$$H(E)=-\sum_E P(E)\log P(E)$$

$$M(E_j;E_{id(j)},\cdots,E_{i1(j)})=\sum_E P(E_j,E_{id(j)},\cdots,E_{i1(j)})\log\frac{P(E_j|E_{id(j)},\cdots E_{i1(j)})}{P(E_j)}$$

위 식(6)으로부터 $I(P(E),P_a(E))$를 최소화하는 것은 $\sum_{j=1}^{K+1}M(E_j;E_{id(j)},\cdots,E_{i1(j)})$를 최대화하는 것과 같음을 알 수 있다. 여기서, $\sum_{j=1}^{K+1}M(E_j;E_{id(j)},\cdots,E_{i1(j)})$는 $d$차 평균 상호 정보의 전체 합을 의미하고, 나머지 항인 $\sum_{j=1}^{K+1}H(E_j)$과 $H(E)$은 상수이기 때문이다. $\sum_{j=1}^{K+1}H(E_j)$와 $H(E)$은 각각 주어진 변수의 엔트로피를 의미한다. 이러한 이론적 배경을 바탕으로 2차 이상의 $d$차 의존관계에 의한 최적의 곱 근사 분포를 구하는 알고리즘을 기술하면 다음과 같다.

입력 : $w$개의 샘플 데이타 $S^1,S^2,\cdots,S^w$

출력 : $d$차 의존관계에 의한 최적의 곱 근사 분포

방법 :

1. 샘플로부터 2차, 3차, …, $(d+1)$차 구성 분포를 계산, 평가한다.
2. 평가된 구성 분포로부터 1차, 2차, …, $d$차 평균 상호 정보를 계산한다.
3. 계산된 평균 상호 정보로부터 1차, 2차, …, $d$차 의존관계에 의한 구성 분포의 평균 상호 정보의 최대 합을 구하고, 관련 곱 근사 분포를 결정한다.

maxTweight=0;
**for** n1=1 **to** 1차 의존관계 분포의 수 **do**
T1 = 0;
1차 의존관계 분포 중의 하나를 선정;
T1 = 선정된 1차 의존관계 분포의 평균 상호 정보;
**for** n2=1 **to** 2차 의존관계 분포의 수 **do**
T2 = T1;
2차 의존관계 분포 중의 하나를 선정;
T2 += 선정된 2차 의존관계 분포의 평균 상호 정보;
…
**while** (탐색되지 않은 인식기의 수 > 0) **do**

탐색되지 않은 인식기중의 하나를 선정;
선정된 인식기를 지닌 $d$차 의존관계에 의한 최적 구성 분포를 선정;
$Td$ += 선정된 $d$차 의존관계 분포의 평균 상호 정보;
**end**
...
T2 += T3;
**end**
T1 += T2;
maxTweight = *MAX*(maxTweight,T1);
maxTweight와 관련된 1차, 2차, …, $d$차 의존관계에 의한 분포를 저장;
**end**
최대인 maxTweight와 관련되어 저장된 1차, 2차, …, $d$차 의존관계의 분포 얻기;
4. 위에서 구한 1차, 2차, …, $d$차 의존관계에 의한 분포를 최적의 곱 근사 분포로 출력한다.

위 알고리즘에서 **while** 루프의 위에 있는 … 부분은 2차 의존관계와 관련된 **for** 루프의 기술된 부분이 3차 의존관계에서부터 ($d$-1)차 의존관계까지 반복적으로 삽입됨을 의미한 것이며, **while** 루프의 아래에 있는 … 부분은 역으로 ($d$-1)차 의존관계로부터 3차 의존관계까지 계산된 평균 상호 정보의 합이 지속적으로 더해져 왔음을 생략한 것이다.

## 4. 베이지안 방법을 이용한 다수 인식기의 결합

3.장에서 제안된 알고리즘을 기반으로 $d$차 의존관계에 의한 최적의 곱 근사 분포를 구한 다음, 이 장에서 정의된 베이지안 방법을 이용한 결합 방식으로 다수 인식기의 결정이 독립 가정 없이 결합된다. 1차 의존관계에 의한 결합 방법을 살펴 보면, ($K$+1)차 확률 분포에 있는 개개의 가설 결정 $C_i$에 대해서, 베이지안 정리와 ($K$+1)차 확률 분포 대신에 1차 의존관계에 의한 최적의 곱 근사 분포를 사용하면 식 (7)과 같이 가설 결정 $C_i$의 포스테리어 확률 $P$를 계산할 수 있다.

$$Bel(C_i) = P(C_i \in C \mid E_1, \cdots, E_K) = \frac{P(C_i \in C, E_1, \cdots, E_K)}{P(E_1, \Lambda, E_K)}$$
$$= \frac{\prod_{j=1}^{K+1} P(E_{n_j} \mid E_{n_{i(j)}})}{P(E_1, \ldots, E_K)} \approx \eta \cdot \prod_{j=1}^{K+1} P(E_{n_j} \mid E_{n_{i(j)}}) \quad (7)$$

위 식(6)에서 $\eta$는 $\sum_{i=1}^{L} Bel(C_i) = 1$ 조건을 만족시키기 위한 상수이고, $(n_1, \cdots, n_K, n_{K+1})$는 1부터 ($K$+1)까지의 정수의 미지 순열로 최적의 의존관계 트리로부터 결정된다. 다수 인식기를 베이지안 방법으로 결합하기 위하여 식(7)로부터 계산된 개개의 $Bel(C_i)$ 값을 기준으로 식 (8)과 같은 $R(V)$ 규칙에 따라서 결합된 결정을 확정하거나 기각하게 된다.

$$R(V) = \begin{cases} C_i, & \text{if } Bel(C_i) = \max_{C_j \in C} Bel(C_j) \\ L+1, & \text{otherwise} \end{cases} \quad (8)$$

다음은 고차 의존관계를 고려한 2차 의존관계에 의한 결합 방법을 살펴 본다. 2차 의존관계에 의해 구해진 최적의 곱 근사 분포를 식(7)에서 ($K$+1)차 확률 분포 대신에 사용하면, 식(9)과 같은 가설 결정 $C_i$의 포스테리어 확률 $P$를 계산할 수 있다.

$$Bel(C_i) = P(C_i \in C \mid 1 \ldots \Lambda, E_K) = \frac{P(C_i \in C, E_1, \Lambda, E_K)}{P(\cdots \Lambda, E_K)}$$
$$= \frac{\prod_{j=1}^{K+1} P(E_{n_j} \mid E_{n_{i2(j)}}, E_{n_{i1(j)}})}{P(1 \cdots \Lambda, E_K)} \approx \eta \cdot \prod_{j=1}^{K+1} P(E_{n_j} \mid E_{n_{i2(j)}}, E_{n_{i1(j)}}) \quad (9)$$

위 식(9)에서 $\eta$는 또한 $\sum_{i=1}^{L} Bel(C_i) = 1$ 조건을 만족시키기 위한 상수이고, $(n_1, \ldots, n_K, n_{K+1})$는 1부터 ($K$+1)까지의 정수의 미지 순열로, [9] 또는 3.2절에서 제안된 알고리즘을 통하여 결정된다. 2차 의존관계를 기반으로 다수 인식기를 베이지안 방법을 이용하여 결합하려면, 식(8)과 같은 $R(V)$ 규칙을 적용하면 된다.

2차 이상의 고차 의존관계를 고려한 $d$차 의존관계에 의한 최적의 곱 근사 분포를 구한 다음에, 그 근사 분포를 식(9)에서와 같이 적용하여 사용하면, 식(10)과 같은 가설 결정 $C_i$의 포스테리어 확률 $P$를 계산할 수 있다.

$$Bel(C_i) = P(C_i \in C \mid \ldots \Lambda, E_K) = \frac{P(C_i \in C, E_1, \Lambda, E_K)}{P(E_1, \Lambda, E_K)}$$
$$= \frac{\prod_{j=1}^{K+1} P(E_{n_j} \mid E_{n_{\ldots}} \ldots \Lambda, E_{n_{i1(j)}})}{P(\ldots \Lambda, E_K)} \approx \eta \cdot \prod_{j=1}^{K+1} P(E_{n_j} \mid E_{n_{id(j)}}, \Lambda, E_{n_{i1(j)}}) \quad (10)$$

위 식(10)에서 $\eta$는 또한 $\sum_{i=1}^{L} Bel(C_i) = 1$ 조건을 만족시키기 위한 상수이고, $(n_1, \cdots, n_K, n_{K+1})$는 1부터 ($K$+1)까지의 정수의 미지 순열로, 3.2절에서 제안된 알고리즘을 통하여 결정된다. 이 경우에도 마찬가지로 $d$차 의존관계를 기반으로 다수 인식기를 베이지안 방법으로 결합하려면, 동일하게 식(8)과 같은 $R(V)$ 규칙을 적용하면 된다.

## 5. 실험 결과 및 분석

이 장에서는 필기 숫자 인식에 사용되는 개별 인식기 *E1, E2, E3, E4, E5*의 테스트 데이타 세트 T에 대한 인식 결과와 함께, 이들 인식기를 결합하여 수행한 실험 결과를 소개한다. $d$차 의존관계에 의한 최적의 곱 근사 분포를 결정하거나 다수 인식기를 결합하는데 있어서, 훈련 데이타 세트 A, B를 사용하였으며, 인식기의 기각 결과를 클래스에 포함시켰다. 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 |
|---|---|---|
| *E1* | 96.00 | 0.00 |
| *E3* | 94.15 | 0.00 |
| *E3* | 84.45 | 12.25 |
| *E4* | 88.15 | 10.40 |
| *E5* | 90.95 | 8.15 |

실험에 사용되는 결합 방법은 BKS 방법[5], 조건부 독립 가정에 의한 베이지안 방법[4,8], 1차 의존관계에 의한 베이지안 방법[1,9], 2차 의존관계를 포함하는 제안된 프레임워크에서의 $d$차 의존관계에 의한 베이지안 방법 등이 있으며, 이들의 성능 비교는 테스트 데이타 세트 T에 대해서 수행되었다. 다만, 가용한 테스트 시스템의 환경 조건에 의해 2차 의존관계까지만을 고려한 결합 방법을 사용하여 실험하였다.

첫번째 실험은 5개의 인식기로부터 3개의 인식기를 선택하여 이들을 각 결합 방법에 따라 결합하고, 그 인식 성능을 비교하는 실험이다. 모두 10가지의 인식기 집합이 생성되어 실험되었고, 그 결과는 표 2와 같다. 표 2에서 가장 좋은 인식률은 보인 방법은 대부분의 인식기 집합에서 2차 의존관계에 의한 베이지안 방법이다.

표 2 3개의 인식기를 결합한 다수 인식기의 성능

| 인식기 집합 / 결합 방법 | *E1,E2,E3* | | *E1,E2,E4* | | *E1,E2,E5* | | *E1,E3,E4* | | *E1,E3,E5* | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 인식률 | 기각률 | 인식률 | 기각률 | 인식률 | 기각률 | 인식률 | 기각률 | 인식률 | 기각률 |
| BKS | 96.00 | 2.95 | 95.25 | 2.70 | 96.00 | 2.65 | 95.60 | 2.30 | 95.90 | 2.60 |
| 조건부 독립 가정에 의한 베이지안 | 95.70 | 0.00 | 95.90 | 0.00 | 96.55 | 0.00 | 96.65 | 0.00 | 97.00 | 0.00 |
| 1차 의존관계에 의한 베이지안 | 95.70 | 0.00 | 95.90 | 0.00 | 96.55 | 0.00 | 96.65 | 0.00 | 97.00 | 0.00 |
| 조건부 1차 의존관계에 의한 베이지안 | 96.65 | 0.00 | 96.75 | 0.00 | 97.50 | 0.00 | 97.15 | 0.00 | 96.85 | 0.00 |
| 2차 의존관계에 의한 베이지안 | 96.65 | 0.00 | 96.75 | 0.00 | 97.50 | 0.00 | 97.15 | 0.00 | 96.85 | 0.00 |

| 인식기 집합 / 결합 방법 | *E1,E4,E5* | | *E2,E3,E4* | | *E2,E3,E5* | | *E2,E4,E5* | | *E3,E4,E5* | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 인식률 | 기각률 | 인식률 | 기각률 | 인식률 | 기각률 | 인식률 | 기각률 | 인식률 | 기각률 |
| BKS | 96.75 | 1.60 | 94.85 | 2.70 | 95.35 | 3.40 | 96.10 | 1.70 | 95.25 | 3.15 |
| 조건부 독립 가정에 의한 베이지안 | 97.50 | 0.00 | 95.85 | 0.00 | 96.80 | 0.00 | 97.15 | 0.00 | 96.25 | 1.30 |
| 1차 의존관계에 의한 베이지안 | 97.10 | 0.00 | 95.85 | 0.00 | 96.80 | 0.00 | 96.80 | 0.00 | 96.00 | 1.30 |
| 조건부 1차 의존관계에 의한 베이지안 | 97.50 | 0.00 | 96.30 | 0.00 | 97.00 | 0.00 | 96.50 | 0.00 | 96.65 | 1.30 |
| 2차 의존관계에 의한 베이지안 | 97.50 | 0.00 | 96.45 | 0.00 | 97.15 | 0.00 | 97.25 | 0.00 | 96.65 | 1.30 |

두번째 실험은 5개의 인식기로부터 4개의 인식기를 선택하여 이들을 각 결합 방법에 따라 결합하고, 그 인식 성능을 비교하는 실험이다. 모두 5가지의 인식기 집합이 생성되어 실험되었고, 그 결과는 표 3과 같다. 표 3에서 가장 좋은 인식률은 보인 방법은 대부분의 인식기 집합에서 2차 의존관계에 의한 베이지안 방법이다(조건부 1차 의존관계에 의한 베이지안 방법도 2차 의존관계의 특수한 경우로 포함하기 때문이다). 결합되는 인식기의 갯수가 3개에서 4개로 증가하면서 BKS 방법의 기각률도 더불어 증가하면서 인식률이 하락하는 양상을 보였다. 이것은 BKS 방법에서 사용하는 누적 빈도수 테이블이 제대로 테스트 데이타를 반영하지 못한다는 증거이기도 하다.

표 3 4개의 인식기를 결합한 다수 인식기의 성능

| 인식기 집합 / 결합 방법 | *E1,E2,E3,E4* | | *E1,E2,E3,E5* | | *E1,E2,E4,E5* | | *E1,E3,E4,E5* | | *E2,E3,E4,E5* | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 인식률 | 기각률 | 인식률 | 기각률 | 인식률 | 기각률 | 인식률 | 기각률 | 인식률 | 기각률 |
| BKS | 93.60 | 5.35 | 94.25 | 5.20 | 94.75 | 4.45 | 94.85 | 3.85 | 94.10 | 4.90 |
| 조건부 독립 가정에 의한 베이지안 | 96.15 | 0.00 | 97.00 | 0.00 | 97.00 | 0.00 | 97.45 | 0.00 | 97.25 | 0.00 |
| 1차 의존관계에 의한 베이지안 | 96.15 | 0.00 | 97.00 | 0.00 | 96.55 | 0.00 | 97.00 | 0.00 | 96.80 | 0.00 |
| 조건부 1차 의존관계에 의한 베이지안 | 97.55 | 0.00 | 97.80 | 0.00 | 97.70 | 0.00 | 97.80 | 0.00 | 97.40 | 0.00 |
| 2차 의존관계에 의한 베이지안 | 97.55 | 0.00 | 97.80 | 0.00 | 97.50 | 0.00 | 97.80 | 0.00 | 96.95 | 0.00 |

세번째 실험은 5개의 인식기를 모두 선택하여 결합하고, 그 인식 성능을 비교하는 실험이다. 그 결과는 표 4와 같다. 표 4에서 가장 좋은 인식률은 2차 의존관계에 의해서 최적의 곱 근사가 결정되는 특수한 경우인 조건부 1차 의존관계에 의한 베이지안 방법이고, 그 다음도

역시 2차 의존관계에 의한 베이지안 방법이다. BKS 방법은 결합되는 인식기의 갯수가 증가하면서 관찰되지 않은 결정 벡터의 수가 증가하게 되어 기각률이 높아지면서 인식 성능이 저하되는 현상을 보였다.

표 4 5개의 인식기를 결합한 다수 인식기의 성능

| 인식기 집합 / 결합 방법 | *E1,E2,E3,E4,E5* | |
|---|---|---|
| | 인식률 | 기각률 |
| BKS | 92.90 | 6.75 |
| 조건부 독립 가정에 의한 베이지안 | 97.35 | 0.00 |
| 1차 의존관계에 의한 베이지안 | 97.00 | 0.00 |
| 조건부 1차 의존관계에 의한 베이지안 | 98.00 | 0.00 |
| 2차 의존관계에 의한 베이지안 | 97.80 | 0.00 |

이상과 같은 실험 결과를 통하여 2차 의존관계에 기반한 베이지안 방법이 1차 의존관계나 조건부 독립 가정에 의한 베이지안 방법, BKS 방법 보다 더욱 우수한 성능을 보임을 알 수 있었다. 조건부 1차 의존관계를 포함한 2차 의존관계에 의한 베이지안 방법은 대부분의 인식기 집합에 있어서 1차 의존관계에 의한 베이지안 방법 보다 좋은 인식률을 보여 주어, 가능한 고차 정보를 사용하는 것이 유익함을 알 수 있었다. 결합되는 인식기의 갯수가 증가하면서 인식률이 하락하는 BKS 방법은 많고 충분한 훈련 데이타 세트로부터 누적 빈도수 테이블이 생성되지 않은 이유라고 볼 수 있으며, 그러한 데이타가 보장되는 경우에 사용하는 것이 바람직하다고 볼 수 있다. 실험 결과를 요약하면, 제안된 의존관계 기반의 프레임워크가 독립 가정 없이 다수 인식기를 결합하여 인식 성능의 실질적인 향상에 기여했음을 실험 결과로부터 얻었다고 볼 수 있다.

## 6. 결 론

본 논문에서는 인식기의 독립 가정 없이 다수 인식기를 확률적으로 결합하기 위한 의존관계 기반의 프레임워크를 제안하였다. 이 프레임워크는 2차 이상의 의존관계를 고려하여 최적의 곱 근사 분포를 결정할 수 있도록 새로이 제안된 알고리즘과 이 근사 분포를 기반으로 베이지안 방법을 사용하여 다수 인식기를 결합하는 결정 결합기로 구성된다. 또한, 제안된 프레임워크는 독립 가정에 의한 베이지안 방법과 BKS 방법의 장점을 취하도록 일반화한 방법으로서 공간 저장량과 인식 성능에서 효율적임을 보여 주었다. $d$차 의존관계에 의한 베이지안 방법으로 다수 인식기를 결합하는 것이 다른 방법보다 우수한 인식 성능을 보였음을 CENPARMI 데이타베이스를 이용한 필기 숫자 인식 실험으로부터 알 수 있었다.

## 참 고 문 헌


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

[2] Hull, J. J., Commike, A., and Ho, T. K., "Multiple algorithm for handwritten character recognition," In *Proceedings of the 1st Int. Workshop on Frontiers in Handwriting Recognition*, pp. 117-129, 1990.

[3] Suen, C. Y., Nadal, C., Mai, T. A., Legault, R., and Lam, L., "Recognition of totally unconstrained handwritten numerals based on the concept of multiple experts," In *Proceedings of the 1st Int. Workshop on Frontiers in Handwriting Recognition*, pp. 131-143, 1990.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] Kang, H.-J. and Lee, S.-W., "A Dependency-based Framework of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals," In *Proceedings of 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.2, pp. 124-129, 1999.

[12] Ho, T. K., Hull J. J., and Srihari, S. N., "Decision Combination of Multiple Classifier Systems," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(1), pp. 66-75, 1994.

[13] Woods, K., Kegelmeyer, W. P. Jr., and Bowyer, K., "Combination of Multiple Classifiers Using Local Accuracy Estimates," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(4), pp. 405-410, 1997.

[14] Kittler, J., Hatef, M., Duin, R. P. W., and Matas, J., "On Combining Classifiers," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(3), pp. 226-239, 1998.

강 희 중

1986년 서울대학교 전자계산기공학과 졸업(학사). 1988년 한국과학기술원 전산학과 졸업(석사). 1997년 한국과학기술원 전산학과 졸업(박사). 1986년 ~ 1998년 삼성전자주식회사 기업통신개발그룹 선임연구원. 198년 ~ 2000년 고려대학교 인공시각연구센터 연구조교수. 2000년 ~ 현재 한성대학교 정보전산학부 교수. 관심분야는 인공지능, 패턴인식, 그룹웨어 임.

이 성 환

정보과학지 논문지 : 소프트웨어 및 응용 제 27 권 제 5 호 참조