

## 확률적 자율 학습을 위한 베이지안 모델

### Bayesian Model for Probabilistic Unsupervised Learning

최준혁\* · 김중배\*\* · 김대수\*\*\* · 임기욱\*\*\*\*

Jun-hyeog Choi, Joong-bae Kim, Dae-su Kim, and Kee-wook Rim

\* 김포대학 컴퓨터계열

\*\* 한국전자통신연구원 컴퓨터·소프트웨어 기술연구소

\*\*\* 한신대학교 컴퓨터학과

\*\*\*\* 선문대학교 산업공학과

#### 요 약

Bishop이 제안한 Generative Topographic Mapping(GTM)은 Kohonen이 제안한 자율 학습 신경망인 Self Organizing Maps(SOM)의 확률 버전이다. GTM은 데이터가 생성되는 확률 분포를 잠재 변수, 혹은 은닉 변수를 사용하여 모형화한다. 이것은 SOM에서는 구현될 수 없는 GTM만의 특징이며 이러한 특징으로 인하여 SOM의 한계들을 극복할 수 있게 된다. 본 논문에서는 이러한 GTM 모형에 베이지안 학습(Bayesian learning)을 결합하여 작은 오분류율을 가지는 분류 알고리즘인 베이지안 GTM(Bayesian GTM)을 제안한다. 이 알고리즘은 기존의 GTM의 빠른 계산 처리 능력과 데이터에 대한 확률 분포, 그리고 베이지안 추론의 정확성을 이용하여 기존의 분류 알고리즘보다 우수한 결과를 얻게 된다. 본 논문에서는 기존의 분류 알고리즘에서 많이 실험하였던 학습 데이터를 통하여 이를 확인하였다.

#### Abstract

GTM(Generative Topographic Mapping) model is a probabilistic version of the SOM(Self Organizing Maps) which was proposed by T. Kohonen. The GTM is modelled by latent or hidden variables of probability distribution of data. It is a unique characteristic not implemented in SOM model, and, therefore, it is possible with GTM to analyze data accurately, thereby overcoming the limits of SOM. In the present investigation we proposed a BGTM(Bayesian GTM) combined with Bayesian learning and GTM model that has a small mis-classification ratio. By combining fast calculation ability and probabilistic distribution of data of GTM with correct reasoning based on Bayesian model, the BGTM model provided improved results, compared with existing models.

**Key Words** : Generative Topographic Mapping, Bayesian Inference, Slice Sampling

#### 1. 서 론

신경망 분야에서 Kohonen은 주어진 입력에 대한 정확한 출력을 모르는 상황에서 스스로 학습할 수 있는 Self Organizing Maps(SOM)[4][6]을 제안하였다. SOM 알고리즘은 문서 분류, 텍스트 마이닝(text mining) 등 매우 다양한 분야에서 좋은 성능을 나타내며 현재까지 폭 넓게 사용되어 왔다. 하지만 SOM 알고리즘은 데이터와 모형에 대한 명확한 규칙에 의하지 않고 휴리스틱한 방법에 지나치게 의존하는 블랙 박스와 같은 구조 때문에 모형에 대한 정확한 해석을 필요로 하는 경우에는 여러 한계에 부딪치게 된다[2].

특히 패턴 분류와 같은 경우에는 이러한 문제가 더욱 극명하게 나타난다. Bishop은 이러한 SOM의 한계를 극

복하기 위하여 학습 데이터에 대한 분포의 개념을 도입하여 모형에 이론적 규칙을 뒷받침할 수 있는 SOM에 대한 확률 버전으로서 Generative Topographic Mapping(GTM)[2]을 제안하였다. GTM 모형은 데이터의 차원에 비해 상대적으로 매우 작은 차원을 갖는 잠재 변수(latent variable), 혹은 은닉 변수(hidden variable)로 표현되는 사상을 통하여 낮은 차수의 공간(space)에서 데이터에 대한 확률 분포(probability density)를 모형화할 수 있는 개념을 도입하여 SOM의 여러 한계를 극복할 수 있는 바탕을 마련하였다. 특히 GTM은 EM(Expectation Maximization) 알고리즘을 사용한 학습을 통하여 잠재 공간(latent space)에서 데이터 공간(data space)으로의 비선형 변환(non-linear transformation)을 수행하는 함수를 결정한다.

본 논문에서는 이러한 GTM 모형에 베이지안 추론(Bayesian inference)을 적용하여 패턴을 분류하는 모형을 제안하였다. 이러한 베이지안 추론에 의한 베이지안 GTM 모형은 기존의 신경망의 과대적합(over-fitting)과 모형 설명의 어려움[1][7]을 극복하고 분류의 정확성을

접수일자 : 2001년 11월 1일

완료일자 : 2001년 12월 1일

감사의 글 : 본 연구는 2001학년도 김포대학의 연구비 지원에 의하여 연구되었음

향상시킨다. 본 논문의 2장에서는 일반적인 GTM 모형에 대해 알아보고, 3장에서는 본 논문에서 제안하는 베이저안 GTM에 대한 구조를 설명하고, 4장에서는 베이저안 GTM과 기존의 분류 알고리즘과의 분류 성능에 대해 비교한 실험 결과를 보이고, 마지막으로 5장에서는 베이저안 GTM의 향후 연구 계획에 대해서 알아본다.

## 2. GTM 모형

많은 변수들 간에는 상당한 상관 관계가 존재한다. 이러한 것들을 모형화하는 한 가지 방법은 잠재 변수(latent variable)를 이용하여 자료의 분포를 모형화하는 것이다. GTM은 잠재 변수에 의해 모형을 표현한다. 잠재 변수 모형은 고차원의 공간을 잠재 변수를 이용하여 저차원의 공간으로 변환시킨다. 자료 공간에서 잠재 공간으로 선형 변환을 통해 차원을 축소시키는 인자 분석 등은 잠재 변수 모형의 한 예가 될 수 있다. 잠재 변수 모형의 중요한 적용은 바로 자료의 시각화를 통한 군집 분석이다. 많은 시각화 모형은 D-차원의 자료 공간으로부터 2차원의 시각 공간으로의 투영으로 정의된다. 반면에 GTM 모형은 잠재 공간에서 자료 공간으로의 사상으로 정의된다. GTM은 자료 공간에서 잠재 공간으로 차원을 축소해 비선형 변환을 사용하는 비선형 잠재 변수 모형이다. 이 때, 모수의 추정에는 EM (Expectation Maximization) 알고리즘을 사용한다. 결국 GTM은 모형에 대한 설명력이 부족한 SOM의 이론적 한계를 극복하고 좀 더 정확한 데이터의 군집화를 수행하게 된다.

### 2.1 Latent variable

잠재 변수 모형은 작은 L차원의 잠재 변수들,  $x=(x_1, \dots, x_L)$ 를 이용하여 D 차원 공간의 데이터,  $t=(t_1, \dots, t_D)$ 의 확률 분포,  $p(t)$ 를 구한다. 이때 잠재 변수 공간에서 데이터 공간으로의 사상(mapping)은 모수 집합,  $W$ 를 갖는 비선형 함수(non-linear function),  $y(x; W)$ 를 이용하여 이루어 진다. 이러한 사상에서 잠재 변수 차원의 크기 L은 데이터 차원의 크기 D에 비해서 매우 작게 된다. 따라서 D차원의 데이터들은 L차원의 잠재 변수에 의해 서로 유사한 패턴끼리 분류되게 된다. 여기서 잠재 변수 공간에서 이들의 확률 분포,  $p(x)$ 를 정의하게 되면 이것이 잠재 변수들의 사전 확률 분포(prior probability distribution)가 된다. 이러한 사전 확률 분포는 GTM에 베이저안 학습을 이용할 수 있는 요소를 제공하여 준다. 잠재 변수,  $x$ 와 모수 집합  $W$ 의 확률 분포  $y(x; W)$ 가 주어지면 분산  $\beta^{-1}$ 를 갖는 데이터  $t$ 의 확률 분포를 유도해 낼 수 있고, 이러한 분포는 다음과 같이 일반적으로 가우시안 분포(gaussian distribution)를 갖게 된다.

$$p(t|x, W, \beta) \sim N(t|y(x; W), \beta^{-1}I)$$

위 식의 가우시안 분포는 다음과 같은 형태의 지수족(exponential family)을 이루게 된다.

$$N(t|y(x; W), \beta^{-1}I) = \left(\frac{\beta}{2\pi}\right)^{-D/2} \exp\left\{-\frac{\beta}{2} \sum_d (t_d - y_d(x; W))^2\right\}$$

여기서 주어진  $W$ 에 대해서  $t$ 공간의 데이터 분포는 다음과 같이 잠재 변수  $x$ 의 분포에 대한 적분을 통하여 얻게 된다.

$$p(t|W, \beta) = \int p(t|x, W, \beta)p(x)dx$$

잠재 변수의 분포  $p(x)$ 는 다음과 같은 델타 함수(delta function)의 형태를 취하게 된다.

$$p(x) = \frac{1}{K} \sum_k \delta(x - x_k)$$

이러한 잠재 변수의 함수를 통해 최종적으로 다음과 같은 식의 데이터 분포를 얻게 된다.

$$p(t|W, \beta) = \frac{1}{K} \sum_k p(t|x_k, W, \beta)$$

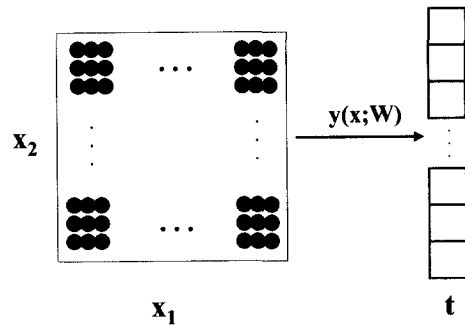


그림. 1 GTM Mapping

$W$ 와  $\beta$ 의 가능도 함수(likelihood function)는 다음과 같이 구할 수 있다.

$$L(W, \beta) = \sum_{n=1}^N \ln p(t^n|W, \beta)$$

여기서, 모수 집합,  $W$ 는 위의 식에 대한 최대 가능도 추정량(Maximum Likelihood Estimator:MLE)으로서 구한다. 이 MLE는 잠재 변수의 사전 확률 분포인  $p(x)$ 를 이용하여 잠재 변수 공간으로부터 데이터 공간으로의 사상,  $y(x; W)$ 를 구하게 된다.

### 2.2 EM 알고리즘

GTM 모형에서 최종적으로 구한 사상,  $y(x; W)$ 에서의 특정 모수의 결정은  $W$ 에 대한 미분 함수가 된다. 즉,  $L(W, \beta)$ 를 최대화하는  $W$ 와  $\beta$ 를 결정하기 위하여 공역경사도(conjugate gradients), 혹은 뉴턴 방법(Newton methods)과 같은 비선형 최적화(non linear optimization)의 기법을 사용하게 된다. 그러나 두 개 이상의 서로 다른 형태의 모형이 결합된 혼합 모형(mixture model)은 이러한 일반적인 방법으로는 모수의 최적화가 어렵다. 이러한 혼합 모형의 최적화는 EM(expectation maximization) 알고리즘을 사용한다[1].  $W$ 와  $\beta$ 에 대한 초기값이 주어지면 GTM을 위한 EM 알고리즘은 우선 E-step을 수행한다. 이 단계에서는 다음 식과 같은 가우시안 혼합 모형의 신뢰도(responsibilities)를 계산한다.

$$r_{kn} = p(x_k | t_n, W, \beta) = \frac{p(t_n | x_k, W, \beta) p(x_k)}{\sum_k p(t_n | x_k, W, \beta) p(x_k)}$$

여기서  $r_{kn}$ 은  $k$ 번째 잠재 변수에서의  $n$ 번째 데이터에 대한 가우시안 혼합 모형의 신뢰도를 나타낸다. 이 신뢰도를 통하여  $k$ 번째 잠재 변수에 의해 생성된  $n$ 번째 데이터에 대한 사후 확률 분포를 계산할 수 있게 된다. 다음으로 M-step인 두 번째 단계에서는  $W$ 와  $\beta$ 의 모수 방정식을 신뢰도 함수의 갱신을 통하여 계산한다. 본질적으로는 잠재 변수들의 각 모수들이 학습하는 데이터의 특성을 가장 잘 반영하게 되는 쪽으로 갱신되게 된다. 이때 오차 함수(error function)와 같은 목적 함수(objective function)를 사용할 수 있다.

### 2.3 GTM 알고리즘

- 1) Generate latent points  $\{x_k\}$ ,  $k = 1, \dots, K$ .
- 2) Initialize the set of parameters,
  - 2.1) Initialize  $W$  randomly or using PCA.
  - 2.2) Initialize  $\beta$
- 3) If desired, select a value for  $\alpha$ .  
where,  $\alpha$  is the criterion of convergence.
- 4) Repeat EM algorithm until ( $< \alpha$  ).
  - 4.1) E-step
  - 4.2) M-step

## 3. Bayesian GTM

### 3.1 베이지안 학습

추정, 혹은 검증하고자 하는 모수(parameter)를 확률 변수로 취급하고 이에 대한 사전 정보를 모수의 확률 분포 형태로 표현하는 방법이 베이지안 방법(Bayesian methods)이다. 이러한 베이지안 방법에서 관심있는 변수의 분포는 과거의 경험(학습)에서 얻은 변수의 특성과 실험에서 얻은 관측값을 사용하여 수정, 재조정될 수 있다. 본 논문에서 각 패턴의 특성을 설명해 주는 모수가 베이지안 방법에서의 관심있는 변수라고 하면 베이즈 정리에 의해 모수의 사후 확률 분포(posterior probability distribution)에 근거하여 모수의 통계적 추론이 이루어지고 이러한 과정이 반복되어 최종적인 패턴의 특성에 대한 분포의 모수가 갱신되어가는 과정을 베이지안 학습(Bayesian learning)으로 본 논문에서는 정의하였다. 베이지안 학습의 기본 형태는 다음과 같은 사후 확률 분포를 구하는 베이지안 방법의 반복 구조를 따른다.

$$p(w|D, X) = \frac{p(D|w, X)p(w|X)}{p(D|X)}$$

여기서,  $w$ 는 모수를 나타내고  $X$ 는 학습 데이터를 나타내며,  $D$ 는 패턴을 분류하기 위하여 주어진 데이터를 표시한다. 즉,  $p(w|D, X)$ 는 학습 데이터  $X$ 로 계산된 사후 확률 모형이고, 이 확률 모형을 통하여 특정 입력 데이터  $D$ 가 특정 패턴  $w$ 로 분류되는 확률을 구하게 된다. 이 확률값이 가장 큰 패턴으로 주어진 입력 데이터는 분류된다. 여기서 각 패턴의 특성을 나타내는 사전 확률 분포  $p(w)$ 는 일반적으로 다음과 같은 형태의 지수

족(exponential family)인 가우시안 분포를 사용하였다.

$$p(w) = \frac{1}{Z_w(a)} \exp(-aE_w)$$

이러한 지수족을 사용하는 이유는 베이지안 방법이 공액 사전 확률 함수(conjugate prior)를 사용하게 되면 사후 확률 분포의 수학적 계산이 매우 용이하기 때문이다. 하지만 대부분의 데이터에 대한 사전 확률 분포는 이러한 정형화된 지수족을 따르지 않게 되어 임의로 데이터에 대한 분포를 지수족으로 가정하게 되면 데이터에 대한 특성을 충분히 반영하지 못하게 되어 결국은 모형의 성능을 저하시키게 된다. 그러므로 본 논문에서는 사전 확률 분포가 지수족을 포함한 어떠한 분포로 정의되어도 사후 확률 분포를 계산할 수 있는 알고리즘으로 제안 알고리즘을 구축하였다. 즉 임의의 사전 확률 분포에 대한 사후 확률 분포의 계산을 마코프 체인 몬테 카를로(Markov Chain Monte Carlo;MCMC) 적분을 통해 수행하였다.

### 3.1 모형의 구조

$D$ -차원 자료 벡터  $D$ 는  $L$ -차원의 잠재 변수 벡터  $x$ 와 잡음(noise) 벡터  $e$ 를 이용하여 다음 식 (1)과 같이 표현된다.

$$D = y(x, W) + e \tag{1}$$

여기서,  $W$ 는 모수 행렬(parameter matrix)이고  $y$ 는 잠재 변수와 가중치의 함수로 표현된다. 즉, 데이터 집합은 개개의 데이터가 속하게 되는 집단에 대한 특성을 가장 잘 표현하는 잠재 변수,  $x$ 를 생성하게 되고 이러한  $x$ 에 의한 분포를 통해 역으로 각 데이터가 속하게 되는 패턴을 사상시키게 된다. 이러한 과정을 통하여 개개의 학습 데이터는 각 패턴의 잠재 변수를 모형화 하게 되고 이러한 확률 분포를 통하여 새로운 데이터를 포함한 데이터들의 패턴을 정확히 분류하게 된다. 또한 각 자료는 베이즈 법칙(Bayes' rule)[3]에 의해 식 (2)와 같은 사후 확률식을 갖는다.

$$p(x|D, W) = \frac{p(D|x, W)p(x)}{p(D|W)} \tag{2}$$

이 사후 확률식으로부터 구체적인 확률값은 Slicing sampling[5]에 의해 추출된 표본을 이용한 Markov Chain Monte Carlo(MCMC) 적분[8]에 의해 구해진다. 이렇게 확률식의 모수가 추정되면 최종 갱신 분포를 이용하여 본격적인 패턴 분류 작업이 수행된다.

### 3.2 Slice Sampling

Slice sampling[9]은 Gibbs sampling, Metropolis algorithm과 유사한 MCMC 방법이다. 이 알고리즘은 균등 분포(uniform distribution)를 사용함으로써 Gibbs sampling이나 Metropolis algorithm에 보다 원하는 분포로부터 간편하게 표본들을 뽑을 수 있다. Slice sampling은 [그림 2]의 각 단계를 반복 수행한다.

Gibbs sampling이나 Metropolis algorithm과 같은 기존 방법들의 문제점은 임의 보행 현상(random walk behaviors)이 발생한다는 것이다. Slice sampling도 마찬가지로 임의 보행 현상을 피할 수는 없지만 기존의 방법들보다는 임의 보행 현상에서 많이 벗어날 수 있고 적용이 쉽다.

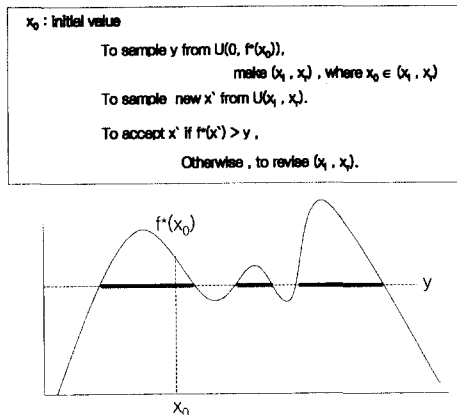


그림. 2 Slice sampling 알고리즘

그러나 단계의 길이가 너무 작으면 슬라이스를 만들 때까지 불필요하게 많은 시간이 소요되고, 너무 크면 슬라이스를 수정하는데 불필요한 작업을 반복해야 하기 때문에 슬라이스의 크기 결정에 신중해야 한다.

본 논문의 베이저안 GTM에서 분류함수로 사용되는 사후 확률 분포의 최종 확률 계산을 위한 몬테 카를로 시뮬레이션에 사용되는 표본의 추출에 이 Slice sampling을 사용한다.

### 3.3 Bayesian GTM 알고리즘

기본적으로 GTM 모형은 잠재 변수들의 사전 확률 분포와 모수와 분산의 가능도 함수 구조를 가지고 있기 때문에 베이저안 추론의 적용을 위한 매우 유연한 구조를 가지고 있다. 우선적으로 데이터의 벡터 구조로부터 각 잠재 변수들의 확률 분포에 대한 모수를 결정하고 분산 값이 주어지게 된다. 다음으로 잠재 변수를 이용하여 잠재 변수 공간으로부터 데이터 공간으로 사상하는 함수를 결정하고 이러한 사상의 함수는 베이저안 학습을 통하여 갱신하게 된다. 이렇게 모든 데이터를 학습하여 최종적으로 얻게 된 사후 확률 분포를 이용하여 데이터의 패턴 분류 작업이 이루어진다. 이러한 과정을 정리하면 다음과 같이 된다.

- 1) Initialize
  - 1.1) Set data vectors.  
 $D = (D_1, D_2, \dots, D_D)$
  - 1.2) Prior of weights.  
 $W \sim f(\theta)$
- 2) Determine Posterior probability
  - 2.1) Mapping to Latent variable.  
 $D = y(x, W) + e$
  - 2.2) Compute posterior probability distribution.  
$$p(x|D, W) = \frac{p(D|x, W)p(x)}{p(D|W)}$$
- 3) Classification
  - 3.1) Classification by posterior probability distribution.  
this phase needs the slice sampling for MCMC simulation.
  - 3.2) Replace until converge.

## 4. 실험 및 결과

본 논문의 Bayesian GTM의 패턴분류 실험을 위하여 군집의 결과가 알려져 있는 Fisher의 붓꽃 자료와 digit data를 이용하여 기존의 분류 기법으로 주로 사용되고 있는 CHAID, CART와 본 알고리즘의 성능을 오분류율 측도로 비교하였다. 본 실험에서 사용한 데이터의 분포는 가우시안 분포(Gaussian distribution)를 사용하였다. 이는 많은 데이터가 가우시안 분포를 따르고 있고, 베이저안 추론에 대한 적용이 용이하면서도 정확도가 높기 때문이다.

### 4.1 Fisher의 붓꽃 자료를 이용한 실험

표 1. Iris data의 분류 결과

	Chi	Gini	BGTM
1	0.088	0.074	0.049
2	0.074	0.056	0.027
3	0.069	0.052	0.007
4	0.066	0.066	0.072
5	0.085	0.051	0.024
6	0.029	0.029	0.005
7	0.183	0.050	0.040
8	0.083	0.067	0.001
9	0.015	0.045	0.005
10	0.094	0.063	0.053
11	0.086	0.052	0.007
12	0.046	0.062	0.021
13	0.070	0.018	0.008
14	0.098	0.066	0.039
15	0.095	0.079	0.022
16	0.062	0.015	0.005
17	0.167	0.074	0.064
18	0.051	0.051	0.001
19	0.095	0.054	0.051
20	0.136	0.061	0.017

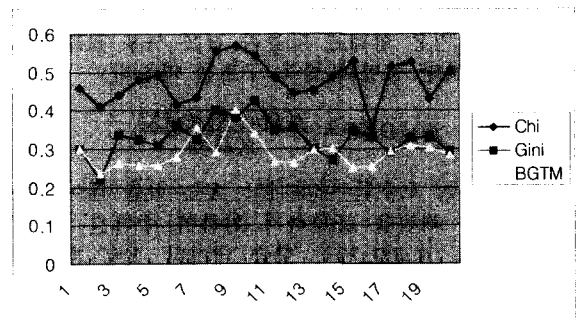


그림. 2 Iris data 오분류율 비교

Bayesian GTM과 기존의 대표적인 분류 알고리즘들을 비교하기 위하여 Fisher의 Iris 자료[11]를 이용하였다. 이 자료는 붓꽃의 특성치에 관한 자료로서 4개의 입력 변수를 갖는다.

입력 변수가 되는 꽃받침의 길이, 꽃받침의 넓이, 꽃잎의 길이, 꽃잎의 넓이에 따라 붓꽃의 종류를 Sectosa, Versicolor, Verginica의 세 집단으로 분류하였다. 즉, 4개의 입력 변수와 1개의 목표 변수로 이루어진다. 초기 가중치는 가우시안 사전 분포로 하고, 베이즈의 규칙에 의해 갱신되는 사후 확률 분포도 마찬가지로 사후 확률 분포가 된다. 그리고 각 개체의 분류를 위해서 사후 확률 분포로부터 얻게 되는 최종 분류 예측 분포의 확률 계산은 Slice sampling을 사용하였다. 이에 대한 실험 결과 표 1 과 그림 2에 나타나 있다.

표 2. Digit data의 분류 결과

	Chi	Gini	BGTM
1	0.460	0.299	0.300
2	0.411	0.219	0.237
3	0.442	0.338	0.263
4	0.482	0.325	0.258
5	0.493	0.310	0.256
6	0.415	0.362	0.279
7	0.434	0.325	0.355
8	0.553	0.400	0.292
9	0.570	0.384	0.402
10	0.541	0.425	0.340
11	0.488	0.350	0.266
12	0.448	0.356	0.263
13	0.455	0.299	0.298
14	0.487	0.269	0.300
15	0.529	0.345	0.251
16	0.352	0.330	0.253
17	0.513	0.290	0.293
18	0.526	0.329	0.310
19	0.433	0.330	0.303
20	0.506	0.294	0.285

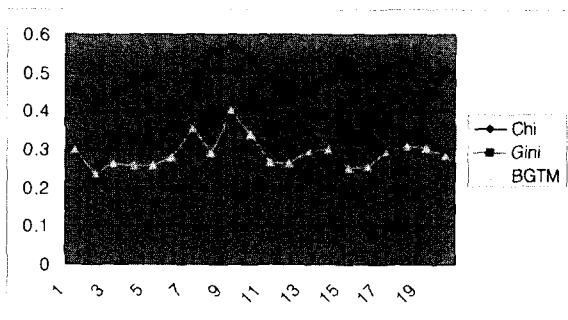


그림 3 Digit data 오분류율 비교

20번의 Iris data의 반복 실험의 결과를 통해서 각 알고리즘의 오분류율의 평균을 비교해 보면 CHAID(Chi), CART(Gini) Bayesian GTM이 각각 0.0846, 0.0541, 0.0257로서 기존의 분류 알고리즘에 비해서 본 논문의 Bayesian GTM의 오분류율이 훨씬 작음을 알 수 있다. 이 결과를 통하여 베이지안 추론을 적용한 Bayesian GTM 모형이 기존의 다른 분류 알고리즘에 비해 좋은 분류 결과를 보임을 알 수 있다.

#### 4.2 Digit data를 이용한 실험

고장난 전자계산기로부터 얻어진 자료[11]로서 디지털 숫자의 각 위치를 나타내는 7개의 입력 변수와 0에서 9까지의 숫자를 인식하는 1개의 목표 변수로 이루어진다. 이 실험을 위한 Bayesian GTM 구조도 앞의 Iris data와 같은 구조를 갖는다. 이 digit data의 Bayesian GTM 실험 결과는 아래의 <표 2>와 그림 3에 나타나 있다.

앞의 Iris data의 결과와 마찬가지로 digit data의 실험에서도 Bayesian GTM 모형이 기존의 분류 알고리즘인 CHAID(Chi)나 CART(Gini)에 비해서 오분류율이 훨씬 작음을 알 수 있다. 즉 CHAID(or Chi), CART(or Gini) Bayesian GTM이 각각의 20번 반복 실험에서 오분류율의 평균이 0.4768, 0.3289, 0.2901로 나타났다.

### 5. 결론

1980년대 중반 Kohonen이 제안한 SOM 모형은 Unsupervised Learning에서 매우 많이 쓰이는 신경망 모형이다. 하지만 SOM은 모형을 정의하는 이론적 근거가 부족하였다. 이러한 단점을 보완하기 위해 Bishop은 1990년대 후반에 SOM의 확률 버전(probabilistic version)인 GTM을 제안하였다. 이 모형은 잠재변수를 사용하여 다차원의 자료를 저차원으로 사상(Mapping)시켜 자료를 군집화하는 알고리즘이다.

본 논문에서는 이러한 GTM 모형에 Bayesian 추론을 적용하여 분류 알고리즘으로 사용하였다. 이 분류 알고리즘은 오분류율이 기존의 알고리즘에 비해 낮고 비교적 쉽게 적용할 수가 있었다. 향후, 본 논문의 Bayesian GTM을 범주형 데이터를 포함하는 모든 데이터 영역(domain)으로 확장, 발전시킨다면 각종 생체인식 시스템에도 적용되어 작은 오분류율로 패턴 인식 시스템을 관리할 수 있으리라 기대한다.

### 참 고 문 헌

- [1] Bishop, C. M. "Neural Networks for Pattern Recognition", Oxford University Press, 1995.
- [2] Bishop, C. M., Svensen, M., Williams, C. K. I. "GTM : a Principled Alternative to the Self Organizing Map", Proceedings 1996 International Conference on Artificial Neural Networks, ICANN 96, Volumell12. pages 165-170, Bochum, Germany, 1996.
- [3] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. "Bayesian Data Analysis", Chapman & Hall, 1995.

[4] Teuvo Kohonen, "Self-Organizing Maps", Springer, 1997.  
 [5] Neal, R. M. "Bayesian Learning for Neural Networks", Springer, 1996.  
 [6] Abhijit S. Pandya, Robert B. Macy. "Pattern Recognition with Neural Networks in C++", IEEE Press, 1995.  
 [7] Ripley, B. D. "Pattern Recognition and Neural Networks", Cambridge University Press, 1996.  
 [8] Martin A. Tanner, "Tools for Statistical inference", Springer, 1996.  
 [9] R. M. Neal, "Bayesian Learning for Neural Networks", Springer, 1996.  
 [10] Thomas Leonard, John S. J. Hsu, "Bayesian Methods", Cambridge University press, 1999.  
 [11] UCI Machine Learning Repository, <http://www1.ics.uci.edu/~mllearn>

**저 자 소 개**



**최준혁(Choi Jun-hyeog)**

1990년 : 경기대학교 전자계산학과 졸업 (이학사)  
 1995년 : 인하대학교 대학원 전자계산공학과 졸업(공학석사)  
 2000년 : 인하대학교 대학원 전자계산공학과 졸업(공학박사)

1997년~현재 : 김포대학 컴퓨터계열 조교수

관심분야 : 정보검색, 데이터마이닝, 신경망, 유전자 알고리즘 등  
 e-mail : jhchoi@kimpo.ac.kr



**김중배(Kim Joong-bae)**

1986년 : 고려대학교 공과대학 산업공학과 졸업(공학사)  
 1988년 : 한국과학기술원 산업공학과 졸업(공학석사)  
 1998년~현재 : 한국과학기술원 산업공학과 박사과정

1988년~1991년 : 대한항공(주) 시스템부  
 1991년~현재 : 한국전자통신연구원 컴퓨터소프트웨어기술연구소 모바일응용서버연구팀장

관심분야 : 인터넷 정보검색, 미들웨어, 시스템 소프트웨어 등  
 e-mail : jjkim@etri.re.kr



**김대수(Kim Dae-su)**

1977년 : 서울대학교 사대수학과 학사  
 1986년 : 미국 Univ. of Mississippi, Computer Science, M. S.  
 1990년 : 미국 Univ. of South Carolina, Computer Science, Ph. D.  
 1991년~1993년 : 한국전자통신연구원 컴퓨터연구단 선임연구원

1993년~현재 : 한신대학교 컴퓨터학과 교수

관심분야 : 신경망, 퍼지, 인공지능, 지능시스템, 에이전트, 융합 모델링

Phone : +82-031-370-6784

Fax : +82-031-372-3343

E-mail : daekim@hanshin.ac.kr



**임기욱(Rim, KeeWook)**

1977. 2. : 인하대학교 공과대학 전자공학과 졸업  
 1987. 2. : 한양대학교 전자계산학 석사  
 1994. 8. : 인하대학교 전자계산학 박사  
 1977~1983 : 한국전자기술연구소 선임연구원  
 1983~1988 : 한국전자통신연구원 시스템소프트웨어 연구실장

1988~1989 : 미 캘리포니아 주립대학(Irvine) 방문연구원

1989~1997 : 한국전자통신연구원 시스템연구부장

주전산기(타이컴) III, IV 개발 사업책임자

1997~2000 : 정보통신연구진흥원 정보기술전문위원

2000~현재 : 선문대학교 산업공학과 교수

관심분야 : 실시간 데이터베이스시스템, 운영체제, 시스템구조 등  
 e-mail : rim@omega.sunmoon.ac.kr