

자기조직화 지도를 위한 베이지안 학습

전성해¹⁾ 전홍석²⁾ 황진수³⁾

요약

Kohonen이 제안한 자기조직화 지도(Self Organizing Maps : SOM)는 매우 빠른 신경망 모형이다. 하지만 다른 신경망 모형과 마찬가지로 학습 결과에 대한 명확한 규칙을 제시하지 못할 뿐만 아니라 지역적 최적값으로 빠지는 경우가 종종 있다. 본 논문에서는 이러한 자기조직화 지도의 모형에 대한 설명력을 부여하고 전역 최적값으로 수렴할 수 있는 예측 성능을 갖는 모형으로서 자율학습 신경망에 베이지안 추론을 결합한 자기조직화 지도를 위한 베이지안 학습(Bayesian Learning for Self Organizing Maps ; BLSOM)을 제안한다. 이 방법은 기존의 자기조직화 지도가 지역적 해에 머물러 있는 것에 비해서 언제든지 전역적 해로 수렴함이 실험을 통하여 밝혀졌다.

주요용어: 자기조직화 지도, 베이지안 학습, 군집화, 자율학습 신경

1. 서론

신경망 분야에서 Kohonen은 주어진 입력에 대한 정확한 출력을 모르는 상황에서 스스로 학습할 수 있는 자기조직화 지도(Self organizing maps)를 제안하였다. 이 모형은 다른 신경망과 마찬가지로 모형에 대한 정확한 해석이 어렵다. 또한 사전에 군집 수를 알면서 군집화를 수행하는 계층적 군집분석이나 k-평균 군집방법에 비해 군집의 정확성이 떨어진다. 하지만 군집화를 수행하는 계산속도는 매우 빠르다. 통계학 분야의 베이지안 추론(Bayesian inference)에서 자료를 얻기 이전의 모수에 대한 정보인 사전 분포(Prior distribution)는 관측된 자료에 의한 우도 함수(Likelihood function)와의 결합에 의해 사후 분포(Posterior distribution)로 갱신된다. 이러한 베이지안 추론이 신경망에 적용되면 한 개의 값에 의한 가중치 갱신이 분포에 의한 갱신으로 바뀌어진다. 이러한 베이지안 추론에 의한 신경망은 기존의 신경망의 과대적합(over-fitting)과 모형 설명의 어려움을 극복하고 예측의 정확성을 향상시킨다. Utsugi는 자기조직화 지도의 통계적 근사 모형을 만들어 베이지안 추론과 결합시킨 베이지안 자기조직화 지도(Bayesian self organizing maps : BSOM)를 제안하였다. 하지만 이 모형은 자기조직화 지도에 직접 베이지안 추론이 적용되지 않았다. 그리고 자기조직화 지도와 마찬가지로 한 개의 값에 의한 가중치 갱신이 이루어진다. 하지만 본 논문

1) (121-742) 서울특별시 마포구 신수동 1, 서강대학교 컴퓨터학과 박사과정

E-mail: shjun@ailab.sogang.ac.kr

2) (402-751) 인천광역시 남구 용현동 253, 인하대학교 통계학과 교수

E-mail: hsjorn@anova.inha.ac.kr

3) (402-751) 인천광역시 남구 용현동 253, 인하대학교 통계학과 부교수

E-mail: jshwang@anova.inha.ac.kr

에서 제안하는 자기조직화 지도를 위한 베이지안 학습(Bayesian learning for self organizing maps) 알고리즘은 Utsugi의 베이지안 자기조직화 지도와는 다르게 Kohonen의 자기조직화 지도 안에서 직접 베이지안 학습이 적용되어 가중치 갱신이 분포를 통해 이루어진다. 따라서 본 논문의 자기조직화 지도를 위한 베이지안 학습 알고리즘은 베이지안 자기조직화 지도와는 다른 관점에서 자기조직화 지도에 대한 모형의 설명력과 분석의 정확성을 향상시킨다.

2. 제안 배경

Kohonen이 제안한 SOM은 대표적인 자율 학습 신경망이다. 다른 신경망 학습 알고리즘에 비해서 SOM은 자료에 숨겨져 있는 기하학적 구조를 이끌어 내는 능력이 있다. 하지만 다른 신경망 모형들과 마찬가지로 SOM도 블랙 박스와 같은 신경망의 각 층들 간의 연결 학습을 통해서 결과를 계산하기 때문에 출력 결과에만 유일한 관점이 맞추어져 있다. 즉 근본적으로 학습의 결과에만 모형이 적용되기 때문에 결과값에 대한 과정 설명과 같은 높은 수준의 추론은 어렵다. 따라서 주어진 새로운 입력 자료에 대한 출력이 어떠한 규칙을 통해 계산되는지에 대한 명확한 설명이 어렵다. 학습 자료의 군집화에 있어서 SOM은 군집의 수를 사전에 결정해 주어야 하는 통계학의 계층적 군집분석과는 달리 주로 2차원 형상 지도의 차원의 크기만 결정해 주면 신경망이 스스로 학습하여 최적의 군집을 형성한다. 이는 자료의 정확한 특성을 사전에 알지 못해 군집의 수를 잘못 결정하면 그릇된 결론을 내릴 위험이 있는 계층적 군집 분석의 문제점을 해결한다. SOM과 BSOM에 의해서 최종적으로 구축된 모형은 한 개의 가중치 값으로 고정된 모수를 갖는 모형이 된다. 이 모형은 일반적으로 최적의 모형이라는 보장은 없다. 따라서 최종 구축된 모형에 대한 최적값의 계산이라는 보장을 주기 위해서 한 개의 고정된 값의 모수로 모형을 고정 시키지 않고 여러 가능성을 포함하는 분포의 개념을 모형의 모수에 도입해야 한다. 본 논문에서 제안하는 자기조직화 지도를 위한 베이지안 학습(Bayesian Learning for Self Organizing Maps ; BLSOM) 알고리즘은 이러한 SOM과 BSOM의 문제점들을 해결한 모형이다. 즉 SOM의 단점인 모형에 대한 설명력 부족을 베이지안 추론 방법으로 해결하였고, 사전에 군집의 수를 결정해야 하는 계층적 군집화의 문제점에 대해서는 SOM의 형상 지도를 이용하여 해결하였다. 그리고 BSOM의 최종적으로 구축된 모형에서 출력층의 각 노드가 한 개의 가중치 값들로 고정되는데 비해 BLSOM은 고정된 가중치 값이 아니고 가중치가 자신이 속하게 되는 분포를 갖게 된다. 이것은 인간의 뇌구조를 모형화하는 연결주의적 인공지능의 구현에 유용하게 사용된다. 즉 인간이 사고를 하는 과정은 인간이 갖고 있는 하나의 개념과 또 다른 소규모의 개념들이 연관되어 있고 이러한 개념들이 서로 영향을 주는 관계에 있다. 연결주의적 인공지능은 하나의 경험과 그것에 따른 하위 경험들, 또 이 하위 경험들에 영향을 미칠 수 있는 다른 하위 경험들에 의해 도출되는 새로운 정보 관계를 이용하여 기존의 인공지능 기법에서 불가능했던 유사성과 연관성에 의한 유연한 추론을 통해 인간과 같은 사고를 하는 할 수 있는 인공지능을 모형화 할 수 있게 된다. 각각의 경험들을 구성하는 모든 자료는 매우 많다. 따라서 이 모든 학습 자료를 모형을 통해 학습하기에는 불가능 할 정도로 많다.

따라서 각 경험들의 자료를 모두 학습하는 것 보다는 각 경험들을 가지고 있는 분포를 추정 후 새로운 정보를 위한 학습이 필요할 경우 베이지안 추론에 의한 사전 분포로서 표현하여 각 경험을 대표하면 모든 자료를 학습하지 않으면서도 주어진 하위 경험을 모두 고려한 새로운 모형을 구축하게 된다. 이러한 일련의 과정을 BLSOM이 해결한다. 또한 시간이 지날수록 자료가 거대하게 쌓이는 자료에서 정보를 찾아내야 하는 데이터 마이닝 작업을 인간이 직접 담당하기에는 한계가 있게 된다. 이 때 인간이 담당하는 데이터 마이닝 작업을 BLSOM에 의해 구현된 지능화된 인공지능이 담당한다면 훨씬 효율적으로 자료 마이닝 작업을 담당할 수 있게 된다. 이러한 시스템이 데이터 마이닝 에이전트이다.

3. Utsugi의 BSOM

베이지안 자기조직화 지도(Bayesian Self Organizing Maps ; BSOM)는 1990년대 중반 이후에 Utsugi에 의해 제안되었다(Utsugi 1996, 1997). Utsugi는 자기조직화 지도 모형에 베이지안 추론을 적용하기 위한 시도를 하였다. Utsugi는 이러한 모형을 만들기 위해서 우선 Kohonen의 자기조직화 지도에 해당하는 통계적 근사 모형(statistical approximate model)을 가우시안 혼합 분포를 이용하여 만들어 기존의 베이지안 추론을 적용하였다. 즉 Utsugi는 SOM을 통계적 일반화 변형 모형(Generalized Deformable Model ; GDM)[Yuille, 1990]과 같은 구조의 가우시안 평활 사전 확률 분포(Gaussian smoothing prior probability distribution)를 갖는 가우시안 혼합 모형(Gaussian mixture model)으로 근사시켰다. 이 근사 모형의 모수는 MAP(maximum a posterior) 추정 알고리즘으로서 계산된다. Utsugi의 BSOM의 이론적 구조는 SOM의 통계적 근사 모형과 모수에 대한 추정 알고리즘의 2가지로 구성된다. BSOM의 자료에 대한 우도 함수는 가우시안 분포로 한다. 즉 m 차원의 입력 자료, $x_i = (x_{i1}, \dots, x_{im})'$ ($i = 1, \dots, n$)는 r 차원의 가중치를 모수로 갖는 식 (3.1)의 형태의 우도 함수를 갖는다,

$$l(X|w, \beta) = \prod_{i=1}^n \sum_{s=1}^r \frac{1}{r} f(x_i|w_s, \beta). \quad (3.1)$$

여기서 $f(x_i|w_s, \beta)$ 는 평균이 w_s 이고 분산이 $\frac{1}{\beta}$ 인 식 (3.2)의 가우시안 분포이다.

$$f(x_i|w_s, \beta) = \left(\frac{\beta}{2\pi}\right)^{\frac{m}{2}} \exp\left(-\frac{\beta}{2}\|x_i - w_s\|^2\right). \quad (3.2)$$

BSOM은 기하학적 공간(topological space)을 따라 평균(mean, centroids)의 평활 변화(smooth variation)를 나타내는 가우시안 평활 사전 확률 분포(gaussian smoothing prior probability distribution)를 갖는다. 이러한 사전 확률 분포의 평활성(smoothness)은 가중치에 대한 사전 정보의 표현을 좀 더 일반적으로 표현하기 위한 것이다. 또한 가중치의 사전 정보에 대한 평활성 뿐만 아니라 기하학적 공간에서 적당한 이산화 차분 조작 행렬(discretized

differential operator matrix), 를 정의하여 식 (3.3)과 같은 평활 사전 분포를 만들었다.

$$f(w|\alpha) = \prod_{j=1}^n \left(\frac{\alpha}{2\pi}\right)^{\frac{1}{2}} (\det^+ D'D)^{1/2} \exp\left(-\frac{\alpha}{2}\|Dw_{(j)}\|^2\right). \quad (3.3)$$

여기서, $l = \text{rank} D'D$ 이고, $\det^+ D'D$ 는 $D'D$ 의 양의 고유치(positive eigenvalues)의 곱을 나타낸다. α 는 평활 상수(smooth constant)를 나타낸다.

식 (3.1)의 학습 자료에 대한 우도 함수와 식 (3.3)의 가우시안 평활 사전 확률 분포를 이용하여 식 (3.4)과 같은 사후 확률 분포를 구한다.

$$f(w|X, \alpha, \beta) = C f(X|w, \beta) f(w|\alpha) \quad C : \quad (3.4)$$

식 (3.4)의 분포에 로그를 취하면 식 (3.5)의 로그 사후 확률 분포를 구할 수 있다,

$$\log f(w|X, \alpha, \beta) = C + \log f(X|w, \beta) + \log f(w|\alpha). \quad (3.5)$$

BLSOM 모형의 모수의 추정에는 두 개의 알고리즘이 동시에 필요하다. 우선 모형의 가중치에 대한 추정에 사용되는 MAP 추정 알고리즘이 있다. 그리고 가중치에 대한 하이퍼 파라메타(hyper parameters)를 최적화 하는 알고리즘이다. Map 추정 알고리즘의 가장 간단한 형태는 가중치의 사후 확률 분포를 위한 기울기 상승(gradient ascent) 알고리즘이다. 이 알고리즘에서 가중치는 식 (3.6)의 로그 사후 확률 분포의 최고 급경사 방향(direction of steepest ascent)으로 갱신된다.

$$d(w) = \frac{\partial}{\partial w} \log f(X, w|\alpha, \beta, D). \quad (3.6)$$

그래서 최종 갱신된 가중치의 추정값은 $d(w)$ 를 최대로 하는 값이 된다. 또한 Map 추정을 위한 EM 알고리즘을 사용할 수 있다. 이 알고리즘은 근사 헤시안(hessian)을 사용한 근사 뉴턴-랩슨(Newton-Raphson) 알고리즘으로 간주되며 식 (3.7)과 같이 나타난다.

$$H_{EM}(w) = -E_Y \left(\frac{\partial^2}{\partial w \partial x'} \log f(X, Y|w, \beta) \Big| X, w, \beta \right) + H_p. \quad (3.7)$$

EM 알고리즘의 갱신 규칙은 식 (3.8)과 같이 뉴턴-랩슨 규칙을 따른다.

$$w^{(k+1)} = w^{(k)} + H_{EM}^{-1}(w^{(k)})d(w^{(k)}). \quad (3.8)$$

여기서 $w^{(k)}$ 는 k 번째 단계에서의 임시 추정치(temporary estimate)이다. EM 알고리즘은 대응되는 상향 경사로 알고리즘보다 더 우수한 갱신 방향(directions)을 가지고 있다. 그러나 EM 알고리즘은 수렴값에 빠르게 도달하지 못 한다. 이러한 문제를 해결하기 위해서 EM 알고리즘에 대한 가속화 방법이 필요하다(Jordan, 1995). Utsugi가 제안한 BSOM에서는 Kohonen이 제안한 SOM이라는 신경망 모형에 베이지안 추론을 적용한 것이 아니라 SOM과 유사한 통계적 모형을 찾아 베이지안 추론을 적용하였다. 하지만 다음 절에서 제안하는 본 논문의 BLSOM은 SOM 모형에 직접 베이지안 추론을 결합하려는 시도를 하였다.

BSOM의 최종 모형에서 가중치는 로그 사후 확률 분포를 최대화 하는 한 개의 고정된 값으로 결정된다. 때문에 최종적으로 구축된 모형에 대한 출력 결과는 동일한 입력 데이터에 대해서는 항상 같게 된다. 따라서 구축된 모형이 전역적 최적값이 아니라 지역적 최적값이 되는 경우에는 주어진 자료에 대해서 부정확한 결과를 구하게 된다. 이는 SOM이나 BSOM이 공통적으로 갖고 있는 문제이다. 이러한 문제를 해결하기 위해서 본 논문에서는 자기조직화 지도를 위한 베이지안 학습(Bayesian Learning for Self Organizing Maps : BLSOM) 알고리즘을 제안한다.

4. 제안 BLSOM의 구조와 알고리즘

본 논문에서 제안하는 BLSOM은 기존의 SOM이나 BSOM 모형의 가중치 갱신이 한 개의 고정된 값에 의존하는 것과는 다르게 가중치가 속한 확률분포의 모수가 학습 자료에 의해 갱신되는 구조를 갖는다. 따라서 새로운 자료의 군집화를 위한 신경망 모형의 가중치 결정시에 최종 갱신된 가중치 확률분포로부터 생성되는 가중치를 사용하기 때문에 SOM이나 BSOM과는 달리 동일한 입력 자료에 대해서 항상 같은 결과만을 제공하지는 않는다. 따라서 한 번 국지적 최적값에 빠지면 최적의 군집 결과를 제공하기 어려운 SOM이나 BSOM과는 달리 지역적 최적값에 빠진 경우에도 가중치 확률 분포로부터 생성되는 새로운 가중치를 이용하여 언제든지 전역적 최적값에 도달할 수 있다. BLSOM은 모형의 가중치 갱신이 주어진 자료를 학습하기 이전의 모형에 대한 정보인 사전 확률 분포가 자료에 의한 우도 함수와의 결합을 통하여 사후 확률 분포가 되어 한 개의 학습 자료에 의해 학습이 끝난 갱신된 가중치 확률 분포를 이루게 된다. BLSOM 알고리즘은 크게 3단계로 이루어진다. 우선 초기화 단계이다. 승리 노드(winner node)를 결정하기 위한 측도로서 사용되는 개 개의 입력 자료와 가중치간의 유클리디안 거리의 사용을 위하여 입력 자료에 대한 표준화를 수행한다. 또한 이 단계에서 형상지도의 차원과 각 노드에 해당하는 초기 가중치의 분포를 결정한다. 사전에 군집 수를 결정해야 하는 기존의 k-means 방법이나 계층적 군집 방법과는 달리 군집이 형성될 수 있는 형상지도의 차원을 결정하는 방법이 상대적으로 객관적이다. 예를 들어 형상지도의 차원을 5'5로 결정한다면 경쟁적 학습을 거쳐 최종적으로 생성되는 군집 수는 1개에서 25개 까지의 가능성 중에서 가장 최적인 군집 수로 형성된다. 물론 형상 지도의 차원을 크게 하면 최적 군집 수는 증가할 수 있다. 하지만 군집에 대한 정확한 정보가 없이도 객관적인 군집화를 할 수 있는 것이 SOM의 경쟁 학습의 가장 큰 장점이며 BLSOM도 SOM의 이러한 장점을 수용하고 있다. 또한 이 단계에서 형상 지도의 각 노드의 가중치에 대한 사전 확률 분포를 결정한다. 입력 자료가 평균이 0이고 분산이 1인 정규 분포를 따르도록 표준화 되었고 유사성 측도로서 유클리디안 거리를 이용하기 때문에 식 (4.1)의 정규 분포를 이용한다.

$$f(x) \sim N(\mu, \sigma_w^2) \quad (4.1)$$

이 때 모수인 평균, μ 와 분산, σ_w^2 은 적절하게 주어진다. 4절 이후의 실험에서는 평균과 분산을 각각 0과 1로 하였다. 이는 입력 자료가 표준정규분포를 따르도록 표준화 시켰기 때문이다. 물론 각 모수에 대한 하이퍼 파라메타를 고려 할 수 있다. 이 때 하이퍼 파라메타의

역할은 지도 학습(supervised learning) 모형인 다층 신경망(Multi-layer perceptron) 모형에 서는 중요하고, 보통 σ_w^2 는 위치 모수가 α 이고 척도 모수가 β 인 식 (4.2)의 감마 분포를 따르는 분포를 갖을 때 우수한 결과가 나타나는 것으로 알려져 있다(Neal 1996).

$$\sigma_w^{-2} \sim \Gamma(\alpha, \beta), f(\sigma_w^{-2}) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad (4.2)$$

제안하는 BLSOM은 자율 학습(unsupervised learning) 신경망 모형이지만 Neal의 연구 결과를 이용하여 하이퍼 파라메타에 대한 분포를 고려하였으나 실험결과는 차이가 없는 것으로 나타났다. 두 번째 단계에서는 각 입력 자료와 각 가중치에 대한 유클리디안 거리를 계산하여 가장 작은 거리를 갖는 가중치의 노드를 승리 노드로 결정하여 이 노드를 중심으로 가중치의 확률 분포를 베이지안 학습에 의해 갱신한다. 마지막으로 이러한 베이지안 학습을 통한 가중치 확률 분포의 갱신을 주어진 정지 조건을 만족할 때까지 수행한다. 보통 정지 조건은 사전에 반복 학습의 회수를 결정하거나 가중치 갱신 허용 한계를 결정하여 허용 한계 이내에서 가중치 갱신이 일어나면 학습을 멈추게 된다.

/ BLSOM 알고리즘 /

단계1 : 초기화 단계 (n은 자료의 크기)

1.1 입력 벡터의 정규화

$$x_i^{normal} = \left(\frac{x_{i1} - \mu_1}{\sigma_1}, \dots, \frac{x_{ip} - \mu_p}{\sigma_p} \right) = (x_{i1}^{normal}, \dots, x_{ip}^{normal})$$

$$x_i^{normal} \sim N(0, 1), i = 1, \dots, n$$

1.2 형상 지도 각 노드의 가중치에 대한 사전확률 분포 결정

1.2.1 사전확률 분포 $f(\cdot)$ 의 분포 형태 결정 (정규분포, 감마분포, ...)

$$w \sim f(\theta)$$

선택적으로 모수(θ)는 hyper-parameter(ϕ)를 갖을 수 있다. $\theta \sim g(\phi)$

단계2 : 승자 노드의 결정 단계 (m: 형상지도(feature map) 차원(dimension)의 크기)

2.1 주어진 현재의 사전 확률 분포로부터 각 가중치를 표본 추출

2.2 x_i^{normal} 와 w_j 의 거리(Euclidean distance) 계산

$$dist(x_i^{normal}, w_j) = \sqrt{(x_{i1}^{normal} - w_{j1})^2 + \dots + (x_{ip}^{normal} - w_{jp})^2}$$

p : 입력벡터의 차원 $i = 1, \dots, n, j = 1, \dots, m$

2.3 승리 노드(winner node)의 결정

$dist(x, w_k) < dist(x, w_j)$ 을 만족하면 w_k 는 승리노드가 됨

단계3 : 각 가중치 분포의 갱신

3.1 베이스 규칙(Bayes' rule)을 이용하여 승리 노드의 사후확률 분포 결정

3.2 결정된 현재의 사후확률 분포(current posterior)를 새로운 입력 값에 대한 사전 확률 분포(new prior)로 사용

주어진 조건이 (가중치의 갱신 허용 한계 치, 사전에 정한 반복 학습 수) 만족 될 때까지 단계2와 단계3을 반복한다

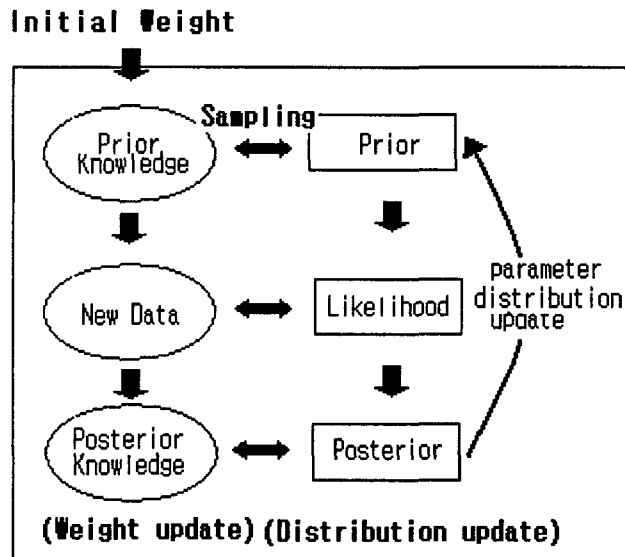


그림 4.1: BLSOM의 가중치 갱신 절차

일반적인 신경망 모형은 주로 주어진 자료에 대한 한 개의 결과값을 구하는데 그 목적이 있지만 제한된 시간 내에 과거의 모든 자료를 이용하여 만족할 만한 모형을 찾아내야 하는 휴리스틱스 모형에는 적합하지 않다. 왜냐하면 과거의 모든 자료를 학습하려면 시간이 너무나 많이 걸리기 때문이다. 따라서 BLSOM의 분포에 의한 갱신이 효과적일수 있다. 또한 하나의 고정된 값으로서 가중치가 갱신되기 때문에 최종적으로 구축된 모형이 최적이지 아닐 경우 바꿀 수 없게 된다. 즉 동일한 입력에 대해 항상 같은 결과만을 제공한다. 하지만 BLSOM은 언제든지 가중치 확률 분포로부터 가중치를 생성하여 새로운 군집 모형을 만들 수 있다. 그림 4.1은 BLSOM과 가중치 갱신 절차를 보이고 있다.

5. 실험 결과 및 해석 I

본 논문에서는 군집분석, 분류분석, 판별분석 등에서 기존에 사용되었던 자료와 인공으로 생성시킨 자료를 대상으로 한 실험을 통하여 SOM과 본 논문에서 제안하는 BLSOM을 비교하였다. 모형의 평가 기준은 사전에 알고 있는 자료의 군집 수와 SOM과 BLSOM의 학습에 의한 군집 수와의 차이를 이용하였다. 실험을 위한 프로그램은 C++언어로 코딩되었고 컴파일러로는 Microsoft Visual C++ 6.0을 사용하였다. 프로그램의 수행은 Microsoft Windows 95 환경에서 펜티엄 II 350MHz CPU를 탑재한 시스템 성능에서 이루어졌다. 주어진 자료를 이용한 실험을 하기 위해서는 앞 장에서 다룬 BLSOM의 사전 확률 분포, 우도 함수 그리고 사후 확률 분포의 구체적인 구조를 결정하여야 한다. 본 논문의 실험은 가우시안 분포를 사용하였다. 이는 많은 자료가 가우시안 분포를 따르고 있고, 베이저안 추론에 대한 적용을 용이하게 하기 위함이다. 따라서 실험을 시작하기 전에 가우시안 분포의 사전 확률 분포, 우도 함수 그리고 사후 확률 분포에 대한 결정을 해야 한다. Kohonen의 SOM은 자료에 대한 자율 학습(unsupervised learning)인 군집화(clustering)에 이용된다.

SOM 모형과 BLSOM 모형의 성능 비교를 위하여 Glass 자료를 이용하여 실험을 하였다. German(1987)의 Glass 자료는 총 214개의 학습 자료로 이루어져 있다. 입력 변수들은 굴절률 변수와 성분을 나타내는 8개의 변수들(나트륨, 마그네슘, 알루미늄, 실리콘, 칼륨, 칼슘, 바륨, 철)로 총 9개로서 이루어져 있다. 1개의 목표 변수는 glass의 형태를 나타내는 변수로서 실제 자료는 6개의 군집을 이루고 있다. Glass 자료에 대한 SOM과 BLSOM의 비교 실험을 위한 확률 분포는 가우시안 분포를 이용하였다. 즉 출력층의 각 노드의 가중치 분포는 식 (5.1)과 같은 가우시안 분포로 결정하였다.

$$f(x) \sim N(\mu, \sigma_w^2) \quad (5.1)$$

이 식에서 평균 μ 는 다시 평균, 0과 분산, 1의 하이퍼 파라메타를 갖는 가우시안 분포를 갖게 하였다. 이는 BLSOM에 사용되는 입력 자료들이 학습을 하기 전에 모두 입력 변수들이 표준화되기 때문에 이 입력 변수들의 값들과의 유클리디안 거리를 구해야 하기 때문에 입력값과 가중치의 척도를 맞추기 위함이다. 그리고 분산, σ_w^2 는 형상 모수가 α 이고 척도 모수가 β 인 식 (5.2)의 감마 분포를 따르는 하이퍼 파라메타를 갖는다.

$$\sigma_w^{-2} \sim \Gamma(\alpha, \beta), f(\sigma_w^{-2}) = \frac{1}{\Gamma(\alpha)\beta^\alpha} (\sigma_w^{-2})^{\alpha-1} e^{-\frac{\sigma_w^{-2}}{\beta}} \quad (5.2)$$

본 실험에서는 α 와 β 를 각각 3과 2로 결정하였다. 이는 임의 값으로 결정하였다. 지도 학습(supervised learning)과 같이 정확한 결과값을 사전에 알고 목적 함수(objective function)을 최소화 하는 경사 하강(gradient descent)방법에서는 하이퍼 파라메타 분포가 예측의 정확도에 영향을 미치지 때문에(Neal, 1996) 제안 모형에 이러한 하이퍼 파라메타의 분포를 고려해 보았으나 본 논문의 자율 학습의 경우에는 크게 영향을 주지 않는 것으로 나타났다. 입력 벡터, x 의 분포는 식 (5.3)의 가우시안 분포를 따른다. 이

식에서 분산, σ_x^2 은 1로서 고정된 값을 사용한다.

$$l(x|w) \sim N(w, \sigma_x^2) \tag{5.3}$$

9개의 입력 변수를 고려하여 형상 지도의 차원을 9'5로 결정하고 실험하였다. SOM을 이용한 학습 결과는 표 5.1의 3개의 군집을 형성하였다. 이것은 실제 6개의 군집과는 차이를 보인다.

표 5.1: 9X5 Maps의 SOM

Cluster	Node	빈도	백분율(%)
1	5	127	59.3
2	23	36	16.8
3	25	51	23.9

BLSOM을 이용한 학습 결과는 표 5.2의 2개의 군집부터 표 5.3의 13개의 군집까지 나타났다. 물론 정확한 6개의 군집을 형성하는 학습결과도 나타남을 볼 수 있다

표 5.2: 9X5 Maps의 BLSOM(2)

Cluster	Node	빈도	백분율(%)
1	0	201	93.9
2	7	13	6.1

반복된 BLSOM을 이용한 군집화 결과는 표 5.4과 표 5.5 그리고 그림 5.1과 그림 5.2에 나타나 있다. 표 5.4과 그림 5.1의 10번의 BLSOM 실험에서는 원래의 군집 수와 같거나 비슷한 결과인 5개, 6개, 7개의 군집 결과가 전체의 50%를 차지하고 있는 것으로 보아 BLSOM이 원래의 군집 수와 비슷한 결과를 비교적 많이 형성하는 것을 실험을 통해 확인할 수 있다. 표 5.5과 그림 5.2의 100번의 BLSOM 실험에서는 원래의 군집의 개수와 같거나 비슷한 결과인 5개, 6개, 7개의 군집 결과가 전체의 55%를 차지하고 있는 것을 볼 수 있다. 이것은 반복 실험의 회수를 증가시킬수록 비교적 원래의 정확한 군집 수와 비슷한 5개, 6개, 7개가 군집결과가 더 많이 나타나고 있음을 알 수 있다. 결국 SOM의 고정된 학습 결과에 비해 BLSOM은 반복 실험의 회수를 증가 시킴으로써 정확한 군집 수에 수렴되는 특성을 가지고 있다.

6. 실험 결과 및 해석 II (재 표본추출 방법을 이용한 모형의 비교)

SOM과 BLSOM을 비교하는 또 다른 방법으로서 재 표본추출에 의해 생성된 표본을 학습 자료로 이용하였다. 재 표본 추출방법에 의해 주어진 자료에서 표본을 추출하여 가중치를

표 5.3: 9X5 Maps의 BLSOM(13)

Cluster	Node	빈도	백분율(%)
1	0	14	6.5
2	5	5	2.3
3	15	17	7.9
4	20	48	22.4
5	25	17	7.9
6	26	12	5.6
7	28	4	1.9
8	30	19	8.9
9	33	41	19.2
10	37	12	5.6
11	40	23	10.7
12	42	1	0.5
13	41	1	0.5

표 5.4: 9X5 Maps의 BLSOM의 10번의 수행결과

군집수	실험 결과수	백분율(%)
4	1	10
5	2	20
6	2	20
7	1	10
8	2	20
10	1	10
11	1	10

결정하는 학습을 반복하여 SOM과 BLSOM의 차이를 비교하였다. 자료는 Fisher의 붓꽃 자료를 사용하였다. 이 자료는 꽃의 종류에 따라 3개의 군집(setosa, versicolor, virginica)을 갖는 150개의 자료로 이루어졌다. 각 군집의 개수는 50개이다. 재

표본추출방법에 의한 SOM의 실험을 위하여 150개의 자료에서 100개의 표본을 단순 임의 추출하여 SOM의 학습을 수행하였다. 100개의 자료에 대해서 SOM을 학습시킨 결과는 표 6.1와 같이 5개의 군집을 형성하고 있다.

동일한 자료를 SOM과 같은 차원의 형상 지도에서 BLSOM으로 학습한 군집화 결과가 표 6.2에 나타나 있다. SOM에 비해서 정확한 군집 수인 3개의 결과가 나타남을 알 수 있다.

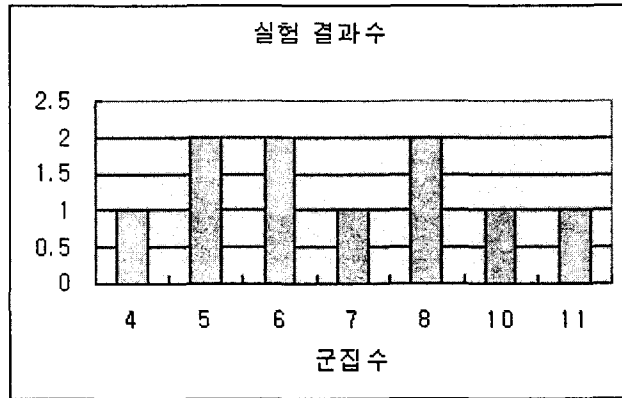


그림 5.1: 9X5 Maps BLSOM의 10번 수행결과

표 5.5: 9X5 Maps의 BLSOM의 100번의 수행결과

군집수	실험 결과수	백분율(%)
2	1	1.0
3	4	4.0
4	6	6.0
5	18	18.0
6	21	21.0
7	16	16.0
8	16	16.0
10	10	10.0
11	5	5.0
12	1	1.0
13	2	2.0

형상지도의 차원을 4X20으로 크게 한 후에 SOM 학습을 수행했을 때의 결과는 표 6.3와 같이 11개의 군집으로 나타났다. 표 6.1와 표 6.3의 결과에서처럼 SOM에서 형상 지도의 차원을 크게 하면 학습결과의 군집 수는 증가하게 된다.

마찬가지로 형상 지도의 차원을 4X20으로 증가시킨 후에 BLSOM을 학습한 결과가 표 6.4에 나타나 있다. 같은 크기의 형상 지도를 갖는 SOM에 비해 정확한 군집의 개수에 근접함을 알 수 있다. 하지만 표 6.2의 결과와 표 6.4의 결과를 비교해 볼 때 BLSOM도 형상지도의 차원이 증가하면 결과 군집의 수가 증가함을 알 수 있다.

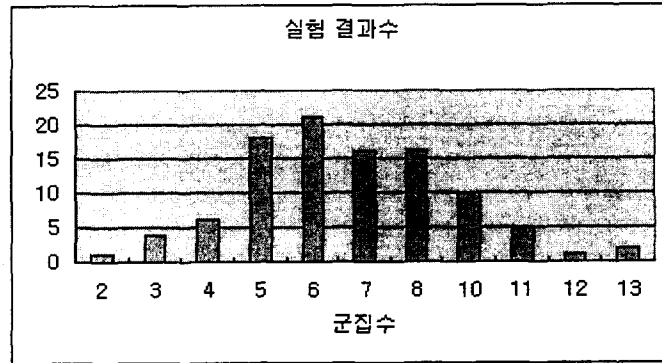


그림 5.2: 9X5 Maps BLSOM의 100번 수행 결과

표 6.1: 4X10 Maps의 SOM

Cluster	Node	빈도	백분율(%)
1	0	12	12.0
2	9	25	25.0
3	11	17	17.0
4	23	4	4.0
5	29	42	42.0

표 6.2: 4X10 Maps의 BLSOM

Cluster	Node	빈도	백분율(%)
1	0	19	19.0
2	8	53	53.0
3	23	28	28.0

총 150개의 붓꽃 자료에서 100개의 표본을 추출하여 SOM과 BLSOM 학습을 수행하는 실험을 100번 수행 한 결과가 표 6.5과 그림 6.1에 나타나 있다.

SOM에 비해서 BLSOM이 정확한 군집의 수에 근접하는 결과가 많이 나타남을 알 수 있다. 동일한 자료에 대해 형상 지도의 차원을 4X20으로 증가시킨 후에 같은 실험을 수행한 결과가 표 6.6과 그림 6.2에 나타나 있다.

역시 SOM에 비해 BLSOM이 정확한 군집 수에 근접하고 있음을 알 수 있다. 그리고 형상 지도의 차원이 커짐에 따라 결과로서 계산되는 군집의 수가 SOM과 BLSOM에서 모두 증가하고 있음을 알 수 있다. 하지만 SOM은 10개 이상의 군집 수를 얻는 경우가 많은

표 6.3: 4X20 Maps의 SOM

Cluster	Node	빈도	백분율(%)
1	0	17	17.0
2	2	13	13.0
3	3	7	7.0
4	9	2	20.0
5	10	1	10.0
6	18	19	19.0
7	22	7	7.0
8	28	6	6.0
9	35	6	6.0
10	36	18	18.0
11	39	4	4.0

표 6.4: 4X20 Maps의 BLSOM

Cluster	Node	빈도	백분율(%)
1	1	14	14.0
2	12	13	13.0
3	22	29	29.0
4	29	21	21.0
5	39	23	23.0

것에 비해 BLSOM은 군집 수가 증가하는 하였지만 역시 정확한 군집 수에 집중되어 있음을 알 수 있다.

7. 결론

BLSOM은 신경망 모형인 Kohonen의 SOM에 직접적으로 베이지안 추론을 적용하여 학습한다. 이것은 기존의 SOM에서는 구현되지 못한 방법이다. 특히 BLSOM을 사용하게 되면 최종적으로 구한 신경망의 가중치가 고정된 값을 갖는 것이 아니라 가중치가 속하게 되는 분포를 갖게 된다. 따라서 동일한 새로운 입력 자료에 대해서도 서로 다른 결과값이 계산되어질 수 있다. 결국 최적의 신경망 모형을 구하지 못해 지역적 최적값에 머물러 있게 되는 문제를 해결할 수 있다. 또한 이론적 전개가 미흡한 신경망 모형에 대해서 결과가 계산되어지는 규칙에 대한 이론적 설명을 가능하게 하는 출발이 된다. SOM에 베이지안 사전 확률 분포의 개념을 도입함으로써 계속적으로 진화하는 신경망 모형에서

표 6.5: 1000개의 표본을 통한 100회의 4X10 Maps 학습결과

Cluster	실험 결과수		백분율(%)	
	SOM	BLSOM	SOM	BLSOM
2	14	20	14.0	20.0
3	15	34	15.0	34.0
4	13	17	13.0	17.0
5	10	16	10.0	16.0
6	13	9	13.0	9.0
7	19	3	19.0	3.0
8	8	1	8.0	1.0
9	5	0	5.0	0.0
10	3	0	3.0	0.0

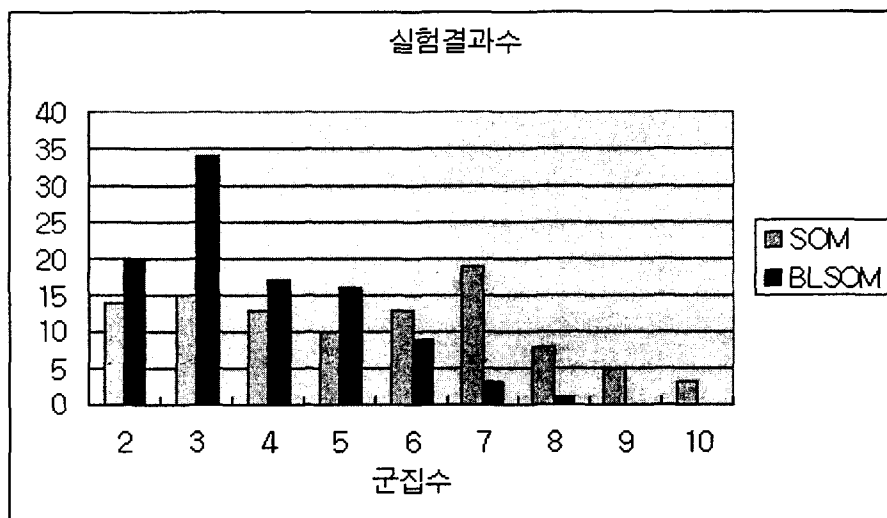


그림 6.1: 100개의 표본을 통한 100회의 4X10 Maps 학습결과

과거의 자료와 현재의 자료를 동시에 학습해야 할 때 과거의 자료는 사전 확률 분포의 정보로만 이용함으로써 신경망 학습의 계산적 부담을 덜어주는 방법으로 사용될 수 있다. 즉 이전 자료는 베이저안의 사전 확률 분포로서 간단하게 모형에 적용되고 현재 자료는 우도 함수로서 학습을 하게 된다. 이러한 사전 확률 분포와 우도 함수를 결합하여 구하여진 사후 확률 분포를 이용하여 새로운 자료에 대한 군집화를 수행하게 된다. 이것은 인간의 뇌 구조를 모형화 하는 연결 주의적 인공지능 모형에서 대규모 데이터의 학습에 필요한 많은 계산 시간에 대한 어려움을 해결할 수 있는 방안이 될 수 있다.

표 6.6: 100개의 표본을 통한 100회의 4X10 Maps 학습결과

Cluster	실험 결과수		백분율(%)	
	SOM	BLSOM	SOM	BLSOM
2	4	9	4.0	9.0
3	9	28	9.0	28.0
4	11	24	11.0	24.0
5	7	19	7.0	19.0
6	13	14	13.0	14.0
7	11	2	11.0	2.0
8	16	3	16.0	3.0
9	10	1	10.0	1.0
10	9	0	9.0	0.0
11	5	0	5.0	0.0
12	5	0	5.0	0.0

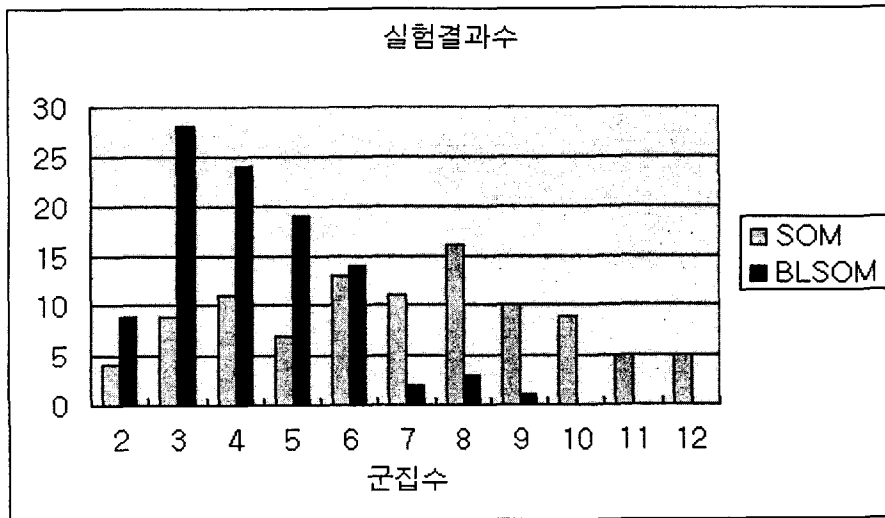


그림 6.2: 100개의 표본을 통한 100회의 4X20 Maps 학습결과

결론적으로 BLSOM은 신경망 모형과 통계적 이론의 수리적 결합을 통해 새로운 확률 신경망 모형을 고려했다. 실험 결과에서 기존의 SOM 모형에 비해서 정확한 학습 결과를 보여줌을 알 수 있다. 앞으로 다층 신경망 등을 포함한 모든 신경망 모형과 유전자 알고리즘, 기타 기계 학습 모형에도 본 논문의 시도가 계속적으로 적용되어 학습 알고리즘의 성능 향상을 기대할 수 있다.

참고문헌

- [1] Bishop , C. M. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press.
- [2] Bishop, C. M., Svensen, M., Williams, C. K. I. (1996). GTM : a Principled Alternative to the Self Organizing Map. *Proceedings 1996 International Conference on Artificial Neural Networks*, ICANN 96, Volume1112. pages 165-170, Bochum, Germany.
- [3] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, Chapman & Hall
- [4] Gerd Gigerenzer (2000), *Adaptive Thinking rationality in the real world*, Oxford University Press.
- [5] Jordan M. I., Xu L., (1995), Convergence results for the EM approach to mixture of expert architectures, *Neural Networks* 8, 1409-1431.
- [6] Kohonen, Teuvo (1997), *Self Organizing Maps*, Springer
- [7] Neal, R. M. (1996), *Bayesian Learning for Neural Networks*, Springer
- [8] Nils J. Nilsson (1998), *Artificial Intelligence : A New Synthesis*, Morgan Kaufmann Publishers, Inc.
- [9] Pandya, A. S. Robert B. Macy (1995), *Pattern Recognition with Neural Networks in C++*, IEEE Press
- [10] Ripley, B. D. (1996), *Pattern Recognition and Neural Networks*, Cambridge University Press
- [11] Tanner, Martin A. (1996), *Tools for Statistical inference*, Springer.
- [12] Utsugi A. (1996), Topological selection for self-organizing maps, *Network: Computational in Neural Systems* vol. 7, No. 4, 727-740.
- [13] Utsugi A. (1997), Hyperparameter selection for self-organizing maps, *Neural Computation*, Vol. 9, No. 3, 623-635.
- [14] Yuille, A. L. (1990), Generalized deformable models, statistical physics and matching problems, *Neural Computation*, 2: 1-24

[2001년 10월 접수, 2002년 5월 채택]

Bayesian Learning for Self Organizing Maps

Sung-Hae Jun ¹⁾ Hongsuk Jorn ²⁾ Jinsoo Hwang ³⁾

ABSTRACT

Self Organizing Maps(SOM) by Kohonen is very fast algorithm in neural networks. But it doesn't show sure rules of training results. In this paper, we introduce to Bayesian Learning for Self Organizing Maps(BLSOM) which combines self organizing maps with Bayesian learning. So it supports explanatory power of models and improves prediction. BLSOM has global optima anywhere but SOM has not. This is proved by experiment in this paper.

Keywords: Self Organizing Map; Bayesian Learning; Clustering; Self Learning Network

-
- 1) Department of Computer Sciences, Seogang University, #1 Sinsudong Mapogu, Seoul, 121-742, Korea
E-mail: shjun@ailab.sogang.ac.kr
 - 2) Department of Statistics, Inha University #253 Yonghyundong Namgu, Incheon, 402-751, Korea
E-mail: hsjorn@anova.inha.ac.kr
 - 3) Department of Statistics, Inha University #253 Yonghyundong Namgu, Incheon, 402-751, Korea
E-mail: jshwang@anova.inha.ac.kr