



EM 알고리즘 및 메타휴리스틱을 통한 다이나믹 환경에서의 베이지안 네트워크 학습 전파 프레임워크

Learning and Propagation Framework of Bayesian Network using Meta-Heuristics and EM algorithm considering Dynamic Environments

추상현* · 이현수*[†]
Sanghyun Choo and Hyunsoo Lee[†]

*국립 금오공과대학교 산업공학부
[†]School of Industrial Engineering, Kumoh National Institute of Technology

요 약

기 구축되어있는 베이지안 네트워크에서 다이나믹한 환경 변화가 발생 할 때, 관련된 베이지안 네트워크의 파라미터는 새롭게 형성된 데이터의 패턴에 적용하여 새로운 파라미터로 변경되어야 한다. 이때, 새로운 파라미터는 베이지안 네트워크의 인과관계를 고려하여 변경되어야 한다. 본 논문에서는 Expectation Maximization(EM)알고리즘과 Meta-Heuristics 기법 중 하나인 Harmony Search(HS)알고리즘을 이용한 다이나믹한 파라미터 업데이트 프레임워크를 제안한다. 일반적으로, EM 알고리즘은 숨겨진 파라미터를 추정하는데 유효한 알고리즘이지만 지역 최적값에 수렴한다는 단점을 가지고 있다. 이 문제를 해결하기 위해서 본 논문은 Maximum Likelihood Estimator(MLE)의 파라미터가 글로벌 최적값을 지향하도록 하기 위하여 메타휴리스틱 방법론의 하나인 HS를 적용한다. 제안된 방법은 EM 알고리즘의 단점을 보완하고 글로벌 최적값에 수렴하는 MLE의 파라미터를 추정하여 다이나믹하게 변화하는 환경에서도 사용 가능한 베이지안 네트워크의 학습 및 전파 프레임워크를 제시한다.

키워드 : 베이지안 네트워크, 다이나믹 파라미터 학습, 최대우도추정, EM 알고리즘, 하모니 서치

Abstract

When dynamics changes occurred in an existing Bayesian Network (BN), the related parameters embedding on the BN have to be updated to new parameters adapting to changed patterns. In this case, these parameters have to be updated with the consideration of the causalities in the BN. This research suggests a framework for updating parameters dynamically using Expectation Maximization (EM) algorithm and Harmony Search (HS) algorithm among several Meta-Heuristics techniques. While EM is an effective algorithm for estimating hidden parameters, it has a limitation that the generated solution converges a local optimum in usual. In order to overcome the limitation, this paper applies HS for tracking the global optimum values of Maximum Likelihood Estimators (MLE) of parameters. The proposed method suggests a learning and propagation framework of BN with dynamic changes for overcoming disadvantages of EM algorithm and converging a global optimum value of MLE of parameters.

Key Words : Bayesian Network, Dynamic Parameter Learning, Maximum Likelihood Estimator, Expectation and Maximization, Harmony Search

Received: Sep. 29, 2016
Revised : Oct. 18, 2016
Accepted: Oct. 19, 2016
[†]Corresponding authors
hsl@kumoh.ac.kr

1. 서 론

베이지안 네트워크(Bayesian Network; BN)[1]은 확률변수에 대한 동시 확률 및 그 구조를 그래프로 표현한다. 베이지안 네트워크는 Directed Acyclic Graph(DAG)와 확률 분포로 구성된다. 베이지안 네트워크의 그래프 안에는 노드와 노드를 연결하는 아크가 있고, 이것은 각 확률 변수와 변수들의 연관관계를 나타낸다. 그리고 베이지안 네트워크에서 노드들 간의 연관관계는 조건부 확률 테이블(Conditional Probability Table; CPT)에 의해 표현된다. 이와 같은 특징으로 베이지안 네트워크는 다이나믹한 환경에 적용가능하며, 이에 관한 많은 연구들[2-6]이 진행되고 있다. Bomin Choi, Jungsik Lee 그리고 Myung-Mook Han[2]은 베이지안 네트워크를 개선하여 IDS 모델의 탐지율을 향상시켰으며, Sun-Jung Yeon, Hye-Jeong Hwang 그리고 Sang-Yong Lee[3]는 베이지안 네트워크를 이용한 일정관리 에이전트를 제안하였다. Woo-Young Chung 과 Eun-Tai Kim[4]은 베이지안 네트워크의 구조학습을 통하여 스마트 홈에서 상황인식을 할 수 있는

이 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 지역신산업선도인력양성사업 성과임(No. 2016H1D5A1908116)
This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

시스템을 제안하였고, Jeong-Sik Hwang, Su-Young Pi, Chang-Sik Son 그리고 Hwan-Mook Chung[5]은 베이지안 네트워크를 구성한 신경망 모델을 통해 소비자의 구매 패턴 예측하였다. Jin-San Yang과 Byoung-Tak Zhang[6]은 전자상거래 고객들의 성향을 베이지안 네트워크를 통해 분석하였다.

베이지안 네트워크의 두 가지의 학습 요소로는 구조 학습과 파라미터 학습이 있다. 베이지안 네트워크에서 데이터에 대한 최적의 모델을 찾아야 하는 이유는 세 가지로 정리 될 수 있다. 첫째, 정확하지 않은 구조는 인과성과 도메인 구조에 대한 잘못된 가정을 내포 할 수 있다. 둘째, 불필요한 아크가 있는 경우에는 정의해야 하는 확률 파라미터의 수가 증가하여 모델의 복잡도가 증가하고 효율성이 떨어지게 된다. 셋째, 필요한 아크가 결여되어 부족한 경우에는 정의 가능한 확률 파라미터의 수가 줄어들게 되고 다른 파라미터를 아무리 정확하게 정의해도 올바른 결과를 얻을 수 없다. 이러한 이유로 데이터에 대한 최적의 모델을 찾는 것이 중요하고 현재까지 많은 알고리즘이 연구되고 있다. 구조 학습 방법 알고리즘으로는 K2 알고리즘[8], Minimum Description Length(MDL)[9]방법 등이 있다.

베이지안 네트워크의 파라미터 학습의 경우 구조를 알고 있고 데이터가 완전하다면 베이지안 네트워크의 구조를 학습하는 것 보다 쉽다. 그 이유는 베이지안 네트워크 구조에 의해 CPT 및 관련된 모든 데이터를 가지고 있어 관측된 데이터의 빈도를 계산한다면, MLE 분포에 대한 최적의 파라미터를 구할 수 있기 때문이다. 하지만 데이터 결여(Missing Data)와 같이 데이터의 불완전성이 내포된 경우 이에 대한 고려가 필요하게 되므로 파라미터 추정 문제는 다소 복잡해진다. 주로 베이지안 네트워크에서 파라미터를 학습하는 방법으로는 EM 알고리즘이 사용되며 이에 관한 많은 연구들[10~15]이 진행되었다.

Liao와 Ji[10]는 베이지안 네트워크에서 EM 알고리즘을 사용하여 불완전한 데이터를 통한 학습 방법을 제시하였고, S.Geman과 D. Geman[11]는 깁스 샘플링과 EM 알고리즘을 사용하여 결측된 데이터에 대한 파라미터를 추정하고 이미지 복원을 하는 방법을 제시하였다. Ramoni와 Sebastiani[12]는 불완전한 데이터 집단에서 조건부 확률 분포의 파라미터를 학습하기 위해서 Robust Bayesian Estimator(RBE) 방법을 사용하여 기존의 EM 알고리즘 방법보다 성능이 좋음을 보였다. G. Elidan과 N. Friedman[13]는 데이터가 임의적으로 결여되었을 때 Information-bottleneck EM(IB-EM) 알고리즘을 적용하여 EM 알고리즘의 단점을 보완하는 파라미터 추정 방법을 제시하였다. B. Thiesson[14]는 Conjugate Gradient Acceleration(CGA) 알고리즘이 적용된 EM 알고리즘을 사용하여 연산속도를 높이는 파라미터 추정 방법을 제안하였다. E. Bauer, D. Koller 그리고

Y.Singer[15]는 EM 알고리즘에 온라인 업데이트 규칙을 적용하여 파라미터 추정하였다. 대부분의 연구에서 베이지안 네트워크의 파라미터 추정은 EM 알고리즘에 의해서 이루어지고 있다.

하지만 EM 알고리즘의 특성상 추정된 파라미터는 지역 최적값에 수렴하며 이를 극복하기 위하여 글로벌 최적값을 구하기 위한 Meta-Heuristics 방법론을 적용할 필요성이 있다.

본 논문에서는 Meta-Heuristics 방법 중 HS 알고리즘을 적용한 파라미터 학습 방법을 제안한다. 제안 기법은 기존의 구조화된 베이지안 네트워크에서 새로운 환경변화로 인하여 파라미터들이 변화될 경우 이를 추정하여, 이때 글로벌 최적값을 지향하기 위한 방법이며, 사용된 주요 프레임워크로는 EM을 통한 파라미터 MLE모델링 및 본 모델의 글로벌 최적값을 구하기 위한 HS기법의 적용으로 구성된다. 이를 테스트하기 위한 Gaussian Mixture Model(GMM)에 기반한 다이나믹 환경변화가 내재화된 베이지안 네트워크를 통해 파라미터를 추정하면서 제안된 기법의 유효성 및 성능을 보여준다.

본 논문의 구성은 다음과 같다. 다음 장에서 배경 및 관련연구에 대해 살펴보고, 3장에서 제안 기법을 소개하며 4장에서 제안기법의 실험결과를 제시하고, 마지막으로 결론 및 향후 연구방향을 제시한다.

2. 배경 및 관련 연구

2.1 Bayesian Network

베이지안 네트워크[16]는 DAG로서 불확실한 상황을 확률 수치로 표현하고, 복잡하고 애매한 인과관계를 미리 정의된 CPT에 의해 네트워크로 이루어진 노드들 간의 확률 관계를 효율적으로 표현 가능한 모델이다.

베이지안 네트워크에서 주어진 증거 집합 E에 대한 추론 결과 I에 대한 확률을 구하면 다음과 같다.

$$P(I|E) = \frac{P(E|I)P(I)}{P(E)} = \frac{P(I,E)}{P(E)} \quad (1)$$

이때 조건부 확률은 체인 규칙(Chain Rule)에 의해 식 (2)와 같이 나타낼 수 있다.

$$\begin{aligned} P(x_1, \dots, x_n) &= P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \dots \\ &= \prod_{i=1}^n p(x_i|x_{i-1}, \dots, x_1) \end{aligned} \quad (2)$$

그림 1의 베이지안 네트워크 모델에 식(2) 체인 규칙을 적용하면, 다음식과 같다.

$$P(A, B, C, D, E) = P(A)P(B)P(C|A, B)P(D|C)P(E|C) \quad (3)$$

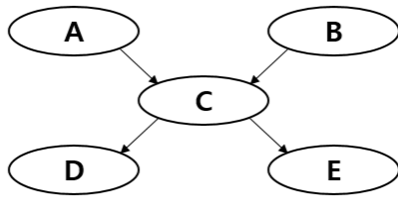


그림 1. 베이지안 네트워크 예시
Fig. 1. An example of a Bayesian Network.

2.2 Expectation Maximization Algorithm

EM 알고리즘[17]은 숨겨진 확률변수 z 를 가지는 확률 모델의 파라미터를 추정하기 위해 사용된다. EM 알고리즘은 두 가지의 단계로 구성되어 있다. 첫 번째, Expectation 단계는 숨겨진 변수 z 의 기댓값을 구하는 단계이고, 그 식은 다음과 같다.

$$\begin{aligned} \gamma(z_{nk}) &= p(z_k = 1 | x_n) = \frac{P(z_k = 1)P(x_n | z_k = 1)}{\sum_{j=1}^K P(z_j = 1)P(x_n | z_j = 1)} \\ &= \frac{\pi_k N(x_n | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j N(x_n | \mu_j, \Sigma_j)} \end{aligned} \quad (4)$$

식 (4)에서 함수 γ 는 할당 확률을 나타내고 이는 해당 데이터의 전체 혼합 확률밀도함수에서의 비율을 나타낸다.

두 번째, Maximization 단계는 Expectation 단계에서 계산된 숨겨진 변수 z 의 값을 사용하여 파라미터를 새롭게 추정하는 단계이고, 식은 다음과 같다.

$$\hat{\mu}_k = \frac{\sum_{n=1}^N \gamma(z_{nk}) x_n}{\sum_{n=1}^N \gamma(z_{nk})} \quad (5)$$

$$\hat{\Sigma}_k = \frac{\sum_{n=1}^N \gamma(z_{nk}) (x_n - \hat{\mu}_k)(x_n - \hat{\mu}_k)^T}{\sum_{n=1}^N \gamma(z_{nk})} \quad (6)$$

$$\hat{\pi}_k = \frac{\sum_{n=1}^N \gamma(z_{nk})}{N} \quad (7)$$

식 (5), (6), (7)에서 각각의 파라미터 $\hat{\mu}_k, \hat{\Sigma}_k, \hat{\pi}_k$ 는 Expectation 단계에서 계산된 할당 확률을 적용한 추정 파라미터들을 가리킨다. 전자의 제시된 식 (4), (5), (6), (7)의 값은 최우추정량 (MLE)에 의해 계산된다.

Maximization 단계에 의해 추정된 파라미터들을 사용하여

다시 숨겨진 변수 z 의 기댓값을 계산하는 Expectation 단계를 수행할 수 있다. 이처럼 Expectation 단계와 Maximization 단계를 반복함으로써 보다 정확한 파라미터를 추정하게 된다.

2.3 Harmony Search

HS는 음악에서 잘 조합된 하모니에 의해 좋은 소리를 들을 수 있듯이, 같은 원리를 진화 알고리즘에 적용한 Meta-Heuristics 알고리즘이다[18].

숙련된 음악가가 즉흥연주를 위한 선택은 3가지가 있다. (1) 유명한 음악의 어떤 한 부분을 기억에 의해 정확히 연주(harmony memory); (2) 유명한 음악의 잘 알려진 부분을 비슷하게 연주(pitch adjusting); (3) 무작위로 구성하여 연주(randomization); 제시된 3가지의 선택이 있고 이를 바탕으로 HS 알고리즘이 구성되어 있다 [19].

그림 2는 HS 알고리즘의 슈도 코드를 나타내고, 식 (9), (10)은 각각 그림 2에서의 Randomization 확률과 PA 확률을 의미한다.

```

Pseudo Code
·Begin
·Objective function f(x)
·Generate initial harmonics (real number arrays)
·Define pitch adjusting rate(rpa), pitch limits and bandwidth
·Define harmony memory accepting rate(raccept)
·while (t < max number of iterations)
·Generate new harmonics by accepting best harmonics
·Adjust pitch to get new harmonics (solutions)
·If (rand > raccept), choose an existing harmonic randomly
·else if (rand > rpa), adjust the pitch randomly within limits
·else generate new harmonics via randomization
·end if
·Accept the new harmonics (solutions) if better
·end while
·Find the current best solutions
·End
  
```

그림 2. 하모니 서치의 슈도 코드
Fig. 2. Pseudo code of the Harmony Search

$$P_{random} = 1 - r_{accept} \quad (9)$$

$$P_{pitch} = r_{accept} * r_{pa} \quad (10)$$

2.4 Gaussian Mixture Model(GMM)

본 연구에서는 베이지안 네트워크의 테스트를 위하여 GMM 기반의 노드로 구성된 모형을 다룬다. K개의 확률밀도함수의 선형 결합으로 GMM[17]의 전체 확률밀도함수는 다음 식과 같이 표현 가능하다.

$$P(x) = \sum_{k=1}^K P(z_k)P(x|z) = \sum_{k=1}^K \pi_k N(x|\mu_k, \Sigma_k) \quad (11)$$

식 (11)에서 $P(x|z)$ 는 혼합 모델의 확률밀도함수를 나타내고, 가우시안 확률밀도함수뿐만 아니라 다른 함수도 사용가능하다. z_k 는 k번째 성분임을 나타내는 확률변수이고, $P(z_k)$ 는 k번째 성분이 전체 혼합 확률밀도함수에서 차지하는 비율을 나타낸다. 또한 식 (11)의 각 변수들은 식 (12)와 같은 성질을 만족한다.

$$z_k \in \{0, 1\}, \sum_k z_k = 1, \quad (12)$$

$$P(z_k = 1) = \pi_k, \sum_{k=1}^K \pi_k = 1, 0 \leq \pi_k \leq 1,$$

$$k = 1, 2, \dots, K(\text{Number of Cluster})$$

GMM의 2차원이상의 확률밀도함수를 구하기 위해서는 식 (13)과 같은 확률밀도함수를 사용하면 된다.

$$N(x | \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{((x-\mu)^T \Sigma^{-1} (x-\mu))}{2}} \quad (13)$$

3. EM 알고리즘 및 Harmony Search 기반 베이지안 네트워크 학습 프레임워크

일반적인 베이지안 네트워크에서 임의의 노드에 기존과 다른 양상의 데이터가 관측되면 새로운 데이터에 적합한 파라미터를 다시 추정해야 한다. 임의의 노드에 새로운 양상의 데이터가 추정되었다는 것은 노드의 특징을 구성하는 확률이 바뀌거나 확률의 파라미터가 바뀌었음을 의미하기 때문이다. 본 연구에서는 확률 모형은 유지하되, 그 파라미터가 바뀌는 경우를 대상으로 한다. 이 경우, 새로운 데이터가 관측이 되었음에도 불구하고 해당 노드의 파라미터를 재 추정하지 않는다면 그 모델에 의해 분석되거나 예측되는 수치들은 잘못된 결과를 도출할 수 있으며, 전체 베이지안 네트워크의 성능 저하를 초래하게 된다.

그림 3은 본 연구에서 제안하는 파라미터 추정 프레임워크의 전체적인 구성을 나타낸다.

3.1 다이나믹 환경하에서 단일노드의 파라미터 변화 추적 및 업데이트

본 연구에서는 각 노드들의 확률분포가 GMM을 따르는 베이지안 네트워크를 그 대상으로 한다. 베이지안 네트워크에서 임의의 노드에 대한 GMM은 다음과 같이 표현할 수 있다.

$$GMM_{Q^*} \sim \sum_{k=1}^K \pi_{Q^*(k)} N(x_{Q^*} | \theta_{Q^*(k)}) \quad (14)$$

식 (14)에서 Q 와 $Q^{(k)}$ 는 베이지안 네트워크에서의 Q번째 노드와 각 GMM에서의 k 번째 군집을 나타낸다.

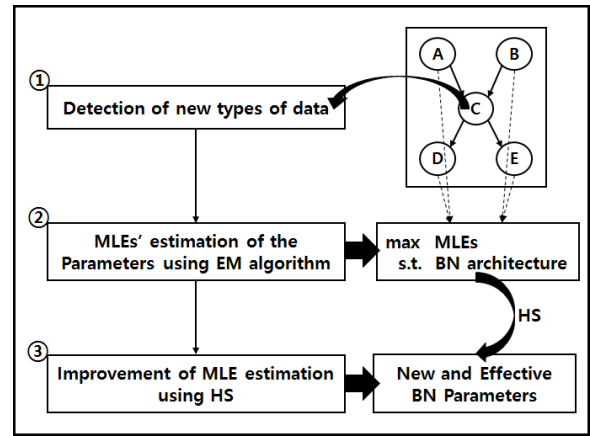


그림 3. 다이나믹 환경 변화에 따른 베이지안 네트워크 파라미터 학습 프레임워크
Fig. 3. Dynamic parameter update in BN

새로운 데이터가 베이지안 네트워크에서 임의의 노드에서 관측되면 해당 노드에 대한 GMM 식은 (15)와 같이 표현할 수 있다.

$$GMM_{Q^*} \sim \sum_{k=1}^K \pi_{Q^*(k)} N(x_{Q^*} | \theta_{Q^*(k)}) \quad (15)$$

식 (15)에서 Q^* 은 새로운 데이터가 관측되어진 노드를 나타낸다. 식 (15)에 의하여 식 (7), (8), (9), (10)은 다음과 같이 전개된다.

$$\gamma(z_{nk})^* \equiv p(z_k = 1 | x_{Q_n}^*) = \frac{\pi_{Q^*(k)} N(x_{Q_n}^* | \mu_{Q^*(k)}, \Sigma_{Q^*(k)})}{\sum_{j=1}^K \pi_{Q^*(j)} N(x_{Q_n}^* | \mu_{Q^*(j)}, \Sigma_{Q^*(j)})} \quad (16)$$

$$\widehat{\mu}_{Q^*(k)} = \frac{\sum_{n=1}^N \gamma(z_{nk})^* x_{Q_n}^*}{\sum_{n=1}^N \gamma(z_{nk})^*} \quad (17)$$

$$\widehat{\Sigma}_{Q^*(k)} = \frac{\sum_{n=1}^N \gamma(z_{nk})^* (x_{Q_n}^* - \widehat{\mu}_{Q^*(k)}) (x_{Q_n}^* - \widehat{\mu}_{Q^*(k)})^T}{\sum_{n=1}^N \gamma(z_{nk})^*} \quad (18)$$

$$\widehat{\pi}_{Q^*(k)} = \frac{\sum_{n=1}^N \gamma(z_{nk})^*}{N} \quad (19)$$

식 (16)은 새로운 패턴의 데이터가 관측되었을 때 해당 데이터를 설명하는 GMM 내의 k번째 가우시안 분포의 해당 확률을 나타내고, 식 (17), (18), (19)는 이를 사용하여 추정된 각 파라미터를 의미한다. 여기서 Q_n 은 특정 노드의 GMM 분포하의

n번째 데이터 셋을 의미한다.

3.2 전체 베이지안 네트워크 모델을 고려한 EM알고리즘 기반 파라미터 MLE 추정 모델

임의의 노드에서 파라미터가 재 추정되면 재 추정된 노드와 연결이 되어있는 모든 노드는 파라미터의 변화에 영향을 받는다. 또한 재 추정된 노드와 연관이 있는 모든 노드들도 베이지안 네트워크의 연결관계에 의하여 파라미터의 변화에 영향을 받을 수 있다. 그러므로 파라미터의 변화에 대한 영향을 고려하여, 이 영향을 가장 잘 표현할 수 있는 파라미터를 추정하는 것은 중요하다. 베이지안 네트워크에서 영향을 가장 잘 표현하기 위해서는 베이지안 네트워크의 구조를 따르면서, 그 파라미터들의 추론식을 최대화 하는 파라미터를 찾아야 한다.

재 추정된 노드와 연결되어 있는 모든 노드에 대한 새로운 최적의 파라미터를 추정하기 위해서 HS 알고리즘을 사용할 수 있고 HS 알고리즘의 목적함수는 다음 식처럼 표현될 수 있다.

$$L(\theta) = \prod_{all\ Q} p(Q|Q_{pa}; \theta) \quad (20)$$

$$\begin{aligned} \hat{\theta} &= \underset{\theta}{\operatorname{argmax}} [\log L(\theta)] \\ &= \underset{\theta}{\operatorname{argmax}} \left[\log \prod_{all\ Q} p(Q|Q_{pa}; \theta) \right] \end{aligned} \quad (21)$$

식 (20)은 베이지안 네트워크의 구조를 고려한 MLE(Maximum Likelihood Estimation)을 계산하기 위한 우도함수이고, 식 (21)은 이를 통해서 추정된 MLE를 의미하며, Q_{pa} 는 Q 노드의 부모노드를 나타낸다. 또한 그림 3과 같은 베이지안 네트워크의 구조일 때 식 (20)과 (21)는 다음과 같이 표현할 수 있다.

$$L(\theta) = P(E|C; \theta_{EC})P(D|C; \theta_{DC})P(C|A, B; \theta_{CA,B})P(A; \theta_A)P(B; \theta_B) \quad (22)$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} [L(\theta)] \quad (23)$$

3.3 HS 알고리즘을 통한 베이지안 네트워크 MLE들의 전역 추정

HS 알고리즘은 다음과 같은 이점[18]을 가진다. 첫째, 복잡한 연산이 없고 초기의 가정이 존재하지 않는다. 둘째, HS 알고리즘의 특성상 HM을 사용하여 적은 공간으로 가능성이 높은 해집합을 많이 확보한다. 이처럼 HS 알고리즘은 여러 Meta-Heuristic 알고리즘 보다 낮은 제약과 빠른 연산 속도로 최종 해결안에 대한 전역 최적값을 계산할 수 있다.

본 연구에서 제안하는 HS 알고리즘은 그림 2에서 제시한 슈도 코드 알고리즘을 기반으로 한다. HS 알고리즘에서 사용되는 목적함수는 식 (21)과 같이 베이지안 네트워크의 확률식에 로그를

적용한 값이며, 파라미터는 이를 최대화 하는 값으로 도출한다. EM 알고리즘을 통해 재 추정된 파라미터 값은 HS 알고리즘에 의해서 지역 극소값을 벗어나게 된다.

4. 실험 및 분석 결과

4.1 실험 모델 및 가정

본 연구에서의 제시한 프레임워크의 검증을 위하여 사용된 베이지안 네트워크 모델은 총 5개의 노드로 이루어져 있고 구성은 그림 4와 같다. 또한 각 노드의 확률분포는 GMM을 따른다.

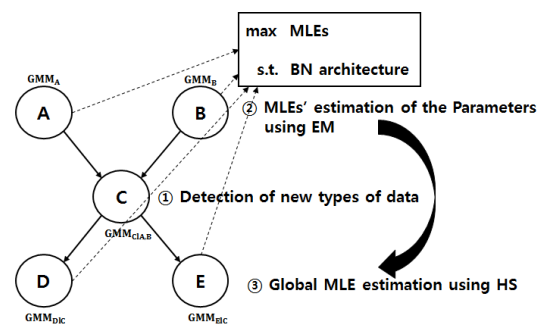


그림 4. 실험을 위한 베이지안 네트워크와 파라미터 업데이트 프레임워크

Fig. 4. Exemplary Bayesian Network and its parameter updating framework

다음은 본 실험을 위한 가정 및 절차를 보여준다. 첫 번째, 구성되어 있는 베이지안 네트워크 모델에서 노드 C에서 새로운 데이터가 관측이 된다고 가정한다. 둘째, 이를 통해, 베이지안 네트워크 추론식에서 각 노드의 확률은 GMM의 식 (24)와 같은 패턴의 확률밀도함수를 따른다.

$$\begin{aligned} P(A) &\sim GMM_A \\ P(B) &\sim GMM_B \\ P(C|A, B) &\sim GMM_{C|A,B} \\ P(D|C) &\sim GMM_{D|C} \\ P(E|C) &\sim GMM_{E|C} \end{aligned} \quad (24)$$

셋째, 새로운 데이터가 관측된 노드 C의 파라미터는 3.1절 및 3.2절에서 기술한 EM 알고리즘에 의해 추정된 후 베이지안 네트워크의 노드 C와 연결된 모든 노드의 파라미터 추정을 요구한다. 제시한 세 가지의 절차에 의해 실험을 실행한다.

4.2 실험 방법 및 결과분석

실험은 제안된 GMM기반의 베이지안 네트워크 모델을 사용하여 실행하였다. 제시된 모델의 노드 C에서는 EM 알고리즘에 의해

새로운 데이터에 적합한 파라미터가 추정된다. 새롭게 추정된 노드 C에 의하여 노드 C와 인과 관계를 맺고 있는 모든 노드 A, B, D, E 또한 파라미터를 재 추정한다. 노드 A, B, D, E는 제안된 모델에서의 HS 알고리즘을 적용하여 전역최적화에 가까운 파라미터를 추정한다.

하지만, 본 연구에서 제시한 알고리즘은 EM 및 HS에 기반한 학습을 다루므로, 반복수에 따른 과적합 문제가 발생할 여지를 남겨둔다.

그림 5는 각 실험모델의 파라미터에 의한 $\log L(\theta)$ 값을 나타낸다. 그림 5(A), 5(B), 5(C), 5(D)는 각 각 학습 반복수가 10, 100, 1000, 10000회에서의 $\log L(\theta)$ 값을 나타내고, 붉은색점이 뜻하는 것은 가장 큰 $\log L(\theta)$ 값을 가지는 모델을 의미한다. 그림 5(A)는 설정된 학습 반복수가 너무 작을 경우에 충분한 학습을 하지 못하여 목적함수에 적합한 파라미터를 추정하지 못함을 알 수 있다. 그림 5(D)는 학습 반복수를 과도하게 설정하여 과적합이 발생하여 실제 모델보다 더 높은 $\log L(\theta)$ 값을 도출되는 오류를 보여준다. 따라서 파라미터 학습을 위하여 충분한 학습을 하고 과적합 되지 않는 학습 반복수를 설정하는 것이 중요하다. 본 연구에서는 그림 5(C)와 같은 반복을 통하여 $\log L(\theta)$ 값이 실제 모델과 유사함을 보여주었다.

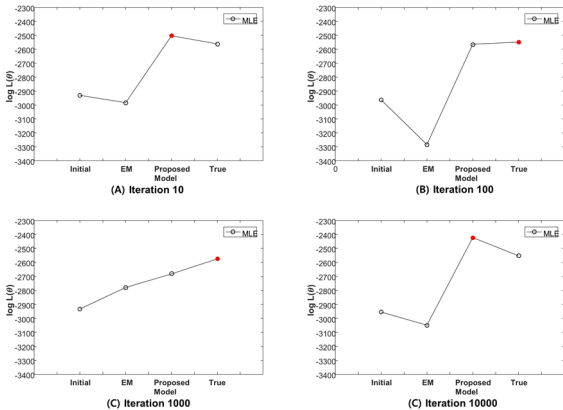


그림 5. 학습반복수에 의한 각 실험모델의 $\log L(\theta)$ 값
Fig. 5. of each Experiment Model by Iterations

제안된 모델의 유효성을 검증하기 위해 본 연구에서 제시한 프레임워크를 적용 (실험 모델 #1; 그림 6-(a))하였다. 이 프레임워크의

유효성을 증명하기 위하여 업데이트 되지 않은 초기 베이지안 네트워크의 파라미터 (실험모델 #2; 그림 6-(b))들과 베이지안 네트워크의 노드 구성을 고려하지 않은 채 각 단일노드에서 EM 알고리즘만을 적용한 파라미터 추정값 (실험모델 #3; 그림 6-(c))을 비교 분석하였다.

그림 6(b)는 노드 B의 초기 확률밀도함수와 다이나믹 환경하에서 업데이트된 후속 확률밀도함수를 나타내며, 두 확률밀도함수의 파라미터들이 전혀 일치하지 않음을 보여준다. 이러한 양상이 베이지안 네트워크의 한 노드에서 관측되었을 경우, 관측된 노드에 의해 영향을 받은 다른 노드들의 파라미터를 또한 학습 또는 재 추정할 필요가 있음을 의미한다.

그림 6(c)는 실험모델 #3인 단일노드 B에서 베이지안 네트워크를 고려하지 않고 EM 알고리즘만을 사용해서 추정한 확률밀도함수와 실제 데이터의 확률밀도함수의 비교를 보여준다. 그림의 결과 두 확률밀도함수가 완전히 일치하지는 않지만 실험모델 #2에서 보여준 초기의 파라미터 보다 좀 더 정확한 파라미터를 추정하였음을 알 수 있다.

그림 6(a)는 노드 B에서 본 연구에서 제안된 모델을 사용하여 추정한 확률밀도함수와 실제 데이터의 확률밀도함수를 나타낸다. 그림 결과 제안된 모델은 초기의 파라미터와 EM 알고리즘을 사용한 파라미터 보다 정확한 파라미터를 추정함을 알 수 있다.

표 1. 각 실험모델의 대응표본 t-검정 결과
Table 1. Result of Paired-Sample T-test in Each Experiment Model

	Experiment Model #1	Experiment Model #2	Experiment Model #3
p-value	0.2318	0.0155	0.1717
t-statistics	1.2030	2.4636	1.3760

실제 모델과 각 모델과의 유사도를 판정하기 위해서 대응표본 t-검정을 5% 유의수준(α)으로 실시하였다. 표 1은 실제 모델과 각 모델의 데이터를 사용한 검정결과인 p-value와 검정통계량을 보여준다. 검정결과 실험모델 #1과 실험모델 #3는 p-value의 값이 유의수준의 값보다 크므로 귀무가설이 채택이 된다. 즉, 실제 모델과 제안된 모델의 파라미터에 의한 데이터는 차이가 없다고 할 수

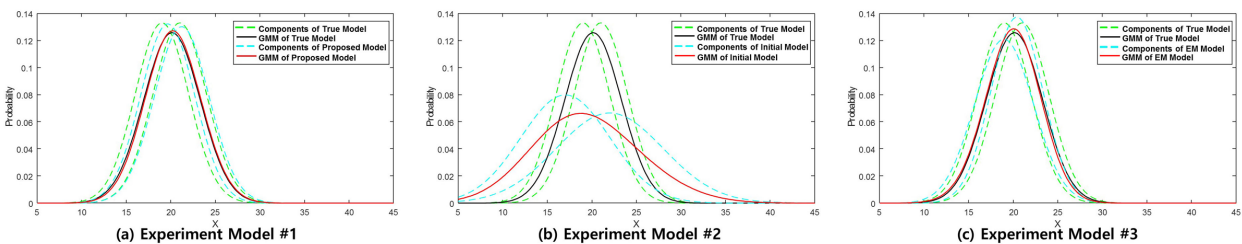


그림 6. 다이나믹한 환경 변화에서 노드 B의 확률 업데이트
Fig. 6. Node B's probability update from the dynamic changes in environments

있고, 실제 모델과 EM 알고리즘만을 사용한 모델 또한 파라미터에 의한 데이터는 차이가 없다고 할 수 있다. 그리고 실험모델 #2의 p-value는 유의수준보다 작은 값을 가지므로 귀무가설이 기각되어, 실제모델과 초기의 파라미터를 가진 모델의 데이터 간의 차이는 있다고 할 수 있다. 그러므로 실험모델 #1과 실험모델 #3은 유의수준 5% 하에서 실제 모델과 유사함을 보여주며, 본 연구가 제시한 실험모델 #1이 보다 우수함을 검증한다.

본 실험에서는 제안된 파라미터 학습 기법으로 글로벌 최적값에 가까운 파라미터를 추정하고, 이를 실제 모델과의 비교 그림 및 대응표본 t 검정을 통해 검증하였다.

5. 결론 및 향후 발전방향

본 연구에서는 기존의 구조화 된 베이지안 네트워크에서 새로운 패턴의 데이터가 관측되었을 때 해당 노드에서 전체 베이지안 네트워크 모델을 고려한 EM 알고리즘 기반의 MLE의 파라미터를 추정하고, 이를 다시, HS 알고리즘에 기반하여 글로벌 최적값에 가까운 파라미터를 추정하는 프레임워크를 제시하였다. 본 연구에서 제안하는 방법은 다음과 같은 이점이 있다. 첫째, 구축된 베이지안 네트워크를 사용하는 응용 모델에서 새로운 패턴의 데이터의 관측으로 야기되는 전체 베이지안 네트워크의 모수 추정에 적용될 수 있다. 즉, 제안된 모델은 베이지안 네트워크의 구조를 유지하며, 그 구성 노드의 확률분포에서 업데이트되는 파라미터를 추정할 수 있고 해당 노드와 연관되어 있는 타 노드 또한 변화된 환경에 적용된 파라미터를 재 추정할 수 있다. 둘째, Meta-Heuristics 알고리즘을 사용하여 지역 최적값이 아닌 글로벌 최적값을 지향하는 파라미터를 추정할 수 있다. 제시된 방법은 베이지안 네트워크 모델에서 노드들 간의 관계를 가장 잘 표현할 수 있는 모수를 탐색하며, HS 알고리즘을 통하여 보다 정확한 파라미터를 추정하는 효과를 가진다.

향후 연구 분야로는 제안한 방법을 기 구축된 다양한 베이지안 네트워크에 적용하여 다이나믹한 베이지안 네트워크를 학습하고 응용하고자 한다. 또한, 베이지안 네트워크에서 파라미터 추정을 위해 다양한 Meta-Heuristics 알고리즘을 적용하고 비교 분석하여 파라미터 추정 효율을 높이고자 한다.

References

[1] Jiang Su, Harry Zhang, Charles X. Ling, Stan Matwin, "Discriminative Parameter Learning for Bayesian Networks," *Proceedings of the 25th International Conference*

on Machine Learning, pp. 1016-1023, 2008

[2] Bomin Choi, Jungsik Lee, Myung-Mook Han, "IDS Model using Improved Bayesian Network to improve the Intrusion Detection Rate", *Journal of Korean Institute of Intelligent Systems*, vol. 24, no.5, pp. 495-503, 2014

[3] Sun-Jung Yeon, Hye-Jeong Hwang, Sang-Yong Lee, "Scheduling Management Agent using Bayesian Network based on Location Awareness", *Journal of Korean Institute of Intelligent Systems*, vol. 21, no.6, pp. 712-717, 2011

[4] Woo-Young Chung, Eun-Tai Kim, "Context-aware application for smart home based on Bayesian network", *Journal of Korean Institute of Intelligent Systems*, vol. 17, no.2, pp. 179-184, 2007

[5] Jeong-Sik Hwang, Su-Young Pi, Chang-Sik Son, Hwan-Mook Chung, "A Purchase Pattern Analysis Using Bayesian Network and Neural Network", *Journal of Korean Institute of Intelligent Systems*, vol. 15, no. 3, pp. 306-311, 2005

[6] Jin-San Yang, Byoung-Tak Zhang, "Analysis of Web Customers Using Bayesian Belief Networks", *Journal of Korean Institute of Intelligent Systems*, vol. 11, no. 1, pp. 16-21, 2001

[7] Geum-Seong Hwang, Sung-Bae Cho, "Learning of Bayesian Network," *Korea Robotics Society*, vol. 3, no. 4, pp. 15-27, 2006

[8] G.F. Cooper, E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, no. 4, pp. 309-348, 1992.

[9] J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 465-471, 1984.

[10] Wenhui Liao, Qiang Ji, "Learning Bayesian network parameters under incomplete data with domain knowledge," *Pattern Recognition*, vol. 42, no.11, pp. 3046-3056, 2009

[11] S. Geman, D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6 pp.721-741, 1984

[12] M. Ramoni, P. Sebastiani, "Robust learning with missing data," *Machine Learning*, vol. 45, no. 2, pp. 147-170, 2001

- [13] G. Elidan, N. Friedman, "The Information Bottleneck EM Algorithm," *In Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence*, pp. 200-209, 2003
- [14] B. Thiesson, "Accelerated quantification of Bayesian networks with incomplete data," *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, pp. 306-311, 1995
- [15] E. Bauer, D. Koller, Y. Singer, "Update rules for parameter estimation in Bayesian networks," *In Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, pp. 3-13, 1997
- [16] Richard E. Neapolitan, *Learning Bayesian Networks*, Prentice Hall, 2003
- [17] Il-Chul Moon, *K-Means Clustering and Gaussian Mixture Model*, Lecture Note in Korea Advanced Institute of Science and Technology, 2010
- [18] Z. W. Geem, J. H. Kim, G. V. Loganathan, "A new heuristic optimization algorithm: harmony search," *Simulation*, vol. 76, pp.60-68, 2001
- [19] X.S. Yang, "Harmony Search as a Metaheuristic Algorithm," *Music-Inspired Harmony Search Algorithm: Theory and Applications* (Editor Z.W. Geem), *Studies in Computational Intelligence, Springer*, vol. 191, pp. 1-14, 2009

저자 소개



추상현(Sanghyun Choo)

2016년~현재 : 금오공과대학교 산업공학과 석사과정

2016년 : 금오공과대학교 산업공학부 공학사

관심분야 : Data Mining, Machine Learning, Deep Learning, Big Data

Phone : +82-54-478-7681

E-mail : sjksh@kumoh.ac.kr



이현수(Hyunsoo Lee)

2011년~현재 : 금오공과대학교 산업공학부부교수

2006년 : Texas A&M University

산업시스템공학과 박사

2002년 : POSTECH 산업공학과 공학 석사

관심분야 : Nonlinear Control and Optimization, Virtual Intelligence, Data Engineering

Phone : +82-54-478-7661

E-mail : hsl@kumoh.ac.kr