

비모수적 베이지안 모델을 이용한 서브패턴 추출기법 설계

허민오⁰ 이상우 장병탁
서울대학교 컴퓨터공학부¹
{moheo, slee, btzhang}@bi.snu.ac.kr

A Subpattern Extraction Method Design based on Nonparametric Bayesian Modeling

Min-Oh Heo⁰ Sang-Woo Lee Byoung-Tak Zhang
School of Computer Science & Engineering, Seoul National University

요 약

데이터를 구성하는 중요한 패턴들을 찾고자 하는 연구는 데이터의 특성을 파악하려는 작업에서 가장 중요한 부분 중 하나이다. 이러한 방법론들에서 주요 요소들의 개수는 미리 알기 어려운 요소이며, 주요 요소들의 곱 형태로 모분포를 표현하려는 모델인 Product of Experts (PoE)에서도 마찬가지이다. 데이터의 일부만을 설명하는 서브 패턴은 일부 feature에 대한 규칙만을 다루는 것으로, 좀더 유연한 모델링이 가능하며, 이산변수의 경우 규칙으로 변환이 가능하다. 본 고에서는 비모수적 베이지안 모델링을 통해 PoE에서 자동적으로 서브패턴을 찾는 기법을 제안한다. 또한, 제안 방법에 대한 실험설계의 예로서 PoE의 하나인 Restricted Boltzmann Machine을 학습한 결과와 분석을 보였다.

1. 서 론

데이터를 이루는 주요 요소를 찾고자 하는 연구는 노이즈 제거, 데이터 압축, 데이터 모델링과 같은 데이터 분석 및 활용에 있어 필수적인 요소이자 절차이다. 특히, 주성분분석(PCA), Gaussian Mixture와 같이 데이터들이 주요 요소들의 weighted sum으로 표현되는 것을 가정한 방법이 널리 사용되고 있다[1]. 이는 Mixture of Experts로 볼 수 있으며, 이와 대조적인 모델로서 Product of Experts (PoE)가 존재한다. 이들은 각각 요소 간의 OR 연산 기반 확률모델, 요소 간의 AND 연산 기반 확률모델로 볼 수 있으며, PoE는 각 Expert보다 더 날카로운 분포도 표현할 수 있게 된다[2].

실제 문제에서는 사전에 데이터를 구성하는 요소의 수를 알 수 없는 경우도 빈번하다. 비모수적 베이지안 모델(Nonparametric Bayesian (NPB) Model)은 latent factor의 개수를 무한히 늘릴 수 있도록 확장된 방법으로, 은닉 변수의 수를 정하기 어려운 경우에 흔히 고려되는 방법이다. 지금까지 군집화[3], 문서 내 주제 추출 [4], 은닉 마코프 모델[5], 최근 역강화학습[6]에까지 여러 문제에 적용되었다.

본 고에서는 데이터를 구성하는 factor 집합을 추출하는 새로운 방법을 제안한다. 이 방법에서는 데이터 생성에 대한 다음과 같은 가정을 부여한다. 즉, 각 데이터는 데이터 일부를 나타내는 규칙들과 그에 대한 적절한 가중치로부터 각각 확률적으로 생성되며, 이 규칙들은 가능하면 확률적으로 AND 조건이 되는

것을 선호한다. 또한, Hypernetwork 모델[7]에서 언급된 ‘매치되는 규칙을 차례로 조합하는 과정을 통한 데이터 생성과정’에 필요한 조각을 찾는 문제와 깊이 연관된다.

규칙의 수는 사전에 알기 어려우므로 NPB 방법을 이용하는 것이 가능하고, 규칙을 제안하는 NPB 방법으로 Indian Buffet Process(IBP)[8]를 사용할 수 있다. 이 규칙 간에 AND적 특성을 부여하기 위해 관측 모델로서 Product of Experts 모델을 사용하고자 한다.

이후, 본 논문은 다음과 같이 구성된다. 2 장에서는 데이터에 대한 확률적 규칙(서브패턴) 추출문제 및 추출 방법을 소개하고, 3 장에서는 제안된 방법에 대한 기초 모델로서 RBM 학습 결과를 보이고 실제 수행할 실험방법을 소개한 후, 4 장에서 결론을 맺는다.

2. 제안 방법

2.1 규칙 추출을 위한 비모수적 베이지안 모델

[6]에서 역강화학습을 위한 NPB를 이용한 feature construction 방법이 소개되었다. 이 방법은 데이터로부터 강화학습에서 쓰이는 보상함수를 학습하기 위해 설계되었으며, 본 고에서는 이 방법을 기초로 하여 PoE에 적용할 수 있도록 모델을 그림 1과 같이 설계하였다.

규칙 생성부분으로서, F개의 feature로 이루어진 N개의 데이터에 대하여 K개의 규칙 r이 가중치 w에 의해 생성되었다고 가정한다. u는 각 instance 생성에 규칙들이 참여하였는지 여부를 나타내는 랜덤변수이다. 또한, v는

각 feature가 가질 수 있는 값을 나타내는 Dirichlet-categorical 랜덤 변수이며, z 는 각 feature가 규칙 r 에 참여하는지를 나타내는 이진 변수 행렬이며 IBP를 따라 생성된다. 여기서, K 는 데이터의 특성에 따라 무한히 늘어날 수 있다.

생성된 규칙으로부터 데이터를 표현하는 부분으로서, 다음과 같이 PoE 모델을 따른다. 그림 1에서 관측변수는 회색으로 표기된 변수인 x 이며, 1 또는 -1을 갖는 이진 변수로 정의한다. x 와 관련된 변수들 간의 한계확률분포는 아래와 같다.

$$P(x_{nf}, r, w, u_n) = \frac{1}{Z} \exp\left(\sum_k r_{fk} \cdot w_k \cdot u_{kn} \cdot x_{nf}\right)$$

상기 식에서 Z 는 모든 경우에 대해 더해서 얻는 normalization 상수(또는 partition function라고도 함)이다. 또한, x 에 대한 조건부 확률은 아래와 같이 유도된다.

$$P(x_{nf} = 1 | r, w, u_n) = \frac{1}{1 + \exp(-2 \cdot \sum_k r_{fk} \cdot w_k \cdot u_{kn})}$$

또한, v , z 와 r 과의 관계는 결정적으로 정의된다. 즉, z 에서 선택된 변수에 v 에서 샘플링 된 값을 택하여 r 이 결정된다. 즉, r 과 z 는 동일한 자료구조를 가지되, z 가 1인 원소에 대하여 해당 v 값을 가져와서 완성한다.

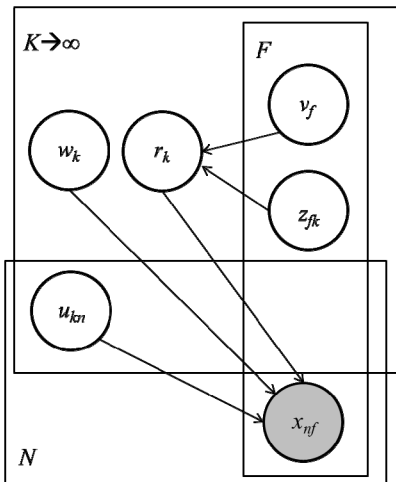


그림 1 subpattern 추출을 위한 비모수적 베이زي안 모델

2.3 추론 방법

NPB 모델은 일반적인 경우에 은닉변수에 대한 Posterior 분포를 구하는 분석적 방법이 존재하지 않으므로, 근사적 방법을 주로 이용한다. 흔히 사용되는 방법은 다수의 샘플을 통해 근사적으로 원하는 분포를 얻는 Markov Chain Monte Carlo (MCMC) 방법이다. 실제적으로는 MCMC의 한 방법이며 일반적으로 적용하기 용이한 Gibbs Sampling이 주로 첫 번째 시도로서 흔히 테스트 되며, Metropolis-Hastings algorithm도 자주 사용된다.

Gibbs sampling은 모든 변수에 대한 샘플링을 한 변수마다 차례로 진행하되, 샘플링 하려는 변수 외의 나머지 모든 변수는 조건으로 둔다. 조건에 랜덤변수가 포함되기 위해서는 값이 주어져야 하는데, 가장 최근에 샘플링 한 값을 이용한다.

데이터로부터 제안모델의 r , w 에 대한 사후 확률을 얻기 위해 [6]에서 제안된 방법에 따라 다음 단계를 순차적으로 수행하여 수렴할 때까지 반복한다.

- 1) v , u , z 의 샘플링에 Gibbs sampling을 이용한다.
- 2) 새로운 규칙 $r^+ = \langle K^+, Z^+, V^+, U^+, w^+ \rangle$ 을

위해 다음과 같이 IBP를 이용하여 샘플링을 수행한다. 즉, K^+ 는 포아송 분포를 따라 샘플링 한다. Z^+ 는 $f \times K^+$ 이진 행렬이 되는데, K^+ 열에 대해 f 번째 원소는 1이고 나머지는 0으로 설정한다. V^+ 와 w^+ 는 해당 사전확률로부터 샘플링 한다. 이와 같이 얻은 r^+ 를 다음의

acceptance ratio ρ_z 를 이용해 선택한다. v 와 z 에 의해 r 이 결정되므로 아래 식에서는 r 대신 U , Z 로 표기하였다.

$$\rho_z = \frac{P(X | [W, w^+], [U, U^+], [Z, Z^+], [V, V^+])}{P(X | W, U, Z, V)}$$

- 3) w 와 새로운 z 를 추가함에 있어서는 Metropolis-Hastings update를 이용한다. w 의 update는 다음의 acceptance ratio ρ_w 를 이용한다.

$$\rho_w = \frac{P(X | W^{new}, U, Z, v)P(w^{sampled})}{P(X | W, U, Z, v)P(w_{kf})}$$

3. 실험 결과

최근 대두되고 있는 대표적인 PoE 모델은 Restricted Boltzmann Machine (RBM)이다. RBM은 Deep Learning의 기반 요소 모델로 자주 이용되고 있으며, 다양한 변종이 존재한다[9]. 또한, Hypernetwork과 표현 상에 큰 연관관계가 있는 것이 [10]에서 보여졌으며, RBM의 표현을 일부 바꾸어 규칙과 가중치가 부여된 형식을 제안하였다. 상기 제안 방법에 대한 실험 소개에 앞서, 제안 방법보다 단순한 모델인 RBM을 학습한 결과를 먼저 보이도록 한다.

3.1 Gibbs sampling을 이용한 Restricted Boltzmann Machine 학습

RBM을 MCMC를 통해 학습한 결과를 예비결과로서 제공한다. 그림 2와 같이 RBM과 동일한 표현력을 갖도록 모델을 정의하고 h 를 위해 Gibbs sampling을 이용하고 w 를 위해 Metropolis-Hastings update를 이용하였다. K 는 20으로 설정하였다. 본 실험을 위해

UCI repository에 있는 optical hand-written digit data를 이용하였다. 이 데이터 집합은 손으로 적힌 숫자 데이터이며, training data가 3823개, test data가 1797개이다. 각 instance는 8 x 8 형태의 -1 또는 1의 값으로 이루어져 있다. 그림 3은 첫 50개의 instance를 보이고 있다.

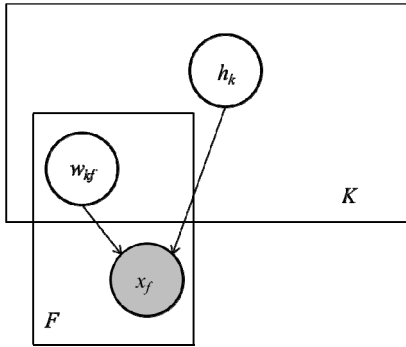


그림 2. RBM과 동일한 표현력을 갖는 모델



그림 3 학습 데이터 예. 8 x 8 크기의 이진 이미지

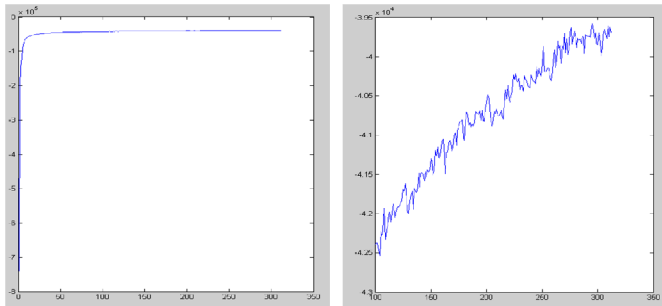


그림 4 (좌): 모델이 주어졌을 때 데이터 집합에 대한 확률. 샘플링 횟수가 늘어남에 따라 증가한다. (우): 100 step 이후에 대한 동일 그래프

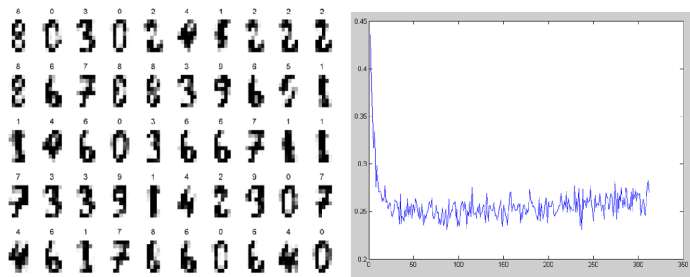


그림 5 (좌): 모델에 의해 생성된 숫자의 50개의 예, (우): 재생성 에러율 추이

3.2 제안 방법에 대한 실험 설계

3.1에서 사용한 데이터를 그대로 적용하고, 이미 시행한 유사한 분석방법들 외에도 각 데이터를 이루는 요소들과 가중치를 실제 패턴과 비교하여 확인할 수 있다.

4. 결론

본 고에서는 데이터를 나타내는 서브패턴을 추출하는 방법을 제시하였다. [7]에 소개된 방법이 제안된 방법에 대한 빠른 알고리즘으로 작동하도록 하는 기반이 될 수 있을 것을 기대한다. 이는 [7]에서 제안된 ‘데이터 조각으로부터 패치를 무작위로 추출하여 규칙 후보로서 사용하도록 하는 아이디어’를 제안 방법에서의 규칙의 좋은 후보로서 사용할 수 있을 것으로 생각되기 때문이다.

감사의 글

이 논문은 정부(산업통상자원부)의 재원으로 한국산업기술평가관리원 지원을 받아 수행된 연구이며(10035348-mLife, 10044009), 정부 (미래창조과학부)의 재원으로 한국연구재단의 지원(2010-0017734, Videome)을 일부 받았음.

참고문헌

- [1] Simon Haykin, *Neural Networks and Learning Machine*, Prentice Hall, 2009.
- [2] G. E. Hinton, Training Products of Experts by Minimizing Contrastive Divergence, *Neural Computation* vol.14, no.8 pp. 1771-1800, 2002.
- [3] S. Richardson and P. J. Green, On Bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society. Series B (Methodological)*, 59:731-792, 1997.
- [4] A. Ahmed, Q. Ho, C. H. Teo, J. Eisenstein, A. J. Smola, and E. P. Xing, Online Inference for the Infinite Topic-Cluster Model: Storylines from Streaming Text, *AISTAT 2013*, 2011.
- [5] M. J. Beal, The Infinite Hidden Markov Model, *NIPS*, pp.140-150, 2001.
- [6] J. Choi, K.-E. Kim, Bayesian Nonparametric Feature Construction for Inverse Reinforcement Learning, *IJCAI 2013*, 2013.
- [7] B.-T. Zhang, Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory, *IEEE Computational Intelligence Magazine*, vol.3, no.3, pp.49-63, 2008.
- [8] Z. Ghahramani, T. Griffiths, and P. Sollich, Bayesian Nonparametric Latent Feature Models, *Bayesian Statistics*, vol. 8, pp.1-25, 2007.
- [9] Y. Bengio, Learning Deep Architectures for AI, *Foundations & Trends in Machine Learning*, vol.2, no.1, pp.1-127, 2009.
- [10] 김권일, 허민오, 이상우, 장병탁, 실수값 인자 데이터의 비지도 학습을 위한 에너지 기반 하이퍼네트워크 모델, 2012 한국컴퓨터종합학술대회(KCC2012)논문집, 제39권 1(B), pp. 480-482, 2012.06.