

Reversible Jump MCMC

1, 1,2

1

2

{syha, btzhang}@scai.snu.ac.kr

Data Mining Using Reversible Jump MCMC and Bayesian Network Learning

Sun-Young Ha¹, Byoung-Tak Zhang^{1,2}

¹Interdisciplinary Program in Cognitive Science

²School of Computer Engineering and Science
Seoul National University

가

(Bayesian network classifier) 가

RBF
Monte Carlo
BN Augmented Naïve-Bayes Classifier

Reversible Jump Markov Chain
Selective

1.

(datamining)

가 [1][2]. 가

2. Selective Augmented Naïve-Bayes classifier

2.1 Bayesian Network Classifier

(Bayesian Network Classifier)

가 가
Naïve-Bayes, Tree Augmented Naïve-Bayes Classifier(TAN)[11], BN
Augmented Naïve-Bayes Classifier(BAN)[12] [12]

BN Augmented Naïve-Bayes Classifier(BAN)

$O(N^2)$

가 BAN Classifier

2.2 BN Augmented Naïve-Bayes Classifier(BAN)

[12] BAN CBL

CBL algorithm

mutual information test($O(N^2)$ 가)

Drafting, Thickening Thining

Drafting

information

mutual

mutual information (1)

(Bayesian networks)

가

[3].
Directed Acyclic Graph(DAG)

가 [4].

Selective Augmented Naïve-Bayes Classifier

¹ Chow-Liu algorithm [13]

conditional mutual

information $I(X_i, X_j | C)$

X_i, X_j C d-separation

가

conditional mutual information

[12][13]

$$I(X_i, X_j | C) = \sum_{x_i, x_j} P(x_i, x_j, c) \frac{P(x_i, x_j | c)}{P(x_i | c)P(x_j | c)} \quad (1)$$

Thickening edge 가 , Thinning edge d-separate 가
 missing value 가 (target value)

2.3 Reversible Jump MCMC

MCMC (sampling based) 가 model selection
 feature selection 가 [10]
 Reversible Jump MCMC 가 Metropolis-Hastings
 Markov Chain sampling [10]
 (birth move), (death move), (height move),
 (position move) 가
 $(k = \dots) b_k + d_k + h_k + p_k = 1$
 (proposal distribution) $q(u)$

$$b_k + d_k, \dots, b_k + d_k + p_k \quad (2)$$

$$b_k = \min \left\{ 1, \frac{p(y|A')}{p(y|A)} \frac{p(k+1)}{p(k)} \frac{q(u')}{q(u)} \right\} \quad (2)$$

k 가 h
 k s

$$(2) \quad \frac{1}{b_k} \quad h$$

$$(3) \quad \frac{1}{b_k} \quad h$$

$$h_k = \min \left[1, \frac{p(y|A')}{p(y|A)} \times \frac{h'^{\alpha}}{h^{\alpha}} \times \exp\{-\beta(h' - h)\} \right] \quad (3)$$

$$p_k = \min \left[1, \frac{p(y|A')}{p(y|A)} \times \frac{(s_{j+1} - s'_j)(s'_j - s_{j-1})}{(s_{j+1} - s_j)(s_j - s_{j-1})} \right] \quad (4)$$

² (uniform distribution)
³ Jacobian

2.4 Selective BAN

Selective BAN

2.4.1 BAN

Reversible Jump MCMC 2.2

2.4.2 Reversible Jump MCMC

$$p(k) = 1 / \frac{k!}{(k - k_{\max})! k_{\max}!} c \frac{\lambda^k}{k!} \quad (3) \quad \text{truncated Poisson} \quad (5)$$

k_{\max} 가 ,
 2.3 가 ,
 가

3.

3.1

KDD2000 Competition
 Question 3 Legcare
 (,
 12
 (heavy spender) 가
 465
 1782 80%

1420 missing value

3.2

Selective BAN $\lambda = 3, k_{\max} = 30, \alpha = 1, \beta = 200, c = 0.25$
 4000 burn-in 40000
 Selective BAN
 Decision Tree,
 Neural Network, Bayesian Neural Network with ARD
 가 KDD 2000

Competition Competition 가

4.

4.1

Selective BAN 1
 Bayesian Neural Network ARD
 ARD feature subset selection
 KDD
 가 가
 가

1. 12

(%)

| Selective BAN | Decision Tree | Neural Network | Bayesian Neural Network with ARD |
|---------------|---------------|----------------|----------------------------------|
| 84.20 | 82.50 | 82.08 | 84.33 |

2

Selective BAN

KDD

KDD

2.

| | Selective BAN | Decision Tree | Bayesian Neural Network with ARD |
|------------------|---------------|---------------|----------------------------------|
| Friend Promotion | ○ | ○ | ○ |
| Time effect | ○ | × | × |
| My coupon | ○ | × | ○ |
| Send e-mail | ○ | ○ | ○ |
| Buy for others | ○ | ○ | ○ |
| Product view | ○ | ○ | ○ |
| Work dress | ○ | × | × |
| Geography | ○ | × | × |
| Causal socks | ○ | × | × |
| Leg care | × | × | × |
| Year of date | × | × | × |
| Pantyhose | × | × | × |

2 Selective BAN
BAN

1. Friend coupon

Family

heavy spender

2. Northeast U.S.

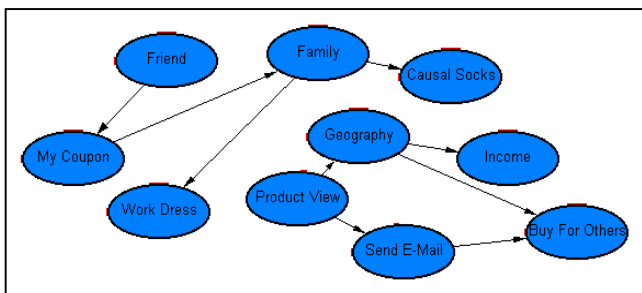
heavy spender

3. Product view

Northeast U.S.

e-mail heavy spender

My



2. Selective BAN

()

4.2

KDD

30

가

30

Selective BAN

Selective BAN

가

missing value

가

가

가

가

가

[6]

(CI-98-006800),
(1999-001-E01025),

(AITRC)

- [1] Berry, M. and Linoff, G., *Data Mining Techniques*, Wiley, New York, 1997.
- [2] Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., *Classification and Regression Trees*, Chapman & Hall, New York, 1984.
- [3] Heckerman, D., "Bayesian networks for knowledge discovery," In *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996.
- [4] Heckerman, D., Meek, C., and Cooper, G., "A Bayesian Approach to Causal Discovery," *Technical Report MSR-TR-97-05*, Microsoft Research, February, 1997.
- [5] Bishop, C., *Neural Networks for Pattern Recognition*, Oxford, 1995.
- [6] Koller, D. and Sahami, M., "Hierarchically classifying documents using very few words," *Proceedings of the 14th International Conference on Machine Learning (ICML)*, Morgan Kaufmann, pages 170-178, 1997.
- [7] D. Koller and M. Sahami, "Toward optimal feature selection," *Proceedings of the 13th International Conference on Machine Learning (ICML)*, Morgan Kaufmann, pages 284-292, 1996.
- [8] Yang, J. and Honavar, V., "Feature subset selection using a genetic algorithm," In: *Feature Extraction, Construction and Selection - A Data Mining Perspective*. Motoda and Liu (eds.) Kluwer Academic Publishers, Chapter 8:117-136, 1998.
- [9] Sykacek, P., "On input selection with reversible jump Markov chain Monte Carlo sampling," In *Advances in Neural Information Processing Systems 12*, S.A. Solla, T.K. Leen and K.-R. Mueller (eds.), MIT Press, pages 638-644, 2000.
- [10] Green, P., "Reversible Jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika* 82, 711-732. 1995.
- [11] Friedman, N., Geiger, D., and Goldszmidt, M., "Bayesian Network Classifiers," *Machine Learning* 29:131-163, 1997.
- [12] Cheng, J. and Greiner, R., "Comparing Bayesian network classifiers," *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, 1999.
- [13] Cheng, J., Bell, D.A., and Liu, W., "Learning belief networks from data: an information theory based approach," *Proceedings of the Sixth ACM International Conference on Information and Knowledge Management*, 1997.