

Classification of Gene Expression Data by Ensemble of Bayesian Networks

Kyu-Baek Hwang, Jeong-Ho Chang, and Byoung-Tak Zhang

School of Computer Science and Engineering, Seoul National University

DNA (gene expression data), (gene expression pattern), (probabilistic graphical model), (Bayesian network), (ensemble of Bayesian networks), Bayes, support vector machine (SVM), naïve

1. DNA (gene expression data) (expression level) (gene expression pattern) [4], (probabilistic graphical model) (Bayesian network)[6] (random variable) (joint probability distribution) (conditional independency) (conditional probability) (input feature) class

2. (Bayesian Networks)

2.1 $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$ (local probability distribution) G DAG \mathbf{X} (directed acyclic graph) G (conditional independence assertions) \mathbf{X} [6].

$$P(\mathbf{X}) = \prod_{i=1}^n P(X_i | \mathbf{Pa}_G(X_i)) \quad (1)$$

2.2

Complete data G (maximum likelihood estimation) [6]. $D = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ G D $Score(G;D)$ \log

(ensemble of Bayesian networks) Bayesian model averaging [4]

$$\begin{aligned} \log Score(G;D) &= \text{penalizing term} + \log \text{likelihood} \\ &= \text{penalizing term} + \sum_{t=1}^M \log P_G(\mathbf{x}_t) \\ &= \text{penalizing term} \\ &\quad + \sum_{t=1}^M \sum_{i=1}^n \log P(X_i = x_{ti} | \mathbf{Pa}_G(X_i) = \mathbf{pa}_{ti}) \end{aligned} \quad (2)$$

\mathbf{x}_t $P_G(\mathbf{x}_t)$ G \mathbf{x}_t x_{ti} \mathbf{pa}_{ti} penalizing term G (prior probability) $P(G)$

2.3

[2].

가 , 가 가
 가 가 가
 class C ,
 A_1, A_2, \dots, A_{n-1}
 (2) log likelihood

$$\log \text{likelihood} = \sum_{t=1}^M \log P_G(A_1 = a_{t1}, A_2 = a_{t2}, \dots, A_{M-1} = a_{(M-1)t})$$
 (3)

$$+ \sum_{t=1}^M \log P_G(C = c_t | A_1 = a_{t1}, A_2 = a_{t2}, \dots, A_{M-1} = a_{(M-1)t})$$
 (3)
 term
 (2)
 가 [2].

naive Bayes
 가
 [1].

3. 3.1 (Bayesian learning for Bayesian networks)

(Bayesian learning) (:
) (prior probability) 가
 (posterior probability)
 maximum likelihood (ML)
 maximum a posteriori (MAP) 가
 over-fitting

(4)

$$P_G(C | A_1, \dots, A_{n-1}) = \alpha \cdot P_G(C, A_1, \dots, A_{n-1})$$

$$= \alpha \cdot \int \{P(C | \mathbf{P}a_G(C)) \cdot \prod_{i=1}^{n-1} P_B(A_i | \mathbf{P}a_G(A_i))\} P(\Theta | D) d\Theta$$

 normalization constant , G

(5)

$$P_G(C | A_1, \dots, A_{n-1}) = \alpha \cdot P_G(C, A_1, \dots, A_{n-1})$$

$$= \alpha \cdot \sum_{G \in \mathbf{g}} P_G(C, A_1, \dots, A_{n-1}) \cdot P(G | D)$$

$$= \frac{\alpha}{P(D)} \cdot \sum_{G \in \mathbf{g}} P_G(C, A_1, \dots, A_{n-1}) \cdot P(G) \cdot P(D | G)$$

 g 가 G
 가 가 $O(2^{n^2})$
 n 6 가 가
 G (5)

3.2

(5)

(6)

$$P_G(C | A_1, \dots, A_{n-1}) = \alpha \cdot P_G(C, A_1, \dots, A_{n-1})$$

$$\approx \alpha \cdot \sum_{G \in \mathbf{g}_{high}} P_G(C, A_1, \dots, A_{n-1}) \cdot P(G | D)$$

$$= \frac{\alpha}{P(D)} \cdot \sum_{G \in \mathbf{g}_{high}} P_G(C, A_1, \dots, A_{n-1}) \cdot \text{Score}(G; D)$$

 \mathbf{g}_{high} $P(G|D)$ 가 \mathbf{g}
 (5) \mathbf{g}_{high}

multimodality가

[5].

가 , K
 (n - 1) / K
 class

For $k = 1, 2, \dots, K$
 - $n - 1$, (n - 1) / K
 . (iteration
)
 - class
 - greedy
 search algorithm

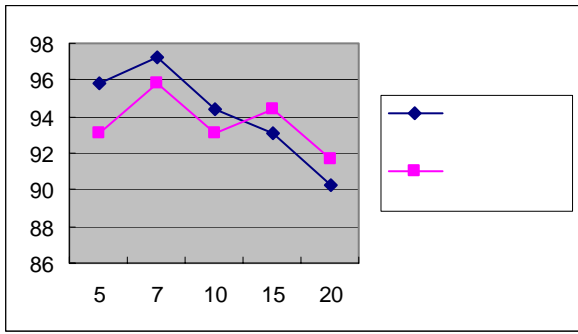
4.

[4]

		72	
leukemia (ALL)	25	47	acute lymphoblastic
			acute myeloid leukemia (AML) ,
			7,129
			[4]
			가 50

0 1
 $\text{Score}(G; D)$ BD (Bayesian Dirichlet) [3]가
 K 5, 7, 10, 15, 20

4.1
 가 가
 leave-one-out cross validation (LOOCV)



1. K
(LOOCV, %)

가 1 가 가 . K 가 5, 7, 10 가 7 가 97.22% ($K = 15, 20$) class Class Markov blanket[7] Markov blanket

1. K (Average #) class Markov blanket (Total #)

	5	7	10	15	20
Total #	35.00	37.24	34.50	37.51	36.72
Average #	14.14	12.53	9.85	8.99	8.11

K Markov blanket 36 K 가 가 가 4.2

[4] weighted voting scheme, c4.5, naïve Bayes, support vector machine (SVM) weighted voting 가 2 가 SVM 98.61% 가 97.22% SVM

2.

		(LOOCV)
Weighted voting		95.83%
(c4.5)		95.83%
naïve Bayes		95.83%
		97.22%
		95.83%
		97.22%
Support vector machine		97.22%
		98.61%
		95.83%
		97.22%

5.

MCMC(Markov chain Monte Carlo)

BK21
IMT-2000, NRL BrainTech

[1] Domingos, P. and Pazzani, M., On the optimality of the simple Bayesian classifier under zero-one loss, *Machine Learning*, vol. 29, no. 2, pp. 103-130, 1997.
 [2] Friedman, N., Geiger, D., and Goldszmidt, M., Bayesian network classifiers, *Machine Learning*, vol. 29, no. 2, pp. 131-163, 1997.
 [3] Friedman, N. and Goldszmidt, M., Learning Bayesian networks with local structure, *Learning in Graphical Models*, Jordan, M.I., (ed.), pp. 421-459, Cambridge, MA: MIT Press, 1999.
 [4] Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., and Lander, E.S., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, vol. 286, no. 5439, pp. 531-537, 1999.
 [5] Haykin, S., *Neural Networks: A Comprehensive Foundation*, Upper Saddle River, NJ: Prentice-Hall, 1999.
 [6] Heckerman, D., A tutorial on learning with Bayesian networks, *Learning in Graphical Models*, Jordan, M.I., (ed.), pp. 301-354, Cambridge, MA: MIT Press, 1999.
 [7] Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Francisco, CA: Morgan Kaufmann Publishers, 1988.