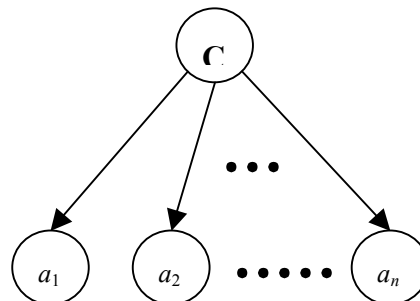


The Influence of Prior Knowledge on Bayesian Networks Learning

kbhwang@scai.snu.ac, {btzhang, ytkim}@cse.snu.ac.kr

: (prior knowledge)
가

(prior knowledge)



1:

tfidf

: $X = \{X_1, \dots, X_n\}$

3009

$\alpha_{111}, \alpha_{112}$

$$p(\mathbf{x} | \boldsymbol{\theta}_s, S^h) = \prod_{i=1}^n p(x_i | \mathbf{pa}_i, \boldsymbol{\theta}_i, S^h)$$

\mathbf{pa}_i X_i ()
 $\boldsymbol{\theta}_i$ X_i 가

가 (unrestricted multinomial distribution)
 X_i 가 r_i 가

$\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{iq_i})$, $\boldsymbol{\theta}_{ij} = (\theta_{ij2}, \theta_{ij3}, \dots, \theta_{ijr_i})$ 가 (, $q_i = \prod_{X_i \in \mathbf{pa}_i} r_i$, $\theta_{ij1} = 1 - (\theta_{ij2} + \theta_{ij3} + \dots + \theta_{ijr_i})$.)

θ_{ijk} (Dirichlet distribution)

$$p(\boldsymbol{\theta}_{ij} | S^h) = \text{Dir}(\boldsymbol{\theta}_{ij} | \alpha_{ij1}, \dots, \alpha_{ijr_i})$$

$$\alpha_{ijk} (k = 1, \dots, r_i)$$

D

$$p(\boldsymbol{\theta}_{ij} | D, S^h) = \text{Dir}(\boldsymbol{\theta}_{ij} | \alpha_{ij1} + N_{ij1}, \dots, \alpha_{ijr_i} + N_{ijr_i})$$

$N_{ijk} (k = 1, \dots, r_i)$ D
 X_i 가 \mathbf{pa}_{ij} k 가

:
[1] C
 a_1, \dots, a_n

Reuters-21578

acq

297 , 1456
8754

: [1], [2]

가 가
[1] 가 [2] 가
가 negative 가

(500, 500)

가

가 가

$(\alpha_{111}, \alpha_{112})$	(1, 1)	(50, 50)	(500, 500)
(%)	86.24	86.34	86.91

1:

가

$(\alpha_{111}, \alpha_{112})$	(15, 3)	(90, 18)	(800, 200)
(%)	86.21	86.21	86.31

2:

가

[1] Heckerman, D., A Tutorial on Learning with Bayesian Networks, Technical Report MSR-TR-95-06, Microsoft Research, Redmond, WA, 1995