퍼지 및 지능시스템학회 논문지 2007, Vol. 17, No. 3, pp. 316-321

베이지안 망 연결 구조에 대한 데이터 군집별 기여도의 정량화 방법에 대한 연구

Quantitative Annotation of Edges in Bayesian Networks with Condition-Specific Data

정성원[1], 이도헌[2], 이광형[3]
Sungwon Jung[1], Doheon Lee[2] and Kwang H. Lee[3]

[1] 대전광역시 유성구 한국과학기술원 바이오시스템학과
E-mail: swjung@biosoft.kaist.ac.kr
[2] 대전광역시 유성구 한국과학기술원 바이오시스템학과
E-mail: dhlee@biosoft.kaist.ac.kr
[3] 대전광역시 유성구 한국과학기술원 바이오시스템학과, AITrc
E-mail: khlee@biosoft.kaist.ac.kr

요 약

본 연구에서는 베이지안 망 구조 학습에서, 학습 데이터의 특정 부분집합이 학습된 망의 각 연결 구조(edge)의 형성에 기여하는 정도를 정량화하는 방법을 제안한다. 생물학 정보의 분석 등에 베이지안 망 학습을 이용하는 경우, 제안된 방법은 망의 각 연결 구조의 형성에 특정 군집 데이터가 기여하는 정도의 정량화가 가능하다. 제안된 방법의 유효성을 보이기 위해, 벤치마크 베이지안 망을 이용하여 제안된 방법이 망 연결 구조에 대한 데이터 군집별 기여도를 효과적으로 정량화할 수 있음을 보인다.

키워드 : 베이지안 망, 정량적 주석화, 조건부 데이터

Abstract

We propose a quatitative annotation method for edges in Bayesian networks using given sets of condition-specific data. Bayesian network model has been used widely in various fields to infer probabilistic dependency relationships between entities in target systems. Besides the need for identifying dependency relationships, the annotation of edges in Bayesian networks is required to analyze the meaning of learned Bayesian networks. We assume the training data is composed of several condition-specific data sets. The contribution of each condition-specific data set to each edge in the learned Bayesian network is measured using the ratio of likelihoods between network structures of including and missing the specific edge. The proposed method can be a good approach to make quantitative annotation for learned Bayesian network structures while previous annotation approaches only give qualitative one.

Key Words : Bayesian network, quantitative annotation, condition-specific data

1. 서 론

베이지안 망은 확률변수들 사이의 확률적 조건부 의존 관계를 망 형태의 모습으로 나타내는 모델이다[1]. 한 베이지안 망 $B=(G,\boldsymbol{\Theta})$은 확률변수들 사이의 조건부 의존 관계를 나타내는 directed acyclic graph (DAG) G와, 조건부 확률분포를 기술하는 매개변수들의 집합 $\boldsymbol{\Theta}$로 구성된다. 어떤 시스템으로부터 관찰된 데이터 $\boldsymbol{D}$를 이용하여 시스템 각 요소들 사이의 관계를 분석하고자 하는 경우, 주어진 데이터로부터 최적의 베이지안 망을 학습함으로써 각 요소들 사이의 관계를 분석하는 방법이 널리 사용되고 있다[2][3]. 특히 생물학 데이터와 같이 대상 시스템이 확률적인 특성과 노이즈를 갖고 있는 경우, 베이지안 망 학습을 통해 목표 시스템을 분석하는 방법이 널리 사용되어지고 있다[4][5].

베이지안 망 학습을 통해 목표 시스템을 이루는 각 요소들 사이의 관계를 분석하는 경우, 베이지안 망 구조를 이루는 각 edge들이 어떤 의미를 내포하는가를 정의하는 주석화 작업은 결과를 올바르게 분석하는 데에 있어 필수적인 부분이다. 그러나 베이지안 망 edge를 주어진 학습 데이터를 이

접수일자 : 2007년 4월 1일
완료일자 : 2007년 5월 25일
감사의 글 : 본 논문은 과학기술부 시스템생물학 연구사업(2005-00343)의 지원과 정보통신부 연구비 C109006020001의 지원으로 수행되었음. 연구시설은 정물술 바이오정보전자센터의 도움을 받았음.

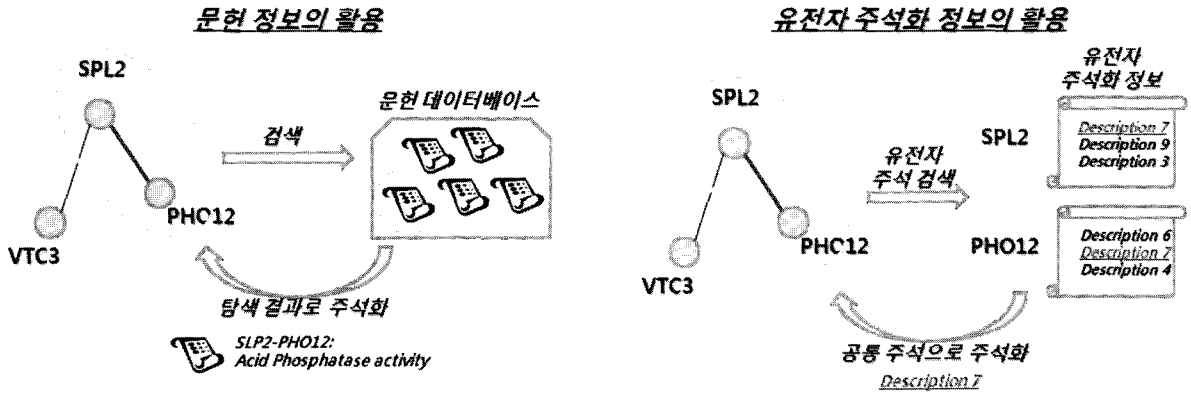


그림 1. 망 구조에 대한 기존의 정성적 주석화 방법의 예
Fig. 1. Examples of previous qualitative annotation approaches for network structures.

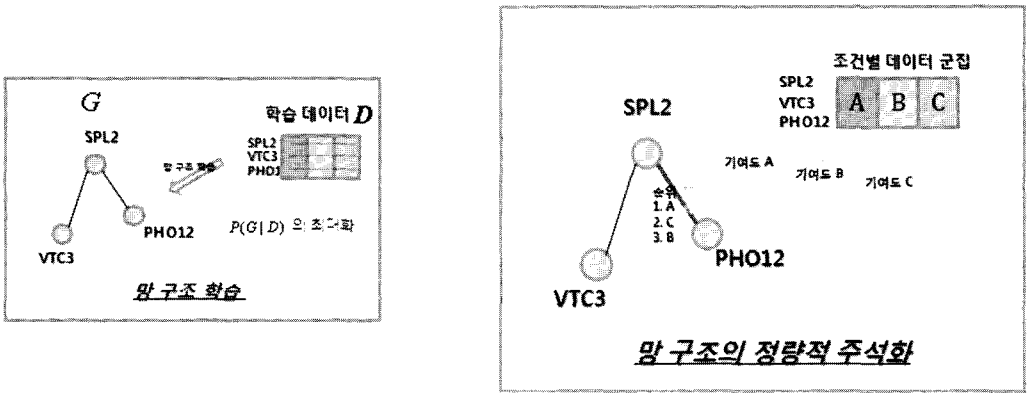


그림 2. 제안된 정량적 주석화 방법의 개요. 망 구조 학습에 이용된 데이터는 조건별 데이터 군집으로 이루어진 것으로 가정.
Fig. 2. Proposed approach for quantitative annotation of edges in network structures. Source data for learning network structures is assumed to be composed of condition-specific data sets.

용하여 정량적으로 주석화하는 방법이 현존하지 않는 관계로, 기존의 베이지안 망을 활용한 연구들의 경우 정성적 주석화만을 수행하고 있는 것이 현실이다(그림 1). 본 연구에서는 주어진 데이터가 서로 중복되지 않는 부분집합들로 이루어진 경우, 각 부분집합이 베이지안 망의 각 edge의 형성에 기여하는 정도를 정량화하는 방법을 제안한다. 제안된 데이터 군집의 edge에 대한 기여도의 정량화 방법을 사용하여, 학습된 베이지안 망의 각 edge의 형성에 어떤 데이터 군집이 보다 더 기여하는지를 정량화하는 것이 가능하며 이를 통해 해당 edge가 보다 높은 기여도를 보이는 데이터 군집의 생성 환경에 의해 결정된다는 것을 파악할 수 있다.

본 논문은 다음과 같이 구성된다. 2장에서는 제안된 데이터 군집의 edge에 대한 기여도의 정량화 척도를 기술한다. 3장에서는 알려진 베이지안 망을 벤치마크로 사용하여 제안된 기여도 정량화 척도를 통해 신뢰성 있는 정량적 주석화가 가능함을 보인다. 마지막으로 4장에서는 본 연구의 요약 및 향후 과제에 대한 고찰이 언급된다.

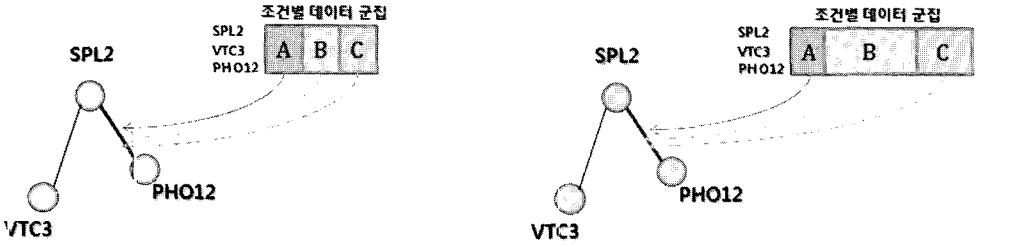
2. 제안된 데이터 군집별 기여도의 정량화 척도

서로 중복되지 않고, 공집합이 아닌 부분집합 $\boldsymbol{D}_j$의 집합으로 이루어진 관찰된 데이터 집합 $\boldsymbol{D}=\boldsymbol{D}_1\cup\boldsymbol{D}_2\cup\cdots\cup\boldsymbol{D}_d$가 있다고 하자. 베이지안 망 구조 학습의 일반적인 절차는 주어진 $\boldsymbol{D}$를 이용하여 최대의 $P(G|\boldsymbol{D})$를 갖는 G를 찾는 것이다. 본 연구의 목표는 G에 있는 한 edge e_i의 존재를 위해 각 $\boldsymbol{D}_j$가 기여하는 정도를 정량화하는 것이다 (그림 2).

한 edge e_i가 G에 포함됨으로서 주어진 $\boldsymbol{D}$에 대해 얻게 되는 이득의 정도는 다음과 같이 계산되어질 수 있으며

$$\frac{P(G|\boldsymbol{D})}{P(G_{/e_i}|\boldsymbol{D})} \tag{1}$$

역으로 이것을 $\boldsymbol{D}$가 e_i의 형성에 기여하는 정도로 생각할 수 있다. 즉, $\boldsymbol{D}$가 주어진 G에서 e_i의 형성에 기여하는 정도를 다음과 같이 정의한다.



각 데이터 군집의 크기가 동일한 경우 -> 기여도의 상호 비교 가능

각 데이터 군집의 크기가 다른 경우 -> 단위 데이터당 기여도로의 정규화 필요

그림 3. 조건별 데이터 군집의 크기가 서로 다른 경우, 기여도의 정규화가 필요.
Fig. 3. Normalization of contributions is need when the sizes of condition-specific data sets are different.

$$\frac{P(G|\boldsymbol{D})}{P(G_{/e_i}|\boldsymbol{D})} \tag{2}$$

다음과 같은 Bayesian theorem에 의해,

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)} \tag{3}$$

식 (2)는 다음과 같이 표현되어질 수 있다.

$$\frac{P(\boldsymbol{D}|G)P(G)}{P(\boldsymbol{D}|G_{/e_i})P(G_{/e_i})} \tag{4}$$

그래프 구조에 대한 prior probability가 모두 동일하다고 가정하면, 즉 G와 $G_{/e_i}$에 대한 prior probability $P(G)$와 $P(G_{/e_i})$가 같다고 가정함을 통해 $\boldsymbol{D}$가 G에 대해 e_i의 형성에 기여하는 정도를 다음과 같이 정의할 수 있다.

정의 1. 주어진 DAG G와 관찰된 데이터의 집합 $\boldsymbol{D}$에 대해, $\boldsymbol{D}$가 G에 속한 e_i의 형성에 기여하는 정도를 다음과 같이 정의한다.

$$cont(e_i;G,\boldsymbol{D})=\frac{P(\boldsymbol{D}|G)}{P(\boldsymbol{D}|G_{/e_i})} \tag{5}$$

정의 1에서 정의된 데이터의 edge에 대한 기여도는 다음과 같은 특성을 갖는다.

정리 1. 관찰된 데이터의 집합 $\boldsymbol{D}$가 서로 중복되지 않고, 공집합이 아닌 부분집합 $\boldsymbol{D}_j$의 집합으로 이루어져 있다고 가정하자 ($\boldsymbol{D}=\boldsymbol{D}_1\cup\boldsymbol{D}_2\cup\ldots\cup\boldsymbol{D}_d$). 모든 $\boldsymbol{D}_j$가 서로 독립이라고 가정하면 다음 식이 성립한다.

$$cont(e_i;G,\boldsymbol{D})=\prod_{j=1}^{d}cont(e_i;G,\boldsymbol{D}_j) \tag{6}$$

증명) 모든 $\boldsymbol{D}_j$는 서로 독립이므로,

$$P(\boldsymbol{D}|G)=\prod_{j=1}^{d}P(\boldsymbol{D}_j|G)$$

$$P(\boldsymbol{D}|G_{/e_i})=\prod_{j=1}^{d}P(\boldsymbol{D}_j|G_{/e_i})$$

따라서, 식 (5)는 다음과 같이 전개된다.

$$\begin{aligned}cont(e_i;G,\boldsymbol{D})&=\frac{P(\boldsymbol{D}|G)}{P(\boldsymbol{D}|G_{/e_i})}\\&=\frac{\prod_{j=1}^{d}P(\boldsymbol{D}_j|G)}{\prod_{j=1}^{d}P(\boldsymbol{D}_j|G_{/e_i})}=\prod_{j=1}^{d}\frac{P(\boldsymbol{D}_j|G)}{P(\boldsymbol{D}_j|G_{/e_i})}\\&=\prod_{j=1}^{d}cont(e_i;G,\boldsymbol{D}_j)\end{aligned}$$

(증명 끝)

정의 1에서 정의한 $\boldsymbol{D}$의 e_i에 대한 기여도는 주어진 구조 조건부 데이터의 확률을 계산하는 부분에서 데이터의 수, 즉 $|\boldsymbol{D}|$에 의해 영향을 받는다. 즉, $\boldsymbol{D}_j$와 $\boldsymbol{D}_k$의 크기가 서로 다른 경우 ($|\boldsymbol{D}_j|\neq|\boldsymbol{D}_k|$) G의 한 edge e_i에 기여하는 정도를 서로 동등하게 비교하기 위해서는 집합 전체의 e_i에 대한 기여도를 비교하는 것 보다 집합 내의 한 데이터 경우가 기여하는 정도를 비교하는 것이 바람직하다 (그림 3).

$\boldsymbol{D}$ 내의 모든 데이터 경우 d_j가 서로 독립이라고 가정하면 정리 1에서의 증명과 같은 과정을 통해, 다음을 알 수 있다.

$$\begin{aligned}cont(e_i;G,\boldsymbol{D})&=\frac{P(\boldsymbol{D}|G)}{P(\boldsymbol{D}|G_{/e_i})}\\&=\prod_{j=1}^{|\boldsymbol{D}|}\frac{P(d_j|G)}{P(d_j|G_{/e_i})}\end{aligned} \tag{7}$$

즉, $\boldsymbol{D}$의 e_i에 대한 기여도는 $\boldsymbol{D}$에 속한 모든 데이터 경우들 각각의 기여도의 곱에 해당하므로, $cont(e_i;G,\boldsymbol{D})$의 조화평균을 구함으로써 $\boldsymbol{D}$의 e_i에 대한 평균적인 단위 기여도를 구할 수 있다.

정의 2. 주어진 $\boldsymbol{D}$에 속한 각 데이터 경우들 d_j가 모두 서로 독립이라고 가정하자. 이 때, $\boldsymbol{D}$의 주어진 G의 한 edge e_i의 형성을 위한 단위 기여도를 다음과 같이 정의한다.

$$u-cont(e_i;G,\boldsymbol{D})=\sqrt[|\boldsymbol{D}|]{cont(e_i;G,\boldsymbol{D})} \tag{8}$$

제안된 정량화 척도를 통해, 관찰된 학습 데이터가 그래프 구조 G의 형성 과정에서 각 edge e_i에 기여하는 정도를 수치적으로 비교하는 것이 가능하다. 또한 형성된 그래프 구조 G에 포함된 각 edge들이 관찰된 학습 데이터 중 어느 부분

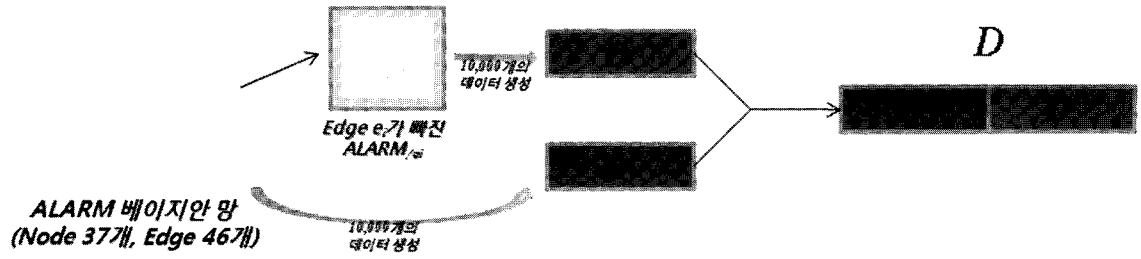


그림 5. 실험에 사용될 두 조건별 데이터 군집 D+와 D-의 생성.
Fig. 5. Generation of two condition-specific data sets D+ and D- from ALARM Bayesian network.

에 의해 주도적으로 형성된 것인지에 대해 정량적으로 주석화할 수 있다.

3. 실험 및 결과

3.1 실험 환경

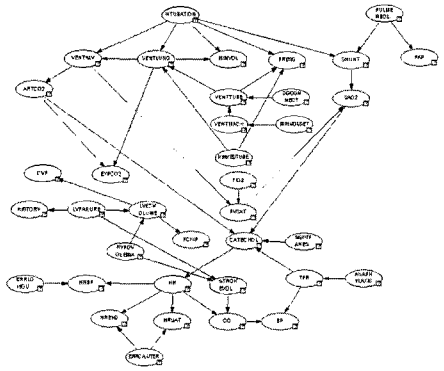


그림 4. ALARM 베이지안 망
Fig. 4. ALARM Bayesian network

제안된 정량화 척도 $u-cont$를 사용하여 그래프 구조에 속한 edge들의 올바른 정량적 주석화가 가능함을 보이기 위해, 그림 4와 같은 ALARM 베이지안 망을 사용한 실험을 수행하였다. ALARM 베이지안 망은 37개의 확률변수와 46개의 edge로 구성되어 있으며 베이지안 망에 대한 다양한 연구에서 벤치마크로 널리 사용되어지고 있다.

먼저, ALARM에 속한 각 edge e_i의 형성에 기여하는 데이터 집합을 만들기 위해, ALARM으로부터 10,000개의 데이터로 구성된 집합 $\boldsymbol{D}^+$를 생성하였다. 그리고 e_i의 형성에 기여하지 않는 데이터 집합을 만들기 위해, ALARM으로부터 e_i를 제거한 ALARM$_{/e_i}$를 각각 만든 후 10,000개의 데이터를 생성하여 $\boldsymbol{D}_i^-$를 구성하였다. ALARM$_{/e_i}$를 만들 때 e_i를 제거함으로써 e_i의 끝 부분에 연결되는 확률변수 v의 조건부 확률분포 표의 수정이 필요하다. 이 과정에서 v의 prior probability가 ALARM에서의 v와 동일한 값을 갖도록 확률분포 표를 수정하였다. 이를 통해, ALARM$_{/e_i}$에서의 모든 확률변수들은 ALARM에서와 동일한 prior probability를 갖지만 e_i가 표현하는 조건부 확률적 의존성은 갖지 않게 된다 (그림 5).

이와 같은 과정으로 ALARM의 각 e_i에 대해, e_i가 존재하는 환경에서 생성된 $\boldsymbol{D}^+$와 존재하지 않는 환경에서 생성된 $\boldsymbol{D}_i^-$가 기여하는 각각의 $u-cont(e_i;\text{ALARM},\boldsymbol{D}^+)$, $u-cont(e_i;\text{ALARM},\boldsymbol{D}_i^-)$값을 비교하였다. 일반적으로 $u-cont(e_i;\text{ALARM},\boldsymbol{D}^+) > u-cont(e_i;\text{ALARM},\boldsymbol{D}_i^-)$가 성립한다면, 제안된 기여도의 정량화 척도가 각각의 edge e_i가 형성되는데에 기여하는 데이터 집합을 구분하고 비교하는 데

표 1. D^+와 D^-의 각 edge에 대한 정규화 기여도의 비교
Table 1. Normalized contributions of D^+ and D^- for each edge.

i	$\log(u-cont(e_i;G,\boldsymbol{D}^+))$	$\log(u-cont(e_i;G,\boldsymbol{D}_i^-))$
1	0.1064	-0.0029
2	0.7653	-0.0055
3	0.0868	-0.0063
4	0.67	-0.0107
5	0.1022	-0.0039
6	0.7769	-0.0066
7	-0.0006	-0.0033
8	0.0841	-0.0038
9	0.6745	-0.0118
10	0.1326	-0.0054
11	0.0231	-0.0004
12	0.1133	-0.0007
13	0.0188	-0.0003
14	0.6243	-0.0021
15	0.276	-0.0017
16	0.4584	-0.0039
17	0.2606	-0.0013
18	0.6102	-0.0023
19	0.3026	-0.0015
20	0.5269	-0.0011
21	0.4158	-0.0084
22	0.2068	-0.0037
23	0.1448	-0.0002
24	0.3671	-0.001
25	0.0886	-0.0007
26	0.1062	-0.0012
27	0.6742	-0.0028
28	0.6129	-0.0013
29	-0.0081	-0.0084
30	0.0776	-0.0068

31	0.2244	-0.01
32	0.0804	-0.0078
33	0.6707	-0.0022
34	0.0703	-0.001
35	0.4073	-0.0006
36	0.3354	-0.0011
37	0.0286	-0.001
38	0.0108	-0.0002
39	0.6208	-0.0023
40	0.142	-0.0014
41	0.6278	-0.0024
42	0.1486	-0.0015
43	0.4624	-0.0032
44	0.4529	-0.0027
45	0.4234	-0.004
46	0.303	-0.0043

에 사용되어질 수 있음을 의미하게 된다.

3.2 결과

ALARM 베이지안 망에 존재하는 46개의 edge 각각에 대해, 제안된 데이터 기여도의 정량화 척도를 $\boldsymbol{D}^{+}$와 $\boldsymbol{D}_i^{-}$ 각각을 통해 계산한 결과는 표 1 및 그림 6과 같다. 실험에서는 계산의 편의성을 위해, 각 기여도 값의 log를 취한 값을 계산하였다.

결과에서 보는 바와 같이, 모든 edge들의 경우에 대해 $u-cont(e_i;\mathrm{ALARM},\boldsymbol{D}^{+}) > u-cont(e_i;\mathrm{ALARM},\boldsymbol{D}_i^{-})$ 가 성립함을 알 수 있다. 이를 통해 제안된 데이터별 edge에 대한 기여도의 정량화 척도가 올바른 결과를 제공함을 알 수 있다. 표에 나타난 값은 제안된 기여도 값의 log이므로, 0보다 큰 기여도 값은 각 edge의 형성에 해당 데이터가 긍정적으로 기여하고 있음을 의미하며, 0보다 작은 기여도 값은 각 edge가 형성되지 않는 방향을 선호하는 데이터임을 의미한다.

또한 각 기여도 값의 절대값 크기에 따라 기여하는 정도의 상대적인 비교가 가능하다. 즉, 보다 큰 기여도 값을 보이는 경우 해당 데이터 집합이 그 edge의 형성에 보다 더 크게 기여하고 있음을 의미한다. 예를 들어 e_1과 e_2의 경우, 동일한 데이터 $\boldsymbol{D}^{+}$이지만 이 데이터가 e_1의 형성에 기여하는 정도보다 e_2의 형성에 기여하는 정도가 보다 큼을 의미한다. e_7과 e_{29}의 경우 두 edge들이 존재하는 상황에서 생성된 $\boldsymbol{D}^{+}$의 기여도가 거의 없거나 혹은 미미하게 반대 방향으로 기여하는 결과가 나타났다. 이것은 두 edge가 나타내는 확률변수 간의 조건부 확률적 의존성의 강도가 지극히 미미함으로 인해 생성된 10,000개의 데이터만으로는 두 edge의 형성에 필요한 조건부 확률적 의존성이 잘 드러나지 않았기 때문이다.

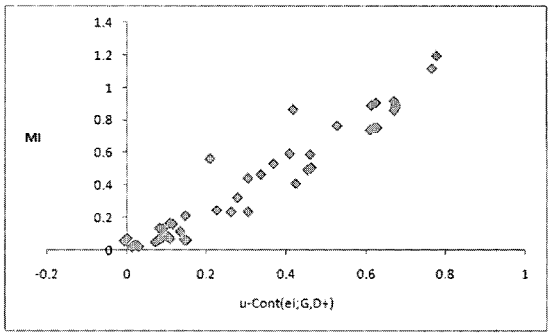


그림 7. 제안된 데이터의 edge에 대한 기여도 정량화 척도와 mutual information과의 상관관계

그림 7은 46개의 edge 각각에 대해 $\boldsymbol{D}^{+}$ 데이터의 단위 기여도와, 해당하는 edge로 연결되는 두 확률변수 사이에서의 mutual information값을 $\boldsymbol{D}^{+}$로부터 계산한 결과를 비교한 그래프이다. 제안된 데이터의 edge에 대한 단위 기여도 값은 mutual information과 0.96의 correlation 값을 보이며 밀접하게 연관되어 있음을 알 수 있다.

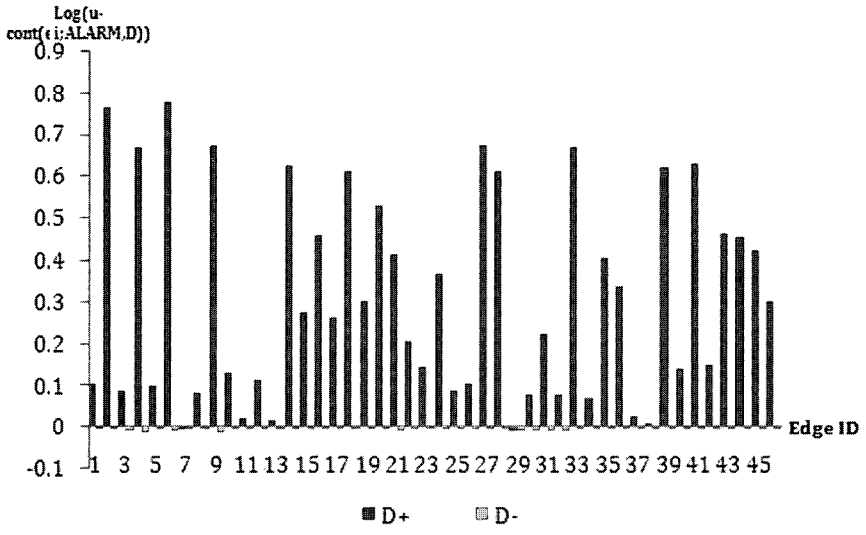


그림 6. D^{+}와 D^{-}의 각 edge에 대한 정규화 기여도의 비교
Fig. 6. Normalized contributions of D^{+} and D^{-} for each edge.

본 결과를 통해, 제안된 데이터별 edge에 대한 기여도의 정량화 척도를 사용하여 한 edge의 형성에 대해 여러 데이터 집합 중 어느 데이터 집합이 보다 잘 기여하는지를 판단할 수 있을 뿐 아니라, 어떤 하나의 데이터 집합이 어느 edge의 형성에 보다 더 기여하는지에 대한 정보도 제공함을 알 수 있다.

4. 결론 및 향후 과제

본 연구에서는, 주어진 데이터로부터 DAG 형태의 상호연관 관계 구조를 학습한 결과에 대해 각 edge의 형성에 대한 데이터의 정량적 기여도를 측정하기 위한 척도를 제안하였다. 벤치마크 베이지안 망을 이용하여 제안된 척도가 각 데이터 군집별 기여도를 정확하게 정량화하여 비교할 수 있게 해 줌을 알 수 있다. 또한 제안된 척도를 사용하여 어떤 하나의 데이터 군집이 각 edge의 형성에 얼마나 기여하는지에 대한 상대적인 비교 또한 가능함을 실험을 통해 입증하였다.

향후 과제로서, 현재 제안된 정량화 척도의 경우 주어진 그래프 구조 G에 대해서의 데이터별 기여도만을 측정하는 것이 가능하나, 특정 그래프 구조가 주어지지 않은 상황에서 임의의 두 확률변수 사이에 edge가 존재하도록 기여하는 정도를 정량화하는 척도로의 확장이 필요하다.

참 고 문 헌

[1] R. E. Neapolitan, "Learning Bayesian Networks," Pearson Prentice Hall, 2004.

[2] N. Friedman, I. Nachman and D. Pe'er, "Learning Bayesian Network Structure from Massive Datasets: The "Sparse Candidate" Algorithm," In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, pp. 206-215, 1999.

[3] D. Heckerman, D. Gerger and D. M. Chickering, "Learning Bayesian Network: The Combination of Knowledge and Statistical Data," Machine Learning, Vol. 20, pp. 197-243, 1995.

[4] N. Friedman, M. Linial, I. Nachman and D. Pe'er, "Using Bayesian Networks to Analyze Expression Data," Journal of Computational Biology, Vol. 7, pp. 601-620, 2000.

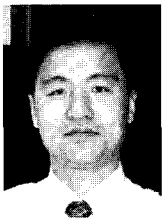
[5] P. H. Lee and D. Lee, "Modularized Learning of Genetic Interaction Networks from Biological Annotations and mRNA Expression Data," Bioinformatics, Vol. 21, No. 11, pp. 2739-2747, 2005.

저 자 소 개

정성원(Sungwon Jung)

1998년 : 한국과학기술원 전산학과 졸업
2000년 : 한국과학기술원 전산학과 졸업 (공학석사)
2007년 : 한국과학기술원 전산학과 졸업 (공학박사)
2007년~현재 : 한국과학기술원 IBM-KAIST 바이오컴퓨팅 연구센터 연구원

관심분야 : 기계학습, 바이오정보학, 인공지능, 의료정보학
Phone : 042-869-5356
Fax : 042-869-8680
E-mail : swjung@biosoft.kaist.ac.kr



이도헌(Doheon Lee)

1990년 : 한국과학기술원 전산학과 졸업
1992년 : 한국과학기술원 전산학과 졸업 (공학석사)
1995년 : 한국과학기술원 전산학과 졸업 (공학박사)
1994년~1995년 : 한국전자통신연구원 위촉연구원
1996년~2002년 : 전남대학교 교수
2002년~현재 : 한국과학기술원 바이오시스템학과 교수

관심분야 : 바이오 데이터 마이닝, 바이오 시스템 모델링, 바이오정보학
Phone : 042-869-4316
Fax : 042-869-8680
E-mail : dhlee@biosoft.kaist.ac.kr

이광형(Kwang H. Lee)

1978년 : 서울대학교 산업공학과 졸업
1980년 : 한국과학기술원 산업공학과 졸업 (공학석사)
1985년 : INSA de Lyon 전산학과 졸업 (공학박사)
1985년~현재 : 한국과학기술원 교수
2000년~현재 : 한국과학기술원 미래산업 석좌교수

관심분야 : 바이오정보학, 인공지능, 퍼지 시스템
Phone : 042-869-4313
Fax : 042-869-8680
E-mail : khlee@biosoft.kaist.ac.kr