

베이지안 네트워크의 불확실성 감소를 위한 확률분포의 분산 감소 방법

정성원⁰ 이도현¹ 이광형¹
한국과학기술원 전자전산학과 전산학 전공⁰
한국과학기술원 바이오시스템학과¹
{swjung⁰, dhlee, khlee}@biosoft.kaist.kr

Reducing Uncertainty of Bayesian Networks by Reducing Variances of Probability Distributions

Sungwon Jung⁰ Doheon Lee¹ Kwang H. Lee¹
Division of Computer Science, Department of Electrical Engineering and Computer Science, KAIST⁰
Department of BioSystems, KAIST¹

요 약

베이지안 네트워크는 주어진 변수들 사이의 확률적 의존성을 분석하는 데에 널리 사용되어지고 있는 모델이다. 이러한 베이지안 네트워크의 활용에 있어서 베이지안 네트워크의 확실성을 분석하는 방법의 필요성이 대두되어지고 있다. 특히 규모가 큰 베이지안 네트워크 모델을 특정하는 상황에서 주어질 수 있는 학습 데이터의 수가 제한되는 경우나, 주된 관심사가 베이지안 네트워크의 일부 부분에 한정되는 경우에 베이지안 네트워크의 확실성에 대한 분석은 유용하게 사용될 수 있다. 본 논문에서는, 베이지안 네트워크에 존재할 수 있는 불확실성을 언급한 후, 베이지안 네트워크 내의 변수들이 갖는 확률분포의 분산을 이용해 베이지안 네트워크의 불확실성을 정의하는 방법을 제안한다. 간단한 베이지안 네트워크의 예시 모델을 이용하여 제안된 베이지안 네트워크의 불확실성 분석 방법이 유용할 수 있음을 보인다.

1. 서 론

베이지안 네트워크는 변수들 사이의 조건부 확률을 표현하기 위한 모델로서, 변수들 사이의 확률적 의존성을 분석하기 위한 용도로 많이 사용되어지고 있다. 베이지안 네트워크 B 는 (G, θ) 로 표현되며, G 는 변수들을 node로, 변수들 간의 의존성을 방향성 있는 edge로 표현한 directed acyclic graph (DAG)이다. 또한 θ 는 G 로 나타내어지는 변수들 간의 조건부 확률분포를 기술하기 위한 확률값들의 집합이다. 베이지안 네트워크의 사용 초기에는 전문가 시스템과 같은 곳에서의 응용을 위해 인간 전문가에 의해 만들어진 베이지안 네트워크로부터의 추론에 대한 연구가 많이 이루어졌다. 그러나 최근 베이지안 네트워크를 활용한 연구의 주된 방향 중 하나는 특정 시스템을 이루는 요소들 사이의 관계를 규명하기 위하여, 주어진 학습 데이터로부터 변수들 사이의 확률적 의존성을 학습을 통해 분석하는 것이다.

주어진 학습 데이터로부터 변수들 사이의 확률적 의존성 분석을 위해 베이지안 네트워크를 활용하는 경우에 대두될 수 있는 문제점들 중 하나는, 학습된 베이지안 네트워크의 불확실성의 분석에 대한 점이다. 베이지안 네트워크에 존재하는 불확실성을 정의하는 접근 방법에는 여러 가지가 있을 수 있으나, 본 연구에서는 특정 베

이지안 네트워크 내의 한 변수 V_i 가 갖는 조건부 확률분포의 불확실성을 통해 베이지안 네트워크의 불확실성을 다루고자 한다. 즉, 본 연구에서 베이지안 네트워크 B 에 속한 변수 V_i 가 갖는 불확실성은, V_i 가 갖는 조건부 확률분포 $P(V_i|Pa(V_i))$ 가 갖는 불확실성을 통해 표현된다. 만약 V_i 가 갖는 조건부 확률분포의 불확실성이 높다면, G 내부에 포함된 V_i 와 V_i 의 부모 변수들 $Pa(V_i)$ 사이의 네트워크 구조 또한 불확실함을 의미한다.

이러한 베이지안 네트워크의 불확실성에 대한 분석이 유용할 수 있는 경우는 여러 가지가 존재할 수 있으나, 특히 다음과 같은 경우 유용하게 활용되어질 수 있다. 첫째, 학습하고자 하는 베이지안 네트워크의 규모가 큰 상황에서 학습에 필요한 추가 데이터의 습득이 제한되는 경우이다. 이러한 경우, 현재 갖고 있는 학습 데이터만으로 베이지안 네트워크를 학습한 뒤, 어떤 추가 학습 데이터의 습득이 보다 우수한 베이지안 네트워크 학습에 유리한 것인지에 대한 분석이 필요할 수 있다. 예를 들어 유전자 칩 데이터로부터 유전자 사이의 관계를 베이지안 네트워크로 학습하는 경우가 이에 해당한다. 일반적으로 실험 대상인 유전자의 수는 수백 ~ 수천에 이르는 것에 비해, 사용 가능한 학습 데이터(유전자 칩 실험 결과)는 수 개 ~ 수십 개 정도로 극히 제한된 경우가 많다. 또한 유전자 칩 실험의 비용이 비교적 큰 점으로 인해, 추가 실험의 방향을 올바르게 결정하는 것은 중요한

문제가 된다. 두 번째로, 베이지안 네트워크를 이용한 시스템 분석에 있어 주된 관심사가 베이지안 네트워크의 특정 영역에 제한되는 경우이다. 이러한 경우 주 관심의 대상 부분에서의 불확실성을 줄이는 방향으로 추가 학습 데이터를 형성하는 것이 가능하다면 동일한 비용으로 베이지안 네트워크 중 관심 영역에 대해 최적화된 실험이 가능하게 된다.

이러한 베이지안 네트워크의 활용에 있어서 추가 실험 데이터의 습득에 관련된 연구로는 Tong, et. al.의 연구 [1][2]와 같은 베이지안 네트워크의 active learning이 있으며, 베이지안 네트워크에 있는 edge의 entropy와 확률분포의 KL-divergence를 각각 네트워크 구조와 확률분포값의 불확실성으로 해석하는 접근 방법을 취하고 있다. 또한, Pournara, et. al.[3]과 Yoo, et. al.[4]의 연구에서는 그러한 active learning을 적용한 유전자 네트워크의 학습 및 추가 유전자 칩 실험의 제안 방법을 언급하고 있다.

본 연구에서는, 베이지안 네트워크의 불확실성을 다루기 위한 방법으로서 기존 연구와는 달리 확률분포 값의 분산을 이용하여 불확실성을 활용한 접근 방법을 제안한다. 본 논문은 다음과 같이 구성된다. 2장에서는 제안하고자 하는, 확률분포 값의 분산을 이용한 베이지안 네트워크의 불확실성에 대한 정의를 언급한다. 또한 불확실성을 줄이기 위한 제어 실험의 후보를 제안하는 방법을 제안한다. 3장에서는 제안된 제어 실험 후보 방법을 적용하여 베이지안 네트워크의 불확실성이 감소될 수 있다는 점을 간단한 예시를 통해 보이며, 4장에서 본 연구의 요약 및 향후 과제를 제시하도록 한다.

2. 확률분포의 분산을 이용한 불확실성의 정의

2.1 한 변수에 대한 불확실성의 정의

베이지안 네트워크 $B=(G, \theta)$ 가 주어졌다고 가정하자. G 는 주어진 변수의 집합 V 에 포함된 변수들을 노드로 갖는 directed acyclic graph이다. 이 때, V 에 포함된 변수들 중 하나인 V_i 에 대해, V_i 가 갖는 값의 수가 r_i 라고 가정하자. 또한, G 에서의 V_i 의 부모 변수들의 집합을 $Pa(V_i)$ 로 표현하며, V_i 에 대한 부모 변수 값들 조합의 가능한 모든 수를 q_i 라 한다. 즉, q_i 는 다음과 같다.

$$q_i = \prod_{j=1}^{|Pa(V_i)|} r_j, \quad V_j \in Pa(V_i) \quad (1)$$

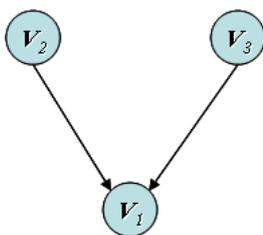


그림 1

	$V_2=0$		$V_2=1$	
	$V_3=0$	$V_3=1$	$V_3=0$	$V_3=1$
$V_1=0$	f_{110}	f_{120}	f_{130}	f_{140}
$V_1=1$	f_{111}	f_{121}	f_{131}	f_{141}

표 1

이 때, θ 에서의 V_i 의 조건부 확률값을 기술하기 위한 확률 값의 표는 $q_i r_i$ 개의 셀을 갖게 된다. 예를 들어, 그림 1과 같은 변수 세 개로 구성된 베이지안 네트워크가 있다고 가정하자. 또한 세 변수 V_1, V_2, V_3 는 $r_1 = r_2 = r_3 = 2$ 인 이진 변수라고 가정하자. 이 때, V_1 이 갖는 조건부 확률 값의 표는 표 1과 같으며, 셀의 수는 $4 \times 2 = 8$ 이 된다.

V_i 가 부모 변수 값들 조합의 k ($\leq q_i$) 번째 경우에 대해 l ($\leq r_i$) 번째 값을 가지는 경우의 확률값을 나타내는 확률변수를 f_{ikl} 이라고 하자. 이 때 $f_{ik1}, f_{ik2}, \dots, f_{ikr_i}$ 가 $a_{ik1}, a_{ik2}, \dots, a_{ikr_i}$ 의 매개변수를 갖는 Dirichlet 분포 $Dir(f_{ik1}, f_{ik2}, \dots, f_{ik(r_i-1)}; a_{ik1}, a_{ik2}, \dots, a_{ikr_i})$ 를 따른다고 가정하자. 이 때, f_{ikl} 의 분산 $v(f_{ikl})$ 은 다음과 같다.

$$v(f_{ikl}) = \frac{a_{ikl}(N - a_{ikl})}{N^2(N + 1)}, \quad N = \sum_{j=1}^{r_i} a_{ikj} \quad (2)$$

이 때, V_i 의 정규화 Dirichlet 분산합계 (normalized sum of Dirichlet variances)를 다음과 같이 정의한다.

정의 1. 정규화 Dirichlet 분산합계 (Normalized sum of Dirichlet variances)

모든 k ($\leq q_i$) 에 대해 $f_{ik1}, f_{ik2}, \dots, f_{ikr_i}$ 가 $a_{ik1}, a_{ik2}, \dots, a_{ikr_i}$ 를 매개변수로 갖는 Dirichlet 분포 $Dir(f_{ik1}, f_{ik2}, \dots, f_{ik(r_i-1)}; a_{ik1}, a_{ik2}, \dots, a_{ikr_i})$ 를 따르는 변수 V_i 가 있다. 이 때, V_i 의 정규화 Dirichlet 분산합계 (normalized sum of Dirichlet variances) $nsDv(V_i)$ 는 다음과 같이 정의된다.

$$nsDv(V_i) = \frac{\sum_{k=1}^{q_i} \sum_{l=1}^{r_i} v(f_{ikl})}{q_i r_i} \quad (3)$$

제안된 V_i 의 정규화 Dirichlet 분산합계는, V_i 가 갖는 일정한 넓이의 확률분포 공간에 존재하는 확률 분산값들의 정규화된 합계이다. 어떤 V_i 의 확률값을 추정하기 위해 일반적으로 관찰된 데이터로부터 그림 1과 같은 주어진 확률적 의존 관계에 V_i 가 갖는 확률값들을 추정하게 된다. 이 때, V_i 가 갖는 부모 변수들의 수가 많고 부모 변수들이 갖는 값의 수가 많을 수록 V_i 의 확률 값 표 안의 셀 수는 늘어나게 되는 반면, 각 셀에 해당하게 되는 학습 데이터의 경우의 수는 그만큼 줄어들게 된다. 또한 어떠한 확률변수의 확률분포를 추정하는 데에 사용되는 관찰된 데이터의 수가 적을수록 추정된 확률분포의 분산은 일반적으로 크게 된다. 따라서 정규화를 하지 않은

분산합계는 부모 변수들의 수나 부모 변수들의 값 조합의 경우의 수에 영향을 받게 되므로, 정규화를 통해 V_i 가 갖는 불확실성을 일반적으로 정의하였다.

2.2 불확실성을 줄이기 위한 추가실험의 제안 방법

앞 2.1장에서는 V_i 에 대한 정규화 Dirichlet 분산합계를 정의함으로써 한 변수 V_i 에 대한 불확실성을 정의하였다. 본 장에서는 V_i 에 대한 불확실성을 줄이기 위한 추가 학습 데이터의 습득에 있어서, $Pa(V_i)$ 에 속한 한 변수 V_p 의 값을 m ($\leq r_p$)로 제한한 경우의 추가 데이터를 얻는 상황을 가정한다. 일정한 수의 추가 데이터를 얻는다고 가정하는 경우, 어떤 V_p 에 대해 어떤 m 으로 조건을 제한한 추가 데이터를 얻는 것이 현재까지 학습되어져 있는 베이저안 네트워크 내에서 V_i 의 불확실성을 효과적으로 줄일 수 있는지를 논의한다.

이를 위해, 먼저 V_i 의 $V_p=m$ 조건부 정규화 Dirichlet 분산합계 (conditional normalized sum of Dirichlet variances) 를 다음과 같이 정의한다.

정의 2. 조건부 정규화 Dirichlet 분산합계 (Conditional normalized sum of Dirichlet variances)

$V_p = m$ ($V_p \in Pa(V_i), 1 \leq m \leq r_p$) 에 해당하는 V_i 의 부모 변수 값 조합의 모든 경우에 대해, V_i 의 확률값을 나타내는 확률변수 f_{inl} ($1 \leq l \leq r_i$) 들의 집합을 F 라 하자. 이 때, V_i 의 $V_p = m$ 조건부의 정규화 Dirichlet 분산합계(Conditional normalized sum of Dirichlet variances) $nsDv(V_i; V_p = m)$ 은 다음과 같이 정의된다.

$$nsDv(V_i; V_p = m) = \frac{\sum_n \sum_{l=1}^{r_i} v(f_{inl})}{q_i^{r_i}}, \quad f_{inl} \in F \quad (4)$$

입력: 베이저안 네트워크 $B=(G, \theta), V_i$
 결과: V_i 의 확률분포 표에서 가장 불확실성이 가장 큰 부분에 해당하는 V_p ($\in Pa(V_i)$)와 m ($1 \leq m \leq r_p$)

Step 1: V_i 의 모든 부모 변수 V_p ($\in Pa(V_i)$) 와 V_p 가 가질 수 있는 값 m ($1 \leq m \leq r_p$) 에 대해, $nsDv(V_i; V_p = m)$ 을 계산한다.

Step 2: 가장 큰 $nsDv(V_i; V_p = m)$ 값을 보이는 V_p 와 m 을 찾는다.

알고리즘 1

일반적으로, 추정된 확률분포의 분산이 큰 이유 중 하나는 추정에 사용된 학습 데이터의 양이 적기 때문이며, 그러한 큰 분산 값은 확률값의 확신도를 떨어뜨리게 된다[5]. 따라서, 알고리즘 1과 같이 V_i 에 대해 가장 높은 조건부 정규화 Dirichlet 분산합계를 보이는 V_p 와 m 을 찾은 후, 불확실성이 가장 컸던 $V_p=m$ 의 경우에 대한 추가

데이터를 요구하여 V_i 의 불확실성을 효과적으로 줄이고자 한다.

3. 정규화 Dirichlet 분산 합계를 이용한 분산 감소 예

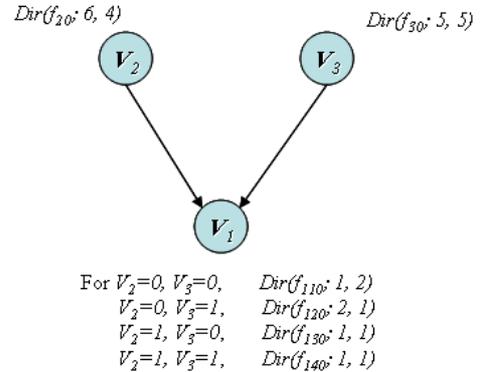


그림 2

	$V_2=0$		$V_2=1$	
	$V_3=0$	$V_3=1$	$V_3=0$	$V_3=1$
$V_1=0$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{2}$
$V_1=1$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$

표 2. V_i 의 조건부 확률 값 표

	$V_2=0$		$V_2=1$	
	$V_3=0$	$V_3=1$	$V_3=0$	$V_3=1$
$V_1=0$	0.056	0.056	0.083	0.083
$V_1=1$	0.056	0.056	0.083	0.083

표 3. V_i 확률변수들의 분산 값 표

앞 장에서 제안한 정규화 Dirichlet 분산합계를 이용한 분산 감소 방법의 예를 보이기 위해, 그림 2와 같이 10개의 학습 데이터를 이용하여 조건부 확률분포를 추정된 베이저안 네트워크를 예시로 사용한다. 이 때 V_i 의 조건부 확률 값 표는 표 2와 같으며, 각 확률값은 해당 확률변수 f_{ikl} 의 기댓값 $E(f_{ikl})$ 이 된다. 각 경우에 대해 V_i 의 확률변수 f_{ikl} 의 분산 $v(f_{ikl})$ 을 표시하면 표 3과 같다. 이 때, V_i 의 정규화 Dirichlet 분산 합계를 계산하면 다음과 같다.

$$nsDv(V_i) = \frac{1}{8}(0.056 \times 4 + 0.083 \times 4) = 0.0695$$

V_i 에 알고리즘 1의 절차를 적용하는 과정에서 계산되는 V_2, V_3 각각에 대해 0과 1의 값을 가지는 경우 조건부의 V_i 의 조건부 정규화 Dirichlet 분산 합계를 계산하면 다음과 같다.

$$nsDv(V_1; V_2 = 0) = \frac{1}{8}(0.056 \times 4) = 0.028$$

$$nsDv(V_1; V_2 = 1) = \frac{1}{8}(0.083 \times 4) = 0.0415$$

$$nsDv(V_1; V_3 = 0) = \frac{1}{8}(0.056 \times 2 + 0.083 \times 2) = 0.03475$$

$$nsDv(V_1; V_3 = 1) = \frac{1}{8}(0.056 \times 2 + 0.083 \times 2) = 0.03475$$

계산에 의해 V_1 의 불확실성을 가장 크게 하는 부분은, 조건부 정규화 Dirichlet 분산 합계가 가장 큰 $V_2=1$ 인 경우임을 알 수 있다. 이러한 이유 중의 하나는, $V_2=1$ 인 경우에 확률값의 추정을 위해 사용된 학습 데이터의 수가 4개로, 다른 경우에 비해 가장 적었기 때문이기도 하다.

이제 그림 2에 주어진 베이지안 네트워크에 다음과 같은 네 가지 추가 조건부 데이터를 각각 더해 확률분포를 새로이 추정하는 경우를 고려하자.

- $V_2=0$ 인 경우의 추가 데이터 D_1
 $(V_1, V_2, V_3) = (0, 0, 0) (1, 0, 0) (0, 0, 1) (1, 0, 1)$
- $V_2=1$ 인 경우의 추가 데이터 D_2
 $(0, 1, 0) (1, 1, 0) (0, 1, 1) (1, 1, 1)$
- $V_3=0$ 인 경우의 추가 데이터 D_3
 $(0, 0, 0) (0, 1, 0) (1, 0, 0) (1, 1, 0)$
- $V_3=1$ 인 경우의 추가 데이터 D_4
 $(0, 0, 1) (0, 1, 1) (1, 0, 1) (1, 1, 1)$

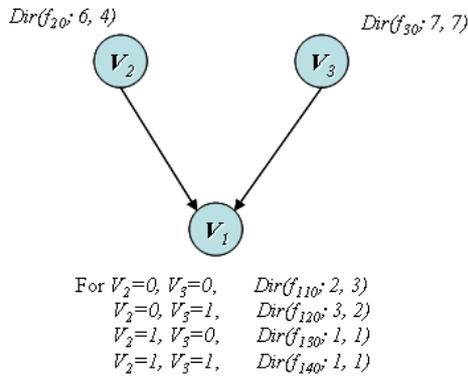


그림 3. D_1 으로 갱신한 결과

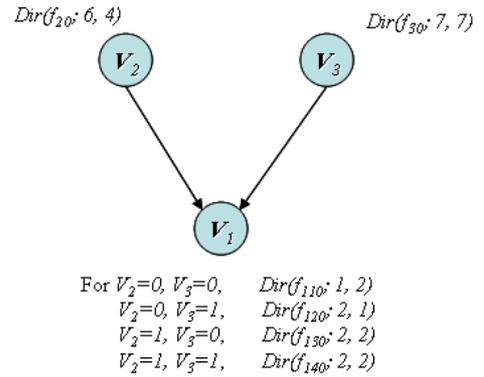


그림 4. D_2 로 갱신한 결과

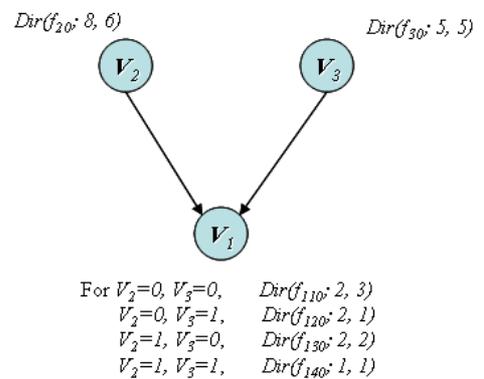


그림 5. D_3 로 갱신한 결과

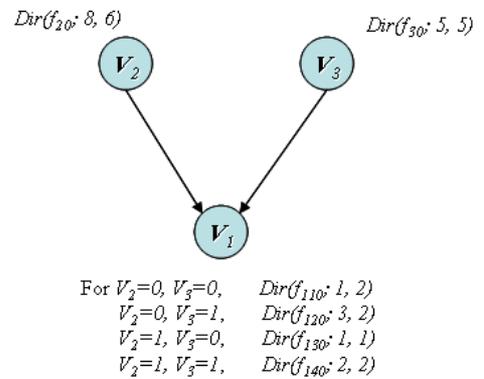


그림 6. D_4 로 갱신한 결과

	$V_2=0$		$V_2=1$	
	$V_3=0$	$V_3=1$	$V_3=0$	$V_3=1$
$V_1=0$	0.04	0.04	0.083	0.083
$V_1=1$	0.04	0.04	0.083	0.083

표 4. D_1 으로 갱신한 후의 V_1 확률변수들의 분산 값 표

	$V_2=0$		$V_2=1$	
	$V_3=0$	$V_3=1$	$V_3=0$	$V_3=1$
$V_1=0$	0.056	0.056	0.05	0.05
$V_1=1$	0.056	0.056	0.05	0.05

표 5. D_2 로 갱신한 후의 V_1 확률변수들의 분산 값 표

	$V_2=0$		$V_2=1$	
	$V_3=0$	$V_3=1$	$V_3=0$	$V_3=1$
$V_1=0$	0.04	0.056	0.05	0.083
$V_1=1$	0.04	0.056	0.05	0.083

표 6. D_3 로 갱신한 후의 V_1 확률변수들의 분산 값 표

	$V_2=0$		$V_2=1$	
	$V_3=0$	$V_3=1$	$V_3=0$	$V_3=1$
$V_1=0$	0.056	0.04	0.083	0.05
$V_1=1$	0.056	0.04	0.083	0.05

표 7. D_4 로 갱신한 후의 V_1 확률변수들의 분산 값 표

D_1, D_2, D_3, D_4 는 각각 공평한 비교를 위하여 각 f_{in} 에 대해 한 개 씩의 추가 데이터를 배정함으로써 만들어졌다. 이 네 개의 추가 조건부 데이터를 이용하여 확률변수의 추정을 갱신한 네 가지 베이지안 네트워크의 결과는 각각 그림 3, 4, 5, 6에 해당하며 각 경우에 대한 V_1 의 확률변수들의 분산 값 표는 표 4, 5, 6, 7과 같다. 이때 확률변수 값의 갱신은 Cooper, et. al.이 사용한, 조건부에 해당하는 변수의 확률분포는 갱신하지 않는 방법 [6]을 사용하였다. D_1, D_2, D_3, D_4 네 가지 경우 추가 데이터로 갱신한 후의 V_1 의 정규화 Dirichlet 분산 합계는 각각 다음과 같다.

$$nsDv(V_1|D_1) = \frac{1}{8}(0.04 \times 4 + 0.083 \times 4) = 0.0615$$

$$nsDv(V_1|D_2) = \frac{1}{8}(0.056 \times 4 + 0.05 \times 4) = 0.053$$

$$nsDv(V_1|D_3) = \frac{1}{8}(0.04 \times 2 + 0.056 \times 2 + 0.05 \times 2 + 0.083 \times 2) = 0.05725$$

$$nsDv(V_1|D_4) = \frac{1}{8}(0.56 \times 2 + 0.04 \times 2 + 0.083 \times 2 + 0.05 \times 2) = 0.05725$$

앞서 계산했던 원래의 $nsDv(V_1)=0.0695$ 와 비교했을 때, V_1 에 대한 조건부 정규화 Dirichlet 분산합계를 이용하여 추정한 $V_2=1$ 의 경우에 대해 조건부 추가 데이터를 사용했을 때 V_1 의 불확실성이 가장 많이 줄어든 것을 알 수

있다. 이를 통해 본 연구에서 제안한 정규화 Dirichlet 분산합계의 활용이 베이지안 네트워크의 불확실성 감소에 유용할 수 있음을 알 수 있다.

4. 결론 및 향후 과제

본 논문에서는 베이지안 네트워크의 불확실성을 다루기 위한 방법으로, 변수의 확률 값에 대한 확률변수의 분산을 이용하는 방법을 제안하였다. 베이지안 네트워크 내의 특정 변수 V_i 에 대해 정규화 Dirichlet 분산합계를 변수의 조건부 확률분포의 불확실성의 척도로서 제안하였으며, 특정 부모 변수 값 조합의 경우에 대한 조건부 정규화 Dirichlet 분산합계를 제안하였다. 또한 제안된 조건부 정규화 Dirichlet 분산합계를 이용한, 추가 데이터 습득의 방향을 제시하는 방법 또한 제안되었다. 예로 든 베이지안 네트워크에 대해, 제안된 베이지안 네트워크의 불확실성 척도 방법이 유용할 수 있음을 보였다.

향후 과제로는, 보다 일반적인 베이지안 네트워크 구조에서의 제안된 정규화 Dirichlet 분산합계의 특성 파악이 필요하다. 또한 추가 실험을 제안하기 위한 조건부 정규화 Dirichlet 분산합계의 활용은 현 단계에서는 최적화된 결과를 보장하지 못하는 단점이 있다. 향후 이러한 점들에 대한 보완이 필요하며, 추후 베이지안 네트워크의 active learning 방법 또한 연구 대상이 될 수 있다.

Acknowledgement

본 논문은 과학기술부 국가지정연구실사업(2005-01450)의 지원으로 수행되었음. 연구시설은 정문술 바이오정보전자센터의 도움을 받았음.

참고문헌

- [1] Simon Tong and Daphne Koller, "Active Learning for Parameter Estimation in Bayesian Networks", NIPS, 2001.
- [2] Simon Tong and Daphne Koller, "Active Learning for Structure in Bayesian Networks", Proceedings of International Joint Conference on Artificial Intelligence, 2001.
- [3] Iosifina Pournara and Lorenz Wernisch, "Reconstruction of gene networks using Bayesian learning and manipulation experiments", Bioinformatics, Vol. 20, no. 17, pp. 2934-2942, 2004.
- [4] Changwon Yoo and Gregory F. Cooper, "An evaluation of a system that recommends microarray experiments to perform to discover gene-regulation pathways", Artificial Intelligence in Medicine, Vol. 31, pp. 169-182, 2004.
- [5] Richard E. Neapolitan, "Learning Bayesian Networks", Pearson Prentice Hall, NJ, 2004.
- [6] Gregory F. Cooper and Changwon Yoo, "Causal Discovery From a Mixture of Experimental and

Observational Data," in Proceedings of the Fifteenth Conference of Uncertainty in Artificial Intelligence, San Mateo, California, 1999.