

베이지안 네트워크에 기반한 스마트 홈에서의 상황인식 기법개발

Context-aware application for smart home based on Bayesian network

정우용¹, 김은태², 김동연³

¹ 서울시 서대문구 연세대학교 전기전자공학과

E-mail: wychung@yonsei.ac.kr

² 경기도 안성시 한경대학교 전자공학과

E-mail: dykim@hknu.ac.kr

³ 서울시 서대문구 연세대학교 전기전자공학과

E-mail: etkim@yonsei.ac.kr

요 약

본 논문은 스마트 홈에서 베이지안 네트워크에 기반을 둔 보편성을 가지는 상황인식 시스템의 구현방법을 제안한다. 상호정보를 사용하여 베이지안 네트워크의 구조 학습을 하고, 보다 효율적인 데이터 처리를 위해서 퍼지 클러스터링을 사용하는 방법을 도입한다. 마지막으로 시뮬레이터를 통하여 자료 취득 및 상황인식의 결과를 보인다.

Key Words : context-aware, Bayesian network, smart home, Fuzzy clustering, mutual information

1. 서 론

컴퓨팅 디바이스와 네트워크 기술의 발전은 우리가 주변으로부터 다양한 정보를 보다 쉽게 제공받을 수 있는 환경을 제공하였다. 이러한 정보들은 다양한 사물들이 상황정보를 인식하고 판단하여 우리에게 보다 능동적이고 유용한 서비스를 제공할 수 있는 기반이 된다. 하지만 이러한 상황인식 서비스는 대부분이 사용자나 개발자가 직접 명시한 사항에 대해서만 가능한 것이 현실이다. 따라서 본 논문에서는 스마트 홈에서 베이지안 네트워크에 기반을 둔 보편성을 가지는 상황인식 시스템을 제안한다.

베이지안 네트워크는 많은 변수들 간의 확률 관계를 비교적 축약된 형태로 표현하는데 유용한 모델로 확률적 추론, 예측, 의사 결정 등에 적합한 모델이다. 이러한 특성으로 확률적 성격이 강한 상황인식 연구 분야에서 베이지안 네트워크가 많이 사용되고 있다 [1][2]. 본 논문에서도 이러한 베이지안 네트워크를 보다 쉽게 구성할 수 있도록 상호정보를 사용한 방법을 제안한다. 또한 베이지안 네트워크를 구현하는데 보다 효율적으로 하기 위하여 각 노드들이 가질 수 있는 값을 이산화하는데, 이산화의 방법으로 클러스터링을 통한 보다 효과적인

이산화 방법을 제시한다.

본 논문의 구성은 다음과 같은 순서로 구성 되어 있다. 제2장에서는 본 논문의 기본이 되는 베이지안 네트워크에 관하여 간략히 기술한다. 제3장에서는 베이지안 네트워크를 구성하는 과정에서 상호정보를 이용한 방법을 소개하고, 연속데이터를 이산화하는 방법으로 퍼지 클러스터링을 사용한 방법을 제시한다. 제4장에서는 제안된 알고리즘을 통한 시뮬레이션 결과를 통하여 유효성에 관하여 기술한다. 마지막으로 제5장에서는 결론으로 마무리한다.

2. 베이지안 네트워크

베이지안 네트워크는 그래프 모델 중에서 방향성 비순환 그래프(Directed Acyclic Graph, DAG)에 기반을 둔 확률적 모델이다[3]. 방향성 비순환 그래프 모델에서 각 노드를 연결하는 아크는 직접적인 인과 관계를 나타내는 것으로 생각할 수 있기 때문에 때로는 베이지안 네트워크를 인과 네트워크라고 부르기도 한다. 방향성 비순환 그래프는 크게 유한개의 노드들의 집합(V)과 노드 간을 연결하는 경로(E)로 나누어진다.

이 방향성 비순환 그래프의 각 노드는 확률 변수를 나타내며, 조건부 확률표가 존재한다. 또한 각 노드 V_i 는 이 노드의 부모가 주어졌을 때 V_i 의 후손이 아닌 노드들과 조건부 독립이 된다고 할 때 Markov 조건이 성립된다. 때로는 베이지안 네트워크에서 조건부 독립이 드러날 수 있도록 원인과 결과를 연관 시킬 수 있는 경우가 있는데, 이처럼 인과 관계에 대한 직관적인 개념으로 베이지안 네트워크를 구성하면 내재된 조건부 독립 가정에 적합한 네트워크가 된다. 이와 같은 성질을 통해서 그림 1과 같은 베이지안 네트워크의 예는 다음과 같이 식 (1)로 표현된다.

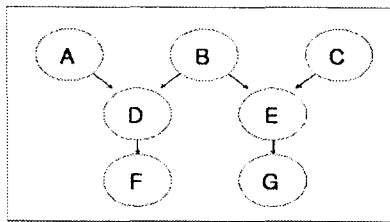


그림 1. 베이지안 네트워크 구조 예

$$P(A, B, C, D, E, F, G) = P(A)P(B)P(C) P(D|A, B)P(E|B, C)P(F|D)P(G|E) \quad (1)$$

이와 같이 V_1, V_2, \dots, V_k 를 베이지안 네트워크를 이루고 있는 노드라고 할 때, 이 네트워크가 조건부 독립이라는 가정 하에 이 네트워크의 모든 노드에 대한 결합 확률을 일반화하면 식 (2)와 같이 주어진다.

$$P(V_1, V_2, \dots, V_n) = \prod_{i=1}^n P(V_i | \pi_i) \quad (2)$$

여기서 π_i 는 그래프에서 V_i 의 부모집합이다. 식 (2)는 베이지안 네트워크가 의미하는 부분 순서와 일관성을 유지하는 연쇄 법칙 순서를 사용하는 모든 변수의 결합 확률을 구하는 연쇄 법칙에 조건부 독립을 적용함으로써 쉽게 유도할 수 있다.

3. 상황인식 시스템

본 논문의 상황인식 시스템은 베이지안 네트워크를 통하여 일정한 학습을 한 후에 센서 정보로부터 거주자의 행동을 파악하는 것이다. 3장에서는 이러한 상황인식 시스템을 구성하기 위해서 베이지안 네트워크와 데이터의 처리하는 과정을 보다 상세하게 설명한다.

3.1 상호정보를 통한 구조학습

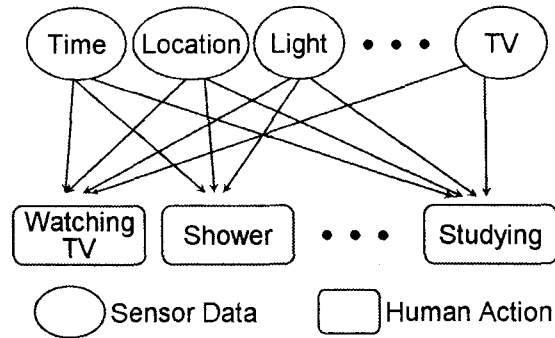


그림 2. naive Bayesian network

상황인식 시스템의 기본이 되는 베이지안 네트워크 구조의 학습은 각 센서들이 상호 독립이라 가정하고 수행하였다. 따라서 우리는 그림 2와 같이 베이지안 네트워크를 복잡한 구조가 아닌 단순한 형태의 베이지안 네트워크 (naive Bayesian network)을 구성하였다 [4]. 하지만 센서의 종류가 다양한 환경에서는 이러한 모든 정보를 사용할 수가 없기 때문에 중요한 부분을 추출하여야 한다. 즉, 행동이 나타나는데 중요한 역할을 하는 센서정보들을 추출하여야 하는 것이다.

이와 같은 과정은 스마트 홈에서 거주자의 활동을 관찰하고 그에 따른 센서정보의 자유도를 통하여 구현하였다. 즉, 어떤 활동을 할 때 어떤 센서들이 동작을 하고 그 중에서 가장 많이 영향을 주는 것은 어떤 것인지를 확인하여 구조를 결정하는 것이다. 거주자의 활동을 관찰하면서 센서의 자유도 정보를 확인하는 것을 식으로 표현하면 다음과 같다.

$$\begin{aligned} I(Sensor; Act) &= H(Sensor) - H(Sensor|Act) \\ &= H(Sensor) - H(Sensor, Act) - H(Act) \\ &= \sum_{s \in Sensor} \sum_{a \in Act} p(s, a) \log_2 \left(\frac{p(s, a)}{p(s)p(a)} \right) \end{aligned} \quad (3)$$

여기서 $p(s, a)$ 는 이산 랜덤 변수 s, a 의 결합 확률분포 함수이며, 상호정보 I 는 각 활동과 센서들 간의 상관관계를 나타낸다고 할 수 있다.

본 논문에서는 위와 같은 상호정보 값을 사용하여 일정한 수의 연결을 유지하는 형태로 베이지안 네트워크를 구성하였다. 일반적으로 노드의 수가 많을수록 더욱 정밀한 추론이 가능하지만 학습 샘플의 수가 더 많이 필요하게 될 뿐만 아니라 너무 특정화되는 경향이 있다. 따라서 가능한 노드간의 연결이 최소화하여 보다 효율적으로 구성하였다.

3.2 퍼지 클러스터링을 통한 이산화

스마트 홈에서 얻을 수 있는 다양한 정보들 중에서는 시간이나 위치, 온도 등의 연속된 데이터가 존재한다. 이러한 정보를 보다 효율적으로 사용하기 위해서는 이산화 과정이 필요하다. 이산화의 가장 간단한 방법은 연속된 값을 균일한 값으로 나누는 것이다. 하지만 각 집마다 정보기기의 위치가 모두 다르고, 거주자의 활동시간도 다르기 때문에 균일한 간격으로 이산화를 하면 필요가 없는 정보를 추출하거나 잘못된 결과를 가져올 수 있다.

따라서 본 논문에서는 퍼지 클러스터링을 사용하여 스마트 홈에서의 이산화 과정을 제안한다. 이산화과정은 분할클러스터링 알고리즘의 하나로 Bezdek에 의해 제안된 FCM (fuzzy c means)를 사용하여 각 정보들을 데이터가 아닌 클러스터의 중심들로 표현하였다 [5]. FCM에서 주어진 데이터 집합이 $X = \{x_1, x_2, \dots, x_n\}$ 이고 퍼지 클러스터링의 중심 벡터가 $V = \{v_1, v_2, \dots, v_c\}$ 일 때, 목적 함수는 각 데이터 x_j 와 각 클러스터 중심 v_i 과의 거리와 클러스터 소속 정도(membership) 값으로 정의된다.

$$J_{FCM}(X, U, V) = \sum_{j=1}^n \sum_{i=1}^c (\mu_{ij})^m d^2(x_j, v_i) \quad (4)$$

여기서 μ_{ij} 는 x_i 와 x_j 번째 클러스터에 대한 소속 정도를 나타내며 (c×n)의 소속행렬 $U = [\mu_{ij}]$ 의 원소이다. $d^2(x_j, v_i)$ 는 유클리디안 거리(Euclidean distance)의 제곱이며, 매개 변수 m 은 각 데이터의 소속 정도에 대한 퍼지 값을 나타내며 1 보다 큰 값을 사용한다.

$J_{FCM}(X, U, V)$ 을 U 에 대하여 최소화 하면 다음과 같이 μ_{ij} 를 구할 수 있다.

$$\mu_{ij} = \frac{\left\{ \frac{1}{d^2(x_j, v_i)} \right\}^{\frac{1}{m-1}}}{\sum_{k=1}^c \left\{ \frac{1}{d^2(x_j, v_k)} \right\}^{\frac{1}{m-1}}} \quad (5)$$

$J_{FCM}(X, U, V)$ 에 식 (5)의 μ_{ij} 을 적용하면 다음과 같이 식 (6)으로 간단히 표현될 수 있다.

$$J_{FCM}(X, V) = \sum_{j=1}^n \left(\sum_{i=1}^c d^{1-m}(x_j, v_i) \right)^{1-m} \quad (6)$$

이와 같이 FCM을 통하여 연속된 데이터에 대하여 이산화를 수행하였는데 베이지안 네트워크의 조건부 확률표를 보다 간단하게 구성할

수 있도록 클러스터의 수를 비교적 작게 설정하였다. 시간의 경우에는 6-9개의 클러스터로 구성하였고, 위치의 경우에는 그림 3과 같이 9-16개의 클러스터로 구성하였다.

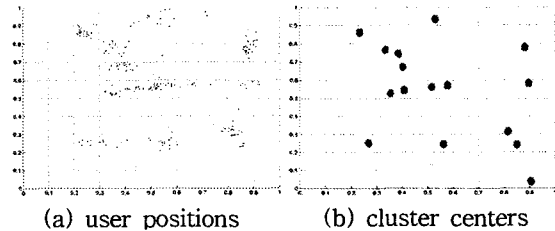


그림 3. FCM을 이용한 이산화 결과

4. 시뮬레이션 및 결과 고찰

4.1 상황인식 시뮬레이터

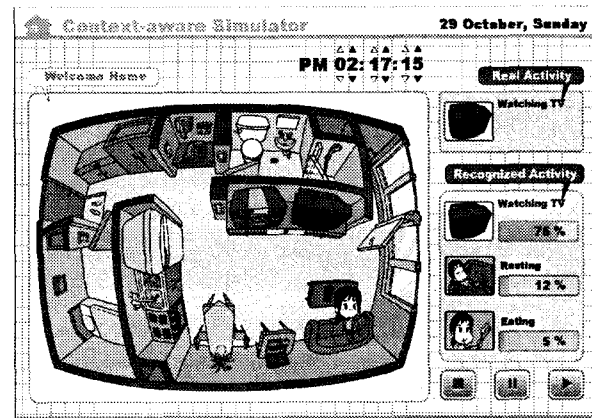


그림 4. 상황인식 시뮬레이터

사용자에 대한 센서자료와 활동자료는 그림 4와 같은 플래시에 기반을 둔 인터페이스를 만들어서 수집하였다. 수집 과정은 가상의 집에서 거주자가 활동하고, 각 활동에 대한 각종 정보는 데이터베이스에 저장 된다. 이때 센서들은 스위치를 켜게 되면 자동으로 저장되지만, 사용자의 활동정보는 직접 입력해주어야 하며, 선택할 수 있는 활동의 종류는 한 가지가 된다. 그러면 활동들은 각각 할당된 최소시간이 기본적으로 지나게 되며, 그 후에도 같은 행동을 하게 되면 사용자가 임의의 시간을 설정하여 그 내용이 데이터베이스에 저장될 수 있도록 한다.

본 시뮬레이션에서는 사용자가 거부감을 가지지 않을 정도의 시간인 10일간의 비슷한 일상의 시뮬레이션 내용을 저장하여 10개의 폴드(fold)로 구성하여 10겹 교차검증(10-fold cross-validation)을 수행하였다.

4.2 실험결과

각 노드간의 연결 수를 임계값으로 사용하여 노드간의 연결을 수정하였으며 이때의 성능을 비교하면 다음과 같다 (표 1, 그림 5).

표 1. 상황인식 성능 비교

노드간 연결수	상황인식의 정확도		노드간 연결수	상황인식의 정확도	
	균일분포	클러스터링		균일분포	클러스터링
61	0.7634	0.8384	71	0.7723	0.8295
62	0.7634	0.8260	72	0.7776	0.8313
63	0.7669	0.8293	73	0.7776	0.8447
64	0.7669	0.8302	74	0.7776	0.8420
65	0.7669	0.8376	75	0.7803	0.8536
66	0.7652	0.8322	76	0.7839	0.8367
67	0.7652	0.8286	77	0.7910	0.8509
68	0.7652	0.8438	78	0.7910	0.8311
69	0.7643	0.8170	79	0.7892	0.8402
70	0.7643	0.8465	80	0.7892	0.8322

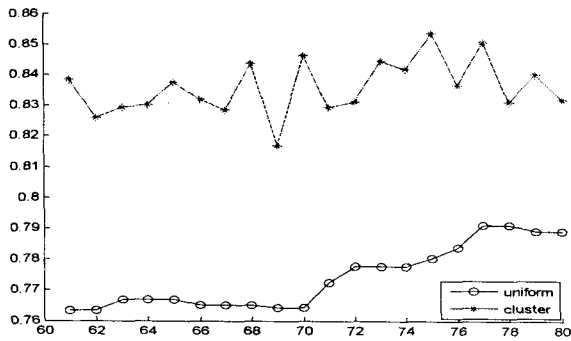


그림 5. 상황인식 성능 비교

그림 5를 살펴보면 클러스터를 사용한 이산화의 경우에 보다 좋은 효과를 나타내는 것을 알 수 있으며, 베이지안 네트워크의 구조를 단순화 시켰음에도 비교적 좋은 성능을 나타내는 것을 알 수 있다. 위의 상황인식을 각 활동에 따라 살펴보면 그림 6과 같다.

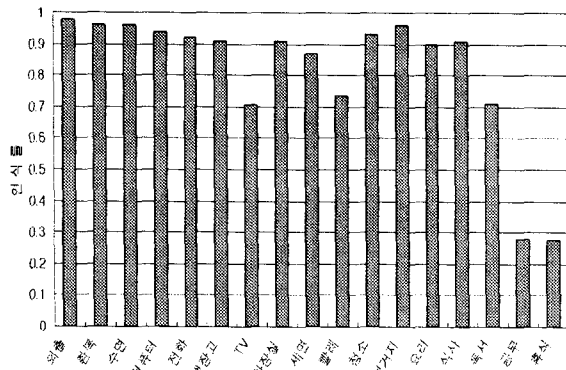


그림 6. 각 활동에 따른 인식률

그림 6에서 살펴볼 수 있듯이 제안된 시스템에서 대부분의 경우에는 잘 인식하지만 모호성이 있는 휴식이나 독서, TV시청과 같은 경우에는 비교적 낮은 인식률을 가진다.

5. 결 론

본 논문에서는 상호정보를 이용한 구조학습을 통한 베이지안 네트워크의 구성과 클러스터링을 통한 이산화와를 사용하여 스마트 홈에서 센서 네트워크로부터의 상황인식을 할 수 있는 시스템을 제안하였다. 시뮬레이션 결과에서도 확인할 수 있듯이 센서정보를 통하여 사람이 직접보고서 판단할 수 있는 경우에는 제안된 방법으로도 쉽게 상황이 인식될 수 있었다. 즉, 제안된 방법은 개발자가 모든 사항을 관찰하고 상황인식을 할 수 있는 시스템을 구축하는 것과 비슷한 결과를 나타낸다. 이것은 상황인식에서 자동화할 수 있는 부분이 더욱 늘어나게 된 것을 의미한다.

한편 상황인식을 하는 절차나 상황인식을 사용하여 서비스를 제공할 때 편의성과 효율성만을 생각하는 것이 아니라 불편함, 간섭, 보안과 같은 사항도 같이 고려하여야 할 것이다.

감사의 글

이 논문은 산업자원부 성장동력기술개발사업 (Hybrid Intelligent Service Station 기술개발)의 지원에 의해 연구되었음

참 고 문 헌

[1] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home setting using simple and ubiquitous sensors," in Proceedings of PERVASIVE 2004, vol. LNCS 3001, pp.158-175

[2] A. Ranganathan, J. Al-Muhtadi, R.H. Campbell, "Reasoning about uncertain contexts in pervasive computing environments," IEEE Trans. on Pervasive Computing, Vol. 3, No. 2, pp. 62-70, 2004

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

[4] F. Fleuret, "Fast binary feature selection with conditional mutual information," The Journal of Machine Learning Research, vol. 5, Dec 2004