

FCM 이산화를 이용한 스마트 홈에서 행동 모델링

Intelligent Modeling of User Behavior based on FCM Quantization for Smart home

정우용, 이제헌, 윤숙현, 조영완, 김은태*

(Woo-yong Chung, Jae-hun Lee, Suk-hyun Yon, Young-wan Cho, and Eun-tai Kim)

Abstract : In the vision of ubiquitous computing environment, smart objects would communicate each other and provide many kinds of information about user and their surroundings in the home. This information enables smart objects to recognize context and to provide active and convenient services to the customers. However in most cases, context-aware services are available only with expert systems. In this paper, we present generalized activity recognition application in the smart home based on a naive Bayesian network(BN) and fuzzy clustering. We quantize continuous sensor data with fuzzy c-means clustering to simplify and reduce BN's conditional probability table size. And we apply mutual information to learn the BN structure efficiently. We show that this system can recognize user activities about 80% accuracy in the web based virtual smart home.

Keywords : activity recognition, context-awareness, Bayesian network, smart home, fuzzy clustering

I. 서론

컴퓨팅 디바이스와 네트워크 기술의 발전은 우리가 주변으로부터 다양한 정보를 보다 쉽게 제공받을 수 있는 환경을 제공하였다. 이러한 환경을 통하여 축적된 데이터는 다양한 사물들이 환경의 변화와 함께 거주자의 활동을 인식할 수 있도록 한다. 즉, 서비스 공급자가 보다 편안하고 능동적인 콘텐츠를 제공할 수 있게 된다. 이와 같은 상황인식 및 활동인식에 관련된 연구는 관련된 연구로서는 사용자에게 걸려온 전화를 사용자와 가까운 위치에 있는 전화로 착신해주는 call forwarding[1], 간단한 수많은 센서를 집의 다양한 곳에 부착하여 거주자의 활동 및 가정의 정보를 파악하는 활동인식[2] 등이 있다. 하지만 이러한 상황인식 시스템의 대부분이 사용자나 개발자가 직접 명시한 사항에 대해서만 인식이 가능한 것이 현실이다.

따라서 본 논문에서는 베이지안 네트워크 및 퍼지 클러스터링을 기반으로 다양한 환경에 스스로 적용할 수 있는 거주자의 활동 인식 기법에 대하여 제안한다. 제안된 기법은 환경 정보를 FCM 클러스터링을 기반으로 효율적으로 이산화하고, 이를 바탕으로 베이지안 네트워크를 학습시키고 추론 시키는 방법이다. 여기서 베이지안 네트워크는 많은 변수들 간의 확률관계를 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[2,3].

또한, 가정에서 사용되는 수많은 센서들은 기존의 방법으로 베이지안 네트워크를 학습할 수가 없게 된다. 따라서 본 논문에서는 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 구성되어 있다. 제2장에서는 본 논문의 기본이 되는 베이지안 네트워크에 관하여 간략히 기술한다. 제3장에서는 FCM 이산화에 기반한 베이지안 네트워크를 구성하는 과정으로 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.

II. 베이지안 네트워크

먼저 활동인식을 위한 추론방법인 베이지안 네트워크에 대하여 살펴본다. 베이지안 네트워크는 그래프 모델 중에서 방향성 비순환 그래프(Directed Acyclic Graph, DAG)에 기반을 둔 확률적 모델이다[4]. 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 노드(V_i)는 확률변수를 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. 때

* 책임저자(Corresponding Author)

논문접수 : 2007. 1. 25., 채택확정 : 2007. 2. 17.

정우용, 이제헌, 윤숙현, 김은태 : 연세대학교 전기전자공학부
(wychung@yonsei.ac.kr/aznable17@yonsei.ac.kr/dbstnrgus@yonsei.ac.kr/etkim@yonsei.ac.kr)

조영완 : 서경대학교 컴퓨터공학과(ywcho@skuniv.ac.kr)

※ 본 논문은 산업자원부 성장동력기술개발사업(Hybrid Intelligent Service Station 기술개발)에서 지원하여 연구하였음.

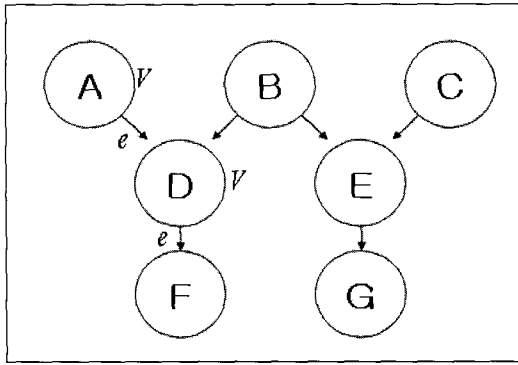


그림 1. 베이저안 네트워크 구조 예.
Fig. 1. An example of Bayesian network structure.

로는 베이저안 네트워크에서 조건부 독립이 드러날 수 있도록 원인과 결과를 연관 시킬 수 있는 경우가 있는데, 이처럼 인과 관계에 대한 직관적인 개념으로 베이저안 네트워크를 구성하면 내재된 조건부 독립 가정에 적합한 네트워크가 된다. 이와 같은 성질을 통해서 그림 1과 같은 베이저안 네트워크의 예는 다음과 같이 (1)로 표현된다.

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

이와 같이 V_1, V_2, \dots, V_k 를 베이저안 네트워크를 이루고 있는 노드라고 할 때, 이 네트워크가 조건부 독립이라는 가정 하에 이 네트워크의 모든 노드에 대한 결합 확률을 일반화하면 (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. (2)는 베이저안 네트워크가 의미하는 부분 순서와 일관성을 유지하는 연쇄 법칙 순서를 사용하는 모든 변수의 결합 확률을 구하는 연쇄 법칙에 조건부 독립을 적용함으로써 쉽게 유도할 수 있다.

III. FCM 기반 활동인식 시스템

본 논문의 활동인식 시스템은 축적된 거주자의 활동과 센서들의 정보를 통하여 베이저안 네트워크를 학습하는 단계와 이를 바탕으로 거주자의 활동을 추론하는 단계로 구성되어있다. 3장에서는 이러한 활동인식 시스템을 구성하기 위하여 베이저안 네트워크를 구성하는 방법과 베이저안 네트워크에 사용될 데이터의 처리하는 과정을 보다 상세하게 설명한다.

1. 상호정보를 통한 베이저안 네트워크의 구조학습

베이저안 네트워크의 구조는 다양한 형태가 있다. 하지만 가정에서의 스마트 디바이스에 대한 존속관계를 알기가 힘들므로 활동인식 시스템의 베이저안 네트워크 구조의 학습은 각 센서들이 상호 독립이라 가정하고 수행하였다. 따라서 베이저안 네트워크를 복잡한 구조가 아닌 단순한 형태의 베이저안 네트워크(naive Bayesian network)을 구성할

수 있게 된다[5]. 하지만 다양한 종류의 센서가 많은 환경에서는 단순 베이저안 네트워크임도 불구하고 조건부 확률표가 너무 커지게 된다. 이와 같은 문제를 해결하기 위해서 각 노드간의 연관성을 파악하여 서로 영향을 많이 주는 부분을 추출하여야 한다. 즉, 행동이 나타나는데 중요한 역할을 하는 센서 정보들을 추출하여 각 노드간의 연결을 결정해 주어야 하는 것이다.

이처럼 각 노드간의 상관성을 파악하기 위한 방법으로는 [6]에서와 같이 상호정보를 사용할 수 있다. 이 방법은 스마트 홈에서 거주자의 활동을 관찰하고 그에 따른 센서정보의 자유도를 통하여 구현하는 것이다. 즉, 어떤 활동을 할 때 어떤 센서들이 동작을 하고 그 중에서 가장 많이 영향을 주는 것은 어떤 것인지 확인하여 구조를 결정하는 것이다. 거주자의 활동을 관찰하면서 센서의 자유도 정보를 확인하는 것을 식으로 표현하면 다음 (3)과 같다[7].

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

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

본 논문에서는 [6]에서와 같이 위와 같은 상호정보 값을 사용하여 일정한 수의 연결을 유지하는 형태로 베이저안 네트워크를 구성하였다. 일반적으로 노드간 연결수가 많을수록 더욱 정밀한 추론이 가능하지만 학습 샘플의 수가 더 많이 필요하게 될 뿐만 아니라 너무 특정화되는 경향이 있다. 또한 연결이 많게 되면 확률분포표가 너무 크게 되기 때문에 더 많은 저장 공간이 필요하게 된다. 따라서 가능한 노드간의 연결이 최소화하여 보다 효율적으로 구성하였다.

2. Fuzzy c-Means(FCM)를 통한 이산화

스마트 홈에서 얻을 수 있는 다양한 정보들 중에서는 시간이나 위치, 온도 등의 연속된 데이터가 존재한다. 이러한 정보를 베이저안 네트워크에서 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 필요한 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[8]. FCM에서 주어진 데이터 집합이 $X = \{x_1, x_2, \dots, x_n\}$ 이고 퍼지 클러스터링의 중심 벡터가 $V = \{v_1, v_2, \dots, v_n\}$ 일 때, 목적 함수는 각 데이터 x_j 와 각 클러스터 중심 v_i 와의 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 소속정도를 나타내며 $(c \times n)$ 의 소속행렬 $U = [\mu_{ij}]$ 의 원소이다. v_i 는 클러스터센터이며, $d^2(x_j, v_i)$ 는 유클리디안 거리(Euclidean distance)의 제곱이며, 매개 변수 m 은 각 데이터의 소속 정도에 대한 퍼지 값을 나타내며 1보다 큰 값을 사용한다.

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

$$J_{FCM}(X, V) = \sum_{j=1}^n \left(\sum_{i=1}^c d^{1-\frac{2}{m}}(x_j, v_i) \right)^{1-m} \quad (6)$$

위의 FCM을 구현하는 과정은 그림 2와 같다.

이와 같이 퍼지 클러스터링을 통하여 가정에서 얻어진 위치 및 시간 값의 거리에 대하여 클러스터를 구성하고, 각 클러스터에 이진 값을 부여하는 형태로 연속된 데이터에 대하여 이산화를 수행하였다. 즉, 위치 정보를 5개의 클러스터로 구성한다면 1~5까지의 이산화된 위치 정보로 변환된다.

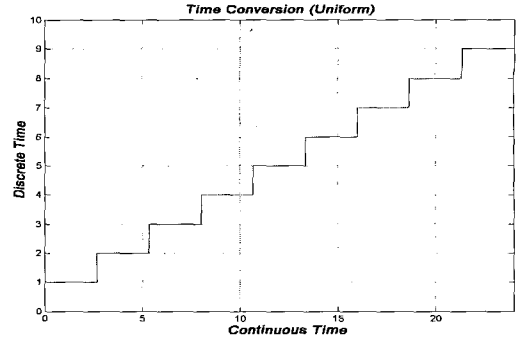
본 논문에서는 베이지안 네트워크의 조건부 확률표를 보다 간단하게 구성할 수 있도록 클러스터의 수를 비교적 작게 설정하였다. 시간의 경우에는 그림 3과 같이 6~9개의 클

- Step 1: For given data set $X = x_1, \dots, x_n, x_i \in R^p$, fix $c \in 2, 3, \dots, n-1, m \in 1, \infty$ and initialize $U^{(0)} \in M_{fc}$
- Step 2: At iteration $l, l = 0, 1, 2, \dots$, compute the c mean vectors

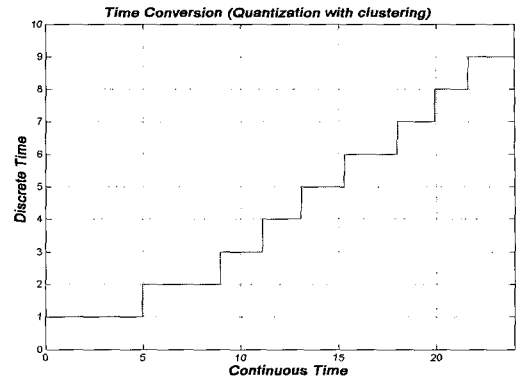
$$v_i^{(l)} = \frac{\sum_{k=1}^n (u_{ik}^{(l)})^m x_k}{\sum_{k=1}^n (u_{ik}^{(l)})^m}, \quad 1 \leq i \leq c$$
- Step 3: Update $U^{(l)} = [u_{ik}^{(l)}]$ to $U^{(l+1)} = [u_{ik}^{(l+1)}]$ using

$$u_{ik}^{(l+1)} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i^{(l)}\|}{\|x_k - v_j^{(l)}\|} \right)^{\frac{2}{m-1}}}, \quad \begin{matrix} 1 \leq i \leq c \\ 1 \leq k \leq n \end{matrix}$$
- Step 4: If $\|U^{(l+1)} - U^{(l)}\| < \epsilon$, stop; otherwise, set $l = l+1$ and go to Step 2

그림 2. Fuzzy c-Means(FCM)알고리즘.
Fig. 2. Fuzzy c-Means algorithm.



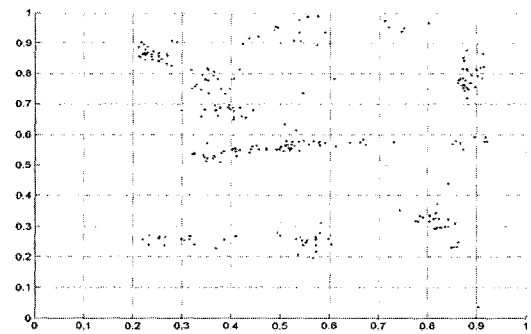
(a) Uniformly quantized time



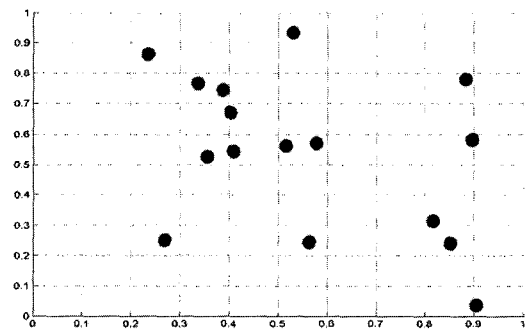
(b) Quantized time with FCM

그림 3. FCM을 이용한 시간의 이산화 결과.

Fig. 3. Quantization time with FCM.



(a) Records of user positions



(b) Cluster centers for user positions

그림 4. FCM을 이용한 사용자 위치의 이산화 결과.

Fig. 4. Quantization of user positions with FCM.

러스터로 구성하였고, 위치의 경우에는 그림 4와 같이 9-16 개의 클러스터로 구성하였다. 그림 3에서와 같이 시간의 경우에는 사용자가 활발하게 활동하는 일과시간에는 수면중인 새벽보다는 보다 세분화되어 나누어진 것을 볼 수 있다. 거주자 위치의 경우에는 그림 4와 같이 각 활동이 나타나는 점을 중심으로 각 클러스터의 센터가 위치하게 되고, 이에 따라서 이산화탄소 수렴함으로써 보다 효율적인 이산화 결과를 가져올 수 있게 된다.

IV. 시뮬레이션 및 결과 고찰

1. 상황인식 시뮬레이터

사용자에 대한 센서자료와 활동자료는 그림 6과 같은 플래시에 기반을 둔 인터페이스를 만들어서 수집하였다[6]. 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 최소 시간이 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.

2. 실험결과

각 노드간의 연결 수를 임계값으로 사용하여 노드간의 연결을 수정하였으며 [6]에서와 같이 균일한 간격으로 이산화 한 것과 성능을 비교하면 다음과 같다(그림 6, 표 1).

그림 6을 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 7과 같다.

그림 7에서 살펴볼 수 있듯이 제안된 시스템에서는 균일

한 간격으로 이산화한 경우보다 높은 성능을 나타내는 것을 볼 수 있다. 하지만, 대부분의 경우에는 잘 인식하지만 모호성이 있는 독서, 공부와 휴식과 같은 경우에는 비교적 낮은 인식률을 가지는 것을 확인할 수 있다.

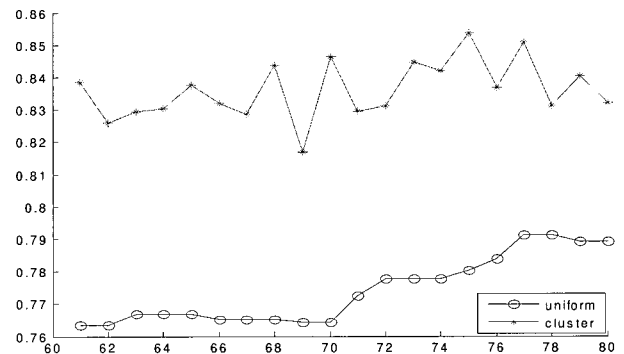


그림 6. 상황인식 성능 비교.

Fig. 6. The performance of context-aware recognition.

표 1. 상황인식 성능 비교.

Table 1. The performance of context-aware recognition.

노드간 연결수	상황인식의 정확도		노드간 연결수	상황인식의 정확도	
	균일 분포[6]	제안된 방법		균일 분포[6]	제안된 방법
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

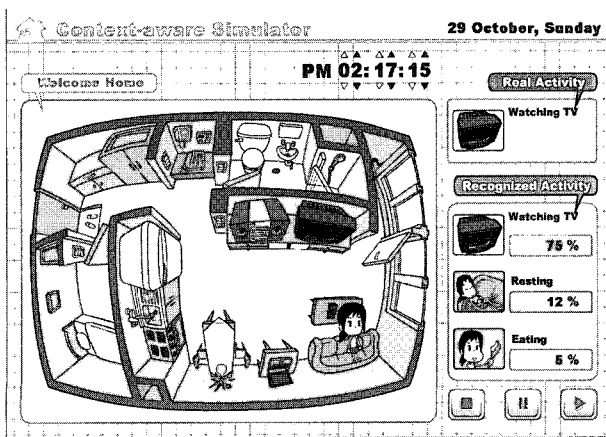


그림 5. 상황인식 시뮬레이터.

Fig. 5. Context-aware simulator.

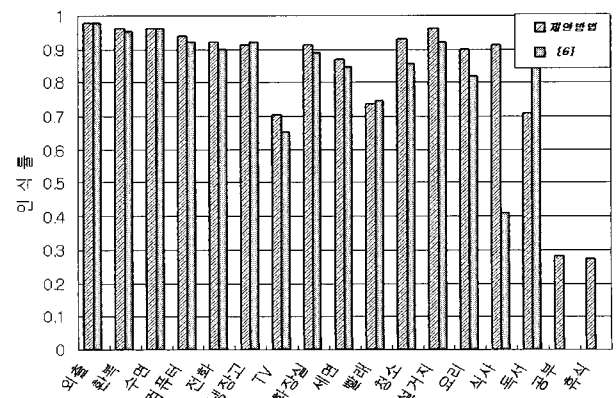


그림 7. 각 활동에 따른 인식률 비교.

Fig. 7. Context-aware recognition rate for each activity.

V. 결론

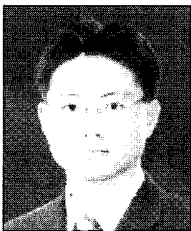
본 논문에서는 제안된 상황인식 기법은 크게 퍼지 클러스터링을 통한 환경 정보의 효율적인 이산화와 상호정보를 이용한 베이지안 네트워크 구조의 학습 및 추론으로 구성되어 있다. 그리고 이와 같이 구성된 상황인식 시스템으로 스마트 홈에서 센서 네트워크로부터의 정보를 바탕으로 거주자의 활동을 인식하였다. 가상 환경에서의 시뮬레이션 결과에서도 확인할 수 있듯이 제안된 기법은 스마트 가전기기 및 센서정보를 통하여 거주자의 활동을 인식하는데 80%정도의 인식률을 가진다.

이와 같은 거주자의 인식률을 각 활동에 따라서 분석해보면, 사람이 단순히 직접 보고서 판단이 가능한 경우에는 제안된 기법으로도 활동 인식이 가능하였다. 즉, 제안된 기법이 개발자가 직접 명시한 경우와 비슷한 성능을 나타낼 수 있는 것이다. 이것은 제안된 기법을 통하여 활동인식에서 자동화할 수 있는 부분이 더욱 늘어나게 된 것을 의미한다고 할 수 있다. 하지만 가시적으로도 판단이 모호한 경우에는 매우 낮은 인식률을 가지므로 이에 대한 보완점이 필요하다.

한편 상황 및 활동인식을 하는 절차나 활동인식을 사용하여 서비스를 제공할 때 편의성과 효율성 이외에 고려하여야 할 사항으로는 사용자가 느끼는 거부감, 개인 사생활 보안 등이 있으며 이러한 문제가 해결되어야 실제로 상황 및 활동인식 서비스를 제공할 수 있을 것이다.

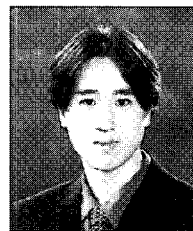
참고문헌

- [1] R. Want, A. Hopper, V. Falcão, and J. Gibbons, "The active badge location system," *ACM Transactions on Information Systems*, pp. 91-102, January 1992.
- [2] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home setting using simple and ubiquitous sensors," in *Proceedings of PERSASIVE 2004*, vol. LNCS 3001, pp. 158-175.
- [3] A. Ranganathan, J. A. Muhtadi, and R. H. Campbell, "Reasoning about uncertain contexts in pervasive computing environments," *IEEE Trans. on Pervasive Computing*, vol. 3, no. 2, pp. 62-70, 2004.
- [4] R. E. Neapolitan, *Learning Bayesian Networks*, Prentice Hall, 2004.
- [5] F. Fleuret, "Fast binary feature selection with conditional mutual information," *The Journal of Machine Learning Research*, vol. 5, Dec 2004.
- [6] 정우용, 김은태, "베이지안 네트워크에 기반한 스마트 홈에서의 상황인식 기법개발," 퍼지 및 지능시스템학회 논문지, 제 17 권 제 2 호, pp. 179-184, 2007.
- [7] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, McGraw Hill, 2002.
- [8] J. Zhang and Y. Leung, "Robust clustering by pruning outliers," *IEEE Trans. on SMC*, vol. 33, no. 6, pp. 983-999, Dec 2003.



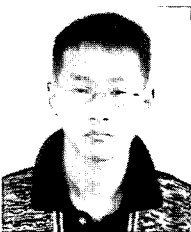
정 우 용

2004년 연세대학교 기계전자공학부 졸업. 2006년 동 대학원 석사. 2006년~현재 동 대학원 박사과정. 관심분야는 상황인식 및 센서 네트워크.



이 제 현

2005년 연세대학교 전기전자공학부 졸업. 2007년 동 대학원 석사. 2007년~현재 동 대학원 박사과정. 관심분야는 지능형 홈 네트워크 및 센서 네트워크.



윤 숙 현

2003년 목포해양대학교 해양경찰학과 졸업. 2006년~현재 연세대학교 전기전자공학과 석사과정. 관심분야는 센서 네트워크.



조 영 완

1991년 연세대학교 전자공학과 졸업. 1993년 동 대학원 석사. 1999년 동 대학원 박사. 2000년~2003년 삼성전자 프린터 디지털 프린팅 사업부. 2003년~현재 서경대학교 컴퓨터공학과 교수. 관심분야는 지능 제어 및 로봇틱스 응용.



김 은 태

1992년 연세대학교 전자공학과 졸업. 1994년 동 대학원 석사. 1999년 동 대학원 박사. 1999년~2002년 한경대학교 제어계측공학과 전임강사. 2002년~현재 연세대학교 전기전자공학과 교수. 관심분야는 계산지능 및 지능형 서비스

로봇과 지능형 홈 네트워크의 적용.