인지과학, 제20권 제4호
Korean Journal of Cognitive Science
2009, Vol. 20, No. 4, 535~554

모바일 컨텍스트 로그를 사용한 속성별 베이지안 네트워크 기반의 랜드마크 예측 모델 학습

이 병 길† 임 성 수 조 성 배
연세대학교 컴퓨터과학과

모바일 장비에서 수집되는 정보는 개인의 기억을 보조하기 위한 수단으로 활용될 수 있지만, 그 양이 너무 많아 사용자가 효과적으로 검색하기에는 어려움이 있다. 데이터를 사람의 기억과 유사한 에피소드 방식으로 저장하기 위해 중요 이벤트인 랜드마크를 탐지하는 것이 필요하다. 본 논문에서는 사용자에게 새로운 서비스를 제공하기 위해서 다양한 컨텍스트 로그 정보로부터 자동으로 랜드마크를 찾아내는 속성별 베이지안 랜드마크 예측 모델을 제안한다. 랜드마크 예측 정확도를 높이기 위해 요일별, 주간별로 데이터를 나누고 다시 수집된 경로에 따른 속성으로 분류하여 학습을 통해 베이지안 네트워크를 생성하였다. 노키아의 로그데이터로 실험한 결과, 베이지안 네트워크를 사용한 방법이 SVM을 사용한 방법보다 예측성능이 높았으며, 주간별 및 요일별로 설계한 베이지안 네트워크에 비해 제안한 방법인 속성별 베이지안 네트워크의 성능이 가장 우수하였다.

주제어 : 컨텍스트 로그, 랜드마크, 속성별 베이지안 네트워크, 베이지안 구조학습

---

* 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단-신기술융합형 성장동력사업의 지원을 받아 수행된 연구임(2009-0093676)
†교신저자: 연세대학교 컴퓨터과학과, 연구 세부 분야: 인공지능
E-mail: byulyi@sclab.yonsei.ac.kr, lss@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

Ⅰ. 서 론

최근 유비쿼터스 기술의 발전과 함께 다양한 모바일 기기와 센서 장비들이 보급되고 있다. 이에 따라서 사람들이 보고, 듣고, 느끼는 것들에 대한 데이터를 모바일 기기와 센서 장비를 통해서 수집할 수 있게 되었다. GPS를 통해서 개인의 위치 정보를 얻을 수 있고, 각종 환경 센서를 이용하여 온도, 습도 등의 환경 정보를 획득할 수 있으며, 또한, 다양한 착용형 센서를 이용하여 사용자의 현재 행동을 추론할 수 있다. 최근 Nokia의 LifeBlog 프로젝트 등 모바일 장비를 통해서 사용자의 일상생활 로그의 분석과 수집, 관리하는 연구가 진행되고 있다[1-7]. 모바일 장비는 개인성이 강한 장비이므로 개인의 기호나 성향에 따라 적응되어 특화될 수 있다. 그리고, 모바일 장비는 사용자가 항상 휴대하기 때문에 사용자의 일상정보를 효과적으로 수집하고 분석하여 사용자의 기억을 증강시키는데 도움을 줄 수 있다[1].

모바일 장비로부터 수집한 라이프 로그에는 개인의 기억을 증강시킬 수 있는 정보가 많이 있지만, 각종 센서로부터 들어오는 많은 수의 정보로부터 사용자가 원하는 정보를 찾는 다는 것은 쉬운 일이 아니다. 본 논문에서는 모바일 장비로부터 수집 가능한 컨텍스트 정보를 활용하여 개인의 하루 생활에서 자동으로 랜드마크를 찾아낸다[8]. 여기에서 랜드마크란 사용자가 경험한 많은 이벤트들 중 특별한 에피소드에 대한 기억에 도움을 주는 핵심 키워드를 의미하며 E. Horvitz의 연구[9]에서 제안된 개념이다. 랜드마크는 인지과학 분야의 추출 큐(Retrieval Cue; Episodic Memory를 불러올 때 중요한 역할을 하는 정보)라는 개념으로부터 파생되었으며, 여기서는 경험 에피소드를 기억할 때 중요한 역할을 하는 정보를 의미한다. 컨텍스트 정보는 에피소드 단위로 재구성되며 사용자의 개인 정보검색을 향상시키는 목적으로 쓰인다. 또한 랜드마크는 이벤트가 발생한 장소 및 시간, 같이 경험한 사람, 그리고 이벤트가 일어난 시점 전후, 또는 일어난 동안에 저장된 이벤트들에 대한 정보에 의해 구성되고 사용자가 에피소드를 기억하는데 도움을 준다.

본 논문에서는 랜드마크를 에피소드 단위로 재구성된 컨텍스트 정보의 효율적인 검색을 위한 핵심 로그데이터로 정의하고, 확률기반 추론 모델인 베이지안 네트워크를 사용한 랜드마크 예측 모델을 제안한다. 이때, 랜드마크 예측의 정확도를 높이기 위하여 수집된 데이터를 요일별, 주간별로 분할하고, 이를 다시 저장 경로

속성에 대해 구별한 후에 베이지안 네트워트 모델을 학습하는 방법을 제안한다. 실제 노키아의 벤치마크 로그데이터에 대하여 실험하여 그 성능을 비교 및 평가한다.

Ⅱ. 배경 및 관련연구

1. 특이성 추출

저장 장치의 용량 증가, 멀티미디어 컨텐츠 증가, 개인 컴퓨터를 비롯하여 모바일 장비의 사용자 증가 등으로 관리해야 하는 정보가 늘어나서 사용자가 과거에 일어났던 기억을 떠올리기가 힘들어지고 있다. 이로 인하여 사용자의 이벤트 캘린더 정보로부터 중요한 정보, 즉 랜드마크를 찾아내는 기법이 연구되고 있다. 마이크로소프트 연구소의 랜드마크 예측과 관련된 연구모델에서는 온라인 캘린더 정보와 사용자 정보를 베이지안 네트워크로 학습하여 랜드마크 예측율에 대한 성능을 평가하였다[9]. 28개의 노드로 구성된 온라인 캘린더 정보는 Beckerman과 Meek의 베이지안 구조 탐색 프로시저를 통해서 학습되었다[10]. Life Browser는 사용자의 일상생활을 요약하여 핵심 사항만을 볼 수 있도록 구성한 것으로 만약 사용자가 요

표 1. 관련 연구 비교

	마이크로소프트 연구소	Life Browser	MIT 미디어연구실
목적	랜드마크 예측 모델 설계	일상생활 요약 사용자 요청자료 검색	특정위치에서의 상황 예측
특징	사용자별로 모델 평가	2종류의 랜드마크 사용	해당 요소들 간의 관계분석 오류 데이터 제거
데이터	온라인 캘린더 사용자 정보	웹페이지, 이메일, 문서 사용자 정보	스마트 폰의 컨텍스트 로그
방법	베이지안 네트워크	-	Hidden Markov Model Gaussian Mixture Model

청한 자료가 있을 경우 대중적인 랜드마크와 개인적인 랜드마크를 검색 조건에 추가하여 검색시간을 단축시켜 좀 더 밀접한 관련이 있는 자료를 결과로 제공해 준다[11]. 이 연구에서 사용한 대중적인 랜드마크란 휴일, 신문 또는 뉴스의 중요한 헤드라인을 의미하고, 개인적인 랜드마크란 개인이 촬영한 사진이나 사용자가 직접 입력한 중요한 캘린더 이벤트로 정의된다. MIT 미디어 연구실에서는 스마트 폰에서 수집할 수 있는 컨텍스트 로그를 사용하여 HMM(Hidden Markov Model) 기반으로 사용자의 생활 패턴을 분석하여 특정한 위치에서의 사용자 모델링 연구를 진행하였다[12]. 이 세 연구에 대한 비교는 표 1과 같다.

마이크로소프트 연구소의 경우는 사용자의 변화에 재빨리 대응할 수가 없을 뿐만 아니라, 사용자 정보를 사용자가 직접 입력해야 하기 때문에 입력이 잘못될 경우 예상하지 못한 결과가 발생되는 문제가 있다. Life Browser는 제한된 환경의 데이터만을 사용해서 일상생활을 요약하는데 상대적으로 과거 기억에 도움을 주기에는 한계가 있다. 마지막으로 MIT 미디어 연구실의 경우 특정위치(집, 직장, 그 외)에서 사용자 상황예측의 대상이 연구소 직원이나 그 근처 다른 연구소 직원들로 한정되어 특별한 방법을 사용하지 않고 단순히 학습된 상황을 사용해 분석 평가한다. 본 논문에서는 실시간으로 들어오는 데이터를 토대로 모바일 장비 환경에서 정확한 랜드마크를 예측하기 위하여 속성별로 다른 베이지안 네트워크를 사용하는 방법을 제안한다.

2. 베이지안 네트워크

베이지안 네트워크는 변수들 간의 원인과 결과 관계를 확률적으로 모델링하기 위한 도구로서 불확실한 환경에서 좀 더 신뢰성 있는 결과를 추론하기 위해 쓰이는 대표적인 방법이다. 이 방법은 그림 1과 같이 변수를 나타내는 노드와 변수 사이의 상관관계 및 인과관계를 나타내는 연결선(arc, edge)으로 이루어져 있는 방향성 비순환 그래프(directed acyclic graph, DAG)이다[13, 14]. 베이지안 네트워크 B는 연결선의 집합 E와 변수의 집합 V로 이루어지며 파라미터 θ를 가진다. 변수들 사이의 연결선은 확률적인 의존관계를 나타낸다. 파라미터 θ는 의존 관계가 있는 변수들 사이의 조건부 확률 값을 나타낸다. 변수의 개수가 n인

경우 $V=\{V_1, V_2, ..., V_n\}$으로 정의되며 연결선의 개수가 m인 경우 $E=\{E_1, E_2, ..., E_m\}$ 으로 정의된다. $V_a \rightarrow V_b$인 연결선의 경우 변수 V_a가 변수 V_b의 부모가 된다. 각 변수의 부모들은 $Pa(V_x)=\{V_i | V_i \rightarrow V_x, V_i \in V\}$로 정의한다. 만약 $Pa(V_x)$가 $\varnothing$이면 V_x는 사전확률만을 가진다. 그렇지 않다면 V_x는 조건부 확률을 가진다.

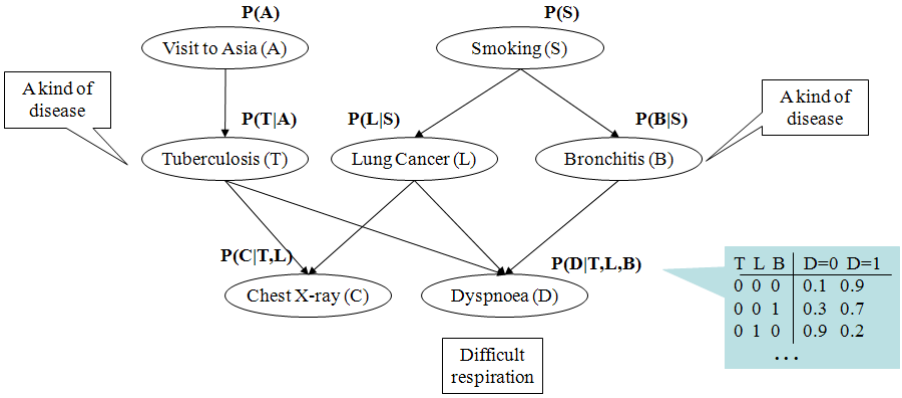

그림 1. 베이지안 네트워크의 예

조건부확률을 나타내는 테이블을 가지고 베이즈 규칙(Bayes' Rule)을 이용해 각 노드의 사후확률을 계산한다. 만약 두 노드 사이의 연결선이 없다면 두 노드는 서로 독립적이라는 의미로 해석한다. 노드의 부모와 확률값이 정해지면 조건부확률 테이블을 가지고 계산을 하게 되는데, 확률기반 규칙 시스템에 비해 훨씬 적은 양의 연산으로 확률 추론이 가능하다. 그림 1에서 결합 확률 값인 $P(A,S,T,L,B,C,D)$는 다음과 같이 계산한다.

$$\begin{aligned} &P(A,S,T,L,B,C,D) \\ &= P(A)P(S)P(T|A)P(L|S)P(B|S)P(C|T,L)P(D|T,L,B) \end{aligned} \quad (1)$$

베이지안 네트워크의 추론은 베이즈의 규칙을 통해 계산하며 다음과 같이 정의

한다.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2}$$

$P(A)$와 $P(B)$는 사전지식이며 $P(B|A)$는 조건부 확률이다. 즉 모든 변수 중에서 관측이 가능한 변수들의 집합을 B라고 하고 관측이 되지 않았고 관심의 대상이 되는 변수들의 집합을 A라고 하면 이미 사전에 $P(B|A)$, $P(A)$, $P(B)$가 주어져 있다고 할 때 $P(A|B)$의 값을 계산할 수 있다.

베이지안 네트워크의 구조를 설계하는 것은 전문가가 변수들 사이의 의존관계 중에서 특히 중요하고 결합 확률에 큰 영향을 미칠 것으로 생각되는 것들만 선택적으로 결정해 주는 과정을 말한다. 구조가 결정되면 전문가는 해당 변수들의 조건부 확률 테이블을 채워 베이지안 네트워크의 설계 과정을 마무리한다. 일반적으로 해당 도메인에 대한 전문적을 지식을 가진 사람이 필요하며 그 사람이 꼭 베이지안 네트워크 구조를 직접 설계할 필요는 없다. 전문가는 해당 도메인에 대한 지식만을 알려주고 베이지안 네트워크 설계자가 직접 구조를 설계할 수 있다. 만약 해당 도메인에 대한 전문지식을 획득하기가 어렵다면변수들의2와 같이 데이터로부터 직접 학습하는 방법이 있다[15]. 이 경우 해당 도메인으로부터 데이터를 수집해야 하는 어려움이 있지만 데이터를 충분히 얻을 수 있는 곳이라면 학습이 좋은 해결책이 될 수 있다.

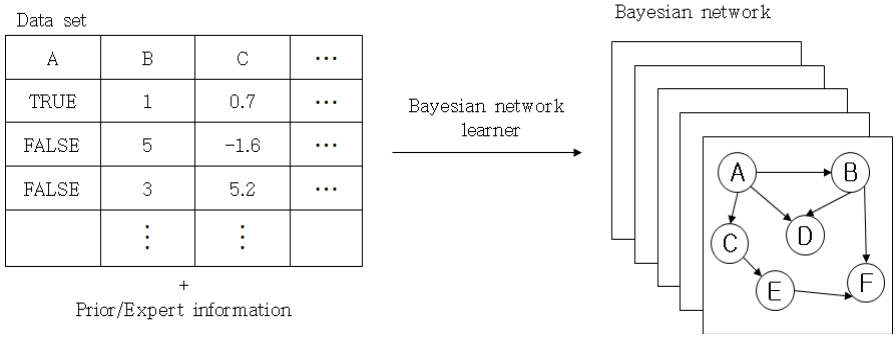

그림 2. 베이지안 네트워크의 학습

Ⅲ. 속성별 베이지안 네트워크를 이용한 랜드마크 예측 모델

그림 3은 본 논문에서 제안하는 랜드마크 예측 과정을 보여준다. 우선 모바일 장비를 사용하여 컨텍스트 로그 정보를 수집한다. 실제로 수집된 로그 정보를 베이지안 네트워크에 적용하고 분류하기 위해서 각 노드의 이름과 상태에 맞게 변경해 주는 연산을 한다. 전처리된 컨텍스트 정보는 베이지안 네트워크의 성능을 향상시키기 위하여 각 요일에 맞게 분류하고 후에 각 속성별(위치정보, 환경정보, 시스템정보)로 분류한다. MIT 연구에 비춰보면 대부분의 사람들은 요일별로 비슷한 생활패턴을 가지고 있기 때문에 요일정보를 데이터의 분류 기준으로 사용하므로 베이지안 네트워크 모델의 노드에서 제외되었고 날짜정보는 불확실한 환경에서도 좀 더 신뢰성 있는 결과를 예측하기 위하여 제외하였다. 요일과 속성별로 분류된 데이터를 사용하여 각각에 대해 구조와 파라미터를 학습한다. 각각 속성별로 학습된 베이지안 네트워크를 요일별로 구조를 통합한다. 이렇게 통합된 베이지안 네트워크는 데이터 입력이 들어오면 동일한 요일에 해당하는 베이지안 네트워크를 선택하여 랜드마크인지 아닌지에 대하여 예측한다.

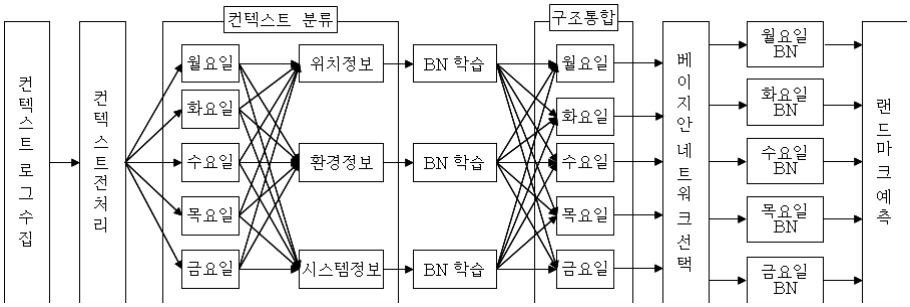

그림 3. 제안하는 랜드마크 예측 방법

일상에서 일어나는 사건들은 매우 다양하여 몇 개의 베이지안 네트워크로 모델링하려면 설계가 복잡해지고 정확한 모델링이 어려워진다. 일반적으로 베이지안 네트워크 학습은 NP-hard 문제로서 정확한 방법으로 해결할 수 없기 때문에 경험적 방법들(heuristic methods)로 주어진 문제를 해결해야 한다. NP-hard 문제는 처리

하고자 하는 데이터의 수가 증가함에 따라서 고려해야 하는 경우의 수가 기하급수적으로 증가하므로, 데이터의 수를 줄이게 되면 시스템의 연산량을 줄이며, 정확도도 높일 수 있다. 본 논문에서는 베이지안 네트워크의 학습을 효율적으로 수행하고 예측성능을 높이기 위하여 데이터를 요일별로 나누고, 베이지안 네트워크 구조를 속성별로 나누어 학습하는 방법을 제안한다. 제안하는 방법은 입력되는 정보를 각 요일에 맞게 분류하고 후에 이를 속성별(위치정보, 환경정보, 시스템 정보)로 분류한다. MIT 연구[12]에 비춰보면 대부분의 사람들은 요일별로 비슷한 생활 패턴을 가지기 때문에 데이터를 분류하는 기준으로 사용하여 예측에 대한 성능을 높이기 위해서 독립된 BN을 사용한다. 또한 요일정보를 베이지안 네트워크에 포함해서 구조학습을 진행하는 경우, 복잡도(아크의 수, 계산량)가 높아지고 예측에 더 많은 시간이 소요된다. 따라서 본 논문에서는 요일별로 분류하여 BN을 생성하며, 같은 이유에서 속성별로 BN을 구성하고, 각각 속성별로 학습된 베이지안 네트워크를 다시 요일별로 구조를 통합한다.

본 논문에서는 예측모델 학습방법으로 탐욕적 탐색 방법을 최적화한 알고리즘인 K2 알고리즘을 사용하였다[16]. 기본적인 알고리즘은 그림 4에서 보듯이 그리디 기반 휴리스틱 알고리즘 B[10]를 사용하여 네트워크의 구조를 생성한다. $A[i,j]$는 x_j를 x_i의 부모로 연결한 경우의 점수에서 이전 점수를 뺀 차이값을 나타내는 행렬인데, $A[i,j]$값이 가장 큰 i,j쌍을 찾아서 $x_j \rightarrow x_i$ 아크를 하나씩 추가한다. 이때 방향성 사이클이 생기는 것을 막기 위해서 x_j의 조상 x_a가 x_i의 자손 x_b의 자식으로 연결될 수 없도록 $A[a,b]$ 값을 $-\infty$로 조정한다. 그리고 아크의 연결로 인해 바뀐 $f(x_i, Pa(x_i))$의 값을 반영하기 위해 연결 가능한 아크 $x_k \rightarrow x_i$에 대한 $A[i,k]$ 값을 새롭게 갱신한다. 점수 메트릭으로는 그림 5와 같은 K2 메트릭을 사용하여, 데이터와 그래프간의 적합도를 계산한다. K2 알고리즘은 가장 짧은 경로를 탐색할 수 있는 협동적 행동에 근거하여 처음에는 다양한 경로를 조사하고 설정하지만 데이터에 의한 학습이 계속 진행될수록 가장 짧은 경로만을 설정하게 되는 최적화 방법으로 최근 많이 사용되고 있는 방법 중 하나이다.

이와 같은 방법을 사용하여 컨텍스트 로그 데이터를 학습한 속성별 베이지안 네트워크는 크게 위치 추론 베이지안 네트워크, 센서 추론 베이지안 네트워크, 어

1. *Initialization:*
(a) for $i=1$ **to** n **do:** $Pa(x_i)=\varnothing$
(b) for $i=1$ and $j=1$ **to** n **do:**
if $(i \neq j)$ **then** $A[i,j]=f(x_i,\{x_j\})-f(x_i,\varnothing)$

2. *Loop:*
(a) repeat
i. Select two indexes (i,j), such that,
(i,j) = $\arg max_{i',j'} A[i',j']$
ii. **if** $(A[i,j]>0)$ **then** $Pa(x_i)=Pa(x_i)\cup\{x_j\}$
iii. $A[i,j]=-\infty$
iv. **for all** $x_a \in Ancestors(x_j)\cup\{x_j\}$
and $x_b \in Descendants(x_i)\cup\{x_i\}$ **do:**
$A[a,b]=-\infty$
v. **for** $k=1$ **to** n **do:**
if $(A[i,k]>-\infty)$
then $A[i,k]=f(x_i,Pa(x_i)\cup\{x_k\})-f(x_i,Pa(x_i))$
until $\forall i,j\,(A[i,j]\leq 0$ or $A[i,j]=-\infty)$

* *f*: scoring metric
* A[i,j]: adjacency matrix
* Pa(x_i): x_i의 부모

그림 4. 베이지안 네트워크 구조학습 알고리즘

$$f(x_i,Pa(x_i))=\prod_{j=1}^{q_i}\frac{(r_i-1)!}{(N_{ij}+r_i-1)!}\prod_{k=1}^{r_i}N_{ijk}!$$

* r_i: number of possible values of the variable x_i
* q_i: number of possible configurations for the variables in Pa(x_i)
* N_{ijk}: number of case in D in which variable x_i has its kth value
* $N_{ij}=\sum_{k=1}^{r_i}N_{ijk}$

그림 5. K2 평가함수

플리케이션 추론 베이지안 네트워크로 분류된다. 특히 위치요소(cell id, location area code)는 시간요소(day period, hour, half hour)에 많은 영향을 받기 때문에 속성을 분류할 때 위치 추론 베이지안 네트워크에 포함하여 학습한다.

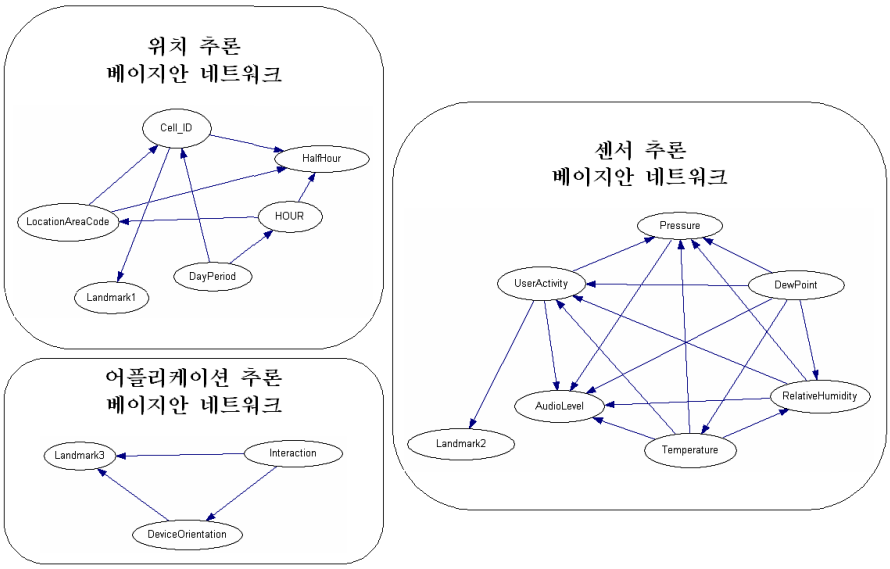

그림 6. 속성별 베이지안 네트워크의 예

위치 추론, 어플리케이션 추론, 그리고 센서 추론을 위한 속성별 베이지안 네트워크는 각각 랜드마크1, 랜드마크2, 랜드마크3의 노드를 추가로 가지며, 이 노드들은 각 속성에 대한 랜드마크 여부(참 또는 거짓)를 상태값으로 가진다. 베이지안 네트워크의 노드를 속성에 따라서 분류한 후, 구조 학습을 통해서 각 속성별 베이지안 네트워크를 얻는다. 그림 6은 학습된 속성별 베이지안 네트워크의 예를 보여준다. 이렇게 얻어진 세 개의 속성별 베이지안 네트워크는 최종 추론 노드인 랜드마크와 랜드마크1, 랜드마크2, 랜드마크3의 세 노드를 연결하여 하나의 베이지안 네트워크로 통합한다. 그림 7은 이러한 과정을 거쳐 최종적으로 통합된 베이지안 네트워크의 한 예이다.

본 논문에서 사용한 예측 방법은 베이지안 네트워크에서 일반적으로 사용되는

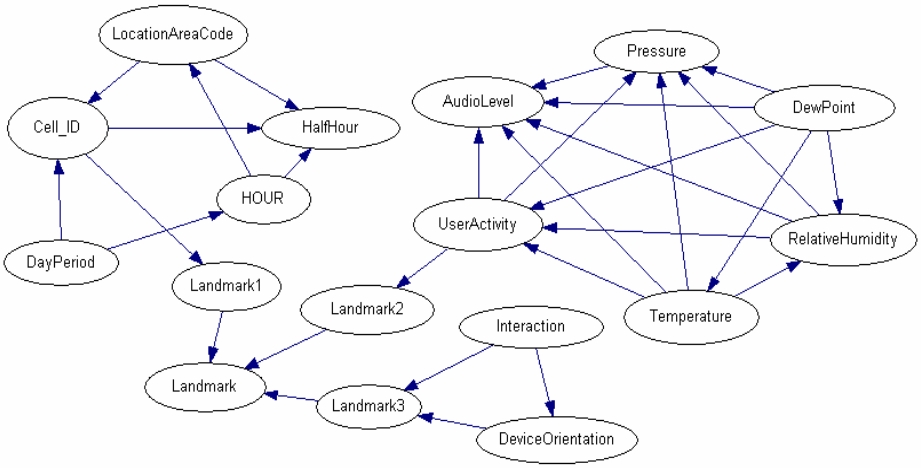

그림 7. 통합된 속성별 베이지안 네트워크 구조의 예

정션트리 알고리즘을 사용한다. 이 알고리즘은 그래프 이론과 확률 이론 사이의 연결 분석에 기반을 두는 알고리즘으로 실질적인 DAG 대신에 정션트리 데이터 구조를 사용하여 랜드마크를 예측한다[17]. 예측 과정은 수집된 컨텍스트 로그를 시간과 속성별로 정리하여 베이지안 네트워크의 증거변수로 설정해서 확률 연산을 진행한다.

Ⅳ. 실험 및 결과

1. 실험환경

실험에는 노키아 데이터 집합을 사용하였다. 노키아 데이터 집합은 43개의 세션으로 6주에 걸쳐 토요일과 일요일을 제외한 27일(11/3~12/12)동안 기록된 데이터이다[18]. 총 16개의 속성(표 2)으로 구성된 약 21만개의 데이터를 전처리를 통하여 13개의 속성, 약 3만 여개의 데이터로 재구성하고, 랜드마크 여부를 레이블링 하였다. 입력된 데이터는 결측치(missing value)가 존재하지 않는다.

표 2. 노키아 데이터 집합

Node	Type	Value
TIME		
Day Name	STRING	Monday, Tuesday, Wednesday, Thursday, Friday
Day Date	INT	1-31
Month	INT	1-12
Day Period	STRING	Night, Morning, Afternoon, Evening
Hour	INT	1-24
Half Hour	INT	1 (hour:45-hour:15), 2 (hour:15-hour:45)
GPS		
Cell ID	INT	0-124
Location Area Code	INT	0-5
SENSOR		
Temperature	INT	1-10
Relative Humidity	INT	1-10
Dew Point	INT	1-10
Pressure	INT	1-10
User Activity	STRING	Standing, Walking, Running, Sitting, Sleeping
Average Audio Level	INT	1-5
SYSTEM		
Device Orientation	INT	1-6
Interaction	STRING	Application, Web page, Communication
Landmark		
Landmark	Boolean	True, False
Landmark1	Boolean	True, False
Landmark2	Boolean	True, False
Landmark3	Boolean	True, False

선행실험을 통해 날짜정보와 요일정보를 포함해서 베이지안 네트워크를 학습할 경우 복잡도(아크의 수, 계산량)가 높아지고 수행시간(구조 학습 및 예측에 소요되는 시간)이 길어지는 것에 비하여 성능 향상에는 많은 영향을 주지 못하였다. 따라서 이 두 요소를 베이지안 네트워크의 노드에서 제외하고 실험을 진행한다.

데이터는 요일에 맞게 분류한 경우(요일별)와 로그가 발생한 시간 순서(주간별)로 분류한 경우로 나누고 이를 다시 속성별로 학습한 경우와 그렇지 않은 경우로 분류하여 비교 실험을 진행하였다. 다른 기계학습 도구와 성능을 비교하기 위해서 분류 문제를 해결할 때 널리 사용되고 있는 SVM(Support Vector Machine) 방법을 사용하였다. 각 요일에 맞게 분류된 데이터 집합은 학습을 통하여 설계된 각각의 베이지안 네트워크에 파일 수만큼 상호평가(Cross-Validation; CV)를 통하여 성능을 측정한다[19]. 월요일과 수요일 데이터의 경우는 5주의 데이터를 제외한 5개의 데이터가 있고, 금요일 데이터는 4주의 데이터를 제외한 5개의 데이터가 있으므로 5-fold CV를 적용하고, '화,' '목'의 경우는 6개의 데이터가 모두 있으므로 6-fold CV를 적용한다. 주간별로 학습한 경우는 학습 데이터를 1주, 2주, …, 5주까지 점차적으로 늘려가면서 베이지안 네트워크를 설계하여 마지막 6주에 해당하는 데이터를 테스트 데이터로 사용하여 랜드마크 예측에 대한 성능을 평가하였다(표 3). SVM의

표 3. 실험방법에 따른 데이터 분류

	요일별					주간별
	월	화	수	목	금	
1주	○	○	○	○	○	학습
2주	○	○	○	○	○	
3주	○	○	○	○	○	
4주	○	○	○	○	×	
5주	×	○	×	○	○	
6주	○	○	○	○	○	테스트
실험방법	5-fold CV	6-fold CV	5-fold CV	6-fold CV	5-fold CV	

(○: 데이터 있음, ×: 데이터 없음)

경우에는 속성별로는 구분하지 않고 요일별과 주간별로만 구분하여 실험을 진행하였다.

제안한 방법에 대한 속성별 베이지안 네트워크 모델은 각각에 대해서 요일별 모델 총 108개(요일에 따른 속성별 모델: 27×3=81개, 통합된 속성별 모델: 27개)와 주간별 모델 20개(주에 따른 속성별 모델: 5×3=15개, 통합된 속성별 모델: 5개)를 생성하여 실험에 사용하였다. 비교 실험 모델로 마이크로 소프트연구소의 방법[8]과 같이 속성을 분류하지 않고 학습한 베이지안 네트워크 모델과 SVM 모델은 모델별로 각각 32개(요일별 27개, 주간별 5개)를 실험에 사용하였다.

2) 랜드마크 예측 성능 비교

속성을 분류하지 않고 학습된 예측모델(그림 8)은 데이터 집합을 통하여 상관관계가 다양하게 발생하기 때문에 구조가 복잡하지만 속성별로 구조를 학습하고 통합한 모델(그림 7)의 경우에는 사전에 복잡한 데이터의 연관관계를 전처리한 효과를 가져와 상대적으로 단순하게 구성된다. 단순히 예측 모델의 구조만 비교하더라도 복잡도가 낮아진다는 것을 알 수 있다.

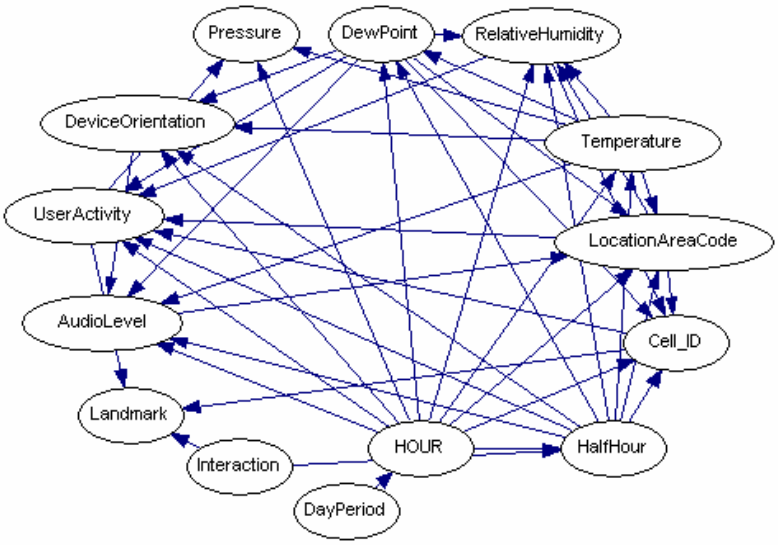

그림 8. 학습된 일반적인 베이지안 네트워크의 예

그림 9의 요일별 랜드마크 예측율을 보면 본 논문에서 제안한 방법이 SVM 방법과의 예측율에서는 큰 차이를 보이지는 않지만 상대적으로 표준편차는 작게 나타나 정확도 면에서 높은 안정성을 보여준다. 데이터를 분류하지 않고 학습한 베이지안 네트워크의 경우 제안한 방법보다 예측율과 안정성 면에서 모두 성능이 낮다는 것을 알 수 있다.

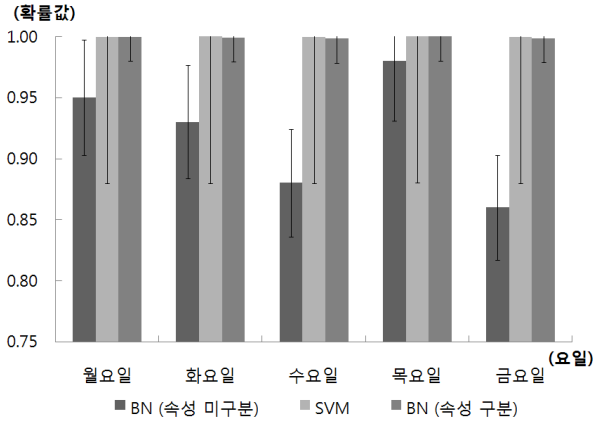

그림 9. 요일별 랜드마크 예측율

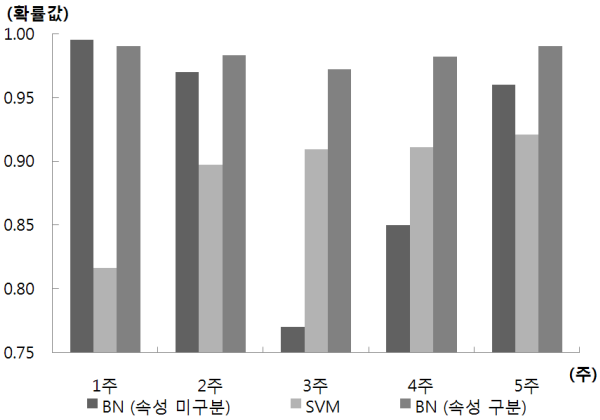

그림 10. 주간별 랜드마크 예측율

그림 10에서 속성별로 학습하지 않은 베이지안 네트워크와 SVM에 대한 주간별 비교를 보면 1~3주의 경우 베이지안 네트워크의 경우 SVM과 다르게 예측 정확도 성능이 점점 떨어지는 것을 알 수 있다. 이는 학습된 데이터의 양이 적고 다양한 구성으로 로그가 기록될 경우 정확한 예측이 불가능하다는 것을 보여준다. 하지만 본 논문에서 제안한 방법은 데이터의 양이 많고 적음에 상관없이 안정적으로 높은 성능을 보여준다. 이는 속성별로 분류한 데이터가 로그의 다양성을 낮춰 랜드마크 예측의 정확도를 높여주기 때문으로 해석할 수 있다.

표 4. 속성별 vs 비속성별 베이지안 네트워크 성능비교

	요일별		주간별	
	속성구분	속성미구분	속성구분	속성미구분
예측율 (%)	99.912	91.9	98.34	90.8
표준편차 (%)	2.4	5	-	-
학습 및 예측 평균 소요시간 (시:분:초)	00:00:17×5 =00:01:25	00:06:13×5 = 00:30:45	00:03:41	01:43:59
아크의 수 (개)	31.6	44.4	33.2	50.2
계산량	17,428.7	533,486.4	25,834	1,913,170

표 4에서 데이터의 속성을 구분하지 않고 학습한 예측 모델과 속성을 구분하고 학습한 예측 모델의 성능은 모든 면에서 큰 차이를 보인다. 데이터의 속성을 구분하지 않은 예측 모델이 제안한 방법에 비해 요일별 모델은 학습 및 예측 소요시간(약 21.7배), 아크의 수(약 1.4배), 계산량(약 30.6배)이 높고, 주간별 모델은 학습 및 예측 소요시간(약 28.2배), 아크의 수(약 1.5배), 계산량(약 74.0배)에서 성능이 떨어지는 것을 알 수 있다. 이에 비해 예측율과 정확도는 높은 것을 알 수 있다. 전체적으로 요일에 맞는 예측 모델을 사용한 경우가 주간별 모델보다 복잡도가 낮고 성능이 높은 것을 알 수 있다. 이러한 결과를 보이는 이유는 MIT 미디어 연구실에서 수행한 연구 결과에 비추어 보면 사람의 생활 패턴은 요일별로 유사하게 나타나기 때문으로 추정된다.

표 5. 베이지안 네트워크 vs SVM 성능비교

	요일별		주간별	
	BN	SVM	BN	SVM
예측율 (%)	99.912	99.974	98.34	89.10
표준편차 (%)	2.4	12	-	-
학습 및 추론 평균소요시간	00:01:25	01:41:54	00:03:41	01:24:50

표 5의 비교를 보면 전체적으로 베이지안 네트워크 모델이 SVM 모델에 비해 성능은 비슷하지만 오차 범위와 소요되는 시간면에서 높은 성능을 나타내는 것을 알 수 있다.

V. 결론 및 토의

랜드마크 예측과 관련된 기존 연구들을 보면 여러 장비들을 사용하지 않고 한정된 장비만을 사용하여 실험을 진행하였기 때문에 장비를 추가하거나 모델을 일반화 시킬 경우 성능이 떨어지거나 기존 모델을 사용하지 못하고 새롭게 설계해야 하는 문제가 발생할 수 있지만 본 논문에서 제안한 방법은 여러 장비를 각 속성별로 분류하여 구조를 설계했기 때문에 기존 연구에 비해 모델 확장에 있어 좀 더 용이하다. 또한 랜드마크 예측을 특정한 날에 초점을 맞추지 않고 일반적인 상황에 맞는 모델을 설계하기 위하여 요일별 랜드마크 예측 모델을 제안하였고, 실험을 통해 그 유용성을 확인하였다.

향후 과제로는 본 논문에서 분석된 랜드마크 항목을 일반 사용자에게 쉽게 보여주고 이해할 수 있도록 하는 응용 프로그램 개발과 수집된 데이터를 실시간으로 처리할 수 있는 컨텍스트 전처리 시스템의 개발이 필요하다. 그리고 보다 많은 장비를 활용하여 다양한 정보를 수집해서 랜드마크 예측에 대한 정확도를 향상 시킬 필요가 있다. 또한 탐지된 랜드마크를 활용하는 연구 역시 진행되어야 할 것이다. 그리고 속성 구분이 랜드마크 추출에 미치는 영향을 분석하기 위해서, 속성 구분

SVM을 통한 랜드마크 모델을 구축하여 제안하는 속성 구분/미구분 BN과의 비교 평가가 필요하다.

참고문헌

[1] M. Raento, A. Oulasvirta, R. Petit, and H. Toivonen, “ContextPhone: A prototyping platform for context-aware mobile applications,” *IEEE Pervasive Computing*, vol. 4, no. 2, pp. 51-59, 2005.

[2] A. Krause, A. Smailagic, and D. P. Siewior다, “Context-aware mobile computing: Learning context-dependent personal preferences from a wearable sensor array,” *IEEE Trans. on Mobile Computing*, vol. 5. no. 2, pp. 113-127, 2006.

[3] R. DeVaul, M. Sung, J. Gips, and A. Pentland, "MIThril 2003: Applications and Architecture," *Proc of 7th IEEE Int. Symposium on Wearable Computers*, pp. 4-11, 2003.

[4] P. Zheng and L. M. Ni, “The rise of the smart phone,” *IEEE Distributed Systems Online*, vol. 7, no. 3, pp. 1-14, 2006.

[5] P. Korpipaa, J. Mantyjarvi, J. Kela, H. Keranen, and E.-J. Malm, "Managing context information in mobile devices," *IEEE Pervasive Computing*, vol. 2, no. 3, pp. 42-51, 2003.

[6] J. Gemmell, L. Williams, K. Wood, R. Lueder, and G. Bell, “Pervasive capture and ensuing issues for a personal lifetime store,” *Proc. of the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences*, pp. 48-55, 2004.

[7] D.P. Siewior다, A. Smailagic, J. Furakawa, A. Krause, N. Moraveji, K. Reiger, J. Shaffer, and F.L. Wong, "SenSay: A context-aware mobile phone," *Proc. 7th Int. Symp. of Wearable Computers*, pp. 248-249, 2003.

[8] R. M. Shiffrin and M. Steyvers, “A model for recognition memory: REM retrieving effectively from memory,” *Psychonomic Bulletin & Review*, vol. 4, no. 2, pp. 145-166, 1997.

[9] E. Horvitz, S. Dumais, and P. Koch, “Learning predictive models of memory

landmarks," *CogSci 2004: 26th Annual Meeting of the Cognitive Science Society,* Chicago, pp. 267-272, August 2004.

[10] D.M. Chickering, D. Heckerman, and C. Meek, "A Bayesian approach to learning Bayesian networks with local structure," *Proc. of Uncertainty in Artificial Intelligence*, pp. 80-89, 1997.

[11] M. Ringel, E. Cutrell, S. Dumais, and E. Horvitz, "Milestones in time: The value of landmarks in retrieving information from personal stores," *In Proceedings of the 9th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT)*, pp. 184-191, September 2003.

[12] N. Eagle and A. Pentland, "Reality mining: Sensing complex social systems," *J. of Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255-268, June 2005.

[13] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.

[14] K. B. Korb, and N. E. Nicholson, *Bayesian Artificial Intelligence*, Chapman & Hall/CRC, 2003.

[15] G. F. Cooper and E. A. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, no. 4, pp. 309-347, 1992.

[16] L. M. de Campos, J. M. Fernrandez-Luna, J. A. Gamez, and J. M. Puerta, "Ant colony optimization for learning Bayesian networks," *Int. Journal of Approximate Reasoning*, pp. 291-311, 2002.

[17] U. Kjaerulff, "Inference in Bayesian networks using nested junction trees," *Learning in Graphical Models 89,* pp. 51-74, 1998.

[18] http://www.soft.uni-linz.ac.at/Research/Context_Database/index.php.

[19] Y. Bengio and Y. Grandvalet, "No unbiased estimator of the variance of k-fold cross validation," *Journal of machine learning research*, vol. 5, pp. 1089-1105, September 2004.

1 차원고접수 : 2008. 10. 14
2 차원고접수 : 2008. 12. 17
최종게재승인 : 2009. 10. 20

(*Abstract*)

Learning Predictive Models of Memory Landmarks based on Attributed Bayesian Networks Using Mobile Context Log

Byung-Gil Lee Sungsoo Lim Sung-Bae Cho
Department of Computer Science, Yonsei University

Information collected on mobile devices might be utilized to support user's memory, but it is difficult to effectively retrieve them because of the enormous amount of information. In order to organize information as an episodic approach that mimics human memory for the effective search, it is required to detect important event like landmarks. For providing new services with users, in this paper, we propose the prediction model to find landmarks automatically from various context log information based on attributed Bayesian networks. The data are divided into daily and weekly ones, and are categorized into attributes according to the source, to learn the Bayesian networks for the improvement of landmark prediction. The experiments on the Nokia log data showed that the Bayesian method outperforms SVMs, and the proposed attributed Bayesian networks are superior to the Bayesian networks modelled daily and weekly.

Keywords : Context log, Landmarks, Attributed Bayesian network, Bayesian structure learning