

모바일 디바이스 사용자의 행동 패턴 분석 및 요약

정명철⁰ 조성배

연세대학교 컴퓨터학과
mcjung@scslab.yonsei.ac.kr⁰, sbcho@cs.yonsei.ac.kr

Analysis and Summary of User's Behavior Patterns in Mobile Devices

Myung-Chul Jung⁰, Sung-Bae Cho
Department of Computer Science, Yonsei University

요약

최근 모바일 디바이스의 기능이 다양해지면서 현대인에게 없어서는 안 될 필수품이 되었다. 모바일 디바이스의 사용영역이 넓어지면서 늘어나는 개인 정보의 활용에 대한 관심이 집중되고 있다. 본 논문에서는 모바일 디바이스에서 사용자의 행동 패턴 분석 및 요약용 위한 지능형 에이전트를 제안한다. 사용자의 다양한 행동 및 상태 패턴 분석을 위해 협력적 모듈 베이지안 네트워크를 사용한다. 협력적 모듈 베이지안 네트워크는 비슷한 유형의 패턴끼리 모듈로 설계해 상호 협력적으로 작동하여 사용자의 특이성을 추론한다. 사용자에게 기억에 남을 만한 특이성을 선택하기 위해 Noisy-OR gate를 적용하여 계산한 특이성 간의 연결 강도와 특이성의 우선순위를 바탕으로 사용자의 하루 동안의 행동을 요약하여 구성한다. 추론을 위한 프로토타입을 작성하고 시나리오를 바탕으로 제안한 방법의 유용성을 보인다.

1. 서론

최근 모바일 디바이스의 기능이 다양해지면서 현대인에게 없어서는 안 될 필수품이 되었다. 전화기능이나 SMS기능 외에 사진을 찍거나 음악을 듣고 TV를 보는 등 다양한 일을 모바일 기기 하나로 할 수 있게 되었다. 디지털 기기들이 결합하는 현상을 디지털 컨버전스(Digital Convergence)라고 하며, 모바일 디바이스의 사용영역이 넓어지면서 기기에 저장되는 개인 정보의 양도 늘어나게 되었다. 또한 모바일 디바이스는 하루 종일 사용자와 함께 하기 때문에 사용자의 일상에 대한 다양한 정보를 수집할 수 있다. 사용자의 다양한 정보를 활용하는 방법에 대한 많은 연구 중 기억 보조자로서의 모바일 디바이스의 역할을 부각시켜 사용자가 과거의 특정 시점에 발생한 일을 기억할 수 있도록 돕는 시스템에 대한 연구가 활발히 진행 중이다.

사용자의 생활 속의 정보들을 효율적으로 분석하고 재가공하여 사용자의 정보 관리를 위한 서비스에 관한 연구가 활발히 진행되고 있다. R. Miiikkulainen은 스크립트 기반의 기억 정보를 계층적 SOM(Self-organizing Feature Maps)으로 분류, 저장, 복원하는 에피소딕 메모리(Episodic Memory)를 제안하였다 [1]. E. Horvitz 등은 사람의 인지 행동을 베이지안 네트워크(Bayesian Network)로 모델링하고 응용 소프트웨어에 적용하였다 [2]. 컴퓨터 상의 온라인 캘린더 정보를 토대로 기억의 실마리가 될 수 있는 핵심 이벤트 정보를 자동으로 탐지하는 학습 모델을 제시하였다. Nokia의 LifeBlog는 휴대폰에서 찍은 사용자의 사진, 동영상, SMS, MMS 등을 사용자의 코멘트를 붙여 연대기순으로 사용자의 일상을 정리를 해주는 서비스이다 [3]. 생성된 블로그는 PC나 휴대폰으로 접속하여 손쉽게 웹블로그를 통해 타인과 자신의 일상을 공유할 수 있다.

본 논문에서는 모바일 디바이스에서 사용자의 로그 정보를 바탕으로 사용자의 행동 패턴 분석 및 요약용 위한 지능형 에이전트를 제안한다. 사용자의 로그 수집을 위해 모바일 디바이스에서 가능한 로그 정보들을 정의하고 수집된 저수준의 로그 정보를 행동 추론 모델에 입력으로 사용하기 위해 전처리 단계를 거친다. 사용자의 다양한 행동 및 상태 패턴 분석을 위해 협력적 모듈 베이지안 네트워크를 제안한다. 사람의 일상을 및

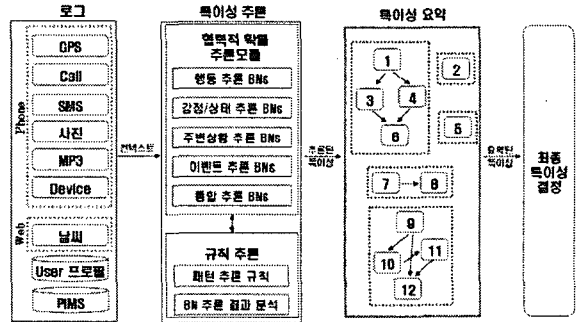


그림 1. 제안하는 방법의 개요

개의 베이지안 네트워크로 설계를 하려면 구조가 굉장히 복잡해 설계하기 힘들어진다. 이를 극복하기 위해 베이지안 네트워크를 비슷한 유형의 패턴끼리 모듈로 설계해 상호 협력적으로 작동하도록 구성한다. 각 모듈의 기본 추론 단계와 모듈간 상호 협력 추론 단계를 거쳐 특이성이 결정되도록 한다. 하루 동안 로그 정보를 바탕으로 뽑힌 많은 특이성들이 사용자에게 모두 유용한 정보는 아니다. 이 중에서 기억에 남을 만한 정보를 뽑아내기 위해 Noisy-OR gate를 이용한 특이성 간의 연결 강도와 특이성의 우선순위를 바탕으로 사용자의 하루 동안의 이야기를 구성하는 방법에 대해 설명한다. 그림 1은 제안하는 방법의 개요를 나타낸다.

2. 행동 패턴 분석

2.1 특이성 추론

사용자의 특이성 추론을 위해서 수학적 확률 분포 모델인 베이지안 네트워크를 사용하였다. 베이지안 네트워크는 도메인 지식을 쉽게 반영할 수 있는 장점을 가지며, 방향성 비순환 그래프(directed acyclic graph, DAG)의 형태를 취한다[4]. 그래프에서 노드는 변수를, 노드간의 연결은 확률적인 의존관계를 의미한다. 베이지안 네트워크는 분류 문제를 속성 노드와 결과

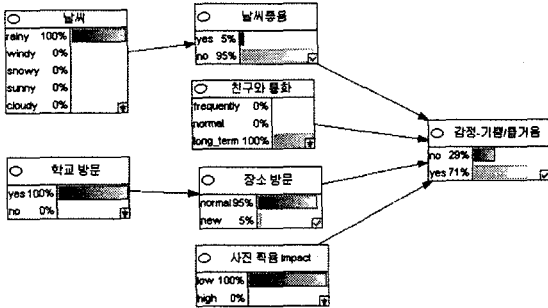


그림 2. 즐거움 추론을 위한 베이저안 네트워크 예
노드간의 확률 관계를 가정하며, 이로 인해 인과 관계를 갖는 추론 모델링에 장점을 갖는다.

그림 2는 즐거움에 관련된 특이성을 추론하기 위한 베이저안 네트워크 일부이다. 저수준 로그로부터 분석된 장소 정보를 바탕으로 날씨, 학교 방문, 친구와의 긴 통화, 사진 찍기를 베이저안 네트워크의 증거값으로 입력한다. 결과는 즐거움이 71% 확률로 추론 되었다.

2.2 협력적 모듈 베이저안 네트워크

사람의 일상의 행동 및 감정 상태는 광범위해 몇 개의 베이저안 네트워크로 모델링하려면 설계가 복잡해서 정확한 모델링이 어려워진다. 본 논문에서는 제한적 성능의 모바일 환경에서 베이저안 네트워크의 사용 및 학습을 효율적으로 수행하기 위한 협력적 모듈 베이저안 네트워크를 제안한다. 비슷한 종류의 특이성별로 분류하여 베이저안 네트워크를 설계하고 모듈화 된 베이저안 네트워크는 모듈별 확률 연산을 수행하고 결과 확률을 다른 모듈의 증거로 재사용하도록 한다. 그림 3은 협력적 모듈 베이저안 네트워크의 연산 과정을 보여준다.

협력적 모듈 베이저안 네트워크는 전문가 설계의 정확도를 높이기 위해 비슷한 유형의 특이성들끼리 그룹화해 베이저안 네트워크를 구성하였다. 베이저안 네트워크를 이용한 확률 추론은 크게 행동 추론, 감정 및 상태 추론, 주변상황 추론, 이벤트 추론으로 분류된다. 각 분류별 베이저안 네트워크의 구체적인 추론 내용은 표 1과 같다.

3. 사용자의 행동 특이성 요약

3.1 인과 관계 추론

특이성 간의 연결 관계는 시나리오를 구성하는데 중요한 척도가 된다. 인과 관계의 연결 강도를 계산하기 위해 Noisy-OR gate 확률 테이블을 이용한 노드간 연결강도 추출 방법을 제안한다. Noisy-OR gate의 확률 계산은 일반 베이저안 네트워크의 확률 테이블의 근사에 의한 독립적 확률 연산 모델이다. 일반적으로, Noisy-OR gate는 설계 및 학습 비용을 줄이기 위해 사용하는 베이저안 확률 테이블 계산 방법이다[5]. 노드의 원인이 여러 개가 들어올 경우 각각이 독립적으로 반영된다는 가정 하에 개별적인 인과 확률만 정의를 한다. 설계자는 식(1)과 같이 p_i 값에 대해 정의를 한다.

$$p_i = \Pr(x_i | x_1, x_2, \dots, x_i, \dots, x_{n-1}, x_n) \quad (1)$$

식(2)는 여러 원인에 대한 확률값을 자동으로 연산하게 된다. 이때, Leak 은 모든 원인이 발생하지 않았을 경우의 확률이 된다.

$$\Pr(x_i | X) = 1 - (1 - p_{i, Leak}) \prod_{i: x_i \neq Leak} (1 - p_i) \quad (2)$$

그림 4는 노드 "짜증 SMS"의 인과 노드인 "바쁜시간"과 "스팸문자"와의 연결 강도 계산을 위해 Noisy-OR이 적용된 베이

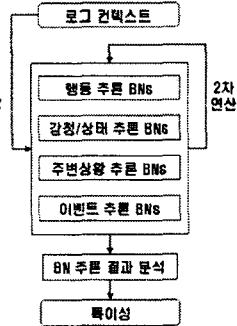


그림 3. 협력적 모듈 베이저안 네트워크의 작동 순서

표 1. 협력적 모듈 베이저안 네트워크의 설계 내용

BN 종류	BN	노드개수	추론 내용
행동 추론	집	22	집에서 하는 일
	휴식	15	휴식을 취하는 방법
	연락	26	전화와 SMS 관련
	사진	23	찍은 사진의 종류
	일터	10	회사에서 하는 일
	운동	34	다양한 운동 중에 실제 한 운동
	학교	32	학교에서 일어난 일
	유흥	13	누구와 뭐하고 놀았는지
	외식	20	외식의 종류
	쇼핑	8	쇼핑을 했는지 여부
	모임	17	모임의 종류와 누구와 만났는지
	병원	5	일원 또는 병원안 여부
	관람	19	운동, 영화, 공연 등의 관람 여부
	자연	15	산책, 등산, 낚시 등의 행동 여부
	종교	10	종교 활동 여부
	교통	14	이동 수단
	음악	13	음악을 들으면서 하는 일
이동행동	9	걷기, 뛰기 여부	
감정 및 상태 추론	놀람	3	놀랐는지
	당황	4	당황하는 상황
	바쁨	11	어떤 일로 바쁜지
	우울	15	기분이 우울 할 때의 상황
	즐거움	23	즐거운 상황
	짜증	10	짜증나는 상황
	추움	5	추웠는지
	피곤	8	피곤한 상황
화남	6	화났는지	
주변 상황 추론	시간적	35	일반적인 시간대의 의미
	공간적	10	어떤 장소에 있는지
	날씨	26	날씨 분석
	기기 상태	14	장치 상태 분석
이벤트 추론	그룹 상태	20	혼자인지 특정 그룹과 함께 인지
	기념일	13	기념일 정보
	행사	34	특별한 행사

저안 네트워크 테이블의 예이다. "짜증 SMS"의 원인인 "스팸문자"와 "바쁜시간" 정보를 활용해 노드 사이의 연결 강도를 계산하는데 "짜증 SMS"가 "Yes" 일 확률은 스팸문자가 "Yes" 일 때는 63%, "바쁜시간"이 "Yes"일 때는 53%가 되고 이 값이 연결 강도가 된다.

3.2 특이성 선택

본 논문에서 설계한 협력적 모듈 베이저안 네트워크를 사용하면 하루에도 수십 개의 특이성이 나올 수 있다. 특이성 중에서 사용자에게 기억에 남을 만하면서 남들에게 자신의 일상생활의 요약을 쉽게 이해시킬 수 있는 특이성을 선택해야 한다. 위

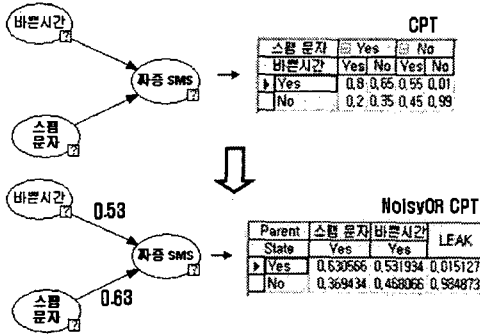


그림 4 연결 강도 계산

조건을 충족시키기 위해서 다음 두 가지 원칙을 가지고 사용자의 요약된 하루 일과의 시나리오를 구성 할 필요가 있다. 첫째, 개별적인 특이성 정보가 사용자에게 기억에 남을 만한 행동이어야 한다. 둘째는, 시나리오로 선택된 특이성을 연결했을 때 이야기의 흐름을 유지해야 한다.

사용자의 일상을 요약하기 위해서 각 특이성 노드별 가중치와 정확도, 각 노드 사이의 인과 관계 강도를 바탕으로 시나리오 구성한다. 우선, 추론된 특이성 간의 연결 관계를 바탕으로 연결 관계에 있는 특이성들을 그룹화 한다. 그룹화된 특이성들은 식(3)을 적용하여 인과 관계를 바탕으로 가중치를 적용한다.

$$np = np + \sum_{i \in stk} st_i \times lp \times gw \quad (3)$$

np : 노드 자신의 우선순위
 st : 연결강도
 lp : 링크로 연결된 노드의 우선순위
 gw : 그룹의 가중치

기본적으로 각 특이성의 우선순위를 기준으로만 특이성을 선택하다 보면 요약된 이야기의 흐름이 부자연스러운 경우가 많다. 요약된 이야기의 흐름이 논리적으로 일관성을 갖게 하기 위해서는 관련된 특이성들끼리 묶여야 좋은 이야기가 구성이 된다. 따라서 우선순위가 높으면서 이와 관련된 특이성들을 선택하기 위해 연결 관계에 있는 특이성의 우선순위를 높여주기 위해 가중치를 전파하는 방법을 사용하였다. 가중치 전파가 완료되고 최종적으로는 우선순위 값을 기준으로 필요한 만큼의 특이성을 선택한다. 선택되는 특이성의 개수는 사용자의 선호도를 바탕으로 한다.

4. 시나리오 작동 예

사용자의 시나리오는 그림 5와 같다. 간단히 살펴보면, 사용자는 아침에 학교에 가서 수업을 듣고 레스토랑에서 식사를 한다. 식사를 마치고 산책을 하는 도중에 오랜만에 친구에게 전화 연락이 온다. 친구는 공연 티켓이 생겼다고 함께 가자고 한다. 대학로의 공연장으로서 친구와 즐겁게 공연을 보고 간단히 맥주 한 잔을 했다. 오랜만에 친구들을 보니 반가움 감정이 생긴다. 위 시나리오에서 사용자에게 가장 기억에 남을 만한 특이성은 공연이다. 공연을 중심으로 이전에 친구에게 걸려온 전화와 공연 후 친구들과의 술자리, 이 때 느끼는 반가움의 감정 등이 일기의 내용이 되는 특이성으로 구성 될 수 있을 것이다.

그림 11을 보면 실제 추론된 특이성은 11개고 이 때 선택된 특이성은 6개이다. 6개 중에는 사용자에게 기억에 남을 공연과 관련된 특이성 정보들이 4개다 선택이 되었고 2개는 우선순위를 따라 선택이 되었다.

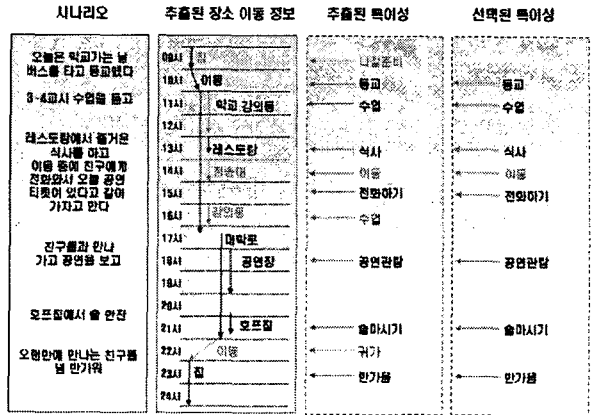


그림 5. 시나리오 및 선택된 특이성

5. 결론 및 향후 연구

모바일 디바이스는 이제 현대인에게 없어서는 안 될 필수품이 되었고 그 안에 담겨지는 사용자 정보의 양도 하드웨어의 발전에 따라 개인이 관리하기 힘들 정도로 많은 정보가 저장되었다. 본 논문에서는 모바일 디바이스의 사용자 정보를 바탕으로 사용자의 행동 패턴 분석 및 요약을 위한 지능형 에이전트를 제안하였다. 사람의 일상 생활을 분석하기 위해서 하나의 베이지안 네트워크로는 한계가 있어 협력적 모듈 베이지안 네트워크를 제안하여 사용자 행동 패턴 분석을 위한 모델링을 하였다. 베이지안 네트워크로부터 얻어지는 확률값과 각 특이성 간의 관계에 대한 연결 강도를 Noisy-OR gate를 적용하여 획득하였다. 특이성의 우선순위와 연결강도를 바탕으로 특이성간의 그룹화 및 최종적으로 적합한 특이성을 선택하여 사용자의 하루 행동을 요약하였다. 제안한 방법을 바탕으로 모바일 디바이스에서 작동하는 프로토타입을 구현하고 시나리오를 바탕으로 작동 가능함을 보였다.

향후 연구로는 현재 수집된 로그 정보로 다양한 사용자의 행동 패턴을 분석하는 부분에는 한계가 있다 이를 극복하기 위해 로그 수집을 위해 사용되고 있는 센서의 추가와 사용자 어플리케이션 로그의 종류를 더 확장해야한다. 또한 확장된 로그 정보를 바탕으로 현재 추론하지 못하는 사용자 행동을 위한 베이지안 네트워크의 확장이 필요하다.

참고문헌

- [1] R. Mikkilainen, "Script recognition with hierarchical feature maps," *Connection Science*, vol. 2, pp. 83-101, 1990.
- [2] E. Horvitz, S. Dumais, P. Koch. "Learning predictive models of memory landmarks," *CogSci 2004: 26th Annual Meeting of the Cognitive Science Society*, Chicago, August 2004.
- [3] Nokia LifeBlog, <http://www.nokia.com/lifeblog>
- [4] Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.
- [5] A. Onisko, M. J. Druzdzel, and H. Wasyluk, "Learning bayesian network parameters from small data sets: Application of noisy-OR gates," *12th European Conference on Artificial Intelligence (ECAI-2000)*, 2000.