

모바일 기반 라이프로그를 이용한 사용자 행동 예측 기법[†]

(A user behavior prediction technique using
mobile-based Lifelog)

방재근¹⁾, 김병만²⁾
(Jae-Geun Bang and Byeong Man Kim)

요약 최근 많은 사람들이 스마트폰을 이용해 다양한 방법으로 원하는 정보를 추천 받고 있다. 그와 관련해 추천을 위한 많은 어플리케이션이 존재하지만, 현재 사용자 상황에 맞는 정보를 추천해 주는 것은 없다. 자동으로 사용자의 상황에 맞는 추천을 하기 위해서는 사용자의 과거 행위이력으로 부터 미래의 행위를 예측할 필요가 있다. 이에 본 논문에서는 스마트폰을 이용해 사용자의 현재 상황을 수집하고, 수집된 데이터를 라이프로그를 분석하여 구축한 베이저안 네트워크에 적용하여 현 행동을 판별한 후 연관분석을 통해 사용자가 미래에 하게 될 행동을 예측하는 방법을 제안한다. 5명의 실제 학생과 5명의 가상의 직장인에 대해서 실험 및 분석해 본 결과 그 유용성을 확인할 수 있었다.

핵심주제어 : 라이프로그, 베이저안 네트워크, 연관 분석

Abstract Recently the desired information has been recommended to many people in a number of ways using the smartphone. Though there are many applications for that purpose, but most applications does not consider the user's current situation. In order to automatically recommend the information considering the user's situation, it is necessary to predict the future behavior of the user from the records of the past behavior of the user. Therefore, in this paper, we propose a method that predicts the user's future behavior through association analysis based on the user's current behavior which is identified by applying the user's current situation data collected via a smartphone to the Bayesian network built from the user's life log. From the experiments and analysis for five students and five virtual workers, the usefulness of the proposed method is confirmed .

Key Words : Lifelog, Bayesian network, Association analysis

1. 서론

미국 시장조사기관 스트래티지 애널리틱스(Strategy

Analytics)의 보고서에 따르면 지난해 한국의 스마트폰 보급률은 67.6%로 전 세계 국가 중 가장 높은 것으로 나타났다. 이는 세계 평균(14.8%)의 4.6배에 이를 만큼 현재 국내 대다수의 사람들이 스마트폰을 이용하고 있다[1]. 스마트폰을 이용하여 전화통화와 문자메시지 전송, 음악과 동영상 감상, 사진촬영과 같은 일상

[†] 이 연구는 금오공과대학교 학술연구비에 의하여 지원된 논문임.

1) 금오공과대학교 컴퓨터소프트웨어공학과, 제1저자

2) 금오공과대학교 컴퓨터소프트웨어공학과, 교신저자

생활 속에서 필요한 서비스를 제공받고, 다양한 어플리케이션을 이용하고 있다. 그 중에서도 쇼핑몰, 배달 음식점, 위치정보 등 추천에 관련된 어플리케이션을 많이 이용하고 있다. 하지만 이와 같은 추천 어플리케이션들은 사용자의 현재 상황을 고려하지 않아 추천의 질의 떨어지는 문제점이 있다.

상황인식 응용들은 사용자의 의식적인 입력이 없어도 특정한 시간과 장소에서 사용자의 행동을 추정하여 정보를 전달하거나 요청한다. 이 때 사용자의 행동을 예측하기 위해 센서를 사용한다[2]. Randell & Muller[3]는 Tourist Guide 프로젝트에서 사용자가 여행지의 특정한 위치에서 사용자의 활동 상태에 따라 여행지 정보와 서비스를 제공하는 연구를 수행 하였다. 모바일에 다양한 센서를 활용하면 현재 사용자의 행위 및 일정에 따라 능동적으로 다양한 서비스를 제공할 수 있다.

라이프로그(Lifelog)라는 용어는 1945년 7월 미국국 과학기술연구소 소장인 Vannevar Bush가 Atlantic Monthly지에 기고한 “As We May Think”라는 기고문에서 처음 언급 하였다. Vannevar Bush[4]는 이마에 부착하는 호두알만한 크기의 카메라와 소형 녹음기를 가지고 개인이 보고 듣고 말하는 것을 영구 기록하여 기억을 확장 하고자 하는 생각을 하였다. 현재는 저장 공간의 가격이 낮아지고 스마트폰(Smart Phone)의 출현으로 자연스럽게 소형화된 고성능 센서가 일반 대중에게 보급 되어 보다 쉽게 라이프로그를 수집할 수 있게 되었다.

라이프로그(Lifelog)에 관련한 연구는 국내외에서 다양하게 진행되고 있으나, 대부분의 연구들이 스마트폰(Smart Phone)과 웨어러블(Wearable) 형태의 장비를 이용하여 라이프로그를 수집 할 수 있는 방법들과 상황에 맞는 적절한 서비스를 제공하는데 초점이 맞추어져 있다. 사용자에게 상황에 맞게 적절한 내용을 자동으로 추천해 주기 위해서는 먼저 현 상황을 인식하고 이를 기반으로 사용자의 다음 행위를 사용자의 과거이력으로부터 예측할 필요가 있다.

이에 본 논문에서는 모바일 기기를 이용해 사용자의 현재 상황을 인식하고 수집된 라이프로그를 기반으로 베이지안 네트워크를 통하여 현 행위에 대한 카테고리를 판별한 후 연관분석을 통해 사용자가 미래에 하게 될 행동에 대해 예측하는 방법을 제안한다.

2. 관련 연구

이영설[5]은 모바일 기기, 그 중에서도 스마트폰에 저장된 개인정보를 이용하여 실질적으로 사람의 행동 패턴을 예측하거나 분석할 수 있는지 검증하였고, 행동 이론의 요소를 바탕으로 설계된 콘텍스트 계층 구조를 이용하여 확률 모델을 설계하였다. 스마트폰을 이용해 수집하는 정보로 행동, 통화내역, GPS 정보, MP3 정보, 사진 정보, 장소 정보 SMS 정보 등을 이용 하였고, 설계된 콘텍스트 계층 구조에 적용 하여 사용자가 취할 수 있는 여러 행동 중 가장 높은 가중치를 가진 행동을 추측 하였다.

홍성용[6]은 빅데이터 마이닝을 이용해 IT학습자의 집단지성 학습경로를 분석 예측하고 추천 할 수 있는 방법을 제시하고 실현 가능성을 위한 시험 및 구현을 하였고, IT학습자들의 학습 성향을 지속적으로 저장관리하며 연관분석을 통한 학습 단계의 예측과 추천을 가능하도록 제시했다. 결과로 IT학습자들에게 좀 더 유용한 지식과 학습경로 결정에 도움을 줄 수 있을 것으로 기대하고, 대용량의 빅데이터를 실시간적으로 처리하여 제공하기 위해서 100M 이하의 단위로 분산 병렬처리를 하였다.

장영완[7, 8]은 모바일 기반에서 수집된 라이프로그 레코드를 설명할 수 있는 태그를 자동으로 부착할 수 있도록 각각의 라이프로그를 이동활동, 청각활동, 시각활동, 기록활동, 말하기활동의 5가지 사용자 활동으로 분류하고 관련된 일련의 사용자의 활동을 지속상황, 감지상황, 선택상황의 세 가지 상황으로 분류하여 각각의 라이프로그를 설명할 수 있는 대표 키워드를 부여하는 방법을 제안했다. 제안하는 방법을 활용하여 라이프로그의 수집시 별다른 연관성 없이 대량으로 무분별하게 수집되는 데이터에 대하여 자동으로 레코드의 의미를 설명하는 키워드를 태깅할 수 있고, 자동으로 부착된 태그를 이용하여 데이터를 탐색하거나 시간의 흐름에 따라 기록된 데이터의 의미를 명확하게 함으로써 향후 과거 이력을 이용한 상황 추론과 같은 과거에 발생하였던 사건을 이용하여 현재 상황을 추론할 수 있다. 또한 유클리디안 거리 유사도를 이용하여 스마트폰의 사용자의 움직임을 인식하고 판별하기 위한 방법을 제안 및 실제 스마트폰을 이용해 실험 하였다.

카네기멜론 대학의 K. Dey[9]는 Mem - exerciser

를 개발하였으며 이를 기억력 손상환 자에게 적용 및 실험 하였다. 이 시스템은 카메라, 오디오 레코더, GPS, 가속도, 조도, 온도센서를 장착한 로깅장비와 이에 대하여 위치, 운동, 음성, 얼굴 등을 맥락단서로 하여 손쉽게 검색, 재생할 수 있는 프로그램으로 구성된다.

홍민성[10]은 스마트폰 디바이스 사용과 이동 상황을 판별하는 알고리즘을 제안하여 가속도 센서 데이터를 통해 사용자의 이동 상황을 판별하는 실험을 하였다.

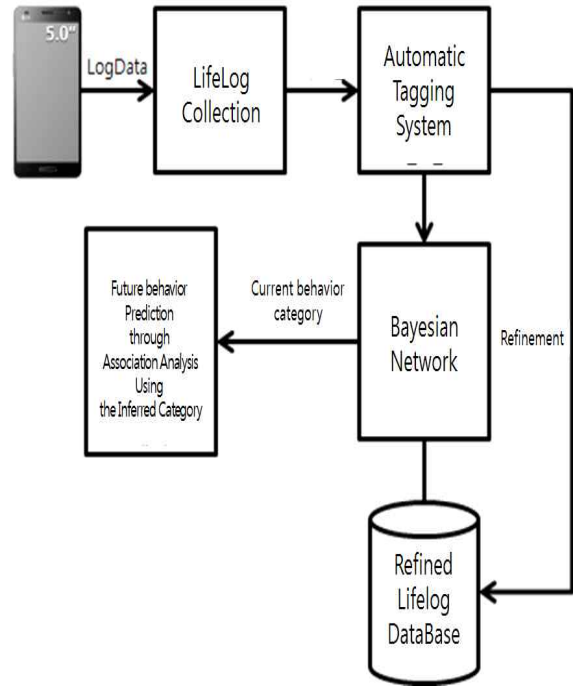
정도준[11]은 실내 환경에서 발생할 수 있는 사람의 행동을 인식하는 방법을 제안했다. 시선 방향을 머리와 얼굴의 관계를 표현한 베이지안 네트워크 모델을 이용하여 추정 하였고, 이벤트와 사람의 행동을 나타내는 시나리오를 인식하여 사람의 상태변화를 인식하고, 이벤트들의 조합과 제약 사항을 이용해 규칙 기반으로 행동을 인식하였다.

본 연구는 선행 연구자들의 연구를 참고하여 모바일 기기를 이용해 사용자 라이프로그를 수집하고, 수집된 라이프로그들을 베이지안 네트워크 방법과 그 후 연관 분석을 통해 사용자가 미래에 하게 될 행동에 대해 예측 하는 방법을 제안한다.

3. 제안하는 방법

3.1 시스템 구조

<Fig. 1>은 제안하는 맞춤형 행동 예측 기법을 도식화한 것이다. 먼저 사용자의 현재 행동을 예측해야 하므로, 대표적인 모바일기기인 스마트폰을 이용해 사용자의 라이프로그를 수집해야 한다. 본 논문에서는 선행 연구인 “모바일 사용자 상황정보를 이용한 퍼스널 라이프로그 자동 태깅 방법”[7]을 적용해 얻은 일화 데이터를 라이프로그로 사용한다. 사용자 라이프로그는 1분에 한 번씩 로그데이터로 기록을 남기고 하루에 한번 자동태깅 시스템을 통해 일화 데이터를 생성한다. 일화 데이터는 다음 <Table 1>과 같은 데이터 형식을 갖는다.



<Fig. 1> Behavior prediction based on mobile lifelog

<Table 1> Original Lifelog

| Time | Activity | Situation | Tag |
|-----------------------------------|---|---|--------------------|
| A.M 00:00 ~ P.M 23:00 | Moving, Listening, Watching, Writing, Talking | Persistence, Detection, Selection | Tag Information |

<Table 1>의 이동활동으로 알 수 있는 정보는 현 사용자의 움직임이 정지, 걷기, 뛰기, 차량이동 중인지 알 수 있고, 청각 활동, 시각 활동, 기록 활동, 말하기 활동으로 현재 핸드폰의 사용 중인 기능에 대해 알 수 있다. 이 시스템에서 관심을 갖는 핸드폰의 기능은 노래듣기, 동영상보기, 사진촬영, 전화송수신, 문자메시지 송수신이며, 이중 노래듣기, 동영상보기, 전화송수신은 청각활동에 해당되며, 동영상보기, 사진촬영, 문자메시지 송수신은 시각활동에, 문자메시지 송수신, 사진촬영은 기록활동에, 전화송수신은 말하기활동에 해당된다. 나머지 지속상황, 감지상황, 선택상황, 태그 정보에 대해서는 본 논문에서는 사용하지 않는다.

<Table 2> Refined lifelog

| Moving | Time | Phone Function | Location |
|----------------------------|-----------|-------------------------------------|----------|
| Stop, Walk, Run, Car | A.M | Music, Video, Picture, Call, SNS | Location |
| | 00:00 ~ | | |
| | P.M 23:00 | | |

본 논문에서는 <Table 1>과 같이 정의된 기존의 사용자 라이프로그를 <Table 2>와 같이 정제된 라이프로그를 사용한다. 정제된 라이프로그는 사용자 별로 데이터베이스에 저장되며 저장된 이 라이프로그를 사용하여 행동 카테고리를 예측하기 위한 베이지안 네트워크 모델이 구축되고 이 모델을 사용하여 사용자의 현 행동에 대한 카테고리가 추론된다. 그 후 추론된 행동 카테고리를 이용하여 연관분석을 통해 사용자가 미래에 하게 될 행동을 예측한다.

3.2 사용자의 현재 행동 추론

3.2.1 베이지안 네트워크

사용자의 현재 행동을 추론하기 위해 본 논문에서는 베이지안 네트워크 모델을 구축한다. 베이지안 네트워크는 각각 수치화된 가중치(weight), 상태(state) 간의 천이(transition)확률 등으로 지식을 표현할 수 있고, 상태 분할이 올바르게 이루어졌을 경우, 주어진 데이터에 대한 확실적인 특성을 효과적으로 모델링할 수 있는 장점이 있다. 또한, 불확실한 상황에서 추론[12]을 위한 대표적인 모델로써, 방향성 비순환 그래프 모양을 가진다. 베이스 정리를 통해 미리 계산된 사건 확률 값을 바탕으로 최후 사후 확률 값을 추론한다. 베이스 정리(1)는 다음과 같다.

$$P(E) = \frac{p(E|H_i) * p(H_i)}{\sum_{k=1}^n p(E|H_k) * p(H_k)} \quad (1)$$

$i = 1, 2, 3 \dots n$

여기서 사건 E는 어떤 사건이 일어났을 경우를 나타내고, 조건부 H는 추론 하고 싶은 대상을 의미한다. E

를 이동활동 중 정지활동이라 가정하고 H₁운동, H₂ 공부, H₃수면이라 가정했을 때, 현재 사용자가 정지 상태일 경우 운동, 공부, 수면을 하게 될 확률 값을 추론할 수 있다. 또한 사건 E는 1가지 외에도 <Table 2> 에서 정의된 라이프로그들이 모두 사건 E가 될 수 있다.

3.2.2 데이터 카테고리화

행동이나 장소는 너무 다양하여 그대로 사용할 경우 모델을 구축하는데 어려움이 있어 본 논문에서는 행동과 장소를 카테고리화하여 사용하였다. 행동을 통계청의 생활시간조사 행동분류표[13]를 참조하여 <Table 3>와 같이 총 41개의 행동카테고리로 분류하였다. 장소는 총 17개의 카테고리로 분류하였다. 장소 카테고리는 행동 카테고리를 기반으로 카테고리화 하였다.

베이지안 네트워크 모델과 연관분석을 위해서는 각 일화에 해당하는 행동과 장소가 필요한데 본 논문에서는 일화 생성 시 사용자로부터 직접 입력받도록 하였다. 이때 입력받는 행동과 장소로 <Table 3>의 카테고리를 이용한다.

<Table 3> Behavior, location category table

| Class | Category |
|------------------|---|
| behavior (42) | sleep, meal, snack, hygiene, appearance, cure, class, regular work, non-regular work, part-time job, online lecture, lecture, study, cooking, dish-washing, washing, clothing arrangement, clear, item arrangement, voluntary service, newspaper, TV, DVD, ball games, health, jogging, take a walk, religion, book, web surfing, online game, movie, theater, sport view, shopping, date, stop, walk, run, car, bus, train |
| | house, school, library, hospital, restaurant, beauty parlor, laundry, church, bookstore, mart, department store, cafe, load, schoolyard, downtown, office, park |
| location (17) | |

3.2.3 세부 행동 추론을 위한 확률 모델 설계

행동 카테고리 확률 모델이 가지고 있는 노드는 크게 입력노드, 연관노드, 목표노드로 구성 된다. 입력노드는 연관노드와 목표노드를 추론하기 위한 사건의 사전 확률 값이 들어가는 노드를 의미한다. 연관노드는 입력노드와 목표노드의 중간에 위치한 노드로서 목표노드와 연관성이 가장 높은 노드를 의미하며 한 가지 입력노드를 고려하는 것이 아닌 여러 가지의 입력노드를 고려하여 추론 값이 결정되며 결정된 연관노드 추론 값을 바탕으로 다양한 목표노드를 추론 할 수 있다. 연관노드의 값은 <Table 2>에서 추출된 활동 필드에 있는 값으로 지정하였다. 목표노드는 실제 추론되는 노드를 의미하며, 연관노드 추론 값을 참조하여 추론된다. 행동 카테고리 확률 모델의 구성은 <Table 4>와 같이 설계를 할 수 있다.

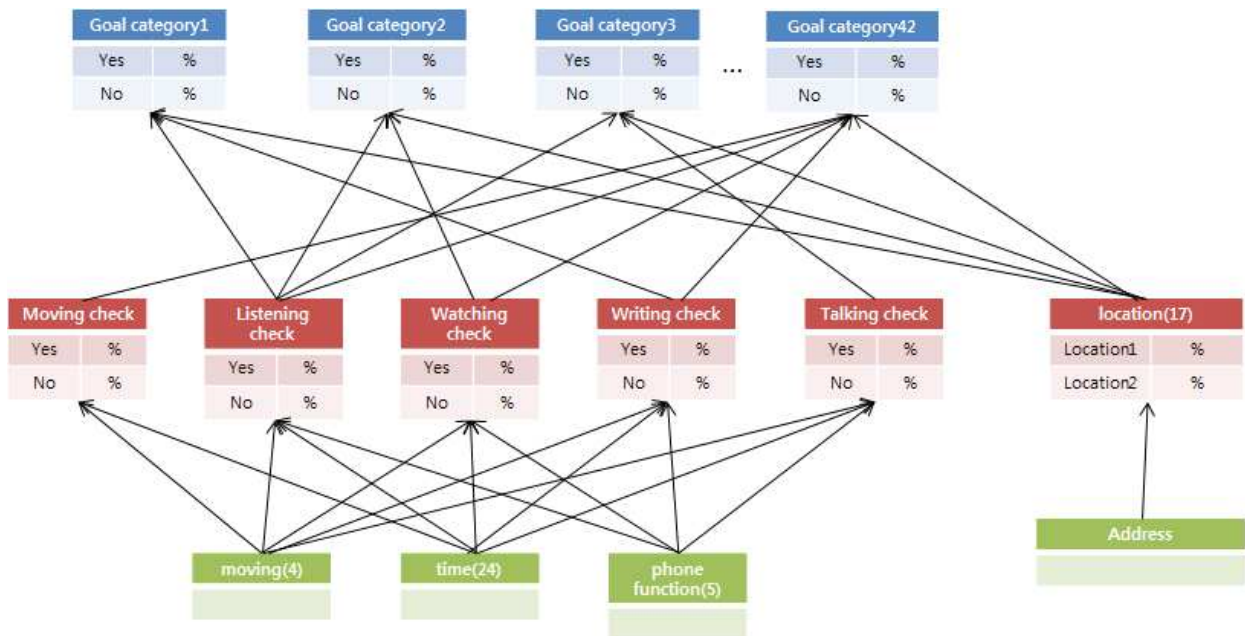
<Fig. 2>는 현재 각 노드들이 연결된 상태를 보여주고 있다. 하위에 있는 노드들이 상위의 노드를 가리키는 형태로 모델링이 되었다.

Location을 제외한 연관노드들은 사용자의 특성에 따라 연결되는 목표노드가 달라진다. 예를 들어 자가 학습에 관한 행동을 하고 있다고 가정한다면, A사용자는 자가 학습을 할 때 주로 집에서 하고 혼자 노래를 듣고 동영상 강의를 시청 한다고 하면 관련 연관

<Table 4> Behavior category bayesian network design

| Node | Content |
|-----------------------|---|
| input node (34) | Moving Activity (이동활동-4), Time(시간-24), Phone function (핸드폰기능-5) |
| association node (21) | Moving check (이동여부), Listening check (청각 여부), Watching check (시각 여부), Writing check (기록 여부), Talking check (말하기 여부) |
| goal node (42) | User behavior category (사용자 행동 카테고리-42) |

노드는 청각여부와 시각여부가 연결이 될 것이고, B사용자는 자가 학습을 할 때 주로 전화, 문자 송수신이



<Fig 2> Bayesian network example

나 사람들과 같이 한다고 하면 청각여부, 시각여부, 말하기여부, 기록여부와 연결 될 것이다. 이처럼 각 사용자의 행동마다 목표노드와 연관노드 연결이 달라 질 수 있다.

입력노드와 연관노드의 연결은 모두 연결이 되지만, Phone function 입력노드는 각각의 핸드폰 기능의 정해진 연관노드와 연결된다. <Table 5>를 참조하면 그와 관련된 자세한 정보를 볼 수 있다. 예를 들어 핸드폰 기능 중 동시에 노래를 들으면서 동영상을 볼 순 없다. 물론 핸드폰 자체는 동영상을 보면서 노래를 들을 수 있도록 여러 프로세스를 실행할 수 있지만, 사용자 입장에선 동영상보기, 노래듣기 중 하나는 집중할 수 없기 때문에, <Table 5>와 같이 정해진 연관노드와 연결된다.

<Table 5> Relation regarding Association node and Input node (Phone function)

| Association node | Input node (Phone function) |
|------------------|-------------------------------|
| Listening check | music, movie, call |
| Watching check | picture, SNS, movie |
| Writing check | SNS, picture |
| Talking check | call |

입력노드 location은 사용자의 특성에 무관하게 연관노드 location에만 연결되고 연관노드 location도 사용자 특성에 무관하게 전체 목표노드와 연결된다. 주소로부터 장소카테고리를 추출하는 방법은 입력노드 location에 “경상북도 구미시 대학로 61” 값이 입력이 되었으면, 베이스 정리를 통해 예측할 수 있다. 예를 들어 “경상북도 구미시 대학로 61” 값에 사용자 라이프로그 데이터베이스를 참조한 결과 “학교”와 “집”에 대해서만 카테고리를 예측 한다고 할 때, 각 장소 카테고리는 학교 59%, 집 41%의 비율을 구할 수 있고, 조건부 확률을 계산해본 결과 학교→“경상북도 구미시 대학로 61”는 82%, 집→“경상북도 구미시 대학로 61”는 5% 로 계산할 수 있다.

베이스 정리를 통해 계산과정은 다음과 같다.

※ 학교의 확률 = (0.82 * 0.59) / 모든 확률의 합

※ 집의 확률 = (0.05 * 0.41) / 모든 확률의 합

입력된 주소 값이 참조 데이터베이스에서 찾지 못할 경우, 조건부 확률 값을 구할 수 없다. 그래서 조건부 확률값은 모두 동일한 비율($\frac{1}{7}$)을 가지게 되고, 나중에 사용자가 직접 수정함으로써, 조건부 확률 값을 계산할 수 있다.

3.3 연관 분석을 통한 사용자의 미래 행동 예측

페이지안 네트워크를 통해 현 행동에 대한 카테고리라 추론이 되면 과거 이력 데이터를 기반으로 다음에 사용자가 취할 행동에 대한 추론이 필요하다. 본 논문에서는 이를 위해 사용자가 어떤 행동 시 다음에 하게 될 경향이 강한 행동을 찾아내 그룹화 한 후 그들 사이의 연관관계를 규칙으로 표기하는 기법인 연

<Table 6> Future behavior prediction algorithm

```

Future behavior prediction algorithm
u_id = Current User Id;
u_X = Predicted current behavior
a_Y[] = All behavior <Table 3>

// 사용자 라이프로그 DB 연동
UserLifeLogDAO dao
    = new UserLifeLogDAO();

// 사용자의 모든 LifeLog 개수
u_count = dao.getUserAllLifeLogCount(u_id);

// 현재 예측된 행동의 라이프로그 개수
X =
dao.getCountCurrentBehavior(u_id, u_X);

for(int i = 0 ; i < Y.size ; i++) {
// 모든 행동의 행동 라이프로그 개수
Y[i] =
dao.getCountCurrentBehavior(u_id, a_Y[i]);

// 지지도, 신뢰도, 향상도 계산
Support[i] = getSupport(X, Y[i], u_count);
Confidence[i] = getConfidence(X, Y[i]);
Lift[i] = getLift(Support[i],
X / u_count, Y / u_count);
}
    
```

관분석을 이용하였다.

예를 들어, “수면”과 “자가 학습”사이의 연관분석을 한다면, 사용자 라이프로그에서 “수면”→“자가 학습”에 관한 정보를 찾아 지지도(Support)와 신뢰도(Confidence), 향상도(Lift)에 관한 비율을 구할 수 있을 것이다. 미래 예측의 알고리즘은 다음 <Table 6>과 같다.

여기서 지지도(Support)는 사용자 라이프로그 중 수면 후 자가 학습을 했던 비율이고, 신뢰도(Confidence)는 사용자가 수면을 했던던 라이프로그 중 수면 후에 자가 학습을 한 비율을 의미하고, 향상도(Lift)는 예측에 있어서 무작위 추측에 비해 규칙이 얼마나 우수한지를 보여주는 비율이고 값이 1보다 크면 예측력이 있다고 평가한다. 따라서 수면과 자가 학습이 서로 독립이라면 향상도는 1이 되는데 향상이 없다는 의미로 판단 할 수 있다. 각 구하는 공식[6]은 (2) (3) (4)과 같다.

$$\text{지지도}(S) : \frac{X \cap Y}{N} \quad (2)$$

$$\text{신뢰도}(C) : \frac{X \cap Y}{X} \quad (3)$$

$$\text{향상도}(L) : \frac{S}{P(X)P(Y)} \quad (4)$$

여기서, X는 “수면”이 나오는 라이프로그 수, Y는 “자가 학습”이 나오는 라이프로그 수, N은 전체 라이프로그 수이다. $P(X)$ 는 수면의 확률, $P(Y)$ 는 자가

학습의 확률을 나타낸다.

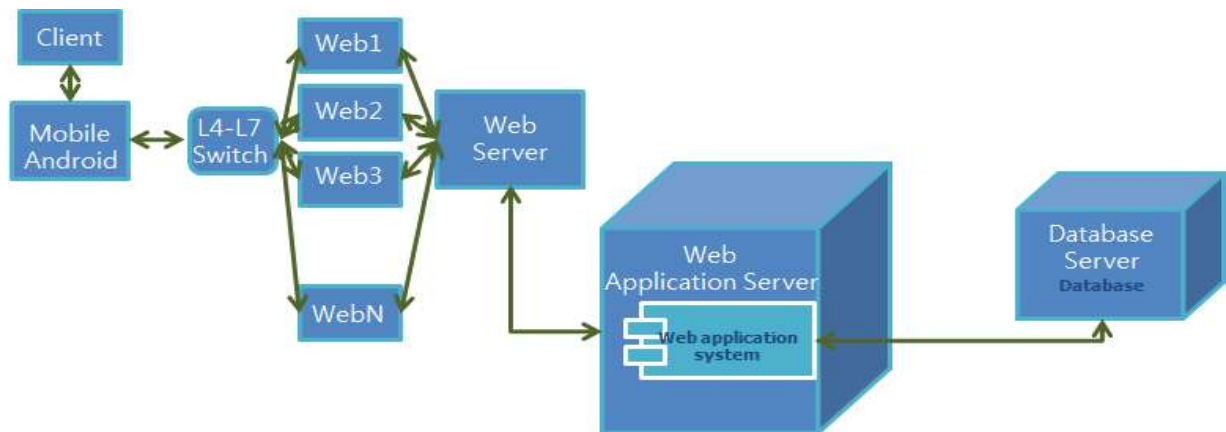
4. 구현과 실험

본 논문에서는 실험을 위해 SM-N900 스마트폰을 이용하였고, 학부생 5명이 실험에 참여 하였다. 학생들의 생활 패턴이 비슷한 관계로 학생 외에 직업을 가진 실험자를 모집해야 되는데 현실적으로 모집하는데 힘들어 본 논문에서는 가상으로 직장인 5명을 생성하였다. 학생들은 약 1달간 실제로 데이터를 수집하였고, 가상의 직장인들은 가상으로 생활패턴을 입력하였다. 학생들에게 수집된 데이터에서 장소와 행동에 대해서는 직접 입력 할 것을 요구하였다.

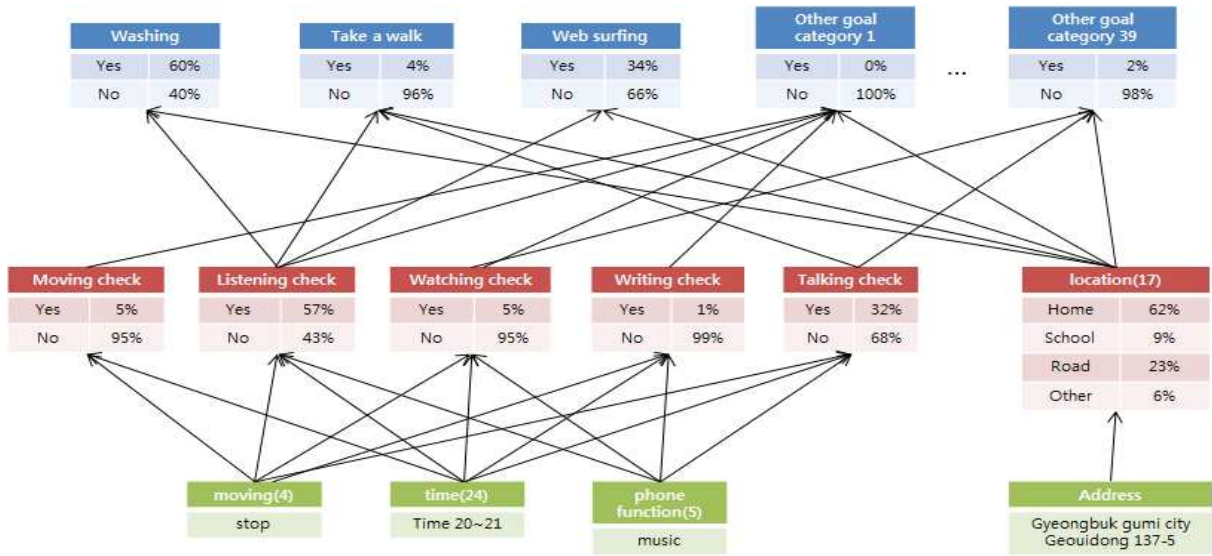
전체 시스템 구조는 다음 <Fig 3>과 같다. 사용자는 웹을 통해 수집된 데이터의 장소와 행동을 직접 입력 할 수 있고, 현재 스마트폰을 통해 추출된 Moving, Time, Phonefunction, location 정보를 Web Application Server로 전송 한다. 전송된 정보와 사용자 과거 라이프로그를 참조해 베이지안 네트워크의 입력 노드로 사용하여 현재 행동을 예측 후, 미래 행동을 예측 한다.

본 논문에서 특정상황에 대한 예측 과정을 보여주기 위해 1명의 학생과 1명의 가상의 직장인의 예시를 제시하였다.

학생과 가상의 직장인의 주요 생활패턴은 다음 <Fig 4>과 <Fig 5>와 같다. 학생과 가상의 직장인이 항상 같은 패턴으로 활동 하는 것이 아니기 때문에



<Fig. 3> System Architecture



<Fig. 6> Bayesian network of an undergraduate

주요 패턴만 나타낸 것이다.

| | |
|------------|-------------|
| Time 0~9 | sleep |
| Time 9~10 | washing |
| Time 10~13 | class |
| Time 13~14 | meal |
| Time 14~18 | class |
| Time 18~19 | meal |
| Time 19~20 | bus |
| Time 20~21 | washing |
| Time 21~22 | TV&DVD |
| Time 22~24 | watching |
| Time 22~24 | web surfing |

<Fig. 4>
An undergraduate

| | |
|------------|--------------|
| Time 0~8 | sleep |
| Time 8~9 | washing |
| Time 9~12 | regular work |
| Time 12~13 | meal |
| Time 13~18 | regular work |
| Time 18~19 | meal |
| Time 19~20 | Health |
| Time 20~22 | book |
| Time 22~24 | web surfing |

<Fig. 5>
A virtual worker

학생의 현 상황(정지, 시간 20~21시, 노래듣기)으로부터 BN(베이지안 네트워크)를 통하여 <Table 7>에 추론된 결과를 보면 사용자가 “씻기”행동을 하고 있음을 판단할 수 있다. 또한, 추론된 사용자의 현재 행동으로부터 연관분석을 통해 1시간 후의 행동을 예측해보면 <Table 8>에 보는바와 같이 “DVD 시청”을 할 확률이 높음을 알 수 있다. <Table 8>의 연관분석 결과는 대표로 2개만 나타냈다.

가상의 직장인의 현 상황(정지, 시간 20~21시, 노래듣기)으로부터 BN을 통하여 <Table 9>에 추론된 결과를 보면 사용자가 “독서”행동을 하고 있음을 판단할 수 있다. 또한, 추론된 사용자의 현재 행동으로부터 연

관분석을 통해 1시간 후의 행동을 예측해보면 <Table 10>에 보는바와 같이 똑같이 “독서”를 할 확률이 높음을 알 수 있다. <Table 10>의 연관분석 결과는 대표로 2개만 나타냈다.

<Fig 6>, <Fig 7>은 각 1명의 학생과 1명의 가상의 직장인의 베이지안 네트워크를 나타낸다. 입력노드는 연관노드에 연결되고 목표노드는 사용자에게 행동에 따라 연결되는 연관노드가 달라진다.

사용자 라이프로그 데이터베이스를 참조하여<Fig 6>의 Location 연관노드 제외한 노드들의 사전확률을 계산해 보면, 이동여부 18%, 청각여부 26%, 시각여부 20%, 기록여부 17%, 말하기여부 19%의 사전확률 값이 계산 되었다. 조건부확률을 계산해 보면, 이동여부의 경우 이동여부→정지 6%, 이동여부→20~21시 32%, 이동여부→노래듣기 53%이고, 청각여부의 경우 청각여부→정지 61%, 청각여부→20~21시 49%, 청각여부→노래듣기 95%이고, 시각여부의 경우 시각여부→정지 54%, 시각여부→20~21시 66%, 시각여부→노래듣기 0%이고, 기록여부의 경우 기록여부→정지 34%, 기록여부→20~21시 21%, 기록여부→노래듣기 0%이고, 말하기여부의 경우 말하기여부→정지 29%, 말하기여부→20~21시 54%, 말하기여부→노래듣기 0% 였다. 여기서, 핸드폰기능의 노래듣기 기능은 <Table 5>와 같이 연관노드인 이동여부, 청각여부에만 포함 되고, 나머지 연관노드는 참조되지 않는다. 베

이즈 정리 특성상 참조되지 않으면 0%가 나올 수 있어, 그것을 방지하기 위해 다음 아래와 같은 공식을 적용한다.

기본 가중치 공식 :

$$1 - \alpha) * w + \alpha \quad (5)$$

w : 중치
 α : 기본가중치 (0.1)

위 식(5)의 α 는 기본가중치로 본 논문에서는 0.1의 기본가중치 값을 주었다. 0%의 가중치 값이 계산되어도 최소 10%의 기본가중치를 적용해준다는 의미이다. 본 논문에서 기본 가중치 공식은 Location을 제외한 연관노드에만 적용 했다.

각각의 조건부확률 값을 기본 가중치 공식을 적용 후, 베이즈 정리를 통해 계산한 과정은 다음과 같다.

- ※ 이동여부의 확률 = $(0.26 * 0.154 * 0.388 * 0.577)$ / 모든 확률의 값
- ※ 청각여부의 확률 = $(0.344 * 0.649 * 0.541 * 0.955)$ / 모든 확률의 값
- ※ 시각여부의 확률 = $(0.28 * 0.586 * 0.694 * 0.1)$ / 모든 확률의 값
- ※ 기록여부의 확률 = $(0.253 * 0.406 * 0.289 * 0.1)$ / 모든 확률의 값
- ※ 말하기여부의 확률 = $(0.271 * 0.361 * 0.586 * 0.1)$ / 모든 확률의 값

계산 결과 이동여부 5%, 청각여부 57%, 시각여부 5%, 기록여부 1%, 말하기여부 32%로 계산되었다.

Location 연관노드의 경우 각 장소 카테고리는 집 42%, 학교 35%, 도로 17%, 나머지 장소 카테고리 합은 6%의 비율을 구할 수 있고, 조건부 확률을 계산해 본 결과 집→“Gyeongbuk gumi city Geuidong 137-5”는 93%, 학교→“Gyeongbuk gumi city Geuidong 137-5”는 12%, 도로→“Gyeongbuk gumi city Geuidong 137-5”는 46% 로 계산되었다. 나머지 장소 카테고리들은 조건부 확률 비율이 낮아 따로 계산되는 과정은 보여주지 않고, 나머지 장소 카테고리 계산 결과를 모두 더한 비율로 나타냈다.

베이즈 정리를 통해 계산과정은 다음과 같다.

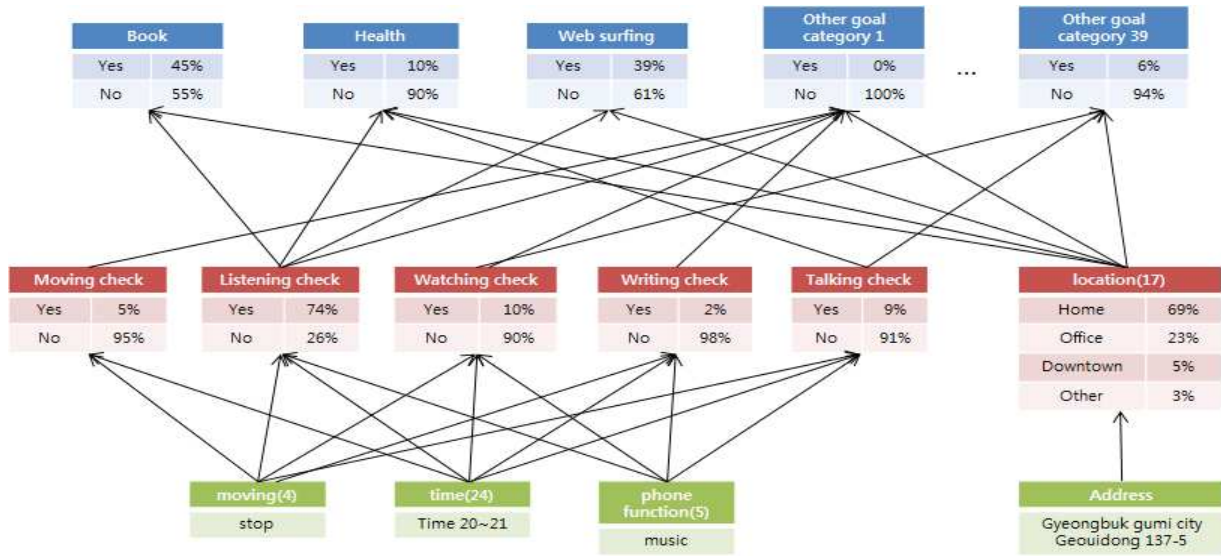
- ※ 집의 확률 = $(0.59 * 0.93)$ / 모든 확률의 합
- ※ 학교의 확률 = $(0.35 * 0.22)$ / 모든 확률의 합
- ※ 도로의 확률 = $(0.45 * 0.46)$ / 모든 확률의 합
- ※ 나머지 장소 카테고리의 확률(14) = 남은 14개 장소 카테고리의 합 / 모든 확률의 합

계산 결과 대략 집은 62%, 학교 9%, 도로는 23%, 나머지 장소 카테고리의 확률은 모두 합쳐 6%로 계산 되었다.

연관노드의 사전비율은 이동여부 5%, 청각여부 57%, 시각여부 5%, 기록여부 1%, 말하기여부 32%, 장소는 집이 62%, 학교 9%, 도로가 23%, 나머지 장소 들은 모두 6%이 계산 되었다. 목표노드의 조건부확률을 계산해 본 결과 “씻기”행동의 경우 씻기→청각여부 85%, 씻기→집 100%, “걷기”행동의 경우 걷기→말하기여부 90%, 걷기→청각여부 62%, 걷기→도로 95%, “웹서핑”의 경우 웹서핑→청각여부 51%, 웹서핑→집 95% 로 계산 되었다. Other goal category는 조건부 확률 비율이 낮아 따로 계산되는 과정은 보여주지 않고, 나머지 행동 계산 결과를 모두 더한 비율로 나타 냈다.

베이즈 정리를 통해 계산과정은 다음과 같다.

- ※ 씻기의 확률 = $(0.85 * 0.56 * 0.67 * 1)$ / 모든 확률의 합
- ※ 걷기의 확률 = $(0.9 * 0.32 * 0.62 * 0.56 * 0.95 * 0.25)$ / 모든 확률의 합
- ※ 웹서핑의 확률 = $(0.51 * 0.56 * 0.95 * 1)$ / 모든 확률의 합
- ※ Other goal category(39) = 남은 39개 행동 확률의 합 / 모든 확률의 합



<Fig. 7> Bayesian network of a virtual worker

<Table 7> Predicted result of current situation for an undergraduate

| Goal node | Rate | Goal node | Rate |
|------------------|------|----------------------|------|
| sleep | 0% | lecture | 0% |
| meal | 0% | study | 1% |
| snack | 0% | cooking | 0% |
| hygiene | 0% | dish-washing | 0% |
| appearance | 0% | washing | 60% |
| cure | 0% | clothing arrangement | 0% |
| class | 0% | clear | 0% |
| regular work | 0% | item arrangement | 0% |
| non-regular work | 0% | voluntary service | 0% |
| part-time job | 0% | newspaper | 0% |
| DVD | 1% | shopping | 0% |
| ball games | 0% | TV | 0% |
| health | 0% | stop | 0% |
| jogging | 0% | walk | 0% |
| take a walk | 4% | run | 0% |
| religion | 0% | car | 0% |
| book | 0% | bus | 0% |
| theater | 0% | train | 0% |
| movie | 0% | online lecture | 0% |
| online game | 0% | date | 0% |
| web surfing | 34% | sport view | 0% |

계산 결과, 씻기 60%, 식사 4%, 웹서핑 34%, 나머지 상황 2%로, 앞으로 학생이 할 행동이 “씻기”로 예측 되었다.

<Table 8> Association analysis result for an undergraduate

| Situation | Support | Confidence | Lift |
|---------------------|---------|------------|-------|
| washing→DVD | 9% | 72% | 1.573 |
| washing→web surfing | 2% | 46% | 2.002 |

사용자 라이프로그 데이터베이스를 참조하여<Fig 7>의 Location 연관노드 제외한 노드들의 사전확률을 계산해 보면, 이동여부 10%, 청각여부 27%, 시각여부 21%, 기록여부 11%, 말하기여부 31%의 사전확률 값이 계산 되었다. 조건부확률을 계산해 보면 이동여부의 경우 이동여부→정지 4%, 이동여부→20~21시 19%, 이동여부→노래듣기 61%이고, 청각여부의 경우 청각여부→정지 61%, 청각여부→20~21시 22%, 청각여부→노래듣기 97%이고, 시각여부의 경우 시각여부→정지 65%, 시각여부→20~21시 37%, 시각여부→노래듣기 0%이고, 기록여부의 경우 기록여부→정지 29%, 기록여부→20~21시 18%, 기록여부→노래듣기 0%이고, 말하기여부의 경우 말하기여부→정지 36%,

말하기여부→20~21시 41%, 말하기여부→노래듣기 0%였다.

- ※ 이동여부의 확률 = $(0.19 * 0.136 * 0.271 * 0.649)$ / 모든 확률의 값
- ※ 청각여부의 확률 = $(0.343 * 0.649 * 0.298 * 0.973)$ / 모든 확률의 값
- ※ 시각여부의 확률 = $(0.289 * 0.685 * 0.433 * 0.1)$ / 모든 확률의 값
- ※ 기록여부의 확률 = $(0.199 * 0.361 * 0.262 * 0.1)$ / 모든 확률의 값
- ※ 말하기여부의 확률 = $(0.379 * 0.424 * 0.469 * 0.1)$ / 모든 확률의 값

계산 결과 이동여부 5%, 청각여부 74%, 시각여부 10%, 기록여부 2%, 말하기여부 9%로 계산되었다.

Location 연관노드의 경우 각 장소 카테고리는 집 39%, 직장 41%, 시내 17%, 나머지 장소 카테고리 합은 3%의 비율을 구할 수 있고, 조건부 확률을 계산해 본 결과 집→“Gyeongbuk gumi city Geoidong 137-5”는 89%, 직장→“Gyeongbuk gumi city Geoidong 137-5” 29%, 시내→“Gyeongbuk gumi city Geoidong 137-5”는 14%로 계산되었다. 나머지 장소 카테고리들은 많고 비율도 낮아 따로 계산되는 결과는 보여주지 않았다.

베이스 정리를 통해 계산과정은 다음과 같다.

- ※ 집의 확률 = $(0.39 * 0.89)$ / 모든 확률의 합
- ※ 직장의 확률 = $(0.41 * 0.29)$ / 모든 확률의 합
- ※ 시내의 확률 = $(0.17 * 0.14)$ / 모든 확률의 합
- ※ 나머지 장소 카테고리의 확률(14) = 남은 14개 장소 카테고리의 합 / 모든 확률의 합

계산 결과 대략 집은 69%, 직장은 23%, 시내는 5%, 나머지 장소 카테고리의 확률은 모두 합쳐 3%로 계산되었다.

연관노드의 사전비율은 이동여부 5%, 청각여부 74%, 시각여부 10%, 기록여부 2%, 말하기여부 9%, 장소는 집이 69%, 직장 23%, 시내가 5%, 나머지 장소들은 모두 3%이 계산 되었다. 목표노드의 조건부확률

을 계산해 본 결과 “독서”행동의 경우 독서→청각여부 82%, 독서→집 91%, “헬스”행동의 경우 헬스→청각여부 95%, 헬스→시내 55%, “웹서핑”의 경우 웹서핑→청각여부 71%, 웹서핑→집 92%로 계산 되었다. Other goal category는 조건부 확률 비율이 낮아 따로 계산되는 과정은 보여주지 않고, 나머지 행동 계산 결과를 모두 더한 비율로 나타냈다.

베이스 정리를 통해 계산과정은 다음과 같다.

- ※ 독서의 확률 = $(0.74 * 0.82 * 0.69 * 0.91)$ / 모든 확률의 합
- ※ 헬스의 확률 = $(0.74 * 0.95 * 0.23 * 0.55)$ / 모든 확률의 합
- ※ 웹서핑의 확률 = $(0.74 * 0.92 * 0.69 * 0.71)$ / 모든 확률의 합
- ※ Other goal category(39) = 남은 39개 행동 확률의 합 / 모든 확률의 합

계산 결과, 독서 45%, 헬스 10%, 웹서핑 39%, 나머지 상황 6%로, 앞으로 가상의 직장인이 할 행동은 “독서”로 예측 되었다.

<Table 9> Predicted result of current situation for a worker

| Goal node | Rate | Goal node | Rate |
|------------------|------|----------------------|------|
| sleep | 1% | lecture | 0% |
| meal | 0% | study | 1% |
| snack | 0% | cooking | 0% |
| hygiene | 0% | dish-washing | 0% |
| appearance | 0% | washing | 0% |
| cure | 0% | clothing arrangement | 0% |
| class | 0% | clear | 0% |
| regular work | 0% | item arrangement | 0% |
| non-regular work | 0% | voluntary service | 0% |
| part-time job | 0% | newspaper | 1% |
| DVD | 1% | shopping | 0% |
| ball games | 0% | TV | 2% |
| health | 10% | stop | 0% |
| jogging | 0% | walk | 0% |
| take a walk | 0% | run | 0% |

| | | | |
|-------------|-----|----------------|----|
| religion | 0% | car | 0% |
| book | 45% | bus | 0% |
| theater | 0% | train | 0% |
| movie | 0% | online lecture | 0% |
| online game | 0% | date | 0% |
| web surfing | 39% | sport view | 0% |

<Table 10> Association analysis result for a worker

| Situation | Support | Confidence | Lift |
|--------------------------|---------|------------|-------|
| book → book | 8% | 69% | 1.890 |
| book → web surfing | 3% | 42% | 1.121 |

실험에 참여한 학생 5명에게 예측된 결과가 얼마나 정확한지 스마트폰을 통해 피드백을 받았다. 피드백은 1달간 진행되었고, 하루에 자는 시간을 제외한 약 10번의 피드백을 받도록 하였다. 1인당 약 300개의 피드백을 받을 수 있었고, 5명의 실험 결과 1511개의 피드백을 받을 수 있었다. 정확도는 <Table 11>에 나타났다.

<Table 11> Student feedback accuracy

| | Accuracy |
|--------------|----------|
| success rate | 60.1% |
| fail rate | 39.9% |

행동 예측의 실패 원인을 분석해본 결과 데이터 수집 과정에서 학생들의 주중과 주말의 행동이 많이 다르다는 것을 알 수 있었고, 수집된 데이터들이 주중 행동 데이터에 많이 치우쳐 있어 주말 행동에 대해선 정확한 행동을 예측 할 수 없었다. 정확률을 높이기 위해 입력 노드의 시간 값을 요일, 월, 년으로 세분화하고 다양한 환경의 데이터를 입력노드로 사용함으로써 정확도 상승을 기대 할 수 있다.

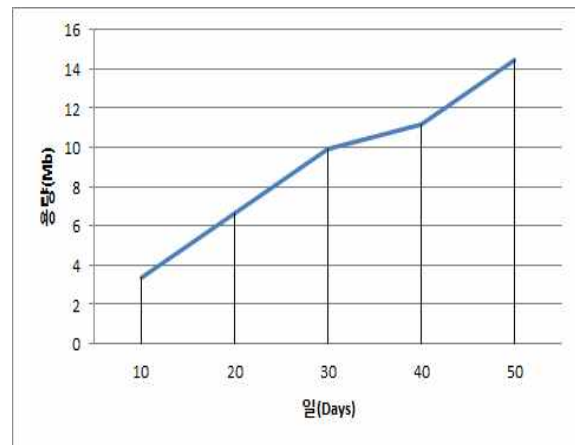
5. 평 가

평가는 한 사람의 사용자 패턴을 저장 하는데 얼마

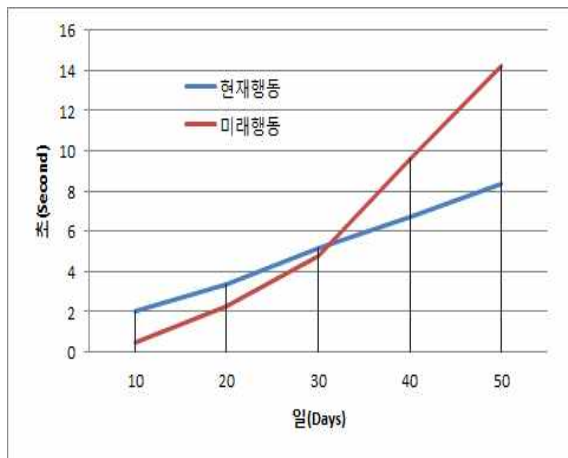
나 필요한지, 그리고 현재 행동과 미래 행동을 예측 하는데 얼마나 시간이 소요되는지 평가했다.

현재 1명의 사용자를 대상으로 과거 라이프로그를 50일 동안 20,365개를 수집했고, 10일 기준으로 얼마나 용량이 증가하는지 메가바이트(Mb)단위로 보여준다. 하루에 약 407개의 라이프로그를 저장 할 수 있었고, 그 수치는 다음 <Fig 8>과 같다. 실험 결과 10일 기준으로 3.38Mb의 공간을 차지했고, 50일 기준으로 14.47Mb의 공간을 차지했다. 이를 1년(365일) 기준으로 봤을 때, 대략 105.63Mb의 공간을 차지 할 것으로 보인다. 본 논문에서 스마트폰이 저장하는 사용자 패턴은 하루치지만, 최근 스마트폰의 보조기억장치 용량이 늘어나는 것을 봤을 때, 1년의 사용자 패턴도 스마트폰이 저장 할 수 있다는 결론을 낼 수 있다.

또한, 현재 행동과 미래 행동을 추론하는데 걸리는 시간을 10일 기준으로 측정해 보았다. 10일 기준으로 현재 행동 2초, 미래행동 0.5초의 시간이 소요되었고, 50일 기준으로 현재 행동 8.4초, 미래 행동 14.25초의 시간을 소요했다. 이를 1년(365일) 기준으로 봤을 때, 대략 현재 행동은 61.32초, 미래 행동은 167.5초의 시간이 소요 될 것으로 보인다. 미래 행동은 데이터가 많아질수록 시간의 폭이 현재 행동보다 커짐을 알 수 있었지만, 본 논문에서 미래 행동은 현재 행동의 한 시간 후를 의미 즉, 한 시간의 여유가 있어서 미래 행동을 예측 하는데 충분한 시간이라는 결론을 낼 수 있다.



<Fig. 8> Storage for a user's pattern



<Fig. 9> Time for inferring a user's current and future behavior

6. 결론 및 향후 연구

본 논문에서는 행동 카테고리 베이지안 네트워크를 이용해 현재 상황을 인지 후 연관분석을 통해 사용자가 미래에 하게 될 행동에 대해 예측하는 방법을 제안하였다. 최근 많이 개발되는 웨어러블(Wearable) 기기와 연동하여 더욱 다양한 분석이 가능 하고, 사용자들에게 많은 유용함을 줄 것으로 보인다.

향후, 추천 종류에 관한 확률 모델을 추가로 정의하고 이를 본 논문에서 제안한 방법과 결합하여 실제 추천 되는 항목을 결정하여 사용자에게 콘텐츠를 제공하는 연구를 진행할 것이다. 또한 Apache Hadoop 시스템에 이러한 추천기법을 접목 시켜 방대한 빅 데이터를 처리, 정제, 분석함으로써 빠르고 정확하게 사용자에게 추천할 수 있는 연구를 진행할 계획이다.

본 시스템은 학생 위주로 설계하였기 때문에 학생이 아닌 일반인의 경우, 해당 행동과 장소가 존재 하지 않을 수 있다. 따라서, 제안한 시스템을 실용화하기 위해서는 논문에 제시된 행동과 장소 외의 행동과 장소를 고려하고, 추가적으로 실제 사용자에게 필요한 행동과 장소를 피드백 받아 시스템을 확장할 필요가 있다.

References

- [1] "Strategy Analytics" Handset Sales Forecast by Type: Phablet, Superphone, Smartphone, Feature Phone & Basic Phone, Apr 17 2014
- [2] G. D. Abowd and E. D. Mynatt, "Charting Past, present, and future research in ubiquitous computing", ACM Transactions on Computer-Human Interaction, Vol 7, No.1 pp. 29-58, 2000.
- [3] C. Randell and H. L. Muller, "The Well Mannered Wearable Computer", Personal and Ubiquitous Computing, pp. 31-36, 2002.
- [4] Vannevar Bush, "As We May Think", Atlantic Monthly, pp.112-124, Jul. 1945.
- [5] Young-Seol Lee, Sung-Bae Cho "Constructing probability model to infer activities using Activity Theory and mobile life log" Korean Institute Of Information Technology, Vol 36, No.1 pp397-401, 2009
- [6] Seong-Yong Hong "Collective Intelligence of IT Learners Learning Path Prediction and Recommendations using Big Data Mining", Korean Institute Of Information Technology, pp197-199, 2013
- [7] Young-Wan Jang, Byeong-Man Kim, Chang-Bae Moon, Yoon-Sik Shin "Auto Tagging Using Mobile-Based User's Context Information for Personal Lifelog" Korean Institute Of Information Technology Vol 40, No.5 pp.236-247, 2013
- [8] Young-Wan Jang, Byeong Man Kim, Sung Bong Jang, Yoon Sik Shin "Application of Euclidean Distance Similarity for Smartphone-Based Moving Context Determination" Journal of the Korea Industrial Information Systems Research Vol 19, No.4 Aug. 2014
- [9] M. L. Lee and A. K. Dey, "Providing Good Memory Cues for People with Episodic Memory Impairment," the 9th International ACM SIGACCESS Conference on Computer and Accessibility, Tempa, USA, pp.131-138, 2007
- [10] Min-Sung Hong and Nam-Hee Mok "A method of determining the user's state of

movement based on the smart device usage”
Journal of the Korea Industrial Information
Systems Research Vol 18, No.6 Aug. 2013

- [11] Do-Joon Jung, Jeong-Oh Yoon “Human Activity Recognition using Model-based Gaze Direction Estimation” Journal of the Korea Industrial Information Systems Research Vol 16, No.4 Nov 2011
- [12] G.D. Kleiter, “Propagating imprecise probabilities in Bayesian networks”, Artificial Intelligence, Vol 88, No.1-2 pp.143-161, 1996
- [13] “Life time behavior taxonomy table”
www.kostat.go.kr StatisticsKorea 2009



방 재 근 (Jae-Geun Bang)

- 학생회원
- 2009년 ~ 현재 : 국립금오공과대학교 컴퓨터소프트웨어공학과 학사과정
- 관심분야 : 인공지능, 상황인식컴퓨팅, 추천시스템



김 병 만 (Byeong Man Kim)

- 1987년 : 서울대학교 컴퓨터공학과(공학사)
- 1989년 : 한국과학기술원 전산학과(공학석사)
- 1992년 : 한국과학기술원 전산학과(공학박사)
- 1992~현재 : 국립금오공과대학교 소프트웨어공학과 교수
- 1998년~1999년 : 미국 UC, Irvine 대학 방문교수
- 2005년~2006년 : 미국 콜로라도 주립대학 연구교수
- 관심분야 : 인공지능, 정보검색, 정보보안

논문 접수 일 : 2014년 09월 19일

1차수정완료일 : 2014년 11월 13일

게재확정일 : 2014년 11월 17일