

키그래프 기반 스마트폰 사용자 정보 관리

User's information management based on KeyGraph for Smartphone

정명철, 조성배
연세대학교 컴퓨터과학과

Myung-Chul Jung, Sung-Bae Cho

Dept. of Computer Science, Yonsei University

E-mail : mcjung@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

요 약

최근 디지털 기기들이 융복합화 되며 다양한 기능을 갖춘 제품이 출시되고 있고, 유비쿼터스 환경을 이끌어갈 제품으로 스마트폰이 주목받고 있다. 스마트폰에서 실행되는 다양한 사용자 행동 정보들 중에 중요한 의미를 지닌 행동들을 찾아내고 기록해서 사용자의 중요했던 기억을 찾는 데 도움을 줄 수 있다. 스마트폰을 사용하면서 얻어지는 행동 정보들은 양이 방대하여 기존의 시간대별 또는 카테고리식의 나열은 사용자가 필요한 정보를 쉽게 얻기에는 한계가 있다. 따라서 본 연구에서는 키그래프를 사용해 사용자 행동 정보를 분석하고 핵심 이벤트 중심의 사용자 정보 관리 방법을 제안한다. 키그래프는 단순히 행동 정보의 빈도수가 많다고 높은 점수를 부여하는 것이 아니라 빈도수가 낮더라도 중요한 이벤트 정보를 발견할 수 있다. 시나리오를 바탕으로 만들어진 사용자 정보를 분석해서 시간의 흐름에 따른 핵심 이벤트와 그와 관련된 이벤트를 찾아내는 서비스를 생성하고 제안한 시스템의 유용성을 보인다.

1. 서론

최근 모바일 컴퓨팅 기기들은 전화나 SMS 뿐 아니라 일정이나 연락처 관리, MP3듣기, 사진찍기, 게임 등 다양한 기능들을 제공하고 있다. MP3 플레이어나 디지털 카메라 같이 각각의 디지털 기기들이 이동전화 단말기나 PDA와 결합하고, PDA와 전화 기능이 결합하여 스마트폰이 되는 디지털 융복합화가 활발히 진행되고 있다. 휴대용 단말기의 기능이 다양해지고 활용도가 늘어나면서 사람들에게 모바일 단말기는 필수적인 도구로 자리를 잡았다.

모바일 도메인을 포함한 다양한 사용자 정보들을 효율적으로 분석하고 재가공하여 사용자에게 도움을 줄 수 있는 서비스를 위한 연구가 활발히 진행되고 있다. R. Miiikkulainen은 에피소딕 메모리(Episodic Memory)라는 주제로 스크립트 기반의 기억 정보를 계층적 SOM(Self-organizing Feature Maps)으로 분류, 저장, 복원하는 방법을 제안하였다 [1]. 계층마다 정해진 정보들을 저장함으로써 기억의 복원 속도를 높이고 일부분의 정보가 소실되더라도 계층적 구조의 도움으로 기억을 복원할 수 있다. E. Horvitz 등은 사람

의 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 시간 순서대로 정리해 주고 이를 블로그로 생성해주는 서비스이다 [3]. 생성된 블로그는 PC나 휴대폰으로 접속하여 사진을 업로드하고 자신의 블로그를 관리할 수 있다.

본 논문에서는 스마트폰에서 사용자의 정보를 효율적으로 관리하기 위한 방법을 다음과 같이 제안한다. 전화기록, SMS 기록, 멀티미디어 사용 기록, 위치 정보 등과 같은 사용자 로그를 추출하고 이를 하나의 공통된 형태로 통합한다. 통합된 사용자 로그는 키그래프를 적용하여 사용자 정보간의 연결 관계와 핵심 정보를 찾아내는 분석 과정을 거치게 된다. 키그래프는 이벤트들의 빈도수와 이벤트 사이의 연결 관계를 바탕으로 빈도수는 적지만 중요하고 의미있는 이벤트를 찾아내는 알고리즘이다. 분석이 완료된 사용자 정보는

중요한 의미를 지니는 정보를 중심으로 재구성 된다. 이는 단순히 시간 순서에 따른 이벤트의 나열이나 전화, 사진, 멀티미디어 등 카테고리 별로 분류하는 방법과 비교해 제안하는 방법은 사용자에게 중요도가 높은 이벤트와 연결 관계를 갖는 이벤트들로 구성되기 때문에 사용자에게 따른 유연한 정보 관리 방법을 제공할 수 있다.

2. 키그래프

문서 인덱싱 방법의 하나로 제안된 키그래프(KeyGraph)는 문서 내에서 중요한 단어를 추출하고 단어들 사이의 연결 구조를 도식화 해준다 [4]. 본 연구에서는 사용자의 이벤트들 가운데 사용자에게 중요한 의미를 갖는 이벤트를 찾아내고 이벤트들 사이의 연결 관계를 표현하는 방법으로 키그래프 알고리즘을 사용한다.

2.1 키그래프의 의미

키그래프는 핵심적이면서 쉽게 관찰되지 않는 이벤트를 찾아내는데 유용한 방법이다. 시간 순서에 따른 연속된 이벤트들이 나열된 문서 예제 D는 식 (1)과 같다.

$$D = \text{Call20\#Pic2\#Pic7\#SMS11\#} \quad (1)$$

$$\text{EB24\#MM12\#Pic7\#Call20\#SMS11\#}$$

$$\text{MM5\#SMS11\#}$$

$$\text{EB24\#MM5\#}$$

여기서, 'Type-number#' 는 사용자 정보로부터 얻어진 각각의 이벤트이다. Call20#과 Pic2#은 20번째 전화 사용 기록과 2번째 사진 기록을 의미한다. 식 (1)에 포함된 점('.')은 이벤트 환경의 큰 변화가 일어났음을 의미한다. 여기서 변화의 정의는 장소의 이동이나 스케줄의 변화이다.

상호 발생 빈도가 높은 이벤트들을 모아 클러스터를 만들면 이것은 키그래프의 기본 클러스터가 된다. 즉, 이벤트 시퀀스에서 출현 빈도가 높은 이벤트가 추출되고 같은 단위 이벤트 시퀀스 안에서 자주 나타나는 이벤트 쌍들이 서로 연결 관계를 갖게 된다. 그림 1에서 Call20#, Pic7#, SMS11#는 기본 클러스터를 형성하고 있다. 이렇게 서로 연결된 이벤트들은 하나의 기본 클러스터가 되고 이벤트 발생의 공통적 원인을 갖는다는 것을 의미한다.

발생 빈도가 낮아 클러스터에는 포함이 되지 못하였지만 핵심이 되는 이벤트가 있다. 예를 들어, 그림 1에서 'MM5'는 각 클러스터들과 연결 관계를 가짐으로써 전체 그래프에서 핵심 이벤트가 된다. 핵심 이벤트는 이벤트 관계의 구조적인 측면에서 발생 빈도는 낮지만 중요도가 높아 사용자가 큰 의미를 가질 수 있는 이벤트가 된다.

2.2 키그래프 알고리즘 적용 과정

이벤트 문서 D에서 발생 빈도가 높은 노드를 찾는

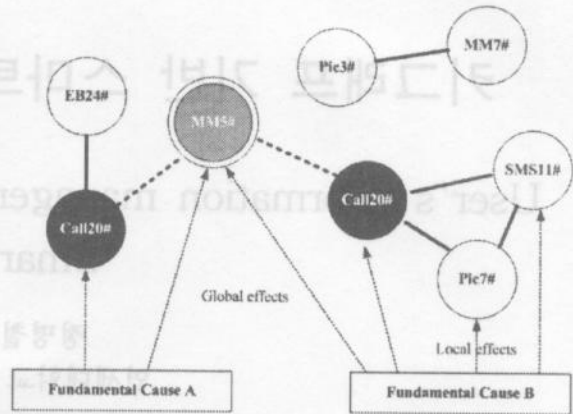


그림 1. 키그래프 예제

다. 그리고 식(2)의 $local(w_i, w_j)$ 를 적용하여 연결 관계가 높은 노드들의 쌍을 찾아내고 연결하면 그림 1에 보이는 실선이 그려진다. 각 링크는 지역적 영향을 받는 이벤트 사이의 쌍방향 관계이다. 식(2)에서 $|w|_s$ 는 가장 가까운 두 점 's' 사이의 단위 시퀀스 s에서 이벤트 w의 출현 빈도수를 의미한다. w는 각 이벤트를 의미하고, g는 w에 의해 구성되는 클러스터를 의미한다.

$$local(w_i, w_j) = \sum_{s \in D} |w_i|_s |w_j|_s \quad (2)$$

클러스터들이 구성된 후 식(3) $global(w, g)$ 을 적용하여 각 클러스터와 각 이벤트 노드 사이의 연결 강도를 계산한다.

$$global(w, g) = \frac{\sum_{s \in D} |w|_s |g - w|_s}{\sum_{s \in D} \sum_{w \in s} |w|_s |g - w|_s} \quad (3)$$

$$where |g - w|_s \equiv \begin{cases} |g|_s - |w|_s & \text{if } w \in g \\ |g|_s & \text{if } w \notin g \end{cases}$$

$|g|_s$ 는 단위 시퀀스 s에서 클러스터 g가 나타나는 횟수이다. 각 클러스터로부터 연결 강도가 강한 이벤트가 핵심 이벤트로 간주된다. 여기서는, $global(w, g)$ 의 값을 계산해 가장 큰 값을 갖는 이벤트 w가 핵심 이벤트로 선택되고, 그림 1에서는 MM5#가 선택된다. 핵심 이벤트와 연결되는 클러스터들의 연결선은 점선으로 그려진다.

3. 사용자 정보 관리

본 연구에서 제안하는 시스템은 크게 사용자 로그 수집 및 통합, 키그래프 적용, 키그래프 분석 및 서비스 생성으로 구성된다. 그림 2는 제안하는 시스템의 전체적인 구조를 보여준다. 사용자 로그 수집은 종류별로 사용자 로그들을 수집하는 것이다. 통합 로그 생성은 수집된 종류별 로그를 키 그래프로 분석하기 위해 하나의 형식으로 통합하는 과정이다. 키그래프 적

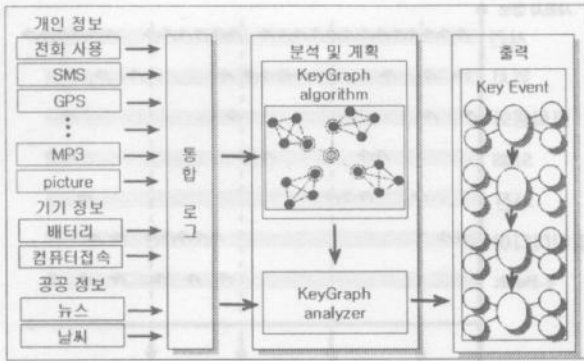


그림 2. 제안하는 시스템의 구조

용은 앞서 2장에서 설명한 키그래프 알고리즘을 통합된 사용자 로그에 적용하여 키그래프를 만들게 된다. 키그래프 분석은 키그래프에서 핵심 이벤트와 이벤트 사이의 연결 정도를 분석한다. 서비스 생성은 키그래프 분석 단계에서 찾아낸 핵심 이벤트를 중심으로 연결된 이벤트를 정의하고 시간 도메인 상에서 재구성하는 작업을 하게 된다.

3.1 통합 로그 구성 및 키그래프 적용

각 사용자 로그 수집 모듈을 통해 종류별로 모아진 사용자 로그들은 키그래프에 적용하기 위한 전처리 과정을 거치게 된다. 로그 분석 및 관리를 쉽게 하기 위해 각 로그들을 통합로그라는 포맷으로 변환한다. 통합로그는 시각, 장소, 이벤트 내용, 로그의 가중치로 구성되어 있다. 시각은 이벤트가 발생한 시각을 나타내고 장소는 GPS로부터 얻어진 위도, 경도 로그 값을 이미 정의되어 있는 장소 정보로 매칭을 하여 얻는다. 이벤트는 종류별 로그로부터 얻어진 구체적인 이벤트 정보이다. 가중치는 로그 종류마다 기록되는 형태와 값의 경중의 차이가 있어 이를 보완해준다. 예를 들어, 전화사용 10분은 이벤트가 긴 시간 일어났고 중요할 가능성이 높은 이벤트가 되지만, MP3를 10분 동안 들은 것은 그다지 중요한 이벤트로 생각되지 않는다. 이와 같이 로그별 차이를 줄여주기 위해 가중치를 설정을 해주게 된다.

통합된 로그 정보를 키그래프에 적용하기 위해서는 앞서 2장에서 설명한 이벤트 시퀀스 형태의 문서 D로 키그래프 모듈에게 넘겨줘야 한다. 각 이벤트는 시간 도메인, 장소 도메인에 의해 이벤트 시퀀스 구조가 결정된다. 시간 도메인 상에서는 이벤트 사이의 시간 간격이 일정치 이상이면 중요한 변화요소로 판단한다. 장소 도메인에서는 같은 장소에서 일어난 이벤트들끼리 하나의 이벤트 셋으로 묶어서 이벤트를 구성한다. 각 주요 도메인에 의한 영향력을 고려해서 만든 이벤트 문서 D를 스마트폰에 구현된 키그래프 모듈에 입력하고 에이전트는 키그래프 형태로 결과를 전달 받게 된다. 그림 3은 각 로그들을 이벤트 문서 D로 만드는 과정을 도식화 하였다.

3.2 키그래프 분석 및 서비스 생성

입력한 이벤트 문서 D의 결과로 반환된 키그래프는

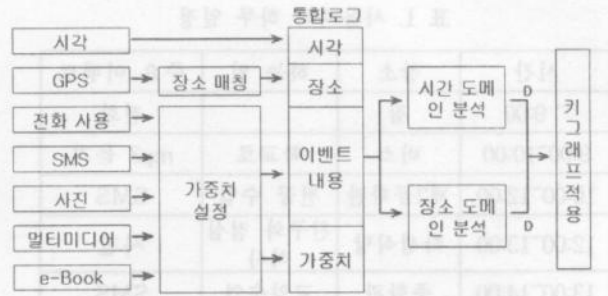


그림 3. 키그래프에 적용하기 위한 로그 전처리

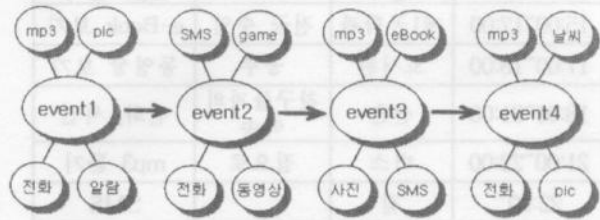


그림 4. 서비스 생성의 예

12개의 핵심 이벤트를 가지고 있다. 그리고 핵심 이벤트와의 연결 노드들, 즉 연관된 이벤트와의 연결 강도를 알 수 있다. 키그래프 분석을 통해 12개의 핵심 이벤트와 그와 연결된 이벤트 셋을 가지고 각 핵심 이벤트의 시간 순서에 따라 정렬을 한다. 정렬된 결과를 스마트폰 상에 보여줘 사용자가 쉽게 자신의 중요한 이벤트를 중심으로 검색하고 연결된 이벤트를 볼 수 있도록 지원한다. 그림 4는 시간 순서에 따른 서비스 생성 결과의 예를 보여준다.

4. 실제 적용 예

본 연구에서는 제안한 에이전트의 유용성을 입증하기 위해 가상의 시나리오를 작성하고 시나리오 상에서 에이전트의 추론 과정 및 서비스 생성 결과를 살펴보았다. 시나리오는 스마트폰을 적극적으로 사용하는 20대의 남자 대학생의 하루 일정이고 표 1과 같이 정의하였다.

설정된 시나리오의 일정에는 11곳의 장소를 이동하며 아침부터 저녁까지의 사용자 이벤트를 대상으로 한다. 스마트폰 에이전트는 사용자의 행동을 관찰하다가 행동 정보가 변경이 될 때마다 사용자 로그 기록을 남긴다. 하루 일정이 종료가 되면 3장에 언급했던 과정을 통해 사용자의 핵심 이벤트를 중심으로 일과를 재구성하여 사용자에게 의미 있는 행동 정보들을 알려주게 된다.

그림 5는 시나리오상의 사용자 이벤트를 키그래프로 구성한 것이다. 열두 개의 검은색 노드는 전체 키그래프의 핵심 노드이고, 이와 연결된 노드 또는 따로 클러스터링 된 노드들이 있다. 많은 이벤트들이 물려있는 좌측 아래쪽 그래프에는 여덟 개의 핵심 이벤트를 포함해 26개의 이벤트가 연결되어 있다.

그 중 핵심 이벤트로 뽑힌 이벤트를 분석해 보면, Pic15, Pic16, Pic17은 오랜만에 만난 친구와 점심 식사를 하면서 스마트폰으로 찍은 사진 정보를 의미한

표 1. 사용자의 하루 일정

시간	장소	하는 일	주요 이벤트
~ 9:00	집		전화
9:00~10:00	버스	학교로	mp3 듣기
10:00~12:00	제2공학관	전공 수업	SMS
12:00~13:00	학생식당	친구와 점심 식사	사진
13:00~14:00	종합관	교양수업	SMS
14:00~15:00	공대 앞 벤치	휴식	e-Book 보기, mp3 듣기
15:00~17:00	제1공학관	전공 수업	e-Book 보기
17:00~18:00	도서관	공부	동영상 보기
18:00~21:00	신촌	친구들과의 약속	전화, 사진
21:00~22:00	버스	집으로	mp3 듣기
22:00~	집		SMS

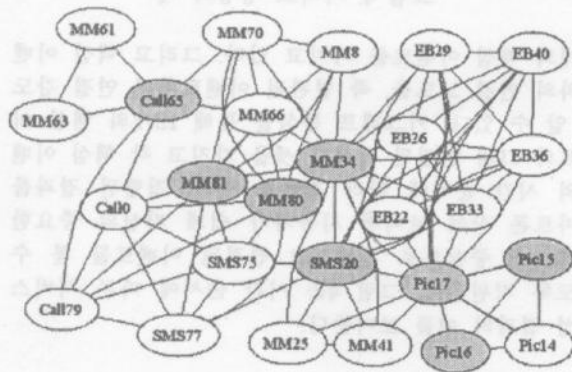


그림 5. 시나리오를 바탕으로 만들어진 키그래프

다. SMS20은 중요한 저녁 모임을 안내하는 친구의 SMS 이다. {MM34, MM80, MM,81}은 사용자가 하루 동안 가장 인상 깊게 들었던 MP3 트랙을 의미한다. Call65는 사용자가 친구들과의 모임을 끝마치고 집으로 돌아오면서 여자 친구에게 전화를 걸었던 정보이다. 이 시간에 MP3 음악을 많이 들으면서 와서 MP3 정보와 연결 관계가 높았고 핵심 이벤트로 선정이 되었다.

그림 6은 키그래프를 분석하여 나온 핵심 이벤트 열 두 개를 시간 순서로 배열하고 그 연결 관계에 있는 이벤트를 보여주는 최종 결과의 일부분이다. 사용자 정보를 바탕으로 시간 축에서 핵심 이벤트를 추출하고 핵심 이벤트와 연결된 강도가 센 이벤트들을 하위 이벤트로 표현해준다.

5. 결론 및 향후 연구

본 연구에서는 스마트폰에 저장된 사용자의 이벤트 정보를 사용자 중심에서 재구성하고 관리하기 위한 방법으로 로그를 수집 및 통합하고 이를 키그래프에 적용하여 핵심 이벤트를 찾아내는 방법을 제안하였다.

스마트폰 상에서 가용한 사용자의 정보를 정의하고 이를 키그래프에 적용하기 위한 사용자 로그 통합과

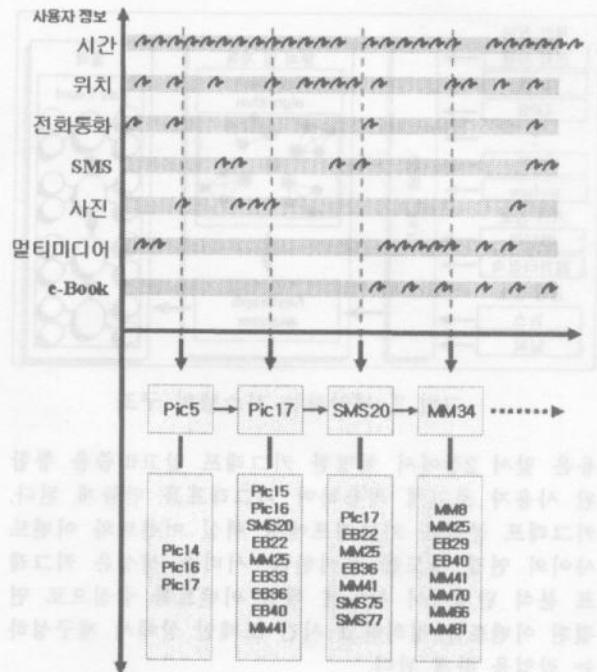


그림 6. 사용자 정보를 바탕으로 시간에 따른 핵심 이벤트 추출 및 연결 이벤트

가중치 설정을 하였다. 키그래프 알고리즘을 적용해 나온 결과를 분석하여 핵심 이벤트를 중심으로 사용자의 이벤트를 재구성하였다. 제한한 방법의 유용성을 보이기 위해 가상의 시나리오를 설정하고 그에 따른 서비스 생성 결과를 시나리오를 바탕으로 분석해 보았다. 키그래프를 적용하여 생성된 결과는 사용자에게 중요한 기억으로 남은 이벤트들을 찾아내었다. 성능에 대한 객관적인 평가에까지는 이르지 못하였지만 제한한 방법의 가능성을 보여주었다.

향후 연구로 하루의 데이터가 아닌 수십일 이상의 비교 가능한 데이터 셋을 수집하고 다른 방법과 비교 평가하는 작업이 필요하다. 사용자가 중요하다고 생각하는 이벤트를 선정하고 비교 분석하는 사용자 테스트를 통해 제한한 방법의 유용성을 평가하고 한계점 및 개선점을 파악하여 다음 연구를 진행할 것이다.

참고문헌

- [1] R. Miikkulainen, "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] Y. Ohsawa, N. E. Benson and M. Yachida, "KeyGraph: Automatic indexing by co-occurrence graph based on building construction metaphor", *Proc. Advanced Digital Library Conference (IEEE ADL'98)*, pp. 12-18, 1998.