스마트폰 사용자의 컨텍스트 수집과 생성

이영설[0] 정명철 조성배
연세대학교 컴퓨터과학과
tiras@sclab.yonsei.ac.kr[0] mcjung@sclab.yonsei.ac.kr sbcho@cs.yonsei.ac.kr

Collection and Construction of User's Context in Smart Phone

Young-Seol Lee[0], Myung-Chul Jung and Sung-Bae Cho
Dept. of Computer Science, Yonsei University

요 약

스마트폰의 성능이 향상되고 많은 기능이 추가됨에 따라 스마트폰에 저장되는 사용자 정보도 증가하고 있다. 이렇게 저장된 사용자 정보는 사용자가 경험한 일이나 상황을 추론하기 위한 판단의 근거가 될 수 있다. 본 연구에서는 스마트폰으로부터 GPS 위치 정보, 전화 통화 기록, SMS 정보, 날씨 정보, MP3 음악듣기 정보, 사진 찍은 정보, 사진을 본 내역, 충전상태에 관한 정보를 수집하고, 통계적 분석, 임팩트 분석을 통하여 수집된 정보로부터 컨텍스트를 추출한다. 추출된 컨텍스트를 통하여 사용자의 상태에 대한 추론이 가능하며 추론 모델이나 서비스를 위해 입력으로 제공될 수도 있다.

1. 서 론

제 3세대 이동전화 서비스의 등장과 개인용 컴퓨팅 환경의 필요성이 증가함에 따라 등장한 스마트폰은 컬러 LCD 스크린과 대용량의 메모리, 메모리 카드를 이용한 넓은 저장 공간, Wi-Fi, Bluetooth, 적외선 통신 등 PC와 데이터를 연동하기 위한 통신 기능, 그리고 스케줄러, 주소록, 미디어 플레이어, E-Book 등 다양한 기능을 가지고 있다. 많은 기능들이 스마트폰에 탑재되면서 스마트폰에 저장되는 사용자 정보도 증가하고 있으며 이런 정보를 이용하기 위한 연구도 활발히 진행되고 있다.

헬싱키 대학의 Mika Raento[2]는 스마트폰 기반의 컨텍스트 수집 프레임워크를 개발하였다. 이 프레임워크는 GSM Cell ID, Bluetooth 디바이스, 그리고 Bluetooth를 통한 GPS 데이터와 전화 통화 내역, 받는 SMS 정보, 그리고 그 외에 미디어 정보 등을 통합하여 수집하고 그것들을 자동으로 서버로 전송하는 형태로 이루어져 있다. 이 프레임워크에서 얻어진 컨텍스트는 주로 미디어나 다른 정보에 부가정보로서 제공될 수 있다. VTT Technical Research Centre의 Panu[3] 등은 모바일 디바이스에서 컨텍스트에 이용될 로그를 수집하고 수집된 로그를 전처리하여 특징을 추출한다. 이 실험에 사용된 모바일 디바이스는 GPS 정보나 디바이스 내부 프로세스 정보를 사용할 뿐만 아니라 마이크로폰, 온도, 습도, 광센서 등의 여러 센서를 내장하고 있어 더욱 효과적으로 컨텍스트 로그를 수집한다.

본 논문에서는 일상생활 속에서 스마트폰에 저장되는 사용자 정보를 수집하는 방법을 정리하고 저수준 컨텍스트를 획득하는 방법을 제안한다. 컨텍스트는 정보의 빈도, 지속시간, 발생 간격을 통계적으로 분석하고 임팩트 수치를 계산하여 구한다. 또한 2주간 3명의 실험자를 대상으로 스마트폰에서 정보를 수집하고 수집된 정보를 처리하여 컨텍스트를 추출하여 제안한 방법의 가능성을 보인다. 얻어진 저수준 컨텍스트는 사용자가 하루 동안 경험한 일들을 추론하기 위한 모델이나 서비스에 입력으로 이용될 수 있다. 전체 시스템의 개요는 그림 1에서 보여주는 것처럼 스마트폰의 사용자 정보를 수집하는 부분과 수집된 정보를 처리하여 저수준 컨텍스트를 생성하는 부분으로 나뉜다.

2. 사용자 정보 수집

표 1은 스마트폰에서 수집되는 사용자 정보의 내용을 설명한다. GPS 정보로부터 사용자가 방문한 장소에 대한 정보를 얻을 수 있으며, Call과 SMS 정보로부터 사용자가 통화한 내역, 빈도를 구할 수 있다. 인터넷으로부터 얻을 수 있는 날씨 정보는 사용자가 하루 중에 경험한 기분이나 상태에 영향을 미칠 수 있다. MP3 음악은 사용자의 감성에 영향을 주고 사진을 찍은 내역은 사용자가 기억하고 싶은 일이 있음을 알 수 있다. 사진을 본 내역은 사용자가 즐겨 보는 사진이나 그림에 대한 정보를 제공해 준다.

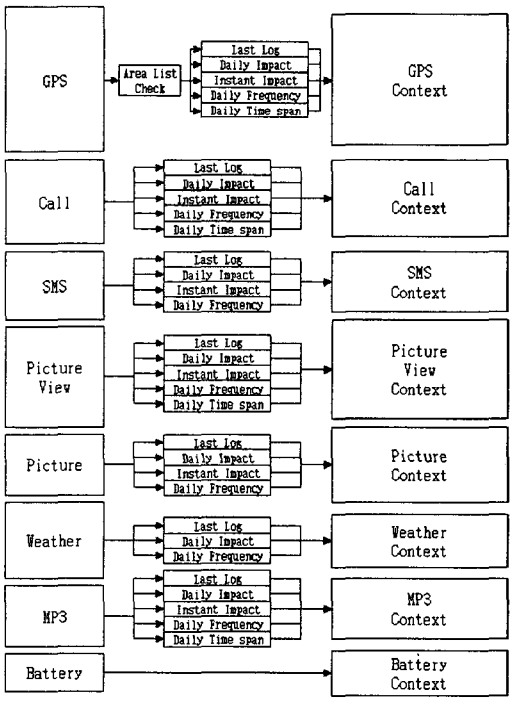

그림 1. 시스템 개요

표 2는 사용자 정보의 수집 방법과 주기를 설명한다. 사용자 정보는 수집되는 정보의 종류에 따라서 수집 주기가 다르다. 사진보기나 MP3 듣기 내역의 경우에는 이미지 뷰어나 MP3 플레이어 프로그램을 이용할 때마다 수집된다. SMS와 Call, 사진 찍은

내역과 날씨 정보는 하루에 한번 수집된다. GPS 위치 정보와 충전상태 정보는 매 1초 마다 수집된다.

표 1. 수집되는 로그 정보

로그 종류	얻을 수 있는 정보
GPS	위도, 경도, 이동속도, 진행방향, 날짜, 시간
Call	상대방 전화번호, 송신/수신/부재 여부, 통화 시작/종료시간
SMS	상대방 전화번호, 송신/수신 여부, 발신/수신 시간
사진보기	사진파일명, 사진보기 시작한 시간, 사진 닫은 시간
사진	사진파일명, 사진 파일 생성날짜
날씨	날씨, 시정(km), 전운량(%), 현재기온(°C), 불쾌지수(%), 체감온도(°C), 강수량(mm), 적설(cm), 습도(%), 풍향, 풍속(m/s), 해면기압(hPa)
MP3	노래제목, 시작시간, 종료시간, mp3가 시작한 위치
충전상태	현재 충전량, 충전중인지 여부, 현재 시각

표 2. 로그 수집 방법과 수집 주기

로그 종류	수집 방법	수집 주기
GPS	GPS 모듈로부터 수집	매 1초마다 로그를 남김
Call	저장된 통화 내역을 수집	매 1일 마다 로그를 남김
SMS	저장된 SMS 내역을 수집	매 1일 마다 로그를 남김
사진보기	이미지 뷰어를 사용하여 수집	매 수행 시 로그를 남김
사진	사진파일 생성내역 수집	매 1일 마다 로그를 남김
날씨	인터넷에서 날씨 정보를 수집	매 1일 마다 로그를 남김
MP3	MP3 플레이어로부터 수집	매 수행 시 로그를 남김
충전상태	백그라운드 프로그램으로 수집	매 1초마다 로그를 남김

3. 컨텍스트 생성

수집된 사용자 정보가 추론 모델이나 서비스에 이용되기 위해서는 저수준 컨텍스트를 생성할 필요가 있다. 특히 GPS 정보는 스마트폰에 저장된 기록만으로는 사용자의 위치를 판단할 수 없고 기록되는 정보량이 많아 추론 모델이나 서비스에서 직접 이용하기가 어렵다. 본 논문에서는 통계 정보와 임팩트를 이용하여 저수준 컨텍스트를 생성한다. 또한 GPS 정보로부터 사용자의 위치를 분석하고 필요한 정보를 추출하는 방법을 제안한다.

3.1. 통계적 분석

저수준의 정보에서 빈번하게 발생하거나 오래 지속되는 사건, 혹은 평소에 거의 발생하지 않는 사건이라면 사용자에게 의미 있는 일이 될 수 있다. 따라서 다음 3가지 내용에 대해 통계적 분석을 한다.

- 하루 동안 발생한 빈도(횟수)
- 하루 동안 지속된 시간의 총량
- 마지막 발생 후 지난 시간

3.2. 임팩트 분석

사용자에게 의미 있는 일을 판단할 때 단순히 빈도나 지속 시간만으로 판단하기는 어렵다. 사건의 집중도가 사용자에게 더욱 의미 있을 가능성이 높다. 임팩트는 특정 사건이 발생한 빈도에 대한 집중도를 나타내 주기 위한 것이다.

그림 2에서는 임팩트 발생의 예를 그래프로 표현하고 있으며, 표 3에서는 임팩트가 증가하는 기준과 감소하는 기준을 나타내고 있다. 일정 시간 내로 이벤트가 연속적으로 발생되면 임팩트 수치는 점점 증가하게 된다. 즉, 임팩트 수치가 이벤트의 빈도와 집중도가 높은 시점을 파악할 수 있다.

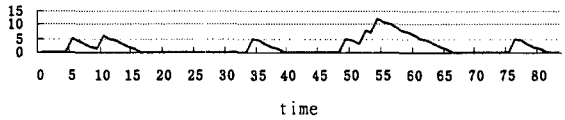

그림 2. 임팩트 그래프의 예

표 3. 임팩트 수치의 증감

로그 종류	임팩트 수치 증가	임팩트 수치 감소
GPS	if(GPS Event) Impact +1	if(이벤트 후 1시간) Impact -1
Call	if(Call Event) Impact +1	if(이벤트 후 1시간) Impact -1
SMS	if(SMS Event) Impact +1	if(이벤트 후 20분) Impact -1
사진보기	if(사진 보기) Impact +1	if(이벤트 후 5분) Impact -1
사진	if(사진 찍기) Impact +1	if(이벤트 후 30분) Impact -1
날씨	적용되지 않음	적용 안됨
MP3	if(음악 듣기) Impact +1	if(이벤트 후 30분) Impact -1
충전상태	적용되지 않음	적용 안됨

3.3. 위치 분석

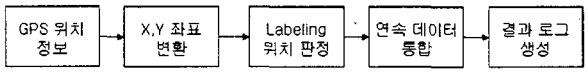

그림 3. GPS 정보를 이용한 사용자 위치 분석 과정

그림 3은 제안하는 시스템에서 GPS 위치 정보를 분석하는 과정을 나타내고 있다. 먼저 GPS 위치 정보를 도분초에서 X,Y 좌표로 환산한 후 환산된 좌표가 레이블링된 장소인지 확인하여 사용자가 방문한 장소의 목록을 만들어 낸다. 사용자의 방문 장소목록이 만들어지면, 목록에서 연속된 데이터를 통합하여 최종적인 사용자 위치 컨텍스트를 추출한다.

본 연구에서는 사용자가 방문한 장소들을 레이블링하여 사용자의 위치를 파악한다. 레이블링할 장소가 넓은 지역일 경우에는 그 지역을 다각형으로 표시하고 사용자가 그 다각형의 내부에 있을 경우 그 장소를 방문한 것으로 판단한다. 좁은 지역이나 건물의 경우에는 건물이나 지역의 중심으로부터의 거리가 오차범위 이하일 경우 건물이나 지역을 방문한 것으로 판단한다. 다각형의 내부로 판단하는 방식은 좁은 건물이나 지역에서는 컨텍스트의 추출 정확도가 떨어지고, 중심점으로부터의 거리로 판단하는 방식은 넓은 지역에 적용될 경우 레이블링 되지 않은 지역도 레이블링된 장소로 판단하게 만들 확률이 높다.

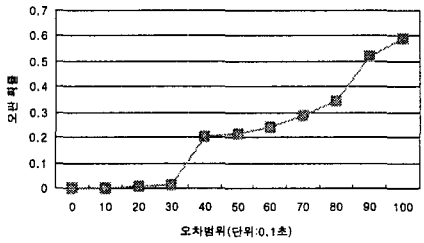

그림 4. 오차범위에 따른 오판 확률

그림 4는 넓은 지역에 중심점 방식을 사용할 경우 장소를 잘못 판단할 확률을 보여주고 있다. 다각형 판정을 사용할 경우 오판 확률이 0 이지만 중심점 방식을 사용할 때는 오차 범위가 커짐에 따라서 오판 확률이 점차 증가함을 알 수 있다. 장소의 크기가 커질수록 잘못 판단하는 면적도 증가하게 된다.

장소가 건물이나 좁은 지역일 경우에는 먼저 건물을 다각형으로 표시하고 다각형을 이루는 점들(x_1, y_1) ~ (x_n, y_n)의 좌표에서 다각형의 중심점(x_m, y_m)의 좌표를 계산한다.

$$x_m = \sum_{i=1}^{n}(x_i)/n, \quad y_m = \sum_{i=1}^{n} y_i/n \qquad (1)$$

그리고 중심점으로 부터의 거리가 건물의 반경 R보다 작을 경우, 그 건물을 방문한 것으로 판단한다.

$$R=\sqrt{(x_m-x_1)^2+(y_m-y_1)^2}+30 \qquad (2)$$

R은 중심점에서 다각형의 한 점까지의 거리 + 0.3초(건물에 들어가기 전에 GPS가 끊기는 경우를 고려한 오차범위) 이다.

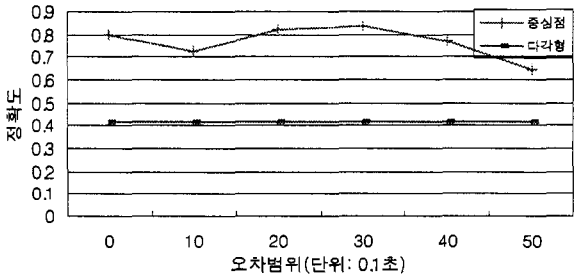

그림 5 좁은 지역에서의 위치 판단의 정확도

그림 5.에서 나타낸 정확도는 컨텍스트로 추출된 방문 장소 목록과 실제로 사용자가 방문한 장소목록이 일치하는 비율을 나타낸 것이다. 정확도는 추출된 방문 장소 목록의 개수/(실제 방문 장소 개수 + 방문하지 않았으나 방문으로 판단된 개수)로 계산된다. 오차범위가 0.3초일 때 가장 좋은 결과를 보였다.

4. 컨텍스트 생성 예

본 논문에서 제안한 컨텍스트 생성 시스템의 목적은 사용자의 정보를 스마트폰으로부터 수집하고, 수집된 정보를 바탕으로 추론 모델이나 서비스에 이용할 수 있는 컨텍스트를 생성하는 것이다. 이를 위해서 2주간 3명의 실험자를 대상으로 정보를 수집하고 수집한 정보를 분석하여 저수준 컨텍스트를 생성하였다. 그림 5는 스마트폰 상에서 수집된 통화내역이고 그림 7은 제안한 방법을 통해 분석된 저수준 컨텍스트이다. 그림 8은 저수준 컨텍스트를 이용하여 사용자의 상황을 추론하는 예이다. 하루동안 전화통화 35번, 장소 변화 20번, 현재 위치는 일터, SMS 40번, 시간은 9시인 점을 미루어 보아 사용자는 바쁨 상황임을 추론 할 수 있다. 이처럼 제안하는 방법을 통해 생성된 컨텍스트는 다른 추론 모델의 판단 근거로 사용 될 수 있다.

```
#시작시간, 종료시간, 수신/발신/부재, 전화번호
135812, 135812, 0, 060-601-0999
174221, 174302, 2, 018-253-9521
230149, 230158, 1, 02-900-9217
223904, 223936, 2, 011-888-3862
223837, 223856, 2, 011-888-3862
```

그림 6. 수집된 통화내역 정보 파일의 내용

```
DSTS:9 #하루 통화시간 (송신)
DRTS:379 #하루 통화시간 (수신)
DTTS:388 #하루 통화시간의 총합(수신/발신)
DSTSP:02-900-9217,9 #특정번호의 통화시간(송신)
DRTSP:02-900-9217,0 #특정번호의 통화시간(수신)
DSF:1 #전화를 건 횟수
DAF:1 #부재중 전화 횟수
DSIP:02-900-9217,0,0,0,0,0,0,0,0,...  #송신에 대한 하루 임팩트
MTTS:011-888-3862,338 #가장 오래 통화한 번호(송/수신)
MSF:02-900-9217,1 #가장 통화빈도가 높은 번호 (송신)
```

그림 7.통화 내역을 분석하여 생성한 컨텍스트의 일부

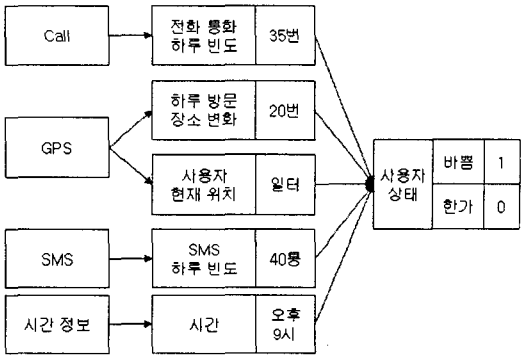

그림 8. 컨텍스트를 통한 추론의 예

5. 결 론

본 논문에서 개발한 시스템은 스마트폰에서 사용자 정보를 수집하고 수집된 정보에 통계적 분석과 임팩트를 적용하여 저수준 컨텍스트를 추출하고 추출된 저수준 컨텍스트를 통해 사용자 상태에 대한 추론이 가능함을 보였다. 향후 연구에서는 스마트폰에 다양한 센서를 추가하여 보다 정확하고 세세한 추론이 가능한 정보를 얻도록 할 것이다. 또한 사용자 정보를 저장할 때의 안정성 제고, 저수준 컨텍스트를 추출에 소요되는 시간의 단축, 그리고 방문 위치 판정의 정확도를 높이는 방법도 함께 고려해야 할 것이다.

참고 문헌

[1] P.Zheng and L.M.Nio,"The rise of the smart phone" *IEEE Distributed Systems Online,* Vol. 7, No. 3

[2] Mika Raento, Antti Oulasvirta, Renaud Petit,and Hannu Toivonen, "ContextPhone - A prototyping platform for context-aware mobile applications" *IEEE Pervasive Computing,* 51-59, 2005.

[3] Panu Korpipaa, Jani Mantyjarvi, Juha Kela, Heikki Keranen, and Esko-Juhani Malm, "Managing context information in mobile devices," *IEEE Pervasive Computing, Volume 2 Number 3, July-September 2003,* pp. 42-51, *ISSN 15361268*