

## 소셜 미디어를 기반으로 한 사용자 생활패턴 분석

서 승 현<sup>○</sup>, 정 철

소프트웨어 대학원, 정보과학 기술대학, 한국과학기술원

hefos@kaist.ac.kr, chul.chung@kaist.ac.kr

### Pattern Analysis of Social Media based User Daily Life

Seung Hyun Seo<sup>○</sup>, Chul, Chung

College of Information Science and Technology, KAIST

#### 요 약

사용자 행동 분석(프로파일링)은 “Social” 영역으로 점점 확대되고 있다. 대부분의 소셜 미디어는 대중과 사용자를 프로파일링 하기 위한 자신들만의 기술을 보유하고 있다. 예를 들어, 페이스북은 서비스 내 활동을 기반으로 한 프로파일링 알고리즘을 보유하고 있으며, 포스퀘어의 경우는 위치를 기반으로 한 프로파일링 기술을 보유하고 있다. 프로파일링은 위와 같이, 다양한 데이터를 기반으로 수행할 수 있다. 특히 지리정보는 다른 정보들과 합쳐지면서 점점 중요해지고 있다. 프로파일링은 대상에 따라서 개인과 그룹 프로파일링의 두 영역으로 나누어지는데 [1], 본 논문에서 제시하는 시스템은 개인의 영역에 초점을 맞춘 연구라고 할 수 있다. 본 연구의 목적은 개인 사용자의 소셜 미디어 사용습관에 시간에 따른 행동 반경과 위치정보를 가미하여 정량적으로 분석해 보는 것 이며, 소셜 미디어, 특히 페이스북에서 사용자 프로파일링 위하여 수집할 정보와 분석결과에 대하여 논하고 이를 위한 시스템을 제시한다.

#### 1. 서 론

사용자들이 어느 시점에 소셜미디어에서 “Post”, “Comment”, “Like”와 같은 활동을 가장 많이 하는지에 대한 연구는 이미 진행되고 있다. 이러한 종류의 연구는 대중을 대상으로 한 비즈니스 영역에 따른 분류까지 진행되어 있다 [6].

스마트폰이 보급이 가속화됨에 따라 소셜 미디어에서 위치정보의 영향이 커졌다. 스마트폰에 내장 되어 있는 GPS와 네트워크센서의 영향으로 위치정보를 쉽게 얻을 수 있게 되었으며, 이러한 기술은 LBS(Location Based System)이 확산되는데 큰 영향을 미쳤다고 할 수 있다. 페이스북은 Gowalla와 Instagram을 인수하면서 위치 서비스를 확장하려는 움직임을 보이고 있지만, 현재 빠르게 확장되지 못하고 있는 상태이다.

온라인 소셜 네트워크는 점점 위치기반으로 변해가고 있는 추세이다. 그 이유는 소셜 네트워크 안에서 위치정보는 유저간의 “tie-strength”, “link-prediction”, “추천”과 같은 여러 가지 목적으로 쓰일 수 있기 때문이다 [6,7]. 소셜 미디어에서 위치 정보와 관련된 연구는 “Social Graph”와 관련되어, 사람들과의 관계에 대한 연구로서 진행되어 오고 있다 [3,4].

위치 추적에 대한 연구는 GPS 뿐만 아니라 대체 센서를 이용한 연구들도 활발히 진행되고 있다. Foursquare의 전신인 Dodge Ball은 SMS를 사용하여 서비스를 제공했었고, 3G [2], GSM, Wi-Fi, 가속센서를

이용한 위치기반 연구들도 계속 진행되어오고 있다.

보통 LBS는 GPS를 사용하지만, 본 연구의 모바일 시스템에서는 안드로이드 시스템의 3G, GPS, Wi-Fi를 사용한다. 본 연구에서 제시하는 시스템에서는 모바일 폰의 에너지 문제를 해결하기 위하여 구글 I/O 2011 모범 실무(Best Practice)를 벤치마킹 하였다.

본 연구에서 제시하는 Geo-tagging 시스템은 두 가지 자료출처(Data source)를 갖는다. 위치 감지를 위한 모바일 측면과 페이스북 데이터 수집을 위한 서버를 사용한다. 두 가지 시스템에서 수집한 데이터를 비 동기 Geo-tagging을 통하여 분석하며 사용자의 행동패턴을 분석해 보았다.

스마트폰과 더불어 페이스북이 우리 삶에 큰 영향을 끼치고 있다는 것은 누구나 동의 하지만, 과연 어느 정도로 일상생활에 영향을 미치는지 측정하기는 굉장히 힘들다. 본 연구의 실험결과는 페이스북이 우리 생활에 미치는 영향을 측정하고, 페이스북 활동정보에 위치정보가 추가되었을 때의 파급효과를 예상해 볼 수 있는 자료가 될 수 있을 것이다. 또한 개인의 사용 성향과 페이스북이 쓰이는 용도를 유추해 볼 수 있을 것이다.

본 논문은 앞서 진행된 연구들 시스템에 적용하여 정보를 수집하고, 이전에 집중한 범위와는 다르게 사용자 본인의 행동 분석에 초점을 맞춘다. 앞서 언급한 연구들과 같이 소셜 미디어에서 사용자 프로파일링을 진행하는 점은 같지만, 연구의 초점을 관계가 아닌

개인의 행동에 맞추고 시간과 위치정보를 연결하려는 시도를 하고 있다는 점이 다르다고 할 수 있다.

## 2. 관련연구

### 2.1 사용자 프로파일링

사용자 프로파일링은 보통 지식기반(Knowledge-based)과 행동기반(behavior-based)로 나누어진다 [1,10,11]. 또한 정보과학(Information Science)에서는 프로파일링 대상에 따라서 개인과 그룹으로 분류하고 있는데 [15], 본 연구는 행동기반 개인 프로파일링으로 볼 수 있다.

### 2.2 위치감지와 소셜 미디어

위치정보가 중요한 이유는 따로 정보를 가공하지 않더라도 위치정보만으로 사람의 행동반경을 알 수 있다는 것이다. 행동반경과 더불어 사진에 추가된 위치정보의 중요성은 앞선 연구에서도 이미 언급되어 왔다 [14]. 또한 위치정보는 소셜 미디어 내의 다른 정보들과 결합되면 프로파일링의 질을 높이고 확장할 수 있다 [4,5,13,14]. 본 연구에서는 소셜 미디어의 사용과 위치정보를 비 동기적인 방법으로 결합하여 사용자의 행동패턴을 분석하는 시스템을 디자인한다. 또한 소셜 미디어를 자주 사용하는 장소와 시간을 정형화 하여 재해석하고, 분석하는 것에 초점을 맞춘다.

### 2.3 에너지 소모

모바일 폰을 사용한 프로파일링 연구에서는 배터리에 관한 연구가 상당히 진행되어 왔다. 그 중, 본 시스템은 Google I/O 2011에 모범 실무(Best Practice)로 소개된 Framework를 기본으로 사용한다 [8]. 또한 [9] 에서 제시한 LBS를 위한 4가지 원칙 (SS, SR, SP, SA)중 SS(Sensing Substitution)와 SP(Sensing Piggybacking)을 적용한다.

## 3. 시스템 디자인

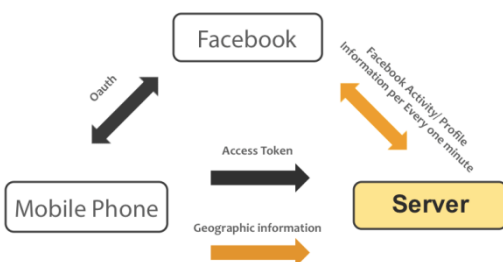


그림 1: 시스템 흐름.

본 시스템의 자료출처는 사용자의 모바일폰, 페이스북

서버이다. 전체적인 시스템의 흐름은 다음과 같은 흐름을 갖는다. 1. 모바일폰에서 페이스북으로 인증과정(Oauth)을 거쳐 Access token을 얻는다. 2. 얻은 Access token을 Geo-tagging 서버로 전송한다. 3. Geo-tagging 서버는 전송받은 Access token을 사용하여 사용자의 Facebook 사용내역을 수집한다. 4. 모바일폰은 30분마다 Geo-tagging 서버로 위치정보와 위치감지 센서의 종류를 전송한다. (그림 1)

### 3.1 App 디자인

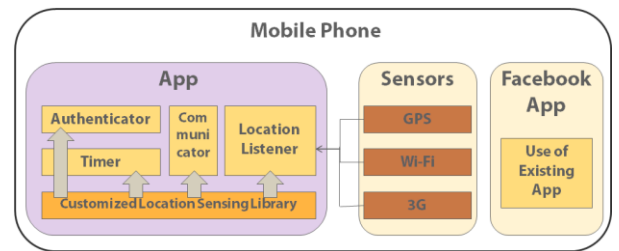


그림 2: App.

모바일폰에서는 위치정보만을 수집한다. 센서는 GPS, Wi-Fi, 3G를 사용하며, 최초 1회 인증 외에 연구에 사용되는 App을 사용해야 하는 제약 사항은 두지 않았다. 배터리 문제로 인하여 GPS는 자주 사용하지 않기 때문에 대부분 Wi-Fi, 3G를 사용할 것으로 판단하여, 위치 감지에 사용된 센서정보도 함께 서버로 전송한다. 위치감지 빈도는 30분으로 설정했다.

### 3.2 서버 디자인

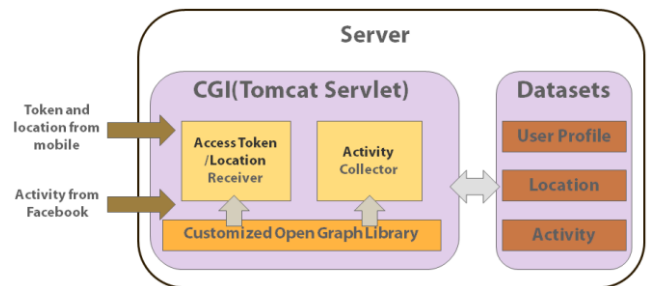


그림 3: 서버 패키지

서버에서는 사용자 당 최초 1회 Access token을 전송 받고 유저에 대한 데이터 베이스를 구축한다. 페이스북의 특성상 친구, 그룹과 같은 데이터는 변경 될 수 있기 때문에 위치정보를 받는 30분마다 정보를 업데이트 한다. 정보는 수집이 종료된 시점부터 1분 후 다시 수집을 시작한다. 정보를 수집하는 방법에 있어서 여러 가지 제약사항들이 존재지만 본 논문에서는 정기적으로 수집하는 정보와 이에 관한 제약사항만 언급하기로 한다. 아래 정보는 사용자가 직접적으로 참여한 정보들이다.



그림 4: 위치 정보의 군집화

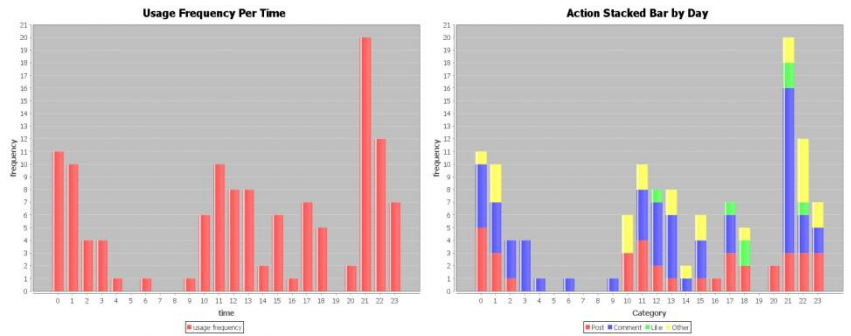


그림 5: 시간대별 소셜 미디어 사용 패턴

- **포스트(Post):** 자신의 혹은 친구의 담벼락에 글을 작성하는 활동으로 댓글이 작성될 때마다 업데이트된다. 제약사항으로는 일정 시간이 지난 글은 업데이트 되지 않는다.
- **댓글(Comment):** 자신의 글 또는 친구의 담벼락 글(Object)에 종속되는 글을 작성하는 활동으로 업데이트되지 않는다.
- **좋아요(Like):** 위의 두 가지 활동과는 다르게 작성시간(Created time)과 업데이트 시간이 존재하지 않는다. 이전 정보 수집시간과 현재 정보 수집시간을 토대로 유추한다.
- **소셜 플러그인(Social-plugin):** 제 3진영(Third party)의 웹사이트나 App을 사용한 활동을 지칭한다. 일부 정보는 수집하는데 어려움이 있다.

#### 4. 결론 및 향후 연구.

그림 3,4는 관찰기간 동안의 행동 전체를 패턴화 한 것이다. 밤시간에 수집된 위치정보가 집중된 장소를 주거지, 낮 시간에 집중된 장소를 직장으로 유추할 수 있으며 이동경로를 간단하게 유추할 수 있었다. 위치정보에 페이스북 사용정보를 Geo-tagging 하여 사용자마다 소셜 미디어의 사용 패턴을 알아 내는 것뿐만 아니라 생활 패턴 및 사용자의 속성을 유추해 볼 수 있다. 예를 들어 위의 그림에서는 6시쯤 기상, 9시 출근 및 회사에 도착하여 페이스북을 쓰기 시작한다는 것을 유추 해 볼 수 있으며 위치 별 활동의 차이도 알아볼 수 있다. 이외 여러 가지 성향의 추론들이 가능해 진다.

본 연구결과의 한계는 사용자의 위치정보는 3G와 Wi-Fi를 통해 수집한 결과로 오류가 있을 수 있다는 점, 그리고 정보를 수집할 수 있는 기간이 충분히 길지 않아 전체적인 패턴만을 알아볼 수 있었다는 점이다. 그리고, 이동경로의 경우 지하철을 이용하는 경우 오류확률이 높아질 수 밖에 없다는 점이 한계점이다. 또한, 페이스북을 적극적으로 사용한 정보만 수집할 수

있기 때문에, 사용자가 적극적으로 활동하지 않은 정보는 배제 할 수 밖에 없었다.

본 연구는 향후 다음과 같은 형태로 발전할 수 있을 것이다. 1. 위치를 공유하는 그룹과 위치를 공유하지 않는 그룹간의 차이에 관한 연구. 2. 온라인/오프라인 그룹과 사용자간의 “Tie-strength”차이에 관한 연구. 3. 텍스트 분석을 통한 위치의 변화에 따른 유저의 행동과 감정에 관한 연구. 4. 소셜 미디어 사용 패턴의 변화를 토대로 한 사용자의 비정형적 행동의 감지에 대한 연구.

#### 5. 참고문헌

- [1] STUART E. MIDDLETON et al., Ontological User Profiling in Recommender Systems, University of Southampton, ACM, January 2004.
- [2] Ram Keralapura et al., Profiling Users in a 3G Network Using Hourglass Co-Clustering, MobiCom'10, September 20-24, 2010
- [3] Nathan Eagle et al., Inferring friendship network structure by using mobile phone data, MIT Media Lab ,PNAS ,September 8, 2009 ,vol. 106 ,no. 36
- [4] Yu Zheng et al., Recommending Friends and Location Based on Individual Location History, Microsoft Research Asia and Tsinghua University, ACM, 2011
- [5] Jukka-Pekka Onnela et al., Geographic Constraints on Social Network Groups, PLOS ONE , April, 2011
- [6] STRATEGIES FOR EFFECTIVE FACEBOOK WALL POSTS: A STATISTICAL REVIEW, BUDDY MEDIA, 2011
- [7] David J. Crandall et al., Inferring social ties from geographic coincidences, PNAS ,December 28, 2010 ,vol. 107, no. 52
- [8] <http://code.google.com/p/android-protips-location/>, 2011 Google I/O
- [9] Zhenyun Zhuang et al., Improving Energy Efficiency of Location Sensing on Smartphones , Mobisys'10, ACM, June 15, 2010
- [10] Alfred Kobsa, User Modeling:Recent Work, Prospects and Hazards, Alfred Kobsa, Department of Information Science, University of Konstanz, GERMANY.
- [11] SEBASTIANI, Machine learning in automated text categorization, ACM Computing Surveys, 2002
- [12] Jeffrey McGee et al., A Geographic Study of Tie Strength in Social Media, Texas A&M University, October, 2011
- [13] Justin Cranshaw et al., Bridging the Gap Between Physical Location and Online Social Networks, UbiComp '10, Carnegie Mellon University, Sep 26-Sep 29
- [14] Jongsea Park et al. Geo-Web, KAIST, Samsung Electronics, Kiss 2010.
- [15] [http://en.wikipedia.org/wiki/Profiling\\_practices](http://en.wikipedia.org/wiki/Profiling_practices). wikipedia, profiling practice