

소셜 네트워크 영향력 측정 모델 제안

이승희*, 박영호*

*숙명여자대학교 멀티미디어학과
e-mail : arishine@sookmyung.ac.kr
yhpark@sm.ac.kr

A Proposal of a Model for Measuring Influences of Social Networks

Seung-Hee Lee*, Young-Ho Park*

*Dept of Multimedia Science, Sookmyung Women's University

요 약

최근, 스마트폰 가입자 수가 급격히 증가하는 가운데, 사람과 사람의 관계에 중심을 두는 서비스인 SNS가 급격히 성장하고 있다. 모바일 디바이스를 사용한 소셜 네트워크가 접목되어 사용자들 사이에 자연스러운 의사 교류 관계가 형성되며, 장소나 이벤트를 기반으로 빠르고 끊임없이 정보들이 발생하게 되었다. 그러나 너무 많이 발생하는 불필요한 데이터들로 인해 의미 있는 정보, 필요한 정보를 파악하는 데 어려움이 생긴다. 이러한 문제점을 해결하기 위해, 소셜 네트워크 서비스를 통해 제공되는 정보들 가운데 영향력 있는 정보가 무엇인지 판단하는 연구가 중요시 되고 있다. 본 논문에서는 소셜 네트워크 영향력 측정을 위해 장소를 기반으로 한 동적 소셜 커뮤니티에서의 소셜 네트워크 액티비티를 중심으로 이벤트에 대한 영향력을 측정하는 방법에 대해 고찰하고, 새로운 영향력 측정 방안에 대해 제안한다.

1. 서론

최근 1년간 스마트폰 사용자가 급증하고 있다. 스마트폰과 같은 모바일 디바이스는 장소의 제약에서 벗어나 동적인 커뮤니케이션을 유발하는 특징을 가지고 있다. 이는 3G 활성화와 Wi-Fi 지역 확대, 그리고 GPS 탑재와 같은 기술적 환경으로 인해 가능해졌으며, 이러한 무선 통신의 발달과 위치기반 서비스(Location-based Services) 발달로 관련 서비스들이 등장하여 위치 정보에 대한 정보 요구가 높아지고 있다[4]. 위치 정보와 데이터 공유 등 모바일 사용자간의 사회적 상호작용을 바탕으로 한 소셜 네트워크 서비스들이 점차 늘어나고 있는 실정이다.

언제 어디서나 장소와 시간에 구애 받지 않고 모바일 디바이스를 통해 자신들의 동일한 생각이나 의사를 공유하고 있지만, 때로는 장소와 만남을 통해서 존재를 확인하고 정보를 교환하고자 하는 모임도 활성화 되어 특정 장소에 의미를 부여하고 관계성을 향상시키고자 한다. 이러한 장소를 기반한 이벤트에 주목할 필요가 있다.

동호회 모임, 예술 활동과 같은 이벤트에, 이제는 모바일 디바이스를 사용한 소셜 네트워크를 통해 사람들 간의 자연스러운 의사 교류 관계를 형성하여 많은 관심사를 빠르게 공유할 수 있게 되었다. 끊임없이 발생하는 이러한 동적인 커뮤니케이션들은 빠르게 많은 정보들을 서로 공유할 수 있지만, 너무 많이 발생하는 불필요한 데이터들로 인해 어떤 것이 의미 있고 자신이 필요한 정보인지 파악하기 어렵다는 문제점이 있다.

이러한 문제점 때문에 소셜 네트워크 서비스에서 제공하는 정보들 가운데 더 영향력이 있는 정보가 무엇인지 판단하는 연구가 중요시 되고 있다. 웹 서비스에서 어느 사이트가 영향력이 높고 신뢰성 있는 사이트인지 판단하는 연구들이 있었다면, 최근에는 소셜 네트워크 서비스에서 영향력 있는 사용자를 통해 연구나 마케팅에 적용하기 위한 연구들이 진행되고 있다. 특별히 파급효과가 가장 크다고 여겨지고 있는 트위터에서의 팔로워 수, 작성한 트윗의 개수, 리트윗 수 등과 같은 요소로 영향력을 측정하는 연구가 있다[1, 2].

영향력이 있는 사용자, 즉 인플루엔셜이 누구인가를 찾아 영향력 있는 대상, 장소, 모임 등에 집중하는 것이 마케팅의 관점에서 효과가 있고 비용이 절감된다고 여겨지고 있다. 또한 트위터 팔로워 수, 페이스 북에서의 소통을 통한 영향력, 서비스 가입자수나 검색 빈도에 대한 영향력도 중요하지만, 온라인 상에서만 영향력이 아니라 장소가 가지는 실제성을 부여하고 현실 속에서 소셜 커뮤니케이션을 통해 영향력을 측정할 수가 있다. 장소를 중심으로 한 대표적인 영향력 측정의 연구 예로, 각 매장들을 소셜 네트워크를 구성하는 노드로 보고 매장 간의 관계를 링크로 모델링 하여 매장의 확산 기여도를 영향력이라고 보는 스타벅스 네트워크를 활용한 소셜 네트워크 영향력 모델 연구[7]가 있다. 그러나 이런 기존 연구들로는, 앞에서 중요하게 여겼던 소셜 네트워크 서비스로 이루어지는 동적인 모임에 대해서 영향력을 측정할 수 없으며, 소셜 네트

워크 액티비티들이 어떠한 효과를 주는 지를 밝힐 수 없었다. 또한 이러한 영향력에 대해서 정성적인 질의의 답을 구하는 데 어려움이 있다. 본 고찰에서는 소셜미디어 서비스에서 영향력을 측정하는 사례를 분석하고, 이를 보완할 영향력 모델을 소개하는데 초점을 두고자 한다.

2. 영향력 측정 관련 연구

본 장은 관련 연구로서, 기존의 소셜 네트워크에서의 영향력 측정 연구들에 대해 기술한다.

2.1 트위터 인플루엔셜

소셜 네트워크에서 영향을 많이 미치는 사람을 인플루엔셜(영향자, Influential)이라고 한다. 입소문 마케팅에서 가장 중요한 것 중 하나가 바로 누가 인플루엔셜인가를 찾아내는 것이다[1].

인플루엔셜을 찾기 위한 다양한 연구들 가운데, 특히 최근에는 현재 가장 영향력 있다고 여겨지는 소셜 네트워크 서비스인 트위터를 통해서 인플루엔셜을 측정하기 위한 연구들이 많이 있다. 트위터에서는 팔로워(follower) 수에 따른 인플루엔셜과 작성한 트윗의 개수, 리트윗(retweet) 수 등과 같은 요소로 영향력을 측정할 수 있다.

트위터의 유행으로 이를 이용한 각종 서비스들이 제공되고 있는데, 특별히 트위터 사용자들의 순위를 제공하는 트위터 카운터(twittercounter)[3]와 트위터 그레이더(twittergrader)[4]와 같은 서비스는 가 트위터 사용자 중 영향력 있는 사용자가 누구인지 파악할 수 있는 자료가 된다[5].

먼저 트위터에서 영향력을 측정하는 요소로 사용자의 팔로워 수(Indegree influence), 리트윗 횟수(Retweet influence), 멘션 수(Mention influence)의 다양한 기준을 방대한 데이터에 적용하여 비교한 것이 있다. 이 연구에서는 이 세 가지의 액티비티를 사용자의 영향력 타입을 나타내는 것으로 설명하였다[6].

팔로워 수에 의한 인플루엔셜을 정의하는 경우, 팬이나 지지자가 많은 유명 인사나 연예인과 같은 사용자가 영향력 있는 사용자로 랭크 되며[7], 다른 사람이 올린 정보를 가져와서 본인의 팔로워들에게 그대로 혹은 의견을 덧붙여 전달하는 리트윗의 경우는 유명인이 될 수도 있지만 주로 뉴스매체나 오피셜 공식 트위터가 주로 랭크된다.

다음으로 TunkRank[2] 연구가 있다. 이는 Daniel Tunkelang이 트위터에서 트윗이 리트윗 될 확률을 고려한 영향력 측정 연구이다. 다음 그림 1과 같은 식을 통해서 영향력을 측정한다.

$$Influence(X) = \sum_{Y \in Followers(X)} (1 + p * Influence(Y)) / |Following(Y)|$$

그림 1. TunkRank의 영향력 측정 연구

Influence(X)는 사용자 X가 작성한 트윗을 읽을 사람들의 수의 기대값이며, Followers(X)는 X를 팔로우하는 사용자의 집합이고 Following(Y)는 Y가 팔로우 하는 사용자의 집합이다. Y가 X의 팔로워라면 $1/|Following(Y)|$ 의 확률로 X가 작성한 트윗을 읽는다고 가정한다. Y가 X의 트윗을 읽을 때 p의 확률로 리트윗을 한다. 이런 가정 아래 Influence(X)는 X가 쓴 트윗을 읽게 되는 사람의 수를 나타내게 된다.

기존의 소셜 네트워크 영향력 측정 연구에서는 주로 영향력을 판단하는 기준을 영향력 있는 사용자인 인플루엔셜에 중점을 두어 연구가 진행되었다. 소셜 네트워크에서의 인플루엔셜은 네트워크의 종류나 측정하고자 하는 영향력 의미에 따라 달라질 수 있으므로 한마디로 정의하기는 어려우나, 일반적으로 다른 사용자에게 영향을 주고 주요한 정보를 전달하는 사용자를 이야기한다.

트위터와 같은 마이크로 블로그 형태의 소셜 네트워크 서비스는 인플루엔셜의 영향력을 측정하기 위해 사용자들 간의 관계도와 트윗의 양 등을 통해서 판단할 수 있는 요소가 있으나, 사용자가 실제로 방문하는 장소, 또는 실제로 참여하는 모임과 같은 경우는 그 영향력을 측정하는 기준이 모호하다는 단점이 있다.

2.2 스타벅스 인플루엔셜

실제 장소를 기반으로 하여 영향력을 측정하는 연구로서, 스타벅스의 네트워크를 활용한 소셜 네트워크 영향력 모델 연구[8]가 있다. 이 연구에서는 각 매장을 소셜 네트워크를 구성하는 하나의 노드로 보고, 매장 간의 거리를 매장간 관계(Link)로 모델링 하여, 매장의 영향력을 측정할 수 있는 모델을 연구하고 분석하였다.

이 연구에서는 스타벅스 전파에 영향을 주는 주요 요소로 이노베이션, 커뮤니케이션 채널, 시간, 소셜 시스템을 이야기 하였으며, 영향력을 측정하는 요소로 방문자 수에 따른 가중치, 매장간 거리에 따른 가중치를 두어 인플루엔셜 모델을 제안하였다. 스타벅스 네트워크 영향력 모델 연구에서 제안한 내용은 다음과 같다. 한 매장이 가지고 있는 영향력은 그 매장에 연결된 모든 다른 매장으로 전파되고, 이는 다시 다른 매장에 연결된 모든 매장으로 계속 전파된다. 따라서 임의의 매장이 가지고 있는 영향력은 재귀적으로 정의할 수 있으며, 영향력 전파 단계가 거듭될수록 감쇄효과에 의해 결국 일정한 값으로 수렴하게 된다. 이때 원매장의 영향력은 매장이 확대되는 가중치를 고려해서 다른 매장들에 전파된다. 다이내믹 인플루엔셜 모델에서 나타난 식에서 볼 수 있듯이 영향력을 e(effect)라고

하고, 가중치 계수의 곱을 통해서 계산하였다. 방문자 수에 따른 가중치 계수 V (현매장, 다른 매장)는 각 매장 방문자수를 모르는 상황이므로 1로 가정하여 나타낸다. 매장 간 거리에 따른 가중치 계수는 W (현매장, 다른 매장)로 표현하고, 기존 주변 매장 오픈 시점과의 차이에 따른 영향력 보정계수는 f_{year} 이다. 이웃 매장과 거리 증가함에 따른 영향력 감쇄효과 f_{step} 를 0.5로 가정하고, 각 노드가 가진 초기 영향력 e 를 1로 가정하였다. 스타벅스 네트워크 영향력 모델 연구의 경우, 매장 방문자 수와 주변 매장, 그리고 매장 간의 거리를 가중치로 주어 영향력을 측정하였다면, 시간, 사용자, 거리 등을 가중치로 둔 것을 유사하게 볼 수 있으나, 제안하는 모델에서는 이벤트가 발생한 장소의 반경 범위, 이벤트가 지속되는 시간, 사용자의 수, 메시지 빈도수 등을 주요하게 보고 이를 영향력 측정에 활용되는 액티비티 데이터로 여긴다. 앞에서 설명한 기존 연구들에서는 영향력 있는 사용자, 영향력을 가지는 매장에 대한 영향력 크기만 측정할 수 있는 단점이 있다.

3. 제안하는 영향력 모델

본 논문에서 제안하는 모델은 장소와 시간, 소셜 시스템을 대상으로 하고, 소셜 네트워크에서의 누군가의 영향력을 구하는 것이 아닌 소셜 네트워크 액티비티를 이용하여 이벤트의 영향력을 측정하는 연산으로하는 모델링 방안이다. 이를 위해, 본 논문에서 설명하는 이벤트의 영향력에 대해 정의하고, 장소, 시간, 이벤트, 사용자 4가지 요소를 의미하는 Quad Relation Factors를 제안한다. 이를 통하여, 이벤트의 속성과 영향력을 나타낼 수 있는 장소와 시간을 중심으로 나타내는 Human Activity Graph로 모델링 하는 방안을 설명한다.

장소, 시간, 이벤트, 사용자 네 가지의 연관 요소들을 통해서 장소를 기반으로 하여 영향력에 대한 질의 연구에 대해서 설명한다. 이 네 가지 주요 요소들은 소셜 네트워크 액티비티를 통하여 영향력을 가늠할 수 있도록 하는 요소로, 이들 정보의 중요도를 통해서 다양한 질의 처리를 할 수 있다. 질의에 따라 이 네 가지의 Quad Relation Factors만이 아니라, 장소-이벤트, 장소-사용자, 이벤트-시간, 이벤트-사용자 등 두 가지의 요소에서의 중요도를 파악하는 방법과, 장소-이벤트-시간, 장소-이벤트-사용자 등 세 가지의 요소에서 중요도를 파악할 수 있다. 이러한 서브 그래프에서의 중요도를 계산하여 Top-k를 구할 수가 있으며, 이는 어떠한 모임의 영향력을 나타내는 데 근거가 되는 요소들이다. 장소에서 지역적인 특성을 갖고 동적인 이벤트를 통해 특정 모임들이 얼마나 생성되고, 어떻게 분산되어 있는지 밀집되어 있는지를 파악할 수 있는 요소들이고, 사용자들의 행동 특성이 나타날 수 있다.

이제, 본 장에서는 제안하는 Human Activity Graph에 대하여 설명한다. Human Activity Graph는 이벤트의 속성과 영향력을 반영한 그래프로써, 위도와 경도로 나타낼 수 있는 장소와 시간 축을 중심으로 3차원 그래프를 구축

하여 표현한다. 이는 특정 지역에서 발생한 이벤트가 시간의 흐름에 따라 표현되는 것으로, 단순히 작은 지역만을 표현하는 것이 아니라 위도, 경도로는 전지구상의 위치를 표시할 수 있고, 또한 시간도 특정 기간의 시간이 아닌 0부터 ∞ 까지의 시간을 나타낼 수 있으므로 Human Activity Graph 라고 명명한다.

이러한 Human Activity Graph는 좀 더 확장하면 인문학, 역사학과 연계되어 시간과 공간을 중심으로, 기존에 어떤 지역에서 어떤 사건이 있었는지를 나타내어 인류의 생활이나 문화, 다양한 활동들을 그래프로 표현하는 연구로 발전시킬 수 있다.

본 논문에서 제안하는 내용은, 장소 기반의 소셜 네트워크 활동을 통해서 장소, 시간, 이벤트, 사용자의 데이터를 수집하여 이를 통한 영향력을 측정하는 것을 포커스 하여 Human Activity Graph로 표현한다. 이러한 내용은 x, y, z 축에서의 시간과 위도, 경도와 매핑된다. 장소와 시간을 중심으로 동적으로 생성되고 소멸되는 이벤트 자체는 그래프 상에서 원기둥 모양과 같은 3차원의 도형으로 표현된다. 이는 발생하는 이벤트가 특정 장소를 중심으로 얼마만큼의 반경으로 위치되어 있는지 크기를 나타내는 원과, 이벤트의 생성에서 소멸 시점까지를 시간 축을 따라서 나타내므로 이벤트의 지속 시간에 따라 도형의 길이가 달라지는 내용을 반영한다.

이러한 그래프 모델에서는 도형의 색을 이벤트의 카테고리 구분 용도로 사용한다. 이벤트의 카테고리, 즉 어떠한 분야, 종류의 이벤트인지를 표현한다. 따라서 그래프에서 색으로 분류되는 도형의 분포를 보면, 그 위치한 지역을 판단하여 해당하는 이벤트가 어느 카테고리에 속하는지, 어느 지역에서 어떤 이벤트가 주로 발생하는 특징을 가지는 지 파악할 수 있는 자료가 될 수 있다. 예를 들어, 빨간 색상 도형이 축구 경기, 노란 색상 도형이 대규모 집회와 같은 이벤트를 나타낸다고 가정한다. 제안하는 모델에서는 특정 색상의 도형이 모여있는 지역은 특정 이벤트(예, 축구 경기)가 많이 발생하는 지역이고, 도형의 크기는 넓은 지역은 아니지만 시간에 따라 자주 발생하는 이벤트라고 판단할 수 있다. 또한 남쪽에서는 넓은 범위로 긴 시간 지속되는 집회가 발생한다는 것을 볼 수 있다. 그리고 도형의 크기가 바로 이벤트의 크기, 즉 영향력의 크기를 나타낸다. 앞에서 언급한 것과 같이 장소에서의 반경과, 이벤트 지속 기간 등의 내용으로 인하여 3차원 도형의 부피로서 크기가 결정된다. 장소의 반경이 커질수록, 시간 단위가 증가할수록 도형의 부피가 커지므로, 이를 이벤트의 크기가 증가하는 것으로 계산해 볼 수 있다. 이처럼 도형의 분포, 크기, 모양 등을 통하여 이벤트의 영향력을 파악할 수 있다. 그리고 “가장 영향력 있는 이벤트가 발생한 지역?”, “특정 이벤트가 군집적으로 발생하는 지역은?”, “작은 모임이지만 지속적인 이벤트는?” 과 같은 정성적인 질의의 답에 대해서 정량적으로 계산하여 구할 수가 있다.

Human Activity Graph를 표현하기 위해서는 이벤트의

장소, 시간, 사용자에게 대한 로그 데이터가 필요하다. 이 로그 데이터에서 장소의 위치 정보인 위도, 경도의 값은 GPS 정보로부터 받아 저장되고, 이러한 로그 데이터에서 3차원 인덱스를 사용하여 데이터를 빠르게 검색할 수 있도록 한다.

4. 결론 및 향후연구

본 논문에서는 소셜 네트워크 액티비티를 통하여 이벤트의 영향력을 측정하는 모델을 제안하였다. 영향력을 측정하기 위한 소셜 네트워크 액티비티 데이터는 장소, 시간, 이벤트, 사용자 네 가지 요소인 Quad Relation Factors를 기준으로 설정하였다. 정성적인 질의를 정량적으로 처리하기 위한 방안으로 장소를 기반으로 하여 사용자들 간의 소셜 관계를 응집시키고, 시간에 따른 이벤트의 영향력을 나타낼 수 있는 Human Activity Graph 모델링 방안을 제안하였다. 또한 이를 이벤트의 크기를 측정하는 Metric들을 제안하여 설명하였다. 이를 기반으로 하부적 시스템과 Metric을 기반으로 더 발전시켜, 더 많은 데이터로 실험을 해야 한다. 본 연구에서 제안한 Human Activity Graph의 세부적인 다차원 질의 처리 내용은 향후 연구에서 계속 진행하고자 한다.

참고문헌

- [1] 박호성, 광해운, 차미영, 문수복, "소셜 네트워크에서의 인플루엔셜 랭킹, " 정보과학회지, Vol.28, No.3, pp.24~30, 2010.
- [2] D. Tunkelang, A Twitter Analog to PageRank, "http://thenoisychannel.com/2009/01/13/a-twitte-ranalog-to-pagerank, " 2009
- [3] 트위터 카운터, "http://twittercounter.com"
- [4] 트위터 그레이더, "http://twitter.grader.com"
- [5] 박지혜, 김보현, 이명준, 권영근, "TwitNet : 트위터 사용자들의 관계를 시각적으로 나타내는 Cytoscape 플러그인 개발, " 한국정보과학회 한국컴퓨터종합 학술발표논문집, Vol.37, No.1, 2010
- [6] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy," In Proc. of International AAAI Conference on Weblogs and Social Media(ICWSM), 2010
- [7] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" WWW '10: Proceedings of the 19th international conference on World wide web, 2010
- [8] Minkyoung Kim, Byoung-Tak Zhang, June-Sup Lee, "Dynamic and Static Influence Models on Starbucks Networks," International Conference on Advances in Social Network Analysis and Mining, 2009