

웹 3.0 시대의 소셜 네트워크 영향력 모델 연구

문효정*, 박영호**

*한국방송통신대학교 미디어영상학과

**숙명여자대학교 멀티미디어과학과

e-mail : m_jiwon@naver.com, yhpark@sm.ac.kr

An Influence Model for Social Networks in Web 3.0

Hyo-Jung Moon*, Young-Ho Park**

*Dept of Media Arts & Sciences, Korea National Open University

**Dept of Multimedia Science, Sookmyung Women's University

요 약

최근, 참여, 공유, 개방의 키워드를 가지고 웹 2.0 시대가 열렸다. 이러한 웹 2.0 시대의 키워드가 익숙해지기도 전에, 우리는 소통, 융합, 개인화라는 키워드를 지닌, 웹 3.0 시대를 맞이하게 되었다. 새로운 개념과 용어들은 이미 우리에게 혁명적 수준으로 주어질 뿐만 아니라, 이 시대를 살아가기 위한 기본 양식으로 자리매김하게 되었다. 본 논문에서는 최근 소셜 네트워크 서비스(SNS)의 대표주자인 트위터 영향력 모델과 소통을 중요시하는 스타벅스 네트워크 영향력 모델을 소개하고, 이와 대비되는 새로운 영향력 표현 모델을 제안하고자 한다.

1. 서론

인터넷이 보편화 되어 감에 따라, 2005년 이후 웹 2.0 시대가 시작되었다. 웹 2.0은 참여, 공유, 개방을 키워드로 모든 사용자가 참여할 수 있는 공간을 제공해 주었고, 이로 인해 아래로부터의 변화를 선도하여 왔다.

그러나, 2010년을 기점으로 웹 3.0의 개념이 등장하게 되었고, 소통, 융합, 개인화라는 키워드를 표방하였다.

첫째, 소통이란, 단방향의 소통에 참여한 다수간 또는 서비스 제공자와 이용자간, 다방향의 의사소통을 지원하는 웹의 필요성을 의미하는 것으로 b2c, c2c, b2b 보다 원활하게 정보를 소통할 수 있게 된다. 또한 우리들이 체험할 수 있는 소통 방식도 텍스트뿐만 아니라 즉 오디오, 비디오 등을 이용한 다양한 멀티미디어 방식들이 사용되는 특징을 나타내고 있다.

둘째, 융합이란, 융합은 정보와 기술의 융합 측면이 있다. 먼저 정보의 융합은 이용자들이 정보를 소유하게 되면 이 정보들을 또 다른 정보와 융합하여 새로운 가치를 지닌 새로운 정보로 재생산하는 것을 말한다. 기술의 융합은 플랫폼의 융합을 의미한다. 현재는 웹, 데스크 탑, 모바일 장치 등이 별도로 동작하였지만, 이제 이들간의 경계가 무너지 융합되어 사용할 수 있는 시스템으로 진화하는 것이다. 따라서 웹에서 실행하던 것들을 그대로 데스크 탑과 핸드폰 등에서 실행하고 서로 연동하여 사용하는 것이 일상화 되어가고 있다.

마지막으로 개인화란, 자신의 정체성을 반영한 정보들을 정리하고 타인들과 소통할 수 있는 방식을 말한다. 대표적으로 페이스 북 같은 매체가 있으며,

앞으로 다가올 개인화는 대량생산되는 정보들로부터 자신이 필요로 하는 정보만을 추출하여 보여줄 수 있는 맞춤형 웹이 될 것이다.

본 연구에서는 이러한 개념을 기반으로 하여 소셜 네트워크 영향력 측정을 위해 소셜 네트워크 액티비티를 중심으로 이벤트에 대한 영향력 측정 방안을 제안한다.

2. 관련연구

본 장에서는 기존 연구들 가운데, 트위터와 스타벅스 네트워크 모델을 중심으로 설명하고, 이들 방안이 제안하는 방법과 무엇이 다른지를 소개한다.

첫째, 트위터 영향력 모델은 특히 최근에는 현재 가장 영향력 있다고 여겨지는 소셜 네트워크 서비스이다. 이것은 대표적인 소셜 네트워크의 사례 연구라고 할 수 있다. 트윗의 영향력을 측정하기 위한 연구들이 많이 있다[1]. 트위터에서는 팔로워 수에 따른 인플루언셜과 작성한 트윗의 개수, 리트윗 수 등과 같은 요소로 영향력을 측정할 수 있다. 트위터의 유행으로 이를 이용한 각종 서비스들이 제공되고 있는데, 특별히 트위터 사용자들의 순위를 제공하는 트위터 카운터 [5]와 트위터 그레이더 [2]와 같은 서비스는 가 트위터 사용자 중 영향력 있는 사용자가 누구인지 파악할 수 있는 자료가 된다[3]. 다음으로 TunkRank[4] 연구가 있다. 이는 Daniel Tunkelang 이 트위터에서 트윗이 리트윗 될 확률을 고려한 영향력 측정 연구이다.

다음 그림 1 과 같은 식을 통해서 영향력을 측정한다. Influence(X)는 사용자 X 가 작성한 트윗을 읽을 사람들의 수의 기대값이며, Followers(X)는 X 를 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.

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

그림 1. TunkRank의 영향력 측정 연구

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 모호하다.

둘째, 스타벅스 네트워크 영향력 모델은 실제 장소를 기반으로 하여 영향력을 측정하는 사례 연구이다. 스타벅스의 매장 네트워크를 활용한 소셜 네트워크 영향력 모델 연구[6]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. 그림 2는 영향력을 측정하기 위한 다이나믹 인플루엔셜 모델이다.

$$E_i(e) = e + \sum_{j \in C_i} E_j \underbrace{(f_{step} f_{year}(i, j) W(i, j))}_{\text{Propagated effect of node } j \text{ to node } i}$$

, where

$C_i = \{j | W(i, j) > 0\}$, a set of conneted nodes

$$W(i, j) = \begin{cases} 1, & \text{if distance}(i, j) < 200m \\ 1/2, & \text{if } 200m \leq \text{distance}(i, j) < 400m \\ 1/3, & \text{if } 400m \leq \text{distance}(i, j) < 600m \\ 0, & \text{if distance}(i, j) > 600m \end{cases}$$

$$f_{year}(i, j) = \begin{cases} \frac{1}{1 + e^{2(\text{year}(i) - \text{year}(j))}}, & \text{if year}(i) < \text{year}(j) \\ 0, & \text{otherwise} \end{cases}$$

$$f_{step} = 0.5$$

∴ Effect of node i = $E_i(1)$

그림 2. 스타벅스의 동적 영향력 모델

스타벅스 네트워크 영향력 모델 연구에서 제안한 내용은 다음과 같다. 한 매장 i가 가지고 있는 영향력은 i에 연결된 모든 매장 j로 전파되고, 이는 다시 j에 연결된 모든 매장들로 계속 전파된다. 따라서 임의의 매장이 가지고 있는 영향력은 재귀적으로 정의할 수 있으며, 영향력 전파 단계가 거듭될수록 감쇄효과에 의해 결국 일정한 값으로 수렴하게 된다. 이때 매장 i의 영향력은 매장이 확대되는 가중치를 고려해서 매장 j에 전파된다.

다이나믹 인플루엔셜 모델에서 나타난 식에서 볼 수 있듯이 영향력을 e(effect)라고 하고, 가중치 계수의 곱을 통해서 계산하였다. 방문자 수에 따른 가중치 계수 V(i, j)는 각 매장 방문자수를 모르는 상황이므로 1로 가정하여 나타낸다. 매장간 거리에 따른 가중치 계수는 W(i, j)로 표현하고, 기존 주변 매장 오픈 시점과의 차이에 따른 영향력 보정계수는 fyear이다. 이웃 매장과의 거리가 증가함에 따른 영향력 감쇄효과 fstep를 0.5로 가정하고, 각 노드가 가진 초기 영향력 e를 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, 본 연구에서는 이벤트가 발생한 장소의 반경 범위, 이벤트가 지속되는 시간, 사용자의 수, 메시지 빈도수 등을 주요하게 보고 이를 영향력 측정에 활용되는 액티비티 데이터로 여긴다.

3. 이벤트 영향력 측정 방안

본 장에서는 이벤트의 영향력에 대해 정의하고, 장소, 시간, 이벤트, 사용자 4가지 요소를 의미하는 Quad Relation Factors를 그림 3과 같이 제안한다.

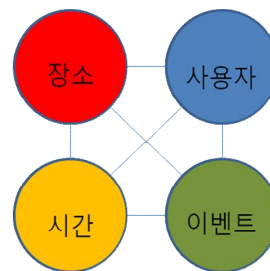


그림 3. 4가지 영향력 측정 모델 제안

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 Factors 만이 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, 질문에 가장 가까운 k 개의 답을 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 밀집 되어 있는지를 파악할 수 있는 요소들이고, 사용자들의 행동 특성이 나타날 수 있다. 이 때 영향력을 측정하기 위한 내용은 장소는 이벤트에 대한 장소의 반경, 시간은 이벤트의 지속 시간이 되며, 이벤트는 지역에 따른 발생 빈도 수, 사용자는 한 이벤트에 대한 사용자 참여 수가 수치화 될 수 있다. 이와 연관하여, 사용자 활동 그래프 (User Activity Graph) 모델링 방안을 제안한다. 사용자 활동 그래프 는 이벤트의 속성과 영향력을 반영한 그래프로서, 위도와 경도로 나타낼 수 있는 장소와 시간 축을 중심으로 3 차원 그래프를 구축하여 표현한 것이다. 아래의 그림 4 와 같이 x, y, z 축에서의 시간과 위도, 경도로 확인할 수 있다. 장소와 시간을 중심으로 동적으로 생성되고 소멸되는 이벤트 자체는 그래프 상에서 원기둥 모양과 같은 3 차원의 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.

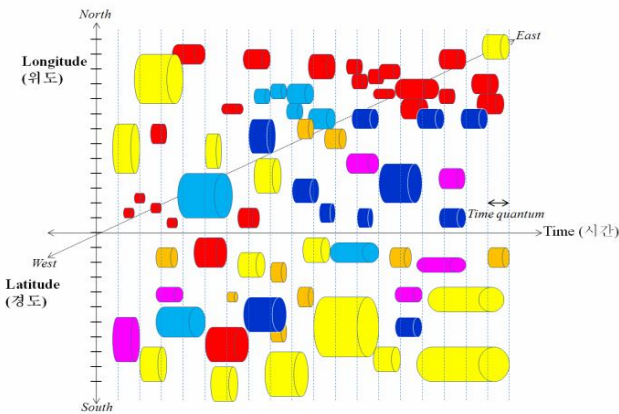


그림 4. 사용자 활동 그래프 제안

그림 4 에서 표시한, 도형의 색은 이벤트의 카테고리, 즉 어떠한 분야, 종류의 이벤트인지를 표현한다. 따라서 그래프에서 색으로 분류되는 도형의 분포를 보면, 역사적으로 과거 위치한 지역을 판단하여 해당하는 이벤트가 어느 카테고리에 속하는 지, 어느 지역에서 어떤 이벤트가 주로 발생하는 특징을 가지는 지 파악할 수 있는 자료가 될 수 있다. 또한, 또한, 원통형 도형의 크기는 바로 이벤트의 크기, 즉 영향력의 크기를 나타낸다. 이처럼 도형의 분포, 크기, 모양 등을 통하여 이벤트의 영향력을 파악할 수 있다.

4. 영향력 모델에 대한 질의 방안

같이 장소에서의 반경과, 이벤트 지속 기간 등의 내용으로 인하여 3 차원 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?” 또는 “특정 이벤트가 군집적으로 발생하는 지역은?” 또는 “작은 모임이지만 지속적인 이벤트는?” 과 같은 정성적인 질의의 답에 대해서 정량적으로 계산하여 구할 수가 있다. 사용자 활동 그래프 를 표현하기 위해서는 이벤트의 장소, 시간, 사용자에 대한 로그 데이터가 필요하다. 이 데이터에서 장소의 위치 정보인 위도, 경도의 값은 GPS 정보로부터 받아 저장되고, 이러한 로그 데이터에서 다차원 색인을 사용하여 데이터를 빠르고 효과적으로 검색할 수 있다.

5. 결론

본 논문에서는 웹 3.0 시대에 활용 가능한 소셜 네트워크 서비스에서의 이벤트 영향력 측정 방안에 대해 고찰하고, 영향력을 표현할 수 있도록 두가지 모델을 제안하였다. 첫째는 장소, 시간, 이벤트, 사용자 4 가지 요소를 제안하고, 두번째로는 사용자 활동 그래프를 제안하였다. 이를 통해, 사용자를 중심으로 발생하는 이벤트에 대한 영향력을 효과적으로 표현하고, 에 대한 질의 방안을 논의함으로 추후, 이를 기반으로 효과적인 질의 처리 기능이 가능할 수 있는 기반을 제공하게 될 것으로 기대된다.

참고문헌

[1] 박호성, 곽해운, 차미영, 문수복, "소셜 네트워크에서의 인플루엔셜 랭킹, " 정보과학회지, Vol.28, No.3, pp.24~30, 2010.
 [2] 트위터 그레이더, <http://twitter.grader.com>
 [3] 박지혜, 김보현, 이명준, 권영근, "TwitNet : 트위터 사용자들의 관계를 시각적으로 나타내는 Cytoscape 플러그인 개발, " 한국정보과학회 한국컴퓨터종합 학술발표논문집, Vol.37, No.1, 2010.
 [4] D. Tunkelang, A Twitter Analog to PageRank. "http://thenoisychannel.com/2009/01/13/a-twitteranalog-to-pagerank, " 2009.
 [5] 트위터 카운터, <http://twittercounter.com>.
 [6] Minkyong 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.