

사용자 그룹을 고려한 소셜 네트워크 상의 정보 전파에 대한 연구

황성민
전남대학교 전자컴퓨터공학부
email : sungmin1511@gmail.com

김경백
전남대학교 전자컴퓨터공학부
e-mail : kyungbaekkim@jnu.ac.kr

A Study on Information Diffusion in Social Networks Considering User Groups

Sungmin Hwang
Dept. Electronics and Computer
Engineering,
Chonnam National University

Kyungbaek Kim
Dept. Electronics and Computer
Engineering,
Chonnam National University

요 약

온라인 소셜 네트워크 상에서 메시지가 어떻게 사용자로부터 다른 사용자들에게까지 전달되는지 연구하는 분야는 현재 인터넷 인구의 증가와 소셜 네트워크 서비스의 발전에 맞물려서 흥미로운 분야가 되었다. 이를 연구함으로써, 바이럴 마케팅이나 여론 형성 등, 메시지가 최대한 영향력을 발휘하게끔 하는데 도움을 줄 수 있으므로, 메시지 전파의 효율성, 메시지의 발원지 예상 등, 다양한 연구가 지금까지 이루어졌고, 각 연구들은 소셜 네트워크에서의 각기 다른 특징들에 주목하였다. 본 연구는 그 다양한 특징들 중, 소셜네트워크가 다양한 구성원들로 이루어져있고, 그 구성원들은 비슷한 구성원끼리 묶을 수 있다는 점에서 출발하였다. 소셜 네트워크는 수많은 사용자들로 이루어져 있고, 그 사용자들의 개별적인 특징들을 구분한다는 것은 굉장히 어려운 일이다. 따라서 각 사용자들을 추상화 하는 것이 필요하고, 그 중 한 방법은 사용자들을 특징별로 묶는 일이다. 사용자들을 그룹으로 묶는 것을 고려함에 따라, 사용자 그룹들 사이의 관계와 선호도 등을 고려함으로써, 단순한 정보 전달 양상에서 벗어나 자세한 관찰을 하는 것이 가능하다. 또한, 정보 전파 양상에서 그룹의 비율이 미치는 영향에 대해서 관찰하는 것도 가능하다. 본 글에서는 메시지 전파 모델 중 하나인 Independent Cascade Model을 사용하여 그룹을 특정할 수 있는 모델을 제시하며, 각 유저들의 비율이 달라질 경우 발생하는 현상을 실험한다. 제시한 모델을 바탕으로 메시지 전파가 그룹간의 유사도에 영향을 받을 수 있는 지에 대한 앞으로의 연구 또한 제시한다.

1. 서 론

소셜 네트워크상에서의 정보 확산은 현재 많은 관심을 받는 분야 중 하나이다. 전 세계 인터넷 인구의 증가와, Twitter, Facebook, Google+, 등의 소셜 네트워크 서비스의 증가로 소셜 네트워크 상의 정보 확산은 그 전보다 속도와 규모가 훨씬 커지게 되었다. 한 명의 유저에게서 발생한 메시지가 전 세계를 누빌 수 있게 된 것이다. 이런 소셜 네트워크를 이용하여 여론 형성이나, 바이럴 마케팅이 꾸준히 시도되고 있고, 실제로 성과를 거두고 있다. 이에 따라 많은 학자들이 이러한 메시지 전파에 대해 연구하였고, 메시지 전파 모델 [1], 메시지 전파 예측 방법 [6], 메시지 영향력 최대화 [2], 그리고 앞서 언급된 마케팅에 관련된 연구 [4] 등이 이루어졌다.

소셜 네트워크의 구성원 각각은 각기 다른 구분되는 특징들을 지닌다. 나이, 직업 등과 같이 단순한 특징들로도 구분이 될 수도 있고, 좀 더 복잡하게 선호하는 사람 유형, 혹은 선호하는 정보 전달 방법 등도 그 특징 구분의 한 방법이 될 수가 있다. 이러한 특징들을 고려하면, 정보 확산에 대한 예측을 좀 더 자세히 할 수 있지만, 문제는

소셜 네트워크의 규모가 크고, 개개인의 특징을 구분하여 정보 확산의 규모나 마케팅 결과 예측을 하는 것이 불가능하다는 것이다. 따라서 각 구성원들을 어떤 특징들에 따라 구분 지어 어느 정도 추상화를 하는 것이 필요하다. 예를 들어 같은 직업을 갖는 사람들끼리는 서로 소통하는 방법이나, 사고방식 등이 비슷하며, 그들 사이에서의 정보의 신뢰도 또한 높을 수가 있고, 그에 따라 정보 확산도 빠르게, 높은 확률로 이루어질 가능성이 존재한다. 반대로 고등학생과 50대 직장인 그룹 사이에서의 소통은, 서로 간의 사고방식 차이, 사용하는 언어의 차이, 의사소통 방법 및 도구 등의 차이로 인해 정보의 확산이 둔화되거나, 아예 이루어지지 않을 수가 있다. 성별, 성격, 나이, 직업, 등 다수의 요소가 단독으로 혹은 결합되어 정보 전파에 영향을 미칠 수가 있는 것이다. 따라서 이들을 비슷한 경향의 사람들로 묶게 되면, 소셜 네트워크 안에서의 정보 전달 양상을 좀 더 자세히 관찰할 수 있다.

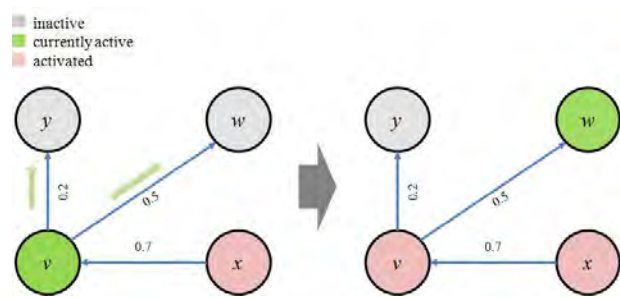
그룹을 고려하게 되면, 정보전달 양상에서 그룹의 비율이 미치는 영향에 대해서도 생각해 볼 수 있다. 소셜 네트워크들이나 커뮤니티를 살펴보면, 각각 특징적인 그

를 비율을 보이는 것을 알 수 있으며, 이러한 특징적 비율은 각각 소셜 네트워크에서의 정보 전달 양상을 좌우한다.

이 연구에서는 앞서 언급된 사용자 그룹과 특징을 고려한 정보 확산 모델을 제시하고, 그 모델에 하나의 그룹 분류 기준을 적용해 메시지 전파 양상을 실험한다. 또한 사용자 그룹 간의 유사성의 정도를 고려하여 정보가 전파되는 모델도 제시한다.

2. 사용자 그룹과 특성을 고려한 Independent Cascade Model

2.1 Independent Cascade Model

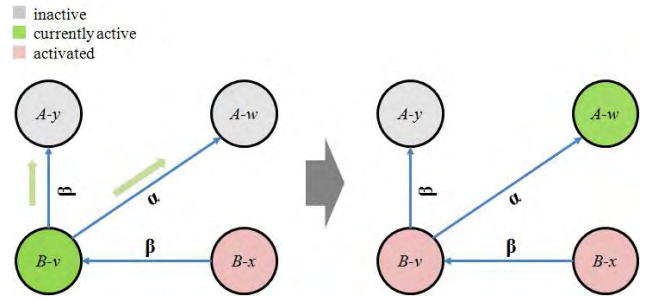


(그림 1) 확률을 기반으로 한 IC모델의 정보 전파 진행 과정

소셜 네트워크 상에서의 정보 확산을 표현하기 위해 Linear Threshold (LT) 모델, Independent Cascade (IC) 모델 [1] 등, 여러 모델들이 지금까지 제시되었다. 그 중 IC 모델은 확률을 바탕으로 한 모델로써, 그룹들 사이에서의 선호도에 따른 전달 성공과 실패를 표현하기에 적합하므로, 이 모델을 바탕으로 그룹을 고려하는 모델이 구성되었다. 그림 1에서와 같이 각 노드들은 활성화, 혹은 비활성화 상태로 분류되며, 한번 노드가 활성화가 되면, 그 노드는 단 한번, 자신의 이웃 노드들을 활성화하려고 시도하게 된다. 노드 v는 x에 의해 활성화 된 직후 그 이웃노드 y와 w를 활성화 하려고 하게 되며 그 성공률은 각 연결의 값인 0.2, 0.5를 각각 갖게 된다. 한번 활성화된 노드는 다시 재 활성화될 수 없다.

2.2 선호하는 정보 전달 매체에 따라 분류한 그룹

그룹을 분류하는 기준에는 여러 가지가 있을 수 있다. 앞서 언급되었던 예와 같이, 직업, 나이, 혹은 성별이 될 수도 있고, 사용자들의 특정 선호도도 예가 될 수 있다. 우리는 사용자들을 정보 전달 매체 선호도로 구분하였다. 다른 특성들도 모델의 사용자 그룹 구분에 적용될 수 있지만 매체 선호도로 구분한 것은 정보 확산에 중요한 속성들을 매체가 가지고 있기 때문이다. 이 모델에서 사용자는 각 매체 선호도를 갖게 되며, 그 선호도에 따라서 어떤 매체를 정보 전달시 사용하고, 어떤 매체를 정보를 받아들일 때 신뢰할지 결정되게 된다.



(그림 2) 사용자 그룹의 정보전달 방식 선호도 차이때 따른 정보 전달 양상

그림 2에서와 같이 이 모델은 IC 모델을 확장하여 구성되었다. 그룹 A는 매체 α 를 다른 매체들에 비해서 신뢰하고 선호하며, 그룹 B는 매체 β 를 신뢰하고 선호한다. 그룹 B에 속한 v가 매체들을 통해 그룹 A에 속한 y와 w를 활성화 하려고 시도하면, y와 w는 그룹 A의 매체 신뢰도를 기준으로 그 정보를 받아들일지 말지를 결정하게 된다. 결국 매체 b를 통해 활성화 하려고 한 y는 높은 확률로 실패하게 되고 매체 a를 통해 활성화 하려고 한 w는 높은 확률로 성공하게 된다.

2.3 그룹 유사성

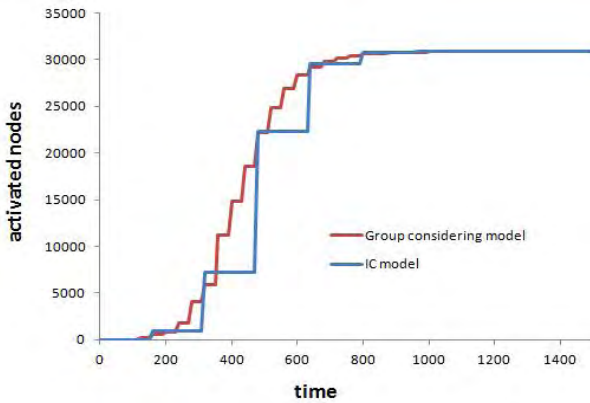
현재 몇몇 연구들이 상업적, 혹은 정치적 목적으로 소셜 네트워크에서의 효율적 정보 확산에 집중하고 있으며, 이 연구들은 어떤 사용자들에게 정보 전달을 해야 효과를 극대화 시킬 수 있을지에 대한 연구도 포함한다 [1][2]. 이러한 효율적 정보 확산을 위해 소셜 네트워크에서의 그룹과 그 비율을 고려한다면 훨씬 정밀한 결과를 얻을 수 있을 것으로 기대된다.

하지만 현재 제시된 모델을 바탕으로 효율성 극대화를 고려하려면 여러 제약이 존재한다. 우선 그룹들 서로간의 매체 선호도와 그 신뢰도 등의 여러 요소들을 고려해야 하며 이들은 효율성 극대화를 계산할 경우, 복잡도를 증가시킨다. 따라서 그룹들 간의 유사성을 계산하여 수치화시키고 단순화시키면 훨씬 단순한 계산을 찾을 수 있을 것이라고 기대된다.

3. 실험 결과

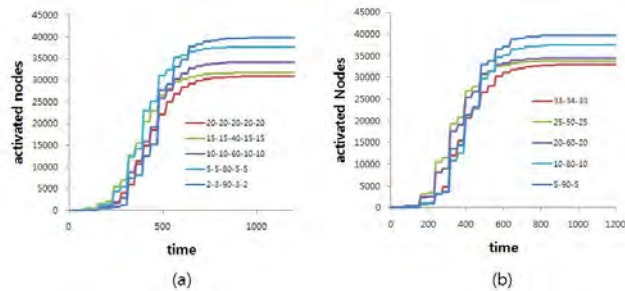
제시된 모델을 이용한 온라인 소셜 네트워크에서의 정보 전파를 측정하기 위해 유저집단들과 선호하는 정보전달 방식 및 각 그룹들의 비율을 고려할 수 있는 시뮬레이터를 구성하였다. 총 5개의 그룹과 5개의 정보전달 방식을 설정하였으며, 소셜 네트워크 그래프는 실제 2008년에 수집된 Slashdot [7] 의 7만개 이상 노드를 갖는 소셜 네트워크 그래프 데이터를 사용하여 구성하였다.

첫 번째로 사용자 그룹을 나누는 것이 중요하다는 것을 보여주기 위해서 평균적인 그룹 하나와 하나의 선호도만 존재하는 IC모델과 제시된 모델을 비교하였다.



(그림 3) 기본 IC 모델과 그룹과 그 특성을 고려한 모델의 정보 전파 양상

그림 3을 보면, 하나의 그룹만 존재하는 기본 모델에 비해서, 사용자 그룹이 다양해 여러 선호도 및 신뢰도가 존재하는 모델은 훨씬 자세한 시간별 정보 전파를 보여주고 있다. 이는 훨씬 정확한 전파 양상을 관찰할 수 있게 한다.



(그림 4) 그룹 비율에 따른 정보 전파 양상

다음으로 그룹 비율을 고려했을 때 나타나는 전파 양상을 실험하여 그룹을 고려하고 그 비율을 고려함이 정보 전파의 범위에 얼마나 영향을 미치는가를 확인하였다. 그림 4는 시간 당 활성화 된, 즉 정보가 전파된 사용자 수를 그래프로 나타낸 것이며, 각 그룹의 비율을 다르게 하여 실험한 결과를 보여준다. 4(a)의 경우 5개의 그룹들 중 한 그룹의 비율이 높아질 경우를 보여주며, 특정 그룹의 비율이 40% 이하일 때는 비슷한 전파 범위를 보여주지만 그 이상의 비율일 경우, 전파 범위가 눈에 띄게 증가하는 것을 보여준다. 4(b)의 경우, 3개의 그룹들 중, 한 그룹의 비율이 증가할 때의 전파 양상을 보여준다. 이 결과를 살펴 보면 정보의 전파 범위는 4(a)에서의 다섯 개의 그룹으로 실험한 결과와 그리 차이가 나지 않음을 알 수 있다. 정보 전파 범위는 그룹의 비율이 어느 정도 수준 이상으로 올라가기 전까지는 전파에 영향을 크게 미치지 않으므로, 그룹들 중 최대의 비율을 갖는 그룹들의 비율을 따라감을 확인할 수 있었다.

실험 결과는 사용자를 그룹으로 나누는 것이 정보 확산을 파악하고 예측하는데 있어서 어떻게 활용될 수 있는가를 보여준다. 하나의 그룹의 비율이 어느 한계치 이상 높

아질 경우, 정보가 퍼질 수 있는 정도는 확실하게 달라지며, 이에 따라 서로 비율이 다른 커뮤니티나 소셜 네트워크 안의 정보 확산을 예측하는데 유용하게 이용될 수 있을 것이라고 기대된다.

4. 결론

소셜 네트워크에서의 사용자 그룹을 고려함에 따라, 기존의 단순화된 사용자에서 벗어나 적절한 다양성을 부여할 수 있게 되었고, 사용자 그룹의 특성에 따라 자세한 정보 전파 양상을 확인할 수 있었다. 사용자 그룹들이 소셜 네트워크에서 차지하는 비율에 따라, 정보 전파의 효율 및 확산 범위에 영향을 얼마나 받을 수 있는지 또한 확인할 수 있었으며, 이 정보는, 새로운 사용자가 유입되거나, 혹은 사용자가 이탈하는 등, 그룹의 비율에 변화가 있을 때, 좀 더 효율적으로 정보 전파를 예상하는 곳에 사용될 수 있을 것이라 기대된다. 사용자 그룹에 대한 고려가, 정보 전달 효율성 증대 연구와 연계해서 이루어질 수 있게 그룹간의 관계와 유사성을 정리해서 수치화, 단순화 시키는 방법에 대한 연구가 뒤따를 예정이다.

Acknowledgements

이 논문은 2014년 정부(교육부 또는 미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2014R1A1A1007734).

참고문헌

- [1] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," Proc. ACM SIGKDD 2003, pp. 137-146.
- [2] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," Proc. ACM SIGKDD June. 2009, pp. 199-208.
- [3] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyone's an influencer: quantifying influence on twitter," Proc. ACM international conference on Web search and data mining, 2011, pp. 65-74.
- [4] S. Bhagat, A. Goyal, and L. V.S. Lakshmanan, "Maximizing product adoption in social networks," in Proc. the fifth ACM international conference on Web search and data mining, February. 2012, pp. 603-612.
- [5] T. Lappas, E. Terzi, D. Gunopulos, and H. Mannila, "Finding effectors in social networks," Proc. the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, July. 2010, pp. 1059-1068.
- [6] A. Najar, L. Denoyer, and Patrick Gallinari, "Predicting information diffusion on social networks with partial knowledge," Proc. 21st international conference companion on World Wide Web, April. 2012, pp. 1197-1204.
- [7] Slashdot. [Online]. Available from <http://Slashdot.org>