

소셜 네트워크에서의 우선순위 인지 영향력 최대화

황성민, 김경백

전남대학교 전자컴퓨터공학과

e-mail : sungmin1511@gmail.com, kyungbaekkim@jnu.ac.kr

Priority Aware Influence Maximization in Social Networks

Sungmin Hwang, Kyungbaek Kim

Department of Electorinics and Computer Engineering,
Chonnam National University

요 약

소셜 네트워크에서의 정보 전파 과정을 살펴보면, 메시지 전파 범위를 예측할 때, 여러 상황 및 조건들을 고려해야 하는 것을 알 수 있다. 다양한 속성들에 초점을 맞춘 연구들이 진행되고 있으며, 본 글에서는 여러 조건들 중, 현실에서의 재난 경보와 연결 지어 볼 수 있는, 시간제한이 있는 메시지 전파 과정에 초점을 맞추었다. 긴급한 상황에서의 사람들의 행동을 고려하기 위해서, 시간제한 정보 전파 모델에 추가로 우선순위를 추가하였으며, 메시지 전달의 우선순위 그룹과 정보 전달 채널의 우선순위를 고려할 수 있으며, 이에 따라 발생하는 순차적인 메시지 전달 또한 고려할 수 있다. 이 글에서는 이 모델이 어떻게 작동하고, 어디에 적용될 수 있으며, 이 모델에서의 메시지 전파 극대화 방법에 대해서 설명한다.

1. 서 론

정보가 어떻게 소셜 네트워크에서 퍼져나가고, 어떻게 이 전파 범위를 최대화 할 것인지는 현실 세계에서의 현상과 연결 지어 흥미로운 연구 분야이다. 소셜 네트워크에서는 한 명의 사람에게서 발생한 메시지가 전체 네트워크 전체에 퍼질 수 있다. 이 같은 소셜 네트워크에서의 메시지 전파는 메시지를 효율적으로 매우 큰 수의 사람들에게 전달하는데 유용하게 사용될 수 있다. 바이럴 마케팅 및 여러 전략에 사용되면서, 몇몇 연구에는 제한된 자원 하에서 어떻게 영향력을 최대화 시킬지에 대해서 연구하였다. 이 영향력 최대화는 처음 정리[1] 된 이후로 많은 연구에서 각각 다른 모델에 적용하여 연구가 되고 있다. 각 연구는 메시지 전파에서 효율적인 영향력 최대화 계산[2], 비용과 효율에 초점을 맞춘 영향력 최대화[3], 큰 규모에서도 가능한 계산 방법[4], 주어진 시간에서의 영향력 최대화[5] 등, 여러 속성에 초점을 맞추고 있으며, 모든 모델은 실제 현상과 연관되어 있다. 이 연구에서는 재난 경보 및 대피 경보와 연결 지어 긴급한 상황에서의 메시지 전파에 초점을 맞추었다.

사람들에게 재난 상황을 알리고 대피 경보를 내리는 긴급한 상황에서는 메시지 전파를 예측하기 위해, 어떻게 사람이 경보에 반응하는지를 고려해야만 한다. 보통의 상황에서라면 사람들은 연결된 사람들에게 메시지로 연락을 하거나, 소셜 네트워크 서비스(SNS)를 사용하여 포스트를 하는 등의 방법을 통해 메시지 전달을 할 것이다. 하지만 긴박한 상황에서는 SNS 포스트 대신 상대방이 메시지

를 받았는지 바로 확인을 할 수 있는, 상호 메시지 전달이 가능한 전화를 할 것이다. 이 경우, 메시지 전달 채널로서의 전화는 통화시간 및 한 번에 연결할 수 있는 사람들 등의 중요한 속성을 갖는다. 전화는 보통 오직 하나의 연결만이 가능하므로, 사람들은 긴급한 상황에서 자신들에게 중요한 사람들에게 먼저 연락을 시도할 것이다. 이 시나리오에서, 채널의 우선순위와 메시지 전달 대상의 우선순위는 메시지 전달에서의 순서를 결정하는 요인이 된다. 이러한 시나리오를 설명하기 위해 우선순위 고려 시간제한 메시지 전달 모델(Priority aware time-constraint model: PATC)을 제안한다.

다른 모델에서는 한 노드에서 다른 대상 노드들로의 전달이 한 번에 이루어지게 되는데, 이 PATC에서는 우선순위 및 우선순위로 인한 순서 발생을 고려해야 하는 특수한 상황 하에서의 메시지 전달에 주목하였다. 통신 채널의 우선순위는 긴급한 상황에서의 선호 순위를 따르며, 주로 상호통신이 가능하여 상황을 확인할 수 있고, 빠른 응답시간을 갖는 채널들이 선호된다. 하지만 이러한 채널들은 메시지 전달 시간이 길며, 한 번에 한번만 전달이 가능한 경우가 대부분이다. 이러한 채널에서의 우선순위는 앞서 언급했던 것처럼 메시지 전달 목표의 우선순위를 갖게 한다. 사람은 가족, 친구, 직장 동료 등, 여러 계층의 그룹에 있으며, 각각의 우선순위는 서로 다르다. 높은 우선순위의 그룹에 속해있는 대상에게는 우선순위가 높은 메시지 전달 방법을 사용해 전달하려고 할 것이며, 우선순위가 낮은 그룹에 속해있는 대상에게는 상호연결이 가능하지

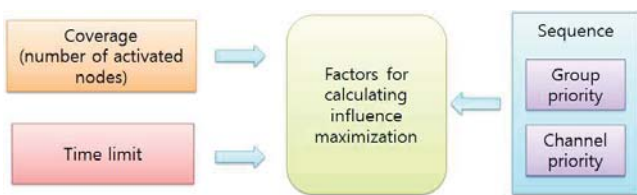
않은 일방적인 메시지이나, 넓은 전파 범위를 가진 채널이 선호될 것이다. 예를 들어 전 지역에 발동된 긴급 대피 경보는 사람들이 자신들의 가족에게 먼저 전화해 상황을 알리게 할 것이며, 그 다음 여유가 생기면 다른 경로를 통해 동료들에게 상황을 알릴 것이다. 이러한 순서의 발생은, 짧은 시간 안에서의 영향력 전파 최대화에 영향을 미칠 수 있는데, 이는 평범한 상황에서의 메시지 전파를 최대화시킬 수 있는 사람이 긴급한 상황에서는 오히려 메시지 전파를 지체시킬 수 있기 때문이다.

2장에서는 관련연구에서의 메시지 전파 모델 및 영향력 최대화 문제에 대해서 소개하며, 3장에서는 PATC 모델에서의 메시지 전파를, 4장에서는 이 모델 하에서의 영향력 최대화 문제 해결을 제안한다.

2. 관련 연구

소셜 네트워크에서의 영향력 최대화 문제는, 메시지 전파를 최대화 하는 초기 노드 셋 S를 찾는 문제이다. 제한된 크기 k의 S를 찾는 문제로써, 다양한 계산방법 및 모델이 제시되고 있다. 영향력 최대화 문제는 처음 정리[1]된 이후로, 많은 연구가 이루어졌다. 처음 연구는 이를 계산하기 위하여, KK-Greedy 알고리즘을 제안하였으며, 이후 이를 최적화한 좀 더 빠른 Greedy 알고리즘인 Cost-Effective Lazy Forward(CELF) 알고리즘이 제안되었다[3]. 이에 더해 효율적인 계산을 위해, Degree Discount 라는 휴리스틱 방법이 사용되기도 하였으며[2], CELF와 결합하여 Mixed Greedy 알고리즘이 제안되었다. 최근에는 시간제한 하에서의 메시지 전달 모델 및 영향력을 최대화 하는 연구가 발표되었으며[5], 시간제한 T안에 가능한 경로들을 계산하여 메시지 전파 최대화 문제에 새로운 조건을 추가하였다.

3. Priority Aware Time Constraint Model



(그림 1) Factors in influence maximization under PATC model

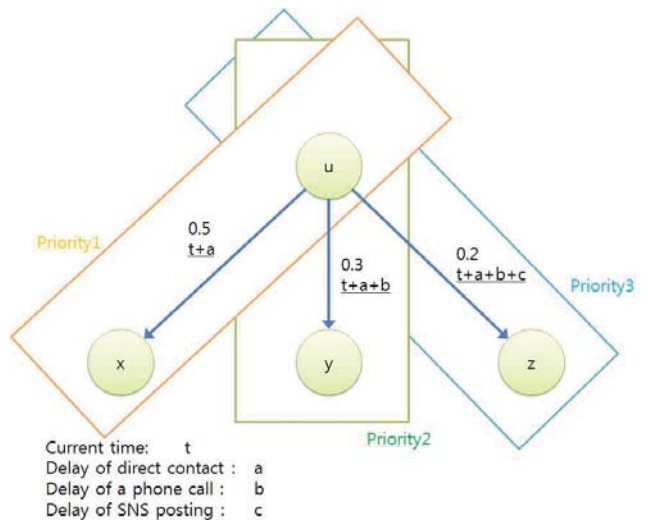
긴급한 상황에서의 메시지 전파 과정을 모델링하기 위해, 그림 1과 같이 기존의 모델에서 고려하는 시간제한 및 영향력 범위에 더해 메시지 전파 순서를 고려하였다. 소셜 네트워크는 사람들로 이루어져 있으며, 각 노드들은 서로 다른 관계를 맺고 있다. 이러한 관계의 차이가 존재하며, 이로 인해 긴급한 상황에서의 메시지 전파 우선순위가 생기게 된다. 또한, 여러 메시지 전달 채널들이 존재하며, 각 대상의 우선순위마다 선호되는 채널의 선호도 및 전달 속

도가 달라진다. 그룹의 우선순위와 채널의 우선순위 둘 다를 고려함으로써, 시간제한 모델에 순서의 개념을 추가하였다.

3.1 채널의 우선순위

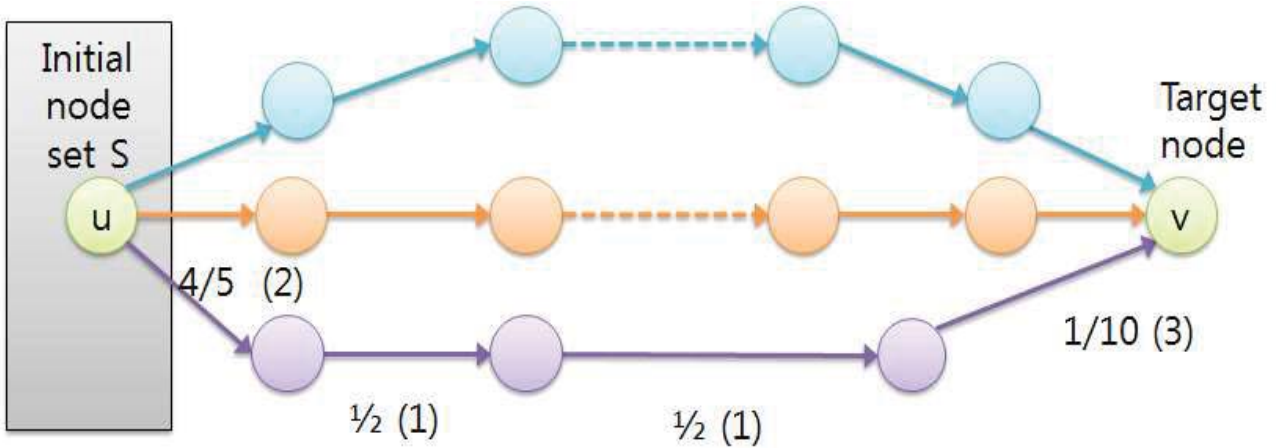
소셜 네트워크에서는 한 쌍의 노드 사이에 존재하는 링크들은 그들이 연결되었고, 메시지 전달 확률이 있다는 것을 보여준다. 메시지를 전달하기 위해서, 전화, 이메일, SNS, 등 여러 방법을 사용하게 되며, 각 채널은 고유한 성격을 가지고 있다. 예를 들자면 전화는 빠른 응답시간 및 높은 반응률을 가진다. 위급한 상황에서는 사람들은 상대방이 메시지를 전달 받았는지 확인할 수 있으며, 앞에서 언급된 장점을 가진 상호 통신 방식을 사용하게 된다. 장점들이 있는 반면, 한 번에 유지할 수 있는 연결 수에 한계가 있으므로, 전화통화가 발생하게 되면, 다른 노드들의 메시지 전달은 그만큼 지연되게 된다.

3.2 그룹의 우선순위



(그림 2) Delays from sequence in PATC model

채널의 우선순위와, 초기에 가능한 연락이 한정됨에 따라, 어떤 사람에게 먼저 메시지를 전송하는지를 고려해야 한다. 사람들은 서로 각각 다른 여러 그룹에 속해있으며, 이들 그룹은 사람들에게 각기 다른 우선순위를 갖게 된다. 가족과 같이 단단하게 연결된 그룹에는 높은 우선순위를 갖게 되며, 단순히 알고 지내는 사람들에게는 낮은 우선순위를 갖게 된다. 따라서 제한된 시간 하에 연락을 해야 하는 경우에는, 높은 우선순위를 가진 순서대로 연락을 취하게 된다. PATC 모델에서는 이를 우선순위에 따라 여러 그룹 레이어로 분리하고 이를 기준으로 메시지 전달을 표현한다. 그림 2에서는 이들 우선순위 및 채널로 인해 발생하는 순서 및 딜레이 변화에 대해 보여주며, 메시지 전달이 발생할 때, 각 경우에 대한 변화를 보여준다. 노드 u가 우선순위 1 계층 그룹에 속해 있는 노드 x로 메시지를 처음으로 전달하게 되며, 이로 인해 발생한 딜레이는 계층 2 그룹에 속해있는 노드 y로의 전달 시간에 영향을 미치



(그림 3) Calculating influence maximization using paths

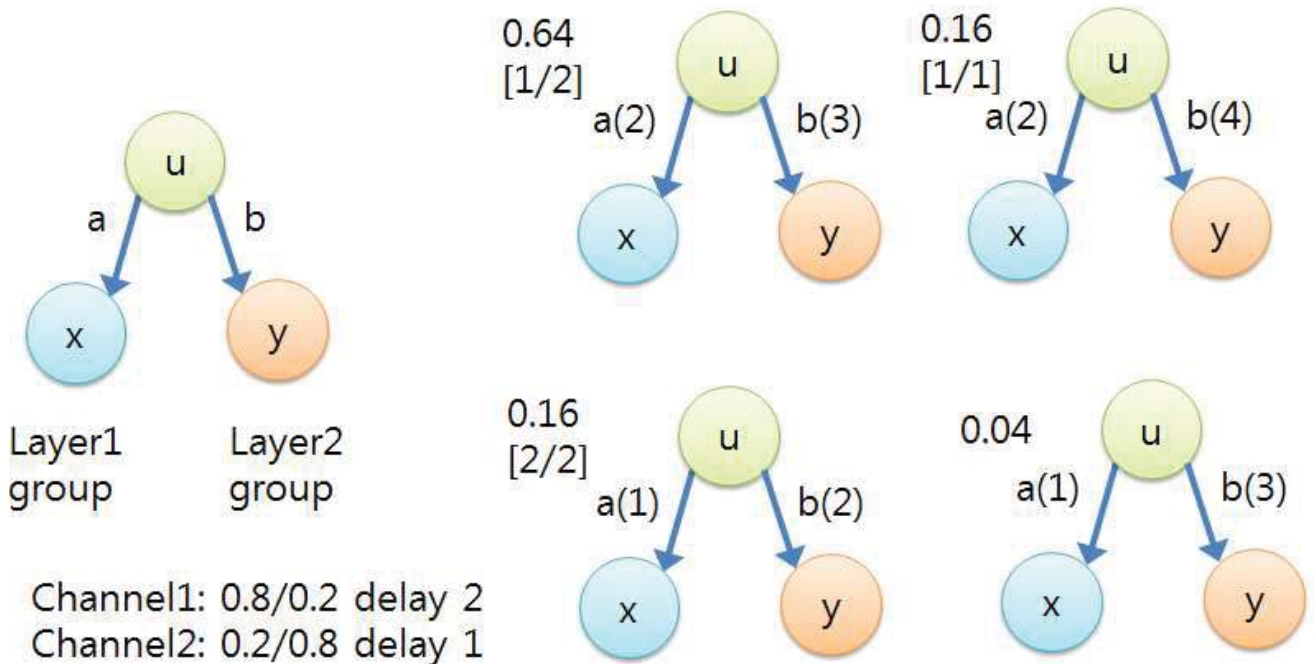
는 것처럼, 앞 단계에서 발생하는 딜레이가 뒷 단계에서 발생하는 딜레이에 영향을 미치게 된다. 이와 같이 단계적으로 발생하는 딜레이는 전과 시간이 짧을 경우, 즉 짧은 시간을 잘게 쪼개서 볼 경우, 큰 영향을 미칠 수 있다.

터 시작하게 되며, 하나씩 영향력을 최대화 시키는 노드를 추가하여 S의 원소의 수가 k에 도달할 때 까지 채운 뒤, 이에 대한 영향력을 확인한다. 모든 노드에 대해 이를 수행하게 되며 이는 매우 많은 수의 계산을 필요로 한다.

4. PAIC모델에서의 영향력 최대화 문제

TC 모델과 같이, PATC모델에서의 영향력 계산은 시작 노드 셋 S로부터 각 노드들까지의 제한시간 내에 가능한 경로들의 확률 합으로 구할 수 있다. 그림 3에서와 같이, 시작 노드 u 로부터 S에 들어있지 않은 모든 각각 노드들까지의 경로를 구한 뒤, 모든 경로의 확률 합을 기준으로 메시지 전파 영향력을 확인한다. 처음 S 하나의 노드로부

그림 4는 이 모델에서 발생한 모든 경우에 대해 가능한 path를 구하는 과정을 보여준다. Channel1이 layer1에서 선호도 0.8, layer2에서 선호도 0.2를 가지며 channel2가 layer1에서 선호도 0.2, layer2에서 선호도 0.8을 가지게 되면, 가능한 경로는 다음과 같이 4개로 분리할 수 있다. 이의 모든 경로에 대해 각각 존재할 수 있는 확률을 바탕으로 경로를 설정하게 된다. 여기에서의 문제는 가능한 모든



(그림 4) Separating paths from original graph

경우에 대한 경로를 고려할 경우, 같은 layer에 존재하는 노드 수가 많을수록, 여러 딜레이 조합을 고려해야 한다는 것이고, 이 모든 경우가 다른 에지의 딜레이에도 영향을 미치게 된다는 것이다. 복잡도가 노드가 가진 이웃 노드의 수가 커지면서 증가하게 되는데, 모든 노드의 순서와 조합이 되므로, 에지의 수가 크면 클수록 급격히 시간복잡도가 증가하게 된다.

앞의 문제를 해결하기 위해, 휴리스틱 방법을 제안한다. 동일한 layer에 속한 노드들에서 발생할 수 있는 딜레이 조합의 경우, 각 노드들이 각각 순서에서 등장할 확률은 모두 같으며, 이를 이용하여, 각 시간 유닛마다의 에지의 딜레이를 가능한 가장 긴 딜레이의 평균으로 계산한다. 하지만 이 경우, 시간 유닛이 정수로 떨어지지 않고, 각각의 노드에 대해서 서로 다른 시간 유닛을 가지게 되므로, 이를 시간 유닛의 한계치 설정에 따라 정한다. 이 방법을 통해, 시뮬레이션 기반 모델에 비교해 빨리 계산이 가능하다.

5. 결론

제안된 모델을 사용함으로써, 특정한 상황에서의 정보 전파 양상에 대해 모델링이 가능하였고, 긴박한 환경에서의 메시지 전달을 최대화 할 때, 짧은 시간 내에서 정보 전달을 관찰할 경우, 순서까지 고려하는 것에 대한 영향을 확인할 수 있었다. 하지만 이로 인해 계산이 더욱 복잡해지므로, 초기 단계의 간단한 휴리스틱 방법을 제안하였다. 낮은 확률의 경로 제거 및 낮은 확률의 딜레이 조합 제거 등을 다양한 값으로 실험하면 보다 빠른 휴리스틱 방법을 찾을 수 있을 것으로 기대한다.

Acknowledgements

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

참고문헌

- [1] Kempe D, Kleinberg J, Tardos É, “Maximizing the spread of influence through a social network”, Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. 2003. pp. 137-146.
- [2] Chen W, Wang Y, Yang S, “Efficient influence maximization in social networks”, Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. 2009. pp. 199-208.
- [3] Leskovec J, Krause, A, Guestrin C, Faloutsos C, VanBriesen J, Glance N, “Cost-effective outbreak detection in networks”, Proceedings of the 13th ACM

SIGKDD international conference on Knowledge discovery and data mining. 2007. pp.420-429.

[4] Wang C, Chen W, Wang Y, “Scalable influence maximization for independent cascade model in large-scale social networks”, Data Mining and Knowledge Discovery. 2012. pp. 545-576.

[5] Liu B, Cong G, Xu D, Zeng Y, “Time constrained influence maximization in social networks”, Proceedings of the 2012 IEEE 12th International conference on Data Mining. 2012. pp.439-448.