

소셜 네트워크를 위한 확산기반 영향력 극대화 기법

응웬트리하이*, 유 명 식^o

Diffusion-Based Influence Maximization Method for Social Network

Tri-Hai Nguyen*, Myungsik Yoo^o

요 약

정보 확산 극대화 문제는 소셜 네트워크에서 정보 확산을 최대화 할 수 있는 Seed 노드 군을 설정하는 것이다. 기존의 Greedy 알고리즘은 최적에 근접한 해를 제시하였으나 높은 계산량의 문제가 있다. 몇몇 Heuristic 알고리즘들이 계산량 감소를 목표로 제안되었으나 정보 확산 성능 측면에서 한계점이 있다. 본 논문에서는 General Degree Discount 알고리즘을 제안하고, 제안된 알고리즘이 계산량 측면 및 정보 확산 성능 측면에서 기존 Heuristic 알고리즘 대비 우수한 성능을 보임을 입증하고자 한다.

Key Words : Social Network, Influence, Diffusion, Greedy, Heuristic

ABSTRACT

Influence maximization problem is to select seed node set, which maximizes information spread in social networks. Greedy algorithm shows an optimum solution, but has a high computational cost. A few heuristic algorithms were proposed to reduce the complexity, but their performance in influence maximization is limited. In this paper, we propose general degree discount algorithm, and show that it has better performance while keeping complexity low.

I. 서 론

영향력 극대화 문제는 소셜 네트워크 분석^[1] 분야에서 중요한 연구주제로, 정보 확산 극대화를 위하여 정보 확산에 기여도가 큰 핵심노드들을 선정하는 문제이다.

Kempe^[2]는 영향력 극대화 문제가 NP-hard임을 보였고, Greedy 알고리즘을 제시하였다. Greedy 알고리즘은 정보 확산 극대화의 정확한 해를 구하기 위하여 MonteCarlo 방법을 사용하였으나 높은 복잡도로 인해 수행속도가 느리고 확장성에 문제가 있다. Greedy 알고리즘의 단점을 보완하고자 Heuristic 알고리즘들이 제안되었다^[3,4]. 이 중에 Degree Discount 방법^[4]은 Chen에 의해 제안되었다. 알고리즘의 복잡도를 줄이기 위하여 해당 노드와 이웃노드들과 연결된 링크의 Out-Degree만 고려하였다. 알고리즘 복잡도를 감소시킨 장점은 있으나, 균일 확산확률에 기반한 Independent Cascade (IC)에 국한되어 있고, Undirected 그래프에만 적용되는 단점이 있다.

본 논문에서는 기존 Heuristic 방식의 단점을 극복하고 정보 확산 성능을 향상시키기 위하여 General Degree Discount 방법을 제안한다. 제안 방식은 노드의 Out-Degree, 정보 확산 확률 및 기 선정된 Seed 노드와의 영향력 등을 고려하였으며, Undirected 및 Directed 소셜 네트워크 그래프에 적용가능하다.

II. General Degree Discount Heuristic

2.1 정보 확산 모델

소셜 네트워크는 그래프 $G = (V, E)$ 로 모델링되며, 이때 $V = \{v_1, v_2, \dots, v_n\}$ 는 모든 노드들을 $E = \{\langle v_i, v_j \rangle | v_i, v_j \in V, 1 \leq i, j \leq n\}$ 는 노드들을 연결하는 모든 에지 (Edge)를 나타낸다. 각 노드는 Active 또는 Inactive 상태를 갖을수 있으며, 노드의 상태는 조건에 따라 Inactive에서 Active로 천이되며, 반대는 성립하지 않는다. Active 노드는 이웃해있는 Inactive 노드들을 Active 상태로 천이시킬 수 있다. S를 초기 Active 노드 군으로 정의할 때 영향력 극대화 문제는 주어진 소셜 네트워크 $G = (V, E)$ 에

* 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학ICT연구센터육성 지원사업의 연구결과로 수행되었음 (IITP-2016-H8501-16-1008)

• First Author : Soongsil University, School of Electronic Engineering, nguyentrihai93@gmail.com, 정회원

o Corresponding Author : Soongsil University, School of Electronic Engineering, myoo@ssu.ac.kr, 종신회원

논문번호 : KICS2016-10-304, Received October 11, 2016; Revised October 14, 2016; Accepted October 14, 2016

서 주어진 확산 모델을 가정하여 정보 확산을 최대화 하는 k 개의 Active 노드 군을 찾는 것이다.

본 논문에서는 정보 확산 모델로 IC 모델을 고려하였다. IC 모델은 정보 확산 확률 $p(u, v) \in [0, 1]$ 과 초기 Seed 노드 군 $S \subseteq V$ 가 주어졌을 때, 매 반복 주기 t 마다 Seed 노드 군에 속한 노드들은 정보 확산 확률에 따라 이웃 Inactive 노드들을 Active 상태로 천이 시킨다. 이때 각 Active 노드들은 단 한 번의 정보 확산 기회를 갖는다. 정보 확산 과정은 더 이상의 Active 노드로의 천이가 없을 때 마치게 된다.

2.2 제안 방식

본 논문에서 제안하는 General Degree Discount (GDD) 방식은 모든 노드들에 대해서 정보 확산 지수 gdd_v 를 계산하고, 정보 확산 지수가 높은 노드들을 Seed 노드 군으로 포함하여 소셜네트워크에서 정보 확산 극대화를 목적으로 한다. GDD 방식은 Undirected 및 Directed 그래프 모두 적용이 가능하여 다양한 소셜네트워크에 적용이 가능하다.

Directed 그래프를 고려한 GDD 방식에서 gdd_v 는 다음과 같이 계산된다. 우선 정보 확산 확률을 $p(u, v) \in [0, 1]$ 로, S 를 Seed 노드군으로, $In(v) = \{v\} \cup \{u \in V | (u, v) \in E\}$ 을

Inbound 이웃노드로,

$Out(v) = \{v\} \cup \{w \in V | (v, w) \in E\}$ 을

Outbound 이웃노드로 표기할 때,

$$gdd_v = d_v \times \prod_{u \in In(v) \cap S} (1 - p(u, v)) \times \sum_{w \in Out(v) \setminus S} p(v, w) \quad (1)$$

이때 d_v 는 노드 v 의 이웃노드의 수이고, $\prod_{u \in In(v) \cap S} (1 - p(u, v))$ 는 노드 v 가 이웃 Inbound 노드 u 로부터 영향을 받지 않을 확률이고, $\sum_{w \in Out(v) \setminus S} p(v, w)$ 는 노드 v 가 Outbound 이웃노드 w 에 영향을 줄 확률이다. 초기 Seed 노드 군이 없을 경우 gdd_v 는 다음과 같이 계산된다.

$$gdd_v = d_v \times \sum_{w \in Out(v) \setminus S} p(v, w) \quad (2)$$

Undirected 그래프를 고려한 GDD 방식에서 gdd_v 는 다음과 같이 계산된다.

$$gdd_v = d_v \times \prod_{u \in N(v) \cap S} (1 - p(u, v)) \times \sum_{w \in N(v) \setminus S} p(v, w) \quad (3)$$

초기 Seed 노드 군이 없을 경우 gdd_v 는 다음과 같이 계산된다.

$$gdd_v = d_v \times \sum_{w \in N(v) \setminus S} p(v, w) \quad (4)$$

이러한 GDD 알고리즘은 다음과 같다.

Algorithm 1

INPUT: A directed or undirected graph $G = (V, E)$ with propagation probabilities $\{p(u, v)\}$, and a seed set size k
OUTPUT: The most influential set S with k nodes

```

1: initialize  $S = \emptyset$ 
2: for each node  $v$  do
3:   compute its out-degree  $d_v$ 
4:   if  $G$  is directed graph
5:     compute its general degree discount  $gdd_v$  by Eq. (2)
6:   else
7:     compute its general degree discount  $gdd_v$  by Eq. (4)
8: end for
9: for  $i = 1$  to  $k$  do
10:  select  $u = \arg \max \{gdd_v | v \in V \setminus S\}$ 
11:   $S = S \cup \{u\}$ 
12:  for each out-neighbor  $v$  of  $u$  and  $v \in V \setminus S$  do
13:    if  $G$  is directed graph
14:      calculate new value of  $gdd_v$  by Eq. (1)
15:    else
16:      calculate new value of  $gdd_v$  by Eq. (3)
17:  end for
18: end for
19: output  $S$ 

```

III. 성능평가

성능평가 실험은 Weighted Independent Cascade (WIC) 모델을 적용하여 수행하였다^[2]. 정보 확산 확률 $p(u, v) = 1/d_u$ 로 가정하였다. 성능평가 대상 소셜 네트워크 모델은 표 1과 같다^[5].

성능평가 비교를 위하여 다음의 세 알고리즘을 고려하였다.

- General Degree Discount (GDD) : 제안 알고리즘
- Degree Discount (DD) : [4]에서 제안된 알고리즘으로 Uniform 정보 확산 확률 사용
- High Degree (D) : 이웃노드 수만을 고려한 정보 확산 알고리즘

그림 1은 PHY 모델에 세 알고리즘을 적용하여 얻

표 1. 대상 소셜 네트워크 모델
Table 1. Social Network Models for Evaluation

Networks	# Nodes	# Edges	Type of graph
PHY	37K	231K	Undirected
EPINIONS	76K	509K	Directed

은 정보 확산 성능을 나타내고 있다. Seed 노드군의 크기가 증가할 때 성능의 차이가 증가함을 볼 수 있다. 그림에서와 같이 Seed 노드군 크기가 50일 때, 제안 알고리즘은 DD 대비 17.64%, D 대비 54.61% 우수한 정보 확산 성능을 보이고 있다.

그림 2는 EPINIONS 모델에 세 알고리즘을 적용하여 얻은 정보 확산 성능을 나타내고 있다. 그림에서와 같이 Seed 노드군 크기가 50일 때 제안 알고리즘은 DD 대비 6.95%, D 대비 8.96% 우수한 정보 확산 성능을 보이고 있다.

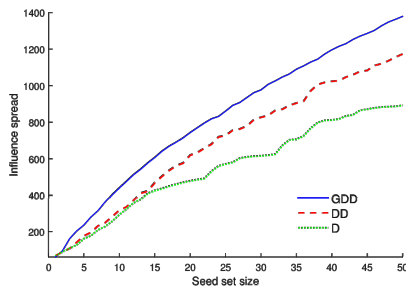


그림 1. PHY 모델에서의 정보 확산 성능
Fig. 1. Influence Spread for PHY mode

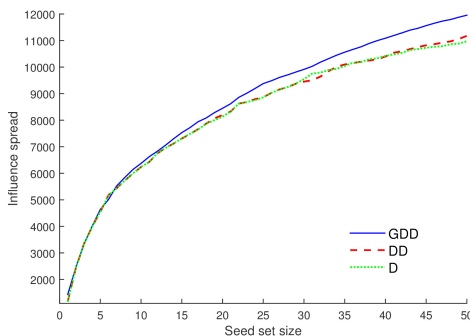


그림 2. EPINIONS 모델에서의 정보 확산 성능
Fig. 2. Influence Spread for EPINIONS

IV. 결 론

본 논문에서는 소셜 네트워크에서 정보 확산 극대화화를 위한 GDD 알고리즘을 제안하였다. GDD 알고리즘은 Directed 및 Undirected 그래프에 모두 적용 가능하여 활용도가 크고, 기존에 제안되었던 Heuristic 알고리즘에 대비하여 우수한 정보 확산 성능을 보임을 제시하였다.

References

- [1] Y. Kim and G. Park, "Topic sensitive_social relation rank algorithm for efficient social search," *J. KICS*, vol. 38, no. 5, pp. 385-393, 2013.
- [2] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 137-146, 2003.
- [3] M. Kimura and K. Saito, "Tractable models for information diffusion in social networks," in *Knowledge Discovery in Databases: PKDD*, Springer Berlin Heidelberg, pp. 259-271, 2006.
- [4] W. Chen, Y. Wang, and S. Siyu, "Efficient influence maximization in social networks," in *Proc. 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pp. 199-208, 2009.
- [5] J. Leskovec, *Stanford Large Network Dataset Collection - Social networks*, Available: <https://snap.stanford.edu/data/>