

Adaptive Influence Maximization in Microblog under the Competitive Independent Cascade Model

Zheng Ding, Kai Niu, and Zhiqiang He

Abstract—With the rapid development of social media technology, many different pieces of information, ideas, products, and innovations are propagating widely in online social networks. Information diffusion has been further researched by many scientists and experts in the past few decades. In this paper, we study the competitive influence propagation in sina microblog under the competitive independent cascade model, which extends the classical independent cascade model. This paper pays attention to the problem that adaptively selecting a certain number of seeds to maximize its influence benefit (IB) under a competitive diffusion model. We call this problem the adaptive influence maximization (AIM) problem. The traditional Monte-Carlo greedy algorithm can select a specified number of seeds, and has a very high complexity. A new efficient algorithm called M-based algorithm is designed to select seeds adaptively, and faster than the Monte-Carlo greedy algorithm.

Index Terms—Adaptive influence maximization, competitive independent cascade model, microblog.

I. INTRODUCTION

As we move further into the information age, we can communicate with each other in online social media more convenient and faster than ever before. So much information surrounds us daily and we are free to choose which pieces of media message to adopt and repost, and which pieces to ignore. This process is called information propagation. The information diffusion model through a social network has been studied by many scholars. For example, susceptible infective Susceptible (SIS) model and susceptible infective removal (SIR) model [1] are first proposed to describe the information diffusion process. It is in both cases that state of nodes transfer from one of S (susceptible), I (infective), R (removal) to one of the two other states with certain probabilities. Simple differential equations are used to describe the percentage of number of nodes in each state. Kempe *et al.* [2] proposed the independent cascade (IC) model and the linear threshold (LT) model, which characterize how information are disseminated throughout the network starting from the initial seeds nodes, which are based on a graph structure and are widely used today.

All the models above only study the propagation of a piece

of information in a social network. We call this single-item diffusion. In fact it is often the case that different information are propagating concurrently in a social network. Information competition is omnipresent in social network, such as the sales competition of similar goods, the campaign between the parties in social media, and the competitive spread of rumors and truth in public. Competitive influence diffusion model has been studied by a lot of papers and books, Chen *et al.* [3] propose the competitive independent cascade (CIC) models and the competitive linear threshold (CLT) model where the different kinds of information diffusion process proceed iteratively in a synchronous way along a discrete time-axis, starting from sets of initial activated nodes. In addition, broecheler *et al.* [4] put forward the weighted generalized annotated program (wGAP) framework for expressing competitive diffusion models. Borodin *et al.* [5] extend the well-studied linear-threshold model in three different ways to model competitive influence diffusion.

The competitive influence maximization problem corresponds to the issue of competitive information diffusion, which aims to expand the influence coverage of a piece of information at the same time that its competitive information is spreading. The literature on the competitive influence maximization problem is abundant. CLDAG algorithm is designed to solve the influence blocking maximization problem under the competitive linear threshold model in the *et al.* [6]. Carnes *et al.* [7] propose two models for competitive information diffusion, and give an efficient algorithm for the influence maximization problem from the follower's perspective.

In this paper, we make a contribution to the competitive influence propagation in microblog under the competitive independent cascade model. Competitive influence maximization problem is redefined as adaptive influence maximization problem which takes the cost of initial seeds into consideration and select the initial seeds adaptively. The method for adaptive influence maximization is worth further researching and widely used in the commercial field. First of all, we introduce the competitive independent cascade model in sina microblog in Section II. Adaptive influence maximization problem is proposed in Section III, while applying the Monte-Carlo greedy algorithm to try to solve the AIM problem. A more efficient algorithm called M-based algorithm is designed to solve AIM problem in Section IV. At last, we make a comparison between these two methods and draw a conclusion about the whole paper.

II. COMPETITIVE INDEPENDENT CASCADE MODEL IN SINA MICROBLOG

Chen *et al.* [3] proposed the competitive independent

Manuscript received March 3, 2015; revised June 2, 2015. This work is supported by National High Technology Research and Development Program of China (863 Program SS2015AA011303), and National Natural Science Foundation of China (61171099).

The authors are with the Key Laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: dzsd2013@163.com, niukai@bupt.edu.cn, hezq@bupt.edu.cn).

cascade model to describe the multi-item diffusion, we apply CIC model to sina microblog and only consider two kinds of competitive information which we refer to as the positive information and the negative information. In this model, a social network is considered as a directed graph $G = (V, E)$, where V is the set of nodes respecting users in sina microblog and E is the set of directed edges respecting influence relationships between users. Each node has three possible states: inactive, positive, negative. The positive and the negative influence probabilities of each edge are the same, that is the propagating probability $p(\cdot)$. A node may change from the inactive state to the positive state or negative state, but cannot change from active state to inactive state, and the state of activated nodes is no longer changes. The propagation process starts from initial seed sets S_0^+ and S_0^- , and the result is to obtain the positive and negative active set S^+ and S^- while the information to complete the diffuse process. The diffusion rule [3] is as follows:

- 1) Given the initial seed sets S_0^+ and S_0^- with $S_0^+ \cap S_0^- = \Phi$ at time step $t = 0$. At every time step $t \geq 1$, nodes in activated seed sets S_{t-1}^+ and S_{t-1}^- try to activate their neighbors.
- 2) Positive activate attempt: every active node $v \in S_{t-1}^+$ try to activate their neighbors $u \in N^in(v) \setminus (S^- \cup S^+)$, v carries out a positive activation attempt with successful probability $p(u, v)$. If the attempt is successful, u is added into a positive successful attempt set for node v , if not, u no longer be activated.
- 3) Negative activate attempt is in the same way with the above process of positive activate attempt. In the case that a node is both positively activated and negatively activated in the same time, the state of the node is set to the positive.
- 4) The competitive diffusion ends with the final active sets S_t^+ and S_t^- until the sets of positive and negative nodes will no longer change, that is, $S_t^+ = S_{t-1}^+$ and $S_t^- = S_{t-1}^-$.

In our model, the diffusion probability $p(u, v)$ of node v to node u is proportional to the similar degree and familiar degree of node u and node v . The similar degree is represented by the microblog text similarity $S(u, v)$ of user u and user v , and the familiar degree is represented by the ratio of common friends $F(u, v)$.

$$p(u, v) = \alpha \times S(u, v) + \beta \times F(u, v), \alpha + \beta = 1 \quad (1)$$

where the parameters α and β indicate the influence weight on diffusion probability of the similar degree and the familiar degree. The math definition of similar degree and familiar degree are as follows,

$$S(u, v) = \frac{|T_u \cap T_v|}{|T_v|}, S(u, v) \in [0, 1] \quad (2)$$

where T_u and T_v are the microblog text's key words of user u and user v , $|T|$ represents the key words' number of text

T .

$$F(u, v) = \frac{|N_u \cap N_v|}{|N_v|}, F(u, v) \in [0, 1] \quad (3)$$

where N_u , N_v represent the neighbor friends of user u , v , $|N_u \cap N_v|$ represents the number of common friends of user u and user v , $|N_v|$ represents the friend number of user v .

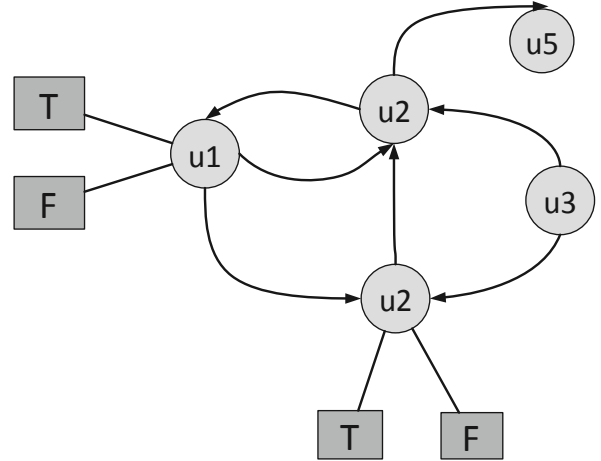


Fig. 1. A directed social graph in which nodes have two attributes: Microblog text T and friends F .

III. ADAPTIVE INFLUENCE MAXIMIZATION PROBLEM

In this section, we first give the definition of the adaptive influence maximization (AIM) problem, then introduce the Monte-Carlo algorithm which is used to solve the AIM problem and analyze its performance.

A. Problem Definition

The adaptive influence maximization problem is a optimization problem in which given a social graph $G = (V, E)$, the above described competitive independent cascade model, a negative seed set S_0^- , the goal is to find a positive seed set $S_0^+ \subseteq V \setminus S_0^-$, such that the positive influence benefit B^+ is maximized. The positive influence benefit B^+ is the positive coverage $\sigma^+(S_0^+, S_0^-)$ minus the cost of selected positive seeds $C(S_0^+)$. $C(S_0^-)$ is proportional to the size of positive seeds $|S_0^+|$.

The mathematical representation is:

$$S_0^{+*} = \arg \max_{S_0^+ \subseteq V \setminus S_0^-} \sigma^+(S_0^+, S_0^-) - C(S_0^+) \quad (4)$$

$$C(S_0^+) = \gamma \times |S_0^+| \quad (5)$$

Compared with the classical influence maximization problem, there is no limit for the number of the required seed set in AIM problem. Besides, we take the cost of seeds into consideration so that make the influence maximization problem have a optimal solution.

The influence benefit function is submodular but not monotonous in our model. We say that a set function $f: 2^V \rightarrow \mathbb{R}$ is submodular if for any subsets $S \subseteq T \subseteq V$ and any element $v \in V \setminus T$,

$$f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T).$$

First of all, we prove the submodularity of the influence benefit function in here. For any subsets $S \subseteq T \subseteq V$ and any element $v \in V \setminus T$, to get

$$(\sigma^+(S \cup \{v\}, S_0^-) - \gamma \times |S \cup \{v\}|) - (\sigma^+(S, S_0^-) - \gamma \times |S|) \geq (\sigma^+(T \cup \{v\}, S_0^-) - \gamma \times |T \cup \{v\}|) - (\sigma^+(T, S_0^-) - \gamma \times |T|) \quad (6)$$

is equal to get

$$\sigma^+(S \cup \{v\}, S_0^-) - \sigma^+(S, S_0^-) \geq \sigma^+(T \cup \{v\}, S_0^-) - \sigma^+(T, S_0^-) \quad (7)$$

because the formula (7) is equivalent to the formula (6) with both sides of the former inequality subtract a constant $\gamma * |v|$.

The above CIC model in sina microblog is homogeneous because the propagation probabilities of each edge are the same for positive and negative influence, and follows TB-FP(1) regulation. The positive influence coverage function is submodular that is proved in chapter 4 of Wei Chen *et al.* [3]. So the inequality (7) is true and inequality (6) is proved. Therefore, The influence benefit function $\sigma^+(S_0^+, S_0^-)$ is sub-modular. It is obvious that influence benefit is not monotonous so that AIM problem can be solved by Monte-Carlo greedy algorithm using the submodularity and monotonicity of influence coverage. Next subsection describes the Monte-Carlo greedy algorithm for the NP-hard problem in detail.

B. Monte-Carlo Greedy Algorithm

Monte-Carlo Greedy Algorithm is used to solve the influence maximization in single information diffusion process in [3], we apply it to solve AIM problem in competitive information diffusion process.

The main idea of Monte-Carlo greedy algorithm is adding the node which makes the influence coverage maximum into the positive seed set until the number of seed set reach the specified number. This algorithm utilize the submodularity and monotonicity property of the influence coverage. Given seed set S_0^+ , we simulate the randomized diffusion process for R times to get the average of the number of the positive active nodes σ_R^+ . We take the average of these counts in order to get high accuracy in our estimate. In the competitive influence diffusion model, the influence spread is a piece of information's activated nodes number while two pieces of information propagate. In this Algorithm, we first determine the number of selected positive seeds k ranging from 1 to the size of the graph N . Then we find the optimal seeds set to maximize the positive influence benefit.

Monte-Carlo Greedy Algorithm can solve the AIM problem, however, the greedy algorithm requires a large number of traversal to choose the optimal node, which cannot be done efficiently. It takes about $O(n^2 \log n)$ time to solve the

AIM problem. When dealing with larger networks, the time consumption is unacceptable. In next section, we propose a faster algorithm to select seed nodes adaptively.

The Monte-Carlo greedy algorithm scheme can be obtained in **Algorithm I** via the following recursion procedures:

Algorithm I Monte-Carlo greedy algorithm for adaptive influence maximization.

Input: G : social graph including arc weights for CIC model; S_0^- : the negative seed set

Output: S_0^+ : the selected positive seed set

- Initialize $S_0^+ \leftarrow \emptyset$;
 - **for** $k = 1 : N$
for $j = 1 : k$
 $u \leftarrow \arg \max_{w \in V \setminus (S_0^+ \cup S_0^-)} \sigma_R^+(S_0^+, S_0^-, G)$
 $S_0^+ \leftarrow S_0^+ \cup \{u\}$
end
 $S_0^+ \leftarrow \arg \max_{S_0^+} B^+$
end
 - **return** S_0^+ ;
-

IV. M-BASED ALGORITHM FOR THE AIM PROBLEM

A. M-Based Algorithm

The positive and negative influence coverage is not only submodular but monotonous. While a node is added in positive seed set, the positive influence coverage may remain the same or increase, the positive influence benefit whose change is continuous will decrease or increase accordingly. It is easy to know the influence benefit function is convex. So the positive benefit has a maximum value, the node which can enhance the positive influence benefit will be selected as the positive seeds by browsing the whole nodes except negative seeds. This idea is based on the maximum value property of the influence spread, so we call this idea M-based algorithm.

The mathematical description of M-based algorithm is as follows: given a social network considered as a directed graph $G = (V, E)$ and a negative seed set S_0^- , we want to select a positive seed set S_0^+ to maximize the positive influence benefit $IB(S_0^+)$. First of all, S_0^+ is set to an empty set. If $v \in V \setminus S_0^-$ can make the positive influence benefit $\hat{IB}(S_0^+)$ increasing, v is added in positive seed set S_0^+ , until there is no node can enhance the positive influence benefit. Because the diffusion process is random, $\hat{IB}(S_0^+)$ is a estimated value which is the average of positive influence benefit in M times competitive diffusion. The number of the positive seeds do not need to specify, we obtain the adaptive seed set in this algorithm. In terms of complexity it takes about $O(n)$ times to solve the AIM problem.

The M-based algorithm process can be obtained in **Algorithm II**:

Algorithm II M-based algorithm for adaptive influence maximization.

Input: G : social graph including arc weights for CIC model; S_0^- : the negative seed set

Output: S_0^+ : the selected positive seed set

- Initialize $S_0^+ \leftarrow \phi$;
- $v \in V \setminus S_0^-$
- if $\hat{IB}(S_0^+ \cup \{v\}) - \hat{IB}(S_0^+) > 0$
- $S_0^+ \leftarrow v$
- $\hat{IB}(S_0^+) = \frac{\sum_{i=1}^M IB(S_0^+)}{M}$;

B. Experiments and Results

In this section, we use a dataset contains the directed user network and their microblog content extracted from sina API in 2014. Each node in the network represents an user, and each directed edge connecting two nodes indicates that the two users have following relationship. The dataset is consist of 592 users. Table I reports the basic statistics of the dataset.

Number of nodes	592
Number of edges	29843
Average degree	50
Maximal degree	327

Every node in the dataset is numbered, the node number ranges from 0 to 591. The negative seeds are selected randomly in this experiment, their number are 22, 102, 233, 389, 402, 440, 449, 503, 553. In our experiments, the seed cost parameter γ is set to 1, Monte-Carlo Greedy algorithm with $R = 10$ under CIC model, M-based algorithm with $M = 1000$ under CIC model, and the result is the average of many times simulation. At the same time, the result curve is smoothed by Origin Pro with the parameter "points of windows" is 20. Simulation computer configuration is Intel CPU with 3.40GHZ and RAM with 5.82G.

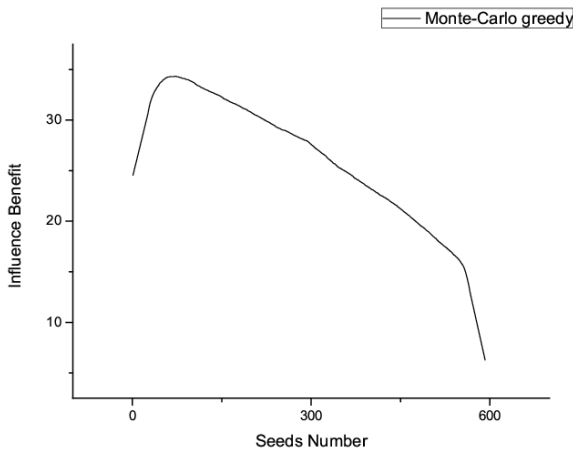


Fig. 2. Experiment result of Monte-Carlo greedy algorithm.

We use above two algorithms to select positive seeds, 72 positive seeds which reach the maximum positive influence benefit 34.343 are selected by Monte-Carlo Greedy algorithm, 70 positive seeds which reach the maximum positive influence benefit 38.572 are selected by M-based algorithm. The program of selecting optimal positive seeds using Monte Carlo Greedy algorithm takes 3.288 hours and the program using M-based algorithm takes 0.853 hours. Fig. 2 depicts the positive influence benefit's change with the increase of the positive seeds number using Monte-Carlo Greedy algorithm.

Fig. 3 shows the positive influence benefit's change using M-based algorithm. Fig. 4 shows the program running times using Monte-Carlo Greedy algorithm and M-based algorithm.

From the result we can see M-based algorithm faster than Monte-Carlo Greedy algorithm, but the influence benefit is slightly smaller. Because the diffusion has random property, the influence benefit may float. And the experiment shows selected seed set is not unique.

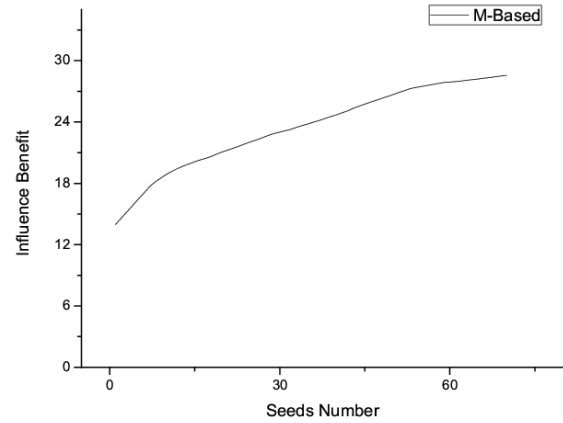


Fig. 3. Experiment result of m-based algorithm.

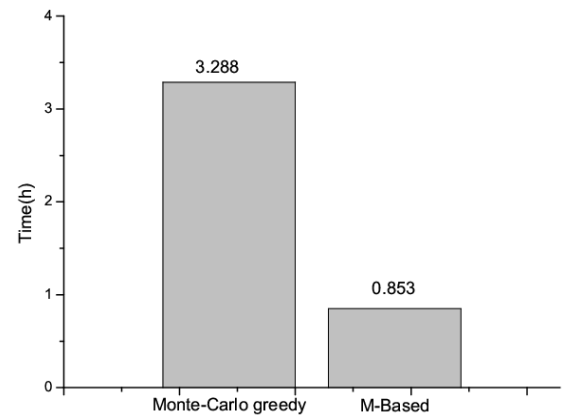


Fig. 4. Running time of Monte-Carlo greedy algorithm and M-based algorithm.

V. CONCLUSIONS

In this paper, we investigate the adaptive influence maximization in competitive diffusion of social network. We design M-based algorithm to overcome the slowness of the Monte Carlo greedy algorithm. However, the problem is still worth continuing research.

REFERENCES

- [1] M. E. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [2] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2003, pp. 137–146.
- [3] W. Chen, L. V. Lakshmanan, and C. Castillo, "Information and influence propagation in social networks," *Synthesis Lectures on Data Management*, vol. 5, no. 4, pp. 1–177, 2013.
- [4] M. Broecheler, P. Shakarian, and V. Subrahmanian, "A scalable framework for modeling competitive diffusion in social networks," in *Proc. 2010 IEEE Second International Conference on Social Computing (SocialCom)*, IEEE, 2010, pp. 295–302.
- [5] A. Borodin, Y. Filmus, and J. Oren, "Threshold models for competitive influence in social networks," *Internet and Network Economics*, Springer, 2010, pp. 539–550.

- [6] X. He, G. Song, W. Chen, and Q. Jiang, "Influence blocking maximization in social networks under the competitive linear threshold model." *SDM*, SIAM, 2012, pp. 463–474.
- [7] T. Carnes, C. Nagarajan, S. M. Wild, and A. V. Zuylen, "Maximizing influence in a competitive social network: a follower's perspective," in *Proc. the Ninth International Conference on Electronic Commerce*, ACM, 2007, pp. 351–360.



Zheng Ding received a B.E. degree in communication engineering from Chongqing University of Posts and Telecommunications, Chongqing, China, in 2012. Currently she is working towards the M.S. degree in information and communication engineering in Beijing University of Posts and Telecommunications, Beijing, China. Her research interests are in the area of data mining and the Complex network dynamics.



Kai Niu received a B.S. degree in information engineering and a Ph.D. in signal and information processing from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1998 in 2003, respectively. Currently he is a professor in the School of Information and Communication Engineering of BUPT. His research interests are in the area of data mining and broadband wireless

communication, particularly on the complex network theory.



Zhiqiang He received the B.E. and Ph.D. degrees (with distinction) in signal and information processing from Beijing University of Posts and Telecommunications, China, in 1999 and 2004, respectively. Since July 2004, he has been with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, where he is currently an associate professor and the director of the Center of Information Theory and Technology. His research interests include signal and information processing in wireless communications, data mining in social network, and underwater acoust.