Maximizing the Spread of Positive Influence by Deadline

Hemank Lamba, Jürgen Pfeffer
Carnegie Mellon University
{hlamba,jpfeffer}@cs.cmu.edu

ABSTRACT

In uence maximization has found applications in various elds such as sensor placement, viral marketing, controlling rumor outbreak, etc. In this paper, we propose a targeted approach to in uence maximization in polarized networks, i.e. networks where we already know or can predict nodes' opinions about a product or topic. The goal is to nd a set of individuals to target such that positive opinion about a speci c topic or the product to be launched is maximized. Another key aspect that is present in most of the existing viral marketing algorithms is that they do not take into account the timeliness of product adoption. In this paper, we present a framework in which we infer users' polarity and activity levels and then select seeds to launch a viral marketing campaign such that positive in uence about the product is maximized by a given deadline.

Keywords

viral marketing, in uence maximization, information di usion

1. INTRODUCTION

The phenomenon of viral marketing rests on exploiting the social connections among individuals to promote awareness for new products. This is equivalent to nding a k-sized set of initial seeds that will maximize the adoption of the product after some di usion process. This problem is also known as *in uence maximization*. In their seminal paper, Kempe et al. formulated this problem as a discrete optimization problem [2]. The paper considered two information propagation models, which were later widely studied in the literature, namely the *Independent Cascade* (IC) and the *Linear Threshold* (LT) model. However, the relevant literature[1] does not take into account the following practical aspects:

- Often companies want to promote their products to certain types of customers as opposed to targeting the entire social network. For instance, a company introducing an application for iPhone users will want to target only those having an iPhone.
- Most of the existing information di usion algorithms assume that once a user gets activated, he/she will also buy the product, which may not actually be the case in practice. In reality, adoption of a product depends on several factors, such as product price, user preferences, product quality, etc.
- Traditional information di usion algorithms generally do not take into account temporal aspects of the diffusion process, for instance, users' activity levels.
- A realistic viral marketing campaign not only requires targeting those users in the social network who will like the product, but also doing so in a timely fashion. For example, if a company is selling tickets to an event, it does not make sense to have information about the event reach an interested person after the event has already taken place.

In this paper, we propose an extension to the traditional in uence maximization framework that addresses the above limitations. We propose a method that will nd the best set of users to maximize the adoption of a product within a given deadline. For our example, we assume that users who prefer and have in the past adopted similar products are likely to adopt the marketed product. We use the users' past activity logs to infer their activity levels (which helps us estimate how active or passive they are in spreading the information to their social peers) as well as their preferences (which helps us estimate whether the user is likely to adopt the product or not).

2. PROBLEM

Information Diffusion Model: We begin by describing the underlying information di usion model that incorporates temporal aspects. Each edge $(u, v) \in E$ is assigned an activity probability $s(u, v)$, which captures how often information ows through the edge. This probability is directly proportional to the activity of both users u and v. At any time step, an activated node u tries to activate all it's currently inactivated edges with probability $s(u, v)$. Once the edge is activated, v gets a single chance to become activated to spread it's own in uence. This can happen with probability p. The process continues until all the edges have

[1] See [4] for a comprehensive literature survey on in uence maximization

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889412 .

been activated or the deadline τ has passed. This model is very similar to the *Independent Cascade with Meetings model*.

Objective Function: If S denotes the set of seed nodes, where $|S| = K$; τ denotes the deadline by which the adoption has to be maximized; and P denotes the set of users who are likely to adopt the product, then the objective function is defined by $\sigma_P(S, \tau)$. It indicates the number of users who were in P and got activated by the deadline τ under the information diffusion model specified above, which was started by choosing S as seed nodes. We are interested in maximizing this objective function.

Since the objective function can be reduced to the traditional influence maximization under IC-model by setting $\tau = \infty$, $P = V$, $s(u, v) = 1 \forall (u, v) \in E$, we can conclude that the objective function is submodular, and we can apply the standard greedy algorithm guaranteeing $(1 - \frac{1}{e})$ optimality.

3. EXPERIMENTS

We conduct our experiment on the social network Flixster[2]. We consider only the largest connected component of the social network, which leaves us with a social graph consisting of $11,643$ nodes and $105,420$ edges. Additionally, we also have activity logs for Flixster containing user ratings of movies. For this dataset, we treat movies as products and choose a movie at random to be marketed.

Estimating product adoption: To infer whether a user will like the marketed movie or not, we consider the user's past ratings for other movies. However, the marketed movie might not have any ratings yet, which would make this a *cold start* problem, which is difficult to handle with traditional collaborative filtering based recommendation algorithms. To avoid such a situation, we provide a random set of 100 user ratings for the marketed movie. These could be seen as user reactions from an advanced private screening of the movie. Alternatively, advanced recommender systems that can deal with cold start could also have been applied. Thus our positive set is comprised of users for which we have ratings greater than 3.5 as given to us by the recommender system output.

Estimating edge activation probability: The edge activation probability $s(u, v)$ is heavily dependent on the activity levels of both the users u and v. The probability is directly proportional to the activity levels of u and v, A_u, A_v. We make use of exponential distribution to model the edge activation probability. This is denoted by

$$s(u, v) = \lambda * \exp(\lambda(A_u + A_v))$$

Baselines: To evaluate our performance, we consider various baseline algorithms. We extend the degree discount heuristic [1] to take into account the targeted set of nodes. For each node $u \in V$, we define its effective degree $d_{eff}(u)$ as the number of neighbors of u that are only in the set P. Starting with empty seed-set $S = \phi$, at each step we add the node to the set S which maximizes $d_{eff}(u) - (2|S_v| + (d_{eff}(u) - |S_v|)|S_v * p)$, where p denotes the activation probability and S_v denotes the neighbors of u that are activated. For greedy, we run the CELF [3] algorithm, followed by 1000 Monte-Carlo simulations. Degree is a heuristic based on the notion of degree centrality that considers high-degree nodes as influential ones. Betweenness is a similar heuristic, but it is based on the notion of betweenness centrality. Activity assumes that nodes which have high activity levels will have a higher probability of activating edges to their neighbors more quickly than others and hence might be influential within the given deadline.

[2]http://www.cs.ubc.ca/~jamalim/datasets/

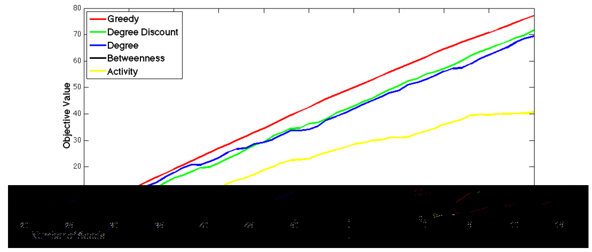

Figure 1: Comparison of our method (greedy) with that of other baseline algorithms. We can clearly see that the greedy approach outperforms the other algorithms.

For all the experiments, we assumed constant activation probability as 0.01, and τ was considered to be 10 time steps. We computed 50 seeds for every model. For computing objective values for baselines, 1000 Monte Carlo simulations were run. We plot the results in Figure 1. We can see from the results that the quality of seed sets based on the greedy algorithm is higher in terms of the expected influence spread by deadline than that of other baseline methods. Surprisingly, the degree discount heuristic performs better than expected compared to the greedy approach. However, the set of nodes chosen by the heuristic is different from the set chosen by the greedy approach. There was an overlap of 19 nodes out of the possible 50 (Jaccard Score=0.234).

4. CONCLUSIONS AND FUTURE WORK

In this paper, we presented an influence maximization framework that maximizes the spread of positive influence about a product within a given deadline. Multiple extensions are possible to this work, such as treating time as continuous instead of discrete, or trying to jointly optimize the spread of the cascade process and the speed of the diffusion, or considering more realistic information diffusion processes by relaxing more assumptions.

References

[1] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *KDD*, 2009.

[2] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.

[3] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *KDD*, 2007.

[4] H. Zhang, S. Mishra, M. Thai, J. Wu, and Y. Wang. Recent advances in information diffusion and influence maximization in complex social networks. *Opport un ist ic Mobile S o arks*, 2014.