

An Efficient and Scalable Distribution of Targets for Optimal IMAX Query Processing with user Interest

Naresh. P¹, Karpagam. V²

¹PG Scholar, ²Professor

Department of Software Engineering (PG)

Sri Ramakrishna Engineering College,

Coimbatore, Tamilnadu.

Abstract: Data Mining is the practice of searching large stores of data to discover patterns and trends that go beyond simple analysis. Influence maximization is a problem of finding seed nodes in a social network in order to maximize the profit of viral marketing in social networks. The influence maximization problem seeks to select k users from a social network through whom the number of influenced users is maximized. In the real time social networks, users have their own interests and are more likely to be influenced through their friends along with similar topics. However influence maximization has the problem with differentiation of particular users from other users. Hence in the existing scenario, fast greedy-based approximation method has been introduced using the expectation model to achieve the objective function. In this scenario, the relationship of paths among the users is discovered by using the concept of expectation model. And the influence maximization (IMAX) aims to find k seed users to maximize the spread of influence among users in social networks. A new efficient expectation model is introduced for influence spread of a seed set. Based on the new expectation model, a greedy-based approximation method processes an IMAX query with efficient incremental updating of the marginal gain of each user. However, it does not consider the diverse distributions of targets such as same community or the same university based on the static profiles of users. Also it is not consider the linear threshold model. With this intension, the focus is on the improvements on query processing for various distribution inferences and linear threshold model concepts. In the linear threshold model, it is avoiding the issues of greedy algorithms effectively. The influence spread in directed acyclic graphs (DAGs) can be computed in linear time, which relies on an important linear relationship in activation probabilities between a node and its in-neighbours in DAGs. Based on the fast influence computation for DAGs we propose the first scalable LIDAG algorithm for influence maximization in the LT model. It is sued for improving the process of IMAX query approach efficiently. Based on the static and dynamic user profiles we can predict the user behaviour and affinities in many web applications such as targeting of online advertising, content personalization and social recommendations. Our results show that our model contributes towards improved behavioural of optimal seeds in targeting nodes.

Key words: Graph algorithms, influence maximization, independent cascade model, social networks, Linear Threshold Model.

1. INTRODUCTION

Influence maximization objective is to select k number of seed users from a online social network such that the number of users influenced by the seeds is maximized, has attracted significant attention form the online marketers due to its widespread applications. In the existing scenario, viral marketing is one of the key applications of influence maximization. In viral marketing, a marketer who wants to promote his company's item will diffuse his product into social networks. Through viral marketing, influence

maximization provides maximum profit from specific user out of all common users in a social network. However, the influence maximization is not always the most effective strategy for viral marketing. Since there can be some items that are used only by the specific users. These specific users can be a certain number of people with a common interest in a given item, some or all people in a same group or community, or some or all users in a class. There is no limit for being specific users.

Consider a Marketer who asked to promote a item which produced by the company which produces the products that can be used only by the specific users. For example, consider a marketer who is asked to promote a video game for teen agers through viral marketing. For that video game product, the specific users are kids, and teen age users who are likely to use it and parents users who wish to purchase it for their children. In this case, the marketer does not need to be worry about the other users because the video game is not useful to them. Instead, it is a better strategy to the marketer to focus on increasing the number of influenced specific users, but influence maximization has the weakness that it cannot distinguish the specific users from the other common users. For handling such targets with influence maximization, the only way is making a homogeneous graph with the targets and executing influence maximization on the graph. However, this approach is not effective on the users who are not targets but can strongly influence the targets, so that the result of this approach should be inaccurate.

2. LITERATURE SURVEY

Jong-Ryul Lee and Chin-Wan Chung proposed model to formulate an influence maximization problem as query processing and call this IMAX query processing. To process an IMAX query processing, a social network is taken as a graph where a node(vertex) represents an individual and an edge represents a relationship between two individuals such as the communication between the friends. Each edge $(u; v)$ has a probability that u influences v . With such probabilities, the information passing through edges is modelled by the independent cascade (IC) model. In the IC model, user u has one-time chance to influence an uninfluenced neighbour v at time $t + 1$ when u is influenced at time t . If u fails to influence v , there is no second chance for u to influence v . However, v can be influenced by another user w when there is an edge from w to v and w is influenced. In addition, if u is influenced at time t , u will not be influenced again after time t . Under the Independent Cascade model, an IMAX query consists of seed set size and target node set, and it asks the seed users to increase the number of influenced users among the users that are specified in the IMAX query. The number of influenced users can be measured by the expected number of influenced users. The IMAX query problem is effective over influence maximization problem in two aspects. One is the targets cannot be distinguished from the other users by the influence maximization problem, so it is not suitable for target viral marketing. However, in the IMAX query problem, we can specify targets explicitly using a target set and we can focus on increasing the influence on those targets. Next, the other one is efficiency. There are lots of users in real world who want to promote many items for improving their profit using online social networks. Since IMAX query processing can be a effective method to promote an item for those users, the total number of potential users of IMAX query processing can be in huge numbers. It means that efficiency is a very important issue for IMAX query processing. The IMAX query processing approach is NP-hard and has sub modularity like influence maximization problem. The techniques used for influence maximization problem can be also used for IMAX query processing approach. However, they are inefficient to process the IMAX query. In contrast to influence maximization, we know target nodes that we want to influence when an IMAX query is given. It means that an efficient method for an IMAX query should identify quickly the nodes that strongly influence the targets of the query with pre-processed data.

3. PROPOSED SYSTEM

In the proposed system, we focus on the improvements on query processing for various distribution inferences and linear threshold model concepts. In the linear threshold model, it is avoiding the issues of greedy algorithms effectively. Using the linear threshold model the influence spread in directed acyclic graphs (DAGs) can be computed in linear time. The computation of influence spread relies on an important linear relationship in activation probabilities between a node and its in-neighbours in directed acyclic graphs. Based on the fast influence computation for Directed

acyclic graphs we propose the first Linear time for improved directed acyclic graph(LIDAG) algorithm and Hybrid of greedy and LIDAG algorithm to compute the threshold value and estimate the influence probability. We propose Improved greedy approach with optimized algorithm to reduce the time complexity and, improve the large scale social networks and to provide optimal seed users for maximal social network.

Based on the static and dynamic user profiles we can predict the user behaviour and affinities in many web applications, content personalization and social recommendations. We propose a new streaming, distributed algorithm which is able to handle millions of users. Our results show that our model contributes towards improved behavioural of optimal seeds in targeting nodes. Thus this scenario is used for improving the user profile through several distributions of targets.

The Linear threshold model:

In this module, we define the linear threshold (LT) model as follows. An (LT) influence graph is a weighted graph $G = (V, E, w)$ where V is set of n vertices (nodes) and $E \subseteq V \times V$ is a set of m directed edges, $w: V \times V \rightarrow [0,1]$ is a weight function such that $w(u, v) = 0$ if and only if (u, v) not belongs to E and $\sum_{u \in V} w(u, v) \leq 1$. In the linear threshold model, when given a seed set $S \subseteq V$, influence cascades in graph G as follows. Initially every vertex v independently selects a threshold λ_v uniformly at random range $[0,1]$ which reflects the lack of knowledge of user's true threshold. The influence cascade in discrete steps $i=0, 1, 2, \dots$ and let S_i denoted the set of vertices activated at step i along with $S_0 = S$. In each step $i \geq 1$, a vertex v belongs to V which is activated. The weighted number of its activated in neighbours reaches its threshold $\sum_{u \in U_0 \subseteq S_{i-1}} w(u, v) \leq 1$. The process stops a step t when $S_t = \text{null}$. In this scenario, we consider $S_i = \text{null}$ for all $i > t$. $\sigma_L(S)$ denote the expected number of activated nodes given the seed set S (i.e. the expected value of $|\bigcup_{i \geq 0} S_i|$), where the expectation is taken among all λ_v values from their uniform distributions. Now we can call $\sigma_L(S)$ the influence spread of seed set S in influence graph G under the linear threshold model.

Linear time for improved directed acyclic graph (LIDAG):

In this module, each node v in graph $G=(V:E:w)$ is associated with a local DAG LDAG(v), a sub graph of G containing v . We cannot apply DAG algorithm directly to the real social networks since they are not typical DAGs. We also refer to LDAG(v) as the LDAG rooted at v . Given a seed set S , we assume that the influence from S to v is only propagated within LDAG(v) according to the LT model. Let $ap(v; S)$ be the activation probability of v when influence is propagated within LDAG(v). Then the influence spread of S in the LDAG influence model, denoted as $\sigma_D(S)$ is given by $\sigma_D(S) = \sum_{u \in E} ap(v, S)$.

The following algorithm describes how the optimal seeds are selected for scalable approaches more effectively. It computes the threshold value and estimates the influence probability for all users in the specified network.

Algorithm:

- $X = \theta; Y = \theta; \forall u \in V, Inf(u, v) = 0; Inf(v, v) = 1$
- While $\max_{u \in V \setminus X} Inf(u, v) \geq \theta$ do
- $x = \arg \max_{u \in V \setminus X} Inf(u, v) \geq \theta$ do
- $Y = Y \cup \{(x, u) \mid u \in X\}$ // adding the edges
- For each node $u \in N^{in}(x)$ do
- $Inf(u, v) += w(u, x). Inf(x, v)$
- End for
- End while
- Return $D = (X, Y, w)$ as the LIDAG (v, θ)

Improved greedy approach with optimized algorithm:

This algorithm is used to provide optimal seed users for maximal social network and it take less computation time for accessing the path. It is efficiently providing the shortest path based on the higher ranks. This improved greedy approach with optimized algorithm is better than previous algorithm.

Algorithm:

Input: Network graph $G(V, E)$, budget b , time T and seed set.
Output: Optimal seed set.

- $S = \text{null};$
- $\delta(v) = (\sigma(SUv) - \sigma(S))/c(v)$, for all v belongs to V
- $S_{max} = \arg \max_{v \in V} \sigma(v)$
- $p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i | x_{pre}(x_i))$
- $\sigma(S) = E(\sigma_C(S)) = \sum_{u \in V} E(x_u) = \sum_{u \in V} P(u)$
- $Max\ Inf\ Path(G, u, v) = \arg \max_{\{p(P) \mid P \in P(G, u, v)\}}$
- *shortest path formula*

$R(v) = \min(-\log(p(P(s, v))))$, for all $s \in S$

- For each $(u, v) \in V_{GR}$ do
- Compute optimal seed $(G, v, \theta) = \{s \in V \mid Max\ Inf(G, s, \theta) \cap Max\ Inf(G, v, \theta) \neq \emptyset\}$
- Update incremental influence spread
- Select maximization user S
- Selected optimal seed set

4. RESULTS

In the result, the existing method and proposed method is compared and estimated by using efficient methodologies. In the existing method, the greedy based approximation method is used for IMAX query processing. In the proposed system, we improve the scalable approaches for enhancing the influence maximization query processing. The proposed hybrid of greedy algorithm and LIDAG and improved greedy approach with optimized algorithm is efficient for providing better performances. From the experimental result we can conclude that the proposed improved greedy approach with

optimized algorithm is superior in terms of higher accuracy and reduction in time complexity metrics.

5. CONCLUSION

IMAX query processing is used to maximize the influence on specific users in social networks. Here the IMAX query model is incorporated with Linear threshold model to give better efficiency under the LDAG algorithm. To summarize, the experimental results demonstrate that our LDAG algorithm performs consistently among the best algorithms in term of the influence spread and can scale to graphs with millions of nodes and edges, while other algorithms either do not scale or do not have stable performance in influence spread. Based on the fast influence computation for DAGs, we proposed the first LIDAG algorithm for influence maximization in the LT model for improving the process of IMAX query approach efficiently.

REFERENCES:

- [1] Jong-Ryul Lee and Chin-Wan Chung, "A Query Approach for Influence Maximization on Specific Users in Social Networks," in IEEE, 2015, VOL. 27, NO. 2.
- [2] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2003, pp. 137–146.
- [3] F.-H. Li, C.-T. Li, and M.-K. Shan, "Labeled influence maximization in social networks for target marketing," in Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust, Int. Conf. Social Comput., 2011, pp. 560–563.
- [4] W. Lu and L. Lakshmanan, "Profit maximization over social networks," in Proc. IEEE 12th Int. Conf. Data Mining, 2012, pp. 479–488.
- [5] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2007, pp. 420–429.
- [6] A. Goyal, W. Lu, and L. V. Lakshmanan, "CELFF++: Optimizing the greedy algorithm for influence maximization in social networks," in Proc. 20th Int. Conf. Companion World Wide Web, 2011, pp. 47–48.
- [7] Y. Wang, G. Cong, G. Song, and K. Xie, "Community-based greedy algorithm for mining top-k influential nodes in mobile social networks," in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 1039–1048.
- [8] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2009, pp. 199–208.
- [9] Q. Jiang, G. Song, C. Gao, Y. Wang, W. Si, and K. Xie, "Simulated annealing based influence maximization in social networks," in Proc. 25th AAAI Conf. Artif. Intell., 2011, pp. 127–132.
- [10] A. Goyal, F. Bonchi, L. Lakshmanan, and S. Venkatasubramanian, "On minimizing budget and time in influence propagation over social networks," Soc. Netw. Anal. Mining, vol. 3, no. 2, pp. 179–192, 2013.
- [11] Barbieri, F. Bonchi, and G. Manco, "Topic-aware social influence propagation models," in Proc. IEEE 12th Int. Conf. Data Mining, 2012, pp. 81–90.
- [12] K. Lerman, and R. Ghosh, "Information contagion: An empirical study of the spread of news on digg and twitter social networks," in Proc. Int. AAAI Conf. Weblogs Soc. Media, 2010, pp. 90–97. 44
- [13] K. Jung, W. Heo, and W. Chen, "Irie: Scalable and robust influence maximization in social networks," in Proc. IEEE 12th Int. Conf. Data Mining, 2012, pp. 918–923.
- [14] S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," in Proc. 3rd Int. Conf. Internet Netw. Econ., 2007, pp. 306–311.
- [15] T. Carnes, C. Nagarajan, S. M. Wild, and A. van Zuylen, "Maximizing influence in a competitive social network: A follower's perspective," in Proc. Int. Conf. Electron. Commerce, 2007, pp. 351–360.