ISSN: 2321-7782 (Online)
e-ISJN: A4372-3114
Impact Factor: 6.047

Volume 5, Issue 6, June 2017

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

A Query Mechanism for Influence Maximization on Definite users in Social Networks

Vidya A. Khairnar[1]
PG student Department of Computer Engineering,
Late G.N.Sapkal college of Engineering,
Savitribai Phule Pune University – India

Prof. J. V. Shinde[2]
Assistant Professor Department of Computer Engineering
Late G.N.Sapkal college of Engineering
Savitribai Phule Pune University – India

Abstract: Influence growth is presented to maximize the profit of viral advertising in social networks. The fault of influence expansion is that it does not differentiate exact users from others, even if some items can be only useful for the specific users. We have proposed an expectation model for the value of the objective function and a fast greedy-based approximation method using the expectation model. For the expectation model, I also investigate a relationship of paths between users. Our experimental results show that our improved greedy algorithm achieves better running time comparing with the improvement of with matching influence spread.

Keywords: Graph algorithms, influence maximization, independent cascade model, IMIP model and social networks.

I. Introduction

Recently the amount of propagation of information is steadily increased in online social networks such as Facebook and Twitter. Social network shows a important role as a medium for the spread of Influence among its members such as Opinions, ideas, information, innovation etc.

Online public networks to use as a selling platform, there are lots of investigates on how to use the spread of influence for viral marketing. One of the research problems is influence maximization (IMAX), which aims to find k seed users to maximize the spread of influence among users in social networks.

Direct Marketing takes the “word-of-mouth” effects to significantly increase the profits (Gmail, Tupperware popularization, Microsoft Origami …)

However, influence maximization is not always the most effective strategy for viral marketing, because there can be some items that are useful to only specific users. These specific users can be a few people with a common interest in a given item, some or all people in a community, or some or all users in a class. There is no limit for being specific users. For example, consider a marketer that is asked to promote a cosmetic product for women through viral marketing. For the cosmetic product, the specific users are female users who are likely to use it and male users who wish to purchase it as a gift for female users. In this case, the marketer does not need to be concerned about the other users because the cosmetic product is not useful to them. Instead, it is a better strategy to focus on maximizing the number of influenced specific users, but influence maximization has the weakness that it cannot distinguish them from the other users.

II. Related Work

F.-H. Li, C.-T. Li and M.-K. Shan [2] proposed the labelled influence maximization problem, which aims to find a set of seed nodes which can trigger the maximum spread of influence on the target customers in a labeled social network. Three algorithms have been proposed. Each user has several predefined labels before query processing. However, it is not flexible to predefine labels to each user before query processing, since a query for targets who do not share any existing label cannot be formulated.

W. Lu and L. Lakshmanan [3] author has presented classical Linear Threshold (LT) model to incorporate prices and valuations, and factor them into users' decision-making process of adopting a product. This paper showed that the Expected profit function under the proposed model maintains sub modularity under certain conditions, but no longer exhibits monotonicity, unlike the expected influence spread function. It sets a probability, however required to check all users one hundred times with different targets set.

P. Domingos and M. Richardson [4] were the first to study the influence maximization as an algorithmic problrm based on a markov random field. This paper proposed the application of data mining to viral marketing.

D. Kempe, J. Kleinberg, and E. Tardos,[5] proposed a greedy method and show that its accuracy is higher than those of other naïve methods. The focus is on more operational models from mathematical sociology and interacting particle systems that explicitly represent the step-by-step dynamics of adoption. They also show that approximation algorithms for maximizing the spread of influence in these models can be developed in a general frame work based on sub modular functions.

J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance,[6] presented an improved greedy method with the Cost-Effective Lazy Forward (CELF) selection. A general methodology for near optimal sensor placement and related problems is presented. They demonstrate that many realistic outbreak detection objectives exhibit the property of "submodularity". The author exploit submodularity to develop an efficient algorithm that scales to large problems, achieving near optimal placements, while being 700 times faster than a simple greedy algorithm.

A. Goyal, W. Lu, and L. V. Lakshmanan,[7] Presented a CLEF++ greedy method by exploiting submodularity. It is a highly optimized approach based on the CELF algorithm in order to further improve the naive greedy algorithm for influence maximization in social networks. CELF++ exploits the property of submodularity of the spread function for influence propagation models to avoid unnecessary re-computations of marginal gains incurred by CELF. It show that CELF++ works effectively and efficiently, resulting in significant improvements in terms of both running time and the average number of node look-ups.

Y. Wang, G. Cong, G. Song, and K. Xie [8] proposed a community-based greedy algorithm for mining top-K influential nodes method based on identifying influence spread in communities. The proposed algorithm encompasses two components: 1) an algorithm for detecting communities in a social network by taking into account information diffusion; and 2) a dynamic programming algorithm for selecting communities to find influential nodes.

W. Chen, C. Wang, and Y. Wang [9] focus on reducing the cost for calculating the influence spread. They propose a greedy method based on randomly generated graphs and a degree-based method wherein the largest effective degree nodes are selected as influential seeds. They also propose Prefix excluding Maximum Influence Arborescence (PMIA) heuristics where seed nodes influence the other nodes along the maximum influence path from a seed node to each node.

Q. Jiang, G. Song, C. Gao, Y. Wang, W. Si, and K. Xie [10] Presented simulated annealing-based methods that are used to escape the confinement problem of the greedy approach. In this paper, a totally different approach based on Simulated Annealing (SA) for the influence maximization problem is proposed. This is the first SA based algorithm for the problem. Additionally, two heuristic methods are proposed to accelerate the convergence process of SA, and a new method of computing influence to speed up the proposed algorithm.

III. PROPOSED SYSTEM

In our proposed system we have implemented Algorithm 1 to 6 as per IMIP model but in Algorithm 7[1]. We have changed some steps for time optimization. As in IMIP model in order to compute marginal gain (step 25-27)[1] system calls Algorithm 5 which recursively calls Algorithm 6 for each node in λ(t) where t Є ϴ*(s), hence to reduce its time complexity in proposed system it does not call algorithm 5 and 6 instead, it recomputed ϴ(t) to calculate marginal gain. As a result it produces same Influence Spread over existing approach with less time complexity.

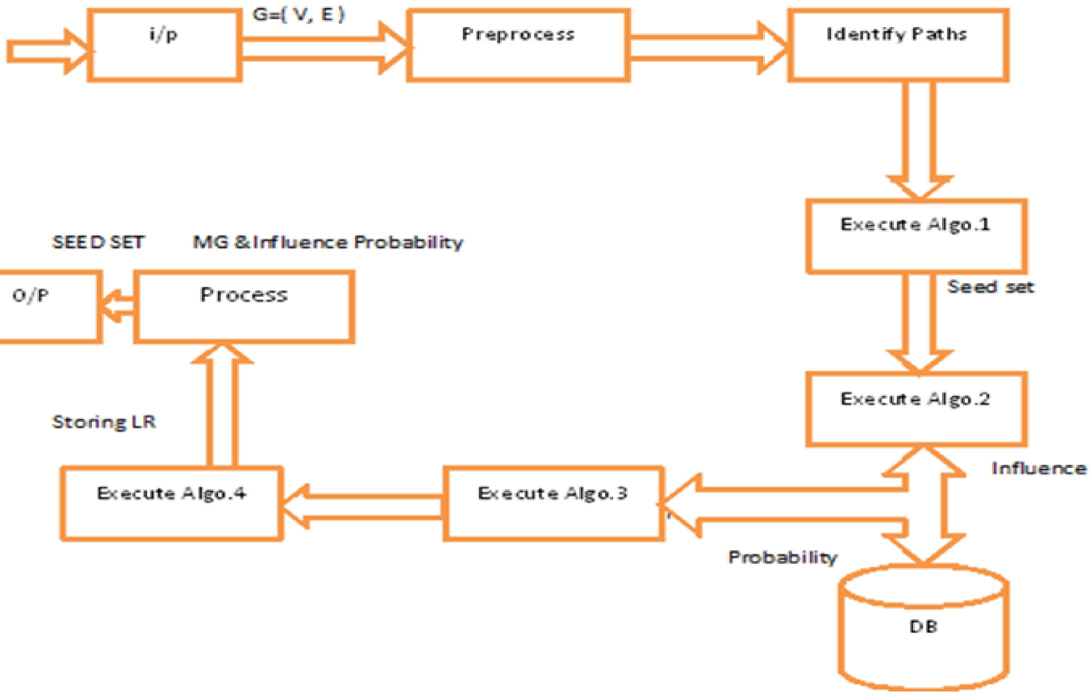

Fig 1: Proposed System Architecture

1. Obtain seed set for the given set of Target using Algorithm 1.

2. Calculate Influence probability for each target node using Algorithm 2.

3. Obtain Candidate set for Target nodes using Algorithm 3.

4. Store LR for Optimum values in Graph, Calculate $\lambda(v)$ in algorithm 4.

5. Obtain Marginal Gain of Seeds with respect t the influence probability of t.

6. Obtain Final seed set S using optimization values i.e. the set of Local influences $\lambda(v)$ & other Parameters described In Algorithm 7 in IMIP Model.[1]

Proposed Algorithm is:

Input:

G = (V, E) an input graph,

T: a set of targets,

K: the size of the output seed set

Output:

S: a seed set

1: begin

```
S=null;
// find candidates and compute u function
Candidates= findLR(G,T)  // Call Algorithm3
foreach c in Candidates
  Δσ^s_T(C) =σ*_T ({C});
for i =1 to k do
S=argmax_{c ϵC-S} Δ σ^s_T(C);
S = S U {s};
foreach c in S do
p(c)=1;
foreach t in θ (s) do
// update the influence probability of t
traverseLIT(s ,t, true);  // call Algorithm 5
if s in T then
for s´ ε λ (s)do
if s´ ε C  then
// reduce Δσ^s_T (s´)  since s is a seed
Δσ^s_T (s´) = Δσ^s_T (s´) -Δσ^s_T (s´,s)
Δσ^s_T (s´,s)  = 0;
foreach t in θ *(s) do
if t ε S or t ∉ T then
continue;
foreach l in λ (t)do
Δ σ ^s_T (l)  =  θ *(s) /// Recompute   using Algorithm 3
```

IV. Result and Analysis

A] Dataset:

For experiments, we have use Synthetic as well as Real Data sets:

TABLE I: Our Synthetic as well as Real Data Sets

Data Set	Wiki -Vote small	Real Data Set 1	Real Data Set 2
Node	5	27	52
Edge	17	100	251

1. Wiki-Vote small: we have used 5 Node & 17 Edges for Synthetic Dataset.

2. Real Data Set1: we have used 27 Node & 100 Edges for RealDataSet1

3. Real Data Set2: we have used 52 Node & 251 Edges for RealDataset2.

B] Result:

In our proposed system we have implemented Algorithm 1 to 6 as per IMIP mode [1] but in Algorithm 7 we have changed

Some steps for time & Memory optimization gives more better results.

Running Time will improve by 25% to 30%

Memory optimization will improve by 30% to 35%

Parameters consider for IMIP (Proposed) model:

Our project, we also considering α, β, h under IMIP model. We also vary each parameter.

TABLE-II: Parameter consider for IMIP (proposed) model

Parameter	Value Larger or Smaller	IS {Influence Spread)	R (Running Time)
α	Larger	Smaller	Shorter
β	Larger	Not changed	Shorter
h	h=1 h>1	Slow High	Shorter

TABLE-III : Time required for each Datasets (in msec.)

Dataset	Existing System Time in msec.	Proposed System Time in msec.
WikiVote1 small	49	39
RealDataSet1	963	221
RealDataSet2	868	800

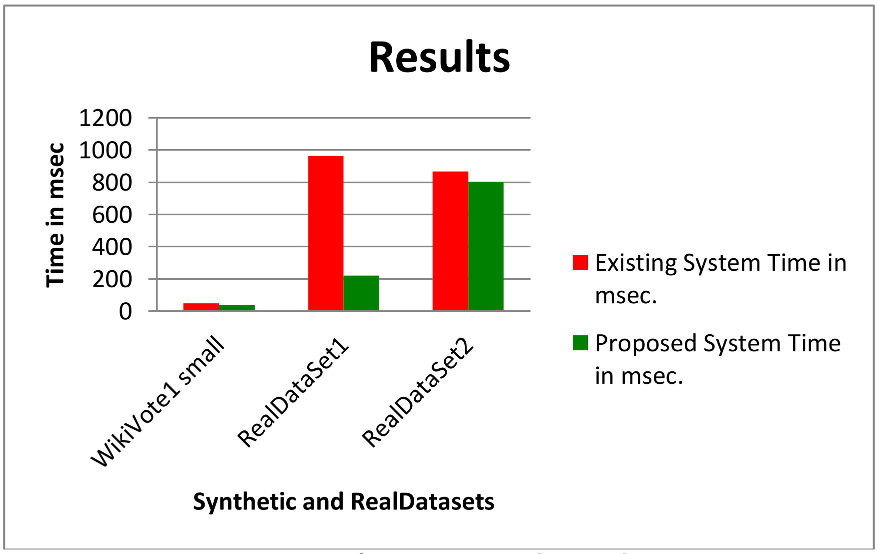

Fig.2. Comparison of Existing System and Proposed System

In our proposed system results shows that our execution time is much better than the existing system.

V. Conclusion

In our proposed system IMAX query processing to the new expectation model to reduce the calculations and number of loops iterations. & finally get the best seeds set for the Influence spread.

In the future, for IMAX query processing, we will consider more various distributions of targets such as users in the same community on the static profiles of users. Next, we will apply IMAX query processing to the Linear Threshold (LT) Model.

ACKNOWLEDGEMENT

First and foremost, I would like to thank my guide Prof. J.V.Shinde for her guidance and support. I would also like to thank to my friends for listening my ideas, asking questions and providing feedback and suggestions for improving ideas. I wish to express my sincere thanks to the Head of department, Prof. N. R. Wankhade also grateful thanks to the departmental staff members for their support.

References

1. Jong-Ryul Lee and Chin-Wan Chung, "A Query Approach for Influence Maximization on Specific Users in Social Networks", in IEEE transactions on knowledge and data engineering, vol. 27, no. 2, February 2015.
2. 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.
3. W. Lu and L. Lakshmanan, "Profit maximization over social networks," in Proc. IEEE 12th Int. Conf. Data Mining, 2012, pp. 479–488.
4. P. Domingos and M. Richardson, "Mining the network value of customers," in Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2001, pp. 57–66.
5. 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.
6. 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.
7. A. Goyal, W. Lu, and L. V. Lakshmanan, "Celf++: optimizing the greedy algorithm for influence maximization in social networks," in Proc. 20th Int. Conf. Companion World Wide Web,2011, pp. 47–48 .
8. 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.
9. W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks,"in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining,2010, pp. 1029–1038.
10. 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. 127–13.

