Learning Influence Probabilities in Social Networks

Prathamesh Jeevan Phuke, Suresh B. Rathod
Dept of Computer Engineering, Sinhgad Academy of Engineering
Kondhwa, Pune, India
prathmeshphuke@gmail.com
sbrathod.sae@sinhgad.edu

ABSTRACT: *The study of spread of data is meanly expanded in online interpersonal organizations, for example, Facebook and Twitter. To use online interpersonal organizations as a promoting stage, there are loads of examinations on the best way to use the proliferation of strength for viral advertising. One of the exploration issues is influence maximization (IMAX), which plans to discover k seed clients to amplify the spread of impact among clients in interpersonal organizations. In this work our contribution is to exhibit the ad of products according to the age of user accordingly. It is executed up being a NP-hard issue by Kempe et al. Since they approaching an mean calculation for the issue, numerous analysts have proposed different heuristic routines. Viral showcasing is one of the time signature utilizations of impact boost. In viral advertising, a capability that an advertiser needs to advance is diffused into immediate communities "by overhearing people's conversations" correspondence. From the point of view of advertising, impact augmentation gives at which point to get the most extreme benefit from every one of the clients in an informal organization through viral showcasing. In any situation, impact amplification is not consistently the excellent technique for viral showcasing, on the grounds that there can be a few things that are profitable to just distinctive clients. These distinctive clients can be a pair individuals with a typical anticipation for a given thing, some or all individuals in a everyone, or some or all clients in a class. There is no prohibition for being distinctive clients. For instance, consider an advertiser particularly approached to progress a restorative item for ladies through viral showcasing. For the corrective item, the particular clients are female clients why should probably utilize it and male clients who prospect to buy it as a disclose for female clients. For this case, the advertiser does not should be worried about alternate clients in light of the case that the restorative item is not helpful to them. Rather, it is a superior approach to concentrate on augmenting the intensity of impacted particular clients, yet impact amplification has the irregularity that it can't get them from alternate clients. The main approach for taking care of such focuses with impact boost is making a linked diagram with the objectives and executing impact expansion on the chart. On the other hand, the aftereffect about methodology ought to be off base, on the grounds that there can be a few clients who are not targets yet rather can unequivocally impact the objectives.*

Keywords: Social network influence, Adaptive seeding strategy, Stochastic sub-modular maximization

Received: 16 April 2017, Revised 19 May 2017, Accepted 26 May 2017

1. Introduction

As of late, the study of spread of data is viciously expanded in online interpersonal organizations, for example, Facebook and Twitter. To use online interpersonal organizations as a promoting generation, there are load of examination on the excellent way to use the proliferation of impact for viral advertising. One of the exploration issues is influence maximization (IMAX), which plans to discover k seed clients to distort the spread of strength among clients in interpersonal organizations. In this work our contribution is to prove the ad of products through the age of user accordingly. It is entire up being a NP-hard issue by Kempe et al. Since they approaching an saving calculation for the issue, numerous analysts have proposed contradictory heuristic routines. Viral showcasing is one of the key utilizations of impact boost. In viral advertising, a capability that an advertiser needs to advance is diffused into informal communities ”by overhearing people’s conversations” correspondence. From the point of view of advertising, impact augmentation gives at which point to gain the most excessive benefit from every one of the clients in an informal organization over viral showcasing. In any case, impact amplification is not consistently the excellent technique for viral showcasing, on the grounds that there can be a few things that are successful to just disparate clients. These distinctive clients can be a pair individuals mutually a typical expectation for a given thing, some or all individuals in a group, or some or generally clients in a class. There is no prohibition for being distinctive clients. For instance, approach an advertiser specifically approached to advance a restorative item for female through viral showcasing. For the corrective item, the particular clients are female clients why should perhaps utilize it and male clients who wish to irregular it as a present for female clients. For this position, the advertiser does not should be worried about alternate clients in light of the case that the restorative item is not successful to them. Rather, it is a superior approach to centralize on augmenting the quantity of impacted particular clients, yet impact amplification has the irregularity that it can’t get them from alternate clients. The main approach for taking care of a well known focuses with impact boost is making a undivided diagram mutually the objectives and executing impact expansion on the chart. On the other hand, the after impact of this methodology ought to be off base, on the grounds that there can be a few clients who are not targets but rather can unequivocally impact the objectives.

2. Literature Survey

In this area, initially specify the notations utilized in this paper, examine some safe primitives utilized in our secure deduplication.

2.1 Adaptive Influence Maximization in Dynamic Social Networks

Author: Guangmo Tong, Weili Wu, Shaojie Tang and Ding-Zhu Du.

For the purpose of propagating information and ideas over a social network, a seeding business aims to find a small set of seed users that are able to maximize the spread of the request, which is termed power maximization problem. Despite a great number of works have perfected this problem, the current seeding strategies are needed to the models that cannot completely capture the characteristics of real-world social networks. In circumstance, due to high-speed data exhibit and large society of participants, the diffusion processes in real-world social networks have multiple aspects of uncertainness.

Limitation: It does not distinguish exact users from others, eventually if some items can be only convenient for the specific users.

Advantage: Maximizing the influence on the persistent users. It can finally distinguish specific users from others.

2.2 Efficient Influence Maximization in Social Networks

Authors: W. Chen, Y. Wang and S. Yang.

Influence maximization is the problem of result a small subset of nodes (seed nodes) in a social network that could maximize the spread of influence. In this paper, we study the effective influence maximization from two matching directions. One is to improve the original greedy algorithm of and its improvement to further Reduce its running time, and the second is to propose new degree discount heuristics that improves influence spread.

Limitations: Influence maximization, would be of interest to many companies as well as individuals that want to market their products, solutions, and progressive suggestions through the powerful word-of-mouth effect called viral marketing. Online social networks provide great opportunities to handle this problem, because they are connecting a huge number of people and they collect a large amount of information about the social network structures and communication dynamics.

Advantage: This system study the effective influence maximization from two complementary directions. One is to improve the original greedy algorithm of [5] and its improvement to further reduce its running time, and the second is to propose new degree discount heuristics that improves influence spread. This project evaluate our algorithms by experiments on two large academic collaboration graphs obtained from the online archival database.

2.3 Influence Blocking Maximization in Social Networks under the Competitive Linear Threshold Model
Author: Xinran Hey Guojie Songy Wei Chenz Qingye Jiangx

In many real world situation, different and frequently opposite opinions, innovations, or products are competing with one to another for their social influence in a networked society. In this paper, this project study competitive influence propagation in social networks under the competitive linear threshold (CLT) model, an extension to the traditional linear threshold model. Under the CLT model, this system focus on the issue that one entity tries to block the influence propagation of its competing entity as significantly as possible by strategically selecting a number of seed nodes that could initiate its own influence propagation. This method call this problem the influence blocking maximization (IBM) problem.

Limitations: The limitation that one entity tries to block the influence propagation of its competing entity as consider as possible by strategically selecting a amount of seed nodes that could initiate its own influence propagation. The system call this problem the influence blocking maximization (IBM) problem.

Advantage: This system design an effective algorithm CLDAG, which utilizes the properties of the CLT model, to address this problem. This project carry out extensive simulations of CLDAG, the greedy algorithm, and other baseline algorithms on real-world and synthetic datasets. Our results display that CLDAG is able to provide best accuracy in par with the greedy algorithm and typically better than other algorithms, while it is two orders of magnitude faster than the greedy algorithm.

2.4 Influence Maximization in Continuous Time Diffusion Networks
Author: Manuel Gomez-Rodriguez, Bernhard Scholkopf.

The problem finding the best set of source nodes in a diffusion network that maximizes the spread of information, influence, and diseases in a limited amount of time depends significantly on the underlying sequential the dynamics of the network. However, this remains largely unexplored to date. To this end, given a network and its temporal dynamics, this system first describe how continuous time Markov chains allow us to analytically compute the average total amount of nodes reached by a diffusion process starting in a set of source nodes. System show that selecting the set of most influential supply nodes in the continuous time influence maximization problem is NP-hard and develop an efficient approximation algorithm with provable near-optimal performance. Limitations: The problem of finding the optimal set of source nodes in a diffusion network that maximizes the spread of information, influence, and diseases in a limited amount of occasion depends dramatically on the original temporal dynamics of the network. However, this still remains largely unknown to date.

Advantages: given a network and its temporal dynamics, we initial describe how continuous time Markov chains allow us to analytically compute the average total number of nodes reached by a diffusion process starting set of source nodes. This system then show that selecting the set of most influential source nodes in the continuous time influence maximization problem is NP-hard and develop an effective approximation algorithm with provable near-optimal performance.

2.5 A 61-million-person experiment in social influence and political mobilization
Author: Robert M. Bond1, Christopher J. Fariss1, Jason J. Jones2, AdamD. I. Kramer3, Cameron Marlow3, Jaime E. Settle1 and James H. Fowler1.

Human behavior is thought to spread during face-to-face social networks, but it is difficult to identify social influence results in observational studies, and it is unknown whether on-line social networks operate in the exact same way 1419. Here we report results from a randomized controlled trial of political mobilization messages delivered to 61 million Facebook users during the 2010 US congressional elections. The results show that the messages straight partial political self-expression, information seeking and real world voting behavior of millions of people.

Limitation: To measure the causal effect of social influence on-line. At the same time, there is increasing interest in the ability to use on-line social networks to study and influence real-world behavior. However, on-line social networks are also made up of

many weak-tie relationships that may not facilitate social influence, and some studies suggest that on-line communication may not be an efficient medium for influence.

Advantages: This system report results from a randomized controlled trial of political mobilization messages delivered to 61 million Facebook users during the 2010 US congressional elections. The results display that the messages directly influenced political self-expression, information seeking and real world voting behavior of millions of people. Furthermore, the messages not only influenced the users who received them but also the users friends, and friends of friends.

3. Proposed Approach Framework And Design

3.1 Architecture

This framework first formally demonstrate the dynamic free Cascade show and present the idea of versatile seeding procedure. At that point, in light of the proposed display, this framework demonstrate that a straightforward avaricious versatile seeding technique finds a successful arrangement with a provable execution ensure. Other than the covetous calculation, an effective heuristic calculation is accommodated better adaptability. Broad investigations have been performed on both this present reality systems and engineered control law systems. The outcomes in this show the prevalence of the versatile seeding methodologies over other gauge techniques.

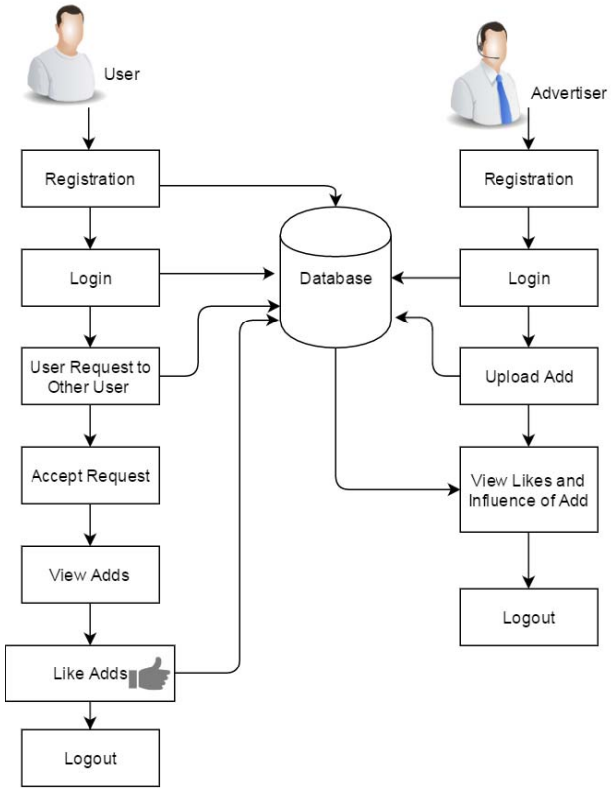

Figure 1. Flow Diagram

The flow of the figure is as follows:

• First the user signups into the system.

• User gets logged in to the system if he/she is valid user.

• There are numerous ads that are uploaded by advertiser in our system.

• Then user may like or share any advertisement if he/she likes it.

• Accordingly count of influence will be recorded using adaptive seeding strategies.

• Spread of influence will be measured by the influence count of particular advertisements.

• Thus user who liked the advertisement uploaded by a particular advertiser will be targeted on priority for future marketing.

• CD model and direct influence models are responsible for maintaining such influence count.

3.2 Seeding Strategies

• **Greedy:** This is the state-of-the art non-adaptive seeding strategy proposed in [4]. In this nodes are selected by using hill climbing algorithm before diffusion process. When implementing DIC model, we fix the propagation probability by its mean as the real propagation probabilities are not available in DIC model before diffusion process. In each seeding step, we select the node that is able to maximize the marginal prot on the observed events.

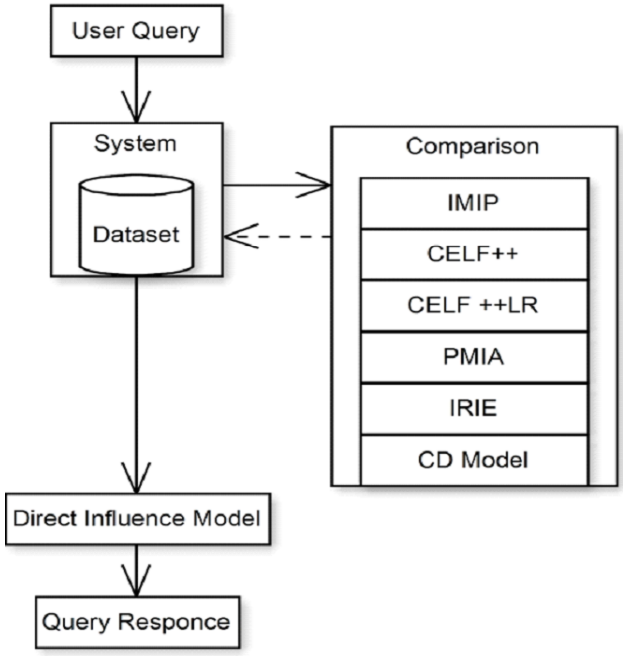

Figure 2. Architecture Diagram

• **Random:** This is a baseline seeding strategy where the seed nodes are selected randomly.

Assuming that the seed nodes are only selected between two spread rounds, we denote the seeding step between round i1 and round i as the ith seeding step, and the rst seeding step is executed before the process of spread. We assume that we need one round to activate the seed nodes selected in each seeding step. In this paper, we preserve step for seeding process and round for diffusion process. Basically, to design an adaptive seeding strategy we consider two problems: (1) how many budgets should we use in each seeding step and (2) which nodes to select [3].

Let S is the Whole System Consist of
S= I, P, O
I = Input
I = U, Q, D
U = User
U = u1,u2.un
Q = Query Entered by user
Q = q1, q2, q3qn

D = Dataset
P = Process:
P = IMIP, CELF, PMIA, IRIE, CD Model
IMIP: Independent maximum influence paths
CELF: Cost-Effective Lazy Forward
PMIA: Prefix excluding maximum influence arborescence

Step1: User enters the Query.

Step2: In comparison process following methods will be performed.

Step3: Independent maximum influence paths (IMIP):
We propose a new efficient expectation model for the influence spread of a seed set based on independent maximum influence paths (IMIP) among users.

Step4: CELF++: is an improved greedy algorithm exploiting Sub-modularity.

Step5: PMIA is a greedy-based algorithm based on maximum influence paths between nodes. In PMIA, parameter u is used to prune out maximum influence paths having low influence.

Step6: IRIE: is one of recent algorithms for influence maximization.

Step7: CD Model: CD is the greedy method using the CD model. The CD model is a probabilistic model based on users historical action logs. We use this method only for the experiment related to the actual influence spread.

Output: Finally the particular result will be shown to user as per his query.

3.3 Propose Work

The rapid information transmission and expansive populace of members, the dispersion forms in true interpersonal organizations have numerous parts of uncertainness. As appeared in the analyses, when considering such uncertainness, the best in class seeding techniques are critical as they neglect to follow the impact dispersion. In this paper, the framework concentrate the techniques that select seed clients in a versatile way. This framework first formally display the dynamic autonomous Cascade demonstrate and present the idea of versatile seeding technique. At that point, in view of the proposed demonstrate, this framework demonstrate that a straightforward voracious versatile seeding technique finds a powerful arrangement with a provable execution ensure. Other than the insatiable calculation, a proficient heuristic calculation is accommodated better adaptability. Broad tests have been performed on both this present reality systems and manufactured power-law systems. The outcomes thus exhibit the prevalence of the versatile seeding systems over other benchmark strategies. In the Linear Threshold Model, a client will receive another thought if the impact from its neighbors has achieved a specific limit, while in the Independent Cascade Model an adopter has a specific likelihood to persuade each of its neighbors. In view of those essential models different propelled models have been created and considered. Among the points with respect to impact dissemination, an imperative one is that how to proliferate data through an interpersonal organization successfully and proficiently. For instance, to promote new items, an organization might want to offer free specimens to an arrangement of introductory clients who will conceivably acquaint the new item with their companions. Because of cost issue, just a predetermined number of tests are accessible and in this manner this venture have a financial plan of the seed clients. A characteristic issue is that how to choose a decent arrangement of seed clients that can amplify the quantity of clients who at long last embrace the objective item. This issue is named as impact augmentation issue initially proposed in [6] in writing.

3.4 Mathematical Model

Let S is the Whole System Consist of
S= {I, P, O}
I = Input
I = {U, Q, D}
U = User

U = {u1,u2.un}
Q = Query Entered by user
Q = {q1, q2, q3qn}
D = Dataset.
P = Process:

Step1: New post (Ad) will be uploaded from advertiser site along with age category wise.

Step2: That uploaded adds will be displayed to that particular age category user. After viewing that add user can like that add and view his friends like to that ad.

Step3: The ranking will be increasing as per user like to that ad.

Step4: Advertiser can view number of likes to the particular ad.

The algorithm used in this project are as follows.
Greedy algorithm:
Algorithm 1 Greedy

1: Input: G = (V, E,FV , FE) and budget B.

2: CurrentBudget 0; A ; 3: y0 = 0; // yi is the p-realization after round i .

4: for each v in V do Sv +;

5: for i = 1 : N do

6: if (CurrentBudget¡B and no nodes can be further activated) then

7: for each v in V A do sv false;

8: while true do

9: v = argmaxvV A Sv

10: if (sv = true) then A A 1 v; break;

11: else sv = xCG(yi1) Prob[x—yi1] NGx (A 1 v)

12: CurrentBudget=CurrentBudget+1;

13: Get yi; // wait for a round of spread

14: y yN

15: Return NGy*(A)

Output: The output will be the response of the user query.

4. Practical Result And Environment

4.1 Hardware and Software Configuration

Hardware Requirements:
Processor : Pentium iv 2.6 GHz

Ram	: 512 MB DD RAM
Monitor	: 15 color
Hard disk	: 20 GB
Keyboard	: Standard 102 keys
Mouse	: 3 buttons

Software Requirements:

Front End : Java

Back End : MYSQL
Tools Used : Eclipse
Operating System : Windows XP/7.

4.2 Result of Practical Work

On this graph showing the time graph between various methods like encryption, decryption, tag generation.

Techniques	Add Likes	View Add Ranking
Greedy	4.3	3.5
Age Category	4	3.6

Figure 3. Performance of File Size with Time

4.3 Contribution Part

Our contribution part is, we are targeting ads to users by dividing them on the basis of age of a user. By implementing this we can target particular users based on their interests, likes and age category as well.

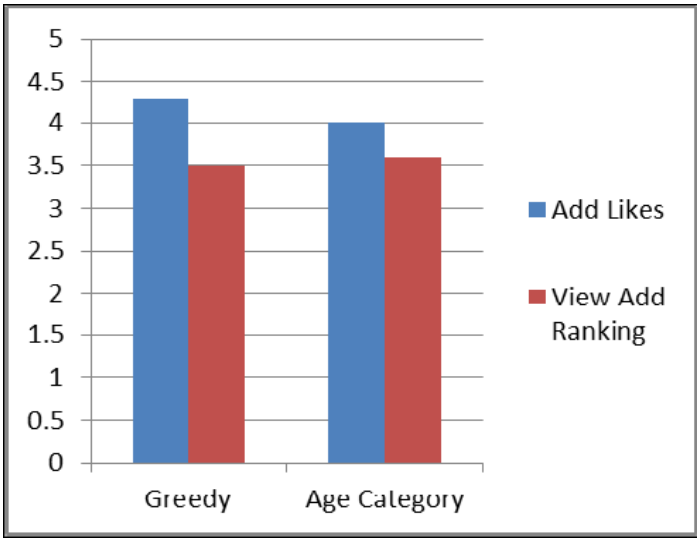

Figure 4. Graph of File Size with Time

5. Conclusion

The lack of influence expansion is that it doesn't recognize distant users from others, despite of the fact that more or less things can be helpful for the distant clients. For such things, it is a superior system to concentrate on expanding the effort on the particular clients. In this paper this system have considered the problem that how to maximize the spread of influence in dynamic social networks. The proposed DIC model is able to capture the dynamic aspects of a actual social network and the uncertainness of the diffusion approach. In the DIC model, a certain node can be seeded for more than as soon as the propagation probability between two users varies following a specific distribution. Based on the DIC model, technique formulate the adaptive seeding strategies by introducing the concept of seeding pattern. This system contribution is to upload include as per age category. In which user age category is divided in different class. That adds will be visible only that particular age group user.

Acknowledgment

With immense pleasure, I publishing this paper as a part of the curriculum of M.E. Computer Engineering. It gives us proud

progress of paper work. We would also like to thank all the Staff Members of Computer Engineering Department, Management, friends and family members, Who have directly or indirectly guided and helped us for the preparation of this paper and gives us an unending support right from the stage the idea was conceived.

References

[1] Guangmo, Tong., Weili, Wu., Shaojie, Tang., Ding-Zhu, Du. (2016). Adaptive Influence Maximization in Dynamic Social Networks, 1063-6692 IEEE.

[2] Mahajan, V., Muller, E., Bass, F. M. (1990). New product diffusion models in marketing: A review and directions for research, *J. Marketing,* 54 (1) 126.

[3] Goldenberg, J., Libai, B., Muller, E. (2001). Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata, *Acad. Marketing Sci. Rev.*, 2001 (9) 119.

[4] Bond et al, R. M. (2012). A 61-million-person experiment in social influence and political mobilization, *Nature,* 489 (7415) p. 295-298.

[5] Fan et al, L. (2013). Least cost rumor blocking in social networks, *In*: Proc. IEEE 33rd Int. Conf. Distrib. Comput. Syst. (ICDCS), p. 540-549.

[6] Kempe, D., Kleinberg, J., Tardos. (2003). Maximizing the spread of influence through a social network, *In*: Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, p. 137-146.

[7] Domingos, P., Richardson, M. (2001). Mining the network value of customers, *In*: Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, , p. 57-66.

[8] Goldenberg, J., Libai, B., Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth, *Marketing Lett*., 12 (3) 211-223.

[9] Chen, W., Wang, C., Wang, Y. (2010). Scalable influence maximization for prevalent viral marketing in large-scale social networks, *In:* Proc 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining,, p. 1029-1038.

[10] Chen, W., Wang, Y., and Yang, S. (2009). Efficient influence maximization in social networks, *In*: Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, p. 199-208.

[11] Chen, N. (2009). On the approximability of influence in social networks, *SIAM J. Discrete Math*., 23 (3) 1400-1415. .