INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2517 - 2523
RECEIVED : 25.10.2017 PUBLISHED : 01.12.2017
Dec 01, 2017

A SURVEY ON SEARCHING SPECIFIC USERS USING QUERY PROCESSING IN SOCIAL NETWORKS

[1]Augustian Issac, M.Narendran[2], Kovvuri Sri Akhil Reddy[3], ChikakollaSaiCharan[4] ,Gobburi Srikanth[5]

[1]Asst. Professor, Dept. of CSE, SRM University, Chennai, India
[2]Assistant Professor, Department of Computer Science and Engineering, SRM University, Ramapuram, Chennai-89, India
[3]B.Tech Student, Department of Computer Science and Engineering, SRM University, Ramapuram, Chennai-89, India
[4]B.Tech Student, Department of Computer Science and Engineering, SRM University, Ramapuram, Chennai-89, India
[5]B.Tech Student, Department of Computer Science and Engineering, SRM University, Ramapuram, Chennai-89, India.

Abstract:
Objectives: To increase the profit of virtual marketing in social networks influence maximization is operated with sub modular functions. It is easier to find the specific users by this operation. It affects on time delay and system will be efficient and accurate.

Methods/Statistical analysis: Basically Influence maximization is applied in social networks depends on two methods CELF, PMIA. Even though these methods give better results by the heuristics and greedy algorithm, it leads to time delay, on the other side cost of system gets increased.

Findings: The consequences that are facing from the existing methods are recovered by implementing the methods like Decision Tree, K-means methods .These methods mainly used to apply the sub modular functions in the process Due to this the data flow of the system gets changed, It was affected by the query processing which separates the searching process in small sub process.

Application/Improvements: The system is improved by formulating query processing to maximize the influence on specific users in social networks by using the IMIP model based on independence between paths .It is faster and higher accuracy.

Keywords: Social Networks, Graph Algorithms, Influence Maximization, Sub Modular Functions Independent Cascade Model.

1. INTRODUCTION

Social networking is the common way of communication Social networking is the grouping of users into specific groups, like small student communities or a neighbourhood subdivision. Although social networking is possible in person, mostly in the offices, colleges, and schools, it is most popular online. This is because unlike most schools, universities ... or offices, the internet is consists millions of individuals who are looking to meet other people.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2517 - 2523
RECEIVED : 25.10.2017 PUBLISHED : 01.12.2017
Dec 01, 2017

Social network is the maps and measures of relationships and flows between people, groups, organizations, computers, and websites and connected other datasets. The nodes in the network are the people and groups while the links represent relationships between the users. Social network provides both a visual and a mathematical analysis of human relationships.

The amount of propagation of information in online social networks such as Facebook and Twitter is steadily increased. To use online social networks, there are lots of researches on how to use the propagation of information for viral marketing. One of the problem in research is influence maximization (IMAX), which aims to find k seed users to maximize the spread of influence among users in social networks. As greedy algorithm is proposed for the problem, many researchers have proposed various different methods.

Viral marketing is one of the key implementation of influence maximization. From the perspective of marketing, influence maximization provides how to get the maximize profit from the users in a social network. Influence maximization is not high effective always to find the specific users. These specific users can be few people with a common interest in a given task, some or all people in a community or in a class. There is no limit for the specific users it is a better on focusing the influenced specific users, it results for the accurate approach.

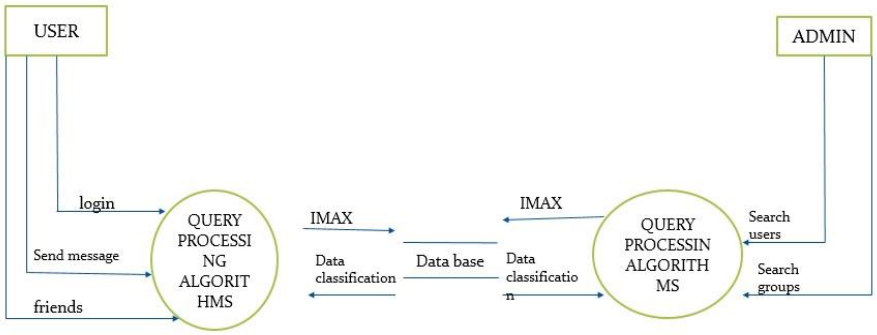

Figure 1: 0 Level Data Flow Diagram

Based on this motivation of finding the specific users, to find the specific users influence maximization is to be query processed, a query contain some labels to specify users who wants to influence other users. It is not mandatory to have the fixed labels before query processing. In another research we can be applied to influence a specific part of a social network. A variation of influence maximization which separates being influenced and adopting in item for profit maximization in this if a user is influenced by other user for an item then the user adopts it with some probability, it needs to check all the users more than one hundred times when we have one hundred items with different sets of targets to overcome the weakness of influence maximization and to provide the flexibility, we formulate IMAX is query has query processing, in IMAX query processing social networks represented by a graph where a node represents an user and an edge represents the relationship between two users such as friendship. Each edge (a, b) has a probability

that a influences b. With such probability on edges the propagation of information is designed by (IC) independent cascade model. In the IC model, user has one time chance to influence an uninfluenced at time t+1 when u is influenced at time t. If u fails to influence b, there is no second chance for a to influence b, where b can influenced by another user c, when there is an edge from c to b and c is influenced. In an IC model, an IMAX query consists of set of seed node size k and target node set T, and it ask k seed users to maximize the number of influenced users among the users operated in the query. The number of influenced users is measured by the expected number of influenced users.

The objective of this to increase the efficiency of searching users in social networks instead of searching through all the data sets by influencing one user to another user, the process divides the problem into sub modular form. It will be easy to find optimal seed node in the social network. The CELF cost effective less forward factor gets minimize as the searching efficiency to find the users is increased. In [1] authors formulate IMAX query processing to maximize the influence on specific users in social networks. Since IMAX query processing is NP-hard and calculating its objective function is #P-hard, we focus on how to approximate optimal seeds efficiently. To approximate the value of the objective function, we propose the IMIP model based on independence between paths. To process an IMAX query efficiently, extracting candidates for optimal seeds is proposed and the fast greedy-based approximation using the IMIP model. experimentally demonstrate that our identifying local influencing regions technique is effective and the proposed method is mostly at least an order of magnitude faster than PMIA and IRIE with similar accuracy In addition, the proposed method is mostly six orders of magnitude faster than CELF++ and the identifying local influencing regions technique makes CELF++ about 3.2 times faster while achieving high accuracy. The[2] process models by which ideas to propagate influence in social networks have been verified through many of the domains along with the diffusion of medical and innovation ,the sudden various strategies adopting into the settings of game theory and effect of the new product promotions . Basic algorithm problems are proposed for the process in social networks , in this process the user is requested to get new product innovatively . The objective of this is to get large no of items by the targeted users . The optimization problem is used to select the high influence users. The framework analysis is used to operate sub modular functions by approach of natural greedy concepts . It also proved by experiments on large networks , heuristics , degree of centrality. Introduced[3] influence maximization problem is to find seed users which increase the influence spread in a social network. In virtual marketing the seed user is used to get the better profits through the effect of word-of-mouth. In real cases many marketers commonly choose the specific type of customers. The influence maximization originally considered specific costumers but not their information , the labeled influence maximization is proposed its objective is to find the seed users which send the influence spread to the specific customers in the labeled social networking. To solve these problems three algorithms are introduced ,first is to improve based on greedy algorithm for specific customers. second is based on novel algorithm it covers maximum proximities of the users to find specific users . By using this owners of those sites can plan and calculate the strategy to advertise products. Recently[4] the study of propagation of influence has increased . In three dimensions are identified , first is the number of seed users which are activate initially ,second is to expect number of users activate at the end and third is the time taken to propagate the influence . By considering two dimensions the third dimension can be optimized. In this time left and covering maximum users comes under influence maximization and problems are optimized alternatively that are motivated by resources and time naturally. In the problem

selecting target users and threshold value h and to find the minimum size of seed users set are activated. IMAX[5], is the problem of finding a small set of seed users in a social network that increases the spread of influence by using influence cascade models. The scalability of IMAX is a main factor for enhancing viral marketing in large scale oriented online social networks. Prior solutions, such as the greedy algorithm and its improvements are slow and also not scalable, where remaining heuristic algorithms doesn't provide consistently high performance on spread of influence . In this paper, a new heuristic algorithm is designed that scalable to millions of nodes and edges in our experiments easily. Our algorithm in this has a simple tunable parameter for the users to maintain the balance between the running time and the spread of influence of the algorithm. The results from extensive simulations on many real-world and synthetic networks shows that at present the algorithm is the best scalable solution to the influence maximization problem: The algorithm measures beyond million-sized graphs where the greedy algorithm leads to infeasible, and around all size ranges, our algorithm performs consistently well in influence spread—it is always the best and in most cases it significantly outperforms all other remaining scalable heuristics . In[6] a water distribution network, we place sensors to detect contaminants quickly? Or else, the blogs should that we read to avoid missing important stories? It seems to be a different problems having common structure: Outbreak detection can be model as selecting users in a network, in order to detect the spreading of information as quickly as possible. In this a common methodology for near optimal sensor and related problems. It shows that many realistic outbreak detection objectives like detection likelihood exhibit the qualities of "sub modularity ". The sub modularity is exploit to develop an efficient algorithm that measures large problems, gets near optimal placements, where it is 700 times faster than a greedy algorithm. We also extracts online packs on the quality of the placements obtained by any algorithm. Our costs measures our approach on several large scale real-world problems, with a model of a water distribution network from real blog data. The achieved sensor placements are near optimal, providing a constant fraction value of the optimal solution. We show that the approach measures, achieving speedups and savings in storage of many orders of magnitude. We a show how the way leads to deeper in both applications, answers multi criteria trade-off, cost-sensitivity and generalization questions.

2. What is CELF and PMIA

A. CELF

It is the cost effective lazy forward used to improve greedy method by exploiting the sub modularity, a community based greedy method is proposed based on identifying the influence spread on communities. Greedy method is proposed on randomly generated graphs and a degree based method where the effective degree nodes as influence seeds .based on this sometimes cost factor may varies goes on increase and decrease. It affects the efficiency and influence of users on other users.

B. PMIA

It is prefix excluding maximum influence arborescence , in this heuristics where seed node influence the other nodes along the maximum influence path from a seed node to the other nodes, if the maximum influence from seed node s to node b includes another seed node in their greedy based algorithm, then their algorithm calculates next maximum influence path from s to b which does not include other set ,

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2517 - 2523
RECEIVED : 25.10.2017 PUBLISHED : 01.12.2017
Dec 01, 2017

the PMIA heuristics are inefficient for IMAX query processing. By the above both problems the searching of specific users in the social networks will be affected. It doesn't allow to get the optimum results.

3. PROPOSED WORK

The proposed work with the techniques like Decision tree, k- means , IMAX to improve the searching experience to find the specific users and the users performing the specific tasks.

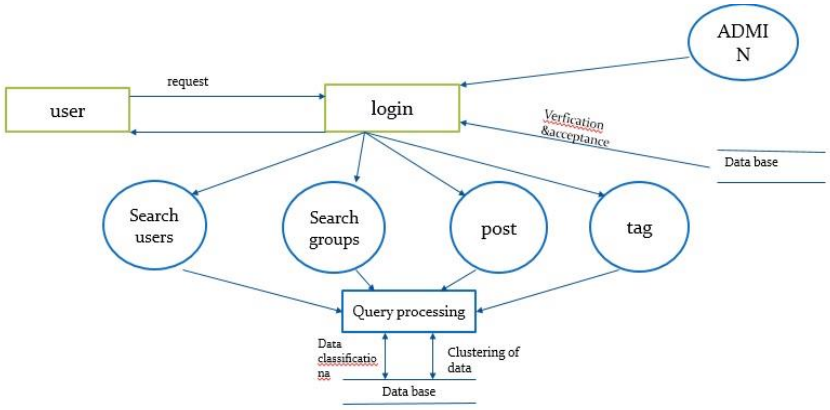

Figure 2: 1 Level Data Flow Diagram

A.DECISION TREE

Decision tree creates data classification or regression models in the form of a tree structure. It divides a dataset into smaller subsets at the same time a decision tree is incrementally developed. The final result of decision is a tree with decision nodes and leaf nodes. The first and primary decision node in a tree is the best predictor of the problem called root node. Decision trees can handle both categorical and numerical..

B.K-MEANS ALGORITHM

K-means clustering is a method of quantization, originally from signal processing, that is popular for cluster analysis in data mining process. It aims to partition n observations into k clusters in which each observation belongs to the cluster item with the nearest mean value, working as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. Clustering is the process of partitioning a group of data points into a small number of clusters. For particular time, the items in a supermarket are clustered in categories (butter, cheese and milk are grouped in dairy based products). Of course this is a qualitative based kind of partitioning. A quantitative approach would be to calculate certain qualities of the products, say percentage of milk and others, and products with high percentage of milk would be grouped together.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2517 - 2523
RECEIVED : 25.10.2017 PUBLISHED : 01.12.2017
Dec 01, 2017

C.IMAX

Influence maximization is a method of searching users in social networks .It is the common method in all the social networks to find the users ,tasks , groups ,lists, events , specific posts. In this method the process will be done basis on the spread of influence . When users in the social network are connected in the form of a graph . In that graph the users are represented as nodes and relationship between them is represented by edges . To find a specific user from the root node of the particular data graph will influence the other user depending on the relevant information of the target user.

D.QUERY PROCESSING

It is a 3 step process that transforms the high level query related problem into a equivalent and more efficient lower level query based problems. First it checks the syntax , verify relations and translate the query into an equivalent expression. After that it generates an optimal evaluation plan for a query plan and finally the query execution engine takes an evaluation plan , and returns the answers to query . It mainly deals with the data base related problems.

4. DATA FLOW DIAGRAM

A. 0 LEVEL:

The level 0 DFD Figure 1, also called a basic system model or context diagram represents the entire software element as a single bubble with input and output element of data indicated by incoming and outgoing arrows, respectively. Both the user and admin process the specific tasks depending on the parameters of IMAX,data classification , clustering.

B. 1 LEVEL:

This level of data flow diagram Figure2 provide more detailed structure. It provides a detailed view of requirements and flow of data from 1 bubble to another. By the above data flow diagrams and techniques are used to overcome the disadvantages and problems that are facing in the present process to find the users in the social networks.

5. CONCLUSION AND FUTURE WORKS

In this paper the IMAX algorithm will be query processed . Every problem will be implemented with the sub modular functions .Each problem divides into a sub smaller parts .To find the specific users in the social networks , one user will influenced by the other user which are related to the same data sets no need to influenced by users of all the datasets. By applying the query processed IMAX algorithm , it used to optimal seed efficiently. As the user is specified from the related data sets , the efficiency of finding the users gets increased . Therefore while coming on the basis of CELF the cost effecting in the social networks gets decreased . And in another direction as the problem is applied by the sub modular functions it is divided into sub parts where the data classification and clustering happens .Therefore on the basis of PMIA the user can be specified optimally no heuristics need to be applied. It minimizes the spread of influence through all the data sets.In future beyond the searching users by influencing among other users the privacy of the

specific individual users will be refined , secure maintenance of the groups in the social networks will be implemented . Efficiency to do the tasks of the user and admin will improved.

6. REFERENCES

To refer a research article:

1. Jong-Ryul Lee and Chin-Wan Chung, "A Query Approach for Influence Maximization on Specific Users in Social Networks," in Proc. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 2, FEBRUARY 2015,pp.340-353.
2. C. Asavathiratham, S. Roy, B. Lesieutre, G. Verghese. The Influence Model. *IEEE Control Systems*, Dec. 2001
3. Amit Goyal ,Francesco Bonchi , Laks V.S. Lakshmanan and Suresh Venkatasubramanian, "On minimizing budget and time in influence propagation over social networks" in a community . Soc . Netw. Anal. Min. DOI 10.1007/s13278-012-0062-z.
4. G. Nemhauser, L. Wolsey, M. Fisher. An analysis of the approximations for maximizing submodular set functions.Mathematical Programming, 14(1978), 265–294.
5. H. Peyton Young. The Diffusion of Innovations in Social Networks. Santa Fe Institute Working Paper 02-04-018(2002).

To Refer A Publication Of Proceedings:

6. 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.