Viral Marketing in Social Network Using Data Mining

Shalini Sharma*,Vishal Shrivastava**
*M.Tech. Scholar, Arya College of Engg. & I.T, Jaipur (Raj.)
**Associate Proffessor(Dept. of CSE), Arya College of Engg. & I.T, Jaipur (Raj.)
**vishal500371@yahoo.co.in, *Shalini.sharma81987@gmail.com

***Abstract*- Data mining is a process of analyzing data from different perspectives and summarizing it into useful information .The proposed work focus on clustering application which is "Collaborative Data mining "which mines different subgroups from some networks in the form of graph such as a social network.**

In this work using social network, two type of clusters are developed using data mining algorithms, one having strong tie and other having a weak tie between members and then the cluster with strong tie is used for discovering the highly influential node for viral advertising using target marketing to increase the profit with less advertising expense.

***Keywords*- Data mining, Collaborative mining, Social network, viral marketing**

*****

I. INTRODUCTION

Data-mining as a truly interdisciplinary subject, can be defined in many different ways. Even the term data mining does not really present all the major components in the figure.

Many people treat data mining as a synonym for another population used term KDD (knowledge discovery device) from data, While other view data mining as merely an essential step in process of KDD.

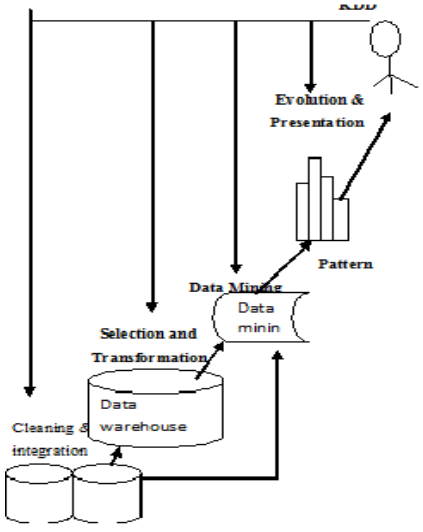

Figure 1-Data Mining Process

1. **Data Cleaning-** To remove noise and inconsistent data.
2. **Data integration**-Where multiple data sources may be combined.
3. **Data Transformation**-Where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operation.
4. **Data Mining-** An essential process where intelligent methods are applied to extract data patterns.
5. **Pattern Evolution-** So indentify **t**he truly interesting patterns representing knowledge based on interestingness.
6. **Knowledge Representations**-Where visualization knowledge representation techniques are used to present mixed knowledge to users.

Analysis and discovery of useful information from World Wide Web poses a phenomenal challenge to the researchers in this area. Such a phenomena of retrieving valuable information by adopting data mining techniques is called Web mining. . Web usage mining refers to the discovery of user access patterns from Web usage logs. Web content mining aims to extract/mine

useful information or knowledge from Web page contents [4].

A. Data mining techniques:-

It is common to describe knowledge discovered during data mining as:-

- Association Rules
- Classification hierarchies
- Clustering

B. Applications of data mining:-Data mining tools take data and construct a representation of reality in the form of a model. The resulting model describes patterns and relationships present in the data *Discovery*—the process of looking in a database to find hidden patterns without a predetermined idea or hypothesis about what the patterns may be.

Predictive Modeling- the process of taking patterns discovered from the database and using them to predict the future.

Forensic Analysis- the process of applying the extracted patterns to find anomalous or unusual data elements.

Retail-Through the use of store-branded credit cards and point-of-sale systems, retailers can keep detailed records of every shopping transaction. This enables them to better understand their various customer segments [3].

Banking-Banks can utilize knowledge discovery for various applications such as Card marketing.

II. LITERATURE SURVEY

For product promotion, event invitations, query responses and newsletters, Email provides an all-in-one solution. Convenient, highly targeted and most importantly, cost-effective, E-Marketing System is an easy choice [6].

Benefits of E-Marketing include:

- Speed: messages are delivered straight to the recipients' inboxes, instantly.
- Reach and Penetration: overcomes geographical parameters that exist with other communication methods.
- Ease and Efficiency: messages can be distributed to multiple recipients at the click of the mouse.
- Low Cost: requires minimal investment to set up an appropriate technical system.
- Targeted: allows you to target specific recipient groups and reach a defined, engaged audience [6].

E-Marketing Strategy- It assists start-up and established companies in building their customer acquisition and retention strategies, it helps to:

- Identify the best target market segments,
- Develop the right marketing programs,
- Attract and retain profitable customers, and
- Put in place the processes and people needed to have a productive and cost-effective marketing team.

Mining Knowledge-Sharing Sites for Viral Marketing-

In many markets, customers are strongly influenced by the opinions of their peers. Viral marketing takes advantage of this to inexpensively promote a product by marketing primarily to those with the strongest influence in the market. The use of relationships between people makes viral marketing potentially more profitable than direct marketing. Further, people typically trust and act on recommendations from friends more than from the company selling the product.

Online Social Networks-

An online social network comprises several main features, which allow for users to: (a) Maintain a personal profile, (b) link the personal profile to the profile of others and (c) provide latter access to the individual one-level social network (without any mediator) numerous users are organized in online social networks.

Nodes -These are the individual actors or users within the networks.

Links- These are the relationships between the actors. The strength of links or ties, as shown in fig. Between nodes in a real world social network can be of two types: strong ties or weak ties.

1. Actors with strong ties usually have some sort of common ground on which they establish their relationships and thus often a subgroup

2. The weak ties are the relationships between members of different groups. They are utilized rarely therefore don't need a lot of management.

Social Network Mining Techniques

Content- based Approach- Content-based approaches select target customers whose interests have a high degree of similarity to the product's content profile.

Collaborative Approach -It is also called structure mining. It looks for relevance among users by observing their ratings assigned to products in a training set of limited size.

Usage Mining-The purpose of the usage mining is to mine the usage of the network to know the connectedness and activeness of that network. It can be used to identify the interaction frequencies of each user with another user to identify the strength of the links.

III. Problem Statement

The proposed work tries to overcome the limitations which are mentioned above. Trying to reduce a cost of advertisement for a product, by finding out the most influential node with the help of proposed two algorithms, first is "Cluster Mining " and second is "Influence Mining ".

Assumptions -

A1- Database of node connectedness is available

A2- Database of node relationship is available.

Proposed Solution- The proposed work, considers a small social network as an example and then-

- First provides weights to all the edges as per relationships between nodes.
- Then find strong cluster and a weak cluster by using proposed algorithm.
- After that finds out mostly influential node which could later be used as target node for acheving target marketing.

Algorithm- suppose we have a graph, which is belongs to social network. A graph G(V,E,W) , where V is set of vertices , E set of edges and W set of weights .

Algorithm 1: Clustering Mining:-

Input- S, Social network Graph (Weighted graph)

Output- G1 and G2 clusters;

G1- Weak cluster;

G2- Strong cluster;

1. Set Gu ←Ø;
2. Select node N, in a graph S

If weight<7 node belongs to the G1 cluster;

Else

Node belongs to the G2 cluster;

3. End if;
4. Return G1 and G2;

In the first algorithm, which having a input S(Social network), finally result in two different kind list of clusters.

Algorithm 2: Influence Mining:-

Input: - G2- Strong Cluster

Output: - Gu node with maximum connectedness in `G2` cluster

1. G2:- N; (number of node)
2. Gu=Ø;
3. N=1; nth node is largest connected node then it is the most influence node;

Cu=n;

Else

N = N+1;

Repeat step 2 until G2=Ø;

G2=G2-nth node;

4. End if;

In this algorithm, Input is Cluster G2 which is a collection of strong tie nodes. And resulted is the most Influential node.

Result and Analysis:- Consider a network example as shown in figure 2.Subgroups are formed on the given graph first without using weights as done by Wan-Shiou et. al. which was based on unweighted undirected network, and then by putting weights on the edges of the graph and the two proposed algorithms are executed.

After executing the first algorithm, two types of subgroups or clusters are formed one representing the weak tie and other representing the strong tie, known as weak and strong clusters respectively.

Example: -

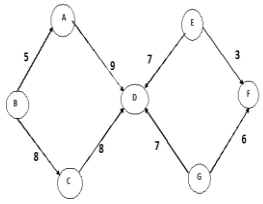

Figure -Weighted directed graph

Then the second algorithm is executed and a influential node is discovered in the strong cluster. Finally the difference in result is analyzed.

1. WEAK CLUSTER TABLE (G1)

Node	Edge	Weight
B	BA	5
E	EF	3
G	GF	6

Table1 - This table shows weak cluster`s record which is calculated by algo.

2. STRONG CLUSTER TABLE (G2)

Node	Edge	Weight
A	AD	9
B	BC	8
D	DE	7
D	DG	7

Table-2 shows the number of nodes in the strong cluster, along with the nodes and relationship with corresponding weights after executing the proposed algorithms for the network example as shown in figure (b) this table shows strong cluster table which is calculated by algo. And lastly find out most

If we consider comparison between directed or undirected weighed graph which is like this:-

- Firstly we can`t get two different kind of clusters. And in the case of directed graph, got a two different kind of clusters. One is strong tie and secondly weakly tie.
- Another difference is, in the directed graph put on different weights according its relationships.
- Lastly, In the case of directed graph, focused on one node or person and we can use target marketing for focused node.

IV. CONCLUSION AND FUTURE SCOPE

Finally it can be said that by, use an example of social network, and mining most influential node which is node with high connectedness. The influential node can be used for viral marketing or target advertising. Thus by using his/her influence the revenue can be increased with less expense on advertising. It helps in target advertising by advertising to the right people in right way at right time, which is the principal of good marketing.

In a future we can use this influential node for product marketing by mining the interested user with the help of content mining. If we having a most influential node then we also knew about all relationships about person to person it means all common people follower of mostly popular people similarly this thing happen with most influential node which explore viral marketing and desired product will be popular through his influence.

V. REFRENCES

[1] 2009, Association for the Advancement of Artificial Intelligence(www.aaai.org)Nitin Agarwal, Huan Liu, Sudheendra Murthy, Arunabha Sen, and Xufei Wang- A Social Identity Approach to Identify Familiar Strangers in a Social Network.

[2] 2002 ACM, Matthew Richardson and Pedro Domingos- Mining Knowledge-Sharing Sites for Viral Marketing.

[3] 2009, Sudipto Das Ömer E˘gecio˘glu Amr El Abbadi- Anonymizing Edge-Weighted Social Network Graphs.

[4] ISSN: 0976-5166, Vol. 1 No. 4 301-305, Mrs. Bharati M. Ramageri- Data Mining Techniques and Applications.

[5] WWW. Prenhall.com/Koter- Marketing Management.

[6]Www. Exa.com.av/Sevices/E-marketing.

[7] Han Kamber Pei (Data Mining).