International Journal of Computer Trends and Technology- volume4Issue2- 2013

An Application of Collaborative Data Mining in Social Network for Viral Marketing

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.)

***Abstract*- Data mining is a process which finds useful patterns from large amount of data. Data mining having various algorithms and techniques. The proposed work focus on clustering application "Collaborative data mining" which aims at mining the subgroups from some networks in the form of graphs, such as social networks. The nodes belonging to the subgroup having strong tie have most of the features in common.**

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.

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

1)Association Rules:- It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories.

2) Classification hierarchies- It is a two-step process. In the first, a model is constructed to describe the characteristics of a set of data classes or concepts. In the second step, the model is used to predict the classes of future objects or data. There are handful techniques for classification. Classification by decision tree and Bayesian classification is another technique.

3)Clustering:-. Clustering is the process of grouping a set of physical or abstract objects into classes of similar objects .So that objects within the same cluster must be similar to some extend, also they should be dissimilar to those objects in other clusters.

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.

E-Marketing-E-Marketing complements your overall marketing strategy perfectly, and offers your advertising campaigns proven, measurable results that can take your current marketing campaign to the next level. E-Marketing offers businesses and other advertising organizations with an easily customized, efficient form of marketing to complement existing advertising strategies.

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.

Application of clustering in strategy marketing- A clustering algorithm like K-means and Kohonen can be applied in order to identify groups which are different from each other, but whose members are very similar to each other. The segmentation results obtained are very useful for marketing concerns and for improving customer services. For instance, the derived strategic segments can be used to derive some high-level business strategies and to perform tactical marketing actions, respectively.

Related work

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.

Anonymizing Edge-Weighted Social Network Graphs

Social network have millions of registered users, and each user is associated with a number of others through friendship, professional association (being members of communities), common interests, and so on. The resulting graph structures have millions of vertices (users or social actors) and edges (social associations). The semantics of the edge weights depend on the application (such as users in a social network assigning weights based on "degree of friendship", "trustworthiness", "behavior", etc.), or the property being modeled (such as detection of communities [7] or modeling network dynamics .

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.

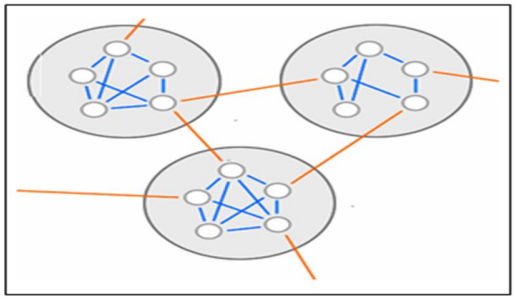

Figure 1 Strong and weak ties among the groups

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. The "nearest-neighbor" users are those that exhibit the strongest relevance to the target user. These users act as "recommendation partners" for the target user and collaborative approaches advertise products to the target user that appear in the profiles of these recommendation partners but not in the target user's own profile. This approach has demonstrated its usefulness in many applications, but it has limitations like failing to advertise newly introduced products that have yet to be rated by users, failing to advertise products to a new user who has yet to provide its rating data, and the prediction is poor if the data is sparse.

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.

Community Mining Algorithms- In context of Social Networks, the task of grouping set of nodes exhibiting similar properties or behavior is referred as Community Mining. The structure of a system always affects its functionality. For example, the topology of social networks affects the spread of infectious disease through a structured population.

A brief overview of the major algorithms proposed for community detection is given below:

Betweenness Algorithm-Betweenness is a measure for community detection in Social Networks. The betweenness algorithm focuses on the edges in the network that occurs more frequently between pairs of vertices in the graph, as compared to the other edges. The betweenness measure finds the edges which when iteratively removed decompose the network into disconnected subgraphs. There are many other algorithms such as- Max Flow Min Cut Algorithm, Modularity Algorithm.

Limitations:-

1. Unable to target potential customer and markets.
2. No proper use of user interaction and communication data.
3. All links are considered of equal weight.
4. Different types of relationships are not considered.

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- SN, Social network Graph (Weighted graph)

Output- C1 and C2 clusters;

C1- Weak cluster;

C2- Strong cluster;

1.Set C1, C2=∅;

2. Select node n, in a graph SN

If weight<7 node belongs to the C1 cluster;

Else

Node belongs to the C2 cluster;

3. End if;

4. ReturnC1 and C2;

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

Algorithm 2: Influence Mining:-

Input:- C2- Strong Cluster

Output:- Cu node with maximum connectedness in `C2` cluster

1. C2:- n; (number of node)

2. Cu=∅;

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 C2=∅;

C2=C2-nth node;

4.end if;

In this algorithm, Input is Cluster C2 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

Example: -

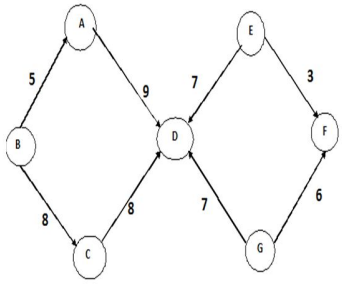

Figure2. Weighted directed graph

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

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).