Volume 1 No. 2, June 2011
International Journal of Information and Communication Technology Research

http://www.esjournals.org

A Framework for Improving Target Marketing Using Collaborative Data Mining Approach

*Vishal Shrivastava, Rajesh Boghey, Bhupendra Verma
Technocrats Institute of Technology, Bhopal, INDIA
*Vishal500371@yahoo.co.in

ABSTRACT

Marketing is one of the most prominent issues of modern trading culture. E-marketing refers to the advertisements published on World Wide Web. Advertising to the right user in the right way at the right moment is critical to the success of any e-marketing campaign. Social network can provide a link structure which can be explored to target a specific group of customers with their appropriate interest on the product rather than irrelevant popping up user's web page for the advertisement. In this paper, an approach is used for mining the interest of users from their profile and determining the groups they belong to**.**

Keywords: *Advertising, target marketing, social network, data mining techniques, collaborative mining*

1. INTRODUCTION

Marketing has already been an old subject of research. In addition, with the advent of World Wide Web, E-Commerce marketing has also been analyzed a lot. However, social network being a new concept and its structure and purpose of user also being different from the normal web user, web-marketing techniques are not much successful for it. Also, with due course of time social networks are attracting large amount of web traffic thus being a huge opportunity for marketers to grab the attention of user by advertising in the right way to the right user at right moment. [1]

In the world of intellectually aware customers, it is more effective to target a specific group of customers with their appropriate interests rather than popping up each user's web page with irrelevant advertisements and this approach is referred to target marketing. This not only helps increasing revenue but also improves customer credibility and retention. Traditional approaches in this direction include manually analyzing transaction records or user behavior and preferences, and targeting advertisements according to that observed preferences. With the advent of technology, various tools were formed to analyze such record, such as transaction record or predefined ratings and predict the interests of users.

Social Network is simply a structure consisting of people or other entities embedded in a social context, with a relationship among those people that represents interaction, collaboration, or influence between entities. Social network views social relationships in terms of nodes and links.

There are few data mining techniques that can be used to mine specific group of customers for target advertisement from any given social network. The approaches can be broadly classified into the following:

1.1 Content-based approach:

It is the process to discover useful information from text, image, audio or video data in the web. This technology has enabled ecommerce to do personalized marketing, which eventually results in higher trade volumes. It characterizes recommendable products by a set of content features, and represents user's interests by a similar feature set. It selects target customers whose interests have a high degree of similarity to the product's content profile. To establish a more accurate content profile for a product, the detailed description of a product must be parsable (e.g., as text), and a set of content features are extracted by some information-extraction or summarization techniques. This approach further evolved as Contextual mining approaches that are also applied to World Wide Web and reached to ontology's mining. These are good in understanding the individual liking of a user however do not considers the group influence. [2]

1.2 Collaborative-based approach

It is the process of using graph theory to analyze the node and connection structure of a web site. It either extracts the groups from the existing user data or uses existing groups from the network and calculates the user network value as a measure of influence of the user in the group. Thus, this approach concentrates on the network value and the group influence of users but do not considers the individual user preferences. [2]

Both the approaches have their advantages and disadvantages. If the advantages of both the approaches are combined by considering both individual preferences and group influence then it can lead to robust advertising methodology. The work considers the combination of both the approaches to earn better results and validates the approach by using network examples.

2. PROPOSED WORK

2.1 Problem Statement

Social network provides a link structure showing the relations among various people in the form of a link. Also it is found that advertising to the right user in the right way at the right moment is critical to the success of any e-marketing campaign because it eliminates the irrelevant popping up of user's web page for the advertisement. So the problem is to mine the specific group of people having good interest in the product whose influence on others in the same group can help in increasing the targeted customer's interest on the product and if the interest increases above any given threshold value then that person can be treated as a relevant customer for the advertisement.

2.2 Proposed Solution

Traditionally methods used in this direction included manually analyzing transaction records or user behavior, and preferences and targeted advertising according to that observed preferences. With the advent of new technologies, various tools were formed to analyze such record, such as transaction record or predefined ratings and predict the interests of users.

2.2.1 Assumptions

Assumption 1: A database containing customer connectedness with the weight of the connectedness indicating the strength of the link is available.

Case 1: Given a social network SN. The social network SN is a weighted directed graph SN = (V, E, W), where V is the set of nodes, and E is the set of edges and W is the set of weights. Each node V_i Є V represents a customer, and each edge (V_i, V_j, W_{ij}), where Vi, Vj Є V, represents the connectedness between V_i and V_j with weight W_{ij}. Also the weight, W_{ij} represents the influence of V_i on V_j.

Assumption 2: A database containing individual ratings indicating the individual interests of customers in specific products is provided.

Case 2: Let P = {P_1, P_2, ..., P_r} be the set of product items, and V = {V_1, V_2, ..., V_n} be the set of customers. Let I be the set of interests vector for each customer I_n= {I_{n1}, I_{n2}, I_{n3}, I_{n4}, ………., I_{nr}}, where I_{nr} is the interest of n^{th} customer in r^{th} product.

2.2.2 Algorithms

For extracting subgroups, the cohesive subgroup algorithm [3], used which is modified for weighted directed graph. Cohesive subgroup algorithm extracts subgroup based on connectedness and common neighbors. It assumes that the strength of a tie between two actors is much greater if the two actors in question have another mutual acquaintance. Cohesion measure to be calculated between numbers of nodes is defined in case 3.

Case 3: For a set S of nodes in a social network SN, the cohesion of S, denoted as Cohesion(S), is defined as

$$\frac{\bigcap \textbf{Neighbors(s)}}{\bigcup \textbf{Neighbors(s)}}$$

where Neighbors(s) denotes a node s Є S and the set of nodes directly connect to s

The cohesion will vary from 0 to 1, 0 for totally disconnected nodes with no common neighbors and 1 for totally connected nodes having all common neighbors. The prior work [3] introduced this algorithm of clustering using cohesion values for unweighted undirected graphs. Here the same algorithm is used for clustering for weighted directed graph. The subgroups are identified with some specific extent of cohesion and average interaction.

Case 4: Given a weighted directed graph, SN = (V, E, W) where V is a set of nodes, E is a set of edges and W are the set of weights. Then the average interaction for the nodes S and set of edges e Є E between the nodes in the subgroup of weighted directed graph, is defined as

$$\textbf{Avg\_interaction} = \frac{\sum (\mathbf{w}_{(i,j)})}{|\mathbf{e}_{i,j}|}$$

where i and j Є S Є V and, $w_{(i, j)}$ is the weight of edge e (i, j) Є E and $|e_{i,j}|$ is the total number of edges in the subgroup of weighted directed graph.

Case 5: A set S of nodes is said to be a cohesive subgroup in a social network SN if Cohesion(S) is no less than user-defined threshold cohesion, tc% [3] and the average interaction between the nodes is no less than threshold interaction, tw%.

Case 6: Given a social network SN, a set S of nodes in SN is maximal if S is not a subset of any other cohesive subgroup in SN.

As cohesion measure shows the connectedness of group, maximal measure shows the disconnectedness from rest of the network, and interaction measure shows the activeness and strength of connectivity, thus these all can be used to improve the quality of the subgroup derived and the further results because of this subgroup formation.

The existing algorithm is modified by introducing a threshold interaction. In this algorithm, potentially cohesive subgroups (or called candidate subgroups) of size k are constructed from joining cohesive subgroups of size *k*−1. The social network is then examined to identify cohesive subgroups of size k from the set of candidate subgroups of the same size. If a cohesive subgroup of size k is identified, all of its subset is removed from cohesive subgroups of size k-1. This procedure is executed

Volume 1 No. 2, June 2011
International Journal of Information and Communication Technology Research

http://www.esjournals.org

iteratively until no further cohesive subgroups are found. Let C_k and L_k denote the set of candidate subgroups and the set of cohesive subgroups of size k, respectively. The maximal interaction based cohesive subgroup discovery algorithm is sketched as Algorithm 1.

Algorithm 1: MiningCohesiveSubgroups (SN)

Input: Social network, SN
Output: Set of cohesive subgroups.

1. Set L_1: = all nodes in SN. [initializes a set L_1 with all nodes in social network]

2. Set n : = 1. [initializes counter]
3. Repeat steps 4 to 7 while $L_n \neq Ø$
4. Set n := n + 1.
5. Call GenerateCandidateSubgroup (L_{n-1}).
6. If Cohesion (a) ≥ tc% and Avg_Interaction ≥ tw%, then:
   Set L_n = {a} where a Є C_n.
   [End of If structure.]
7. Set L_{n-1} : = L_{n-1} – {all subsets of a | a Є L_n}
   [End of Step 3 loop.]
8. Return U L_i.

The detailed algorithm for generating candidate subgroups is sketched as Algorithm 2.

Algorithm 2: GenerateCandidateSubgroup(L)

Input: Set of cohesive subgroups

Output: Set of candidate subgroups.

1. Set C_n : = Ø.
2. Repeat steps 3 to 5 for each pair of cohesive subgroups (A, B) in L.
3. Let s_1 be the first node in A.
4. Let s_2 be the last node in B.
5. If s_1<s_2 and A-{s_1}= B-{s_2}, then:
   Set Candidate: = A U {s_2} and C_n : = C_n U Candidate
   [End of If structure.]
   [End of for loop.]
6. Return C_n.

The network example as shown in figure 1 is used having weights attached to each of the edges. Major differences were observed in the result when the algorithm is run with different threshold cohesion and different threshold interaction. The vector is used to store a graph and interaction vectors to make the calculations simple, as the main objective is to show the effect of weights on the graph and results.

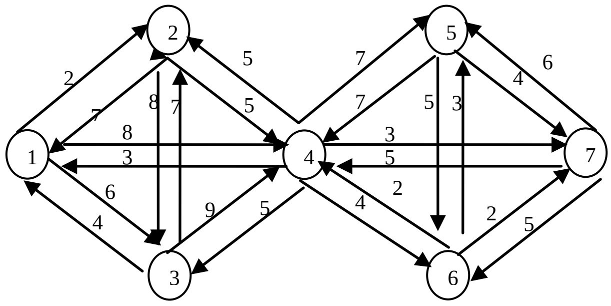

Fig. 1 Weighted directed social network

The initial individual liking of the users for each and every product normalized to the range of (0, 1) is assumed, which can be taken from any rating data for products. These initial liking values of each user are used to find out the weighted average of each user with respect to the other user's interest in the group.

Case 7: Given a weighted directed graph, SN = (V, E, W), the weighted average interest for V_j is given by

$$\textbf{wgt\_avg\_interest}(V_j) = \frac{\sum (w_{(i,j)} * I_{(i,r)})}{\sum (w_{(i,j)})}$$

where i, j Є S Є V, $w_{(i, j)}$ is the weight of edge from V_i to V_j, and $I_{(i, r)}$ is interest of V_i on product r.

Then the weighted average is used for calculating the gained in interest of the user with the group influence effect in the group. And came with new liking data, modified according to the liking data of all the groups to which the user belongs.

After achieving these updated liking values, the customers to be advertised for various products is decided based on threshold interest that is assumed to be 50%, which can be increased or decreased depending on the quality of result required. The pseudo code for selecting customers for advertising a product P_r is listed in Algorithm 3.

Algorithm 3: LikingBasedTargetedAdvertising (SN, P, V, I, S)

Input: Social network, SN in the form of weighted directed graph, a set of products, P, a set of customers, V, a set of individual interest's vector of each customer, I and a set of cohesive subgroups, S

Output: Selects customers for advertising any product P_r.

1. Repeat step 2 for each customer P_r.
2. Set Targeted_Customer(P_r) : = Ø. [Initializes a set of targeted customer.]
   [End of step 1 loop.]
3. Repeat steps 4 and 5 for each customer n.
4. Repeat step 5 for each product r.

5. Set I_{nr}: = I_{nr} + Σ_i (Cohesion (S) x wgt_avg_interest(V_n)). [where n Є S and for all the i subgroups to which the customer belongs.]
[End of step 4 loop.]
[End of step 3 loop.]
6. Repeat step 7 for each product P_r.
7. Set Targeted_Customers (P_r) : = All customers i for which $I_{ir} \geq$ ti%. [where ti% is the required threshold interest.]
[End of step 6 loop.]
8. Return Targeted_Customer (P_r).

This algorithm can be used to advertise new products by obtaining customer's interest through content mining of their profiles, and then it can be used to advertise new customers by knowing the subgroups to which they belong. Moreover, by using weighted average method of modifying existing individual interest of customers, effects of positive and negative weights can be well handled in this algorithm. This makes the algorithm quite robust and effective for weighted directed social networks. Its performance can be optimized by using more sophisticated data structure; however, the two-dimensional arrays are used for storing all vectors to maintain simplicity.

3. CONCLUSION

The work represented a robust algorithm to mine targeted customers for some specific products who can be advertised. The target network contains user profiles, connectedness having relationship between them. The subgroup is mined from the target network and the increased interest of a person due to the influence of other people in the same subgroup is calculated. The technique utilized for mining is collaborative mining. The content mining is assumed to be used for determining initial individual interest of user for some products. The relationships are considered as weights between the people by using structure mining. This algorithm does not require transaction details of customer and the customer's interest, using their initial ratings for the products, can be modified to use them.

Thus, this algorithm can be used over multiple combinations of data, can be applied to negative and positive weights, scalable and robust and uses combination of all the techniques to mine the right set of customers.

REFERENCES

[1] Armstrong G., and Kotler P., "Marketing: An Introduction", Prentice-Hall, 1999. (Upper Saddle River, NJ)

[2] Kwok-Wai Cheung, James T. Kwok, Martin H. Law, Kwok-Ching Tsui, "Mining Customer Product Ratings For Personalized Marketing", ACM, Volume 35 , Issue 2, May 2003.

[3] Wan-Shiou Yang, Jia-Ben Dia, Hung-Chi Cheng, Hsing-Tzu Lin, "Mining Social Networks for Targeted Advertising", Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06) Track 6, vol. 6, pp. 137-147, 2006.