Finding Sales Promotion and Making Decision for New Product Based on Group Analysis of Edge-Enhanced Product Networks*

Yi Huang, Jianbin Tan, and Bin Wu

Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia,
Beijing University of Posts and Telecommunications, Beijing
{huangyi,tanjianbin}@tseg.org, wubin@bupt.edu.cn

Abstract. A novel method is proposed in this paper to find the promotive relationship of products from a network point of view. Firstly, a product network is built based on the dataset of handsets' sale information collected from all outlets of a telecom operator of one province of China, with a period from Jan. 2006 to Jul. 2008. Then the edge enhanced model is applied on product network to divide all the products into several groups, according to which each outlet is assigned to class A or class B for a certain handset. Class A is defined as the outlet which sell the certain handset and contains all of handsets of its group, while other situation for class B which sell the certain handset too. It's shown from the result of analysis on these two kinds of outlets that many handsets are sold better in outlets of class A than that of class B, even though the sales revenue of all these outlets in the time period is close. That is to say the handsets within a group would promote the sale for each other. Furthermore, a method proposed in this paper gives a way to find out the important attributes of the handsets which lead them to br divided into the same group, and it also explains how to add a new handset to an existing group and where would the new handset be sold best.

Keywords: community detection, product network, group, decision making.

1 Introduction

The researches on the competitive relationship between products which reflects the competition for gaining consumers have been done for a long time [1]. Besides the competition relationship among products, another promotive relationship between products is found and displayed in this paper by using the edge enhanced model. The 'small-world' phenomenon in the product space [2] and product competition network [5] has been found in prior researches, and it suggests that the study of sales promotion could be found basing on the theory of complex network.

* This work is supported by the National Natural Science Foundation of China under Grant 60402011 and National Eleven Five-Year Scientific and Technical Support Program under Grant 2006BAH03B05.

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 1654 – 1665, 2009.

Network ideas have been applied with great success to topics as diverse as the Internet and the world wide web, epidemiology, scientific citation and collaboration, metabolism, and ecosystems, to name but a few. Analysis of the network make it possible for us to obtain more veracious information including the importance of the vertexes in the network, evolution of the temporal series network, classification of the vertexes in the network, even the essence of the constitution of the network.

In this paper we apply the edge enhanced model on the dataset of sale information mentioned before, and draw the conclusion that the promotive relationship really exists between products within the same group. We take the research steps as follows: firstly we build the product network on the original sale records, then we divide the vertexes of the network into several groups by the edge enhanced model, after that we analyze on the temporal series groups and find the stable groups. These stable groups represent vertexes promoting each other. It's likely to make a better choice when launch a handset into the market according to the result of these groups since the handsets in one group would promotive each other. And a method to predict which group would promotive the new handset most is proposed in this paper.

The rest of the paper is organized as follows: Section 2 reviews the related work. Section 3 introduces the method of constructing the product network. Section 4 presents the edge enhanced model. Section 5 discusses an approach of seeking the stable group of temporal series groups. Section 6 proposes a method to find the group's characteristic and a method to find the closest group to a new product. Section 7 presents the details of the experiment on the sale dataset and the result of it and Section 8 concludes as well as prospects the future work.

2 Related Work

Hidalgo C A(2007) analyzed the relationship between different kinds of product and the network is constructed by the similarity. It's concluded from his paper that adjusting the construction of the import and export merchandise according with the relationship between different kinds of merchandise might improve the situation of trading and economy of the country [2]. Li Munan(2007) analyzed the relationship between companies producing the similar kind of products [5]. He also did analysis of the basic topology characters of the products competition network.

The algorithm for community detection is involved in this paper in order to seeking for the essence structure of the product network. Girvan and Newman have introduced a betweenness based algorithm by iteratively cutting the edge with the biggest betweenness value to partition the whole network into small communities, by using a proposed Q modularity measure, it can generate the optimized division of the network. To improve the computation efficiency, Newman has also proposed a fast clustering algorithm very recently. Other proposed algorithms include METIS [14], spectral clustering [15], flow simulation, as well as co-clusterfing [3]. Nan Du (2007) proposed a clique-based algorithm which performs well in large graph [7].

3 Product Network

Competition relationship is the main relationship among the homogeneous products. And other subtle relationships are recorded in the sale information from which the whole research work begins.

We define a group as the set of products which promote to each other. In the section 6 a method would be proposed to introduction one efficiently way finding such groups and whether they are GSP (Groups of Similar Products) or GDP (Groups of Different Products) which are mentioned in the last paragraph.

3.1 Construction of the Product Network

Taking the products as the nodes of network, adding an edge e_{ab} between node a and node b if r_{ab} (Pearson's correlation coefficient) > θ.

$$r_{ab} = \frac{\text{covariance}(a,b)}{\text{standard\_deviation}(a)\times\text{standard\_deviation}(b)} = \frac{s_{ab}}{s_a s_b} \quad (1)$$

a_k and b_k are the sold amounts of product a and product b at the k_{th} outlet which has sold both of them in the considered period. We can find out that r_{ab} is between -1 and 1 while value 1 means whole coefficient and value -1 means negative coefficient. So the edge e_{ab} means node a and node b would increase synchronously to some extent which also presents the potential promotion between a and b. And the weight of the edge $W_{e_{ab}}$ is the number of outlets which sell both a and b in the same period.

According the feature of the groups we want to gain, it could be concluded that the products would promote the others within the group and present little or no promotion to the other products in the different groups. These kinds of groups are similar with communities in social network.

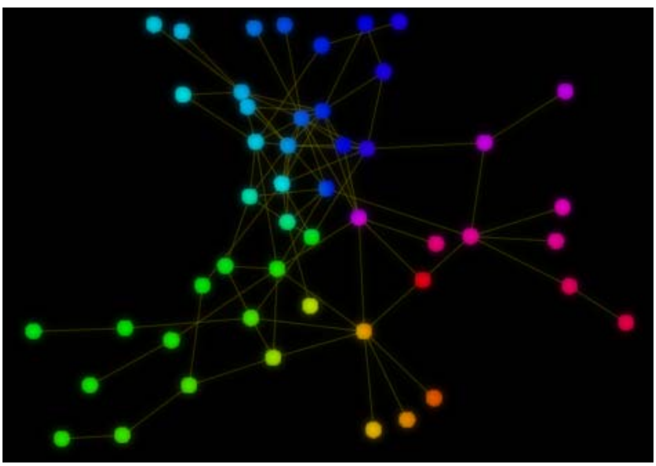

Fig. 1. The product network of the handsets sale in Jan 2007

3.2 Community of Network

Newman proposed the concept of community structure in complex network in 2002[4]. Community is the set of the nodes in network, and the nodes in community connect closely to each other while not between different communities. This set is also called cluster, cohesion group. Although there is not an abroad acceptant, uniform and measurable definition of community, it is the real symbol of the arrangement and model structure in complex systems. Due to the community is always related to social network, we will use the group instead of community to represent the structure of the non-social network in the rest of this paper.

4 Edge Enhanced Model

4.1 Positive Enhance

The groups in the product network mentioned above are such sets of vertexes in which promotive relationship exists between vertexes. As the edges in product network stand for the promotive relationship, the edges are more possible to exist between vertexes within one group. And the edges are enhanced at a probability of P to illustrate this existence. While an edge with low weight is built at a probability of P when there is no edge between some two vertexes of one group to illustrate the possible promotive relationship between them. The probability of P is in inversely proportion of the size of the group and the weight of the edge, for the possibility is lower to fortify an edge in a large group or an edge strong enough. The new network is more likely to show the real promotive relationship than the original network because the edges with bigger weight are those which illustrate the promotive relationship better or more appropriately.

Let E represent the set of all the edges of network G, and set { c_i } (i=1,2,3…) stand for the set of groups found by GN algorithm. For each pair (x_a , x_b) , $x_a \in c_i, x_b \in c_i$, (i=1,2,3…) , if $\exists\, e_{ab} \in$ E then e_{ab} is enhanced at a probability P, otherwise an edge e_{ab} is added to G and E at a probability P. The new network is G′, and the new set of edges is E′.

4.2 Negative Enhance

The negative enhance is proposed to restrict the topology status of the vertexes in the same group to be equivalent approximately. The ranking values of vertexes are gained by Pagerank algorithm, and based on the values could we weaken the edges of vertexes on far different status. An edge is weakened at a probability of P’ when the vertexes of this edge are far different on the ranking values. While the P’ is in direct proportion of this difference, representing the chanciness of this edge. The grater difference there is, the less possible for this edge to exist and the more possible for this edge to be weakened.

Let E represent the set of all the edges of network G, and V represent the set of all the vertexes of network G. And v_a (a$\in$V) is ranking value of a. For each edge $e_{ab}\in$E, weakening it at a probability P′ according to the difference of v_a and v_b mentioned above. The new network is G′, and the new set of edges is E′.

4.3 The Edge Enhanced Model Based on Community Detection

Edge enhanced model is proposed below based on all mentioned above.

1. Constructing the weighted product network according to the sale information;
2. Negative enhance;
3. Dividing the product network enhanced by step 2 into several groups;
4. Positive enhance;
5. Going back to step 2 until no more change to the partition of the groups.

The network usually converges to the network composed by several isolate vertexes or subgraphs each of which represent a group. The communities of Enron's email communication detected by GN algorithm and edge enhanced model are shown below.

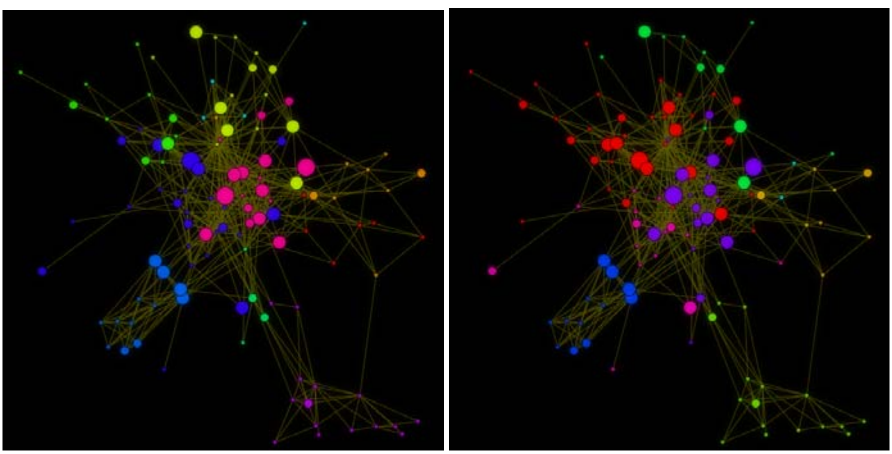

Fig. 2. The result of Community Detection of Enron's email communication. The color of the vertex represents the belonging community and the size of the vertex represents the position of this people in Enron, for example the biggest vertexes are the CEOs of Enron and the smallest vertexes are the employees or traders of Enron. The first one is the result of edge enhanced model and the second one is the result of GN. It's easy to be found out that the vertexes in the same community are more similar on size in the result of edge enhanced model than those in the result of GN, and it means the result of edge enhanced model could gain the result closer to the realistic world than that of GN.

Generally speaking, communities detected by edge enhanced model are more well-proportioned on size and the vertexes of them are more equal on topology status, and the edges in the final convergence of the network are those more realistic ones.

5 Temporal Stable Group Detection

To find the stable temporal group structure, a method using the distance between groups to seek the stable group is proposed as below.

1. Detecting the partition of groups $\{c_i\}$, i=1,2,… of snapshot Ti in the temporal series $\{T_1, T_2, ..., T_n\}$ using the edge enhanced model;
2. $\forall x, y \in c_i$,i=1,2,…,let $d_{xy}=0$ be the distance between vertexes x and y; $\forall x \in c_i, \forall y \in c_j$, i,j = 1,2,…, i $\neq$ j, let $d_{xy} = d_{c_i c_j}$ be the distance between vertexes x and y while $d_{c_i c_j}$ represents the minimum distance between group c_i and c_j;
3. For each snapshot Ti using step 1 and step 2 to gain the average distance $\overline{d}$ for each pair of vertexes;
4. The final stable temporal groups are those cliques with the vertexes close to each other.

This method is effective for the variable product network as well as relative stable social network. The temporal groups of social network are not changing obviously make the distance between vertexes within the same group remain at a stable low level, which promote them to be divided into one final group. While considering to the variable product network the occasional promotive relationship is screened or revised when it's detected on the temporal series.

6 Characteristic of a Group

In the network, a node represents an entity which is full of characters. Take the handset as an example. A handset might have such characters like color, size, design style, produced date, abundant functions, supporting services, etc. For different groups the key characters which make the nodes form these groups are impossible the same. So we should find out the key characters for each group which is named Characteristic.

The values of characters are of different format. For example, the size of a handset is a number, while the color of a handset is a word. Firstly, we reflected all the attributes to the values in the unified range [x_0, x_1]. So for each product we have a attribute-value table as follows:

Table 1. Attribute-Value table for a handset

Attribute	value	Reflected value
COLOR	RED	3
PRODUCED DATE	2008	7
……	……	……

After each product has its own characters and the corresponding values, how can we calculate the Similarity of two products? We used a formula like Jaccard Coefficient as follows:

$$J_{pro} = \frac{\sum_{k=1}^{n} w_{j_k} \times \left(X - abs\left(x_{j_k} - y_{j_k} \right) \right)}{\sum_{k=1}^{n} w_{j_k} \times \left(X - abs\left(x_{j_k} - y_{j_k} \right) \right) + \sum_{l=1}^{m} w_{j_l} \times X} \tag{2}$$

The n in (2) is the number of attributes on which both x and y have the corresponding values, and the m in (2) is the number of attributes on which only one of x and y has the corresponding values. The X is the maximum value in range $[x_0, x_1]$ and this is why we should unify all the value of attributes in this range. The w_j in (2) is the weight of the j_{th} attribute, it means how important this attribute is and whether it's a positive attribute or negative attribute for this group to be distinguished from other groups, and sum of w_j $\sum_j |w_j| = 1$. Moreover, the w_j here might be minus to make J_{pro} of two products larger while the discrepant attributes contribute to the cause forming the group most.

The problem now is how to gain the vector of w_j. Let's review the cause of forming the groups in section 3. Then it could be concluded that the products in the same groups might be either similar or different. If they are similar, the characters of this group to be distinguished from other groups are those similar attributes, while they are different, the characters of this group to be distinguished from other groups are those discrepant attributes. The later characters can't be gained if using the decision tree as the method of characteristic. So we want the J_{pro} within the group is small enough and the J_{pro} between the products of different group is big enough. In this paper, the genetic algorithms is used to generate the hypo-best vector of w_j.

The parameters p, r and m used in the genetic algorithms in the experiments will be shown in the table 2. The Fitness function is shown as follows and the Fitness-threshold defined in this paper is the maximum execute times.

$$Fitness=\frac{\frac{1}{n\times m}(J_{pro\sim x_1y_1}+J_{pro\sim x_1y_2}+...+J_{pro\sim x_1y_m}+J_{pro\sim x_2y_1}+...+J_{pro\sim x_ny_m})}{\frac{1}{n\times(n-1)}(J_{pro\sim x_1x_2}+J_{pro\sim x_1x_3}+...+J_{pro\sim x_1x_n}+J_{pro\sim x_2x_3}+...+J_{pro\sim x_{n-1}x_n})}\ . \qquad (3)$$

In function (3) the n represents the count of the products in this group and the m represents the count of the products out of this group. The x_i represents the i_{th} product of this group and y_i represents the i_{th} product out of this group. Therefore, for each group there is a certain vector of w_j achieved by the genetic algorithms to shown the most distinguishing characters of this group. A larger $|w_j|$ means a more important attribute and if $w_j < 0$ it means the products in this group might different on this attribute while $w_j > 0$ it means the products are similar on this attribute.

After it's known that how the groups formed, we can predict which outlet or outlets should a new product to be sent to in order to profit most. The algorithm is shown as follows:

Table 2. The algorithm of predicting

For each group G_i Loop

Calculate $\overline{J_{pro}}$ = average(sum of J_{pro} of product in G_i);

Calculate $\overline{J_{pro\sim xG}}$ = average(sum of J_{pro} between new product x and product in G_i) ;

End loop;

Get the index x of group G_x if nearest ($\overline{J_{pro}}$, $\overline{J_{pro\sim xG}}$) satisfied;

Return the group G_x .

In the algorithm the function nearest($\overline{J_{pro}}$, $\overline{J_{pro\sim xG}}$) is the function to judge whether $\overline{J_{pro}}$ and $\overline{J_{pro\sim xG}}$ are close enough. Either ratio method or the subtract method is a simple but useful way. In this paper, the subtract method is chosen.

It will profit most when make the new product sell at the outlets which contain all the products of G_x according to the algorithm above.

7 Experiments

The product network of each month's sale of handsets is built basing on the principle mentioned in section 3. The groups of the product network are detected by GN algorithm and edge enhanced model monthly. After the temporal series of groups are gained, the method mentioned in section 5 is used to detect stable groups in the temporal series. For each group all the outlets could be divided into class A and class B. The outlets of class A for a group are those which contain all the handsets of this group, while the ones of class B for this group are those which do not contain all the handsets of this group. The details of sale information of handsets are compared between class A and class B on the average sold amount. Using the genetic algorithms to gain each character vector of w_j for each group. Then predicting the best sell outlets for several new products with the method mentioned in section 6 and checking the sell information between class A and class B to testify this hypothesis.

7.1 Description of Data Used in Experiment

1. The data used in the experiment come from the handset sale records of outlets in a province operator during the period from January ,2006 to July, 2008.
2. The handset fetching for analysis every month absolutely meet the condition of "marketable sale exists in this month" as above proposed. Analysis is not done to the handset without marketable sale.
3. There are 293 outlets which bear the handset sale qualification and subject to this company, 43 different handset brands and 320 kinds of different models exist in the experiment data.
4. Parameters used in the experiment:

Table 3. Value of parameters

Para	p	r	m	*Fitness-threshold*
Value	10	10%	11%	10 (times)

7.2 Results of the Experiment

For each handset we divide all the outlets into class A and class B following the principle mentioned previously. The compare of the result of GN algorithm and the result of EE model shows in fig. 3. Nearly all the handsets are sold better at the outlets of class A than class B checking in the result of EE model, while the result of GN algorithm doesn't perform so well. So it's more possible for the groups of EE model to stand for the set of promotive products.

Taking two groups out of all the groups for example of the Characteristic Process, which represent the GSP and the GDP respectively. Using the method in section 6, the

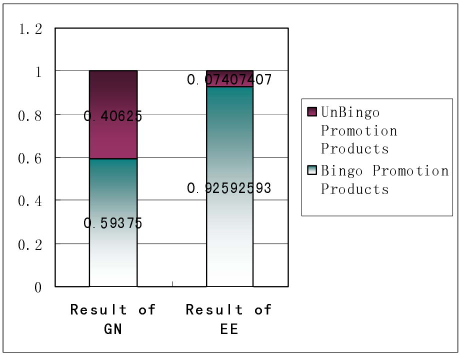

Fig. 3. Blue part of the histogram represents the proportion of better sold handsets in class A compared to class B. While the red part represents the proportion of the less sold handsets in class A compared to class B.

vector of w_j is obtained and we display the top 5 attributes depend on the absolute value of w_j as the figures follows. The same attributes are painted in the same color for each figure. And the large vertexes stand for the handsets while small ones stand for the attributes.

Seeing from figure 4 and figure 5, the attributes somehow shows the conducts of the customers. While the C&R Tree gets the rules like “Don’t have the encryption techniques; SM group sending 10 pieces; Standby time 260 hours” for the first group

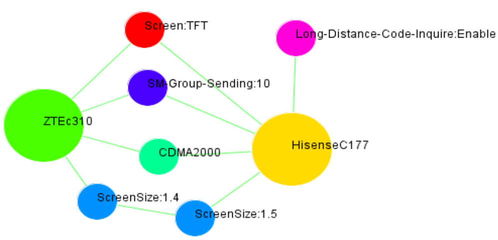

Fig. 4. The top 5 attributes of group {HisenseC177, ZTEc310}

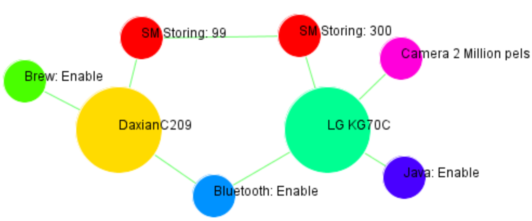

Fig. 5. The top 5 attributes of group {LGKG70C, DaxianC209}

and "Don't have the encryption techniques; Listing dates are 2006 and 2007" for the second group, and these rules are not convictive at all.

There are fifteen new handsets be put into the market from Jan 2008, and we can testify how the method mentioned in this paper works. Using the algorithm in section 6, all the fifteen handsets are assigned to the closest groups, and the average sale amount of each handsets is compared between outlets of class A and outlets of class B as follows:

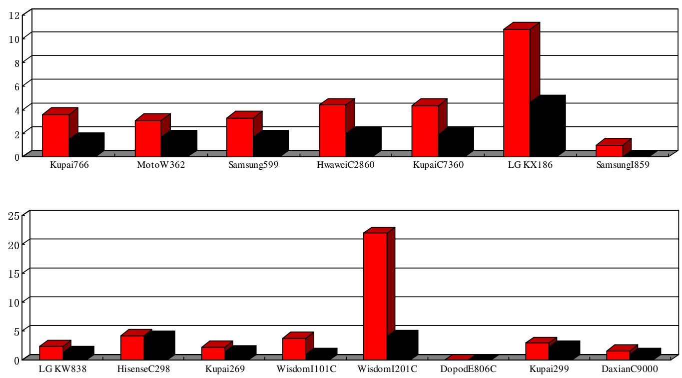

Fig. 6. Comparing of the average sale amount for each new handset between outlets of class A and outlets of class B from Jan, 2008

We could see from the figure 6 that the new handsets are all sold better at the outlets which contain all the handsets within the assigned group than other outlets apparently except for the DopodE806C which is assigned to a group but there isn't any outlet sold it and all the products within its assigned group. Therefore the precision for Prediction is 93.3%.

8 Conclusion

This paper proposed an edge enhanced model on the basis of community detection, it make the network evolve to essential structure by positive enhance and negative enhance. The group detected by edge enhanced model has a more important character than any group before, which is the nodes in the same group have the similar topology status in network. This paper analyzes the handsets' sale records by using edge enhanced model and finds the promotive products groups. We also testify the result and arrive at the conclusion that the handsets locating in the outlet which sells handsets within the same group could promotive sale mutually. The paper also involved the aspects like new product launch, the strategy of balancing new and old products. People could launch the new product purposive into the outlets which contains the closest group for the new products based on similarity, in order to spreading the new products effectively and quickly.

References

1. Bass, F.M.: A new product growth model for consumer durables. Management Science 15(5), 215–227 (1969)
2. Hidalgo, C.A., Klinger, B., Barabási, A.-L., et al.: The Product Space Conditions the Development of Nations. Science 317, 482–487 (2007)
3. Borgs, C., Chayes, J., Mahdian, M., Saberi, A.: Exploring the community structure of newsgroups. In: Proc. 10th ACM SIGKDD Intl. Conf. Knowledge Discovery and Data Mining (2004)
4. Newman, M.: Detecting community structure in networks. The European Physical Journal B 38, 321–330 (2004)
5. Jianmei, L.M.Y., Weiliang, T.: Research on features of product competition network based on complex network. Electronic measurement technology 30 (2007)
6. Radicchi, F., Castellano, C., Cecconi, F., et al.: Defining and identifying communities in networks. Proc. Natl. Acad. Sci. 101(9), 2658–2663 (2004)
7. Du, N., Wu, B., Pei, X., Wang, B., Xu, L.: Community Detection in Large-Scale Social Network. In: KDD workshop (2007)
8. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking:Bringing order to the web. Technical report, Stanford University, Stanford, CA (1998)
9. Barabási, A.L., Albert, R.: Emergence of Scaling in Random Networks. Science 286(5439), 509–512 (1999)