Optimization of Data-mining-based E-commerce Enterprise Marketing Strategy

WenJing Jia¹, Han Zhou¹-², Yan Shi¹

¹ Baoding Vocational and Technical College, Baoding 071000, China.
² Agricultural University of Hebei Province, Baoding 071000, China.

Abstract

With the continuous development of information technology, the Internet has been playing an important role in promoting the development of various traditional industries due to its openness, fast speed and rich content. In commerce and trade, the e-commerce mode emerging on the basis of the Internet overturns the traditional offline face-to-face trading model. Through e-commerce models such as ABC, B2B, B2C, C2C, B2M, M2C, B2A, C2A (C2G) and O2O, people can achieve online and online-offline transactions, greatly promoting the development of e-commerce. In the context of the highly developed e-commerce, network marketing, as an important product of e-commerce, directly affects the sales and brand image of e-commerce enterprises and is crucial to the development of e-commerce enterprises; therefore, e-commerce enterprises must attach great importance to marketing, apply the data mining technology and optimize their network marketing strategies, so as to promote the development of the whole e-commerce industry.

Keywords: Data Mining Technology, E-Commerce, Network Marketing.

1. RESEARCH OVERVIEW

1.1 Research Background

1.1.1 Background of E-Commerce

As early as in 1990-1993, when electronic data exchange just started, Chinese researchers already had a certain knowledge of e-commerce and began their research. During 1993-1997, the three Golden Projects - Golden Gate, Golden Card and Golden Tax – made major progress and laid a solid foundation for the future development of e-commerce. In 1997, the earliest e-commerce platform in China - Chinese Goods Ordering System (CGOS) - was officially put into operation, indicating that e-commerce officially became an integral part of the financial trading system in China. Due to various subjective or objective factors, the CGOS developed quite slowly, and did not complete its first online transaction until one year after its establishment. After 2000, e-commerce gradually formed its overall landscape dominated by B2B platforms and entered a stable development period. During this time, e-commerce platforms led by Taobao and Jingdong Mall developed rapidly and became leading enterprises in China, which contributed a lot to the development of the e-commerce industry.

1.1.2 Background of Data Mining

In the context of big data, the amount of data in all fields is increasing. According to relevant research, the amount of data generated every seven days now is the sum of the data amount generated over the past several thousand years. Data mining is a process of using certain algorithms to search for useful hidden information from a large amount of data. Because of its ability to effectively analyze and mine big data, this technology has been widely applied in various fields. In e-commerce, with the help of data mining technology, e-commerce enterprises can gain a better understanding of consumer spending habits.
1.2 Literature Review

Network marketing has the following characteristics: 1. Network marketing does not exist in isolation. Network marketing is an important part of the production-sales chain in an enterprise and an important way to increase its sales. It cannot exist outside the overall marketing environment, and in some sense, network marketing is an extension of the traditional marketing model in the network. 2. Network marketing is not online sales. The ultimate goal of marketing is to achieve, promote and expand sales and establish brand images. Network marketing is an inevitable product in the development of online sales, but it is not online sales, but a means to increase exposure of products and sales promotions. (Tian et al., 2014). Network marketing and e-commerce are two different concepts, but it is easy to confuse these two as they are closely related in nature. E-commerce has a broader scope. It is essentially a trading activity based on the Internet, reflecting all aspects and ways of commodity trading. However, e-commerce is not a complete transaction process, but an activity of information transmission in the process of production and transaction to achieve and facilitate transactions (Zhang and Fan, 2014). Network marketing has a variety of marketing tools such as search engine marketing, search engine optimization, email marketing, instant messaging marketing, viral marketing, BBS marketing, blog marketing, Weibo marketing, WeChat marketing, video marketing, article marketing, experiential micro-marketing, O2O marketing, we-media marketing, new media marketing and so on. Due to its fast propagation, wide audience and vividness, network marketing is able to facilitate the exchanges between buyers and sellers and effectively reduce marketing costs. At the same time, network marketing relies on the Internet platform, requiring very low maintenance and other costs, which thus can effectively reduce the operating costs and increase the economic efficiency of enterprises (Yang Y.M. 2014).

2. NETWORK MARKETING INFLUENCE MODEL

Network marketing has a direct influence on the sales of goods in an enterprise and even its corporate image. Among the many network marketing tools, viral marketing can apply to many areas and has widespread coverage. The core idea is to first market a product to a small number of consumers and then to influence similar consumers through these consumers so that the marketing effect gradually spreads out like a virus until it reaches the stable state, i.e. no new consumer can be influenced (Zhao and Yang, 2015). This method is based on some assumptions: first, in network marketing, each consumer is considered as a node, and each node has two states - active and inactive; second, an activated node cannot be transformed into an inactive one, while an inactive node can be transformed into an activated one; and at last, the method to influence an inactive node is to increase the number of active nodes around it so that the inactive node will develop a certain activation tendency and in the end complete the activation (Chen and Zhang, 2013).

2.1 Influence Linear Threshold Model

The influence linear threshold model for viral marketing is shown as follows:

$$\sum_{w \in \text{activeneighborof} v} p_{w,v} \geq \theta_v$$  \hspace{1cm} (1)

where, \(p_{w,v}\) is the strength of the influence that node \(w\) has on node \(v\). In general, the sum of weights of all sides is a value within the interval \([0,1]\), whereas \(\theta_v\) represents the threshold of \(v\), whose value also ranges within the interval \([0,1]\). In the linear threshold model, the node activation process is as follows: first, there must be a set of initial activated nodes, and during each round of activation, check whether the status and conditions of the inactive node are satisfied. If they are satisfied, it is activated; and if not, it is recorded as an inactive node. Repeat this process until there is no inactive node, and by then, it means the propagation is stable (Wang et al., 2013).

2.2 Submodular Function

Assuming there is a finite set defined as \(N\), in a subset of the finite set \(N\), the function \(f\) that maps the set to real numbers is called the submodular function. Suppose the submodular function \(f\) satisfies the following conditions:

$$f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T), \forall v, \forall S \subseteq T$$  \hspace{1cm} (2)

where, \(v\) is a set of \(N\) elements, and \(S\) and \(T\) are \(N\) two subsets. This condition can also be called diminishing
return, that is to say, the profit of one element \( v \) after being added to \( S \) is not less than that of any other element added to \( S \). The equivalent condition is as follows:

\[
f(S) + f(T) \geq f(T + S) + f(T \cap S), \forall S, T
\]  

(3)

Assuming that function \( f \) is a non-negative and monotonically increasing function and a submodular function of a finite set, if there exists a subset \( K \) which maximizes \( f \) in set \( S \), then this proof will be the key and difficult part of NP (Dan and Tian, 2013).

2.3 Diversity Index

Diversity index is a quantitative index measuring the types and distribution uniformity of a certain factor in a data set. In network marketing, as the number of nodes, i.e., the number of consumers, continues to increase, the types and uniformity of the factors also increase (Wang, 2013). The diversity index can be divided into simple and complex ones. The simple diversity index is mainly abundance, which is commonly used in ecology, and the complex diversity index is defined as follows:

\[
q_D = \left( \sum_{i=1}^R p_i^q \right)^{\frac{1}{1-q}}
\]

(4)

where, \( R \) represents the abundance of factors, \( p_i \) represents the proportion of factor \( i \) in the whole set, and \( q \) represents the parameter of the factor. When \( q=1 \), the above formula is meaningless, and if \( q \) is infinitely close to 1, there exists a limit for \( D \). The calculation method is as follows:

\[
1^D = \exp(-\sum_{i=1}^R p_i \ln p_i)
\]

(5)

In the above formula, when \( q=0 \), the value of diversity is equal to the abundance; in other words, diversity is, to a certain extent, the abundance. Among the diversity indices, Shannon entropy is the most commonly used. This concept derives from the information theory and is widely used in ecology. The formula is as follows:

\[
H = -\sum_{i=1}^R p_i \log_2 p_i
\]

(6)

2.4 Maximization of Diversity Influence

How to maximize the influence of network marketing to achieve better marketing effects is the main focus of the virus marketing strategy. To study the issue, it is necessary to establish diversity measuring indices and prove that the indices are non-monotonic and sub-modular (Zhou, 2015). The indices needed to study this problem are shown in Figure 1:

**Figure 1. Main Mathematical Symbols Used**

<table>
<thead>
<tr>
<th>mathematical notation</th>
<th>Meaning</th>
<th>mathematical notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S )</td>
<td>Seed set</td>
<td>( w_{ij} )</td>
<td>Membership degree</td>
</tr>
<tr>
<td>( \sigma(S) )</td>
<td>Collective influence</td>
<td>( w_i )</td>
<td>Vector of C</td>
</tr>
<tr>
<td>( \mu_i^S )</td>
<td>Activation probability</td>
<td>( D(\mu^S) )</td>
<td>Population diversity</td>
</tr>
<tr>
<td>( \mu^S )</td>
<td>The vector of ( V )</td>
<td>( D(S) )</td>
<td>Diversity of sets</td>
</tr>
<tr>
<td>( C )</td>
<td>Number of node classes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In network marketing, maximizing influence and maximizing diversity are often not consistent but rather contradictory. Therefore, it is necessary to set the function as an explicit combination of the two, with the following structure:

\[
\max F(S) = (1 - \gamma) \frac{\sigma(S)}{\sigma} + \gamma \frac{D(S)}{D}
\]

(7)
s. t. \( S \subseteq V, |S| = K \) \hspace{1cm} (8)

In the above formula, \( \mu^S \) represents the vector of the length \(|V|\). The \( i \)-th term \( \mu^S \) also shows the probability of node \( i \) being activated when the set is \( S \). In addition, \( \sigma(S) \) represents the influence of the seed set \( S \) on other items, and \( D(\mu^S) \) indicates the diversity of consumers that can be influenced by network marketing under the same conditions. \( \sigma, D \) are normalization factors, which are the maximum values that molecules can choose. \( \gamma \in [0,1] \) represents a parameter that can be used to adjust the diversification of the audience for pilot network marketing. When \( \gamma = 0 \), the problem expressed by the above formula is the problem of maximizing the original influence, and the larger \( \gamma \) is, the greater the influence will be. The problem of maximizing the influence over diversity can be converted to the coin problem. Suppose that for each side \((v, w)\), a coin is used to make the decision, and that the probability that the coin is heads up is \( p_{vw} \), and that \( X \) is the result of coin tossing for all sides \((v, w)\), the first thing to prove is as follows:

\[
\forall S \subseteq T, v \notin T, M^S_{j,x} - M^{T(v)}_{j,x} \geq M^T_{j,x} - M^S_{j,x} \hspace{1cm} (9)
\]

Generally speaking, there are three situations:

First, suppose there exists only one activation path between node \( j \) and \( S \), and that for \( S \), any superset must have an activation path connected to \( j \), the two sides of the inequality are 0, and the inequality holds.

Second, suppose there is no activation path between node \( j \) and \( S \), but that there is an activation path between the other node \( v \) and \( j \), the coordinate of the inequality is 0 and the maximum value of the right side is 1. Therefore, the inequality holds.

Third, suppose there is no activation path between node \( j \) and \( S \), nor is there any between nodes \( v \) and \( j \), both sides of the inequality are 0. Therefore, irrespective of whether there is an activation path between \( T \) and \( j \), the inequality can all hold.

3. STUDY OF THE DATA-MINING-BASED NETWORK MARKETING STRATEGY FOR E-COMMERCE ENTERPRISES

3.1 Data-Mining-Based Network Marketing Strategy Model

The data-mining-based network marketing strategy model is shown in Figure 2:

![Figure 2. Network Marketing Model Based on Big Data](image)

In the background of big data, the data-mining-based network marketing strategy first involves obtaining relevant data resources from logs, forums, Weibo, social networks and transactions, etc., incorporating the collected information resources into the network marketing model, and sorting them out through the algorithm library, which includes sorting tools like classifiers, clustering algorithms and association rules, and then, through big data calculation, it needs to analyze the sorted information, and at last, through big data, it needs to conduct association rule mining marketing, commodity geomarketing, social network marketing, user behavior analysis marketing, personalized referral marketing and the big data analysis marketing based on modern
communication tools (Liu and Cheng, 2015).

3.2 Data-Mining-Based Network Marketing Measures

3.2.1 Data-Mining-Based Commodity Association Mining Marketing

Commodity association mining marketing is the method of selling two different types of products with a certain association. For example, beer and diapers are almost two unrelated products, but when these two products are put together on shelves, sales of both products are greatly increased. According to survey, most American women are housewives, who do not shop frequently. Most of them would ask their husbands to buy diapers on their way home from work, but alcohol consumption is a major hobby for male Americans, so many people would also buy beers while buying diapers. This is the most important reason why these two are put together to increase sales. From this, it can be learnt that, if we apply the data mining technology to explore the association between goods and uses this association to develop a marketing strategy, it will effectively increase goods sales (Zhang, 2016).

3.2.2 Data-Mining-Based Commodity Geomarketing

Consumers in different geographic locations have different consumption preferences, so different marketing methods may be adopted for consumers in different provinces. For example, Taobao did a bra survey in 2011, which showed that sales of bras in the north (especially in Xinjiang) were significantly higher than those in the south, indicating that people in the north had more demands for bras than in the south. What is more, women in the south generally buy A-cup bras, while those in the north are more inclined to choose B- or C-cup ones. Therefore, appropriate marketing strategies can be developed according to these different geographical features (Zang, 2016).

3.2.3 Data-Mining-Based Social Network Marketing

With the rapid development of information technology, social apps such as QQ, Weibo and WeChat are used more and more frequently. Marketing on social apps will not only gain a large audience base and achieve good marketing effects, but also save a lot of money on construction and maintenance (Hu, 2011). For example, when Xiaomi launched a new Red Mi phone, it spread and forwarded the message on Qzones and among QQ friends, forming a virus effect, making the sales of the Red Mi phone far beyond the expectation. Here is another example. By combining network and star resources, Renren carried out a yogurt promotion activity for music lovers in colleges and universities. While providing a platform for young people to pursue their dreams, this activity also played a positive role in marketing Renren and the yogurt product, reflecting the great significance of social network marketing.

3.2.4 Data-Mining-Based User Behavior Analysis Marketing

User behavior analysis marketing means collecting users’ search history, shopping habits and purchased goods, etc. and carrying out marketing accordingly. For example, Suning, after acquiring PPTV, independently developed Yunxi software, which can search and analyze consumers’ shopping behaviors and history records, predict the goods that consumers need to buy and find potential consumers. By sending ads to these consumers, it can effectively increase the sales of the goods. This method is more targeted and can stimulate the desire of potential users to purchase, so it serves as an important network marketing strategy (Li and Jing, 2011).

3.2.5 Data-Mining-Based Personalized Referral Marketing

In this big-data era, the information and data provided by multimedia have become the main source of big data. The traditional print ads have been gradually developing towards acousto-optic ads. On social apps, if we search the people nearby, we will also find many people doing all kinds of marketing, which shows that social network has become a major base for network marketing. On social apps, consumers can set up or join a group or social circle according to their own needs, where they can post product information accordingly. With the help of recommendation algorithm, personalized marketing strategies can be developed to achieve good marketing effects (Zhao et al., 2013).
REFERENCES


Dan T.T., Tian G. (2013). Chinese marketing research papers in English -- a summary of contents and sources of the description and analysis of China market, based on the 2013 (17), 5-16.

Hu Z.J. (2011). Commercial bank marketing system dependent and theoretical model in different system conditions of China's commercial banks marketing research, technology economy and management research, based on the (9), 65-69.


Wang R.Z. (2013). Construction path of Higher Vocational Marketing Specialty Based on "school enterprise cooperation" -- Taking marketing specialty of Shenzhen Polytechnic as an example, vocational and technical education, 34 (11), 20-22.


Zang N.N. (2016). The sale of electricity companies how to build the system reform of electric power marketing system China, energy, 38 (4), 44-47.


Zhao Q.W., Yang Q. (2015). Research on the "Trinity" training mode of marketing professional talents -- Taking the marketing specialty construction of Chongqing University as an example, modern educational technology, 25 (2), 120-125.


Zhou X.Y. (2015). Research on three dimensional marketing practice teaching system from the perspective of innovation and entrepreneurship training, education observation (first half), 4 (10), 89-92+112.