

Association for Information Systems AIS Electronic Library (AISeL)

WHICEB 2017 Proceedings

Wuhan International Conference on e-Business

Summer 5-26-2017

Characteristics Description of Potential User Segments on the E-Commerce Website oriented to Precision Marketing

Tianyu Zhou

School of Economics and Management, Nanjing University of Science and Technology, Nanjing, 210094, China,
tian68600@163.com

Li Li

School of Economics and Management, Nanjing University of Science and Technology, Nanjing, 210094, China,
lily691111@126.com

Follow this and additional works at: <http://aisel.aisnet.org/whiceb2017>

Recommended Citation

Zhou, Tianyu and Li, Li, "Characteristics Description of Potential User Segments on the E-Commerce Website oriented to Precision Marketing" (2017). *WHICEB 2017 Proceedings*. 36.
<http://aisel.aisnet.org/whiceb2017/36>

This material is brought to you by the Wuhan International Conference on e-Business at AIS Electronic Library (AISeL). It has been accepted for inclusion in WHICEB 2017 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Characteristics Description of Potential User Segments on the E-Commerce Website oriented to Precision Marketing

Tianyu Zhou¹, Li Li²

School of Economics and Management, Nanjing University of Science and Technology, Nanjing,
210094, China

Abstract: In the increasingly competitive environment between e-commerce companies, for more accurate implementation of marketing strategies, e-commerce websites often choose to subdivide the consumer market of the enterprise to identify site users' characteristics to find their needs. In this paper, we subdivide consumer market from the four dimensions of behavior, geography, demography and psychology and propose a model to describe the characteristics of potential user market segments. Based on the web log data and user transaction data, a classification algorithm is used to analyze user behavior data in Web log to find the potential user segments and the user's descriptive characteristics in user transaction data are clustered to obtain the distribution of consumer characteristics under various product categories, then we use the product categories in e-commerce website as an intermediary to give every single potential user in potential user market segments the descriptive characteristics, which can provide data support for the realization of precision marketing. The proposed model is applied to the actual data of a certain insurance e-commerce platform, and based on the results, we gain some implications for marketing of the e-commerce website.

Keywords: Multi-dimensional market segmentation, Characteristics description of market segmentation, Precision marketing, Potential user

1. INTRODUCTION

The rapid development of Internet technology and the popularity of intelligent terminal devices provide a good environment for the development of e-commerce, online shopping has become an indispensable part of people's life. "2016 Annual Report and Guide on China's Consumers Online Shopping" released by China Electronic Commerce Research Center points out that today's e-commerce retail market has entered a steady growth in sales in the mature period, China's online retail transaction size has exceeded 5 trillion for the first time this year. Followed by the growing competition between the e-commerce companies, the report points out that the e-commerce market share of the top ten compared with the same period last year fell by 2.2% in the first half of 2016, online retail market is showing a leading market share of Taobao and Jingdong with multi category e-commerce companies being roughly in the same fierce competition ^[1]. In such a fierce competitive environment, how to implement the marketing strategy more targeted than the same type of e-commerce website and stimulate more users to consume is becoming the focus of each e-commerce business.

Marketing as a part of the business process, to some extent, determines the consumer's impression of the product and will also have a significant impact on the final sales of the product. After all, the essence of marketing is to give consumers a reason to buy the products. When e-commerce is not yet popular, traditional enterprises will carry out the promotion such as advertising and paper coupons in order to stimulate consumer purchase intention. Considering the characteristics and needs of consumers are not the same, although these methods may have some effect, no differences in the implementation of marketing measures will lead to low efficiency and high cost marketing problems. Therefore, in order to achieve targeted marketing, it is necessary to

¹ Email address: tian68600@163.com

² Email address: lily691111@126.com

subdivide the consumer groups of enterprises to obtain the main characteristics and needs of different groups, which can help marketing manager plan different marketing strategies so as to achieve better marketing results. In the actual market environment, the consumer market segmentation is a common way to divide the consumer groups, before the popularity of the Internet has not been widespread, companies usually choose the method called offline consumer market segmentation, that is, through the questionnaires, interviews and other ways to get the user's demographic information, to divide the consumer groups. With the development of the e-commerce enterprise, the user data generated by a user's browsing and trading behavior during a visit to the site can more accurately reflect the characteristics of users compared with the data of the questionnaire from consumers, which provides data support for e-commerce enterprises to do online market segmentation. This research will focus on how to provide a new idea of online market segmentation for e-commerce enterprises based on these data.

With the rise of the concept of precision marketing in recent years, marketing strategy being targeted to one single user whose characteristics and needs has been known is undoubtedly a very ideal situation for marketing planning. And in the existing market segmentation study, the market segmentation method that subdivides the potential user groups according to the user's website click data such as behavior variables has been gradually concerned due to its ability to accurately identify the individual consumers. Liu *et al.* used Naive Bayesian classification algorithm with weighted features to process Web log data to obtain a high-accuracy potential user segmentation model^[2], yet this segmentation method does not have a systematic characteristics description for the predicted potential users. In addition, some researchers in order to better reflect the characteristics of users in the market segmentation usually adopt many kinds of subdivision variables. Kim and Ahn proposed a clustering algorithm based on genetic algorithms to segment the online fitness market by dealing with users' psychological factors such as weight loss expectations and demographic information. And they found the users' needs according to the main demographic characteristics and psychological characteristics of these segments from the clustering results^[3]. Although this method can help understand user characteristics better, if we can further find users with the potential shopping intention in segments, it will absolutely make a huge boost for the realization of precision marketing. So, this research proposes a model to describe the characteristics of the potential user segments of the e-commerce market. The model combines the user's behavior variables, demographic variables, psychological variables and geographic variables to segment the consumer market and describe its characteristics. In order to combine the identified potential users who are interested in shopping but have not yet purchased and the consumer segments that have generated purchasing behavior, we take the product category of the e-commerce website which are attributes two groups both have as an intermediary, helping the potential user segments and the descriptive characteristics of the consumers under product categories to create a mapping relationship, so as to achieve the model goal. Moreover, the feasibility of the proposed model is verified by the actual user data of an insurance e-commerce platform in Nanjing. Finally, according to the modeling results and the marketing 4P theory, we gain some implications on the marketing strategy of the company from three aspects: product, channel and promotion.

The paper is structured as follows. We review the relevant literature in Section 2. And in Section 3, a specific explanation about the model to describe characteristics of the potential user segments is given. Besides, we show the data preparation process and modeling results through the actual data of the enterprise. Section 4 provides some thoughts one the site's marketing decisions according to the results of data mining in the previous section combined with the 4P marketing theory. In Section 5, we give the conclusion and the prospect of the research.

2. LITERATURE REVIEW

2.1 Consumer market segmentation

In order to better understand the characteristics and needs of consumers, enterprises usually choose the way of market segmentation to subdivide the consumer market into several segments, and then infer the needs of consumers through the main characteristics of each market segment. In the study of market segmentation, the market segmentation variables are usually divided into four kinds: geographical, demographics, psychological and behavioral variables.

Geographical segmentation variables can help subdivide the consumer market into geographical units, such as countries, provinces, cities and so on, and enterprises can choose one or several regions to operate to reduce operating costs and find more suitable regional marketing plan. However, the cross-regional feature of the Internet and the improvement of logistics system reduce the cost of e-commerce company multi regional operations, almost no e-commerce companies use one single geographical segmentation variables to subdivide their consumer market. Demographic segmentation variables are often used by researchers, including age, gender, income, occupation, race, education, etc. The advantage of this variable is that it is convenient to measure and can be easy-understanding when expressing user characteristics. For example, when a company knows the vast majority of consumers in the market segments are women, the company will obviously choose a fresh and lovely style in line with the needs of women's aesthetic in product design and packaging. Accordingly, in many fields such as tourism, sports and fitness, the researchers regarded the demographic variables as one of the variables when subdividing the market^{[3], [10]}. Yet market segmentation only using demographic variables may not be effective in segmenting consumer market for the psychological needs of consumers in the same demographic segment may not be the same. And some researches have also demonstrated the important role of psychological factors in market segmentation. Liu and Mattila found the Airbnb market share is being higher because of its user segmentation considering user's sense of power, thus Airbnb uses different advertising slogans to attract consumers according to level of the sense of power in marketing promotion^[4]. Psychological factors in the existing study of market segmentation is usually into consideration. Barnes *et al.* took consumer psychological factors as segmentation variables and chose six categories including personality, trust, perceived risk, attitude towards online shopping, shopping pleasure, and purchase intention to segment online consumers, which made consumers divided into three segments called risk-averse doubters, open-minded online shoppers, and reserved information-seekers^[5].

Because of the above three kinds of segmentation variables having the feature of describing user characteristics itself, the main characteristics of the consumers in segments divided by these variables are obvious, we classify the three kinds of variables as descriptive variables. However, the behavior segmentation variables are different from those of the first three, the single behavior itself can not represent the characteristics of the user while the combination of different behavior can show the user's behavior habits. Soopramanien and Robertson divided the online users into buyers, visitors and non-Internet shoppers in accordance with the user's online behavior^[6]. Considering the privacy and security problems, e-commerce website users are very cautious when needing to fill in their own demographic and geographic information, which makes it difficult to find descriptive variables to segment market. On the contrary, the Internet user will inadvertently leave a large number of behavioral data, such as the number of pages, page stay time, search terms, browsing time, etc. So the segmentation variables in the market segmentation for e-commerce websites usually include behavioral variables. Among all the researches on market segmentation using behavioral variables, there is one kind of research mining the user behavior to construct a classifier so as to predict user behavior. This research is of great help to the implementation of the precise marketing because its market segments can be targeted to one single consumer. Guo and Fang predicted potential users by using support vector machine classification algorithm

based on Web log data^[7]; In order to accurately predict the user's complaints about the company's marketing channels, Ozyirmidokuz *et al.* used the C5.0 decision tree algorithm to train data^[8]; In order to achieve the recognition of the micro-blog machine users, Liu *et al.* constructed the micro-blog user's feature vector through the analysis of the characteristics of machine users. Considering the imbalance of the data, the random forest classification algorithm was chosen to design the recognition model of the micro-blog machine user, and a high precision model was obtained after the verification^[9]. In this kind of research, the choice of algorithm and the feature extraction of user data will determine the accuracy of the final classifier.

In the existing market segmentation research, the researchers usually use clustering algorithm combined with various types of variables to segment market. Based on the use of travel social media, Amaro, Duarte and Henriques employed kmeans clustering to process variables of user's browsing behavior, tourism expectation, and the attitude of using the website, which helped provide support for marketing according to the user's demographic information in clustering results^[10]; Wu and Chou developed a soft clustering method using a latent mixed-class membership clustering approach to classify online customers based on their purchasing data across categories to improve the accuracy of the multi-dimensional data clustering^[11]. Taking into account the fact that clustering algorithm can subdivide the consumer market and reflect the main characteristics of the segments effectively while the classification algorithm can achieve the prediction of user behavior and make market segments targeted to one single user, this research will combine the classification algorithm and the clustering algorithm, using geographic, psychological, demographic and behavioral variables to give descriptive characteristics to all the potential users in the potential user segments. As the research object of this study, the insurance e-commerce platform users are more or less demand for insurance products. After observing the actual business data of the site, we intend to predict the potential market segments by the behavior variables and describe the segments characteristics by the three variables of geography, demography and consumer psychology, where the geographic variable refers to the provinces of consumers, the demographic variables include the consumers' gender and age, and the psychological factor we selected is the transaction price which can reflect consumer acceptance of the product price.

2.2 Data driven precision marketing

As a link in the process of enterprise management, marketing is the key to decide whether the enterprise can make profit effectively. With the increasingly fierce market competition, Philip Kotler put forward the concept called "Precision Marketing", saying that precision marketing means that companies need to be more precise, measurable and high investment return on marketing communications and should pay more attention to the results and actions of the marketing communication plan, as well as focus on the direct sales communication investment. Liu gave a definition on the "Precision Marketing"^[12]: "Precision marketing is a marketing communication that can achieve high investment return from the target market of different consumer groups. Modern techniques, methods, and targeted strategies are adopted according to different consumption psychology and behavior characteristics of consumers in the target market through the combination of quantitative and qualitative analysis." In this paper, we believe that whether precision marketing can achieve precision depends on two points, the first is the ability to accurately identify the target user, the second is to have a more accurate understanding of the characteristics and needs of users.

Reviewing the practice of precision marketing in various industries, Guan, Liu and Wang constructed a Logistic model for the recognition of individual customer's identity of commercial bank by the proxy method of the multi-characteristics of customer's behavior. The balancing process of the model between the scale predictive accuracy and the individual predictive accuracy of target group ensured the accuracy of financial business decision. The average recognition rate of the target customers can reach 80%, which provided a base for financial product development aiming at customer group with specific identity^[13]; In order to provide some

marketing decision support to Taiwan's Adidas sports, Liao, Chen and Hsu prepared questionnaire and built analysis database based ontology based on Ontology, which helped understand the relationship between consumer behavior, media publicity and brand spokesperson by Apriori algorithm of association rules and find out user preference by k-means clustering algorithm^[14]; Tang, Liao and Sun proposed a novel framework with a three-stage procedure to discover the correlation between contexts of mobile users and their activities based on sequential pattern mining technique for understanding customers' preferences and quickly advertising customized products or services. The framework overcomes the problem of redundant rules that occur when multiple dimensions of contextual information are used in the prediction^[15]. It can be seen that the difference of the data feature and business process among the industries or even the enterprises make it a problem worth considering to choose the appropriate algorithm to realize precision marketing.

By understanding the specific business processes and data features of the insurance e-commerce platform, we choose the clustering and classification algorithm to achieve the model proposed in this paper. In the actual process, the model predicts potential user market segments through the user's browsing behavior and identify the insurance product category where the potential user belongs to based on the browsing product history of these potential users. Meanwhile, the main descriptive characteristics of the corresponding insurance product category are found out through the clustering analysis of user transaction data. Then the insurance product category as an intermediary will help obtain potential users with descriptive characteristics in potential segments. So, when the website users recognized as potential users visit the website again, the website can infer their needs according to their descriptive characteristics given by the model, which helps the company plan relevant marketing measures aiming at those users to achieve precise marketing.

3. METHOD

3.1 Model explanation

This research takes the actual data of a certain insurance e-commerce platform as the research data, and the model proposed to describe characteristics of the potential user segments contains two parts, which are the market segmentation of potential user and distribution of user characteristics under the product category sold by the e-commerce website. Then the two parts will achieve association through the product category of the website as an intermediary, so as to describe characteristics of the potential user segments. The part on the distribution of user characteristics under the product category classifies the user descriptive characteristics of the transaction data according to the insurance product category by QL language, then the clustering analysis of characteristics under product categories can be used to obtain the distribution of user's descriptive characteristics under product categories. In the part about the market segmentation of potential user, the web log data is trained by the classification algorithm to obtain the potential user classifier to predict potential users, and the web log data needs data preprocessing such as data cleaning, session identification and feature extraction before classification algorithm training. Afterwards, for the potential users predicted by the classifier, we infer the insurance product category that is of interest to them according to their browsing product records. In consideration of the user characteristics under insurance product categories have been derived, the task of describing the characteristics of each potential user in potential user segments is done. Figure1 is the flow chart of the model idea.

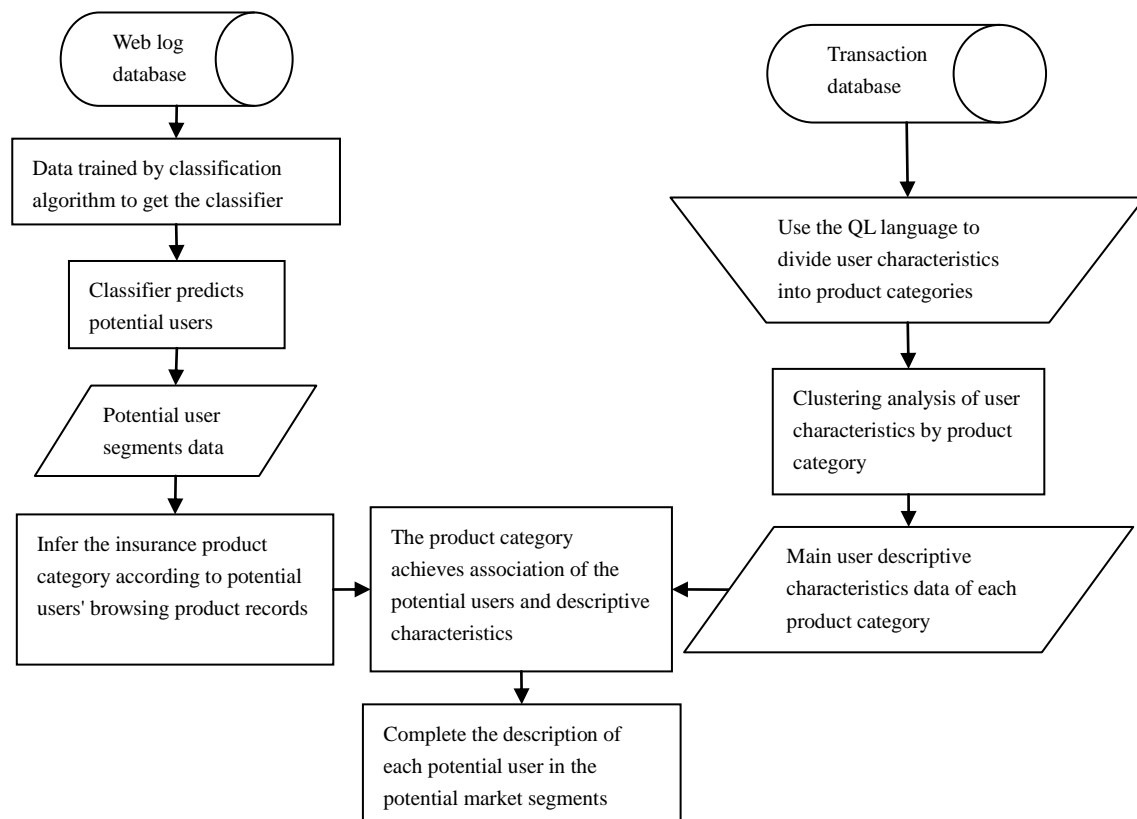


Figure 1. The model to describe characteristics of the potential user segments

3.2 Data preparation

1) Web log data

The basic data of this research is the Web log data and user purchased data on the mobile terminal. Web log data, which is called clickstream, records the user clicks on the site. Before using the classification algorithm to train data, we need to clean and transform the original data of the user clicks to get a data set which can represent the user's browsing behavior to a large extent. The data preparation process for the mobile terminal Web log data of the site has taken the following steps:

(1) Original data acquisition: We obtain data fields from Web log such as user_IP, visit_time, visit_date, vinfo, reference, request, login_id, cookie, agent, etc.;

(2) User identification: In order to accurately distinguish different users, taking into account the users' mobile terminal machine code is inaccessible, we distinguish users according to the user_IP, cookie and agent. If the cookie field is not empty, the user can be distinguished by the cookie. If the cookie field is null, then we will combine user_IP field and agent field to distinguish users. There is a need to explain that cookie means the web cache information generated by user browsing behavior and agent means the web browser used;

(3) Session identification: The user's session time is divided by the time threshold set between the two access requests, and the time threshold is set to 800 seconds;

(4) Feature extraction: We extract features based on the actual situation of the website such as number of pages, average time of each pages viewed, number of product pages viewed, average time of product pages viewed, time quantum of visiting, number of "Study" pages viewed, average time of "Study" pages viewed, number of "Topic" pages viewed, average time of "Topic" pages viewed, number of "Toptag" pages viewed, average time of "Toptag" pages viewed.

2) User purchased data

In order to obtain the descriptive characteristic field of the user in the past mobile transaction record, we associate three data tables, namely, the insurance application table, product table and order table. Table 1 shows the fields to be obtained and the fields used to filter the data.

Table 1. The fields to be obtained and the fields used to filter the data

Product table	Insurance application table	Order table
Product_ID(association)	Product_ID(association)	Order_ID(association)
<i>Product category='XX'</i>	Order_ID(association)	<i>Order Source= 1 或 6</i>
	<i>Insurance Application Status ≥ 2 and ≤ 6</i>	
	<i>Applicant Email address does not contain "xyz" or "made-in-china"</i>	
	Gender(target)	
	Date of birth(target)	
	Province(target)	
	Price(target)	

The four fields marked with "target" in the table are the target fields we want to achieve, and the four tilted fields are used to filter the data. Moreover, the fields marked with " association" are association fields associating three tables. Finally, the four descriptive characteristics of the users who have completed the payment on the mobile terminal and do not work for this e-commerce insurance company are obtained under the corresponding product category, which include users' gender, date of birth, province, and amount of payment. The data cleaning steps for the extracted descriptive fields are as follows:

(1) Considering the clustering effect and the implementation of marketing, the province field is divided into Northeast, North China, Central China, East China, South China, Southwest, Northwest and Hong Kong, Macao and Taiwan regions in accordance with the geographical location of the province, then these regions are encoded using Arabic numerals;

(2) Because the range of the payment price field value is so large that the effect of the clustering analysis will be affected, we standardize it to reduce the range value;

(3) For convenience, we change the date of birth to the form of age.

3.3 Data modeling

1) Distribution of user characteristics under the product category

It is considered that the results obtained by using K-medoids clustering algorithm are less affected by outliers than the results obtained by K-means algorithm which is a classical algorithm. Also, the clustering results from K-medoids algorithm are more understandable than that from hierarchical clustering algorithm, pedigree clustering algorithm, density clustering algorithm. Therefore, we choose the K-medoids algorithm in the clustering analysis of user characteristics under product categories. As the two most commonly used algorithms of the K-medoids algorithm, the Pam algorithm is usually used to deal with small sample data while the Clara algorithm can process the larger sample data due to its sampling technique. In the clustering analysis, we use the Pam function and Clara function of the Cluster package in R to deal with the pretreated data directly. After returning the clustering results, we can continue to use plotting function of R to show the effectiveness of the clustering results, which can help us choose Pam or Clara algorithm. It is worth mentioning that, in the face of property insurance category data, the clustering effect processed by Clara algorithm is better than the clustering effect processed by Pam algorithm. So, we choose the Clara algorithm rather than Pam algorithm to deal with the small sample data of property insurance. Table 2 shows the clustering modeling results.

Table 2. Clustering Modeling Results

Product category	Data size	Algorithm	Clustering result (gender, age, price, region, sample size)	Clustering effect
Overseas travel insurance	10756	Clara	①female, 26, 100, South China, 3451 ②female, 35, 35, East China, 5732 ③male, 50, 90, South China, 1933	①0.51 ②0.54 ③0.49
Domestic travel insurance	33089	Clara	①female, 25, 1, East China, 12568 ②female, 36, 33, East China, 12707 ③male, 47, 7, East China, 7814	①0.60 ②0.50 ③0.55
Property insurance	1928	Clara	①female, 27, 13, East China, 765 ②female, 33, 13, East China, 802 ③female, 43, 13, East China, 361	①0.52 ②0.53 ③0.64
Traffic accident insurance	9234	Clara	①female, 25, 20, East China, 2988 ②female, 33, 20, East China, 3459 ③male, 43, 8, East China, 2787	①0.52 ②0.57 ③0.25
Comprehensive accident insurance	2367	Pam	①male, 26, 100, East China, 892 ②male, 35, 100, East China, 898 ③male, 46, 100, East China, 577	①0.49 ②0.52 ③0.37
Health insurance	1449	Clara	①male, 28, 200, Central China, 562 ②male, 35, 130, Central China, 609 ③male, 47, 130, South China, 278	①0.43 ②0.47 ③0.47

Since the number of clusters is set to 3 when using the K-medoids algorithm, the user descriptive characteristics under each product category are divided into three clusters. It is worth mentioning that if the value of the clustering effect is 1, clusters are separated from each other. Besides, if the value of the clustering effect is 0, clusters are very close, and if the value of the clustering effect is between -1 and 0, the data record is assigned to the wrong cluster. Combined with the clustering results and clustering effects, the main user characteristics under the overseas travel insurance category can be described as a middle-aged woman at the age of 35 in East China whose acceptable price is about 35. The main user characteristics under the domestic travel insurance category can be described as a young woman at the age of 25 in East China whose acceptable price is about 1. The main user characteristics under the property insurance category can be described as a middle-aged woman at the age between 33 and 43 in East China whose acceptable price is about 13. The main user characteristics under the traffic accident insurance category can be described as a woman at the age between 25 and 33 in East China whose acceptable price is about 20. The main user characteristics under the comprehensive accident insurance category can be described as a man at the age between 26 and 46 in East China whose acceptable price is about 100. The main user characteristics under the health accident insurance category can be described as a man at the age between 35 and 47 in East China or South China whose acceptable price is about 130.

2) Market segmentation of potential users

When using the classification algorithm to train the Web log data that has been pretreated and transformed, the classification variable in the training data is whether the user bought the product or not. Taking into account the number of users who purchased account for a small fraction of the total number, so we consider the data as unbalanced data, making us choose the random forest algorithm to train data. Table 3 shows 5 randomly selected results of the prediction of the user's behavior features by the potential user classifier.

Table 3. Sample results predicted by potential user classifier

No.	Page-n	Page-t	Product page-n	Product page-t	Visit-t	Study page-n	Study page-n	Topic page-n	Topicp age-t	Toptag page-n	Toptag page-t	P user
31	12	78.3	2	127.2	3	2	10.27	6	30.33	1	7.45	Yes
34	4	28	1	12.5	2	0	0	0	0	0	0	No
110	11	56.5	3	63.67	4	0	0	1	13.5	2	14.5	Yes
190	8	45.6	1	20.33	1	0	0	1	1.43	0	0	No
97	4	18	0	0	2	1	5.6	0	0	1	2.33	No

The "NO." in the first column in Table 3 is the unique identifier of the user, and user's behavior features show in the second column to the twelfth column, where we make "number of pages viewed " short as "page-n" and make " average time of viewing pages " short as "page-t" for convenience. Besides, the "P user" in the last column is the result of the prediction of the potential user classifier. The accuracy of the potential user classifier is 97.3% by test, which can be used to predict the potential users preferably.

3) Characteristics description of potential user segments

Through the prediction of the potential user classifier, the web browser users are selected to obtain the potential market segments. For each potential user in the potential market segment, we determine the product category of the user's interest by the product category of the product page viewed. Since the main user characteristics under each product category have been obtained, the product category can be used as an intermediary to realize the mapping of each potential user and the descriptive characteristics, making the characteristics description of all potential users in the potential user segment come true. Table 4 shows a random sample of 5 potential users and their corresponding descriptive characteristics.

Table 4. The sample of characteristics description of potential users

NO. (User identification number)	Referrer-site	Login-id	Product category of the product page viewed	Product category of interest	Descriptive characteristics
160	m.xyz.cn	-1	Overseas travel insurance	Overseas travel insurance	Middle-aged women at the age of 35 in East China whose acceptable price is 35.
186	m.xyz.cn	-1	Overseas travel insurance	Overseas travel insurance	Middle-aged women at the age of 35 in East China whose acceptable price is 35.
98	m.xyz.cn	-1	Domestic travel insurance	Domestic travel insurance	Young women at the age of 25 in East China whose acceptable price is 1.
178	m.sogou.com	-1	Property insurance	Property insurance	Middle-aged women at the age of 35 and 43 in East China whose acceptable price is 13.
96	m.baidu.com	-1	Overseas travel insurance	Overseas travel insurance	Middle-aged women at the age between 35 in East China whose acceptable price is 35.

The " Referrer-site " in the second column refers to the source page of the potential user, and if the field value is "m.xyz.cn", it means the user enters the website by inputting the URL. Moreover, if the field value is other websites, it means the user enters the website through other websites. The "Login-id" represents the user's login account, and the value being -1 indicates that the user is not logged in. The content in the fourth to the sixth column indicates that the product category acts as an intermediary to associate descriptive characteristics with potential users.

4. MARKETING STRATEGY IMPLICATIONS

The section 3 presents the process and results of data mining for the actual data of the insurance website under the guidance of the model. Based on the modeling results, we obtain some implications on the marketing strategy of this website. Picture 2 is the flow chart of the precision marketing program based on the characteristics of the potential users.

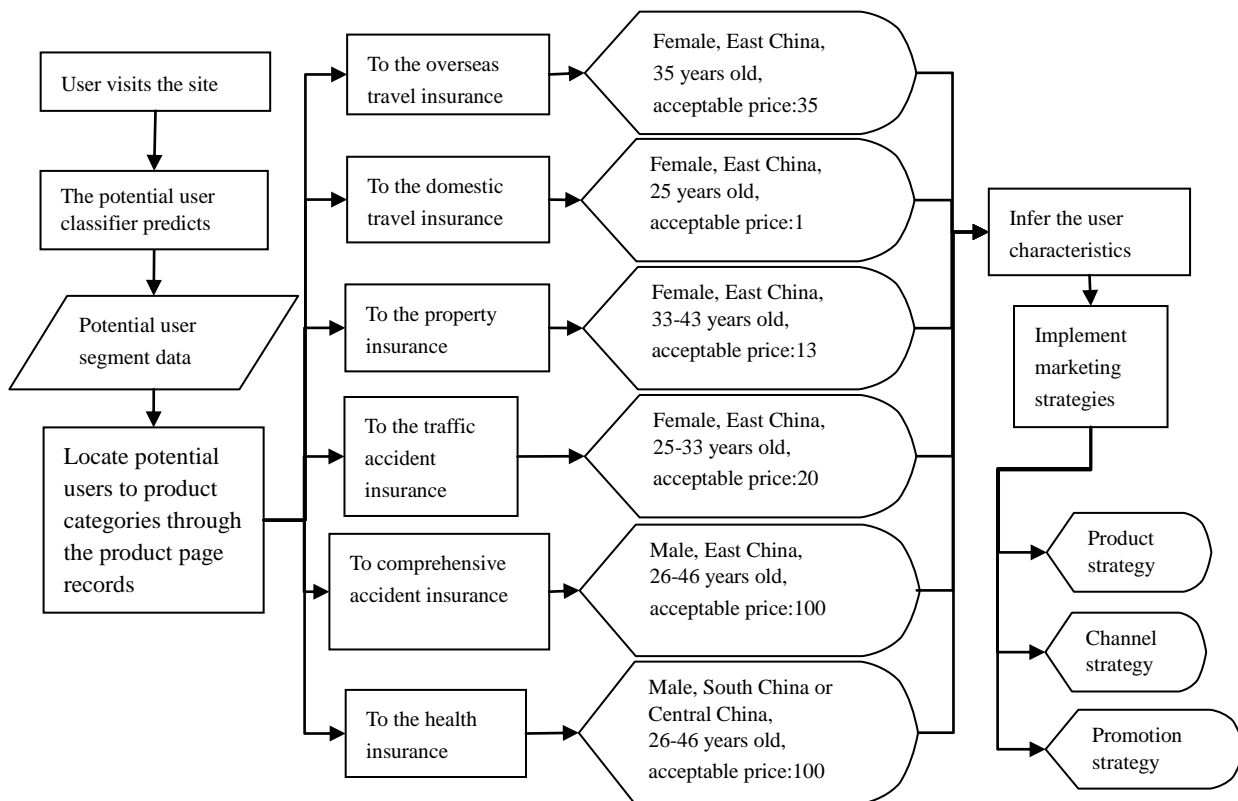


Figure 2. Precision marketing program based on the characteristics of the potential users

As can be seen from the figure 2, when the user is browsing the web site, the potential user classifier is used to find the potential users. For the identified potential users, the website locates potential users to the insurance product category based on their product page browsing records. Due to the main characteristics of users under the insurance product categories have been known, the descriptive characteristics of potential users can be obtained, so as to push the marketing measures to meet their needs. With regard to the specific measures for the implementation of marketing measures, we give some marketing thoughts from three angles of product, channel and promotion combined with 4P theory.

4.1 Product strategy

Taking into account the fact that the products sold by the e-commerce website are the specific insurance agreements and claims or customer service when an accident occurs or the customer has doubts about the insurance products, the description of the product in the product details page can have a great impact on the user's purchase decision. However, the columns shown on the product page of this website, such as the applicable people, product interpretation, sharing case and so on can only provide some basic introduction about the product. What's more, when customers have questions about the product, they will call the customer service to ask for related information, and customer service performance can largely determine whether the user to buy. The marketing implications are as follows:

(1) In the columns on the product pages, such as the applicable people, product interpretation, sharing case and so on, the information offered should add more description associated with the user characteristics without

violating the terms of the insurance products, thus prompting users to have emotional resonance. For example of "overseas travel insurance", a case that one woman named Zhang at the age of 35 from Jiangsu has bought about 35 yuan related products can be presented in the "sharing case" column;

(2) According to the potential user's gender and region, the website can provide users with the customer service who is in different gender and from the same region, increasing the possibility of moving users. For example of "domestic travel insurance", when the potential users interested in domestic travel insurance are asking the customer service insurance details, as the main characteristics under this product category are the women whose acceptable price is about 1 yuan, the customer service can be a Shanghai man with a charming voice and recommend the low-priced products when asked product selection.

4.2 Channel strategy

The appropriate channel strategy can help users find their own content with high efficiency when browsing the entire site. We find that the current layout of this insurance website is not targeted in consideration of user needs and the sort of the product is according to sales. So the marketing implications are as follows:

(1) The user's acceptable price and product sales should be combined to sort the product list. For example of "property insurance", the acceptable price of the main characteristics under this product category is about 13 yuan. Therefore, to the potential users interested in property insurance, the products whose price is about 13 yuan will rank in the forefront and be sorted by the previous sales when they click on the product list;

(2) Product recommendation on the website home page is supposed to consider potential users' acceptable price.

4.3 Promotion strategy

Promotion strategy includes the push strategy and the pull strategy. The insurance website in the face of individual consumers mainly use the pull strategy, and the most typical type of pull strategy is advertising and coupons. So, the marketing implications are as follows:

(1) The website should add a characteristics description associated with potential users when pushing advertising projects to attract potential users. For example of "traffic accident insurance", the main characteristics under this product category are women at the age between 25 and 33 whose acceptable price is about 20. So, to the potential users interested in the traffic accident insurance, the website can plan an advertising project in which 20 yuan or other low-priced products can be recommended and some words like "affordable" women prefer will be used and cute color should be adopted as well.

(2) The website can push coupons based on information about potential user's acceptable price. For example, the acceptable price of the main characteristics of overseas travel insurance is about 35 yuan, so the website can push the 5-yuan coupon available only when the product is over 40 yuan.

5. CONCLUSION AND PROSPECT

In order to make the precise marketing of e-commerce websites possible, we combine the two parts which are the potential user market segmentation based on behavioral segmentation variables and the market segmentation based on the descriptive segmentation variables from the angle of multi-dimensional market segmentation in this paper, and we propose a novel model to describe characteristics of the potential user segment. In this model, the main user characteristics under different product categories are obtained by K-medoids algorithm, and the potential user classifier is trained by the random forest algorithm. To the potential user predicted, we infer the insurance product category that is of interest to them according to their browsing product records, so as to give descriptive characteristics to potential users. Under the data support of the insurance e-commerce platform in Nanjing, the idea of the model proposed in this paper has been into practice, helping us gain some implications for marketing of the insurance e-commerce website. It is to be observed that

although the data is derived from the certain field of insurance, the two data sources which are the log data and the user transaction data as well as the product category considered as the key of the model all exist in the multi-category e-commerce websites. So, the model is also applicable to other types of e-commerce websites. In future research, we hope to add behavioral characteristics to potential users on the basis of the model proposed to increase the level of detail of characteristics description, leading to more accurate marketing strategies.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71271115.

REFERENCES

- [1] China e-Business Research Center. (2016-11-04). 2016 Annual Report and Guide on China's Consumers Online Shopping. http://www.100ec.cn/zt/upload_data/wlxf.pdf.
- [2] Liu Lei, Chen Xingshu, Yin Xueyuan, Duan Yi, Lü Zhao. (2011). Network user identification based feature weighting naive Bayesian classification algorithm. *Journal of Computer Applications*, 36(12): 3268-3270. (in Chinese)
- [3] Kim K, Ahn H. (2008). A recommender system using GA K-means clustering in an online shopping market. *Expert Systems with Applications*, 34(2): 1200-1209.
- [4] Liu S Q, Mattila A S. (2017). Airbnb: Online targeted advertising, sense of power, and consumer decisions. *International Journal of Hospitality Management*, 60: 33-41.
- [5] Barnes S J, Bauer H H, Neumann M M, et al. (2007). Segmenting cyberspace: a customer typology for the internet. *European Journal of Marketing*, 41(1/2): 71-93.
- [6] Soopramanien D G R, Robertson A. (2007). Adoption and Usage of Online Shopping: An Empirical Analysis of the Characteristics of "buyers" "browsers" and "non-internet shoppers". *Journal of Retailing and Consumer Services*, 14(1): 73-82.
- [7] Guo Peipei, Fang Zhaoben. (2010). Web log mining and potential customer discovery based on SVM. *Journal of Industrial Engineering and Engineering Management*, 24(1): 129-133. (in Chinese)
- [8] Ozyirmidokuz E K, Uyar K, Ozyirmidokuz M H. (2015). A Data Mining Based Approach to a Firm's Marketing Channel. *Procedia Economics and Finance*, 27: 77-84.
- [9] Liu Kan, Yuan Yunying, Liu Ping. (2015). A Weibo Bot-users Identification Model Based on Random Forest. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 51(2): 289-300. (in Chinese)
- [10] Amaro S, Duarte P, Henriques C. (2016). Travelers' use of social media: A clustering approach. *Annals of Tourism Research*, 59: 1-15.
- [11] Wu R S, Chou P H. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. *Electronic Commerce Research and Applications*, 10(3): 331-341.
- [12] Liu Zhengyu. (2007). Study on Precision Marketing Method. *Journal of Shanghai Jiaotong University*, 41(S1): 143-146+151. (in Chinese).
- [13] Guan Zhixin, Liu Yin, Wang Qiuwen. (2014). The Recognition of Individual Customer's Identity of Commercial Bank—A Study Based on the Logistic Model. *Finance Forum*, 19(7): 70-75. (in Chinese).
- [14] Liao S H, Chen J L, Hsu T Y. (2009). Ontology-based data mining approach implemented for sport marketing. *Expert Systems with Applications*, 36(8): 11045-11056.
- [15] Tang H, Liao S S, Sun S X. (2013). A prediction framework based on contextual data to support mobile personalized marketing. *Decision Support Systems*, 56: 234-246.