Cross-border Consumer Research Based on Cluster Analysis of Aliexpress Behavioral Data

Yilun Wu and Lei Yao[*]
Business school, Beijing Institute of Fashion Technology, Beijing, China
[*]Corresponding author

Abstract—**This article selects the search data of Aliexpress cross-border e-commerce platform in wedding dress industry. The behavior of consumers will leave traces on the platform. Usually expressed as search index, purchase index and the number of compared products. Through the analysis of correlation, regression and cluster, discover the relationship among consumer behavior of search, comparative and purchase. Consumer is divided into three categories by the cluster analysis of behavior data. The corresponding recommendations will be given to each categories.**

Keywords-cross-border e-commerce; behavioral data; consumer research

I. CONSUMER BEHAVIOR PATH AND DATA GENERATION

In general, the demand for the purchase is a prelude to consumer behavior, then they search for the desired goods on the e-commerce platform. By browsing and comparing the product details page of similar products, they have a buying behavior.

Three main consumer behaviors can be found in the path of consumer purchase behavior, namely search behavior, browsing behavior, comparative behavior and purchase behavior. At the same time, these four acts will be recorded in the cross-border E-commerce platform. The background of the site will record a user search behavior as keyword A has been searched for once, then calculate the number of products related to the keyword A is B. The user compares the products to be searched and decides whether to click into the details page. The system also records whether the user clicks the search results. After the user search keywords, whether the user purchase it will also be recorded. After thousands of people search, click and buy, all the record are going to generate as a variety of indicators. According to the key word to be found, the background can be counted out the visit number of keywords, generated search popularity index. Search index is based on search frequency. According to the number of hits to generate hits index. According to the number of key words to generate the number of payment index. According to the products number searched out by the keywords to generate competition index. In summary, the search behavior of the platform user generated search popularity index and search index. The user comparison behavior is related to the competition index and hits index. Purchase behavior generates payment index. Therefore, the relationship between the indicators can be inferred from the relationship between the behaviors of cross-border e-commerce consumers.

II. DATA SOURCES AND PROCESSING OF ALIEXPRESS

Aliexpress platform records consumer behavior data constantly, calculates the corresponding index daily and publishes them in the platform of the data analysis panel for vendors on the platform to view, download and analysis. There are three columns of data in the panel, which are trade information, search word analysis and product selection experts. The search word analysis is based on the keyword statistics of corresponding consumer behavior data. Platform support to download summary data of. The latest 30 days. Therefore, based on the monthly data from April 2016 to July, this paper analyzes the relationship between the indicators of the behavior data.

At the beginning of each month, you need to download the latest 30 days' search data of search word analysis column for the wedding category, then summarize the data for 4 months. At the same time, assuming that if the corresponding indicators are not 0, the word is the key word to retain the integrity of the user behavior. The following analysis is based on the analysis of the high- quality cross-border consumer behavior data. Excluding the index value of 0, finally retain 114 high-quality keywords.

III. DATA ANALYSIS RESULTS OF ALIEXPRESS

A. Correlation Analysis

Carrying out a pairwise combinatorial correlation analysis of all variables, to judge whether there is linear correlation based on Pearson coefficient and to see if there is significant correlation by significance level. So that you could see if there is a correlation between consumer behaviors. Choose the better representative index to carry on the further regression analysis. The greater the absolute value of the Pearson correlation coefficient, the stronger the correlation will be. The closer the correlation coefficient is to 1 or -1, the stronger the correlation is. The closer the correlation coefficient is to 0, the weaker the correlation is. The correlation coefficient is 0.8-1.0 means that the relation is strong, 0.2-0.4 means weak and 0-0.2 means irrelevant.

TABLE I. BILATERAL PEARSON CORRELATION COEFFICIENT BETWEEN VARIABLES

		Search pop.	Search freq.	Click rate	Hits pop.	Payment covers..	Payment Pop.	Compe t. index
Search popularity	Pearson	1	.990**	.017	.991**	-.185*	.858**	-.158
	Sig.(Bi)		.000	.857	.000	.049	.000	.094
Search frequency	Pearson	.990**	1	.054	.995**	-.190*	.877**	-.131
	Sig.(Bi)	.000		.570	.000	.043	.000	.163
Click rate	Pearson	.017	.054	1	.089	.098	.107	.341**
	Sig.(Bi)	.857	.570		.349	.302	.259	.000
Hits popularity	Pearson	.991**	.995**	.089	1	-.173	.896**	-.123
	Sig.(Bi)	.000	.000	.349		.066	.000	.193
Payment conversion	Pearson	-.185*	-.190*	.098	-.173	1	-.105	.018
	Sig.(Bi)	.049	.043	.302	.066		.268	.848
Payment popularity	Pearson	.858**	.877**	.107	.896**	-.105	1	-.056
	Sig.(Bi)	.000	.000	.259	.000	.268		.553
Competition index	Pearson	-.158	-.131	.341**	-.123	.018	-.056	1
	Sig.(Bi)	.094	.163	.000	.193	.848	.553	

Note：**. .01 level (bilateral) significant correlation. *. 0.05 level (bilateral) significant correlation.

It is not difficult to see from the results of correlation analysis that there is a strong correlation between the two indicators that represents of consumer search behavior, the number of clicks to indicate consumer clicks and the Payment popularity of the purchase behavior. It shows that there is an association between different behaviors. Click rate and payment conversion ratio have weak correlation with most other indicators, but click rate and search popularity shows a strong correlation. There is a strong correlation between payment conversion and competition index. Competition index is also has a weak correlation with most of other indicator. So that we can choose search frequency to represent search behavior, choose hits popularity to represent click behavior and choose payment popularity to represent payment behavior. Meanwhile, competition index is not involved in regression analysis.

B. Regression Analysis

According to the results of correlation analysis, there is a linear relationship between the number of clicks and search index, so that a linear regression equation is established between the two variables. Early behavior is independent variables. Post action is the dependent variable. According to the correlation coefficient of Pearson, there is a strong correlation between the hits popularity and the payment conversion. There is a strong linear correlation between the payment conversion rate and the competition index.
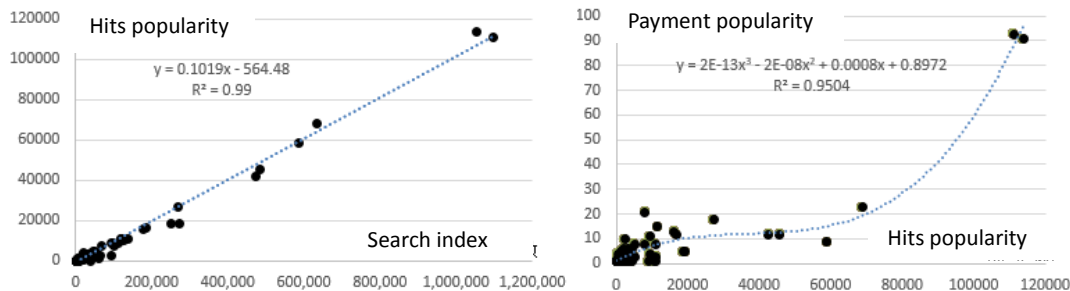

FIGURE I. REGRESSION ANALYSIS

According to the scatter plot of two pairs of variables, it is found that the search index and the hits popularity have the obvious linear regression relationship. While the hits popularity and the payment popularity show polynomial curve regression relationship. With the hits popularity as the dependent variable and the search index as independent variables, we can establish a monadic regression function. At the same time, with the payment popularity as the dependent variable and the hits popularity as the independent variable, we can set up a univariate cubic regression equation.

C. Cluster Analysis

Search index which represents search behavior, click popularity that represents click behavior, competition index indicates comparison behavior and the payment popularity stands for purchase behavior will conduct a K- mean clustering analysis. First, we need to draw up a tree diagram based on systematic cluster analysis, so we can infer that the data can be divided into three to five categories. Then, the original data are divided into three, four and five categories successively by K-means clustering analysis. At last the four factor as the independent variable and the category number as the dependent variable to conduct ANOVA, the clustering results of four

indexes which are divided into three categories is the most effective results.

TABLE II. FINAL CLUSTERING CENTER

	1st Class	2nd Class	3rd Class
Search index	542915	1072974	35717
Click popularity	53834	112279	3046
Payment popularity	14	92	3
Competition index	38	86	98
N cases	4	2	108

Final clustering center table for three categories K- means clustering analysis is shown in Table 2. The central value and cases number of each class is given in the Table 2. ANOVA result is given in the Table 3, the last column shows that indicators have reached a significant level in addition to the index of competition.

TABLE III. ANOVA RESULT

		df	F	Sig.
Search index	SSA	2	505.291	.000
	SSE	111		
	SST	113		
Click popularity	SSA	2	670.182	.000
	SSE	111		
	SST	113		
Payment popularity	SSA	2	1.767	.176
	SSE	111		
	SST	113		
Payment popularity	SSA	2	590.839	.000
	SSE	111		
	SST	113		

IV. CHARACTERISTICS AND CLASSIFICATION OF CONSUMER BEHAVIOR

A. Characteristics of Consumer Behavior

According to regression analysis, the search behavior, browsing behavior, comparative behavior and purchase behavior of consumers are interactive. Due to the positive relationship of search index and click popularity, the frequency of clicks behavior increase follow the increasing number of searching behavior, which shows that the more consumers search the greater the possibility of browsing. Consumers' purchasing behavior is positively influenced by browsing and searching behavior, the more you browse the more likely you are interested in, and the more likely you are to buy. When the number of consumers at a low level, the purchase volume is very little. The number of purchases increases with the increase of the amount of browsing. Consumers generally do not buy on the platform with little browse behavior. When the number of visits increased slightly to a moderate level, browsing behavior hardly generates purchase. When the volume reached a high level, search behavior can lead to a large number of consumers purchase.

B. Consumer Classification Based on Cluster

Through the interpretation of the final cluster center table, combining search terms, can summarize the characteristics of each classification and infer classification of consumer shopping habits.

The first class of consumers have little optional product to choose from, as for the word they searched has a smallest competitive index. Through moderate search and browsing, bring a moderate amount of buying behavior. They are consumers with a clear target who are going to search for some specific product. Part of the consumer is from minority language countries.

The second class of consumers have comparatively abundant optional product. Through huge amount of search and browsing, bring a lot of buying behavior. They are browsing and search oriented consumers. They search the popular vocabulary and compare a variety of products by constantly browsing the product on the platform, which result in a large number of purchases.

The number of consumers in the third class is the largest. There are a lot of products to choose from. Consumer search and browsing is very low, resulting in very low volume of purchase. This type of consumer is not interested in purchasing on the platform. They also search the popular word, but the searching result are not attractive.

V. CONCLUSION

Comprehensive analysis of the above classification results, the number of the third class is the largest, indicating the existence of a large number of consumers on the platform is not interested in the product. Consumers just like to occasionally stroll on the platform and then leave soon, so that products on the platform need to be optimized. Consumers with a clear target and browsing oriented consumers are the crowd is able to produce the purchase. These people are someone who need to be emphasized on by seller on the platform. Sellers need to optimize the display of pictures to attract consumers and to design better product descriptions to facilitate consumer search.

REFERENCES

[1] Zhang Wenshan, Wang Jun, Xu Shanchuan. Cluster analysis of electricity supplier user demand -- a case study of Taobao [J]. Modern library and information technology,2015,03:67-74.

[2] Donald J. Hempel, Subhash C. Jain. House Buying Behavior: An Empirical Study in Cross-Cultural Buyer Behavior [J] .American Real Estate and Urban Economics Association, 1978, (1): 1-21.

[3] Hong Chenyi,Piao Zaixiu,Huan Chunhua. An analysis of the impact of Internet information search intention on consumer behavior [J]. International Business (Journal of University of International Business and Economics),2003,06: 57-61.

[4] Liu Meilian. Research on consumer behavior in e-commerce[D] Huazhong University of Science and Technology,2005.

[5] Zhang Guofa, Shao Shuqin. Co-integration analysis of Taobao search index and turnover index: a case study of down jacket [J]. Modern economic information,2014,03:293-296.