

Research on Data Selection of Cross-Border Retail E-Commerce Enterprises from the Perspective of Consumer Search Behavior—Take AliExpress, a Cross-Border E-Commerce Platform, as an Example

Zizhao Zhan^{1,a}, Zhan Li^{2,b*}, and Mingli Guo^{3,c}

¹Alibaba Business College, Hangzhou Normal University, Hangzhou, Zhejiang 311121, China

²Alibaba Business College, Hangzhou Normal University, Hangzhou, Zhejiang 311121, China

³Alibaba Business College, Hangzhou Normal University, Hangzhou, Zhejiang 311121, China

*Corresponding author. Email: google0011@sina.com

Keywords: *cross-border retail, search behavior, data selection, C2B model*

ABSTRACT. In the digital economy era, under the background of increasingly short-term and fragmented consumer demand, cross-border retail e-commerce has gained tremendous development space. When cross-border retail companies go overseas, facing the complex overseas environments and differentiated product demands, data selection can help companies accurately grasp consumer demand. This article studies consumer search behavior data on AliExpress on the cross-border retail e-commerce platform. Based on the analysis of cross-border retail consumer search behavior and the traffic paths it generates, both two indicators were analyzed by Pearson correlation and linear regression analysis. Based on the quantitative relationship between the data indicators, from the perspective of search behavior, data selection research is conducted. Data selection has the nature of the C2B model, providing consumers with more personalized products to meet consumers' differentiated needs, and changing the weak position of consumers in the traditional single selection method. Enterprises meet consumer needs through data, and provide consumers with precision products and services. Therefore, in the context of the rapid development of the digital economy and the digital empowerment of enterprises to promote the transformation of the industry, the research on data selection of cross-border retail e-commerce companies has certain feasibility and practical value.

1. INTRODUCTION

Under the trend of global digital economy development, the demand for bulk commodities is obviously insufficient, consumer demand is developing towards small trade, and trade demand is becoming more short-term and fragmented. Combined with the development of Internet e-commerce related technologies, cross-border online retail is gaining a huge space for development. More SMEs engaged in foreign trade retail business have entered the global market through cross-border retail e-commerce platforms[1]. The size of the global cross-border retail e-commerce market is expected to reach nearly US \$ 1 trillion in 2020. At that time, the total number of global cross-border retail e-commerce consumers will exceed 943 million, forming a group of strong digital consumers[2].

Cross-border export retail e-commerce business is in a golden development period. Through the cross-border retail e-commerce platform, small and medium-sized enterprises combine e-commerce and international trade, so that consumers and enterprises can be connected across borders. Global online retail e-commerce consumers can find products that match their needs in a broad market. Driven by China's small and medium-sized commodity manufacturing cost advantages and supply chain advantages, the Asia-Pacific region will become the largest cross-border import and export e-commerce region. In the age of consumption classification, with a large number of small and medium-sized foreign trade companies engaged in cross-border retail e-commerce businesses, products which are made in China have ushered in new international opportunities and sold to different overseas markets. AliExpress, as China's largest cross-border e-commerce platform, provides a one-stop service for logistics, payment and customer management marketing, in addition to providing a trading

platform for SMEs[3]. At present, AliExpress has covered 230 countries and regions, and the number of overseas transaction buyers has exceeded 1.5 billion[4].

In the cross-border retail e-commerce business, SMEs face the complicated overseas market environment and the consumption habits and preferences of different regions. How can they more effectively and accurately grasp the user's demand for products, improve the user purchase experience, and reduce merchant costs become the first issue cannot to avoided. How to choose the product that suits you best, how to stand out from homogeneous competitors and improve the conversion rate, the key to solving these problems is to meet the needs of consumers. Data selection has the nature of the C2B model, changing the weak position of consumers' passive acceptance in the traditional selection mode. Consumers have become the core of corporate selection. Companies provide personalized products and services based on the differences between consumers. With the help of consumer search behavior data provided by platform service providers, consumers can perceive consumers' real-time and real search demands and consumer demands in cross-border retail e-commerce activities. Therefore, data selection based on consumer search behaviors empowers companies to link consumption through data. Enterprises surround consumers with demand-oriented and provide personalized product services to cross-border retail e-commerce consumers, while also reducing product demand mismatched risks and costs.

2. Research status of cross-border retail e-commerce company selection

Research on selection of cross-border retail e-commerce is subject to the influence of various cross-border e-commerce platform rules, and domestic research on selection strategies is at an immature stage. In 2016, Pingping Wu proposed a company's marketing and product selection strategy based on the platform on the basis of positioning different target markets[5]; In 2018, Shuyan Zhu proposed seven strategies on the export of cross-border products from the perspectives of cultural background factors, platform technology factors, price and profit factors, risk factors, quality brand factors, and cross-border B2C factors[6]; In 2018, Zhifang Hu put forward suggestions for selecting products from the perspectives of selection value, platform technology, product market, and platform rules[7]. This article adopts the C2B model to change the weak position of consumers in the traditional selection model. Focusing on consumer demand, focusing on the search behavior of cross-border retail e-commerce consumers, this paper proposes data selection for cross-border retail e-commerce enterprises.

The so-called search behavior is the purposeful information-gathering behavior of people, and it is the process of people searching for information purposefully based on the needs of something or cognition[8]. Because of the characteristics of initiative and purpose, search behavior has extensive research value in network marketing. Once consumers have a demand for a certain product or service, they will collect relevant information based on the demand. Beel et al. They believe that the different motivations and uncertainties of demand and purpose, the complexity of product attributes, and the arrangement of search results will affect the results of people's information acquisition[9]. Broder et al. They divided people's use of search engines into three categories based on consumer search purposes: navigational, informational, and transactional[10]. Behind the different search behaviors are consumers' further segmented search goals and needs. In the field of purchase search behavior, Moorthy et al. They found that consumers will often search the Internet before purchasing activities begin[11].

In the era of web3.0, search behavior has become the most common behavior of consumers. Therefore, in the cross-border retail e-commerce business, from the perspective of consumer search behavior, SMEs conduct search behavior data of consumers on the AliExpress platform. Analyze and focus on the real demands of consumers, and put forward data selection recommendations for SMEs in cross-border B2C e-commerce business.

3. Cross-Border Retail E-Commerce Consumer Search Behavior

3.1. Cross-border retail e-commerce consumers' search behavior at different stages

In cross-border retail e-commerce activities, consumers show regular behavior choices and follow established behavior patterns. As shown in Fig.1, consumer behavior patterns are divided into three phases: understanding phase, browsing phase, and purchasing phase[12]. Consumers show specific search demands and behaviors at different stages. Consumer search behavior is essentially a manifestation of consumers' demands and purposes at all stages of their online shopping behavior model: there are search demands that want to understand product information; search demands that analyze and compare product performance; and search demands that hope to achieve precise transactions.

In the understanding stage, the starting point of the consumer's search behavior is to generate a certain problem, and then to solve the problem. In the process of finding a method, the consumer will favor those that are easy to understand and comprehensive about the problem solving method. Therefore, consumers' search terms will focus on some keywords that directly reflect the nature and functionality of the product. After the consumer entered the browsing stage, at this time, he had found a solution to the problem, clarified the type of product he needed to purchase, and started a comparative study of product price or brand characteristics. Compared to the previous stage of understanding, consumers' current search demands are to analyze and compare the brand and cost-effectiveness of the products. The search keywords will also focus on the product prices, brands and other keywords. After clarifying the specific target products, consumers entered the purchase phase. During the purchase phase, consumers' behavior patterns include actions such as putting products into a shopping cart, paying, and actually completing a purchase. At this stage, consumers' search demands are directly hoped to complete the purchase of the target product through a short operation path, and to perform accurate keyword search on the product. The search results generated by the search behavior at this stage often have a high conversion intention.

Research on the search behavior of consumers at different stages can help companies to perceive the search demands of consumers at different stages, further grasp the specific needs of consumers, and conduct data-based selection analysis based on this to achieve the purpose of precision marketing.

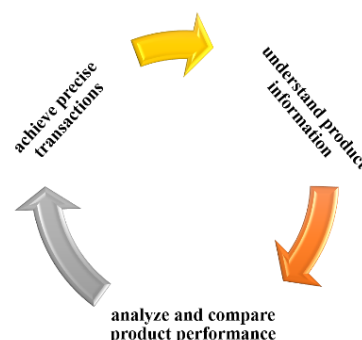


Figure 1 Consumer behavior model of cross-border retail e-commerce

3.2. Cross-border retail e-commerce consumer search behavior traffic path

In the cross-border retail e-commerce business, after understanding consumers' search behaviors at different stages, they must further track the traffic paths generated by consumer search behaviors on corporate websites. The traffic indicators generated by the consumer behaviors provided by enterprises through the website can map the consumer's search behaviors at different stages and their related demands, and provide data support for corporate selection strategies. Every relevant search behavior of cross-border retail e-commerce consumers on the platform is classified and recorded by the platform, presented in the form of data, and organized into continuous time hot search data tables. As shown in Fig. 2, companies can use these data for analysis and research to track and discover the

commonality of consumer search behaviors and consumer demand on the platform, and then perceive their more precise needs at each consumption stage. Provide practical data support for companies to select products.

Search Analysis			
Timeframe (USA Time Zone) 2019-12-13 ~ 2019-12-19			
<div>Hot search</div> <div>Soaring word</div> <div>Zero word</div>			
Keyword	Whether brand original words	Search popularity	
купальник	N	114,639	
купальник женский	N	66,133	
bikini	N	91,890	

Figure 2 AliExpress Hot Search Data

According to the data classification in the AliExpress data aspect, it can be known that the traffic paths generated by corporate stores to obtain cross-border retail e-commerce consumer search behavior are: search results page, product details pages, store homepages, and shopping cart page. As shown in Fig.3, the traffic paths of these four types of pages are the mapping of consumers' search behaviors and demands on the platform to a certain extent. In the understanding stage, that is, the stage of understanding product information, consumers need to search and collect product information as comprehensively as possible, so a large amount of search page click data and product detail page click data will be generated. Consumers at the browsing stage analyze and compare the prices of the same type of products through search behaviors, resulting in click traffic such as product detail pages, store homepages, and search results pages. After clarifying the products that need to be purchased, consumers in the purchase stage, in addition to the source of shopping cart click traffic brought by the search behavior of the shopping cart settlement stage, may also generate traffic clicks on store favorite pages and store homepages. These both clicks help consumers complete product purchases quickly. The AliExpress platform aggregates all traffic data generated by the cross-border retail e-commerce consumer search behavior and further divides it into: search popularity, search index, click-through rate, competition index, browsing-payment conversion rate and other indicators. For companies, this data is the aggregation of consumer search behaviors and search demands under the specified category. This data is analyzed statistically and a selection strategy is formulated based on this data. From the root of the data, a selection strategy that caters to consumers' search behaviors can meet the needs of consumers to the greatest extent, and can further reduce the cost and waste caused by invalid selection of enterprises.

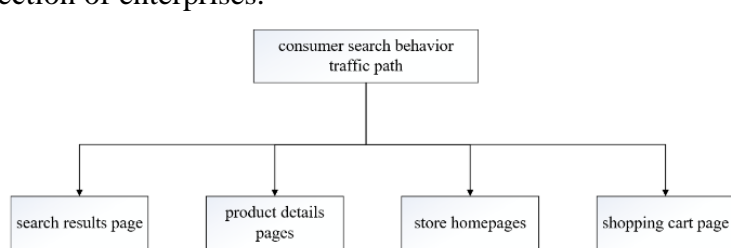


Figure 3 Cross-border retail e-commerce consumer search behavior traffic path

4. Consumer Search Behavior Data Analysis

4.1 Data preprocessing

The data comes from AliExpress's backstage-"data aspect" hot search statistics in the sports and entertainment category in the past 30 days. To ensure the reliability of the experimental data, the data is processed as follows: Firstly, the selected data time range is 2019.02.23 to 2019.03.25, which

avoids the traffic congestion and excessive invalid traffic caused by the AliExpress platform promotion at the end of March ; Secondly, delete the indicators that contain the brand in the data, ensure that the data used are keywords that do not contain brand information, and exclude the impact of brand effects on search results; Furthermore, the indicator "country distribution" in the data is deleted. This data processing analysis does not involve country distribution; Finally, the option of zero indicators in the data is deleted to ensure that the data indicators used in the analysis contain realistic behavioral significance. After the above processing, 501 data indicators are finally left.

The pre-processed data reflects the actual indicators of sports and entertainment in the past thirty days, and each piece of data contains practical significance.

4.2 Correlation analysis

The data indicators are processed, and Pearson correlation analysis is used between each two indicators. Finding valuable indicators and further linear regression analysis, the specific regression relationship between the indicators is obtained.

Table 1 Pearson correlation analysis results

Correlation						
		Search popularity	Search index	Click-Through-Rate	Competition index	View Pay Conversion Rate
Search popularity	Pearson correlation	1	.929**	.030	-.163**	-.063
	Saliency (bilateral)		.000	.497	.000	.160
	N	501	501	501	501	501
Search index	Pearson correlation	.929**	1	.061	-.045	-.049
	Saliency (bilateral)	.000		.171	.313	0.271
	N	501	501	501	501	501
Click-Through-Rate	Pearson correlation	-.030	.061	1	.251**	.357**
	Saliency (bilateral)	.497	.171		.000	.000
	N	501	501	501	501	501
Competition index	Pearson correlation	-.163**	-.045	.251**	1	.066
	Saliency (bilateral)	.000	.313	.000		.138
	N	501	501	501	501	501
View Pay Conversion Rate	Pearson correlation	-.063	.049	.357**	.066	1
	Saliency (bilateral)	.160	.271	.000	.138	
	N	501	501	501	501	501

**Significant correlation at the 0.1 level (both sides).

As can be seen from Table 1, the significant correlation between the search index and search popularity at the 0.1 level (both sides) is extremely strong (0.929). Because the search index refers to the frequency of searches, this correlation indicates that consumers' search behavior for hot search products is repetitive and convergent. For highly popular products on the head, it will attract consumers to repeat in depth multiple times. It is understood that to a certain extent, the characteristics and properties of head products will affect consumers' initial impressions of such products. The subsequent regression analysis will further reveal the specific regression relationship between search popularity and search index.

The significant correlation between search popularity and competition index at the 0.1 level (both sides) is a very weak correlation (0.163). Because the competition index is composed of the ratio of the search index to the current number of babies, the very weak correlation between the search popularity and the competition index indicates that although the head products have sufficient popularity, they are subject to factors such as conversion rate and profit margin. The product market has set a threshold, but a large number of blue ocean products in the long tail market have attracted the attention of enterprises.

The significant correlation between click-through-rate and the competition index at the 0.1 level (both sides) is a weak correlation (0.251). On the one hand, it shows that consumer demand is diversified, and products with high competition indexes often do not get high click-through rates due to fierce homogeneous competition. Within the business field of vision.

A significant correlation between click-through rate and conversion rate at the 0.1 level (both sides) is a weak correlation (0.357). A high click-through rate does not mean that a high conversion rate is

generated. It means that after consumers have clicked to understand or analyze the product, the fate of subsequent actions is diversified, and it also indicates that there is a high conversion rate but click popularity General product.

4.3 Correlation analysis

Select two indicators of search popularity and search index to make a scatter plot. The scatter plot is shown in Fig. 4. It can be seen from the figure that these two variables have a strong linear relationship, and one-variable linear regression can be used to fit the two variables.

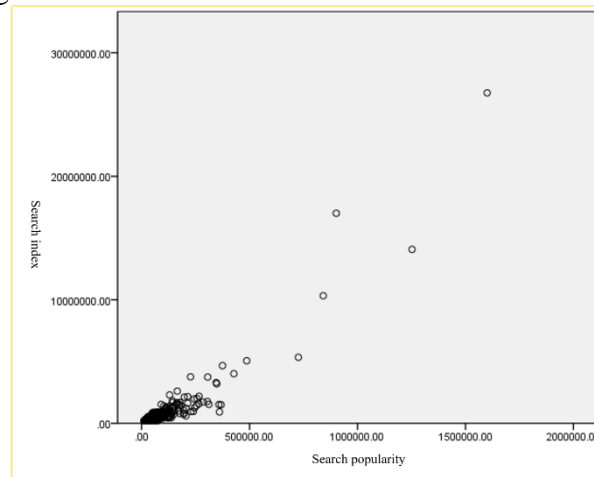


Figure4 Scatter plot

As shown in Table 2, it can be known that the goodness of fit test of the regression equation shows the R (correlation coefficient), R^2 (decisive factor), R^2 (adjusted coefficient of determination), and estimated standard error in the table. This information reflects the strength of the linear correlation between search popularity and search index. $R = 0.929$, this shows that there is a strong correlation between search popularity and search index. $R^2 = 0.862$, This illustrates that x (search popularity) can explain 97.9% of the variance of y (search index). The meaning of the notes below the table: a . Predictor is "Search Popularity"; b . The dependent variable is the "search index".

Table 2 Goodness-of-fit table

Model summary^b

Model	R	R^2	Adjusted R^2	Estimated standard error
1	.929 ^a	.862	.862	624893.4417

a. Predictor is "Search Popularity"

b. The dependent variable is the "search index"

Table 3 is the regression coefficient table. The table shows the constant terms of the regression model, B (non-standardized regression coefficient), and its standard error and standardized regression coefficient values, t (the value of the statistic) and significance level. As can be seen from the table, the constant term value of the regression model is -353914.196 , The value of the regression coefficient of the independent variable "search popularity" is 12.790 .

The significance level of the regression coefficient is 0.000 , significantly less than 0.05 , the initial hypothesis of the T (student's t test) should be rejected. This also illustrates the significance of the regression coefficients and shows that it is appropriate to establish a linear model.

Table 3 Goodness-of-fit table

Coefficient ^a					
Model		Non-standardized coefficient		Standard coefficient	Sig.
		B	Standard error	Trial version	
1	(Constant)	-353914.196	33304.726		-10.627 .000
	Search popularity	12.790	.229	.929	55.914 .000

a. Dependent variable is search index

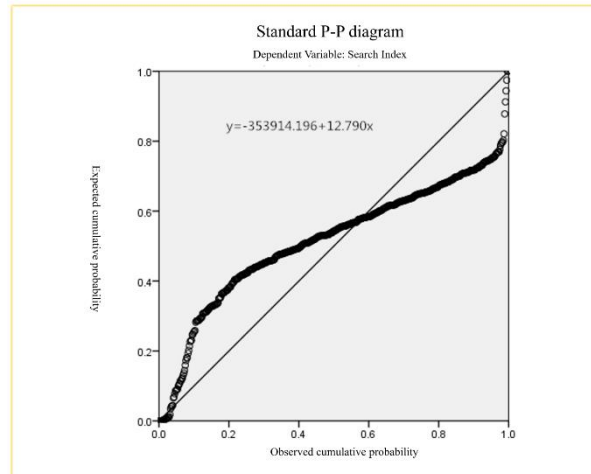


Figure 5 Standard P-P diagram

The strong correlation between search index and search popularity indicates that consumers' search behaviors in the product understanding and analysis stages are similar and repetitive. Convergence means that consumers will have the same tendency to perceive the value of goods. Approved hot search products will most likely be favored by more consumers. Repeatability means that after entering the product analysis stage after completing the product understanding stage, consumers may also return to the product understanding stage, and carry out repeated consideration and comparative analysis. From the regression analysis results of search popularity and search index, we know that the relationship between the two is in line with the regression equation: $y = -353914.197 + 12.790x$. y (dependent variable) is search index, x (independent variable) is search popularity.

From the very weak correlation between search popularity and the competition index, we can see that some of the products of interest in the consumer's vision are waiting to be mined by the enterprise, indicating that some of the hot search products in the market are not very competitive.

There is a weak correlation between click-through-rate and competition index. This means that in the process of product understanding, analysis, and purchase by the platform, consumers show the characteristics of diversified needs. Product options with high click-through rates that can provide consumers with value information are spread across a wide range and have a large Some are not discovered by SMEs.

There is a weak correlation between clickthrough rate and conversion rate. It shows that consumers' click behavior during the understanding phase and analysis phase does not directly flow to the purchase phase after the product is understood and compared to a certain degree of analysis. At the same time, it also shows that some products can generate buying behaviors for consumers after a few clicks.

5. Consumer Cross-Border Retail E-Commerce Company Selection Suggestions

5.1 Follow head products on Hot Search List

Companies need to pay attention to the top products on the hot search list. Such products focus on the hot issues that consumers have paid attention to in this field in the recent period.

Hot search products on the hot search list have their own different attributes. Because the source of product popularity is too extensive, companies cannot accurately grasp it, so it is a good method to select products by paying attention to hot search products under categories. Hot search products on the hot search list have their own different attributes. Because the source of product popularity is too extensive, companies cannot accurately grasp it, so it is a good method to select products by paying attention to hot search products under categories. On the one hand, it can grasp the product's popular trend and trend; on the other hand, due to the linear relationship between search popularity and search index, selecting suitable hot search products to be placed in the store will bring sufficient traffic and exposure to the business store and increase the business brand awareness can also improve the exposure of unpopular products.

5.2 Mining long tail products on the market

Companies do not always be able to accurately sense consumer demand for products. If companies fail to respond to current hot search products in a timely manner, or if they cannot meet the supply of hot search products, they need to consider mining products in the long tail market.

The extremely weak correlation between search popularity and the competition index is a sign that there are long-tail products on the market that have not yet been discovered by companies that can meet the potential needs of consumers. Companies tap the long-tail products in the market, and at the same time tap the potential demand of consumers. Therefore, SMEs choose long-tail products in the product layout, which can accurately meet the long-tail needs of consumers and generate cost-effective product conversions.

5.3 Targeting Blue Ocean Market Products

One of the major problems faced by enterprises when conducting cross-border retail e-commerce business is homogeneous competition. Once enterprises are engaged in product homogeneous competition, it will lead to price wars between enterprises and further reduce profit margins. So companies need to be constantly aware of the products in the blue ocean market.

The so-called blue ocean market refers to a market where there is currently no competition and there is a large potential for demand and profit. The weak correlation between click-through rate and competition index shows that the potential demand of consumers is diverse, but not all demand is valued by the market. There are a large number of markets with potential demand and sufficient profit margins that have not yet appeared in the vision of enterprises. Enterprises need to pay enough attention to the products in the blue ocean market, be the first to enter the blue ocean market to become a leader, and reduce their competitive pressure.

5.4 Grasp high conversion products

The so-called "explosive" single products often appear in the market. The product cycle of such products comes from the fact that their products have specifically touched the widespread consumer demand. In the initial stage, low-exposure, low-click, high-conversion Phenomenon, increase its exposure through follow-up marketing methods to make it a "explosive" traffic product. Just like the weak correlation between click-through rate and conversion rate. Compared with traditional marketing thinking that high clicks bring high conversions, once a product that meets the precise needs of consumers appears, it will immediately produce a product with a high conversion rate. Especially in the cross-border retail e-commerce market, due to differences in geographical, language, and cultural backgrounds, unexpectedly high conversion rates for single products are likely to appear. Therefore, enterprises must seize such cost-effective products to accurately meet consumer needs.

6. Summary

This article is in the context of the digital economy. This article is a data selection study that addresses the pain points of cross-border retail e-commerce product exports that do not match the needs of overseas consumers. Data selection is not only an embodiment of digitally empowered enterprises, but also a way for cross-border retail e-commerce export enterprises to optimize product services and worthy of in-depth research.

Cross-border retail e-commerce companies face complex overseas environments and differentiated product needs. Data selection can help companies accurately grasp consumer demand. Based on the research on consumer search behavior data on AliExpress on the cross-border retail e-commerce platform, based on the analysis of cross-border retail consumer search behavior and the traffic paths it generates, every two data indicators provided by the platform Correlation analysis and linear regression analysis were performed between the indicators. By analyzing consumer search behaviors, the company proposes suggestions for selecting products in cross-border retail e-commerce businesses, which changes the weak position of consumers in traditional selection strategies and helps companies get closer to the needs of target consumers.

References

- [1] Analysis of the impact of B2C cross-border e-commerce platform on China's foreign trade: Alibaba AliExpress as an example [D]. Capital University of Economics and Business, 2016
- [2] AliResearch, Accenturestrategy. Global Cross-Border B2C E-commerce Market Outlook: Digital Consumption Reshapes Commercial Globalization [EB/OL]. <http://www.aliresearch.com/Blog/Article/detail/id/20477.html>, 2015-06-11
- [3] Yan Liu. Research on the Comprehensive Mode of Cross-border E-commerce——Taking AliExpress and Dunhuang as Examples [J]. *Science & Technology Economy Market*, 2017(2):102-103.
- [4] Yunmiao Wang, Lili Fu. Comparative study of the two major cross-border e-commerce platforms of American Amazon and China AliExpress for the Canadian market [J]. *Market Modernization*, 2016(18):39-40.
- [5] Pingping Wu. Analysis on export selection strategies of cross-border e-commerce enterprises [J]. *Times Finance*, 2016(30):235-236.
- [6] Shuyan Zhu. Research on Cross-border E-commerce Export Selection Strategies from a Multi-factor Perspective [J]. *Jiangsu Commercial Forum*, 2018(9):27-30.
- [7] Zhifang Hu. Several strategies and techniques for successful selection of cross-border e-commerce sellers [J]. *Practice in Foreign Economic Relations and Trade*, 2018, No.355(08):69-72.
- [8] Wilson T D . On user studies and information needs[J]. *Journal of Documentation*, 1981, 37(6):658-670.
- [9] Bell D R, Lattin J M. Shopping Behavior and Consumer Preference for Store Price Format: Why “Large Basket” Shoppers Prefer EDLP[J]. *Marketing Science*, 1998, 17(1):66-88.
- [10] Broder A Z , Glassman S C , Manasse M S , et al. Syntactic clustering of the Web[J]. *Computer Networks and ISDN Systems*, 1997, 29(8-13):1157-1166.
- [11] Moorthy S, Ratchford B T, Talukdar D. Consumer Information Search Revisited: Theory and Empirical Analysis | Journal of Consumer Research | Oxford Academic[J]. *Journal of Consumer Research*, 1997, 23(4):263-277.

- [12] Mike Moran, Bill Hunt, Moran, et al. Search Engine Marketing, Inc. ———Driving Search Traffic to Your Company's Web Site Second Edition[M]. *Publishing House of Electronics Industry*, 2007