Multi-category Comparative Analysis of Factors Affecting E-commerce Sales

Mengyao Xu
School of Information Management, Wuhan University, Hubei, China, 2018201040026@whu.edu.cn

Daomeng Guo
School of Information Management, Wuhan University, Hubei, China

Yiran Li
School of Information Management, Wuhan University, Hubei, China

Multi-category Comparative Analysis of Factors Affecting

E-commerce Sales

Mengyao Xu[1*], *Daomeng Guo*[1], *Yiran Li*[1]

[1]School of Information Management, Wuhan University, Hubei, China

Abstract: With the continuous development of e-commerce, more and more types of goods are sold online, so merchants should develop different sales strategies for different types of goods. This paper firstly selects 15 variables to build a stepwise regression model. In the analysis of influencing factors on sales of products in different categories, we find that there are significant differences in the impact of the number of appended reviews and pictures reviews on the sales of utilitarian and hedonic products. In the analysis of influencing factors on sales of products in the same category, we find that the factors influencing the sales of different clothing products are also different to some extent. At last, we put forward some suggestions on adjusting price and title length, and writing product details. This paper is more detailed in variable selection and product classification than some previous studies. It is meaningful for merchants to optimize sales plans and improve product sales.

Keywords: e-commerce sales, influencing factors, product classification, text similarity calculation

1. INTRODUCTION

In recent years, with the continuous development of information technology, e-commerce transactions are rapidly increasing. By the end of 2018, global e-commerce sales have reached 2.8 trillion US dollars, accounting for 11.9% of the global retail sales, and clothing has had the highest online sales among all kinds of products. In the face of opportunities and challenges, e-commerce merchants need to strengthen the understanding of the factors that affect the sales volume, which can better serve consumers and improve the revenue.

Scholars have conducted various studies on the factors affecting sales volume. In addition to price [1], [2], online review [3], [4], service [5], [6] and other common variables, this paper specially adds the related variables of product details and product title. In product details, scholars have focused more on the impact of the pictures [7], [8] and the interaction with text [9]. This paper focuses on description text in the product details, and extracts the number of adjectives, the number of nouns, similarity between details and reviews.

In this paper, we select 15 variables and study the similarities and differences of the factors affecting the sales of products in different categories and in the same category. The results of the study have significance to optimize product display and improve product sales. This paper introduces some references and theoretical hypothesis in the second and third parts, data processing and model construction in the fourth and fifth parts, and significance and conclusion in the sixth and seventh parts.

2. LITERATURE REVIEW

2.1 Product type

After Hirschman and Holbrook (1982) divided the product into two categories of hedonic and utilitarian [10], more and more scholars began to study the different influences of hedonic products and utilitarian products on consumers. Many scholars believed that the judgment of hedonism and utilitarianism directly affected the preference for online retail products and further affected the shopping intention in the future [11]. For utilitarian products, the objective connection between features and utility makes it easier to obtain product information,

---

while for hedonic products, consumers more rely on subjective impression and reduce dependence on product features or information [12]. Of course, hedonism and utilitarianism are not necessarily the two ends of the one-dimensional scale, and different products can be high or low in hedonic and utilitarian attributes [13].

2.2 Factors affecting sales volume

After sorting out previous studies, we find that the factors influencing online purchase behavior mainly include the following:

Table 1. Literature review

Variable	Author	Viewpoint
Price	Jadhav and Khanna, 2016 Chevalier and Goolsbee, 2003	Low price and promotion are important factors that affect online shopping.
Online review	Mo and Li et al., 2015 Babić and Sotgiu et al., 2016	Positive reviews, pictures reviews and appended reviews have a positive impact on consumer purchase behavior. The negative effects are not significant.
Product score	Tamimi and Sebastianelli, 2015 Babić and Sotgiu et al., 2016	Digital and star ratings have a significant impact on product sales.
Reputation	Babić and Sotgiu et al., 2016 Muda and Mohd et al., 2016	Electronic word of mouth and perceived reputation have a strong impact on product sales.
Service score	Chen and Wu et al., 2016 Lim and Heng et al., 2016	Service quality is positively correlated with consumer satisfaction, which can affect consumer loyalty.
Logistics score	Hu and Huang et al., 2016 Kawa, 2017	Logistics services have an impact on consumer satisfaction and are increasingly important in e-commerce.
Product display	Pappas, 2016	Providing detailed and accurate information about products can increase purchase expectations.
Product title	Li and Wang et al., 2015	Extract product features from online reviews for product title optimization.

3.   THEORY AND HYPOTHESIS

3.1   Horizontal analysis and vertical analysis

We refer to the differences of products in different categories as horizontal differences (e.g, mobile phones, necklaces), and the differences of products in same category as vertical differences (e.g, shirts and jeans belong to the clothing category).

In the analysis of horizontal difference, according to Hirschman and Holbrook (1982), we classify the products into two categories: hedonic products and utilitarian products [10]. Hedonic products provide more pleasure and excitement, consumers would pay more attention to themselves. While utilitarian products are mainly instrumental and functional, consumers need to spend much more time to compare properties differences between products. What's more, many scholars have analyzed the factors influencing consumer purchasing behavior from the perspective of product types [17], [18]. Therefore, we put forward the hypothesis 1:

H1: The influencing factors of sales are different in utilitarian products and hedonic products.

In the vertical difference analysis, considering that clothing account for a relatively high proportion of online physical products retail sales, we choose clothing as our vertical analysis objects. Considering that the vertical difference of clothing may be small, we propose hypothesis 2:

H2: The influencing factors of sales are basically the same in clothing category.

3.2 The influencing factors of sales

According to the previous research results and combined with the actual situation, the following 15 variables are selected as independent variables from 5 aspects to analyze their impact on sales.

Table 2. Selection of variables

Variable type	Selected variable
Price related	Price
Product reviews related	The number of cumulative reviews, The number of appended reviews The number of pictures reviews, Similarity of details and reviews
Store evaluation related	Product rating, Description rating, Service rating, Logistics rating
Product display related	The number of product pictures, The number of words in product details The number of adjectives in product details, The number of nouns in product details
Product title related	The length of the product title Similarity of details and title

We hypothesized H3 for the independent variable:

H3a: Variables except price have a positive impact on sales volume.

H3b: Price has a negative impact on sales volume.

4. DATA ACQUISITION AND DATA PROCESSING

4.1 Data acquisition

Ordinary taobao stores lack data such as product rating, the number of appended reviews, the number of pictures reviews, so we choose products in the Tmall stores as our research objects. We obtain the data of the top 800 products in the web ranking of 15 types of products. The data collection time is from June 10, 2019 to June 15, 2019. There are 12,000 products, and the data volume is about 100G.

In the analysis of horizontal difference, we select 10 types of products according to the classification of utilitarianism and hedonism. Utilitarian products are: notebook, professional SLR camera, mobile phone, guitar and suitcase. Hedonistic products are: diamond necklace, clothing, facial cleanser, nuts and beer. In the vertical difference analysis, in order to control the influence of season and style type on clothing sales, we select six types of clothing: shirt, jeans, dress, wind coat, sweater and down jacket.

4.2 Data processing

4.2.1 Calculation of text similarity

We get all the pictures in the product details, count the number of pictures, and use Baidu AI platform to identify the text information in the pictures. Then, we use the 'jieba' package in python to segment the text information in the product description, and count the number of all words, adjectives and nouns. Finally, TF-IDF algorithm is used to calculate the similarity between description text and title and similarity between description text and reviews.

4.2.2 Calculation of other variables

First, price variable processing: because a small number of the same product may use different prices to distinguish its low version and high version (such as: notebook, mobile phone). Faced with such products, we can only choose to use the average price to represent the price of such products. Secondly, picture processing: we calculate the size of the picture according to the height of the picture. Every 1000 pixels of height count as

one picture, and the results remain 1 decimal place. The number of pictures calculated in this way is closer to the number of pictures perceived by the naked eye.

4.2.3 Data cleaning and data processing

In order to reduce the influence of outliers and dimensions on the results of the model, we remove outliers to clean the data, and then standardize the data.

5. MODEL CONSTRUCTION

We put all the known 15 influencing factors into the model. In order to eliminate multicollinearity, the correlation test of 15 independent variables is conducted, and correlation between the number of words and the number of adjectives and nouns is greater than 0.7, so we delete the number of words. A stepwise regression model is used to filter the remaining independent variables, at last, we obtain the final model.

5.1 Horizontal analysis of products of different categories

We want to analyze whether each variable has the same influence on the sales of utilitarian and hedonic products, so we summary the final results into the following table:

Table 3. Horizontal comparison of factors influencing sales volume

Classification	Utilitarian products					Hedonic products				
Category	Notebook (7391)	SLR camera (3007)	Mobile phone (2905)	Guitar (2686)	Suitcase (1464)	Diamond necklace (4150)	Clothing (1232)	Beer (276)	Nuts (163)	Facial cleanser (139)
Product Price	-0.206***	-0.009***	-0.136***	-0.136***	-0.104***	-0.150***	-0.251***	-0.089***	-0.076**	
The number of cumulative reviews	0.298***	0.844***	0.118***	0.609***	0.187***	0.576***	0.667***	0.387***	0.217***	0.116***
Product rating						-0.069***	0.024**			
The number of appended reviews		-0.167**	0.258***		0.294***	-0.348***	-0.574***	-0.571***		
The number of pictures reviews	0.167*		0.270***		0.175*	0.386***	0.484***	0.926***	0.381***	0.726***
Description rating	0.055*		-0.074**		0.105***	0.137***	0.072***			
Service rating		-0.065*	0.140***			-0.111***				
Logistics rating		0.073*			-0.207***		-0.059***		0.072**	-0.101***
The number of product pictures			0.084**	0.006**			-0.038***		0.116***	0.098***
The number of adjectives	0.059*		0.143**			-0.089**		0.172***		-0.085***
The number of nouns		-0.056**	-0.151***			0.096**	-0.067***	-0.170***		
Similarity of details and reviews		-0.049*	-0.088***		0.105***	0.055**	0.096***			-0.058**
Similarity of details and title					-0.108***		-0.033**		-0.085***	
The length of title		0.124***	-0.130***	-0.057*	0.049*	0.045*	-0.059***			
R^2	0.3	0.56	0.48	0.47	0.55	0.53	0.56	0.55	0.35	0.67

P-value: 0.01***   0.05**   0.1* Their average price is shown in parentheses

On the whole, the model passes the F test, and coefficients pass the T test. Except for notebook and nut, the R^2 of other models is greater than 0.45, which has certain explanatory significance. The robustness test of the model showed that among the 14 variables, product price, the number of cumulative reviews, the number of appended reviews, the number of pictures reviews, the number of product pictures, the similarity of details and reviews have the strongest robustness. Among these, product price has a significant negative impact on product

sales, which is consistent with our hypothesis H3b. The number of cumulative reviews and pictures reviews has a significant positive impact on sales. The influence of product pictures number is mainly positive, while the influence of nouns number is mainly negative. However, product rating, service rating and similarity of details and title only have an impact on the sales of a few products, and the impact is not significant on the whole.

The differences between utilitarian products and hedonic products are mainly reflected in the number of appended reviews and pictures reviews. Influence of the number of appended reviews on hedonic products is mainly negative, while that on utilitarian product is mainly positive. The positive impact of the number of pictures reviews on sales of hedonic products is greater than that of utilitarian products. These differences fit our hypothesis H1.

In addition, we find that except for professional SLR and clothing, the negative influence of the price on sales showed a trend of the higher price and greater influence.

5.2 Vertical analysis of products of the same category

5.2.1 Analysis of influence factors

In order to analyze the vertical differences of factors influencing sales volume, we select 6 subcategories of clothing for analysis. We summary the model results and obtain the following table:

Table 4. Vertical comparison of factors influencing sales volume

Category	Jeans (476)	Shirt (777)	Sweater (1031)	Dress (1264)	Wind coat (1378)	Down jacket (2374)
Product Price	-0.067***	-0.167***	-0.218***	-0.202***	-0.141***	-0.121***
The number of cumulative reviews	0.434***	0.752***	0.761***	0.497***	0.664***	0.294**
Product rating			0.070***			0.091***
The number of appended reviews	-0.376***	-0.571***	-0.968***	-0.165***	-0.698***	-0.278**
The number of pictures reviews	0.645***	0.498***	0.689***	0.404***	0.667***	0.219**
Description rating			0.172***	0.055**	0.107***	
Service rating				-0.095***		-0.086**
Logistics rating	-0.149***	-0.045*				
The number of product pictures	-0.054***		0.155***		0.049**	0.062*
The number of adjectives				0.045**		-0.080**
The number of nouns			-0.092***			
Similarity of details and reviews			0.129***	0.050**		0.223***
Similarity of details and title	0.063***		-0.097***			
The length of title	-0.041**	-0.081***	-0.069**		-0.061**	0.118***
R^2	0.69	0.64	0.53	0.73	0.59	0.17

P-value: 0.01*** 0.05** 0.1* Their average price is shown in parentheses

With the exception of the down jacket, the R^2 of models is all greater than 0.5, the model fitting effect is good. Product price, the number of cumulative reviews, the number of appended reviews, the number of pictures reviews and the length of title have the strongest robustness. The number of cumulative reviews, the number of pictures reviews, similarity of details and reviews have a positive impact on sales. Product price, the number of appended reviews and the length of title have a negative impact on sales. Interestingly, only description rating has a positive impact on sales, while service rating and logistics rating have a weak negative impact.

From the perspective of the vertical difference, the influencing factors of different clothing types are still

significantly different, so we reject our hypothesis H2. We find that sweaters, dresses, wind coat and down jackets are more significantly influenced by description rating, the number of product pictures, and similarity of details and reviews than shirts and jeans. This may be due to the fact that sweaters, dresses, wind coat and down jackets require more product information, while shirts and jeans have a more uniform style type.

5.2.2 Analysis of price influence

In the previous horizontal analysis, we believe that the higher the price of the product, the greater the negative impact of the price on the sales volume. To verify this conclusion, we reprocess the data. Considering the particularity of down jackets sold in summer, we exclude down jackets, integrate the data of other clothing, and divide them into three equal parts according to the price level.

From the results of high price, middle price and low price clothing, we find that the negative impact of price on product sales is not that the higher the price, the greater the impact. In order to further study and observe this phenomenon, we further subdivide products into 6 categories, ranging from lowest price to highest price, from clothing 1 to clothing 6. The results show that the negative influence of price on product sales presents the u-shaped curve change. For products with low overall price, the elasticity of sales volume to price is greater while the price is lower. For products with high overall price, the elasticity of sales volume to price is greater while the price is higher. Sales of products in the middle range are barely affected by price.

Table 5. The influence of price on clothing sales at different prices

Category	Low price clothing	Middle price clothing	High price clothing	Clothing 1	Clothing 2	Clothing 3	Clothing 4	Clothing 5	Clothing 6
Product Price	-0.189***	-0.053**	-0.136***	-0.185***	-0.094***			-0.100**	-0.133***

P-value: 0.01***   0.05**   0.1* Here we just show the result of the price variable

6.   ANALYSIS AND DISCUSSION

6.1 Discussion of results

As for the price variable, the result in this paper is basically consistent with the results of previous studies. It is believed that the price elasticity of online sales is very high. Furthermore, this paper shows that the degree of the negative impact of price on sales is affected by product type and product price level. As for the number of cumulative reviews, this paper is consistent with Tamimi and Sebastianelli's idea that the number of reviews can promote product sales [18]. But they thought that product rating had a positive impact on consumers' purchase intentions, whereas this paper argues that the influence of product rating is insignificant. The different opinions may be related to the purchase channel of consumers, product rating is only displayed on the computer, but not on the mobile phone. Some scholars believed that store service and store logistics had a significant positive impact on sales [5], [6], but this paper finds that service rating and logistics rating have no significant impact on some products, and even have a negative impact. We guess the stores with lower sales volume may hope to improve the store sales volume by improving the service quality and logistics speed. Product pictures have a positive effect on the sales of some products and a negative effect on the sales of others. Li et al. believed that this may be caused by excessive information load of some products' pictures [9].

6.2 Theoretical and practical significance

Firstly, this paper analyzes in detail the similarities and differences of the factors influencing sales volume of products of different categories and products of the same category. In addition, this paper classifies different

categories of products into utilitarian and hedonic, which is more detailed than previous studies. The results show that the influence of appended reviews on utilitarian products is positive, while that on hedonic products is negative, and pictures reviews have a greater positive impact on hedonic products than utilitarian ones. Therefore, when selling utilitarian products, consumers should be encouraged to add reviews, and while when selling hedonic products, consumers should be encouraged to display photos.

Secondly, in line with previous viewpoints, this paper also proves that price has a significant negative impact on sales volume. Furthermore, this paper further proves that the influence degree is affected by the price of the product itself. The results show that in the field of clothing, the influence degree changes in a u-shaped curve, that is, products with middle price have the lowest price elasticity. Therefore, we think that when selling products with middle price, the merchants can raise the product price appropriately and lose part of the sales volume in order to gain more profit.

Thirdly, the results show that the similarity of details and reviews has a positive impact on sales volume, while the number of nouns has a negative impact, indicating that merchants should pay more attention to consumer demand when describing product details and reduce the use of nouns. Moreover, the impact of title length on clothing sales is negative. Therefore, the title length should be reduced and controlled when naming clothing products.

7.   CONCLUSION AND PROSPECT

7.1   Conclusion

In the horizontal analysis, the number of cumulative reviews and pictures reviews have a positive impact on sales. The price has a negative impact on product sales. What's more, product rating and services rating only have an influence on sales of a few categories of products, which is not significant on the whole.

As for the differences between utilitarian products and hedonic products, first of all, the impact of the number of appended reviews on hedonic products is mainly negative, while the impact on utilitarian products is mainly positive. Secondly, the positive impact of the number of pictures reviews on sales of hedonic products is greater than that of utilitarian products.

The influence of each variable on clothing sales is similar to that on other categories of products, and the number of cumulative reviews, pictures reviews and similarity of details and reviews have a positive impact on sales. Product price, the number of appended reviews and length of title have negative impact on sales volume. From the perspective of price alone, the negative impact of price on clothing product sales presents a u-shaped curve. For products with low overall price, the elasticity of sales volume to price is greater while the price is lower. For products with high overall price, the elasticity of sales volume to price is greater while the price is higher. But sales of mid-priced products are barely affected by price.

There are also vertical differences in the factors affecting the sales of clothing products. Sweaters, dresses, wind coats, and down jackets are more significantly influenced by description rating, the number of product pictures, and the similarity of details and reviews than shirts and jeans.

7.2   Limitation

Due to the limitation of funds, time and manpower, this paper still has many limitations. First, the sample size is limited. Only 10 categories of products are selected in the horizontal analysis of this paper, which could not completely represent all utilitarian and hedonic products. The sample size should be further expanded to increase universality and authenticity in future studies. Second, the values of some variables are not precise enough. Since the experimental data in this paper are from web crawling, the sales displayed on the web page are often rounded, such as: 5000+, 10000 +. In future studies, we should try to cooperate with businesses to

obtain first-hand real data, or use online public data sets to increase the accuracy of the results.

REFERENCES

[1]   Chevalier J, A Goolsbee. (2003).Measuring Prices and Price Competition Online: Amazon.com and Barnesand Noble.com. Quantitative Marketing and Economics, 1(2): 203-222.

[2]   Jadhav V, M Khanna. (2016).Factors influencing online buying behavior of college students: a qualitative analysis. The Qualitative Report, 21(1): 1.

[3]   Tamimi N, R Sebastianelli. (2015).The relative importance of e-tailer website attributes on the likelihood of online purchase. Internet Research, 25(2): 169-183.

[4]   Mo Z, Y Li, P Fan. (2015).Effect of Online Reviews on Consumer Purchase Behavior. Journal of Service Science and Management, 08(03): 419-424.

[5]   Lim, Y S, et al. (2016).Customers' online website satisfaction in online apparel purchase: A study of Generation Y in Malaysia. Asia Pacific Management Review, 21(2): 74-78.

[6]   Chen, H, et al. (2016).Exploring key factors in online shopping with a hybrid model. SpringerPlus, 5(1): 1-19.

[7]   Yoo J, M Kim. (2014).The effects of online product presentation on consumer responses: A mental imagery perspective. Journal of Business Research, 67(11): 2464-2472.

[8]   Li X, M Wang, Y Chen. (2014).The Impact of Product Photo on Online Consumer Purchase Intention: an Image-Processing Enabled Empirical Study, in PACIS, Pacific Asia Conference on Information Systems (PACIS): AIS Electronic Library. 325.

[9]   Li, M, et al. (2016).The moderating role of information load on online product presentation. Information & Management, 53(4): 467-480.

[10]   Hirschman E C, M B Holbrook. (1982).Hedonic Consumption: Emerging Concepts, Methods and Propositions. Journal of Marketing, 3(46): 92-101.

[11]   Dr Nili, M, D Delavari, N Tavassoli. (2013).Impacts of Utilitarian and Hedonistic Values of Online Shopping on Preferences and Intentions of Consumers. International Journal of Academic Research in Business and Social Sciences, 5(3): 82.

[12]   Noble, S M, D A Griffith, M G. (2005).Weinberger, Consumer derived utilitarian value and channel utilization in a multi-channel retail context. Journal of Business Research, 58(12): 1643-1651.

[13]   Okada, E M. (2005).Justification Effects on Consumer Choice of Hedonic and Utilitarian Goods. Journal of Marketing Research, 42(1): 43-53.

[14]   Gorodnichenko Y, V Sheremirov, O Talavera. (2018).Price Setting in Online Markets: Does IT Click? Journal of the European Economic Association, 16(6): 1764-1811.

[15]   Marchand A, T Hennig-Thurau, C Wiertz. (2017).Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. International Journal of Research in Marketing, 34(2): 336-354.

[16]   Pappas N. (2016).Marketing strategies, perceived risks, and consumer trust in online buying behaviour. Journal of Retailing and Consumer Services, 29: 92-103.

[17]   Dr Gurunathan, K B, M Krishnakumar. (2013).Factors Influencing Apparel Buying Behaviour in India: A Measurement Model. PARIPEX - Indian journal of research, 3(2): 218-224.

[18]   Sanad, R.A. (2016). Consumer Attitude and Purchase Decision towards Textiles and Apparel Products. World Journal of Textile Engineering and Technology, 17(2): 16-30.