



# Exploring The Influence of E-Commerce Multi-Modal Data on Online Shopping Based on Stepwise Regression Model

Yongjie Chu<sup>1\*</sup> and Jingjing Cao<sup>1</sup>

<sup>1</sup>*School of management, Nanjing University of Posts and Telecommunications, Nanjing, China  
chuyongjie@njupt.edu.cn (\*corresponding author), b18110321@njupt.edu.cn*

## Abstract

Online product reviews (OPR) is an important reference information for consumers for online shopping, and has become one of the most persuasive information influencing purchasing decisions of consumers. Researchers have used different methods to explore the impact of various aspects of OPR on consumer shopping behaviour. OPR contains multi-modal data, such as number, text and image, few studies investigated the impact of multi-modal OPR data on the consumers' online shopping behaviour. This study aims to analyse the importance of different modal data of OPR to consumers' purchasing decisions by using real product data sets collected from an E-commerce platform. The stepwise regression model is employed to quantitatively estimate the relationship between the variables of OPR and the product purchases. Through rigorous analysis, the results show that the number and sentiment of initial reviews, as well as the images in OPR have a significant impact on online shopping. The research results provide managerial implications for the e-sellers to make full use of the online product review.

**Keywords:** *Online product review, multi-modal data, online purchase behaviour, stepwise regression*

## 1 INTRODUCTION

Online shopping has been one of the most popular ways for buying goods, as it offers a wide range of products and convenient shopping process. Compared with offline shopping, consumers cannot touch or try a product when shopping online, and thus it is impossible for the consumer to accurately evaluate the performance of the product, especially for those experience products, such as fashions. In order to reduce the uncertainty of online experience, almost all e-commerce platforms, such as Amazon, Taobao, Jingdong, etc., provide a channel for the consumers to write the product reviews after buying or using a product. Basically, the online product reviews (OPR) are a series of information about products, services, corporate brands and other personal experiences released by consumers on the Internet after shopping, which cover multi-modal data, such as number of stars or product rating, text, and pictures, etc. The number of stars or product rating represent the overall performance of a product. The comment text shows the usage experience of buyers, including not only praise for the strength of products, but also criticism for the

weakness of products. Besides, some buyers choose to upload the pictures of a product they bought to display the real product without beautification and decoration. Therefore, the OPR helps the potential consumers increase their perception of products and reduce the uncertainty.

The OPR now are the supplement of product display information provided by sellers, which will affect consumers' shopping decisions. To investigate the impact of product reviews on consumer behaviour, many scholars have carried out extensive research. Harrison-walker (2001) [2] pointed out that OPR own cumulative and storage effects, having an important impact on the purchase behaviour of potential consumers. Zimmermann et al. (2018) [10] studied the impact of differences in product ratings on product pricing and product demand. Besides, according to various consumption fields, the relationship between OPR and demand or purchase is extensively studied. Zhu and Zhang (2010) [11] examined the impact of OPR on the sales of game products; Sparks and Browning (2011) [8] took hotels as the research object to study the impact of online reviewers' interpersonal relationship, positive and negative comment emotions, as well as the consistency

between rating stars and text comments on the product demand. Bigne et al. (2020) [1] studied the online comments of restaurants, and discussed the impact of the OPR number and rating score on the restaurant revenue. Meanwhile, there are studies exploring the impact of OPR from new perspectives. Wei and Lu (2013) [9] compared the impact of celebrity endorsement and online reviews on female consumers' purchase behaviour, Li et al. (2019) [4] used the joint emotional topic model to study the direct and indirect impact of online comments on product sales, they demonstrated that only the positive aspects of OPR have a direct and indirect impact on product sales, while the negative aspects do not show significant impact on sales.

In addition, with the popularity of webcast sales, the consumer purchase behaviour in webcast shopping are further analysed. To examine the influence of images other than only text and ratings on the consumer online shopping behaviour, scholars employed computer vision techniques to extract visual information, and integrated it with empirical study. Khosla et al. (2014) [3] and Li et al. (2020) [5] discussed the effect of pictures released by social network users in predicting consumer preferences and improving user participation. Similarly, Liu et al. (2020) [6] and Rietveld et al. (2020) [7] studied the role of user generated social media images on the brand and product marketing.

Current research mainly focuses on the impact of text reviews and product ratings on consumers' purchase and decision. Though some scholars use visual analysis to study the impact of images on consumers and emphasize the importance of visual content in marketing, the research on the visual information in OPR has just started. At present, it is lack of the impact analysis of visual information and multi-modal data of OPR on fashion consumers' online purchase behaviour. Based on the large amount of OPR data getting from one popular E-commerce platform, in this paper, we employ stepwise regression model and natural language process technique, to investigate the impact of multi-modal data (text, ratings and images, etc.) of OPR on the consumers' shopping behaviour of fashion products.

## 2 RESEARCH FRAMEWORK

### 2.1 Definition of Variables

This paper aims to explore the impact of multi-modal data of OPR on consumer purchasing behaviour. The multi-modal data of OPR contains product ratings, initial reviews, and additional reviews. The product rating, like the product rating star, show the overall satisfaction of a consumer. An initial review is the first or the only production review released by a consumer after a purchase, which states the shopping and use experience of the product. A supplementary review, also called the additional review, is the product review released by a

consumer after the initial review, which is a supplement statement about the product, including good or bad aspects of the product. Production reviews, including the first and second reviews, contain not only text data, but also images uploaded by consumers. In the following, according to the multi-modal data of OPR, we introduce the variables mining from OPR in this study.

#### (1) Number of OPR

The number of online product review is the count of reviews, including initial and additional reviews. The number of initial reviews is an approximate measure of the number of customers buying the product, which also reveals the popularity of the product. The additional reviews are generally released by the consumers after they experienced or used the product for a period of time. Compared to the initial reviews, the additional reviews may provide more authentic and objective experience, therefore, they are more convincing. The more of the additional reviews, the better perception of a product is obtained by the potential customers.

According to the true online shopping experience investigation and the related literature, we select the number of initial reviews and the number of additional reviews as the factors to measure the impact of text-based reviews on consumer behaviour.

#### (2) Sentiment of OPR

When potential consumers plan to buy a fashion product, take a coat as an example, except for the product display information on the platform, they also read the OPR to find out the information about the style, fitness and delivery to evaluate the overall performance of the product. Therefore, the sentiment of OPR reflects the satisfaction of consumers on the products, which may give influence on the potential consumers.

According to the OPR data we gained, we select the initial and additional product reviews for sentiment analysis, and examine impact of the sentiment of OPR on the online shopping behaviours of potential consumer.

#### (3) Number of images in OPR

The presence or absence of images in a product review can reflect the authenticity and usefulness of that review, especially in an online shopping environment where consumers are separated from the product. The product images provided by consumers in product reviews help to enhance the perceived trust of potential consumers, and therefore may have an impact on the final purchase decision of consumers.

Compared to the text in a product review, consumers are more likely to browse the images, since images are more intuitive. As a result, products with more image reviews are more likely to be popular with consumers. In this paper, the number of image in product reviews is

chosen as a measure to evaluate the impact of visual OPR on consumer online shopping behaviour.

#### (4) Rating of product

The E-commerce platform provides a rating for each product. The rating of a product is calculated based on all consumer ratings, which is a simple and direct indicator of the product overall performance assessed by the buyers. The rating of a product range from 1 to 5, the higher rating implies the better reputation of the product, as well as the higher consumer satisfaction. Therefore, a high rating of product will present good impression for potential consumers.

#### (5) Sale price of product

According to consumer theory, the price of a product has an important effect on its sales. The price not only reflects the quality of the product, but also the adjustment of the price has a great influence on the consumer's purchasing behaviour. Therefore, in order to reasonably and fairly examine the influence of the above variables on the online shopping behaviour of consumers, we take the product price as the control variable.

**Table 1** The definition of variables

Variable	Symbol
Number of initial reviews	InitCount
Sentiment value of initial reviews	InitSent
Number of additional reviews	AddCount
Sentiment value of additional reviews	AddSent
Number of pictures in the reviews	ImgCount
Price of product	Price
Rating of product	Rating

## 2.2 Empirical Model

In order to analyse the influence of multi-modal OPR data on consumers' online shopping behaviour, the sales volume of products is taken as the dependent variable to represent consumers' online shopping behaviour. Using stepwise regression method, we construct three regression models to explore the impact of the number of reviews, the sentiment of reviews and the pictures in reviews on consumer behaviour. In the regression models, product price and product score are the control variables.

$$Daysales_{it} = \beta_1 InitCount_{it} + \beta_2 AddCount_{it} + \beta_3 Price_{it} + \beta_4 Rating_{it} + \alpha + \varepsilon \quad (1)$$

Model (1) takes price and rating as the basic control variables, analyses the influence of the number of initial reviews and the number of additional reviews on the online shopping. Based on the model (1), we establish the following regression model by adding two variables:

$$Daysales_{it} = \beta_1 InitCount_{it} + \beta_2 AddCount_{it} + \beta_5 InitSent_{it} + \beta_6 AddSent_{it} + \beta_3 Price_{it} + \beta_4 Rating_{it} + \alpha + \varepsilon \quad (2)$$

Model (2) used price, product rating, number of initial reviews and number of additional reviews as control variables to analyse the impact of the sentiment of initial reviews and additional reviews on the online shopping behaviour.

The above two regression models aim at analysing the effect of text-based OPR on consumers' online purchasing behaviour. With the fast online shopping, consumers are reluctant to spend much time investigating products before buying them. Compared with text-based OPR, visual cues in OPR can provide consumers with more intuitive product information. Therefore, we take the product price and the rating as the control variables, explore influence of the picture quantity in OPR on the consumer purchase behaviour. The regression model used is as follows:

$$Daysales_{it} = \beta_1 ImgCount_{it} + \beta_2 Price_{it} + \beta_3 Rating_{it} + \alpha + \varepsilon \quad (3)$$

Furthermore, by considering the text-based and visual-based OPR simultaneously, we establish the following regression model by combining regression model (2) and model (3).

$$Daysales_{it} = \beta_1 InitCount_{it} + \beta_2 AddCount_{it} + \beta_5 InitSent_{it} + \beta_6 AddSent_{it} + \beta_7 ImgCount_{it} + \beta_3 Price_{it} + \beta_4 Rating_{it} + \alpha + \varepsilon \quad (4)$$

The regression model (4) can be used to analyse the impact of multiple OPR variables on the online shopping behaviour.

## 3 EMPIRICAL ANALYSIS

### 3.1 Data Description and Processing

We used a commercial data collection tool, Octopus catcher, to crawl the data of top 100 clothing products of a brand, including the product name, online review content, the cumulative number of reviews and product ratings. The volume of data is more than 22000.

Through data cleansing, including the removal of duplicate data, invalid data, system default data and other irrelevant data, the number of valid data is 16048, of which the number of additional comment data is 1207.

With the help of HOWNET, the Chinese emotion dictionary, we construct an emotion dictionary of OPR. It includes positive emotion words (positive), positive evaluation words, negative emotion words (negative), negative evaluation words and degree words. The emotion values range from -5 to 5, with positive and negative indicating the emotional orientation of the features in the comment text. The emotion value greater

than 0 indicates positive comments, while less than 0 expresses negative comments. The absolute value of the emotion value indicates the emotional intensity, the greater the absolute value, the greater the emotional intensity, which indicates the satisfaction or dissatisfaction degree of consumers with the product.

### 3.2 Regression Results and Analysis

In this paper, we use SPSS to conduct the statistical analysis. With the linear regression model (1), through stepwise regression analysis, we obtain the significance analysis of the number of initial reviews and the number of additional reviews in OPR on the online shopping behaviour. The results are shown in Table 2.

**Table 2** The significance analysis of the number of reviews on online shopping behaviour

Variable	B	S.E.	Beta	t	Sig.	VIF
InitCount	19.101	3.426	0.82	5.576	0.000	3.963
AddCount	-26.782	9.788	-0.322	-2.736	0.011	2.536
Price	-15.925	5.585	-0.286	-2.851	0.008	1.852
Rating	636.825	310.369	0.181	2.052	0.05	1.431
R <sup>2</sup>	0.641					

Table 2 shows that the regression fitness coefficient is 0.641, the significance level is less than 0.05, the VIF is less than 5, implying that there is no collinearity. The Beta coefficient of InitCount and Rating are all positive, indicating that the number of initial reviews and product rating have a significant positive impact on consumer purchasing behaviour. Since the number of reviews represents the popularity of a product, it means that the popular products are attractive to potential consumers. The coefficient of AddCount and price is negative, implying that the number of additional reviews and price have a negative impact on consumer purchasing behaviour. For fashion goods, especially clothing, the additional reviews may reflect the real quality of the product, as they are posted by consumers after the usage of a product for a period of time. Additionally, most of the consumers post the additional review when they are unsatisfied with the product. This suggests that negative

OPR has a greater impact on consumer behaviour, and the consumer buying behaviour theory states that negative information has a significant impact on consumer behaviour, because it reduces consumers' perceived value and willingness to buy. For online shopping, the information is asymmetry, consumers are more likely to give up a product when they read the negative reviews.

To further explore the influence of OPR sentiment on consumer shopping behaviour, we use the regression model (2) and obtain the significance analysis of OPR variables on online shopping behaviour, such as the number of initial reviews, the number of additional reviews, the sentiment of initial reviews, and the sentiment of additional reviews. The result is listed in Table 3.

**Table 3** The significance analysis of the number and sentiment of OPR on online shopping behaviour

Variable	B	S.E.	Beta	t	Sig.	VIF
InitCount	0.737	0.118	0.634	6.252	0.000	6.496
InitSent	832.196	382.959	0.211	2.173	0.039	5.941
AddCount	-8.11	3.877	-0.18	-2.092	0.047	4.659
AddSent	694.631	330.381	0.197	2.103	0.046	5.537
Price	-2.864	1.213	-0.167	-2.36	0.026	3.15
Rating	-	-	-	-	-	-
R <sup>2</sup>	0.795					

On the basis of the regression model (1), the sentiment variable is added to the model (2). Table 3 shows that the regression fitness of model (2) is improved from 0.641 to 0.795 when the sentiment variable is

considered. The significance level of the initial review sentiment and the additional review sentiment is less than 0.05, which indicates that there is a significant correlation between the review sentiment and the product purchase.

According to the the Beta coefficient, the effect of the number of initial and additional reviews on purchasing behaviour is consistent with the results of Model (1) , both the initial review sentiment and the additional review sentiment have positive impact on the purchasing behaviour. Therefore, the sentiment embedded in OPR is the important information which consumers pay attention to when shopping online. In addition, Table 3 shows that product rating does not have a significant influence on

purchasing when the review sentiment is taken into account, which reveals that compared with the rating, the review sentiment has a greater impact on consumers' shopping behaviour.

Using the regression model (3), we analyse the effect of the number of images in OPR on online shopping behaviour. The analysis result is displayed in Table 4.

**Table 4** The significance analysis of image quantity on online shopping behaviour

Variable	B	S.E.	Beta	t	Sig.	VIF
ImgCount	13.719	3.879	0.463	3.537	0.001	1.839
Price	-17.714	7.722	-0.319	-2.294	0.030	2.068
Rating	1436.052	495.212	0.305	2.9	0.007	1.188
R <sup>2</sup>	0.720					

Table 4 shows that VIF is less than 5, indicating there is no collinear problem. The regression fitness is 0.720 after adding the variable of image quantity, the regression coefficient of ImgCount is 3.537. For fashion products, especially clothing, when observing the pictures shared by the buyers, the potential consumers will reduce the trust risk. The good purchasing experience of other consumers can promote the purchasing intention of

potential consumers. As a result, visual OPR has a greater impact on consumer behaviour than price and average ratings.

By considering the visual OPR and textual OPR at the same time, this paper investigates 5 factors mined from the multi-modal data that may affect consumer behaviour. Table 5 illustrates the analysis result.

**Table 5** The significance analysis of multi-modal OPR on online shopping behaviour

Variable	B	S.E.	Beta	t	Sig.	VIF
InitCount	0.718	0.122	0.617	5.893	0.000	6.915
InitSent	968.42	347.862	0.245	2.784	0.01	4.883
AddCount	-10.586	3.582	-0.234	-2.955	0.007	3.962
AddSent	-	-	-	-	-	-
ImgCount	5.197	2.505	0.192	2.075	0.048	5.396
Price	-3.43	1.164	-0.2	-2.946	0.007	2.889
Rating	-	-	-	-	-	-
R <sup>2</sup>	0.962					

According to the stepwise regression method, the best-fitting model contains four variables, i.e., InitCount, InitSent, AddCount and ImgCount. After fusing text OPR and visual OPR variables, the fitness of the model is improved to 0.962. As can be seen from Table 5, InitCount, InitSent and ImgCount have significant positive correlation with consumer purchasing, while AddCount has a significant negative impact on consumer purchasing. This is consistent with the conclusions drawn from the previous models. Interestingly, the sentiment of the additional reviews and the product ratings does not show significant impact on consumer purchasing behaviour. When consumers make a purchase decision

based on multi-modal data, on the one hand, because the number of additional reviews is relatively small, the number of additional reviews is more important than its sentiment; on the other hand, compared with product rating, OPR including rich content can provide more reference information for consumers when they make a purchase decision.

## 4 CONCLUSIONS

This work explores the impact of different modal data of OPR on consumer purchasing behaviour. Using the data obtained from E-commerce platform, this paper first

analyses the variables that affect shopping behaviour, then constructs the econometric models and makes an empirical analysis. The results show that consumers pay more attention to the sentiment polarity of text data and visual data when they make a purchase decision. According to the conclusions, consumers should be encouraged to post images or short videos in their product reviews, which helps reduce the perceived risk of potential consumers. The additional reviews have a negative impact on consumer shopping behaviour, and thus the sellers in the E-commerce platform are suggested improve their services and products according to the additional reviews.

Multi-modal OPR contains abundant information. In this paper, we only consider the overall sentiment of reviews, do not analyse the product features or corresponding sentiment embedded in reviews. In addition, this paper considers impact of only the number of images on shopping behaviour, but does not consider the quality of the image. In the future research, we will make the further discussion in these aspects.

## ACKNOWLEDGEMENTS

This research was partially supported by the Humanities and Social Science Research on Youth Fund Project of Ministry of Education of China (Grant No. 19YJC870003), the Natural Science Foundation of Jiangsu Province (Grant No. BK20210576), and sponsored by NUPTSF (NY221107).

## REFERENCES

- [1] Bigne E, Chatzipanagiotou K, Ruiz C, 2020. Pictorial content, sequence of conflicting online reviews and consumer decision-making: The stimulus-organism-response model revisited. *Journal of Business Research*, 115:403-416.
- [2] Harrison-Walker LJ, 2001. The measurement of word-of-mouth communication and an investigation of service quality and customer commitment as potential antecedents. *Journal of Service Research*, 4(1):60-75.
- [3] Khosla A, Das Sarma A, Hamid R, What makes an image popular? In: *Proceedings of the 23rd international conference on World wide web*, 2014. pp 867-876.
- [4] Li X, Wu C, Mai F, 2019. The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information & Management*, 56(2):172-184.
- [5] Li Y, Xie Y, 2020. Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of Marketing Research*, 57(1):1-19.
- [6] Liu L, Dzyabura D, Mizik N, 2020. Visual listening in: Extracting brand image portrayed on social media. *Marketing Science*, 39(4):669-686.
- [7] Rietveld R, van Dolen W, Mazloom M, Worrying M, 2020. What you feel, is what you like influence of message appeals on customer engagement on Instagram. *Journal of Interactive Marketing*, 49:20-53.
- [8] Sparks BA, Browning V, 2011. The impact of online reviews on hotel booking intentions and perception of trust. *Tourism management*, 32(6):1310-1323.
- [9] Wei PS, Lu HP, 2013. An examination of the celebrity endorsements and online customer reviews influence female consumers' shopping behavior. *Computers in Human Behavior*, 29(1):193-201.
- [10] Zimmermann S, Herrmann P, Kundisch D, Nault BR, 2018. Decomposing the variance of consumer ratings and the impact on price and demand. *Information Systems Research*, 29(4):984-1002.
- [11] Zhu F, Zhang X, 2010. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2):133-148.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

