AMCIS 2022 Proceedings

Conference Theme Track - Innovative Research Informing Practice

Aug 10th, 12:00 AM

User-Generated Content and Online Product Search - The Case of the Indian Automobile Industry

Madhuri Prabhala
Indian Institute of Management, Calcutta, madhurip16@iimcal.ac.in

Indranil Bose
NEOMA Business School, indranil.bose@neoma-bs.fr

User-Generated Content and Online Product Search – The Case of the Indian Automobile Industry

Emergent Research Forum (ERF)

Madhuri Prabhala
Indian Institute of Management, Calcutta
madhurip16@iimcal.ac.in

Indranil Bose
NEOMA Business School, Reims, France
indranil.bose@neoma-bs.fr

Abstract

The individual's online search is done privately and is considered an accurate measure of purchase intention. Online product information influences purchases and can be expected to influence purchase intention. Using Google's volume of online search trends and 23,000 online automobile reviews scraped from a popular Indian online forum, the relationship between product reviews and online product information search is examined. The reviews are divided into three groups based on the numerical rating associated with each review. Latent Dirichlet Allocation is used in each group to extract discussion topics from the review text. The study shows that online review valence significantly impacts online search volume. Further, while the topics of discussion across the 3 rating groups overlap, their impact on online search volume is different. The study demonstrates the need to look at rating information more closely and the influence of online UGC on online information search.

Keywords

Online Reviews, Online Search, Topic Modelling, Latent Dirichlet Allocation.

Introduction

User-generated content (UGC) influences purchases (Chevalier & Mayzlin, 2006; Duan et al., 2008) of both search and experience products (Ghose & Ipeirotis, 2011). Apart from UGC alone, the interplay between online product reviews and the volume of online product search also influences product sales (Geva et al., 2017). As online product search is generally done in the privacy of the individual's home or personal devices, it is considered a "true and honest" measure of purchase intention (Wu & Brynjolfsson, 2009). Product purchase intention is an antecedent of product purchase. The firm would gain by understanding this honest purchase intention. Therefore, the first research question asks how online product reviews influence the volume of online product search. When processing the online review information, individuals consider the numerical rating associated with each review. If the product is a high-involvement product, more time is spent reading through the text to understand the topics discussed. These topics involve discussion of the product features and experiences of the previous users. If the information is negative, it will have a more severe impact on the final decision-making when compared to more positive or neutral information (Chevalier & Mayzlin, 2006). While the topics of discussion in UGC significantly predict product purchases (Li et al., 2019), the influence of review topics on the volume of online product search is not established. Given the differences in the influence of negative and positive information, the second research question asks how the topics of discussion in negative rated reviews differ in their influence on purchase intention compared to discussion topics in neutral and positive rated reviews. The study uses the volume of online search for automobiles to measure the market purchase intention and seeks answers to these questions in the Indian automobile market.

Literature and Hypothesis

Review Valence

User-generated review rating is termed review valence. Academic research on review valence reflects inconsistent results. In specific contexts, the significant influence of review valence on sales is supported (Duan et al., 2008), while in others, it is not (Chen et al., 2011). However, the influence of review valence on purchase intention is not well established. Given that the volume of online searches measures purchase intention (Geva et al., 2017), the below hypothesis is proposed.

H1: The average online rating of a product's reviews will positively influence the online search volume for the product.

Review Content – Review Topics

The topics of discussion in UGC influence review helpfulness (Vallurupalli & Bose, 2020), product sales (Li et al., 2019), and predict customer purchase intention (Bulut, 2014). A review with a numerical rating above 3 is considered a positive review, while below 3 is negative. Literature has identified that positive and negative UGC are correlated differently with the final decision to purchase a product, and negative reviews have a higher impact, even in the presence of positive reviews (Chevalier & Mayzlin, 2006). Given this background, the topics present in negative reviews and positive reviews will differ in their influence on product outcomes. The below hypothesis is therefore proposed.

H2: The topics in different rating levels will have different influences on the volume of online search for the product.

Control Variables

Previous period sales, review volume, and review length are introduced as controls to account for the variables examined in prior work (Duan et al., 2008; Geva et al., 2017; Ghose & Ipeirotis, 2011). Previous period search volumes are introduced to account for past search experience (Zhang et al., 2014).

Methodology

Data and Variables

The Indian automobile industry, divided into ten segments like Compact, Exotic, Coupe, and so on, depending on the car features such as its engine, length, and price, forms the context for the study. This study is based on the Compact segment, which makes up around sixty percent of the industry. The monthly car sales of 21 cars with non-zero sales between January 2016 to December 2018 are extracted from the CRISIL database. Around 23 000 reviews have been scrapped from Mouthshut.com, India's popular product discussion forum. Cars with less than 100 reviews in the analysis period are removed from the analysis. Review valence is calculated by taking the average of the daily review rating for each car model at a monthly level. The average number of syllables of all reviews of a specific car published in the month is calculated for review length. The monthly volume of online search for each car model is the Google search index that is publicly accessible and is the measure of purchase intention.

Determining the Review Topics

All the reviews are divided into three groups based on the valence associated with each review. Reviews that have a numerical rating of 1,2 are group 1. Reviews with a numerical rating of 3 are in group 2, and those with a numerical rating of 4,5 are in group 3. The review topics are extracted using Latent Dirichlet Allocation (LDA) for each group separately. The optimum number of topics in each group is determined using the perplexity scores (Blei et al., 2003). The LDA is applied using Python's scikit-learn library (Beysolow II, 2018). Each review is a distribution over the identified number of topics, and each topic is a distribution over the individual words. Each review is assigned a probability score for each of the Topics. The sum of these scores adds up to 1. The review is classified to the topic assigned the highest probability score for that review. Finally, each group's number of reviews for each topic for a month is calculated. The results are discussed in the empirical analysis section.

Model and Equations

As the dataset is a panel data with car and month as the unit of analysis, the volume of search is defined as a function of the car model fixed effects, the measures of the hypothesized variables, and the control variables. The equation to be estimated is presented below.

$$\ln(search_{i,t}) = \gamma_o + \gamma_1 \ln(val_{i,t-1}) + \gamma_2{}^X \ln(theme^x{}_{i,t-1})$$
$$+ \lambda_1 \ln(sales_{i,t-1}) + \lambda_2 \ln(vol_{i,t-1}) + \lambda_3 \ln(len_{i,t-1}) + \lambda_4 \ln(search_{i,t-1}) + \varepsilon^s{}_{i,t} \quad (1)$$

The equation estimates the influence of review valence ($\gamma1$) and the rating group-wise review topics ($\gamma2$) on the volume of online search of car model i in the month t. The influence is estimated when controlling for the influence of the previous month's sales ($\lambda1$), review volume ($\lambda2$), review length ($\lambda3$), and the previous month's online search volume ($\lambda4$). All variables are log-transformed (Chevalier & Mayzlin, 2006).

Empirical Analysis

Extraction of Themes – Latent Dirichlet Allocation

The numerical rating associated with each review is used for dividing all the reviews into three groups. The groups are not of equal size. Group 1, with reviews having a rating of 1 or 2, has 2203 reviews. Group 2, with reviews having a rating of 3, has 3,451 reviews, and Group 3, with reviews of rating 4,5, has 10,621 reviews. This distribution shows that the review dataset is skewed towards a higher rating. Different topic models with 2, 3, 4, and 5 topics (Vallurupalli & Bose, 2020) are developed for each group, and the associated perplexity score is noted. The number of topics in the model with the lowest perplexity score is the optimum number of topics for that group.

Figure 1 gives the plots of the perplexity scores against each model for each of the three groups. The figure indicates that the lowest perplexity scores for Group 1 and Group 2 are for the 2 Topics model, and for Group 3, it is for the 3 topics model. Accordingly, the optimum number of topics is chosen as 2 for Group 1 and 2 and 3 for Group 3.

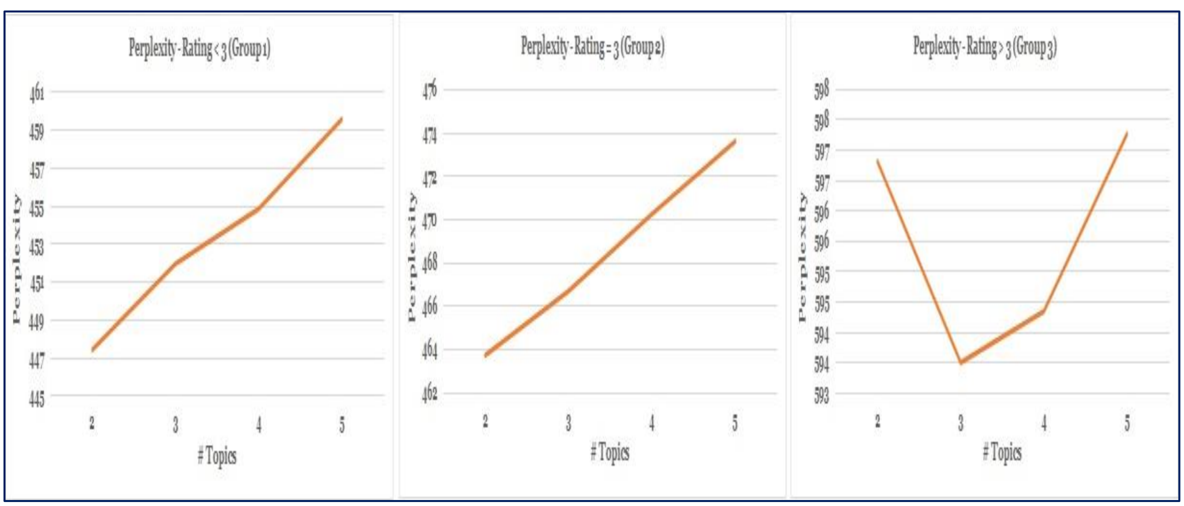

Figure 1. Optimal Number of Topics

The identified Topics, the associated keywords, and the topic name given based on the keywords are presented in table 1. Similar topics come up across the 3 different groups.

Rating Group	Topic Number	Number of Reviews	Percentage of Reviews	Topic Keywords	Topic Name

Group 1 (Rating < 3)	1	1616	10%	look, buy, space, mileage, bad, engine, price, seat	Price and Comfort
	2	587	4%	service, problem, buy, time, drive, issue, purchase, month	Service and Maintenance
Group 2 (Rating = 3)	1	2538	16%	good, drive, space, comfortable, nice, family, look, price	Family and Comfort
	2	913	6%	engine, diesel, feature, petrol, rear, design, power, new	Fuel and Performance
Group 3 (Rating > 3)	1	7545	46%	good, drive, space, comfortable, nice, family, look, nice	Family and Comfort
	2	1514	9%	drive, feel, quality, look, time, new, vehicle, service	Service and Maintenance
	3	1562	10%	engine, diesel, petrol, feature, rear, power, come, variant	Fuel and Performance

Table 1. Topics Identified from LDA

Model Results

The equations are estimated using ordinary least squares with product-level fixed effects (Chevalier & Mayzlin, 2006). The results are presented in Table 2. Online review valence significantly influences online search volume. While the topics across the three groups are similar, the influence of these topics on online search volume is different.

DV = Search	Estimate	Std Error
Review Valence	0.11**	0.05
Rating <3 Price and Comfort	0.19**	0.09
Rating <3 Service and Maintenance	0.25**	0.09
Rating =3 Family and Comfort	0.002	0.07
Rating =3 Fuel and Performance	-0.13**	0.05
Rating >3 Family and Comfort	-0.15^	0.08
Rating >3 Service and Maintenance	-0.12^	0.08
Rating >3 Engine and Features	-0.03	0.08
Previous Period Sales	-0.03***	0.01
Review Volume	0.001	0.01
Review Length	0.04*	0.02
Previous Period Search	0.76***	0.02
Car Model Fixed Effects	Yes	
Adj R - Square	*0.9415*	

'***' p < 0.001, '**' p < 0.01, '*' p < 0.05, '^' p < 0.10

Table 2. Model Results – Search

Discussion of Findings

The findings based on table2 are summarized in table 3.

Hypothesis	Finding	Implication
H1: The average online rating of a product's reviews will positively influence the online search volume for the product.	The hypothesis finds support at 1 % significance.	The higher the online monthly average numerical rating for a car model, the more online search volume for the product. Online review valence positively impacts the volume of online product search.
H2: The topics in different rating levels will have different influences on the volume of online search for the product.	Both the topics in Group 1 (Rating < 3) impact the online search volume at 1 % significance.	The topics discussed in various rating groups are similar. However, their influence on

	One of the 2 topics in Group 2 (Rating = 3) impacts the online search volume at 1% significance. 2 of the 3 topics in Group 3 (Rating > 3) weakly impact the volume of online search (10% significance). The hypothesis is thus supported.	online search differs across the rating groups.

Table 3. Summary of Results

Conclusion

The study begins with two research questions that seek to understand how the online review valence and discussion topics in the review text influence the online product search for product information. The analysis first identifies that online review valence significantly influences online search volume. Secondly, the study sees that the discussion topics in the positive, neutral, and negative review groups overlap. Finally, despite the overlap of topics, the influence of these topics on online search volume differs and is dependent on the review rating groups. As information uncertainty decreases, product information search shifts from active to passive search over time, leading to differences in UGC impact (Huang et al., 2017). The current study further contributes to the academic literature on search and UGC by exploring how product reviews can impact online search volume. Doing so also directs firms to understand the factors influencing the online search for their products.

REFERENCES

Beysolow II, T. (2018). *Applied Natural Language Processing with Python*. Apress Media LLC.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(4–5), 993–1022.

Bulut, A. (2014). TopicMachine: Conversion prediction in search advertising using latent topic models. *IEEE Transactions on Knowledge and Data Engineering*, *26*(11), 2846–2858.

Chen, Y., Wang, Q., & Xie, J. (2011). Online social interactions: A Natural Experiment on Word of Mouth Versus Observational Learning. *Journal of Marketing Research*, *48*(2), 238–254.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345–354.

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems*, *45*, 1007–1016.

Geva, T., Oestreicher-Singer, G., Efron, N., & Shimshoni, Y. (2017). Using Forum and Search Data for Sales Prediction of High-Involvement Projects. *MIS Quarterly*, *41*(1), 65–82.

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498–1512.

Huang, J., Boh, W. F., & Goh, K. H. (2017). A Temporal Study of the Effects of Online Opinions: Information Sources Matter. *Journal of Management Information Systems*, *34*(4), 1169–1202.

Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information and Management*, *56*(2), 172–184.

Vallurupalli, V., & Bose, I. (2020). Exploring thematic composition of online reviews: A topic modeling approach. *Electronic Markets*, *30*(4), 791–804.

Wu, L., & Brynjolfsson, E. (2009). The future of prediction: How google searches foreshadow housing prices and quantities. *Thirtieth International Conference on Information Systems*, 1–14.

Zhang, K. Z. K., Zhao, S. J., Cheung, C. M. K., & Lee, M. K. O. (2014). Examining the influence of online reviews on consumers ' decision-making: A heuristic – systematic model. *Decision Support Systems*, *67*, 78–89.