Association for Information Systems

AIS Electronic Library (AISeL)

SIGHCI 2021 Proceedings

Special Interest Group on Human-Computer Interaction

12-12-2021

Would You Recommend This Product? A Topic Modeling Approach to Understanding Review Behaviors

Long The Nguyen

Zach Sheffler

Follow this and additional works at: https://aisel.aisnet.org/sighci2021

This material is brought to you by the Special Interest Group on Human-Computer Interaction at AIS Electronic Library (AISeL). It has been accepted for inclusion in SIGHCI 2021 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Would You Recommend This Product? A Topic Modeling Approach to Understanding Review Behaviors

Long The Nguyen
University of Massachusetts Amherst
longtnguyen@som.umass.edu

Zachary J. Sheffler
University of Massachusetts Amherst
zsheffler@isenberg.umass.edu

ABSTRACT

Online shopping platforms often highlight reviews to aid consumers' decision-making process. The current research proposes that highlighted review should match between the reviewers' and the browsing consumers' purchasing goals (profiles). Using Latent Dirichlet Allocation (LDA), an unsupervised machine learning method for topic modeling, we uncovered the hidden profiles that show a reviewer's original purchasing goal, whether utility-oriented or hedonic-oriented. Subsequent analysis revealed that utility- and hedonic-oriented reviewers differ in certain review-writing and rating behaviors. The paper contributes to the literature by suggesting a new way to understand reviewers' profiles from text data and resulting review behaviors. We also make a practical recommendation for shopping platforms in highlighting more relevant reviews.

Keywords

e-commerce, online reviews, text analysis, Latent Dirichlet Allocation (LDA), topic extraction.

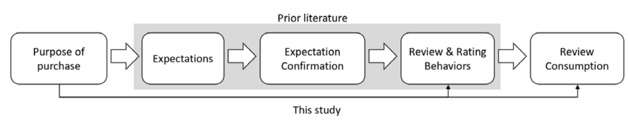
INTRODUCTION

Online reviews help consumers in the decision-making process (Mudambi and Schuff 2010) and aid online shopping platforms in driving up traffic and sales (Chevalier and Mayzlin 2006). As a result, platforms are often concerned with presenting reviews to consumers in accessible manners. Amazon, for instance, highlights certain reviews as "featured" on top of thousands of reviews. These highlights are often selected with criteria like reviewer status, review valence, and the helpful votes a review received (Wu 2017). Nevertheless, to the extent of our knowledge, platforms' criteria to highlight reviews do not include original purposes of purchasing.

Consumers express expectation confirmation – how well the product experiences align with their pre-purchase expectations – in their online review and rating decisions (Ho et al. 2017). Therefore, consumers with different purchasing purposes likely differ in their expectation confirmation and subsequent review writing and rating behavior. Consequently, presenting a current consumer with reviews communicating irrelevant sets of expectations and experience may prove counterproductive to platforms.

In this research-in-progress paper, we aim at understanding (1) how review-writing and rating behaviors differ among reviewers with different purchasing goals ("profiles"), (2) how we may understand reviewer profiles from their reviews, and (3) how a (mis)match between consumer and review profiles influence consumers' evaluation of review. We adopt an unsupervised machine learning method, Latent Dirichlet Allocation (LDA), to uncover reviewer profiles, and conducted a number of analyses with some review behaviors of interest. In doing so, we make a methodological contribution by demonstrating an unsupervised machine learning approach to understand reviewers' purchasing goals, and theoretical contribution by examining how reviewer profiles contribute to their review behaviors. In the following section, we offer a brief review of the literature, introduce the methodological approach, before discussing our initial results and future plans.

LITERATURE REVIEW AND HYPOTHESES



Our study is built upon the theoretical underpinning of shopping behaviors and review-writing behaviors. We consulted two relevant literature streams to develop our theoretical model and hypotheses.

The first stream of the literature documents two fundamental shopping orientations: utility and hedonic orientation (Baker and Wakefield 2012). While hedonic orientation refers to the satisfaction of joy, fun, and other more subjective and personal values, utilitarian purposes are cognitive, functional choices that cater to necessities (Babin et al. 1994, Moore 2015). Though some consumers may express both orientations, they are often found to be either hedonic- or utility-oriented (Baker and Wakefield 2012). Such orientations are found to influence shopping behaviors important to businesses, such as intention to visit (Baker and Wakefield 2012) or to repeat a purchase (Chiu et al. 2014). In an online shopping context, consumers pay attention to utility, functional features as well as hedonic aspects that appeal to emotion like aesthetic performances (Liu et al. 2020).

The second stream of research documents product rating and reviewing behaviors as a means to express experience with the products. Among the many antecedents, product (dis)confirmation is among the important drivers (Ho et al. 2017; Hu et al. 2017). Specifically, consumers build up an expectation about a product and compare it against actual experience (Ho et al. 2017). Expectation (dis)confirmation, a (mis)match between expectation and the actual experience, influences the review and rating behaviors (Ho et al. 2017). In an online shopping experience, failing to meet consumers' utility expectations and hedonic expectations significantly lower their satisfaction (Chiu et al. 2014).

As consumers write about their experience with a product, its characteristics and whether they meet expectations are elaborated in online reviews' textual content, which is read by future consumers. For instance, Benbunan-Fich (2020) documented rich descriptions of a wearable device's feature failures in its online reviews. Nevertheless, these reviews are only perceived as helpful by a consumer if they provide information relevant to his or her specific decision making (Mudambi and Schuff 2010), which involves purchasing orientation, as in what the consumer looks for in the product. For instance, a consumer who is looking for utility features that serve specific needs may find reviews describing a hedonic experience, such as cosmetic quality, good look, and joy, irrelevant. On the other hand, a consumer with hedonic purposes, such as decoration, may deem the aforementioned reviews helpful.

Synthesizing the two literature streams, we posit that a misalignment between reviewers and current consumers' purchasing purpose risks reducing online reviews' positive impact for several reasons. First, as reviews are expressions of experience, a reviewer with different purchasing profiles may engage in different review writing and rating behaviors. Given a product with a blend of hedonic cosmetic values and utilitarian functions, we expect utility-oriented reviewers to use more complex language in order to describe the various functions of the product, comparing to hedonic-oriented reviewers.

H1: Reviewers with more utility-oriented profiles write reviews with more language complexity.

In terms of rating behavior, given the same product with both hedonic and utility values, a reviewer with a higher utility orientation may leave more favorable ratings for several reasons. First, among the various functions of the product, it is more likely that some features meet the utility-oriented reviewer's expectations and lead to a more favorable rating. Second, utility shopping orientation is cognitive, functional, and involves collecting information (Babin et al. 1994). Shoppers purchasing to meet specific necessities likely search, compare features across products, and purchase one that objectively fits them best.

H2: Reviewers with more utility-oriented profiles are more likely to give a positive rating.

Given the reviewers with different profiles will engage in different review writing and rating behaviors for the same product. Current consumers with their own orientation will perceive matching reviews as more helpful.

H3a: More utility-oriented reviews will be rated as more helpful by utility-oriented consumers than hedonic-oriented consumers.

H3b: More hedonic-oriented reviews will be rated as more helpful by hedonic-oriented consumers than utility-oriented consumers.

To the extent of our knowledge, online shopping platforms (e.g., Amazon.com) use algorithms to highlight "featured" reviews based on the reviewers' status ("top reviewer" or "verified") or the number of helpfulness votes received (Wu 2017). Alternatively, they allow consumers to sort for the newest reviews or filter reviews by individual keywords that frequently appear. As a result, we also propose a computer-assisted, automated approach to uncover reviewer profiles contained in the review text to enable current consumers to filter for reviews with the relevant profiles, not just the keywords.

Computer-assisted text analysis techniques are able to uncover useful insights from a large quantity of data in a relatively objective manner (Adamopoulos et al. 2018). For instance, the dictionary methods have been applied to extract various emotions (Yin et al. 2014), personality traits, and review sentiment (Adamopoulos et al. 2018) from the textual content. Automated approaches, such as topic extraction or topic modeling, are recently applied in IS studies involving unstructured data (Abbasi et al. 2018; Shi et al. 2016). On top of the advantages listed above, topic modeling using unsupervised machine learning does not impose strict, predefined rules, therefore can uncover underlying topics based on the natural patterns of words (Humphreys and Wang 2018; Shi et al. 2016). The following section describes our data collection, analytical approach, and initial results.

METHODOLOGY

Data Collection

In March 2020, we collected a small archival sample of online reviews for smartwatches from Best Buy. The selections of the shopping platform and product category were deliberate choices. First, we selected smartwatches for their balance of both hedonic and utilitarian values. Besides utility functions that serve specific needs such as notifications, sport, and activity tracking, smartwatches are also fashionable hedonic items. Second, as we are interested in the positive ratings, BestBuy is appropriate because they apply a binary rating scale asking if a reviewer would recommend a product or not. We deem this recommendation mechanism more suitable to our purpose, comparing to the common five-star rating scale, as the latter suffers from serious rating biases and inflation that makes the distinction between positive and negative ratings obscure in the mid-range of the scale (i.e., 2 or 3 stars) (Breinlinger et al. 2019). After filtering out the observations that are the sellers' replies to original reviewers, we are left with 2296 usable reviews for 74 products.

Measurement

We operationalize the review writing and rating behaviors with observed variables in the dataset. First, review ratings are represented by each review's recommendation choice, in which a "yes" stands for a positive rating, and a "no" encodes a negative rating. Second, review complexity is measured by calculating the Flesch-Kincaid readability grade score (Kincaid et al. 1975), in which a larger score indicates that a text is harder to read. The score was computed using the R package *quanteda*. Reviewer profiles were extracted using an unsupervised machine learning approach for topic extraction, which is described in greater detail below.

Topic Extraction

The paper adopts an unsupervised machine learning method, Latent Dirichlet Allocation (LDA) to uncover underlying topics in the textual reviews. LDA is a parsimonious approach to the analysis of latent topics in textual data (Blei et al. 2003). LDA holds that the probability of a word's appearance in a document (i.e. a product review) is dependent on the presence of the topic it represents in that document. As a result, LDA extracts a topic based on the unique probability vectors of words representing the topic (Büschken and Allenby 2016). For an in-depth introduction to the technicality of LDA, we would refer readers to Tirunillai and Tellis (2014). The analysis was conducted in Knime software version 4.2.

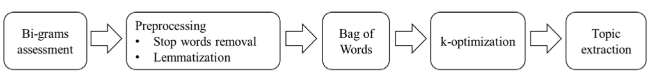


Figure 2. Topic Extraction Procedure

Several document preparation steps were taken before topic extraction, including bi-gram assessment, preprocessing, and creating bags-of-words (BoW). Bi-gram is a specification of N-gram that creates pairs of every two words in a document. Frequently occurring word pair that could be meaningful for analysis was combined into a single compound word (i.e. "heart rate" to "heart-rate") to avoid losing their combined meaning in later steps. Next, the reviews went through part-of-speech tagging, in which each word was given tags for its role in the sentence either as a verb, noun, adjective, and so on. For the purpose of this project, because topics are most likely represented with nouns and noun phrases, only words with the "noun" family tags went on to preprocessing. In preprocessing, stop words (i.e. "a", "the", "of") were removed before the remaining words were lemmatized to their original forms based on the Stanford Core Natural Language Processing (NLP) library. Next, BoWs were created to individualize words from each review, which allowed for subsequence analyses using terms' occurrence frequencies and their connections to topics. Terms appearing less than twice in the whole dataset were deemed infrequent terms and not included in the optimization to identify the number of topics (k-optimization).

As LDA is a probability-based topic extraction method, k-optimization was conducted using the elbow method. This method determines the number of topics at which the joint probability of topics and words (measured in log-likelihood) stop improving noticeably. Specifically, a series of possible values for k from 1 to 40 are tested, and the parameters α (represents the document to topic distribution) and β (representing the topic to word distribution) were respectively set at 0.1 and 0.01, following the general recommendation in the text analysis literature (Steyvers and Griffiths 2006, Kaplan and Vakili 2015, Huang et al. 2018).

RESULT

Reviewer Profiles

The optimization process resulted in 4 interpretable, little-overlapped latent topics. Based on the term frequency, the topic extraction process assigned to each review the probabilities that it belongs to the four topics. Each review is then assigned the topic with the highest probability. The most frequently appeared 15 terms for each topic, which are presented in the word clouds in Figure 3, help us interpret the reviewer profiles.

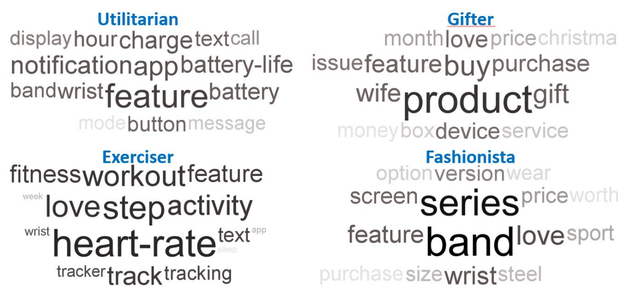


Figure 3. Topic Term Word Clouds

These word clouds represent 4 distinctive reviewer profiles, namely *Utilitarian*, *Gifter*, *Exerciser,* and *Fashionista.* The *Utilitarian* profile is characterized by terms representing basic functions of the products (i.e., feature, app, notification) that help consumers in their day-to-day activities like messaging, texting, calling, which are also mentioned in the most frequent terms for this topic. Users in the *Gifter* profile typically bought the product for their loved ones (i.e., wife) as gifts for special occasions (i.e., Christmas), and thus they pay attention to value (i.e., money), and post-purchase services. The *Exercisers* emphasize workout-related features of the product such as heart-rate monitors and trackers, and they discuss how those functions help their fitness activities too. The *Fashionista* profile is represented by terms referring specifically to value (i.e., price, worth) cosmetic elements like material (i.e., steel), look (i.e., size, band), and others (i.e., version, option).

These four profiles match well with our expectation of utility-oriented versus hedonic-oriented consumer types. The *Exerciser* group appears to be the most utility-oriented, seconded by *Utilitarian*. *Fashionista* is the most hedonic-oriented group. Nevertheless, *Gifter* reviewers appear to go either way, as they may have bought the product for a hedonic-focused or utility-focused and have written reviews accordingly. These categorization results enter the initial hypotheses testing as dummy variables with *Exerciser* chosen as the reference group.

Hypotheses Testing

For an initial analysis, we specified two models controlling for the product effects, using the R package *lme4*, and follow up with pairwise contrasts to compare the reviewer profiles in terms of the two behaviors using the *multcomp* package. To test H1, review complexity was regressed against reviewer profiles, after controlling for product effects. To test H2, a mixed-effect logistic regression model was specified with the recommended rating as the dependent variable (DV), and product random intercepts and reviewer profiles as independent variables (IV). The regression and contrast results in Table 1 and Table 2 respectively show that there are noticeable differences between some reviewer profiles in the hypothesized directions.

DV	Fixed Effects	Est.	S.E.	p-value
Complexity	Intercept	6.010	.109	**<.001**
	Gifter	-.286	.148	.054
	Fashionista	-.132	.144	.359
	Utilitarian	.355	.187	.059
Rating	Intercept	2.938	.204	**<.001**
	Gifter	-.891	.219	**<.001**
	Fashionista	-.432	.023	.065
	Utilitarian	-.824	.255	**.001**

Table 1. Regression Results

Contrasts*	Complexity		Ratings	
	Est.	p-value	Est.	p-value
Gifter – Exerciser	-.286	.212	-.891	**<.001**
Fashionista – Exerciser	-.132	.793	-.431	.252
Utilitarian – Exerciser	.355	.228	-.824	**.007**
Fashionista – Gifter	.154	.761	.459	.172
Utilitarian – Gifter	.640	**.007**	.067	.993
Utilitarian – Fashionista	.487	**.061**	-.392	.427
*Tukey-adjusted multiple contrasts				

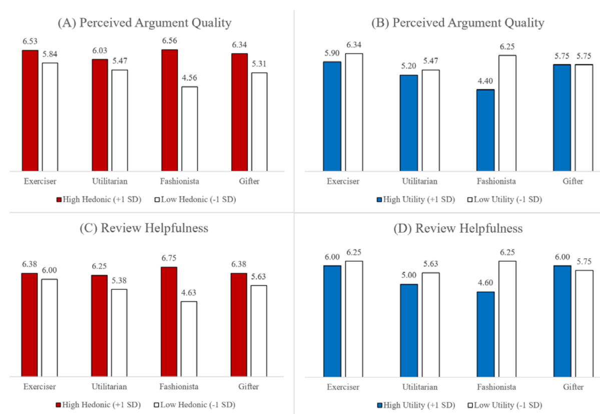
Table 2. Pairwise Contrasts

Specifically, in terms of review complexity, *Utilitarian* reviewers write more complex reviews than those with *Gifter* and *Fashionista* (marginally) profiles. In terms of recommendation, *Exerciser* reviewers will have a higher log-odds of rating the product positively comparing to *Gifter* and *Utilitarian* reviewers.

DISCUSSION, FUTURE PLAN, AND CONCLUSION

Our results provide empirical evidence about the feasibility of adapting an unsupervised machine learning technique to uncover hidden reviewer profiles in the textual content of the reviews. We also contribute to the online review generation by exploring the influences of purchasing goals (profiles) on subsequent review and rating behaviors. In detail, we found that more utility-oriented consumers will later write reviews with more complexity and are more likely to give the product a positive rating.

The next step to test H3a, H3b is underway, utilizing an online, repeated measure experiment, in which each participant provides their shopping orientation, then rates the argument quality and helpfulness of four reviews in a fully randomized order. The reviews are selected to be most representative of their "profile", after controlling for similar word count and complexity. Early results are shown in Figure 4 below. *Fashionista* and *Gifter* reviews' helpfulness and argument quality are rated higher by high hedonic-oriented shoppers (panel A and C), and *Fashionista* review is also rated highly in both measures by low utility-oriented shoppers (panel B and D). Constructs are measured using question items from established and validated sources (e.g., Babin et al. 1994), using a 7-point Likert-like scale. Pilot data were collected on Amazon Mechanical Turk (MTurk).



Besides the methodological and theoretical contribution, these findings are practically relevant. Online shopping platforms and sellers should consider tailoring highlighted reviews to match current consumers' shopping purposes or allow consumers to filter for reviews that match their own profiles.

The current research is not without limitations. First, despite having over 2,000 observations, the dataset is still a narrow sample of a single product category (smartwatch). However, this small sample is efficient for us to test the feasibility of the research-in-progress, and a future plan is in place to include additional product categories and reviews for robustness. Second, our initial analyses were conducted with simple models. Our next steps will involve examining more complex relationships (i.e., why some reviewers give "yes" recommendations despite giving two out of five stars). Also, more aspects of the written reviews documented in the literature, such as sentiment and emotion, shall be included in future models. Third, while the differences between the uncovered reviewer profiles generally support the hypotheses, some profiles need further examination. For instance, *Exercisers* and

Utilitarians, while both are utility-oriented, have differences in rating behaviors as big as that between *Exercisers* and *Fashionista*, a more hedonic-oriented profile. *Gifters*, on the other hand, show only a marginal difference with *Utilitarians* in review complexity. Further analysis taking into account various review characteristics may shed light on these groups' differences or lack thereof.

REFERENCES

1. Abbasi, A., Zhou, Y., Deng, S., and Zhang, P. 2018. "Text Analytics to Support Sense-Making in Social Media: A Language-Action Perspective," MIS Quarterly (42:2), MIS Quarterly, pp. 427-A38.
2. Adamopoulos, P., Ghose, A., and Todri, V. 2018. "The Impact of User Personality Traits on Word of Mouth: Text-Mining Social Media Platforms," Information Systems Research (29:3), pp. 612–640.
3. Babin, B. J., Darden, W. R., and Griffin, M. 1994. "Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value," *Journal of Consumer Research* (20:4), Oxford University Press / USA, pp. 644–656.
4. Baker, J., and Wakefield, K. L. 2012. "How Consumer Shopping Orientation Influences Perceived Crowding, Excitement, and Stress at the Mall," *Journal of the Academy of Marketing Science* (40:6), pp. 791–806.
5. Benbunan-Fich, R. 2020. "User Satisfaction with Wearables," *AIS Transactions on Human-Computer Interaction*, pp. 1–27.
6. Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," Journal of Machine Learning Research (3), pp. 993–1022.
7. Breinlinger, J., Hagiu, A., and Wright, J. 2019. "The Problems with 5-Star Rating Systems, and How to Fix Them," Harvard Business Review. (https://hbr.org/2019/07/the-problems-with-5-star-rating-systems-and-how-to-fix-them).
8. Büschken, J., and Allenby, G. M. 2016. "Sentence-Based Text Analysis for Customer Reviews," Marketing Science (35:6), pp. 953–975.
9. Chevalier, J. A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," Journal of Marketing Research (43:3), pp. 345–354.
10. Chiu, C.-M., Wang, E. T. G., Fang, Y.-H., and Huang, H.-Y. 2014. "Understanding Customers' Repeat Purchase Intentions in B2C e-Commerce: The Roles of Utilitarian Value, Hedonic Value and Perceived Risk," *Info. Systems Journal* (24:1), pp. 85–114.
11. Ho, Y.-C. (Chad), Wu, J., and Tan, Y. 2017. "Disconfirmation Effect on Online Rating Behavior: A Structural Model," Information Systems Research (28:3), pp. 626–642.
12. Hu, N., Pavlou, P. A., and Zhang, J. 2017. "On Self-Selection Biases in Online Product Reviews," MIS Quarterly (41:2), pp. 449–471.
13. Huang, A. H., Lehavy, R., Zang, A. Y., and Zheng, R. 2018. "Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach," Management Science (64:6), pp. 2833–2855.
14. Humphreys, A., and Wang, R. J.-H. 2018. "Automated Text Analysis for Consumer Research," Journal of Consumer Research (6), p. 1274.
15. Kaplan, S., and Vakili, K. 2015. "The Double-Edged Sword of Recombination in Breakthrough Innovation," Strategic Management Journal (36:10), pp. 1435–1457.
16. Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., and Chissom, B. S. 1975. "Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel," Naval Technical Training Command Millington TN Research Branch.
17. Liu, F., Lim, E. T. K., Li, H., Tan, C.-W., and Cyr, D. 2020. "Disentangling Utilitarian and Hedonic Consumption Behavior in Online Shopping: An Expectation Disconfirmation Perspective," *Information & Management* (57:3), p. 103199.
18. Moore, S. G. 2015. "Attitude Predictability and Helpfulness in Online Reviews: The Role of Explained Actions and Reactions," Journal of Consumer Research (42:1), pp. 30–44.
19. Mudambi, S. M., and Schuff, D. 2010. "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.Com," MISQ(34:1), pp.185-200.
20. Shi, Z., Lee, G. M., and Whinston, A. B. 2016. "Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence," MIS Quarterly (40:4), MIS Quarterly, pp. 1035-A53.
21. Steyvers, M., and Griffiths, T. 2006. "Probabilistic Topic Models," in Handbook of Latent Semantic Analysis, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 427–448.
22. Tirunillai, S., and Tellis, G. J. 2014. "Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation," Journal of Marketing Research (51:4), SAGE Publications Inc, pp. 463–479.
23. Wu, J. 2017. "Review Popularity and Review Helpfulness: A Model for User Review Effectiveness," Decision Sup. Sys. (97), pp. 92–103.
24. Yin, D., Bond, S. D., and Zhang, H. 2014. "Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews1," MIS Quarterly (38:2), pp. 539–560.