

When More is More and Less is More: Depth and Breadth of Product Reviews and Their Effects on Review Helpfulness

Jaebong Son
California State University,
Chico
json@csuchico.edu

Arash Negahban
California State University,
Chico
anegahban@csuchico.edu

Youngsu Lee
California State University,
Chico
ylee54@csuchico.edu

James R Connolly
California State University,
Chico
JConnolly@csuchico.edu

Dalen T Chiang
California State University,
Chico
DChiang@csuchico.edu

Abstract

With the growth of online shopping coupled with mobile technology, user-generated product reviews have become an important source of information for product diagnosticity. A significant academic endeavor has been made to comprehend what information factors of reviews help prospective customers better diagnose products. One such factor is review depth that is estimated by the number of a review's words. We propose review breadth as an additional factor based on a review's number of topics—the more review breadth, the more diverse information. By conducting the statistical and predictive analyses, we demonstrate that review breadth reliably measures a review's information. This study makes academic and practical contributions. For academic researchers, review breadth is worth considering as a factor to estimate a review's information over and above review depth. Based on the two information factors of review breadth and review depth, practitioners can recommend more helpful product reviews to their prospective customers.

1. Introduction

The Internet and Web technologies have formed the foundation of online shopping and electronic commerce (or e-commerce) [1]. By operating online stores, companies can expand their marketplace without being restricted by time and space. In the comfort of home, customers can easily explore and conveniently purchase products online. However, due to lack of opportunities to diagnose products in person,

concerns about product uncertainty are raised [2]. One viable means to alleviate such concerns is to utilize fellow consumers' product reviews or also known as "peer-generated product evaluation" [3-5].

Consumer product reviews (hereafter 'reviews' or 'product reviews') have been emphasized in the e-commerce context, since they are written on diverse matters (e.g., product quality, service satisfaction, company reputation, etc.) [3]. The literature on e-commerce and product reviews has revealed that consumers' feedback for products significantly influences prospective consumers' purchase decisions and thus products' sales [6-9]. TripAdvisor's research on hotel reviews pointed out the following: (a) more than 70% of its users have experiences in writing reviews; (b) its users usually read other users' reviews before booking hotels; (c) more than a half of its users are unwilling to book a hotel that has no reviews [10]. With the increasing importance of product reviews, the majority of online business-to-consumer companies provide a dedicated section of fellow consumers' reviews for their prospective customers [11]. Indeed, product reviews become a significant source of information for both prospective customers and e-commerce companies.

However, not all reviews convey equally valuable information, and accordingly individual reviews can be differently appreciated [12]. After reading a review, prospective customers can evaluate whether the review carries helpful information by pressing the "yes" or "no" button. The aggregation of "yes" votes indicates the extent to which reviews are informative (or *review helpfulness*)—the more helpful votes a review receives, the more informative it is considered to be [5]. Review helpfulness is a simple, but important feature of product reviews that can help prospective customers efficiently find helpful reviews for product

diagnosticity [11]. In fact, a review receiving more helpful votes has a higher probability of being exposed to other customers. Therefore, a growing body of literature has attempted to understand what makes reviews informative and thus helpful. Studies focusing on reviewer-related factors found that reviewers' profiles and credentials significantly influence review helpfulness [e.g., 13, 14, 15]. Studies examining review content-related factors revealed that review depth, valence, and readability reliably affect the helpfulness of reviews [e.g., 5, 13, 16].

Regarding product diagnosticity, in particular, review content-related factors are more relevant than reviewer-related ones, because the essential facet of review content is germane to estimating the amount of information that individual reviews carry. For example, each review's number of words is a commonly used indicator of *review depth*—the more words, the more in-depth review content [5, 17]. We argue, however, that while how in-depth information a review conveys is one factor to gauge a review's informativeness, how diverse subjects (or topics) a review carries (or *review breadth*) is another feature worth considering. In this study, we investigate review breadth as another measure of a review's informativeness.

The rest of the paper is organized as follows. We review the literature on product reviews and develop the research hypotheses. Then, we discuss the research methodology, including data collection, statistical and predictive analyses followed by the results and interpretation. We conclude by discussing the findings, limitations, and implications for future research.

2. Literature Reviews

As peer-consumers' subjective appraisal of product reviews, review helpfulness represents the perceived utility of information conveyed in individual reviews and thus serves as a benchmark to evaluate reviewer- and review content-related factors [11, 18].

Reviewer-related factors reveal reviewers' background, identity, and expertise [13]. Forman et al. reported that when reviewers disclosed their identity-descriptive information (e.g., real name and location), their reviews were considered more helpful than those without reviewers' identity disclosure [14]. Ghose and Ipeirotis found a positive relationship between review helpfulness and reviewer experience that is estimated by the total number of individual reviewers' past reviews [9]. Interestingly, however, for top reviewers, such an experience was not a significant determinant of review helpfulness [18]. Li et al.'s study stated that reviewers' social relationship (or social capital) significantly affected review helpfulness [15]. Due to

the anonymous nature of the Internet, reviewers' identity, review experience, and social relationship become important cues that influence the helpfulness of reviews. In fact, reviewer-related factors are mainly used to assess the credibility of reviews' conveying information, because reviewers are considered as the information source of such reviews [16, 18].

Review content-related characteristics are directly associated with information conveyed in reviews [5]. By examining the stylistic elements of product reviews, Schindler and Bickart reported that humor in reviews increased review helpfulness, while slang, misspelling, and grammatical mistakes (e.g., ungrammatical word sequence) negatively affected the helpfulness of reviews [19]. Cao et al. investigated reviewers' writing style by relying on the average length of sentences, the length of pros and cons, the length of titles, etc. [20]. They found that (a) as a review included more cons, its helpfulness increased and that (b) the length of titles negatively affected helpful votes. The star rating is another review-content feature used as a cue of a review's sentiment (e.g., a 5-scale rating). For example, a review is considered to convey negative information (i.e., negative sentiment), as its star rating closes to 1. With a rating of 5, a review is considered to convey positive information (i.e., positive sentiment). By conducting a meta-analysis study of the relationship between product reviews and review helpfulness, Hong et al. delineated both positive and negative effects of review sentiment on the helpfulness of reviews [4]. Wu observed that positive reviews had higher helpful votes than negative reviews [21]. On the contrary, Kwok and Xie stated that prospective customers perceived negative reviews more helpful than positive ones [22]. Mudambi and Schuff further researched review sentiment by examining its curvilinear relationship with review helpfulness and its conditional relationship with the types of goods (e.g., experience and search goods) [5]. They found that for experience goods, moderate reviews were more helpful than either extremely positive or negative reviews. A sentence-level sentiment analysis was also performed by Willemsen et al. [23], stating that negative and positive information together in a review increased its helpful votes. By counting positive and negative words in individual reviews, Baek et al. observed that as a review included more negative words, its helpfulness increased [16].

Along with the above-mentioned factors of product reviews, the amount of information conveyed in individual reviews has been used as a strong determinant of review helpfulness. Most research on product reviews leveraged review depth (i.e., review length) as a proxy to quantify each review's information [e.g., 5, 13, 16]. However, we contend that

review depth alone is incomplete to evaluate a review's information and thus that review breadth, as a complementary factor, should be used together with review depth. In the next section, we develop the research hypotheses to investigate review depth and review breadth together in association with review helpfulness.

3. Hypothesis Development

Information search is an essential part of the purchase decision-making process, when information asymmetry exists between prospective customers and products [24]. Information asymmetry is reduced through information acquisition [25]. Hence, prospective customers expend an effort to find more information to be better informed of the products of their interest—the more product information they gather, the higher product diagnosticity they can achieve [26]. In the literature on e-commerce and product reviews, the length of reviews is used to gauge the extent to which individual reviews are informative and thus helpful. Pan and Zhang claimed that longer reviews are more convincing than shorter reviews, because the former carries more information than the latter [27]. Similarly, Mudambi and Schuff concluded that when a product review is written longer, it conveys more in-depth information, and thus it is considered more informative [5].

By aligning the previous studies, we posit that reviews' in-depth information alleviates information asymmetry and thus reduces product uncertainty, positively affecting the helpfulness of reviews. Therefore, we expect the following relationship:

Hypothesis 1 (H1): Review depth has a positive relationship with review helpfulness.

Along with review depth, we use review breadth that defines information as a type of “telling.” In other words, what subjects or topics that reviews are written about are information. In fact, information in a message is encoded in an agreed-on set of signals—e.g., words, letters, etc.—and, therefore, a message's information is extracted by decoding such signaling components [28]. Since decoding is a process of understanding a message, the message's topics can be considered as an outcome of the decoding process. From this perspective, we utilize a message's topic(s) as another measure for information. As a result, the more topics a review discusses, the broader aspects of products a review is supposed to diagnose, possibly indicating the degree of a review's breadth of information. We argue that when all things are equal, reviews with more topics are believed to be more

informative than those with fewer topics. Therefore, our second hypothesis states:

Hypothesis 2 (H2): Review breadth has a positive relationship with review helpfulness.

Interestingly, however, review breadth is in somewhat inverse proportion to review depth. That is, as a review broadens its coverage by adding more topics, its individual topics may lose specific details. It seems unavoidable that a review with more topics has less details per topic than another review with fewer topics, when both reviews have the same length—as the number of topics in a review increases, the number of words assigned to each topic decreases, producing less detailed topics. Based on the above speculation, we assume a conditional relationship between review breadth and review depth as follows:

Hypothesis 3 (H3): Review breadth weakens the effect of review depth on review helpfulness.

4. Research Methodology

4.1. Data and variables

We examine the hypotheses by employing the reviews of eight products available on Best Buy's website. Each product is categorized into either credence, experience, or search goods to reflect the different nature of products [29, 30]. In total, 32,070 reviews posted between year of 2014 and 2017 are used for the empirical analysis.

The dependent variable of this study is review helpfulness that indicates the extent to which each review is considered helpful by other consumers. The key independent variables are review depth and review breadth. Review depth of a review is measured by its number of words—the more words, the deeper depth [5]. Review breadth of a review is represented by the number of topics conveyed in that review—the more topics, the wider breadth. Table 1 shows descriptive statistics of independent and dependent variables.

We use MALLETT (*MAchine Learning for Language Toolkit*) to identify topics in reviews [31]. This Java library provides a *latent Dirichlet allocation* (LDA)-based topic modeling technique to discover documents' thematic structures or topics [32]. To have consistent and interpretable results from the LDA topic modeling, we leverage *n*-gram noun phrases [33] based on a series of text analyses: (1) lemmatizing each word into standard forms in order to minimize inflectional or derivationally related forms of words by utilizing the *Stanford CoreNLP* toolkit [34]; (2) eliminating stop

Table 1. Variable description

Variables	Explanation	Mean	S.D.	Range
Dependent				
<i>Helpfulness_i</i>	The number of helpful votes in review <i>i</i>	.131	2.16	0-153
Independent				
<i>Depth_i</i>	The number of words in review <i>i</i>	18.6	16.9	5-518
<i>Breadth_i</i>	The number of topics in review <i>i</i>	1.86	.81	1-6
Control				
<i>Rating_i</i>	Review <i>i</i> 's rating	4.7	.64	1-5
<i>Rating_i²</i>	The quadratic effect of review <i>i</i> 's rating			
<i>Experience</i> <i>Credence</i>	Contrast codes to indicate the product types—Experience (2) over Credence (-1) and Search (-1); Credence (1) over Search (-1)			
<i>Days_i</i>	Difference in days between the first review of a product and review <i>i</i>	419.2	327.8	0-1331
<i>Popularity_i</i>	The number of reviews of the product that review <i>i</i> is about	5310	2272	1830-7460
<i>Past_Reviews_i</i>	The number of past reviews of review <i>i</i> 's author	5.479	12.77	0-320
<i>Past_Helpful_Review</i> <i>YN_i</i>	A dummy variable to indicate whether review <i>i</i> 's author has at least one helpful review (i.e., a review with more than one helpful vote)—no helpful review (0); helpful review(s) (1)			

words (i.e., noisy words); (3) extracting each word's part-of-speech tag to identify multi-word noun phrases (or *n*-gram noun words).

One important procedure of the LDA topic modeling is to determine the optimal number of topics (or *k* topics), since it is one of the unsupervised techniques. Perplexity is a common measure to evaluate the goodness-of-fit of topic models—a lower perplexity represents a more generalized topic model [32]. By generating 99 topic models per product (e.g., topic models having 2 to 100 topics), we evaluate each model's generalizability by perplexity and then choose one topic model per product whose generalizability is optimized. Table 2 shows each product's optimal number of topics.

Table 2. The Number of topics per product

Product	Type	<i>k</i> Topics
Office Home & Student	Experience	16
Kaspersky Internet Security		23
Ultra+ 32GB	Credence	21
EVO+ 64GB SDXC		17
α a6000 Mirrorless Camera	Search	38
EOS Rebel T6 Camera		41
iPhone 7+ 128GB		27
Galaxy S8+ 64Gb		26

4.2. Statistical analysis

The dependent variable of this study is a count of the number of times a review was identified as helpful—*Helpfulness*. It has been reported that ordinary least squares produce inconsistent, biased estimates due to the discrete distribution of a count-

dependent variable [35, 36]. Therefore, we leverage specific statistical procedures that are designed to properly deal with the discrete distribution, such as Poisson or negative binomial models [37]. A negative binomial model is preferred to a Poisson model, when a count-dependent variable is overdispersed (i.e., a substantially larger variance than mean) [38], a similar violation of the homoscedasticity assumption [39]. The likelihood-ratio test of alpha confirms that the negative binomial model is preferred to the Poisson model [35].

$$\begin{aligned}
 \text{Helpfulness}_i = & \beta_0 + \underbrace{\beta_1 \text{Depth}_i + \beta_2 \text{Breadth}_i}_{\text{Reviews' Information}} \\
 & + \beta_3 \text{Depth}_i \times \text{Breadth}_i \\
 & + \underbrace{\beta_4 \text{Rating}_i + \beta_5 \text{Rating}_i^2 + \beta_6 \text{Experience}_i}_{\text{Controls-reviews}} \\
 & + \underbrace{\beta_7 \text{Credence}_i + \beta_8 \text{Days}_i + \beta_9 \text{Popularity}_i}_{\text{Controls-reviews cont'd}} \\
 & + \underbrace{\beta_{10} \text{Past_Reviews}_i}_{\text{Control-reviewers cont'd}} + \varepsilon_i
 \end{aligned}$$

Figure 1. Empirical model

When a count-dependent variable includes excessive zeros, both Poisson and negative binomial models can underestimate standard errors and p-values [40, 41]. In our review data, only 1,200 reviews out of 32,066 received at least one helpful vote. That being said, the rest of 30,866 reviews did not receive any helpful vote. We handled both issues of overdispersion and excessive zeros by utilizing the zero-inflated negative binomial (ZINB) model [42, 43]. The Vuong test on the empirical model shown in Figure 1 suggested the use of the ZINB model to estimate our

review data (see Model 4 of Table 3) [44]. Last, from the variance inflation factor (VIF) analysis, we were informed that multicollinearity is not a problem for the empirical model (Mean=1.80; Minimum=1.06; Maximum=3.81). None of the VIFs exceed the acceptable level of 5 [45].

includes the control variables. We controlled product types, review extremity, and reviewers' past reviews [5]. Model 2 adds Model 1 review depth. Model 3 adds Model 2 review breadth. Model 4 adds Model 3 an interaction term between review depth and breadth. We examined this interaction term by centering all

Table 3. Results of the hierarchical regression analysis

Variables \ Models	Model 1	Model 2	Model 3	Model 4
Main				
<i>Depth_i</i>	-	0.0473*** (0.00337)	0.0435*** (0.00376)	0.0564*** (0.00414)
<i>Breadth_i</i>	-	-	0.155* (0.0735)	0.247*** (0.0700)
Interaction				
<i>Depth_i × Breadth_i</i>	-	-	-	-0.0119*** (0.00112)
Control				
<i>Rating_i</i>	-0.314 (0.185)	-0.120 (0.156)	-0.145 (0.155)	-0.117 (0.149)
<i>Rating_i²</i>	0.0937 (0.0567)	0.111* (0.0559)	0.104 (0.0539)	0.106* (0.0530)
<i>Experience_i</i>	-0.00261*** (0.000227)	-0.00226*** (0.000233)	-0.00226*** (0.000229)	-0.00228*** (0.000226)
<i>Credence_i</i>	-0.343*** (0.0525)	-0.265*** (0.0449)	-0.257*** (0.0446)	-0.265*** (0.0460)
<i>Days_i</i>	-1.637*** (0.136)	-1.030*** (0.118)	-1.026*** (0.118)	-1.020*** (0.120)
<i>Popularity_i</i>	0.000185*** (0.0000507)	0.000120* (0.0000472)	0.000126** (0.0000471)	0.000123* (0.0000480)
<i>Past_Reviews_i</i>	-0.0203*** (0.00578)	-0.0182*** (0.00490)	-0.0174*** (0.00476)	-0.0196*** (0.00492)
<i>Constant</i>	-2.838*** (0.248)	-3.144*** (0.243)	-3.183*** (0.243)	-3.132*** (0.246)
Inflate				
<i>Past_Helpful_Review_YN_i</i>	-22.02*** (0.166)	-23.37*** (0.143)	-27.54*** (0.142)	-26.59*** (0.147)
<i>Constant</i>	1.181*** (0.0877)	0.843*** (0.104)	0.838*** (0.104)	0.897*** (0.0984)
Inalpha				
<i>Constant</i>	2.232*** (0.0914)	1.868*** (0.0900)	1.855*** (0.0894)	1.756*** (0.0902)
Model summary				
<i>McFadden's R²</i>	0.117***	0.168***	0.170***	0.175***
<i>Likelihood Ratio</i>	LR(9)=1658.9***	LR(10)=2406.5***	LR(11)=2417.1***	LR(12)=2487.6***
<i>Vuong test</i>	-	-	-	z=8.47***
<i>n</i>	32066 (Nonzero obs.: 1200, Zero obs.: 30866)			

† Unstandardized coefficients with robust errors in parentheses are shown (*p<0.05, **p<0.01, ***p<0.001).

We performed the hierarchical regression analysis by defining the following four models: Model 1 only

numerical variables from their means [46]. By following Cameron and Trivedi's recommendation

[35], we estimated these empirical models with robust standard errors. The regression results of the models are summarized in Table 3.

To evaluate the hypotheses, we utilize Model 4, as it includes all the necessary variables. We found that review depth positively affects review helpfulness, supporting H1. An additional word significantly increases the number of helpful votes by 5.8% on average ($\beta_{\text{Depth}}=0.0564^{***}$), while holding the other variables constant in the model. Moreover, the effect of review depth on the helpfulness of reviews is consistent through Models 2, 3, and 4.

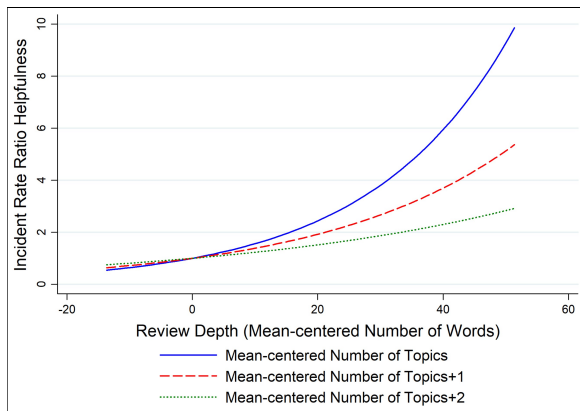


Figure 2. Interaction plot between review depth and breadth¹

It turned out that review breadth increases review helpfulness. In other words, a review’s helpfulness increases by 28.5% on average, as it conveys an additional topic, given that the other variables in this model are held constant ($\beta_{\text{Breadth}}=0.247^{***}$). Therefore, H2 is supported. As we expected, review breadth significantly diminishes the positive relationship between review depth and review helpfulness ($\beta_{\text{Depth} \times \text{Breadth}}=-0.0119^{***}$), supporting H3. The interaction relationship is graphed in Figure 2.

4.3. Predictive analytics

By statistically modeling the review data, we examined how well review breadth contributes to explaining the degree to which reviews are helpful over and above review depth. In order to further support review breadth as a reliable measure for a review’s information, we analyze the same data by predictive analytics techniques. In other words, we investigated review breadth’s predictive capability to foretell whether *unobserved* reviews will be considered to be helpful.

Each review’s helpfulness is defined by a ratio of the number of helpful votes, compared with the total number of votes (helpful and unhelpful)—a review is considered helpful, when its ratio is greater than or equal to 0.5. Otherwise, it is considered unhelpful. By using this helpfulness ratio and reviews’ information components of *Rating*, *Words*, and *Topics*, we formulated the following three predictive models. Predictive Model 1 (PM1) as a baseline includes *Rating*. Predictive Model 2 (PM2) adds PM1 *Words* (i.e., review depth). Predictive Model 3 (PM3) enhances PM2 by including *Topics* (i.e., review breadth). The predictive models were trained by two classification algorithms of Bayesian networks (Bayes) and support vector machines (SVM) to build classifiers by the following steps [47]; (1) 2000 reviews were chosen by a stratified sampling scheme in order to build unbiased classifiers [48]—randomly selected 1000 helpful reviews and 1000 unhelpful reviews; (2) the selected reviews were randomly split into a training set (70%) and test set (30%); (3) each algorithm built each model’s classifier by using a training set and evaluated it by using a test set (or unobserved reviews); (4) the above steps were repeated five times to generalize the performance of a classifier by avoiding overfitting (called *n*-fold cross validation [47]). We assessed classifiers’ performance by the Area Under a

Table 4. Prediction accuracy of classifiers by AUC

5-fold cross validation		1	2	3	4	5	Average	Difference
Algorithms	Models							
Bayes	PM1 (Rating)	0.509	0.516	0.549	0.542	0.526	0.5284	-
	PM2 (PM1+Words)	0.531	0.528	0.556	0.570	0.556	0.5482	0.0198 [†]
	PM3 (PM2+Topics)	0.609	0.624	0.642	0.644	0.649	0.6336	0.0854 ^{††}
SVM	PM1 (Rating)	0.491	0.489	0.51	0.536	0.486	0.5024	-
	PM2 (PM1+Words)	0.576	0.626	0.579	0.620	0.624	0.6050	0.1026 [†]
	PM3 (PM2+Topics)	0.639	0.667	0.652	0.693	0.684	0.6670	0.0620 ^{††}

[†] Difference in AUC average between PM1 and PM2; ^{††} Difference in AUC average between PM2 and PM3.

¹ The number of words (i.e., review depth) is restricted to two standard deviations (2SD) that include 95% of reviews.

ROC Curve (AUC) shown in Table 4. While a ROC (Receiver Operating Characteristic) curve evaluates the overall performance of a classifier by two dimensions

of the true positives and the false positives,² AUC derived from a ROC curve is a single measure of the prediction performance.

The range of AUC values is between 0.0 and 1.0—the higher AUC value, the more accurate the classifier is; a 0.5 AUC value indicates random guessing. Figure 3 shows three models’ ROC curves by the two algorithms on the last iteration of cross validation. These ROC curves visually show that over and above *Rating* and *Words*, *Topics* constantly improved the prediction accuracy of classifiers (i.e., PM3).

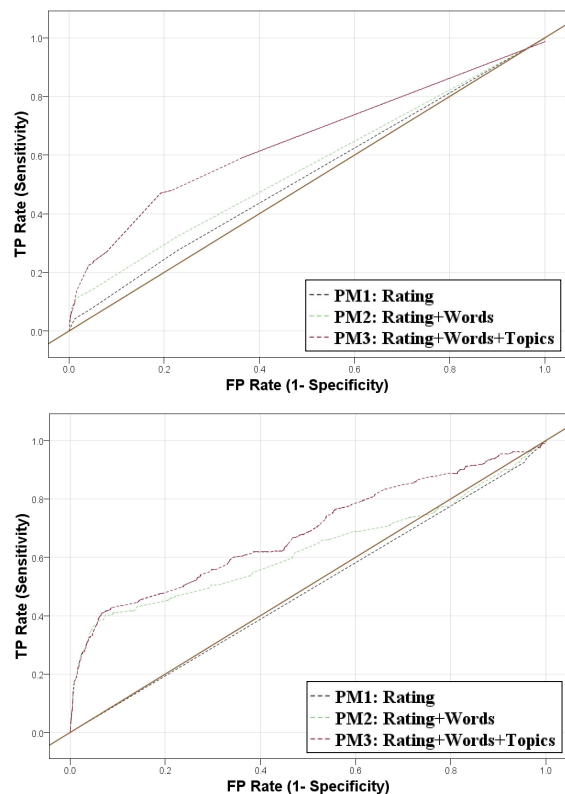


Figure 3. ROC curves by Bayesian network (above) and by SVM (below)

In this regard, we found from the 5th iteration of cross validation in Table 4 that the AUC values of PM3 support the above observation, in that *Topics* enhanced Bayes’ prediction accuracy by 9.3% from 55.6% to 64.9% and that of SVM by 6%. This constant improvement was shown throughout all five iterations regardless of which algorithm was used, resulting in an average increase in Bayes’ and SVM’s prediction accuracy by 8.5% and 6.2%, respectively.

² *True positive*: a review is predicted to be helpful, when it is helpful; *false positive*: a review is predicted to be helpful, when it is unhelpful.

5. Discussion

This study has scrutinized reviews’ information by depth and breadth. Review depth is represented by a simple count of words [5]. Review breadth is defined by a chunk of information, called topics—as a review conveys more topics, it is considered to have more information (and thus informative). By the statistical analyses, we investigated review breadth in concert with review depth. By the predictive analytics, we further corroborated evidence for review breadth as a predictor of whether or not reviews are considered to be helpful.

The statistical results support the hypotheses regarding how a review’s depth, breadth, or both influences its helpfulness. First, the length of reviews matters. The longer a review is, the more informative it is considered to be, increasing the helpfulness of reviews. Second, the number of topics conveyed in a review is also an indicator of its informativeness. When a review carries more topics, its informativeness increases, resulting in more helpful votes. However, review breadth diminishes the positive effect of review depth on the helpfulness of reviews. As a matter of fact, writing a review to convey both in-depth and wide-ranging information of a product is a challenging task. To put it differently, unless a review’s length becomes longer in proportion to an increase in topics, each topic will be less specific, negatively influencing the overall depth of a review. We checked our surmise by examining the average length of reviews per topic increase. Figure 4 demonstrates that the review length does not proportionally increase from 1 through 4 topics, while that relationship becomes more proportional between 4 and 6 topics. It appears that roughly 63% of the reviews in our dataset convey 2, 3, or 4 topics, supporting our explanation on the interplay between review depth and breadth.

We are further convinced from the predictive analytics that review breadth is a significant factor of reviews, because it constantly improves the prediction accuracy of reviews to be helpful or not. Simply put, without review breadth, we will have far less accurate classifiers.

The limitations of the study open opportunities for future research. First, we leveraged topic modeling to estimate individual reviews’ breadth of information. While topic modeling is a convenient way to algorithmically analyze a large volume of product reviews, it would be a necessary task to check how reliably human beings agree with what this topic modeling algorithm produces. Future research can employ actual (or potential) consumers to examine the reliability of topics produced by the LDA algorithm. Second, this study mainly focuses on the information

perspective of reviews in association with review helpfulness. Even though we partly addressed reviewers' factors influencing review helpfulness, there exist other factors that are known to influence review helpfulness, such as reviewers' expertise and social relationships [4], the readability of reviews [9], etc. Including such relevant factors in empirical models can improve our understanding regarding the current findings. Third, investigating an optimal review length per topic will be of interest. Last, although this study uses reviews of eight products available on Best Buy's website, the generalizability of the empirical results can be further strengthened by including product reviews from different e-commerce sites.

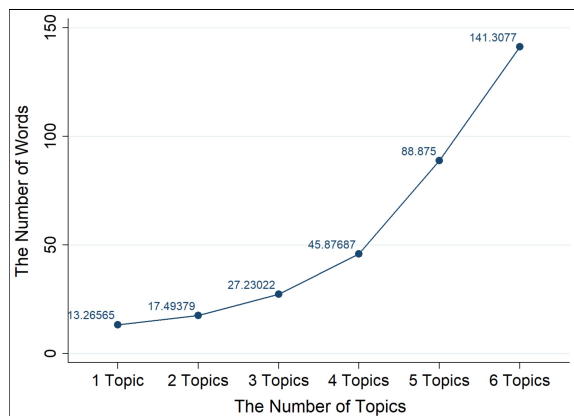


Figure 4. Average number of words by the number of topics

6. Conclusions

According to a 2018 Consumer Research report [50], 82% of consumers read more than 3 reviews before purchasing electronics. When buying clothing, 68% of consumers read more than 3 reviews. Indeed, user-generated product reviews are important, as fellow consumers' personal experiences and opinions can help prospective customers increase product diagnosticity [9, 13, 49]. This study delves into reviews' information by depth and breadth, each of which is expected to be differently manifested in relation to review helpfulness. We also present how to measure review breadth by leveraging the LDA algorithm.

This study contributes to the extant literature of product reviews and e-commerce. Along with review depth, review breadth is a reliable indicator that evaluates the extent to which individual reviews are informative. The statistical results show that the effect of review breadth on review helpfulness is significant over and above review depth. Furthermore, we confirm from the predictive analytics that review breadth is an

effective predictor of whether reviews are considered to be helpful. All things taken together, a product review becomes more helpful, as it conveys more in-depth or breadth information. However, its information depth and breadth must be balanced to each other.

This study also makes several practical contributions. For reviewers, it is recommended to write a review by balancing between in-depth and breadth information of a product. To make a review convey in-depth information per topic, its length must increase as its topics become diverse. For e-commerce sites, it can be an effective strategy to reveal individual product reviews with the number of topics. In so doing, prospective customers can quickly and easily grasp more informative reviews, efficiently alleviating their product concerns. Enhancing search functionality for product reviews by including review breadth is a feasible recommendation. Instead of relying on a few keywords to search for product reviews, a topic-based search function can facilitate potential customers to discover more relevant and helpful reviews.

All things considered, this study sheds light on an additional factor to estimate a product review's information. Unlike a traditional measure of review depth (or a simple count of words), review breadth uses a chunk of information (or topics) that is identified by the LDA algorithm, which mimics how human beings understand texts [32]. Therefore, we conclude that review breadth is an important factor that can improve our current understanding of user-generated product reviews.

7. References

- [1] W. H. Dutton, "The Internet and social transformation: Reconfiguring access," *Transforming enterprise: The economic and social implications of information technology*, pp. 375-397, 2005.
- [2] P. A. Pavlou, H. Liang, and Y. Xue, "Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective," *MIS quarterly*, pp. 105-136, 2007.
- [3] C. Dellarocas, "The digitization of word of mouth: Promise and challenges of online feedback mechanisms," *Management science*, vol. 49, no. 10, pp. 1407-1424, 2003.
- [4] H. Hong, D. Xu, G. A. Wang, and W. Fan, "Understanding the determinants of online review helpfulness: a meta-analytic investigation," *Decision Support Systems*, vol. 102, pp. 1-11, 2017.
- [5] S. M. Mudambi and D. Schuff, "Research note: What makes a helpful online review? A study of customer reviews on Amazon. com," *MIS quarterly*, pp. 185-200, 2010.
- [6] B. Gu, J. Park, and P. Konana, "Research note—the impact of external word-of-mouth sources on retailer sales of

- high-involvement products," *Information Systems Research*, vol. 23, no. 1, pp. 182-196, 2012.
- [7] K. L. Xie, Z. Zhang, and Z. Zhang, "The business value of online consumer reviews and management response to hotel performance," *International Journal of Hospitality Management*, vol. 43, pp. 1-12, 2014.
- [8] X. Li, L. M. Hitt, and Z. J. Zhang, "Product reviews and competition in markets for repeat purchase products," *Journal of Management Information Systems*, vol. 27, no. 4, pp. 9-42, 2011.
- [9] A. Ghose and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 10, pp. 1498-1512, 2011.
- [10] "TripAdvisor Network Effect and the Benefits of Total Engagement." TripAdvisor.
<https://www.tripadvisor.com/TripAdvisorInsights/w828>
(accessed June 07, 2019).
- [11] Y. Wan and M. Nakayama, "The reliability of online review helpfulness," *Journal of Electronic Commerce Research*, vol. 15, no. 3, p. 179, 2014.
- [12] R. Rietsche, D. Frei, E. Stoeckli, and M. Söllner, "NOT ALL REVIEWS ARE EQUAL-A LITERATURE REVIEW ON ONLINE REVIEW HELPFULNESS," in *European Conference on Information Systems*, Stockholm-Uppsala, Sweden, 2019: AIS.
- [13] N. Hu, N. S. Koh, and S. K. Reddy, "Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales," *Decision support systems*, vol. 57, pp. 42-53, 2014.
- [14] C. Forman, A. Ghose, and B. Wiesenfeld, "Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets," *Information Systems Research*, vol. 19, no. 3, pp. 291-313, 2008.
- [15] Q. Li, J. Cui, and Y. Gao, "The influence of social capital in an online community on online review quality in China," in *2015 48th Hawaii International Conference on System Sciences*, 2015: IEEE, pp. 562-570.
- [16] H. Baek, J. Ahn, and Y. Choi, "Helpfulness of online consumer reviews: Readers' objectives and review cues," *International Journal of Electronic Commerce*, vol. 17, no. 2, pp. 99-126, 2012.
- [17] A. Y. Chua and S. Banerjee, "Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth," *Journal of the Association for Information Science and Technology*, vol. 66, no. 2, pp. 354-362, 2015.
- [18] A. H. Huang, K. Chen, D. C. Yen, and T. P. Tran, "A study of factors that contribute to online review helpfulness," *Computers in Human Behavior*, vol. 48, pp. 17-27, 2015.
- [19] R. M. Schindler and B. Bickart, "Perceived helpfulness of online consumer reviews: The role of message content and style," *Journal of Consumer Behaviour*, vol. 11, no. 3, pp. 234-243, 2012.
- [20] Q. Cao, W. Duan, and Q. Gan, "Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach," *Decision Support Systems*, vol. 50, no. 2, pp. 511-521, 2011.
- [21] P. F. Wu, "In search of negativity bias: An empirical study of perceived helpfulness of online reviews," *Psychology & Marketing*, vol. 30, no. 11, pp. 971-984, 2013.
- [22] L. Kwok and K. L. Xie, "Factors contributing to the helpfulness of online hotel reviews: does manager response play a role?," *International Journal of Contemporary Hospitality Management*, vol. 28, no. 10, pp. 2156-2177, 2016.
- [23] L. M. Willemsen, P. C. Neijens, F. Bronner, and J. A. De Ridder, "'Highly recommended!' The content characteristics and perceived usefulness of online consumer reviews," *Journal of Computer-Mediated Communication*, vol. 17, no. 1, pp. 19-38, 2011.
- [24] G. A. Akerlof, "The market for "lemons": Quality uncertainty and the market mechanism," in *Uncertainty in economics*: Elsevier, 1978, pp. 235-251.
- [25] S. Viswanathan, J. Kuruzovich, S. Gosain, and R. Agarwal, "Online infomediaries and price discrimination: Evidence from the automotive retailing sector," *Journal of Marketing*, vol. 71, no. 3, pp. 89-107, 2007.
- [26] Y. Hong and P. A. Pavlou, "Product uncertainty in online marketplaces in China: an econometric model," in *America's conference on information systems, Lima, Peru*, 2010.
- [27] Y. Pan and J. Q. Zhang, "Born unequal: a study of the helpfulness of user-generated product reviews," *Journal of Retailing*, vol. 87, no. 4, pp. 598-612, 2011.
- [28] D. O. Case and L. M. Given, *Looking for information: A survey of research on information seeking, needs, and behavior*. Emerald Group Publishing, 2016.
- [29] P. Nelson, "Information and consumer behavior," *Journal of political economy*, vol. 78, no. 2, pp. 311-329, 1970.
- [30] R. B. Ekelund, F. G. Mixon, and R. W. Ressler, "Advertising and information: an empirical study of search, experience and credence goods," *Journal of Economic Studies*, vol. 22, no. 2, pp. 33-43, 1995.
- [31] S. Graham, S. Weingart, and I. Milligan, "Getting started with topic modeling and MALLET," The Editorial Board of the Programming Historian, 2012.
- [32] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993-1022, 2003.
- [33] A. Poulston, M. Stevenson, and K. Bontcheva, "Topic models and n-gram language models for author profiling," in *Conference and Labs of the Evaluation Forum*, Toulouse, France, 2015.

- [34] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The Stanford CoreNLP natural language processing toolkit," in *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 2014, pp. 55-60.
- [35] A. C. Cameron and P. K. Trivedi, *Regression Analysis of Count Data, 2nd edition* (Econometric Society Monograph). Cambridge University Press, 2013.
- [36] W. Gardner, E. P. Mulvey, and E. C. Shaw, "Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models," *Psychological bulletin*, vol. 118, no. 3, p. 392, 1995.
- [37] R. B. O'hara and D. J. Kotze, "Do not log-transform count data," *Methods in Ecology and Evolution*, vol. 1, no. 2, pp. 118-122, 2010.
- [38] J. M. Hilbe, *Negative binomial regression*. Cambridge University Press, 2011.
- [39] A. C. Cameron and P. K. Trivedi, *Essentials of count data regression* (A companion to theoretical econometrics). 350 Main Street, Malden, MA 02148, USA: Blackwell Publishing Ltd, 2001, p. 18.
- [40] C. Gourieroux, A. Monfort, and A. Trognon, "Pseudo maximum likelihood methods: Theory," *Econometrica: Journal of the Econometric Society*, pp. 681-700, 1984.
- [41] A. C. Cameron and P. K. Trivedi, "Econometric models based on count data. Comparisons and applications of some estimators and tests," *Journal of applied econometrics*, vol. 1, no. 1, pp. 29-53, 1986.
- [42] D. Lambert, "Zero-inflated Poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, no. 1, pp. 1-14, 1992.
- [43] S. J. Long, J. S. Long, and J. Freese, *Regression models for categorical dependent variables using Stata*. Stata press, 2006.
- [44] Q. H. Vuong, "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica: Journal of the Econometric Society*, pp. 307-333, 1989.
- [45] E. K. David A. Belsley, Roy E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons, 2005.
- [46] L. S. Aiken, S. G. West, and R. R. Reno, *Multiple regression: Testing and interpreting interactions*. Sage, 1991.
- [47] T. Wendler and S. Gröttrup, *Data mining with SPSS modeler: theory, exercises and solutions*. Springer, 2016.
- [48] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerging artificial intelligence applications in computer engineering*, vol. 160, pp. 3-24, 2007.
- [49] B. Bickart and R. M. Schindler, "Internet forums as influential sources of consumer information," *Journal of interactive marketing*, vol. 15, no. 3, pp. 31-40, 2001.
- [50] "2018 Consumer Research: Why Product Experiences are What Win Consumer Trust in the Digital Age," Salsify, <https://www.salsify.com/>, 2018, vol. 2019. [Online]. Available: <https://www.salsify.com/hubfs/2018%20Consumer%20Research%20Report/2018%20Consumer%20Research%20Report%20-%20Final.pdf>