Decision Support Systems 50 (2011) 511–521

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach

Qing Cao [a,1], Wenjing Duan [b,*], Qiwei Gan [a,2]

[a] Area of Information Systems and Quantitative Sciences, Rawls College of Business, Texas Tech University, Lubbock, TX 79409-2101, United States

[b] Department of Information System & Technology Management, School of Business, Funger Hall, Suite 515, The George Washington University, 2201 G Street, NW, Washington, DC 20052, United States

ARTICLE INFO

Article history:
Received 2 March 2010
Received in revised form 18 August 2010
Accepted 4 November 2010
Available online 11 November 2010

Keywords:
Online user review
Helpfulness
Text mining
Ordinal logistic regression
Latent Semantic Analysis

ABSTRACT

The "helpfulness" feature of online user reviews helps consumers cope with information overloads and facilitates decision-making. However, many online user reviews lack sufficient helpfulness votes for other users to evaluate their true helpfulness level. This study empirically examines the impact of the various features, that is, basic, stylistic, and semantic characteristics of online user reviews on the number of helpfulness votes those reviews receive. Text mining techniques are employed to extract semantic characteristics from review texts. Our findings show that the semantic characteristics are more influential than other characteristics in affecting how many helpfulness votes reviews receive. Our findings also suggest that reviews with extreme opinions receive more helpfulness votes than those with mixed or neutral opinions. This paper sheds light on the understanding of online users' helpfulness voting behavior and the design of a better helpfulness voting mechanism for online user review systems.

1. Introduction

Understanding the role of online user reviews in e-commerce has become an increasingly important subject for both academics and practitioners [8,10,30]. Online user reviews are regarded as digitalized word of mouth [7] and found to be influential on product sales and consumer decision-making [9]. The huge amount of information available on the Web has created information overload among online users [2,14]. The information overload spans two dimensions. First, as a result of the Internet's unprecedented spread and the virtually unlimited online shelf space of e-tailers, the number and types of products available online have grown exponentially. Online consumers often find they lack the knowledge and time to make the best possible decision out of numerous competing products on various websites. Online user review systems, in this sense, provide a venue for consumers to share their opinions and experience on products. The venue, in turn, offers valuable resources of information for potential buyers to make more efficient and rational purchase decisions.

Second, even though the information overload created by the availability of a large amount and variety of products online could be mitigated by referring to online user reviews to some extent, the utter volume of available online user reviews as well as the great variations in the review content and quality create another big obstacle to consumers who wish to take full advantage of the reviews [21]. For example, the number of reviews for any average ranked book on Amazon.com can easily reach more than several hundred, whereas for popular titles such as *Harry Potter and the Deathly Hallows*, the number of reviews can be in the thousands. It is virtually impossible for consumers to read all the reviews before making purchase decisions, especially for products that have been reviewed by hundreds and sometimes thousands of customers with inconsistent opinions. Instead, what consumers really need could be just a few of the most "helpful" reviews. Many websites encourage users to evaluate the "helpfulness" of user reviews by simply asking anyone who read the review to vote on the question "Was this review helpful to you?". In order to highlight such a feature, many websites display reviews based on the helpfulness voting. For example, the default setting to display the search result on Amazon.com is to rank the reviews according to the helpfulness vote. This feature allows consumers to quickly find the most helpful reviews and makes the consumer decision-making process less tedious and more efficient, thus attracting more consumers and improving the reputation of the e-commerce websites. It is estimated that this simple question "Was this review helpful to you?" brings in about $2.7 billion additional revenue to Amazon.com [27].

However, the helpfulness voting is not a panacea. Not all online reviews received helpfulness votes; instead, a large portion of online user reviews on many popular websites, such as Amazon.com and CNET Download.com (CNETD), do not receive any helpfulness votes.

* Corresponding author. Tel.: +1 806 742 3165; fax: +1 806 742 3193.
E-mail addresses: qing.cao@ttu.edu (Q. Cao), wduan@gwu.edu (W. Duan), qiwei.gan@ttu.edu (Q. Gan).
[1] Tel.: +1 806 742 3919; fax: +1 806 742 3193.
[2] Tel.: +1 202 994 3217; fax: +1 202 994 5830.

0167-9236/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2010.11.009

Without helpfulness votes, the helpfulness voting mechanism does not work effectively. Before reading a review, a user could obtain some rough ideas about the helpfulness of the review by looking at its helpfulness votes. If there is no vote on a review, a user cannot evaluate its helpfulness before reading the review. Hence, even if there is a helpfulness voting mechanism, without actual votes, it cannot facilitate users to locate the most helpful reviews effectively. In addition, since consumers are expected to pay greater attention to the most helpful reviews, less helpful reviews become less attractive to consumers. This may create a vicious cycle in which the more helpful reviews attract more readers and hence receive even more votes on helpfulness, while the less helpful reviews attract fewer readers and hence are even less likely to receive helpfulness votes. As a result, users may ignore these reviews and pay more attention to reviews with more helpfulness votes. Consumer decision-making facilitated by the helpfulness votes, therefore, can be skewed without considering when the review is posted and what the context is.

In order to take all the user reviews into consideration in evaluating their impact, previous studies employed various approaches to assess or predict the helpfulness of reviews without any helpfulness votes [10,16,21]. The assessment and predictions are usually based on the number of helpfulness votes in conjunction with other content characteristics of reviews that have received at least some helpfulness votes. The accuracy of the predictions relies heavily on the correct and precise evaluation of the helpfulness features of reviews that have received helpfulness votes. However, the definition and measurement of the helpfulness features of reviews are neither clear nor consistent in the existing literature [10,20].

In this paper, we approach this problem from a different perspective in order to avoid the ambiguous definition and measurement issues of helpfulness. Instead of predicting a helpfulness level for reviews that have no votes, we investigate the factors that determine the number of helpfulness votes a particular review receives (which includes both "yes" and "no" votes). Our objective is to understand why some reviews receive many helpfulness votes while others receive few or no votes at all. To the best of our knowledge, there is no prior study addressing this most basic yet critical question. However, our approach and prior studies are not mutually exclusive, and in fact, our approach complements previous studies. More understanding of what drives the helpfulness voting would help e-commerce websites improve the design of user review systems to encourage more helpfulness votes on online user reviews. Consequently, as more reviews receive more helpfulness votes, users would benefit from a larger collection of reviews and better aggregation of information, thus resulting in less bias derived from smaller numbers of votes. In addition, the more helpfulness votes a review receive, the more accurate the true helpfulness level of the review can be predicted with whatever methods employed in the previous studies. As a result, in this paper, we take the initiative to explore what characteristics of online user reviews influence the number of helpfulness votes.

Using the reviews collected from a well-known website, we empirically examined the effects of various characteristics, namely, basic, stylistic, and semantic characteristics[3] of online user reviews on the number of helpfulness votes that the reviews have received. Basic characteristics include information that can be easily observed and straightforward, such as the reviewer's rating of the product, the review's posting time, etc. Stylistic characteristics represent some key features of reviewers' writing styles. For example, one reviewer might prefer to write short sentences using simple words, while another tends to write long sentences using sophisticated words. Semantic characteristics refer to the meaning of the words in the review, that is, words may have different influences on readers' propensity to vote. For example, a review containing the words "it is a wise investment on this software" might attract more votes than a review saying "it is good software." In other words, a review with the words "wise investment" may have a higher likelihood of receiving votes than a review using the single word "good." However, it is hard to quantify semantic characteristics due to the large amount of text to be processed. To overcome this challenge, we extracted semantic characteristics from the review texts by employing a text mining methodology, and then compared and contrasted the effects of the basic, stylistic and semantic characteristics on the number of helpfulness votes the reviews received. Previous studies have documented mixed results on the impact of semantic characteristics in predicting helpfulness of online user reviews [10,16,20]. However, our findings show that semantic characteristics are more influential than the other characteristics in affecting how many helpfulness votes the reviews receive. This finding distinguishes our approach from previous studies on the effects of semantic characteristics.

This study contributes to the behavioral research by providing a new perspective in understanding online user voting behavior. Our findings show that the reviews with the most extreme opinions receive more helpfulness votes than those with mixed or neutral, indicating that people tend to pay more attention to extreme opinions. This finding also provides major implications to marketing practitioners in that "extremeness" of opinions could be used to attract customers' attention.

Our research expands the text mining research literature by employing text mining methodology in studying online user reviews. Online user reviews contain very rich textual information, but it is difficult to quantify the often ambiguous textual information from reviews, especially for large amounts of text [23]. Previous research has employed content analysis [15,17], but this methodology is extremely time-consuming when the amount of text is very large. As an alternative to this arduous method, text mining techniques have gained more and more attention from academic researchers. However, methods to quantify textual information and the real value of text mining techniques in this regard remain unclear [10,20]. Our study applied the text mining approach by extracting the semantic characteristics from reviews' texts, and further utilized these characteristics along with basic and stylistic characteristics in ordinal logistic regression models to examine what factors determine the number of helpfulness votes a review receives. Our findings suggest that the semantic feature of reviews has a significant impact on the number of helpfulness votes the reviews receive. This finding demonstrates the importance of employing more viable text mining techniques in uncovering the information content and exploring the influence of online user reviews.

The rest of the paper proceeds as follows. In the next section, we review the related studies. We then describe the data, develop our research methodology, and present the empirical result. In the last section, we discuss the results and implications, and conclude the paper by discussing the limitations and identifying areas for future research.

2. Literature review

2.1. Helpfulness of online user reviews

Extant studies have documented various definitions and measurements of helpfulness of online user reviews. Many websites simply ask users a question "Is this review helpful to you?" The answer can then be merely "yes" or "no." Users' answers to this question are typically summarized in the form of a proportion, for example, "20 out of 100 people think this review helpful," which is normally referred to as the helpfulness of review by previous studies [10,11,16,21]. Although this paper focuses on examining whether users would submit their votes to this question or not, it is important to review and

[3] These characteristics have been extensively used in the previous studies [10,16,20].

understand the definition and measurement of helpfulness in previous research.

Several approaches have been suggested to measure, model, and predict helpfulness. Helpfulness is defined as the percentage of the helpfulness votes, which is the number of helpful votes divided by the total number of votes [16]. For example, for a review with "20 out of 100 people think this review helpful," the helpfulness is quantified as 0.2. Liu et al. [20] manually coded the helpfulness of reviews as good, fair, and bad according to informativeness, readability and subjectivity of reviews. Another study proposed a binary measure of helpfulness by transforming the raw percentage of voting to 0 or 1, for "unhelpful" and "helpful", respectively, based on whether the raw percentage exceeds a benchmark cutoff value [10]. In this study, the final definition of helpfulness is simply a dummy variable indicating whether the review is helpful to the reader or not. All these existing definitions and measurements of helpfulness of online user reviews essentially aim to estimate to what extent reviews are helpful for consumer decision-making. Except for the definition in [20], other definitions require viewers' voting on whether they feel a specific review is helpful or not. Nevertheless, a prominent phenomenon observed in most e-commerce websites is that not all reviews receive helpfulness votes. In fact, a large percentage of online reviews in many websites have never received a single vote in regards to their helpfulness. Kim et al. [16] found that 38% of the 20,919 reviews for all MP3 player products on Amazon.com received three or fewer helpfulness votes. Our data contain about 3500 reviews at CNETD, and 51% of them have never received a single vote on helpfulness.

It is then legitimate to ask the question: "Are those reviews without any helpfulness votes in fact unhelpful?" Besides the content and quality of reviews, there are many other reasons that a specific review does not receive any helpfulness vote. For example, the review could be posted recently at the time researchers were collecting the data, and therefore few users had the opportunity to read and evaluate the review. Some reviews may simply be too long to hold the attention of users, and as a result, they do not vote on helpfulness. Other characteristics of reviews, such as format and tone, could also prevent them from receiving helpfulness votes from users. All the aforementioned factors, however, do not necessarily suggest that the reviews do not contain valuable information about the product even without any helpfulness votes.

A few recent studies attempted to address the question of how to ascertain the helpfulness of reviews that lacked votes by exploring the characteristics of reviews with many helpfulness votes using data mining techniques, and then employing predictive models to determine the helpfulness of reviews without helpfulness votes. Kim et al. [16] assessed the helpfulness of online reviews for two categories of products (MP3 players and digital cameras) using several review characteristics to rank the reviews in terms of their helpfulness. They looked into several characteristics of online reviews: the structural (the length of words, number of sentences, the average sentence length, etc.), lexical (term frequency and inverse document frequency), syntactic (the percentage of tokens that are nouns, the percentage of tokens that are verbs, etc.), semantic (positive and negative sentiment words), and meta-data features (the reviewer's rating on the product). They found that the three most useful characteristics to rank the reviews in terms of helpfulness are the length of the review, lexical features, and reviewer's rating, whereas other structural and syntactic features have no significant impact on the prediction of helpfulness.

Forman et al. [10] investigated the reviews for three types of products on Amazon.com (audio and video players, digital cameras, and DVDs). They examined three characteristics: reviewer information, readability of the review text, and subjectivity of the review text. They found that reviews with a mixture of subjective and objective elements are more helpful. They also found that the readability of reviews has a positive impact on perceived helpfulness, and spelling errors have a negative impact on helpfulness. Furthermore, their findings suggest that the reviewer information, the review subjectivity features, and the review readability features are "interchangeable" in predicting helpfulness of reviews. In other words, any of these three features can have very similar predicting power respectively.

Liu et al. [20] also examined reviews of digital cameras on Amazon.com and explored features such as readability, subjectivity, and informativeness (that is, how much information contained in the review, for example, the number of words, the number of sentences, the number of brand names, the number of product features, and etc.). Their purpose was to use these features to detect low-quality (unhelpful) reviews. Using a Support Vector Machines (SVM) classification model, they classified the quality of reviews and found that the model performs well.

Liu et al. [21] examined movie reviews on IMDB website and attempted to predict the helpfulness of reviews that did not receive any votes. They constructed a model to predict helpfulness using features such as reviewer expertise, writing style, and timeliness of the review. Their study differed from others in that they only included the reviews with at least 10 helpfulness votes in building the predictive model. Their predictive model was based on radial basis functions and was found to outperform a linear regression model.

To the best of our knowledge, extant studies on investigating the helpfulness of online reviews used only reviews with helpfulness votes in their data analyses, leaving out those reviews without votes so that their helpfulness level can be predicted. However, we noticed that no prior studies have addressed a basic yet important question: Why do some reviews receive many helpfulness votes while others receive a few or no votes? Are the underlying causes the same for all the reviews that did not receive any helpfulness vote? If the answer is no, we should not treat all the reviews without any helpfulness votes the same in predicting their helpfulness. Instead, more understanding of why they did not receive helpfulness votes will help discriminate between the helpful and unhelpful reviews and allow better predictive models to be constructed. More importantly, more understanding of why different reviews receive various number of helpfulness votes would help us improve the design of the online user feedback systems, in which significantly more helpfulness votes will be cast for as many reviews as possible, thus providing a better idea the true helpfulness of online user reviews. Therefore, in this paper, we take the initiative to investigate why some reviews do not receive helpfulness votes by exploring characteristics of online user reviews and their effects on influencing helpfulness votes.

2.2. Text analysis of online user review

While earlier studies primarily addressed the relationship between online user review valence/volume and product sales [4,9], there is an emerging research area that pays more attention to the detailed text information generated in reviews. Content analysis was utilized to quantify the feedback text comments on eBay [23]. The findings suggest that the rich content of feedback text comments plays an important role in building a buyer's trust in a seller. More recently, text mining is gaining popularity in IS research and various text mining techniques have been developed to quantify textual information [19,26,29]. Recently, for example, latent semantic analysis was used to discover the "intellectual core" of IS research [26]. In that research, five core research areas in IS discipline were identified, by analyzing abstracts of all research papers published from 1985 to 2006 in three top IS research journals (that is, *MIS Quarterly*, *Information Systems Research*, and *Journal of Management Information Systems*) [26]. Wei et al. [29] also used latent semantic indexing to cluster similar organizational documents in multilingual forms in order to better manage knowledge in organizations.

The helpfulness of reviews, by and large, is closely related to the detailed text information contained in the reviews, that is, whether the information itself is helpful for the viewers to make purchase decisions. Though many extant studies, as discussed above, simply used the ratio of helpfulness votes to represent the information, a few studies have realized the importance of delving deeper into detailed text analysis [10,16,20,21]. For example, users' opinions are often extracted from online reviews to predict product sales [3]. Although various characteristics of reviews were explored in the previous research, the impact of a number of textual characteristics, such as structural, syntactic, and semantic characteristics, on viewers' assessments of helpfulness is yet to be examined due to the difficulty and complexity of the text analysis. Additionally, there is no consistent evidence of their distinct influence on the helpfulness of reviews. In particular, there are mixed results on the influence of semantic characteristics. Therefore, it is imperative to further explore effects of the textual characteristics in online user reviews in the study of helpfulness.

Our paper differs significantly from previous studies in that we explore the effects of various characteristics of online user reviews on the amount of helpfulness votes reviews receive. Another key aspect of our study that differs significantly from prior research is that we also include reviews without helpfulness votes in our data analysis. Furthermore, we incorporate in our model both textual and non-textual characteristics of reviews. Finally, we also compare effects of different combinations of those characteristics on helpfulness votes.

3. Research methods

3.1. Data collection

Data for this research were collected from CNET Download.com (CNETD: http://www.download.com), which is a leading and representative online platform of the software market. CNETD, a part of CNET network, is a library of more than 50,000 free or free-to-try software programs for Windows, Mac, mobile devices, and Webware. CNETD evaluates and categorizes software programs into different groups to facilitate user search. In addition to providing a detailed description of product features for each software program, CNETD updates the cumulative number of downloads on a daily basis. Furthermore, the most recent week's download number is also displayed for each software product. CNETD provides an ideal environment for this study since none of the parties (users, CNETD, and software owners) would benefit directly from the increase of the software download, hence there is no or low incentive for any of them to manipulate the user reviews as a disguised "promotional chat" [22].

CNETD offers a widely accepted user feedback system for online users to share their opinions and experiences. The user review system includes detailed comments and an overall evaluation indicated by a five-star user rating system. A particular advantage of CNETD's online feedback system is that the website displays the entire history of all user reviews posted for a particular listing, which offers a unique opportunity for this study to examine the dynamics of online user reviews.

We collected the entire history of review data up to May 2009 for software programs in one of the largest groups of Windows software (Enterprise Computing), which includes a wide variety of various categories (please refer to Table 1 for more details on the categories included). For each software program, we collected all the reviews posted, with each record consisting of: reviewer's user ID, review post time, the title of the review, pros, cons, summary, and the number of helpfulness votes received out of the total number of votes. In addition, we also collected the average user rating and the total number of ratings (many users only submit a rating, but not detailed reviews) for each software program. Our sample consists of 87 software programs, which belong to 28 unique categories that are large enough to provide a diversified coverage of various software programs. The total number of user reviews for the 87 software programs are 3460. Table 1 lists the number of software programs and reviews in each category. Table 2 presents the distribution of the number of reviews received by the software programs.

Table 1
Number of software programs and reviews by category.

Category	Number of software	Number of reviews	Reviews percentage (%)
Antivirus software	4	772	22.31
Auction tools	1*	28	0.81
Automation software	1	6	0.17
BIOS and system updates	2	41	1.18
Backup software	4	69	1.99
Bookmark managers	1	12	0.35
CD-ROM	1	98	2.83
Collaboration tools	1	8	0.23
Diagnostic software	8	912	26.36
Display and video	1	8	0.23
Download managers	1	1	0.03
E-commerce	1	15	0.43
Encryption software	11	267	7.72
FTP software	4	45	1.30
Fax-modems and ISDN	2	18	0.52
File compression	3	127	3.67
File management	8	243	7.02
File sharing	1	7	0.20
Marketing tools	2	27	0.78
Miscellaneous	4	61	1.76
Network	6	109	3.15
Online form tools	1	10	0.29
Printers	1	45	1.30
Search tools	2	30	0.87
Software removal	2	32	0.92
Sound and multimedia	1	83	2.40
System utilities	12	370	10.69
Tools and editors	1	16	0.46
Total	87	3460	100.00

*Note: There is only one software program in some categories in our randomly-selected sample, however, there are more than one software programs in these categories at CNET.

As mentioned earlier, a large number of reviews in our sample did not receive a single helpfulness vote. Table 3 shows the distribution of reviews by the number of helpfulness votes. Note that one category contains reviews that received seven or more helpfulness votes. This is because as the number of helpfulness votes for a review increases, the number of such reviews decreases sharply. For example, there are nine reviews that received ten helpfulness votes, whereas only three reviews received 15 votes, and there are very few reviews with more. In analyzing the number of votes that reviews received, we combined the number of reviews that received seven votes and above into one group representing the reviews with "many" helpfulness votes, in order to ensure enough observations.

3.2. User reviews at CNETD

Fig. 1 shows a typical user review at CNETD. The average user star rating of the software program is shown in the upper left corner of Fig. 1. The individual user's review is presented in the following part in Fig. 1. It begins with an individual reviewer's star rating and is followed by version, title, reviewer's name, posting date, pros, cons, and

Table 2
Number of software programs by number of reviews.

Number of reviews (n)	Number of software	Percentage (%)
$n \le 10$	23	26.4
$10 < n \le 30$	42	48.3
$30 < n \le 100$	16	18.4
$n > 100$	6	6.9
Total	87	100.0

Table 3
Reviews by number of helpfulness votes.

Number of helpfulness votes	Number of reviews	Percentage (%)
0	1785	51.59
1	700	20.23
2	363	10.49
3	215	6.21
4	93	2.69
5	57	1.65
6	42	1.21
>=7	205	5.93
Total	3460	100.00

summary of the review. The number of helpfulness votes for the review is listed on the upper right part of Fig. 1 (circled).

We categorize the available information from the reviews into three types of characteristics, namely, basic, stylistic, and semantic characteristics, which have been extensively used in previous explorative studies [10,16,20,21] in which researchers integrate all information available in user reviews attempting to discover interesting patterns and trends. In our study, we aim to examine whether those three types of characteristics have any impact on the number of helpfulness votes that reviews receive and are especially interested to see if semantic characteristics are influential in encouraging helpfulness votes.

3.2.1. Basic characteristics of review

The first type of information is what we can directly observe from a review (e.g., the information shown in Fig. 1), including (1) whether the reviewer wrote about "pros," (2) whether the reviewer wrote about "cons", (3) whether the reviewer wrote anything in the summary, (4) how many days since the posting date, and (5) the "extremeness" level of the review, which can be roughly estimated as the absolute value of the difference between the reviewers' rating and the average of all user ratings.[4] We categorize this type of review information as "basic".

3.2.2. Stylistic characteristics of review

The second type of information is the stylistic characteristics of a review, which represent key features of reviewers' writing style that cannot be easily derived by simply browsing the review texts. In particular, we examine such review characteristics as the average number of words in a sentence, the average number of characters per word, etc. Table 4 lists the characteristics we have examined.

3.2.3. Semantic characteristics of review

The third type of information in a review is semantic characteristics, which are related to the substance of the review. We argue that the substance of reviews plays a major role in viewers' decision to vote or not. For example, if a review does not make any sense (that is, it is meaningless or irrelevant to the software reviewed), viewers may not take the time to vote because they see the review as simply not worth extra time required to do so.

However, examining the exact meaning of text is extremely difficult and often subjective. As such, we turn to a more practical way to parse the meaning of reviews with the help of Latent Semantic Analysis (LSA). LSA is a statistical approach to analyzing relationships between a set of documents and terms in these documents that produces a set of meaningful patterns related to the documents and terms [6]. "*LSA represents the words used in it, and any set of these words – such as a sentence, paragraph, or essay – either taken from the original corpus or new, as points in a very high dimensional semantic space*" [18, p.262]. For example, a concept produced by LSA may represent features of some of documents. In our context, a concept like "professionally written" may be a key feature of reviews with many helpfulness votes. LSA has been employed in various studies in IS field [19,26,29]. Our study, however, takes a different approach from previous work. Rather than trying to identify what semantic characteristic(s) causes viewers to vote on the helpfulness of reviews, we examine the more fundamental question of whether semantic characteristics as a whole have any impact on the number of helpfulness votes that reviews receive. The semantic characteristics of the reviews are analyzed by LSA, and more detailed information about LSA and the text mining techniques employed to perform LSA are discussed in the next section.

3.3. Text mining methodology

The LSA-based text mining methodology employed in this paper is shown in Fig. 2. The left side of Fig. 2 shows steps of text mining process, while the right side illustrates the outcome of each step. This process includes text preprocessing, parsing, term reduction, singular value decomposition (SVD), and factor analysis. Details of each step are explained in the following sections.

3.3.1. Text preprocessing

In this step, we calculate the number of words, number of sentences, words per sentence, number of characters per word, and number of words of different length for the title, pros, cons, and summary of each review. This preprocessing provides the stylistic characteristics of the review. In order to obtain an overall view of the review, we combined the title, pros, cons and summary of each review into one text block. In other words, we treat title, pros, cons and summary as a whole text item rather than four separate parts.

3.3.2. Parsing

In the parsing step, we used SAS® Enterprise Text Miner to perform stemming, part-of-speech tagging and term identification. The purpose of stemming is to treat words such as "take, taking, took, taken, takes" as one term because they come from a common root word. Part-of-speech tagging identifies the part-of-speech of each term and classifies the term as a noun, verb, adjective, or other part-of-speech. This step is important because each part of speech has a different function in conveying the meaning of a sentence. Term identification is used to regard a word in the text as one term after stemming and part-of-speech tagging. Homonyms, words spelled the same but belonging to different parts-of-speech, are counted as multiple terms; multiple words with the same root, however, are counted as only one term. For example, "drive" can be either a verb or a noun. If both usages occur in the same text, they are treated as two different terms rather than just one. In other words, if the verb "drive" and the noun "drive" occurred in one review, they are to be identified as two terms. The purpose of term identification is to construct a so-called "term-by-frequency matrix" with each row referring to each review text and each column representing each term. We illustrate the concept in Table 5, wherein each cell of the matrix is a term frequency, that is, the number of times that a term (column) appears in a particular review (row). For example, the term "windows" appears in review 1 twice. However, one of the problems of this matrix is that when there are too many reviews and too many terms, dimensions of the matrix become extremely large, which makes it extremely difficult to conduct computations using this matrix. In our

[4] One reviewer suggests to separate extremeness of user reviews into negative extremeness and positive extremeness and test their effects on the number of helpfulness votes in two situations: when reviews have relatively low ratings (e.g., below 2.5/5) and when reviews have relatively high ratings (e.g., above 2.5/5). Our test result shows that both negative and positive extremeness have positive effects on the number of votes in both situations except that positive extremeness has no statistically significant effect on the number of votes when reviews have relatively low ratings.

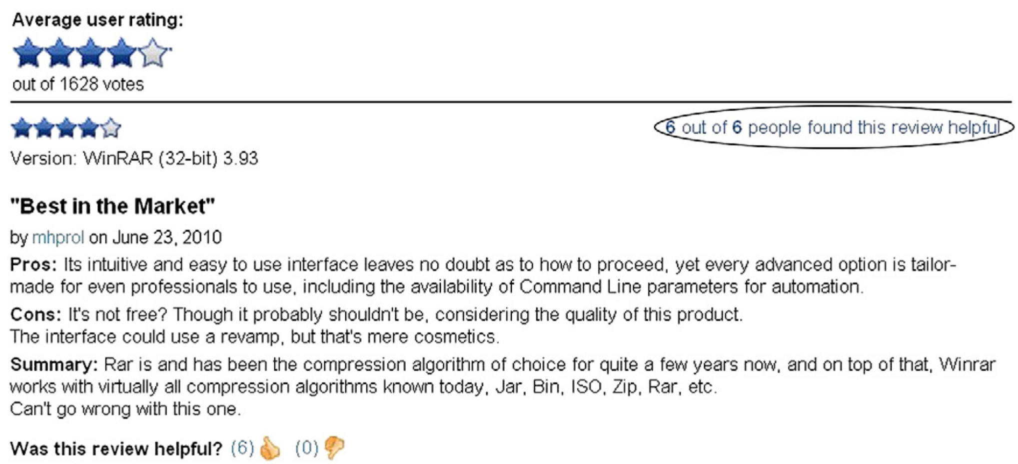


Fig. 1. Screen shot of online user review at www.download.com (CNETD).

data set, 16,168 terms were indentified in 3460 reviews, resulting in a 3460 × 16,168 matrix. Hence, another procedure is required in LSA text mining procedure to reduce the number of terms, that is, the number of columns in the matrix; the number of rows will remain the same, as we want to know the difference between reviews.

3.3.3. Term reduction

One of the objectives of LSA is to discriminate one text from another in a semantic sense. In our study, first, we try to discriminate reviews with many helpfulness votes from those with none or very few. Since relatively meaningless words such as "a, an, the" in the reviews are not useful in discrimination, we compiled a list of the meaningless words (generally called "stop words") and eliminated them from the Term-by-Frequency Matrix, which reduced the number of columns.

Second, we deal with synonyms in the text. Since synonyms by definition convey equivalent meanings, they are not very useful in discriminating reviews. We compiled a synonyms dictionary to treat synonyms as one single term. As a result, many synonyms are consolidated into single terms in our analysis, resulting in fewer columns in the matrix. A considerable number of terms were removed from the matrix after the term reduction procedure. Originally there were 16,168 terms in our dataset and after term reduction, only 3457 remained.

However, only using term frequency cannot discriminate reviews effectively. One term that appears very frequently in one review may also appear so in other reviews. The most commonly used terms may appear frequently in almost all review texts, and thus are not useful in distinguishing one review from others. Conversely, the less frequent and unique terms are more useful. For example, in reviews of software programs, the term "computer" may appear in many reviews, including reviews that receive many helpfulness votes and reviews that do not receive any votes, and thus it is not very useful to use the term "computer" to discriminate reviews. On the other hand, the word "donno" (a misspelling of "don't know") only appears in reviews that receive no helpfulness votes, therefore, it is useful in distinguishing these reviews from those that received many votes. In order to solve this problem, the term frequencies are transformed by TF-IDF (Term Frequency-Inverse Document Frequency, here, document refers to review) weighting [25]. TF-IDF weighting is used to place less weight on more frequent terms and more weight on less frequent terms. Eq. (1) is the standard formula of TF-IDF weighting.

$$w_{ij} = tf_{ij} * idf_i \tag{1}$$

where w is the weighted frequency, tf_{ij} is the term frequency of term i in review j, $idf_i = \log_2(N/n_i) + 1$, N is the number of reviews, and n_i is

Table 4
Stylistic Characteristics.

Description
Number of words in review
Number of sentences in the review
Average characters per word
Average words per sentence
Number of words in pros
Number of words in cons
Number of words in summary
Number of words in title
Number of 1-letter words in the review
Number of 2 to 9-letter words in the review, respectively
Number of 10 or more-letter words in the review

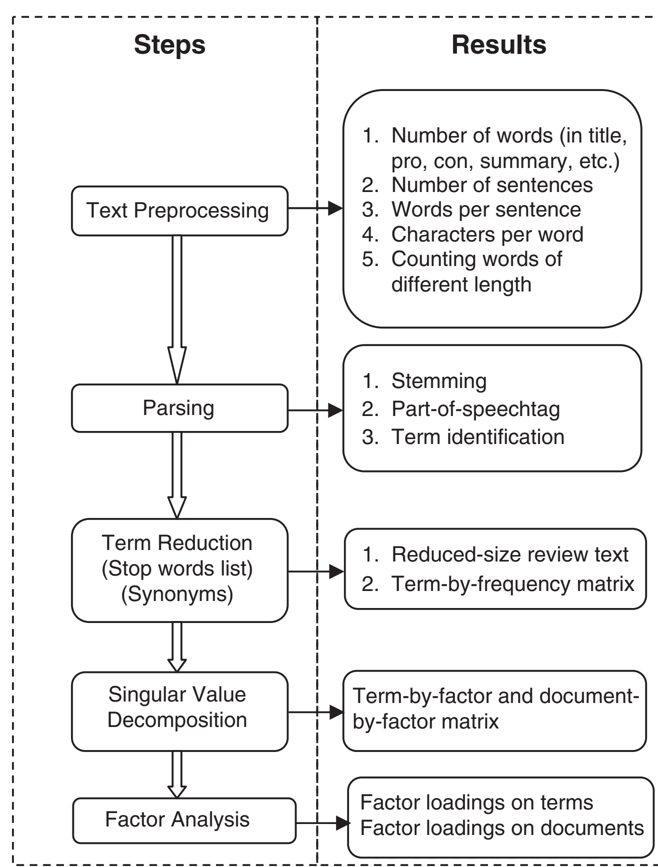


Fig. 2. Text mining methodology.

Table 5
An example of term-by-frequency matrix.

	Windows	Bug	…	Open source
Review 1	2	1		7
Review 2	7	0	…	5
…	…	…	…	…
Review 3460	0	5	…	0

the frequency of term i in all reviews [12,25]. When a term appears frequently across reviews, the inverse document frequency, *idf*, is small, resulting in a smaller weight, and vice versa.

3.3.4. Singular value decomposition (SVD)

Although the term reduction process reduced a large number of terms, there were still too many terms remaining (3457 terms in our case). Singular value decomposition (SVD) was then implemented to reduce the dimensionality (column, that is, terms) of the transformed term-by-frequency matrix. In fact, SVD sums up terms into groups and therefore further reduces columns of the matrix. With SVD, a matrix can be decomposed into the product of three matrices. "*One matrix describes the original row entities as vectors of derived orthogonal factor values, another describes the original column entities in the same way, and the third is a diagonal matrix containing scaling values such that when the three matrices are multiplied, the original matrix is reconstructed*" [18, p. 263]. Limited by the space and scope of this paper, we will not discuss the details of SVD; more information regarding SVD can be found in [18].

3.3.5. Factor analysis

Following the SVD process, we perform the factor analysis, where terms are grouped into factors and are given appropriate loadings. In this study, terms with similar mathematical properties in the term-by-frequency matrix are summed, forming an individual factor (referred to as an "SVD factor" hereinafter). Each SVD factor represents a summarization of words in reviews with similar properties in a higher dimension, which is distinct from the other SVD factors. In general, the number of factors is often chosen subjectively by researchers based on features of data [26]. The number of factors represents the number of semantic characteristics from the reviews. As the purpose of this study is to examine whether semantic characteristics as a whole affect the number of helpfulness votes that reviews receive, we tried several numbers of factors, namely, 50, 100, 150, and 200 for exploratory and demonstration purposes.

For demonstration purposes, we report the 100-factor solution in this paper. We have the document-by-factor matrix with each row referring to each document and each column referring to each one of the 100 factors. In other words, for each review we had 100 SVD factor loadings. The 100 SVD factors summarize semantic structure that links terms with reviews in which they appear. The 100 SVD factor loadings for each review are what we mean by "semantic characteristics" of reviews in our research context. To put it differently, each review now has 100 new variables to describe its semantic characteristics, and these 100 variables will be used in the subsequent analyses.

3.4. Empirical models

Using ordinal logistic regression (OLR) models, we investigate relationships between three types of characteristics of online reviews and the number of helpfulness votes that those reviews receive. OLR is an extension of the binary logistic regression model where the dependent variable can accommodate more than two categories. OLR uses cumulative logits with ordinal dependent variables. The dependent variable in our study is not whether or not a review receives helpfulness vote(s) as in the binomial logistic regression, but a "helpfulness rank" based on the number of votes a review receives. An OLR model is particularly suitable for our study because we are interested in not only whether a review receives at least one vote or not ("yes" vs. "no" question, which is binary), but also what review characteristics lead to more helpfulness votes (the effect of characteristics is cumulative as the number of votes the review receives is increases). Eqs. (2) and (3) depict the basics of our OLR approach.

$$g(\Pr(Y \le i|x)) = \alpha_i + \beta' x, \quad i = 1, 2, \ldots, k \tag{2}$$

where Y is the dependent variable, the ranks are denoted by 1, 2 … k, α_1, α_2,…,α_k are k intercept parameters, β is the vector of slope parameters, β' is the transpose of β, and x is the vector of independent variables. The function $g = g(\mu)$ is called the link function that allows the (μ) assumed response to be linearly related to the independent variables. $\Pr(Y \le i|x)$ is the probability that Y is smaller or equal to i, conditioning on x. The log-odds scale has been used as the link function as in the form of Eq. (3) [5,28].

$$\begin{aligned} g_i(\Pr(Y \le i|x)) &= \ln \frac{P_r(Y \le i|x)}{P_r(Y > i|x)} = \ln \frac{P_r(Y \le i|x)}{1 - P_r(Y \le i|x)} \\ &= \ln \frac{\varnothing_1(x) + \varnothing_2(x) + \cdots + \varnothing_i(x)}{1 - (\varnothing_1(x) + \varnothing_2(x) + \cdots + \varnothing_i(x))} = \alpha_i + \beta' x, \end{aligned} \tag{3}$$

$$i = 1, 2, \ldots, k$$

where $\varnothing_i(x)$ is the probability of being in class i given x.

In our study, Y is the number of votes on helpfulness. "0" to "6" denotes "0" to "6" number of votes respectively while "7" denotes "7 or more" votes. The distribution of the number of reviews with different number of votes was presented earlier in Table 3. As noted previously, we consolidated the reviews with 7 or more votes into one vote category to increase the power of the statistical analysis.

In order to examine what and how different characteristics of reviews influence the number of votes they receive, we construct five OLR models with various combinations of the three types of characteristics, basic, stylistic and semantic. We explore whether the effect of various single characteristics remains unchanged and whether the addition of semantic characteristics has any additional effect on the performance of the model. Table 6 presents the five model descriptions. Models 1, 2, and 3 employ only one of three characteristics respectively. Model 4 uses both basic and stylistic characteristics. In model 5, we add semantic characteristics to model 4. By comparing models 1, 2, and 3, we could see individual impact of three types of characteristics on the number of helpfulness vote. By comparing models 4 and 5, we investigate the marginal impact of additional characteristics on the number of helpfulness vote.

There are a large number of independent variables in the aforementioned five models. In particular, there are 100 SVD factor variables in models 3 and 5, which have much more variables than models 1, 2 and 4. Adding more variables in the model may increase the fit of the model by giving a smaller residual sum of squares; however it may also decrease the model's predictive power. In order to make the models more parsimonious and comparable, we employ the stepwise variable selection method. The stepwise selection method is a model selection process in which a new variable is added into the model in each step, and then a test is conducted to check if any variables can be deleted without appreciably increasing the residual sum of squares (RSS) [13]. Thus, the final model should have a reasonable fit with fewer variables. For instance, we have only 18 independent variables in model 5 in Table 9.

3.5. Model comparison criteria

To compare the models, we choose three widely used fit indices: misclassification rate, Akaike's Information criterion (AIC), and lift ratio. Table 7 shows the mathematical formulas of these three criteria.

Misclassification rate is often used to see how inaccurate the classification is. It calculates the proportion of wrong classifications in total classifications. The larger the misclassification rate, the less accurate is the classification, and the poorer performance of the model is.

Akaike's Information criterion (AIC) is a measure of goodness-of-fit proposed by Akaike [1]. It describes the tradeoff between bias and variance in model construction, that is, between complexity and precision of the model. It penalizes the addition of more variables into the model while maintaining reasonable precision. The smaller the AIC, the better the model is. We therefore use AIC index to compare the models.

The lift ratio is a widely used model accuracy measure in data mining literature. It is a measure of the performance of a model in segmenting the population. The lift ratio of a subset of the population is the ratio of the predicted response rate for that subset to the predicted response rate for the population. It measures the performance of an estimated model as compared to random selection from the population. As a rule of thumb, the larger the lift ratio, the better the performance of the model will be.

Table 7
Model comparison criteria.

Criteria	Formula
Misclassification Rate	$MR = \frac{\text{number of incorrectly classified cases}}{\text{total number of cases}}$
AIC	$AIC = 2k - 2\ln(L)$ *where k is the number of* parameters in the model, and L is the maximized value of the likelihood function for the estimated model.
Lift Ratio	$Lift = \frac{P\{C_i \mid Sample\}}{P\{C_i \mid Population\}}$ where $P\{C_i \mid Sample\}$ is the portion of observations contained in class C_i relative to the biased sample population.

4. Results and discussion

Table 8 summarizes the empirical results of comparing the five OLR models. Model 5, which combines all three characteristics of reviews, has the lowest misclassification rate and AIC index with the highest lift ratio; therefore it has the best performance among all models. The results indicate that integrating semantic characteristics into the model along with basic and stylistic characteristics significantly enhances the performance of the model. Such a finding suggests that semantic characteristics play a very important role in influencing the number of helpfulness votes a review receives. Model 4, which uses only basic and stylistic characteristics, ranks the second in two performance criteria (misclassification rate and AIC). Model 2, which uses only stylistic characteristics performs the worst in two criteria (Misclassification rate and lift ratio), with the third criterion (AIC) just a little better than Model 3. These findings indicate that stylistic characteristics are the least critical criterion to encourage helpfulness votes from users.

As discussed previously, model 5, which contains all three types of characteristics, has the best fit. In order to examine exactly what characteristics have the key impact on the number of helpfulness votes, we investigate model 5 further to include the parameter estimates, standard error, Wald test statistics for the parameters and corresponding P values. Table 9 shows results of the investigation. After applying stepwise OLR, only 18 variables are included in model 5, including one control variable (whether the software is free or free-to-try), three basic characteristic variables (days since posting, difference between reviewer rating and average rating, and whether summary has content or not), three stylistic characteristic variable (number of 4-letter words, number of words in cons, and number of words in title), and eleven semantic variables.

Eight findings are summarized below:

First, we find that the P values for the estimates of control, basic, and stylistic variables are all smaller than 0.01, indicating they are statistically significant at 99% confidence level. The P values for six of eleven semantic characteristics are also smaller than 0.01, demonstrating statistical significance at 99% confidence level. The P values for the other five semantic variables are smaller than 0.05, illustrating statistical significance at 95% confidence level.

Second, we note that the estimate for “number of 4-letter words” is positive, indicating that with more 4-letter words, the review is likely to receive more helpfulness votes. The rationale lies in the fact that 4-letter words are easier to read and understand than longer words, and as such it appears that users prefer easy-to-read reviews and tend to vote on them more often.

Third, it is also interesting to find that the estimate for “number of words in cons” is positive, indicating that the more words in the cons part of the review, the more helpfulness votes the review is likely to receive. Online purchase is a risky venture for most consumers; hence they tend to pay more attention to the negative part of reviews on the product as shown in “cons” part [26]. As a result, more words in “cons” part of the review may encourage more people to read it and then vote on it. This finding is consistent with widely known “negativity bias” effect in psychology, which states that there is a generally bias in humans to give greater weight on negative entities [24].

Fourth, we find that the estimate for “number of words in title” is negative, demonstrating that the more words in the title part of the review, the fewer votes it is likely to receive. A possible explanation would be that too much information contained in the title may discourage people from reading the entire review before voting on it.

Fifth, it is also worth noting that the estimate for “the difference between reviewer rating and the average rating” is positive, which

Table 6
Model descriptions.

Model	Description of explanatory variables (x)	# of independent variables
Model 1	Use only basic characteristics of the reviews (Basic)	5
Model 2	Use only stylistic characteristics of the reviews (Style)	19
Model 3	Use only semantic characteristics of the reviews (SVD factor loadings)	100
Model 4	Use basic characteristics + stylistic characteristics	24
Model 5	Use basic characteristics + stylistic characteristics + SVD factor loadings	124

Table 8
Fit statistics for model comparison.

Model	Misclassification rate	Akaike's information criterion	Lift ratio
Model 1: Basic characteristics (Basic)	0.48301	9863.15	5.92
Model 2: Style characteristics (Style)	0.48589	9920.73	4.32
Model 3: SVD factors (SVD)	0.48560	9990.31	4.64
Model 4: Basic + Style	0.48243	9792.87	5.76
Model 5: Basic + Style + SVD	*0.47753*	*9599.26*	*7.36*

Table 9
Parameter Estimates of OLR Model 5 (Basic + Style + SVD).

Parameter	Estimate	Standard error	Wald Chi-square[a]	Pr > ChiSq
Intercept 7	−2.0271	0.2364	73.53	<.0001
Intercept 6	−1.4494	0.2302	39.64	<.0001
Intercept 5	−0.9653	0.2275	18.00	<.0001
Intercept 4	−0.6245	0.2264	7.61	0.0058
Intercept 3	−0.0302	0.2254	0.02	0.8936
Intercept 2	0.6781	0.2256	9.03	0.0026
Intercept 1	1.7121	0.2273	56.71	<.0001
Number of 4-letter words[b]	0.5575	0.0510	119.46	<.0001
Days since posting	−0.3595	0.0324	122.90	<.0001
Number of words in cons	0.1042	0.0300	12.05	0.0005
Number of words in title	−0.1814	0.0634	8.19	0.0042
Difference between reviewer rating and average rating	0.3347	0.0572	34.25	<.0001
Software is free	−0.2899	0.0386	56.26	<.0001
No summary content	0.5634	0.0647	75.81	<.0001
_SVD_1[c]	−3.7324	0.2728	187.16	<.0001
_SVD_17	−1.2906	0.3493	13.65	0.0002
_SVD_24	−0.9857	0.3179	9.61	0.0019
_SVD_25	−1.0768	0.3782	8.11	0.0044
_SVD_45	0.9274	0.3663	6.41	0.0113
_SVD_47	−0.8456	0.3472	5.93	0.0149
_SVD_6	0.8156	0.2550	10.23	0.0014
_SVD_61	−0.7684	0.3634	4.47	0.0345
_SVD_67	−1.1182	0.3878	8.32	0.0039
_SVD_71	−0.6839	0.3190	4.60	0.0320
_SVD_73	0.7856	0.3280	5.74	0.0166

[a] The Wald test is used to test the statistical significance of each coefficient estimate.
[b] The Italic parameters are log transformed.
[c] The "_SVD_" is the prefix of SVD factor loadings. "_SVD_1" refers to the loading on the first SVD factor, and "_SVD_17" refers to the loading on the 17th SVD factor, and so on. 11 out of 100 SVD factors are found to be significantly related to the number of helpfulness votes.

indicates that the larger the difference between reviewers' rating and the average rating on the review, the more votes the review is likely to receive. This difference represents the extremeness of the review. A review that is drastically different from the average reviews (more extreme) is more likely to stand out and attract significantly more attention from the users. We do not separate extreme positive and extreme negative reviews in this study, since our results suggest that the "extremeness" of the reviews plays an important role in influencing the helpfulness votes, regardless of positive or negative.

Sixth, it is unusual for us to find that the estimate for "days since posting" is negative, indicating that the longer the review has been posted, the fewer votes it is likely to receive. One of the possible explanations is that users prefer the more recent reviews in that the reviewers of the more recent reviews could integrate information in earlier reviews and tend to be more accurate.

Seventh, the results also show that the estimate for "whether the software is free" is negative, indicating that if the review is about free software, it appears to receive fewer votes. In other words, if the review is about free-to-try (one needs to pay to continue to use it after a trial period of time) software, it seems to receive more votes. Although some software programs are free-to-try and with no commitment to buy, some reviewers will eventually purchase them if they like the program after the free trial. Hence, there is a financial "stake" in their reviews which in turn, are taken more seriously by viewers. Therefore, these reviews appear to receive more helpfulness votes. One of potential problems is that the software owners, especially for free-to-try software, may have the incentive to manipulate user reviews. However, we believe this should not be a concern in our study. First, we are not studying whether the reviews will result in more downloads, which may potentially lead to more purchase. Our focus is on why users vote on the helpfulness of a specific review. Even though a review maybe from a software owner, that wouldn't influence the incentive for users to vote on the helpfulness of that review. Second, assuming software owners would post very positive reviews for their product, which may lead to more downloads from users, but eventually whether to buy the product still depends on user experience of trying the software in the first place. Hence, more fraudulent reviews from same owners may quickly incur suspense from both users and CNET, which may result in very negative responses as happened in Amazon.com in their early stage of adopting a user review system. Finally, we have included a dummy variable to control whether the software is free or free-to-try, which shall tease out the potentially different impact of these two types. For using the free-to-try software, users have no obligations to purchase the software only if they want to do so. If they would like to purchase the software after downloading the trial (light) version, they have to go to the software owner's own site and transaction will not involve CNET. Therefore, in nature, these two types of software have no difference on CNET since the only decision users have to make on CNET website is to download or not.

Finally, the estimates for some of the SVD variables are positive while others are negative, indicating that certain words have positive impact on encouraging helpfulness votes while other words have negative impact. In our study, we do not attempt to explore the details of SVD factors and instead examine semantic characteristics as a whole with an emphasis on their effect on the number of helpfulness votes that reviews receive. The empirical results show that semantic characteristics do have a significant impact on the number of helpfulness votes, with certain words encouraging more votes while other words discouraging votes.

5. Conclusions

In this paper, we examine a previously ignored yet important research question concerning the online user reviews: Why do some reviews not receive any votes on their helpfulness, while other reviews receive many votes? The helpfulness voting mechanism works effectively only when online user reviews receive helpfulness votes. We address this question by investigating the impact of various characteristics of online user reviews on the number of helpfulness votes that reviews receive. We categorize characteristics of online reviews into three types, namely, basic, stylistic and semantic. Text mining techniques and ordinal logistic regression models are employed to investigate more than 3400 online reviews of 87 different software programs from CNET Download.com. A number of practical and research implications can be derived from this study.

This study is complementary to the previous work on the helpfulness of online user reviews. Helpfulness votes on user reviews help online users locate the most potentially helpful reviews more efficiently and effectively for their decisions. Previous studies made efforts to predict the helpfulness of reviews without any helpfulness votes for users to make informed decisions. This study approaches this problem from a different perspective to facilitate users' decision-making. Rather than predicting the helpfulness, this study examines the factors influencing the number of helpfulness votes reviews receive. These perspectives are important in helping online users get quality information efficiently. The major contribution of this study would be to get more understanding on what are most important factors to attract more helpfulness votes.

This study also has significant implications for website designers in that it can guide them in designing helpfulness voting mechanisms that may garner more helpfulness votes. Reviews with more helpfulness votes provide users more information about other people's opinions about the helpfulness of reviews. Our findings indicate that some design features encourage voting while others curtail viewers' voting intention. For example, semantic characteristics have the most impact on the number of helpfulness votes that reviews received. Websites could provide incentives to encourage reviewers to write more meaningful comments. In addition, websites could provide more ranking options (e.g. based on the extremeness of opinion) to rank the reviews instead of ranking the most recent reviews first.

Our findings may have important implications for other behavioral researchers by providing a new perspective on the online user voting behavior. In addition to confirm the widely known "negative bias" effect [24], our results show that the reviews with the most extreme opinions have a higher probability of getting more votes, suggesting that people tend to pay more attention to those extreme opinions. The effect of "extreme opinions" on other's attention has not been much discussed in previous information systems research and might be an interesting area for future research.

In addition, the text mining methodology employed in this study was found to be effective in extracting semantic characteristics from the review text. Our findings suggest that the semantic characteristics of reviews have more significant impact than other characteristics on the number of helpfulness votes that reviews receive.

This exploratory study has several limitations. Although we found that semantic characteristics play an important role in encouraging helpfulness votes, we limited the semantic characteristics to an aggregated level and did not delve into the specific semantic characteristics, and which characteristics encourage more helpfulness votes. In addition, when extracting semantic characteristics using latent semantic analysis, we arbitrarily chose four sizes of SVD factors and did not investigate each factor's effect; the rationale for this is that the present study is a "proof-of-concept" designed to demonstrate our research purpose. Another limitation of this study is that our analysis is based on online reviews of software applications and as a result, it might not be generalizable to reviews for other online products or services. However characteristics we used in this study are universal for online user reviews and as such the analytical methodology can be easily replicated and applied to other online products or services.

The aforementioned limitations also call for future research. First, it would be interesting to delve deeper to investigate each specific semantic characteristic of online reviews. Each SVD factor can be carefully examined to determine whether it conveys meaningful information, and whether reviews with many helpfulness votes are more associated with SVD factors with more meaningful information. This approach would allow us to discover *patterns* that potentially encourage more helpfulness votes. Second, it would also be interesting to examine and determine the number of SVD factors we should use in the analysis. As suggested by Sidorova et al. [26], it is important to test different numbers of SVD factors using multiple datasets, and then choose the numbers of factors that carries the greatest weight to investigate in the particular research context. Another interesting future work to extend the current study is to investigate the "extremeness" effect on online users' behavior. In this study, our finding suggests that reviews with extreme opinions may have significant impact on users' helpfulness votes. More investigation is needed in future work to investigate to what extent the valence (positive or negative) of reviews influences the users' perception, and to what extent such a impact varies over the product life cycle. In addition, the extremeness level of reviews could be further separated into positive and negative extremeness. More detailed investigation in future studies shall include data from other industries to take a closer look at the impact of positive and negative extremeness of reviews.

This study can also be extended in several other directions. We focus only on investigating user reviews in this study and we intentionally left out editorial staff reviews at CNET. CNET's editorial staff reviews a small number of the software programs, with an emphasis on popular software programs. The portion of software programs that has been reviewed by CNET editorial staff is less than 10%. The sample size will be reduced significantly if we only look at software programs that have CNET ratings. Considering the small percentage of software programs with CNET ratings and the randomness of our sample, we do not examine the influence of CNET rating in this study. Yet it is definitely interesting and important to study the difference of the impact of these two sources of reviews in future work. In addition, online user reviews posted for the same software may share common features. In other words, reviews posted for the same software are not independent. The more recent reviews tend to be more accurate and "mature" since they could "borrow" and "integrate" the contents and opinions that have been raised in the earlier reviews, thereby may attract more votes on helpfulness. It would be very interesting to explore this non-independency of online user reviews in future research.

Acknowledgment

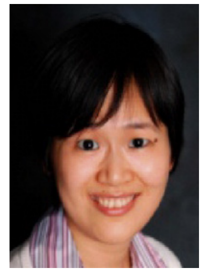
We thank Gerald Pinnes for his assistance in data collecting and processing. We also thank Dr. Donald Jones for his suggestions on earlier version of the manuscript.

References

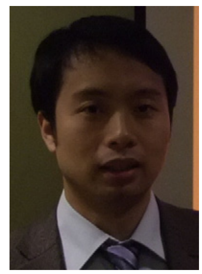
[1] H. Akaike, A new look at the statistical model identification, IEEE Transactions on Automatic Control 19 (6) (1974) 716–723.
[2] E. Brynjolfsson, M.D. Smith, Frictionless commerce? A comparison of internet and conventional retailers, Management Science 46 (4) (2000) 563–585.
[3] P.Y. Chen, S. Wu, J. Yoon, The impact of online recommendations and consumer feedback on sales, The 25th International Conference on Information Systems, Washington, DC, 2004, pp. 711–724.
[4] J.A. Chevalier, D. Mayzlin, The effect of word of mouth on sales: online book reviews, Journal of Marketing Research 43 (3) (2006) 345–354.
[5] D.R. Cox, E.J. Snell, Analysis of Binary Data, Chapman & Hall, London, 1989.
[6] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, R. Harshman, Indexing by latent semantic analysis, Journal of the American Society for Information Science 41 (6) (1990) 391–407.
[7] C. Dellarocas, The digitization of word of mouth: promise and challenges of online feedback mechanisms, Management Science 49 (10) (2003) 1407–1424.
[8] W. Duan, B. Gu, A.B. Whinston, Do online reviews matter? — An empirical investigation of panel data, Decision Support Systems 45 (4) (2008) 1007–1016.
[9] W. Duan, B. Gu, A.B. Whinston, The Dynamics of Online Word-of-Mouth and Product Sales — An Empirical Investigation of the Movie Industry, Journal of Retailing 84 (2) (2008) 233–242.
[10] C. Forman, A. Ghose, B. Wiesenfeld, Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets, Information Systems Research 19 (3) (2008) 291–313.
[11] A. Ghose, P.G. Ipeirotis, Estimating the Socio-Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics(November 26, 2008). Available at SSRN, http://ssrn.com/abstract=1261751.
[12] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, San Francisco, 2006.
[13] R.R. Hocking, The analysis and selection of variables in linear regression, Biometrics 32 (1976) 1–49.
[14] Q. Jones, G. Ravid, S. Rafaeli, Information overload and the message dynamics of online interaction spaces: a theoretical model and empirical exploration, Information Systems Research 15 (2) (2004) 194–210.
[15] H.H. Kassarjian, Content analysis in consumer research, Journal of Consumer Research 4 (1) (1977) 8.
[16] S.M. Kim, P. Pantel, T. Chklovski, M. Pennacchiotti, Automatically assessing review helpfulness, Conference on Empirical Methods in Natural Language Processing, Morristown, NJ, 2006, pp. 423–430.
[17] R.H. Kolbe, M.S. Burnett, Content-analysis research: an examination of applications with directives for improving research reliability and objectivity, Journal of Consumer Research (1991) 243–250.
[18] T.K. Landauer, P.W. Foltz, D. Laham, An introduction to latent semantic analysis, Discourse Processes 25 (1998) 259–284.
[19] K.R. Larsen, D.E. Monarchi, D.S. Hovorka, C.N. Bailey, Analyzing unstructured text data: using latent categorization to identify intellectual communities in information systems, Decision Support Systems 45 (4) (2008) 884–896.
[20] J. Liu, Y. Cao, C.Y. Lin, Y. Huang, M. Zhou, Low-quality product review detection in opinion summarization, Joint Conference on Empirical Methods in Natural

Language Processing and Computational Natural Language Learning, 2007, pp. 334–342.
[21] Y. Liu, X. Huang, A. An, X. Yu, F. Giannotti, D. Gunopulos, F. Turini, C. Zaniolo, N. Ramakrishnan, WuX. , Modeling and predicting the helpfulness of online reviews, Eighth IEEE International Conference on Data Mining, Piscataway, NJ, 2008, pp. 443–452.
[22] D. Mayzlin, Promotional chat on the internet, Marketing Science 25 (2) (2006) 155–163.
[23] P.A. Pavlou, A. Dimoka, The nature and role of feedback text comments in online marketplaces: implications for trust building, price premiums, and seller differentiation, Information Systems Research 17 (4) (2006) 392.
[24] P. Rozin, E.B. Royzman, Negativity bias, negativity dominance, and contagion, Personality and Social Psychology Review 5 (4) (2001) 296–320.
[25] G. Salton, A. Wong, C.S. Yang, A vector space model for automatic indexing, Communications of the ACM 18 (11) (1975) 620.
[26] A. Sidorova, N. Evangelopoulos, J.S. Valacich, T. Ramakrishnan, Uncovering the intellectual core of the information systems discipline, MIS Quarterly 32 (3) (2008) 467–482.
[27] SpoolJ. , The Magic Behind Amazon's 2.7 Billion Dollar Questionavailable online at, http://www.uie.com/articles/magicbehindamazon/2009.
[28] S.H. Walker, D.B. Duncan, Estimation of the probability of an event as a function of several independent variables, Biometrika 54 (1) (1967) 167–179.
[29] C.P. Wei, C.C. Yang, C.M. Lin, A latent semantic indexing-based approach to multilingual document clustering, Decision Support Systems 45 (3) (2008) 606–620.
[30] L. Xinxin, L.M. Hitt, Self-selection and information role of online product reviews, Information Systems Research 19 (4) (2008) 456–474.

Qing Cao is the Jerry Rawls Professor of MIS in the Department of Information Systems and Quantitative Sciences for the Rawls College of Business at TTU in September. Cao holds a Ph.D. from the College of Business Administration at the University of Nebraska and he also earned an MBA from University of Wisconsin System and a B.S. in Mechanical Engineering from Shanghai JiaoTong University.
Cao's research interests include software quality metrics, e-commerce and m-commerce, IT outsourcing, supply chain information management, and artificial intelligence applications. Cao is the recipient of the University of Missouri at Kansas City Trustee's Faculty Research Award (2005). He has published more than 30 research papers in journals such as *Journal of Operations Management, Decision Sciences, Communications of ACM, International Journal of Production Research, Decision Support Systems, IEEE Transactions on Systems, Man, and Cybernetics, Information and Management, European Journal of Operational Research, Computers and Operations Research, Annals of Operations Research, Journal of Production and Innovation Management, International Journal of Project Management, Journal of Database Management, International Journal of Production Economics*, and among others. Cao serves as a member of the editorial review board at *Computer & Operations Research, International Journal of Service Sciences, International Journal of Information and Operations Management Education, International Journal of Information Systems and Change Management, and International Journal of Information Technology and Management.*He served as the Associate Program Chair, 2008 Decision Sciences Institute Annual Meeting.



Wenjing Duan, Assistant Professor of Information Systems & Technology Management, received her Ph.D. in Information Systems from University of Texas at Austin in 2006. Wenjing's research interests glide the intersections between Information Systems, Economics, and Marketing. Among her primary research interests are the economics of e-commerce, online communities and social networks, the Internet marketing, and online Intermediaries. Her ongoing research seeks to analyze issues in the following areas: the impact of consumer-generated content on e-commerce and electronic market; the dynamics and sustainability of online communities and social networks; and the impact of online intermediaries and band equity in electronic markets. Wenjing has published in MIS Quarterly, Communications of ACM, Journal of Retailing, and Decision Support Systems. She is also the recipient of the NET Institute Research Grant and serves on the Editorial Board of the Decision Support Systems.



Qiwei Gan is currently an MIS PhD candidate in the Rawls College of Business at Texas Tech University. Gan is also a PhD candidate in Finance at Southwestern University of Finance and Economics in China. Gan earned his MA in Applied Linguistics from Southwestern University of Finance and Economics. His current research interests include business intelligence, data mining, text mining, IT adoption and post-adoption in Health care, IT governance, and IT strategy. He has published in International Journal of Society Systems Science, and Chinese Journal of Finance and Economics, and also presented at AMCIS. He has been awarded the title of "SAS Student Ambassador (honorable mention)" in SAS Global Forun 2008.