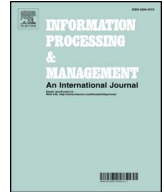




Contents lists available at ScienceDirect

# Information Processing and Management

journal homepage: [www.elsevier.com/locate/infoproman](http://www.elsevier.com/locate/infoproman)

## An analysis of review content and reviewer variables that contribute to review helpfulness

MSI Malik<sup>a</sup>, Ayyaz Hussain<sup>a,b,\*</sup><sup>a</sup> Department of Computer Science & Software Engineering, International Islamic University, Islamabad, Pakistan<sup>b</sup> College of Computing & Informatics, Saudi Electronic University, Saudi Arabia

### ARTICLE INFO

#### Keywords:

Stochastic GB  
 Helpfulness per day  
 Auxverb  
 Amazon reviews  
 eWOM  
 Machine learning

### ABSTRACT

Review helpfulness is attracting increasing attention of practitioners and academics. It helps in reducing risks and uncertainty faced by users in online shopping. This study examines unin-vestigated variables by looking at not only the review characteristics but also important indicators of reviewers. Several significant review content and two reviewer variables are proposed and an effective review helpfulness prediction model is built using stochastic gradient boosting learning method. This study derived a mechanism to extract novel review content variables from review text. Six popular machine learning models and three real-life Amazon review data sets are used for analysis. Our results are robust to several product categories and along three Amazon review data sets. The results show that review content variables deliver the best performance as compared to the reviewer and state-of-the-art baseline as a standalone model. This study finds that reviewer helpfulness per day and syllables in review text strongly relates to review helpfulness. Moreover, the number of space, aux verb, drives words in review text and productivity score of a reviewer are also effective predictors of review helpfulness. The findings will help customers to write better reviews, help retailers to manage their websites intelligently and aid customers in their product purchasing decisions.

### 1. Introduction

It is established that internet is a suitable venue for exchanging ideas, views and opinions on almost every real-life subject (Chen & Zimbra, 2010). Among them, online customer reviews (OCRs) have become electronic word of mouth for the current generations (Bjering, Havro, & Moen, 2015). Particularly OCR on products and services are significant for users seeking independent opinions. Service providers and manufacturers take online reviews very important because reviews are major source of products' and services' evaluations, they may show genuine concerns of online customers and provide useful market intelligence (Forman, Ghose, & Wiesenfeld, 2008; Huang, Chen, Yen, & Tran, 2015). Therefore, retailers and marketers are utilizing online reviews to understand the purchasing trends and attitudes of online consumers (Li & Hitt, 2010). In the neutral sense, we can say that online reviews are an opportunity or thread for merchants in their businesses (Yan, Wang, & Chau, 2015).

To investigate factors that make online reviews helpful in product purchase decisions for buyers are an important research question (Mudambi & Schuff, 2010). On review websites, thousands of customer reviews are being posted for popular products and this trend is rapidly increasing, which creates information overload problem (Liu, Huang, An, & Xu, 2008; Samha, Li, & Zhang, 2014). To resolve this problem, many websites such as Amazon.com drives in the customer feedback mechanism "Was this review helpful to

\* Corresponding author at: Department of Computer Science & Software Engineering, International Islamic University, Islamabad, Pakistan.  
 E-mail addresses: [msi\\_id@yahoo.com](mailto:msi_id@yahoo.com) (M. Malik), [ayyaz.hussain@iiu.edu.pk](mailto:ayyaz.hussain@iiu.edu.pk) (A. Hussain).

you (Yes or No)". The user reviews could be ranked by the number of attracted votes (Anderson & Magruder, 2012). Review helpfulness is the ratio of the number of helpful votes to the number of total votes received by a review. This ratio is referred as the helpfulness and it is the target variable in this study.

Helpful reviews are useful to online users for better and well-informed decisions. Websites that contain more helpful reviews offer necessary potential information to customers and more likely to attract buyers seeking information. Such reviews thus maximize consumers' satisfaction (Kohli, Devaraj, & Mehmood, 2004; Qazi et al., 2016). Early studies utilized quantitative variables of reviews to determine the helpfulness of online reviews. The variables are rating, length of review, thumb up/down etc. (Otterbacher, 2009; Pang, Lee, & Vaithyanathan, 2002). However, recent studies have paid more attention to qualitative determinants along with quantitative measures such as reviewer experience and impact, search goods, linguistic characteristics and cumulative helpfulness of reviewers to determine helpfulness (Huang et al., 2015; Krishnamoorthy, 2015; Mudambi & Schuff, 2010; Qazi et al., 2016).

## 2. Research questions and contributions

Helpfulness is a multi-faceted concept and various types of predictors' are being explored. Related studies investigated order effect (Zhou & Guo, 2017), sentiment (Chua & Banerjee, 2016), discrete emotions (Malik & Hussain, 2017), review effectiveness (Wu, 2017) and cognitive script (Ngo-Ye, Sinha, & Sen, 2017) for review helpfulness. However, review helpfulness has not been widely investigated as a function of significant psychological, linguistic, summary language and text complexity indicators of review content. These dimensions could offer new insights into review helpfulness. In addition, prior studies have not widely explored the influential dimensions (cumulative helpfulness per day, productivity per day) of reviewers. These reviewer characteristics could be more effective predictors for review helpfulness. The combination of these review and reviewer dimensions could provide additional insights into review helpfulness.

In the majority, prior studies have used basic machine learning algorithms for building an effective helpfulness prediction model. However, there is always demand for an effective machine learning algorithm that would be more powerful and robust. Stochastic gradient boosting (SGB) is one of the most popular and widely used machine learning models today. It is adaptable, easy to interpret, and produces highly accurate results. However, implementation of SGB is computationally complex (Ye, Chow, Chen, & Zheng, 2009). This study addresses three research questions:

- Which type of variables (review content or reviewer) effectively determines the helpfulness of online reviews?
- Does any set of variables exist that are most influential and have a strong relationship to review helpfulness?
- Which ML method delivers the best predictive accuracy to build an effective helpfulness prediction model?

The main objective of the current research is to examine the relationship between proposed review content and reviewer variables and helpfulness of online reviews and to highlight the influential predictors for review helpfulness. Three real-life Amazon review data sets are used to evaluate the contribution of variables. Six popular machine learning methods and three evaluation metrics are used for experimentation. An effective prediction model for online reviews is build using stochastic gradient boosting method. Theoretically, the results of this study added contributions to the prior research by highlighting influential variables of review content and their contributions to review helpfulness. More specifically, the research sheds light on the strong relationship between psychological, linguistic, summary language and text complexity variables and review helpfulness. Additionally, the results of the current research have extended the results found in prior studies (Lee & Choeh, 2014; Mudambi & Schuff, 2010) by exploring the effective dimensions of the reviewer (cumulative helpfulness per day and productivity score) to examine their influences on review helpfulness. Major contributions are:

1. We are the first one that adopts stochastic gradient boosting method to build an effective helpfulness prediction model for online reviews.
2. Eleven review content and two reviewer variables are proposed. In addition, three Amazon review data sets (one crawled and two publicly available) and six ML models are used for experimentation.
3. This study derived a mechanism including an example to extract the values of proposed review content (linguistic, psychological and summary language) variables from review text.
4. The findings indicate that helpfulness per day, syllables, space words; aux verb and productivity score are the most effective variables to determine review helpfulness.
5. This research facilitates e-commerce retailers and managers in minimizing the processing costs for better organization of their product reviews.

The rest of the article is organized as follows: Related work is presented in Section 3. Section 4 describes the research model (problem definition, variables, data set's description etc.) in detail followed by Section 5, which presents series of experiments, results, and findings. Then Sections 6 and 7 provides discussion and implications of the current research. Section 8 provides concluding remarks and directions for future work.

## 3. Related work

Among the many characteristics associated with online customers' reviews, review helpfulness is most important. Prior studies

Download English Version:

<https://daneshyari.com/en/article/4966381>

Download Persian Version:

<https://daneshyari.com/article/4966381>

[Daneshyari.com](https://daneshyari.com)