# A Quantitative Study of Impact of Incentive to Quality of Software Reviews

Mingwei Tang[1], Zhiwei Xu[1], Yuhao Qin[1], Cui Su[1,*], Yi Zhu[1], Feifei Tao[2], and Junhua Ding[3]
[1]Nanjing Audit University, Nanjing, Jiangsu, P.R.China
[2]Hohai University, Nanjing, Jiangsu, P.R.China
[3]University of North Texas, Denton, Texas, U.S.A
tmw@nau.edu.cn, 15751833228@163.com, 806350899@qq.com, 15737239019@163.com, 449091513@qq.com, tff@hhu.edu.cn, Junhua.Ding@unt.edu
*corresponding author

*Abstract*—**Online reviews like product reviews are important references for a potential user to learn the basic information of a product. However, some reviewers were rewarded for writing the reviews, which may impact the objectiveness of the reviews. But on the other hand, the reviews written by reviewers who weren't rewarded could be in low quality. Research study already showed incentivized reviewers may give a higher overall score than non-incentivized reviewers (or called organic) do, but does that pattern also apply to review content? In this paper, a quantitative comparison study is conducted to investigate the differences between the incentivized reviews and organic reviews of software products. Four pairs of comparison including overall score, sentiment preference, correlation, and similarity are performed by using statistical and text mining methods. The results show there is no statistically significant difference between the incentive reviews and organic reviews except the sentiment of total, "Problems and Benefits" and "Summary" part of a review text. The results are unexpected since the reviews collected from the website for reviewing software product already filtered low quality reviews. It demonstrates that the incentivized action might not be necessary to produce biased reviews and it may be an effective way to attract more reviews since the website include more than 75% incentivized reviews. The paper also analyzed the possible reasons from the feature of reviewers' position in a company, a review's indicator, and reviewers' common actions. Based on the analysis, this study suggests that a potential user may pay attention to some quality dimensions of a review to mitigate the bias risk from the reviews.**

***Keywords: online review; quantitative study; sentiment analysis; correlation; similartiy***

## I. INTRODUCTION

Online reviews are a kind of reviews about products whose information are posted online. It is an important reference to people who prepare to purchase a specific product. Online reviews usually contain one product's overall score, brief description, function introduction, user experience and so on. That may help people make decision if they should buy this product or not. However, with the advent of web with handy access, the product manufactures or sellers could conveniently invite people to write reviews for their products in their own sales systems or the third-party E-commerce platforms. Generally, these invited people are paid for their reviews. On the other hand, if the strategies of these systems or platforms allow, anyone could write reviews for those products. Such reviews behavior separates the product reviews into two parts. One part is incentivized while another is organic.

Among the incentivized reviewers, they definitely have actual users of these products. However, there also have some professional reviewers who don't have strong relations with these products. It is similar among the organic reviewers. Especially, there might have some malicious reviewers invited by the competitors among the organic reviewers. Therefore, even if some systems have clearly annotated these two kinds of reviews, it is still difficult for the potential consumers to identify which kind of reviews have higher quality and can provide more valuable information to them. It is uncertain that how the incentivized reviewers affect the quality of product reviews. To alleviate the misleading caused by some sham or highly subjective reviews, studies on quality evaluation of product reviews has become increasingly popular these years.

Text mining is the main approach to evaluate the quality of reviews [1]. It tries to extract the actual opinions from the reviews by analyzing the words' features. Sentiment and subjectivity analysis are the general ways to discover the meaningful opinions from reviews [2]. By this way, it could evaluate the reviews' quality one by one. However, it is not significantly useful for a user to know one review's quality. If the difference between incentivized reviews and organic reviews could be dug out, it could help users find the correct orientation to get the valuable reviews. Thus, a multi-dimension contrastive investigation on incentivized reviews and organic reviews is necessary to discover the difference between the two kinds of reviews.

In this paper, we propose using overall score, sentiment preference, correlation and similarity as the four indicators to evaluate the difference between incentivized and organic reviews. The concrete computation methods of the four indicators are presented. Then, four comparison experiments are conducted. Statistical and text mining methods are used to analyze the difference of the two kinds of reviews. According to the results, it explains the reasons why the four indicators perform. Finally, the paper provides suggestions on which part of a review should be focused.

The remainder of this paper is organized as follows. Section II contains a review of related works. Section III presents the concrete computation methods of the four indicators, and four experiments are displayed in this section. In Section IV, the results are discussed in detail. Section V

2767-6684/22/$31.00 ©2022 IEEE
DOI 10.1109/DSA56465.2022.00016

presents the threats to validity of the research. Section VI makes a conclusion and highlights the future work.

## II. RELATED WORKS

Existing related studies mainly focus on the effects of online reviews, using text mining method to discover the value of online reviews, and quality evaluation on online reviews.

### A. Effects of online reviews

High credibility online reviews are definitely important to potential consumers [3]. It may provide decision support for potential consumers [4]. Therefore, the detailed effects of online reviews are discovered by several studies. Online reviews are generally used to rank products. Sedef Çalı et al. used sentiment analysis method to convert the online reviews into performance scores which are used to rank the alternative products [5]. Jian-Wu Bi et al. used interval type-2 fuzzy numbers to propose a new approach for product ranking by representing the sentiment analysis results of online reviews [6]. Prasad Vana et al. investigated how individual affect consumers' purchase likelihood by exploiting the variation in review positions [7]. Zhi-Ping Fan et al. proposed a product ranking information fusion framework based the online reviews [8]. How online reviews influences customers' or sellers' behavior is another research orientation. Bettina von Helversen et al. investigated how product attributes, average consumer ratings, and single affect-rich positive or negative consumer reviews influenced hypothetical online purchasing decisions of younger and older adults [9]. Wen Song et al. developed a game-theoretic model to explore how online reviews impact a third-party seller's decision to sell on an open retail platform and the platform retailers profit [10]. Hoon S. Choi et al. investigated the determinants of online review helpfulness. Since online reviews is mandatory disclosed, how the incentivized or organic online reviews influences consumer becomes popular [11]. Thomas Reimer et al. used attribution theory to analyze how online reviews incentives influence the recommendation audience [12]. Steven Stanton et al. found that the different incentivize online reviews generation methods yielded the significantly different moral judgments, which then predicted consumers' attitudes toward the resort and the resort's image [13]. Su Jung Kim et al. examined the different characteristics and effects of sponsored and organic online reviews [14]. Jin Ai et al. explored the effects of incentive-driven online reviews on receivers' trust utilizing norm conflict and stakeholder perspectives [15]. Geng Cui et al. found mandatory disclosures for incentives have a positive effect on review helpfulness and sales [16]. Xiaorong Wang et al. proposed a moderated moderation model to explore the interactive effect between sponsorship disclosure of positive reviews, emotional intensity, and tie strength on online review [17]. Additionally, online reviews are also used in competitor analysis [18], or to study the interaction effects between reviews and consumers [19]. These studies show that online reviews could play a pivotal role for consumers to make purchase decision.

### B. Text mining on online reviews

Since an online review is a kind of text, text mining methods are used to try to discover the actual opinion of online reviews. Sentiment analysis is a generally used text mining approach on online reviews [20]. M. Salehan et al. used a big data-based sentiment mining approach to investigate the predictor of readership and helpfulness of online reviews [21]. Feng Zhou et al. combined affective lexicons and a rough-set technique to predict sentence sentiments for individual product features with acceptable accuracy [22]. Hong Hong et al. conducted a meta-analysis to examine the determinant factors of perceived review helpfulness in order to reconcile the contradictory findings about their influence on perceived review helpfulness [23]. Ana Costa et al. used a data mining approach to check whether or not a new review published was incentivized [24]. João Guerreiro et al. used machine learning based classification method to identify drivers of explicit recommendations of online reviews [25]. Xue Li et al. proposed an evolutive preference analysis method to handle the dynamic online ratings [26]. It could help firms verify whether their advertised products match the preferences of target consumers. Hao-Chiang Koong Lin et al. extracted characteristic keywords from collected consumer reviews for affective polarity analysis [27]. Jee Young Lee used machine learning method to analyze user requirement issues to improve the quality of software [28]. E. Kauffmann et al. used sentiment analysis to create a fake review detection framework [29]. J. Zhang et al. proposed an unsupervised aspect sentiment analysis method to measure customers' preferences [30]. Ning Zhang et al. applied LDA topic model and sentiment analytics method to extract the repeat purchase intention variables from online reviews [31]. Guohou Shan et al. developed 22 features of review inconsistency based machine learning model for online reviews detection [32]. Yubao Zhang et al. proposed a novel detection approach based on co-review graphs to identify the incentivized reviews [33]. Zhen He et al. applied text-mining approaches and Integrated-Degree based K-shell decomposition to convert online reviews into competitive insights including competitor identification, product ranking, product comparison and so on [34]. Miriam Alzate et al. used a lexicon-based approach to extract brand image and brand positioning from online reviews [35]. Jisu Yi et al. applied a word-level bigram analysis to derive product attributes from review text and examined the influence of the number of attributes on the review's helpfulness votes [36]. Hongjie Deng et al. combined CNN and BiLSTM to analysis the emotional tendency in online reviews [37]. There studies dug the effects of online reviews furtherly.

### C. Evaluation on quality of online reviews

Online reviews have so many positive effects for potential consumers. However, that should be based on the high quality online reviews. Therefore, quality evaluation is becoming increasingly important nowadays. Several studies evaluate the quality of online reviews from its features. The approaches of information science are the general used methods [38]. Jo Mackiewicz et al. evaluated the review quality from credibility, informativeness, and readability by using a survey based quantitative study [39]. Kem Z.K. Zhang et al. proposed a dual-process based heuristic-systematic model to identify the important factors to consumers' purchase decision-making [40]. Raffaele Filieri used a grounded theory approach to evaluate the level of trustworthiness of online reviews [41]. Guohou Shan et al. studied the potential inconsistency between product ratings and review content so as to better assist potential consumers with making purchase decisions [42]. Seongsoo Jang et al. developed a hierarchical log-linear model to evaluate the importance of the functional and emotional content in online reviews [43]. Mohammad Sadegh Nasiri et al. proposed a customer satisfaction model to evaluate the significant factors in consumer perceived value about purchasing refurbished smartphones [44]. Lili Zheng et al. used a five-factor communication process framework to classify the literature on online reviews to provide understanding of the multi-featured nature and complexity of online reviews to related researchers [45]. Data science related approaches are also employed. Erick Kauffmann et al. used NLP based approach to evaluate the authenticity of online reviews [46]. Guanxiong Huang et al. used a computational-experimental approach to evaluate the trustworthiness of online reviews from textual features [47]. Satanik Mitra et al. used machine learning based qualitative approach to evaluate the helpfulness of online reviews from semantic and syntactic features of review contents [48]. Xun Wang et al. employed a Bayesian-based inference method to mine the harshness of online reviews [49]. Since incentivized online reviews attracts public attention. Evaluations are conducted to find the difference between incentivized and organic online reviews. Maria Petrescu et al. applied exchange theory to analyze the relationship between incentivized reviews and the satisfaction ratings assigned by consumers to a product [50]. Mingyue Zhang et al. used the accountability theory to evaluate how incentives with reevaluation mechanism actually influence reviewers' behaviors [51]. Bogdan Anastasiei et al. assessed the relationships between perceived argument quality and perceived source expertise to evaluate which kind of online reviews (incentivized or non-incentivized) has more influence on customers' perceptions [52]. In these regards, no matter what kinds of approaches, the features of online reviews are the main evaluation objects.

In summary, current studies of online reviews are becoming increasingly popular and important. These studies proved online reviews are helpful for users and can be mined to get more valuable information. However, most of the studies focused on quality evaluation and reutilization. Fewer of them paid attention to the comparison on the contents of incentivized and organic online reviews from their textual contents. What are the differences between the two kinds of online reviews? How does the incentivized online reviews impact on the quality of product reviews? They are still uncertain. Obviously, the answers of the two questions will be helpful for potential consumers. That is the research purpose of this paper.

## III. METHODS AND RESULTS

### A. Research Hypotheses

Generally, the incentivized reviewers are paid for their reviews. Perhaps, the manufactures or sellers may ask them to give a relatively high overall score. In spite of this, they couldn't completely control what a reviewer will write about the specific products. Some reviewers may comment the products realistically while some may not. However, whatever which kind a reviewer is, the contents of the online reviews usually contain a reviewer's sentiment preference on the product or even the review itself. Thus, sentiment analysis might discover more actual feeling than review score. Additionally, it is possible that manufactures or sellers may ask a review to write reviews as what they want. If that is true, the correlation which means the similarity between a pair of items in incentivized reviews data is not the same with that of organic online reviews data. At the same time, the similarity between the two groups should be different either.

According to the analysis, we propose the following four hypothesis.

Hypothesis H1. The overall score of the incentivized and organic online reviews is the same.

Hypothesis H2. The sentiment preference of the incentivized and organic online reviews is the same.

Hypothesis H3. The internal similarity(correlation) of the incentivized and organic online reviews is the same.

Hypothesis H4. The external similarity between the incentivized and organic online reviews is the same.

### B. Experiment Design

#### 1) Design collection

To test the four hypotheses, we collected the online reviews data from www.g2.com. It is a third-party independent website which provides professional review service for the most popular software products. Since G2 also provides paid reviews reports for users, it has several strategies to ensure the reality and quality of online reviews. All the reviews users can read are published by the real users of a software product and have already been filtered by G2. For example, G2 will validate the identification of a user and put logic in place to ensure the reviews shown on product pages are the most helpful to buyers. Comparing to e-commerce platform like Amazon, or the software company's selling platform like Microsoft online store, the reviews on G2 should be relative objectiveness, independent and high quality. Currently, there are already over 1,677,700 authentic, timely online reviews from real users.

We fetched the data from G2 by web crawl and used html parser to parse all the reviews data into JSON format and finally stored them into a csv file. Table I shows a review example. It shows that an online review in the G2 contains "overall score", "Summary", "Like", "Dislike", "Recommendation", and "Problems and Benefits (P&B)" about one product. Even more importantly, it has annotated the online reviews as incentivized or organic clearly. The field of "Verification1" is used to annotate if the user is verified user. "Verification2" to "Verification 4" are used to annotate the review type. We got three type of values which are "Review source: G2 invite on behalf of seller", "Incentivized Review" and "Organic Review". According to the "Verification2" to "Verification4" fields, we collected totally 18,000 items of online reviews data from the "CRM" software category in G2. Half of them are incentivized while another half are organic.

TABLE I

An Example of a Review

| Field | Content |
|---|---|
| User | Rhobinson M |
| Identity | Senior Area Sales Manager |
| Companysize | Mid-Market (51-1000 emp.) |
| Verification1 | Validated Reviewer |
| Verification2 | Review source: G2 invite on behalf of seller |
| Verification3 | Incentivized Review |
| Verification4 | |
| Score (Overall score) | 5 |
| Date | 19-Nov-20 |
| Summary | "Wrike streamlines our teamwork remotely enhancing collaboration." |
| Like | Its now 5 years since i started using Wrike in my daily routines at workplace i have got unlimited things i like about Wrike. One thing this platform is built with simple dynamics making so easier to use. Another thing its designed in a web like interface that is very user friendly and highly customizable. I like that this platform offers a free version for 14-day trial . Works perfectly on mobile devices and other devices like laptops and tablets. Hence more portable. Intuitive customer support team readily available. Review collected by and hosted on G2.com. |
| Dislike | At least not one thing i have found not working accordingly. Incase of any question i always rely by the customer support team. Review collected by and hosted on G2.com. |
| Recommendation | Wrike is very user friendly, intuitive and above all very easy to use. Its very inexpensive affordable by all potential. Highly trained customer service team to tackle any technology. So i would recommend anyone with projects and tasks problem consider Wrike. Review collected by and hosted on G2.com. |
| Problems&Benefits(P&B) | I have been solving numerous problems and getting better results using Wrike. Like in my case i am able to keep a real time track of projects. I am also able to make projects and organiz them in the folders i have made and create subtitles for each folder containing projects. There is easy access of the projects. Wrike has got a very documented calendar to give you remainders on important projects and tasks. Review collected by and hosted on G2.com. |

2) *Experiment setup*

According to the approaches we plan to use in this study, the experiment setup plan is shown in Table II.

TABLE II

Experiment Plan

| Online Reviews Data | Experiment (comparing) | Method |
|---|---|---|
| 1,000 incentivized vs 1,000 organic | Overall score | A/B Testing |
| | Sentiment | |
| | Internal similarity (Correlation) | Mean value comparison |
| 9,000 incentivized vs 9,000 organic | External similarity | Distribution comparison<br>Mean value comparison<br>Standard deviation comparison |

Firstly, we selected 1,000 pairs (2,000 items) data of incentivized and organic online reviews as the dataset of the first three experiments. Since two items of reviews can only have a similarity value, similarity is not suitable to use A/B testing. The classical statistical methods were employed to test the differences of the two kinds of similarity. More importantly, in external similarity experiment, we used the whole dataset. Because the bigger the dataset, more meaningful the external similarity comparison.

### C. *Overall score comparison*

1) *Design and procedure*

Overall score is a quantity score for a product. According to the rules of A/B testing, the procedure of this experiment is shown in Fig. 1.

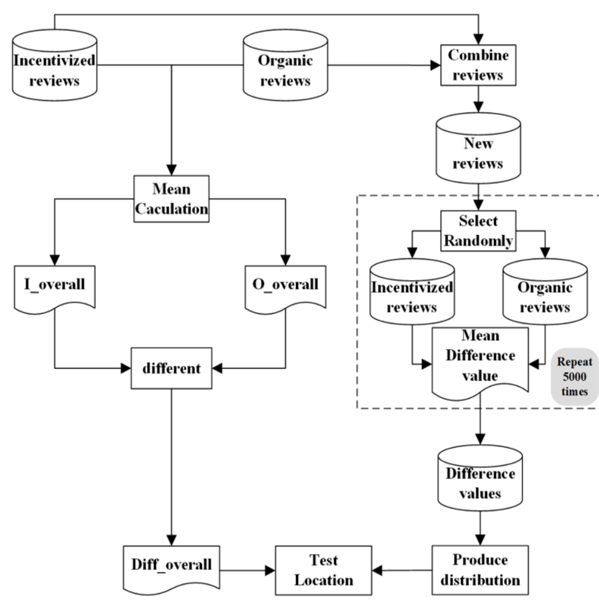


Fig. 1. The procedure of overall score comparison experiment

Firstly, calculate the mean score of 1,000 incentivized reviews, and mark the result as I_overall, while mark the mean score of 1,000 organic reviews as O_overall. Calculate the difference value of the two average score, mark the result as Diff_overall.

Secondly, combine the 2,000 reviews to get a new collection. Then, choose a new 1,000 pairs of online reviews randomly from the new online reviews collection. However, this kinds of incentivized or organic online reviews are not really incentivized or organic. They are just chosen randomly from the new collection and manually marked as incentivized or organic. Based on the new pair, a new difference value of mean overall score can be computed. Repeat this random selection and difference value calculation for 5,000 times, then 5,000 difference values can be got.

Use the 5,000 values to produce a distribution, calculate the mean and standard deviation value of the 5,000 values. Then test where the Diff_overall is located in the distribution. If the P value is set to 0.05, and Diff_overall is out of two times of standard deviation, then H1 is not held, but alternative of H1 is held. Otherwise, H1 is true.

2) *Results*

According to procedure, the results of the experiment are shown in Fig. 2.

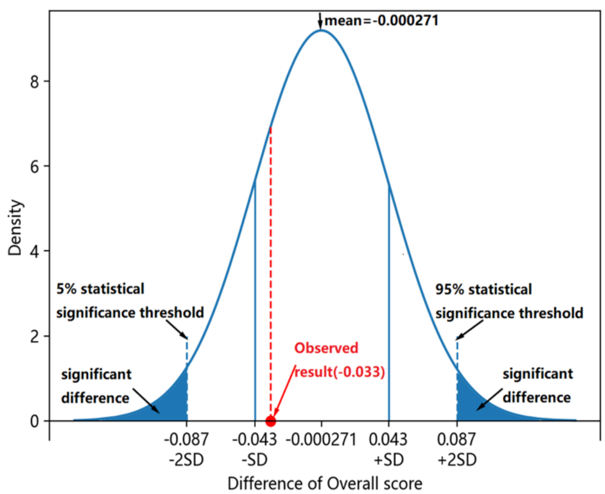


Fig. 2. The difference of overall score

The result shows the observed result is in the confidence interval. It means that on overall score, there is no significant difference between the incentivized and organic online reviews. Then H1 is true.

### D. *Sentiment comparison*

1) *Design and procedure*

Sentiment comparison aims at discovering the difference of reviewers' emotional preferences in the two groups. Emotion is implicit in the text and can be reflected by some specific words such as "fine", "free", "hate" and et al. Thus, before the sentiment comparison, we need to analyze the sentiment of online reviews text from the meaning of the words and express the sentiment in quantity [53].

After comparison and identification, we choose TextBlob to compute the sentiment value. TextBlob is an NLP-based python package that could output a sentiment value of a sentence or paragraph [54]. Take the "Summary" shown in Table I as example, the computation procedure of its sentiment value is shown in Fig. 3.

```
In [9]: from textblob import TextBlob

In [10]: summary=TextBlob("Textblob is amazingly simple to use. What
great fun!")

In [11]: summary.sentiment.polarity
Out[11]: 0.39166666666666666
```

Fig. 3. The screenshot of "Summary" sentiment

The sentiment calculation by TextBlob is relatively easy. Just provide a sentence as the input, it will calculate its sentiment value automatically. It will return three values which are polarity, subjectivity and intensity. In this study, polarity is necessary. Therefore, we just invoke the "sentiment.polarity" to get the polarity as the sentiment value.

The polarity value goes from "-1" to "1". "-1" means the completely negative sentiment while "1" means the completely positive sentiment. It is calculated based on the sentiment lexicon which is embedded in TextBlob. In the lexicon, there include sets of words that have sentiment meanings. One word will be defined in several contexts, and in each context, the word has a sentiment value. When load the word, TextBlob will locate the word in the lexicon, average the word's all sentiment value and return it as its final sentiment value. However, if negative words like "no", "not" and et al. appear in a sentence with the target word, then the sentiment value will be multiplied by -0.5. If some modifiers like "very", "extremely" and et al. appear, the sentiment value will be the product of polarity and intensity. On this basis, we loaded each part of a review by sentences, and average all the polarity values as the final sentiment value of each part. Then average all parts' sentiment values to get the total sentiment of a review. The main computing procedure of a review's sentiment is shown in Fig. 4.

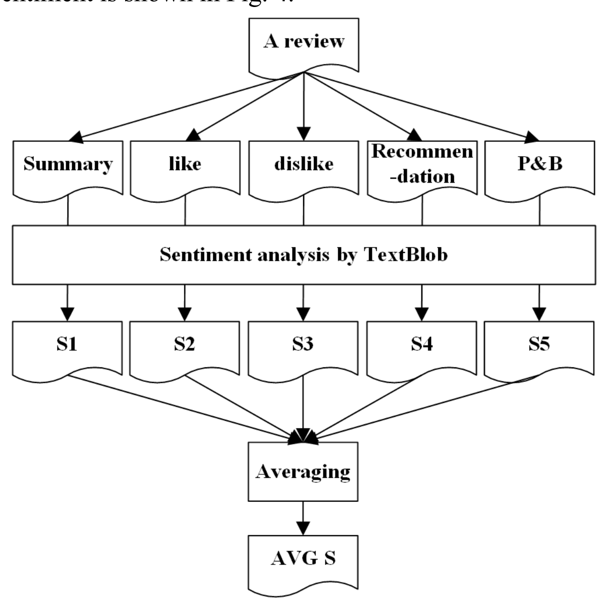


Fig. 4. The computing procedure of a review's sentiment

2) *Results*

The Fig. 5 shows the result of total sentiment.

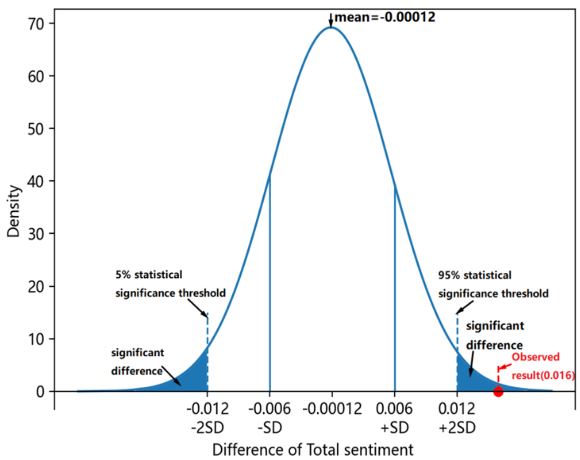


Fig. 5. The difference of total sentiment

Fig.5 shows there is significant difference between the two kinds of online reviews. Then H2 is false. And the results are same on the “P&B” and “Summary”. They are shown in Fig. 6 and Fig. 7.

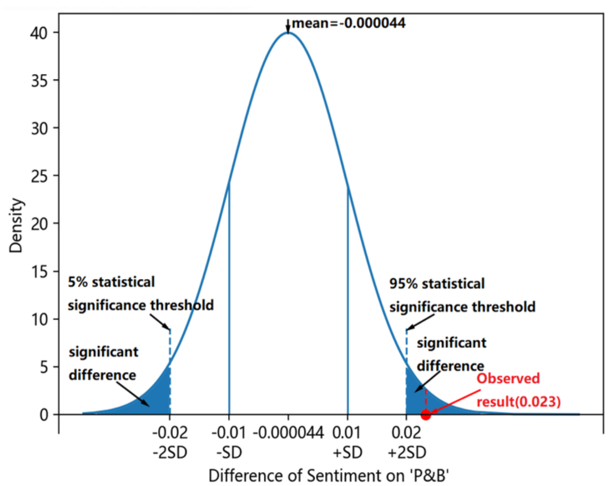


Fig. 6. The difference of sentiment on “Problems and Benefits (P&B)”

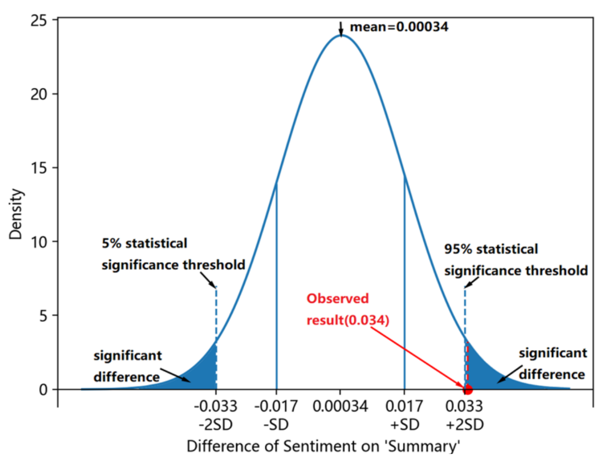


Fig. 7. The difference of sentiment on “Summary”

However, there is no significant difference on “Like”, “Dislike” and “Recommendation”. The H2 is true on these dimensions. The results are shown in Fig. 8 to Fig. 10.

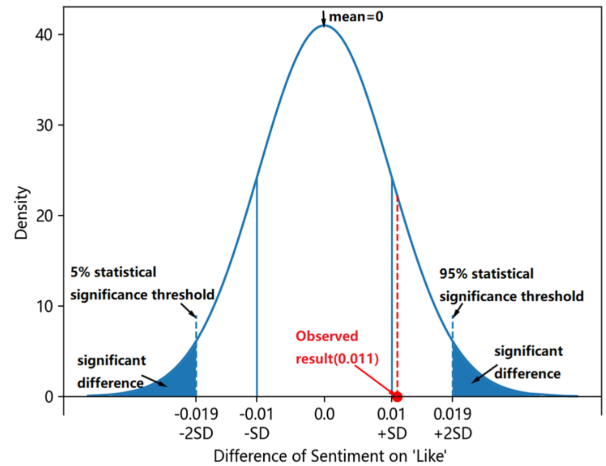


Fig. 8. The difference of sentiment on “Like”

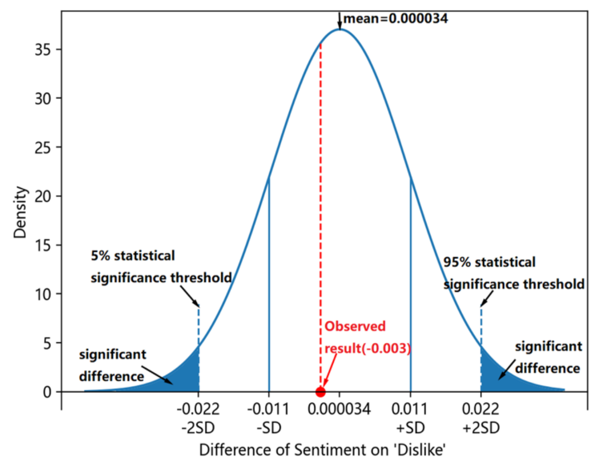


Fig. 9. The difference of sentiment on “Dislike”

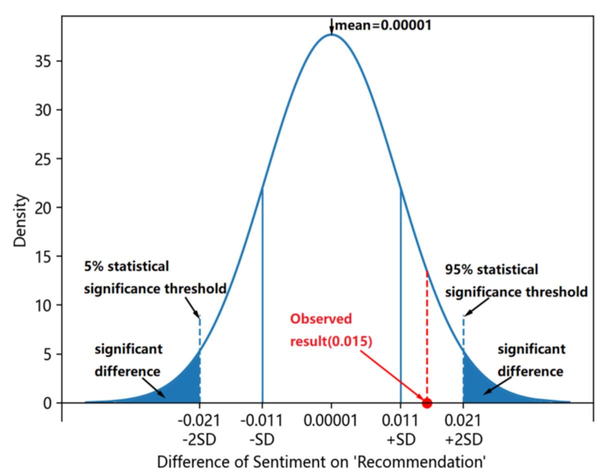


Fig. 10. The difference of sentiment on “Recommendation”

## E. Correlation comparison

### 1) Design and procedure

Correlation describes the similarity between the internal data of a collection. It could be computed by averaging all the similarity between any two items of a collection. For one collection which has 1,000 reviews, the correlation computing procedure is shown in Fig. 11.

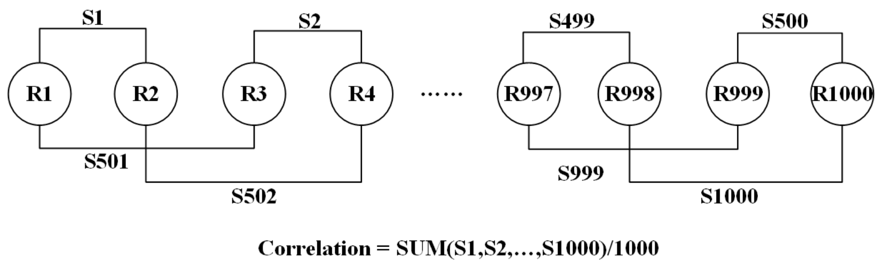


Fig. 11. The correlation computing procedure

This correlation algorithm hasn't computed all the similarity between any two reviews. It computes the similarity by pairs and by two intervals. That can improve the computing efficiency and still could get a relative accurate value. In this study, spaCy which is an NLP python package is used to compute the similarity between any two reviews. The similarity algorithm is based on the TF/IDF cosine similarity. Its value goes from "0" to "1". "0" means completely irrelevant while "1" means completely same.

By this method, the correlation of the incentivized and organic online reviews could be got respectively. Then checking the difference value between the two correlation values to test H3. If the absolute value of the difference is greater than zero, then H3 is false. Otherwise, H3 is true.

### 2) Results

After computation, the correlation of the 1,000 incentivize online reviews is 0.79317 while that of the 1,000 organic online reviews is 0.77463. The difference value is 0.01854. Although it is greater than zero, it is really small. Therefore, H3 could be treated as true.

## F. Similarity comparison

### 1) Design and procedure

The final experiment is to check the similarity between the two kinds of online reviews in contents. In this experiment, the similarity is computed between one incentivized review and one organic reviews. if only in this way, it is meaningless to compare two collections' similarity. Although the two reviews may not for the same products, if the size of the dataset is large enough. It is still statistically significant. Therefore, we extended the size of dataset to 18,000. On this dataset, the main procedure is shown in Fig. 12.

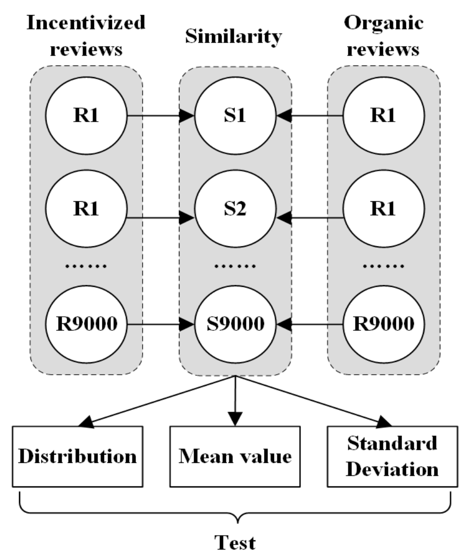


Fig. 12. The main procedure of computing similarity

After computing, 9,000 similarity values will be got. Then compute the distribution, mean value and standard deviation of the values respectively to test H4. The testing logic of H4 is shown in the decision tree.

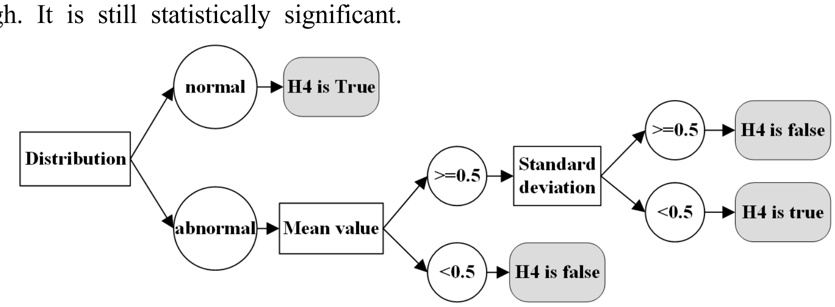


Fig. 13. The decision tree for testing H4

2) *Results*

The distribution of the mean values between the 9,000 pairs of incentivized and organic online reviews is shown in Fig. 14.

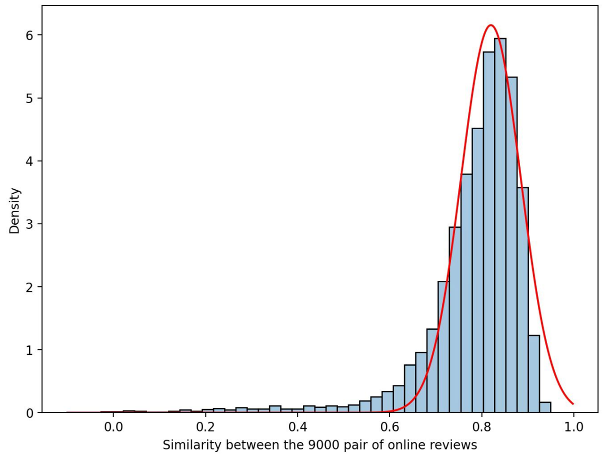


Fig. 14. The distribution of the similarity values

As is shown in Fig. 14, it is a right skewed distribution. The mean value of the similarity is 0.78463 and the standard deviation is 0.11198. According to the logic shown in the decision tree, H4 is true.

## IV. Discussion and Conclusion

In common sense, the incentivized reviewers will give more positive reviews than the organic reviewers. However, the results are not as expected. Only the sentiment scores on the total, “P&B” and “Summary” are close to common sense. Except that, there is no significant difference between the incentivized and organic online reviews. There are three possible reasons for this.

Firstly, G2 is a high-quality and independent review website. It is different with Amazon, Microsoft online store. In Amazon, a user could publish any reviews. Amazon can’t ensure the quality of a user’s review. While in Microsoft online store, Microsoft may only display the reviews which are good for them. However, G2 is independent, and its purpose and strategies try to make review objectiveness. It looks like an independent newspaper which publish comments about something. It will choose the actual users of some specific software strictly as the incentivized reviewers. Among these reviewers, there are chief of staff, manager, director, strategist, co-founder and even CEO. Most people of this crowd may have good moral and professional ethic. That results in the objectiveness and truthfulness of reviews on G2.

Secondly, some review indicators may not reflect the actual feeling of the reviewers. Especially on the overall score, both the average scores of the two kinds of online reviews are around 4.3. According to rule of G2, reviewers select the number of stars to give the overall score. It is not an accurate review method which can express the reviewer’s actual feeling. In common sense, only if the software is really terrible, reviewers will comment it less than four stars. Additionally, the software listed in G2 are the most popular software in market. Therefore, they have similar performance in market. That’s why they have similar overall scores. It is similar on the sentiment of textual review contents. Except total, “P&B” and “Summary”, there is no significant difference on the sentiment of “Like”, “Dislike” and “Recommendation”. The total sentiment is got by averaging all the sentiment values of the five dimensions. It is a balanced value. It could provide an overall sentiment preference on one software. However, it couldn’t provide more detailed information for potential users. “P&B” is abbreviation for problems and benefits. It means “What problems are you solving with the product? What benefits have you realized?”. This review indicator is clearer than other indicators. And when the reviewers are rational and objective, they will and they can take such clear indicator seriously. Besides, the average “P&B” sentiment value of the incentivized online reviews is greater than that of the organic online reviews. And the average words count of the two kinds of reviews are computed too. The results are listed in TABLE III.

TABLE III

The Average Words Count Comparison on “P&B”

| Reviews Data | Mean value | Median value |
|---|---|---|
| incentivized | 37.22 | 31 |
| organic | 35.37 | 29 |

The results show that there are about only two words difference in the average length of each review. Therefore, the difference in “P&B” between the two kinds of reviews is to some extend caused by the different contents. That means on the “P&B” dimension, the incentivized reviewers comment it with more positive words than the organic reviewers do. Therefore, “P&B” is more valuable especially in the website of G2. Although there is significant difference on the “Summary” dimension, the observed value is only a little over the confidence interval. And the average words count comparison are listed in TABLE IV.

TABLE IV

The Average Words Count Comparison on “Summary”

| Reviews Data | Mean value | Median value |
|---|---|---|
| incentivized | 6.36 | 6 |
| organic | 6.72 | 6 |

The results show the average length of the organic reviews is a little longer than that of the incentivized reviews. From this point of view, there is still minor difference between the two kinds of reviews. Therefore, there is no significant reference value on “Summary”.

Thirdly, there exist commonalities among the action of reviewers especially in the same software category. The two kinds of reviewers may use similar words to make reviews for a software. That results in the high correlation between the inside of each online reviews collection and also the high similarity between the two kinds of online reviews collection.

By the analysis above, at least in the "CRM" category of G2 website, "P&B" is more valuable. It is suggested that a potential user could pay more attention to the "P&B" part of an online review. Comparing the "P&B" part of the two kinds of online reviews, they may find some differences. That may help them purchase the software product.

## V. Threats to Validity

The current study is based on the high score reviews on G2. However, if the reviews as not so good, then the results may be largely different. Besides, the experiment data is from the online reviews of the "CRM" software category on G2. If more categories' reviews could be compared, maybe there will be more interesting in the results. Factually, there are many other websites which also could provide fruitful online reviews. The cross-site comparison may discover the information of universal values which can provide more directly help for a potential user.

## VI. Conclusion and Future Work

This study used statistical and text mining methods to make a multi-dimension comparison between the incentivized online reviews and organic online reviews from overall score, sentiment, correlation and similarity. The comparison results show there is nearly no significant difference between the two kinds of online reviews except the sentiment of total, "P&B" and "Summary" part of a review text. That means the incentivized action has little impact on the quality of products reviews. The results are unexpected since G2 has already filtered low quality reviews. On this basis, it demonstrates that the incentivized action might not be necessary to produce biased reviews and it may be an effective way to attract more reviews since the website include more than 75% incentivized reviews. Finally, the study explained the possible reasons from review sources, the feature of reviewers' position in a company, a review's indicator, and reviewers' common actions. Based on the analysis, this study suggests that a potential user may pay attention to some quality dimensions of a review to mitigate the bias risk from the reviews.

As is analyzed in Section V, the future plan will try to collect more reviews data from multiple categories of G2. And furthermore, the reviews data will be classified by overall score. Multiple comprehensive comparisons on high overscore and low overscore from different categories will be conducted. Many more interesting and valuable things are expected to be found.

## Acknowledgment

This work is funded by the Significant Project of Jiangsu College Philosophy and Social Sciences Research (NO: 2021SJZDA153); the Youth Project of National Natural Science Foundation of China (NO: 62006121); the Qing Lan Project of Jiangsu College; Postgraduate Research & Practice Innovation Program of Jiangsu Province (NO: KYCX21_ 1951); Jiangsu Water Conservancy Science and Technology Foundation (No:2020014).

## References

[1] C. Tang and L. Guo, "Digging for gold with a simple tool: Validating text mining in studying electronic word-of-mouth (eWOM) communication," Marketing Letters, vol. 26, no. 1, pp. 67–80, 2015.

[2] P. Zhao, J. Wu, Z. Hua, and S. Fang, "Finding eWOM customers from customer reviews," Industrial Management & Data Systems, vol. 119, no. 1, pp. 129–147, 2019.

[3] I. Pranata and W. Susilo, "Are the most popular users always trustworthy? The case of Yelp," Electronic Commerce Research and Applications, vol. 20, pp. 30–41, 2016.

[4] M. Akbarabadi and M. Hosseini, "Predicting the helpfulness of online customer reviews: The role of title features," International Journal of Market Research, vol. 62, no. 3, pp. 272–287, 2020.

[5] S. Çalı and Ş. Y. Balaman, "Improved decisions for marketing, supply and purchasing: Mining big data through an integration of sentiment analysis and intuitionistic fuzzy multi criteria assessment," Computers & Industrial Engineering, vol. 129, pp. 315–332, 2019.

[6] J. W. Bi, Y. Liu, and Z. P. Fan, "Representing sentiment analysis results of online reviews using interval type-2 fuzzy numbers and its application to product ranking," Information Sciences, vol. 504, pp. 293–307, 2019.

[7] P. Vana and A. Lambrecht, "The effect of individual online reviews on purchase likelihood," Marketing Science, vol. 40, no. 4, pp. 708–730, 2021.

[8] Z. P. Fan, G. M. Li, and Y. Liu, "Processes and methods of information fusion for ranking products based on online reviews: An overview," Information Fusion, vol. 60, no. December 2019, pp. 87–97, 2020.

[9] B. von Helversen, K. Abramczuk, W. Kopeć, and R. Nielek, "Influence of consumer reviews on online purchasing decisions in older and younger adults," Decision Support Systems, vol. 113, no. June, pp. 1–10, 2018.

[10] W. Song, W. Li, and S. Geng, "Effect of online product reviews on third parties' selling on retail platforms," Electronic Commerce Research and Applications, vol. 39, no. September 2019, p. 100900, 2020.

[11] H. S. Choi and S. Leon, "An empirical investigation of online review helpfulness: A big data perspective," Decision Support Systems, vol. 139, no. September, p. 113403, 2020.

[12] T. Reimer and M. Benkenstein, "Not just for the recommender: How eWOM incentives influence the recommendation audience," Journal of Business Research, vol. 86, pp. 11–21, 2018.

[13] S. J. Stanton, J. Kim, J. C. Thor, and X. Deng, "Incentivized methods to generate electronic word-of-mouth: Implications for the resort industry," International Journal of Hospitality Management, vol. 78, pp. 142–149, 2019.

[14] S. J. Kim, E. Maslowska, and A. Tamaddoni, "The paradox of (dis) trust in sponsorship disclosure: The characteristics and effects of sponsored online consumer reviews," Decision Support Systems, vol. 116, pp. 114–124, 2019.

[15] J. Ai, D. Gursoy, Y. Liu, and X. Lv, "Effects of offering incentives for reviews on trust: Role of review quality and incentive source," International Journal of Hospitality Management, vol. 100, no. October 2021, p. 103101, 2022.

[16] G. Cui, Y. Chung, L. Peng, and W. Zheng, "The importance of being earnest: Mandatory vs. voluntary disclosure of incentives for online product reviews," Journal of Business Research, vol. 141, pp. 633–645, 2022.

[17] X. Wang, F. Xu, X. (Robert) Luo, and L. Peng, "Effect of sponsorship disclosure on online consumer responses to positive reviews: The moderating role of emotional intensity and tie strength," Decision Support Systems, vol. 156, p. 2022, 2022.

[18] J. Jin, P. Ji, and R. Gu, "Identifying comparative customer requirements from product online reviews for competitor analysis," Engineering Applications of Artificial Intelligence, vol. 49, pp. 61–73, 2016.

[19] S. Li, F. Li, and S. Xie, "Do online reviews have different effects on consumers' sampling behaviour across product types? Evidence from

the software industry," Journal of Information Science, p. 016555152096539, 2021.

[20] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093–1113, 2014.

[21] M. Salehan and D. J. Kim, "Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics," Decision Support Systems, vol. 81, pp. 30–40, 2016.

[22] F. Zhou, J. R. Jiao, X. J. Yang, and B. Lei, "Augmenting feature model through customer preference mining by hybrid sentiment analysis," Expert Systems With Applications, vol. 89, pp. 306–317, 2017.

[23] 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.

[24] A. Costa, J. Guerreiro, S. Moro, and R. Henriques, "Unfolding the characteristics of incentivized online reviews," Journal of Retailing and Consumer Services, vol. 47, pp. 272–281, 2019.

[25] J. Guerreiro and P. Rita, "How to predict explicit recommendations in online reviews using text mining and sentiment analysis," Journal of Hospitality and Tourism Management, vol. 43, no. July, pp. 269–272, 2020.

[26] X. Li, H. Liu, and B. Zhu, "Evolutive preference analysis with online consumer ratings," Information Sciences, vol. 541, pp. 332–344, 2020.

[27] H. C. K. Lin, T. H. Wang, G. C. Lin, S. C. Cheng, H. R. Chen, and Y. M. Huang, "Applying sentiment analysis to automatically classify consumer comments concerning marketing 4Cs aspects," Applied Soft Computing, vol. 97, p. 106755, 2020.

[28] J. Y. Lee, "User Review Mining: An Approach for Software Requirements Evolution," International journal of advanced smart convergence, vol. 9, no. 4, pp. 124–131, 2020.

[29] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making," Industrial Marketing Management, no. July, pp. 1–15, 2019.

[30] J. Zhang, X. Lu, and D. Liu, "Deriving customer preferences for hotels based on aspect-level sentiment analysis of online reviews," Electronic Commerce Research and Applications, vol. 49, no. March, p. 101094, 2021.

[31] N. Zhang, R. Liu, X.-Y. Zhang, and Z.-L. Pang, "The impact of consumer perceived value on repeat purchase intention based on online reviews: by the method of text mining," Data Science and Management, vol. 3, no. June, pp. 22–32, 2021.

[32] G. Shan, L. Zhou, and D. Zhang, "From conflicts and confusion to doubts: Examining review inconsistency for fake review detection," Decision Support Systems, vol. 144, no. August 2020, p. 113513, 2021.

[33] Y. Zhang, S. Hao, and H. Wang, "Detecting incentivized review groups with co-review graph," High-Confidence Computing, vol. 1, no. 1, p. 100006, 2021.

[34] Z. He, L. Zheng, and S. He, "A novel approach for product competitive analysis based on online reviews," Electronic Commerce Research, 2022, doi:10.1007/S10660-022-09534-Y.

[35] M. Alzate, M. Arce-Urriza, and J. Cebollada, "Mining the text of online consumer reviews to analyze brand image and brand positioning," Journal of Retailing and Consumer Services, vol. 67, no. April, p. 102989, 2022.

[36] J. Yi and Y. K. Oh, "The informational value of multi-attribute online consumer reviews: A text mining approach," Journal of Retailing and Consumer Services, vol. 65, no. July 2020, p. 102519, 2022.

[37] H. Deng, D. Ergu, F. Liu, Y. Cai, and B. Ma, "Text sentiment analysis of fusion model based on attention mechanism," Procedia Computer Science, vol. 199, pp. 741–748, 2021.

[38] C. C. Chen and Y. De Tseng, "Quality evaluation of product reviews using an information quality framework," Decision Support Systems, vol. 50, no. 4, pp. 755–768, 2011.

[39] J. Mackiewicz and D. Yeats, "Product review users' perceptions of review quality: The role of credibility, informativeness, and readability," IEEE Transactions on Professional Communication, vol. 57, no. 4, pp. 309–324, 2014.

[40] K. Z. K. Zhang, S. J. Zhao, C. M. K. Cheung, and M. K. O. Lee, "Examining the influence of online reviews on consumers' decision-making: A heuristic-systematic model," Decision Support Systems, vol. 67, pp. 78–89, 2014.

[41] R. Filieri, "What makes an online consumer review trustworthy?," Annals of Tourism Research, vol. 58, pp. 46–64, 2016.

[42] G. Shan, D. Zhang, L. Zhou, L. Suo, J. Lim, and C. Shi, "Inconsistency investigation between online review content and ratings," in Twenty-fourth Americas Conference on Information Systems, AMCIS 2018, 2018, pp. 2–11.

[43] S. Jang, J. Chung, and V. R. Rao, "The importance of functional and emotional content in online consumer reviews for product sales: Evidence from the mobile gaming market," Journal of Business Research, vol. 130, no. December 2019, pp. 583–593, 2021.

[44] M. S. Nasiri and S. Shokouhyar, "Actual consumers' response to purchase refurbished smartphones: Exploring perceived value from product reviews in online retailing," Journal of Retailing and Consumer Services, vol. 62, no. June, p. 102652, 2021.

[45] L. Zheng, "The classification of online consumer reviews: A systematic literature review and integrative framework," Journal of Business Research, vol. 135, no. November 2019, pp. 226–251, 2021.

[46] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "Managing marketing decision-making with sentiment analysis: An evaluation of the main product features using text data mining," Sustainability, vol. 11, no. 15, pp. 1–19, 2019.

[47] G. Huang and H. Liang, "Uncovering the effects of textual features on trustworthiness of online consumer reviews: A computational-experimental approach," Journal of Business Research, vol. 126, no. December 2019, pp. 1–11, 2021.

[48] S. Mitra and M. Jenamani, "Helpfulness of online consumer reviews: A multi-perspective approach," Information Processing & Management, vol. 58, no. 3, p. 102538, 2021.

[49] X. Wang, T. Zhou, X. Wang, and Y. Fang, "Harshness-aware sentiment mining framework for product review[Formula presented]," Expert Systems With Applications, vol. 187, no. September 2021, p. 115887, 2022.

[50] M. Petrescu, K. O'Leary, D. Goldring, and S. Ben Mrad, "Incentivized reviews: Promising the moon for a few stars," Journal of Retailing and Consumer Services, vol. 41, pp. 288–295, 2018.

[51] M. Zhang, X. Wei, and D. D. Zeng, "A matter of reevaluation: Incentivizing users to contribute reviews in online platforms," Decision Support Systems, vol. 128, p. 2022, 2020.

[52] B. Anastasiei, N. Dospinescu, and O. Dospinescu, "Understanding the adoption of incentivized word-of-mouth in the online environment," Journal of Theoretical and Applied Electronic Commerce Research, vol. 16, no. 4, pp. 992–1007, 2021.

[53] X. Liu, "Full-Text Citation Analysis : A New Method to Enhance," Journal of the American Society for Information Science and Technology, vol. 64, no. July, pp. 1852–1863, 2013.

[54] S. G. Kanakaraddi, A. K. Chikaraddi, N. Aivalli, J. Maniyar, and N. Singh, "Sentiment Analysis of Covid-19 Tweets Using Machine Learning and Natural Language Processing," in Proceedings of Third International Conference on Intelligent Computing, Information and Control Systems, 2022, pp. 367-379.