In Search of Plagiarism Behaviors: An Empirical Study of Online Reviews

Zhuolan Bao(✉) and Michael Chau

School of Business, Faculty of Business and Economics,
The University of Hong Kong, Hong Kong, China
mbao@connect.hku.hk, mchau@business.hku.hk

Abstract. Plagiarism has been a common phenomenon in the online environment. But the plagiarism phenomenon in online product reviews has been an uninvestigated issue, in spite of various studies in the product review domain. Readers of the plagiarized review could obtain the needed information for understanding the quality or other features of the products through reviews. Besides, the positivity or negativity of the plagiarized review might have different influences on people's perception towards the target products. In this study, we first probe into the current situation of online review plagiarism. Then drawing on the frameworks of attribution theory, in order to examine the effects of plagiarism, we develop and test a model of reviews' diagnosticity perception. The moderation effects of review rating and sentiment are also investigated on the relationship of review plagiarism and review perception. A research plan to analyze reviews collected from Amazon.com is discussed and the results are expected to shed lights upon the understanding of plagiarism behaviors for online reviews, contributing to both theoretical and managerial implications.

Keywords: Online product reviews · Plagiarism · Diagnosticity perception · Review helpfulness · Attribution theory

1 Introduction

With the development of electronic markets, there are an increasing number of people purchasing products online. The vast amount of online product reviews impact people's decision making process of online shopping behaviors [1].

Qualitative feedback in online marketplaces provides a signal of the seller as well as the product for potential buyers who pay attention to the signals when they make purchase decisions [2, 3]. Past studies suggest multiple functions of product reviews. For example, online reviews reflect seller reputation and boost product sales [4, 5]; they also help inform future consumers and reduce product uncertainty in the purchase experience [6, 7].

Review feedback could also have diagosticity value on potential buyers [8]. On an e-market platform, review helpfulness rating can reflect the review diagnosticity [9]. In understanding people's perception and decision making process, researchers are driven to investigate the causal relationship between user-generated review features

M. Chau et al. (Eds.): PAISI 2015, LNCS 9074, pp. 102–112, 2015.
DOI: 10.1007/978-3-319-18455-5_8

and review helpfulness. Earlier studies addressed review length, rating valence, review volume and reviewer characteristics as the determinants of a helpful review [9, 10]. Recently, more researchers are digging into the role of review content in influencing review helpfulness. The characteristics of reviews, such as readability, subjectivity [11], discrete emotions [12] and emotion intensity [13], have been taken into consideration.

However, despite the quality of reviews, the mechanism of voluntary feedback information provides incentives and opportunities for people to post free- ride, and possibly "Pollyanna"(disproportionately positive) information [4], hence influencing people's judgment to the product.

Plagiarism has been intensively studied in academic and business contexts, raising a number of moral, legal and ethical concerns [14-16]. Following prior work, we define plagiarism through the Merriam-Webster dictionary entry – "to commit literary theft: present as new and original an idea or product derived from an existing source" [17].

In our research, our focal issue is in the scope of online markets, the plagiarized reviews, which are reviews copied from reviews already posted by others. Since shorter reviews are more likely to be similar to each other, and providing less information as well as power to affect perceptions, we limit our study to the plagiarism behaviors that appear in longer reviews. David and Pinch [18] studied the incentive of people posting plagiarized reviews on the internet. Their empirical findings suggest that the numerous cases of review re-use are either to promote the sales of a specific item or to increase the reviewer's own credibility.

One might argue that in online markets, customers don't mind too much about the plagiarism behaviors for that online reviews are not as important as academic publications. But since online opinion leaders are becoming more influential and people care more about their online reputation [19], the paper, by studying the dishonest behaviors, also serves the objective and intend to quantify the impact of plagiarism on potential customers of e-markets.

In our research, we go beyond the search on the antecedents to the plagiarism behaviors, but the consequences it would make to influence people's judgment when purchasing products online.

For example, when people are going to buy a camera online, they need to read prior buyers' experience of the camera, paying more attention to the features that they care about, such as lens, sensor, megapixels and battery. If they find some of the reviews for a particular camera are exactly the same or of great similarity, all talking about "Great optimal zoom, long battery life, lightweight and fun to use", they might stop for a moment and think what causes the coincidence. The process of perception would influence their judgment of the product reviews, hence change their evaluation of the camera. Also, they might argue that the sentiment of the review in the example would also make a difference. The perception of the plagiarized review's value may vary when the review is positive or negative.

The influence of plagiarism might be neglected if it is rare to see. But evidence has shown that many reviews are not authentic, that users are applying various techniques to game the system, and that this phenomena might be widespread [18]. Since there has been little research on review plagiarism in the domain, this research addresses the issue by revealing the existence and examining the impact on the diagnosticity judgment of potential customers.

Given the need for studying plagiarism behaviors and potential of customer reviews, we draw on attribution theory and on past research to develop a model of understanding the impact of plagiarized reviews. We then empirically test the model using actual review data from Amazon.com. The analysis gives rise to a better understanding of the consequences of plagiarism behaviors in online purchase decision. Finally, we conclude with a discussion of the managerial implications of the research.

2 Theoretical Foundation and Hypotheses

2.1 Plagiarism Behaviors

There are no universal standards for posting information online, so digital information could be produced under low costs and may be easily plagiarized [20, 21]. In online markets, people post product reviews for many reasons, such as the concerns for other customers, self-enhancement, and so on [22]. The plagiarism behaviors are free of regulation for online markets. For example, Amazon.com states in their policy that they give users complete freedom in posting reviews and do not intervene in the process. The empirical findings of David and Pinch [18] suggest that the numerous cases of review re-use result from certain beneficial causes. However, the impact of this phenomenon is left to be studied.

The attribution theory provides a relevant foundation to address the effects of plagiarism behaviors in the online review context. The common idea of attribution theory is that people interpret behavior in terms of its causes and these interpretations play an important role in determining reactions to the behavior [23, 24].

One of the typical antecedents of attribution is motivation. If an action affects the benefits of the perceivers, there would be greater likelihood that people will infer from it [25]. The motivation for online customers to read reviews is for gathering valuable information to diagnose the quality of the target products. People expect the product reviews to be real and honest, so that they can make smart purchase decisions. The quality and honesty of volunteered review content will influence the benefits of buyers [26].

However, plagiarized reviews, verbatim or with variations, hinder people from obtaining the true knowledge or experience of prior buyers, therefore, inferences of dishonesty are made. Skowronski and Carlston [27] found that dishonest behaviors on one occasion lead to a general dishonesty perception of a person for that the dishonesty is regarded stable. In the case of online reviews, the plagiarism behavior indicates a poor reputation of the reviewer and further be diagnosed as unhelpful. Given these inferences, plagiarized reviews are more likely to be neglected, abandoned or voted as unhelpful.

The discussion suggests the following set of testable hypotheses:

Hypothesis 1:

- H1 - There is plagiarism phenomenon in posting review content on online markets.

Hypothesis 2:

- H2a - A plagiarized review is less likely to be voted.
- H2b - A plagiarized review is less likely to be rated helpful.

2.2 Review Sentiment and Rating Valence

There are several explanations for the positivity, negativity and extremeness biases in the context of online markets, such as frequency-as-information, cognitive processing theory and attribution-based frequency theory. According to the frequency-as-information, rarer information indicates more informative value [28]. Since in online marketplaces, positive reviews outnumber negative reviews [29], negative reviews are perceived more diagnostic and more valuable. From the perspective of cognitive processing, as the arousal level impact the processing capacity for elaboration upon a persuasive message, the peripheral cues are more influential on the persuasion effects [30]. Therefore, potential buyers perceive higher endorsement of persuasion by reviews with a high arousal level than ones with moderate. By attribution-based frequency theory, given that social norms make positive information more prevalent [31], the positive information is less attributed to the underlying stimulus and is therefore less influential [32].

Attribution theory also provides a rationale for differences in perceiving positive plagiarized reviews and negative plagiarized reviews. The interaction between prior beliefs and the new information involves sequential processes of causal inferences [23]. When evaluating persuasive messages, consumers assess the extent to which the communication is due to intrinsic (the endorser's liking to the product) versus extrinsic (monetary) incentives for the endorser [32, 33]. Negative plagiarized reviews are rarer and generally perceived as more influential, therefore, comparing with positive plagiarized reviews, negative plagiarized reviews are more likely to be attributed to the intrinsic incentives of the reviewer and diagnosed as helpful. Since the positivity and negativity of product reviews can be observed from either numeric ratings or the sentiment of review content, we take both of them into consideration.

Together, among the plagiarized reviews, review sentiment and ratings shift people's belief about the cause of the plagiarized reviews, hence perceiving different levels of diagnosticity value of the plagiarized reviews. Therefore, our research also proposes the following testable hypotheses.

Hypothesis 3:

- H3 – A plagiarized review with positive sentiment is less likely to be rated helpful than a plagiarized review with negative sentiment.

Hypothesis 4:

- H4 – A plagiarized review with a high numeric rating is less likely to be rated helpful than a plagiarized review with a low numeric rating.

Fig. 1 summarizes our theoretical model.

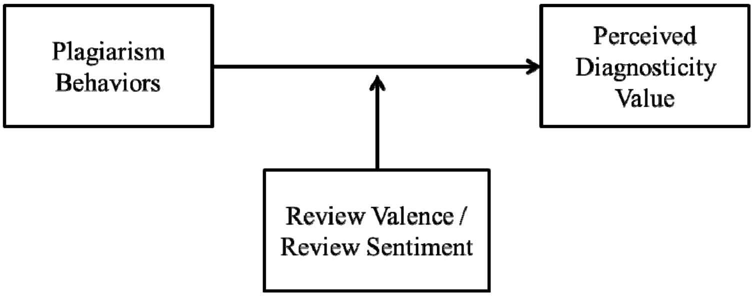

Fig. 1. Theoretical framework

3 Research Methodology

3.1 Data Collection

The data we plan to use to test the hypotheses are actual Amazon.com reviews collected by the Stanford Network Analysis Project (accessible at http:// snap.stanford.edu/index.html). The raw data span a period of 18 years, including around 35 million reviews up to March 2013. Reviews include product and user information, score ratings, helpfulness votes, and a plaintext review. Since the raw data were crawled from online pages, we first filter out the invalid and redundant records. We choose products of books, electronics and music for our study, as the products are more of experience attribute and their reviews often contain expert knowledge and true experience of the reviewers [34].

We further select our data set with several constraints. We select products which were launched between three months in 2012 and all the reviews of the products within one year after their launch. Then we choose products that have more than 20 proper reviews, for that products with few reviews indicate fewer people have left their experience, which offers insufficient opportunity to make comparisons between plagiarized and unplagiarized reviews for a product.

For each selected product, we obtain all the posted reviews. Each review contains the following data of the review.

- The star rating that the reviewer gave to the product.
- The price of the product being reviewed.
- The total number of people who voted to the helpfulness level.
- The number of people who voted that the review was helpful.
- The plain text of the review.

At last, we obtain our final data set while excluding products of which all the reviews did not have anyone voted for helpfulness.

3.2 Variable Operationalization

We operationalized the variables of our model using the Amazon.com data set. We used two dependent variables, review voting and review helpfulness, as adapted from the past research [9, 35]. We used a binary variable to measure review voting, representing whether a review received any votes for being helpful. A value of "0" denotes that the review received no votes, and "1" denotes that the review received at least one vote (VOTING). For the second dependent variable, review helpfulness, we measured it by the proportion of helpful votes received (HELPFULPERCENT).

The explanatory variables are review plagiarism, review sentiment, and score rating.

We develop a language-processing tool to detect the plagiarized reviews in the data set. The tool only applies to detecting reviews with more than 10 words, since reviews with few words are more likely to be similar with each other, such as "Good product" and "Good purchase experience", which are less meaningful for our study. Also, short reviews are often regarded as low-quality reviews, which probably lead to a lower level of helpfulness perception and mix the effects of plagiarism and short reviews. The review plagiarism is measured by a ratio variable representing proportion of plagiarized sentences of a review (PLAGIARISM).

Review sentiment is measured by a dictionary provided by the Linguistic Inquiry and Word Count (LIWC)[1]. LIWC is widely adopted in fields of psychology and linguistic as a text-mining tool. It was developed by Pennebaker, Francis [36] and designed to calculate the degree to which people use different categories of words across a wide array of texts. We calculated the positive/negative sentiment as the proportion of positive/negative words appeared in each review (POSITIVE_SENTIMENT/ NEGATIVE_SENTIMENT).

1 http://www.liwc.net/howliwcworks.php#index2

Table 1. Descriptions of variables

Variable	Description
PLAGIARISM	The proportion of plagiarized sentences in a review
POSITIVE_SENTIMENT	The proportion of positive words appeared in each review
NEGATIVE_SENTIMENT	The proportion of negative words appeared in each review
RATING	The star rating of a review
WORD_COUNT	The number of words in a review
READABILITY	The Gunning Fog Index of a review
UNIQUENESS	The uniqueness of a review's textual content
TOTAL_VOTES	The total number of votes on a review's helpfulness
PRICE	The price of product being reviewed
VOTING	Whether a review receives any vote for being helpful
HELPFULNESS	The proportion of helpful votes received

Score rating is the star rating of a review, representing the reviewer's evaluation towards the certain product (RATING).

We also include word count, readability, uniqueness, total number of votes and product price as control variables. Word count is measured by the number of words in the review (WORD_COUNT). We measure the readability by the Gunning Fog Index, which considers the number of complex words in its formula in order to estimates the years of formal education needed to understand the text (READABILITY). A review with a higher Gunning Fog Index value is more difficult to understand [37]. Researchers have found the uniqueness or innovativeness of product reviews positively influences review diagnosticity [38]. Since plagiarized reviews, by definition, provide less uniqueness information of the products, to avoid the explanation that reviews were perceived less helpful because of its ordinary instead of its nature of plagiarism, we include the uniqueness of review content as a control variable (UNIQUENESS). We also include the number of total votes (TOTAL_VOTES) and product price (PRICE) in the control variables in order to eliminate the effect of the number of voters and the product value on the perception of review diagnosticity. Table 1 summarizes the variables in our study.

3.3 Analysis Method

In Hypothesis 1, we want to demonstrate the existence of review plagiarism phenomenon. We expect a common existence of plagiarism in reviews which contain more than 10 words. Therefore, we expect that the variable PLAGIARISM is positive for reviews with more than 10 words. We use the statistical results of review plagiarism detection to test Hypothesis 1.

The regression analysis in the second stage follows the approach of Mudambi and Schuff [9], Yin et al. [12] and Kuan et al. [35] by adopting Probit regression and Tobit regression.

The model of our empirical study is given by the specifications below.

$$\text{Voting} = \beta_1 * \text{PLAGIARISM} + \beta_2 * \text{POSITIVE\_SENTIMENT} + \beta_3 * \text{NEGATIVE\_SENTIMENT} + \beta_4 * \text{RATING} + \beta_5 * \text{WORD\_COUNT} + \beta_6 * \text{READABILITY} + \beta_7 * \text{UNIQUENESS} + \beta_8 * \text{TOTAL\_VOTES} + \beta_9 * \text{PRICE} + \xi \quad (1)$$

$$\text{Helpfulness} = \beta_1 * \text{PLAGIARISM} + \beta_2 * \text{POSITIVE\_SENTIMENT} + \beta_3 * \text{NEGATIVE\_SENTIMENT} + \beta_4 * \text{RATING} + \beta_5 * \text{WORD\_COUNT} + \beta_6 * \text{READABILITY} + \beta_7 * \text{UNIQUENESS} + \beta_8 * \text{TOTAL\_VOTES} + \beta_9 * \text{PRICE} + \xi \quad (2)$$

$$\text{Helpfulness} = \beta_1 * \text{PLAGIARISM} + \beta_2 * \text{POSITIVE\_SENTIMENT} + \beta_3 * \text{NEGATIVE\_SENTIMENT} + \beta_4 * \text{RATING} + \beta_5 * \text{PLAGIARISM} * \text{POSITIVE\_SENTIMENT} + \beta_6 * \text{PLAGIARISM} * \text{NEGATIVE\_SENTIMENT} + \beta_7 * \text{WORD\_COUNT} + \beta_8 * \text{READABILITY} + \beta_9 * \text{UNIQUENESS} + \beta_{10} * \text{TOTAL\_VOTES} + \beta_{11} * \text{PRICE} + \xi \quad (3)$$

$$\text{Helpfulness} = \beta_1 * \text{PLAGIARISM} + \beta_2 * \text{POSITIVE\_SENTIMENT} + \beta_3 * \text{NEGATIVE\_SENTIMENT} + \beta_4 * \text{RATING} + \beta_5 * \text{PLAGIARISM} * \text{RATING} + \beta_6 * \text{WORD\_COUNT} + \beta_7 * \text{READABILITY} + \beta_8 * \text{UNIQUENESS} + \beta_9 * \text{TOTAL\_VOTES} + \beta_{10} * \text{PRICE} + \xi \quad (4)$$

For specification (1), we use Probit regression since the dependent variable is binary. For specification (2)-(4), we consider Tobit regression appropriate, for that the dependent variable, i.e., the reviews' helpfulness percentage, is constructed as a ratio and the value is bounded in range according to its nature [39].

To test Hypothesis 2, we expect that review plagiarism has a negative effect on review voting and review helpfulness. We include two moderation effects in Hypothesis 3. Specifications (3) and (4) are to test the moderation effects of review sentiment and review rating respectively, on the relationship between plagiarism and helpfulness level.

As a robustness check, we apply propensity score matching method to evaluate the moderation effects. A propensity score is the probability of a unit being assigned to a particular treatment given a set of observed covariates [40]. In order to test Hypothesis 3, for each product, we classify our reviews within the product into three groups: plagiarized with positive sentiment (Group A), plagiarized with negative sentiment

(Group B) and plagiarized with neutral sentiment (Group C). We will find the similar plagiarized review units of two groups by matching their propensity scores of all the other variables and use group C as a baseline. According to our hypothesis, we consider the average sentiment of the review content as the treatment. Then we estimate the effects of treatment by comparing the helpfulness levels of the matched review units. Similarly, to test Hypothesis 4, we divide our reviews into different groups by their score ratings.

4 Potential Contribution and Conclusion

This paper seeks to investigate the role of plagiarized review in online marketplaces. And our study contributes to both theory and practice. We develop our hypotheses with the framework of attribution theory and found negativity bias among the plagiarized reviews. The result will be consistent with the notion of negativity bias commonly found in online product reviews [32, 41]; however, the explanation of the bias in our research will complement the evaluation of the diagnosticity value in product reviews.

Besides, the results of our research will quantify and shed light upon the impact of plagiarism in the contexts other than the contexts which have already been intensively studied, such as business and academic research. Our findings will be influential for future studies of online reputation. For online customers who eager to become an opinion leader in cyberspace, the paper can also offer some insights.

Moreover, our findings of plagiarized review will provide incentives for sellers or online markets to detect the plagiarism behaviors of their product reviews, as they not only lead to less diagnostic understanding of their products, but also cast doubt about the causes of the existing reviews, especially in a searching environment of information overload. Also, in terms of the moderation effects, since negative plagiarized reviews provide informative messages, by analyzing them, marketers can recognize their defects and make adjustment accordingly.

There are some limitations in our study. First, in order to highlight the role of online reviews, we select products with more of experience attributes, so that potential buyers have the incentives to learn from review messages for their own benefits. However, reviews for search products are suggested influential in many studies, later research can address this issue with a larger scope of products.

Second, in our study, we will only test the effects of plagiarism behaviors within the same product, but the type of plagiarism behaviors could be different. For example, the review duplication could happen to relevant products.

Third, little is mentioned about the antecedents of plagiarism behaviors in this paper. Due to the importance and impact of the phenomenon, more investigation on the motivation of plagiarism behaviors should be done in future studies.

Acknowledgments. This research is supported in part by a HKU Seed Funding for Basic Research grant. We thank Reno Ha of the University of Hong Kong for his help in data processing and analysis.

We also thank the Stanford Network Analysis Project for making the Amazon data set available for research.

References

1. Aral, S., Ipeirotis, P.G., Taylor, S.J.: Content and context: Identifying the impact of qualitative information on consumer choice. Available at SSRN 1784376 (2011)
2. Ghose, A., Ipeirotis, P.G.: Designing novel review ranking systems: predicting the usefulness and impact of reviews. In: Proceedings of the Ninth International Conference on Electronic Commerce. ACM (2007)
3. Schindler, R.M., Bickart, B.: Perceived helpfulness of online consumer reviews: the role of message content and style. Journal of Consumer Behaviour 11(3), 234–243 (2012)
4. Resnick, P., Zeckhauser, R.: Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. Advances in Applied Microeconomics 11, 127–157 (2002)
5. Dellarocas, C., Gao, G., Narayan, R.: Are consumers more likely to contribute online reviews for hit or niche products? Journal of Management Information Systems 27(2), 127–158 (2010)
6. Dellarocas, C.: The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Management Science 49(10), 1407–1424 (2003)
7. Dimoka, A., Hong, Y., Pavlou, P.A.: On product uncertainty in online markets: Theory and evidence. MIS Quarterly 36 (2012)
8. Pavlou, P.A., Liang, H., Xue, Y.: Understanding and mitigating uncertainty in online exchange relationships: a principal-agent perspective. MIS Quarterly, 105–136 (2007)
9. Mudambi, S.M., Schuff, D.: What makes a helpful online review? A study of customer reviews on Amazon.com. MIS Quarterly 34(1), 185–200 (2010)
10. Forman, C., Ghose, A., Wiesenfeld, B.: Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. Information Systems Research 19(3), 291–313 (2008)
11. Ghose, A., Ipeirotis, P.G.: Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. IEEE Transactions on Knowledge and Data Engineering 23(10), 1498–1512 (2011)
12. Yin, D., Bond, S., Zhang, H.: Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. MIS Quart. (2013)
13. Peng, C.-H., et al.: How and When Review Length and Emotional Intensity Influence Review Helpfulness: Empirical Evidence from Epinions.com (2014)
14. Martin, D.E., Rao, A., Sloan, L.R.: Plagiarism, integrity, and workplace deviance: A criterion study. Ethics & Behavior 19(1), 36–50 (2009)
15. Martin, D.E.: Culture and unethical conduct: Understanding the impact of individualism and collectivism on actual plagiarism. Management Learning 43(3), 261–273 (2012)
16. Schmelkin, L.P., et al.: A multidimensional scaling of college students' perceptions of academic dishonesty. The Journal of Higher Education, 587–607 (2008)
17. Dictionary, M.W.O.L.: What is plagiarism? (2007)
18. David, S., Pinch, T.J.: Six degrees of reputation: The use and abuse of online review and recommendation systems. Available at SSRN 857505 (2005)
19. Li, F., Du, T.C.: Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs. Decision Support Systems 51(1), 190–197 (2011)

20. Metzger, M.J.: Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. Journal of the American Society for Information Science and Technology 58(13), 2078–2091 (2007)
21. Rieh, S.Y.: Judgment of information quality and cognitive authority in the Web. Journal of the American Society for Information Science and Technology 53(2), 145–161 (2002)
22. Yoo, K.H., Gretzel, U.: What motivates consumers to write online travel reviews? Information Technology & Tourism 10(4), 283–295 (2008)
23. Kelley, H.H., Michela, J.L.: Attribution theory and research. Annual Review of Psychology 31(1), 457–501 (1980)
24. Kelley, H.H.: Attribution theory in social psychology. In: Nebraska Symposium on Motivation. University of Nebraska Press (1967)
25. Jones, E.E., Davis, K.E.: From acts to dispositions the attribution process in person perception. Advances in Experimental Social Psychology 2, 219–266 (1965)
26. Otterbacher, J.: Managing information in online product review communities: a comparison of two approaches. In: Proc. 16th European Conference on Information Systems (ECIS 2008). Association for Information Systems (2008)
27. Skowronski, J.J., Carlston, D.E.: Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. Journal of Personality and Social Psychology 52(4), 689 (1987)
28. Skowronski, J.J., Carlston, D.E.: Negativity and extremity biases in impression formation: A review of explanations. Psychological Bulletin 105(1), 131 (1989)
29. Hu, N., Zhang, J., Pavlou, P.A.: Overcoming the J-shaped distribution of product reviews. Communications of the ACM 52(10), 144–147 (2009)
30. Sanbonmatsu, D.M., Kardes, F.R.: The effects of physiological arousal on information processing and persuasion. Journal of Consumer Research, 379–385 (1988)
31. Wojciszke, B., Brycz, H., Borkenau, P.: Effects of information content and evaluative extremity on positivity and negativity biases. Journal of Personality and Social Psychology 64(3), 327 (1993)
32. Chen, Z., Lurie, N.H.: Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth. Journal of Marketing Research 50(4), 463–476 (2013)
33. Folkes, V.S.: Recent attribution research in consumer behavior: A review and new directions. Journal of Consumer Research, 548–565 (1988)
34. Huang, P., Lurie, N.H., Mitra, S.: Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods. Journal of Marketing 73(2), 55–69 (2009)
35. Kuan, K.K., et al.: What Makes a Review Voted? An Empirical Investigation of Review Voting in Online Review Systems. Journal of the Association for Information Systems 16(1), 48–71 (2015)
36. Pennebaker, J.W., Francis, M.E., Booth, R.J.: Linguistic inquiry and word count: LIWC [Computer software]. LIWC.net, Austin (2007)
37. Goes, P.B., Lin, M., Yeung, C.-M.A.: "Popularity Effect" in User-Generated Content: Evidence from Online Product Reviews. Information Systems Research (2014)
38. Pan, Y., Zhang, J.Q.: Born unequal: a study of the helpfulness of user-generated product reviews. Journal of Retailing 87(4), 598–612 (2011)
39. Greene, W.H.: Econometric analysis. Pearson Education, India (2003)
40. Rosenbaum, P.R., Rubin, D.B.: The central role of the propensity score in observational studies for causal effects. Biometrika 70(1), 41–55 (1983)
41. Wu, P.F.: In Search of Negativity Bias: An Empirical Study of Perceived Helpfulness of Online Reviews. Psychology & Marketing 30(11), 971–984 (2013)