Mobile Applications Online Review and Rating Research: Heuristic Decision

Israt Jahan[1, 2], Sayeed Sajal[2] and Kendall Nygard[3]

[1]Department of Computer Science, Nueta Hidatsa Sahnish College
220 8th Ave E New Town, ND 58763, ijahan@nhsc.edu

[2]Department of Math & Computer Science, Minot State University
500 University Ave W, Minot, ND, 58707, sayeed.sajal@ndus.edu

[3]Department of Computer Science, North Dakota State University
1340 Administration Ave, Fargo, ND 58105, kendall.nygard@ndsu.edu

Abstract

With the incredible growth and potential of online consumer reviews and ratings, online reviews of different applications are now playing a significant role in consumer attitude and buying behaviors. These online stores help users discover the required applications as well as leave a review. For developers, it is very important to get positive reviews and high ratings to ensure that an application has a feasible future. Ratings and reviews add value to both the developer and potential new users by providing a crowd-sourced indicator of application quality. A couple of studies showed that responding to a review often has a positive effect on the rating that is given by the user to an app also in uses by the user. This study reviewed and analyzed in Google Inc related to review and rating data. This study presented a brief fusion of research by investigating content-related characteristics of different applications online reviews in different market segments. In summary, we developed a model to determine the top 2 factors that motivate buyers to buy mobile applications.

1. Introduction

Every day people browse different mobile applications and download applications according to their needs and their choice and also based on online customer reviews and rating. Now a day's online consumer reviews and rating have become a significant source of information that assists consumers to make a purchase decision about similar products, which may be confusing. This developed because of the advance of the internet and electronic commerce [1]. The review may be positive or negative or neutral in nature. But with the help of review and rating, important information can be obtained which can guide and help the customer to choose and buy the right product among the same category products. In the collected data we find the argument feature that is informative of application review which is the systematic factor that influences consumer or user to purchase the application or download it as free and based on this people rate the application. On the other hand, we also find the perceived quantity of reviews that have a direct influence on purchase an intention.

2. Background

A couple of research paper published based on review and rating. It could be a consumer products reviews and ratings, restaurants reviews and ratings or any kind of application like health app reviews and rating, etc. Among all of them, it shows that online review has a powerful influence on purchase power. An industry survey report from Channel Advisor [2] found that 90% of shoppers read the reviews, and 83% among them believe that reviews and influence their purchase decision. In another study, Duan et al. [3] found that there is no impact on movie revenues based on rating and reviews. According to the meta-analyses, the number of volume of reviews are the most important features influencing sales and vary the attitudes of taking the decision [4-6]. The more positive reviews mean a good application which increases sales and attitudes whereas negative reviews reduce them [7]. It's reported that negative review carries more weight than positive reviews on consumer attitude for buying decision.

According to Cheung and Thadani [8], online reviews generally focus on two types of online

contexts: online consumer review sites (e.g., Epinions.com and Tripadvisor.com) and online shopping sites (Amazon.com and eBay.com). The impacts of online reviews may vary from online contexts.

A systematic review is often applied in the biomedical or healthcare context, but it can be applied in any field of research. Dual-process theories have extensively explained how individuals react to a piece of information. There are two most widely used dual-process theories. These are Elaboration likelihood model as ELM and the Heuristic-systematic model As HSM (Chaiken, 1980; Davis & Tuttle, 2013; Petty & Cacioppo, 1986; Yang, 2015; Zha et al., 2018). Information cues are proposed by both the model, such as heuristic (peripheral) cues and systematic (central) cues, are the keys to initiate information processing (Chaiken & Maheswaran, 1994; Petty & Wegener, 1999). The length of a message or the appearance of the spokesperson which we called information cues; influence a person's judgment of the information (Chen & Chaiken, 1999; Petty & Cacioppo, 1986). In contrast, individuals, who are motivated to process the information, are more likely to use systematic or central cues. In this condition, systematic or central information cues, such as the strength of the argument, affect one's evaluation of the message (Eagly & Chaiken, 1993; Petty & Cacioppo, 1984).

3. Research Methodology

Figure 1 shows us the steps that we follow to do the research.

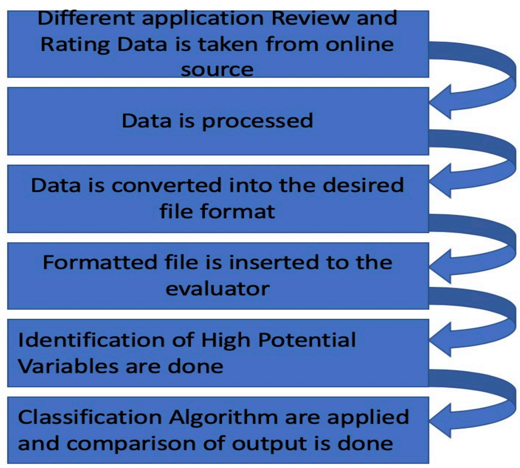

Figure 1: Flowchart of proposed work

A. Data Collection

Data set is downloaded from Kaggle data [20] repository which contains 13 attributes and 9368 instances. The attributes present in the data set are; Apps, Category, Rating, Reviews, Size, Installs, Types, Content Rating, Gens, Last updated, Current version, Android version.

B. Data Pre-processing

Data transformation or clean the data is called data pre-processing. In this research, Kaggle's data set is used for data association mining. It is necessary to identify the features which will support the higher accuracy in classification. According to the No Free lunch theorem, there is no fixed algorithm to provide high accuracy. Data pre-processing is necessary to improve machine learning. Classification and clustering accuracy is mostly dependent on the proper representation of data. Correlation-based feature selection is used to reduce the number of features [20].

C. Machine Learning Algorithm

Machine learning (ML) means, make the machine learn by processing the data with various machine learning algorithm and statistical models that describe the relationship of the data. In most of the cases, deep learning provides better accuracy [14-18]. For any application, it is important to apply a few machine learning algorithms to find out the best-suited model. Machine learning algorithms can be grouped under Bayes, Rule-Based, Neural network and Decision tree.

- **Naïve Bayes**

Naïve Bayes describes the conditional probability of an event based on previous knowledge that is related to the event. A given problem instance to be classified, represented by a vector representing some *n* features (independent variables), it assigns to this instance probabilities for each of the possible outcomes or *classes* [9,10]. Following equation mathematically describe the Bayes' theorem;

$$P(A\backslash B) = P(B\backslash A)\, P(A)/P(B)$$

where A and B are events and $P(B) \neq 0$.

- $P(A\backslash B)$ is a conditional probability where of event A occurring given that B is true.
- $P(B\backslash A)$is also a conditional probability where of event B occurring given that A is true.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B independently of each other which is known as the marginal probability[9,10].

Naïve Bayes theorem is the best machine learning algorithm when the features are independent of one another.

➢ **Decision Tree**

A decision tree is a decision support tool that is used as a tree of the decision and their possible outcomes. By finding the optimum way to arrange the various nodes, the decision tree arrives. Information gain or gain ratio are ways to identify the best dataset node. The decision tree model which uses information gain is ID3 and the gain ratio is J48 [11].

➢ **Sequential minimal optimization (SMO)**

SMO is an iterative algorithm for solving the optimization problem. SMO breaks this problem into a series of smallest possible sub-problems, which are then solved analytically. To increase operational speed by use of a single threshold value SMO is a useful technique. [12-13]

4. Experiment

The dataset acquired from Kaggle [19], first of all, we do the data preprocessing. In the data set it was found that some attributes were redundant, Apps, Category, Installs, Types, Content Rating, Gens, Last updated, Current version, Android version where some of them represent the same information. To avoid redundancy of attributes only one the representation was kept.

For simplicity, any review of 4 stars and above (rating:>4) is assigned a positive sentiment, while 2 and below (rating:<2) is considered negative. And then chose the attributes that we needed for our study. In the data set, there are so many different categories and each category has lots of different apps. To make it simple we chose two different categories; HEALTH_AND_FITNESS and FOOD_AND_DRINK.

To get better results from the most dataset, Machine learning algorithm such as J48, NaïveBayes, and Multilayer perceptron are used. So, in this study, we used the same classifications by using the algorithm in WEKA which is a free online data mining tool published by Waikato University. The data set is preprocessed, Features are selected, trained and tested using WEKA. The algorithm found to get a better result is SMO where its precession and Correctly Classified Instances are high.

By observation we were able to see that factors affecting consumer are both number of reviews and rating and after that paid or free version. But during the decision-making process, whether download the app or not, it found that it depends on the positive or negative review. People like mostly those apps which have highly positive reviews though the number of reviews is less with those positive reviews rate them highly.

5. Results

The study compares the results of the proposed design to the existing based on Accuracy, Precision, Recall, ROC Area. Here in both categories, we compared all the parameters in different machine learning techniques.

The results according to the HEALTH_AND_FITNESS Apps are shown in Table 1.

Technique	Correctly Classified Instances	TP Rate	FP Rate	Precision	ROC Area
Naïve Bayes	62.96%	0.621	0.167	0.989	0.786
J48	95.96%	1	1	0.96	0.429
SMO	96.63%	1	0.833	0.966	0.583

The results according to the FOOD_AND_DRINK Apps are shown in Table 2.

Technique	Correctly Classified Instances	TP Rate	FP Rate	Precision	ROC Area
Naïve Bayes	79.63%	0.811	1	0.977	0.443
J48	98.15%	1	1	0.981	0.157
SMO	98.15%	1	1	0.981	0.5

6. Conclusion

For examining the influence of online reviews on consumers' decision making, we conduct this study in the context of online consumer review and rating on the basis of paid and free apps. Google Inc.'s Google Play has offered developers a single, low cost, and powerful distribution mechanism. These online stores help users discover new apps as well as leave a review. By providing a crowd-sourced indicator of app quality. Ratings and reviews add value to both the developer and potential new users. It is highly appreciated and important to get positive reviews and high ratings to ensure that an app has a viable future. Some users tend to leave short but helpful reviews, and the rating, as well as the category, influence the length of a review. Some users leave longer messages which also indicate that they rate an app poorly, and the depth of feedback in certain categories is significantly higher than for others. Drawing upon the literature of dual-process theories, we develop a research model to identify factors that are important to consumers' purchase decision-making. We find that argument quality of online reviews which is characterized by perceived informativeness and perceived persuasiveness, has a significant effect on consumers' purchase intention. In addition, we find that source credibility and perceived quantity of reviews have direct impacts on purchase intention. The positive influences on argument strength.

References

1. Examining the influence of online reviews on consumers' decision-making: A heuristic-systematic model.

2. ChannelAdvisor, 2011 Consumer Survey: Global Consumer Shopping Habits, ChannelAdvisor, 2011.

3. Dellarocas, The digitization of word of mouth: promise and challenges of online feedback mechanisms, Management Science 49 (10) (2003) 1407–1424.

4. J.W. Lian, D.C. Yen, Online shopping drivers and barriers for older adults: age and gender differences, Computers in Human Behavior 37 (2014) 133–143.

5. K. Floyd, R. Freling, S. Alhoqail, H.Y. Cho, T. Freling, How online product reviews affect retail sales: a meta-analysis, Journal of Retailing 90 (2) (2014) 217–232, https://doi.org/10.1016/j.jretai.2014.04.004.

6. N. Purnawirawan, M. Eisend, P. De Pelsmacker, N. Dens, A meta-analytic investigation of the role of valence in online reviews, Journal of Interactive Marketing 31 (2015) 17–27, https://doi.org/10.1016/j.intmar.2015.05.001.

7. Y. Liu, Word of mouth for movies: its dynamics and impact on box office revenue, Journal of Marketing 70 (3) (2006) 74–89

8. C.M.K. Cheung, D.R. Thadani, The impact of electronic word-of-mouth communication: a literature analysis and integrative model, Decision Support Systems 54 (1) (2012) 461–470.

9. Bayes Theorem. Source: https://en.wikipedia.org/wiki/Bayes%27_theorem Access Date: March 27, 2019

10. Lewis, David D. "Naive (Bayes) at forty: The independence assumption in information retrieval." European conference on machine learning. Springer, Berlin, Heidelberg, 1998

11. Decision tree – Wikipedia, Source: https://en.wikipedia.org/wiki/Decision_tree Access Date: March 27, 2019

12. Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993

13. Study of the SMO Algorithm Based on Data Mining in Shot-Term Power Load Forecasting Model,2008

14. Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. Journal of Personality and Social Psychology, 39(5), 752–766.

15. Davis, J. M., & Tuttle, B. M. (2013). A heuristic-systematic model of end-user information processing when encountering IS exceptions. Information & Management, 50(2–3), 125–133.

16. Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. Journal of Personality and Social Psychology, 46(1), 69–81.

17. Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. Advances in Experimental Social Psychology, 19, 123–162.

18. Eagly, A., & Chaiken, S. (1993). The psychology of attitude. Fort Worth, TX: Harcourt, Brace, & Janovich

19. Dataset, Source: https://www.kaggle.com/datasets, Access Date: March 10, 2019

20. D. H. Deshmukh, T. Ghorpade, and P. Padiya, "Improving classification using preprocessing and machine learning algorithms on nslkdddataset," inCommunication, Information & ComputingTechnology(ICCICT), 2015 International Conference on. IEEE, 2015, pp.1-6