International Journal of Research in Engineering, Science and Management
Volume-2, Issue-8, August-2019
www.ijresm.com | ISSN (Online): 2581-5792

Sentiment Analysis of the Customer Reviews and Opinion Verdict on Probabilistic Analysis

Gayika Singh[1], Brij Kishore[2]
[1]*Student, Department of Computer Science, Apex Institute of Engineering and Technology, Jaipur, India*
[2]*Assistant Professor, Dept. of Computer Science, Apex Institute of Engineering and Technology, Jaipur, India*

***Abstract*: This paper is intended to examine and analyze the market trends and the scenarios occurring according to the product sales and reviews and same goes with the sentiments of the users regarding the services provided by the service providers. The use of concept of Knowledge Discovery in Database and Opinion mining has been used to give the feedback which will eventually help the seller/service provider in analyzing market and its need for better products and services.**

***Keywords*: Opinion Mining, KDD, Data Mining, Probabilistic Approach.**

1. Introduction

This paper is postulating the opinion or the sentimental analysis of the reviews by the customers. With the pervasiveness of online discussion forum and social media platform, user generated text containing opinions on some hot issues has increased significantly. Due to the large amount of such emotional reviews and posts on Internet, it is impossible for users to digest such information manually. Therefore, automatically mining opinion from online texts, aiming at discovering user concerned topics and the corresponding opinion, becomes essential. In general, opinion mining aims to extract a quintuple <e, a, s, h, t> from texts, where 'e' is the entity or the target, a is the aspect of the entity e, h is the opinion holder, t is the time when the opinion holder expresses her opinion on the entity 'e', and 's' is the opinion which h holds to the aspect a of the entity 'e' at 't'. For example, opinion mining processes the review text "I bought a new iPhone X today, the screen is great, but the voice quality is poor" and outputs two quintuples <iPhone X, screen, and great, I, today> and <iPhone X, voice quality, poor, I, today>. However, not all the opinion mining tasks need to extract all the five elements in quintuple. For example, sentiment analysis cares more about the sentiment polarity s of the text, stance detection aims to identify the opinion s to the specific target e, and product aspect mining focus on extracting the aspect a and corresponding opinion s from text.

In the last decade, stance detection and product aspect mining have attracted many scholars. Following the general opinion mining framework, stance detection can be formalized as the task of extracting tuple <e, s> ('e' means target and s represents opinion) without considering other elements. Stance detection focuses on detecting the user stance (favor, against) on a particular debate topic or hot debated event. It is similar to sentiment analysis, but with big difference. In specific, sentiment analysis aims to identify the sentiment polarity (positive, negative) of the text while stance detection cares about the stance on the target. For example, the tweet "Jeb bush is the only sane candidate in this republican lineup, I support him" will be assigned positive by sentiment analysis, but extracted with 'against' stance to the topic "Rahul Gandhi as Prime Minister" by stance detection. Research on stance detection can be categorized into four groups based on debate settings, such as congressional floor debates, company-internal discussions, online forums ideological debates and hot-event oriented debates on social media. The latter two are open domain and flexible, therefore more challengeable. For example, the debate forum such as convinceMe.net has a wide range of debate topics from the playful (e.g. 'Cats vs. Dogs', 'Mac vs. PC', 'Superman vs. Batman' and etc) to the ideological (e.g. 'Death penalty', 'Exist God', 'Gay marriage', 'Healthcare' and etc). Furthermore, the participants prefer to use colorful and emotional language to express their viewpoints, such as the tweet "It looks like they like Hilary more... and that plain stupid" related to the hashtag "#2014India election#". Therefore, we concentrate on the prevalent work of stance detection for online debate forums and social media in this survey.

2. Literature Survey

For the opinion mining, the polarity classification, it is the main job in sentiment analysis as well as opinion mining such as neutral, negative and positive polarities in the opinion [4]. For example, of positive opinion, "Coke tastes good" indicates that coke is tasty, a positive statement. However, "Coke drinks are not healthy if consumed in large quantity" indicates that Coke drinks are harmful to body if they drink a lot, a negative statement. As for the neutral statement, "Coke is a carbonated drink". This paper explains the sentiment analysis taxonomy or techniques based on the information from [1][2][3][6]. If you wish to have detailed proposed sentiment analysis taxonomy, refer to this paper [5]. The opinion can be categorized as regular and comparative opinion [7]. Example of regular opinion, "The display quality is crisp." indicates that the aspect of "picture quality” is referred directly and it gives a positive polarity [7]. As for the indirect regular opinion, “After applying the cream, my skin broke out completely.” indicates that “the cream” indirectly state the cream is bad for the skin and hence it gives a negative polarity [7]. As for the comparative opinion, “The processor speed and screen resolution of S6 is better than IPhone 6; however, the metal body of IPhone 6 is more attractive than S6.” indicates S6 has two positive opinions and one negative opinion [7]. It implicitly states the processor One Plus is better than Yureka [7].

According to this author [8], [13], document level classifies the opinion document as giving either of the two of a positive or negative statement. It takes into consideration of the entire document as the basic information unit (discussing about a topic). As for the sentence level [8], [15], it classifies it in positive and negative opinions if a sentence is subjective. Thus, it categorizes sentiment expressed in every sentence. Lastly, the aspect level [8], [14], the people can provide not the same opinions for not the same aspects of the same entity. It categorizes the sentiment with respect to the specific aspects of entities. As for the machine learning [9], refer to Table 1 [10] for types of machine learning techniques are applied in sentiment analysis especially during classification. The lexicon-based approaches are dependent on the obtainability of a sentiment lexicon [6], which it is a group of previously created and known sentiment words. These approaches could be classified into two dissimilar sets: (i) dictionary based, which it is using dictionaries as lexical and (ii) corpus-based, which it is using semantic methods or statistical to search sentiment polarity [6]. An emotion can be defined [11] as “feeling states with physiological, cognitive, and behavioural components.” As for the cognitive structure of emotion you may look into [12] by Ortony, Clore, and Collins. Refer to Yue et al. [16], they constructed the two-dimensional structure of emotions. It has low positive affect, high positive affect, low negative affect and high negative affect to classify the emotions. We follow the definition of opinion or sentiment from where it is represented as a quintuple as in formula (1):

$(e_i, a_{ij}, s_{ijkt}, h_k\ t_t)$……………….(1)

In which e_i is the i^{th} entity, a_{ij} is the j^{th} aspect of i^{th} entity, h_k is the k^{th} opinion holder, t_t is the time when the opinion is expressed, s_{ijkt} is the opinion or sentiment towards the jth aspect of the i^{th} entity from opinion holder h_k at time t_t.

Method	Attributes
Supervised Learning	A group of labeled data that will be learned. Labeled training data is compulsory. It is the common form for the learning.
Unsupervised Learning	A group of unlabeled data that will be learned. In the data independent of class label, it finds the unseen of the relationships. Clustering is the very common form.
Semi-supervised Learning	Unlabeled data and labeled data that will be learned. Only needs a relatively lesser set of labeled data which is added with a big amount of unlabeled data. With the a lot of unlabeled data exist, review spam is perfect for the cases.

3. Proposed work

The proposed algorithm works over the sentiment analysis of customer reviews. The data that has been taken for the analysis and for the testing purposed is being sourced from UC Irvine Machine Learning Repository. The text data is fetched into the MATLAB workspace for the analysis of the review individually, for testing purposes the records that have been taken, are of the cars sales during the period 2016, 2017 and 2018. The reviews are individually taken into the account and is examined on the basis of the keywords used in it and the result is generated about the overall review of the product.

Algorithm steps that are being responsible for the whole process are stated below:

1) Clear the MATLAB workspace.
2) Fetch the data from the excel sheet of the desired year, which is been provided by the user through a dialogue box.
3) Count the number of reviews present in the database for the subsequent year.
4) Generate the frequency rate for the use of the positive words and negative words.
5) Calculate the probabilistic count and features for the entered input dataset.
6) Generate the mass opinion for the input data and its practicality.

4. Result

The proposed algorithm analyses the input sentiments and generates the verdict for each review as shown in figure 1.

Sample Review: ‘My daughter has a 2000 Accent and is wonderful. My wife has driven an Elantra for 5 years and it has been bullet proof. I just purchased an Accent and it drive great.’

```
Out of 33.000000 words 2.000000 are positive and 0 are negative.
Verdict is Positive Review.
>>
```

Fig. 1. Verdict Result for sample text review

5. Conclusion and Future Scope

The proposed work is generic and can used in any industry, the addition of the mathematical modelling to the input dataset enhances the accuracy and also shows the sentiments of the users in a very convenient and briefly. Also the addition of the Machine Learning based algorithms may help this work in generating more and more reliable opinions in the long run as updating database is crucial but also helpful in understanding the changes occurring globally.

References

[1] A. Yadollahi, A. Shahraki and O. Zaiane, "Current State of Text Sentiment Analysis from Opinion to Emotion Mining", ACM Computing Surveys, vol. 50, no. 2, pp. 1-33, 2017.

[2] 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.

[3] B. Liu, Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press, 2015.
[4] S. Almatarneh and P. Gamallo, "A lexicon based method to search for extreme opinions", PLOS ONE, vol. 13, no. 5, pp. 1-19, 2018.
[5] R. Rodrigues, C. Camilo-Junior and T. Rosa, "A Taxonomy for Sentiment Analysis Field", International Journal of Web Information Systems, pp. 00-00, 2018.
[6] N. Silva, L. Coletta and E. Hruschka, "A Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning", ACM Computing Surveys, vol. 49, no. 1, pp. 1-26, 2016.
[7] V. Patel, G. Prabhu and K. Bhowmick, "A Survey of Opinion Mining and Sentiment Analysis", International Journal of Computer Applications, vol. 131, no. 1, pp. 24-27, 2015.
[8] A. D'Andrea, F. Ferri, P. Grifoni and T. Guzzo, "Approaches, Tools and Applications for Sentiment Analysis Implementation", International Journal of Computer Applications, vol. 125, no. 3, pp. 26-33, 2015.
[9] M. Ahmad, S. Aftab, S. Muhammad and S. Ahmad, "Machine Learning Techniques for Sentiment Analysis: A Review", International Journal of Multidisciplinary Sciences and Engineering, vol. 8, no. 3, pp. 27-32, 2017.
[10] M. Crawford, T. Khoshgoftaar, J. Prusa, A. Richter and H. Al Najada, "Survey of review spam detection using machine learning techniques", Journal of Big Data 2:23, vol. 2, no. 1, pp. 1-24, 2015.
[11] C. K. Hsee, E. Hatfield, J. G. Carlson, C. Chemtob, “The effect of power on susceptibility to emotional contagion”, Cognition and Emotion 4, no. 4 pp. 327–340, 1990.
[12] A. Ortony, G. Clore and A. Collins, The Cognitive Structure of Emotions. Cambridge University Press, 1988.
[13] S. Behdenna, F. Barigou and G. Belalem, "Document Level Sentiment Analysis: A survey", EAI Endorsed Transactions on Context-aware Systems and Applications, vol. 4, no. 13, pp. 1-8, 2018.
[14] K. Schouten and F. Frasincar, "Survey on Aspect-Level Sentiment Analysis", IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 813-830, 2016.
[15] A. B. Sayeed, J. Boyd-Graber, B. Rusk, and A. Weinberg, “Grammatical structures for word-level sentiment detection,” In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, pp. 667- 676, 2012.
[16] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, “A survey of sentiment analysis in social media,” Knowledge and Information Systems, pp. 1-47, 2018.