

Sentiment Analysis of Reviews and Ratings in Retail Shop

Suresh Babu P [1], Akilan R [2], Arun Kumar N [3], Nithis kumar T N [4]

[1] Assistant professor, Department of IT, Velammal college of Engineering and Technology, Madurai, Tamilnadu, India

[2] [3] [4] UG Students, Department of IT, Velammal college of Engineering and Technology, Madurai, Tamilnadu, India

Abstract - Sentiment analysis or opinion mining is one of the major tasks of NLP (Natural Language Processing). Sentiment analysis has gain much attention in recent years. In this paper, we aim to tackle the problem of sentiment polarity categorization, which is one of the fundamental problems of sentiment analysis. A general process for sentiment polarity categorization is proposed with detailed process descriptions. Data used in this study are online product reviews. In this paper, we have done sentiment analysis on product reviews to find the most purchased and populated item on each category. And this can be useful for merchant to improve the retail shop trades.

Key Words: Sentiment analysis, Sentiment polarity categorization, Natural language processing, Product reviews, Populated item.

1. INTRODUCTION

Every single day huge amount of information, reviews or opinions are getting stored in the websites of social media or e-services in the form of raw data. Sentiment analysis is also known as “opinion mining” or “emotion Artificial Intelligence” and alludes to the utilization of natural language processing (NLP), text mining, computational linguistics, and bio measurements to methodically recognize, extricate, evaluate, and examine emotional states and subjective information. Sentiment analysis is generally concerned with the voice in client materials; for example, surveys and reviews on the Web and web-based social networks. The basic idea of sentiment investigation is to detect the polarity of text documents or short sentences and classify them on this premise. Sentiment polarity is categorized as “positive”, “negative” or “impartial” (neutral). It is important to highlight the fact that sentiment mining can be performed on three levels as follows:

1. Document-level sentiment classification: At this level, a document can be classified entirely as “positive”, “negative”, or “neutral”.
2. Sentence-level sentiment classification: At this level, each sentence is classified as “positive”, “negative” or unbiased.
3. Aspect and feature level sentiment classification: At this level, sentences/documents can be categorized as

“positive”, “negative” or “non-partisan” in light of certain aspects of sentences/archives and commonly known as “perspective-level assessment grouping”.

The main objective of this paper is to do sentiment analysis of reviews to find the most purchased item on each category. And to segregate the review based on polarity. And also to help the merchant to identify the most profitable item on each category.

1.1 Classification of Sentiment analysis

The approach made in sentimental analysis can be categorized based on the techniques used, structure of dataset and level of rating, etc. These categorization are again sub-categorized as below:

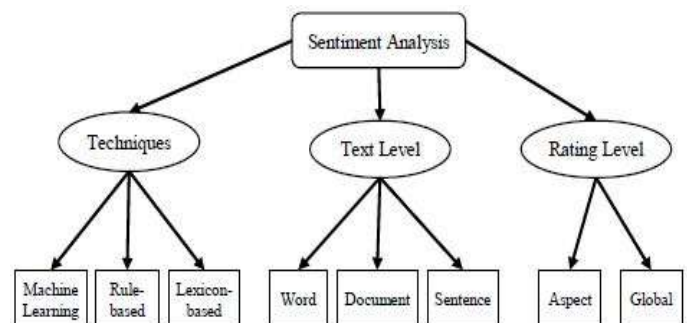


Figure -1: Classification of Sentiment Analysis

Technically sentimental analysis can be done by either.

1. Machine Learning: Dataset are to be trained beforehand. Using standard machine algorithms polarities are detected [6].
2. Rule based: Extracts information from dataset and try to asses them according to the polarity of words. There are different rules such as negation words, idioms, dictionary polarity, emoticons etc. [13].
3. Lexicon-based: Using Semantic orientation i.e. measurement of opinion and subjectivity of a review or comment it generates sentiment polarity (either positive or negative) [7].

2. EXISTING WORKS

Social networking web sites like Facebook and Twitter, to mention few, have changed how information is disseminated or shared instantly. Moreover, they became a platform for social meetings and information exchange. In the process, there is possibility that people express their valuable opinions on products and services. Thus social feedback is made available readily. Any organization cannot afford to have a blind eye on this feedback. However, it is challenging to collect social media content and use it for discovering business intelligence. This is the problem addressed in this paper with Indian Railways case study. [1]. One fundamental problem in sentiment analysis is categorization of sentiment polarity. Given a piece of written text, the problem is to categorize the text into one specific sentiment polarity, positive or negative (or neutral). Based on the scope of the text, there are three levels of sentiment polarity categorization, namely the document level, the sentence level, and the entity and aspect level. And the sentimental analysis has been done on Amazon.com products reviews, tweets about Indian Railway, Movie reviews and Analysis of tweets on Twitter. The disadvantages in this existing system are Opinion orientation (polarity), Comparisons and reviews with Emoji. And now-a-days the proposed systems are comes up with fine grained analysis and fuzzy logic. But they are not completely in field. There are still in development stage. The reputation of a product, brand or company which can be seen as a summary of companies past customer experiences, customer perceptions and business actions, is reflected on or shaped by reviews and forms usually a strong basis for purchasing decisions. Personal purchasing motives include the physical activity or the information seeking while social motives include communications. These motivations are not the same as for a consumer's simple information search in an online environment but differ due to the consumer's intention of purchasing products while searching for reviews. Consumers see online communities as helpful and at the same time they are empowering for the consumers. As the motives for consumers to seek reviews have been identified, it will be analysed which diverse online reviewing platforms are used by consumers. Further, it will be characterized which review design characteristics on platforms might influence consumers' product choice. Retail websites, as well as independent consumer review platforms, personal blogs and video-sharing platforms are classified by the literature as platforms containing customer-generated content in the form of consumer reviews [17].

3. PROPOSED SYSTEM

Here we proposed a system for analysing the reviews of retail shop products to identify the most purchased item on each category. The following diagram represents the proposed architecture for our system.

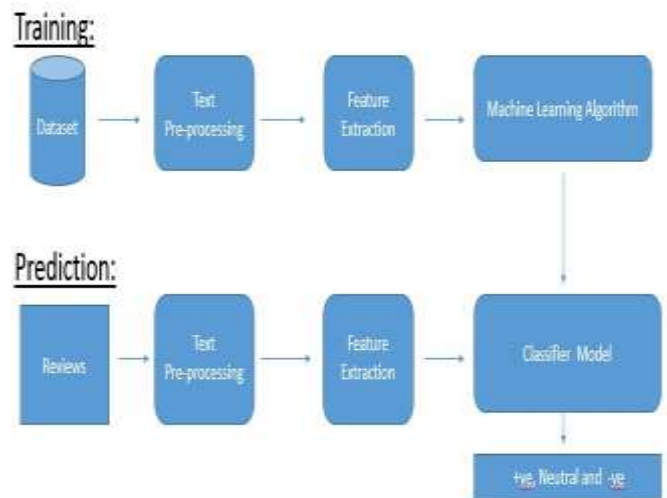


Figure -2: Proposed model representation for sentiment analysis of reviews in retail shop

Here the above architecture represents the sentiment analysis of reviews and rating in retail shop. Here the architecture consists of two processes. One process is named as Training Process and another process is called as Prediction process. In the first process (ie) training process, the model can be trained with the keywords and that the keywords are labelled as positive, negative and neutral keywords. And more keywords can be added for the training process and that can be labelled as mentioned as above. And the second and the last process is prediction process. Here the operation can be done on real users reviews. First the user can write the reviews for the particular product and post the review on the product wall. The model will take the review the prediction process. The model will first do the text-preprocessing and then core task (ie) feature extraction. Here the model can extract the features (ie) keywords used in the reviews. The model can compare the extracted keyword with trained keywords and identify whether the extracted keyword is positive or negative or neutral. Then the review can be labelled as positive or neutral or negative based on the extracted keyword. Thus the model can matches the supervised learning algorithm process. And based on the reviews posted for the particular product, the rating can be given to the product. The process of displaying the rating can be discussed later in this paper. And the diagram shows the star level that can represent how good the product is.

Star Level	General Meaning
★	I hate it.
★★	I don't like it.
★★★	It's okay.
★★★★	I like it.
★★★★★	I love it.

[11].

Figure -3: Meaning for Star Level.

4. ALGORITHM AND WORKING PROCESS

Machine learning techniques are broadly classified into supervised and unsupervised categories. The supervised methods need training phase while the other category does not need training and uses some sort of similarity measure to learn and perform intended operation [1]. In this paper we have used the Support Vector Machine (SVM) algorithm. Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate the reviews based on the training data. For eg, suppose we want to classify the reviews, the model should accurately classify the review, so such a model can be created using SVM algorithm. We will first train our model with lots of positive, neutral and negative keywords, and then we test it strange creature. So as support vector creates a decision boundary between these data and choose extreme cases (support vectors), it will see the extreme case of reviews. On the basis of the support vectors, it will classify whether it is a positive or neutral or negative. As we said that the rating logic can be discussed here. There is a simple logic behind it. (ie) weighted averaging. Consider the positive reviews can be labelled as 5 and 4 stars, the neutral reviews can be labelled as 3 stars and the negative reviews can be labelled as 2 and 1 stars. For eg, consider the reviews for a particular product, the below shows the total number of ratings that the product gets,

5 star - 121

4 star - 92

3 star - 36

2 star - 19

1 star - 8

4.1 Calculation for overall rating:

That is the weighted average, where you weigh each rating with the number of votes it got:

$$\frac{(5*121 + 4*92 + 3*36 + 2*19 + 1*8)}{(121+92+36+19+8)}$$

$$(121+92+36+19+8)$$

$$= 4.08$$

(ie) 4.1 rating

This is the simple logic used to calculate the rating for a particular product. This is logic used in the amazon website.

5. EXPERIMENTAL RESULTS

Experiment are made with the support vector machine method against keywords that are used in reviews of products from its Retail shop website. The below diagram shows the product with description and rating.

S.No	Product	Description	Rating
1.	Soap 1	Hygienic Soap	★★★★★ 4.1

Figure -4: A Product that contains its description and rating

The above diagram shows the product with its description and rating. This is just an example that the retail shop website that contains the product details. So like this there are so many brands of soap will be available in the shop. Each soap has its own description and rating. That all brands will come under the category the "Soap". So, in this paper we have depicted that the customers will have different opinions on each brands. And the opinion can be displayed through rating and that rating may differs from one another one. So we depicts that with Bar-chart that contains the rating in the y-axis and the different brands of soaps are placed in the x-axis. The below diagram shows the bar-chart with the rating and different brands of soaps.

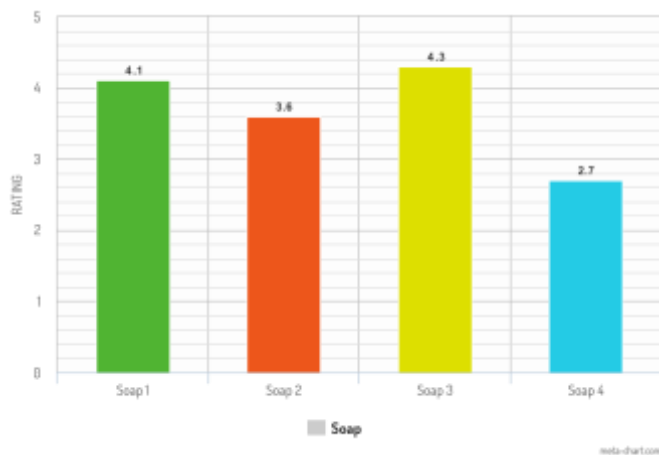


Figure -5: Bar-chart contains the rating for the different brands of soap

The above diagram shows the bar-chart contains the rating for the different brands of soap and usually the bar-chart can be used for statistical purpose and useful for the merchant to identify the most populate item among the category. This bar-chart can be used for other categories of products so that can be useful for improving the retail shop trades.

6. CONCLUSION AND FUTURE WORKS

In this paper, a methodology for sentiment classification has proposed. The case study considered is product reviews in retail shop. It explores supervised learning methods called SVM in sentiment classification of product reviews of retail shop. Since reviews carry valuable social feedback and opinion on the particular product, this study provides useful insights on sentiment classification. It considers positive, negative and neutral sentiments. The proposed framework has two phases such as training and testing. First the user can write the reviews for the particular product and post the review on the product wall. The model will take the review the prediction process. The model will first do the text-preprocessing and then core task (ie) feature extraction. Here the model can extract the features (ie) keywords used in the reviews. The model can compare the extracted keyword with trained keywords and identify whether the extracted keyword is positive or negative or neutral. Then the review can be labelled as positive or neutral or negative based on the extracted keyword. In future we consider fuzzy logic for further refinement of sentiment classification with Highly Positive, Positive, Neutral, Negative and Highly Negative. And we planned to do sentiment analysis for all possible retail shop products and made it bigger.

7. REFERENCES

[1] D.Krishna Madhuri (2019). A Machine Learning based Framework for Sentiment Classification: Indian Railways

Case Study. International Journal of Innovative Technology and Exploring Engineering (IJITEE).

[2] Najma Sultana, Pintu Kumar, Sourabh Chandra, sk Safiful Alam (2019). Sentiment Analysis for Product Review. International Journal of Soft Computing.

[3] Xiang Zhang, Junbo Zhao and Yann LeCun. (2015). Character-level Convolutional Networks for Text Classification, p1-9.

[4] Mengxiang Li, Chuan-Hoo Tan (2013). Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features. International Journal of Electronic Commerce.

[5] Malay Bhatt, Dinesh J. Prajapati (2012). Detection and summarization of genuine review using Visual Data Mining. International Journal of Computer Applications.

[6] Collomb, C. Costea, D. Joyeux, O. Hasan and L. Brunie, "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation". Available at: <https://liris.cnrs.fr/Documents/Liris-6508.pdf>.

[7] Mika V. Mantyla, Daniel Graziotin and Miikka Kuutilla (2018). "The Evolution of Sentiment Analysis-A Review of Research Topics", Computer Science Review, Vol. 27, No. 1, pp. 16-32.

[8] Abdullah Alsaeedi, Mohammad Zubair Khan (2019). Study on Sentiment Analysis Techniques of Twitter Data. International Journal of Advanced Computer Science and Applications.

[9] Murilo Thomaz, Paulo Tromboni de Souza Nascimento, Abraham S. O. Yu, Lilian Schreiner (2016). Purchasing in global new product development: evidences from the cosmetics Industry. Conference Paper

[10] Amel Ziani, Nabih Azizi, Didier Schwab, Monther Aldwairi, Nassira Chekkai, Djamel Zenakhra, Soraya Cheriguene. (2018) Recommender System Through SentimentAnalysis. Available at: "https://hal.archives-ouvertes.fr/hal-01683511".

[11] Fang and Zhan (2015). Journal of Big Data 2:5 DOI 10.1186/s40537-015-0015-2.

[12] Gurneet Kaur, Abhinash Singla (2016). Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET).

[13] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng and Christopher Potts, (2011) "Learning Word Vectors for Sentiment Analysis", Proceedings of 49th Annual Meeting of the Association for Computational Linguistics, pp. 1-7.

[14] Understanding Sentiment Analysis: What It Is and Why It's Used Understanding Sentiment Analysis: What It Is and Why It's Used, Available at: "<https://www.brandwatch.com/blog/understanding-sentiment-analysis/>"

[15] J. Ashok Kumar¹, S. Abirami². "An Experimental Study of Feature Extraction Techniques in Opinion Mining" (2015), IJSCAI.

[16] Nishit Shrestha and Fatma Nasoz. "Deep Learning Sentiment Analysis of Amazon.Com Reviews and Ratings" (2019), IJSCAI.

[17] Nina Isabel Holleschovsky, Efthymios Constantinides. "Impact of online product reviews on purchasing decisions" (2016), Conference Paper.