International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 1296-1300 www.ijaem.net ISSN: 2395-5252

Tracking and Understanding the General Perception of Consumers using Sentiment Analysis

Tanmayee Tushar Parbat B.E IT, Pune
Rohan Benhal BBA IT, Pune
Honey Jain B.E IT, Pune

Submitted: 10-08-2021 Revised: 22-08-2021 Accepted: 25-08-2021

ABSTRACT
In this modern era of globalization, e-commerce has become one of the most convenient ways to shop. Every day people buy many products through online and post their reviews about the product which they have used. These reviews play a vital role in determining how far a product has been placed in consumers' psyche. So that the manufacturer can modify the features of the product as required and on the other hand these will also help the new consumers to decide on whether to buy the product or not. However, it would be a tedious task to manually extract overall opinion out of enormous unstructured data. This problem can be addressed by an automated system called 'Sentiment Analysis and Opinion Mining' that can analyse and extract the users' perception in the whole reviews. In our work we have developed an overall process of 'Aspect or Feature based Sentiment Analysis' by using a classifier called TF-IDF (Term Frequency-Inverse Document Frequency) in a novel approach. It is proved to be one of the most effective ways to analyse and extract the overall users' view about the particular feature and whole product as well.

Keywords: Sentiment Analysis,TF-IDF (Term Frequency-Inverse Document Frequency)

I. INTRODUCTION

The long-term sustainability of companies depends, to a great extent, on their ability to properly meet customer needs. In fact, the aim of satisfying customers is to create brand value, which is a key factor for a company's sustainability [1]. Accordingly, many companies invest huge amounts of money on marketing research to gather information about consumer preferences and demands. From this information, it is crucial to understand what consumers think about the products they buy in order to develop appropriate branding and positioning strategies. Reference [2] stated that a powerful brand name can influence the consumer decision-making process and can positively impact brand sustainability. Specifically, marketing managers need to know how a brand is perceived by its target market relative to other brands in the category and in relation to the most relevant attributes defined for its category. In fact, brand image is built through consumer opinions on specific product characteristics. With global access to the internet, a large amount of data is generated, thereby providing a promising way to discover consumer opinion about products that are bought and experienced. Organizations want to take advantage of these data and convert them into relevant information that allows them to make better decisions, and this is possible by analysing all available data ("big data"). Internet users collaborate daily in the generation of huge amounts of data, thereby becoming one of the most important sources of big data. By writing blogs, participating in social media, or reviewing products online, internet users are constantly generating content. Consumer comments in online forums have proven to be a useful source for revealing consumer insights [3], and this user-generated content (UGC) represents a promising alternative source for potentially identifying customer needs [4]. Thus, mining this UGC and analysing the sentiments of the comments expressed by consumers might be useful for companies. Actually, many researchers highlight the importance of factoring in UGC to aid in decision-making in the marketing field. Particularly, brand management can be one area of interest, as online reviews might have an influence on brand image and brand positioning, including design decisions. In the same line, Fan et al. [5] argued that this type

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 1296-1300 www.ijaem.net ISSN: 2395-5252

of analysis might help manufacturers not only to find out what consumer demands or requirements are, but also to facilitate the design of new products and the improvement of products already available on the market.

II. RELATED WORK

In this section, firstly, the main studies regarding sentiment analysis are reviewed. Secondly, the recommendation systems using metrics of sentiment analysis are treated.

N. Srivats Athindran et al [1] he tweets pertaining to the two smartphones Vivo Nex and Oppo FindX are used as individual customer feedback and the sentiment for each tweet is classified using a hybrid model which is a combination of the Lexicon Based Sentiment analysis and the Naive Bayes algorithms, thus obtaining better accuracy. Then, the overall sentiments for each of the two smartphones is compared which provide a bird's eye view of the user perceptions about the two mobiles. Lastly, we dig further to compare the user sentiments of the individual features in the two smartphones which acts as a powerful feedback mechanism for companies, which they could use to make immediate corrections or utilize this for improving the design of their subsequent models.

G. Hu, et al [2] demonstrate the value of such an analysis in order to assess the impact of brands on social media. We hope that this initial study will prove valuable for both researchers and companies in understanding users' perception of industries, brands and associated topics and encourage more research in this field.

D. V. N. Devi, et al [3] This problem can be addressed by an automated system called 'Sentiment Analysis and Opinion Mining' that can analyse and extract the users' perception in the whole reviews. In our work we have developed an overall process of 'Aspect or Feature based Sentiment Analysis' by using a classifier called Support Vector Machine (SVM) in a novel approach. It is proved to be one of the most effective ways to analyse and extract the overall users' view about the particular feature and whole product as well.

M. Ngaboyamahina et al. [4] this study aims to manage a large volume of information to rank the institutions and provide a practical solution for competitive, marketing analysis, and track the improvement of customer satisfaction within both the public and private sectors to boost the excellent service delivery in Rwanda.

C. Pino, et al. [5] present GeoSentiment, a tool for effective assessment and visualization of event-related sentiments in geographically confined populations. GeoSentiment is developed as a web tool application and provides to stakeholders an easy-to-use and powerful means to investigate how events are perceived by people and which factors may influence such perception. GeoSentiment relies on different services for a) retrieving and mining official statistical information as well as the most-common social networks and b) performing sentiment analysis. It is provided with an interactive interface, which enables rendering and deep exploration of all the processed data and results.

Rahul, V. Raj et al [6] The product provider also gets to know about the user's opinion over a product. This can help the company to improve its marketing strategy and quality of product in their favour. Sentiment analysis uses various semantic approaches like on these online reviews to extract as much feature it can and categorize the type of opinion. Some techniques also help in rating the product value based on user's opinion. This paper is a literature survey including various authors and their sentiment techniques on product or online review.

M. Shukla, et al [7] The framework measures perception of a brand in comparison to peer brands with in-memory distributed algorithms utilizing supervised machine learning techniques. Experiments performed with open data and models built with storylines of known peer brands show the technique as highly accurate and effective in capturing brand perception.

M. G. Sarowar, et al [8] E-commerce reviews and comments about specific products disclose consumer's perceptions as well as attitudes. This attitude expressed by the consumer's seem to be most useful for the new customers who is interested on any product. Meanwhile, an ever-increasing number of reviews and comments are being stored daily and the amount of people buying goods online are increasing in a great extent. User emotions record associated with every product is beneficial for both the makers as well as customers.

Proposed methodology

Tweet cleaning is the first step towards data transformation. This task consists of three subtasks to accomplish this process. The overview of the subtasks is given following. Remove retweets: All tweets which contain the commonly used string “RT” (denoting retweet or repost of tweet) in the body of the tweet are removed. Remove uninformative tweets: short length tweets which will not contain informative data have been removed. A minimum length of 20 characters is used as tweet length. Remove non-English tweets:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 1296-1300 www.ijaem.net ISSN: 2395-5252

Each word used in tweet is compared with common English words list and remove with less than 15% of content matching. • Remove duplicate tweets: After comparing tweets with one another, those tweets whose 90% content has been matched are discarded. Converting all Tweets into Lower case: next step had to convert all tweets to lower case in order to bring the tweets in a consistent form. By doing this, data has been used to perform further transformation and classification without having to worry about non-consistency of data. This task is done using lower () function in RStudio. This function converted all alphabets in lower case and solved case sensitivity problem by making all data consistent in lower case. Removing emoticons and punctuations: next step is to remove emoticons and punctuations because they were not needed in the analysis. One might ask why emoticons have been removed, because when they were being extracted, they appeared in the form of square boxes instead of proper emoticons. These garbage values of emoticons were removed by using "replace" keyword. Replace keyword was used in a way that it replaced all emoticon symbols and values with an empty space making data clear from useless emoticon mess. Removing URL'S: The very next step is to remove the URL'S, as they provide no information during analysis. URL's show links to other webpages and websites. These were of no use so from all tweets these were removed by using resub () in RStudio, which replaced all sentences and sub parts of sentences started from http with blank spaces.

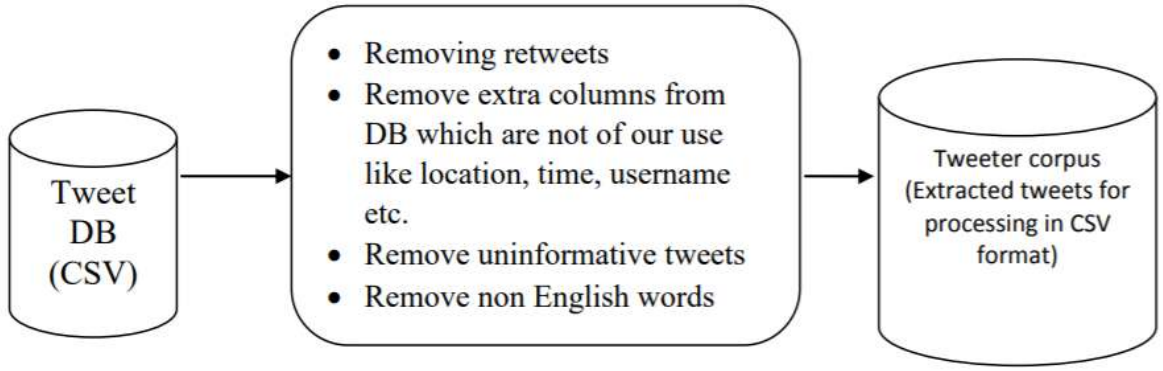

Figure 1: Proposed flowchart

TF-IDF (Term Frequency-Inverse Document Frequency)

The TF-IDF score is used in sentiment analysis, to balance the weight of words based on frequency of occurrence in document. The emotions are calculated by the presence of term and also frequency of term. Term presence is whether the term is present or not and represent by 1 or 0 while the term frequency is how many times the attribute occurs (1, 2, 3, etc.) that is registered. The use of presence (absence) or count (frequency) affects the resulted emotion of tweet. Other thing, which should consider is whether every attribute has the same importance for a Twitter post. For example, does 'this' have the same importance as 'filthy' for a given Twitter post? TF-IDF (Term Frequency, Inverse Document Frequency) is a way of measuring the importance of term in a tweet. The TF-IDF (TI) of an attribute a in Twitter post t can be calculated by Where is the frequency of the attribute in tweeter's post(t), T is the total number of tweets in training data, Ta is the total number of tweets that included attribute t. Attributes are words selected from tweets, this will assign less weight to most frequent words and will assign high frequency to those which occur rarely in tweets because in an article if a word or phrase appears with high frequency but in other articles it appears rarely that can considered that word or phrase has ability to distinguish between categories. After calculating TF-IDF next step is to create Term Document Matrix (TDM) or Document Term Matrix (DTM). This is a simple and easy way to represent text for future computation. It can be exported from a Corpus to use as a bag-of-words mechanism, which results in a matrix with document IDs as rows and terms as columns where matrix elements are representing the term frequencies. TDM is an inverse of DTM where Rows are represented by documents and columns are represented by the words contained in respective documents.

III. CONCLUSION AND FUTURE WORK

The paper firstly describes the process of sentiment analysis on three groups of tasks and the specific challenges that each task tackles. It was pointed out that the data resource has to be adequately prepared, which is usually

accomplished by adding metadata to a set that is, annotating the texts according to polarity, the goal of the sentiment – aspect but also by attaching other semantic, syntactic or lexical information to the source data. Annotations have to be accurate and relevant to the task in order to achieve effective training by the data-mining algorithm, so adequate annotation is critical for the sentiment analysis. It is pointed out that text classification (to positive, negative or neutral polarity of the expressed opinion in the comments) could be conducted at different levels of analysis (at the level of the document, sentence, word or phrase). The paper particularly highlights the importance of properly selected and robust supervised and unsupervised methods. It also underlines the specific difficulties in analysis that arise from the potential existence of negation with varying range of influence, irony that is hard to detect and implicitly expressed sentiment in objective sentences. It is also pointed out that the implementation of sentiment analysis could be even more complex because it requires combining with other tasks (summarizing opinions, finding opinions, ranking products according to some expressed opinion, etc.) and methods from the specific domain of interest.

REFERENCE

[1]. N. Srivats Athindran, S. Manikandaraj and R. Kamaleshwar, "Comparative Analysis of Customer Sentiments on Competing Brands using Hybrid Model Approach," 2018 3rd International Conference on Inventive Computation Technologies (ICICT), 2018, pp. 348-353, doi: 10.1109/ICICT43934.2018.9034283.

[2]. G. Hu, P. Bhargava, S. Fuhrmann, S. Ellinger and N. Spasojevic, "Analyzing Users' Sentiment Towards Popular Consumer Industries and Brands on Twitter," 2017 IEEE International Conference on Data Mining Workshops (ICDMW), 2017, pp. 381-388, doi: 10.1109/ICDMW.2017.55.

[3]. D. V. N. Devi, C. K. Kumar and S. Prasad, "A Feature Based Approach for Sentiment Analysis by Using Support Vector Machine," 2016 IEEE 6th International Conference on Advanced Computing (IACC), 2016, pp. 3-8, doi: 10.1109/IACC.2016.11.

[4]. M. Ngaboyamahina and S. Yi, "The Impact of Sentiment Analysis on Social Media to Assess Customer Satisfaction: Case of Rwanda," 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA), 2019, pp. 356-359, doi: 10.1109/ICBDA.2019.8713212.

[5]. C. Pino, I. Kavasidis and C. Spampinato, "GeoSentiment: A tool for analyzing geographically distributed event-related sentiments," 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), 2016, pp. 270-271, doi: 10.1109/CCNC.2016.7444775.

[6]. Rahul, V. Raj and Monika, "Sentiment Analysis on Product Reviews," 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2019, pp. 5-9, doi: 10.1109/ICCCIS48478.2019.8974527.

[7]. M. Shukla, A. Fong, R. D. Santos and C. Lu, "DERIV: Distributed In-Memory Brand Perception Tracking Framework," 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), 2016, pp. 387-393, doi: 10.1109/ICMLA.2016.0069.

[8]. M. G. Sarowar, M. Rahman, M. N. Yousuf Ali and O. F. Rakib, "An Automated Machine Learning Approach for Sentiment Classification of Bengali E-Commerce Sites," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019, pp. 1-5, doi: 10.1109/I2CT45611.2019.9033741.

[9]. Anand Singh Rajawat SumitJain KanishkBarhanpurkar, (2021) Fusion protocol for improving coverage and connectivity WSNs, IET Wireless Sensor Systems, https://doi.org/10.1049/wss2.12018

[10]. X. Liu, H. Xu, P. Xue, X. Liu and Z. Xu, "Research on the Influence of E-commerce Non-subjective Fault Behavior on Consumers' Repurchase Intention," 2020 International Conference on Communications, Information System and Computer Engineering (CISCE), 2020, pp. 221-226, doi: 10.1109/CISCE50729.2020.00050.

[11]. M. G. Sarowar, M. Rahman, M. N. Yousuf Ali and O. F. Rakib, "An Automated Machine Learning Approach for Sentiment Classification of Bengali E-Commerce Sites," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019, pp. 1-5, doi: 10.1109/I2CT45611.2019.9033741.

[12]. Goyal S.B., Bedi P., Rajawat A.S., Shaw R.N., Ghosh A. (2022) Multi-objective Fuzzy-Swarm Optimizer for Data

Partitioning. In: Bianchini M., Piuri V., Das S., Shaw R.N. (eds) Advanced Computing and Intelligent Technologies. Lecture Notes in Networks and Systems, vol 218. Springer, Singapore. https://doi.org/10.1007/978-981-16-2164-2_25

[13]. Rajawat A.S., Barhanpurkar K., Goyal S.B., Bedi P., Shaw R.N., Ghosh A. (2022) Efficient Deep Learning for Reforming Authentic Content Searching on Big Data. In: Bianchini M., Piuri V., Das S., Shaw R.N. (eds) Advanced Computing and Intelligent Technologies. Lecture Notes in Networks and Systems, vol 218. Springer, Singapore. https://doi.org/10.1007/978-981-16-2164-2_26

[14]. Amato A., Cozzolino G., Giacalone M. (2020) Opinion Mining in Consumers Food Choice and Quality Perception. In: Barolli L., Hellinckx P., Natwichai J. (eds) Advances on P2P, Parallel, Grid, Cloud and Internet Computing. 3PGCIC 2019. Lecture Notes in Networks and Systems, vol 96. Springer, Cham. https://doi.org/10.1007/978-3-030-33509-0_28

[15]. Bedi P., Goyal S.B., Rajawat A.S., Shaw R.N., Ghosh A. (2022) A Framework for Personalizing Atypical Web Search Sessions with Concept-Based User Profiles Using Selective Machine Learning Techniques. In: Bianchini M., Piuri V., Das S., Shaw R.N. (eds) Advanced Computing and Intelligent Technologies. Lecture Notes in Networks and Systems, vol 218. Springer, Singapore. https://doi.org/10.1007/978-981-16-2164-2_23

[16]. Dutta A., Das S. (2021) Tweets About Self-Driving Cars: Deep Sentiment Analysis Using Long Short-Term Memory Network (LSTM). In: Gupta D., Khanna A., Bhattacharyya S., Hassanien A.E., Anand S., Jaiswal A. (eds) International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, vol 1165. Springer, Singapore. https://doi.org/10.1007/978-981-15-5113-0_40

[17]. Rajawat A.S., Rawat R., Mahor V., Shaw R.N., Ghosh A. (2021) Suspicious Big Text Data Analysis for Prediction—On Darkweb User Activity Using Computational Intelligence Model. In: Mekhilef S., Favorskaya M., Pandey R.K., Shaw R.N. (eds) Innovations in Electrical and Electronic Engineering. Lecture Notes in Electrical Engineering, vol 756. Springer, Singapore. https://doi.org/10.1007/978-981-16-0749-3_58.

[18]. Tiruwa A., Yadav R., Suri P.K. (2020) Sentiment Analysis: An Effective Way of Interpreting Consumer's Inclinations Towards a Brand. In: Suri P., Yadav R. (eds) Transforming Organizations Through Flexible Systems Management. Flexible Systems Management. Springer, Singapore. https://doi.org/10.1007/978-981-13-9640-3_12.

[19]. Luna-Nevarez C. (2018) An Exploratory Analysis of Consumer Opinions, Ethics, and Sentiment of Neuromarketing: An Abstract. In: Krey N., Rossi P. (eds) Back to the Future: Using Marketing Basics to Provide Customer Value. AMSAC 2017. Developments in Marketing Science: Proceedings of the Academy of Marketing Science. Springer, Cham. https://doi.org/10.1007/978-3-319-66023-3_32