

Twitter Mining using R

Omkar Jagdale¹, Vikrant Harmalkar¹, Swati Chavan¹ and Prof. Nidhi Sharma²

¹ UG Scholar, ² Professor

Department of Computer Engineering

Bharati Vidyapeeth College of Engineering

Navi Mumbai, Maharashtra, India.

ABSTRACT

With the growing availability of user-generated contents (UGC), such as discussion forums, blog sites, Internet forums and social networks, public have multiple ways to mention their reviews, comments and make them available to everybody. Publicly open opinions provide valuable data for decision-making processes. Therefore, the computational treatment of sentiment and opinions has been regarded as a challenging field of research that can serve diverse purposes. In this, the various methods of mining in multiple ways such as web and data mining are used to retrieve data from web sites and to optimize we need to go through the queries of data mining.

The proposed development of Opinion Mining is fundamentally intended to develop a system where users can get an optimized result for the different opinions on different products or services available on different e-commerce websites. This project mainly deals with evaluating different opinions so that we can get a quick idea of different views expressed by different users. Here, the data mining concepts are used which mainly deals with mining the UGC from different e-commerce websites which are being used in our daily routine life and after we extract the required UGC we need to prepare the definite opinion result using different data mining techniques.

KeyWords: *Opinion, User Generated Content, Mining*

1. INTRODUCTION

Opinion Mining involves development of system to explore user's opinions made in blog posts, comments, reviews or tweets, about the product, current affairs, policy or etc. Opinion Mining is the method of finding the opinion of the users who are actually the part of the organization using the product generated by them. As the world turned into E-World the mode of expression is dramatically changed for example Nowadays we use wide varieties of smiles and symbol for expressing feeling in texting. Large amount of social communication can be observed on internet and innovative terms have been coined for various ways of communication like twitting, posting, texting, etc. folks like to connect with others via internet, they want to share their opinions, reactions, likes, dislikes, views, reviews, feelings, sentiments etc. people are glad to share their personal life via social media, the popularity of social media has increased so much and so quickly that even nobody bothers about what they are sharing and is this good to share our personal life with strange people? Is there any necessity to share our photos, videos or our daily happenings on internet? So, finding the sentiment, feeling behind this activity is also an important job for understanding the psycho-socio status. So, from that text, mining the opinions of people and finding their sentiments, feelings, views, reaction

and emotions have become tough and challenging task. Opinion Mining is the technique that evaluates people opinion, emotions, attitudes, views, sentiments and feelings from specified text. Opinion mining is widely use technique and research field which is use in data mining and text mining. This research has got caught so much interest even outside of computer science to the management science and social science because of its importance to business and society. The rising importance of opinion mining coincides with the progress of social media such as blogs, twitter, discussion forums, reviews sites, discussion groups, WhatsApp, facebook, and blogs.

2. RELATED WORK

The Internet has made huge volumes of information accessible to the normal non-technical user at home, in business and in education. For many people, having access to this information is no longer just an advantage, it is essential. For mining of opinions from sentences we need to use text classification algorithms which will classify the words captured from the reviews, comments, etc. Michael G Madden proposed a methodology for induction of Bayesian network structure for categorization of sentimental words and it is called Partial Bayesian Network. K2 framework is used to implement this structure. Author explained Partial Bayesian Network is feasible for small data sets as complexity of K2 algorithm is exponential to the no. of variables. Xiaowen Ding, Bind Liu discussed about customer reviews of the product, they also mentioned about the sentimental words that show state of mind whether it is desirable or undesirable. Authors have used method based on holistic lexicon for answering the problem by manipulating exterior evidences and linguistic conventions of expressions used in natural language. Sentimental classification can be achieved by using graph based approaches. Authors categorized the sentence into positive, negative and objective of the sentence. Intra-document and inter-document evidence are two outside sentence features and these features are described by Zhao Bind, Liu Ting. Then to increase the efficiency of sentence sentiment classification, a graph based propagation approach is presented to integrate these inside and outside sentence features.

Opinion mining can also be done effectively from various blogs, discussion forums, etc. Jack G. Conrad, Frank Schilder said the scope & opinion mining of blogs which is increased in legal domain. Authors first developed the weblog collection. The weblog was containing entries made to the blogs that discuss legal search tools. Then authors afterward inspected the performance of language modelling method arranged for both subjectivity analysis and polarity analysis. Opinion mining can also be used to simply decide the polarity of any review. S.M. Kim and E. Hovy discussed about how polarity ontology can be described to show user opinions. Sentences describing the individual's opinion can be extracted from tweets, reviews and determine whether each opinion sentence is positive or negative.

3. PROBLEM DEFINITION

With more and more common users feeling ease with using Internet, a key advantage of social media is that we can understand the good and bad things, people express about the specific brand or personality. The bigger your business gets difficult it becomes to keep a grip on how everyone thinks about your brand. For large businesses, firms with thousands of mentions every day on social media, discussion forums, review sites and blogs, it is extremely difficult to do this manually even it can be problematic. To overcome this problem, opinion mining software is essential. This software can be used to evaluate the people's opinion about particular product or object.

The objective of this project is to show how opinion mining can be beneficial in improving for decision making about user specified word. The learning algorithm will learn what our reviews, mentions or comments are from statistical data and then will determine the polarity. After that it will deliver computed opinion from overall reviews of the user specified word which will help user in understanding the polarity of the required topic. The project aims to implement these in web app, while making our lives better and our experience richer and efficient.

4. IMPLEMENTED SYSTEM

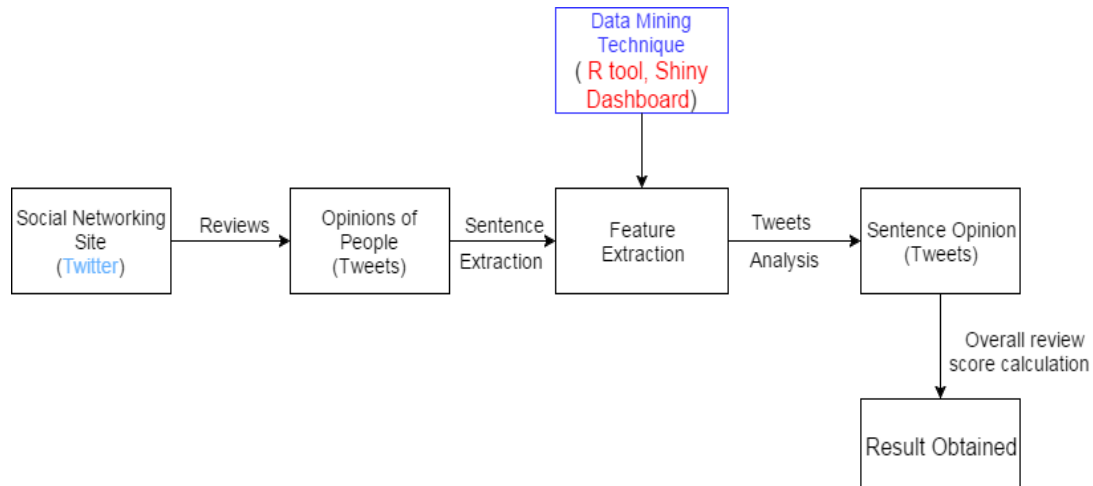


Figure4.1: Workflow of Twitter Mining

The implemented system provides the related tweets for which user is demanded. The system also takes the number of tweets the user want, the duration in which tweets are tweeted. If the user did not choose the option then the default values are taken for the mining. After submitting the requirements, the system will search the related tweets and display the summery of tweets along with the polarity which is represented in the form of graph. There are various machine learning tools are available out of those R tool and Shiny Dashboard is use in our system. R tool is use for the analysis purpose while Shiny dashboard is use for the GUI (Graphical User Interface) purpose. The R tool has inbuilt data dictionary with their respective sentimental values. The sentimental words are identified and the polarity of the word is calculated using the Naive Bayes Classifier algorithm. In Naïve Bayes, the sentimental words will be fetched from the available reviews and then those words will be matched with our data dictionary and then the corresponding value from -5 to +5 will be mapped with those sentimental words. The polarity obtained from data dictionary will be evaluated to calculate the overall score (rating) of the topic. If there are multiple attributes for one topic, then overall score will be calculated by considering every attribute’s score.

4.1 Challenges in Opinion Mining

- Domain Independence: The major challenge faced by opinion mining and sentiment analysis is the domain dependent nature of sentiment words. One features set may give very good performance in one domain, at the same time it performs very poor in some other domain.

- Detection of spam and fake reviews:

The web contains both authentic and spam contents. For effective Sentiment classification, this spam content should not be considered for processing. This can be done by detecting duplicates, by spotting outliers and by considering reputation of reviewer.

- Way of Expressing the Opinion:

The people don't always mention their opinions in the same way, it might be specified in many different ways. The way of thinking and expressing may vary from person to person so accordingly the opinion of each individual is different.

- Use of Abbreviations and short forms:

Now a day's people are using shortcuts, abbreviation, synonyms, special symbols regularly so fetching the accurate opinion from that is too difficult.

5. GRAPHICAL USER INTERFACE (GUI)

The web app interface has Title panel and side bar layout as shown in fig 5.1. The sidebar layout has two columns one is side bar panel and another is main panel. The title panel contains the title of the page while side bar panel provides user input box where user specifies the word for which he/she wants to calculate polarity through text box. The side panel also contains the slide bar for specifying the count of tweets along with the duration in which tweets are tweeted below that user has the choice of which type of tweets he wants (example Recent, popular, mixed). The main panel contains the tweets related to search word, it also provides the word cloud and polarity graph. For example, if we want to calculate the polarity for the word 'ipl' then user will search it through text box and will select the count of the tweets, duration and the type of the tweets he wants. After that the mining process will start and the requested tweets will appear in the main panel. The main panel has many choices in which form the user want to see the result. Our system provides summery of tweets, represents the result in terms of pie chart, graph and also provide the word cloud for the recommended word.

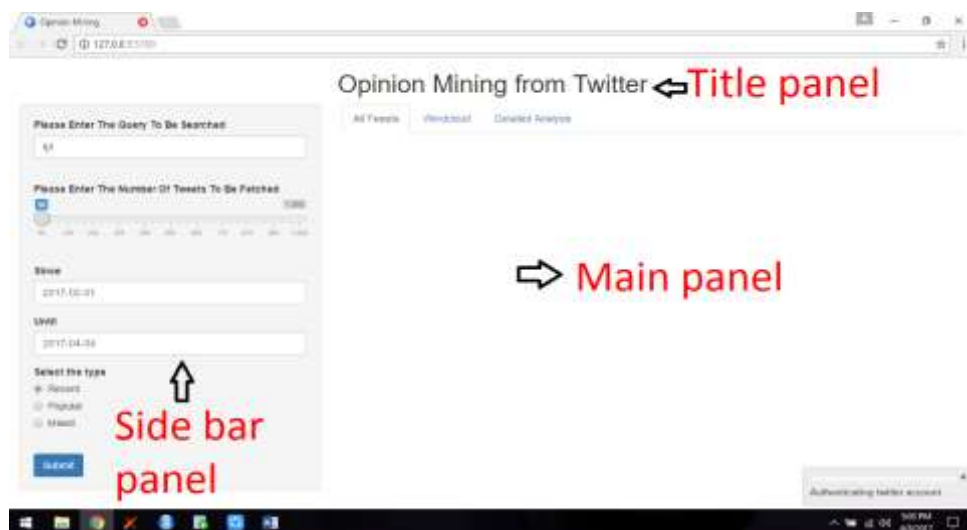


Figure 5.1: GUI (Graphical User Interface)

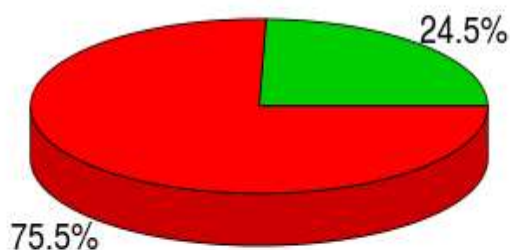


Figure 5.4: Pie chart

6. CONCLUSION

Opinion mining is an evolving arena of data mining used to extract the precious knowledge from vast volume of customer reviews, mentions, comments and feedback on any product or topic etc. The opinions of the user are extracted at the three levels of analysis i.e. at the document, sentence and aspect level. It is observed that opinion mining is widely used for sentimental calculations of the reviews from the twitter data, comments from the social networking sites. In present scenario, Opinion Mining can also be carried out on a set of reviews and set of discovered feature expressions extracted from reviews. Naive Bayesian algorithm is the technique for current methods, useful for constructing better summary based on feature based opinions as positive, negative or neutral and it is the most efficient method.

ACKNOWLEDGMENT

We are heartily thankful of our project guide Prof. Nidhi Sharma for her guidance and support during project development. Also, we are thankful to BVCOE principal and all staff for their good support.

REFERENCES

- [1] Weishu Hu, Zhiguo Gong, Jingzhi Guo, "Mining Product Features from Online Reviews", IEEE International Conference on E-Business Engineering.
- [2] Wenqian Shang, Youli Qu, Houkuan Huang, Yongmin Lin, and Hongbin Dong, "A Role-based Customer Review Mining".
- [3] XU Xueke, CHENG Xueqi, TAN Songbo, LIU Yue, SHEN Huawei, "Aspect-Level Opinion Mining of Online Customer Reviews".
- [4] Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi, and Tao Li, "Dual Sentiment Analysis: Considering Two Sides of sOne Review".