

A Tool for Brand Analysis with Twitter Data

Vaibhav Lanke¹, Amrish Raghuwanshi², Siddhant Gholap³, Paresh Nankar⁴

Students, Computer Engineering, PVPIT, Pune, India^{1,2,3,4}

Abstract: As one of the largest Social Media in providing public data every day, Twitter has attracted the attention of researcher to investigate, in order to mine public opinion, which is known as Sentiment Analysis. Consequently, many techniques and studies related to Sentiment Analysis over Twitter have been proposed in recent years. However, there is no study that discuss about sentence pattern of positive/negative sentence and neither subjective/objective sentence. In this paper we propose POS sequence as feature to investigate pattern or word combination of tweets in two domains of Sentiment Analysis: subjectivity and polarity. Specifically we utilize Information Gain to extract POS sequence in three forms: sequence of 2-tags, 3-tags, and 5-tags. The results reveal that there are some tendencies of sentence pattern which distinguish between positive, negative, subjective and objective tweets. Our approach also shows that feature of POS sequence can improve Sentiment Analysis accuracy.

Keywords: Dimensions, Reach, Engagement, ROI.

I. INTRODUCTION

Social media plays an increasingly important role in how customers both express their interests and discover relevant products. It is, as a result, important that businesses can respond to trends and interests expressed in social media and engage both existing and potential customers. In comparison with most traditional advertising channels, the content in social media can be used to better target potential customers. In this paper, we develop a solution that determines product relevance based on a customer's interests as expressed through social media.

Today's modern life is completely based on Internet. Now a day's people cannot imagine life without Internet. From last few years people share their views, ideas, information with each other using social networking sites. Such interchanges might include diverse sorts of substance such as text, image, audio and video data.

We concentrate on engaging users of Twitter, a popular micro-blogging social medium site, with huge amount of reviews and feed backs for various products. Twitter is particularly suited for this purpose because the communications is largely public. Twitter also represents several unique challenges. First, Twitter messages, or tweets, are limited to 140 characters, and are often written in different languages with poor spelling and abbreviated sentence structure. Each tweet, therefore, represents only some information about the user. Also same sentence in Twitter can have multiple meanings so it is difficult to gauge the sentiment of user accurately.

More and more data is getting generated and thus it becomes quite feasible to use this data in analytics. As the amount of the data generated is very huge (normally in Petabytes), we term the data as Big Data. With quite accurate techniques coming into play for Sentiment Analysis the task of analyzing user sentiments is also gaining precision and thus can be used efficiently.

Analyze sentiments for a particular brand. Sentiments are categorized as Negative and Positive. Verify the feedback from Twitter users after launching the new feature or offer or release. Launch a campaign based on these sentiments. Gather data from Twitter for a particular brand. Sentiment Analysis also referred as Opinion Mining on Big Data is a current topic for research as this leverages an opportunity to use already present data in open and analyze the same. Also there are open libraries available for Sentiment Analysis for example SentiWordNet that uses WordNet for its dictionary. This paper focuses on using Sentiment Analysis techniques for building and monitoring a brand through tweets available on Twitter.

Twitter provided us the API in Twitter4Js which can be used get Twitter data for past 4-7 days. This data can be filtered to get the tweets for a particular brand. This classified data then can be used to carry out the sentiment analysis for a brand. The launch of new features or offers can be done through Twitter and then the user feedback can be checked by analyzing this data. The tool also helps in targeting specific user category like users form a particular locality, age, gender or existing customers etc.

II. LITERATURE SURVEY

In paper [1], This paper introduces a cloud-based tool for brand monitoring in social media, namely SBM tool. The proposed prototype bases its analysis on Twitter contents, crawled and managed through the computational resources provided by Windows Azure's PaaS service model. Moreover, our sentiment and author influence analyses are based on a simple algorithm allowing us to fully test the potentiality of our prototype. We believe that our tool is extremely promising and a valid asset for companies looking to enhance their electronic commerce. However, in order to develop an advanced and public beta version of the SBM tool, an extensive work should be done in the

future. As previously mentioned, we will change the sentiment and influence analysis algorithms in order to enhance the accuracy of researchers' analysis and to handle more complex sentiments. We will create new services such as the possibility to understand not only users feelings but also to discriminate between good and bad features of a product. In addition, we will also improve the web crawler layer, extending the brand search to other social media such as Facebook and Youtube. Finally, to further improve the performance of the SBM tool, we will test the Windows Azure Table Storage service, a NoSQL data-store for structured and non-relational data.

In this paper [2], This paper proposes the SMA framework as a means of understanding and explaining how SMA brings value to organizations. The framework was developed by synthesizing concepts from two theories: organizational motivation theory and the resource based view of the firm. One limitation of the study is that it is based on one case study. Although we argue that the Bankco case study is revelatory, more case studies are required to further refine and develop the framework. Our next step is to further develop the SMA benefits framework using expert interviews and more industry case studies. Further research his planned to conduct a survey based on constructs within the model.

In this paper [3], As social media have become a topic of interest for many industries, it is important to understand how social media data can be harvested for decision making at the industry level. Currently, the majority of social media studies focus on individual companies or organizations. There are few studies performing social media competitive analysis on the leading companies in an industry in a systemic way. As an exploratory study, this case study made a contribution by using text mining to perform competitive analysis for the user-generated data on Twitter and Facebook in three major pizza chains. Results from the text mining and social media competitive analysis show that these pizza chains actively engaged their customers in social media such as Twitter and Facebook. They used the social media not only to promote their services, but also to bond with their customers. Findings from this study suggest that social media plays an important role in sustaining a positive relationship with customers.

In this paper [4], we discuss about the use of POS Sequence in Sentiment Analysis over Twitter in two domains: subjectivity and polarity. To achieve the most optimum POS sequence in uncovering sentence pattern, we conducted the study in three variations of POS sequence ($n = 2$, $n = 3$, and $n = 5$). In addition, we performed sentiment classification by incorporating AFINN Lexicon and POS Sequence. In our first experiment, the results reveal that subjective tweets tend to have word combinations consisted of adverb and adjective. This is in line with subjective utterance purpose

that expresses emotion or private point of view. In contrast, the objective tweets tend to have word combination of nouns which basically aims to express a fact or neutrality rather than emotion. Whereas, in polarity domain, the negative tweets tend to have word combination of affirmation words which often appear as negation word. In the second experiment, the results show that features of POS sequence are able to boost the accuracy in incorporation between AFINN and POS sequence. It affirms that POS sequence can be utilized for performing Sentiment Analysis over Twitter.

III. MOTIVATION AND PROBLEM STATEMENT

Social Media like Twitter have gained a huge popularity in current age and many users for a brand are connected through this social media net. Twitter is also a popular player in Social Media arena and quite a lot of people are connected through this social media network. Data from this social media are generated in huge amount daily which is freely available. The data is generated from the users that post their sentiments for a particular brand. This data which is available from social media can be used to carry out an analysis to gather user sentiments. The field of Sentiment Analysis is thus gaining a lot of limelight as it provides an effective way of analyzing this data.

Recently, several companies recognize the benefit derived from the analysis of social networks, (also defined media) data. The user-generated contents (e.g. individual opinions, media information, etc.) are becoming a useful resource for researchers to understand what people think about a certain topic, brand or product. Through the analysis of the contents produced by social network's users, companies can manage their brand reputations. Moreover, companies can create adhoc marketing campaigns and advertisements analysing people feelings. The benefits derived from users' data analysis allow companies to increase their own electronic commerce, satisfying user's needs. Currently, a large number of start ups or companies have created social media monitoring services, in order to facilitate the monitoring of several social networks. For example, a list of companies and social media monitoring tools is provided in and , describing systems like Brandwatch and Sysomos, two tools designed for business marketing. These social media monitoring tools allow finding insights into a brand's visibility on social networks, searching and streaming social media data using some criteria such as keywords or languages. These tools also provide useful functions to analyse the data and understand people feelings and opinions.

IV. PROPOSED SYSTEM

The system is designed to enable a brand to gather the sentiment flowing among users by collaborating data from Twitter. The Twitter data is accumulated on the fly to get a sentiment analysis done for a particular brand. A brand can thus judge the current market scenario from the

sentiments floating by the Twitter user that respond for a brand. The brand campaign is thus strategized based on these sentiments from the user.

V. SYSTEM ARCHITECTURE

This project is created in a three tier architecture. The rest tier is the layer which is visible to the end-user. The input is given and the output is displayed in this layer. It is the client and it is connected to the next layer using connectivity drivers. Second tier is the important layer which performs all the logical operations in this project. This is the middle layer and it acts as a medium between rest and third layer. It is connected to the third layer by the database connectivity. The third tier is the storage layer. This layer has the cache database which stores and manages the data which comes into the application.

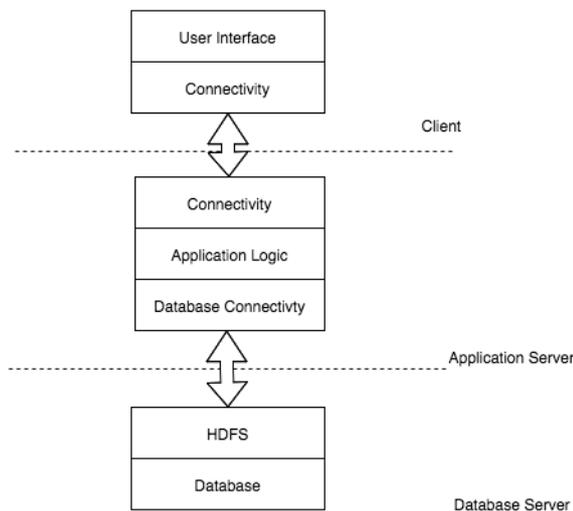


Fig1: System Architecture

In particular, **Content-Based Messages Filtering (CBMF)**: For content-Based Messages filtering, we first filter out duplicate tweets and facebook comments, non-English tweets and non English facebook comments, and tweets that do not contain hash tags. From the remaining set (about 4 million), we investigate the distribution of hash tags and identify what we hope will be sets of frequent hash tags that are indicative of positive, negative and neutral messages. These hash tags are used to select the tweets that will be used for development and training [1].

- **PRESENTATION LAYER**: It is the topmost and interactive layer in this architecture. Input from Twitter is fed into the system from this layer. The input data after getting processed is sent to the presentation layer. Output in the form of tables and graphs are displayed in this layer.

- **APPLICATION TIER**: This layer lies between presentation and data tier. All the logical operations are performed here. Apache Scala is the language used to build this component. The analysis of data is done by searching for adjectives in the tweets and classifying these

tweets as negative, positive or neutral. The polarity of the tweets attained through classifier is rounded off in four sets 0, 1, -0.5 and 0.5. The analysis is carried out parallel with the help of RDDs provided by Scala.

- **DATA TIER**: Twitter data is collected using streaming api interface provided by this language. Twitter provides Twitter4j API which provided tweet data for a particular topic. The count of tweets can be customized along with the region for tweet collection and language as well.

VI. EXPECTED RESULT

The final result are shown in the tool in graphical format. The classification time for 1000 tweets was around 1.2 min. The tweet sentiment will be categorized in 5 classes as shown in the table Table I. The 50 tweets out of which were manually classified which gave an accuracy on 76%.

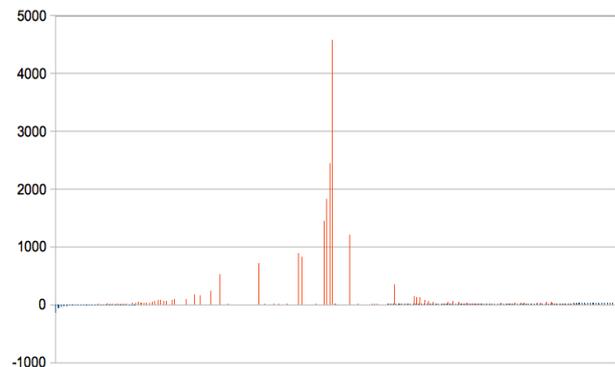


Fig2: Result Display Page

The Sentiment Analysis done by SentiWordNet claims in between 12-70 percent depending upon the quality of text used and training data set. The expected outcome of the sentiment analysis done in the technique used here aims to achieve 70-80 percent precision.

VII. FUTURE SCOPE

The tool focuses on Twitter data analysis as the Twitter limits the data size to 140 characters which give us an advantage of size limited dataset. Also another advantage of Twitter is it has only text data. The tool but can be extended to Facebook or other such social media tool which have an added complexities of video data, images and sound files as well. Also the Sentiment Analysis layer can be extend to use multiple algorithm like a combination of Naive Bayes and SVM instead of using a single algorithm which can increase the accuracy of the proposed model.

VIII. CONCLUSION

The paper thus describes a simple method for analyzing Twitter tweets filtered from some specific keywords that describe a brand and gives us an insight of the sentiments for a brand. Thus the implementation of Naive Bayes



algorithm can be used for analyzing tweets and using this information for building brand reputation on Twitter media.

ACKNOWLEDGMENT

We take this golden opportunity to owe our deep sense of gratitude to our project guide **Prof. Y.B. Gurav**, for his instinct help and valuable guidance with a lot of encouragement throughout this paper work, right from selection of topic work up to its completion. Our sincere thanks to Head of the Department of Computer Engineering and Information Technology **Prof. B. K. Sarkar** who continuously motivated and guided us for completion of this paper. I am also thankful to all teaching and non-teaching staff members, for their valuable suggestions and valuable co-operation for partially completion of this work. We specially thank to those who helped us directly-indirectly in completion of this work successfully.

REFERENCES

- [1] Antonio Tedeschi, Francesco Benedetto, "A Cloud-Based Tool for Brand Monitoring in Social Networks", 2014 International Conference on Future Internet of Things and Cloud.
- [2] Nargiza Bekmamedova, Graeme Shanks, "Social Media Analytics and Business Value: A Theoretical Framework and Case Study", 47th Hawaii International Conference on System Science, 2014
- [3] Wu He, Shenghua Zha, Ling Li, "Social media competitive analysis and text mining: A case study in the pizza industry", International Journal of Information Management 33 (2013) 464– 472, 2013
- [4] Fajri Koto, and Mirna Adriani, "The Use of POS Sequence for Analyzing Sentence Pattern in Twitter Sentiment Analysis", 29th International Conference on Advanced Information Networking and Applications Workshops, 2015.
- [5] Sheng Yu and Subhash Kak, "A Survey of Prediction Using Social Media".