

Survey Paper on Twitter Sentiment Analysis Using Portar Stemming Algorithm

Nishad Patil¹, Tingre Sayali², Thorat Kalyani³, Shivshetty Swapnil⁴, Patil Shwetal⁵

^{1,2,3,4}Department of Computer Engineering, MMIT, Lohegaon, Pune, India

⁵(Guide)Department of Computer Engineering, MMIT, Lohegaon, Pune, India

Abstract: *Twitter and other social networking sites have recently evolved as the main source of data for analysis of a product or a particular thing over the time. These sites have millions of user's everyday producing trillions of data day in and day out. Twitter being the most common micro blogging site help us fetch millions of tweets that are generated every minute, these tweets can be used for performing sentiment analysis and also for checking what the user are saying about a particular product i.e. positive, negative or neutral. The system will be fetching live tweets and generating result as positive, negative and neutral.*

Keywords: hashtag, tweets, hadoop, sentiment analysis, twitter, portar stemming

1. Introduction

Today's modern life is totally based on Internet. Now a day's people cannot imagine life without Internet. Also, OSNs are just a part of modern life. From last few years people share their views, ideas, information with each other using social networking sites. OSN users have ability to keep in touch with his/her friends by exchanging different types of information or messages also. Social Media is one of the most significant information exchange technology of the 21st century. People of all ages use social media to share their views and opinions with friends or the wider social web. Social media, such as Twitter or Facebook they can share their ideas / views and opinions of users related to any public topics. Consequently, sentiment analysis of social media content may be of interest to different public sector organizations, especially in the security and law enforcement sector.

This paper mainly focuses on twitter data i.e. tweets from twitter. Twitter data has been applied to address a wide range of applications (e.g., political election prediction and disease tracking); however, no studies have been conducted to explore the interactions and potential relationships between twitter data and social events available from government entities. Put into a larger social context, such research is important because government entities often work within limited resources to serve their constituents. The results from the above analysis can facilitate government entities and public service organizations to better understand the people they serve and the effect of their actions, as well as to identify potential issues in a timely manner. Our Approach is to download Twitter messages for a particular #hashTag and perform sentiment analysis i.e. to find positive, negative or neutral sense of that tweet using hadoop framework. Each #hashTag may have 1000 of comments and new comments are added every minute, in order to handle so many tweets we are using apache hadoop framework.

2. Related Work

Papers related to twitter sentiment analysis Dmitry Davidov, Oren Tsur & Ari Rappoport. provided a supervised sentiment classification framework which is based on data from Twitter.. By utilizing 50 Twitter tags and 15 smileys used as sentiment labels, this framework avoids the need for labor intensive manual annotation, allowing identification and classification of the diverse sentiment types of short texts. They evaluate the contribution of different feature types of sentiment classification and show that their framework successfully identifies sentiment types of untagged sentences. They utilized 50 Twitter tags and 15 smileys as sentiment labels which allow them to build a classifier for dozens of sentiment types for short textual sentences. In their study they use four different feature types (punctuation, words, n-grams and patterns)^[1] for sentiment classification and evaluate the contribution of each feature type for this task. They showed that their framework successfully identifies sentiment types of the untagged tweets.

Luciano Barbosa provided a 2-step sentiment analysis classification method for Twitter, which first classifies messages as subjective and objective, it further distinguishes the subjective tweets as positive or negative. To better utilize these sources, he verified the potential value of using and combining them, providing an analysis of the provided labels, examine different strategies to combine these sources in order to obtain the best outcome; and, proposed a more robust feature set that captures more abstract representation of tweets, composed by meta-information associated to words and specific characteristics of how tweets are written.

Sascha Narr, Michael H'ulfenhaus and Sahin Albayrak provided examined a language-independent sentiment classification approach. They trained a classifier to label the sentiment polarity specifically of tweets. They used a semi-supervised emoticon heuristic to generate labeled training data. For any language, their approach requires only raw tweets of that language for training and no additional adjustments or intervention. They trained classifiers on tweets of 4 different languages: English, German, French and

Portuguese. For their evaluation, they collected thousands of human-annotated tweets in these 4 languages using Amazon's Mechanical Turk2.

Preslav Nakov, Zornitsa Kozareva, Alan Ritter, Sara Rosenthal, Sara Rosenthal, Theresa Wilson. proposed SemEval-2013 Task 2: Sentiment Analysis of Twitter, which included two subtasks: A, an expression-level subtask, A and B, a message level subtask. They used crowdsourcing on Amazon Mechanical Turk to label a large Twitter training dataset along with additional test sets for Twitter and SMS messages for both subtasks. The primary goal of our SemEval-2013 task 2 has been designed for promoting research that will lead to a better understanding of how sentiment is conveyed through Tweets and SMS messages. Toward that goal, authors created the SemEval Tweet corpus, which contains Tweets (on both training and testing) and SMS messages (for testing only) of sentiment expressions annotated with contextual phrase-level polarity as well as an overall message-level^[4] polarity

Anna Jurek, Yaxin Bi, Maurice Mulvenna provided a lexicon based approach for analysing the sentiments of tweets on twitter. They have provided a algorithm that provides the intensity of the sentiments rather than the positive and negative label. They evaluated evidence-based combining function that supports classification process in cases when positive and negative words co-occur in a tweet. They have illustrated a case study of the relation in between sentiment of twitter post related to English defence league

Erik Cambria provided approach for concept level sentiment analysis for automatic analysis of online opinions of various user's using natural language text by machines to go beyond the mere word level sentiment analysis of texts and provide approaches for opinion mining and sentiment analysis that enable's a more efficient passage from textual information to machine process able data.

3. Conclusion

Sentiment analysis will help the companies get a review about their products form the customers and also from expert on a large scale through social networking sites, it will also help various analysts to see and check what are the sentiments of the people on a particular social topic. At present the system only gives us information about how many people are talking about the particular topic on the social networking sites or how much the topic is trending in the social network, we don't get an idea whether the opinion of the people is positive negative or neutral in order to gather this information we need to perform sentiment analysis on the data. Our proposed system will be a combination of the various systems described in this paper it is expected to perform more efficiently than any other system. Our system is expected to evaluate the sentiments of the users which will help companies improve their product quality.

References

- [1] Dmitry Davidov, Oren Tsur & Ari appoport "Enhanced Sentiment Learning Using Twitter Hashtags and Smileys," Coling 2010: Poster Volume, pages 241–249, Beijing, August 2010.
- [2] Luciano Barbosa, "Robust Sentiment Detection on Twitter from Biased and Noisy Data" Coling 2010: Poster Volume, pages 36–44, Beijing, August 2010
- [3] Sascha Narr, Michael H'ulfenhaus and Sahin Albayrak "Language-Independent Twitter Sentiment Analysis"
- [4] Preslav Nakov, Zornitsa Kozareva, Alan Ritter, Sara Rosenthal, Sara Rosenthal, Theresa Wilson. " SemEval-2013 Task 2: Sentiment Analysis in Twitter" Preslav Nakov, Zornitsa Kozareva, Alan Ritter, Sara Rosenthal, Sara Rosenthal, Theresa Wilson. " SemEval-2013 Task 2: Sentiment Analysis in Twitter"
- [5] Anna Jurek, Yaxin Bi, Maurice Mulvenna " Twitter Sentiment Analysis for Security-Related Information Gathering" 2014 IEEE Joint Intelligence and Security Informatics Conference pages 48-55
- [6] Erik Cambria, " Sentiment Knowledge Discovery in Twitter Streaming Data,"

Author Profile

Nishad Patil B.E. final year student from MMIT, lohegaon, Pune affiliated to Savitribai Phule Pune University.

Sayali Tingre B.E. final year student from MMIT, lohegaon, Pune affiliated to Savitribai Phule Pune University.

Swapnil Shivshetty B.E. final year student from MMIT, lohegaon, Pune affiliated to Savitribai Phule Pune University

Kalyani Thorat B.E. final year student from MMIT, lohegaon, Pune affiliated to Savitribai Phule Pune University.

Shwetal Patil (GUIDE) lecturer in MMIT, Lohegaon , Pune affiliated to Savitribai Phule Pune University