

A Survey Paper on Twitter Opinion Mining

Geetanjali S. Potdar¹, Prof R. N. Phursule²

¹Pune University, Department of Computer Engineering, JSPM's Imperial College of Engineering & Research, Wagholi, Pune, India

²Pune University, Department of Computer Engineering, JSPM's Imperial College of Engineering & Research, Wagholi, Pune, India

Abstract: Million people have primary focus on Social media platforms to share their own thoughts and opinions in regards to their day to day life, business, celebrity entertainments, politics etc. Opinion Mining defined as an Intersection of information retrieval techniques to deal with the opinions expressed in a document. The main goal is solving the problems related to opinions about Politian in newsgroup posts, products, review sites comments on the facebook post, tweeterfeel blogging and twitter etc. Due to the high usage of internet to share thoughts and opinion rich web resources such as discussion representation, review sites, blogs and news web applications available in digital format , much of the current research is give the attention on the area of sentiment analysis. A wide range of features and methods for training sentiment classifiers for Twitter datasets have been researched in recent years with varying results. This paper presents the methods for opinion extraction, classification techniques and summary details.

Keywords: Opinion Mining, Twitter, Sentiment Analysis on opinion, Social Media, classification

1. Introduction

Due to the growth of internet is 100 % per year many user access and share data on Internet every day. Twitter, with nearly 600 million users and over 250 million messages per day, has quickly become a gold mine for organizations to monitor their reputation and brands by extracting and analyzing the sentiment of the Tweets posted by the public about them, their markets, and competitors. Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. social network is a graph consisting of nodes and links used to represent social relations on social network sites [17]. In Fig. 1 node represents entities and link represents the link between entities.

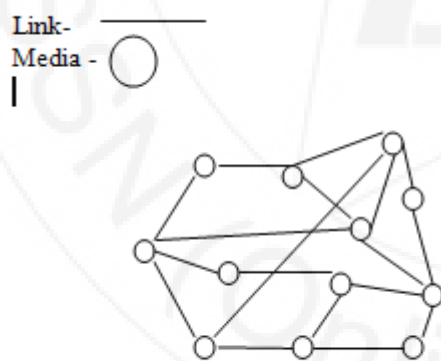


Figure 1: Social Media Network showing nodes and links

Sentiment analysis over Twitter data and other similar micro-blogs faces several new challenges due to the typical short length and irregular structure of such content. Following are some challenges faced in sentiment analysis of Twitter feeds [1]

- Named Entity Recognition (NER):- NER is the method of extracting entities such as people, organization and locations from twitter corpus.

- Anaphora Resolution:- the process of resolving the problem of what a pronoun or noun phrase refers to. "We both had a dinner and went for a walk, it was awful". What does "It" refers to?
- Parsing:- the process of identifying the subject and object of the sentence. The verb and adjective are referring to what?
- Sarcasm:- Sarcasm means what does a verb actually stand for? Does 'bad' mean bad or good?

A generic framework of sentiment analysis where a sentiment engine receives feedback (data) from different channels and then a unique algorithm categorizes (positive/negative) them by assigning scores. The results can be used to draw various types of graphs which are presented in the dashboard is presented in Fig 2

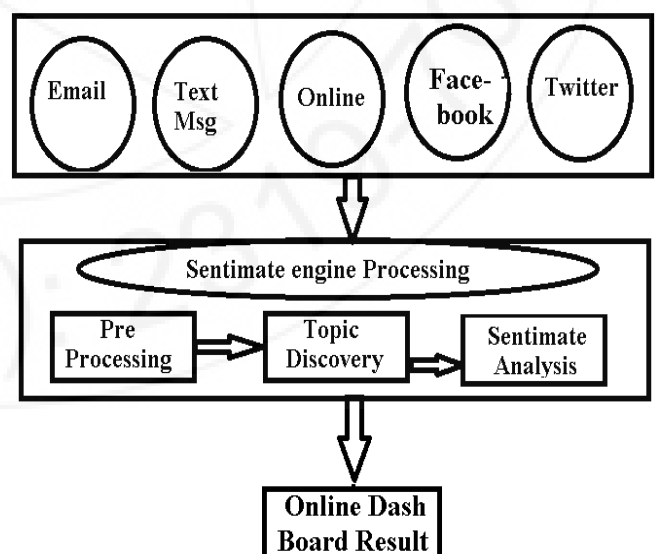


Figure 2: Shows the details of Generic sentiment analysis framework.

2. Literature Survey

1] TOM: Twitter opinion mining framework using hybrid classification scheme, *Decision Support Systems* (Sept 2013)

This paper explains the Introduces and implements a hybrid approach for determining the Sentiment of each tweet. The comparison with other techniques to prove the effectiveness of the proposed hybrid approach. It resolves the data sparsity issue using domain independent techniques.

2] Twitter as a Corpus for Sentiment Analysis and Opinion Mining, *Alexander Pak, Patrick Paroubek.* (2014)

This paper use a dataset formed of collected messages from Twitter. Twitter contains a very large number of very short messages created by the users of this micro blogging platform. The contents of the messages vary from personal thoughts to public statements. Present a method to collect a corpus with positive and negative sentiments, and a corpus of objective texts. Our method allows collecting negative and positive sentiments such that no human effort is needed for classifying the documents. Objective texts are also collected automatically.

3] Sentiment Analysis and Opinion Mining: A Survey, *IJARCSSE June 2012*

The performance of different methods used for opinion mining is evaluated by calculating various metrics like precision, recall and F-measure. Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. The two measures are sometimes used together in the F1 score (also F-score or F-measure) is a measure of a test's accuracy.

4] Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter *IEEE Computer Society 2014.*

The paper states the temporal Latent Dirichlet Allocation and co-occurrence analysis to analyze and validate the relationship between topics in tweets and related events on the twitter. In co-occurrence analysis, all twits are stored in database then visualize the user network. The limitations of LDA are compensated by co-occurrence retrieval. The author explained how co-occurring terms on a particular social issue can be a useful feature for content analysis. The topic model and co-occurrence shows the various aspects of social issues on the Twitter.

5] Sowmya Kamath S, Anusha Bagalkotkar, Ashesh Khandelwal, Shivam Pandey, Kumari Poornima," Sentiment Analysis Based Approaches for Understanding User Context in Web Content", 978-0-7695-4958-3/13, 2013 IEEE.

The paper explains the natural language procedure for decision making and classifying the text on web content. The Sentiment analysis technique works with the help of natural

language procedure as well as computational linguistics. Sentiment approach extracts information about comments expressed by any one regarding with particular subject. This helps other to classify them in order to review classification, any product review mining. The paper contains different approaches regarding with sentiment analysis in the user's context.

6] Diman Ghazi, Diana Inkpen & stan Szpakowicz "Hierarchical Approach to Emotion recognition & classification in texts" 2010 Springer.

The author stated the automatic classification of text using neutrality, polarity and emotions hierarchy. The novel method used, construct a hierarchy for classification on the basis of relation between polarity and text emotions. The polarity is analyzed as positive and negative. They have also compared it to corpus-based and lexical-based feature sets to outperform the flat approach of analyzing.

7] "Wikipedia Based Semantic Smoothing For Twitter Sentiment Classification" by Dilara Totunoglu, Gurkan Telsereen, Ozgun Sagturk & Murat C.Ganiz *IEEE(2013).*

The naïve bays algorithm is used for classification of text when datasets are large. But in case of twits, it has limitations. In this paper the smoothing technique is explained. They have extended twitter sentiment140 dataset Wikipedia article titles, categories and redirects. The semantic smoothing approach statistically maps topic signatures i.e. important concepts in all documents into single-word features and then using expectation maximization algorithm feelings of twits are classified into positive and negative.

8] Survey on Opinion Mining and Summarization of User Reviews on Web by Prof. D.D. Londhe 2014.

This paper discussed different opinion classification and summarization approaches, and their outcomes. This study shows that machine learning approach works well for sentiment analysis of data in particular domain such as movie, product, hotel etc., while lexicon based approach is suitable for short text in micro-blogs, tweets, and comments data on web. Due to applications of opinion detection in various domains such as product, travel, movie etc, it is emerged as a popular topic in web mining.

9] Opinion Mining: Issues and Challenges (A survey) *Baroque Azam Department of Computer Engineering National University of Sciences and Technology, Volume 49- No.9, July 2012*

The author stated that Opinion mining is equally important for companies and helps them to know what customers think about their products. Therefore company's managers can take decisions based on the opinion of customers about their products. According to the customers need and their opinion relationship by giving them exactly what they need. The companies can find regular users and irregular user and retain customers; they can save on production costs by utilizing customer requirements. In this paper we have discussed the

issues which are faced to do opinion mining from web data and the related work that has been done to deal with these issues.

10] Survey on Mining Subjective Data on the Web by Mikalai Tsytsarau · Themis Palpanas 08 Sept 2011.

We now turn our attention to the following interesting question: whether the subjective data that exist on the web carry useful information. Information can be thought of as data that reduce our uncertainty about some subject. According to this view, the diversity and pluralism of information on different topics can have a rather negative role.

3. Overall Survey

3.1 Opinion Mining

It is kind of web content mining. A If a set of text documents (D) are given, that have opinions on an object, opinion mining intends to identify attributes of the object on which opinion have been given, in each of the document $d \in D$ and to find orientation of the comments i.e. whether the comments are positive or negative.

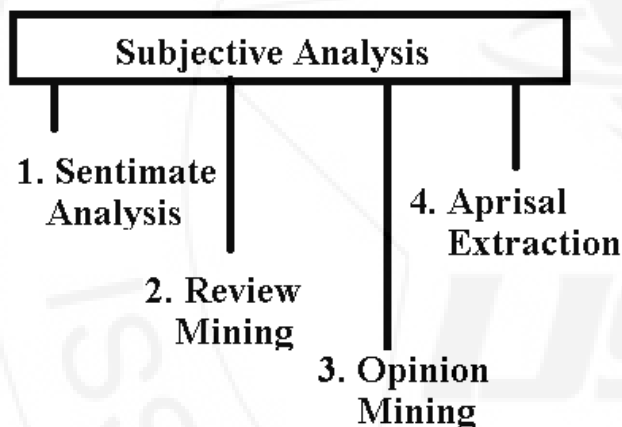


Figure 3: Related techniques of Opinion Mining.

4. Future Enhancement

Future research directions include the development of a web application in order to compare the performance of Opinion Mining with other applications like Tweet Feel & Sentiment140 and the use of supervised learning algorithms to further increase the accuracy.

5. Conclusion

During the past decade, we have focus on Web text data mining. In this work, we presented an overview of a special class of web mining, that of Subjectivity Analysis. More specifically, we reviewed the most prominent approaches for the problems of Opinion Mining and related work analysis like sentimate analysis, Review analysis and aprisal Extraction.

References

- [1] TOM: Twitter opinion mining framework using hybrid classification scheme, Decision Support Systems (Sept 2013)
- [2] Twitter as a Corpus for Sentiment Analysis and Opinion Mining, Alexander Pak, Patrick Paroubek. (2014)
- [3] Sentiment Analysis and Opinion Mining: A Survey, IJARCSE June 2012
- [4] Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter sIEEE Computer Society 2014.
- [5] Sowmya Kamath S, Anusha Bagalkotkar, Ashesh Khandelwal, Shivam Pandey, Kumari Poornima, "Sentiment Analysis Based Approaches for Understanding User Context in Web Content", 978-0-7695-4958-3/13, 2013 IEEE.
- [6] Diman Ghazi, Diana Inkpen & stan Szpakowicz "Hierarchical Approach to Emotion recognition & classification in texts" 2010 Springer.
- [7] "Wikipedia Based Semantic Smoothing For Twitter Sentiment Classification" by Dilara Totunoglu, Gurkan Telsezen, Ozgun Sagturk & Murat C.Ganiz IEEE(2013).
- [8] Survey on Opinion Mining and Summarization of User Reviews on Web by Prof. D.D. Londhe 2014.
- [9] Opinion Mining: Issues and Challenges (A survey) Baroque Azam Department of Computer Engineering National University of Sciences and Technology, Volume 49- No.9, July 2012
- [10] Survey on Mining Subjective Data on the Web by Mikalai Tsytsarau · Themis Palpanas 08 Sept 2011.