

# Paper on Opinion Mining to Predict Election Result

Ketan Patel<sup>1</sup> Kharadi Brijal G<sup>2</sup>

<sup>1</sup>Associate Professor <sup>2</sup>ME Student

<sup>1,2</sup>Department of Computer Engineering

**Abstract**— Use of social networking sites, as a Twitter, for different purpose continues to grow its first appearance. This social net is a micro blogging site to share short messages on a variety of topics. In particular, political elections are a very interesting field to change. Million or thousand people have primary focus on Social media platforms to share their own think or opinion in to their day to day life, business, celebrity entertainments, politics etc. Opinion Mining is defined as an Intersection of information retrieval techniques to deal with the opinions expressed in a document. Our results show that Twitter is used for Political discussion, and that the references to the different political parties correlate, significantly, with the votes of the electors. The main goal is solve the problems related to opinions about Politian in newsgroup posts, of products, review sites comments on the, twitter blogging and twitter etc. Due to the high usage of internet to share thoughts and opinion rich web sources as a review sites, blogs and news web applications available in digital format.

**Key words:** Opinion Mining, Machine Learning

## I. INTRODUCTION

The proposed work is to collect information from social networking sites like Twitter and the same is used for sentiment analysis. Outline people, appraisals, feelings or sentiments toward entities, events and their properties. Impelled by this growth, companies, media, review groups are progressively seeking ways to mine Twitter for information about what people think and feel about a particular product or serve.

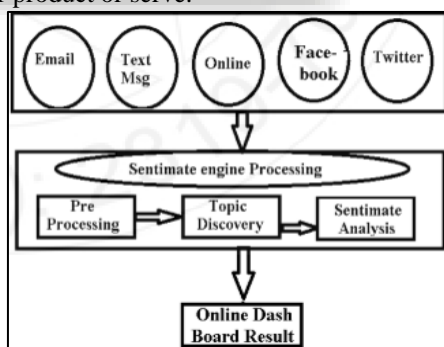


Fig. 1: The details of Generic opinion mining framework.

Now the age of Internet has changed the way people express their views, opinions.. Millions of people are using social network sites like Face book, Twitter, Google Plus, etc. to express their emotions, opinion ,feelings and sharing views about their daily lives. Twitter generates gigantic data that cannot be handled manually hence the requirement of automatic categorization. Twitter data is a valuable source of information for marketing intelligence and trend analysis in all industries. Through the online communities, we get an interactive media where consumers inform and influence others through forums. Social media is generating a large volume of sentiment rich data in the form of tweets, status updates, blog posts, comments, reviews,

etc. Micro blogging today has become a very popular communication tool among Internet users. Millions of messages are appearing daily in popular web-sites.

Micro blogging platforms are used by different people to express their opinion about different topics, thus it is a valuable source of people's opinions.

Twitter's audience varies from regular users to celebrities, company, representatives, and politicians.

## II. MOTIVATION OF WORK

Opinion mining is the one of the most popular trends in today's world. Now Current works in this field which uses a mathematical approach using algorithms for opinion polarity are based on a classifier trained using a collection of annotated text data. Before training, data is preprocessed so as to extract only the main content .Some of the classification methods have been proposed are Naïve Bays, Support Vector Machines, K-Nearest Neighbors etc. Continuous research is being done to determine most efficient method for opinion mining.

With the population of blogs and social networks, opinion mining and sentiment analysis became a field of interest for many researches. The authors applied SVM and CRF learners to classify sentiments at the sentence level and then investigated several strategies to determine the overall sentiment of the document. As the result, the winning strategy is defined by considering the sentiment of the last sentence of the document as the sentiment at the document level.

## III. APPROACHES FOR SENTIMENT ANALYSIS

Machine Learning Approaches Machine learning based approach uses classification technique to classify text into classes. There are mainly two types of machine learning techniques

### A. Unsupervised Learning

It does not consist of a category and they do not provide with the correct targets at all and therefore rely on clustering.

### B. Supervised Learning

It is based on labeled dataset and thus the labels are provided to the model during the process. These labeled dataset are trained to get meaningful outputs when encountered during decision- making. The success of both this learning methods is mainly depends on the selection and extraction of the specific set of features used to detect sentiment. The machine learning approach applicable to sentiment analysis mainly belongs to supervised classification.

In a machine learning techniques, two sets of data are needed:

- Training Set
- Test Set

A number of machine learning techniques have been formulated to classify the tweets into classes. Machine learning techniques like Naive Bayes (NB), maximum entropy (ME), and support vector machines (SVM) have achieved great success in sentiment analysis. Machine learning starts with collecting training dataset. Next we train a classifier on the training data. Once a supervised classification technique is selected, an important decision to make is to select feature. They can tell us how documents are represented. The most commonly used features in sentiment classification are

- Term presence and their frequency
- Part of speech information
- Negations
- Opinion words and phrases

#### IV. CONTRIBUTION

- 1) We present a method to collect a corpus with positive and negative sentiments, and a corpus of objective texts. Our method allows to collect negative and positive sentiments such that no human effort is needed for classifying the documents.
- 2) We perform statistical linguistic analysis of the collected corpus.
- 3) We use the collected corpora to build a sentiment classification system for micro blogging.
- 4) We conduct experimental evaluations on a set of real micro blogging posts to prove that our presented technique is efficient and performs better than previously proposed methods.

##### A. What is Opinion mining to Predict Election Result data?

Opinion mining and sentiment analysis refer to the identification and the aggregation of attitudes or opinions expressed by internet users towards a specific topic. due to the limitation in terms of characters and the use of informal language, the state-of-the-art approaches of sentiment analysis present lower performances in Twitter than that when they are applied on longer texts. Sarcasm is when a person conveys implicit information, usually the opposite of what is said, within the message he transmits, We also study the importance of detecting sarcastic tweets automatically, and demonstrate how the accuracy of sentiment analysis.

Social media constitute a challenging new source of information for intelligence gathering and decision making. Twitter is one of the most popular social media sites and often becomes the primary source of information. Twitter messages are short and well suited for knowledge discovery.

#### V. SCOPE FOR FUTURE WORK

##### A. Deployment in Cloud

It is an arduous task for average system to process, analyze and store huge amount of data sets. Hence there is a requirement of powerful machines which are made available through cloud platforms (IAAS or VPC). Deploying the whole system in cloud provides hassle-free access to it. B. Secure e-Voting System: A secure mechanism is required for expressing people's consensus on policy initiatives and electoral procedures in countries that follow direct democracy. Direct democracy involves people's opinion in

government decision making by conducting regular referendums involving people casting their votes in polling booths. These referendums require physical presence of voters in polling stations in frequent intervals causing disruption in their daily routines. Whereas a secure model having similar architecture can provide a mechanism to cast people's vote from their home.

#### VI. PROPOSED SYSTEM

Proposed system various techniques have been used to do sentiment analysis or opinion mining of tweets. The proposed system contains various phase of development. A dataset is created using twitter posts of movie reviews. As we know that tweets contains slang words and misspelling. So we perform a sentiment level sentiment analysis on tweets. This is done in three phases. In the first phase preprocessing is done. Feature vector is created using relevant features. Finally using different unsupervised learning techniques, tweets are cluster into positive and negative level. The supervised approach can be categorized as corpus-based methods as it uses labeled data to train sentiment classifiers. Given the difficulties of supervised sentiment analysis, it is conceivable that unsupervised approach to sentiment classification is even more challenging.

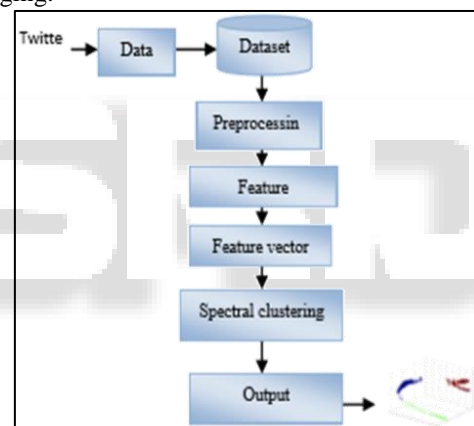


Fig. 1: System model for opinion mining on twitter data.

##### A. Applications of Opinion Mining

Sentiment Analysis has many applications in various Fields.

- 1) Applications that use Reviews from Websites: Today Internet has a large collection of reviews and feedbacks on almost everything. This includes product reviews, feedbacks on political issues, comments about services, etc. there is a need for a sentiment analysis system that can extract sentiments about a particular product or services. It will help us to automate in provision of feedback or rating for the given product, item, etc. This would serve the needs of both the users and the vendors
- 2) Applications as a Sub-component Technology A sentiment predictor system can be helpful in recommender systems as well. The recommender system will not recommend items that receive a lot of negative feedback or fewer ratings. In online communication, we come across abusive language and other negative elements. These can be detected simply by identifying a highly negative sentiment and correspondingly taking action against it.

- 3) Applications in Business Intelligence It has been observed that people nowadays tend to look upon reviews of products which are available online before they buy them. And for many businesses, the online opinion decides the success or failure of their product. Thus, Sentiment Analysis plays an important role in businesses. Businesses also wish to extract sentiment from the online reviews in order to improve their products
- 4) Applications across Domains: Recent researches in sociology and other fields like medical, sports have also been benefitted by Sentiment Analysis that show trends in human emotions especially on social media.
- 5) Applications in Smart Homes are supposed to be the technology of the future. In future entire homes would be networked and people would be able to control any part of the home using a tablet device. Sentiment Analysis would also find its way in IoT. By tracking public views, important data regarding sales trends and customer satisfaction can be extracted.

## VII. CONCLUSION

Opinion Mining is a very wide branch for research. We have covered some of its important aspects. The same architecture could be used for a variety of applications designed to look at Twitter data, such as identifying spam accounts, or identifying clusters of keywords. Taking the system even further, the general architecture can also be expanded to other social media platform usages like Facebook, movie reviews, personal blogs, etc. Evidently, taking into account all the constraints, this method is one of the most efficient ways to perform opinion mining in real time. we can conclude that Twitter may be a valid tool to predict the outcome of political elections. Of course, using it alongside many other well-known tools, such as polls, and with due precautions, given that we cannot forget we are relating two events with very different characteristics. In this way, we think that as future work, there are some aspects which we must take into account. One of them is try to find out whether the mentions are positive or negative.

## REFERENCES

- [1] Alexander Pak, Patrick Paroubek Universite´ de Paris-Sud, Laboratoire LIMSIS-CNRS, Baˆtiment 508, F91405,Orsay,Cedex,Francealexpak@limsi.fr,pap@limsi.fr. Twitter as a Corpus for Sentiment Analysis and Opinion Mining
- [2] Vishal A. Kharde Department of Computer Engineering ,Pune Institute of Computer Technology, Pune University of Pune (India), S.S. Sonawane Department of Computer Engg, Pune Institute of Computer Technology, Pune University of Pune (India), Sentiment Analysis of Twitter Data: A Survey of Techniques
- [3] Narahari P Rao#1, S Nitin Srinivas#2 and Prashanth C M\*3 # B.E, Dept of CS&E, SCE, Bangalore, India \* Prof & HOD Dept of CS&E, SCE, Bangalore, India, Real Time Opinion Mining of Twitter Data
- [4] Geetanjali S. Potdar1, Prof R. N. Phursule2 1Pune University, Department of Computer Engineering, JSPM's Imperial College of Engineering & Research,

Wagholi, Pune, India , A Survey Paper on Twitter Opinion Mining.

- [5] Muqtar Unnisa Deccan College of Engineering and Technology Darussalam Hyderabad TS , Ayesha Ameen Associate Professor IT Dept Deccan College of Engineering and Technology Darussalam Hyderabad TS , Syed Raziuddin, PhD Professor & HOD of CSE Dept Deccan College of Engineering and Technology Darussalam Hyderabad TS , Opinion Mining on Twitter Data using Unsupervised Learning Technique.
- [6] Juan M. Soler \*, Fernando Cuartero\*, Manuel Roblizo† \*Instituto de Inform´atica de Albacete †Facultad de Educaci´on Universidad de Castilla-La Mancha, Campus Universitario. 02071-Albacete, Spain, Twitter as a Tool for Predicting Elections Results