

# Review for Twitter Sentiment Analysis

Anand Kumar Seth, Arshdeep Singh, Mudit Jain

**Abstract**— Social networking sites, micro-blogging sites like twitter have millions of people share their thoughts daily in the form of tweets. As tweet is characteristic short and simple manner

of expression. So the sentiment analysis of Twitter is discussed in this review paper. The Sentiment Analysis perceive as domain of text mining, natural language processing. The study of sentiment analysis can be performed in various aspects. This paper demonstrates sentiment analysis types and techniques used to perform extraction of sentiment from tweets.

**Index Terms**— Opinion Mining, Sentiment Analysis, Text Mining, Twitter

## I. INTRODUCTION

The social networking sites like Twitter, Facebook, MySpace, and YouTube have gained so much popularity now days. They allow people to build connection networks with other people in an easy and timely way and allow them to share various kinds of information and to use a set of services like picture sharing, blogs, etc. Twitter have generated an unequaled public collection of opinions about every global entity that is of interest known as Tweets. Tweets are also called as the microblog because of its short text. Twitter may provision for an excellent channel for opinion creation and presentation. The various research projects that apply sentiment analysis to Twitter corpora in order to extract general public opinion regarding political issues.[1] Their search carried out to use sentiment analysis to gauge the public mood and detect any rising antagonistic or negative feeling on social media. Sentiment mining in twitter can be used for real time applications that range from applications in business (marketing intelligence; product and service bench marking and improvement), applications as sub-component technology (recommender systems; summarization; question answering) to applications in politics.[2] The area of Sentiment Analysis intends to comprehend the opinions and distribute them into the categories like positive, negative, neutral. Till now most sentiment analysis work has been done on review sites [3]. Review sites provide with the sentiments of products or II. SENTIMENT ANALYSIS AND OPINION MINING On micro-blogging websites most of the messages which are user-generated gives textual information, so recognizing their sentiments have become an important issue. The research in this field started

movies, thus, restricting the domain of application to solely business. In the field of twitter sentiment analysis tweets give

us a richer and more varied resource of opinions

sentiments that can be about anything.

with sentiment classification, which can treat the problem as a text classification problem.

In the world Textual information can be classified into two main categories as facts and opinions. Facts are also known as objective statements about entities and events. Opinions are also known as subjective statements that reveal people's sentiments or views about the entities and events. The main difference in facts and opinion is facts are same for all but different people have unlike opinions on the same event or entity. Sentiment analysis is also called as opinion mining which is the field of study as well as analyzing people's

opinions, sentiments, appraisals, evaluations, attitudes, and emotions regarding entities like products, services, organizations, individuals, issues, topics, events for knowledge discovery.

The types of sentiment classification are as follows:

- **Word Level Classification:** In this classification is done on the source of the words which specify the sentiment about the target event. The word may be noun, adjective or adverb. This type of classification gives accurate classified sentiments.
- **Phrase Level Classification:** The phrase refers to mixture of two or more words which are closely related to each other. This classification may give sarcastic classification in the form of good as well as bad category. The classification is done by discovering the sentence phrase denoting the opinion. It sometimes gives incorrect results when a negation word is added in front of the phrase denoting opinion.
- **Document Level Classification:** In this level of classification, single document is considered as opinionated text for analysis. The single review as well as single issue from this document is considered for classification. But sometimes it is not useful in case of blogs as well as forums because customers may compare one product with the other products which has similar characteristics. So the document may consist of the unrelated sentences which don't seem to be opinion about the event. [4]

### III. RELATED WORK

The first investigation of tweet sentiment was done by Go et al. [5] in which they utilized emoticons to annotate tweet with sentiment label. The next study by Agarwal et al. used manually annotated tweets with sentiment and perform unigram model to do classification [6]

The Sentiment Analysis tasks can be done at several levels of granularity, namely, word level, phrase or sentence level, document level and feature level. In the word level sentiment analysis, word sentiments fall into dictionary-based approaches and corpus-based approaches. Further, to automate sentiment analysis, different approaches have been applied to predict the sentiments of words, expressions or documents. These include Natural Language Processing (NLP) and Machine Learning (ML) algorithms [7].

Barbosa et al. [8] classified tweets as objective or subjective and then the subjective tweets were classified as positive or negative. The feature space used included features of tweets like retweet, hashtags, link, punctuation and exclamation marks in conjunction with features like prior polarity of words and POS of words.

In other studies, Pak and Paroubek [9] used a dataset formed of collected messages from Twitter. This paper shows how to automatically collect a corpus for sentiment analysis and opinion mining purposes. This work is able to determine positive, negative and neutral sentiments of documents. The classifier is based on the multinomial Naïve Bayes classifier that uses N-gram and POS-tags as features.

Wang et al. utilized hashtags to perform graph-based classification, while Cui et al. analyzed the emoticon of tweets with graph propagation algorithm for emoticon weighting [10]

Xia et al. [11] used an ensemble framework for Sentiment Classification which is obtained by combining various feature sets and classification techniques. In their work, they used two types of feature sets as Part-of-speech information and Wordrelations as well as three base classifiers i.e. Naive Bayes, Maximum Entropy and Support Vector Machines.

Sentiment analysis has been handled as a Natural Language Processing task at many levels of granularity. Starting from being a document level classification task (Turney, 2002; Pang and Lee, 2004), it has been handled at the sentence level (Hu

Feature Selection: Feature selection consists of following features:

#### A. Lexicon and Micro-blogging as Feature

Based on the subjectivity of the word we can classify the words into positive, negative and neutral lexicons. We have to compare each word with predefined WordNet libraries. By the presence of abbreviations and intensifiers we can classify tweets in positive, negative and neutral. Online available slang dictionaries can be used for emotions and abbreviations.

#### B. Part of speech as Feature

Part of speech features i.e. nouns, adverbs, adjectives, etc. in each tweets are tagged.

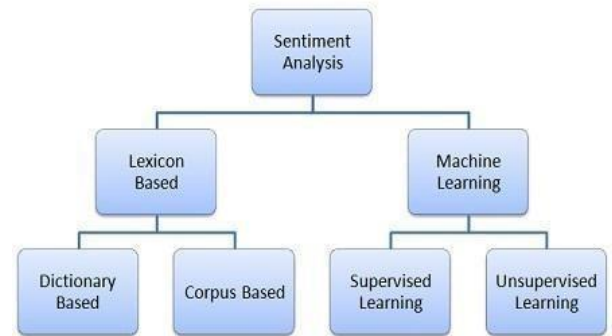


Fig.1: Sentiment Analysis Methods Methods:

**Keyword recognizing:** In this technique the text is categorized based on the pres-ence of fairly unambiguous words present in it. Therefore the words or keywords present in the text have the importance for sentiment analysis.

**Lexical affinity:** For a particular emotion, Lexical affinity assigns arbitrary words a probabilistic similarity.

**Statistical method:** It calculates the valence or target of affective keywords and word co-occurrence frequencies on the base of a large training corpus. In early work it was aim to classify entire document into overall affirmative or negative. These systems mainly depend on supervised learning approaches which depend on manually labeled data. The examples of such systems are movie or product review databases.

**Machine Learning Methods:** One of the important approaches is Machine Learning (ML) used for sentiment analysis mostly belongs to supervised classification in general and text classification techniques in particular. Thus, it is called “supervised learning”. In a machine learning based techniques, two sets of documents are

and Liu, 2004; Kim and Hovy, 2004) and more recently at the phrase level (Wilson et al., 2005; Agarwal et al., 2009).

### IV. SENTIMENT ANALYSIS IN TWITTER

**Data Collection:** Microblogging sites like Twitter, Facebook is used as a source of text. The tweet gathering for major companies is done by using keywords and dates to scrap them. Twitter API used to gather text data.

**Text cleaning:** Case Normalization, Tokenization, Retweet Removal, URL Removal, Stop-words removal, punctuation removal, Reconstruction of Abbreviations, Stemming etc. are few of the techniques used to clean the text.

needed which training set and a test set. The classifier uses a training set to learn the differentiating characteristics of documents, and a test set is used to check the performance of the automatic classifier. Machine learning techniques like

Naive Bayes (NB), maximum entropy (ME), and support vector machines (SVM) have been used in sentiment analysis [12].

**Sentiment Generation:** It is important to find the sentiment of the statement; there are many smart and reliable algorithms to tag statements with complex or simple sentiments. Some sentiment algorithm would also give sentiments like anxiety, sadness etc. while the most commonly used are positive, negative and neutral. Natural language processing is used to tag each word with a sentiment and the overall score or sentiment that the

statement would get depends on the word sentiments.

**Prediction:** In this the step prediction is given by sentiments which are analyzed from the text e.g. positive, negative or neutral. By using this we can predict about different events like product review, election situation etc.

## V. CHALLENGES IN SENTIMENT ANALYSIS

Sentiment Analysis is a very challenging task. Following are some of the challenges [13] faced in Sentiment Analysis of Twitter.

### Identifying subjective parts of text:

Subjective parts represent sentiment-bearing content. The same word can be treated as subjective in one case, or an objective in some other. This makes it difficult to identify the subjective portions of text.

For **example:**

1. The language of the Mr. Dennis was very crude.

2. Crude oil is obtained by extraction from the sea beds.

The word „crude“ is used as an opinion in first example, while it is completely objective in the second example.

### Domain dependence [24]:

The same sentence or phrase can have different meanings in different domains. For Example, the word „unpredictable“ is positive in the domain of movies, dramas ,etc, but if the same word is used in the context of a vehicle's steering, then it has a negative opinion.

### Sarcasm Detection:

Sarcastic sentences express negative opinion about a target using positive words in unique way.. **Example:**

“Nice perfume. You must shower in it.”

The sentence contains only positive words but actually it expresses a negative sentiment.

### Thwarted expressions:

There are some sentences in which only some part of text determines the overall polarity of the document.

**Example:**

“This Movie should be amazing. It sounds like a great plot, the popular actors , and the supporting cast is talented as well. “ In this case, a simple bag-of-words approaches will term it as positive sentiment, but the ultimate sentiment is negative.

### Explicit Negation of sentiment:

Sentiment can be negated in many ways as opposed to using simple no, not, never, etc. It is difficult to identify such negations .

**Example:**

“It avoids all suspense and predictability found in Hollywood movies.”

Here the words suspense and predictable bear a negative sentiment, the usage of „avoids“ negates their respective sentiments.

### Order dependence:

Discourse Structure analysis is essential for Sentiment Analysis/Opinion Mining.

**Example:**

A is better than B, conveys the exact opposite opinion from, B is better than A.

### Entity Recognition:

There is a need to separate out the text about a specific entity and then analyze sentiment towards it.

**Example:**

“I hate Microsoft, but I like Linux”.

A simple bag-of-words approach will label it as neutral, however, it carries a specific sentiment for both the entities present in the statement.

### Building a classifier for subjective vs. objective tweets.

Current research work focuses mostly on classifying positive vs. negative correctly. There is need to look at classifying tweets with sentiment vs. no sentiment closely.

### Handling comparisons.

Bag of words model doesn't handle comparisons very well.

**Example:**

"IIT"s are better than most of the private colleges", the tweet would be considered positive for both IIT"s and private colleges using bag of words model because it doesn't take into account the relation towards "better".

**Applying sentiment analysis to Facebook messages.** There has been less work on sentiment analysis on Facebook data mainly due to various restrictions by Facebook graph api and security policies in accessing data.

### Internationalization [16,17].

Current Research work focus mainly on English content, but Twitter has many varied users from across.

## VI. APPLICATIONS OF SENTIMENT ANALYSIS

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. Thus 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 and in turn their reputation and help in customer satisfaction .

#### 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*

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. Recently there has been lot of research going on Internet of Things(IoT). Sentiment Analysis would also find its way in IoT. Like for example, based on the current sentiment or emotion of the user, the home could alter its ambiance to create a soothing and peaceful environment.

Sentiment Analysis can also be used in trend prediction. By tracking public views, important data regarding sales trends and customer satisfaction can be extracted.

## VII. CONCLUSION

The different techniques of data analysis of twitter are discussed including machine learning and lexicon-based approaches. The analysis of Twitter data is being done in various perspectives to mine the sentiment. This study defined

the concept of opinion in a sentiment analysis of Twitter. Sentiment analysis deals with opinions classified into positive, negative and neutral. The study shows that machine learning methods, such as SVM and Naive Bayes have the highest accuracy and can be considered as the baseline learning methods as well as in some cases lexicon-based methods are very effective. More work in future is needed to improve performance measures.

## REFERENCES

- [1] Amit Narote, sohail Shaikh, saville periera, nitin jadhav, platini Rodrigues, "Sentimental analysis on Twitter", International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137.
- [2] [www.nlp.stanford.edu/courses/cs224n/patlai.pdf](http://www.nlp.stanford.edu/courses/cs224n/patlai.pdf)
- [3] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011 Workshop on Languages in Social Media, 2011, pp. 30–38.
- [4] V. S. Jagtap, Karishma Pawar, "Analysis of different approaches to Sentence-Level Sentiment Classification", International Journal of Scientific Engineering and Technology Volume 2 Issue 3, PP: 164-170. [5] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision." In CS224N Project Report, Stanford, 2009, pp. 1-12.
- [6] Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau, "Sentiment Analysis of Twitter Data", Department of Computer Science Columbia University New York, NY 10027 USA. [7] A. Kumar and T. M. Sebastian, "Machine learning assisted Sentiment Analysis", Proceedings of International Conference on Computer Science & Engineering.
- [8] L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitter from Biased and Noisy Data". COLING 2010: Poster Volume, pp. 36-44.
- [9] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320–1326.
- [10] Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, Ming Zhang, "Sentiment analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach", School of EECS, Peking University, Beijing, China.
- [11] R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," Information Sciences an International Journal, vol. 181, no. 6, pp. 1138–1152, 2011.
- [12] Neethu M,S and Rajashree R, "Sentiment Analysis in Twitter using Machine Learning Techniques", 4th ICCCNT 2013 at Tiruchengode, India. IEEE – 31661.