

Sentiment Analysis of Twitter Data Set: A Reviews

Foziya Rizwan ¹, Mr. Mahesh ²

M. Tech (CSE) ¹, Assistant Professor ²

Ganga Institute of Technology and Management, Kablana, Haryana ^{1,2}

Abstract: Twitter is a platform that is widely used to express your opinions and express feelings on different occasions. Emotional analysis is a way to analyze data and restore the emotions it embodies. The Twitter Sentiment Analysis is an application to analyze the sentiment of Twitter data (Tweets) in order to extract the sentiments transmitted by the user. Research in this area has evolved over the past few decades. The reason behind this is that the shape of a hard tweet makes processing difficult. The tweet format is very small, which creates a completely new problem, such as the use of slangs, acronyms, and so on. In this paper, our goal is to review some of the research on emotion analysis research on Twitter, describe the methods and applications used, and describe Python-based methods.

Keyword: Twitter Data, Python, Sentiment Analysis, Opinion Mining, Data Analysis

1. INTRODUCTION

The analysis of meaning is a task for natural language processing and information extraction aimed at obtaining the author's feelings expressed by comments, questions and positive or negative requests by analyzing a large number of documents. In general, emotion analysis is designed to determine the position of the speaker or author on the general tone of the subject or document. In recent years, the explosive growth of Internet usage and the exchange of public opinion has been the driving force behind today's sentiments analysis. The web is a huge repository of structured and unstructured data. Analyzing these data to extract public opinion and potential emotions is a difficult task.

Sentiments Analysis - Also known as Opinion Extraction. Essentially, it is the process of determining the emotional tone behind a series of words that are used to understand the attitudes, opinions and emotions expressed in online references. **Sentiments** analysis is very useful in monitoring social media because it allows us to understand the broader public opinion behind certain topics. The application of emotion analysis is broad and powerful. The ability to extract ideas from social data is a practice commonly used by organizations around the world.

2. SENTIMENT ANALYSIS

Sentiments analysis (also known as prospecting) refers to the use of natural language processing and the analysis of computational texts and linguistics to identify and extract personal information in source materials. **Sentiments** analysis is widely used in a variety of application and social media reviews, from marketing to customer service. In general, **Sentiments** analysis aims to determine the position of the speaker or author in relation to the general context of a topic or document. The attitude may be appreciation or appreciation

(see evaluation theory), **Sentiments** I state (i.e. emotional state of author writing), or expected **Sentiments** communication (i.e., **Sentiments** impact of the author). This hope to have readers, every day, come a lot of data from social networks, blogs and other media, are spread on the World Wide Web.



Fig 1: Sentiment Analysis –for reviews

This huge data contains very crucial opinion related information that can be used to benefit businesses and other aspects of commercial and scientific industries. Manual tracking and extraction of this useful information is not possible, thus, Sentiment analysis is required. Sentiment Analysis is the phenomenon of extracting sentiments or opinions from reviews expressed by users over a particular subject, area or product online. It is an application of natural language processing, computational linguistics, and text analytics to identify subjective information from source data. It clubs the sentiments in to categories like positive or negative. Thus, it determines the general attitude of the speaker or a writer with respect to the topic in context.

3. THE NETWORK: TWITTER

Twitter is an online social networking service and micro blogging service that enables its users to send and read text-based messages called "tweets". Tweets are publicly visible by default, but senders can restrict the message delivery to a limited crowd. Twitter is one of the largest microblogging service having over 326 million registered users as of 10/06/2018. There is a large mass of people using twitter to express sentiments, which makes it an interesting and challenging choice for sentiment analysis. When so much attention is being paid to twitter, why not monitor and cultivate methods to analyze these sentiments. Twitter has been selected with the following purposes in mind.

Twitter is an Open access social network.

- Twitter is an Ocean of sentiments (limited within 140 characters, i.e. high sentiment density).
- Twitter provides user friendly API making it easier to mine sentiments in real-time.



Fig 2- Networking on Twitter

4. BACKGROUND

1. Nakov et al. (2016) article discusses the fourth year of "Moral Analysis of Twitter Tasks". SemEval-2016 Task 4 contains five sub-tasks, three of which are completely different from earlier versions. The first two sub-tasks were the reintroduction of the years in previous years and required them to predict the general sentiments and emotions of a subject in a tweet. These three new sub-tasks focus on two different types of Twitter emotional rating. The first variable uses a five-point measure that gives the classification function an ordinal advantage. The second alternative focuses on the correct assessment of the universality of each interest category, a task called quantitative measurement in supervised learning literature. The task is still very popular, attracting 43 teams.

2. Pontiac et al. (2016) article describes the joint task SemEval 2016 for side-by-side analysis (ASSA), continuing tasks 2014 and 2015. In the third year, the mission provided 19 training groups and 20 data sets tested for 8 languages and 7 domain names, Joint Assessment. Of these data sets, 25 are used for the wholesale level and 14 for the ASSA text level, and the latter is presented for the first time as a sub-task in SemEval. The job attracted 245 participants from 29 teams.

Rosenthal et al. (2017) article describes the fifth year of sentiment analysis in the Twitter task. SemEval-2017 Task 4 continues to restart subtasks for SemEval-2016 Task 4, which include determining the general feeling of tweet, the mood of the quintessential subject and the five-point ranking, and determining the theme in some emotional distribution tweets: again in the order of two points and five points. Compared to 2016, we made two changes: (1) we introduced the new Arabic language for all subtasks, and (2) provided information from the Twitter user profile that published the targeted tweets. This task is still very popular, there are 48 teams taking part this year.

Rosenthal et al. (2015) this article, we describe the 2015 frequency of the shared SemEval task to analyze sentiment in Twitter. This is by far the most important task of sharing emotional analysis, with more than 40 teams participating each year over the past three years. This year's joint task quiz includes five sub-tasks for emotional prediction. Two of them were repeating in previous years: (a) feelings expressed in terms of tweets, and (b) total emotions in tweets. We also include three new subtasks to predict (C) the mood of the subject in one tweet, (D) the general mood of the subject in a set of tweets, and (e) the previous polarization of the term.

5. Pontiac et al. (2015) SemEval-2015 Task 12, a continuation of SemEval-2014 Task 4, aims to encourage post-

sentence search or classification of text level feelings to achieve side-based emotion analysis. The goal is to determine opinions about the expression of a particular entity (for example, a laptop) and its aspects (such as price). The task provides manual comment comments in three areas (restaurants, laptops, hotels), as well as a shared assessment process. Attracted 93 participants from 16 teams.

6. Schouten & Frasincar (2016) the field of sentiment analysis, where the collection, analysis and gathering of sentiments of text drew much attention in the last few years. The corresponding growth in this area has led to the emergence of different subfields, each dealing with different levels of analysis or research issues. The survey focuses on emotional analysis at the side level, with the aim of finding and summarizing the emotions of the entities mentioned in the document or side. An in-depth view of the current state of the art shows tremendous progress in finding goals (which can be entities or some aspects) and corresponding objectives. Emotions. The side-level analysis of intelligence produces very accurate emotional information useful for applications in a variety of areas. Whether the current solution is classified depends on whether the current solution provides side detection, emotion analysis, or both. Additionally, errors based on the type of algorithm being used are provided. For each discussion of the study, including reported performance. In order to facilitate the quantitative assessment of the various methods proposed, standardization of the evaluation method, including the use of shared data sets, must be standardized. Rich fatty aspects, which focus on the concept of emotional analysis are discussed and identified as one of the promising research trends in the future.

7. Saxena et al. (2019) recent years, text mining and analysis of morale received considerable attention. Emotional analysis is one of the main tasks of NLP. Data are often unorganized and noisy, so the task of obtaining information is both complex and costly. There is an increasing need to develop different methods and models for word processing and efficient information extraction. One way to extract information is to extract text and analyze emotion. This article describes various techniques used in text exploration and emotional analysis that separates all tasks. It also classifies the meaning analysis at the wholesale level and analyzes the level of the document.

8. Clombis et al. (2011) paper, we examine the usefulness of language features in discovering emotions in Twitter messages. We evaluate the utility of existing vocabulary resources and the ability to capture information about the informal and creative languages used in Webbo. They use a supervisory approach to solve the problem, but we use topic tags in your Twitter data to create training data.

5. EXPERIMENTAL INVESTIGATION

Analysis of sentiment analysis of twitter data through python and compare it to R-Tool performance. The aim while performing sentiment analysis on tweets is basically to classify the tweets in different sentiment classes accurately. In this field of research, various approaches have evolved, which propose methods to train a model and then test it to check its efficiency. Performing sentiment analysis is challenging on Twitter data, as we mentioned earlier. Here we define the reasons for this:

- **Limited tweet size:** with just 140 characters in hand, compact statements are generated, which results sparse set of features.
- **Use of slang:** these words are different from English words and it can make an approach outdated because of the evolutionary use of slangs.
- **Twitter features:** it allows the use of hashtags, user reference and URLs. These require different processing than other words.

- **User variety:** the users express their opinions in a variety of ways, some using different language in between, while others using repeated words or symbols to convey an emotion.

All these problems are required to be faced in the preprocessing section. Apart from these, we face problems in feature extraction with less features in hand and reducing the dimensionality

6. METHODOLOGY

Methods of Sentiment Analysis



Fig 3: Methods of Sentiment Analysis

• **Data Collection** Consumers usually express their sentiments on public forums like the blogs, discussion boards, product reviews as well as on their private logs Social network sites like Facebook and Twitter. Opinions and feelings are expressed in different way, with different vocabulary, context of writing, usage of short forms and slang, making the data huge and disorganized. Manual analysis of sentiment data is virtually impossible. Therefore, special programming languages like 'R' are used to process and analyze the data.

• **Text Preparation** Text preparation is nothing but filtering the extracted data before analysis. It includes identifying and eliminating non-textual content and content that is irrelevant to the area of study from the data.

• **Sentiment Detection** At this stage, each sentence of the review and opinion is examined for subjectivity. Sentences with subjective expressions are retained and that which conveys objective expressions are discarded. Sentiment analysis is done at different levels using common computational techniques like Unigrams, lemmas, negation and so on.

• **Sentiment Classification** Sentiments can be broadly classified into two groups, positive and negative. At this stage of sentiment analysis methodology, each subjective sentence detected is classified into groups-positive, negative, good, bad, like, dislike.

• **Presentation of Output** The main idea of sentiment analysis is to convert unstructured text into meaningful information. After the completion of analysis, the text results are displayed on graphs like pie chart, bar chart and line graphs.

7. CONCLUSION

The aim while performing sentiment analysis on tweets is basically to classify the tweets in different sentiment classes accurately. In this field of research, various approaches have evolved, which propose methods to train a model and then test it to check its efficiency. Performing sentiment analysis is challenging on Twitter data, as we mentioned earlier. Here we define the reasons for this:

- **Limited tweet size:** With just 140 characters in hand, compact statements are generated, which results sparse set of features.

• **Use of slang:** These words are different from English words and it can make an approach outdated because of the evolutionary use of slangs.

• **Twitter features:** It allows the use of hashtags, user reference and URLs. These require different processing than other words.

• **User variety:** The users express their opinions in a variety of ways, some using different language in between, while others using repeated words or symbols to convey an emotion.

All these problems are required to be faced in the preprocessing section. Apart from these, we face problems in feature extraction with less features in hand and reducing the dimensionality of features.

References

1. Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F., & Stoyanov, V. (2016). SemEval-2016 task 4: Sentiment analysis in Twitter. In *Proceedings of the 10th international workshop on semantic evaluation (semeval-2016)* (pp. 1-18).
2. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A. S., & Hoste, V. (2016, June). Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)* (pp. 19-30).
3. Rosenthal, S., Farra, N., & Nakov, P. (2017, August). SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017)* (pp. 502-518).
4. Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S., Ritter, A., & Stoyanov, V. (2015). Semeval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)* (pp. 451-463).
5. Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., and routsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)* (pp. 486-495).

6. Schouten, K., & Frasincar, F. (2016). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3), 813-830.
7. Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing & Management*, 52(1), 5-19.
8. Kouloumpis, E., Wilson, T., & Moore, J. (2011, July). Twitter sentiment analysis: The good the bad and the omg. In *Fifth International AAAI conference on weblogs and social media*.

