JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Sentiment Analysis of text using SVM

Dipanjali Digambar Chavare

Prof V.D. Jadhav

Abstract- Users of Twitter, one of the most well-known social networking platforms, are free to express their ideas, opinions, and feelings. These tweets are compiled and analyzed to get sentiment data regarding the terrorist assault in Uri. The present study gathers tweets on the Uri attack and examines their polarity and feelings. Tweets are mined for polarity and emotions using text mining techniques. Five thousand tweets are pre-processed and recoded to create a collection of frequently used words. Sentiment analysis is a machine learning technique that looks for positive or negative polarity in texts. By using text samples that represent a range of emotions, machine learning technology can be trained to automatically identify sentiment without the need for human intervention.

Index Terms- twitter, text mining, Sentiment Analysis, Positive, Negative words.

I. INTRODUCTION

Social awareness and the popularity of social networking services like Twitter are rising at the same time. Anyone can tweet about any event on Twitter, a popular and important social media network. On an open platform, people can freely express their ideas, opinions, and feelings. People use Twitter because it's more socially significant, the internet is less expensive, and portable gadgets are more reasonably priced. Most of them tweet about different things that happen. In the age of social networking, people express their feelings and ideas using Twitter. Twitter is therefore incredibly data-rich. We know the duration of every tweet.

Opinion mining, also referred to as sentiment analysis, is the process of analyzing attitudes and feelings found in textual data. Sentiment analysis determines the thoughts and feelings of every individual on a specific event. To build a system or model for sentiment analysis, we need to offer a text document or other analysis-ready material.
give a summary of the main points of the document.
Twitter sentiment analysis is one of the most challenging and recent study areas. Because social media platforms like Twitter provide a substantial amount of text sentiment data in the form of tweets, it might be useful to ascertain people's thoughts or sentiments about particular events.

Opinion mining and sentiment analysis are helpful for reviews of movies, merchandise, customer service, comments about any event, etc. This makes it easier to decide whether or not a certain good or service is chosen. Finding out what other people think about a scenario or a particular person can be useful as well, since it can assist decide whether a text is positive, negative, or neutral. Text can be categorized into different moods with the help of a kind of
Sentiment analysis is a type of text analysis. Sentiment analysis, also referred to as opinion mining, is the process of determining the sentiment or emotional tone expressed in a text, such as positive, negative, or neutral. Machine learning algorithms are often used for sentiment analysis since they are capable of automatically identifying patterns and attributes from data.

II. Literature Review

A hybrid classification technique has been used for the sentiment categorization of movie reviews. Many feature sets and classification algorithms, including Naive Bayes and Genetic algorithms, have been combined in order to assess performance based on accuracy. The study's findings show that hybrid NB-GA outperforms base classifier in terms of efficiency and effectiveness, and that GA outperforms NB in terms of efficiency. [1]

Text mining needs to take the polarity of a document into account. [2] has addressed the subject of tree kernel engineering in the future. This procedure yields better results than the others. Two classification models—a two-way classification and a three-way classification—are defined by the paper's author. Emotions can be classified as positive or negative in a two-way categorization, or as positive, negative, or natural in a three-way classification. For tweet representation, the author takes into account the tree kernel technique. A tree kernel model was the most feature-based and accurate model. By 4%, the experiment performs better

than the unigram model. [2]

A hierarchical sentiment analysis approach can be used for cascaded classification [3]. The author cascaded three classifications—positive against negative, polar against non-polar, and objective against subjective—to produce a hierarchical paradigm. This method was assessed against a four-way (objective, neutral, positive, and negative) categorization model. The comparison shows that the 4-way classification strategy is not as effective as the hierarchical model. [3]

A domain-specific feature-based model for movie reviews has been developed by the author [4]. In this instance, text movie reviews are analyzed using an aspect-based technique, and a sentiment label is applied.The sentiment score for a certain film is then determined by averaging all of the factors across many assessments. The author uses a SentiWordNet-based technique for sentiment analysis and feature extraction at the document level. The algorithmic outcome is contrasted with the results from the Alchemy API. The comparison shows that the Alchemy API technique is not as effective as the feature-based strategy. In short, document-based sentiment results are inferior to aspect-based sentiment results. [4]

A substantial corpus of about 300,000 tweets has been compiled by the author [5] for sentiment analysis and opinion mining. Using a sentiment classification model, tweets can be categorized as good, negative, or neutral. The collected corpus was divided into three groups using this technique: Positive emotions, such joy, happiness, or amusement; Negative emotions, like despair, wrath, or disappointment; and Neutral emotions, which are texts that don't include any emotions. Tree Tagger is used for POS-tagging to distribute emotions.

Information from consumer marketing campaigns is used to forecast future trends and get feedback on items. Because there is a lot of data on customer reviews, the author uses the Hadoop environment to analyze sentiment. Hadoop clusters were constructed as part of a data analysis exercise. Tweets that were neutral, positive, or negative were categorized [6].

Twitter data is analyzed using Hadoop's HIVE and FLUME tools as well. The FLUME application is used to extract data, which is then saved in HDFS format. The HIVE tool is used to extract and analyze data from HDFS-style storage. The HIVE tool facilitates the analysis of many

Additionally, Twitter data is automatically classified into positive, negative, and neutral categories based on the search phrases used in tweets concerning customer reviews. The Parts of Speech (POS) polarity approach and the tree kernel are used by the paper's author. A manual dictionary of emotions and an online lexicon are the two types of resources utilized in the research. The author used a variety of methods for feature extraction and categorization.[8]

In 2019, Saad and Yang [9] aimed for a comprehensive ordinal regression-based machine learning algorithmic analysis of tweet sentiment. The recommended methodology includes a step called pre-processing tweets, and an efficient feature was produced by using the feature extraction model. To categorize the sentiment analysis, methods such as SVR, RF, Multinomial Logistic Regression (SoftMax), and DTs were employed. Furthermore, the suggested model was investigated with Twitter data.

III. Sentiment Analysis

Using the sentiment package, you can analyze two different analysis of positive and negative words. Word clouds were used to record frequently recurring words. These words were used frequently and a feeling was added. For sentiment analysis, these new words and sentiments are added to the sentiment file. Bayes algorithm is used at now. Each word is compared to terms in the sentiment file by the sentiment analysis algorithm, which then assigns a count to each sentiment. Finally, it can show the number of sentiments for each. The current work also identifies text polarity. Positive, negative, or neutral polarity will all exist. In this experiment, new words were found using a word cloud and then given a polarity. Similar to sentiment analysis, it counts the polarity of the text file by comparing each word to the polarity word file. Each polarity's count is displayed last.

Dataset Collection

Tweets about the Product data were retrieved to create the dataset. To retrieve tweets, a tweeter application was developed to obtain Secret. R Studio and the tweeter application are connected via these keys. After connecting, the search term "Product data" produced a collection of 5000 tweets. During pre-processing, duplicate tweets were eliminated, resulting in a final dataset with just 1788 tweets.

Data Preprocessing

The final dataset was made up of raw tweets that needed to be pre-processed in order to yield meaningful information. During processing, stop words and commonly used words such as numbers, prepositions, names, basic verbs, and conjunctions were eliminated from tweets. In sentiment analysis, these words are essentially useless. The following is a list of preprocessing steps.

Filtering: In this phase, tweets are cleaned by eliminating user names, emoticons, and other special words from them. Tokenization: During this stage, tweets are divided up into several tokens.

Stopwords Removal – Stopwords are simply particular common words that are eliminated from tweets because they

lack any analytical significance.

stemDocument – This step gets rid of common word ends like "ing", "es", "s" and so forth.

Remove White Spaces - There are numerous white spaces throughout each text. White spaces are removed in this stage.

Convert to lower case: Text is changed to lower case once all superfluous terms have been removed.

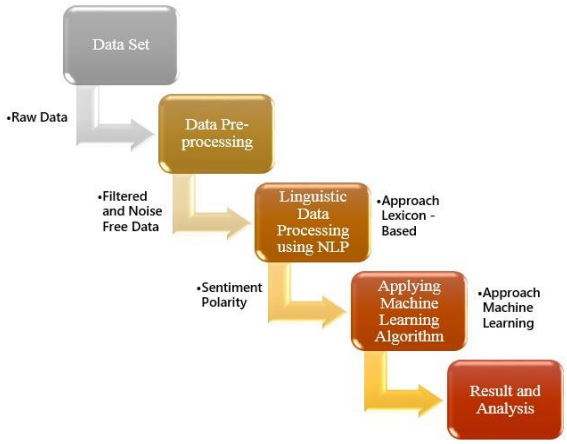

Fig 1 Flow Chart

Logistics Regression

Logistic regression is a popular machine learning technique for binary classification applications. It represents the relationship between a dependent variable and one or more independent factors by estimating the likelihood that the dependent variable belongs to a particular class.

Natural Language Processing

An interdisciplinary topic within linguistics, computer science, and artificial intelligence called "natural language processing" studies how computers and human language interact, with a focus on how to programmed computers to handle and analyze massive amounts of natural language data. Artificial intelligence (AI) includes the field of natural language processing (NLP). In many different commercial disciplines and places, personal assistants employ this technology. This technology processes the user's speech in accordance with its proper understanding by dissecting it. Due to the fact that it is a very current and successful strategy, there is a huge demand for it right now. In the nascent field of natural language processing, advancements like smart device interoperability and interactive human conversations have already been made possible.

Support Vector Machine (SVM)

The supervised (feed-me) machine learning method SVM is useful for problems involving both regression and classification. Regression predicts a continuous value, whereas classification predicts a label or group. By locating the hyper-plane that separates the classes we plotted in n-dimensional space, SVM carries out classification.

CONCLUSION

Machine learning-based sentiment analysis of text has shown to be a useful method for comprehending and extracting sentiment from textual data. Large amounts of labelled data can be used to train machine learning algorithms, which can then use those patterns and features to forecast the future or classify data. We can automatically classify and analyze the sentiment expressed in text, whether it is good, negative, or neutral, by using machine learning models to sentiment analysis tasks. This technique has applications across a number of industries, including brand reputation management, market research, and social media monitoring.

REFERENCES

[1] M.Govindarajan, Sentiment Analysis of Movie Reviews using Hybrid Method of Naive Bayes and Genetic Algorithm , International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970), Volume-3 Number-4 Issue-13 December-2013

[2] Apoorv Agarwal, BoyiXie Ilia Vovsha, Owen Rambow, Rebecca Passonneau, Sentiment Analysis of Twitter Data ‘

[3] Apoor v Agarwal, Jasneet Singh Sabharwal, End-to-End Sentiment Analysis of Twitter Data, Proceedings of the Workshop on Information Extraction and Entity Analytics on Social Media Data, pages 39–44,COLING 2012, Mumbai, December 2012.

[4] V.K. Singh, R. Piryani, A. Uddin,P. Waila, Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification, Conference Paper March 2013, DOI: 10.1109/iMac4s.2013.6526500

[5] Alexander Pak, Patrick Paroubek, Twitter as a Corpus for Sentiment Analysis and Opinion Mining

[6] Ajinkya Ingle, Anjali Kante, ShriyaSamak, Anita Kumari, Sentiment Analysis of Twitter Data Using Hadoop, International Journal of Engineering Research and General Science Volume 3, Issue 6, November-December, 2015, ISSN 2091-2730, www.ijergs.org

[7] Sangeeta, Twitter Data Analysis Using FLUME & HIVE on HadoopFrameWork,Special Issue on International Journal of Recent Advances in Engineering & Technology (IJRAET) V-4 I-2 For National Conference on Recent Innovations in Science, Technology & Management (NCRISTM) ISSN (Online): 2347-2812, Gurgaon Institute of Technology and Management, Gurgaon 26th to 27th February 2016

[8] VarshaSahayak, VijayaShete, ApashabiPathan, Sentiment Analysis on Twitter Data, International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163, Issue 1, Volume 2 (January 2015)

[9] http://www.rstudio.com accessed 10-Feb-2016 [10] https://cran.r-project.org/web/packages/ accessed 15-Feb-2016