Twitter- Treasure of Sentiments

Arti Bansal

Computer Application, Guru Kashi University, (India)

ABSTRACT

With the growth of social media, Sentiment Analysis is now a very attractive research area for researcher. Social media have been used by millions of people for sharing their thought daily. Twitter is a social micro blogging service where users post real time reactions to and opinions about everything. In the real world, if we want to review any product or any issue then twitter provide us lot of information about that and people can know about sentiments of other people easily. This paper tells about the Twitter uses and sentiment which are spread over the twitter by people.

Keywords : Twitter, sentiment, social sites, tumblr

I.INTRODUCTION

Twitter is an online news and social networking service where users posts and interact with messages called tweets. The launch of twitter in 6 Oct 2006 by Jack Dorsey, Noah Glass, Biz Stone and Evan Willians in San fransico, California. Registered users can post tweets but those who are unregistered can only read them. Tweets were originally restricted to 140 characters but on November 7, 2017 the limit was doubled to 280 characters. Twitter is a social microblogging service used by many people for opinions sharing on different topics on a single platform. People use slang and misspell words for tweet. In research area, twitter data is analyzed for opinion mining regarding political issues and product or movie reviews and other areas also. Sentiment analysis has recently received a lot of attention in the NLP field. SA goal is to whether the opinion in a Twitter is positive or negative. SA is also useful in social media monitoring as it allows gaining an overview of the wider public opinion behind certain topic. SA is a way to evaluate tweets to determine if the expression is positive, negative or neutral.

II.RELATED WORK

Burns A. (2016) reviews the origin of twitter. He traces the origin and gradual development of the platform and outlines some of the key contemporary uses of Twitter. He define the synchronous communication between multiple participants who are digitally co-present Twitter have launched in march 2006, initially influenced by SMS but the limitation of Twitter messages not more than140 characters. Retweets were commonly proceeded by "RT@ username" to acknowledge that the messaged send from username. In 2015, company also introduced new feature "quoted tweet" that generates a URL linking to the original tweet page on the Twitter website. In 2007, Hashtags are suggested for brief keyword preceded by the hash symbol. Twitter posts generally text based, further addition to Twitter insertion of images, videos and links to other types of content by URL

pointing to the location. Twitter is particularly well suited to the rapid dissemination and subsequent discussion and evaluation of news report. Twitter as a back channel to broadcast contents or live contents, from popular entertainment through sports to conference. The ecosystem of third party developers and service providers which has emerged around Twitter constitutes a further node in this network.

Devi Lakshmi N. and Rividya K.S. analyze twitter data using unsupervised learning techniques such as K-means clustering and DBSCAN clustering. In first phase pre processing is done and then feature vector is created using relevant feature. In third phase, different unsupervised learning techniques are used to cluster into positive, negative and neutral. It divides the document into sentence and categorize each sentence using word list of each category. They does not require training set and test set. When the data set is small, K-means produce good clustering results. It cannot produce effective results when dataset is large. DBSCAN gives good results in small as well as large datasets. DBSCAN performed outstanding in terms of Accuracy.

Gurini feltoni et.al. recommended user to identify relevant people to follow among million of users of social network. They defined a novel weight function that takes into account sentiment, volume and objectivity related to the user interests. User recommendation are based on their definition of a similarity measure between two users. The results show the benefits of their model compared with some state of art methods. After that they use NLTK for Naïve Bayes classifier to train the classifier. After training the classifier is optimized by removing some unimportant features. At last they compared the result between languages using tweets were human annotated that using Amazon mechanical Truck services. The classifier can be trained effortlessly on new languages, given only raw training data.

Ismail H.M. *et.al.* (2016) compare the performance of different machine learning classifier for twitter sentiment analysis. For this analysis STS (Stanford Testing Documents) dataset is used. They analyze unigram as well as bigram as feature spaces. They analyze TF representation of data set. They evaluate the performance of multinomial NB, Bernoulli NB and SVM in sentiment mining. They choose WEKA for evaluating the performance of the selected classifiers. The overall accuracy for bigram datasets. Training time for unigrams dataset is in general less than bigrams. Multinomial NM produced the best results with frequency unigram dataset. Unigram as a form of representing dataset feature proved to be more effective in the context of Twitter sentiment analyzes as they produce less sparse dataset.

Ivan hibernal et.al. deep study of machine learning methods for sentiment analysis of Czech social media. They focus on document level sentiment analysis performed on three different data sets such as facebook, movie and product reviews. They first preprocess the data by streaming of data and then features are extracted by TF-IDF and then classifiers are performed for getting results. They used Maximum Entropy and SVM and use pure Java framework for machine learning. They significantly outperformed the baseline in three class classification and achieved F-measure 0.69 using a combination of features and preprocessing techniques.

Kumar Akshi and Jaiswal Arunina study twitter and TUMBLR are two social networking websites for expressing their emotions and opinions. Twitter has limited to 140 characters for tweet but Tumblr have no limitation for blog. They used NB, SVM, MLP, DT, KNN, Fl for sentiment analysis of Twitter and Tumblr posts. In this paper, the higher accuracy and precision is achieved using SVM for both Twitter and Tumblr.

Naer sascha et.al. Analyses twitter data of mixed language by using Naïve Bayes classifier. They collect thousand of tweets for four different languages: English, German, French and Portuguese. They assume that any one tweet only contains one opinion at a time. They used Naïve Bayes classifier on N-gram features to classify sentiments in tweets. Then data is tokenized using an extended version of O'Connor's regular expression based tweet tokenize. They use noise polarity class using ui and uj.

Pak Alexander and Paraoubek Patrick evaluate on posts for negative and positive sentiments. Corpus is collection of text posts with positive, negative and neutral sentiments. They use English language. After the collections of corpus, Ziff's law is applied for the plotting of word frequency distribution. After the distribution of words frequency features are extracted to train sentiment classifier. They experimented with unigram, bigram and trigram. Trigram should better capture patterns of sentiment expressions and unigram should provide a good coverage of data. They used SVM, Naïve Bayes and CRF classifier for classification. They used Shannon entropy and salience for calculation. As result best performance is achieved by bigrams. Bigram provide good balance between coverage and ability to capture the sentiment expression.

Wand yilin et.al. Analysis the social media images for sentiment classification using unsupervised sentiment analysis (USEA) framework for social media images. The proposed method does not assume the availability of label information but employ auxiliary textual information.

III.ANALYSIS

In this paper, we take the data of US airlines from Dataworld.com for analysis. In this paper data is about negative sentiments of customers of US airlines which book their tickets and travel from US airlines. Negative sentiments such as late flight, can't tell, cancelled flight, lost luggage, bad flight, flight booking problems, flight attendant problem, loglines and damage are some problems which are faced by customers of US airlines. In the diagram data is arranged in the form of bar diagram for representing the problems of customers. Customer service issues are the main problem in US airline.
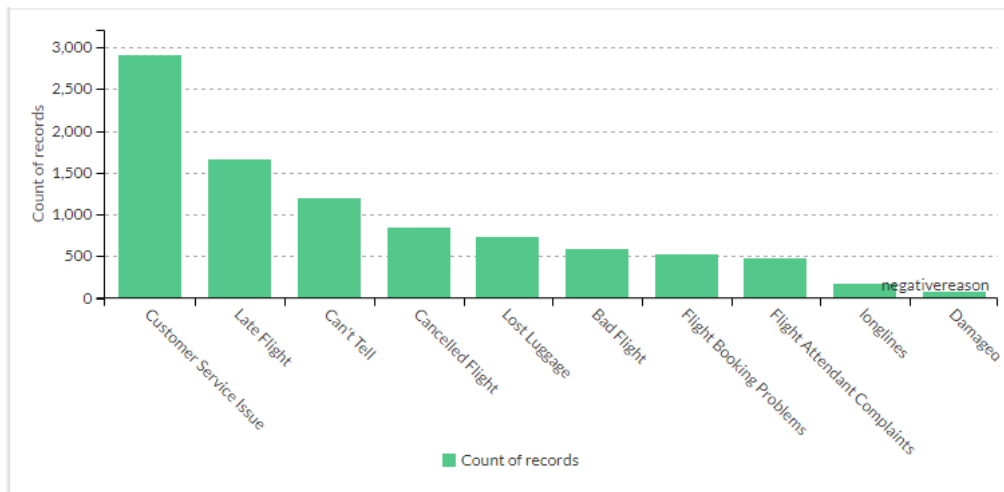
IV.CONCLUSION

This paper study about the origin and feature of the Twitter. Twitter is a micro blogging service which is used by many users for expressing the sentiments. It provides a common platform to various user for sharing different opinion. This paper analyze the US airline data by negative sentiments.

REFERENCE

[1]Bruns, Axel. "Real-Time Applications (Twitter)." Handbuch Soziale Praktiken und Digitale Alltagswelten (2016): 1-9.

[2]Devi, N. Lakshmi. "Analyzing Twitter Data Using Unsupervised Learning Techniques." *Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org* 7.11 (2017).

[3]Gurini, Davide Feltoni, et al. "A Sentiment-Based Approach to Twitter User Recommendation." *RSWeb@ RecSys*. 2013.

[4]Habernal, Ivan, Tomáš Ptáček, and Josef Steinberger. "Sentiment analysis in czech social media using supervised machine learning." *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*. 2013.

[5]Ismail, Heba, Saad Harous, and Boumediene Belkhouche. "A Comparative Analysis of Machine Learning Classifiers for Twitter Sentiment Analysis."*Research in Computing Science* 110 (2016): 71-83.

[6]Kumar, Akshi, and Arunima Jaiswal. "Empirical Study of Twitter and Tumblr for Sentiment Analysis using Soft Computing Techniques." *Proceedings of the World Congress on Engineering and Computer Science*. Vol. 1. 2017.

[7]Narr, Sascha, Michael Hulfenhaus, and Sahin Albayrak. "Language-independent twitter sentiment analysis." *Knowledge discovery and machine learning (KDML), LWA*(2012): 12-14.

[8]Pak, Alexander, and Patrick Paroubek. "Twitter as a corpus for sentiment analysis and opinion mining." *LREc*. Vol. 10. No. 2010. 2010.

[9]Wang, Yilin, et al. "Unsupervised Sentiment Analysis for Social Media Images." *IJCAI*. 2015.