Real-Time Sentiment Analysis of Twitter Data

Shaik Hamed¹, Mohammed Riyan Ali Hashmi², Mohammed Firasath³, Pathan Ahmed Khan⁴

¹, ², ³ UG Student, ⁴ Assistant Professor, Department of Computer Science & Engineering, ISL Engineering College, Hyderabad, Telangana, India

Abstract: The Internet has been a promising forum for online learning, exchanges of ideas and sharing views with the emerging social networking era and its development. Social media provide a large quantity of feeling information in emails, blogs, status updates and posts. This article uses the most common twitter microblogging site. The Twitter emotion analysis is an extension of Twitter (tweet) feel analysis to extract views and feelings from users. The key aim is to examine how methods of text analysis may be used to delve through some of the information in a series of articles that concentrate on tweet language patterns, tweet twitter volumes. Experimental assessments demonstrate the efficiency and precision of the proposed machine classifiers. Python implements the proposed algorithm. The paper focuses about how a feeling study has been designed and a large number of tweets extracted. In this development prototyping is used. Results characterise the view of consumers by tweets as 3 types of positive and 3 types of hate as seen in a pie chart and in an HTML tab.

Keywords: Machine Learning, Natural Language Processing, Python, Sentimental Analysis

I. INTRODUCTION

The use of microblogging websites like Twitter has grown enormously in recent years. Driven by that growth, businesses and media groups are looking more and more for ways to get Twitter to know what people think and feel about their goods and services. Companies like Twitratr (twitrratr.com), twitfeel (www.tweetfeel.com), and Social Note are those that advertise sentiments interpretation from Twitter as one of their services. (www.socialmention.com).

While there has been a good deal of investigation into how feelings are conveyed in genres such as web reviews and news stories, there has been much less explored in the context of casual language and microblogging's message-length constraints. Features like automated part-of-speeching tags and tools like feeling lexicons have been useful for sentiment analysis in other areas, but are they still useful for Twitter sentiment analysis? We start to look at this issue in this article. The sentiment analysis refers to the wide field of linguistic treatment which covers the analysis of the views, sentiments and emotions expressed in the text. Pathology Analysis (SA) or Viewpoint Mining (OM) helps to educate people about their beliefs, perceptions and feelings. The object may describe people, activities or subjects. In the field of sentiment analysis there has been a huge amount of testing. But most of them concentrated on classifying structured and broader text details as evaluations. Since social networking and microblogging platforms are well known, and since there are an enormous number of data from these tools, sentiment analysis testing programmes have seen a change in domain gradually. The use of microblogging websites has grown enormously in recent years. Popular websites such as Twitter have developed into a source of diverse knowledge. This knowledge diversité is due to the importance of these microblogs as forums where people post in real time their views on a broad range of subjects, debate current matters and share experience with their goods and services in their everyday lives. Encouraged through microblogging websites, companies explore ways to obtain insights about Twitter on how people respond to its goods and services. A lot of investigation was underway into how feelings are conveyed in structured patterns of text, including product or film reviews and news stories, but in light of the informal language and message duration limitations of microblogging it was less studied how feelings are expressed.

A. Problem Statement

Sentiment analysis of the field of microblogging is a comparatively recent field of study and so further research still needs to be done. A good amount of previous similar work on user feedback, papers, online blogs / items and an overall sentence mood analysis has been completed. Their differences are largely due to the 140 characters cap per tweet that forces users to share their viewpoint in a relatively short document. The best results reached in the classification of feelings are based on guided learning methods like the Naive Bayes and Support Vector Machines, but manual markings are very costly. Unchecked and semi-controlled methods have been some work done and a great deal of progress has been made. Diverse researchers who assess new features and classification methods also only compare their findings to their basic performance. Correct and formal comparisons of these findings are required across various features and classification techniques so that the best characteristics and most effective classification techniques are selected for such applications.
B. Objective

The aims of the survey are to research sentiment analysis in microblogging with a view to analysing the input provided by a customer of the product of an organisation; and also to create a software for the evaluation of a product by the customer, which enables an organisation or person to feel and interpret large numbers of tweets in a useful format.

II. METHODOLOGY

The proposal was split into two stages. First, a review of literature and the history of the method is carried out. Literature studies include research on different approaches and methods of sentiment analysis commonly used. Prior to development, specifications and functionalities for application are specified in Phase 2. Often, software design and interface design are defined and how the programme can communicate. Several applications such as Python Shell 2.7.2 and Notepad are used to build the Twitter Sentiment Analysis programme.

III. LITERATURE SURVEY

Analyzes of feelings are a growing field in natural language processing, with studies ranging from the characterization of documents (Pang and Lee 2008) to the analysis of the polarity of words and sentences (e.g. (Hatzivassiloglou and McKeown 1997; Esuli and Sebastiani 2006).

Given the constraints on tweets, the classifying of Twitter's feelings is quite similar to sentence level sentiments analysis, for example (Yu and Hatzivassiloglou 2003 and Kim and Hovy 2004); but Twitter sentiment analysis is very different due to the informal and technical vocabulary used in tweeting, and the very existence of the microblogging domain. It is open how well the characteristics and strategies use them on more well-formed data will transfer to the microblogging domain.

There were some Twitter-sensitive articles in the past year alone (Jansen et al. 2009; Pak and Paroubek 2010; O'Connor et al. 2010; Tumasjan et al. 2010; Bifet and Frank 2010; Barbosa and Feng 2010; Davidov, Tsur, and Rappoport 2010). Other scientists have started to study the usage of speaker-part functions, but the findings are mixed. Microblogging features (e.g. emoticons) are popular, however the utility of current sentimental services based on non-microblogging data has not been investigated extensively.

Researchers have already started researching different means of gathering data automatically. Several researchers use emoticons to define their data (Pak and Paroubek 2010; Bifet and Frank 2010). (Barbosa and Feng 2010) use current Twitter sentiment collection pages. (Davidov, Tsur, and Rappoport 2010) also use hashtags for development of training, but they do not, as we do, use the classification of feeling/non-sense in their studies. For this second classification we use WEKA and apply the following Machine Learning algorithms to achieve the best outcome:

A. Opinion Mining

Opinion mining refers to the vast field of the analysis of natural languages, text mining, computational linguistics, including computer-based assessment of the feelings, thoughts, and emotions of the text[8]. Emotional views or attitudes are sometimes referred to as a feeling instead of reason[8]. Therefore, loans for a peer review or sentiment analysis counterpart. [9] said mining opinions have a variety of fields of application, including accounting, law, science, entertainment, education, telecommunications, policy and marketing. Many social media outlets earlier provided online users with a chance to communicate their ideas and viewpoints and post them[10].

B. Twitter

Twitter is a popular microblogging service that provides users with short-term, tweet-limited information [2,3], [11]. Users write tweets with opinions on different subjects related to their everyday lives. Twitter is an excellent forum for the public opinion on particular topics to be extracted[9,10]. The main corpus of emotion analysis, which relates to the use of opinion mining or natural language treatment is a series of tweets[1].

Twitter has soon become a powerful asset for companies, with 500 million users and millions of posts a day, to boost their brands' reputations through the collection and analysis of the public's twitters' sentiments about their products, the services industry and even rivals. [2] emphasised that, based on opinions created by the social media, the mammoth expansion of the global Web, the worldwide web provides the most promising, comprehensive, easy-to-read media for sentiment analysis, including super volumes of opinion papers, tweets, journals, blogs, and other comment groups and forums.
C. Microblogging with E-commerce

A microblogging website like Twitter, like a traditional blogging site, is only shortened by individual entries[13]. Twitter has reduced itself to a small number of terms intended to quickly transmit or share information[7]. But the promise of microblogging as an eCommerce tool has been initiated by small and large companies[3]. However, a period has been devised to encourage international trade by using an external microblogging network as Twitter marketing[3]. It does, in reality, have been established for a few years. The moment the social, community-oriented capabilities of the micro-blogging site are sharing open up e-commerce, a new light spot can be seen where businesses can create brand images, relevant products, boost retail sales, discuss positive engagement with the customer and other market practises [2,3] [2] [14]. [14]. [13] said, in reality, the firms producing those goods began polling the microblogs to get a feeling for a commodity. Often such businesses investigate user reactions and respond to microblog users[14].

D. Social Media

[14] described social media as a web-based community of applications that can develop and share user generated content based on Web2.0 ideological and technical foundations. [15] [16] In a debate about Internet Start, internet users have shown that the overall duration spent on mobile and social media has rising by 37 percent in 2012 to 121 billion minutes, up from 88 billion minutes in 2011, and is spending more time using the social networks. In contrast, companies can locate and connect with customers on social networking sites; enterprise can prove harm to competitiveness caused by social networking[17]. The fact that the public will quickly access social media will harm the spread of private information to the real world[11]. On the contrary[18] addressed the advantages of social media participation to develop the prestige of the organisation and offer job prospects and monetary profits. [15], [35] have said that social networks are used in advertising by promotional agencies, quest and recruitment practitioners, online and electronic commerce for social learning. In e-commerce it is referred to the buying and selling of online products or services that can be obtained via social media, which are convenient to Twitter because of its availability for 24 hours, its easy customer support and worldwide reach[19]. One of the reasons that companies want to use more social media is to learn about customer behaviour, market analytics and the possibility to use. Know from client reviews and insights.

IV. CLASSIFICATION TECHNIQUE

Many tasks are classified, but our groups are fortunately much easier to define than Borges’ tasks. This classification presents the native classification of Bayes algorithms showing an important issue of classification: the categorization of the texts, the task of classifying the whole text, allocating a categorization mark derived from those labels. One popular role of categorising texts, analysis of feelings, and analysis of ex sentiment Sentimental traction, positive or negative orientation to any object that a writer communicates. When a film, book or product is shown on the internet, it reflects the feeling of the author towards the product and the editorial or political document expresses a feeling towards a candidate or political action. For the selling of any commodity, automatically calculating customer feelings is key while calculating public feelings is important in political terms and even for estimation of the market. A binary classification is the easiest version of feeling analyses a

\[ P_{NE}(c|d) = \arg \max_c P(c) \prod_{m=1}^{m} P(f|m)^m / P(d) \]

The equation above refers to f as a function, the feature count (f) as ni(d) as well as a tweet in d. m marks no. of characteristics here. The P(c) and P(f|c) parameters are calculated by maximum probability calculations, and smoothing is used in unseen functions. Use Naïve Bayes machine learning techniques to train and distinguish. we can use the Python NLTK library.
V. SENTIMENT ANALYSIS APPROACH

The feeling is expressed on the comments or on the tweets for several useful metrics [20]. In fact, [12] and [36] indicated that a feeling can be classified as negative and positive terms in two categories. The emotion analysis is a normal computing methodology for language processing in a collection of tweets to measure an explicit thought or feeling [7]. The study of sentiments applies to the general approach for excluding polarity and subjectivity from semantic orientation, which refers to words and text or phrases in terms of polarity [19]. The lexicon method and machine-learning approach have two key ways for automatically extracting emotion [19-23].

A. Input Keyword

Raw data is obtained through the 'tweepy' library of Python, providing a single twitter streaming API bundle. The SampleStream and FilterStream API are two ways of viewing tweets. SampleStream essentially provides a brief random sampling of all streaming tweets in real-time. FilterStream offers tweets that meet a number of requirements. The tweets given can be filtered in three ways:

1) Track/search keyword in the tweets
2) Specific user of Twitter by their name
3) Tweets from a certain place(s) (only for geo-tagged tweets).

B. Tweets Retrieval

As human labelling is an expensive process, we sort out the tweets to be labelled to make the greatest variety of tweets without losing their generality. The below are the applicable filtering criteria:

1) Retire Retweets (any tweet containing the "RT" string)
2) Remove tweets really short (tweet with length less than 20 characters)
3) Remove tweets not English • (by comparing the words of the tweets with a list of 2,000 common English words, tweets with less than 15 percent of content matching threshold are discarded)
4) Exclude equivalent tweets (by contrasting each tweet with any tweet you delete tweets with a content of more than 90% that matches any other tweet).

About 30% of tweets exist on average for human labelling per survey after this filtering and 10,173 tweets have been labelled.

C. Data Preprocessing

Data Preprocessing can be done in three steps, i.e., Tokenization, Normalization, and Parts of Speech Tagging.

1) Tokenization: It is the act of breaking down a stream of text into words, symbols, and other significant elements known as "tokens." Whitespace and/or punctuation characters may be used to distinguish tokens. This is done so that we can examine tokens as separate components of a tweet. Emoticons and abbreviations (for example, OMG, WTF, BRB) are defined and handled as individual tokens during the tokenization process.

2) Normalization: The use of abbreviations in a tweet is noticed and abbreviations are replaced by their original meaning during the normalization step (e.g. BRB -> be right back). Often, We recognise casual intensifiers like all-caps (e.g., I LOVE the show!!!) and repetitions of character (e.g., I have a mortgage!! happyyyyyyy’). Both caps are in a lower case and repetitive letter instances are replaced by one character. Finally, it is noticed that special twitter tokens are present (such as #hashtags, user tags, and URLs) and the token frame holders are replaced. Our hope is to increase efficiency through this normalisation.

3) Part of Speech Tagging: The POS-Tagging method is the assignment of a tag for a phrase in which grammatical aspect of the language the term includes, i.e. noun, verb, adjective, adverb, conjunction coordination, etc. We have the counting functions of each tweet for the number of verbs, adverbs, adjectives, nouns, etc.

D. Classification Algorithm

Let's develop a Twitter data sentiment analysis to demonstrate how you can incorporate such an algorithm into your apps. First of all, we will choose a subject, then collect tweets using that keyword and analyse feelings on the tweets. We would eventually get an overview of whether or not people see the issue favourably.

1) Gather Tweets: First, choose a topic you wish to analyze. Inside sentiment-analysis.js, you can define input to be whatever phrase you like. In this example, we'll use a word we expect to return positive results.

2) Perform Sentiment Analysis on Tweets: After gathering and cleaning our data set, we are ready to execute the sentiment analysis algorithm on each tweet. Then, we will calculate an average score for all the tweets combined.
We have iterated in no retweets via each tweet to give this feedback to the Sentiment Analyzing algorithm in the above algorithm. We then applied the performance result to a total score vector using the result of the API call. We follow up how many tweets we have with the vector score count and then we measure the final point by averaging the total score as it exceeds the same amount as the number of tweets we decided to evaluate. This final outcome is a very pessimistic, negative, favourable, optimistic and very positive sentiment in the range [0–4].

E. Classified Tweets
We marked the tweets in three groups based on the tweets’ feelings: positive, negative, and neutral. To support our labellers in the labelling process we gave the following guidelines:

1) Healthy: If the whole tweet is positive / happy / cheerful or if something with good connotations is listed.
2) Positive: And if the tweet expresses more than one emotion, the good feeling is more prevailing.
3) Example: 'I'm in the USA for another four years after I'm in Shithole Australia.'
4) Negative: Nonetheless, whether the whole tweet is negative/bad/bad, or whether anything with negative connotations is stated. Even, if there is more than one feeling in the tweet, but the bad feeling is more prominent. Example: 'Now this iPhone is dull, I want to get an Android: S.'

VI. RESULTS AND DISCUSSION
We will discuss our findings for objective/subjective/negative classifications for the first time. This are the first steps of our approach to classification. On all these outcomes we use only the shortlisted functions. This means that we have 5 characteristics in the objective/subjective classification and we have 3 in the positive/negative classification. We use the classification algorithm Naïve Bayes for both these results, since that is the algorithm used in our current classification method at the first stage. In addition, both estimates are a 10-times cross-validation result.

The performance, however, is classified into 2 encoded and unencoded types. Part of the output is displayed in an ID shape, such as string ID, according to a security problem. Analysis sentiment. The tweets are categorised according to the lexicon dictionary in positive and negative languages. The values of each word are allocated. In .txt,.csv and Html the results are seen. The outcome is seen in a pie diagram that represents each percentage positive, weakly positive, strongly positive, neutral, hate, weakly hate, strongly hate feelings. However, a top ten positive and hate Hash Tags can be listed here.

As seen in Fig.1 & 2, each percentage positive, weakly positive, strongly positive, neutral, hate, weakly hate, strongly hate feeling hazard 1000 tweets tags in various colours is shown in the bar map.
VII. CONCLUSION AND RECOMMENDATION

Sense analysis of Twitter is developed in order to analyse the experiences of consumers who are important for business growth. The programme, in conjunction with natural language processing methods, employs a machine-based learning approach which is more accurate to analysis. This results in an optimistic and negative feeling, as is seen in a pie map and HTML page. However, because of limitations of Django that only works on the server Linux or LAMP the software is intended to be created as a web application. It cannot then be realised. Therefore it is recommended in the future study to further improve this aspect.

REFERENCES

[1] M. Rambocas, and J. Gama, “Marketing Research: The Role of Sentiment Analysis”. The 5th SNA-KDD Workshop’11. University of Porto, 2018.
[2] A. K. Jose, N. Bhatia, and S. Krishna, “Twitter Sentiment Analysis”. National Institute of Technology Calicut, 2019.
[3] P. Lai, “Extracting Strong Sentiment Trend from Twitter”. Stanford University, 2012.
[4] Y. Zhou, and Y. Fan, “A Sociolinguistic Study of American Slang,” Theory and Practice in Language Studies, 3(12), 2209-2213, 2013. doi:10.4304/tpls.3.12.2209-2213
[5] M. Comesaña, A. P. Soares, M. Perea, I. Fraga, and A. Pinheiro, “Author’s personal copy Computers in Human Behavior ERP correlates of masked affective priming with emoticons,” Computers in Human Behavior, 29, 588-595, 2013.
[6] A. H. Huang, D. C. Yen, & X. Zhang, “Exploring the effects of emoticons,” Information & Management, 45(7), 466-473, 2008.
[7] D. Boyd, S. Golder, & G. Lotan, “Tweet, tweet, retweet: Conversational aspects of retweeting on twitter,” System Sciences (HICSS), 2010. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5428313
[8] T. Carpenter, and T. Way, “Tracking Sentiment Analysis through Twitter.” ACM computer survey. Villanova: Villanova University, 2010.
[9] D. Osimo, and F. Mureddu, “Research Challenge on Opinion Mining and Sentiment Analysis,” Proceeding of the 12th conference of Fruct association, 2010, United Kingdom.
[10] A. Pak, and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining,” Special Issue of International Journal of Computer Application, France: Universite de Paris-Sud, 2010. [11] S. Lohmann, M. Burch, H. Schmauder and D. Weiskopf, “Visual Analysis of Microblog Content Using Time-Varying Co-occurrence Highlighting in Tag Clouds,” Annual conference of VISVISUS. Germany: University of Stuttgart, 2012.
[11] H. Saif, Y. He, and H. Alani, “Semantic Sentiment Analysis of Twitter,” Proceeding of the Workshop on Information Extraction and Entity Analytics on Social Media. United Kingdom: Knowledge Media Institute, 2011.
[12] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, “Sentiment Analysis of Twitter Data,” Annual International Conferences. New York: Columbia University, 2012.
[13] J. Zhang, Y. Qu, J. Cody and Y. Wu, “A case study of Microblogging in the Enterprise: Use, Value, and Related Issues,” Proceeding of the workshop on Web 2.0., 2010

[14] G. Kalia, “A Research Paper on Social Media: An Innovative Educational Too”, Vol.1, pp. 43-50, Chitkara University, 2013.

[15] Internet World Start, “Usage and Population Statistic”, Retrieved 10 15, 2013 from: http://www.internetworldstats.com/stats.htm

[16] A.M. Kaplan, and M, Haenlein, “Users of the world, unite! The challenges and opportunities of Social Media,” France: Paris, 2010.

[17] Q. Tang, B. Gu, and A.B. Whinston, “Content Contribution in Social Media: The case of YouTube”, 2nd conference of social media. Hawaii: Maui, 2012.

[18] M.Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, “LexiconBased Methods for Sentiment Analysis,” Association for Computational Linguistics, 2011.

[19] M. Annett, and G. Kondrak, “A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blogs,” Conference on web search and web data mining (WSDM). University of Alberta: Department of Computing Science, 2009.

[20] P. Goncalves, F. Benevenuto, M. Araujo and M. Cha, “Comparing and Combining Sentiment Analysis Methods”, 2013.

[21] E. Kouloumpis, T. Wilson, and J. Moore, “Twitter Sentiment Analysis: The Good the Bad and theOMG!”, (Vol.5). International AAAI, 2011.

[22] S. Sharma, “Application of Support Vector Machines for Damage detection in Structure,” Journal of Machine Learning Research, 2008.

[23] A.Sharma, and S. Dey, “Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis,” Association for the advancement of Artificial Intelligence, 2012. [25] J. Spencer and G. Uchyigit, “Sentiment or: Sentiment Analysis of Twitter Data,” Second Joint Conference on Lexicon and Computational Semantics. Brighton:University of Brighton, 2008.

[24] A. Blom and S. Thorsen, “Automatic Twitter replies with Python,” International conference “Dialog 2012”.

[25] B. Pang, and L. Lee, “Opinion mining and sentiment analysis,” 2nd workshop on making sense of Microposts. Ithaca: Cornell University. Vol.2(1), 2008.

[26] M. Hu, and B. Liu, “Mining and summarizing customer reviews,” 2004.

[27] P. Nakov, Z. Kozareva, A. Ritter, S. Rosenthal, V. Stoyanov, T. Wilson, Sem Eval-2013 Task2:Sentiment AnalysisinTwitter (Vol.2, pp. 312-320) , 2013. [29] J. Wu, J., Wang, & L. Liu, “Kernel-Based Method for Automated Walking Patterns Recognition Using Kine matics Data”, 5th Workshop on Natural Language Processing. China: Xi’an Jiaotong University. 2006

[28] T. D. Smedt, and W. Daelemans, “Pattern for Python,” Proceeding of COLING. Belgium: University of Antwerp, 2012.

[29] A. Sweigart, “Invent your own computer games with Python. 2nd edition, 2012. Retrieved from http://inventwithpython.com/

[30] C. Seberino, “Python. Faster and easier software development,” Annual Conference. California: San Diego, 2012.

[31] A.Lukaszewski, “MySQL for Python. Integrate the flexibility of Python and the power of MySQL to boost the productivity of your applications,” UK: Birmingham. Packt Publishing Ltd, 2010.

[32] V. Nareyko, “Why python is perfect for startups,” Retrieved 01 10, 2014 from: http://opensource.com/business/13/12/why-python-perfect-startups

[33] A. Hawkins, “There is more to becoming a thought leader than giving yourself the title”. Retrieved 10 18, 2013. from: http://www.thesocialmediashow.co.uk/author/admin/

[34] R. Prabowo, and M. Thelwall, “Sentiment Analysis:A Combined Approach,” International World Wide Web Conference Committee (IW3C2), 2009. UnitedKingdom:Universityof Wolverhampton. [37] H. Saif, Y. He and H. Alani, “Alleviating Data Scarcity for Twitter Sentiment Analysis”. Association for Computational Linguistics, 2012.
INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call: 08813907089 (24*7 Support on Whatsapp)