Analysis of Behavior Extraction on Social Life Issues Using Sentiment Analysis: A Review

Adnan Beg

B.Tech. Student, Computer Science and Engineering Department, Integral University, Lucknow, India

Abstract: *At present time most of the world can be found on internet. The popularity of the social networking sites has been growing fast, parallel to emerging technologies, with increase in the number of users to the social networking sites that actively express their opinions on these sites. To forecast the sentiment analysis we have used the data stored on social sites stockpiling. Our goal is to retrieve data from social sites and analyze the emotion of a particular person on certain topic. For our analysis we have used the data available on Twitter, as Twitter has a large amount of data that people post, which gives the output beyond the polarity but those polarities can be used in product profiling, trend analysis and forecasting. This paper presents an overview of past and current research on twitter sentiment analysis and presents better ideas for future work.*

1. Introduction

People make judgments about the world around them where they are living. They create positive and negative attitudes about people, products, places and events according to their understandings. These types of attitudes can be considered as sentiments. Sentimental analysis also known as opinion mining is the process to gain an understanding of the attitudes, opinions and emotions expressed by people on social networking sites. Sentimental analysis is found extremely useful in social media monitoring as it allows us to gain an overview of the opinions of wider public behind certain topics. Twitter has 310 million monthly active users, who exchange information through tweets of length approx. 140 characters. Tweets on Twitter are delivered in real time from any geographical area and by anyone irrespective of the user backgrounds such as religion, political inclination or the level of education or civilization. In this study, the author has used these tweets tweeted on Twitter on a certain topic, to analyse whether the sentiments of the people are positive, negative or neutral. As web data like tweets on twitter is large amount of data generated by people so it's become important to work for an efficient intelligent systems that can do refinement of data and analysis of tasks intelligently and efficiently. A significant research work can be seen since 1980 on different aspects of twitter sentiment analysis. Now we will see major work of these researches.

2. Literature Survey

2.1 Sentence based analysis [1]

In this paper the author had studied text to speech analysis of data. They study about processing of data and ability to render natural expressive speech. The work is mainly about the techniques that are used in natural expressive speech and how data of natural expressive speech is being used. Next we will see the process of natural expressive speech using the diagram shown by author. This work basically focuses on production of synthetic speech and adaption of sentiment analysis [11] procedure that can then be used as input feature for expressive speech synthesis. So, we can say that the work of this paper revolves around data to speech synthesis.

This paper is mainly concerned with the identification that how text and web data can effect in different aspects and tries to categories data on the basis of its effects. This work was the first attempt to adapt conventional sentiment analysis work. In this work author had used small training data for the study so the classification can be improved in terms of speed and accuracy when applied to large data sets. The conclusion to this research paper is using a fast scripting language the functionalities the author wants to introduce can be achieved up to a level.

2.2 Automated Topic Modeling and Sentiment Analysis of Tweets on Spark R [2]

In this paper the author had proposed an automated topic modeling technique LDA which helps in identifying the attracted matter of arguments from huge amount of tweets co-related to any two leading and famous political leaders of India. In this paper, the author takes a spark R framework on which the topic modeling technique is performed to upgrade to the speed and performance for vast social data processing and its analysis in real time. At that point sentiment analysis, those tweets are finished by utilizing dictionary based ways to deal with arrange the general population's sentiment against the two leaders. The result shows that the proposed method has improved the speed compared to the normal R tool.

2.3 RA summarization of short comments [3]

Along a timeline (TASC-t) for dynamic tweets and experimented on 6 topics from published corpuses demonstrating that usefulness of the TASC algorithm. Their future work emphasized on accurate labeling of data and a good

In this paper author proposes a framework to summarize the opinions for some product and item based on user votes. Basically, it is the study of generating a RA summary of short comments, which is a decomposed view of the overall ratings for the major aspects so that a new user can make a best choice according to the experience and perspective of the old user.

The output is that overall rating is decomposed into several small aspects each having support information showing the

confidence on the rating of the aspects. The future work strongly emphasizes on the comparative study of product reviews and seller services for better marketing. This rated mechanism can be applied with product reviews summarization to produce more accuracy in generated output text.

2.4 Field adaptive classification of tweets [4]

In this paper author had studied classifications of tweets in a manner that demonstrate field adaptation manner of the tweets. As topics in twitter are very diverse, it is impossible to train a universal classifier for all topics [10] and twitter lacks data labeling so it has becomes more difficult to make them classify into different fields. The TASC algorithm that the author had used updates the topic-adaptive features based on the collaborative selection of the data that is unlabeled, which helps in selection of more reliable tweets to boost the performance.

The algorithm that the author uses for classification of tweets into different feature Fields is shown in below table 1. The below table shows results of sample data showing total Tweets and their behavior in different categories. They also design the adapting model

Table 1: Corpus Statistics [2] classifier to model dynamic tweets [11] into feature field

Topics	Positive	Negative	Neutral	Total
Apple	1911	377	581	1,149
Microsoft	93	138	671	902
Google	218	61	604	883
President Debate	1,465	729	1,109	3,213
Twitter	68	78	647	793
Taco Bell	902	596	2,099	3,597

3. Other Related Works

Other research work includes cross domain sentiment classification [5] .Using this feature the augmentation and selection according to the information gain criteria for the cross domain sentiment classification work performed better compared to other approaches. Sentiment work is also done on some real time themes like sentiment performance in characterizing debate performance [6]. They have shown that visuals and metrics which can be used to inform the design of visual analytic systems for social media events. Such sections are able to identify key sections of a debate performance.

Sentiment classification is also done on microblogs [7]. Microblogs as a new textual domain offers a unique proposition for sentiment analysis. Their short document length suggests that any sentiment they represented is compact and explicit. In some way classifying sentiments in microblogs [10] is easier then blogs and make a number of observations pertaining to the challenge of SL for sentiment analysis on microblogs. Apart from this, work is also done in emoticons areas of twitter [8]. In these emoticons smoothed language models are building for the data cleaning and classification. This model is known as ESLAM. Extraction of emoticons [9] is a category of sentiment analysis, where a lot of work is done to extract different behavior represented by a tweet and text. For this a supervised multi-engine classifier approach has used by the author to identify emotion topic(s) from the English blog sentences.

4. Conclusion

As social media website like twitter is so vast that getting all information is almost infeasible but possible steps can be taken to get most of the useful information from it. This can be achieved through sentiment analysis as there are so many areas in sentiment analysis field like data refinement, topic modeling of sentiments and summarization of sentiment tweets which are still unrecognized and untouched. These areas build strong chances for innovation of new outcomes, techniques and methods if we look forward to explorer these areas.

References

[1] Alexandre Trilla and Francesc Alías;"Sentence-Based Sentiment Analysis for Expressive Text-to-Speech" IEEE transaction on audio, speech and language processing, February 2013, Page 223-233.

[2] Monish, P., Kumari, S., & Babu, C. N. (2018, July). Automated Topic Modeling and Sentiment Analysis of Tweets on Spark R. In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.

[3] Yue Lu, ChengXiang Zhai, Neel Sundaresan; "Rated aspect summarization of short comments" In www 2009, pp 131-140.

[4] Shenghua Liu, Xueqi Cheng, Fuxin Li, and Fangtao Li;"Topic-Adaptive Sentiment Classification on Dynamic Tweets" IEEE Trans. Knowl. Data Eng. 27(6): 1696-1709 (2015).

[5] Y. He, C. Lin, and H. Alani, "Automatically extracting polarity bearing topics for cross- domain sentiment classification," in Proc. 49th Annu. Meeting Assoc. Comput. Linguistics: Human Language Technol.-Volume 1, 2011, pp. 123–131.

[6] N. A. Diakopoulos and D. A. Shamma, "Characterizing debate performance via aggregated twitter sentiment," in Proc. SIGCHI Conf. Human Factors Comput. Syst., 2010, pp. 1195–1198.

[7] Adam Bermingham & Alan Smeaton; "Classifying Sentiment in Microblogs: Is Brevity an Advantage?" ACM, New York, NY, USA, 1833-1836.2010.

[8] K.-L. Liu, W.-J. Li, and M. Guo, "Emoticon smoothed language models for twitter sentiment analysis." in Proc. 26th AAAI Conf. Artif. Intell., 2012, pp. 1678–1684.

[9] D. Das and S. Bandyopadhyay, "Extracting emotion topics from blog sentences: Use of voting from multi-engine supervised classifiers," in Proc. 2nd Int. Workshop Search Mining User-Generated Contents 19th ACM Int. Conf. Inform. Knowl. Manage., 2010, pp. 119–126.

[10] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment in twitter events," J. Am. Soc. Inform. Sci. Technol., vol. 62, no. 2, pp. 406– 418,2011.

[11] Aggarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in Proc. Workshop Lang. Soc. Media, 2011, pp. 30–38.