

Comparative Analysis of Classifiers in Sentiment Mining on the Basis of Performance Matrix

Stuti Mehla¹

Research Scholar, Computer Science & Engineering
Maharishi Markandeshwar University, Mullana
Ambala, Haryana
e-mail: stuti21mehla@gmail.com

Dr. Sanjeev Rana

Professor, Department of Computer Science & Engineering
Maharishi Markandeshwar University, Mullana
Ambala, Haryana
e-mail: dr.sanjeevrana@mmumullana.org

Abstract— Social media has put its feet in late nineteenth century but with the advancement in technology social networking platform has logarithmic increment. Social platform has reached in that stage where every user post or blog their aspects and experiences. These views posted by user can be made useful by private and public enterprises, political parties, medical organizations to conduct surveys and society expectations and experiences. To discover the useful data from a huge amount of posted views can be made useful by every organization which depend upon customers and the process of extraction of beneficial data from posted data is termed as sentiment mining or text mining. Different classifiers are used for sentiment analysis and in this research paper a comparison among classifiers is conducted and the best classifier is predicted through review of various research paper on the basis of performance matrix.

Keywords: Sentiment analysis, Big Data, text mining, IOT, Naive Bayes, SVM, LR, EL

I. INTRODUCTION

From nineteenth to present time technology has come to a great extent, every field related to it has a huge rise. Computer science is also come in that category which is revolutionized due to boom in technology.

Due to enhancement in technology social media has come into existence where every user share its experiences either by posting or blogging. These shared experiences and views can be made useful for prediction, analysis and visualization. Views are termed as sentiments and are coming at an exploded rate through facebook, twitter or from different blogging sites and it can be made useful to predict whether a given mass is satisfied with a certain product or opinion about some political party or some organization which serve a mass. So, to find the useful sentiment from the pool of sentiments or views can be termed as sentiment mining or sentiment discovery. In this process required sentiments are discovered from large sentiments. In Sentiment Analysis, sentiments can be defined as opinions and it can be termed as opinion mining [1]. Generally opinions categorized into different categories i.e positive, negative and neutral and by categorization these refined knowledge can be made useful for feedback analysis of a product and public rating for some specific domain i.e mass review about

some movie, automobile, political parties.[1,2]. Opinion mining is categorized into two categories: topic dependent and domain dependent[2].

Traditional statistical techniques can't be used for sentiment mining because huge data is coming at high speed and usually not in any predefined manner. So to overcome this drawback classifier techniques, machine learning algorithms have come into existence which work on the basic principle of data collection, data cleaning, feature extraction and then feeding to classifiers and machine learning algorithms. Sentiment Analysis or opinion mining is also different on twitter data and with machine learning algorithms [3]. There are two drawbacks in using machine learning algorithms first is ambiguity: results are not good when a user uses machine learning algorithms. second is twitter uses unofficial words for communication which becomes hinderance for preprocessing experiment.

II Related Work

A lot of research work has been carried out on sentiment analysis and different types of literatures has been reviewed on the same. Some of research work is explained following.

In [4] paper a review of sentiment analysis techniques is carried out. In this review work it has been observed that there are two approaches for sentiment analysis: lexicon based approach and learning based methods. One

of the major drawback of these approaches is they are domain specific and depend on source. So, a new approach has been proposed which can overcome these issues. In [5] paper authors have developed a system that is used for opinion analysis. In this system tweets are processed and then preprocessing them. In this research work opinion analysis is categorized into two categories one is through machine learning algorithms and second is using hybrid approach. Machine learning algorithms are used are Naïve Bayes, SVM, Maximum entropy and in hybrid approach u-gram technology is mixed with SVM.

In [6] paper authors have developed a system which analyse the sentiments by extracting the feature weights from twitter data set and then analysing them through classifiers ie Naïve Bayes, Maximum entropy and through SVM.

In [3] this paper researcher have come up with drawbacks of opinion mining ie preprocessing of slangs and second is ambiguity and give stress on analysis of tweets is different in machine learning algorithms is different from opinion mining. In the proposed approach stream of tweets is preprocessed and categorized into positive, negative and neutral tweets and results are carried out on the basis of precision and recall.

In [8] authors has researched and find out the reason of disease with there symptoms by mining by social media data. In [7] authors did a lot of research in field of sentiment analysis by experimenting on twitter data using classifiers. In [9] different classifiers are used for sentiment analysis and classify them into positive, negative and neutral tweets and then using classifiers for training of those data.

Summary of previous research work:

1. Through these research works, we have concluded that sentiment analysis can be done through machine learning approach and through classifiers.

IN DIAGRAM IT HAS BEEN CLEARLY DEFINED THAT SENTIMENT ANALYSIS IS ENTIRELY DIFFERENT FROM MACHINE LEARNING APPROACH BECAUSE SENTIMENT ANALYSIS IS DONE ON SOCIAL MEDIA DATA WHICH IS FULL OF SLANGS AND EXPRESSIONS.

2. Sentiment analysis is done through machine learning algorithms and they can be hybridized which can enhance there value of

accuracy, precision, recall[5] and techniques. It has also observed

That there are different techniques ie lexicon based approach and learning based algorithms for analysis [4]

III CONCLUSION:

Through research work it has be concluded that sentiment analysis is one of the emerging field because it directly communicate to public which share and post its opinion. This can be made useful for any type of opinion, analysis and review. In this research review an analysis has been carried out b/w different techniques and brief comparison is done between techniques.

References:

- [1]. Malhar Anjaria, Ram Mahana Reddy Guddeti "Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning" 978-1-4799-3635-9/14/\$31.00 ©2014 IEEE
- [2]. Preslav Nakov¹ • Sara Rosenthal² • Svetlana Kiritchenko³ "Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts" Springer Science+Business Media Dordrecht 2016
- [3]. Balakrishnan Gokulakrishnan * 1, Pavalanathan Priyanthan *2, Thiruchittampalam Ragavan * "Opinion Mining and Sentiment Analysis on a Twitter Data Stream" 978-1-4673-5530-8/12/\$31.00 ©20 12 IEEE
- [4]. Ming Hao, Christian Rohrdantz*, Halldór Janetzko*, Umeshwar Dayal "Visual Sentiment Analysis on Twitter Data Streams"
- [5]. T. K. Das¹, D. P. Acharjya² and M. R. Patra³ "Opinion Mining about a Product by Analyzing Public Tweets in Twitter" and 978-1-4799-2352-6/14/\$31.00 ©2014 IEEE
- [6]. Geetika Gautam , Divakar yadav "Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis" 978-1-4799-5173-4/14/\$31.00 ©2014 IEEE
- [7]. Praveen Kumar, 2Tanupriya Choudhury, 3Seema Rawat, 4Shobhna Jayaraman "Analysis of Various Machine Learning Algorithms for Enhanced Opinion Mining using Twitter Data Streams" 978-1-5090-3411-6/16 \$31.00 © 2016 IEEE DOI 10.1109/ICMETE.2016.19
- [8]. Sunghoon Lim, Conrad S. Tucker, Soundar Kumara "An unsupervised machine learning model for discovering latent infectious diseases using social media data"
- [9]. Harshali P. Patil , Dr. Mohammad Atique, "Sentiment Analysis for Social Media: A Survey" 978-1-4673-8611-1/15/\$31.00 ©2015 IEEE.