NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 3, ISSUE 1, JAN.-2016

COMPARISON OF CLASSIFIERS FOR SENTIMENT ANALYSIS

Varsha D. Jadhav
P.E.S.College of Engineering, Aurangabad

S.N. Deshmukh
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad

ABSTRACT

The need of social media has vividly changed people's life with more and more sharing their thoughts, expressing opinions, and in the hunt for support on social media websites such as Twitter, Facebook, blogs etc. Twitter, an online social networking and micro blogging service, which enables users to send and read text-based posts, known as *tweets*, with 140-character limit. Newspapers and blogs express opinion of news entities (people, places, things) while exposure to recent events. We present a system which extracts the sentiments from the online posts of twitter about news event. Our system shows sentiment identification, which expresses opinion associated with each entity. Also it consists of scoring phase, which assigns scores to each entity, on which the tweets are classified. Finally, we compare Maximum Entropy, Decision tree, Support vector machine and Naives Bayes classifiers.

INTRODUCTION

News can be good or bad, but it is infrequently neutral. Even though natural language text understanding remains beyond the power of machine, the statistical analysis of quite simple sentiment can provide amazingly meaningful sense of how the latest news impacts significant entities. In this paper we present comparison of machine learning classifiers for sentiment analysis of user opinion towards news through comments and tweets using Maximum Entropy, Decision tree, Support vector machine and Naives Bayes. Senti Word classifier will calculate the positivity and negativity points of the opinionated words contained in the tweet based on the Senti Word scores. For each unigram extracted from the tweet to be classified, its corresponding Senti Word score would be fetched. Separate scores would also be maintained with respect to positive and negative unigrams contained in the tweet. The score is maintained between -4 to 4. These scores will then be used to classify the tweets as positive, negative and neutral using the machine learning classifiers should present original research contributions.

RELATED WORK

Ghazaleh Khodabandelou et al [1] highlighted difference between process mining and intention mining. In process mining some technique is used to process models by analyzing event logs where no apriori information is available and some α -algorithm may be used to model the behavior of the actor. In intention mining actor's intention is identified from event logs and produce intentional process models. Novel approaches on modeling and inferring users actions in a computer is proposed [2] using two linguistic features-keyword and concept features. Bo Pang et.al [3] considered the problem of classifying documents not by topic, but by overall sentiment. They employed Naïve Bayes, maximum entropy classification, and support vector machines, which do not perform well on sentiment classification. Jeonghee Yi[4] present Sentiment Analyzer (SA), which detects all references for the given subject, and determines sentiment in each of the references using natural language processing (NLP) techniques. Minqing Hu and Bing Liu[5] focus on mining opinion features in customer reviews. Lada Adamic et.al[6] studied the linking patterns and discussion topics of political blogger. Lloyd.L.et.al [7] analyzed the Lydia project. The project seeks to build a relational model of people, places, and things through natural language processing of news sources and the statistical analysis of entity frequencies and co-locations. Jeonghee Yi, et.al [8] describes the fully functional system environment and the algorithms, and reports the performance of the sentiment miner. The performance of the algorithms was verified on online product review articles. Andrew Mehler, et.al [9] developed a model of estimating the frequency of reference of an entity in any given city from the reference frequency centered in surrounding cities, and techniques for evaluating the spatial significance of this distribution. Levon Lloyd et.al [10] compared the prevalence of popular topic time series of 197 entities in the two corpora. Anindya Ghose et.al [11] Proposed two ranking mechanism for ranking product reviews: a consumer-oriented ranking mechanism ranks the reviews according to their expected helpfulness, and a manufacturer-oriented ranking mechanism ranks the reviews according to their expected effect on sales. Rohitha Goontilake [12] focuses on the effect of news that surfaces throughout the day in the stock market. Nikolay Archak et.al [13] developed a novel hybrid technique combining text mining and econometrics that models consumer product reviews as elements in a tensor product of feature and evaluation spaces. Mikhail Bautin et.al [14] analyzed entity sentiment in of newspapers in nine languages, and in five languages of a parallel corpus. Kanayama Hiroshi et.al [15] developed a high-precision sentiment analysis system at a low development cost, by making use of an existing transfer-based machine translation engine. Rada Mihalcea et.al[16] investigate methods to automatically generate resources for subjectivity analysis for a new target language by leveraging on the resources and tools available for English, which in many cases took years of work to complete. Soo-Min Kim et.al [17] presents a system that, given a topic, automatically finds the people who hold opinions about that topic and the sentiment of each opinion. The system contains a module for determining word sentiment and another for combining sentiments within a sentence. Andrea Esuli et.al [18] confronts the task of deciding whether a given term has a positive connotation, or a negative connotation, or has no subjective connotation at all.

EXPERIMENTATION

CORPUS COLLECTION

Twitter API is used to collect a corpus of text posts and a dataset is formed of three classes: positive, negative and neutral sentiments. On the birth anniversary of Dr. A.P. J Abdul Kalam (APJ) on 15th October 2015, the data on twitter was collected. Data related to fourth phase of Bihar elections (4PBE) is collected on 26th October 2015. Dataset for fifth phase of Bihar elections (5PBH) is collected on 5th November 2015.

FEATURE EXTRACTION

The collected dataset is used to extract features that will be used to train the sentiment classifier. Experimentation is carried out using n-gram binary features. The process of obtaining n-grams from the Twitter post is as follows.

i)Filtering: Remove URL links e.g. http://t.co/46iM8j8pkt,

ii) Tokenization: we segment text by splitting it by spaces and punctuation marks, and form a bag of words. However, we make sure that short forms such as “don’t”, “I’ll”, “she’d”will remain as one word.
iii) Removing stop words: we remove articles (“a”, “an”, “the”) from the bag of words.
iv)Removing punctuation, numbers and unnecessary spaces: e.g. Photoset httptco46iM8j8pkt after pre processing is obtained as Photoset httptcoiMjpkt Missing values: NA is assigned to the missing values.
v) Converting to lower case: All the letters in the sentences are converted into lower case.
vi) Constructing n-grams: we make a set of n-grams out of consecutive words. A negation (such as “no” and “not”) is attached to a word which precedes it or follows it. For example, a sentence “I do not like fish” will form two bigrams: “I do+not”, “do+not like”,“not+like fish”

SCORING THE TWEETS
To score each tweet, score.sentiment( ) function is used to iterate through the input text. It strips punctuation and control characters from each line using R’s regular expression-powered substitution function, and matches against each word list to find matches.
The score.sentiment ( ) function assigns score to the tweets using the formula as
Score = sum (pos.matches) – sum(neg.matches)
The score is maintained between -4 to 4.
4 and 3 represent very positive
-4 and -3 represent very negative
2 and 1 represent positive
-2 and -1 represent negative
If the score turns out to be zero, it is classified as neutral.

MACHINE LEARNING METHODS

We experimented the four standard algorithms: Maximum Entropy, Decision tree, Support vector machine and NaivesBayes.

To implement these machine learning algorithms, we need to find a source which categorizes words by sentiment. Hu and Liu’s “opinion lexicon” categorizes nearly 6,800 words as positive or negative and can be downloaded from Bing Liu’s web site:http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar. The lexicon consists of two text files, one containing a list of positive words and the other containing negative words. Each file begins with some documentation, which we need to skip and is denoted by initial semi-colon (“;”) characters.

RESULTS AND DISCUSSION

The classifiers are compared based on the accuracy measures such as Mean error (ME), Root mean square error (RMSE), Mean absolute error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE).

The simplest measure of forecast accuracy is called *Mean Absolute Error (MAE)*. MAE is simply the mean of the absolute errors. The absolute error is the absolute value of the difference between the forecasted value and the actual value. MAE tells us how big of an error we can expect from the forecast on average. Cort J. Willmott et.al [19] indicates that MAE is the most natural measure of average error magnitude than RMSE. Evaluations and inter-comparisons of average model performance error should be based on MAE. Table 1 shows the MAE for the datasets APJ,4PBE and 5PBE.

Table 1. MAE for the datasets APJ, 4PBE, 5PBE.

Classifier	**MAE**		
Dataset	**APJ**	**4PBE**	**5PBE**
Maximum Entropy	599.7809	1499.639	**332.7687**
Decision Tree	599.6181	1499.655	332.8491
SVM	599.8012	**1499.51**	332.8662
Naive Bayes	**599.4314**	1499.607	332.797

The plot for the sentiment for the dataset APJ is shown in fig.1

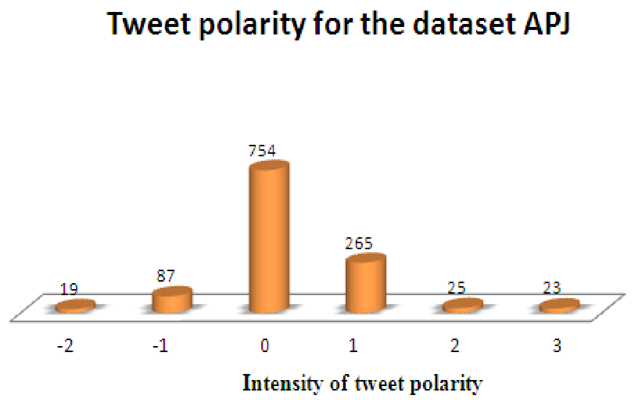

Fig.1 Tweet polarity for the dataset APJ

The plot for the sentiments of dataset 4PBE is shown in fig.2

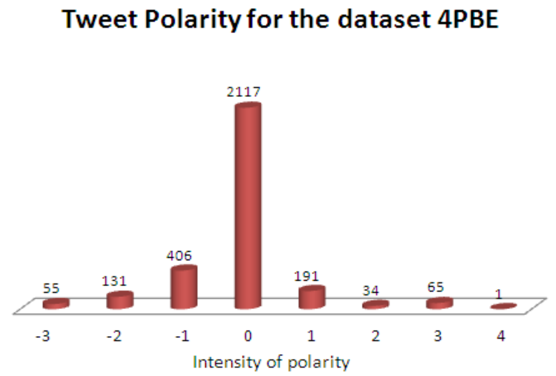

Fig.2 Tweet polarity for the dataset 4PBE

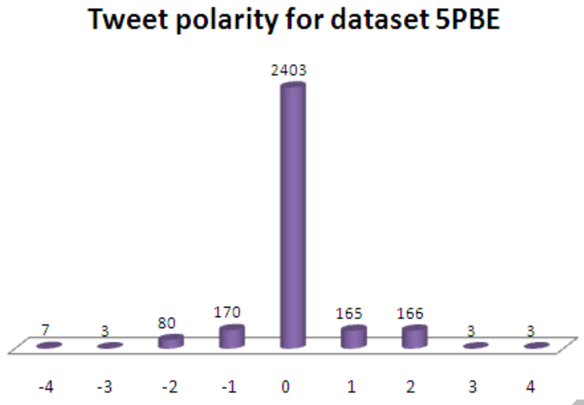

Fig. 3 Tweet polarity for the dataset 5PBE

CONCLUSION AND FUTURE SCOPE

Smaller is the MAE the more the accuracy. The results show that the performance of the classifiers is same. There is marginal difference in the MAE. The performance of the classifiers was made for three datasets (APJ, 4PBE, 5PBE). In the APJ dataset NaiveBayes performs best. In the 4PBE SVM is showing best performance, whereas in the 5PBE Max Entropy performs best. In future we plan to extract the emotions and polarity of the text data using Bayes classifier because it is simple and intuitive method.

REFERENCES

[1] Ghazaleh Khodabandelou, Charlotte Hug, Rebecca Deneckère, Camille Salinesi, (2011). Process Mining *Versus* Intention Mining, © *Springer- Verlag* Berlin Heidelberg.

[2] Zheng chen, Fan lin, Huan liu, In liu, Wei-ying ma, Liu Wenyin, (2002).User Intention Modeling in Web Applications Using Data Mining *World Wide Web: Internet and Web Information Systems*, 5, *Kluwer Academic Publishers. Manufactured in The Netherlands.* 181–191

[3] Bo Pang and Lillian Lee, Shivakumar Vaithyanathan, (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques, *Appears in Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP).*

[4] Jeonghee Yi,Tetsuya Nasukawa, Razvan Bunescu, Wayne Niblack, (2003). Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques, *Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03) 0-7695- 1978-4/03 © 2003 IEEE*.

[5] Minqing Hu and Bing Liu, (2004).Mining Opinion Features in Customer Reviews, *Copyright ©, American Association for Artificial Intelligence.*

[6] Lada Adamic, Natalie Glance, The Political Blogosphere and the 2004 U.S. Election: Divided They Blog.

[7] Lloyd L, Kechagias D, Skiena S, (2005). News and Blog Analysis with Lydia, *A System for Large-Scale News Analysis* presented at String Processing and Information Retrieval: *12th International Conference, SPIRE 2005*.

[8] Jeonghee Yi, Wayne Niblack, (2005).Sentiment Mining in Web Fountain, *Proceedings of the 21st International Conference on Data Engineering (ICDE2005)1084-4627/05 © IEEE*

[9] Andrew Mehler, Yunfan Bao, Xin Li, Yue Wang, and Steven Skiena, (2006).Spatial Analysis of News Sources, *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER* GRAPHICS, Vol. 12, No. 5,

[10] Levon Lloyd, Prachi Kaulgud, Steven Skiena, (2006).Newspapers vs. Blogs: Who Gets the Scoop? *Copyright ©, American Association for Artificial Intelligence*.

[11] Anindya Ghose, Panagiotis G. Ipeirotis, (2007). Designing Novel Review Ranking Systems: Predicting the Usefulness and Impact of Reviews, ICEC'07, Minneapolis, Minnesota, USA. Copyright, ACM 978-1-59593-700-1/07/0008.

[12] Rohitha Goonatilake, Susantha Herath, (2007). The Volatility of the Stock Market and News, *International Research Journal of Finance and Economics* , Issue 11

[13] Nikolay Archak, Anindya Ghose, Panagiotis G. Ipeirotis, (2007). Show me the Money! Deriving the Pricing Power of Product Features by Mining Consumer Reviews, *KDD'07, August 12 San Jose, California, USA*. Copyright 2007 ACM 978-1-59593-609-7/07/0008

[14] Mikhail Bautin, Lohit Vijayarenu, Steven Skiena, (2008). International Sentiment Analysis for News and Blogs, Copyright, *Association for the Advancement of Artificial Intelligence.*

[15] Kanayama Hiroshi, Nasukawa Tetsuya, Deeper Sentiment Analysis Using Machine Translation Technology.

[16] Rada Mihalcea, Carmen Banea, Janyce Wiebe, Learning Multilingual Subjective Language via Cross-Lingual Projections.

[17] Soo-min Kim, Eduard Hovy, Determining the Sentiment of Opinions.

[18] Andrea Esuli and Fabrizio Sebastiani, Determining Term Subjectivity and Term Orientation for Opinion Mining.

[19] Cort J. Willmott, Kenji Matsuura, (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, CLIMATE RESEARCH Clim Res, Vol. 30: 79–82.