© 2018 IJSRST | Volume 4 | Issue 2 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

An Approach to Sentiment Analysis

Wankhede Rohit*, Rajkumar Jagdale
Department of CS & IT, Dr.BAMU, Aurangabad, Maharashtra, India

ABSTRACT

In this paper we illustrate of various sentiment analysis methods. We explore the basic day to day use of various approaches of sentiment classification .This paper helps to take an overview about the sentiment analysis and polarity classification of opinions of given entity. The paper gives a view of the basic evaluation techniques of of sentiment sentiment analysis along with their performance accuracy according to the survey obtained of sentiments given by user. It also describes steps in sentiment analysis-data input, data preprocessing, selecting pre processed dataset, extracting feature list, training and classification, selection of feature list and related sentence, finding synonyms, calculating similarity and showing polarity.

Keywords : Sentiment Analysis, Opinion Minning, Support Vector Machines, Naïve Bayes, Maximum Entrophy.

I. INTRODUCTION

Communication in our day to day life has become a necessary mean. In our today's world communication and expressing of one's point of view over social media like twitter , facebook etc has become a self explaining fact. Opinion mining sometimes referred to as technique which make use of natural language processing text analysis, computational linguistics and biometrics to systematically identify, extract, quantify and study affective states and subjective information. Sentiment analysis is widely applied to voice of customer materials like comments, review over online web applications, social medias, news and other informational medium including marketing and customer services.

Generally speaking sentiment analysis strives to determine attitude and the expression with respect to the some topic or some topic over all contextual polarity or emotional reaction to document, or event interacted by person. knowledge.

II. LITERATURE SURVEY

In the past few years a lot of work has been done on "Sentiment analysis". In the beginning it was assumed for binary classification which estimates and assigns reviews to bipolar classes such as positive and negative. They [3] predicted average semantic orientation containing the adverb and adjective of a phrase deciding whether it is positive or negative. It [4] is base on survey report of summary of report where latent semantic analysis is used to identify opinions. It [5] compares positive and negative opinions. In this [6] demonstrated use of rule based method , based on SVM and baseline of sentiment analysis of Chinese document level. In it [7] calculates the polarity of each word that is positive or negative against all the words depending on sentence structure. They [8] have proposed use of LSA and cosine similarity for better quality of preprocessed data of raw sentence. In [9] part of speech based rules and dependency relation was used extraction and syntactical information from documents.

In [10] performed work on aspect based sentiment analysis for customer review. They [11] worked on social document to illustrate the importance of verbs in classification of sentiments. It [12] illustrated the various forms and use of lexicons in sentiment polarity. They [13] demonstrated use fuzzy logic in sentiment analysis for classification of customer reviews. Authors of [14] has desciribed membership function using fuzzy approach. Paper [19], [20] ,[22]describes the basic techniques for processing twitter data. In [21] downloaded tweets of different events and calculated positive and negative opinion of users.

III. TYPES OF SENTIMENT ANALYSIS

The basic element in sentiment analysis is to classify polarity of a given text a document, aspect or at a sentence level, against the expression of positive, negative or neutral. Moreover at advanced level it looks for expression of anger, happiness, sad etc. Many other methods are based on evaluation of polarity based on scaling ranging from positive to negative. For example in a movie review to predict star rating or a product ranking.

In many cases it is easier to determine sentiment by use of scaling system where words are associated with negative, positive, or neutral sentiment in them which are associated with a number on -10 to +10 scale.(most negative to most positive)or simply from 0 to a positive upper limit .

Subjectivity /Objectivity Detection.

This task usually classifies a given text based on two classes : objective or subjective. The subjectivity of words and text depend on context and objective may usually contain subjective phrases(eg. news article quoting peoples opinion). Phrase based woking is demonstrated in [15].

Apect Based

This refers to analysing and determining the opinions of people based on different features of a particular entity. Example jeans of a particular brands, camera quality of particular cell phone, service of entities which may represent a company. Various entities represent various properties which create some deficiencies in some or the other way..Thus resulting in representing of opinion of individuals of various polarities may be good, bad, neutral. The automatic identification of features can be performed using syntactic methods with topic modeling and Deep Learning.

Sentence level

We can download all our daily tweets on a particular activity and then analyse it's polarity based on positive, negative, or neutral, as well as confidence score. This score represents the accuracy of our prediction and determination in our daily life events. Sentence level can be widely used to find and detect novelty streams in news and as well other social media applications. The data source may be dataset from social medias, twitter, news dataset from google and review of movie etc.

IV. METHODS AND FEATURES

It is classified as following three methods statistical methods, knowledge based techniques and hybrid approaches.

Knowledge based techniques are based on classification of text based on presence of words such as anger, joy, sorrow, disappointment. Statistical methods imply on elements from machine learning such as latent semantic analysis, support vector machines and vector space modeling etc. The phrases are evaluated based on conspiracy that holder of sentiment is checked with respect to some. Grammatical relationship between words are achieved by deep vocabulary parsing, which are obtained by deep parsing of text. Hybrid approaches leverage on both machine learning and knowledge representation such as ontologies and semantic networks , to predict polarity and subtle information. Sentiment analysis

usually operates on large collections of texts, including web pages, online news, internet group discussions, online reviews, web blogs, and social media. Sentiment analysis can also be performed on images and videos. One of the first approach in this case is Sentibank that is utilizing adjective noun pair representation of visual content.

Evaluation

The accuracy is measured based on precision and two target categories of positive and negative texts. However according to survey human raters only agree upto 80% of the time. The most commonly used type of evaluation methods in phrases and text include SVM, cosine measures, Maximum Entropy, and Naïve Bayes algorithms. Following bar graph shows accuracy obtained from survey of following methods. Table below shows the performance of methods of sentiment analysis.

Table1. Sentiment Evaluation Techniques.

Sentiment evaluation techniques	**Performance**	
	Methods	***Accuracy***
1	Naïve Bayes	82.2
2	Maximum entrophy	83.8
3	Support Vector Machines	85.5
4	Semantic Analysis(Wordnet)	89.9

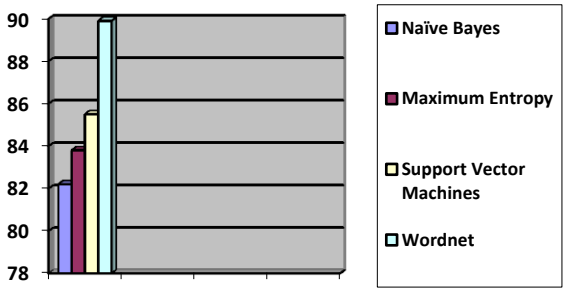

Figure 1. Performance comparison in terms of Accuracy.

Table 2. Sentiment Polarity Classification.

Sentiment Polarity Classification	**Text Analysis**	
	Text	***Polarity Type***
1	I am happy	Good
2	The movie was Good	Good
3	PM said that…..	Neutral
4	The Cell was bad	Bad

A. Naive Bayes

Because of its simplicity in training and classification, it is used as a machine learning technique. It can classify a set of categorised documents as in [16]. It compares set of words to the documents for classification.

C^* = argmaccPNB(c|d)

$$p_{NB}(c/d) = \frac{(P(c)\sum_{i=1}^{n} P(f/c)^{n_i(d)}}{P(d)} \quad \ldots\ldots(1)$$

Where Class c^* enumerates of text d, where, f is feature and ni(d) represents the count of feature fi found in text d. The total count of features are m . The maximum likelihood estimates derives Parameters P(c) and P (f|c) which are incremented by one for smoothing. The input data for Naïve Bayes is pre-processed along with feature extracted. After the training is complete, during classification the polarity of the sentiments can be estimated. For example for

the review comment "I am glad' it will provide Positive polarity in result.

B. Maximum Entrophy

It entrophy is maximized based on the conditional probability distribution. It is similar to logistic regression and handles overlapping feature. It's work as demonstrated. also follows certain feature exception constraints [17].

$$P_{ME}(c/d) = \frac{\exp\ [\sum_i\ \lambda i fi\ (c,d)]}{\Sigma \exp [\ \sum_i\ \lambda i fi(c,d)]} \quad \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Where, c is the class, d is the text, and is a weight vector. The weight vectors decide the significance of a feature in classification. It follows the similar processes as Naïve Bayes, discussed above and provides the polarity of the sentiments.

C. Support Vector Machines

It makes use of input space kernels for analyzing data, defining decision boundaries. Where two sets of vectors of size m each represent the data, and each vector is classified in specific class. The next step follows to find a partition between both classes far from the documents. SVM also supports regression and classification statistical learning theory. As seen in [18].

D. Sentiment analysis

Sentiment analysis is derived from Wordnet. This English database consist of bag of words ,each associated to each other, that is they are similar semantically. The technique is to map the words used by the user to words in database to check similarities. This is usefull for polarity detection. The polarity is displayed according to the requirement of analyser that is in statistical measure, scale or literally.

Following flow diagram shows execution of sentiment analysis

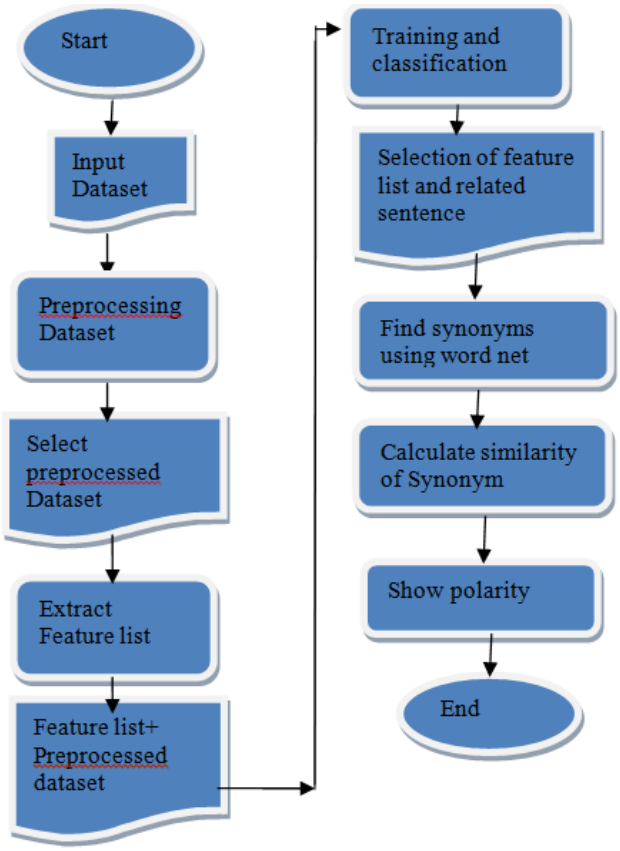

Figure 2. Flow Diagram of working of Sentiment Analysis.

V. APPICATION AND FUTURE SCOPE

Future Directions:

1. In industry: As the need for an advanced and better service has been become mandatory due to growth in advancement. Sentiment analysis gives a direction towards a better resolution , with feature level, business aligned and concerned and away from simplistic keyword based solutions.

2. Considering sources: Detection and exploitation of individual and as a whole in terms of emotion in speech and images (body language and facial) will necessarily come into action. The Solutions for smaller market languages will be demand driven.

3. Predictive modeling :The analysis on various domain and entities will bring out the necessary estimate and prediction.

4. A focus on the intent: The task of finding out the polarity, surely and absolutely aims to strive out on a emotion prediction and analysis on a accurate level.

VI. CONCLUSION

In this paper we have seen the basic sentiment analysis types and methodologies. We took a view of common methods used, Naïve Bayes, SVM, maximum entrophy, sentiment analysis. We also seen their Performance towards accuracy. And the core use of sentiment analysis in today's networking and social sector, where it can be used to predict performance of a particular product or entity based on review of customers. Further future work will gain success in predicting any event or entity based on peoples review.

VII. REFERENCES

[1]. R. Feldman, " Techniques and Applications for Sentiment Analysis ," Communications of the ACM, Vol. 56 No. 4, pp. 82-89, 2013.

[2]. Y. Singh, P. K. Bhatia, and O.P. Sangwan, "A Review of Studies on Machine Learning Techniques," International Journal of Computer Science and Security, Volume (1) : Issue (1), pp. 70-84, 2007.

[3]. P.D. Turney," Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, pp. 417-424, July 2002.

[4]. Ch.L.Liu, W.H. Hsaio, C.H. Lee,and G.C.Lu, and E. Jou," Movie Rating and Review Summarization in Mobile Environment," IEEE Transactions on Systems, Man, and Cybernetics, Part C 42(3):pp.397-407, 2012.

[5]. Y.Luo,W.Huang," Product Review Information Extraction Based on Adjective Opinion Words," Fourth International Joint Conference on Computational Sciences and Optimization (CSO), pp.1309 – 1313, 2011.

[6]. R.Liu,R.Xiong,and L.Song, "A Sentiment Classification Method for Chinese Document," Processed of the 5th International Conference on Computer Science and Education (ICCSE), pp. 918 – 922, 2010.

[7]. A.khan,B.Baharudin, "Sentiment Classification Using Sentence-level Semantic Orientation of Opinion Terms from Blogs," Processed on National Postgraduate Conference (NPC), pp. 1 – 7, 2011.

[8]. L.Ramachandran,E.F.Gehringer, "Automated Assessment of Review Quality Using Latent Semantic Analysis," ICALT, IEEE Computer Society, pp. 136-138, 2011.

[9]. B.Agarwal,V.K.Sharma,andN.Mittal,"Sentiment Classification of Review Documents using Phrase Patterns," International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1577-1580, . 2013.

[10]. J.Zhu, H.Wang, M.Zhu, B.K.Tsou, and M.Ma,," Aspect-Based Opinion Polling from Customer Reviews," T. Affective Computing2(1):pp. 37-49, 2011.

[11]. M.Karamibekr,A.A.Ghorbani,"Verb Oriented Sentiment Classification," Processed of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT), Vol (1): pp. 327-331, 2012.

[12]. A. Neviarouskaya, H.Prendinger, and M.Ishizuka," SentiFul: A Lexicon for Sentiment Analysis," T. Affective Computing 2(1), pp.22-36, 2011.

[13]. L.Liu, X.Nie,and H.Wang," Toward a Fuzzy Domain Sentiment Ontology Tree for Sentiment Analysis," Processed of the 5th Image International Congress on Signal Processing (CISP), pp. 1620 – 1624, 2012.

[14]. R. Srivastava, M. P. S. Bhatia," Quantifying Modified Opinion Strength: A Fuzzy Inference

System for Sentiment Analysis," International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1512-1519, 2013.

[15]. C. Tillmann , and F. Xia, "A phrase-based unigram model for statistical machine translation," Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL, pp.106-108, 2003.

[16]. B.Ren,L.Cheng," Research of Classification System based on Naive Bayes and MetaClass," Second International Conference on Information and Computing Science, ICIC '09, Vol(3), pp. 154 – 156, 2009.

[17]. C.I.Tsatsoulis, M.Hofmann,"Focusing on Maximum Entropy Classification of Lyrics by Tom Waits," IEEE International on Advance Computing Conference (IACC), pp. 664 – 667, 2014.

[18]. Hearst, Marti A., et al. "Support vector machines." IEEE Intelligent Systems and their applications 13.4 (1998): 18-28.

[19]. Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." Contemporary computing (IC3), 2014 seventh international conference on. IEEE, 2014.

[20]. Thejas Mol, Thomal Pretty Babu2"event based sentence level interpretation of sentiment variation in twitter data"

[21]. Rajkumar S. Jagdale, Vishal S. Shirsat, Sachin N. Deshmukh, "Sentiment Analysis of Events from Twitter Using Open Source Tool", International Journal of Computer Science and Mobile Computing, ISSN 2320–088X, Vol.5 Issue.4, April- 2016, pp. 475-485.

[22]. Shu long Tan, Yang Li,Huan Sun, et,al, "Intrepreting public sentiment variation in twitter,"IEEE Trans Knowledge and data Engineering, vol 26,no:5,M ay 2014.