International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 06, June 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

A Survey on Sentiment Analysis for Big Data

Swati Sharma[1], Mamta Bansal[2], Ankur Kaushik[3]

[1]Deptt. of C.S.,Shobhit University.,Deptt. of I.T. , M.I.E.T. ,India)

[2]Deptt. of C.S.E. , Shobhit University.(India)

[2]Deptt.of I.T, M.I.E.T.(India)

ABSTRACT

With the expeditious rate of the internet , number of people are exchanging their thoughts and opinions on numerous issues on microblogging websites . Microblogging websites are those social media sites on which one can post or share their emotions or feelings anytime.Sentiment analysis or opinion mining is very helpful in this field .An exact technique for analysing sentiments will help us to identify sentiments from I-net and identify user's choice. Numerous algorithms are available for Sentiment Mining.Sentiment Mining has three steps of granules i.e. Aspect level , Sentence Level and Document level. Ahead of applying any sentiment mining algorithm , one has to perform the pre-processing.Then on this pre-processed output tokenization of sentences is being done in which sentences are extracted and then the sentiment analysis is being performed by making rules. In this paper ,a number of algorithms for sentiment analysis are analyzed and challenges faced and applications in respect to this field are discussed.

Keywords: Sentiment Analysis, Machine Learning ,Opinion Mining, , Token, Performance Analysis, sentiment, polarity, Naïve Bayes Classifier, MEAD.

I. INTRODUCTION

Sentiment analysis is the procedure of calculating, identifying and grouping views represented in a form of text, specifically in order to identify whether the authorsbehaviour towards a particular taskis positive, neutral or negative.Opinion Mining also refers to NLP
(Natural Language Processing) , biometrics , text analysis and computational linguistics in order to detect , extract and refer subjective information.

Sentiment analysis basically aims to identify the attitude of a writer with respect to a topic or the complete polarity to a document. The behaviour may be aevaluation or judgemental or affective state of the author or the emotional communication or interlocutor.

It is the calculative study of users opinions, views , behaviour and emotions toward an object . Sentiment mining helps to gather positive , negative or neutral information about a product.Then , the highly counted opinions about a product are passed to the user.For promoting marketing, big companies and business magnets are making use of this opinion mining .

Using given studies, sentiment analysis is being performedat three levels i.e. attribute level , sentence level or document level. In Attribute level sentiment analysis ,a sentiment for each entity in a sentence is provided. In sentence level sentiment analysis, the overall sentiment of each and every sentence in a document is provided. In Document level sentiment analysis,the overall sentiment of the complete document is provided.

According to literature survey done in respect to sentiment analysis,there are two techniques i.e. semantic orientation and machine learning which are important.The semantic orientation of a view suggests whether the view is positive , negative or neutral whereas machine learning is a technique of data analysis which automates logical building of a model. The techniques are shown in Fig.1.

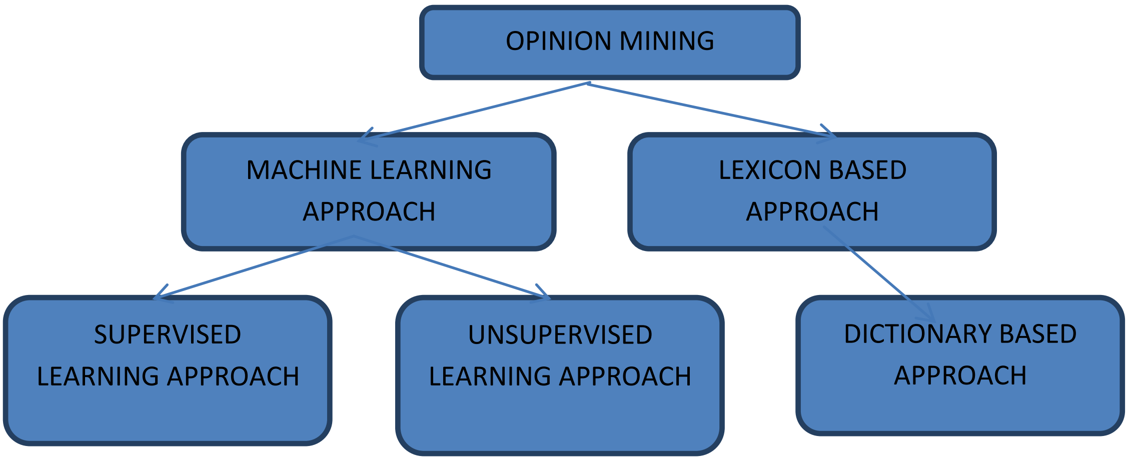
OPINION MINING
MACHINE LEARNING APPROACH
LEXICON BASED APPROACH
SUPERVISED LEARNING APPROACH
UNSUPERVISED LEARNING APPROACH
DICTIONARY BASED APPROACH

Figure 1. Opinion Mining Techniques

In the supervised learning approach of machine learning ,pseudo codes are trained using descriptive examples, as an input in which the desired output is already known. It is basically used in applications where historical data is used to predict forthcoming data. In the unsupervised learning approach of machine learning , there is no historical data.The objective is to investigate the data and to make some useful information within it .

II. RELATED WORK

1. Supervised Learning Approach

According to supervised learning , pseudo codes are trained using descriptive examples, as an input in which the desired output is already known. It is basically used in applications where historical data is used to predict forthcoming data.

Methods of Supervised Learning:

1.1 Decision Tree Classifier

The Decision tree classifier is an elementary and broadly used classification procedure. The basic objective of this classifier is to produce an output of classification problem .The Decision tree classifier produces a series of interrelated questions and each time an answer is received , it produces an another series of question untill a conclusion is reached and recorded. Movie review attributes gathered from IMDb was withdrawn using inverse document frequency and the significance of the word found. Principal component analysis and Classification and Regression Tree (CART) were used for attribute selection based on the significance of the work with respect to the complete document. The classification accuracy obtained by Learning Vector Quantization (LVQ) was 75%. Exploring emotional differences in adolescent age and ground behind these differences using data mining techniques is proposed. By categorizing emotions and using decision tree various emotional variations are recorded. If-then rules are also produced from decision tree.

1.1.1 Linear Classifier

In Linear Classifier support Vector Machine and Neural Network is used .A support vector machine builds a set of hyperplane inan N-dimensional space, which is further used for variety of tasks such as classification , regression etc.A Neural network is a calculative model which is used in machine learning consists of a big collection of simple units known as artificial neurons.

1.1.2 Rule Based Classifier

In Rule based classifier , the complete data set is represented using a set of rules . The right hand side represents a class label whereas the left hand side mentions a condition on the feature set represented in DNF(Disjunctive Normal Form) .

1.1.3 Probabilistic Classifier

The probabilistic classifier consists of Naïve Bayes , Bayesian network and maximum entropy.The Naïve Bayes classifier is the straightforward and generally used classifier . On the basis of distribution of words in a record , it calculates the posterior probability of a class. The main objective of the Naive Bayes classifier is the independence of the attributes . One assumption in this is to predict that all the attributes are fully dependent . The Maximum Entropy Classifier also regarded as conditional exponential classifier translates labelled attributes sets to vectors using encoding . For each attribute , this encoded vector is helpful in calculating weights .

1.2 Dictionary Based Approach

According to Dictionary based approach, manual collection is being performed of seed words in respect to positive or negative orientations. Then this collection is taken further by matching with their synoynms and antonymns by using their online dictionary . New words are concatenated to existing seed list and then the next iteration is initiated.If no new words are found then the iteration is dropped.Cleaning of list is performed by manually inspecting.

Kumar puspesh presented a paper on , “ Multi-document update and opinion summarization “This paper presented an automatic keyphrase extraction techniques .The process of selecting keyphrases from a document comprises of selecting salient words and multi-word units, generally noun compunds no longer than a threshold, from an input document.

ShanmugasundaramHariharan presented a paper on , “Extraction Based Multi Document Summarization using Single Document Summary Cluster “ This paper presented a multi-document summarization approach for sentence selection and producing summaries for each and every single document and combining the sentences in an order.This approach is much better in comparison to MEAD summarizer.

III. CHALLENGES TO OPINION MINING

- Opinion mining in interrogative sentence may have some problem because these sentences have neither positive nor negative sentiments.
- Opinion mining in sarcastic sentences may encounter problems because these sentences doen’t have clear cut meaning.

- Opinion mining in funny sentences not only violate the particular sentence but also change the meaning of the complete document.
- It is not necessary that sentence should contain the sentiment words but still reflect the positive or negative attitude.
- Opinion mining in spam sentence is difficult to predict as they are the sentences posted by the competitor organization in order to increase one's value.

IV. CONCLUSION

This survey paper demonstrated an overview of the current updates in opinion mining .Sentiment Mining has become extremely popular in the field of research technique. A lot of research has already been done but still certain challenges to sentiment mining still exist related to unstructured data . The keen in languages other than english in regard to this areas is raising at a much higher rate but still lack of resources exist in this field.The most repeated lexicon source used is WordNet , exist in languages apart than English. According to this survey, it can be accompolished that supervised techniques provide much better accurate result in comparison to dictionary technique.

REFERENCES

[1] G.Vinodhini, R.M.Chandrasekaran, "Sentiment Analysis And Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, June 2012

[2] Jiawei Han, Micheline Kamber and Jian Pei, "Data mining Concepts and Techniques", Third Edition, Morgan Kaufmann Series in Data management Systems

[3] DikshaSahni, Gaurav Aggarwal, "Recognizing Emotions and Sentiments in Text: A Survey ", International Journal ELSEVIER, 2013 of Advanced Research in Computer Science and Software Engineering , 2015

[4] Charu C. Aggarwal, "Data Mining: The Textbook", Springer, 2015

[5] Rizvaan Irfan, Christine K. King ,"A Survey on Text Mining in Social Networks", The Knowledge Engineering Review, 2004

[6] A.Jeyapriya, C.S.KanimozhiSelvi,"Extracting Aspects And Mining Opinions In Product Reviews Using Supervised Learning Algorithm", IEEE, 2015

[7] JeevanandamJotheeswaran, Dr. Y. S. Kumaraswamy, "Opinion Mining Using Decision Tree Based Feature Selection Through Manhattan Hierarchical Cluster Measure", Journal of Theoretical and Applied Information Technology, 2013

[8] BlessySelvam1 , S.Abirami2, "A Survey On Opinion Mining Framework", International Journal of Advanced Research in Computer and Communication Engineering, 2013

[9] Kai Gao, Hua Xu, JiushuoWanga, "A Rule-Based Approach To Emotion Cause Detection For Chinese Micro-Blogs", ELSEVIER, 2015

[10] ChetashriBhadane,HardiDalal, HeenalDoshi, "Sentiment Analysis: Measuring Opinions", Science Direct, 2015

[11] Weiyuan Li, Hua Xu, "Text-based emotion classification using emotion cause extraction",

[12] Richa Sharma, Shweta Nigam, Rekha Jain, "Polarity Detection at Sentence Level", International Journal of Computer Applications, Volume 86- No 11, 2014

[13] Vikrant Yadav. 2016. thecerealkiller at SemEval-2016 Task 4: Deep learning based system for classifying sentiment of tweets on two point scale. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016), San Diego, US.

[14] V. Sahayak, V. Shete, and A. Pathan, "Sentiment Analysis on Twitter Data", IJIRAE, 2015.

[15] Yunxiao Zhou, Zhihua Zhang, and Man Lan. 2016. ECNU at SemEval-2016 Task 4: An empirical investigation of traditional NLP features and word embedding features for sentence-level and topic-level sentiment analysis in Twitter. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016), San Diego, US.

[16] A. S. Patwardhan, “Multimodal Affect Recognition using Kinect”, ACM TIST (in review), 2016.[17] A. S. Patwardhan, “Augmenting Supervised Emotion Recognition with Rule-Based Decision Model”, IEEE TAC (in review), 2016.[39] A. S. Patwardhan, Jacob Badeaux, Siavash, G. M. Knapp, “Automated Prediction of Temporal Relations”, Technical Report. 2014.