



ROLE OF VARIOUS DATA MINING TECHNIQUES IN SENTIMENTAL ANALYSIS

Milanjit Kaur
Computer Science and Engineering
Lovely Professional University,
Punjab, India

Komal Arora
Computer Science and Engineering
Lovely Professional University,
Punjab, India

Abstract: Data is present in large amount and it became quite hectic to deal with with this amount of data but data mining has made it easy. Data mining has given various concepts which has made the process of sentimental analysis easy. Sentiment analysis involves the usage of text scrutiny and processing in order to recognize the patterns plus gaining the useful information. Various techniques namely Natural Language Processing, Machine learning, Text mining are involved in this process of sentiment analysis. By collecting different opinions, polarity is found out and is categorized as positive, negative or neutral.

Keywords: data, mining, sentimental, analysis, machine, learning, text, mining

I. INTRODUCTION

Data mining allows the users to check the different aspects of the data and assembles them to reach to the valuable conclusion. Data mining involves examination of the data to classify it into various dimensions. Data mining involves the usage of previously known information to predict the pattern and correlating. Data mining have various applications. Sentiment analysis can be used with common sense with the help of septic computing[1]. New opportunities tend to generate if the quality and size of the data is acknowledgeable. The main job of sentimental analysis is to analyze the given text and further classify the text into various categories by keeping the polarity factor in mind. Sentimental analysis is done at the document level, sentence level or feature/aspect level. It categories given data as positive, negative or neutral. Along with polarity, emotion study also pays the attention to the states of emotions. Sentiments found within comments or posts provide useful information that can be used for various purposes. These emotions are divided into three categories in terms of polarity positive negative or neutral or into n-fact scale e.g. good, very good, bad, satisfactory, angry, happy etc. So due to this is termed as a classification task. Sentiment analysis help companies or organizations to know about their product's quality and acceptance of the products by their customers and if in case they are not liked by them then to know the reasons and determining the strategies to improve them. System, clients fill their feedback and then system analyze the system to improve their service [2].

II. LITERATURE SURVEY

In 2015, Santanu Mandala et al. [3] performed a work "Lexicon-Based Text Classification Model to analyses and Predict Sentiments from Online Reviews" in which an algorithm is formulated which shows the significant role in

predicting the emotion of the text. It categorize the text as positive, negative and neutral. This model also determines the method to improve the accuracy of the system. In this model we first ensemble the information from various sources and then we make a list the words on the basis of polarity. We have database which consist of text which reflects the emotion and then we do processing in which words are compared to words placed in Database. The list of the words depicting emotions are splatted further into various lists. Various value are assigned to different words based on the emotions the emotions they are carrying. We can also list the words with their respective merits and demerits with the help of supervised learning. Irony, Pun and sarcasm are the topics which can also use to make our system more realistic.

In 2014, Pal Christian et al. [4] performed a work "Evaluating Feature Sets and Classifiers for Sentiment Analysis of Financial News" in which five algorithms are used to determine the emotion of news. Fewer effort is done on the opinion mining of the news so this research work has been carried to put focus on this term. Various subsets are classified to determine their accuracy. When we use two classifiers in conjunction, then the accuracy gets improved. In this research we conclude that Support vector machine provides the best results. We extract the data from various articles on the basis of the features. We further split the list into various categories. We studied the exact definition of the text which identify the feature related to the category it belongs. Accuracy of 71% is received by using this model. Range of the dataset can be increased to improve the performance of the data. We can also increase the accuracy of the model by adding grammatical properties.

In 2016, Nirbhay Kashia et al. [5] performed a work "Mood Based Classification of Music by Analyzing Lyrical Data Using Text Mining" in which mainly focus is on Audio and lyrical content. Data mining process is performed on the

lyrical content .Firstly Ground set of database is formed out of which we extract the lyrics .Same process is done with Audio dataset .We further compare various types of lyrics .Classification of the lyrics is done with the help of Naïve Bayes. We compare the content of Lyrics with the content of audio. We ensemble the results from various algorithms and then do comaparision.Further we compare the results of Audio based system with the result of Lyrics based systems. We configure the total 18 mood classifications and then assigned a value to each rank. Process of Bagging is performed on the list which ensemble the results of various classifiers. We conclude when the words gets repeated and then audio is played in combination then the results are more accurate. This proposed model has achieved the accuracy of 85%.

In 2014, Xiao Sun *et al*. [6] performed a work “Mining the Impact of Social News on the Emotions of Users Based on Deep Model “in which use of Deep Belief Nets (DBN) has been made to extract the hidden emotions of the news. In this process ,use of three feature selection has been made to build the input vector .The knowledge of text classifier has been used to make three models .We make use of DBN to reduce the dimension of feature selection and to increase the accuracy of the process. We determines the list of Dimension of Original properties with the list of abstracted features. The framework of DBN tends to effect the working of classification process. After all the processing, properties are set to the value of 500.In future we can also focus on the performance of classifiers.

In 2016, Ms.A.M.Abirami and Ms.V.Gayathri [7] performed a work “A Survey on Sentiment Analysis Methods and Approach “in which detailed study of the various works done on the sentimental analysis is done. Different methodologies and approaches are determined and are listed with the help of lists. Detailed study is done on the terms, namely polarity shift, accuracy .Various classification techniques are observed and then comparison is made between all the mechanisms. This paper determines that various features such as Machine Learning and Dictionary based provides the good results .This survey also determines the merits and demerits of various algorithms.

In 2017, Kashif Ali [8] performed a work “Sentiment Analysis as a Service: A social media based sentiment analysis framework “in which it proposed a model that extract the emotions from various services and then convert them into useful information. We focused on the properties of the users which make use of social services. Work has been done I the real world scenario experiments .This research work analyses the performance of various algorithms. Classification of the data is done on the basis of various properties .It eliminate the inconsistent data and work on the data independent of the error. The range of the data can be increased and we can use multiple services for the canalization process. Thee classification of the social services can be improved.

III.COMPARISON OF VARIOUS DATA MINING TECHNIQUES

Table 1: Comparison of various Techniques

Author	Year	Techniques	Advantages	Disadvantages
Santanu Mandala	2015	Classification	1. Common sense knowledge is applied to perform Fine-Grained Sentiment analysis. 2. News mining is done and it is difficult because they avoid usage of direct positive or negative language.	1. Performance achieved sematic parser is less
		Lexicon-Based Text Model		
		Sentiment Analysis		
		Common – Sense Knowledge		
Pal Christian	2014	Sentiment Analysis	1. Analysis of sarcasm and negations is done. 2.Comparision of different text Mining and classification algorithms is done	Data set is not wide
		MALLET		
		Text Mining		
Nirbhay Kashia	2016	Text Mining	1.Area of News has been taken in which less research work is done	1. Analysis of sarcasm and negations in the text is not done.
		Sentimental Analysis		
		Opinion Mining		
		News Analysis		
Xiao Sun	2014	Sentiment Analysis	2. Evaluation of combinations of different classification algorithms and	Less data is taken to avoid complexity
		Text Mining		

		Classification	filtering scheme. 2. Filtering schemes reduce the original Dataset.	
Ms.A.M.Abirami and Ms.V.Gayathri	2016	K-nearest Neighbor	1. Data is of wider range. 2. Compares the performance of Single classifiers with ensemble of classifiers.	Issues such as Polarity shift problem, data sparsely are not covered
		Random Forest		
		Naive Bayesian		
		Opinion Mining		
		Sentiment Analysis		
		classification		
Kashif Ali	2017	Naïve Bayes	Accuracy of different genre and opinions is calculated	Data is not of wider genre
		SVM		
		Opinion Mining		
		Sentiment Analysis		
		Opinion Mining		
		Feature Selection		
		Maximum Entropy		

IV. CONCLUSION

We have concluded that large amount of research has been done facebook,twitter and less research has been conducted in sentimental analysis of news.We studied various data mining techniques that are used for sentimental analysis.

V.REFERENCES

[1] E. Cambria and A. Hussain, *Sentic Computing: Techniques, Tools, and Applications*, Dordrecht, Netherlands: Springer, ISBN: 978-94-007-5069-2

[2] F. Greaves, D.R. Cano, C. Millet, A.Darzi, and L. Donaldson, "Use of Sentiment Analysis for Capturing Patient Experience from Free-Text Comments", *Journal of Medical Internet Research*, 2013, 15:11, e239. Online publication date: 1-Jan-2013

[3] Santana Mandal and Sumit Gupta "A Lexicon –Based Text Classification Model to Analyze and predict Sentiments from Online Reviews", *ACM Trans. Inf. Syst.*, Vol. 21, No. 4, pp. 315–356, 2004.

[4] Pal-Christian et al."Evaluating Feature Sets and classifiers for sentiment Analysis of Financial News", *International conference on Web Intelligence*, 978-1-4799-4143-8/14 ©2014 IEEE.

[5] Nirbhay Kashyap et al "Mood Based Classification of Music by Analyzing Lyrical Data Using Text Mining", *International Conference on Micro –Electronics and Telecommunications Engineering*, 978-1-5090-3411-6/16 ©2016 IEEE.

[6] Xiao Sun et al. "Mining the Impact of Social News on the Emotions of Users Based On Deep Model", 978-1-4799-4719-5/14©2014 IEEE.

[7] Ms.A.M.Abirami and Ms.V.Gayathri "A Survey on Sentiment Analysis Methods and Approach", *Eighth International Conference on Advanced Computing*, 978-1-5090-5888-4/16©2016 IEEE.

[8] Kashif Ali "Sentiment Analysis as a Service: A social media based sentiment analysis framework", *24th International Conference On web Services*, 978-1-5386-0752-7/17 ©2017 IEEE.