

Comparative Study of various Surveys on Sentiment Analysis

Milanjit Kaur¹, Deepak Kumar².

¹Student (M.Tech Scholar), Computer Science and Engineering, Lovely Professional University, Punjab, India.

²Assistant Professor, Computer Science and Engineering, Lovely Professional University, Punjab, India.

ABSTRACT

Content development in the Internet as of late has made an enormous volume of data accessible. Tons of data is generated every single day and we are making no use of it. Data is rich and knowledge is very poor. Data analysis provides the efficient way to take a good decision. By using various Data mining techniques, analysis of sentiments is done. Analysis of sentiments in various fields, namely, twitter, movie reviews and product reviews has been done. In this paper, we have discussed the various researches done in the field of sentiment analysis. Much work has been done in sentiment analysis, even then there are few areas that are still need to be covered. Comparison has been made between various approaches used

1. INTRODUCTION

Sentimental analysis is a process of determining the emotions of the author in the text. Sentimental analysis is a type of Text mining in which opinion of the author is determined. The amount of online data is increasing day by day. Users give their views online on product, movies, and books. So text analysis is needed by every organization to increase their sale. Users give their views and organization do sentimental analysis to determine the emotions behind the user's views. Organization generate policies and make changes in the product according to the views given. In sentimental analysis, polarity of the words is checked by using semantic orientation techniques [1]. Emotions could be positive, negative or neutral. Emotions are determined by finding the relation between the style of the author and sentimental state [2]. Machine learning is also used to determine the emotions in the text [3]. It has become easy for the users to decide whether they should go for the movie or not. The users give their reviews about particular movie online. These reviews are categorized as positive and negative by using sentimental analysis. It has become easy for the users to decide whether they should go for the movie or not with the help of sentimental analysis. Sentimental analysis is a process of finding the sentimental state of the author. Natural processing is done to find out the emotions behind the particular text. Today, Web is populated with large amount of data and it is increasing day by day. It has become difficult for users to choose one product. Sentimental analysis plays a significant role in choosing the product by analyzing the reviews given by the users, Sentimental analysis determine whether the user has appreciate the product or criticize the product.

1.1 Data Mining Role :

Data mining plays an important role in sentimental analysis. Text pre-processing is done in all types of sentimental analysis. Text pre-processing is mainly don to filter the text. Text processing is done prior to classification. Linear Support Vector Machine, a Data Mining Technique, is used for sentimental analysis by classifying the opinion and doing regression. Other data mining technique which is used in sentimental analysis is Naïve Bayes theorem. Naive Bayes theorem is used to find out the likelihood function and then relation between two occasions is determined by making use of hypothesis theorem. Maximum Entropy is also used for sentimental analysis by doing classification. Classification probability is find out which helps to classify the emotions. K-means Clustering makes clusters of the members having similar properties. It makes groups of the texts which persue the same emotions.

1.2 Why News Mining

Research work has been done in other fields, namely, product reviews, movie reviews and twitter, but less work has been done in news mining. Users are free to write their views without being diplomatic. They are free to express their views but in News mining you are not supposed to object on anything freely. You need to be very careful while using language. It should not be clearly positive or negative. Authors are not supposed to be opinionated. So the mining in news articles is more difficult. Text used mainly consists of complex language so it becomes difficult to determine the sentiment in news.

2. Literature Survey

In 2016, Shweta Rana and Archana Singh [4] proposed a work “**Comparative Analysis of Sentiment Orientation Using SVM and Naïve Bayes Techniques**” in which movie reviews are analyzed by taking data set from Internet Movie Database comprised of both positive and negative reviews. Filtration of the content is done by doing text processing. Elimination of the suffixes is done to convert the data into valuable information. Unimportant additions are removed. Support Vector Machine and Naïve Bayes classifier are the two techniques used to classify the data and to solve the regression. According to analysis, among all types of movies, drama related movies are most liked. Rapid Miner is the tool which is used in this experiment.

In 2015, Anurag P.Jain and Vijay D.Katkar [5] performed a work “**Sentiments Analysis Of Twitter Data Using Data Mining**”, in which mining of the twitter data is done to depict the emotions of the user and their sentiments towards politics are depicted. Comparison of single classifier and ensemble classifier is done by using various mining classifiers. Data set comprised of 2,102,52 tweets collected by using Twitter API v 1.1. Preprocessing of data is done to convert the large amount of data into valuable information by removing user information and duplicate data. SentiWordNet is used to analyze the news as positive, negative and neutral. Various Classifiers, namely, k-nearest neighbour, Random Forest, Naïve Bayesian, Bayesnet are used and the best result is provided by K-means Neighbour with accuracy of 99.6456%.

In 2016, Shrawan Kumar Trivedi and Ankita Tripathi [6] performed a work “**Sentiment Analysis of Indian Movie with various feature Selection techniques**”, in which sentiments of the user are analysed by applying feature selection techniques on the movie review. Data is collected from www.imdb.com site. Classification of the reviews is done as good and bad. Preprocessing of the movie reviews is done to convert them into binary representation. Different Feature Selection Techniques, namely, Gain Ratio, Chi-Squared, Relief F, One Rule are used to classify the data. Java and Microsoft excel 10 platform is used to do this experiment. Experiment shows that Relief-F provides the best accuracy.

In 2014, Jinyan Li et al [7], performed a work “**Hierarchical Classification in Text Mining for Sentimental Analysis**”, in which different classification techniques were analysed and used for text mining. Sentiment are analyzed by taking dataset from different news articles. Dataset comprised of 268 articles, out of which some are taken as training data and others as testing data. Different filtering classification techniques, namely, Naïve Bayes, C45, Decision Tree, are used and compared. Three filters are used to evaluate the polarity and others two are used to filter out the unique or high frequency words. Result shows that Max Entropy and Naïve Bayes gives the best result and Decision trees provides the result with poor accuracy.

In 2016, Jagbir Kaur and Meenakshi Bansal[8] performed a work “**Multi-Layered Sentiment Analysis Model for Product Reviews Mining**”, in which reviews on the users on the products are analyzed and then classified as positive, negative or neutral. Dataset is taken online and processing is done. Polarity of the message is analyzed and weightage of particular emotion is listed using Review Analytical Algorithm. Data classified are aggregated to specify the details of particular category. Model is created to compare different models. Model created is compared with existing models. Accuracy is improved from 82% to 99%.

In 2013, Prashant Raina[9] performed a work “**Sentiment Analysis in News Articles Using Sentic Computing**”, in which opinion mining engine is formulated which classified the news articles as positive, negative or neutral. Semantic parser is used to extract the meaningful information from the data. SenticNet and ConceptNet is used to do the sentiment analysis. Data set is comprised of 500 articles taken from different sources. Different parameters, namely, Accuracy, F-measure, Precision, are taken to consideration. Accuracy received is 71% and is more as compared to Wilson et al model.

In 2016, Amir Hamzah and Naniek Widyastuti [10], performed a work “**Opinion Classification using Maximum Entropy and K-Means clustering**”, in which Opinion Classification system was framed by using which different views, comments, advices are classified. Maximum Entropy and K-means Clustering are the two techniques which have been used to analyze the opinion of different users. In this system we have taken the dataset of 2000 comments. TF/IDF is the scheme used for this purpose. Preprocessing is done in which stemming words are eliminated. TF values received by doing stemming are used to train and test the data. Complexity in terms of time and accuracy is measured and K-Means Clustering provides the better result as compared to Maximum Entropy with average precision of 3%.

In 2013, Simon Fong et al. [11] performed a work “**Sentimental analysis Of Online News using Mallet**”, in which MALLET (Machine Learning for Language Toolkit) was used to do opinion mining of the online news. 50 news articles are taken as dataset. Dataset is further divided into training set and testing set. Different classification techniques, namely, Naïve Bays, Maximum Entropy, Decision tree are used to classify the data as positive, negative and neutral. Result obtained shows that Naïve bayes performs better than other classification techniques.

In 2013, S Padmaja [12] performed a work “**Analysis of Sentiment on Newspaper Quotations: A Preliminary Experiment**” in which opinion mining of newspaper by framing a model. Data set comprised of 95 quotes from different newspapers. Data is preprocessed to eliminate the stop words and then objective of the quote is analysed by using SentiWordNet. Polarity is checked by using Sentiment Analyzer. Accuracy received is 0.465 which proved that open domain sentiment analysis is more difficult to achieve.

3. COMPARISON OF VARIOUS APPROACHES

Author	Year	Techniques	Advantages	Disadvantages
Prashant Raina	2013	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 semantic parser is less.
		Sentic computing		
		Sentiment Analysis		
		Common – Sense Knowledge		
Simon Fon, et. al.	2013	Sentiment Analysis	1. Analysis of sarcasm and negations is done. 2. Comparison of different text Mining and classification algorithms is done	Data set is not wide
		MALLET		
		Text Mining		
S Padmaja	2013	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		
Jinyan Li,et al.	2014	Sentiment Analysis	2. Evaluation of combinations of different classification algorithms and filtering scheme. 2. Filtering schemes reduce the original Dataset.	Less data is taken to avoid complexity
		Text Mining		
		Classification		
Anurag P.Jain et.al.	2015	K-nearest Neighbour	1.Data is of wider range. 2.Compares the performance of Single classifiers with ensemble of classifiers.	Issues such as Polarity shift problem,data sparsity are not covered
		Random Forest		
		Naïve Baysin		
		Opinion Mining		
		Sentiment Analysis		
		classification		
Shweta Rana	2016	Naïve bayes	Accuracy of different genre and opinions is calculated	Data is not of wider genre
		SVM		
		Opinion Mining		
		Sentiment Analysis		
Shrawan Kumar,et.al.	2016	Sentiment Analysis	1. Machine learning is used to increase the learning capability of the classifier. 2. Comparative analysis is performed.	Data set is not appropriate for testing different supervised machine learning
		Opinion Mining		
		Feature Selection		
Amir Hamzah	2016	Opinion Classification	Less computational Complexity1	Irony,sarcasm,pun,duality are not covered
		Maximum Entrophy		

Table 3.1: comparison

4. CONCLUSION

We have concluded that considerable measure of work done on Sentiment investigation of motion picture surveys, item audits, twitter, Face book and so forth however there has been less work done on daily paper articles. This survey gives us the knowledge about various sentimental analysis approaches and their respective issues.

5. REFERENCES

1. G.Forman," An Extensive Empirical Study of Feature Selection Metrics for Text Classification" The Journal of Machine Learning .Res.3. 2003:1289-1305.
2. Yessenov, Kaut and Sasa Misailovic, "Sentimen analysis of movie review comments"Methodology (2009):1-17.
3. Bo Pang, Lillian lee. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with respect to Rating Scales, ACL2005:115-124.
4. Shweta Rana and Archana Singh, "Comparative Analysis of Sentiment Orientation Using SVM and Naïve Bayes Techniques", International Conference on Next Generation Computing Technologies, 978-1-5090-3257-0/16 ©2016 IEEE.
5. Anurag P.Jain and Vijay D.Katkar, "Sentiments Analysis Of Twitter Data Using DataMining", International Conference on Information Processing,978-1-4673-7758/15 ©2015 IEEE.
6. Shrawan Kumar Trivedi and Ankita Tripathi, "Sentiment Analysis of Indian Movie with various feature Selection techniques", International Conference on Advances in Computer Applications,978-1-5090-3770-4/16©2016 IEEE.
7. Jinyan Li et al., "Hierarchical Classification in Text Mining for Sentimental Analysis", International Conference on Soft Computing and Machine Intelligence, 978-4673-6751-6/14 ©2014 IEEE.
8. Jagbir Kaur and Meenakshi Bansal, "Multi-Layered Sentiment Analysis Model for Product Reviews Mining", International Conference on Parallel, Distributed and Grid Computing, 978-1-5090-3669-1/16/©2016 IEEE.
9. Prashant Raina, "Sentiment Analysis in News Articles Using Sentic Computing ", International Conference on Data Mining Techniques,978-0-7695-5109-8/13©2013 IEEE.
10. Amir Hamzah and Naniek Widyastuti, "Opinion Classification using Maximum Entropy and K-Means clustering", International Conference on Information ,Communication technology and system,978-1-5090-1381-4/16©2016 IEEE..
11. Simon Fong et al., "Sentimental analysis Of Online News using Mallet",International Conference on Computational and Business Intelligence 978-0-7695-5066-4/13 ©2013 IEEE.
12. S Padmaja, et al. "Analysis of Sentiment on NewspaperQuotations: A Preliminary Experiment ", 4thICCCNT, IEEE, 2013.