INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Sentiment Analysis: A Survey

Suman[1], Jaswinder Singh[2]

[1, 2] Department of Computer Science and Engineering, GJUST, Hisar

Abstract: Sentiment analysis or opinion mining classifies the human's opinion or reviews into the positive, negative and neutral class which are written in form of text about some topic. Social networking sites such as Twitter, Facebook etc are rich in comments, customer reviews, opinion and sentiments. This paper describes the study of different sentiment analysis methods on different web resources such as review sites, blogs, discussion forums and news. It also gives overview of tools and application area of sentiment analysis.

Keywords: Sentiment Analysis, Opinion Mining, Supervised and Unsupervised Learning Approach

I. INTRODUCTION

Due to the rapid growth of the internet, people access and share data on the internet every day. People share their own thoughts and opinions in regards to their day to day life, business, entertainments, polities etc. An enormous amount of information and opinion online is scattered and unstructured with no machine to arrange it. That encourage us to study the field of sentiment analysis. So currently Sentiment analysis is one of the hot topic of research. It is also known as opinion mining. Today, the web is full of opinion of people. Textual data is present on the web in two forms: - facts and opinion (or sentiments). Facts present the objective statements and opinions are subjective statements. Sentiment analysis tracks the opinions, attitudes, feeling, thoughts, views and emotions of people about an entity. The entity can be a product, service, event, issues, person and organization. It classifies a given text into three categories: - positive, negative and neutral. Natural language processing is an area of computer science that deals with the interactions between human (natural) languages and computers. Sentiment analysis is the branch of NLP (Natural Language Processing) for analyzing and examining the public opinion. NLP is related to the area of human– computer interaction. NLP is a method for computers by which human language can be understood and explore in a smart and useful way. NLP is used to analyze text and to understand human language is to understand not only the words but the concepts and how they're linked together to create meaning. "Natural language processing" here refers to the use and ability of systems to process sentences in a natural language such as English, rather than in a specialized artificial computer languages.

II. LEVELS OF SENTIMENT ANALYSIS

Sentiment Analysis can be done at three levels which are sentence level, document level and aspect level [16][31].

A. Sentence Level

Sentence level sentiment analysis classifies polarity of each sentence in a document whether it is positive or negative. This level close to document level but here it accomplished by every sentence. In the case of simple sentences, a single sentence contains a single opinion about an entity. But there will be complex sentences also in the opinionated text. In such cases, sentence level classification is not desirable.

B. Document Level

Document level sentiment analysis classifies polarity of entire document as positive, negative or neutral. It considers the whole document focus about one topic. It aims to classify an opinion or a single review where a single topic is to be studied. The challenge is that all sentences in document may not contribute to opinion about a specific entity.

C. Aspect Level

It is also known as feature level sentiment analysis. The two previous approaches work well when either the whole document or each individual sentence refers to a single entity. However, in many cases, people talk about entities that have many aspects (attributes) and they have a different opinion about each of the aspects.

III. TECHNIQUES/APPROACHES FOR SENTIMENT ANALYSIS

Two techniques are commonly used for sentiment analysis which is supervised learning technique [15] and unsupervised learning approach [30].

A. Supervised Learning Approach

It is also called machine learning approach. Two data sets are required one for training and another for testing. The training data set is used to train the classifier and test data set is used to measure the accuracy of the classifier. Support vector machine, Naïve Bayes classifier and maximum entropy are the most common algorithm for supervised learning.

Table1: Summary of supervised learning approach

Author	Title of the Paper	Techniques	Dataset Used	Accuracy	Year
A. Go, R. Bhayani, and L. Huang [1]	Twitter sentiment classification using distant supervision	Naïve Bayes, Maximum Entropy and SVM	Twitter data	Above 80%	2009
G. Somprasertsri and P. Lalitrojwong [2]	Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization	Maximum Entropy	Online customer reviews	Precision- 72.65% Recall- 78.77% F score- 75.45%	2010
H. Y. Lee, C. ePulze Sdn Bhd, C. Block, and H. Renganathan [3]	Chinese sentiment analysis using maximum entropy	Maximum Entropy	Product reviews	Above 80%	2011
V. Narayanan, I. Arora, and A. Bhatia [4]	Fast and accurate sentiment classification using an enhanced Naive Bayes model	Naïve Bayes	Movie reviews	88.80%	2013
C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro [5]	Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning	Naïve Bayes	Facebook status	Precision-77% Recall- 68% F score- 72%	2013
L. L. Dhande and G. K. Patnaik [6]	Analyzing sentiment of movie review data using Naive Bayes neural classifier	Naïve Bayes, Neural Network	Movie reviews	80.65%	2014
C. Bhadane, H. Dalal, and H. Doshi [7]	Sentiment Analysis: Measuring Opinions	SVM	Customer reviews	78%	2015
Preety and Sunny Dahiya [8]	Sentiment Analysis using SVM and Naïve Bayes Algorithm	SVM, Naïve Bayes, Modified K- Means	Mobile reviews	89.01%	2015
X. Yan and T. Huang [9]	Tibetan Sentence Sentiment Analysis Based on the Maximum Entropy Model	Maximum Entropy	Tibetan sentiment sentences	F value – 82.8%	2015
L. Dey, S. Chakraborty, A. Biswas, B. Bose,	Sentiment Analysis of Review Datasets Using Naïve Bayes and K-NN	Naïve Bayes, K-NN classifier	Hotel reviews, Movie reviews	NB using movie reviews- 80%	2016

and S. Tiwari [10]	classifier			KNN using movie reviews- 70% NB using hotel reviews- 55% KNN using hotel reviews- 50%	
Cut Fiarni, Herastia Maharani and Rino Pratama [11]	Sentiment analysis system for Indonesia online retail shop review using hierarchy Naive Bayes technique	Naive Bayes	online retail shop review	89.21%	2016
D. Bholane Savita and D. Gore [12]	Sentiment Analysis on Twitter Data Using Support Vector Machine	SVM	Twitter data	97.54%	2016
S. Rana and A. Singh [13]	Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques	SVM, Naive Bayes	Movie reviews	-	2016
R. Hegde and S. Seema [14]	Aspect based feature extraction and sentiment classification of review data sets using Incremental machine learning algorithm	SVM, Naive Bayes	Customer reviews	NB-78.44% SVM- 80.34%	2017

B. Unsupervised Learning Approach

It is also known as semantic orientation approach or lexicon based approach. There is no prior training like machine learning approach. It uses the lexical resource like WordNet, SentiWordNet(SWN) which have a collection of words with predefined polarity and score.

Table2: Summary of unsupervised learning approach

Authors	Title of the Paper	Techniques	Dataset Used	Accuracy	Year
Kerstin Denecke [34]	Using SentiWordNet for multilingual sentiment analysis	Unsupervised approach using SWN	Movie reviews	66%	2008
Shaishav Agrawal and Tanveer j Siddiqui [19]	Using Syntactic and contextual Information for Sentiment Polarity Analysis	Unsupervised approach using SWN	Movie reviews	85%	2009
Bruno Ohana and Brendan Tierney [17]	Sentiment Classification of Reviews Using Sentiwordnet	Unsupervised approach using SWN	Film reviews	SWN (term counting)- 65.85% SWN (scores used as	2009

				features)- 69.35%	
Alaa Hamounda and Mohamed Rohaim [18]	Reviews Classification Using SentiWordNet Lexicon	Unsupervised approach using SWN	Product reviews	Sum on review- 67% Average on sentence on review- 68.63%	2011
Diana C. Cavalcanti, Ricardo B. C. Prudencio, Shreyasee S. Pradhan, Jatin Y. Shah and Ricardo S. Pietrobon [20]	Good to be Bad? Distinguishing between Positive and Negative Citations in Scientific Impact	Unsupervised approach using SWN	citations	-	2011
Orimaye Sylvester Olubolu, Saadat M. Alhashmi abd Siew Eu-gene [27]	Sentiment Analysis amidst Ambiguities in YouTube Comments on Yoruba Lanauage (Nollywood) Movies	Unsupervised approach using SWN	YouTube comments	Precision for pos, neg and neu is 85%, 87% and 79% Recall for pos, neg and neu is 91%, 89% and 90%	2011
Fazal Masud Kundi, Aurangzeb Khan, Shakeel Ahmad and Muhammad Zubair Asghar [28]	Lexicon- Based Sentiment Analysis in the Social Web	Lexicon Based Framework	Twitter data	Binary classification- 92% Multiclass classification- 87%%	2014
Purtata Bhoir and Shipla Kolte [21]	Sentiment analysis of Movie Reviews Using Lexicon Approach	Unsupervised approach using SWN	Movie reviews	NB- 71.42% SWN- 53.33%	2015
Rincy Jose and Varghese S Chooralil [22]	Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Word Sense Disambiguation	Unsupervised approach using SWN	Twitter data	78.6%	2015
Alexandra Cernian, Valentin Sgarciu and Bogdan Martin [23]	Sentiment analysis from product reviews using SentiWordNet as lexical resource	Unsupervised approach using SWN	Product reviews	61%	2015

Shoiab Ahmed and Ajit Danti [24]	A Novel Approach for Sentiment Analysis and Opinion Mining based on SentiwordNet using Web Data	Unsupervised approach using SWN	Web data	Kappa- 77.7%	2015
Vivek Sharma, Apoorv Agarwal, Renu Dhir and Geeta Sikka [25]	Sentiments Mining and Classification of Music Lyrics using SentiWordNet	Unsupervised approach using SWN	Music Lyrics	-	2016
Vivek Sharma, Apoorv Agarwal, Renu Dhir and Geeta Sikka [26]	Opinion Mining of News Headlines using SentiWordNet	Unsupervised approach using SWN	News Headlines	-	2016
Priyanka R. Patil and Pratibha S. Yalagi [33]	Sentiment Analysis of movie reviews using SentiWordNet Approach	Unsupervised approach using SWN	Movie reviews	-	2016

IV. TOOLS AND DATA SOURCES FOR SENTIMENT ANALYSIS

There are no dedicated tools for sentiment analysis. Instead, a variety of tools are available for sentiment analysis [29].

A. Python Natural Language Tool Kit (NLTK)
It offers libraries for classification, tokenization, stemming etc and lexical resources such as wordnet, sentiwordnet.

B. RapidMiner
It is written in java. It holds up all steps for machine learning process like data preparation, results visualization, validation and optimization.

C. R
It has a number of packages for sentiment analysis e.g. TM package and R Sentiment.

D. Gate
It stands for General Architecture for Text Engineering. It has a set of modules tokenizer, gazetter, sentence splitter, part of speech tagger and named entities transducer etc.

E. Weka
It is a collection of machine learning algorithm and visualization tools.

F. Lingpipe
It provides a tool kit for analysis and synthesis of natural languages.

There are different data sources available for sentiment analysis [32].
1) *Blogs:* Expressing views through blogs is a common thing in today's world.
2) *Online Reviews:* Online reviews for a product can be gathered from e- commerce websites like www.amazon.com. Movie review data is available online is MDS (multi domain sentiment) data.
3) *Facebook:* Opinion is present in form of short messages like status on Facebook which can be used for the task of sentiment analysis.
4) *Twitter:* Twitter is a popular micro blogging website. It is broadly used to convey the opinion about any entity among all social networking sites. The maximum limit of character is 140 in twitter.

5) *Youtube:* Millions of video are uploaded on youtube. People give their opinion these video by comments. These comments can be extracted by using API.

6) *Email:* Email is the popular communication medium. The number of email users reached to 2672 million as to 2016. Sentiment analysis can be done of personal email, office email, feedback email and business email etc.

7) *Forum:* Forum is an online discussion website where users can express their thoughts and views.

8) *News:* News is also a good source of opinion.

V.   APPLICATIONS OF SENTIMENT ANALYSIS

Sentiment analysis can be used in diverse fields for various purposes. Some of the common ones are given below.

A. The most common application of sentiment analysis is in the area of reviews of consumer products and services.

B. Facebook and Twitter are a main point of many opinion mining applications. The most common application is observing the reputation of a specific brand on Twitter and/or Facebook.

C. Sentiment analysis can provide substantial value to candidates running for various positions. Sentiment analysis also can help the government in assessing their strength and weakness by analyzing opinions from the public.

D. Classifying a movie review to predict star rating.

E. Analysis of restaurant reviews, predicting a rating for various aspects of the given restaurant such as food and atmosphere.

VI.   CONCLUSION

Two mainly used techniques for sentiment analysis are machine learning approach and lexicon based approach. In lexicon based approach, there is no training but it suffers from word sense disambiguation or domain generalization. In machine learning approach there are two phases training and testing. Naïve Bayes, Support Vector Machine and Maximum Entropy are popular algorithms of machine learning approach. Among these SVM is dominant algorithm. It is highly effective for text categorization because text data comes into linear separable category.

REFERENCES

[1] Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Proj. Rep. Stanf., vol. 1, pp. 12, 2009.

[2] G. Somprasertsri and P. Lalitrojwong, "Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization.," J UCS, vol. 16, no. 6, pp. 938–955, 2010.

[3] H. Y. Lee, C. ePulze Sdn Bhd, C. Block, and H. Renganathan, "Chinese sentiment analysis using maximum entropy," Sentim. Anal. AI Meets Psychol. SAAIP, pp. 89-93, 2011.

[4] V. Narayanan, I. Arora, and A. Bhatia, "Fast and accurate sentiment classification using an enhanced Naive Bayes model," International Conference on Intelligent Data Engineering and Automated Learning, 2013, p. 194–201.

[5] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro, "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning," Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on, 2013, pp. 1–6.

[6] L. L. Dhande and G. K. Patnaik, "Analyzing sentiment of movie review data using Naive Bayes neural classifier," Int. J. Emerg. Trends Technol. Comput. Sci. IJETTCS, vol. 3, no. 4, 2014.

[7] C. Bhadane, H. Dalal, and H. Doshi, "Sentiment Analysis: Measuring Opinions," Procedia Comput. Sci., vol. 45, pp. 808–814, 2015.

[8] Preety and Sunny Dahiya, "Sentiment Analysis Using SVM and Naive Bayes Algorith," International Journal of Computer Science and Mobile Computing, vol. 4, no. 9, 2015.

[9] X. Yan and T. Huang, "Tibetan Sentence Sentiment Analysis Based on the Maximum Entropy Model," 10th International Conference on Broadband and Wireless Computing, Communication and Applications,   2015, pp. 594–597.

[10] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets Using Naive Bayes and K-NN Classifier," ArXiv Prepr. ArXiv161009982, 2016.

[11] Cut Fiarni, Herastia Maharani and Rino Pratama, "Sentiment Analysis for Indonesia Online Retail Shop Review using Hierarchy Naïve Bayes Technique," 2016 4th International Conference on Information and Communication Technology (ICoICT), 2016.

[12] D. Bholane Savita and D. Gore, "Sentiment Analysis on Twitter Data Using Support Vector Machine," International Journal of Computer Science Trends and Technology, vol. 4, 2016.

[13] S. Rana and A. Singh, "Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques," Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on, 2016, pp. 106–111.

[14] R. Hegde and S. Seema, "Aspect based feature extraction and sentiment classification of review data sets using Incremental machine learning algorithm,"Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2017 Third International Conference on, 2017, pp. 122–125.

[15] Shailendra Kumar Singh, Sanchita Paul and Dhananjay Kumar, "Sentiment Analysis Approaches on Different Data set Doamin," International Journal of Database Theory and Application, vol. 7, no.5, pp. 39-50, 2014.

[16] Alessia D'Andrea, Fernando Ferri, Patrizia and Tiziana Guzzo, "Approaches, Tools and Applications for Sentiment Analysis Implementation," International Journal of computer Application, vol. 125, no. 3, pp. 26-33, 2015.

[17] Bruno Ohana and Brendan Tierney, "Sentiment Classification of Reviews Using Sentiwordnet," 9th,IT & T Conference , 2009.

[18] Alaa Hamounda and Mohamed Rohaim, "Reviews Classification Using SentiWordNet Lexicon," The online Journal on Computer Science and Information Technology, vol. 2 no.1, 2011.

[19] Shaishav Agrawal and Tanveer j Siddiqui, "using syntactic and contextual Information for Sentiment Polarity Analysis," 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, 2009.

[20] Diana C. Cavalcanti, Ricardo B. C. Prudencio, Shreyasee S. Pradhan, Jatin Y. Shah and Ricardo S. Pietrobon, "Good to be Bad? Distinguishing between Positive and Negative Citations in Scientific Impact," IEEE 23rd International Conference on Tools with Artificial Intelligence, 2011, pp. 156-162.

[21] Purtata Bhoir and Shipla Kolte, "Sentiment analysis of Movie Reviews Using Lexicon Approach," International Conference on Computational Intelligence and Computing Research, 2015.

[22] Rincy Jose and Varghese S Chooralil, "Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Word Sense Disambiguation," International Conference on Control, Communication & Computing India, 2015.

[23] Alexandra Cernian, Valentin Sgarciu and Bogdan Martin, "Sentiment analysis from product reviews using SentiWordNet as lexical resource," International Conference on Electronics, Computers and Artificial Intelligence, 2015.

[24] Shoiab Ahmed and Ajit Danti, "A Novel Approach for Sentiment Analysis and Opinion Mining based on SentiwordNet using Web Data," International Conference on Trends in Automation, Communications and Computing Technology, 2015.

[25] Vivek Sharma, Apoorv Agarwal, Renu Dhir and Geeta Sikka,"Sentiments Mining and Classification of Music Lyrics using SentiWordNet," Colossal Data Analysis and Networking , 2016.

[26] Vivek Sharma, Apoorv Agarwal, Renu Dhir and Geeta Sikka, "Opinion Mining of News Headlines using SentiWordNet," Colossal Data Analysis and Networking , 2016.

[27] Orimaye Sylvester Olubolu, Saadat M. Alhashmi abd Siew Eu-gene, "Sentiment Analysis amidst Ambiguities in YouTube Comments on Yoruba Lanauage (Nollywood) Movies," 21st International Conference on World Wide Web, 2012.

[28] Fazal Masud Kundi, Aurangzeb Khan, Shakeel Ahmad and Muhammad Zubair Asghar, "Lexicon- Based Sentiment Analysis in the Social Web," Journal of Basic and Applied Scientific Research, vol. 4, pp. 238-248, 2014.

[29] Suad Alhojely, "Sentiment Analysis and Opinion Mining: A Survey," International Journal of Computer Applications, vol. 150, no. 6, 2016.

[30] Amrita Kaur and Neelam Duhan, "A Survey on Sentiment Analysis and Opinion Mining," International Journal of Innovations & Advancement in Computer Science, vol. 4, 2015.

[31] Neha Raghuvanshi and Prof. J.M. Patil, "A Brief Review on Sentiment Analysis," International Conference on Electrical, Electronics and Optimization Techniques, 2016.

[32] Jasleen Kaur and Dr. Jatinderkumar R. Saini, "On Classifying Sentiments and Mining Opinions," International Journal of Emerging Trends & Technology in Computer Science, vol. 9, 2014.

[33] Priyanka R. Patil and Pratibha S. Yalagi, "Sentiment Analysis of movie reviews using SentiWordNet Approach," International Journal of Emerging Trends & Technology in Computer Science, vol. 38, 2016.

[34] Kerstin Denecke, "Using SentiWordNet for multilingual sentiment analysis,"IEEE 24th International Conference on Data Engineering Workshop, 2008.

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)