Parallel comparison of sentiment analysis techniques A survey

Dr.Shaukat Wasi
Computer Science Department
Mohammad Ali Jinnah University
Karachi, Pakistan

Sarah Yousuf
Computer Science Department
Mohammad Ali Jinnah University
Karachi, Pakistan

Abstract— *An idea or statement about sentiment analysis is not new, it does not suddenly occur. Sentiment analysis has a great importance in electronic world to capture positive or negative opinion regarding an entity. People express their sentiments or giving their opinion about any product on social media like Facebook, Twitter, Instagram etc so, It is very helpful in decision making to buy any product. There are several techniques have been applied to analyze sentiments. The objective of this survey paper is to highlight different techniques through which sentiment analysis can be performed. It provides the comparative review of lexicon based techniques, machine learning techniques and deep learning techniques that have been used to analyze sentiments by using different types of datasets.*
Keywords— (Sentiment analysis, Lexicon based, Machine Learning, Deep learning)

I. Introduction

We are the part of the electronic world which has a different amount of information in different formats like text, images, videos etc and these categories of information growing tremendously from that information we need to analyze sentiments. There are so many platforms like Facebook, Twitter, Instagram and many others where people express their opinion regarding any product or policy, which is very important for business people. Sentiment analysis is the process to identify people's opinion or behavior towards an entity. There are so many techniques which have been used for sentiment analysis.

The basic procedure of sentiment analysis is often divided into two phases Opinion Extraction and Sentiment categorization. Opinion Extraction yield to fetch opinion words from the target text, whereas Sentiment categorization classifies and ranks the opinionated text phrases based on polarity placement.

By far there is so many research has done to identify the user's opinion. In these research sentiment analysis mainly conducted in three levels document level, sentence level, and attribute level. By the side of other social media, marketers need to observe the information to their products. Techniques to analyze sentiment has been implemented in most of the applications and these applications are widely used. Somewhere these applications need to be improved. Our work is based on a survey which is categorized in Lexicon based, Machine learning and Deep learning techniques to analyze sentiments. Rest of the paper is organized in this way. In section II we have related work, Section III comparative review of lexicon based sentiment analysis techniques with table I, Section IV comparative review of machine learning based sentiment analysis techniques with table II and section V comparative review of deep learning based sentiment analysis techniques with table III respectively.

II. Related Work

Surbhi Bhatia, Manisha Sharma and Komal Kumar Bhatia [1] presents a survey and focus the current methods available in the field of opinion mining. They also discuss a process to analyze the opinions through a useful and efficient system.

Surya Prakash Sharma, Dr. Rajdev Tiwari and Dr. Rajesh Prasad [2] In this paper they researched on Opinion mining with respect to their different levels, architecture, techniques applied, tools used, comparative study of techniques and challenges.

Saqib Iqbal, Ali Zulqurnain, Yaqoob Wani and Khalid Hussain [3] presents a comprehensive analysis of the methods which are used on user behavior prediction. This comparison will give a detailed information, advantages, and disadvantages in the domain of sentiment and opinion mining.

Abhishek Kaushik, Anchal Kaushik and Sudhanshu Naithani [4] presents a detailed discussion of Sentiment analysis methods and tools used.

Dr.B.Radha and V.Meera [5] presents an overview of opinion mining and sentiment analysis tools and techniques. They have discussed that sentiment classifiers are extremely dependent on fields or topics. They also point that different kinds of features and classification algorithms are united in an organized way in order to overcome their singular disadvantage and benefit from each other's merits, and at least improve the sentiment classification performance.

Vishal A. Kharde and S.S. Sonawane[6] presents a survey on sentiment analysis of twitter with techniques. They give a survey and a comparative analysis of existing techniques for opinion mining like machine learning and lexicon-based approaches, together with evaluation metrics. Using various machine learning algorithms like Naive Bayes, Max Entropy, and Support Vector Machine, we provide research on twitter data streams. We have also discussed general challenges and applications of Sentiment Analysis on Twitter.

Kaushik Hande, Prof. A. G. Phakatkar [7] presented a paper which tackles an overview in this area and gives a survey which covers Opining Mining, Sentiment Analysis, techniques, tools, and classification.

Naïve Bayes and Support Vector Machines are the most frequently used machine learning algorithms for solving sentiment classification problem.

Zalak M. Patel and Vishal P. Patel [8] discussed to retrieve the sentiment from a microblogging service Twitter. In this they have faced a problem to find out meaningful tweets that include positive and negative emotions of users. They focused on to extract positive and negative emotions of users from social media. This paper presents the process for opinion retrieval and classification techniques.

Qurat Tul Ain, Mubashir Ali, Amna Riaz, Amna Noureen, Muhammad Kamran, Babar Hayat and A. Rehman [9] focused on latest studies regarding the implementation of deep learning models such as deep neural networks, convolutional neural networks and many others for solving different issues of sentiment analysis such as sentiment classification, cross lingual problems, textual and visual analysis and product review analysis, etc.

Upma Kumari, Dinesh Soni and Dr. Arvind K Sharma [10] presents a comprehensive study of various techniques and tools which have been used in the sentiment analysis process. They proposed a methodology which provides important phases the sentiment of text, whether it is positive or negative.

III. DETAIL REVIEW OF PAPERS

This section has contain comparison of sentiment analysis techniques as described below.

A. Comparative review of lexicon based sentiment analysis techniques

Eman M.G. Younis proposed a lexicon based approach by using twitter microblogs data to perform text mining and sentiment analysis for evaluating online reviews about two giant retail stores in the UK[11].

Muhammad Zubair Asghar, Aurangzeb Khan, Shakeel Ahmad, Maria Qasim, and Imran Ali Khan presented an approach by using three user's reviews datasets, namely drug, car and hotel to improve the performance of sentiment analysis and resolve the issues of data scattered and incorrect classification due to use of noisy text, emoticons, modifiers and domain specific words[12].

Kajal Sarawgi and Vandana Pathak determined the aspect terms present in the individual sentence, searching out their polarities, detect the polarity of sentences and the polarity of individual aspect category by applying lexicon based approach by using reviews of a product like a mobile from Amazon[13].

G. Vaitheeswaran and L. Arocki described a Senti-Lexi based approach is to determine the sentiment knowledge on tweets using lexicon based approach. To improve the accuracy the emoticons are used for polarity calculations by using dataset collection of tweets[14].

Aymun Siddiqua, Tanveer Ahsan and Abu Nowshed Chy proposed an approach by applying rule-based classifier for sentiment analysis by using Twitter data[15].

TABLE I.   TABULAR VIEW OF LEXICON BASED SENTIMENT ANALYSIS TECHNIQUES

S.No	Name and Year	Technique	Purpose	Training Dataset	Stop words Removal	Feature Extraction
1	Eman M.G. Younis (2015) [11]	Lexicon based Approach	Mining Twitter Microblogs	No	Yes	No
2	Muhammad Zubair Asghar, Aurangzeb Khan, Shakeel Ahmad, Maria Qasim, and Imran Ali Khan (2015) [12]	Rule based classifier	Analyzing and classifying user reviews	Yes	Yes	No
3	Kajal Sarawgi and Vandana Pathak (2017) [13]	Lexicon based	Aspect Level Sentiment Analysis	No	Yes	No
4	G. Vaitheeswaran and L. Arockiam (2016)[14]	Lexicon based approach	To enhance the accuracy of sentiment analysis on tweets	No	Yes	Yes
5	Umme Aymun Siddiqua, Tanveer Ahsan and Abu Nowshed Chy (2016) [15]	Rule-Based Classifier, multinomial naivebayes classifier, support vector machine (SVM) classifier, and sequential minimal optimization (SMO) classifier	Combining rule based classifier with ensemble of feature sets and machine learning techniques	Yes	Yes	Yes

B.   *Comparative review of machine learning based sentiment analysis technique*

Dongzhi Wang, Xinwei Yan, Huimin Wang and Xiu Li proposed a novel conceptual platform which is designed to use the potentials of opinion mining in e-commerce area by applying machine learning algorithm and NLP techniques. Opinion mining is done by using dataset E-commerce store reviews[16].

Peiman Barnaghi, John G. Breslin, and Parsa Ghaffari provided a positive or negative sentiment on twitter post using a machine learning method: Bayesian Logistic Regression and Naïve Bayes for text categorization[17].

BiswaRanjanSamal, Anil Kumar Behera, and Mrutyunjaya Panda described an approach for processing the movie reviews. They have used supervised machine learning techniques used to capture labeled data, which already have grouped into the available classes[18].

Chao Wu, Chunlin Li1, Wei Yan, Youlong Luo, Xijun Mao, Shumeng Du and Mingming Li used a machine learning approach to identify opinion leader in the internet forum, they have used page rank algorithm on posts of an internet forum. They compared with Interest-based PageRank algorithm, online time Algorithm, and Experience-based Algorithm, the result explained that the opinion leader page rank algorithm can identify opinion leaders than others in the Internet forum effectively[19].

Pranoti P. Jagtap proposed an approach for identification of such features from unstructured textual reviews. They proposed two different approaches for the identification of opinion features from unstructured textual online in a domain independent and domain dependent corpus. They have used Apriori algorithm for mining the frequent item set[20].

TABLE II.   TABULAR VIEW OF MACHINE LEARNING BASED SENTIMENT ANALYSIS TECHNIQUES

S.No	Name and Year	Technique	Purpose	Training Dataset	Stop words Removal	Feature Extraction
1.	Dongzhi Wang,Xinwei Yan,Huimin Wang and Xiu Li (2015)[16]	Machine Learning and NLP	Sentiment analysis system customers' review in commerce malls.	No	Yes	Yes
2.	Peiman Barnaghi, John G. Breslin and Parsa Ghaffari (2016) [17]	Machine Learning algorithm, Bayesian Logistic Regression and Naïve Bayes	Text categorization of twitter posts.	Yes	Yes	Yes
3.	BiswaRanjanSamal, Anil Kumar Behera and Mrutyunjaya Panda (2017) [18]	Supervised Machine learning algorithms	Categorization of movie reviews	Yes	Yes	Yes
4.	Chao Wu, Chunlin Li1, Wei Yan, Youlong Luo, Xijun Mao, Shumeng Du and Mingming Li (2015) [19]	Machine learning Page Rank algorithm	Identifying opinion leader on internet forum	No	Yes	No
5.	Pranoti P. Jagtap (2016) [20]	Apriori Algorithm	An automatic model which extracts the product features from the large review corpus and give the customer with good product purchase feature.	Yes	Yes	Yes

C.   Comparative review of deep learning based sentiment analysis technique

Shiyang Liaoa, Junbo Wang, Ruiyun Yua, Koichi Sato and Zixue Chengb performed sentiment analysis of twitter database on deep learning techniques to understand the situation in the real world[21].

Huy Nguyen and Minh-Le Nguyen introduced a deep learning framework including a lexicon-based approach to sentence level prediction of sentiment label distribution on three twitter sentiment classification datasets. Their experimental result described that their model can enhance the classification accuracy of sentence-level sentiment analysis in Twitter social networking [22].

Andreea Salina presented an approach to classifying business reviews using word embeddings on the large-scale dataset. They conducted several experiments to express the semantic relationship between business reviews and they used deep learning techniques that prove that the obtained results are competitive with traditional methods[23].

Jan Deriu, Aurelien Lucchi, Valeria De Luca, Aliaksei Severyn, Simon Müller, Mark Cieliebak, Thomas Hofmann, and Martin Jaggi presented an approach for multi-lingual sentiment classification in short texts. They leveraged massive amounts of weakly supervised data in different languages to train a multi-layer convolutional network and demonstrate the importance of using pre-training of such networks[24].

Alexis Conneau, Holger Schwenk, Yann Le Cun and Lo¨ıc Barrault proposed a new architecture very deep convolutional network for text processing which operates directly at the character level and uses only small convolutions and pooling operations. They showed that the performance of this model increases with the depth using up to 29 convolutional layers, they reported enhancements over the state-of-the-art on different public text classification tasks[25].

TABLE III.   TABULAR VIEW OF MACHINE LEARNING BASED SENTIMENT ANALYSIS TECHNIQUES

S.No	Name and Year	Technique	Purpose	Training Dataset	Stop words Removal	Feature Extraction
1.	Shiyang Liaoa, Junbo Wang, Ruiyun Yua , Koichi Satob and Zixue Chengb (2016)[21]	CNN	Sentiment analysis of product	Yes	No	No
2.	Huy Nguyen and Minh-Le Nguyen (2017) [22]	CNN	Sentence level sentiment classifications	Yes	Yes	Yes
3.	Andreea Salinca (2017) [23]	CNN	Classifying business reviews	Yes	No	No
4.	Jan Deriu , Aurelien Lucchi , Valeria De Luca , Aliaksei Severyn , Simon Müller , Mark Cieliebak , Thomas Hofmann, and Martin Jaggi (2017) [24]	CNN	Multi-language sentiment classification in short texts	Yes	Yes	Yes
5.	Alexis Conneau , Holger Schwenk , Yann Le Cun and Lo¨ıc Barrault (2017) [25]	Very Deep convolutional	Text processing at character level	Yes	NO	NO

Conclusion

Sentiment analysis is an apparent approach to fetch data from customer reviews and feedback on any object. A massive amount of work has done in opinion mining to mine opinions in the form of document, sentence or feature, but still, a lot of work still remains to be done. By applying sentiment analysis techniques it can be identified that whether the opinion is contrary or in favor of an entity. There are many other techniques that have not discussed here. This paper is based on a comparative review of sentiment analysis by using lexicon based, machine learning and deep learning techniques.

REFERENCES

[1] Surbhi Bhatia, Manisha Sharma and Komal Kumar Bhatia, "Strategies for Mining Opinions: A Survey".2nd International Conference on Computing for Sustainable Global Development (INDIACom), 2015.

[2] Surya Prakash Sharma, Dr Rajdev Tiwari and Dr Rajesh Prasad, "Opinion Mining and Sentiment Analysis on Customer Review Documents- A Survey", International Conference on Advances in Computational Techniques and Research Practices, Vol. 6, Special Issue 2, February 2017.

[3] Saqib Iqbal, Ali Zulqurnain, Yaqoob Wani and Khalid Hussain , "THE SURVEY OF SENTIMENT AND OPINION MINING FOR BEHAVIOR ANALYSIS OF SOCIAL MEDIA", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.6, No.5, October 2015.

[4] Abhishek Kaushik, Anchal Kaushik and Sudhanshu Naithani , "A Study on Sentiment Analysis: Methods and Tools", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.

[5] Dr.B.Radha and V.Meera ,"A Survey on Research Issues in Opinion Mining", International Journal of Innovative Research in Computer

and Communication Engineering, Vol. 5, Issue 2, February 2017.

[6] Dongzhi Wang,Xinwei Yan,Huimin Wang and Xiu Li ,"Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications 139(11): 5-15, April 2016.

[7] Kaushik Hande , Prof. A. G. Phakatkar ,"Sentiment Analysis using Machine Learning Algorithms: A Survey", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 5, May 2017.

[8] Zalak M. patel and Vishal P. Patel, "A Survey Various Techniques of Sentiment Analysis in Data Mining", International Journal of Engineering Development and Research, Volume 3, Issue 4 2015.

[9]   Qurat Tul Ain, Mubashir Ali, Amna Riaz, Amna Noureen, Muhammad Kamran, Babar Hayat and A. Rehman , "Sentiment Analysis Using Deep Learning Techniques: A Review", International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, 2017.

[10]   Upma Kumari, Dinesh Soni and Dr. Arvind K Sharma, "A cognitive study of Sentiment Analysis Techniques and Tools: A survey", International Journal of Computer Science and Technology, IJCST vol 8, issue 1, Jan-March 2017.

[11]   Eman M.G. Younis , "Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study", International Journal of Computer Applications · February 2015.

[12]   Muhammad Zubair Asghar, Aurangzeb Khan, Shakeel Ahmad, Maria Qasim, and Imran Ali Khan , "Lexicon-enhanced sentiment analysis framework using rule-based classification scheme", PLoS ONE 12(2): e0171649. https://doi.org/10.1371/journal.pone.0171649.

[13]   Kajal Sarawgi and Vandana Pathak, "Opinion Mining: Aspect Level Sentiment Analysis using SentiWordNet and Amazon Web Services" International Journal of Computer Applications (0975 – 8887) Volume 158 – No 6, January 2017.

[14]   G. Vaitheeswaran and L. Arocki , "Lexicon Based Approach to Enhance the Accuracy of Sentiment Analysis on Tweets", International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555, Vol.6, No3, May-June 2016

[15]   Umme Aymun Siddiqua, Tanveer Ahsan and Abu Nowshed Chy, "Combining a Rule-based Classifier with Ensemble of Feature Sets and Machine Learning Techniques for Sentiment Analysis on Microblog", International Conference on Computer and Information Technology, December 18-20, 2016.

[16]   Dongzhi Wang,Xinwei Yan,Huimin Wang and Xiu Li, "A Conceptual Framework of E-Commerce Supervision System Based on Opinion Mining", International Conference on Services Science

[17]   Peiman Barnaghi, John G. Breslin and Parsa Ghaffari ,"Opinion Mining and Sentiment Polarity on Twitter and Correlation between Events and Sentiment", 2016 IEEE Second International Conference on Big Data Computing Service and Applications.

[18]   BiswaRanjanSamal, Anil Kumar Behera and Mrutyunjaya Panda, "Performance Analysis of Supervised Machine Learning Techniques for Sentiment Analysis", 2017 IEEE 3rd International Conference on Sensing, Signal Processing and Security.

[19]   Chao Wu, Chunlin Li1, Wei Yan, Youlong Luo, Xijun Mao, Shumeng Du and Mingming, "Identifying Opinion Leader in the Internet Forum", International Journal of Hybrid Information Technology Vol.8, No.11 (2015)

[20]   Pranoti P. Jagtap, "Domain Relevance Opinion Mining", International Journal of Engineering Science and Computing, 2016.

[21]   Shiyang Liaoa, Junbo Wang, Ruiyun Yua , Koichi Satob and Zixue Chengb, "CNN for situations understanding based on sentiment analysis of twitter data", International Conference on Advances in Information Technology, IAIT2016, 19-22 December 2016.

[22]   Huy Nguyen and Minh-Le Nguyen, "A Deep Neural Architecture for Sentence-level Sentiment Classification in Twitter Social Networking", Published2017 in ArXiv

[23]   Andreea Salinca, "Convolutional Neural Networks for Sentiment Classification on Business Reviews", Proceedings of IJCAI Workshop on Semantic Machine Learning (SML 2017): 35-39.

[24]   Jan Deriu , Aurelien Lucchi , Valeria De Luca , Aliaksei Severyn , Simon Müller , Mark Cieliebak , Thomas Hofmann, and Martin Jaggi , "Leveraging Large Amounts of Weakly Supervised Data for Multi-Language Sentiment Classification", 2017 - 26th International World Wide Web Conference.

[25]   Alexis Conneau , Holger Schwenk , Yann Le Cun and Lo¨ıc Barrault , "Very Deep Convolutional Networks for Text Classification", 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, April 3-7, 2017.