Aspect Level Sentiment Analysis using Machine Learning Approach: A Comprehensive Review

Jyoti Hanvat
M.Tech Scholar
Department of CSE
Vaishnavi Institute of Technology and Science
Bhopal (M.P), India
jyotihanvat99@gmail.com

Sumit Sharma
Professor
Department of CSE
Vaishnavi Institute of Technology and Science
Bhopal (M.P), India
sumit_sharma782022@yahoo.co.in

***Abstract*— Sentimental analysis is now used from product marketing specific to the detection of social behavior. Progress on Facebook, Twitter, Youtube and other microblogging and social networking sites has not only contributed to a change in social sites, but also to the way we use these sites and the way we do it. People are fundamentally changing their feelings and their points of view with the general public. In this paper a detailed study of different approaches for lexicon-based sentiment analysis are discussed. This paper also shows that efficiency of machine learning over traditional lexicon based sentiment analysis.**

Keywords— Lexicon-based approach, Sentiment analysis, Natural Language Processing, Machine Learning.

I. INTRODUCTION

As it is known that internet is one of the most common platforms for communication among opinions of an individuals. This platform explores different fields to express anyone's review, opinions or emotions. One of the most common medium to express emotions/opinions is through reviews. Either it is in the field of movie, news, products, tweets, etc. In each and every field through review anyone can express their sentiments. As most of the people are active online for their most of the time and express their sentiments. So, day by day database of these reviews are increasing and becoming heavy. As per recent study, there are approx. 2.4 billion active online users, who write and read online around the world [1].

Although the scientific domain is as large that there are over 4,000 classified conferences and 5,000 classified journals [2]. In particular, a large fragment of researchers publish their content so that researchers, companies, universities and businesses can use and analyze the data. As a result, a large number of studies and research have followed the trend of increasing online search resources year after year.

Emotions are an important aspect of the interaction and communication between individuals. Exchanging emotions through text messages and personal blog posts presents research with an informal writing challenge.

The extraction of emotions from the text is used to determine the human-machine interaction that governs communication and many others [3]-[6]. Emotions are also expressed through a person's speech, face, and textual emotions. Emotions are also expressed through a word or a series of words. The sentence-level emotion sensing technique plays an important role in tracking emotions or finding clues to generate those emotions. Phrases are the essential units of information in a document. For this reason, the emotion recognition technique at the document level depends on the feeling expressed by the individual sentences of that document, which in turn depends on the emotions expressed by the individual words. Feelings of different texts and classified accordingly in positive, negative or neutral classes.

Sentiment analysis is an innovation that will be very important in the years to come. With opinion mining, you get high quality content. An important part of the early research in this area was based on product evaluations [7] and defined feelings as positive, negative or neutral. Most opinion polls currently focus on online network sources, such as IMDB, Twitter and Facebook, and require that the methods be adapted to meet the growing interest of the public. Furthermore, the analysis of phrases classification at sentence level proves to be a difficult task [8].

II. SENTIMENT ANALYSIS

Sentiment Analysis is considered a classification method. There are 3 main classification levels in SA:

Sentiment analysis at the document level considers the entire document as a data unit for determination of

 DOI: https://doi.org/10.24113/ijosthe.v7i4.134

sentiments. The opinion related to whole document is analyzed.

In sentiment analysis at sentence level, emotions are classifying at sentence level. In such analysis, each and every sentence is considered to express polarity of the sentence towards being positive or negative.

In aspect level-based sentiment analysis, each entity in entire sentence or document is analyzed and polarity of each entity is determined. The polarity of each entity differentiates from sentence to sentence.

III. SENTIMENT ANALYSIS (SA) AND NATURAL LANGUAGE PROCESSING (NLP)

It is an area of the computer science, artificial intelligence, and computational linguistics interested in the interactions between computers and human (natural) languages. Intrinsically, NLP is related to the field of human–computer interaction (HCI). Many challenges in NLP include natural language understanding, that is, enabling computers to deduce meaning from human or natural language input, and others inVolve natural language generation. The term of Natural Language Processing inVolves a wide set of techniques for automated generation, manipulation and analysis of natural or human languages. Despite most NLP techniques inherit largely from Linguistics and Artificial Intelligence, they are also affected by relatively newer domains such as Machine Learning, Computational statistics and Cognitive Science [9]-[10].

- Token: Before any real processing can be done on the input text, it needs to be segmented into linguistic units such as words, punctuation, and numbers or alphanumeric. These units are recognized as tokens.
- Sentence: This refers to an ordered sequence of tokens.
- Tokenization: The operation of splitting a sentence into its constitutive tokens.
- Corpus: This means a body of text, usually including a large number of sentences.
- Part-of-speech (POS) Tag: A word can be categorized into one or more of a set of lexical or part-of-speech classes such as Nouns, Verbs, Adjectives and Articles, to name a few. A POS tag is a symbol representing such a lexical category – NN (Noun), VB (Verb), JJ (Adjective), AT (Article). One of the oldest and most commonly used tag sets is the Brown Corpus tag set.
- Parse Tree: It represents a tree defined over a given sentence that interprets the syntactic structure of the sentence as identified with a formal grammar.
- Part-Of-Speech (POS) Tagging: Given a sentence and a set of POS tags, a mutual language processing task is to automatically specify POS tags to each word in the sentences. For example, given the sentence "The ball is red", the output of a POS tagger would be the /AT ball/NN is/VB red/JJ. State-of-the-art POS taggers can reach to higher accuracy as 96%. Tagging text with part-of-speech turns out to be much beneficial for more complicated NLP tasks such as parsing and machine translation.
- Computational Morphology: Natural languages are made up of a very large number of words built on basic building blocks called morphemes (or roots), with the smallest linguistic units having meaning. Computer morphology interested in discovering and analyzing the internal structure of words using computers.
- Parsing: In the parsing task, a parser builds the parse tree given a sentence. There are some parsers assume the existence of a set of grammar rules to parse but recent parsers are smart enough to deduce the parse trees directly from the given data using complex statistical models. Most parsers also operate in a supervised setting and require the sentence to be POS-tagged before it can be parsed. Statistical parsing is an area of active research in NLP.
- Subjective Sentence: It is a sentence in which the writer expresses his or her feelings or sentiments toward entities, events and their properties. For example: “I like swimming”.
- Objective Sentence: It is a factual sentence about entities, events, and their properties.
- Opinion: It is a belief or judgment based on special knowledge towards a topic. Opinions are sometimes expressed explicitly like: “The voice quality of this phone is amazing.” But sometimes they are hidden in the sentiment of a sentence, for instance; “The earphone broke in two days”. Since the concept of opinion is very wide, sentiment classification mostly concentrates on the general feeling expressed by opinions (Positive / Negative). In fact, positivity or negativity is determining the Polarity of an opinion. In other words, one of the main subtasks of sentiment analysis is determining the polarity of documents or in more details, determining the polarity of each subjective sentence in a document.
- Opinion words: They are words that are commonly used to express positive or negative sentiment. For example: {Beautiful, pretty, love} is Positive sentiment {Ugly, awful, hate} is Negative sentiment.
- Sentiment Orientation-SO (Polarity): It indicates whether the expressed opinion by opinion words is positive, negative or neutral. For example: "The camera takes wonderful pictures" is Positive.
- Opinion Sentence: It is a sentence which contains one or more opinion words. For example: "The story was amazing as was the playing of actors".

- Object / Features: So far, we have used "topic" to refer the main subject in reviews which is going to be discussed. Hence, we call it "Object". In opinionated documents, objects and their components or attributes are going to be reviewed and sentiments toward them are expressed in terms of "opinion words"; these components or attributes are called: "object-features".

IV. LITERATURE REVIEW

Yang et al. [1] proposed a new mood analysis model-SLCABG, which is based on the mood dictionary and combines the conVolutional neural network (CNN) and the attention-based bidirectional recurrent unit (BiGRU). In terms of methodology, the SLCABG model combines the benefits of the sentiment lexicon and deep learning technology and overcomes the shortcomings of the existing sentiment analysis model for product reviews.

Xu et al. [2] proposed method for sentiment analysis for Big Data. This method integrates the subject's semantic information into text visualization via a neural network model. The attention mechanism is introduced into the neural network and a context-sensitive vector is introduced to calculate the weight of each word. Also, to make the model more adaptable, the mood dictionary markup method is used to get the training data.

Meyyappan et al. [3] recommend an innovative method of common sense-based sentiment analysis (domain specific ontology) for ConceptNet-based tourism ontology in Oman.

Xu et al. [4] proposed an improved word representation method that integrates the contribution of sentiment information into the traditional TF-IDF algorithm and generates weighted word vectors. The weighted word vectors are inserted into short-term long-term bidirectional memory (BiLSTM) to efficiently acquire context information, and comment vectors are better represented. The mood trend of the comment is obtained from an anticipatory classifier for neural networks. Under the same conditions, the sentiment analysis proposed is compared with the sentiment analysis methods of RNN, CNN, LSTM and NB. Experimental results show that the proposed sentiment analysis method has higher accuracy, better recall and a higher F1 score. The method proved effective on very specific comments.

Iqbal et al. [5] proposed an integrated framework that bridges the gap between vocabulary-based approaches and machine learning to achieve better accuracy and scalability. To solve the scalability problem that arises with increasing feature sets, a new feature reduction technique based on genetic algorithms (GA) is proposed.

Wongkar et al. [6] designed a framework for twitter data analysis which was conducted on 2019 Republic of Indonesia presidential candidates. The Naïve Bayes method is used to help classify classes or the level of sentiments of society and achieved accuracy value of 75.58%.

Table 1 gives the comparative study of various methods for sentiment analysis.

Author	Description	Results	Drawbacks
Yang et al. (2020)	ConVolutional Neural Network (CNN) and attention-based Bidirectional Gated Recurrent Unit (BiGRU) based product review sentiment analysis.	~93% accuracy	Performance decreases for non-weighted words.
Xu et al. (2019)	Context-aware vector is introduced to calculate the weight of each word and neural network for classification.	~91% precision	Binary classification and sentiment data dictionary is not considered.
Meyyappan et al. (2019)	Domain specific ontology combined feature extraction, lexicon-based approach and conceptual semantic sentiment analysis to determine the sentiment analysis of tweets about Oman tourism.	~83% Precision	Lexicons are not well focused for opinion mining
Xu et al. (2019)	weighted word vectors with BiLSTM for sentiment analysis	~91% precision	BiLSTM consumes a long time in the training model

Iqbal et al. (2019)	lexicon-based and machine learning with genetic algorithm optimization for domains such as terrorism, global conflicts, and social issues.	~90% Accuracy	Other performance parameters are not focused. Limited to criminal sentiment domain.
Wongkar et al. (2019)	Twitter data sentiment analysis using Naïve Bayes.	~76% accuracy	Word frequency is calculated in sentence.
Shen et al. (2018)	Deep Neural Network model with Feature Adaptive Transformation & Combination strategy.	Accuracy=77%	Performance rate is quite low.

VII. CONCLUSION

As sentimental analysis has improved over the past decade, their applications have also improved. Every day, millions of messages appear on popular websites offering microblogging services. The authors of these messages write about their lives, share opinions on various topics, and discuss current topics. Due to the free message format and ease of access to microblogging platforms, Internet users tend to shift from traditional communication tools (such as traditional blogs or mailing lists) to micro-blogging services. blogging. As more and more users post articles about products and services they use or express their political and religious beliefs, microblogging sites become valuable sources of opinions and feelings. After the detailed study and analysis of natural language processing it has been reported that there are multiple domains in which sentiment analysis is needed to be analyzed. There is a lack of modern sentiment analysis tools which provide insights into the multiple domains. There is a need of smart systems to analyze and give accurate and timely prediction of emotions of a person.

REFERENCES

[1] L. Yang, Y. Li, J. Wang and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," in IEEE Access, Vol. 8, pp. 23522-23530, 2020.

[2] G. Xu, Z. Yu, Z. Chen, X. Qiu and H. Yao, "Sensitive Information Topics-Based Sentiment Analysis Method for Big Data," in IEEE Access, Vol. 7, pp. 96177-96190, 2019.

[3] V. Ramanathan and T. Meyyappan, "Twitter Text Mining for Sentiment Analysis on People's Feedback about Oman Tourism," 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), Muscat, Oman, 2019, pp. 1-5.

[4] G. Xu, Y. Meng, X. Qiu, Z. Yu and X. Wu, "Sentiment Analysis of Comment Texts Based on BiLSTM," in IEEE Access, Vol. 7, pp. 51522-51532, 2019.

[5] F. Iqbal et al., "A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction," in IEEE Access, Vol. 7, pp. 14637-14652, 2019.

[6] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), Semarang, Indonesia, 2019, pp. 1-5.

[7] X. Hu, J. Tang, H. Gao, and H. Liu, ''Unsupervised sentiment analysis with emotional signals,'' in Proc. 22nd Int. Conf. World Wide Web, 2013, pp. 607–618.

[8] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, ''Sentiment analysis of Twitter data,'' in Proc. Workshop Lang. Social Media, 2011, pp. 30–38.

[9] M. Pontiki et al., ''SemEval-2016 task 5: Aspect based sentiment analysis,'' in Proc. 8th Int. Workshop Semantic Eval. (SemEval), 2014, pp. 27–35.

[10] P. C. S. Njølstad, L. S. Høysæter, W. Wei, and J. A. Gulla, ''Evaluating feature sets and classifiers for sentiment analysis of financial news,'' in Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. (WI) Intell. Agent Technol. (IAT), Vol. 2, Aug. 2014, pp. 71–78

[11] S. Shikhar, et al. "LEXER: LEXicon Based Emotion AnalyzeR." International Conference on Pattern Recognition and Machine Intelligence. Springer, Cham, 2017.

[12] Low, Lu-Shih Alex, et al. "Content based clinical depression detection in adolescents." Signal Processing Conference, 2009 17th European. IEEE, 2009.

[13] Wang, Xinyu, et al. "A depression detection model based on sentiment analysis in micro-blog social network." Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Berlin, Heidelberg, 2013.

[14] Wang, Xinyu, Chunhong Zhang, and Li Sun. "An improved model for depression detection in micro-blog social network." 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW). IEEE, 2013.

[15] Shen, Tiancheng, "Cross Domain Depression Detection via Harvesting Social Media." Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. 2018.