

A REVIEW ON SENTIMENT AND EMOTION ANALYSIS FROM SOCIAL MEDIA TEXT

Harihar Nath Verma^{1*}, Sanjay Kumar Malik²

^{1*}Research Scholar, USIC&T, Guru Gobind Singh Indraprastha University, Dwarka, Delhi, India,

hnvermagwl@gmail.com

²Professor, USIC&T, Guru Gobind Singh Indraprastha University, Dwarka, Delhi, India,

skmalik@ipu.ac.in

ABSTRACT: *Sentiment analysis is an area in the field of Natural Language Processing (NLP), where text, audio or video data is analyzed for recognition of a person's interest and likeliness. Its basic purpose is to obtain the inclination of people towards a lesson, product, a politician, a service. It also helps in identification of positive, negative or neutral feeling of persons. This can be achieved using a dataset and some sentiment and emotion analysis tool. The dataset itself can be of varying nature, like text-based, audio, video or a combination of multiple nature. Sentiment analysis can be used as customer feedback for a product or service to identify the quality of a product of service. In this paper a survey on various approaches available in the literature for sentiment and emotion analysis has been provided. Such surveys are helpful for businessmen to enhance the quality of their product or service, and hence their business. Nowadays sentiment analysis is widely used in subjective and unstructured dataset also. This paper reviews the approaches for sentiment and emotion analysis over the textual data from social media platforms.*

KEYWORDS: *Sentiment Analysis, Emotion Analysis, Natural Language Processing, Textual Data Mining, Social-Media, Web Mining.*

1. INTRODUCTION

Sentiment is a positive, negative or neutral feeling towards an opinion, a statement, an audio/video or text. Most of the time we perform the sentiment analysis over the text-oriented data but there are approaches which goes beyond text based and include audio, images and video as well, called multimodal approaches. Sentiment analysis also includes the domain of emotion analysis. It is tough to have a proper definition of emotion. However, it is well established and majority of researchers agree that sentiment and emotions include a mixture of behavioral, feeling, expressive, physiological, sensation and phenomenological attributes [1]. An emotion can be defined as one's experiences that he/she feels from a situation or the appraisal of the self. In some extent, emotions and sentiments can be analyzed to develop automated tools. However, the objective of such analysis is to identify the emotion involved, rather than the sentiment, which makes it a tough task as the differences are very subtle between some emotion classes [2].

Emotions have a great role in the learning phase of people, particularly at the stage of adolescent. Much of our day-to-day experiences are affected by the emotions we feel and observe [3]. These feelings and observations are not only from the real incidents. Emotions can be felt during playing a game, online or offline, while enjoying a story, watching a favorite movie or television serial [4]. The research interest

is growing in recognizing the deliberate choices of people towards their emotional states when performing some pleasure activities like during playing an online game or enjoying a horror story or watching a mythological or sci-fi movie.

The number of approaches is rapidly increasing in the area of emotion and sentiment analysis recently. With the inclusion of technology like artificial intelligence methods, machine learning algorithms, data mining approaches for text analysis provides many latest tools for text analytics [5]. Fast computers also made it possible to process a large-dataset for the purpose of emotion analysis from a huge collection of texts. By recognizing emotional and sentimental classes of text one can build a model of different patterns from a given data [6].

Next section reviews the literature over sentiment and emotion analysis. Circumstantially elaborating psychological and computational approaches and current state-of-the art research in the area. Phenomenal work done by authors are discussed. Followed by the challenges during the analysis phase from textual data obtained from social media or available over the semantic web. Penultimate section, then mention some remarkable future research scope in the area.

2. LITERATURE REVIEW

In relation emotion expressions are the essential thing involved in interpersonal communication. Mostly, emotions can be categorized into positive, neutral or negative [6]. Generally, few keywords like happy, joy, excited, pride, shows the positive emotions, keywords like indifferent, not against, okay, impartial, unbiased, dispassionate are used to represent neutral emotions and keywords like sad, fear, depressed and disgust express the negative emotions. In this way, various forms of emotions are used to communicate. A vast source of textual data is available over the social networking platforms such as Twitter, YouTube and Instagram etc., where persons are devoting most of their leisure time in commenting and communicating their emotions. Research literature over sentiment and emotion analysis can broadly be classified into two categories, viz. psychological research and computational models.

2.1. PSYCHOLOGICAL RESEARCH

Psychological research on sentiment and emotion analysis can be further divided into three major categories.

2.1.1. Categorical Approach:

Research following this approach classifies emotions and place them into different sets called classes. These emotions are very well known in nature and recognized universally [7].

2.1.2. Dimensional Approach:

In categorical approach where the emotions are considered as independent, in dimensional approach emotions are considered as dependent and bound to other emotions. It is represented as one-dimensional or multi-dimensional space, which describes how emotions are related with other emotions based on the situation and their degree of occurrence. The degree of occurrence can be considered as high or low [7].

Russell has proposed a circumplex model [8] which represents the emotions in a 2D model. Since emotions are considered as connected, it distinguished as Valence (Pleasantness and Unpleasantness) and Arousal (Activation and Deactivation) where one dimension refers the amount by how much it is matched or dissimilar and the other dimension refers the amount by how much an emotion is positive or negative [6]. Russell's circumplex model is shown in Figure 1.

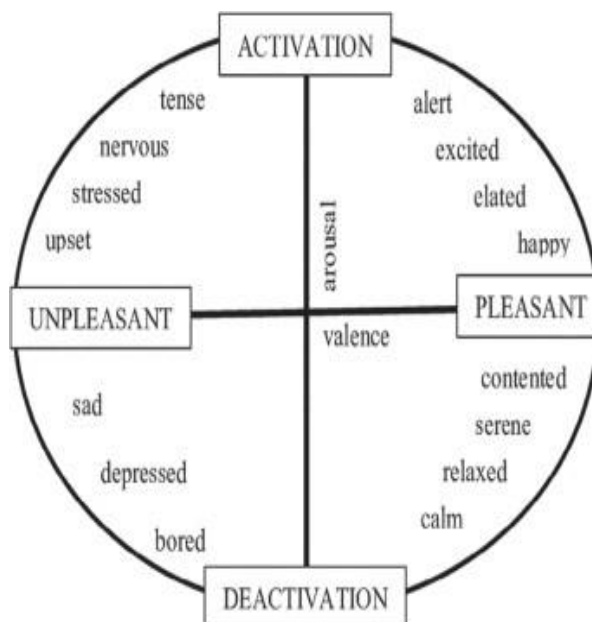


Figure 1: Circumplex model by Russell

2.1.3. Appraisal Based Approach:

This approach contains a model of emotions which is based on the appraisal theory. This theory of appraisal consider that different set of people shows different emotions for the same event. This also depend on the time of the event [9].

2.2. COMPUTATIONAL APPROACHES

The number of different types of emotions are limited in the categorical approach, which makes it hard to resolve all the emotional situation particularly when the emotions are mixed or complex. In that case, dimensional approach can be suited useful. However, that too is not suitable for all basic emotions. Hence few researchers follow another approach which can be used as per the variability of the various emotion class. One such approach is the computational model [7]. Following are the computational based approaches for emotion recognition available in the literature.

2.2.1. Keyword Based Methods:

In this approach emotion features that are in relation with other emotion classes are first selected. These features are then represented using a lexicon. Linguistic rules are checked and verified. It also explores the sentence structures. Then it processes the text for the selected dataset. Keyword based approach uses techniques like tokenization, stop-word removal and lemmatization. Then finally the approach is able to identify the different emotion class for each phrase[10].

C Liu et al. [11] expressed a framework based on a collection of specific keywords mentioning emotions. The flaw is, if a sentence does not contain any such keyword which might be supposed to express some emotion means there is no emotion contained in that sentence. For example, the sentence: "Today, I lost my mobile phone" and the sentences "Alas! Today, I lost my mobile phone", and "I lost my mobile phone today" express the same emotion (sad), but if the emotion is detected based on the keyword "Alas" only, the first and the last sentence without this "Alas" keyword will remain undetected.

Rahman et al. [12] also proposed a technique which detects emotions from sentences and classify them into various emotion classes. They have created twenty-five emotion classes in their proposed model. Their model is also based on the keywords and emoticons. High accuracy level has been achieved.

2.2.2. Corpus Based Methods:

This approach relies upon supervised learning methods to analyze the syntax and semantics of text for emotion detection. In some cases, unsupervised learning is also implemented. Wang et al. [13] developed an optimization architecture for emotion detection. Their model identifies multiple emotions and classify them according to various labels.

2.2.3. Rule Based Methods:

Models based on this approach begins with text preprocessing, followed by various other steps like tokenization, stop word removal, Part-Of-Speech (POS) tagging etc. Concepts of statistics and linguistics have been used along with the computation rules for emotion detection. At later stage of processing, rule-based approach selects the best rules for emotion detection and classify them into various labels.

Dibyendu et al. [14] suggested an emotion detection method based on the semantic rules at the sentence level. The F1 score by their approach is 66.18%. Srinivas et al. [15] suggested a rule-based model for emotion detection from Twitter dataset and classify them among different emotion classes. The accuracy achieved by their proposed model is 85%.

2.2.4. Machine Learning Methods:

These approaches are based on the Machine Learning (ML) algorithms and Natural Language Processing (NLP). Such approaches detect and classify emotion based on the text-based dataset. Both, supervised and unsupervised learning techniques of ML are used. Most widely used techniques are the Support Vector Machine (SVM), Naïve Bayes (NB) and Conditional Random Fields.

For offline and online text data Hasan et al. [16] proposed a method using NB, Decision Tree (DT) and SVM for emotion detection with 90% accuracy. A model by Merav et al [17] has been proposed for the children with communication problem. Their model helps such children in handling societal situation and response. Their approach uses the ML algorithms and SVM method to achieve a recall of 80% and more than 75% precision. Tripto et al. [18] worked upon the YouTube comments, applied ML algorithms and achieved accuracy of 59.2%. Fakhri et al. [19] developed a text-based emotion detection model using supervised learning method of ML and obtained highest accuracy as 64.08%.

In the context of Indian language like Malayalam, R Jayakrishna et al. [20] proposed an ML based approach and used SVM to classify the sentences of a Malayalam novel. This approach obtained a precision of 94% for happy, 92% for sad, 93% for fear, 90% for anger and surprise.

2.2.5. Hybrid Approach:

When multiple approaches are combined and collectively used, it is known as hybrid approach. This approach is rapidly growing over the other individual approaches.

Riahi et al. [21] suggested a hybrid framework based on three different sub-models. In their approach each sum-model analyze input data from different perspective and classify the emotions. In this way they detect the emotion from text. The efficiency of their hybrid model is better than the models based on only one single approach.

Hamed et al. [22] proposed a system combined with Long Short-Term Memory (LSTM) and Convolution Neural Network (CNN) to identify the semantics from a health model. Their dataset contains mostly patient data suffering from the cancer disease. Their framework detects emotions like joy and sad from the sentences. A comparative analysis of recently used hybrid approaches is give in the Table 1:

Table 1: Comparative analysis of recent research using hybrid approaches

Paper	Author(s)	Year of Publication	Journal	Approach/Method	Outcome Accuracy
Sentiment Analysis on Disputed Territory Discrepancies Using Machine Learning-Based Text Mining Approach[23]	ur Rehman, M. and Bashir, M.	2023	VFAST Transactions on Software Engineering	Kernel SVM	80%
Sentiment Analysis of Hausa Language Tweet Using Machine Learning Approach[24]	Sani, M., Ahmad, A. and Abdulazeez, H.S.	2022	Journal of Research in Applied Mathematics	Logistic Regression	86%
Lexicon and Deep Learning-based Approaches in Sentiment Analysis on Short Texts[25]	Taminul Islam, Md. Alif Sheakh, Md. Rezwane Sadik, Mst. Sazia Tahosin, Md. Musfiqur Rahman Foysal, Jannatul Ferdush, and Mehbuba Begum	2024	Journal of Computer and Communications	Lexicon, LSTM and Deep learning	98%
Developing Cross-Lingual Sentiment Analysis of Malay Twitter Data Using Lexicon-Based Approach[26]	Imanina Zabha, N., Ayop, Z., Anawar, S., Hamid, E. and Zainal Abidin, Z.	2019	International Journal of Advanced Computer Science and Applications	Logistic Regression, SVM	96%

It is evident from the table 1 that lexicon-based approaches linked with ML and DL based models give good results. At the same time complexity and other overheads increases. Text based approaches when combined with DL has better performed compared to other textual based approaches available in the literature.

3. CHALLENGES IN EXISTING RESEARCH

Though a plethora of work in the field of emotion and sentiment analysis has been done through different approaches, still there are many unsolved challenges. The literature survey revealed the various approaches applied on classification of emotions [27, 28], failed to achieve the accuracy at a satisfactory level. Sentiment extraction from negative sentences and having double negation words are also challenging, e.g. *I do not dislike fish-curry*. The order of words can also change the meaning of sentiments involve in the written text. For example, *I saw a happy man with a baby* or *I saw a man with*

a happy baby. Both sentences present the happiness involved, but in the former a man is happy whereas in the later one the baby is happy, due to the change of order of the words. Sentiment extraction from such ambiguous statements is a challenging task.

Sarcastic sentences also pose a great challenge in sentiment extraction. Sentences as *Oh, I am really happy to see your result* involves some pre-condition or history. Here one single sentence is not capable to give the meaningful sentiment. Depending on the *result* the sentiment can be positive or negative. In Indian context sentiment analysis based on keyword-based approaches face some bizarre ambiguity as in *Happy is sad*.

Over social media text is mostly written as misspelled or in short form or involving emoticons, which makes the sentiment analysis task difficult.

4. PERFORM EVALUATION

The performance of any model or framework can be evaluated over the following metrics:

- Accuracy = $\frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$
- Precision = $\frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$
- Recall = $\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$
- F1-Score = $\frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$

5. CONCLUSION

Sentiments are subjective impressions instead of objective facts. Different methodologies, strategies, approaches and techniques are used to recognize the different type of sentiments. Broadly sentiment analysis can either be subjectivity identification or it may be aspect-based sentiment analysis. In a former type of analysis, a sentence can be classified into different category. The main challenge in such type of analysis, is that the meaning of a word or a phrase is often reliant on the context in which the statement is written or said.

Emotions are also considered a part of sentiment analysis. Emotions can be captured from the Internet, blogs, tweets, comments, news articles, policy reviews and various social networking websites. Like sentiments emotions can also be analyzed using various techniques, including NLP, statistics, data mining and machine learning methods and can be categorized accordingly based of ones feeling. Organizations often use this information to improve their business.

6. FUTURE SCOPE

Sentiment analysis made the manufacturer to improve the quality of their products. Service providers can improve their services depending on the feedback analysis. It also helps companies to increase the sales and many more. The inclusion of social media platforms makes it possible to explode the sentiment or emotions world-wide. Grown-up popularity of Twitter, Facebook and YouTube among netizens and analysis in the comments section using emotion analysis can predict the future course of action of an individual or as a collective a community etc. To achieve the best possible analysis and prediction it is most important that emotion analysis should be sharp, clear and unambiguous. With the advancement in the technologies from 4G to 5G will remove further barriers in the study of sentiment and emotion analysis.

REFERENCES

- [1] Ronald de Sousa & Andrea Scarantino, (2018) "The Stanford Encyclopedia of Philosophy" Ed. by Edward N. Zalta. Stanford, CA.
- [2] Nirmal Varghese Babu & E. Grace Mary Kanaga, (2021) "Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review", Springer Nature Computer Science 3:74, pp 1–20.

- [3] Mayur Wankhade, Annavarapu Chandra Sekhara Rao & Chaitanya Kulkarni, (2022) "A survey on sentiment analysis methods, applications, and challenges" Springer Nature Artificial Intelligence Review, Vol. 55, pp 5731–5780.
- [4] Pansy Nandwani & Rupali Verma, (2021) "A review on sentiment analysis and emotion detection from text", Springer Nature Social Network Analysis and Mining, Vol. 11, No. 81, pp 1–19.
- [5] Jingfeng Cui, Zhaoxia Wang, Seng-Beng Ho & Erik Cambria, (2023) "Survey on sentiment analysis: evolution of research methods and topics", Springer Nature Artificial Intelligence Review, Vol. 56, pp 8469–8510.
- [6] Kopelman, Shirli, Ashleigh Shelby Rosette & Leigh Thompson, (2006) "The three faces of Eve: Strategic displays of positive, negative, and neutral emotions in negotiations." *Organizational Behavior and Human Decision Processes* 99, pp: 81–101.
- [7] Gunes H & Pantic M, (2010) "Automatic, dimensional and continuous emotion recognition", *Int J Synth Emot (IJSE)* 1(1), pp: 68–99.
- [8] Russell JA, (1980) "A circumplex model of affect", *J Pers Soc Psychol*, 39(6), pp: 1161–1178.
- [9] Kao Edward Chao-Chun, Yang Ting-Hao, Hsieh Chang-Tai & Soo Von-Wun, (2009) "Towards text-based emotion detection a survey and possible improvements", *International Conference on Information Management and Engineering*, IEEE.
- [10] Vishal A. Kharde & S. S. Sonawane, (2016) "Sentiment Analysis of Twitter Data: A Survey of Techniques", *International Journal of Computer Applications*, Vol. 139, No. 11, pp 5–15.
- [11] Liu C, Osama M & DeAndrade A. DENS, (2019) "A dataset for multi-class emotion analysis", arXiv preprint arXiv:1910.11769.
- [12] Islam T, Ema R R & Ahmed H, (2017) "Detecting emotion from text and emoticon", *London Journal of Research in Computer Science and Technology*, pp: 9–13.
- [13] Wang, Yichen & Aditya Pal, (2015) "Detecting emotions in social media: A constrained optimization approach", *Twenty-fourth international joint conference on artificial intelligence*, pp: 996–1002.
- [14] Seal D, Roy UK & Basak R, (2020) "Sentence-level emotion detection from text based on semantic rules", *Proceedings of the Information and Communication Technology for Sustainable Development*, Springer, pp: 423–430.
- [15] Badugu S & Suhasini M., (2017) "Emotion detection on twitter data using knowledge base approach", *International Journal of Computer Applications*, 162(10), pp: 28–33.
- [16] Hasan M, Rundensteiner E & Agu E., (2019) "Automatic emotion detection in text streams by analyzing twitter data", *Int J Data Sci Anal.*, 7(1), pp: 35–51.
- [17] Allouch, Merav, Azaria Amos, Azoulay Rina, Ben-Izchak Ester, Zwilling Moti & Zachor Ditzza A., (2018) "Automatic detection of insulting sentences in conversation", *IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE)*. IEEE, pp: 1–5.
- [18] Tripto, Nafis Irtiza & Mohammed Eunus Ali, (2018) "Detecting multilabel sentiment and emotions from bangla youtube comments", *International Conference on Bangla Speech and Language Processing (ICBSLP)*. IEEE.
- [19] Nasir Ahmad Fakhri Ab, Nee Eng Seok, Ghani Ahmad Shahrizan Abdul, Majeed Anwar P. P. Abdul, Adam Asrul & Furqan Mhd., (2020) "Text-based emotion prediction system using machine learning approach", *IOP Conference Series: Materials Science and Engineering*. Vol. 769. No. 1. IOP Publishing, pp: 1–11.
- [20] Jayakrishnan R, Gopal Greeshma N & Santhikrishna MS., (2018) "Multi-class emotion detection and annotation in Malayalam novels", *Proceedings of the 2018 International Conference on Computer Communication and Informatics.*, pp: 1–5, IEEE.
- [21] Riahi N & Safari P, (2016) "Implicit emotion detection from text with information fusion", *J Adv Comput Res*, 7(2), pp: 85–99.
- [22] Khanpour Hamed & Cornelia Caragea, (2018) "Fine-grained emotion detection in health-related online posts", *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp: 1160–1166.
- [23] ur Rehman, M. & Bashir, M., (2023) "Sentiment Analysis on Disputed Territory Discrepancies Using Machine Learning-Based Text Mining Approach", *VFAST Transactions on Software Engineering*, 11, pp. 17-25.
- [24] Sani, M., Ahmad, A. & Abdulazeez, H.S., (2022) "Sentiment Analysis of Hausa Language Tweet Using Machine Learning Approach". *Journal of Research in Applied Mathematics*, 8, pp. 7-16.
- [25] Taminul Islam, Md. Alif Sheakh, Md. Rezwane Sadik, Mst. Sazia Tahosin, Md. Musfiqur Rahman Foysal, Jannatul Ferdush & Mehbuba Begum, (2024) "Lexicon and Deep Learning-based Approaches in Sentiment Analysis on Short Texts", *Journal of Computer and Communications*, vol. 12, pp: 11–34.

- [26] Imanina Zabha, N., Ayop, Z., Anawar, S., Hamid, E. and Zainal Abidin, Z., (2019) “Developing Cross-Lingual Sentiment Analysis of Malay Twitter Data Using Lexicon-Based Approach”, *International Journal of Advanced Computer Science and Applications* , 10, pp. 346-351.
- [27] Gan Chengguang, Quinghao Zhang & Tatsunori Mori, (2023) “USA: Universal Sentiment Analysis Model & Construction of Japanese Sentiment Text Classification and Part of Speech Dataset”, arXiv:2309.03787v2 [cs.CL], Springer Nature, pp. 1–11.
- [28] Monali Bordoloi & Saroj Kumar Biswas, (2023) “Sentiment analysis: A survey on design framework, applications and future scopes”, *Springer Nature Artificial Intelligence Review*, Vol. 56, pp 12505–12560.