Emoji's Impact on Sentiment Analysis – A Survey

M.Srisankar[1], Dr.K.P. Lochanambal[2]
[1]*Research scholar, Department of Computer Science, Government Arts college, Udumalpet*
[2]*Assistant Professor, Department of Computer Science, Government Arts College, Udumalpet*

***Abstract* - Sentiment Analysis (SA) is a still-developing field of text mining research. It refers to the algorithmic processing of text's ideas, feelings and subjectivity. The usage of emojis on social media has skyrocketed in recent years. As a result, we have focused on the importance of emojis in SA. This survey article aims at offering a comprehensive overview of the most current developments in this field. This study looks at and rapidly covers a variety of recently suggested algorithm advances as well as several SA applications. The contributions of these papers to various SA techniques are grouped into categories. The importance of related fields to SA, such as transfer learning, emotion detection and resource building, which have lately attracted researchers is emphasized. The study's main objective is to present a near-complete overview of SA techniques and related fields, as well as some background information. The sophisticated categorizations of a significant number of recent articles, as well as the description of a recent research trend in sentiment analysis and related disciplines are the main contributions of this work.**

***Index Terms* - Sentiment Analysis (SA), Text and emoji, Twitter data.**

I.INTRODUCTION

Sentiment Analysis is a technique for identifying people's perceptions, judgments, feelings, perspectives, opinions, conclusions and other ideas about anything.[1] It's a way of analyzing messages, photographs, emojis, and other actions to see what other people think about a product, service, company, brand name or reaction to a certain event, social movement or other issue by analysing texts, photos, emojis and other activities. Without individuals expressing their feelings, thoughts and viewpoints, understanding their feelings, opinions and perspectives is difficult. Users may easily learn about people's sentiments you may rapidly learn about people's sentiments if you provide places where they may openly share their opinions and concerns. [2].

An emoticon, such as ;-), represents a facial emotion. It aids the author in expressing his or her feelings, moods, and emotions, as well as adding nonverbal elements to a written message. [3] It helps to draw the reader's attention to the material while also increasing and improving understanding.

Emojis take things a step further, allowing for more expressive communication through the use of modern communication technologies. [4] Emojis are graphic symbols that represent not just face expressions but also concepts and ideas such as festivities, weather, automobiles and structures, food and drink, animals and plants, as well as emotions, feelings, and activities.

Figure 1. Emotions in social media

Figure.1 [5] represents emoticons such as :) ;) :-) and :(is commonly used in online social interactions like social media, instant messaging (e.g., Skype), blogs, forums and other forms of online social interactions. Emoticons in text were often utilized by NLP researchers in tasks like sentiment analysis as features for Machine Learning algorithms or entries of sentiment lexicons for rule-based techniques since they are commonly used in online interactions and are often direct signals of sentiment. [6]

Depending on the online community and tool, different levels of emoticon usage may be elicited. Twitter, a microblogging platform, is one of the most widely used social networking platforms. Having access to its massive amount of user-generated data is essential for researchers and companies to understand user behavior and attitude. We thought it would be interesting to learn about the prevalence of emoticons on Twitter nowadays, how users express and perceive sentiment through emoticons, and whether emoticons can be used as a reliable cue for sentiment polarity classification with access to about 50 million tweets per day via the Twitter Decahose Application programming interface. This application makes use of Artificial Intelligence and Machine Learning. [6]

The structure of the survey is summarized below. Let's take a brief look at the most relevant works on sentiment analysis in Section II. Sections III and IV contain studies on Emojis and tweeter data. In Section V, we will go through some of the most recent machine learning approaches for sentiment analysis. Section VI brings the paper to a conclusion and future work.

II.LITERATURE REVIEW

Sentiment analysis has long been a popular research topic in Natural Language Processing at various degrees of granularity. Sentiment analysis is broken down into three stages: phrase, document, and aspect. [7] By analysing the entire text as a single unit, document-level sentiment analysis determines if a piece of material represents a positive, negative, or neutral opinion. Sentiment analysis indicates if a statement is subjective or objective at the sentence level. If the statement is subjective, it determines whether it reflects a negative or positive viewpoint. All aspects of a document's expressed views are identified using aspect-level sentiment analysis.

Words categorized as good, negative, or neutral feelings form the foundation of sentiment analysis. The majority of sentiment analysis research is focused on English text and lexicon corpora. General Inquirer is a lexicon-based system that uses pointers to connect words with related meanings. SentiWordNet 3.0 is a lexicon-based method and an improvement of WordNet, including positive, negative and neutral numerical ratings based on synsets. [8,9] The SenticNetwhich consists of 14,244 commonsense concepts, uses a concept-level opinion mining method. [10]

III.EMOTICONS AND EMOJIS

Previously, emoticons were used to communicate emotions through text on the web. Emoticons are also characters such as punctuation marks, numerals, and letters that are used to express emotions on the face. The data in this study was trained using language and emoticons. Emojis have taken the role of emoticons in recent years. [11]

Emojis initially appeared in Japan's mobile phones in 1997 and in Apple's operating system in 2011. Some emoji character sets from 2010 have been included into the Unicode Standard. [12] In 2018, the new Emoji Version 11.0 was published, bringing the total number of authorized emoji to 2,823.[13]

Emoji emotions were computed using the sentiment of tweets. In Arabic tweets, the most commonly used emojis were analysed and classified as anger, disgust, joy, and sadness. The majority of sentiment analysis research has focused on text or emoji, rather than both. Emoticons have been used to create sentiment lexicons and train machine learning classifiers in a number of previous researches. It has long been assumed that emoticons are accurate indications of emotion. Several attempts to create a sentiment corpus based on emoticons have been attempted. However, none of the research has looked at the link between emoticons and emotion polarity on social media, or the functions that emoticons play in such situations. [14]

Since so many users interchange emoticons with emojis, and the two frequently have the same implications, they are not the same thing. Emoticons are a visual representation of a facial expression made up entirely of characters available on a keyboard. Two instances are ":(" and ":)". Emojis are visual symbols that represent concepts such as “Each emoji is a Unicode character that cannot be typed on a keyboard. Emoticons and emojis are interchangeable, and emojis are frequently regarded as the most recent generation of emoticons.[15]

Byungkyu and Julia analyzing that emojis usually take the place of emoticons; they are never used simultaneously. When compared to words, emojis convey more emotion. Emojis close to each other have a higher sentiment ratio, which may be used to improve sentiment analysis accuracy. Emojis are often similar to other emojis or current terms such as "lol" or "haha," however these words should not be utilised to replace emojis while performing sentiment analysis. Emoji feelings assist train other English words, thus replacing them with one or a few words results in the best accuracy. [15]

IV.EMOJIS IN TWITTER

Twitter is one of the most significant and well-known social networks in the digital world. Every day, over 695 million people tweet over 58 million tweets. As of September 2016, the number of active users surpassed 342 million [16]. 43 percent of those users send tweets from their mobile phones. This makes Twitter an even better source of sentiment data since people may share their thoughts or views about any event as it happens. Twitter users tweet about almost any issue, which is another benefit of Twitter. People express their thoughts and feelings about both happy and negative situations. The usage of hashtags is yet another significant element of Twitter. Users can categories tweet subjects by using hashtags. Hashtags assist to categorize subjects and make it easier to find tweets about them. The hashtags may also be used to monitor tweets related to a certain event, such as the "Oscars" or "World Cup."

83 human annotators were hired to categorise 1.6 million tweets from 13 different European languages as negative, neutral, or positive. They discovered that Emoji were used in about 4% of the tweets they gathered. Using the emotion score of the plain text, they rated the 751 greatest frequent Emoji characters. They assigned Emoji characters to represent the classes. F Barbieri et al. [17] obtained 10 million English-language tweets in terms of creating skip-gram word embedding models by mapping words and Emoji in the same space.

Many devices show the emoji with various visuals, which may somewhat alter the meaning [18]. Twitter data are slightly different since Twitter has its own emoji visuals called Twemoji. Except for the native Twitter apps for iPhones, Twemoji is shown similarly across all devices. The meaning of emojis may also vary by culture [19]. For example, an emoji with a happy face has a good feeling in China but a negative sentiment in Argentina.

Shiha and Ayvaz studied the use of Emoji characters on social networks, as well as their implications on text mining and sentiment analysis. They looked at several significant positive and unfavorable events for seeing whether there was a difference in Emoji usage between positive and negative occurrences. They also discovered that using Emoji in sentiment analysis improves overall sentiment ratings. While Emoji characters are used to convey both positive and negative ideas, the use of Emoji characters in sentiment analysis appeared to increase the expressivity and overall sentiment scores of positive opinions in their analysis considerably more than negative ones. [20]

There were numerous publications that used Twitter data to train their algorithms with emoticons. Chen et al. (2018) developed an RNN model for Twitter sentiment analysis that incorporates bi-sense emoji embedding. The authors demonstrate using this approach how emojis may also be used to represent more sensitive emotions such as hate, sorrow, joy, and so on. [21]

V.MACHINE LEARNING FOR SENTIMENT ANALYSIS

Supervised learning and unsupervised learning are the two main machine learning approaches utilised in sentiment analysis. The trained dataset is required in supervised learning, but the labelled dataset is not required in unsupervised learning.

Wang and Castanon (2015) investigated the function of emoticons in both the construction of sentiment lexicons and the training of learning classifiers. In terms of sentiment accuracy, the model reported in the study improved by about 15%. The research also found that big groupings of emoticons express complex emotions and should be used with considerable caution. Emoticons are not always constant and often change depending on the environment and the person using them. [22]

Similarly, there has been research done to measure the impacts of emoji in sentiment analysis. The research revealed that using emojis in sentiment analysis resulted in a better sentiment score. The research anticipated that emoji characters had a greater influence on overall feelings of good thoughts than negative ones. This intelligence may be used to track items, enhance services, and forecast impending occurrences. [23]

Guibon, Ochs, and Bellot (2016) proposed that the impacts of emojis are not limited to simply sentiment expression but may also be extended to sentiment augmentation and sentiment alteration. Emoji usage makes them highly confusing and untrustworthy when taken out of context, thus distinct emoji usages may be recognized by comparing their sentiment to the sentiment of the phrase in which they are included. [24]

Chen et al. presented a unique sentiment classification technique that combined the current Chinese sentiment vocabulary with a convolutional neural network (CNN). [25] Their findings demonstrate that the suggested method outperforms the convolutional neural network (CNN) model with only word embedding data [26].

Peng, Cambria, and Hussain (2017) performed a detailed assessment of the literature on Chinese sentiment analysis. According to their findings, there are two techniques to Chinese sentiment analysis: supervised machine learning and unsupervised knowledge-based approaches. [27]

Text segmentation, feature extraction, and sentiment classification are the three distinct phases for machine learning-based sentiment analysis tasks. Text segmentation is the process of dividing a text into meaningful tokens. Feature extraction collects emotion features as well as raw segmented words characteristics and stores them in a bag of words (BoW). Finally, the dataset is put into a machine learning model, which assigns a sentiment score to the provided text. Nave Bayes, maximum entropy, SVM, neural networks, and other algorithms are frequently employed. [28]

Tan and Zhang's empirical study, which classified sentiment in Chinese papers, is a good example of such research [29]. Four characteristics, including mutual information, information gain, chi-square, and document frequency, were examined independently on five different algorithms, including centroid classifier, k-Nearest Neighbor, Winnow classifier, NB, and SVM (SVM). Among these methods, the information gain and SVM features were discovered to produce the greatest results using topic-dependent classifiers.

VI.CONCLUSION AND FUTURE WORK

This survey study provided an overview of current changes to SA algorithms and applications. Many of the recently published and referenced articles were sorted and summarised. These papers contribute to numerous SA-related disciplines by utilizing SA approaches in a variety of real-world applications. After reviewing these publications, it is apparent that improving sentiment classification algorithms is still a work in progress. The most used ML methods for handling sentiment classification problems are Nave Bayes, Support Vector Machines, Neural Networks and Deep learning algorithms. They are regarded as a reference model against which many suggested algorithms are evaluated. According to the research described above, emojis in sentiment analysis may be utilised to improve word embeddings and vector space, as well as construct rich sentiment lexicons. The inclusion of emoticons in addition to text enhances classifier performance. Furthermore, several deep learning models are gradually becoming popular tools for sentiment categorization. It may be studied in the future to incorporate multiple senses of emojis co-occurring with text and to do multi-class analysis rather than simply positive and negative classes to expose the deeper emotional experiences in the text such as humour, irony, frustration, and so on.

REFERENCE

[1] Barbieri, F., Ronzano, F., & Saggion, H. (2016). What does this emoji mean? A vector space skip-gram model for twitter emojis. Proceedings of Language Resources and Evaluation Conference, LREC, Slovenia: Portoroz.

[2] Geetika Vashisht, Manisha Jailia.(2020). Emoticons & Emojis Based Sentiment Analysis: The Last Two Decades!. International Journal of Scientific & Technology Research. Volume 9, issue 03, March 2020. ISSN 2277-8616

[3] Guibon, G., Ochs, M., & Bellot, P. (2016, June). From emojis to sentiment analysis. In WACAI 2016.

[4] Yinxia Lou, Yue Zhang, Fei Li, Tao Qian, and DonghongJi. (2020). Emoji-Based Sentiment Analysis Using Attention Networks. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 19, 5, Article 64 (August 2020), 13 pages. DOI:https://doi.org/10.1145/3389035

[5] Free Icon Library. https://icon-library.com/icon/sentiment-icon-14.html.

[6] Wang, H., & Castanon, J. A. (2015, October). Sentiment expression via emoticons on social media. In *2015 ieee international conference on big data (big data)* (pp. 2404-2408). IEEE.

[7] Fernández-Gavilanes, M., Costa-Montenegro, E., García-Méndez, S., González-Castaño, F. J., & Juncal-Martínez, J. (2021). Evaluation of

online emoji description resources for sentiment analysis purposes. Expert Systems with Applications, 184, 115279.

[8] Cambria E, Olsher D, Rajagopal D. (2014).SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In: AAAI, Quebec City; 2014. p. 1515–21. 22.

[9] Cambria E, Hussain A. (2012). Sentic computing: techniques, tools and applications, Springer briefs in cognitive computation. Berlin: Springer; 2012. 23.

[10] Cambria E, Hussain A. (2012) Sentic PROMs: application of sentic computing to the development of a novel unified framework for measuring health-care quality. Expert Syst Appl. 2012;39(12):10533–43.

[11] Al-Azani, S., & El-Alfy, E. S. M. (2018, April). Emoji-based sentiment analysis of Arabic microblogs using machine learning. In 2018 21st Saudi Computer Society National Computer Conference (NCC) (pp. 1-6). IEEE.

[12] Allen JD, et al. (2014). The Unicode Consortium, the Unicode Standard, Version 7.0. Mountain View, CA; 2014.Available from: http://www.unicode.org/versions/Unicode 7.0.0/

[13] Allen JD, et al. (2015). The Unicode Consortium, The Unicode Standard, Version 8.0. Mountain View, CA; 2015. Available from: http://www.unicode.org/versions/Unicode 8.0.0/.

[14] Al-Azani, S., & El-Alfy, E. S. M. (2018, April). Emoji-based sentiment analysis of Arabic microblogs using machine learning. In 2018 21st Saudi Computer Society National Computer Conference (NCC) (pp. 1-6). IEEE.

[15] Yoo, B., & Rayz, J. T. (2021, April). Understanding Emojis for Sentiment Analysis. In The International FLAIRS Conference Proceedings (Vol. 34).

[16] Statistic Brain. (2016). STATS | Twitter company statistics. Statistic Brain. Retrieved Jan. 30, 2017, from http://www.statisticbrain.com/twitter-statistics/.

[17] Barbieri, F., Ronzano, F., & Saggion, H. (2016). What does this emoji mean? A vector space skip-gram model for twitter emojis. Proceedings of Language Resources and Evaluation Conference, LREC, Slovenia: Portoroz.

[18] Gupta, S.; Singh, A.; and Ranjan, J. 2020. Sentiment analysis: Usage of text and emoji for expressing sentiments. In Kolhe, M. L.; Tiwari, S.; Trivedi, M. C.; and Mishra, K. K., eds., Advances in Data and Information Sciences, 477–486. Singapore: Springer Singapore.

[19] Financieras, C. N. 2020. World emoji day: which are the most commonly used and what meanings have according to cultures. Copyright - CE Noticias Financieras English, Latin America - Distributed by ContentEngine LLC; Last updated - 2020-12-30.

[20] Shiha, M., & Ayvaz, S. (2017). The effects of emoji in sentiment analysis. Int. J. Comput. Electr. Eng.(IJCEE.), 9(1), 360-369.

[21] Chen, Y.; Yuan, J.; You, Q.; and Luo, J. 2018. Twitter sentiment analysis via bi-sense emoji embedding and attention based lstm. In Proceedings of the 26th ACM International Conference on Multimedia, MM ’18, 117–125. New York, NY, USA: Association for Computing Machinery.

[22] Wang, H., and Castanon, J. A. 2015. Sentiment expression via emoticons on social media. In 2015 IEEE International Conference on Big Data (Big Data), 2404–2408.

[23] Ayvaz, S., and Shiha, M. 2017. The effects of emoji in sentiment analysis. International Journal of Computer and Electrical Engineering 9:360–369.

[24] Guibon, G.; Ochs, M.; and Bellot, P. 2016. From Emojis to Sentiment Analysis. In WACAI 2016. Brest, France: LabSTICC and ENIB and LITIS.

[25] Chen, Z., Xu, R., Gui, L., & Lu, Q. (2015). Combining convolution neural network and word sentiment sequence features for Chinese text sentiment analysis. Journal of Chinese Information Processing, 29(6), 172–178.

[26] Kim, Y. (2014). Convolutional neural networks for sentence classification. ArXiv preprint arXiv: 1408.5882.

[27] Peng, H., Cambria, E., & Hussain, A. (2017). A review of sentiment analysis research in Chinese language. Cognitive Computation, 9(4), 423–435.

[28] Dhande, L. L., & Patnaik, G. K. (2014). Analyzing sentiment of movie review data using naive Bayes neural classifier. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 3(4), 313‐320.

[29] Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for Chinese documents. Expert Systems with applications, 34(4), 2622‐2629.