Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

SENTIMENT ANALYSIS TO HANDLE COMPLEX LINGUISTIC STRUCTURES: A REVIEW ON EXISTING METHODOLOGIES

Suvendra Kumar Jayasingh, Raj Kumar Mishra, Srutipragyan Swain Department of Computer Science & Engineering, Institute of Management and Information Technology, Cuttack, BPUT, Odisha, India
Amiya Kumar Sahoo Department of Computer Science & Engineering, AIET, Bhubaneswar
,[1,2,3]sjayasingh,rkmishra,sruti56}@gmail.com, [4]amiya79@rediffmail.com

Abstract: Sentiment analysis is an important field of natural language processing that focuses on identifying and extracting emotions and opinions from text data. With the increasing use of social media and messaging platforms, sentiment analysis has become even more important. In this paper, we present a review of the existing literature on sentiment analysis on single and multilingual mixed words/scripts sentences with or without emojis embedded. But we believe that the presence of emojis in text messages can add an additional layer of complexity to sentiment analysis, as they may convey emotions or sentiments that are not explicitly stated in the text. We summarize the various approaches and techniques used for sentiment analysis in this context, including rule-based methods, machine learning methods, and hybrid approaches. We also discuss the datasets used in these studies and the evaluation metrics used to assess the performance of the sentiment analysis models. Our review shows that sentiment analysis on multilingual mixed word sentences with emojis embedded is still an active area of research, with many challenges to be addressed, including the need for more annotated datasets and the development of more robust models.

Keywords: Sentiment Analysis, Natural Language Processing, Odia, English, Emoji, Mixed Language, Review

INTRODUCTION

Sentiment analysis is a rapidly evolving field that has gained significant attention in recent years due to the growth of social media and the need to understand public opinion because it has the potential to provide valuable insights into public opinion and has applications in various domains, including marketing, politics, and healthcare.

Sentiment analysis, also known as opinion mining, is a process that involves identifying, categorizing, and recognizing the emotions and opinions expressed by users for various services, such as movies, product issues, events, or any attribute as positive, negative, or neutral. The sources for sentiment analysis are social communication channels, including websites such as reviews, forum discussions, blogs, micro-blogs, Twitter, etc. The popularity of sentiment analysis research has increased significantly in recent times, mainly because of the vast amount of opinionated data available in digital forms. For a particular topic or opinion, sentiment analysis involves the mining of data to provide output. Emotion-based summarization and mind extraction are some of the research areas related to sentiment analysis. Sentiment analysis uses Natural Language Processing (NLP) to analyze public opinion about a particular topic or service by following the emotions and feelings expressed. This field is critical for businesses and organizations that need to understand and analyze the sentiments of their customers and stakeholders.

The review paper is organized into four main sections, with each section serving a specific purpose. The objective of the paper is presented in Section 2, which provides a brief background of sentiment analysis, including its definition, history, and applications. This section aims to introduce the reader to the topic and provide a foundation for the subsequent sections. Section 3, which is dedicated to the literature review and this section aims to provide an overview of the existing research and highlight the different approaches used in sentiment analysis. Section 4 is dedicated to the analysis of various methodologies used in sentiment analysis, including supervised, unsupervised, and deep learning-

based methods. The last section, Section 5, focuses on the conclusion and future scope of the review paper. This section summarizes the key findings of the review and presents the potential areas for further research.

II. OBJECTIVES

The objective of sentiment analysis is to automatically identify and classify the emotional states, opinions, and attitudes expressed in text, speech, or other forms of communication. This involves analyzing a large volume of data from various sources, including social media, customer reviews, and news articles, to understand how people feel about a particular product, service, event, or topic. The primary goal of sentiment analysis is to help businesses and organizations make data-driven decisions based on customer feedback, public opinion, and market trends. By analyzing sentiment, companies can identify their strengths and weaknesses, improve their products and services, and make informed decisions about marketing, branding, and public relations.

Sentiment analysis has a wide range of applications, including customer service, brand monitoring, political polling, and market research. It can also be used to track the sentiment of employees, patients, or students, and to monitor social issues such as racism, sexism, and hate speech. In addition, sentiment analysis can be used in conjunction with other machine learning techniques, such as natural language processing, machine translation, and speech recognition, to provide a more accurate and comprehensive analysis of the data. Overall, this review paper aims to provide a comprehensive understanding of the current state of sentiment analysis research, along with its potential applications and future directions.

III. LITERATURE REVIEW

3.1. Studies Abroad:

Oscar Araque et al. [1] have presented their study in two parts. The first part involved the release of two new high-performance and high-coverage lexica that target the English and Italian languages. The English version of DM++ was an improved version of DM2014, constructed using a larger dataset, while the Italian lexicon is entirely new and the first publicly available large-scale emotion lexicon for this language. In the second part, the study extensively benchmarked different setup decisions affecting the construction of the two resources and evaluated their performance on various datasets and tasks, showcasing a wide range of domain, languages, settings, and task diversity. Hamid Bagheri et al. [2] have highlighted the significance of social network analysis and its diverse applications across various fields. It mainly focused on Twitter and employs a Python program to conduct sentiment analysis. The research presented the outcomes of the analysis on various daily topics. To gather the dataset, the study employs the Tweepy library, and for sentiment analysis, the researchers used the TextBlob library. The findings revealed that the number of neutral sentiments was considerably high, indicating a need for improving Twitter sentiment analysis. Maite Taboada et al. [5] have developed a word-based method to extract sentiment from texts. They built upon previous research that focused on adjectives and used Amazon's Mechanical Turk service to collect data and validate their dictionary. The team extended the Semantic Orientation Calculator (SO-CAL) to include other parts of speech and introduced intensifiers while refining their approach to negation. Their results showed a statistically significant improvement over previous iterations of the SO-CAL system. Furthermore, the researchers demonstrated that a manually built dictionary serves as a solid foundation for a lexicon-based approach, which is essential for maximizing the benefits of a system like SO-CAL. Overall; the team's work provides a valuable contribution to sentiment analysis research. SAMINA KAUSAR et al. [7] have used a custom-built crawler in Python to obtain a dataset from Amazon, and then evaluated the performance of six classifiers including Random Forest, Decision Tree, Naive Bayes, Support Vector Machine, Gradient Boosting, and Sequence to Sequence models. They experimented with three combinations of features and found that their review-level categorization approach produced promising results, with an accuracy rate of 81%. This accuracy rate was 3% higher than the average accuracy rate achieved by many

previous techniques, which was 78%. Erik Cambria et al. [8] have compared SenticNet 6 with 15 other popular sentiment lexica across six datasets. They also proposed a new version of SenticNet that combines a top-down approach, which leverages symbolic models such as logic and semantic networks to encode meaning, with a bottom-up approach that uses subsymbolic methods such as biLSTM and BERT to implicitly learn syntactic patterns from data. The researchers believe that coupling symbolic and subsymbolic AI is crucial for advancing from natural language processing (NLP) to natural language understanding. They assert that machine learning is only effective in making educated guesses based on past experiences, as it only encodes correlations and relies on probabilistic decision-making. SANA NABIL et al. [12] have tried to examine the performance of the most widely used sentiment classification algorithms and to compare their results. The author applied several algorithms, including Naive Bayes, Support Vector Machine (SVM), Decision Tree, and Logistic Regression, to analyze the sentiments expressed in Amazon product reviews. The SVM classifier proved to be the most effective, with an accuracy of 100%, while the Decision Tree classifier achieved an accuracy of 75%. Ali Hasan et al. [18] have presented a framework for the collection, sentiment analysis, and classification of Twitter opinions regarding political trends. The process was begun by collecting tweets expressed through hashtags. The tweets were stored in a database, translated into Urdu if necessary, and pre-processed. The author calculated polarity and subjectivity using SentiWordNet, W-WSD, and TextBlob libraries. The Naive Bayes and SVM classifiers were applied to the training set using Weka, leading to the creation of a classification model. Hassan Raza et al. [19] have conducted a sentiment analysis on scientific citations using an annotated corpus consisting of 8736 citation sentences from research papers in the computational linguistics field. They utilized different machine learning classifiers, including NB, SVM, DT, LR, KNN, and RF, along with various features to process the data and optimize classification results. The study revealed that SVM performed better than the other classifiers. Additionally, uni-grams, bi-grams, and tri-grams features were found to perform very well, contributing to the classifiers' ability to achieve high accuracy scores. GUIXIAN XU et al. [23] have proposed an improved word representation method that integrates sentiment information into the traditional TF-IDF algorithm to generate weighted word vectors. These vectors were then feed into a bidirectional long short-term memory (BiLSTM) to effectively capture contextual information and better represent comment vectors. The sentiment tendency of the comment is obtained using a feedforward neural network classifier. The proposed method was compared to other sentiment analysis methods like RNN, CNN, LSTM, and NB and was found to have higher precision, recall, and F1 score.. Chuchu Liu et al. [26] have presented in this research, a deep learning architecture known as the Chinese emoji-embedding Long Short-Term Memory (CEmo-LSTM) model was proposed to investigate the effect of emojis on sentiment analysis. Both the plain text and embedded emojis were utilized as input features, and the influence of emoji introduction and the uncertainty of emoji tags were also analyzed. The results showed that the accuracy of supervised learning algorithms was generally higher compared to unsupervised learning algorithms. Zhenpeng Chen et al. [27] have provided SEntiMoji as a new approach for sentiment analysis in software engineering using emojis. This model was based on DeepMoji, a representation model pre-trained on Tweets that can represent texts with sentiment-aware vectors. In order to incorporate more technical jargon into DeepMoji, GitHub posts were also used. The fine-tuned representation model and manually labeled data were then used to train the final sentiment classifier.

3.2. Studies in India:

Ayushi Mitra [3] discussed various approaches for performing sentiment analysis and presented their accuracy results. She attempted to provide a comprehensive overview of the architecture and explain the functioning of a machine learning model in both the training and prediction phases. To implement the sentiment analysis, the author used the Natural Language Toolkit (NLTK) feature in Python. As part of the implementation, the author utilized sample movie review data to demonstrate the effectiveness of the sentiment analysis approach. Gaurav Mohanty et al. [4] have created a corpus of

2045 Odia sentences sourced from news articles and provided annotations for the dataset. The inter-annotator agreement score for the dataset was found to be significant, with a score of 0.791. The study's primary objective was to create and test the performance of an Odia sentiment lexicon. Gaurav Mohanty et al. [6] have presented a novel approach for creating a SentiWordNet for the low-resource language of Odia, using only existing resources available for Indian languages. Authors believe that the resulting lexicon can serve as a valuable resource for conducting sentiment analysis tasks on Odia data. To validate the effectiveness of their approach, the researchers provided a sample set of data to three manual annotators, who were native Odia speakers and spoke the language on a daily basis. These annotators independently annotated the data as either positive or negative, enabling the researchers to evaluate the accuracy of their SentiWordNet. Overall, the team's approach provides a promising solution for developing sentiment analysis resources for low-resource languages such as Odia. Kamal Sarkar et al. [9] have developed a sentiment analysis system for Indian languages, specifically Hindi and Bengali, using multinomial Naive Bayes as a classifier. The system was trained on labelled training data to classify tweets into positive, negative, or neutral categories. The evaluation was based on accuracy and was conducted for both constrained and unconstrained cases in Hindi and Bengali. The team performed four different runs and found that the performance of their system was comparable to other top-performing systems in the contest. Overall, the researchers' work contributes to the advancement of sentiment analysis in Indian languages and demonstrates the effectiveness of their approach. Pinkesh Badjatiya et al. [10] have utilized a dataset of 16K annotated tweets to compare deep learning methods with state-of-the-art char/word n-gram methods for hate speech detection. Their experiments with multiple baseline classifiers for both the TF-IDF and BoWV approaches showed that deep neural network architectures significantly outperformed existing methods, with an improvement of approximately 18 F1 points. The researchers found that combining embeddings learned from deep neural network models with gradient boosted decision trees led to the best accuracy values. Overall, the team's findings have demonstrated the effectiveness of deep learning methods for hate speech detection and highlight the potential of these approaches for improving upon existing methods. Dipak Kawade et al. [11] have conducted a study to retrieve tweets about the Uri attack and analyze the emotions and polarity expressed in those tweets using text mining techniques. They collected approximately 5000 tweets, pre-processed them, and created a dataset of frequently appearing words. They used R and its text mining packages to mine emotions and polarity in the tweets. The experimental results showed that 94.3% of people expressed disgust towards the Uri attack. The study highlights the potential of text mining techniques for analyzing public sentiment towards specific events or topics on social media platforms such as Twitter. Palak Baid et al. [13] have analyzed movie reviews utilizing different techniques, including Naïve Bayes, K-Nearest Neighbour, and Random Forest. The data used for the analysis was gathered from 2000 user-created movie reviews found on the IMDb (Internet Movie Database) web portal. After conducting the analysis, the Naïve Bayes classifier emerged as the technique that provided the best results. Mukhtiar Singh et al. [14] have collected a corpus of English-Punjabi code-mixed text from micro-blogging websites. They then conducted a sentiment analysis on the text, utilizing the lexicon approach. To determine the sentiment polarity of the English-Punjabi code-mixed dataset at the sentence level, they employed the N-gram approach, which is a statistical technique. The results of their analysis indicated that the accuracy of the sentiment analysis was 83%, with an F-1 measure of 77%. Sanjib Kumar Sahu et al. [15] have developed a methodology for determining the polarity of the Odia language, which they tested on movie reviews using a Naive Bayesian classifier. The results of the experiment showed that the approach they utilized performed well in this domain, achieving an accuracy rate of 90%. S.K. Mohapatra et al. [16] have proposed a solution for detecting hate speech on social media using machine learning techniques. They developed and compared machine learning and text feature extraction methods specifically for hate speech detection in the Odia-English mixed code language. To accomplish this, they utilized a dataset of posts and comments from public Facebook pages. The authors tested four conventional machine learning techniques - logistic regression (LR), decision tree

(DT), gradient boosting (GB), and K-nearest neighbor (KNN) - for the classification of text. Lenka V.S.R. Lakshmi [17] have outlined various approaches and discusses supervised algorithms in machine learning classification for sentiment analysis. It highlighted the importance of opinion mining and sentiment analysis for both companies and users. Pruthwik Mishra et al. [20] have demonstrated that machine learning approaches and neural networks can achieve comparable accuracy in sentiment analysis of code-mixed social media compared to systems that rely on hand-crafted features. They utilized a corpus containing code-mixed sentences and their corresponding sentiments. The study showed that char n-gram TF-IDF vectors outperformed word n-grams. Rupal Bhargava et al. [21] have utilized neural networks to identify the sentiment of a tweet. The study's primary contribution was using all possible combinations up to three hidden layers of the three major neural network layers - recurrent neural network (RNN), convolutional neural network (CNN), and long short-term memory (LSTM) - and analyzing the results in detail. They proposed 39 hybrid models for each of the three Indian languages based on the vector space. All the sequential models were tested using uniform parameters for the layers. The proposed approach used a word-level binary classification approach and vector-space model to predict the sentiment of a tweet. Thakare Ketan Lalji et al. [22] have utilized both Lexicon Based Approach and Machine Learning Approach for sentiment analysis and combined the results of both approaches. Lexicon-based approaches usually perform entity-level sentiment analysis, resulting in high precision but low recall. To improve the performance measurements such as Recall, F-Score, Accuracy, a machine learning algorithm was trained using the polarity given by the lexicon-based approach. The authors hypothesized that the accuracy of this approach would increase with the size of the training data. Vikash Nandi et al. [24] have discussed a sentiment analysis of political reviews on Twitter using a hybrid approach of lexical and feature-based classification. The support vector machine was used as a sentiment classifier, and achieved an accuracy of 93%. The lexical and feature-based algorithms were combined to overcome individual drawbacks and improve classification performance. Positive, negative, and neutral sentiments were identified through this approach. Rohini V et al. [25] have conducted domain-based sentiment analysis on Kannada movie reviews using a machine learning algorithm. They compared the results obtained using a direct Kannada dataset with those obtained using a machine-translated English language dataset. Their training dataset comprised about 100 movie reviews from Kannada websites. Overall, the literature review highlights the diversity of methodologies and techniques in sentiment analysis and the ongoing efforts to improve the accuracy and applicability of sentiment analysis to real-world problems.

IV. ANALYSIS

Through the literature review in Section 3, we have identified several existing methodologies in sentiment analysis, including lexicon-based, machine learning-based, and hybrid methods.

Lexicon-based methods use pre-defined dictionaries or lexicons to analyze the sentiment of text. These dictionaries contain lists of words that are labeled as positive, negative, or neutral. The sentiment of a document or sentence is then determined by summing up the scores of the words in the dictionary. While lexicon-based methods are relatively simple and computationally efficient, they suffer from limited accuracy due to the lack of context and the inability to handle sarcasm or irony. **Machine learning-based** methods, on the other hand, rely on training algorithms to classify text as positive, negative, or neutral based on labeled data. These methods can be supervised or unsupervised and can use various machine learning algorithms, such as Naïve Bayes, Support Vector Machines (SVM), and Random Forest. While machine learning-based methods can achieve high accuracy and handle context better than lexicon-based methods, they require a significant amount of labeled data and can be computationally expensive. **Hybrid** methods combine both lexicon-based and machine learning-based approaches to overcome the limitations of each method. For example, some hybrid methods use machine learning algorithms to determine the sentiment of a sentence or document and then refine the results using lexicon-based approaches. Hybrid methods can achieve higher accuracy than either

lexicon-based or machine learning-based methods alone, but they can also be more complex and computationally intensive.

V. CONCLUSION AND FUTURE SCOPE

In conclusion, this review paper tries to provide a comprehensive overview of the existing methodologies used in sentiment analysis. However, despite the progress made in sentiment analysis, there are still several challenges that need to be addressed. One of the major challenges is the lack of labeled data, which hinders the development and evaluation of sentiment analysis models. Another challenge is the ability to handle complex linguistic structures, such as sarcasm, irony, and metaphor, which can significantly affect the sentiment of a text. As the field continues to evolve, there is a need for continued research and innovation to improve the accuracy and applicability of sentiment analysis. In addition to the challenges mentioned above, future research in sentiment analysis can also focus on analyzing multilinguistic mixed words/scripts expressions, such as Odia-English sentences, which are mostly used in state of Odisha. Moreover, the use of emojis in text has become increasingly common in social media, and integrating emoji embeddings in sentiment analysis models can provide a more comprehensive analysis of public opinion. Therefore, we plan to conduct sentiment analysis on Odia-English sentences with emoji embedding in our future research to contribute to the field of sentiment analysis.

VI. REFERENCES

[1] Oscar Araque, Lorenzo Gatti, Jacopo Staiano, Marco Guerini (2022) DepecheMood++: A Bilingual Emotion Lexicon Built Through Simple Yet Powerful Techniques, IEEE Transactions on Affective Computing, Vol 13(1), 496 - 507.

[2] Hamid Bagherim, Md Johirul Islam (2017) Sentiment analysis of twitter data, ArXiv abs/1711,10377.

[3] Ayushi Mitra (2020) Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset), Journal of Ubiquitous Computing and Communication Technologies (UCCT), Vol 2(3), 145-152.

[4] Gaurav Mohanty, Pruthwik Mishra, Radhika Mamidi (2020) Annotated Corpus for Sentiment Analysis in Odia Language, Proceedings of the Twelfth Language Resources and Evaluation Conference, 2788–2795.

[5] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Stede (2011) Lexicon-Based Methods for Sentiment Analysis, Computational Linguistics, 37 (2), 267–307.

[6] Gaurav Mohanty, Abishek Kannan, and Radhika Mamidi (2017) Building a SentiWordNet for Odia. In Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 143–148.

[7] SAMINA KAUSAR, XU HUAHU, WAQAS AHMAD, MUHAMMAD YASIR SHABIR, WAQAS AHMAD (2020) A Sentiment Polarity Categorization Technique for Online Product Reviews, in IEEE Access, Vol 8, 3594-3605

[8] Erik Cambria, Yang Li, Frank Z. Xing, Soujanya Poria, Kenneth Kwok (2020) SenticNet 6: Ensemble Application of Symbolic and Subsymbolic AI for Sentiment Analysis, In Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM '20). Association for Computing Machinery, New York, NY, USA, 105–114.

[9] Kamal Sarkar., Saikat Chakraborty (2015) A Sentiment Analysis System for Indian Language Tweets, International Conference on Mining Intelligence and Knowledge Exploration.

[10] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, Vasudeva Varma (2017) Deep Learning for Hate Speech Detection in Tweets,In Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 759–760.

[11] Dipak Kawade, Dr.Kavita Oza (2017) Sentiment Analysis: Machine Learning Approach, International journal of engineering and technology, Vol 9, 2183-2186.
[12] SANA NABIL, JABER ELBOUHDIDI, MOHAMED YASSIN CHKOURI (2021) SENTIMENT ANALYSIS OF AMAZON'S REVIEWS USING MACHINE LEARNING ALGORITHMS, Journal of Theoretical and Applied Information Technology, Vol 99 (22).
[13] Palak Baid, Apoorva Gupta, Neelam Chaplot (2017) Sentiment Analysis of Movie Reviews using Machine Learning Techniques, International Journal of Computer Applications (0975 – 8887), Vol 179(7).
[14] Mukhtiar Singh, Vishal Goyal, Sahil Raj (2020) Sentiment Analysis of English-Punjabi Code-Mixed Social Media Content, Proceedings of the 17th International Conference on Natural Language Processing: System Demonstrations,24-25.
[15] Sanjib Kumar Sahu, P. Behera, D. P. Mohapatra, R. Balabantaray (2016) Information Retrieval in Web for an Indian Language: An Odia Language Sentimental Analysis Context, International Journal of Circuit Theory and Applications (IJCTA), Vol 9(22), 249-256.
[16] S.K. Mohapatra, Srinivas Prasad, D.K. Bebarta, T.K. Das, K. Srinivasan, Yuh-Chung Hu (2021) Automatic Hate Speech Detection in English-Odia Code Mixed Social Media Data Using Machine Learning Techniques, Applied Sciences, Vol 11(18):8575.
[17] Lenka V.S.R. Lakshmi (2022) Sentimental Analysis Using Machine Learning Algorithms, International Journal of Research Publication and Reviews, Vol 3(11), 534-539.
[18] Ali Hasan, Sana Moin, Ahmad Karim, Shahaboddin Shamshirband (2018) Machine Learning-Based Sentiment Analysis for Twitter Accounts, Mathematical and Computational Applications, Vol 23(1):11.
[19] Hassan Raza, M. Faizan, Ahsan Hamza, Ahmed Mushtaq, Naeem Akhtar (2019) Scientific Text Sentiment Analysis using Machine Learning Techniques, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol 10(12).
[20] Pruthwik Mishra, Prathyusha Danda, Pranav Dhakras (2018), Code-Mixed Sentiment Analysis Using Machine Learning and Neural Network Approaches, ArXiv abs/1808, 03299.
[21] Rupal Bhargava, Shivangi Arora, Yashvardhan Sharma (2019) Neural Network-Based Architecture for Sentiment Analysis in Indian Languages, Journal of Intelligent Systems, Vol 28(3), 361 - 375.
[22] Thakare Ketan Lalji, Sachin N. Deshmukh (2016) Twitter Sentiment Analysis Using Hybrid Approach, International Research Journal of Engineering and Technology (IRJET), Vol 3(6), 2887-2890.
[23] GUIXIAN XU, YUETING MENG, XIAOYU QIU, ZIHENG YU, XU WU (2019) Sentiment Analysis of Comment Texts Based on BiLSTM, in IEEE Access, Vol 7, 51522-51532.
[24] Vikash Nandi, Suyash Agrawal (2016) Political Sentiment Analysis using Hybrid Approach, International Research Journal of Engineering and Technology (IRJET), Vol 3(5), 1621-1627.
[25] Rohini V, Merin Thomas, Latha .C.A (2016) Domain Based Sentiment Analysis in Regional Language-Kannada using Machine Learning Algorithm, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICACT, Vol 4(22).
[26] Chuchu Liu , Fan Fang , Xu Lin, Tie Cai , Xu Tan , Jianguo Liu, Xin Lu (2021) Improving sentiment analysis accuracy with emoji embedding, Journal of Safety Science and Resilience, Vol 2(4), 246-252.
[27] Zhenpeng Chen, Yanbin Cao, Xuan Lu, Qiaozhu Mei, Xuanzhe Liu (2019) SEntiMoji: An Emoji-Powered Learning Approach for Sentiment Analysis in Software Engineering, Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, 841-852.