Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 6037 - 6049
Received 05 March 2021; Accepted 01 April 2021.

Twitter Sentiment Analysis Using Syntactic Action Rule-Based Decision Regression

Dr.P.Kavitha ,

Assistant Professor, Department Of Computer Science,Sri.Sharadha College for Women, Perambalur-621113

Abstract

Sentiment Analysis (SA) is the most effective technique to determine people's sentiments or emotionsand the text's attitude regarding any event.SA comes with the rapid development of the Twitter microblogging service users can post their messages (tweets) to followers (friends) to find analyzing people's feelings and emotions. According to the sentiment analysis is a classification problem of the attitude of the text of the tweet.Due to increasing tweets' increasing from users' opinions, text detection sentiment is a low accuracy process.The proposed Syntactic Action Rule-based Decision Regression (SARDR) algorithm is done to classify the tweets and extract the actionable patterns based on SA of Twitter to solve this problem.Initially collect the tweets dataset then pre-processing done to remove unwanted data and noise then feature extraction for collect the relevant text data information calculate the verb weightages, finally trained into SARDR algorithm prepared to classify the sentiment of text as positive, negative or neutral.The proposed algorithm is the corpus-based Twitter sentiment analysis generated based on the structure of words and the verbs' correct form.The experimental results are shown to improve the classification accuracy compared with previous algorithms.

Keywords: Sentiment Analysis (SA), Twitter, Syntactic Action Rule-based Decision Regression (SARDR), Pre-processing, feature extraction, classification, verb weights.

1. Introduction

Sentiment analysis, to determine the people's opinion, is one of the recently developed techniques. This is, to extract information from a paragraph of text, is the context mining of text.Sentiment analysis is opinion mining to determine the part of the text's opinions and feelings, which means they use natural language processing.Nowadays, Twitter is a microblogging service users post their posts messages called tweets, with 140 characters.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 6037 - 6049
Received 05 March 2021; Accepted 01 April 2021.

Opinions are always very important, such as politics, arts, entertainment, fashion, opinions on work and related fields and all aspects of events held around the world, emotional fields of people every day.Before social networking site is to appear, such as Twitter, Facebook, Instagram and so on, people relied upon a newspaper, television, the magazine of news, and has relied on major events of the world, now on all people use social media has put the information in hand, it does not depend. A single source of information of their news and opinions.

Social media is changing the face of so many people's lives, and they have given a platform to express their opinion in any event or any of the products. Social media will allow all of the topics to express his preference or aversion, such as the character on Twitter.Sentiment can analyze three levels of document level, sentence level, and aspect level.The document level is to provide positive or negative sentiment from the text.Therefore, text that contains the learning of the comparison cannot be considered at the document level.Sentence level analysis is a subjectivity of the classification process that provide aspect and opinions.The aspect level sentence is the overall performance of the feature level. At the entity level, the core task is to identify the following structures and pay attention to direct opinions and feelings at the vertical and horizontal levels.

2. Related work

R. Li et al. [1] explores the Danmaku live commenting make by users on the video screen. It analyses information and classifies the sentiment using Naïve Bayes (NB) for Danmuku reviews, but it is not suitable for Danmuku data.

K. Lu et al. [2] introduces the Chinese micro-blog Sentiment Analysis (SA) based on multiple sentiment dictionaries and semantic rules, classifying the Chinese micro-blog as positive, negative, or neutral. Nevertheless, it doesn't provide high accuracy during the classification process.

Y. Fang et al. [3] uses the multi-strategy sentiment analysis method with semantic fuzziness for consumer reviews. SA contains opinions at the word sentence and document level. It cannot constitute a well of consumer reviews opinion for SA.

N. Al-Twairesh et al. [4] defines the Arabic Twitter sentiment classification using a feature ensemble model of surface and deep features it evaluates the Arabic tweet. It manually features removing process, so it is a complex task during the classification process.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 6037 - 6049
Received 05 March 2021; Accepted 01 April 2021.

J. Wang et al. [5] defines the refining word embedded to provide a low-dimensional vector representation of the sequence of words has been widely used in various natural language processing tasks using the word vector refinement model.

Jiabao Sheng et al. [6] carry out SA public opinion rule-based semantic natural language explanations to train the sentiment classifier. It is a challenging task for analyzing.C. Clavel et al. [7] explores the existing method of Opinion Mining to Human-Agent Interaction for sentiment detection used in socio-affective human agent strategies.

R. Xia et al. [8] explains the public opinion about product reviews in sentiment analysis. It is a natural language to track the public's emotions related to a specific product or topic [8]. Polarity classification is the most classic sentiment analysis process about product reviews, and its purpose is to consider whether it is divided into positive or negative.

M. J. Cobo et al. [9] describes that opinion mining, sentiment analysis, and the consciousness of emotion in advertising have dramatically increased in the past few years [9]. The existing bibliometric analysis algorithm was used to analyze the relationship between opinion mining and sentiment analysis for contextual advertising.Z. Cui et al. [10] explain the new sentiment dictionary based on expression and tone based on barrage sentiment analysis calculate sentiment values [10].

S. Seo et al. [11] explores multi-modal sentiment analysis as a traditional language-based extension method of sentiment analysis using other relevant modal data. Multi-modal sentiment analysis is usually visual, apply text, and the sentiment prediction acoustic representation.

Z. Chen et al. [12] carry out big data has become an important means of supervision of public opinion in the emotional tendency of the mining era included in a large amount of text on the Internet through the natural language processing technology.

K. Mishevet al. [13] demonstrates Internet, online submission opinions and provides you the opportunity for them to share with other investors [14].Online public opinion posts of sentiment analysis can promote the risk perception of investors' investment decisions and securities firms simultaneously [15].

S. Aloufi [16] et al. demonstrate the sports fans will generate a lot of their opinions and feelings reflect tweets about what is happening in various sporting events.In the popular

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 6037 - 6049
Received 05 March 2021; Accepted 01 April 2021.

soccer game, the author focuses on that representation, and the fan is to post their feelings through Twitter based on goal scoring and penalties.

P. Ghasiya et al. [17] focus on the importance of the COVID-19 pandemic of such current has become vital and more important in the health crisis users post their opinion through Twitter.

3. Proposed Methodology

Sentiment Analysis (SA) is people's opinion about any topic to identify the sentiment, attitude, and state of mind from the text to classify them as positive, negative, and neutral. The proposed implementation presents the classifying the sentiment initially to collect tweets raw data, the first process of pre-processing done to remove unwanted data, then check missing values, then feature extraction for collecting relevant data and tweet weightages from the tweets dataset and finally classify the tweet sentence as positive, negative, or neutral.

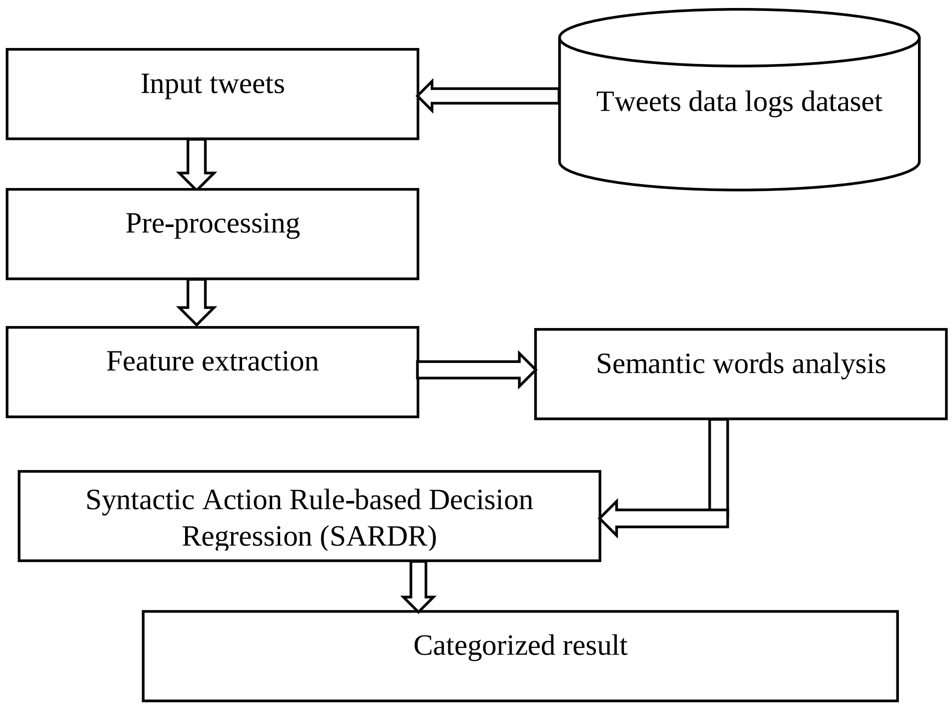

Figure 1: Block diagram for Twitter sentiment analysis

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 6037 - 6049
Received 05 March 2021; Accepted 01 April 2021.

Figure 1 defines the block diagram for Twitter sentiment analysis. The SARDR algorithm can classify the sentiment as positive, negative or neutral from the tweet dataset. Then each semantic words are a synonym of tweets.

3.1 Pre-processing

The tweet contains a lot of data about expressing their opinions in various ways for different users.The pre-processing stage is done to remove unwanted data, noise, URLs trending news with a hashtag, so it removes from the tweets dataset.Tokenization is done by splitting the words from URLs, hashtags and it important to the pre-processing process.Stop word has a high frequency of article appearance. But it is a preposition. There is a need to be removed from the text during the pre-processing done to remove stopwords.

Algorithm steps

Input: Tweets dataset

Output: Pre-processed terms set (P_t)

Begin

Step 1: Initialize the tweets dataset

Step 2: Read tweets dataset , (t=1, 2….n)

Step 3: Read stop word, hashtags from the tweet dataset

Step 4: Remove stop words, hashtags

Step 5: Obtain the pre-processed terms set (P_t)

The above algorithm step provides the tweet dataset's pre-processing for removing noise, unwanted data, stop words, and hashtags.

3.2 Feature extraction

After pre-processing, feature extraction is part of a dimensionality reduction process. The first set of raw data is done reduced to make it easier to segment and manage tweets.In this stage, feature extraction is done to extract all aspects from the processed data set.Feature extraction includes parts of speech, opinion of the words, idioms, synonyms, position of

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 6037 - 6049
Received 05 March 2021; Accepted 01 April 2021.

terms. Parts of speech are like verbs, adjectives of the subjectivity of tweets.Feature extraction is done to identify the sentiment terms of synonyms, verbs from the tweets dataset.

Algorithm steps

Input: pre-processed terms set (P_t)

Output: Feature extraction tweets terms (F_t)

Step 1: Begin

Initialize the P_t terms set

Sentimental text (St) =extract text from P_t

Step 2: Find the keyword terms of analysis

Tweet sentence=$\left(\sum(\text{Relational size (St)}\ Xsplit(.)\right)$

Order by relevant weightages of tweets (Tw)

Step 3: Extract the feature extraction of tweet terms (F_t)

Stop

The above algorithm provides a feature extraction process to extract the best features of the tweets dataset.

3.3 Syntactic Action Rule-based Decision Regression (SARDR)

Syntactic Action Rule has done the text of the possible tweet of the object from the state about decision-making attributed to another state.The proposed is corpus-based sentiment analysis process has been generated based on the structure of words.The proposed SARDR algorithm is done to classifying the accurate result sentence tweet as positive, negative or neutral. Syntactic Action Rule is based if a Tweet sentence has more than two positive words is considered as a positive. If a Tweet sentence has more than two negative words is considered as a negative, otherwise neutral tweet.

Algorithm steps

Input: Feature extraction of tweet terms (F_t)

Output: Classification result

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 6037 - 6049
Received 05 March 2021; Accepted 01 April 2021.

Step 1: Begin

Step 2: Initialize the Feature extraction of tweet terms (F_t)

Read thetweet terms (F_t), (t=1,2…n)

Step 3: Calculate the feature extraction weightages (Tw)

Step 4: Classifying the twitter sentence

If (Tw<0)

Return negative

End if

Else if (Tw>0)

Return positive

Else if

Else

Return Neutral

Step 5:Classification result

Step 6: Stop

The above algorithm presents the classification of the Twitter sentiment analysis process using the SARDR algorithm.

4. Result and discussion

This section describes the proposed implementation the comparing the metrics results are classification accuracy performance, sensitivity, specificity, time complexity. The proposed Syntactic Action Rule-based Decision Regression (SARDR) algorithm compared with other existing methods are Naïve Bayes (NB), Support Vector Machine (SVM)

Table 1: Details of simulation parameters

Parameters	Values

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 6037 - 6049
Received 05 March 2021; Accepted 01 April 2021.

Dataset name	Twitter sentiment analysis(CSV)
Simulation Tool	Anaconda
Simulation Language	Python
Training dataset	1000
Testing dataset	800

Table 1 defines the simulation parameters of the proposed implementation using the anaconda tool in python language.

Table 2: Classification accuracy performance

No of data	SVM in %	NB in %	SARDR in %
100	60	68	74
200	66	72	79
300	71	78	83
400	77	81	88

Table 2 defines the classification accuracy performance of Twitter sentiment analysis. The proposed is to provide higher performance.

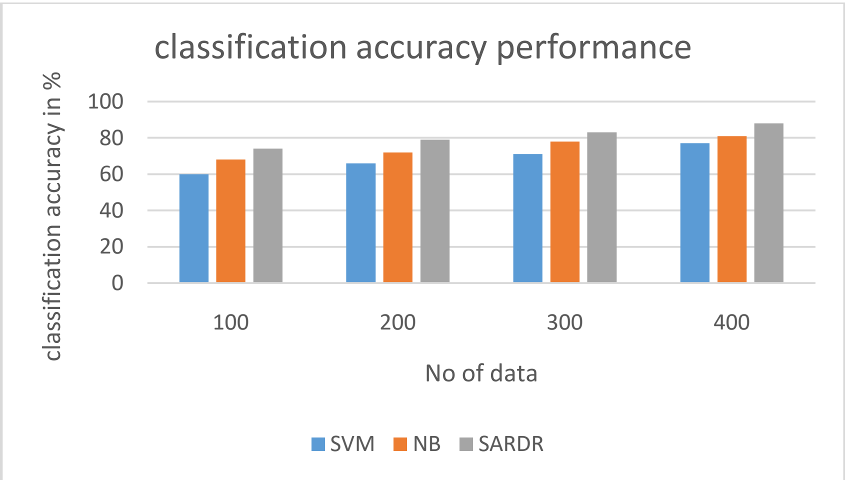

Figure 2: Explorationof classification accuracy

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 6037 - 6049
Received 05 March 2021; Accepted 01 April 2021.

Figure 2 defines the exploration of classification accuracy for the tweet dataset. The proposed Syntactic Action Rule-based Decision Regression (SARDR) algorithm provides a result is 88%. Similarly, the existing methods support Vector Machine (SVM) algorithm classification accuracy performance is 77% Naïve Bayes (NB) classification accuracy performance is 81%.

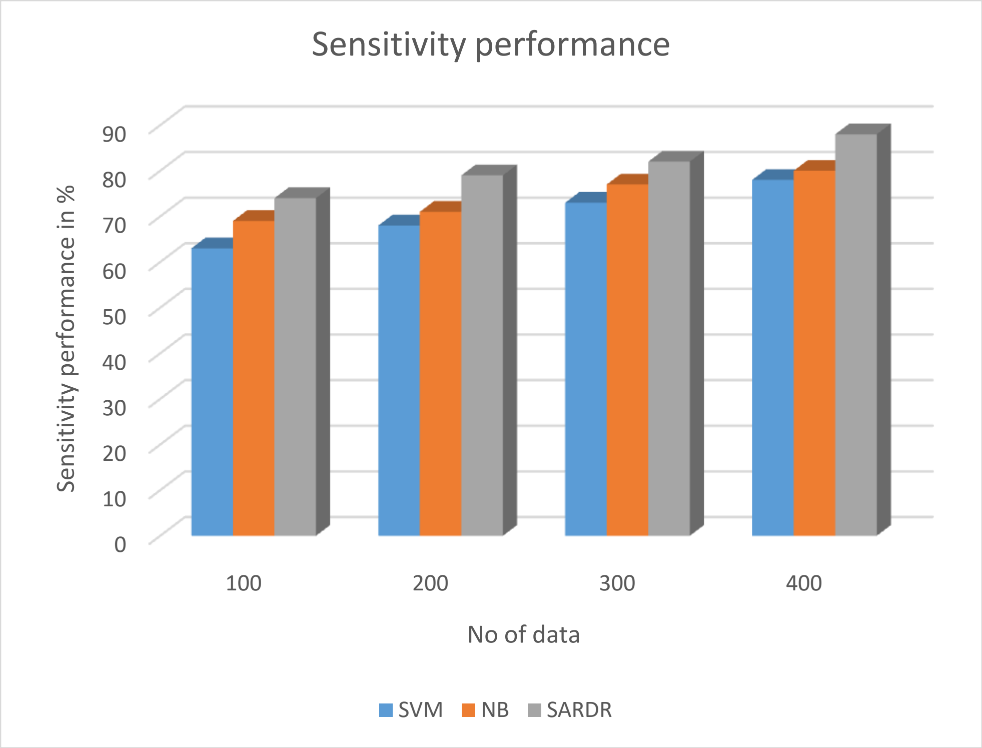

Figure 3: Exploration of Sensitivity performance

Figure 3 defines the exploration of sensitivity performance for tweet dataset,and sensitivity is the test that can identify tweets dataset true positive indicators. The proposed Syntactic Action Rule-based Decision Regression (SARDR) sensitivity performance is 88%, Support Vector Machine (SVM)sensitivity performance is 78%, Naïve Bayes (NB)sensitivity performance is 80%.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 6037 - 6049
Received 05 March 2021; Accepted 01 April 2021.

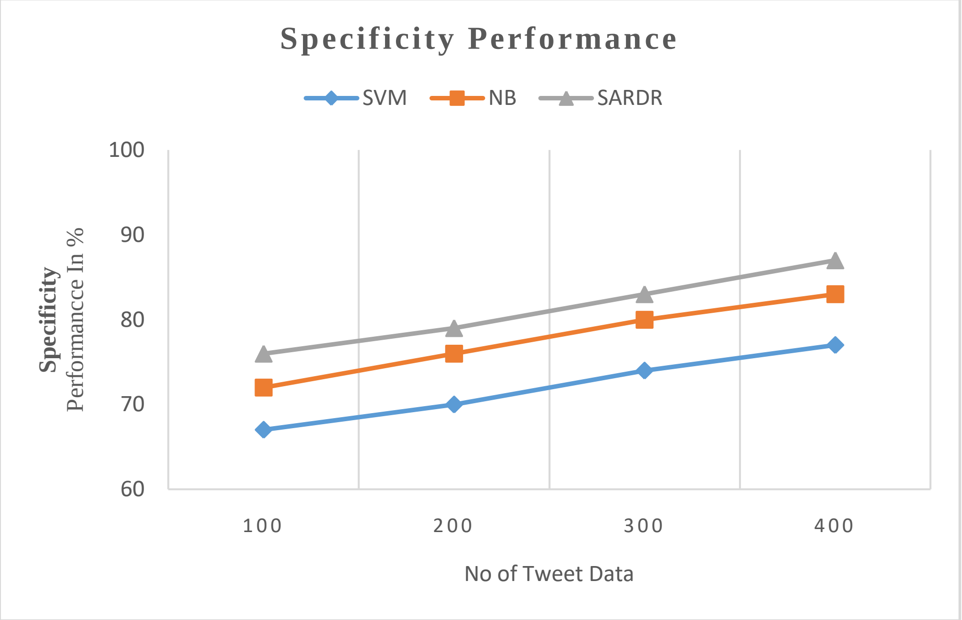

Figure 4: Exploration of SpecificityPerformance

Figure 4 defines the exploration of Specificityperformance of tweets dataset, and sensitivity is the test that can identify tweets dataset true negative indicators. The proposed Syntactic Action Rule-based Decision Regression (SARDR) SpecificityPerformance 87%, Support Vector Machine (SVM) SpecificityPerformance is 77%, NB SpecificityPerformance is 83%.

Table 3 Exploration of Specificity and Sensitivity performance

No of data	**Specificity in %**			**Sensitivity in %**		
	SVM	NB	SARDR	SVM	NB	SARDR
100	67	72	76	63	69	74
200	70	76	79	68	71	79
300	74	80	83	73	77	82
400	77	83	87	78	80	88

Table 3 defines the Sensitivity and Specificity performance compared with different algorithms. The proposed algorithm demonstrates a high result.

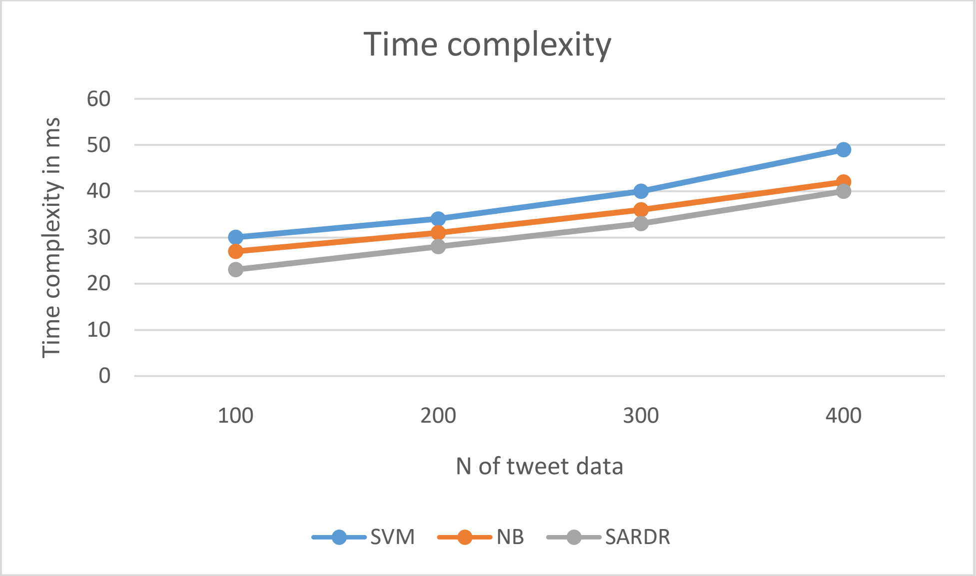

Figure 5: Exploration of time complexity

Figure 5 defines the Exploration of time complexity of tweets dataset with different comparison algorithms. The Proposed Syntactic Action Rule-based Decision Regression (SARDR) time complexity performance 40 ms,Support Vector Machine (SVM)time complexity performance is 49 msNaïve Bayes (NB)time complexity performance is 42 ms.

5. Conclusion

Sentiment analysis is a method to calculate in a given person's view of a piece of text, especially toward the idea is to identify various topics people's tweets are positive, negative or neutral. The proposed Syntactic Action Rule-based Decision Regression (SARDR) algorithm is done to classifying the tweet text. The first process pre-processing stage is done to remove unwanted data, noise, URLs trending news are have a hashtag, so it removes from the tweets dataset. Feature extraction is done reduced to make it easier to segment and manage tweets then. Finally, the proposed SARDR algorithm is done to categorize the result. The proposed SARDR algorithm metrics results are classification accuracy result is 88%,sensitivity performance is 88%,SpecificityPerformance 87%,time complexity performance 40 ms.

Reference

1. Z. Li, R. Li and G. Jin, "Sentiment Analysis of Danmaku Videos Based on Naïve Bayes and Sentiment Dictionary," in IEEE Access, vol. 8, pp. 75073-75084, 2020, doi: 10.1109/ACCESS.2020.2986582.

2. J. Wu, K. Lu, S. Su and S. Wang, "Chinese Micro-Blog Sentiment Analysis Based on Multiple Sentiment Dictionaries and Semantic Rule Sets," in IEEE Access, vol. 7, pp. 183924-183939, 2019, doi: 10.1109/ACCESS.2019.2960655.
3. Y. Fang, H. Tan and J. Zhang, "Multi-Strategy Sentiment Analysis of Consumer Reviews Based on Semantic Fuzziness," in IEEE Access, vol. 6, pp. 20625-20631, 2018, doi: 10.1109/ACCESS.2018.2820025.
4. N. Al-Twairesh and H. Al-Negheimish, "Surface and Deep Features Ensemble for Sentiment Analysis of Arabic Tweets," in IEEE Access, vol. 7, pp. 84122-84131, 2019, doi: 10.1109/ACCESS.2019.2924314.
5. L. Yu, J. Wang, K. R. Lai and X. Zhang, "Refining Word Embeddings Using Intensity Scores for Sentiment Analysis," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 3, pp. 671-681, March 2018, doi: 10.1109/TASLP.2017.2788182.
6. Z. Ke, J. Sheng, Z. Li, W. Silamu and Q. Guo, "Knowledge-Guided Sentiment Analysis Via Learning From Natural Language Explanations," in IEEE Access, vol. 9, pp. 3570-3578, 2021, doi: 10.1109/ACCESS.2020.3048088.
7. C. Clavel and Z. Callejas, "Sentiment Analysis: From Opinion Mining to Human-Agent Interaction," in IEEE Transactions on Affective Computing, vol. 7, no. 1, pp. 74-93, 1 Jan.-March 2016, doi: 10.1109/TAFFC.2015.2444846.
8. R. Xia, F. Xu, C. Zong, Q. Li, Y. Qi and T. Li, "Dual Sentiment Analysis: Considering Two Sides of One Review," in IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 8, pp. 2120-2133, 1 Aug. 2015, doi: 10.1109/TKDE.2015.2407371.
9. P. Sánchez-Núñez, M. J. Cobo, C. D. L. Heras-Pedrosa, J. I. Peláez and E. Herrera-Viedma, "Opinion Mining, Sentiment Analysis and Emotion Understanding in Advertising: A Bibliometric Analysis," in IEEE Access, vol. 8, pp. 134563-134576, 2020, doi: 10.1109/ACCESS.2020.3009482.
10. Z. Cui, Q. Qiu, C. Yin, J. Yu, Z. Wu and A. Deng, "A Barrage Sentiment Analysis Scheme Based on Expression and Tone," in IEEE Access, vol. 7, pp. 180324-180335, 2019, doi: 10.1109/ACCESS.2019.2957279.
11. S. Seo, S. Na and J. Kim, "HMTL: Heterogeneous Modality Transfer Learning for Audio-Visual Sentiment Analysis," in IEEE Access, vol. 8, pp. 140426-140437, 2020, doi: 10.1109/ACCESS.2020.3006563.

12. 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, doi: 10.1109/ACCESS.2019.2927360.
13. K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev and D. Trajanov, "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers," in IEEE Access, vol. 8, pp. 131662-131682, 2020, doi: 10.1109/ACCESS.2020.3009626.
14. F. Ceci, A. L. Gonçalves and R. Weber, "A model for sentiment analysis based on ontology and cases," in IEEE Latin America Transactions, vol. 14, no. 11, pp. 4560-4566, Nov. 2016, doi: 10.1109/TLA.2016.7795829.
15. D. D. Wu, L. Zheng and D. L. Olson, "A Decision Support Approach for Online Stock Forum Sentiment Analysis," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 44, no. 8, pp. 1077-1087, Aug. 2014, doi: 10.1109/TSMC.2013.2295353.
16. S. Aloufi and A. E. Saddik, "Sentiment Identification in Football-Specific Tweets," in IEEE Access, vol. 6, pp. 78609-78621, 2018, doi: 10.1109/ACCESS.2018.2885117.
17. P. Ghasiya and K. Okamura, "Investigating COVID-19 News Across Four Nations: A Topic Modeling and Sentiment Analysis Approach," in IEEE Access, vol. 9, pp. 36645-36656, 2021, doi: 10.1109/ACCESS.2021.3062875.