Neural Computing and Applications (2019) 31:1425–1433
https://doi.org/10.1007/s00521-018-3476-3

S.I. : EMERGING INTELLIGENT ALGORITHMS FOR EDGE-OF-THINGS COMPUTING

Classifying streaming of Twitter data based on sentiment analysis using hybridization

Senthil Murugan Nagarajan[1] · Usha Devi Gandhi[1]

Received: 11 February 2018 / Accepted: 3 April 2018 / Published online: 25 April 2018

Abstract

Twitter is a social media that developed rapidly in today's modern world. As millions of Twitter messages are sent day by day, the value and importance of developing a new technique for detecting spammers become significant. Moreover, legitimate users are affected by means of spams in the form of unwanted URLs, irrelevant messages, etc. Another hot topic of research is sentiment analysis that is based on each tweet sent by the user and opinion mining of the customer reviews. Most commonly natural language processing is used for sentiment analysis. The text is collected from user's tweets by opinion mining and automatic sentiment analysis that are oriented with ternary classifications, such as "positive," "neutral," and "negative." Due to limited size, unstructured nature, misspells, slangs, and abbreviations, it is more challenging for researchers to find sentiments for Twitter data. In this paper, we collected 600 million public tweets using URL-based security tool and feature generation is applied for sentiment analysis. The ternary classification is processed based on preprocessing technique, and the results of tweets sent by the users are obtained. We use a hybridization technique using two optimization algorithms and one machine learning classifier, namely particle swarm optimization and genetic algorithm and decision tree for classification accuracy by sentiment analysis. The results are compared with previous works, and our proposed method shows a better analysis than that of other classifiers.

Keywords Sentiment analysis · Preprocessing · Machine learning · Particle swarm optimization (PSO) · Genetic algorithm (GA) · Decision tree (DT)

1 Introduction

Twitter is the tremendous online social networking site that probably became regular surfing websites by millions of users. Twitter helps its user to express the feelings or thinking regarding some situations of real-world happenings. Various organizations used the sentiment analysis for users to find their mentality and feelings about their products [1]. Twitter analyzes the opinions by using user's posts, blogs, and reviews to help many organizations which are in tie-up with Twitter for enriching the customer opinions, business, politics, and recommender system. The process of examining the intention or polarity of the Twitter user messages is known as sentiment analysis [2, 3]. So, without automatic sentiment analysis and classification, the overall impression of people is hard to identify due to massive sentiment tweets. Several sentiment classifications are emerging, with the works showing interest in tweets.

Figure 1 shows the sentiment analysis process involved in terms of text input, tokenization, stop word filtering, negation handling, stemming, classification, and sentiment class. Using the natural language, the particular issues of text data based on opinions are expressed by authors. The customer insights based on an event with real-time are provided by opinion mining [4, 5]. Twitter data consist of noisy, poor structured sentences and incomplete, ill-formed words, irregular expressions, and non-dictionary terms. The end-user opinions, emotions, and sentiment are expressed for the multiple product reviews using these

✉ Senthil Murugan Nagarajan
senthilmurugan.n@vit.ac.in

Usha Devi Gandhi
ushadevi.g@vit.ac.in

[1] School of Information Technology and Engineering, VIT University, Vellore, India

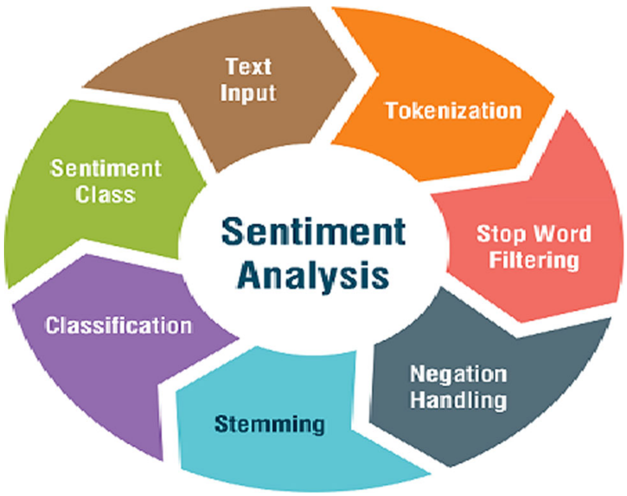

Fig. 1 Process involved in sentiment analysis

websites. Hence, it helps people to better understand and also business-oriented organizations [6, 39]. Several approaches have been proposed for automatic sentiment classification for reviewing documents using natural language processing, machine learning algorithms like Naïve Bayes, support vector machine, maximum entropy, and decision tree. Furthermore, the feature selection methods are combined with these algorithms for predicting polarity of opinions and reviews of each user and emotions are classified as neutral, positive, and negative [7, 40].

The sentiment data are analyzed by using five steps. First step is collecting data using Twitter API and analyzing the particular text data from the dataset and classifying using natural language processing and text analytics as it has a huge amount of data with different slangs and words. The second step includes text preprocessing that analyzes and extracts the particular data and cleans the non-textual contents, emoticons, and numerical values [8]. Thirdly, sentiment analysis is processed detecting emotion from the extracted tweets and its reviews and inspects the opinions. The objective communication is discarded, and individual expressions are retained from the sentences. The fourth step includes the classifier for sentiment classification in terms of "positive" and "negative." Finally, the meaningful information is converted from unstructured text and presented the output which is the key objective of sentiment analysis [9, 10, 11].

In this paper, the collected tweets are processed for sentiment analysis [12]. The above five steps are processed in our dataset by extracting particular text feature and analyzing it. At last, our proposed hybrid algorithm with lexicon based is used for classifying sentiment on extracted data.

2 Related works

Many researchers are working toward the classification of the polarity in Twitter data. As millions of tweets are streaming daily, it is still challenging for researchers to find a better way for classifying polarity and emotions of several users. Many techniques are proposed for sentiment analysis, and classification results are improved, and different steps for preprocessing have been used by many researchers. The limitation of characters in Twitter is 140 that are much used as a resource for opinion mining. Labeling Twitter data is a huge process and hard to complete, so researchers use emoticons for labeling the data. Unfortunately, the noise is created when these data are labeled.

Go et al. [13] used emoticons for labeling the collected Twitter dataset which is 1600,000 tweets. Liu et al. [14] used a combination of emoticons and manual labeling after creating the dataset. Da Silve et al. [15] proposed a method for Twitter sentiment classification using a classifier. For Twitter sentiment classification, meta-level features are used. For polarity and subjectivity classification, words with different dimensions are processed. Kaewpitakkun et al. [16] calculated the scores using an add-on lexicon for obtaining objective and vocabulary words, and also for features, they used a weighting scheme. Saif et al. [17] proposed a method for capturing meanings of words from Twitter in different contexts using sentiment lexicons by adaptation method. Coletta et al. [18] classified tweets with cluster ensemble using an SVM classifier. Lu [19] built a semi-supervised classifier using microblog relations, text similarities, and incorporating social relations. Saif et al. [20] captured patterns of words using his approach for evaluating entity-level and tweet-level sentiment analysis. They improved classification accuracy by incorporating latent semantic relations.

Agarwal et al. [21] found many fixed response features for classification results like bigrams, post tagging, unigrams, ngrams, and hashtags. The concepts like information gain and Chi-square are semantic features of feature selection methods used by Khan et al. [22]. Also this approach carefully analyzed the preprocessing of data that followed supervised machine learning. They used different domains for collecting labeled datasets so that it could not limit machine learning to a particular domain. The different feature sets like information gain with feature frequency and feature presence and cosine similarity with feature presence and feature frequency are used for learning SVM with different datasets in which feature presence is better than the feature frequency. Agarwal et al. [23] showed the results from the datasets using machine learning classifiers and its best, but finding better features is still a challenging task. The features are used as a concept that is named as "Semantic Parser." The minimum redundancy and maximum relevance (mRMR) is used in this approach for selection of features.

Bhadane et al. [24] proposed a method by combining the machine learning approaches with sentiment lexicon for increasing accuracy. Muhammad et al. [25] built a SmartSA system for handling the polarity of words in terms of global and local context. Having more F1 score, it is superior for systems like SVM, Naïve Bayes and baseline lexicons. Addlight and Supreethi [26] analyzed the performance by comparing two algorithms, namely support vector machine and *K*-nearest neighbor, and results showed that SVM classifier outperformed KNN. Saif and He [27] calculated the context of words using the concept of SentiCircles and proved the necessity for improvement in sentiment classification. Jianqiang et al. [28] discussed six different preprocessing techniques and their role in increasing the evaluation measure.

Pang et al. [29] used a movie review dataset and analyzed it using several machine learning algorithms with different feature selection methods. The machine learning techniques are applied directly by using a binary unigram representation of patterns. The occurred number of words is not counted; instead, the presence or absence of words is used for representing trained patterns in the documents. The best performance is reported when the document is applied with machine learning approach using SVM for the review dataset. SVM-based unigram gives the better result. The results give 82.9% accuracy than that of other algorithms. Pang and Lee [30] later studied and proposed subjective sentence separately from the text. The similar subjective label is assumed for two consecutive sentences. For finding the minimum s-t cut in a graph, the task and labels of all sentences are reformulated as subjective and objective. Accuracy around 85% is achieved for Internet Movie Data Base (IMDB).

Mullen and Collier [31] used the semantic orientation of word proposed in previous works for using prior knowledge of documents from the Internet and thesaurus. They achieved 75% of accuracy by using the same dataset used in previous work. Wiebe et al. [32] proposed a method by using the review data for banks, automobiles, travel destinations, and movies for classifying words into positive and negative classes and obtained the overall score for the text. The document is assumed to be positive if it contains more positive terms than negative terms. This type of classification is based on sentence level and document. Each feature opinion cannot be found to be liked or disliked; instead, it only helps to improve the effectiveness of sentiment classification.

Zhang et al. [33] used a decision learning method for customer reviews about the product and feedback for sentiment classification. The discrete-valued target functions are approximated using this decision tree learning method in which the learned function can be represented by using this decision tree. The human readability is improved by re-representing learned trees as sets of if–then rules. This learning method is one of the inductive inference algorithms that are being applied successfully in a wide range of tasks and assess the credit risk of loan applicants by diagnosing medical cases. Chen and Chiu [34] developed a method using a neural network (NN)-based index which effectively classifies sentiment by combining semantic orientation indices and has advantages of machine learning techniques. Tao and Tan [35] evaluated emotional states by using emotional keywords instead of emotional function words. Hu and Liu [36] judged the semantic orientations of adjectives by using antonym sets and adjective synonym sets in WordNet. Ye et al. [37] evaluated reviews for sentiment classification using three supervised machine learning algorithms, namely SVM, Naïve Bayes, and character-based *N*-gram model. At least 80% of accuracy is obtained by these classifiers, and Naïve Bayes gives a lesser than that of other methods.

Zhang et al. [38] combined semantic orientation approach and machine learning into a single framework for sentiment classification by lexicon enhanced method. Particularly, the features for machine learning have added the dimension of words with semantic orientations. The text containing emoticons and opinions are known to be direction-based text which concerned by sentiment analysis. For determining whether the text is subjective or objective, and whether subjective text is positive or negative, it is studied by sentiment classification.

The main contribution of these works concentrated on developing a new model for sentiment classification. Although many machine learning algorithms are proposed for classification of sentiment text, still a hybrid method has to be developed for improving accuracy.

3 Proposed work

In this section, the proposed architecture is shown in Fig. 2. The work is processed in different stages, and it includes first collecting tweets and applying tokenization, then the text is preprocessed and training set data are produced, and finally our hybrid algorithm is applied for sentiment classification. At last, the trained model is applied for streaming of tweets with same preprocessing techniques and acquires classification accuracy. The proposed hybrid algorithms include particle swarm optimization, genetic algorithm, and decision tree. This combination of algorithms gives a better result for classifying sentiment.

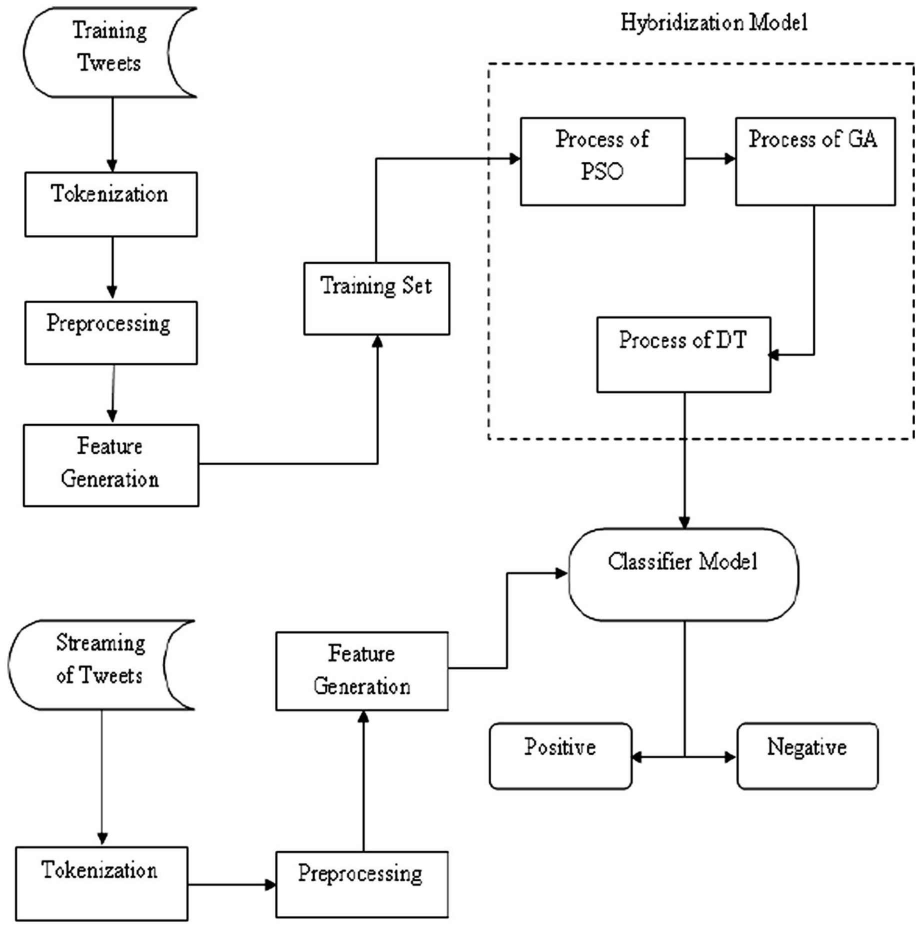

Fig. 2 Overall proposed architecture

(a) Data Collection

The dataset is collected using Twitter Streaming API with URLs. The real-time access is provided by public streaming API with 1% of all public tweets, but no access is granted to protected accounts or direct messages [12, 39]. As tweets are collected in JSON format, each line can be easily parsed as objects. The collected tweets have many attributes and mainly consist of user-based features and tweet-based features [12]. From the proposed work, we focus only on sentiment classification that only acquires text attribute for further process to the proposed work.

(b) Tokenization

Breaking texts into meaningful words, symbols, or phrases is the processes of tokenization. The text mining or parsing process is done by these tokens. Problems can be caused in later phases if any error is made during tokenization. Segmenting text into words is the first step in the majority of text processing. Locating word boundaries helps to form tokenization. Word boundary is known as starting with one word and ending of another word. For any language processing, it is important to do tokenization in the first place. Tokenization can be easy if the words have space. The punctuation symbol is assumed to be white space when tokenization process occurs if there is no existence of any white space. Before doing any kind of processing, the tokens notions are defined. The notion can be methodological or linguistic. An example of tokenization is defined as

Input: I am going for shopping, movie and friends house.

Output:

I	Am	Going	For	Shopping
Movie	And	Friends	House	

(c) Preprocessing of Tweets

Social media sites have many languages which are different from mainstream media found and words in the dictionary. A special "slang," emoticons are employed in social media platforms to emphasize words by repeating some of their letters [40]. Additionally, specific characteristics like markup tweets are employed for languages in Twitter that were reposted by other users with "RT" and also users signs "@" and markup of topics using "#" are used.

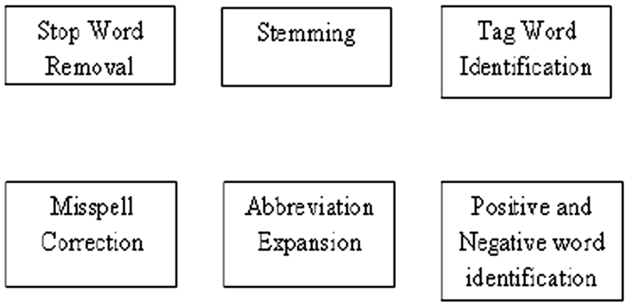

Fig. 3 Preprocessing steps

Figure 3 shows the preprocessing steps that include removal of stop words, abbreviation expansion, correcting misspells in the text, stemming of words, identification of tags, and positive and negative word lists of each tweet.

(d) Feature Generation

After completing the preprocessing step, we move forward for generating features for the preprocessed tweets. In this technique we analyze the frequency of positive and negative words, tag count in the text, analyzing score of positive and negative words, and overall scoring of the words present in it. Here, for learning classifier various features are used and word count is defined as total words present in each tweet after preprocessing is done, tag count is referred as total number @ tags used in each tweet, negative word count is the total number of negative words present in each tweet, positive word count is the total positive words in each tweet, positive score is the total number of positive scores gained after adding the positive adjective, negative score is defined by total number of negative scores obtained when adding each negative adjective, and score is the final total outcome by subtracting negative score for each tweet with positive score.

(e) Hybrid Algorithm for Classification

Training tweets are applied for tokenization and preprocessing and created a training set.

Input: Training set of tweets
Step 1: initialize the population of training set
Step 2: for each particle x in T do
Step 3: If $f(x_t) == f(Pd_t)$ then
$Pd_t = x_t$
End if
Step 4: If $f(x_t) == f(Nd_t)$ then
$Nd_t = x_t$
Else
$Gd_t = (Pd_t, Nd_t)$
End for
Step 5: Stop if condition fulfilled or goto Step 3
Step 6: Update new particles population S
Step 7: while isNotTerminated ( ) do
$P_p(S)$ = P(S).selectPositive;
$P_n(S)$ = P(S).selectNegative;
Step 8: Mutate ($P_p(S)$)
Step 9: Mutate ($P_n(S)$)
Step 10: Evaluate ($P_p(S)$, $P_n(S)$)
Step 11: Built new population ($P_p(S)$, $P_n(S)$)
Step 12: New population as D
Step13: GenDecTree (D, features F)
Step 14: if stopping_condition (D, F) = = True then
Leaf = createNode( )
Leaf.label = Classify (D)
Return Leaf
Step 15: root = createNode( )
Step 16: root.test_codition = findBestSplit(D)
Step 17: Z = {z | z a possible outcome of root.test_condition}
Step 18: for each value $z \in Z$:
a. D_z = {d | root.test_condition (D) = z and $d \in D$}
b. Child = TreeGrowth (D_z, F)
c. Add child to the root and label the edge (root -> positive or negative) as z
Step 19: return root

The above algorithm shows the flow of our proposed hybrid method which is a combination of PSO, GA, and DT.

4 Results and discussion

In this section, the tweets collected are sent for text preprocessing and separating the words into tokens using tokenization method; later, it extracted that features with

Table 1 Ternary classification of tweets

	Accuracy	Precision	Recall	*F*-measure
Positive	0.935	0.894	0.832	0.896
Negative	0.893	0.873	0.865	0.863
Neutral	0.884	0.859	0.814	0.849
Overall	0.904	0.875	0.837	0.869

Table 2 Confusion matrix of classified tweets

Class	Classified to be		
	Positive	Negative	Neutral
Positive	6648	1152	976
Negative	1478	6135	1786
Neutral	1765	1874	4786

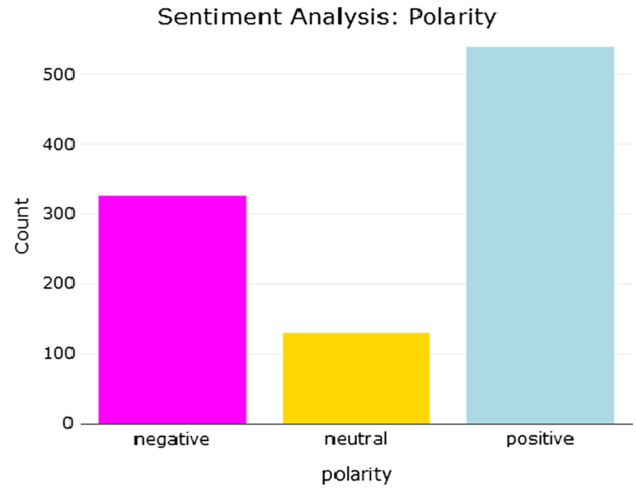

Fig. 4 Polarity based on proposed method

developing the scores and polarity of the tweets. Using our proposed classifier, we evaluate the performance and effectiveness in terms of accuracy, precision, recall, and *F*-measure.

- Accuracy refers the correctness of classification obtained overall. The ratio is measured between the correctly classified instances with a total number of instances.
- Precision refers to the fraction obtained by correctly classified tweets for the given sentiment words, of the total number of tweets classified that are present in this sentiment.
- Recall that the definition of KPI is not different from the accuracy of that particular sentiment.
- *F*-measure is defined as

$$F\text{-Measure} = 2\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

(a) Ternary Classification

Our proposed hybrid algorithm is used to classify the emotion and polarity of the tweets. We grouped these tweets into three classes, namely "positive," "neutral," and "negative." The former class contains tweets from the classes "happiness," "fun," and "love," while later the tweets from the classes have "hate," "anger," and "sadness." Table 1 shows the classification result obtained by proposed hybrid algorithm that combined particle swarm optimization, genetic algorithm, and decision tree. The overall accuracy of our classifier is about 90.4%, and positive accuracy is about 89.3%, and negative is about 93.5%. Noticeably, the overall precision of positive and negative tweets is 87.5%. The recall measure obtained is about 83.7% for positive and negative, and F-measure obtained is 86.9% for both polarities. Table 2 shows the overall confusion matrix that predicts positive tweet correctly as positive and negative tweet correctly as negative. The neutral is classified with an accuracy of 89.3% and *F*-measure with 86.3% in classification of tweets. The overall results obtained are better than that of other classifiers.

Figure 4 shows the classification of the polarity of each tweet present in the collected dataset. It shows how our proposed hybrid algorithm classifies the sentiment of each text present in the tweet. The count represents the increase in dataset from 100 to 500. It also shows that positive, negative, and neutral classes present in this whole dataset.

(b) Classification Based on Multiple-Class

In this subsection, we classify emotion classes based on "anger," "disgust," "fear," "joy," "sadness," "surprise," and "N.A.". Figure 5 shows the emotions classified for this each class using our proposed approach. The analysis shows that the class "anger" has more datasets.

(c) Comparison with other approaches

So far, our proposed works show the overall performance based on accuracy, precision, recall, and *F*-measure. Table 3 shows the comparison of our hybrid method with other classifiers and other existing hybrid approaches. It shows that the performance and accuracy of the proposed work is better when compared to other classifiers and evolutionary algorithms like particle swarm optimization (PSO), genetic algorithm (GA), decision tree (DT), support vector machine (SVM), *K*-nearest neighbor (KNN), and hybrid method of SVM + KNN.

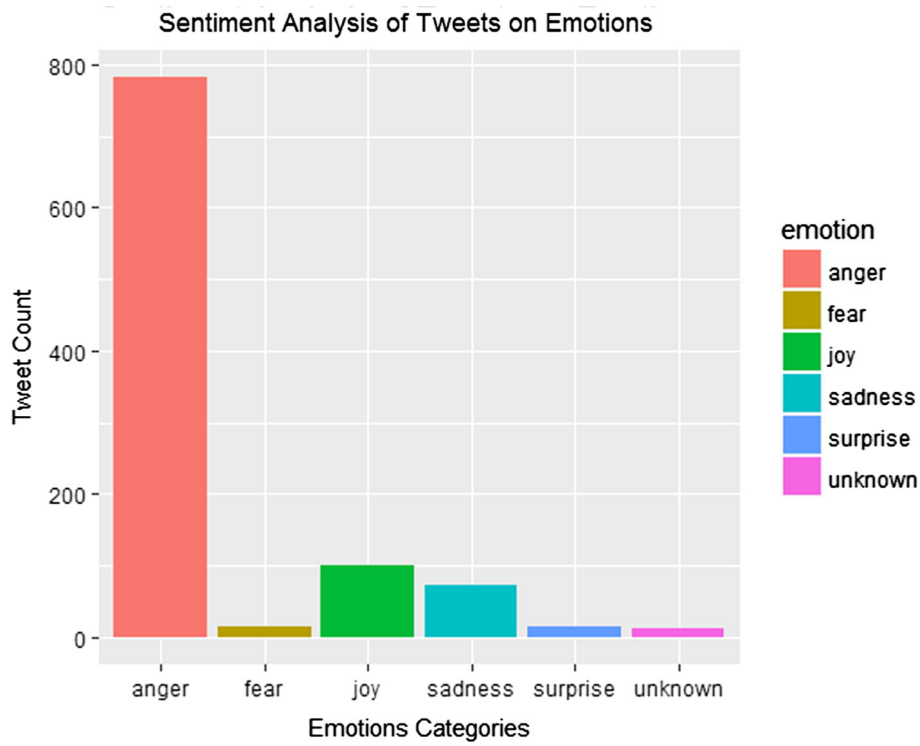

Fig. 5 Classified emotions of proposed work

Table 3 Existing approach versus proposed work

Classifiers	Accuracy	Precision	Recall	*F*-measure
KNN	67	70.5	69.3	67.9
SVM	68	69	68.1	68.7
SVM + KNN	76	68.45	68.14	77.56
DT	80	81.4	81.4	80.95
GA	86	87.3	87.9	87.59
PSO	88	88.7	89.1	86.78
Proposed work	90	91.5	91.7	91.4

Figure 6 shows the overall accuracy of existing classifiers compared with our proposed hybrid model. The KNN has the lowest accuracy of about 67% than any other classifiers. The SVM classifier has 68% of accuracy that is in the second least performance compared with other classifiers. The existing hybrid model which is a combination of SVM and KNN has an improvement in accuracy than the separate analysis of the classifiers, and it has 76% of accuracy. Other approaches like DT, GA, and PSO have better improvement than the previous existing works that have the accuracy rate of 80, 86, and 88%. Furthermore, our proposed hybrid model shows a better accuracy and performance than other existing approaches that have 90.5% accuracy. Table 3 shows the comparison results of overall accuracy, precision, recall, and *F*-measure of each classifier with proposed work.

5 Conclusion

In this paper, a new hybrid method is proposed for classification of sentiment analysis. The combination of PSO, GA, and DT has proved to have better performance when compared to other existing algorithms. This optimization technique with machine learning classifier gives an accuracy of over 90% for the classification of sentiment tweets into "positive," "negative," and "neutral" classes. The proposed model is trained by using two main stages, namely preprocessing and feature generation. The overall accuracy obtained by our proposed work is over 90% compared with other machine learning classifiers and optimization techniques. In future work, the modification of our proposed work with other classifiers could be

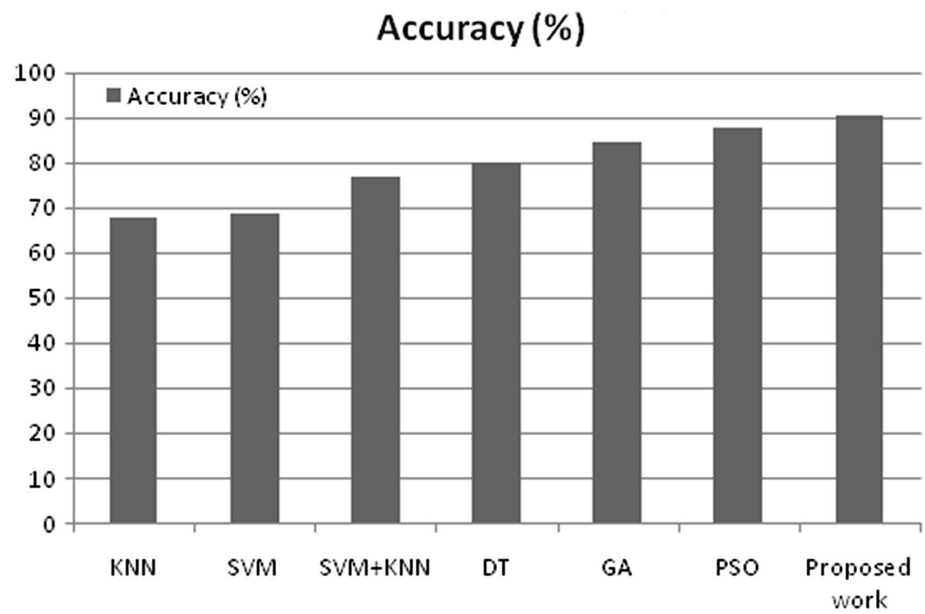

Fig. 6 Overall comparison of accuracy with proposed work

extended and changing the optimization problem could be a major step toward sentiment classification.

Compliance with ethical standards

Conflict of interest This statement is to certify that all authors have seen and approved the manuscript being submitted. We warrant that the article is the author's original work. We warrant that the article has not received prior publications and is not under consideration for publication elsewhere. On behalf of all co-authors the corresponding author shall bear full responsibility for the submission. The author(s) declare that there is no conflict of interest.

References

1. Somani A, Suman U (2011) Counter measures against evolving search engine spamming techniques. In: 2011 3rd international conference on electronics computer technology (ICECT), vol 6, pp 214–217
2. Varatharajan R, Manogaran G, Priyan MK, Sundarasekar R (2017) Wearable sensor devices for early detection of Alzheimer disease using dynamic time warping algorithm. Clust Comput. https://doi.org/10.1007/s10586-017-0977-2
3. Varatharajan R, Manogaran G, Priyan MK, Balaş VE, Barna C (2017) Visual analysis of geospatial habitat suitability model based on inverse distance weighting with paired comparison analysis. Multimed Tools Appl. https://doi.org/10.1007/s11042-017-4768-9
4. Balan EV, Priyan MK, Gokulnath C, Devi GU (2015) Fuzzy based intrusion detection systems in MANET. Proc Comput Sci 50:109–114
5. Manogaran G, Varatharajan R, Priyan MK (2018) Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system. Multimed Tools Appl 77(4):4379–4399
6. Devi GU, Balan EV, Priyan MK, Gokulnath C (2015) Mutual authentication scheme for IoT application. Indian J Sci Technol 8(26):15
7. Priyan MK, Devi GU (2017) Energy efficient node selection algorithm based on node performance index and random waypoint mobility model in internet of vehicles. Clust Comput. https://doi.org/10.1007/s10586-017-0998-x
8. Varatharajan R, Manogaran G, Priyan MK (2017) A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing. Multimed Tools Appl. https://doi.org/10.1007/s11042-017-5318-1
9. Devi GU, Priyan MK, Balan EV, Nath CG, Chandrasekhar M (2015) Detection of DDoS attack using optimized hop count filtering technique. Indian J Sci Technol 8(26):4
10. Gokulnath C, Priyan MK, Balan EV, Prabha KR, Jeyanthi R (2015) Preservation of privacy in data mining by using PCA based perturbation technique. In: 2015 International conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM). IEEE, pp 202–206
11. Kumar PM, Gandhi U, Varatharajan R, Manogaran G, Jidhesh R, Vadivel T (2017) Intelligent face recognition and navigation system using neural learning for smart security in internet of things. Clust Comput. https://doi.org/10.1007/s10586-017-1323-4
12. Manogaran G, Varatharajan R, Lopez D, Kumar PM, Sundarasekar R, Thota C (2017) A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system. Future Gener Comput Syst 80:1
13. Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, vol 1, no 12
14. Liu KL, Li WJ, Guo M (2012) Emoticon smoothed language models for Twitter sentiment analysis. In: Aaai
15. Da Silva NF, Hruschka ER, Hruschka ER Jr (2014) Tweet sentiment analysis with classifier ensembles. Decis Support Syst 66:170–179
16. Kaewpitakkun Y, Shirai K, Mohd M (2014) Sentiment lexicon interpolation and polarity estimation of objective and out-of-vocabulary words to improve sentiment classification on microblogging. In: Proceedings of the 28th Pacific Asia conference on language, information and computing

17. Saif H, He Y, Fernandez M, Alani H (2014) Adapting sentiment lexicons using contextual semantics for sentiment analysis of twitter. In: Presutti V, Blomqvist E, Troncy R, Sack H, Papadakis I, Tordai A (eds) The semantic web: ESWC 2014 satellite events. ESWC 2014. Lecture notes in computer science, vol 8798. Springer, Cham, pp 54–63
18. Coletta LFS, da Silva NFF, Hruschka ER, Hruschka ER (2014) Combining classification and clustering for tweet sentiment analysis. In: 2014 Brazilian conference on intelligent systems (BRACIS), pp 210–215
19. Lu TJ (2015) Semi-supervised microblog sentiment analysis using social relation and text similarity. In: 2015 International conference on big data and smart computing (BigComp), pp 194–201
20. Saif H, He Y, Fernandez M, Alani H (2014) Semantic patterns for sentiment analysis of twitter. In: Mika P et al (eds) The semantic web – ISWC 2014. ISWC 2014. Lecture notes in computer science, vol 8797. Springer, Cham, pp 324–340
21. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R (2011) Sentiment analysis of Twitter data. In: Proceedings of the workshop on languages in social media. Association for Computational Linguistics, pp 30–38
22. Khan FH, Qamar U, Bashir S (2017) A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet. Knowl Inf Syst 51(3):851–872
23. Agarwal B, Poria S, Mittal N, Gelbukh A, Hussain A (2015) Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach. Cognit Comput 7(4):487–499
24. Bhadane C, Dalal H, Doshi H (2015) Sentiment analysis: measuring opinions. Proc Comput Sci 45:808–814
25. Muhammad A, Wiratunga N, Lothian R (2016) Contextual sentiment analysis for social media genres. Knowl Based Syst 108:92–101
26. Mukwazvure A, Supreethi KP (2015) A hybrid approach to sentiment analysis of news comments. In: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), pp 1–6
27. Saif H, He Y, Fernandez M, Alani H (2016) Contextual semantics for sentiment analysis of Twitter. Inf Process Manage 52(1):5–19
28. Jianqiang Z, Xiaolin G (2017) Comparison research on text preprocessing methods on Twitter sentiment analysis. IEEE Access 5:2870–2879
29. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on empirical methods in natural language processing, vol 10, pp 79–86
30. Pang B, Lee L (2004) A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd annual meeting on Association for Computational Linguistics, p 271
31. Mullen T, Collier N (2004) Sentiment analysis using support vector machines with diverse information sources. In: Proceedings of the 2004 conference on empirical methods in natural language processing
32. Wiebe J, Wilson T, Bruce R, Bell M, Martin M (2004) Learning subjective language. Comput Linguist 30(3):277–308
33. Zhang C, Zuo W, Peng T, He F (2008) Sentiment classification for Chinese reviews using machine learning methods based on string kernel. In: Third international conference on convergence and hybrid information technology, ICCIT'08, vol 2, pp 909–914
34. Chen LS, Chiu HJ (2009) Developing a neural network based index for sentiment classification. In: Proceedings of the international multiconference of engineers and computer scientists, vol 1, pp 18–20
35. Tao J, Tan T (2004) Emotional Chinese talking head system. In: Proceedings of the 6th international conference on multimodal interfaces, pp 273–280
36. Hu M, Liu B (2004). Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp 168–177
37. Ye Q, Zhang Z, Law R (2009) Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Syst Appl 36(3):6527–6535
38. Zhang Y, Dang Y, Chen H (2011) Gender classification for web forums. IEEE Trans Syst Man Cybernet Part A Syst Hum 41(4):668–677
39. Manogaran CTG, Priyan M (2017) Centralized fog computing security platform for IoT and cloud in healthcare system. In: Exploring the convergence of big data and the internet of things, p 141, IGI Global
40. Balan EV, Priyan MK, Devi GU (2015) Hybrid architecture with misuse and anomaly detection techniques for wireless networks. In: 2015 International conference on communications and signal processing (ICCSP). IEEE, pp 0185–0189