Sentimental Analysis using Capsule Network with Gravitational Search Algorithm

V. Diviya Prabha* and R. Rathipriya

Department of Computer Science Periyar University, Salem, India
E-mail: diviyaprabha7@gmail.com; rathipriyar@gmail.com
*Corresponding Author

Abstract

Day by day the recent development of communication and the data on the web is increasing tremendously. Moreover, the use of social media among people to express their opinion has greatly increased. Therefore, analyzing this textual data using sentimental analysis techniques can be very helpful in capturing and categorizing people's opinions. This work aims to propose an algorithm which is combination of Capsule Network (CN) with Gravitational Search Algorithm (GSA) to analyze people's sentiments from twitter data. In text data mining, CN works to an excessive extent for sentiment analysis compared with other models. The performance of the proposed approach is studied using existing benchmark datasets and COVID-19 twitter posts. The results showed that the proposed approach could automatically classify the sentiments with high performance. It works better compared to other algorithms and results also encourage further research.

Keywords: Deep learning, capsule network, machine learning, sentimental classification.

1 Introduction

Recent technologies allow people to exchange their opinion through online community. Opinions shared by users in online are an important clue to understand the user's behavior. Analyzing user's opinion and extracting useful information for sentimental analysis is significant in understanding current trend/pulse of the common public. Twitter is most popular social media helps to extract user's information. It allows users to post their opinion and emotions. Sentiments help to categorized people opinions based on tweets posted in social media. This kind of study is necessary to convey people positive or negative opinion on certain topic. Survey states that every minute tweets posted in twitter are exceeding in lakhs (Ekenga and McElwain, 2018). Twitter also allows user to interact with other users by sharing some hash tags. The information available in tweets post are unstructured format and generally difficult to analyze. In this sense, a techniques is needed to analyze tweets and extract tweets in a meaning full way which is not possible by traditional techniques (Ali Hasan et al., 2018). The technique must be scalable and helps to identify sentiment of users in a reliable way. The machine learning techniques are doing better to classify sentiments. But still it is not sufficient when it comes to tremendous amount of data. There are certain machine learning techniques work proves better for text classification. In this work, (Castro et al., 2017) proposed a hybrid machine learning based sentiment classification for converting the urdu to English tweets and identify their sentiment (Meng et al., 2020) classify public sentiment and collect disaster related data in online communities. Analyzing the public opinion and satisfying their needs (Burnap et al., 2013) detects tension such as tension, aggressive and high tension hash tags tweets and classify their sentiments using machine learning approach.

To overcome the limitations nature-inspired optimization are investigated for better sentiment analysis. However, the swarm intelligence takes inspiration from nature algorithm like Gravitational Search Algorithm (GSA) (Prabha and Rathipriya, 2013) is used to search space in a minimal architecture. It works on the concept of force and mass (Rashedi and Nezamabadipour, 2009) every words in the text attracts the other related words here particle act as agents. Search processes of each word are optimized to find the best capsule to classify the text. It also optimizes the values to find the best solution for identifying sentiments. Methods used to categorize tweets are discussed (Gohil and Vuik, 2018) in this methodology. Capsule Network (CN) (Kim and Jang, 2020) improves the performance in text analysis. It is an efficient method to understand text data and converting word to vector.

The contribution in this paper can be précised as follows: (1) Design architecture of CN-GSA used improve the classification accuracy of both long and short twitter post about COVID-19, (2) It helps to classify tweets as five classes such as strong positive, positive, strong negative, negative and neutral, (3) Comparing the efficiency of the proposed model with machine learning and other deep learning model, (4) The proposed model improves in terms of precision, recall, F-measure and accuracy, (5) The model improves the accuracy for different kind of sentiment classification datasets.

In this work, the proposed Capsule Network (CN) with GSA is used to identify important key terms from sentence and optimize from large space text to minimized this will consequently improve the classification performance. This paper is organized as follow. The detail descriptions of proposed model are discussed in Section 3. Comparative analysis of CN with other models is represented in section 4. The results of experiments show that CN outperforms well and good compared with other methods. Finally, Section 5 concludes the paper.

2 Related Work

Over the past decade, the researchers focused on sentimental analysis, especially using machine learning techniques such as (Silva et al., 2018) SVM techniques for text retrieval. The main drawback of this approach is certain patterns are manually given as input which will not fit while using different dataset at same time. It is insufficient for huge dataset also (Tripathy, 2016) discusses the usage of machine leaning techniques for sentiments classification. Although machine learning techniques succeed in performing sentimental classification they are limited to the performance for varied datasets. To overcome this drawback deep learning techniques is solution to the problem.

Deep Learning is a kind of machine learning techniques has significant importance to classify tweets. They automatically extract patterns and provide better performance for varied datasets (Conneau et al., 2016) developed a very deep CNN showing the performance of Deep Learning model for text data (Changyi et al., 2020) also developed a novel approach sparse binary optimization to reduce the data from high dimensional to low dimensional space reducing scarcity of space (Mehreen, 2019) classifying sentiments using forest optimization algorithm to improves the classifier accuracy and helps to extract meaningful words from large context of text. Machine learning analysis of sentiments are also beneficial to limited number of posts

(Yujiao and Fleyeh, 2018). Optimizing sentiments (Jaspreet and Gurvinder, 2017) to identify relevant words and classify using classifiers (Mosa, 2019) discusses the usage of Gravitational Search Algorithm (GSA) for mining large amount of text to maximize the significant sentiments and minimizes the redundancy of terms (Zhao et al., 2019) uses an ensemble method of capsule networks is given as input to the polling layer to improve the better accuracy. The combination of RNN and CNN (Lai, 2015) proposed a new model for text classification. Here, word-embedding is used for both given word and passed through max polling layer (Cheng et al., 2016) it expands the work of LSTM in replace of memory in single cell with possible neural network. CN is better to classifier (Jae young Kim, Sion Jang, 2020) for text and improves the performance for varied datasets. Gender classification (Xian Zhong, Jinhang Liu, Shuqin Chen, 2020) and emotional tags identification achieves better result. Hybrid method of CN proposed a good algorithm for sentiment classification (Yongping Du, Xiaozheng Zhao , 2019). Though, deep learning model perform better the overall meaning of the text in the sentence must not differ. If a word in a sentence is taken as misread information if literally affects the sentiment of post. The sentiment mainly depends on the word that gives meaning to the sentence. Therefore, to understand the meaning of the words CN with GSA is developed. This proposed model performance best for wide-ranging datasets. Since, the extraction of words is important role for sentimental analysis.

The main objective of this paper is to improve the performance of sentiment classification by extracting the relevant words. The proposed model overcomes different data pre-processing techniques and makes it suitable for converting the words to vectors encoding. The motivation of CN is to identify words without choosing the sentiment appropriately (Hu and Cui, 2017). These studies also suggest that CN act as a core network for sentiment classification (Pang et al., 2002) discuss better achievement of using CN to identify sentiments.

3 Proposed Model

3.1 Data Collection

Six different types of datasets are used to test the proposed model. The first three benchmark dataset are collected form UCI repository and next three COVID-19 dataset is collected form twitter. First three dataset consist of imdb dataset, yelp dataset and amazon dataset collected from URL

Table 1 Dataset description

Dataset Name	Class	Size
IMDB dataset	2	1000
Yelp dataset	2	1000
Amazon dataset	2	748
COVID_1	5	6200
COVID_2	5	10,969
COVID_3	5	5200

https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences labeled with binary sentiments. The other dataset are user generated dataset collected from Twitter streaming API (Trupthi and Pabboju, 2017) tweets are extracted using the keyword about COVID-19 posted in English language. The extracted unstructured data is converted into structured data. Furthermore, the data is stored in machine readable format in ".CSV" file. Table 1 shows the description about the datasets and classification classes.

3.2 Data Preprocessing

Data preprocessing was carried out to minimize the unnecessary keywords, noise data and special characters, symbols etc., the purpose of preprocessing is to handle data appropriately. Data stemming techniques was used to remove affixes, hyphen and not important keywords. Lemmatization was considered to analyze group of words into single meaning words. The sentimental classification for first there dataset is fixed label. For, COVID-19 datasets they are classified based on the sentiments polarity. It is classified into strong positive, positive, negative, strong negative and neutral. The sentiments belong to strong positive label if text of polarity rate lies in between 0.5 to 1. Similarly, sentiments are positive if their polarity rate is less than 0.5, sentiments are strong negative if the polarity rate is greater than -0.5, sentiments are negative if the polarity rate occur in between -0.5 to -0.1 and finally sentiment belong to neutral if polarity rate is zero.

3.3 Proposed Algorithm

The CN architecture of the proposed algorithm is represented in Figure 1. The CN works on the concept of grouping the neurons actively represent vector and finally results in capsule layer (Hugo et al., 2020). The probability of detecting the correct feature is based on the vector representation and
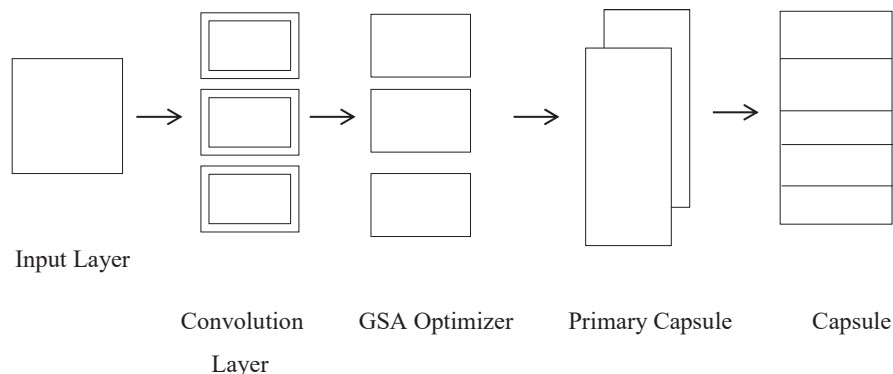
Figure 1   CN architecture.

Input Layer

Convolution Layer   GSA Optimizer   Primary Capsule   Capsule

feature representation. CN model also contains convolution layer used to extract important features from the input vector. Among all the text vectors it works on the function of gravity every particle attracts the other particle with high attraction value to make a better capsule. Representing words in sequential and minimizes the parameter instantiations. GSA optimizer is initialized to CN model for better identification of words. It is used to minimize the distance between the words and helps to choose the best words. It incorporated to capture words and makes the classification easier. The maximization fitness function is used for the purpose of choosing top words. However, final capsule consist of five classes.

Figure 2 explains in detail about manipulation of sample sentence using proposed model. The sentence is "Help prevent spread of COVID 19 by maintain social distancing" as passed input to Input layer denoted by w_1, $w_2 \ldots w_n$ for each word. The sentence is divided into token or words containing of n dimensional value. The tokens of each word are converted to vector is the process of convolutional layer by extracting important features. The convolution layers are carried out to GSA optimizer to find the optimal features to form a capsule. This context information is sufficient to make prediction. The value in the second capsule is high which helps to classify sentiment as strong positive.

The convolution layer is represented as follow let w □ R to be the matrix of input layer with t tokens. In equation 1, W denotes weight belongs to R in convolution layer to feature generation.

$$F_i = f(W * w_i, w_2 \ldots w_n) \tag{1}$$
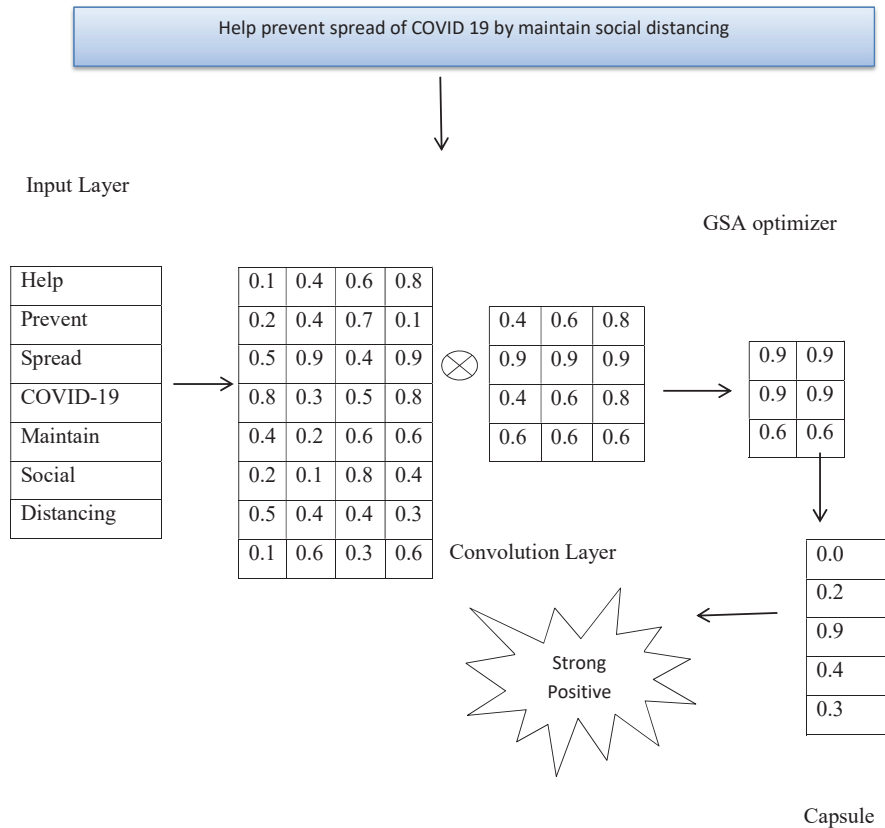
Figure 2 CN-GSA architecture.

The features are extracted based on the weight of word passed to each word in the sentence. Weights given to word dependence on sentiment polarity of word present in the sentence.

Algorithm:

Step 1: Procedure: Input Text $(T_1, T_2 \dots T_n)$ to be tokens $(t_1, t_2 \dots t_m)$

Step 2: Initialize the vectors to each word converting text to vectors $(V_1 \dots V_m)$

Step 3: Apply convolution layer to identify the importance of features with transformation of capsule:

$$\widehat{u_{ij}} = W_i V_i \qquad (2)$$

Step 4: Tokens are considered as agents in GSA each agents are attracted by other agents by force and mass.

$$\text{Agents} = u_i \ldots u_{ij} \tag{3}$$

Step 5: The total force that act as agent's t_i can be weighted randomly.

$$TF = \sum_{i=1}^{n} rand(TF_i) \tag{4}$$

Step 6: Find the best and worst value

$$fitness_{value(i)} = Max(TF)) \tag{5}$$

$$Mass = \frac{fitness_{value(i)} - worst(i)}{best(i) - worst(i)} \tag{6}$$

$$\widehat{M_{ij}} = W_{ij} * Mass \tag{7}$$

Step 7: Applying squashing mass function to achieve the capsule

$$C_j = \frac{\left\|\widehat{M_{ij}}\right\|^2}{1 + \left\|\widehat{M_{ij}}\right\|^2} \frac{\widehat{M_{ij}}}{\|M_{ij}\|} \tag{8}$$

Step 8: Apply the training model and compile the metrics accuracy

The Equation (2) represents the word to vector transformation by extracting important features. The agents in Equation (3) represent the number of tokens in a sentence. Equation (4) describes the process of fitness function calculated with total force value in Equations (6) and (7). Mass function was calculated based on the weighted text value. Squash function is calculated and denoted in Equation (7). Finally the obtained value is divided into train and test function. The performance of the algorithm depends on the accuracy level. Finally this paper suggests text data processing through the convolution layers and GSA layer is successful. As GSA layer has fast converging characteristics it takes less time for powerful computation. This shows the proposed algorithm is able to run in any kind of text dataset without any domain problem. Therefore, this model achieves good results than the existing techniques.

4 Experimental Analysis

In this section, discuss the results of experiments and performance of the proposed model. In Table 2, the Support Vector Classifier (SVC) achieved the best F1 scores. For recall measure Random Forest (RF) and Logistic Regression (LR) achieved best results for negative sentiments. The macro average of Logistic Regression (LR) and Support Vector Classifier (SVC) remains the same at 82%. For weighted average (WA) the SVC achieved best result. The K-Neighbor Classifier (KN) obtains low macro average (MA) and WA compared to other classifier.

As observed from Table 3, for Yelp Dataset sentiment classification is represented. SVC classifier achieved best result in precision, recall and F1 score. RF classifier in MA is minimum compared to other approaches. Performance of SVC is better in both MA and WA values. Table 4 shows SVC performs better for Amazon dataset as it achieved 76% of precision and 77% accuracy for WA. Similarly, 75 % value for F1 in MA and WA.

Table 2   IMDB dataset sentiment classification using machine learning techniques

Classifier	Class	Precision	Recall	F1
RF	0	0.72	0.91	0.80
	1	0.89	0.67	0.77
	MA	0.80	0.79	0.78
	WA	0.81	0.79	0.78
LR	0	0.74	0.92	0.82
	1	0.90	0.70	0.79
	MA	0.82	0.81	0.80
	WA	0.82	0.81	0.80
KN	0	0.83	0.74	0.78
	1	0.78	0.86	0.82
	MA	0.80	0.80	0.80
	WA	0.80	0.80	0.80
SVC	0	0.77	0.88	0.82
	1	0.87	0.76	0.81
	MA	0.82	0.82	0.81
	WA	0.82	0.82	0.81

Table 3 Yelp dataset sentiment classification using machine learning techniques

Classifier	Class	Precision	Recall	F1-Score
RF	0	0.64	0.80	0.71
	1	0.76	0.57	0.65
	MA	0.70	0.69	0.68
	WA	0.70	0.69	0.68
LR	0	0.78	0.86	0.81
	1	0.85	0.77	0.81
	MA	0.81	0.81	0.81
	WA	0.81	0.81	0.81
KN	0	0.80	0.63	0.71
	1	0.71	0.85	0.78
	MA	0.76	0.74	0.74
	WA	0.75	0.74	0.74
SVC	0	0.79	0.87	0.83
	1	0.86	0.79	0.82
	MA	0.83	0.83	0.82
	WA	0.83	0.82	0.82

Table 5 shows the twitter dataset of COVID_1 consist of 5 classes. The results indicate SVC classifier performs better based on the WA and MA values. The precision values of RF and SVC are similar. Results also suggest that KN classifier obtain minimum WA value of 88% precision and F1-Score.

Table 6 illustrates COVID_2 dataset in which RF achieved high results in both MA. SVC produces better results in WA value. LR techniques achieve medium results which is not so high and low results. KN classifier produces minimum classification results. Certain sentiments values in precision achieve 1.00 results represents 100 percentage correctly classified results.

In Table 7, SVC methods achieved high MA using precision measure. For recall RF technique WA achieves best results. SVC performs better in all ways compared to other methods. KNN classifier performs minimum results. The precision, recall and f1-score increases and decreases based on the machine learning methods.

Table 8, discusses the comparison of proposed model with existing model. CN with GSA shows comparable better results than other methods for all

Table 4   Amazon dataset sentiment classification using machine learning techniques

Classifier	Class	Precision	Recall	F1-Score
RF	0	0.75	0.56	0.64
	1	0.57	0.76	0.65
	MA	0.66	0.66	0.65
	WA	0.67	0.65	0.65
LR	0	0.86	0.51	0.64
	1	0.59	0.89	0.71
	MA	0.72	0.70	0.68
	WA	0.74	0.68	0.67
KN	0	0.69	0.80	0.74
	1	0.68	0.55	0.61
	MA	0.68	0.67	0.67
	WA	0.69	0.69	0.68
SVC	0	0.86	0.65	0.74
	1	0.66	0.86	0.75
	MA	0.76	0.76	0.75
	WA	0.77	0.75	0.75

datasets. It is clear from the results that the proposed model is suitable to classify and understand text data easily. It further shows the algorithm is best suitable for benchmark dataset and extracted twitter dataset. Table 9 shows the comparison with other deep learning model, proposed model exhibit superior performance.

Figures 3–5 illustrates the twitter user's sentiments about COVID-19 disease. The experimental result achieves a high range value at neutral sentiments. Sentiments results suggest that the people are not aware of the disease during the period of March 2020. COVID-19 disease awareness must be educated to people to prevent the spread of the virus. The next high level of sentiments is positive. In Figure 5, shows some positive feedback about COVID. Since, the second dataset consist of 10,969 tweets our decision about people sentiment is neutral about the disease. Tweets classification about COVID is resulted in neutral sentiment.

Table 5   COVID_1 dataset sentiment classification using machine learning techniques

Classifier	Class	Precision	Recall	F1-Score
RF	0	0.84	0.99	0.91
	1	0.93	0.80	0.86
	2	0.98	0.60	0.75
	3	1.00	0.64	0.78
	4	1.00	0.25	0.40
	MA	0.95	0.65	0.74
	WA	0.89	0.25	0.88
LR	0	0.86	0.81	0.90
	1	0.84	0.89	0.91
	2	0.81	0.80	0.92
	3	0.82	0.65	0.96
	4	0.00	0.75	0.50
	MA	0.87	0.88	0.75
	WA	0.88	0.87	0.88
KN	0	0.80	0.90	0.85
	1	0.80	0.73	0.77
	2	0.82	0.46	0.59
	3	0.78	0.64	0.70
	4	0.00	0.00	0.00
	MA	0.64	0.55	0.58
	WA	0.80	0.80	0.79
SVC	0	0.86	0.98	0.91
	1	0.93	0.83	0.88
	2	1.00	0.65	0.79
	3	1.00	0.64	0.78
	4	1.00	0.25	0.40
	MA	0.95	0.76	0.75
	WA	0.90	0.75	0.89

Table 6   COVID_2 dataset sentiment classification using machine learning techniques

Classifier	Class	Precision	Recall	F1-Score
RF	0	0.78	0.99	0.87
	1	0.88	0.80	0.84
	2	1.00	0.52	0.68
	3	1.00	0.45	0.62
	4	0.95	0.59	0.73
	MA	0.92	0.67	0.75
	WA	0.85	0.83	0.83
LR	0	0.79	0.96	0.86
	1	0.83	0.83	0.83
	2	0.97	0.41	0.57
	3	1.00	0.30	0.47
	4	0.95	0.56	0.70
	MA	0.91	0.61	0.69
	WA	0.84	0.82	0.81
KNN	0	0.57	1.00	0.73
	1	0.99	0.35	0.52
	2	0.97	0.41	0.57
	3	1.00	0.33	0.50
	4	1.00	0.41	0.58
	MA	0.91	0.50	0.58
	WA	0.80	0.66	0.63
SVC	0	0.86	0.96	0.91
	1	0.89	0.84	0.86
	2	0.98	0.64	0.78
	3	0.00	0.15	0.27
	4	0.00	0.23	0.38
	MA	0.95	0.57	0.64
	WA	0.88	0.88	0.87

Table 7   COVID_3 dataset sentiment classification using machine learning techniques

Classifier	Class	Precision	Recall	F1-Score
RF	0	0.85	0.99	0.91
	1	0.93	0.81	0.87
	2	1.00	0.60	0.75
	3	1.00	0.64	0.78
	4	1.00	0.25	0.40
	MA	0.96	0.66	0.74
	WA	0.89	0.88.	0.88
LR	0	0.80	0.98	0.88
	1	0.91	0.73	0.81
	2	1.00	0.44	0.61
	3	1.00	0.55	0.71
	4	0.00	0.00	0.00
	MA	0.74	0.54	0.60
	WA	0.85	0.84	0.83
KN	0	0.80	0.90	0.85
	1	0.80	0.73	0.77
	2	0.82	0.46	0.59
	3	0.78	0.64	0.70
	4	0.00	0.00	0.00
	MA	0.63	0.46	0.58
	WA	0.73	0.74	0.79
SVC	0	0.84	0.97	0.90
	1	0.88	0.85	0.87
	2	0.96	0.56	0.70
	3	1.00	0.42	0.60
	4	0.89	0.69	0.78
	MA	0.91	0.70	0.77
	WA	0.87	0.87	0.86

Table 8 Comparing proposed work with existing method

Dataset	Techniques	Accuracy
Imdb	RF	0.78
	LR	0.80
	KN	0.80
	SVC	0.81
	Proposed approach	0.90
Yelp	RF	0.68
	LR	0.81
	KN	0.74
	SVC	0.82
	Proposed approach	0.91
Amazon	RF	0.65
	LR	0.68
	KN	0.68
	SVC	0.75
	Proposed approach	0.81
COVID_1	RF	0.88
	LR	0.83
	KN	0.79
	SVC	0.88
	Proposed approach	0.91
COVID_2	RF	0.83
	LR	0.81
	KN	0.65
	SVC	0.86
	Proposed approach	0.91
COVID_3	RF	0.88
	LR	0.83
	KN	0.79
	SVC	0.88
	Proposed approach	0.91

Table 9   Comparing with Deep Learning methods

Techniques	COVID_1	COVID_2	COVID_3
CNN	0.74	0.76	0.75
LSTM	0.79	0.74	0.76
CNN_Static	0.80	0.80	0.79
Bi_LSTM	0.88	0.87	0.85
Emdd+Conv	0.89	0.86	0.86
Proposed Approach	0.91	0.90	0.91
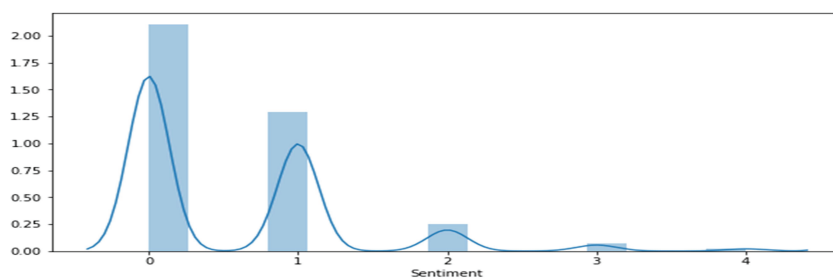
Figure 3   COVID_1 dataset sentiments classification.
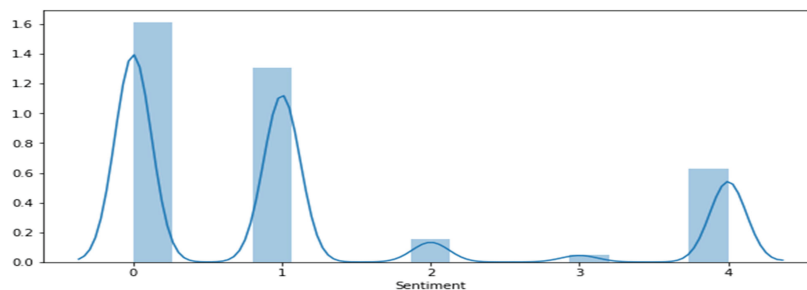
Figure 4   COVID_2 dataset sentiments classification.
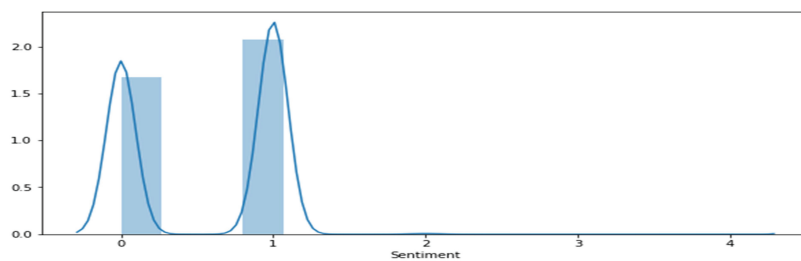
Figure 5   COVID_3 dataset sentiments classification.

5 Conclusion

In this paper, an innovative approach CN with GSA is designed for sentiments classification. Experimental study was conducted to test the performance of the proposed model using six different dataset. From the result it has been observed that the proposed model works better when compared with the other methods. This approach also has the ability to deal with small to large text data. Furthermore, due to the nature of GSA used in CN in handling text data the proposed model increases the classification accuracy for all kind of sentiment dataset. It achieves the accuracy above 90% in all dataset. From the experimental results, CN with GSA achieves better results even if dataset size increases.

For future works, the proposed method is expanded to search multiple objectives while searching, in order to make the algorithm fit to any kind of search process and choose the best algorithm based on domain specific. More layers can be included in convolution layer to increase the classification accuracy.

References

Castro, R.; Kuffó, L.; Vaca, C. Back(2017) Predicting venezuelan states political election results through twitter. In Proceedings of the Fourth International Conference on e Democracy & eGovernment (ICEDEG), Quito, Ecuador, 19–21 April 2017; pp. 148–153.

Ali Hasan , Sana Moin , Ahmad Karim and Shahaboddin Shamshirband.(2018). Machine Learning-Based Sentiment Analysis for Twitter Accounts, Mathematical and Computational Application, 23, 11.

Lingyu Meng, Zhijie Sasha Dong. (2020). Natural Hazards Twitter Dataset,Social and Information Networks.

Pete Burnap, Omer F. Rana. (2013). Detecting tension in online communities with computational Twitter analysis, Technological Forecasting & Social Change, Elsevier.

V. Diviya Prabha, R. Rathipriya (2013). Biclustering of web usage data using Gravitational Search Algorithm, International Conference on Pattern Recognition, Informatics and Mobile Engineering, IEEE.

Esmat Rashedi, Hossein Nezamabadi-pour. (2009) GSA: A Gravitational Search Algorithm, Information Sciences 179, Elsevier.

Hongping Hu, Xiaxia Cui.(2017).Two Kinds of Classifications Based on Improved Gravitational Search Algorithm and Particle Swarm Optimization Algorithm.

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—Volume 10; Association for Computational Linguistics: Stroudsburg, PA, USA, pp. 79–86.

Sunir Gohil, Sabine Vuik. (2018). Sentiment Analysis of Health Care Tweets: Review of the Methods Used, JMIR Public Health Surveillance.

Yujiao, L.; Fleyeh, H. (2018). Twitter Sentiment Analysis of New IKEA Stores Using Machine Learning. In Proceedings of the International Conference on Computer and Applications, Beirut, Lebanon, 25–26 July.

Jaspreet Singh, Gurvinder Singh. (2017). Optimization of sentiment analysis using machine learning classifiers, Human-centric Computing and Information Sciences.

Conneau A, Schwenk, H, Barrault, L; Lee Cam, Y. (2016). Very Deep convolutional networks for natural language processing, axiv.

M. Trupthi; Suresh Pabboju. (2017). Sentiment Analysis on Twitter Using Streaming API, IEEE. https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences

Walter HugoLopez Pinaya, Sandra Vinera. (2020).Convolutional neural networks, Science Direct.

Changyi Ma, Wenye (2020). LiSparse Binary Optimization for Text Classification via Frank Wolfe Algorithm, ACM.

Mehreen Naz, Kashif Zafar. (2019). Ensemble Based Classification of Sentiments Using Forest Optimization Algorithm.

Mohamed AtefMosa, Real-time data text mining based on Gravitational Search Algorithm, Expert System with Applications, Elesiver.

Jaeyoung Kim, Sion Jang, (2020), Text classification using capsules, Neurocomputing, Elsevier.

Lai, S.; Xu, L.; Liu, K.; Zhao, J. (2015). Recurrent convolutional neural networks for text classification. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence AAAI Press: Austin, TX, USA, pp. 2267–2273.

Cheng, J.; Dong, L.; Lapata, M. (2016). Long short-term memory-networks for machine reading. arXiv:1601.06733.

Jaeyoung Kim, Sion Jang. (2020). Text classification using capsules, Neuro-computing.

Xian Zhong, Jinhang Liu, Shuqin Chen. (2020). An emotion classification algorithm based on SPT-CapsNet, Deep Learning and Neutral Computing for Intelligent Sensing and Control, Neural Computing and Applications.

Yongping Du, Xiaozheng Zhao (2019). A Novel Capsule Based Hybrid Neural Network for Sentiment Classification, IEEE.

Zhao W, Ye J, Yang M, Lei Z, Zhang S, Zhao Z. (2018). Investigating capsule networks with dynamic routing for text classification. ArXiv.

Ekenga, C.C.; McElwain, C.-A.; Sprague, N. Examining Public Perceptions about Lead in School Drinking Water: A Mixed-Methods Analysis of Twitter Response to an Environmental Health Hazard. Int. J. Environ. Res. Public Health 2018, 15, 162. [CrossRef] [PubMed].

Bassari Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization.

Silva, J; Coheur, L; Mendes, A.C; Wichert, A. From symbolic to sub-symbolic information in question classification. Artif. Intell Rev. 2011, 35, 137–154.

Tripathy, A.; Agrawal, A; Rath, S.K. Classification of sentiment reviews using n-gram machine learning approach, Expert System Appl. 2016, 57, 117–126.

Neethu, M; Rajasree R. Sentiment analysis in twitter using machine learning approaches and semantic analysis. In Proceedings of the 2014 Seventh International Conference on Contemporary Computing (IC3), Nodia, India, 7–9 August 201, pp. 1–5.

Biographies

V. Diviya Prabha is a Ph.D. student at the Periyar University since 2016. He has received his B.Sc in Computer Science in 2009. She has completed MCA degree and M.Phil from Periyar University during 2012 and 2013. Her Ph.D. work centers on Data Mining and discusses the Text Mining to develop a solution for sentimental analysis in social media.

R. Rathipriya received his B.Sc and M.Sc degrees in Computer Science from Periyar Univeristy, Tamil Nadu, India; M.Phil and MCA degree from Periyar University and Ph.D. degree in Computer Science from Bharathiyar Univeristy, Tamil Nadu, India. Dr. R. Rathipriya is Assistant Professor at Periyar University in Department of Computer Science from 2008. She was Principal Invigestor for UGC MRP funding agency. She is expert in Web Mining, has acquired a solid experience in Bioinformatics.