

A hybrid sentiment analysis approach using black widow optimization based feature selection

Anand Joseph Daniel * and M Janaki Meena**

*Research Scholar in Department of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, India.

** Professor in the School of Computing Science and Engineering at VIT Chennai.

*Corresponding Author: danny02.20099@gmail.com

Submitted : 28-10-2020

Revised : 27-08-2021

Accepted : 12-09-2021

ABSTRACT

With the massive development of Internet technologies and e-commerce technology, people rely on the product reviews provided by users through web. Sentiment analysis of online reviews has become a mainstream way for businesses on e-commerce platforms to satisfy the customers. This paper proposes a novel hybrid framework with Black Widow Optimization (BWO) based feature reduction technique which combines the merits of both machine learning and lexicon-based approaches to attain better scalability and accuracy. The scalability problem arises due to noisy, irrelevant and unique features present in the extracted features from proposed approach, which can be eliminated by adopting an effective feature reduction technique. In our proposed BWO approach, without changing the accuracy (90%), the feature-set size is reduced up to 43%. The proposed feature selection technique outperforms other commonly used Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) based feature selection techniques with reduced computation time of 21 sec. Moreover, our sentiment analysis approach is analyzed using performance metrics such as precision, recall, F-measure, and computation time. Many organizations can use these online reviews to make well-informed decisions towards the users' interests and preferences to enhance customer satisfaction, product quality and to find the aspects to improve the products, thereby to generate more profits.

Key words: Feature optimization; Black Widow Optimization; sentiment analysis; classifier; machine learning.

INTRODUCTION

A lot of customers' purchase decision depends on the experiences people share, their comments, ratings and recommendations about purchased products through online. Many organizations can use these online reviews to make well-informed decisions towards the users' interests and preferences to enhance customer satisfaction, product quality and to find the aspects to improve the products, thereby to generate more profits (Iqbal et al., 2019). Hence sentiment analysis (SA) is essential to determine the public's opinion towards a

certain product by extracting the information from unstructured product reviews so as to determine whether a review is neutral, positive or negative (Yang et al., 2020).

SA is mainly classified as lexicon-based approach, machine learning (ML) approach and hybrid approach (Kausar et al., 2020). The sentiment polarity of dataset is calculated with the words' semantic orientation in lexicon-based approach (Taboada et al., 2011). In ML based approaches, for training the classifier and identify the sentiment, various learning algorithms and labelled datasets are used (Boiy & Moens, 2008). To bridge the gap between ML and lexicon-based approaches, the hybrid approach is employed (Shang et al., 2016). If we take all the probable features of larger documents, then the large size feature vectors will not fit in the memory and cause scalability problem. An approach with improved scalability will be able to sustain its performance even as

it is tested by larger dataset. As the feature-set size develops with larger dataset, there occurs a scalability problem due to noisy features which can be eradicated by adopting an appropriate feature selection technique (Ansari et al., 2019). The main aim of feature selection technique is to create a well-selected feature vectors for improving the performance of classifier with better scalability and minimize the computation time simultaneously. The feature selection through optimization algorithm can automatically select features without manual intervention and eliminate a large quantity of unnecessary features without compromising the accuracy. There are many researchers used optimization algorithms in different fields (Serbanescu et al., 2017, Ravariu et al., 2009, Eberhart & Kennedy 1995).

In this paper, a hybrid technique is proposed for SA, which employs lexicon-based and ML methods to automatically analyze sentiments of online product reviews with Black Widow Optimization (BWO) based feature reduction. In the hybrid approach, lexicon dictionary-based feature vector acts as input to the classifier and the product reviews act as test data to get the sentiment score of each word. According to the information I have, no work has been conducted on the hybrid sentiment analysis region with BWO based optimized feature selection. The proposed optimization reduced 43% features while maintaining the accuracy. Moreover, for evaluating the performance of the proposed work with the BWO based feature selection method, the results are compared with other commonly used PSO (Eberhart & Kennedy, 1995) and GA (Holland & John, 1975) based feature selection methods.

The main intention of this paper is to propose a hybrid SA method by combining lexicon-based and ML approaches in which lexicon dictionary-based feature vector acts as input for LSVM classifier with a novel BWO based feature selection approach. This proposed method is analyzed to show that the accuracy is improved when compared to other feature reduction algorithms. Experiments are conducted to show the performance of the proposed method, with the data of amazon product reviews from five categories such as Electronics, Musical Instruments, Furniture, Wireless products and Watches. The evaluation of the proposed approach is based on conventional performance matrices such as F-measure, recall, precision and accuracy. Moreover, the computation time analysis of proposed BWO based feature selection algorithm is also provided. The proposed BWO feature selection method's performance is evaluated through experiments, both in sentiment analysis-oriented feature selection domain and general feature selection domain.

The remainder of this paper is arranged as follows. Section II details the background in the hybrid SA and optimized feature selection area. Section III explains the proposed methodology. Experiments are provided in Section IV. The experimental results and discussion are given in the Section V. Section VI concludes the paper and discuss the future work.

BACKGROUND

Hybrid Sentiment Analysis

SA methods are mainly classified as lexicon-based approach, ML based approach and hybrid approach (Kausar et al., 2020). In the lexicon-based approach, the polarities and the frequencies of the negative and positive words are examined to get the sentiment of the analyzed text using a predefined dictionary of words (Bollegala et al., 2013, Cho et al., 2014). In the ML based approach, features are generated from the text and these features are used by different learning algorithms to predict a label (Katz et al., 2015). The lexicon-based approach can classify the text using unlabeled training set and the lexicons are defined independently of the text, so that over fitting at any instance can be prevented (Ortigosa et al., 2014). Also, this approach can show robust performance across texts and domains and perform SA on multiple domain datasets (Katz et al., 2015). Though this approach has more merits, it also has disadvantage like the need of manual maintenance. This approach can't identify abbreviations in the non-standard form which are mainly used in twitter posts due to its limited coverage on texts with informal forms (Ducange et al., 2019). The ML approach is very suitable for informal text and unstructured contents such as twitter data. This approach provides more flexibility and hence it eliminates predefined lexicons (Katz et al., 2015).

Although this approach performs well, it requires labelled training dataset that needs manual annotation (Zhang et al., 2015, Palit & Ghosh, 2020). Hence, a hybrid approach is necessary to eliminate the disadvantages but to combine the merits of each methodology. The hybrid approach combines both ML approach and lexicon-based approach together to bridge the gap between them. In Gupta & Joshi, 2019, Sahu & Khandekar, (2020), a hybrid SA approach is introduced, which combines ML method and polarity-based lexicon method for twitter data. A hybrid SA technique on Facebook to automatically analyze sentiments of online product reviews is proposed by Ortigosa et al., (2014). For SA of product reviews from Amazon, a hybrid SA is implemented by Mumtaz & Ahuja, (2017). Govindarajan, (2013), implemented integration classification-based hybrid SA approach with arcing classifier. They analyze performance of NB and GA ensemble classifier in terms of accuracy. Srivastava et al., (2019) offered a hybrid approach by means of Naive Bayes and Random Forest on mining Twitter datasets. Sharma and Anurag, (2020) exploited ranking-based feature selection to textual data in social media with hybrid collaborative learning. But these existing works did not employ any optimization algorithms in the feature selection task to enhance the scalability and accuracy of hybrid SA. In SA based articles, most of the datasets are collected from the reviews from multiple online platforms such as Amazon, Yelp, IMDb, TripAdvisor, Getty Images, Facebook, Twitter, You Tube, Flickr, and reviews of Google restaurant/play/customer Yadav & Vishwakarma, (2020). Our proposed work focuses mainly on product reviews. Since high quality word embedding can be obtained for words from a

large number of unlabeled texts by Glove (Pennington et al., 2014) and Word2vec (Mikolov et al., 2013) model, our proposed work utilizes this model to get initial word embedding.

Optimized Feature Selection in SA

Feature selection in sentiment analysis plays a vital role in system accuracy enhancement. There are many studies done in text classification domain with optimized feature selection. In solving the different optimization problems, the bio-inspired algorithms have incredible capability (Yadav & Vishwakarma, 2020). We can improve the accuracy of the SA process, by using these algorithms in the optimization of feature set. The main classification techniques for the bio inspired algorithm in SA are Naive Bayes, SVM, K-means, decision trees, K-nearest neighbor, neural network, and Logistic regression (Yadav & Vishwakarma, 2020). In our proposed work, four different classifiers' performance matrices are compared. Rathor et al., (2018) and Haque et al., (2018) have presented that SVM can give better accuracy than other classifiers in SA on Amazon product reviews. Since SVM classifier is proven to be robust for sentiment classification Gupta & Joshi, (2019), it is chosen in our paper for the final sentiment classification. Linear SVM (LSVM) is used for classification because it shows effective results on a larger feature space (Mohammad et al., 2013). Kalarani & Selva Brunda, (2018) have proposed the firefly algorithm (FA) for feature reduction and the SVM and ANN classifier have been utilized to classify the sentiments of the movie reviews. The experimental results showed an improved accuracy with reduced training time. Kristiyanti & Wahyudi, (2017) have implemented three algorithms called Principle Component Analysis (PCA), PSO and GA based feature reduction with SVM classifier. The accuracy of SVM classifier has been enhanced with these three algorithms. They have used Amazon products' reviews for classification. When compared with GA and PCA algorithm, PSO algorithm has shown higher accuracy. Thus, PSO based feature selection in SVM shows higher accuracy for classification problems. But PSO based optimization has disadvantages like the multi-objective optimization which requires fine-tuning on parameters' value and shows poor performance for datasets of multi-domain (Jiang et al., 2018).

Iqbal et al., (2019) have proposed a novel GA based feature selection technique to improve the scalability problem that happens as the feature vector size increases in the SA. A hybrid SA approach with reviews from IMDB, Yelp and Amazon has been used. They have analyzed the performance by measuring the recall, accuracy, precision, F1 score and have used six different classifier algorithms. GA based optimization has been employed for feature reduction. The results have proved that GA based feature selection approach has higher accuracy than LSA and PCA methods. This paper proposes a novel BWO based feature reduction in SA. The proposed feature reduction method BWO is encouraged by the black widow spider's unique mating behavior (Hayyolalam & Pourhaji Kazem, 2020, Transpire online., 2020). The BWO algorithm can avoid local optima problem and deliver fast convergence speed by providing a better performance in exploration and exploitation stages. Furthermore, it has the capability to keep the stability between exploitation and exploration by searching a large space to get the

best global solution. Thus, BWO will be a better solution for the feature selection task and that algorithm has the ability to overcome other optimization algorithms such as PSO and GA with better scalability and higher accuracy (Taboada et al., 2011).

PROPOSED METHODOLOGY

This work focuses on sentimental analysis using various optimizations for feature selection. Though the lexical dictionary-based framework for SA shows higher accuracy, they have limited size of lexical databases. This drawback can be eliminated by utilizing a hybrid approach of ML and lexical based sentiment analysis. In this paper, a novel hybrid framework has been proposed with feature selection using BWO. Our proposed BWO based feature selection method with hybrid sentimental analysis is explained in the following subsections.

Hybrid SA with Optimized Feature Selection

Our hybrid method includes ML and lexicon-based dictionary for training the LSVM classifier and bag-of-words as features for testing the LSVM classifier with BWO based feature selection. The proposed method is shown in Figure 1.

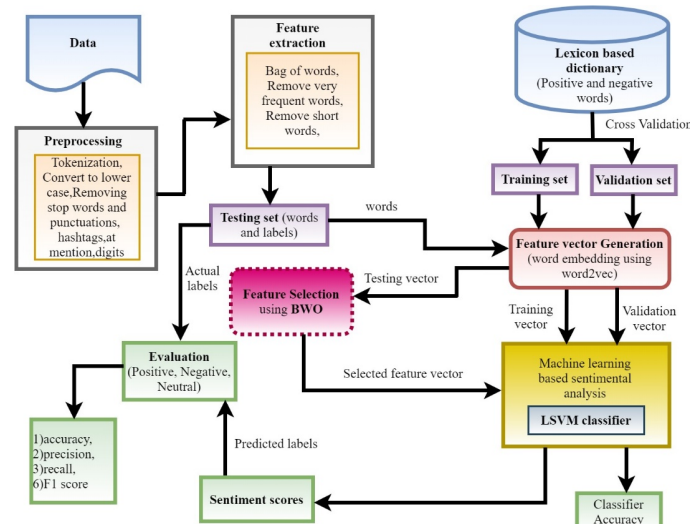


Figure 1 Proposed sentiment analysis framework

In the first step of this process, the amazon product review data is cleaned and preprocessed. The preprocessing and data cleaning comprises tokenization, lower case conversion, removal of stop words, punctuations, hashtags, at-mention and digits. Furthermore very common words and short words are also detached. These preprocessing and data cleaning steps are evidently elucidated in the experiments section. In this approach, ML algorithm with LSVM classifier is used for sentiment classification of the given text. The lexicon-based dictionary with positive and negative words cross validates into training and validation data. The training data is utilized to train the LSVM classifier and validation set is utilized to validate the performance of classifier. The selected features from our product reviews are taken as testing set by performing different modules ranging from data cleaning, pre-processing, feature extraction using Bag-Of-Words and sentiment analysis. These training and testing data set are embedded using Glove model and word2vec method to generate feature vector. This generated feature vector after preprocessing contains 80% of the input data and that will arise scalability problem with bag-of-words approach. As the dataset size grows bigger, this problem worsens. Hence, so as to eliminate this scalability issue, we want to select a proper feature selection method. To minimize this problem due to more feature set, an efficient BWO based feature selection method is proposed in this paper. The most frequent words are removed from feature vector using bag-of-words method. In this method, each

feature vector is modelled on to the BWO model, and for several hundred generations, BWO simulation is run to find the optimal selected feature set. The chosen features are verified by means of LSVM classifier which gives equivalent sentiment scores. Based on the values of sentiment scores the anticipated labels are assessed. Thus the sentiment polarity is decided with associating the actual labels from dataset and anticipated labels from sentiment score which gives better precision for sentiment classification. The performance metrics such as accuracy, recall, precision and F1 scores are assessed at this step.

When using a larger dataset, scalability issue arises as features extracted from bag-of-words model have significantly large feature vector size. This issue can be eliminated with the feature optimization by minimizing its size and keeping the accuracy same. This issue is articulated in this section and its way out is suggested by means of BWO approach which is shown by the flowchart of BWO based feature selection in the Figure 2.

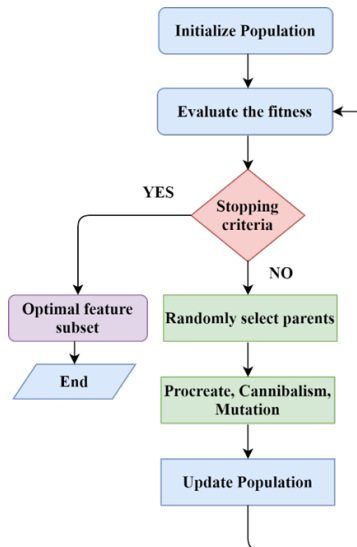


Figure 2 Flowchart of BWO algorithm for feature selection

Problem formulation

If we include all the possible features in the testing set, then the feature vector of larger datasets will larger the feature size. As the feature size get larger that will not fit in the memory and there arises the scalability problem. Thus, to eliminate this problem, an optimized feature selection technique is required. The optimization of this problem is the minimum number of feature selection from the feature-set of large size. As discussed before, BWO has more merits such as achievement of optimized fitness value and early convergence, compared to other algorithms such as GA and PSO and this BWO selected to optimize this problem. This optimization algorithm does not need labelled dataset for sentiment classification.

Mathematical model

In this paper, so as to model our problem, the structure should be taken as an array. Assume W be the test data with all the tokens in a text after pre-processing.

$$W = \{w_1, w_2, \dots, w_n\} \quad (1)$$

For a N_{var} dimensional feature vector W, Widow is an array of $1 \times N_{var}$ representing the feasible solution of the problem. This array is given as:

$$\text{Widow} = [x_1, x_2, \dots, x_{N_{var}}] \quad (2)$$

Here Widow is a D-dimensional array for a D-dimensional feature vector W. A candidate widow matrix of size $N_{pop} \times N_{var}$ is generated with an initial population of spider.

Calculate fitness

So as to employ BWO to this issue, $f(W)$ is taken as function of fitness, that is a main part of BWO based feature reduction. It decides the measures for the good candidates that will survive after mating and best offspring which will survive as strong candidate. Each of the variables $(x_1, x_2, \dots, x_{N_{var}})$ is floating point values. The fitness of widow is calculated by fitness function f evaluation at a widow of $(x_1, x_2, \dots, x_{N_{var}})$. Therefore,

$$\text{Fitness} = f(W) = f(x_1, x_2, \dots, x_{N_{var}}) \quad (3)$$

Feature selection using BWO

The whole population would converge to one optimal value by running the model upto generations of N number. In each generation, different steps such as parent selection, procreate, cannibalism, mutation and fitness evaluations initiate the function of the BWO. The complexity in time for BWO is based on the fitness function. Let N_p be the size of the population and the value of N_p is 40. Based on the procreating rate 'pc', the number of reproductions 'nr' is calculated. The value of 'pc' is 0.8. Then best nr solutions are selected from population and saved in population1. Then parents' pairs are selected randomly from population1 to accomplish the procreating step.

Now, in this algorithm, an array called alpha should be formed to reproduce, until widow array with random numbers comprise. Then offspring is formed by using α with the equation 4 in which x_1 and x_2 represent parents, y_1 and y_2 represent offspring.

$$\left. \begin{aligned} y_1 &= \alpha \times x_1 + (1 - \alpha) \times x_2 \\ y_2 &= \alpha \times x_2 + (1 - \alpha) \times x_1 \end{aligned} \right\} \quad (4)$$

For $N_{var}/2$ times this process is repeated, while numbers of randomly chosen should not be duplicated. D -dimensional children is generated using equation 1. Here, during or after mating, the female black widow eats the male. So that, the father is destroyed. Finally, the mother and children spiders are added to an array and ordered based on their fitness value.

In sibling cannibalism strong spiders eat their weaker siblings. Now cannibalism rating 'pCannibalism' determines the number of survivors. The value of 'pCannibalism' is 0.5. So, based on the cannibalism rate, some of the children are destroyed and remaining solutions are saved in population 2. Then, based on the mutation rate, the number of mutation children 'pmutation' is calculated. The value of 'pmutation' is 0.4. For the number of mutation children, individuals are randomly selected from population 1. Each of the selected solutions randomly mutate two elements in the array and generate a new solution. This solution is saved in population 3. Then population is updated by adding population 2 and population 3.

$$\text{Population} = \text{population 2} + \text{population 3} \quad (5)$$

At last, the exact best solution is produced from the group of solutions, which is utilized for the effective sentiment classification. During the process of feature selection, all the feature is visited once and utilized for optimized features prediction from the set of features. The selected features act as the input to the classifiers for SA from the Amazon dataset.

EXPERIMENTS

This section explains how to evaluate proposed model for SA with optimized feature selection. This section includes the dataset description, data preprocessing steps and performance metrics.

Dataset

The dataset utilized for this work is the product reviews of amazon from five categories such as Electronics reviews, Musical Instruments reviews, Furniture reviews, Wireless products reviews and Watches reviews which contains about 48500 product reviews taken out from Amazon.com (<https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt>). The dataset after manual data cleaning process has been taken as the dataset of this work which comprises 1000 positive reviews and 1000 negative reviews for each category. Each review has the information with rating (0–5 stars), reviewer location and name, a review title and date, a product name, and the text of review. The dataset was unlabeled and it was labelled manually to utilize it in machine learning as testing data. For that purpose, the ratings of product reviews have been divided into two categories, reviews with ratings >3 have been labelled positive and the reviews with rating <3 have been labelled negative. These reviews have to go through data cleaning and pre-processing before given as input to LSVM classifier because these reviews will be in an unstructured form.

Data Cleaning

The first step in the pre-processing process is data cleaning. In this stage, product reviews after manual processing is preserved in the memory for cleaning. Usually, the online review data encloses lots of noises which needs more phases to clean it.

Removal of garbage. This step removes unwanted non-characters consisting of URLs, symbols, and digits using modified regular expressions.

Removal of stop words. Most of the regularly used stop words in English are “an”, “of”, “a”, “you”, “the”, “and”. These words do not have any meaning. So, these words are generally ignored to improve the accuracy of the SA and are collected together and removed from dataset.

Data Pre-Processing

This module includes tokenization, and case conversion.

Tokenization. In this process, the text is separated into keywords, words, symbols, phrases and tokens based on the space of separation. Punctuation marks are also discarded in this process.

Case conversion. All the reviews should be converted to lower case because it has to be in the same case to get processed. At last, a string of meaningful words is obtained.

Performance Metrics

In this paper, four common performance measures were utilized for the performance evaluation of our model: precision, accuracy, recall, and F1 score. The performance measures are defined as below:

TP: It is the number of positive comments classified as positive.

FP: It is the number of negative comments classified as positive.

TN: It is the number of negative comments classified as negative.

FN: It is the number of positive comments classified as negative.

Accuracy: It is the ratio of correctly classified comments over the total number of comments.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision: It is the ratio of correctly classified positive comments over the total number of positive classified comments.

$$\text{precision} = TP / TP + FP \tag{7}$$

Recall: It is the ratio of correctly classified positive comments over all comments actually belonging to that class.

$$\text{recall} = TP / TP + FN \tag{8}$$

F1 score: the weighted average of recall and precision.

$$\text{F1 score} = 2 * \text{precision} * \text{recall} / (\text{precision} + \text{recall}) \tag{9}$$

EXPERIMENT RESULTS AND DISCUSSION

This section gives the experiment results and the discussion. The software used for evaluations of the proposed framework is MATLAB R2018a. The model parameters used to validate the performance of various optimization techniques in SA oriented feature selection domain and general feature selection domain as given in Table 1. In order to make an assessment and to validate the outcomes of BWO, two famous prevailing algorithms GA (Ravariu et al., 2009) as the best evolutionary technique and PSO (Serbanescu et al., 2017) as the best swarm-based algorithm and the results of the BWO are related with all of them (Hayyolalam & Pourhaji Kazem, 2020). The experiment outcomes of BWO based optimized feature selection in general domain have been assessed in numerous scales and associated with standard form of PSO and GA algorithms. Figure 3 exemplifies that BWO finds the global optimum more rapidly than other two optimization algorithms and our BWO algorithm has a sophisticated convergence rate (Hayyolalam & Pourhaji Kazem, 2020). Thus the suggested BWO algorithm is able to keep the poise amid the investigation and manipulation stages and escape from local optima problem in comparison with other two optimization algorithms. Hence, in our paper, BWO based feature reduction technique is chosen for sentiment analysis.

Table 1 The model parameters.

Algorithm	Parameters	Values
PSO	Weight	0.2
	Constant	2
GA	Rate of crossover	0.8
	Rate of mutation	0.01
Proposed BWO	Procreating rate	0.8
	Cannibalism rate	0.5
	Mutation rate	0.4

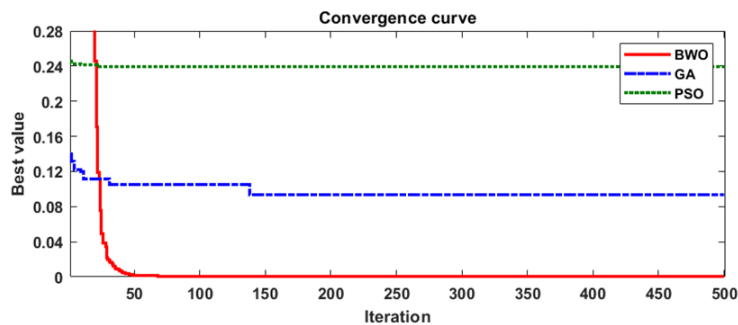


Figure 3 Convergence curve comparison for PSO, GA in general feature selection domain

Table 2 Performance measures comparison for various optimization-based feature selection in our proposed model.

Optimization Techniques	Performance measures of proposed model for furniture review				
	Accuracy	Precision	Recall	F1 score	Computation time
PSO	89.00	92.05	95.29	93.64	38 sec
GA	86.00	93.12	89.41	91.56	35sec
Proposed BWO	90.00	93.00	95.29	94.20	21 sec

Table 2 shows the performance measures comparison for various optimization-based feature selection. This evaluation shows that proposed BWO based approach is more efficient than the other two well-known algorithms of feature selection. Table 3 shows the accuracy, recall, precision and F1 score comparison for four various classifiers in SA. The SA effects of the LSVM classifier is compared with other common ML based classifiers existing works. From these results we can conclude that the classification accuracy of the LSVM classifier is significantly better than other classifiers in existing SA works. Table 4 shows the results of the proposed model with BWO and without BWO. These results show that the performance measures of BWO based optimized feature selection is almost similar to non-BWO based technique with reduced feature size. Thus, BWO based optimal selection of feature in SA doesn't affect the performance compared to non- BWO based approach.

Table 3 Performance measures comparison of different classifiers in sentiment analysis.

Classifier	Accuracy	Precision	Recall	F1 score
LSVM	90.00 %	93.00 %	95.29 %	94.20 %
NB [1]	77.9 %	78.6%	77.8%	78.7%
RF [28]	88.2%	96.7%	96.7%	97%
LR [28]	81.99%	87%	88%	88%

Table 4 Performance measures comparison of proposed hybrid model with BWO and without BWO on Amazon data.

Products	With BWO				Without BWO			
	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
Furniture	90 %	90%	95.29%	94.2%	89.81%	89.21%	94.89%	92.98%
Electronics	90 %	93.2%	92.9%	94.2%	89.44%	92.93%	92.67%	93.23%
Watches	90 %	96.5%	92.2%	94.3%	89.73%	96%	92%	93.26%
Wireless	80 %	83.9%	92.4%	87.9%	79.82%	83.1%	92.13%	87.2%
Musical Instruments	88 %	94 %	92%	93 %	87.88%	93.8%	91.78%	92.35%

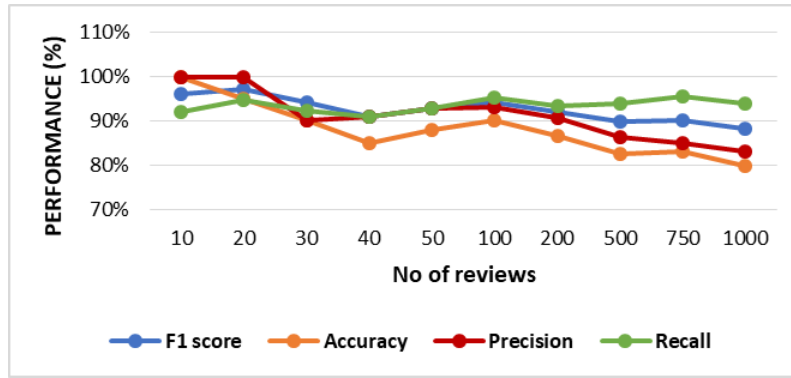


Figure 4 Effects of number of reviews on proposed model

Figure 4 displays the results of the proposed model for various number of reviews. The number of reviews is increased from 10 to 1000 and the corresponding performance is measured by means of the metrics like accuracy, precision, recall and F1 score. As the number of reviews is small, a complete accuracy and precision is attained. When the number of reviews slowly increases from 10 to 50, the feature size is also increased in order that the performance is slightly decreased. The results show that as increase of the reviews for testing data, the model performance slightly decreases and it doesn't make more effect in the performance of the model. Figure 5 shows the feature size before and after the BWO optimization on selection of feature. From this result we can see that performance of both methods is similar but optimization by means of BWO gives a minimized size of feature.

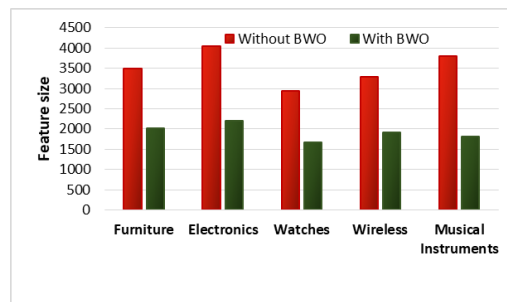


Figure 5 Features size comparison of proposed model with BWO and without BWO optimization

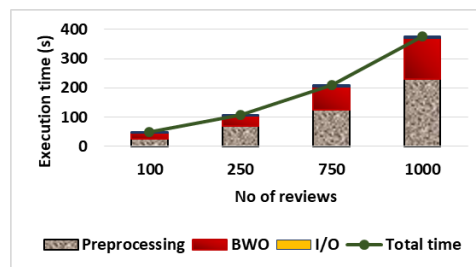


Figure 6 Time consumption and Scalability of different steps in our proposed model

Table 5 Comparative Analysis with related works.

Paper	Dataset	Accuracy
(Xia et al., 2016)	Books	82.9%
	Electronics	86.6%
	DVD	83.7%
	Kitchen	89.1%

(Elli & Yi-Fan, 2016)	Accessories and cellphone	80.11%
		72.95%
(Shaikh & Deepa, 2016)	Camera	62%
		80%
		68%
	Musical instruments	62%
		80%
		68%
	Books	70%
		70%
		80%
Proposed Model	Furniture	90 %
	Electronics	90 %
	Watches	90 %
	Wireless	80 %
	Musical Instruments	88 %

From Figure 6 we can see that the feature-set size after BWO optimization on feature selection has reduced the size of feature size before BWO optimization by almost 43%. This has an important effect on the system scalability. The BWO consists almost 50%-60% of the total execution time. As the reviews increases, the execution time also proportionally increases. Thus we can assess that the scalability of proposed SA approach is improved with cost of time. The Table 5 shows the comparative analysis of our proposed dataset with related

works in terms of accuracy. The related works employed various preprocessing and feature extraction processes. In our approach we employed hybrid approach with optimized feature selection process. From this comparative analysis and above results we can decide that our proposed approach is more effective than other existing optimization algorithms for feature selection task and could give better results than some of the existing works.

CONCLUSION

In this paper, a novel hybrid model with feature selection using BWO algorithm was designed, developed and evaluated. Moreover, our proposed hybrid technique was evaluated for scalability by comparing the execution time. In the total execution time, selection of optimal feature vector using BWO needs 50% to 60% and reduced feature-size up to 43% without changing the accuracy (90%). Experiments were conducted to show the performance of proposed approach, with the data of amazon product reviews from five categories such as Electronics, Musical Instruments, Furniture, Wireless products and Watches. The evaluation of the proposed approach was based on performance matrices such as F-measure, recall, precision and accuracy. The results indicated that our BWO based optimized feature selection method showed improved accuracy than PSO and GA algorithms with reduced computation time of 21sec. Thus, our proposed approach has higher accuracy and better scalability for SA of online reviews. Our proposed work considers only the product data. The main intention of our future work is to extend the proposed work on multi-domain SA with various sources of datasets like reviews of movies, twitter data etc. In our proposed work, the precision is slightly decreased with the increase in the number of data. So, in order to improve the accuracy, our future work can also consider this challenge.

REFERENCES

- Ansari, G., Ahmad, T., & Doja, M. N. 2019.** Hybrid Filter–Wrapper Feature Selection Method for Sentiment Classification. *Arabian Journal for Science and Engineering*.
- Boiy, E., & Moens, M.-F. 2008.** A machine learning approach to sentiment analysis in multilingual Web texts. *Information Retrieval*. 12(5). 526–558.
- Bollegala, D., Weir, D., & Carroll, J. 2013.** Cross-Domain Sentiment Classification Using a Sentiment Sensitive Thesaurus. *IEEE Transactions on Knowledge and Data Engineering*. 25(8). 1719–1731.
- Cho, H., Kim, S., Lee, J., & Lee, J.-S. 2014.** Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews. *Knowledge-Based Systems*. 71. 61–71.
- Ducange, P., Fazzolari, M., Petrocchi, M., & Vecchio, M. 2019.** An effective Decision Support System for social media listening based on cross-source sentiment analysis models. *Engineering Applications of Artificial Intelligence*. 78. 71–85.
- Eberhart, R., & Kennedy, J. 1995.** A new optimizer using particle swarm theory. *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*.
- Elli M S & Yi-Fan. 2016.** Amazon Reviews, business analytics with sentiment analysis.
- Govindarajan, M. 2013.** Sentiment analysis of movie reviews using hybrid method of naive Bayes and genetic algorithm. *International Journal of Advanced Computer Research*. 3(13).
- Gupta, I., & Joshi, N. 2019.** Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic. *Journal of Intelligent Systems*. 0(0). doi:10.1515/jisys-2019-0106.
- Haque, T. U., Saber, N. N., & Shah, F. M. 2018.** Sentiment analysis on large scale Amazon product reviews. 2018 IEEE International Conference on Innovative Research and Development (ICIRD).
- Hayyolalam, V., & Pourhaji Kazem, A. A. 2020.** Black Widow Optimization Algorithm: A novel meta-heuristic approach for solving engineering optimization problems. *Engineering Applications of Artificial Intelligence*. 87. 103249.
- Holland, & John H. 1975.** *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. U Michigan Press.
- Iqbal, F., Maqbool, J., Fung, B. C. M., Batool, R., Khattak, A. M., Aleem, S., & Hung, P. C. K. 2019.** A Hybrid Framework for Sentiment Analysis using Genetic Algorithm based Feature Reduction. *IEEE Access*. 7(4). 14637-14652.
- Jiang, H., Kwong, C. K., Park, W. Y., & Yu, K. M. 2018.** A multi-objective PSO approach of mining association rules for affective design based on online customer reviews. *Journal of Engineering Design*. 29(7). 381–403.
- Kalarani, P., & Selva Brunda, S. 2018.** Sentiment analysis by POS and joint sentiment topic features using SVM and ANN. *Soft Computing*.
- Katz, G., Ofek, N., & Shapira, B. 2015.** ConSent: Context-based sentiment analysis. *Knowledge-Based Systems*. 84. 162–178.
- Kausar, S., Huahu, X., Shabir, M. Y., & Ahmad, W. 2020.** A Sentiment Polarity Categorization Technique for Online Product Reviews. *IEEE Access*. 8. 3594–3605.

Kristiyanti, D. A., & Wahyudi, M. 2017. Feature selection based on Genetic algorithm, particle swarm optimization and principal component analysis for opinion mining cosmetic product review. 2017 5th International Conference on Cyber and IT Service Management (CITSM).

Mikolov, T., Yih, W. & Zweig, G. 2013. Linguistic regularities in continuous space word representations. NAACL. 746–751.

Mohammad S. M., Kiritchenko, S., & Zhu, X. 2013. Building the state-of-the-art in sentiment analysis of tweets. Second Joint, Conference on Lexical and Computational Semantics (SEM). 321–327.

Mumtaz, D., & Ahuja, B. 2017. A Lexical and Machine Learning-Based Hybrid System for Sentiment Analysis. Studies in Computational Intelligence. 165–175.

Ortigosa, A., Martín, J. M., & Carro, R. M. 2014. Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior. 31. 527–541.

Palit, S., & Ghosh, S. 2020. Real Time Sentiment Analysis. International Journal of Synthetic Emotions. 11(1). 27–35.

Pennington, J., Socher, R., & Manning, C. D. 2014. Glove: Global vectors for word representation. EMNLP. 14. 1532–1543.

Rathor, A. S., Agarwal, A., & Dimri, P. 2018. Comparative Study of Machine Learning Approaches for Amazon Reviews. Procedia Computer Science. 132. 1552–1561.

Ravariu, C., Ionescu-Tirgoviste, C., Ravariu, F. 2009. Glucose Biofuels Properties in the Bloodstream, in Conjunction with the Beta Cell Electro-Physiology. Conference: International Conference on Clean Electrical Power (ICCEP 2009). 124.

Sahu, T. P., & Khandekar, S. 2020. A Machine Learning-Based Lexicon Approach for Sentiment Analysis. International Journal of Technology and Human Interaction. 16(2). 8–22.

Serbanescu, M., Placinta, V. M., Hutanu, O. E., et al C. 2017. Smart, Low Power, Wearable Multi-Sensor Data Acquisition System for Environmental Monitoring. 10th International Symposium on Advanced Topics in Electrical Engineering (ATEE). Bucharest. ROMANIA. 23-25. 118-123.

Shaikh T, & Deepa D. 2016. Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews. International Journal of Computer Trends and Technology. 36(4). 225-230.

Shang, L., Zhou, Z., & Liu, X. 2016. Particle swarm optimization-based feature selection in sentiment classification. Soft Computing. 20(10). 3821–3834.

Sharma, S., & Jain, A. 2020. Hybrid Ensemble Learning With Feature Selection for Sentiment Classification in Social Media. International Journal of Information Retrieval Research. 10(2). 40–58.

Srivastava, A., Singh, V., & Drall, G. S. 2019. Sentiment Analysis of Twitter Data. International Journal of Healthcare Information Systems and Informatics. 14(2). 1–16.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. 2011. Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics. 37(2). 267–307.

Transpire online. 2020. Black Widow Optimization (BWO) Algorithm: Mating Behaviour of Black Widow Spider for Solving Engineering Optimization Problems. <https://transpireonline.blog/2020/02/06/black-widow-optimization-bwo-algorithm-mating-behaviour-of-black-widow-spider-for-solving-engineering-optimization-problems/>.

Xia, R., Xu, F., Yu, J., Qi, Y., & Cambria, E. 2016. Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis. Information Processing & Management. 52(1). 36–45.

Yadav, A., & Vishwakarma, D. K. 2020. A comparative study on bio-inspired algorithms for sentiment analysis. *Cluster Computing*.

Yang, L., Li, Y., Wang, J., & Sherratt, R. S. 2020. Sentiment Analysis for E-commerce Product Reviews in Chinese based on Sentiment Lexicon and Deep Learning. *IEEE Access*. 1–1.

Zhang, Y., Hu, X., Li, P., Li, L., & Wu, X. 2015. Cross-domain sentiment classification-feature divergence, polarity divergence or both? *Pattern Recognition Letters*. 65. 44–50.