[1] Shilpi Gupta
[2] Niraj Singhal
[3] Gunjan Ansari

Optimizing Features using BGWO for Aspect Term Extraction

Journal of Electrical Systems

Abstract: - There are two major tasks in the area of sentiment analysis- aspect term extraction and determining the sentiment polarity of extracted aspect terms. One of the major challenges in the domain of aspect term classification is selection of optimal features. In this paper, a binary variant of Grey Wolf Optimization (BGWO) is employed for selecting the most valuable features from the identified linguistic features for classifying the token as an aspect term. The process of GWO starts by selecting three best solutions from the binarized population of grey wolves using the fitness function that maximizes the accuracy and minimizes the feature set size. The position of grey wolves is updated using the stochastic crossover operation on the identified three best solutions. To evaluate the performance of the proposed method for optimal feature selection, an experimental study is conducted on two SemEval datasets of restaurant and laptop reviews using three different classifiers- Logistic Regression, Naive Bayes and Support Vector Machine. The results depict that model could achieve an average f1 score of 86.3% and 71% on laptop and restaurant domains respectively using the optimal feature set. The proposed work is further extended to generate an opinion-based summary of domain-specific aspect terms extracted from the review documents using VADER. VADER is a popular rule-based sentiment analyser that does not require prior training and performs efficiently on social media data. The performance analysis of Sentiment Analysis on few test samples of laptop domain indicates the efficacy of the proposed work.

Keywords: Sentiment Analysis, Aspect Term Extraction, Binary Particle Swarm Optimization, Binary Grey Wolf Optimization, Feature Selection

I. INTRODUCTION

Aspect-Based Sentiment Analysis (ABSA) involves extraction of aspects and prediction and of its sentiment polarity. For example, in the review, "The price is reasonable although the service is poor" [26]. The review contains two opposing viewpoints on the extracted attribute "price" and "service" of the same entity "restaurant". The task of aspect term extraction is to identify the implicit and explicit aspect terms in the given review to determine the overall product quality. In the initial works on ABSA, Toh& Wang (2014) [27] used Conditional Random Field (CRF) on SemEval-2014 dataset for the extraction of aspect terms from Restaurant and Laptop domains. The authors also utilized clustered features for unlabelled data along with the syntactic, semantic and lexical features. Bhamare and Prabhu (2021) [8] used multivariate filter based and selective dependency relations approach for the extraction of aspect terms on the restaurant datasets of SemEval, Kaggle and Yelp. The accuracy and F1 score for the aspect identification task achieved by the authors are 94.75% and 85.24% respectively by using a feature set generated from both the approaches.

Aspect-Based Sentiment Analysis (ABSA) involves extraction of aspects and prediction and of its sentiment polarity. For example, in the review, "The price is reasonable although the service is poor" [26]. The review contains two opposing viewpoints on the extracted attribute "price" and "service" of the same entity "restaurant". The task of aspect term extraction is to identify the implicit and explicit aspect terms in the given review to determine the overall product quality. In the initial works on ABSA, Toh& Wang (2014) [27] used Conditional Random Field (CRF) on SemEval-2014 dataset for the extraction of aspect terms from Restaurant and Laptop domains. The authors also utilized clustered features for unlabelled data along with the syntactic, semantic and lexical features. Bhamare and Prabhu (2021) [8] used multivariate filter based and selective dependency relations approach for the extraction of aspect terms on the restaurant datasets of SemEval, Kaggle and Yelp. The accuracy and F1 score for the aspect identification task achieved by the authors are 94.75% and 85.24% respectively by using a feature set generated from both the approaches.

The selection of features is one of the important step in the field of SA. In real-life applications, data representation frequently employs an excessive number of redundant features. Feature selection is the method of identifying key

[1] Ph.D. Research Scholar, Shobhit Institute of Engineering and Technology (Deemed-to-be University), Meerut, India

[2] Director, Sir Chhotu Ram Institute of Engineering & Technology, C.C.S. University, Meerut, India

[3] School of Computing Science & Engineering, VIT Bhopal University, Bhopal, India

features and deleting irrelevant features from the dataset. Akhtar et al (2017) [1] employed Particle Swarm Optimization (PSO), a wrapper-based feature selection method for extracting relevant features from textual reviews. The authors performed an experimental study on the SemEval dataset of restaurant and laptop reviews and achieved good results in terms of f-measure.

In the initial stage of this work, a linguistic feature set identified by DLIREC [27] is utilized for representation of every token in the given review for the task of aspect term classification. To select the most relevant and non-redundant features, we further employed a binary version of Grey Wolf Optimizer (GWO). This nature-inspired algorithm has gained popularity in recent years due to its ease of implementation and better exploration capability. The experimental study on aspect-term extraction is conducted on two SemEval datasets of laptop and restaurant reviews. The observed results show that the model could achieve f1 score of 86.3% and 71% on laptop and restaurant domains respectively. The study depicts that Parts-of-Speech tag, dependency relations, WordNet, Stop word and Named entity information are relevant features in classifying the token as an aspect term.

In the next stage of the work, the relevant aspect terms are extracted from the reviews using the optimal feature set identified in the initial stage. A list of top-k frequently used aspect terms are selected from the extracted aspect terms for sentiment classification [8]. Each review sentence is scanned further to find the sentiment polarity of the top-k terms using Valence Aware Dictionary for Sentiment Reasoning (VADER). VADER [9] is popularly used open source rule-based sentiment analyser which calculates the polarity score depending on the intensity of words used in the sentence. Recent studies show that it achieved accurate results on social media data without the need of prior training. In addition, the lexicon-based approach used in VADER makes it more efficient in terms of computational cost. The sentiment polarity computed by VADER is used to generate an aspect- based summary of the given review dataset.

The following are the key contributions of the proposed work:

- Selection of optimal feature set for aspect term classification using BGWO
- Generation of aspect-based summary of laptop review dataset using VADER
- Performance evaluation of proposed approach using three different classifiers- Logistic Regression, SVM and Naïve Bayes classifiers on two SemEval datasets of laptop and restaurant reviews in terms of accuracy and f1 score

The rest of the paper is structured as: Section 2 discusses some of the related work in the field of ABSA. The overview of the proposed work employed in our study is described in Section 3. The findings from the experimental results on the review dataset is explained in Section 4. The conclusion of the proposed approach is provided in the last section.

II. LITERATURE SURVEY

There are various approaches utilized by the researchers in the field of ABSA. This section focuses on a comprehensive literature review that includes aspect term extraction and feature selection techniques.

Asriet. al. (2023) [6] utilised PSO approach on drug reviews to choose the optimal features and compare the model's performance with Ant Colony Optimisation (ACO) and Genetic Algorithms (GA). It was found that the PSO model outperformed the other two models in terms of accuracy, precision, recall and F-score. . Gaber et. al. (2023) [14] performed experiment on WUSTL-IIOT-2021 Dataset and proposed a unique intrusion detection model for the categorization of harmful behaviours in IIoT-based network traffic. They utilized PSO and Bat Algorithm (BA) for feature selection. It was found that the BA approach performed well as compare to PSO with RF classifier in terms of accuracy, recall, precision and f1-score. The authors in this chapter [7] discussed the use of Evolutionary Algorithm (EA) and Swarm Intelligence (SI) in the health care system. The contribution of various researchers who have employed SI and EA to reduce training time and enhance classification performance by removing redundant and unnecessary features has been discussed. These approaches seem to be particularly beneficial for feature selection in the classification of medical data. Rachdian et al (2022)[23], the researchers scrapped the reviews of laptop domain by using Selenium scrapping tool. An algorithm was developed for determining customer needs using topic modelling with Non-negative Matrix Factorization (NMF) and lexicon-based method analysis of online reviews. The Term Frequency - Inverse Document Frequency (TF-IDF) approach was utilised for feature extraction and it was found that the screen, battery, pricing, storage, performance, and keyboard are the important features having positive polarity as per the customers review. Almazini&Ku-Mahamud (2021) [3] proposed an improved binary grey wolf optimisation (EBGWO) technique for feature selection in anomaly detection. The experiment was conducted by the authors on NLS-KDD dataset. It was found that the maximum accuracy score of 87.46% was obtained by the proposed approach when compared with binary variant of bat algorithm and particle swarm optimisation approach.

Most existing systems use heuristic-based methods to identify the most relevant set of features. Shahraki et al (2021) [21] proposed an upgraded model of GWO termed as I-GWO. This approach uses the dimension learning based hunting strategy which enriches the balance between the global and local search. Alqaryouti et al (2020) [4] extracted both the implicit and explicit aspect terms from the data generated through government mobile applications using a unified lexicon and rule-based approach. The results of the SVM classifier on the selected features could achieve an accuracy score of 93.01%. Salam & Ali (2020) [24] used a hybrid approach of Extreme learning machine and GWO. They compared their proposed model with PSO and Ant Colony Optimization (ACO) based feature selection methods and it was found that their approach showed better performance in terms of accuracy, precision, recall and f-score on Twitter dataset using SVM classifier. Cat Swarm Optimization with a long short-term memory neural network was utilized by Alarifi et al (2020) [2] for selection of optimal features. The experimental results depict high accuracy as compared to PSO for aspect term extraction.

Despite the fact that several text classifiers exist, the large dimensionality of the feature space poses a significant issue. The extracted features may be noisy, less informative, or redundant that may increase the computational time and reduce the performance of the classifiers. Deng et al (2019) [12] provided an extensive review on filter, wrapper, embedding and hybrid techniques for feature selection utilized in text classification. Faris et al (2018) [13] summarized the contribution of various researchers' on GWO in different fields. The authors mentioned the operations performed in GWO along with its different versions. They also explained the GWO's application and its future directions. Manek et al (2017)[18] utilized a statistical feature selection method named as Gini index along with Support Vector machine on the movie review dataset for aspect based sentiment classification

The empirical approach for the aspect term extraction was proposed by Asghar et al. (2017) [20]. The fusion of corpus-based and lexicon based techniques were employed by the authors for the classification of aspect related sentiments. When compared with other existing methods, their proposed approach was found better in terms of precision. Anand&Naorem (2016) [5] employed clustering, manual labelling and review guided clustering approaches for the selection of aspect terms and their results showed that the manual method was most effective for aspect term extraction as well as the sentiment prediction.

Yonghe et al (2015) [17] had proposed two modified models of PSO on the basis of fixed constriction factor and functional inertia weight. It was observed that the asynchronous inertia weight and constriction factor performs better as compared to synchronous method. Due to imbalance between exploration and exploitation in swarm-based optimization algorithms. Gupta et al (2015) [15] employed an evolutionary algorithm PSO for feature selection on SemEval 2014 dataset. They had achieved the accuracy score of 71.25% and 78.48% respectively on the laptop and restaurant domain. Chandrashekar and Sahin (2014) [10] discussed different feature selection techniques like filter, wrapper and embedded methods in their study and compared the performance of few feature selection methods on seven different datasets using Support Vector Machine (SVM) and Radial Basis Function (RBF). A new dataset was created from three publically available datasets in the work [22] by Pavlopoulos&Androutsopoulos (2014). They used an unsupervised method along with a continuous space vector of the words and phrases for the extraction of aspect terms. Mirjalili et al (2014)[19] proposed GWO algorithm inspired by the nature of grey wolves. The authors compared the performance of GWO with five different heuristic methods named as PSO, DE, GSA, EP and ES and showed that their proposed optimizer could achieve better results on both constrained and unconstrained problems.

Zahran et al (2009) [25] utilized PSO for feature selection on the Arabic dataset. In this work, the parameters of PSO were tuned empirically and the performance was compared with methods such tf-idf, df, chi-square and RBF without feature selection.

III. PROPOSED WORK

The two main tasks focused in this study are optimal feature selection and aspect-based summarization. For selecting optimal features, we utilized the binary variant of GWO approach and the polarity detection of the top extracted frequent aspect terms is performed using VADER. The architecture of the proposed work is shown in fig 1. This subsequent subsection explains the methodology used in this work.

A. Data collection and Pre-processing

In the initial stage, the experiment is performed on publically available SemEval 2014 dataset of two different domains restaurant and laptop for the task of aspect term extraction and polarity detection.The pre-processing step starts with sentence tokenization of each review and then word tokenization to extract every token in a sentence. Lemmatization is performed to extract the root words from the review documents. The Part-Of-Speech (POS) tagging is applied to assign tags to each token of the review. The syntactic analysis of the sentences in the review is performed using Stanford dependency parser.

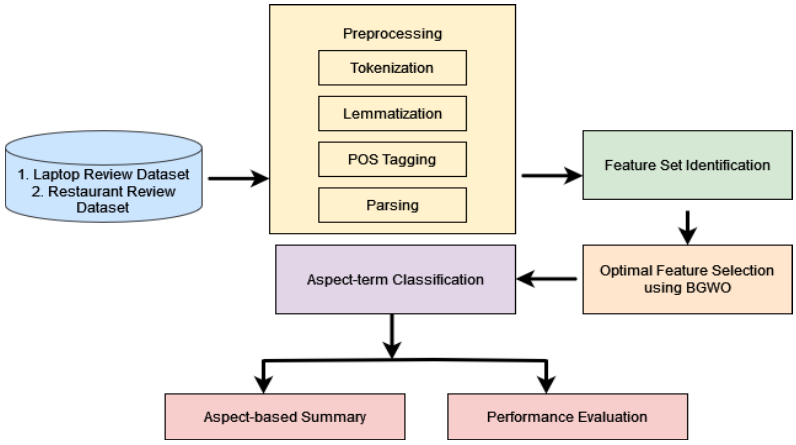


Fig 1: Architecture of proposed approach

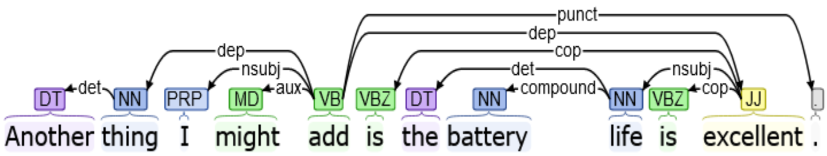
B. Features for Aspect term representation

In order to reduce the model complexity, a set of linguistic features commonly used in the task related to NLP are selected. The features used for the representation of each token in the review documents are as follows:

1) Head word: A high number of constituent words in noun phrases have the potential to be an aspect term. A binary feature used indicates if the current token is a noun phrase's head word. The terms that belong to a noun phrase are assigned value 1 else 0.

2) POS of the head word: The head word defines the nature of the review. The POS tagging of the head word is used as one of the features.

3) Dependency relation: The dependency relation helps to determine the syntactical relationship between the words in the review. The Stanford parser is used in our work to find the relationship between the head word and its dependent token.



For example, in the review sentence, “Another thing I might add is the battery life is excellent”, the token battery life is dependent on excellent with relation as ‘nsubj’. Therefore, only the token that correlates to dependent "nsub" will be activated.

4) Wordnet: Different words that are conceptually related are grouped into synsets in Wordnet[16]. The Synset information allows the model to group tokens with similar senses. Only noun synsets are considered in this study. This feature is especially important for recognising an undetected aspect.

5) Word Cluster: Word clustering is a technique for organising a collection of words into clusters where the words are semantically related to each other. In this work, Brown clustering technique is used which generate the binary valued feature vector of length 5.

6) Stop word: As stop words does not have any significance role to play for the identification of aspect terms, they can be safely ignored without threatening the sentence's meaning. A feature is introduced in this study which will assign a value as 1 or 0 depending on whether or not it is a stop word.

7) Word length: A token's length may be useful in detecting aspect phrases. If the length of the token exceeds a predefined threshold, a binary-valued feature is set to be high. The token having character more than 5 is considered to indicate the high probability of being an aspect term.

8) Chunk information: Aspect terms may contain multiple words. To identify the borders of aspect terms, the concept of chunk information is used which provides the start and end points of the chunk. Each token is assigned a tag as I, O and B depending on the position in the chunk. The prefix B before a tag denotes the start of a chunk. The prefix I denotes that it is included within a chunk. The tag O denotes that the token is not part of any chunk.

9) Semantic orientation (SO) score: An opinion's semantic orientation on a feature indicates whether it is positive, negative, or neutral. The sentiment orientation (SO) score for each word is calculated using Point-wise Mutual Information, which is a measure of token correlation with positive or negative word. The sentiment score[1] for t token is calculated as

SO(t) = PMI(t, pos_rev) – PMI (t, pos_rev) (1)

and$PMI(t, pos_rev) = \log \frac{(freq(t,pos_rev)*N)}{freq(t)*M}$ (2)

where PMI (t, pos_rev) represents the PMI score in relation to the positive review for token t. The value of SO score lies between -5 to 5.

10) Frequent Aspect Term: From the training dataset, the frequent aspect terms are extracted and compiled into a list. A value 1 or 0 is assigned if the current token is the member of the list. The threshold value is considered as 5 in this work.

11) Named entity (NE): It aims to locate and classify named entities in text into different categories. It is considered as one of the features which assign a binary value that triggers when a token is a named entity.

The finalized feature file generated is of size m x n where m and n represents the number of tokens and n is the total number of features. The value of m is 54062 and 49885 for laptop and restaurant respectively. The total number of features used are 20 in the proposed work.

C. Feature Selection using BGWO

In the proposed approach, the binary variant of GWO is used to generate the optimal feature set. This subsection provides the overview to BGWO approach.

1) Introduction to BGWO

Mirjalili et al. [16] introduced GWO in the year 2014. It is one of the meta-heuristic algorithms which is based on grey wolf's leadership structure and their hunting techniques. To imitate the hierarchical leadership, wolves are divided into four categories named alpha, beta, delta, and omega. The first, second, and third best individuals are labelled as alpha, beta, and delta respectively, while the rest of the individuals are labelled as omega. The optimal solution is directed by alpha, beta, and delta. Three best solutions are used to evaluate the likely location of the prey in an iterative searching process. The solution of the proposed approach is constrained to binary {0, 1} values depending on whether the feature is selected or not.

Before hunting the prey, the wolves must first encircle it. The mathematical model for encircling the prey is given by:

$$\overrightarrow{X}(t+1) = \vec{X}p(t) + \vec{A} \cdot \vec{D} \quad (3)$$

where t indicates the iteration number, $\vec{X}p$ represents position of the prey, $\vec{X}$ represents the position of grey wolf. The value of $\vec{D}$, $\vec{A}$ and $\vec{C}$ are calculated in (4), (5) and (6) as:

$$\vec{D} = \left|\vec{C} \times \vec{X}p(t) - \vec{X}(t)\right| \quad (4)$$

$$\vec{A} = 2a.\overrightarrow{r_1} - a \quad (5)$$

$$\vec{C} = 2\overrightarrow{r_2} \quad (6)$$

In eq. (5) and (6), r_1 and r_2 are the random binary vectors. The value of ‘a’ is linearly reduced from 2 to 0. The first three best solutions produced will obligate the remaining search agents to update their locations in accordance with the best search agents' positions. The updated equation will be given by (7) as

$$X_i^{t+1} = Crossover\ (x_1, x_2, x_3) \quad (7)$$

where Crossover (x_1, x_2, x_3) represents cross over between the solutions x_1, x_2 and x_3 and the values of x_1, x_2 and x_3 indicates binary vectors reflecting the influence of wolf movement towards the alpha, beta, and delta grey wolves. The value of x_1, x_2 and x_3 are calculated as shown in (8), (11) and (14) respectively.

$$x_1^d = \begin{cases} 1, if\ (x_\alpha^d + bstep_\alpha^d) \geq 1 \\ 0 \quad , \ otherwise \end{cases} \tag{8}$$

where x_α^d represents the alpha's position vector in d dimension and $bstep_\alpha^d$ indicates binary step in d dimension which is calculated as shown in (9).

$$bstep_\alpha^d = \begin{cases} 1, if\ cstep_\alpha^d \geq r \\ 0 \ , \ otherwise \end{cases} \tag{9}$$

where r is the random number generated from uniform distribution of 0 and 1 and $cstep_\alpha^d$ represents the continuous valued step size for d dimension and can be computed as shown in (10).

$$cstep_\alpha^d = \frac{1}{1+e^{-10(A_1^d D_\alpha^d - 0.5)}} \tag{10}$$

where the value of $D_\alpha^d\ and\ A_1^d$ are evaluated using (4), and (5) for dimension d are

$$x_2^d = \begin{cases} 1, if\ (x_\beta^d + bstep_\beta^d) \geq 1 \\ 0 \quad , \ otherwise \end{cases} \tag{11}$$

where $D_\alpha^d\ and\ A_1^d$ represents the beta's position vector in d dimension and indicates binary step in d dimension as shown in (12).

$$bstep_\beta^d = \begin{cases} 1, if\ cstep_\beta^d \geq r \\ 0 \ , \ otherwise \end{cases} \tag{12}$$

where r is the random number generated from uniform distribution of 0 and 1 and $cstep_\beta^d$ represents the continuous valued step size for d dimension and can be computed as shown in (13).

$$cstep_\beta^d = \frac{1}{1+e^{-10(A_1^d D_\beta^d - 0.5)}} \tag{13}$$

where the value of $D_\beta^d\ and\ A_1^d$ are evaluated using (4), and (5) for dimension d.

$$x_3^d = \begin{cases} 1, if\ (x_\delta^d + bstepx_\delta^d) \geq 1 \\ 0 \quad , \ otherwise \end{cases} \tag{14}$$

wherex_δ^d represents the delta's position vector in d dimension and $bstep_\delta^d$ indicates binary step in d dimension as shown in (15).

$$bstep_\delta^d = \begin{cases} 1, if\ cstep_\delta^d \geq r \\ 0 \ , \ otherwise \end{cases} \tag{15}$$

where r is the random number generated from uniform distribution of 0 and 1 and $cstep_\delta^d$ represents the continuous valued step size for d dimension and can be computed as shown in (16).

$$cstep_\delta^d = \frac{1}{1+e^{-10(A_1^d D_\delta^d - 0.5)}} \tag{16}$$

where the value of $D_\delta^d\ and\ A_1^d$ are evaluated using eq. (4), and (5) for dimension d. To apply crossover on a, b and c solutions, a crossover technique is employed per dimension, as shown in (17).

$$x_d = \begin{cases} a_d, if\ r < 1/3 \\ b_d, \quad \frac{1}{3} \leq \mathrm{r} \leq \frac{2}{3} \\ c_d, otherwise \end{cases} \tag{17}$$

Where, a_d, b_d and c_d are the binary values for the first, second, and third parameters in dimension d, x_d is the d's dimension crossover output, and r is a random number generated from a uniform distribution between 0 to 1.

2) Fitness Function

The selection of the optimal feature set is determined by two factors- accuracy and feature subset size. In order to find the relevant features, the fitness function f used in the proposed approach is given below in (18).

$$f(x) = \alpha\,(1 - P) + (1 - \alpha)\,(N - fs/N) \tag{18}$$

where x is the given input, α is a hyper parameter chosen between 0 and 1, P is the performance of the classifier using the feature set x, fs is size of the feature subset and N is the total number of features. The value of α is used to balance the trade-off between the performance of the classifier and the feature set size.

D. Performance Evaluation

To prove the efficacy of the proposed approach, the performance of the proposed model is evaluated on the basis of different performance metrics that are discussed below:

Precision: It is defined as the number of class predictions that actually belong to the positive class as shown in (19).

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{19}$$

Recall: It is defined as the number of class predictions that are categorised as positive out of all positive samples in the dataset. It can be computed by using (20).

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{20}$$

F-measure: It is computed by taking the harmonic mean between precision and recall as shown in (21).

$$F - measure = \frac{2 * \mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision+Recall}} \tag{21}$$

Accuracy: It is defined as the total number of correct predictions divided by the total number of predictions generated for the given samples. It is calculated by using (22).

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+FN+TN+FP}} \tag{22}$$

TP represents the number of correctly classified positive tuples, FN represents the number of positive tuples that are incorrectly classified, FP stands for the number of wrongly classified negative tuples and TN is the number of accurately classified negative tuples.

E. Aspect-based Summary

The binary variant of GWO explained in the previous subsection is chosen for selecting optimal feature set. The optimal feature set is used to extract the aspect terms from the laptop reviews. The aspect terms found in maximum number of reviews is taken into consideration for generating aspect based summary. The sentiment polarity of top-k aspect terms is computed using VADER. VADER returns the sentiment scores of each phrase related to the identified aspect term in the dataset. It computes the sentiment score of the phrase in four different categories – positive, negative, neutral and compound. Positive, negative and neutral indicates probability of a sentence to be as positive, negative and neutral. Compound score returns normalized score by adding all sentiment's score. Finally, the average of compound score of all phrases is used to generate aspect-based summary.

IV. RESULT AND DISCUSSION

This section focuses on the description of the dataset used, experimental set up and performance of the proposed model. The comparison with state-of-the-art approaches is also discussed in this section.

A. Dataset Description

The benchmark dataset used in this study is SemEval-2014. The dataset contains user-generated reviews from the two domains: restaurant and laptop. Table 1 represents the description of the dataset.

Table 1: Description of dataset

	Restaurant		Laptop	
	Train	**Test**	**Train**	**Test**
Number of Reviews	3044	800	3045	800
Number of Aspect terms	3699	1134	2358	654

The visual representation of most frequent occurring terms of restaurant and laptop domain are shown in fig 2 in the form of word cloud. It has been observed from the fig 2 that 'food', 'place' and 'service', and 'computer', 'laptop' and 'use' are some of frequent terms of restaurant and laptop dataset respectively.

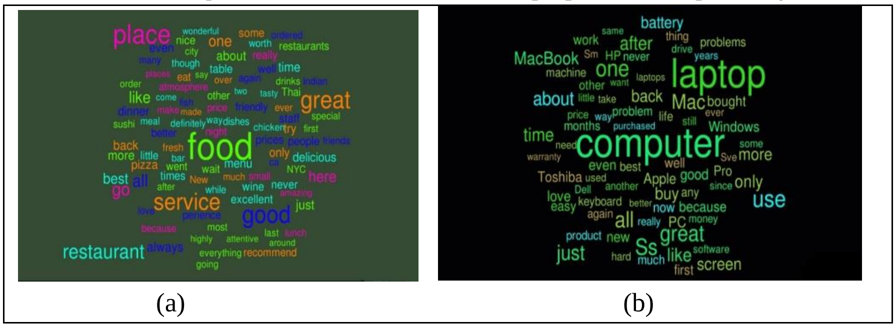


Fig 2. Word cloud of (a) restaurant and (b) laptop reviews

B. Experimental Setup

Anaconda with Python 3.4 is used to implement the modules of our proposed work. For the initial steps of pre-processing and feature extraction, NLTK [11] is utilized. The initial step involves pre-processing the datasets to eliminate relevant information and remove XML tags. In this work, we extract essential details such as Part-of-Speech (PoS), named entity (NE), and lemma from the reviews using the Stanford CoreNLP suite. For handling multi-word token, we employed IOB scheme where 'I', 'B' and 'O' represent the internal, beginning, and outside tokens of the aspect term respectively. The machine learning package of python named scikit-learn (sklearn) is used for implementing aspect classification algorithms. Linear SVM, Bernaulli NB and Logistic regression are used for model training and testing in all experiments using the default settings of sklearn. For implementing the binary version of GWO, code is written in python and combined with sklearn for token classification. Some parameters are established for experimental study in the proposed approach as shown in Table 2.

Table 2: Parameter Settings

Parameters	Value
a	0.6
Total no. of wolves	20
No. of runs	10
No. of iterations	50

C. Performance Evaluation of Aspect Term Classification

The performance of the proposed work is evaluated on SemEval datasets of Laptop and Restaurant reviews in terms of precision, recall, accuracy and f- measure. Table 3 shows the performance of three different classifiers - LR, NB and SVM using BGWO as feature selection method. As observed from Table 3, BGWO with LR classifier could achieve a best precision score of 95% and f-measure of 92% on the laptop dataset. The best accuracy score of 84% is obtained by SVM classifier.

Table 3: Performance metric on laptop and restaurant reviews for aspect term extraction

Approach	Classifier	Laptop				Restaurant			
		Precision	**Recall**	**F- measure**	**Accuracy**	**Precision**	**Recall**	**F- measure**	**Accuracy**
BGWO	LR	**95**	77	**92**	82	**98**	51	62	51
	NB	91	82	76	78	92	75	58	61
	SVM	95	79	91	84	99	74	**93**	**79**

In addition to this, SVM outperforms other classifiers and achieves precision and accuracy score of 99% and 79% respectively on restaurant reviews.

To prove the robustness of the proposed approach, the feature optimizer are run 10 times using Logistic Regression to compute best, average and worst fitness values and are recorded as shown in Table 4. It has been observed from table 4 that the best fitness value is achieved on restaurant reviews.

Table 4: Fitness values on laptop and restaurant dataset

Run	Laptop		Restaurant	
	Fitness score	**No. of optimal features**	**Fitness score**	**No. of optimal features**
1	0.71	12	0.73	10
2	0.65	14	0.74	12
3	0.75	10	0.77	8
4	0.75	10	0.80	8
5	0.76	13	0.73	9
6	0.74	11	0.72	10
7	0.73	12	0.73	10
8	0.72	13	0.76	8
9	0.72	10	0.72	9
10	0.68	13	0.71	11
Best fitness	0.76		0.80	
Average fitness	0.72		0.74	
Worst fitness	0.65		0.71	

Due to randomness of BGWO, features do change over 10 iterations. The average of selected features are recorded as shown in Table 5 on both laptop and restaurant dataset. It has been observed from the table that Parts-of-Speech tag, dependency relations, WordNet, Stop word and Named Entity information are important features in categorising the token as an aspectterm.

Table 5: Optimal feature set (for runs=10)

Dataset	**Head Word (HW)**	**POS of HW**	**Dependency relation**	**Entity**	**Wordnet**	**Word Cluster**	**POS + 1**	**POS + 2**	**POS - 1**	**POS - 2**	**Stop word**	**Word Length**	**Chunk Information**	**SO score**	**Frequent Aspect Term**	**Aspect term**	**Named entity**	**IOB**	**Orthographic IsCap**	**Orthographic IsDigit**
	F1	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	**F10**	**F11**	**F12**	**F13**	**F14**	**F15**	**F16**	**F17**	**F18**	**F19**	**F20**
Laptop		✓	✓		✓			✓		✓	✓		✓	✓		✓	✓	✓		
Restaurant	✓	✓	✓		✓	✓					✓						✓			

*POS + 1, POS + 2 , POS - 1 and POS - 2 represents POS of next word, POS of second next word, POS of the previous word and POS of second previous word respectively

The bar graph in fig. 3 presents analysis of the number of features selected by varying the number of iterations of the laptop and restaurant dataset.

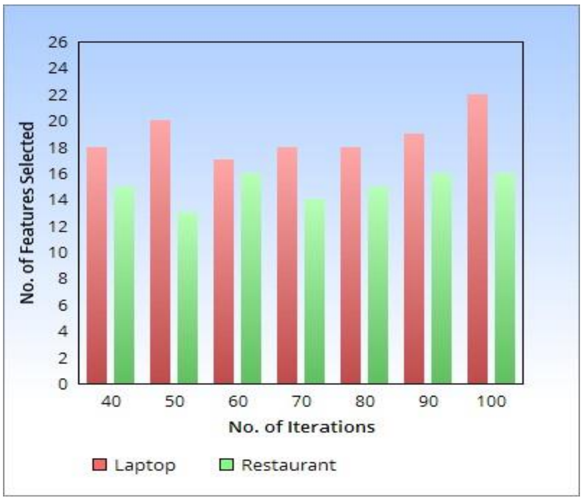


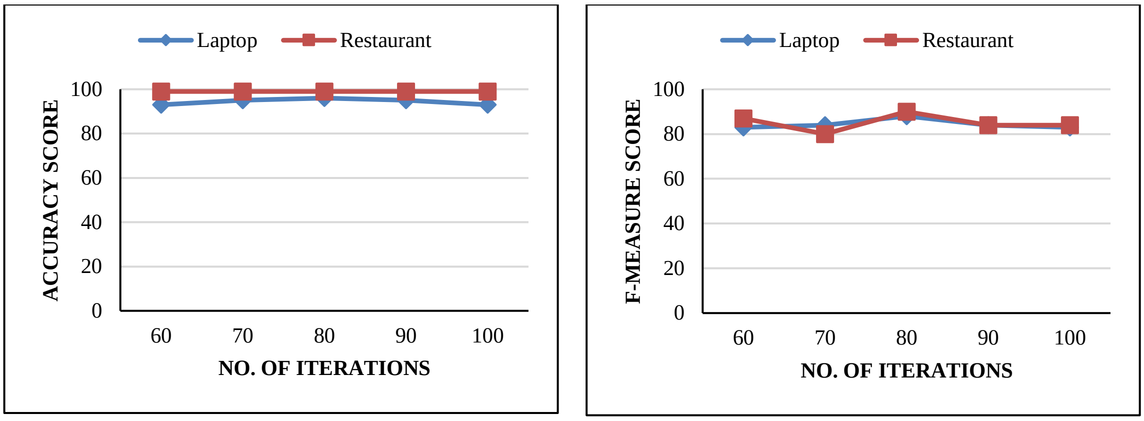
Fig. 3 Analysis of no. of features selection by varying iterations

The feature optimizer was run over 5 times by varying the value of ‘a’ parameter as 0.5, 0.6, and 0.7 on the laptop dataset. It was observed from the table 6 that the maximum accuracy was achieved at a = 0.5 by using SVM classifier.

Table 6: Performance metrics on varying the values of a (for runs=5)

a=0.5	Precision	77	83	86	77	75	92	72	85	85	78	83	87	83	84	78	78	82	86
	Recall	92	73	91	93	71	95	89	72	92	92	73	91	94	79	90	92	74	92
	F-measure	83	76	89	83	73	94	78	76	88	83	76	89	87	81	82	83	77	89
	Accuracy	95	89	96	95	89	98	94	88	96	95	89	96	96	93	95	95	90	97
	Features		18			13			15			16			17			16	
a=0.6	Precision	79	72	88	79	81	79	77	78	90	73	70	71	79	82	89	78	77	84
	Recall	92	63	93	94	70	93	92	67	94	94	79	88	93	71	95	93	70	93
	F-measure	84	65	91	85	74	84	83	70	92	79	74	76	84	75	92	83	72	87
	Accuracy	95	82	97	95	88	95	95	85	97	94	92	93	95	89	97	95	88	96
	Features		13			22			18			10			19			16	
a=0.7	Precision	83	82	73	80	83	76	71	82	91	85	81	83	75	78	70	79	82	79
	Recall	94	73	90	93	75	88	91	70	96	94	79	90	91	76	88	93	75	91
	F-measure	88	76	79	85	78	81	77	74	93	88	80	86	81	71	76	84	76	83
	Accuracy	96	90	94	95	90	94	94	87	98	96	92	95	94	91	93	95	90	95
Classifier		LR	BNB	SVM	LR	BNB	SVM	LR	BNB	SVM	LR	BNB	SVM	LR	BNB	SVM	LR	BNB	SVM
Iteration			1			2			3			4			5			**Average**	

The line graph represents a comparison of accuracy and f-measure score with different no. of iterations in laptop and restaurant reviews. As observed from fig 4, the maximum accuracy is achieved by restaurant dataset.



(a) (b)

Fig. 4 Performance metrics (a) Accuracy score (b) F-measure score with different no of iterations

D. Aspect-based summary of Laptop reviews

In the last set of experiments, the polarity of the top five frequent aspect terms of the laptop dataset is identified using VADER. Table 7 represents aspect terms along with the polarity of few review samples of laptop reviews.

Table 7: Review Samples on laptop domain

Review	Aspect Term	Polarity
Another thing I might add is the battery life is excellent.	Battery	Positive
I continued to take the computer in AGAIN and they replaced the hard drive and mother board yet again.	Hard Drive	Negative
The keyboard feels good and I type just fine on it.	Keyboard	Positive
The screen is very large and crystal clear with amazing colors and resolution.	Screen	Positive
This laptop is a great price and has a sleek look.	Price	Positive

The compound score of VADER is used to generate the aspect-based summary. The polarity of the aspect term is categorized as positive, negative or neutral depending on the value of compound score which lies between -1 to 1. If the value of compound score is greater than 0.05, then the polarity is set to be as positive. If the compound score lies between 0.05 to -0.05, the neutral polarity of the phrase is assigned. The score less than -0.05 indicates negative polarity. Based on human evaluation on test samples of laptop domain, we had found around 60% samples of aspect term is correctly predicted by VADER. The graphical representation of aspect-based summary of top frequent aspect terms of laptop reviews is shown in fig. 5.

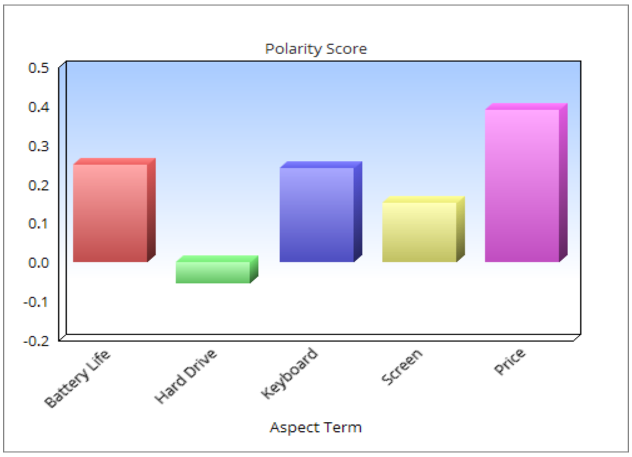


Fig 5. Aspect-based summary of top frequent aspect terms

As observed from the fig. 2, 'price', 'battery life', 'keyboard' and 'screen' are the most liked positive features of the laptop dataset. The negative score of the attribute 'hard drive' represents the feature of laptop not liked by the reviewers.

E. Comparison of proposed work with existing works

Table 8 shows a comparison of our proposed approach for aspect term extraction with other related works in the field of ABSA. The identified feature set and dataset employed in our work is similar to all the related works used for comparative analysis. Toh& Wang [27] identified the feature set for aspect term extraction and categorized them into constrained DLIREC and unconstrained DLIREC. Their experimental results showed that the unconstrained DLIREC achieved good results as compared to the constrained system. Gupta et al [15] used BPSO for selecting relevant features and employed CRF for classification. Akhtar et al [1] extended their work and employed BPSO for feature selection and ensemble approach for classification.

Table 8: Comparison of Proposed work with other related works

Author	Approach Used	Classifier(s) Used	F-Measure
Z. Toh & W. Wang [27]	DLIREC	CRF	Laptop: 73.78 Restaurant: 84.01
Akhtar et al [1]	Feature selection and Construction of classifier ensemble using BPSO	Maximum Entropy (ME), CRF and SVM	Laptop: 72.78 Restaurant: 83.11
Gupta et al [15]	BPSO based feature selection	CRF	Laptop: 81.91 Restaurant: 72.42
Proposed work	BGWO based feature selection	LR, NB and SVM	Laptop: 92 Restaurant: 93

As compared with the other related works, our proposed approach achieves the highest f-measure score of 92% on laptop domain and 93% on restaurant domain with LR and SVM classifier respectively.

V. CONCLUSION

In this paper, a binary version of GWO is utilized for feature selection to improve the performance of aspect term extraction. The proposed approach starts by identifying a set of features that is used for classification of aspect terms. The identified features are then optimized using a binary version of grey wolf optimization using a fitness function that maximizes accuracy and minimizes feature set size. The performance of the proposed model is evaluated using three popular learning algorithms - LR, NB and SVM on SemEval datasets of restaurant and laptop reviews. The experimental findings show that the proposed model achieves better performance in terms of f-measure on both the domains. The aspect-based summary of top five frequent aspect terms for laptop domain is generated. This summary can help the organizations and customers in better decision making. In future, other nature-inspired algorithms could be utilized to improve the efficacy of the proposed model for feature selection. The work can be further extended by utilizing deep learning techniques like BERT and GPT to improve the performance of ABSA. The proposed work can be employed further to generate Aspect-based summary on other domains such as medical and tourism.

REFERENCES

[1] Akhtar MS, Gupta D, Ekbal A, Bhattacharyya P. Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. Knowledge-Based Systems. 2017; 125: 116-135.

[2] Alarifi A, Tolba A, Makhadmeh, ZA, Said W. A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks. The Journal of Supercomputing. 2020; 76(6): 4414-4429.

[3] Almazini H, Ku-Mahamud KR. Grey Wolf Optimization Parameter Control for Feature Selection in Anomaly Detection. International Journal of Intelligent Engineering & Systems. 2021; https://doi.org/10.22266/ijies2021.0430.43.

[4] Alqaryouti O, Siyam N, Monem AA, Shaalan K. Aspect-based sentiment analysis using smart government review data. Applied Computing and Informatics. 2019; https://doi.org/10.1016/j.aci.2019.11.003.

[5] Anand D, Naorem D. Semi-supervised aspect based sentiment analysis for movies using review filtering. Procedia Computer Science. 2016; https://doi.org/10.1016/j.procs.2016.04.070.

[6] Asri AM, Ahmad SR, Yusop NM. Feature Selection using Particle Swarm Optimization for Sentiment Analysis of Drug Reviews, International Journal of Advanced Computer Science and Applications. 2023; http://doi.org/10.14569/IJACSA.2023.0140530.

[7] Awotunde JB, Adeniyi AE, Ajagbe SA, Jimoh RG, Bhoi AK. Swarm Intelligence and Evolutionary Algorithms in Processing Healthcare Data. In: Mishra S, González-Briones A, Bhoi AK, Mallick PK, Corchado JM, editors. Connected e-Health. Studies in Computational Intelligence; 2022. vol 1021. Springer, Cham. https://doi.org/10.1007/978-3-030-97929-4_5.

[8] Bhamare BR, Prabhu J. A supervised scheme for aspect extraction in sentiment analysis using the hybrid feature set of word dependency relations and lemmas. PeerJ Computer Science. 2021; https://peerj.com/articles/cs-347.
[9] Borg A, Boldt M. Using VADER sentiment and SVM for predicting customer response sentiment. Expert Systems with Applications. 2020; https://doi.org/10.1016/j.eswa.2020.113746.
[10] Chandrashekar G, Sahin F. A survey on feature selection methods. Computers & Electrical Engineering. 2014; https://doi.org/10.1016/j.compeleceng.2013.11.024.
[11] D'Aniello G, Gaeta M, La Rocca I. KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis. Artificial Intelligence Review. 2022; https://doi.org/10.1007/s10462-021-10134-9.
[12] Deng X, Li Y, Weng J, Zhang J. Feature selection for text classification: A review. Multimedia Tools and Applications. 2019;https://doi.org/10.1007/s11042-018-6083-5.
[13] Faris H, Aljarah I, Al-Betar MA, Mirjalili S. Grey wolf optimizer: a review of recent variants and applications. Neural computing and applications. 2018; https://doi.org/10.1007/s00521-017-3272-5.
[14] Gaber T, Awotunde JB, Folorunso SO, Ajagbe SA, Eldesouky E. Industrial Internet of Things Intrusion Detection Method Using Machine Learning and Optimization Techniques. Wireless Communications and Mobile Computing. 2023; https://doi.org/10.1155/2023/3939895.
[15] Gupta DK, Reddy KS, Ekbal A. Pso-asent: Feature selection using particle swarm optimization for aspect based sentiment analysis. In: International conference on applications of natural language to information systems. Springer, Cham; 2015. pp. 220-233.
[16] Liu B. Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers; 2012.
[17] Lu Y, Liang M, Ye Z, Cao L. Improved particle swarm optimization algorithm and its application in text feature selection. Applied Soft Computing. 2015; https://doi.org/10.1016/j.asoc.2015.07.005.
[18] Manek AS, Shenoy PD, Mohan MC., Venugopal KR. Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World wide web. 2017; 20(2), 135-154.
[19] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Advances in engineering software. 2014; https://doi.org/10.1016/j.advengsoft.2013.12.007.
[20] Asghar MZ, Khan A, Zahra SR, Ahmad S, Kundi FM. Aspect-Based Opinion Mining Framework using Heuristic Patterns. Cluster Computing. 2017; https://doi.org/10.1007/s10586-017-1096-9.
[21] Nadimi-Shahraki MH, Taghian S, Mirjalili S. An improved grey wolf optimizer for solving engineering problems. Expert Systems with Applications. 2021; https://doi.org/10.1016/j.eswa.2020.113917.
[22] Pavlopoulos J, Androutsopoulos I. Aspect term extraction for sentiment analysis: New datasets, new evaluation measures and an improved unsupervised method. In: Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM), 2014; pp. 44-52.
[23] Rachdian AO, Suryadi D, Fransiscus H. Identification of Customer Needs for Laptop in a Product Reviews using the Aspect-Based Sentiment Analysis Method. 2022; https://dx.doi.org/10.12785/ijcds/1201111.
[24] Salam MA, Ali M. Optimizing Extreme Learning Machine using GWO Algorithm for Sentiment Analysis. International Journal of Computer Applications. 2020; 176 (38): 22-28.
[25] Zahran B, Kanaan GG. Text feature selection using particle swarm optimization algorithm. World Applied Sciences Journal. 2009; 7: 69-74.
[26] Zhang L, Liu B. Sentiment Analysis and Opinion Mining. In: Sammut C., Webb GI, editor. Encyclopedia of Machine Learning and Data Mining. Springer, Boston: MA; 2022. https://doi.org/10.1007/978-1-4899-7687-1_907.
[27] Toh Z, Wang W. Dlirec: aspect term extraction and term polarity classification system. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). pp. 235–240.