International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Novel approach for Aspect Based Opinion Mining using Enhanced Grey Wolf Optimization

Shilpi Gupta[1*], Niraj Singhal[2]

Submitted: 07/05/2024 **Revised**: 19/06/2024 **Accepted**: 26/06/2024

Abstract: The online reviews based on aspects shared by the users have become a common source of customer knowledge on purchasing choices and businesses pursuing opinions. In this work, a new aspect-based opinion mining technique with four major modules: Pre-processing, Aspect sentiment extraction, Aspect grouping and aspect-sentiment classification (ASC) is introduced. Initially, the collected reviews from users are pre-processed via the following steps: Tokenization, Lemmatization, Stemming and Stop word removal. Subsequently, the aspect sentiment extraction is accomplished by following two major phases: POS tagging and Rule based lexicon extraction. The extracted aspects are grouped into primary and second aspect based on the computed semantic similarity score. The mining becomes more precise by giving the dual weightage to the aspects in the review. Accordingly, the weight function is multiplied to both the primary and second aspect. Further, the proposed model tunes the weighting factors of primary and secondary aspects to make the classification more appropriate with respect to opinion. For this, a new Enhanced Grey Wolf with Mutation Operation (EGWMO) model is proposed in this work. Subsequently, the weight optimized primary aspect; weight optimized secondary aspect as well as extracted opinion is considered as the extracted features for final classification via Neural Network. From NN, the sentiments of the reviews are classified as: positive sentiment, extreme, neutral sentiment, extreme positive sentiment or extreme negative sentiment.

Keywords: *Aspect Grouping, Aspect Sentiment Extraction, Data Mining, Grey Wolf Optimization (GWO), NN based Sentiment Classification, Opinion Mining*

1. Introduction

Social networking platforms have experienced a dramatic evolution, enabling internet consumers to express their opinions [1] on health care, goods, policy, policy and services. If a person wants to make a decision about the purchasing of any product, the opinion is generally requested from neighbours, family members, and colleagues. Similarly, opinion polling, surveys and focus groups are conducted by a company when they need inputs from their consumers about their programs and goods [2][3][4]. The extraction, interpretation and description of opinions is carried out using automated opinion mining techniques from a vast number of reviews and as a result, the consumers are easily supported by the interested opinion [5][6].

The key goal of opinion mining[7][8][9] or sentiment analysis (SA) is to classify the feelings, positive and negative views on the target items such as people, subjects or goods. Opinion mining is divided into three levels: sentence-based, document-based and aspect-based. The key role in the aspect-based level is to derive opinion goals (features) and words of opinion [10]. In reality, these opinions are given for specific aspects of the product, such as feature, element or performance, and hence retrieving the word of opinion [11]. Meanwhile, the user's mood is also explored by a fine-grained study of emotion about the attribute called the aspect words. Therefore, the sentiment analysis is basically said to be the means of collecting terms and words of opinion from the review [12][13][14]. The meaning of the word aspects is provided as the significant feature of the product as per the customer's ranking. The combination of optimistic and critical feedback on different facets of the products is referred to product reviews.

Various research works have been developed for the successful extraction of opinion targets [15] and words of opinion. Some of the conventional approaches used for opinion mining are Naive Bayes, k-mean clustering [16], Bayesian Classifier, Support Vector Machine (SVM), and so on. However under certain regards, these classical algorithms neglect the opinion mining, which should be rectified by the introduction of modern methods in the future. Due to the strongly aspect-dependent nature of emotion, the general-purpose sentiment lexicon is always unfavourable. In many situations, the traditional document-level approach does not result with meaningful aspect-based opinion mining [17]. Therefore, it is necessary to create the learning algorithm that does not require a significant number of labelled training data. The primary contribution made by this research work is as follows:

- Introduces a dual weightage (primary and secondary aspect weight) based aspect grouping.

[1] *Shobhit Institute of Engineering and Technology (Deemed-to-be University), Meerut, India. Email: er.shilpibansal@gmail.com*
[2] *Sir Chhotu Ram Institute of Engineering and Technology, Chaudhary Charan Singh University, Meerut, India*

- Fine-tuning of weights (primary and secondary aspect) are carried out by the Enhanced Grey Wolf with Mutation Operation (EGWMO) model

The remaining section of work is organized as: The literature work is discussed in section 2. The proposed Aspect based opinion mining approach is discussed in Section 3. Section 4 covers the steps for pre-processing and Aspect Sentiment Extraction. The proposed aspect grouping and neural network based opinion mining task is discussed in section 5. The acquired results are discussed comprehensively in Section 6. Section 7 discusses the conclusion of the work.

2. Literature Review

2.1 Related Work

This section covers the contributions of different researchers done in the field of aspect based sentiment analysis.

In 2016, Zuo et al. [18] have built a narrative tool called Cross-collection Auto-labeled MaxEnt-LDA (CAMEL) with the goal of transversely balancing aspect-based opinion mining for asymmetric sets. By modelling similar and unique features around the samples, CAMEL was able to gain additional details as all the corresponding views were held for the purposes of comparison analysis. In addition, a modern auto-labelling system called Auto-labeled MaxEn (AME) was proposed for the appearance and opinion of words without intricate human labelling. In addition, the improvement was rendered by introducing a new functionality called word embedding-based similarity. A non-parametric alternative on the basis of the coupled Dirichlet processes (DP) known as CAMEL-DP was also introduced. The thorough investigation of the revamped evidence from the multi-collections of the real world has strengthened the proposed frame work.

In 2017, Asghar et al. [19] had introduced the integrated work consisting of a large range of heuristic patterns for the hybrid sentiment classification portion with extra assistance for intensifiers and negations, the description generator and the extraction aspect. The simulation result has clarified that the output estimate of the applied aspect-based opinion mining method offered an improvement over the other conventional models in terms of recall, F-measure and precision. The proposed framework has given an improved outcome in the comparative result in terms of grouping, generation of a description and extraction of sentiment in the online product reviews. In 2017, Tran and Phan [20] have developed a hybridization technique to mine features and opinion terms on the basis of a sophisticated algorithm, namely double propagation by inserting rules. This was achieved by investigating the semantic relationship of those laws and ontology's of standard speech, and many parts of speech in sentences. The features have been minimized by deploying the HITS algorithm. The experimental result over several data sets has confirmed that the applied approach has produced a consistent and satisfactory outcome.

In 2016, Wang et al. [21] have contrasted the ten strategies for collecting aspect-opinions from seven realms over Chinese companies. Some of the comparative models include MNB-based opinion mining, KNN-based opinion mining, TF-based opinion mining plus POS, RNN-based opinion mining, RFM-based opinion mining, CART-based opinion mining, CRF-based opinion mining, LPM-based opinion mining, SVM-based opinion mining and HMM-based opinion mining. The following outcome has been revealed by the experiment (a) machine learning technique has provided betterment over rule-based methods; (b) algorithm has not dominated over the entire domains; (c) the accuracy of opinion mining was affected negatively by the length of text affects over the rule-based techniques, whereas few of the machine learning models possessed a better mining on long reviews; (e) Best performance was provided by the SVM-based approach among the entire domains on opinion mining.

In 2013, Xueke et al. [22] have concerned the enhancement of aspect-level viewpoint mining for online consumer analysis. Initially, a novel generative theme model namely the JAS model was introduced to jointly undermine aspects and sentimental lexicons based on an aspect-dependent on online consumer feedback. The usefulness of the proposed JAS model inside the functional values of extracted lexicons and learning-dependent emotion lexicons has been successfully demonstrated by the findings of the investigation.

In 2011, Zhu et al. [23] have learnt about the aspect-based opinion polls as a result of unlabelled free-form textual consumer feedback. Initially, to be used for aspect recognition, a suggestion for a multi-faceted bootstrapping approach was made for the analysis of aspect-related terminology of each aspect. Second, a multi-faceted sentence was segmented into several one-sided units using the suggested aspect-based segmentation model for opinion polls. Finally, a detailed description of the opinion polling algorithm based on the aspect was added. The real-time investigative study has found that the suggested methodology could be more effective in terms of aspect-based opinion polling activities.

In 2018, DungVo et al. [24] have introduced a model for mining and summarizing the corresponding thoughts and product aspects from a vast number of product reviews within a given domain. The simulation result has shown that the applied structure has obtained better F1-scores for camera and laptop ratings.

In 2015, Fang et al. [25] have proposed a mmAOM to fix multimodal problems – opinion mining for organizations by exploiting various cross-collection channels of social media.

This approach was introduced by considering the text documents and user-generated images to simultaneously capture the interaction between visual and textual forms, as well as the links between opinions and aspects. The MMAOM was introduced to the multimodal aspect-opinion retrieval and individual association visualization by deciding the aspects and related opinions correlated with the entities. The experiment was performed systematically using individuals in Tripadvisor ratings, news stories and Flickr images. In 2023, Makhadmeh et al [26] provided the analysis on four years paper of Grey Wolf Optimizer (GWO). The authors also discussed the variants of GWO inthis review paper. The authors in [27] presents a text summarising method based on Binary Grey Wolf Optimisation (BGWO). The proposed method outperformed the other existing approach.

2.2 Review

Table 1 reveals the features and challenges of the conventional Aspect-based opinion mining techniques. Various researchers are made to extend the reviews on opinion features and opinion words. However, there are some disadvantages are there in the conventional works that needs rectification. Some of the conventional methods with features and challenges are explained as follows: CAMEL [18] has a capability of integrating complementary information and are suitable for real –time applications. However they are lacks in exploring the AME scheme to all kinds of opinionated texts and needs to provide better precision rate. SWN [19] provides better precision and can classify the aspect-based opinions in multiple domains. Though the limitation in this is given as: Needs elimination in redundant terms and there is a need for efficient post-processing technique. HITS [20] have effective extraction of opinion targets and opinion words as well as have extended opinion lexicons. However, needs more regular expression rules construction and the robustness in sentiment dictionary needs development. RFM [21] is good in extracting long reviews and performs better recall, whereas the challenges of these are poses low accuracy and machine learning algorithms require annotation manually. JAS [22] has provides high precision polarities for opinion words and extremely simple and efficient. Though, the main challenge of this methodology is needs better identification on opinion words and the aspect-aware sentiment polarities needs betterment. Multi-Aspect Bootstrapping Method [23] has better multi-aspect sentence handling is and learning in aspect-related terms. Needs further investigation on effective content-based rating inference techniques and focus on sentiment lexicon domain adaptation for sentiment analysis, these are the two drawbacks of this technique. LDA [24] extracts dependency relations and also detects additional annotations. However, needs improvement over accuracy and sampling procedure. mmAOM [25] poses an effectiveness in entity association visualization. Though, needs improvement in the effectiveness of mmAOM and the flexibility. Further, the major challenge that is focusing in this proposed model is the consideration of both primary and secondary aspect for classifying the sentimental opinion, this is because in many of the conventional works, only the primary aspect is considered alone that may not provide precise classification output.

Table 1. Advantages and Drawbacks of Existing work: A Review

Author [citation]	***Methodology***	***Features***	***Challenges***
Zuo *et al.* [18]	CAMEL	▪ Capable of integrating complementary information ▪ Suitable for real –time applications	▪ Think to explore the AME scheme to all kinds of opinionated texts ▪ Needs better precision
Asghar *et al.* [19]	SWN	▪ Better precision	▪ Needs elimination in redundant terms ▪ There is a need for efficient post-processing algorithm
Tran and Phan [20]	HITS	▪ Effective extraction of opinion targets and opinion words ▪ Opinion lexicons are extended	▪ Needs to construct more regular expression rules ▪ Robustness in sentiment dictionary needs development
Wang *et al.* [21]	RFM	▪ Good in extracting long reviews ▪ Performs better recall	▪ Low accuracy ▪ Machine learning algorithms require annotation manually
Xueke *et al.* [22]	JAS	▪ Provides high precision polarities for opinion words ▪ Extremely simple and efficient	▪ Needs better identification on opinion words ▪ Aspect-aware sentiment polarities needs betterment

Zhu *et al.* [23]	Multi-Aspect Bootstrapping Method	▪ Multi-aspect sentence handling is better ▪ Better learning in aspect-related terms	▪ Investigate effective content-based rating inference techniques
DungVo *et al.* [24]	Latent Dirichlet Allocation (LDA)	▪ Extracts dependency relations ▪ Detects additional annotations	▪ Accuracy needs to be improved ▪ Needs better sampling procedure
Fang *et al*. [25]	mmAOM	▪ Efficient in Multimodal aspect-opinion retrieval	▪ the effectiveness of mmAOM further needs improvement ▪ the flexibility needs enhancement

3. Aspect based Opinion Mining (ABOM) Approach

A new aspect-based opinion mining technique with four major modules: Pre-processing, Aspect sentiment extraction, Aspect grouping and ASC is introduced in this research work.

Step 1: Initially, the collected reviews $\left(Data^{in}\right)$ are pre-processed via the following steps: Tokenization, Lemmatization, Stemming and Stop word removal. The pre-processed data $Data^{pre-process}$ is then subjected to aspect based sentiment extraction.

Step 2: The aspect sentiment extraction is followed by two major phases: POS tagging and Rule based lexicon extraction. The extracted opinions and aspects are denoted as $D^{opinion}$ and D^{aspect}, respectively.

Step 3: The next move is to group the related aspects. The suggested methodology of aspect grouping is based on the weights of the extracted D^{aspect}. Initially, the semantic similarity score is computed for D^{aspect}, from which the aspects with higher similarity score are said to be secondary aspects, while the aspects with lower similarity score are said to be primary aspects. In order to have a more precise review, a weight function W is multiplied with the primary aspect and secondary aspect.

Step 4: Thereby, the dual weighting factor, the weight function of primary aspect $W^{Primary}$ and weight function of secondary aspect $W^{Secondary}$ are fine-tuned using a newly introduced EGWMO model.

Step 5: Finally, the weight optimized primary aspect $\left(W_{opt}^{Primary}\right)$, weight optimized secondary aspect $\left(W_{opt}^{Secondary}\right)$ as well as extracted $D^{opinion}$ are considered as the features F, which are fed as input to the sentiment classification phase.

Step 6: As per the proposed architecture, the sentiment classification is done by NN, where it classifies: positive sentiment, extreme, neutral sentiment, extreme positive sentiment or extreme negative sentiment.

4. Pre-processing and Aspect term Extraction

4.1 Pre- Processing

In general, the online text contains huge count of uninformative parts like the scripts, HTML tags and advertisements, as well as huge noise. Further, at the word level, most of the words do not have a huge impact on the general orientation of it. These words might lead to the problem of higher dimensionality, which reduces the classification accuracy. Therefore, the collected reviews $\left(Data^{in}\right)$ are pre-processed via the following steps: Tokenization, Lemmatization, Stemming and Stop word removal. Fig.1 shows the steps followed in Pre-processing.

a) **Tokenization:** It is the process of splitting the collected raw data $Data^{in}$ into words, sentences called tokens before converting it into vectors. In general, the tokenization aids in interpreting the meaning of the text by evaluating the sequence of the words.

For illustration: If the text in $Data^{in}$ is “It is raining”, them the tokens generated from it is ‘It’, ‘is’, ‘raining’.

b) **Lemmatization** [28]: The lemmatization is done to find the base words for the tokens. The lemmatization is accomplished via morphological analysis and vocabulary that identifies the base form of the word. The lemmatization is similar to stemming, but it takes account of context of the word. For illustration: the lemma of ‘raining’ token is ‘rain’. In addition, the lemma of ‘are’, ‘is’, ‘being’ → ‘be’.

c) **Stemming** [28]: The suffix in the words is removed, and the stemming is good in reducing the count of required computations, and therefore it is preferred in this research work. The stemming of word ‘Eating’ is word ‘eat’.

d) **Stop word removal** [28]: The stop words are the words that have no information. The most commonly said stop words are “the”, “a”, “an”, “in”, which are to be removed to reduce the computational burden. The pre-processed data $Data^{pre-process}$ is subjected to aspect based sentiment extraction.

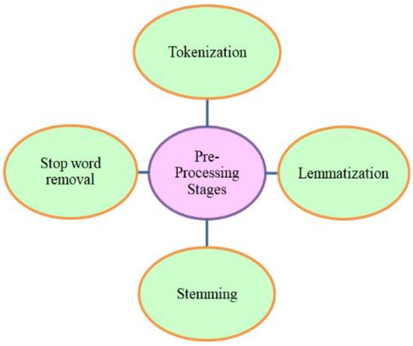


Fig. 1. Steps involved in pre-processing stage

4.2 Aspect Sentiment Extraction

The aspect and sentiment of $Data^{pre-process}$ is detected and extracted. The aspect sentiment extraction is accomplished under two major phases:

Phase1: The aspects and sentiments are extracted from $Data^{pre-process}$ using the POS tagging [19]. The PoS tagging is defined as the process of assigning parts of speech to words of $Data^{pre-process}$. The PSO includes the "nouns, verb, adverbs, adjectives, pronouns, conjunction and their sub-categories".

For illustration: if a sentence "The pictures are absolutely amazing" is fed as input to the POS tagger, then the POS tagger tags the words "amazing" as an adjective and "pictures" as a noun. The sentence can now be represented as "The/DT pictures/NNS are/ VBP absolutely/RB amazing/JJ".

Phase2: Rule based lexicons are used to extract the aspects efficiently [19]. Consider the example: "The pictures are absolutely amazing". In this example, the word "amazing" and the word "pictures" denote the opinion and the aspect, respectively. The aspects that are extracted and the opinions associated in the POS-tagged sentences are stored in an Excel file for further processing. The extracted opinions are denoted as $D^{opinion}$. The extracted aspects are denoted as D^{aspect}, which are then subjected to aspect grouping mechanism.

5. Proposed EGWMO Model for Primary and Secondary Weight Tuning

5.1 Aspect Grouping

The next move is to group the related aspects after the extraction of the aspects. Typically, using different words or phrases, users share their views on the same aspect. For e.g., "battery", "battery life" and "battery use" reflect multiple phrases in the cell phone dataset that relate to the same feature, i.e., "battery." These terms (synonyms) referring to the aspects should be clustered in order to create an accurate aspect-based opinion review. The suggested methodology of aspect grouping is based on the weights of the extracted D^{aspect}, which is computed based on its SSC. The SSC is computed in the Google news vector library via the cosine similarity, which computes the similarity between two vectors by computing the cosine of the angle between the two vectors. The cosine formula is mathematically shown in Eq. (1).

$$\cos = \frac{\sum_{i}^{n} A_i . B_i}{\sqrt{\sum_{i}^{n} (A_i .)^2} * \sqrt{\sum_{i}^{n} (B_i .)^2}}$$

(1)

Here, the notation A_i denotes the topic document and the review document is denoted as B_i.

Then, from the computed results, the aspects with higher similarity score are said to be secondary aspects, while the aspects with lower similarity score are said to be primary aspects. In this research work, if the computed similarity score> 0.2, then it is considered as the secondary aspect, whereas the rests are considered as the primary aspect. The secondary aspect provides the exact review of the product. In order to make the grouping more precise, the weight function W lies within the limit [0, 1] is multiplied with the computed cosine similarity score. Moreover, the weight function of primary aspect $W^{\Pr imary}$ and weight function of secondary aspect $W^{Secondary}$ are fine-tuned using a newly introduced EGWMO model, which is the improved version of standard WOA. Fig. 2 manifests the steps followed in aspect grouping and classification.

5.2 Proposed EGWMO Model for Primary and Secondary Weight Tuning

Grey Wolf Optimizer (GWO) [29] was developed on the motivation gained from the hunting mechanism of grey wolve's. In general, there are four different categories of grey wolves based on their hunting as well as leadership hierarchy. The conventional GWO algorithm encloses amazing facts in position update of the solutions in the search space. Yet, it suffers from the shortcomings like low convergence rate and precision. Dealing with the complex optimization issues require necessary adaptiveness in the existing model [30][31][32][33][34]. Thus, the enhancement of GWO is introduced in this work namely EGWMO model. The steps followed in the proposed EGWMO model is depicted below:

Step 1: The population pop of search agent is initialized. The current iteration is denoted as $iter$ and the maximal count of generation is Max^{iter}. In addition, the value of a (a parameter that lessens from 2 to 0 over $iter$), coefficient vectors X and coefficient vectors Y is calculated as shown in eq. (2), (3) and (4) respectively.

$$a_1 = 2 - 1 * \left(\frac{2}{t_{\max}} \right)$$
(2)

$$X = 2a.r_1 - a$$
(3)

$$Y = 2r_2$$
(4)

Step 2: Calculate the search agent's fitness by using eq. (12).

Step 3: Set v_α as the 1st best search agent, v_η as the 2nd best search agent and other are neglected.

Step 4: While $\left(iter < Max^{iter}\right)$do

Step 5: Revise the position of each search agent based on the first and the second best search agents using eq. (5). Here, the additional weight-ages are given to α , which is the most leading wolf. Here,

$$Z_\alpha = | Y_1 \cdot v_\alpha - v | \quad ,$$

$$Z_\eta = | Y_1 \cdot v_\eta - v |$$

$$v(t+1) = \frac{2v_1 + v_2}{6}$$
(5)

in which,

$$v_1 = v_\alpha - X_1.(Z_\alpha)$$
(6)

$$v_2 = v_\eta - X_2.(Z_\eta)$$
(7)

Step 6: Meanwhile, the solution gets multiplied via the mutation operators. In general, the mutation is a random tweak in the chromosome to get new solutions. The mutation is accomplished to maintain and introduce diversity in the genetic population and is usually applied with a low probability. In EGWMO, the mutation is undergone using Eq. (8).

$$v_{Mut}^i = v_{g.best}^i + K \times v_{ran1}^i - v_{ran2}^i \tag{8}$$

In which, v_{ran1}^i and v_{ran2}^i denotes the selected random vectors from pop. In addition k is a coefficient that lies between the value 0 to 1.

Step 7: Update the value of a, X and Y

Step 8: Compute the fitness of the search agent using Eq. (12).

Step 9: Update the value of v_α and v_η

Step 10: Increment the value of $iter$ by 1

Step 11: Return v_α

Step 12: Terminate

The pseudo-code of EGWMO model is shown in Algorithm 1.

Algorithm 1*:*Pseudo code of EGWMO model
Initialize: pop, Max^{iter}, a, X, Y, $iter$
Compute the fitness of the search agent using Eq. (12).
Set v_α as the 1st most excellent search agent
Set v_η as the 2nd paramount search agent
While $\left(iter < Max^{iter}\right)$do
For each wolf
Update the position of each search agent using Eq. (5).
Perform mutation using Eq. (8).
End for
Compute the fitness of the search agent using Eq. (12).
Update the value of v_α and v_η
$iter = iter + 1$
End while
Return v_α

Finally, the weight optimized primary aspect $\left(W_{opt}^{\Pr imary}\right)$, weight optimized secondary aspect $\left(W_{opt}^{Secondary}\right)$ as well as extracted $D^{opinion}$are the features F fed as the input to classification phase.

$$F = W_{opt}^{\Pr imary} + W_{opt}^{Secondary} + D^{opinion} \tag{9}$$

5.3 NN for Opinion Mining-sentiment analysis

In this work, NN classifier is used to provide the final classification results. The input to NN is the extracted features F . NN model consists of three layers: "input layer with input neuron $inp = 1,2,\cdots N_{inp}$, hidden layer with hidden neurons $hid = 1,2,\cdots N_{hid}$, and output layer with output neuron $out = 1,2,\cdots N_{out}$ ". The NN's network model is shown in Eq. (10) and Eq. (11), respectively. The notation N_{inp} , N_{hid} and N_{out} denotes the count of input, hidden and output neuron, correspondingly. The bias weight of hid and out is depicted as $wg^{N}_{bias.hid}$ .and $wg^{r}_{bias.out}$, respectively. The weight from inp^{th} to hid^{th} is $wg^{N}_{inp,hid}$, and wg^{r}_{zo} is the weight from hid^{th} to out^{th} . In Eq. (12), the error function $F(er)$ is the difference between the predicted output OUT_{pre} and actual outputs OUT_{act} . The activation function in Eq. (10) and Eq. (11) is represented as AF .

$$HIDDEN = AF\left(wg^{N}_{bias,hid} + \sum_{inp=1}^{N_{inp}} wg^{N}_{inp,hidz} .F \right) \tag{10}$$

$$OUT_{pre} = AF\left(wg^{r}_{bias.out} + \sum_{hid=1}^{N_{hid}} wg^{r}_{hid.out} .HIDDEN \right) \tag{11}$$

$$F(er) = \underset{\{wg^{N}_{bias.hid}, wg^{N}_{inp,hid}, wg^{r}_{bias.out}, wg^{r}_{hid.out}\}}{\arg\min} \sum_{out=1}^{N_{out}} \left| OUT_{act} - OUT_{pre} \right| \tag{12}$$

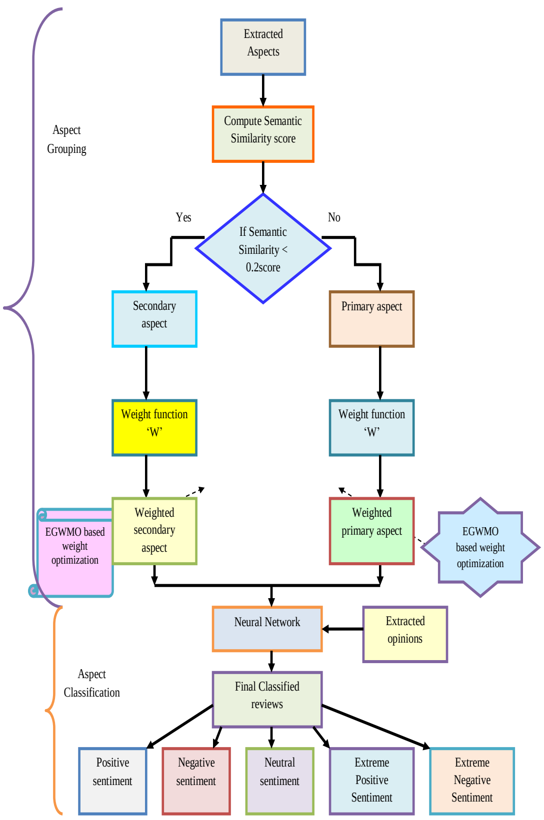


Fig. 2. Aspect Grouping and Aspect Classification

6. Results and Discussion

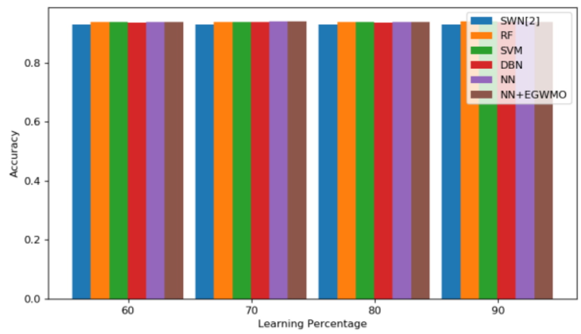
This section covers the implementation details of the proposed work. The dataset used for the evaluation was collected from https://www.kaggle.com/datatattle/covid-19-nlp-text-classification and the implementation of the work is done in python. The review classification is accomplished here using the proposed work (EGWMO+ NN). The performance of EGWMO+ NN was validated over existing classification techniques like SWN [19], RF, SVM, Deep Belief Network (DBN) and NN respectively, and it was undergone by varying the LP. As stated, the computed weight optimized semantic score plays the major role in review grouping. The performance of proposed EGWMO model is then compared with the existing optimization techniques [23] as PSO, MFO and GWO by varying the LP to 60, 70, 80 and 90, respectively. The performance analysis was carried out in terms of Accuracy, Precision, F-measure, Specificity, Sensitivity, False Positive Rate (FPR), False Negative Rate (FNR), Negative Predictive Value (NPV) and Mathews Correlation Coefficient (MCC).

6.1 Analysis on Aspect Classification Performance

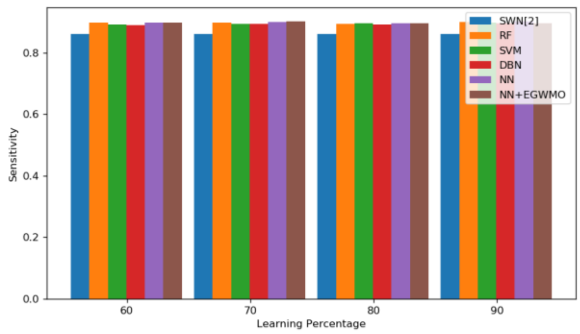
The performance of the EGWMO+ NN is compared over the existing classification techniques like SVN [2], RF, SVM, DBN and NN by means of varying the LP. The acquired results in terms of positive performance: accuracy, sensitivity, specificity and precision are exhibited in Fig.3. In general, a technique is proven to be good for aspect based opinion mining, if its positive measures are as high as possible. With the glance on the recorded accuracy results, it is observed that the accuracy of the EGWMO+ NN is superior to the existing works. At LP=60, the accuracy of EGWMO+ NN is 0.939, which is the highest score when compared to SVN= 0.92892, RF= 0.9390, SVM= 0.937, DBN= 0.936 and NN= 0.939. Similar to this, for every variation in the LP, the accuracy of the EGWMO+ NN model is found to be higher than the entire existing machine learning techniques. Further, at LP=60, the precision of the EGWMO+ NN model is 4.19%, 4.20%, 0.65%, 0.95%, 0.031% better than the existing machine learning techniques like SVN, RF, SVM, DBN and NN, respectively. Moreover, the specificity of the proposed work at LP=90 is 0.956, which is 0.425%, 0.425%, 0.049%, 0.0049% and 0.0941% better than the existing machine learning techniques like SVN, RF, SVM, DBN and NN, respectively. Moreover, on having a review on all the recorded values, the propose work seems to exhibit higher performance than all the existing techniques. Therefore, it is said to be much sufficient for aspect based opinion mining.

In addition, the performance of the proposed work under negative measure is evaluated over the existing works in terms of FNR and FPR, respectively. The corresponding results acquired are shown in Fig. 4. This evaluation is done by varying the count of LP. The FPR of the proposed work at LP=60 is 9.24%, 0.272%, 1.871%, 2.63% and 0.09% better than the existing machine learning techniques like SVN, RF, SVM, DBN and NN, respectively. In addition, at LP=80, the proposed work is 25.984%, 2.718%, 0.4828%, 4.690% and 0.563% better than the existing machine learning techniques like SVN, RF, SVM, DBN and NN, respectively. This outcome ensures the better performance of proposed model on mining aspect based opinions. In addition, the other measures like MCC, NPV and F-measure score shows the highest values as shown in Fig. 5.

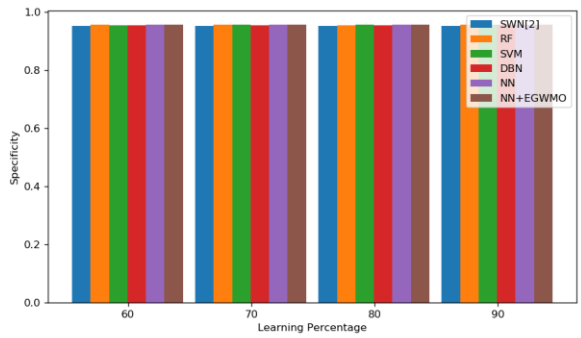
In addition, the overall performance of EGWMO+ NN model is evaluated in terms of positive negative and other measures, and the corresponding results acquired are tabulated in Table 2. The overall sensitivity of EGWMO+ NN model is 4.609%, 4.6095, 0.812%, 0.894% and 0.143% better than the existing machine learning techniques like SVN, RF, SVM, DBN and NN, respectively. Moreover, the overall accuracy of the proposed work is 1.239%, 1.239%, 0.252%, 0.276% and 0.045% better than the existing machine learning techniques like SVN, RF, SVM, DBN and NN, respectively. In addition, while evaluating all the other positive measures like precision and specificity, the proposed work had exhibited the higher values. In addition, the other measures too exhibit the highest performance with the proposed work. The NPV of the proposed work is 0.522%, 0.522%, 0.112%, 0.122% and 0.02% better than the existing machine learning techniques like SVN, RF, SVM, DBN and NN, respectively. On the other hand, the minimization of the classification error is the major objective behind the current research work, which is found to be achieved by the proposed work. The FNR of the proposed work is 10.48%, 1.54%, 2.45%, 2.681% and 0.45% better than the existing machine learning techniques like SVN, RF, SVM, DBN and NN, respectively. This has proved that the proposed dual weightage system makes the mining more precise and optimal with the incorporation of optimization algorithm.



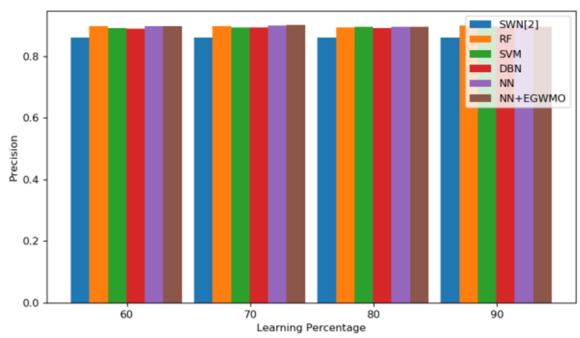
(a)



(b)

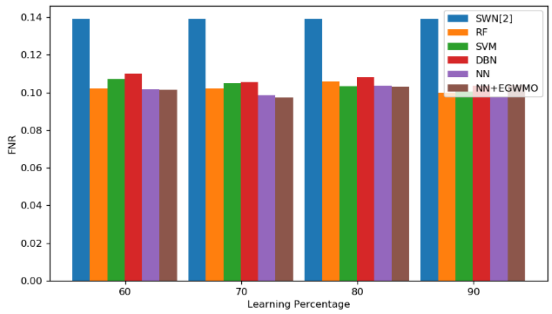


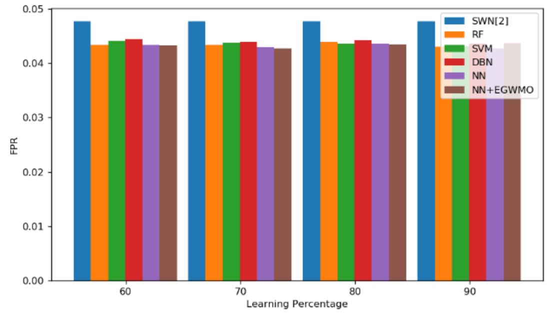
(c)



(d)

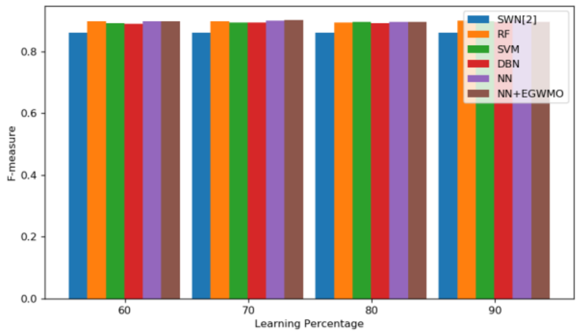
Fig. 3 Performance evaluation of the proposed and extant works in terms of (a) Accuracy, (b) Sensitivity, (c) Specificity and (d) Precision

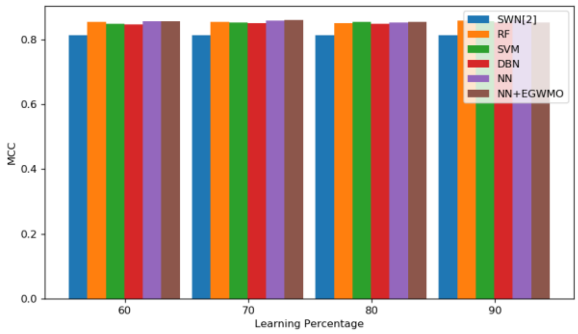




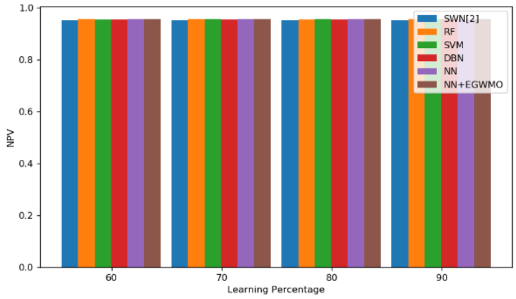
(a) (b)

Fig 4. Performance evaluation of the proposed and extant work in terms of (a) FNR and (b)FPR





(a) (b)



(c)

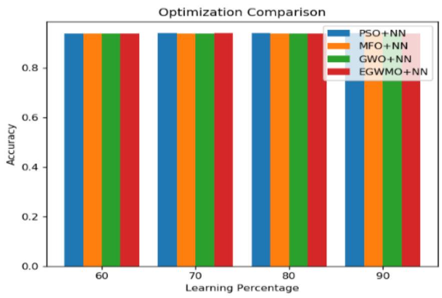
Fig. 5. Performance Analysis of the proposed in terms of (a) F-measure (b) MCC and (c) NPV

Table 2. Performance Analysis of Proposed and Conventional Works

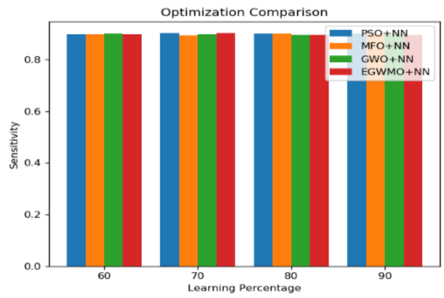
Measures	*SWN[2]*	*RF*	*SVM*	*DBN*	*NN*	*NN+EGWMO*
Sensitivity	0.8608	0.8980	0.8951	0.8944	0.9012	0.9025
Specificity	0.9522	0.9566	0.9562	0.9560	0.9570	0.9573
Accuracy	0.9289	0.9391	0.9382	0.9379	0.9401	0.9406
Precision	0.8608	0.8980	0.8951	0.8944	0.9012	0.9025
F-measure	0.8608	0.8980	0.8951	0.8944	0.9012	0.9025
MCC	0.8131	0.8546	0.8513	0.8505	0.8582	0.8598
NPV	0.9522	0.9566	0.9562	0.9560	0.9570	0.9573
FPR	0.0477	0.0433	0.0438	0.0439	0.0429	0.0427
FNR	0.1391	0.1019	0.1048	0.1055	0.0987	0.0975

6.2 Analysis on Optimized weights of primary and secondary aspects: Proposed Vs Existing Optimization

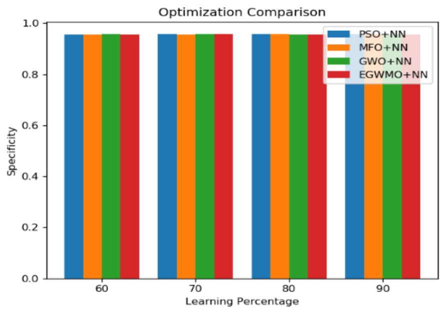
In this research work, the weight of primary and the secondary aspects are fine-tuned using the newly proposed EGWMO model, which is the improved version of standard WOA. In order to prove the betterment of proposed EGWMO model over the existing models like PSO, MFO and GWO, an evaluation is undergone. The corresponding results acquired in terms of positive, negative and other performance are shown in Fig. 7. At LP=60, the accuracy of the EGWMO+NN model is 0.9392, which is found to be the highest value, when compared to PSO+NN= 0.939, MFO+NN 0.93987 and GWO+NN= 0.9391. In addition, the highest accuracy recorded by EGWMO+NN model is 0.9403, which is recorded at the LP=90. Moreover, the sensitivity of the proposed work at LP=60 is 0.9019, while the specificity of PSO+NN= 0.902, MFO+NN= 0.899 and GWO+NN= 0.894. At LP=70, the MCC of the proposed works is 0.522 %, 0.522 % 0.112 %, 0.122 % and 0.020 % better than the existing techniques like PSO+NN, MFO+NN and GWO+NN, respectively. On analysing all the positive measures, it is proved that the proposed work is found to be the best for optimal tuning of weights, which aids in proper opinion mining. Then, the EGWMO+NN model also exhibits higher values are other measures, under all variations in the LP. The NPV of the proposed work is 0.957, which is the highest value than the existing techniques like PSO+NN= 0.9572, MFO+NN= 0.95680 and GWO+NN= 0.9561. On the other hand, the proposed work has recorded least FNR and FPR values for every variation in the LP. This is evident from Fig. 6. The performance of proposed model in terms of F-measure, MCC and NPV is depicted in Fig. 8. As a whole, the proposed work is suggested to be apt for aspect based opinion mining. In addition, the overall performance of the proposed work as well as existing work is evaluated, and the corresponding results acquired are tabulated in Table 3. As per the acquired results, the Sensitivity of the EGWMO+NN is 0.902, which's is 4.60%, 4.6096%, 0.812%, 0.89% and 0.143% better than the existing techniques like PSO+NN, MFO+NN and GWO+NN, respectively. The highest specificity recorded by the proposed work is 0.957. In addition, the Accuracy of the proposed work is 1.239%, 1.23%, 0.252%, 0.27% and 0.045% better than the existing techniques like PSO+NN, MFO+NN and GWO+NN, respectively. Moreover, the NPV of the proposed work is 0.957, which is found to be the highest value than the existing techniques. Moreover, the negative measures are found to be lower for the proposed work. The FNR of proposed work is 0.042, which is found to be the lowest value, and it is 10.48%, 1.544%, 2.45%, 2.681% and 0.45% better than the existing techniques like PSO+NN, MFO+NN and GWO+NN, respectively. Therefore, from the overall results, it is vivid that the proposed optimization model is significant in tuning the weights more appropriately. Hence, it becomes suitable for aspect based opinion mining.



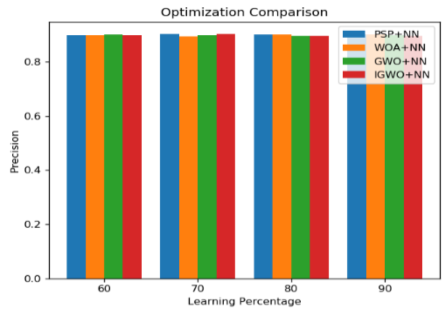
(a)



(b)

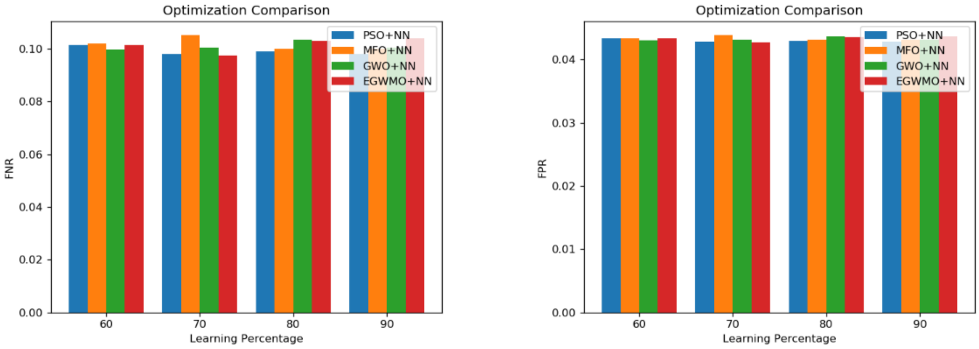


(c)



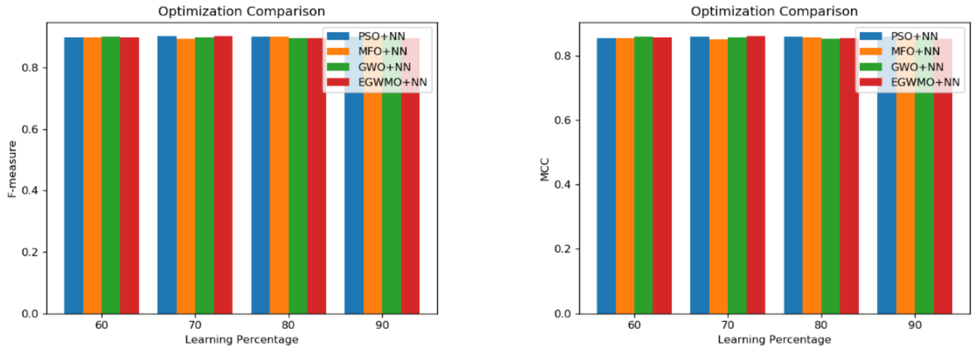
(d)

Fig. 6. Performance evaluation of the proposed and extant tactics in terms of (a) Accuracy, (b) Sensitivity, (c) Specificity and (d) Precision

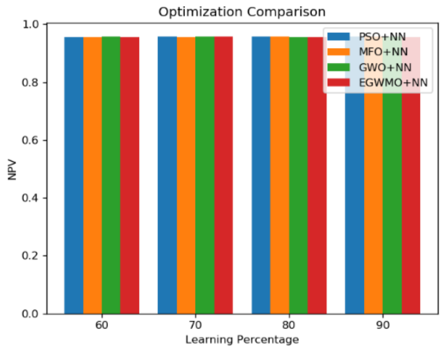


(a) (b)

Fig. 7. Performance evaluation of the proposed and extant tactics in terms of (a) FNR and (b) FPR



(a) (b)



(c)

Fig. 8. Performance evaluation of the proposed and existing works in terms of (a) F-measure, (b) MCC and (c) NPV

Table 3. Overall Performance of Proposed and Conventional Algorithms

Measures	***PSO+NN***	***MFO+NN***	***GWO+NN***	***EGWMO+NN***
Sensitivity	0.9024	0.8994	0.8948	0.9019
Specificity	0.9573	0.9568	0.9561	0.9571
Accuracy	0.9405	0.9395	0.9381	0.9404
Precision	0.9024	0.8994	0.8948	0.9019
F-measure	0.9024	0.8994	0.8948	0.9019
MCC	0.8597	0.8562	0.8509	0.8591
NPV	0.9572	0.9568	0.9561	0.9571
FPR	0.0427	0.0431	0.0438	0.0428
FNR	0.0975	0.1005	0.1051	0.0980

7. Conclusion

In this research work, a new effective Aspect-based Opinion mining technique was developed by following four major modules: “Pre-processing, Aspect sentiment extraction, Aspect grouping and ASC”. Initially, the collected reviews collected reviews were pre-processed via Tokenization, Lemmatization, Stemming and Stop word removal. The aspect sentiment extraction was accomplished with the aid of POS tagging and Rule based lexicon extraction. The semantic similarity score was computed for each of the extracted aspects, from which the aspects with higher similarity score are said to be secondary aspects, while the aspects with lower similarity score are said to be primary aspects. In order to have a more precise review, a weigh function was multiplied with the primary aspects and secondary aspects. The weight function of primary aspect and weight function of secondary aspect were fine-tuned using a newly introduced EGWMO model, which is the improved version of standard WOA. The weight optimized primary aspect; weight optimized secondary aspect as well as extracted opinion is the extracted features, which is utilized to train the Neural Network. From NN, the sentiments of the reviews are classified as: “positive sentiment, extreme, neutral sentiment, extreme positive sentiment or extreme negative sentiment”.

References

[1] Franziska Pröll," The German Media Start-up Opinary: Inventing Tools for Readers’ Opinion-Sharing Online—Reinventing User Engagement on News Websites", Publishing Research Quarterly, vol.33, no.1, pp 14–18, March 2017

[2] NikosTsirakis, VasilisPoulopoulos, PanagiotisTsantilas, IraklisVarlamis," Large scale opinion mining for social, news and blog data", Journal of Systems and Software, vol.127, pp.237-248, May 2017

[3] NecmiyeGenc-Nayebi, AlainAbran," A systematic literature review: Opinion mining studies from mobile app store user reviews", Journal of Systems and Software, vol.125, pp.207-219, March 2017

[4] Lipika Dey, Sk. Mirajul Haque," Opinion mining from noisy text data", International Journal on Document Analysis and Recognition (IJDAR), vol.12, no.3, pp 205–226, September 2009

[5] G. Shi, C. S. Chan, W. J. Li, K. Leung, Y. Zou and Y. Jin, "Mobile Human Airbag System for Fall Protection Using MEMS Sensors and Embedded SVM Classifier," IEEE Sensors Journal, vol. 9, no. 5, pp. 495-503, May 2009.

[6] Q. Huang and J. Yang, "Bayesian adaptive shrinkage in independent component domain," Electronics Letters, vol. 42, no. 1, pp. 58-77, 5 Jan. 2006.

[7] Shih-Jung Wu, Rui-Dong Chiang, Zheng-Hong Ji," Development of a Chinese opinion-mining system for application to Internet online forums", The Journal of Supercomputing, vol.73, no.7, pp 2987–3001, July 2017

[8] Li Chen, Luole Qi," Social opinion mining for supporting buyers’ complex decision making: exploratory user study and algorithm comparison", Social Network Analysis and Mining, vol.1, no.4, pp 301–320, November 2011

[9] P. Manivannan, C. S. Kanimozhi Selvi," Pairwise relative ranking technique for efficient opinion mining using sentiment analysis", Cluster Computing, pp 1–11, 23 February 2018

[10] K. Liu, L. Xu and J. Zhao, "Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model," IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 3, pp. 636-650, 1 March 2015.

[11] Malik Muhammad Saad Missen, Mohand Boughanem, Guillaume Cabanac," Opinion mining: reviewed from word to document level", Social Network Analysis and Mining, vol.3, no.1, pp 107–125, March 2013

[12] Dhekra Najar, Slim Mesfar," Opinion mining and sentiment analysis for Arabic on-line texts: application on the political domain", International Journal of Speech Technology, vol. 20, no. 3, pp 575–585, September 2017

[13] Elena Lloret, Alexandra Balahur, José M. Gómez, Andrés Montoyo, Manuel Palomar," Towards a unified framework for opinion retrieval, mining and summarization", Journal of Intelligent Information Systems, vol. 39, no. 3, pp 711–747, December 2012

[14] G. R. Brindha, P. Swaminathan, B. Santhi," Performance analysis of new word weighting procedures for opinion mining", Frontiers of Information Technology & Electronic Engineering, vol. 17, no. 11, pp 1186–1198, November 2016

[15] X. Zhou, X. Wan and J. Xiao, "CLOpinionMiner: Opinion Target Extraction in a Cross-Language Scenario," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 4, pp. 619-630, April 2015.

[16] H. Xiong, J. Wu and J. Chen, "K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 39, no. 2, pp. 318-331, April 2009.

[17] Rupam Gupta Roy, "Rescheduling Based Congestion Management Method Using Hybrid Grey Wolf Optimization - Grasshopper Optimization Algorithm in Power System", Journal of Computational

Mechanics, Power System and Control, Vol.2,No.1, pp.9-18,2019.

[18] Y. Zuo, J. Wu, H. Zhang, D. Wang and K. Xu, "Complementary Aspect-Based Opinion Mining," IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 2, pp. 249-262, 1 Feb. 2018.

[19] Muhammad Zubair Asghar, Aurangzeb Khan, Syeda Rabail Zahra, Shakeel Ahmad, Fazal Masud Kundi," Aspect-based opinion mining framework using heuristic patterns", Cluster Computing, pp 1–19, 21 August 2017

[20] Thien Khai Tran, Tuoi Thi Phan,"Mining opinion targets and opinion words from online reviews", International Journal of Information Technology, vol.9, no.3, pp 239–249, September 2017

[21] Wei Wang, Guanyin Tan, Hongwei Wang," Cross-domain comparison of algorithm performance in extracting aspect-based opinions from Chinese online reviews", International Journal of Machine Learning and Cybernetics, vol.8, no.3, pp 1053–1070, June 2017.

[22] X. Xu, X. Cheng, S. Tan, Y. Liu and H. Shen, "Aspect-level opinion mining of online customer reviews," in China Communications, vol. 10, no. 3, pp. 25-41, March 2013.

[23] J. Zhu, H. Wang, M. Zhu, B. K. Tsou and M. Ma, "Aspect-Based Opinion Polling from Customer Reviews," IEEE Transactions on Affective Computing, vol. 2, no. 1, pp. 37-49, Jan.-June 2011.

[24] A. Vo, Q. Nguyen and C. Ock, "Opinion–Aspect Relations in Cognizing Customer Feelings via Reviews," in IEEE Access, vol. 6, pp. 5415-5426, 2018.

[25] Q. Fang, C. Xu, J. Sang, M. S. Hossain and G. Muhammad, "Word-of-Mouth Understanding: Entity-Centric Multimodal Aspect-Opinion Mining in Social Media," in IEEE Transactions on Multimedia, vol. 17, no. 12, pp. 2281-2296, Dec. 2015.

[26] Makhadmeh, S. N., Al-Betar, M. A., Doush, I. A., Awadallah, M. A., Kassaymeh, S., Mirjalili, S., & Zitar, R. A. (2023). Recent advances in Grey Wolf Optimizer, its versions and applications. *IEEE Access*.

[27] Das, R., Debnath, D., Pakray, P., & Kumar, N. C. (2023). A binary grey wolf optimizer to solve the scientific document summarization problem. *Multimedia Tools and Applications*, 1-23.

[28] M. Mhatre, D. Phondekar, P. Kadam, A. Chawathe and K. Ghag, "Dimensionality reduction for sentiment analysis using pre-processing techniques," 2017 International Conference on Computing Methodologies and Communication (ICCMC), Erode, 2017, doi: 10.1109/ICCMC.2017.8282676.

[29] Seyedali Mirjalili, Seyed Mohammad Mirjalili, Andrew Lewis, "Grey Wolf Optimizer", Advances in Engineering Software, Vol.69, pp, 46-61, 2014.

[30] Ninu Preetha NS, Brammya G, Ramya R, Praveena S, Binu D; Rajakumar B R, “Grey Wolf Optimisation-based Feature Selection and Classification for Facial Emotion Recognition”, IET Biometrics, Vol. 7, No. 5, pages 490-499, 2018, DOI: 10.1049/iet-bmt.2017.0160

[31] Vinolin V and S Vinusha, "Enhancement in Biodiesel Blend with the Aid of Neural Network and SAPSO", Journal of Computational Mechanics, Power System and Control, Vol.1,No.1, pp.11-17,2018.

[32] R. Yousef, S. Khan, G. Gupta, B. M. Albahlal, S. A. Alajlan, and A. Ali, "Bridged-U-Net-ASPP-EVO and deep learning optimization for brain tumor segmentation," Diagnostics, vol. 13, no. 16, p. 2633, 2023.

[33] Akram et al., "On Layout Optimization of Wireless Sensor Network Using Meta-Heuristic Approach," Computer Systems Science Engineering, vol. 46, no. 3, pp. 1-17, 2023.

[34] S. Khan, G. K. Moorthy, V. T, L. H. Alzubaidi, A. Barno, and V. Vijayan, "Computational Intelligence for Solving Complex Optimization Problems," E3S Web Conf., vol. 399, p. 04038, 2023.