Serialized Multi-Layer Multi-Head Feature Location For Sentiment Analysis In Customer Reviews

Anette Regina[1*], P Sengottuvelan[2]

[1*]Research Scholar, Department of computer science , Periyar University , Salem,
Email: anette1967cs@gmail.com
[2]Professor,Department of Computer Science, Periyar University Centre forPG and Research Studies, Dharmapuri, Tamilnadu, India, Email: sengottuvelan@gmail.com
*Corresponding Author

Received: 10.04.2024 Revised: 16.05.2024 Accepted: 24.05.2024

ABSTRACT
Sentiment analysis seeks to identify the sentiment orientation of a text item (sentence or document), however finer-grained sentiment categorization is the best option because many real-world applications demand a deeper level of analysis. The process of determining emotional polarity for component phrases in a sentence is known as Aspect-level Sentiment Classification (ALSC). In the ALSC job, the high dimensionality problem is typically encountered, and methods for selecting features are presented as a means of addressing this issue. This work proposes a feature extraction strategy based on Serialized Multi-layer Multi-Head Attention (SMMHA) and Bidirectional Encoder Representations from Transformers (BERT) for classification. For feature selection, the Chaotic Cuckoo Search Optimization (CCSO) algorithm is presented. The assessment of every feature is the first step of the CCSO algorithm search, which then chooses the feature with the best efficiency. Semantics, sentiment, readability, structure, and grammar are all included. The CCSO method is a heuristic search algorithm that draws inspiration from a synopsis of the cuckoo's parasitic and reproductive habits. The nests with greater accuracy values are chosen as the number of iterations rises and employed in the sentiment prediction for aspect-based analysis. In addition to producing interactive semantic information between the aspect word and the context, the feature extraction approach captures the context's long-term reliance. For testing reasons, the Amazon Customer Review Dataset is obtained straight from Amazon. Sentiment analysis accuracy, recall, F1-score, and precision are used to gauge the results.

Keywords: Sentiment Analysis, Aspect-level Sentiment Classification (ALSC), Bidirectional Encoder Representations from Transformers (BERT) and Amazon dataset.

1. INTRODUCTION

The explosive development of society and commercial activities has been matched by an equally rapid development of the Internet. Rich text and various kinds of media are just two of the abundance of content that can be found on the Internet. People are exposed to a multitude of facts on a daily basis. Review language on live broadcast websites, movie websites, and e-commerce platforms can be used to determine users' interest in their products, services, or movies. Useful information can be gleaned from these sources to tailor the products and services to the preferences of the users, meeting the demands of related businesses in the process. The method of employing text to collect user opinions on products and services in order to ascertain whether or not users like or dislike a specific item is known as sentiment analysis [1]. Sentiment analysis is an important technique for Customer Relationship Management (CRM) in business. Sentiment analysis falls into the following groups: aspect-level, sentence-level, and document-level, depending on how specific the target of the inquiry is. A subtask of fine-grained sentiment analysis [3, 4] is Aspect-Level Sentiment Classification (ALSC) [2], whereas both of the tasks that come before it are associated with coarse-grained sentiment analysis. The coarse-grained sentiment analysis problem, which has been studied in detail and almost resolved, focuses on the overall sentiment polarity of the entire text item. Especially when a sentence has multiple objects, it cannot extract opinions on that specific topic.

A variety of techniques have been put forth by scholars over the past few decades to finish aspect-level sentiment analysis. The supervised machine learning method performs the best out of all of them [5–6]. Nevertheless, these statistically based techniques waste a lot of time and labor since they rely on meticulously created manual features on large-scale datasets [7]. Without the need for artificial feature engineering, the neural network model can automatically learn the lower dimensional representation of reviews. Important data is lost as a consequence of high dimensionality, resulting in it difficult to record long-term links among feature words and context.

Discovering a single sentiment component using Aspect Based Sentiment Analysis (ABSA) is still far from sufficient for comprehending more comprehensive aspect level opinions, which necessitates not only the extraction of multiple sentiment elements but also the identification of the relationships and dependencies among them. Many frameworks have been developed to address various compound ABSA problems in order to enable aspect-level opinion mining in various contexts, after some groundbreaking efforts. In addition to creating customized models for various jobs, pre-trained language models (PLM) like BERT [8] and RoBERTa [9] have significantly improved a variety of ABSA tasks in recent years. The reliability and extension ability of ABSA models have been increased with PLM serving as the framework. The high dimensionality issue still exists in the sentiment classification task, which hinders their performance [10]. In order to address this issue, feature selection techniques are thus presented [11,12].

Feature selection is used to solve these problems by identifying the most relevant and discriminating characteristics while eliminating erroneous and noisy categorization features. This work proposes the Chaotic Cuckoo Search Optimization (CCSO) algorithm, which looks at the link between each characteristic and classes using five features. Serialized Multi-layer Multi-Head Attention with Bidirectional Encoder Representations of Transformers (SMMHA-BERT) has been suggested for recognizing the emotion of different feature words. It measures the significance of traits throughout recognition and training. The Amazon Customer Review Dataset is tested using a number of tests, and the accuracy, precision, recall, and F1-score are used to assess the conclusions.

2. LITERATURE REVIEW

In [13], a statistical technique for choosing characteristics in sentiment analysis utilizing value by Gini Index technique was developed. This approach also improved the accuracy of sentiment polarity prediction using multiple big movie datasets. When the suggested structure for sentiment analysis utilizing the Support Vector Machine (SVM) classifier is used to movie reviews, the accuracy of the classification has increased when employing this effective and cutting-edge technique in comparison to other feature selection techniques. Akhtar et al. [14] offered a cascaded architecture of choosing features and classifier composition employing Particle Swarm Optimization (PSO) for aspect-based analysis of sentiment. Aspect-based sentiment evaluation involves two processes: sentiment categorization and component term extraction. The smaller, trimmed set of features performs superior to the baseline approach, which employs all of the characteristics for aspect phrase extraction and sentiment classification. Construct a second ensemble based on PSO and arrange it in a sequence beneath the feature selection module. Attributes were identified by looking at the traits of different classifiers and domains. As foundation learning approaches, we use three classifiers: Maximum Entropy (ME), Conditional Random Field (CRF), and Support Vector Machine (SVM). Tests on aspect term extraction and sentiment analysis on two different kinds of domains show the effectiveness of our suggested method.

presented the dependency relation in [15] in order to incorporate it into the convolutional neural network and the bidirectional long short-term memory in order to identify the sentiment characteristic associated with relationship for the aspect term in the dependency parse tree. Tests demonstrate that algorithms are able to distinguish the sentiment polarity of an aspect term by using associated sentiment features. The suggested approaches produce cutting-edge outcomes for neural networks. For component level categorization of a deep Memory Network (MemNet) was recently proposed in [16]. This technique, which find out the sentiment polarity of an feature by employing an LSTM and an ordered neural method, exactly gain the value of all context word. Every computational layer utilized for determining the significance level and text description is a neural attention structure over outside storage. Tests conducted on computer and restaurant datasets show that the proposed method operates much better than LSTM and attention-based LSTM designs, and is on par with the state-of-the-art feature-based SVM systems. The two data sets demonstrate how adding more analytical layers can increase efficacy. The 9-layer deep memory network outperforms the LSTM by a factor of 15 when utilized as a CPU method.

The Attentional Encoder Network (AEN) has been suggested in [17], that expresses the relationship among context and goal using attention-based encoders instead of recurrence. Eliminate the problem of inconsistent labeling and provide a uniform label smoothing frequency. Utilize pre-trained BERT as well to complete this assignment and produce unique, creative outcomes. Tests and research show how effective and light the idea is. By creating an element of component adaptation technique, [18] presented Bidirectional Encoder Representations from Transformers (ALM-BERT), an efficient aspect-level evaluation of sentiment. First, it is shown that a pretrained BERT algorithm could potentially mine additional aspect-level auxiliary data from the setting of the remarks. Additionally, develop an aspect-based sentiment feature extraction technique to understand both the specific characteristics of every component phrase and the dependent on context interaction data. The Target-Dependent versions of the BERTbase concept have been used in [19], with the target terms as the output positions and an optional sentence that incorporates the target. In compared to previous BERT applications, embedding-based models, and classic feature engineering techniques, studies conducted on three data collections demonstrate that the TD-BERT model delivers new state-of-the-art performance. Considering its successful application in many NLP tasks, investigations aim to verify if BERT's context-aware model may achieve a similar degree of effectiveness increase in aspect based sentiment analysis.

A Sparse Attention based Separable Dilated Convolutional Neural Network (SA-SDCCN) was suggested in [20]. It is composed of an output layer, a multichannel embedding layer, a distinct dilated convolution module, and a sparse attention layer. Particularly, the first three sections receive the most of the work. Semantic and sentiment embeddings are added to an embedding tensor in the multichannel embedding layer to create richer representations over the input sequence. Diverse dilation rates are used in the separable dilated convolution module in order to accumulate multi-scale contextual semantic dependencies and examine long-range contextual semantic data. Additionally, the framework's parameters are further reduced by the separable structure. Sentiment-oriented components are identified in the sparse attention layer based on the characteristics of a particular target entity. Ultimately, a few tests conducted on three standard datasets show that SA-SDCCN outperforms state-of-the-art techniques with regard to of higher parallelism and lower computational cost.

A co-extraction model with enhanced word embeddings was given in [21] in order to take advantage of the dependence structures without the need for syntactic parsers. A multilayer dual-attention model based on deep learning is suggested to take advantage of the oblique relationship between aspect and opinion words. Furthermore, in contrast to the Word2Vec approach, the word embeddings are improved by offering unique vector representations to disparate moods. The issue of comparable vector representations of opposing sentiments is addressed by introducing a sentiment refining technique for pre-trained word embedding models. Three benchmark datasets from the SemEval Challenge 2014 and 2015 are used to assess the performance of the proposed approach. The results of the experiment show how successful the model is for aspect-based sentiment analysis when compared to the most advanced models currently in use.In [22], an entirely novel structure for aspect-level sentiment classification was created. To more effectively direct the multi-hop attention procedure, the Deep Selective Memory Network (DSMN) dynamically chooses the context memory by integrating inter-aspect information with the deep memory network.

To gather copious aspect-aware context information, DSMN concentrates on various parts of the context memory across various memory network layers by constructing a selective attention mechanism based on the distance information between an aspect and its context. It not only makes the most use of the inter-aspect data but also develops effective inter-aspect model components that generate structural and semantic information about the surrounding aspects for the target aspect.Examine the benefits of the DSMN framework using three benchmark datasets. The findings of the study demonstrate that the DSMN architecture performs at the cutting edge.The PConvBERT (Prompt-ConvBERT) and PConvRoBERTa (Prompt-ConvRoBERTa) models were introduced in [23]. Both of these models integrate local context features obtained by a Local Semantic Feature Extractor (LSFE) with the global features of BERT/RoBERTa (Robustly Optimized BERT Pretraining Approach). A common solution to the durability issue with models developed using deep learning is the use of adversarial training to boost the stability of the model. Furthermore, Focal Loss is used to lessen the effects of an uneven sample distribution. In order to thoroughly investigate the pre-training model's capabilities, natural language processing is presented to stimulate methods that more effectively address the ALSC challenge. For sentiment classification, masked vector outputs of templates are utilized. Numerous tests using open datasets show that the suggested model works well.

3. PROPOSED METHODOLOGY

Component-based sentiment analysis is a method used to make the sentiment polarity of all view word in a phrase when the sentence and some of the chosen factor words are used as the information being analyzed. Supply a formal statement sentence $S = \{w_1, w_2, \cdots, w_n\}$, where n is the whole word count of S. $A = \{a_1, \cdots, a_i, \cdots, a_m\}$ with size m denotes for an feature vocabulary of length m, where a_i represents the i^{th} aspect word in aspect vocabulary A, and A is a subsequence of sentence S. $P = \{p_1, \cdots, p_j, \cdots, p_C\}$ correspond the candidate sentiment polarities, where C correspond the overall quantity of sentiment polarity and the $p_j$is the j^{th} sentiment polarity. The almost likely sentiment polarity of a granted feature word in a phrase is expected by the aspect-based sentiment analysis method, which may be stated in the following manner:

$$\text{Input:}\begin{cases} S = \{w_1, \dots w_n\}, \\ A = \{a_1, \dots . a_m\}, \end{cases}$$
Output: $p_k = \phi_{max}(a_i, p_j | S)$,
Constraints: $A \in S, m \in [1, N]$
where ϕ is a function that measures how well the aspect word and the sentiment polarity p_j in the sentence S correspond. A stands for the aspect vocabulary. The model outputs the sentiment polarity with the highest degree of matching as a category outcome in the final stage.
Aim and context serve as the foundation for the investigations of aspect-level sentiment polarity assessment. For feature extraction, a method known as Serialized Multi-layer Multi-Head Attention with Bidirectional Encoder Representations from Transformers (SMMHA-BERT) is given. The approach known as Chaotic Cuckoo Search Optimization (CCSO) is presented for content-based feature candidates. Finally, the available feature selection and classification approaches are used to determine the optimal feature pairs for review sentiment analysis. The customer review dataset of Amazon is utilized, and the essential objective is to take out feature expression from every review and utilize classification algorithms to acquire the score for each review.

3.1 Data Collection

The collection of Amazon customer reviews is gathered from the Amazon website. It is comprised of more than 34,000 customer evaluations from Datafiniti's Product Database for Amazon items, including the Kindle, Fire TV Stick, and more. For each creation in the aggregation, there is necessary data about it, a rating, text reviews, and many. It has been widely applied to aspect-based sentiment analysis tasks. Eventually, the dataset was divided into two groups, training and testing, with 17,000 instances each.

3.2 Data Pre-Processing

The preparation of the data for assessment is the first step in the sentiment analysis process. All of the text in the dataset is unprocessed and labeled as either positive or negative. The following are the stages that go into getting the data ready.

Tokenization and Segmentation

For this phase, the Python NLTK package is utilized. In this stage, the only sentence is divided into a few smaller strings known as "Tokens."

Noise Removal

Python does not have a dedicated library for precisely eliminating stop words. Files with a small number of prohibit words that presumably fit the token or are attempted to be removed are regarded as noise.

Lemmatization and Normalization

For carrying out this stemming procedure, "Skip-Gram with Negative Sampling" is presented. There are two variants of the method: CBOW and Skip-Gram. This stage finds words that are comparable in the data or input test. When provided a collection of sentences, the framework iterates through every word and determines whether it can predict the current word by using its neighbors (a technique known as "Skip-Gram") or by using each of these contexts to predict the current word (a technique known as "Continuous Bag of Words" (CBOW).

3.3 Feature Selection

This study integrates the Chaotic Cuckoo Search Optimization (CCSO) algorithm with semantic, sentiment, readability, structure, and syntactic aspects. These attributes have demonstrated resilience in many text mining and aspect-level sentiment analyses.

Semantics features

By modeling term statistics into vectors, semantic features relate to the meaning of words and subject topics from the analyzed material. The following are the five semantic features for the helpfulness prediction task:
UGR and BGR: The term frequency-inverse document frequency (TF-IDF) weighting technique is used in the Unigram Bag-of-Words depiction of an examine, where every vector's constituent relates to a word in the lexicon. The TF-IDF technique is used to evaluate a word's importance in an essay mathematically [24, 25]. Depending on the word's frequency, the score frequency assigned by TF-IDF establishes the word's significance for the document or documents. The following describes the formulas (1-3) that were utilized to determine the TF-IDF value sequentially:

$$\mathrm{tf(w,d) = log(1 + fw,d)} \quad (1)$$
$$\mathrm{idf(w,D) = log\left(\frac{N}{f(w,D)}\right)} \quad (2)$$
$$\mathrm{tf - idf(w,d,D) = tfw,d * idf(w,D)} \quad (3)$$

The equation uses the following symbols: N is the number of documents; d is the document that is given; D is the overall amount of documents utilized; and w is a word in document d. The Term Frequency (TF), where f(w, d) represents the amount of instances word w happens in document d, is determined using the first expression (1). The weighted score of terms that appear less often is increased and the weighted score of terms appearing frequently is decreased by calculating the Inverse Document Frequency (IDF) using formula (2). The log of N documents divided by f(w, d), or the frequency at which the word w appears in document d, yields the IDF formula. To determine the TF-IDF score, the final equation consists of multiplying the TF and IDF results. In the same way, every potential word pair that can be created from adjacent words in a corpus is encoded using the bigram bag-of-words format. The L2 normalization is then used to convert the vector representations into unit vectors. The square root of the sum of the squared vector values is used to determine L2 normalization. L_2 norm is commonly used than other vector norms in classification.

Latent Dirichlet Allocation (LDA): The subject distribution of an overview is learned by a Latent Dirichlet Allocation (LDA) representation. In topic modeling, the corpus is viewed as a collection of subjects, each of which is made up of a collection of words. Reviews of products online may cover a variety of subjects, such as the features of the product, the author's feelings, etc. According to LDA, words are created by fixed conditional distributions over subjects, and topics can be interchanged indefinitely inside a document. The likelihood of a word and topic sequence must thus have the form, according to de Finetti's theorem.

$$\mathrm{p(w,z) = \int p(\theta)\left(\prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n)\, d\theta\right)} \quad (4)$$
$$\mathrm{p(z_1, \ldots, z_N) = p\left(\left(z_{\pi(1),\ldots,\pi(N)}\right)\right)} \quad (5)$$

where θ is the random parameter of a multinomial over topics, π (·) is a permutation function on the integers {1…N}. Conditional probability distribution $\mathrm{p(x_1, \ldots, x_k|\Theta) = \prod_{i=1}^{K} p(x_i|\Theta)}$is simple to articulate, however it is typically impossible to break down the joint distribution.

Skip-Gram with Negative Sampling (SGNS) and Global Vector (GV): The investigation also used the two current categories of word embeddings as characteristics, which is innovative [26]. The goal of global vectors and SGNS is to acquire how words look that are distributed. Each word is mapped into a dense vector space in this configuration, so phrases that are comparable show a closer spatial distance. Therefore, by averaging the embeddings for every review's constituent words—skipping out-of-vocabulary words—each evaluation may be easily transformed into a vector.

Given a word series, $\mathrm{w_1, w_2, \ldots, w_n}$, for training, The skip-gram method desire to lessen its consequent outcome with word embedding learning,

$$\mathcal{L}_{\mathrm{SG}} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{|j|\le c, j\neq 0} \log p\left(w_{i+j}|w_i\right) \quad (6)$$

Considering $\mathrm{w_i}$is a suitable word and $\mathrm{w_{i+j}}$is a circumstance word inside a window of size c. $\mathrm{p(w_{i+j}|w_i)}$described the probability that $\mathrm{w_{i+j}}$described within adjacent of $\mathrm{w_i}$, and is shows up as follows,

$$\mathrm{p\left(w_{i+j}|w_i\right) = \frac{exp\left(t_{w_i}.c_{w_{i+j}}\right)}{\sum_{w\in\mathcal{W}} exp(t_{w_i}.c_w)}} \quad (7)$$

where t_w and c_w are w's embeddings when it behaves as a target and context, respectively. $\mathcal{W}$remain for the vocabulary set. The equation (6) used sigmoid functions and k chosen at random, mention to as negative samples. The following is the result of the objective.

$$\mathcal{L}_{\mathrm{SGNS}} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{|j|\le c, j\neq 0} \psi^{+}_{w_i,w_{i+j}} + k\mathbb{E}_{v\sim q(v)}[\psi^{-}_{w_i,v}] \quad (8)$$

where $\psi^{+}_{w,v} = \log\sigma(t_w.c_v)$, $\psi^{-}_{w,v} = \log\sigma(-t_w.c_v)$,, and σ(x) is the sigmoid function. After that, the gradient descent method is used to update t_{w_i} , $c_{w_{i+j}}$, and $c_{v_1}, \ldots, c_{v_k}$. SGNS training must precompute the the noise distribution q(v) before executing SGD, it must repeatedly scan the whole training set.

Sentiment features
Sentiment characteristics evaluate the degree of subjectivity, valence, and emotion in user-generated material. Because of the varied testing conditions, the emotion features will therefore result in distinct vector renderings.

Multiangle Text Vectorization Mechanism
Word embedding maps every word to a high-dimensional vector space and is mostly used to aid robots in understanding natural language. One language pretraining model that works well with unlabeled text is called BERT.

Aspect-Based Sentiment Feature Extraction Method
The implicit features of the aspect words and their context are extracted using an aspect-based sentiment detection method that draws inspiration from a transformer encoder. This allows for the consideration of auxiliary data present in the component words. The self-focus part of the proposed serialized attention technique serializes information from periodic context and deeper layers. The summing of the feature-level vectors from each head is then used to generate the final aspect-level embedding. Following batch normalization and ReLU activation, the input is fed through the layers of the classifier.

Aspect Feature Location Model
In addition to producing interactive semantic information between the aspect word and the context, the feature extraction approach preserves the context's long-term reliance. Using this information, create a component feature placement model based on the maximum pooling function to emphasize the significance of certain feature words even more.

Readability features
The ease of reading texts is measured by intelligibility. The benefit of the amount of characters, syllables, words, complicated words, and sentences is taken into account while extracting readability qualities.

Structure features
The length and prevalence of particular linguistic types of units are counted using structural characteristics. The structure of a review is represented by the six characteristics listed below. Character count (CHAR), token count (WORD), sentence count (SENT), and the proportion of exclamatory (EXCLAM) and interrogative (INTERRO) phrases are among the characteristics that are considered. Lastly, an evaluation's misspelled word count (MIS) is taken into account.

Syntax features
Throughout the review material, syntactic characteristics take into account particular parts-of-speech kinds and patterns. For aspect-based sentiment analysis, the proportion of those with the most common open-class word categories—nouns (NOUN), adjectives (ADJ), verbs (VERB), percentage of comparison sentences (COMP), and adverbs (ADV)—is computed. A set of keywords and patterns is used in the comparative sentence identification process to match the review sentences.

3.4 Feature selection by Chaotic Cuckoo Search Optimization (CCSO)
Wrapper approach fulfills Chaotic Cuckoo Search Optimization (CCSO) search. The search begins with evaluating every feature in the feature pool and chooses the feature with the best performance. Then, after evaluating every conceivable pairing between the chosen feature and every other feature, the second feature is chosen. Iterations proceed till the results of the predictions cannot be improved by additional features. Consequently, the best feature combination is formed by integrating each of the selected features. The heuristic search technique that motivated it relies on an overview of the cuckoo's parasitic and reproductive habits. The excellent characteristics are kept in the sentiment prediction for aspect-based analysis because as the number of algorithm iterations rises, the nests with better fitness indices are kept, meaning that the good eggs are kept later in the iteration. In order to improve the classifier's accuracy, the CCSO approach's feature selection approach seeks to choose fewer characteristics overall.
The formal definition of feature selection for sentiment analysis is as below. Let $D = \{(D_1, s_1), \dots, (D_n, s_n)\}$ be a collection of n product reviews, where D is the content of a review and s the sentiment (s = 1 positive and s = 0 negative). Every review content $D \in \mathfrak{D}$ is related with a set of features, denoted by $F(D) = \{f_1(D), \dots, f_m(D)\}$ via m various feature extractors {f}. The purpose is to train a binary classifier C that

looks through the characteristic pool F for the best feature combination $\hat{F}$ to estimate usefulness u in a way that

$$\hat{F} = \arg \max_{F' \subseteq F(D)} \sum_{D \in \mathfrak{D}} l\left(s = C(F')\right) \qquad (9)$$

where $l(.)$is an indicator function in this case. The goal of $\hat{F}$'s search should be to find every conceivable combination of features. Based on an overview of the parasitic and reproductive behaviors of cuckoos in the wild, the method creates a heuristic search algorithm with a biological inspiration. The cuckoo egg is defined as a completely newly selected feature solution, and the new feature solution is used to replace the less-than-ideal the solution in the nest when solving the feature selection problem. This allows for the best feature selection from the extracted features by describing each egg in the nest as a solution. The nests with high fitness (accuracy) values are kept as the number of iterations rises. This implies that the excellent features are kept since the good eggs are kept further in the algorithm's iteration.

CS is proposed for choosing features from movie review dataset in conjunction with Levy flying, by idealizing and simplifying the brood behavior of particular cuckoo species. In overall, it follows these three speculative guidelines [27, 28],

1. Using a randomly selected nest, every cuckoo deposits one egg at a time to ensure the best possible selection of attributes from the movie review dataset.,
2. The fitness function can be used to assess the best features (nests), which are impervious to corruption for examine analysis.,
3. The host bird has a given probability $p_a \in [0,1]$of finding the egg, and the number of host nests is fixed.

For cuckoo i, new cuckoo $x^{(t+1)}$ (feature) is generated, a Levy flight is implemented by equation (10),

$$x_{ik}^{(t+1)} = x_i^{(t)} + \alpha \oplus \text{Levy}(\lambda) \qquad (10)$$

when the step size ($\alpha > 0$) is used. Most of the time, α is just set to 1. The symbol $\otimes$ denotes multiplications by entry. Chaotic refers to a type of random-based approach in the CSO. The movie review dataset's extracted characteristics were utilized to initialize the population. Equation (10) indicates that the two primary CS parameters are the step size (α) and discovery rate (pa), which describe the fluctuations of the global optimal step size. The values of these parameters significantly impact the convergence speed and the overall performance of CS. Its convergence is accelerated by the enhanced CS technique with a chaotic variable step size α. All chaotic maps are normalized, therefore their variants are always in [0, 2]. Chaotic Cuckoo Search Optimization (CCSO) is the term used to describe the chaotic map's ability to optimize step size α after normalization [28].As seen in algorithm 1, the full explanation of the CCSO algorithm's pseudo-code has been included..

Algorithm 1. Chaotic Cuckoo Search Optimization (CCSO)

Begin
Step 1: Initialization.
Define the generation counter to 1;
Begin the population P of n host nests
arbitrarily;
Begin value of the chaotic map c_0 randomly,
Define thefinding rate
p_a, and elitism KEEP.
Step 2: While t < MaxGeneration do
Population is sorted by level of fitness.
Maintain the most excellent cuckoos in storage.
Chaotic maps ($\alpha = c_{t+1}$) are used to modify
the size of the steps, and Levy flights are
used to substitute the final solution.
Analyze its fitness F_i.
Select a nest between (say, j) arbitrarily
If ($F_i < F_j$)
Substitute the latest result for j.
End if
A fraction (p_a) of the worse nests is discarded
and new ones are constructed.
Replace worst cuckoo with the KEEP excellent cuckoo.
Evaluate the population and search the present better.

t = t+1.
Step 3: End while
End.

3.5 Serialized Multi-Layer Multi-Head Attention (Smmha)

A collection of N similar layers makes up the serialized attention method. Every layer is made up of two modules stacked together: a feed forward module and a self-attention module. There is a residual contact for each among these components. As in [29, 30], layer normalization is carried out on the input prior to the feed-forward and self-attention modules, separately. To collect and convey the information in a serialized method from one layer to the next, stacking self-attention modules is proposed as an alternative to multi-head attention. In the original multi-head attention, each input chain is separated into multiple similar sub-vectors, called heads. Conversely, deeper aggregation network architecture can increase expressive power by learning and aggregating more discriminative information at numerous levels.

The respective attention module in the recommended serialized focus system allows in a serialized fashion, sanctioning the model to gather temporally contextualized input from deeper layers.Typically, from the n^{th}self-attentioncomponent $\left(n^{th}[1, \ldots., N]\right)$, the weighted mean $\tilde{\mu}$ and weighted standard deviation$\tilde{\sigma}$are get. It is then translated into a featurelevel vectoralso known as a serialized head from layer n—after undergoing an affine transformation. The summation of the feature-level vectors from each head is then used to get the final aspect-level embedding. Classifier layers are then supplied with batch normalization and ReLU activation.

Input-aware self-attention: The focus function maps a query and a set of key-value pairs to an output. Instead of using a single query for all features, use an input-aware unique query for each trait. The query is generated using statistics pooling since the mean and standard deviation can include all of the data. Let us consider an input sequence $[h_1, \ldots, h_{Tc}]$ with $h_{tc} \in \mathbb{R}^d$, where T is the length of the input sequence. In order to create the query q, the system converts the series of inputs in the following manner

$$q = W_q g(h_{tc}) \qquad (11)$$

where is a trainable parameter and is statistics pooling. Regarding key-value pairings, no additional calculation is necessary because the input sequence [h1,..., hTc] is immediately allocated to the value sequence [v1,v2,..., vTc] of d dimensions, hence reducing the total amount of model parameters. By using a linear projection$W_k \in \mathbb{R}^{d_k \times d}$. and an adaptable dimension, the important vector, or ktc, is produced.

$$k_{tc} = W_k h_{tc} \qquad (12)$$

With (v_{tc}, k_{tc}, q) as $\{\text{value}, \text{key}, \text{query}\}$ tuple, the weights are computed via scaled dot-product attention. Every self-attention layer's output is fed into a feed-forward module after the self-attention component, whose job it is to analyze the result and make it more suited as input for the subsequent self-attention layer. ReLU activation occurs during the two linear changes that make up this process.

$$FFW(h) = W_2 f(W_1 h + b_1) + b_2 \qquad (13)$$

where $f(.)$ is a ReLU function,h is the input and the linear transformations vary depending on the layer to layer. $W_1 \in \mathbb{R}^{d_{ff} \times d}, W_2 \in \mathbb{R}^{d \times d_{ff}}$ having inner dimension d_{ff}, or two convolutions with kernel size 1 in another way. Two layers are fed the serialized attention technique feature-level integrating: a regular softmax layer and one fully linked layer. All of the nodes in the softmax layer match the class labels in the training set.

Sentiment Predictor: Multiple self-attention mechanisms are applied to SMMHA-BERT in order to get multi-angle text concealed conveying characteristics. This is followed by aspect feature positioning and contextual interaction of aspect word handling. The comprehensive representation r in is obtained by usingh_{cm}, h_{am}, and h_{af} as shown below.

$$r = [h_{cm}; h_{am}; h_{af}] \qquad (14)$$

The data of r is then preprocessed using a linear function, as illustrated in the following [31],

$$x = W_u r + b_u \qquad (15)$$

where bu stands for bias and Wu for the weight matrix. finally, the probability Pr that the feature word an in a expression has sentiment polarity p is measured utilize the soft max purpose, as seen in the upcoming section.

$$Pr(a = p) = \frac{\exp(x_p)}{\sum_{i=1}^{C} \exp(x_i)} \qquad (16)$$

Where $W_u$stands for the weight matrix, and $b_u$stands for bias. Lastly, the probability Pr that the aspect word an in a phrase has sentiment polarity P is calculated using the soft max function, as seen in the next section.here C stands for the total number of polarity divisions in emotions.

The SMMHA-BERT method is an end-to-end computation procedure in general [32].Reduce the difference among the expected sentiment polarity $\hat{y}$ and the actual sentiment polarity (y). The proposed approach is trained using cross-entropy with L_2 regularization as the loss function. It is described in the following manner

$$\text{Loss} = -\sum_{j}\sum_{i} y_i^j \log \hat{y}_i^j + \lambda||\theta||^2 \qquad (17)$$

where D stands overall training data, and j and I represent the fact of a training data illustration and a sentiment class. λ defines the cause for L_2 regularization, and θdenotes the parameter set of the method .

4. EXPERIMENTAL EVALUATION

In order to assess the efficacy and reliability of the SMMHA-BERT strategy, comparison studies are designed and the study setups are thoroughly described in this part. Furthermore, examine the outcomes of the investigation.

A.Dataset

Use the dataset of Amazon customer reviews while you explore. It was sourced from Datafiniti's Product Database, which has over 34,000 customer reviews of goods offered by Amazon, including the Kindle and Fire TV Stick.There is a text review, a rating, and basic details on every single item in the entire set. Tasks involving sentiment analysis based on aspects have made extensive use of it.

B. Baselines and Evaluation Metrics

Compared the suggested method with additional well-known aspect-based sentiment analysis designs, such as the ones mentioned below, to confirm its efficacy. In [13] is a data-driven technique that utilize numerous attention-based computational layers to calculation the significance of every assumption word.Either the primary process and the BERT-based system AEN-BERT perform well on aspect-based sentiment analysis tasks [14].A feature element placement strategy is created by the ALM-BERT [15] technique, that is based on the efficient component-level sentiment assessment technique.Furthermore, measurements like accuracy, precision, recall, macro-F1 score (F1), and recall as indicators of evaluation are needed in order to impartially evaluate the effectiveness of the suggested strategy in component level sentiment evaluation assignments that are comparable to the ones that are already in use.

Macro-F1, the weighted average of recall and accuracy, is used to correspond the predictive execution of the model in an suitable method. Equation serves as the basis for macro-F1's implementation(20).

$$\text{Pre}_{C_i} = \frac{T_{c_i}}{T_{c_i} + FP_{C_i}} \qquad (18)$$

$$\text{Re}_{C_i} = \frac{T_{c_i}}{T_{c_i} + FN_{C_i}} \qquad (19)$$

$$\text{macro} - F1 = \frac{1}{C}\left(\sum_{i=1}^{C}\left\{\frac{\left(2 * \text{Pre}_{C_i} * \text{Re}_{C_i}\right)}{\left(\text{Pre}_{C_i} + \text{Re}_{C_i}\right)}\right\}\right) \qquad (20)$$

when FN denotes how many items have sentiment polarity T is the overall amount of samples that have been correctly recognized as being part of sentiment polarity I, FP is the amount of samples which were incorrectly recognized as having to sentiment polarity I, and I is the sentiment polarity that was incorrectly recognized as being distinct. C represents the quantity of emotion polarity categories, Pre_{C_i} indicates the sentiment polarity precision, and Re_{C_i} denotes the sentiment polarity recall. The calculation of accuracy (Acc) is based on equation (21).

$$\text{Acc=SC/N} \qquad (21)$$

MemNet,ALM, AEN, and SMMHA are only a few of the classifiers to which performance is compared via feature set in Table 1.

Table 1. Metrics Comparison Of Classification Methods Vs. Feature Set

Metrics (%)	Classifier	Semantics	Sentiment	Readability	Structure	Syntax	Average
Precision	**MemNet**	89.50	92.81	84.52	81.51	88.10	87.28
	AEN	91.51	93.82	88.92	90.77	92.58	91.52
	ALM	92.18	94.62	89.57	91.75	93.45	92.31
	SMMHA	94.88	95.87	92.79	93.59	95.43	94.51
Recall	**MemNet**	89.14	90.15	87.65	87.67	91.09	89.14

	AEN	90.48	91.75	89.22	88.87	92.57	90.58
	ALM	92.17	93.52	91.84	90.16	93.88	92.31
	SMMHA	93.90	95.71	93.54	97.19	95.52	95.17
F1-score	**MemNet**	89.32	91.48	86.085	84.59	89.59	88.21
	AEN	90.99	92.78	89.07	89.82	92.57	91.05
	ALM	92.17	94.07	90.70	90.95	93.66	92.31
	SMMHA	94.39	95.79	93.16	95.39	95.47	94.84
Accuracy	**MemNet**	88.95	91.71	90.36	90.33	92.90	90.85
	AEN	90.55	93.45	92.70	91.48	94.72	92.58
	ALM	91.54	94.82	93.27	92.98	95.94	93.71
	SMMHA	92.21	95.82	94.56	93.85	96.92	94.67

Table 2 shows the performance comparison of classifiers such as MemNet, AEN, ALM, and SMMHA together BERT and feature set.

Table 2. Metrics Comparison Of Bert+ Classification Methods Vs. Feature Set

Metrics (%)	**Classifier**	**Semantics**	**Sentiment**	**Readability**	**Structure**	**Syntax**	**Average**
Precision	**MemNet**	90.75	89.42	91.56	87.69	86.63	89.21
	AEN	91.62	93.24	90.84	92.97	91.95	92.14
	ALM	93.78	94.92	93.73	93.98	94.45	94.17
	SMMHA	95.63	95.87	94.97	95.44	96.80	95.74
Recall	**MemNet**	85.87	90.71	91.22	86.65	89.63	88.82
	AEN	89.15	92.74	92.85	90.46	93.22	92.71
	ALM	92.44	94.33	93.65	93.89	94.75	93.81
	SMMHA	93.24	95.46	94.48	94.27	95.62	94.61
F1-score	**MemNet**	88.31	90.06	91.39	87.17	88.13	89.01
	AEN	90.38	92.99	91.84	91.71	92.58	92.42
	ALM	93.11	94.62	93.69	93.93	94.60	93.99
	SMMHA	94.43	95.66	94.72	94.85	96.21	95.17
Accuracy	**MemNet**	89.25	92.21	91.47	92.66	93.32	91.78
	AEN	91.74	93.46	92.79	93.32	94.45	93.15
	ALM	93.79	94.61	95.46	94.8	95.78	94.88
	SMMHA	95.11	95.84	96.71	96.85	97.24	96.35

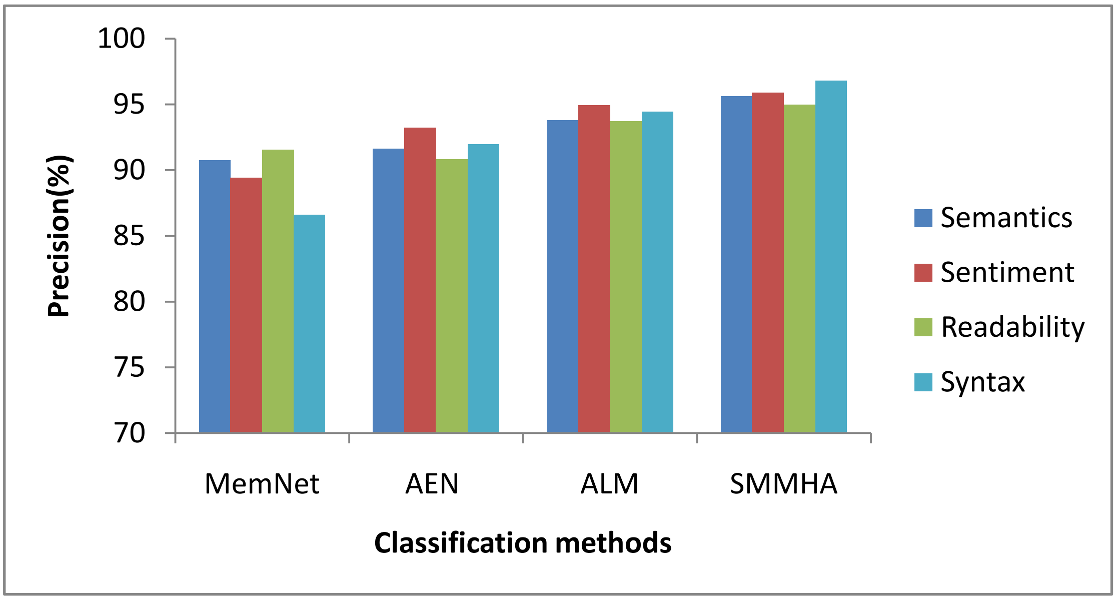

Figure 1. Precision Comparison Of Bert+ Classification Methods With Feature Set

Figure 1 shows the precision comparison of BERT with classification methods among feature sets. From the results it concludes that the proposed system gives higher precision results of 96.80%, the other classifiers such as MemNet, AEN, and ALM give lesser precision value of 86.63%, 91.95%, and 94.45% for syntax features. The SMMHA model will give higher precision results of 95.63%, 95.87%, 94.97%, and 95.44% for semantics, sentiment, readability, and structure.

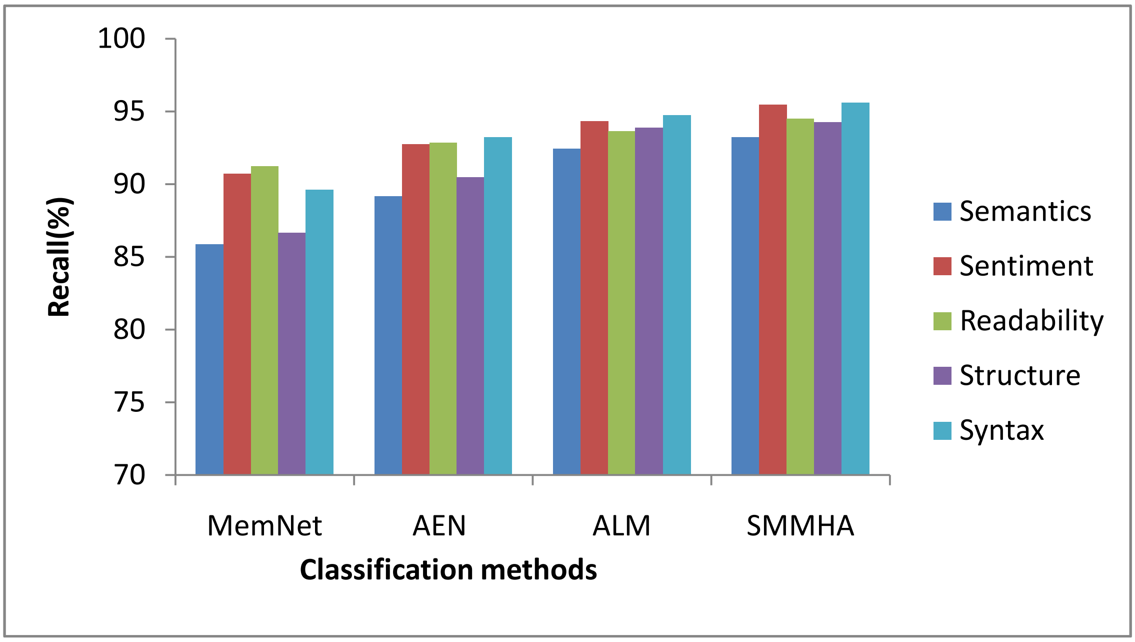

Figure 2. Recall Comparison Of Bert+ Classification Methods With Feature Set

Recall results comparison of BERT with classification methods among feature sets are illustrated in figure 2. Proposed system gives higher recall results of 95.62%, the other classifiers such as MemNet, AEN, and ALM give lesser recall value of 89.63%, 93.22%, and 94.75% for syntax features.

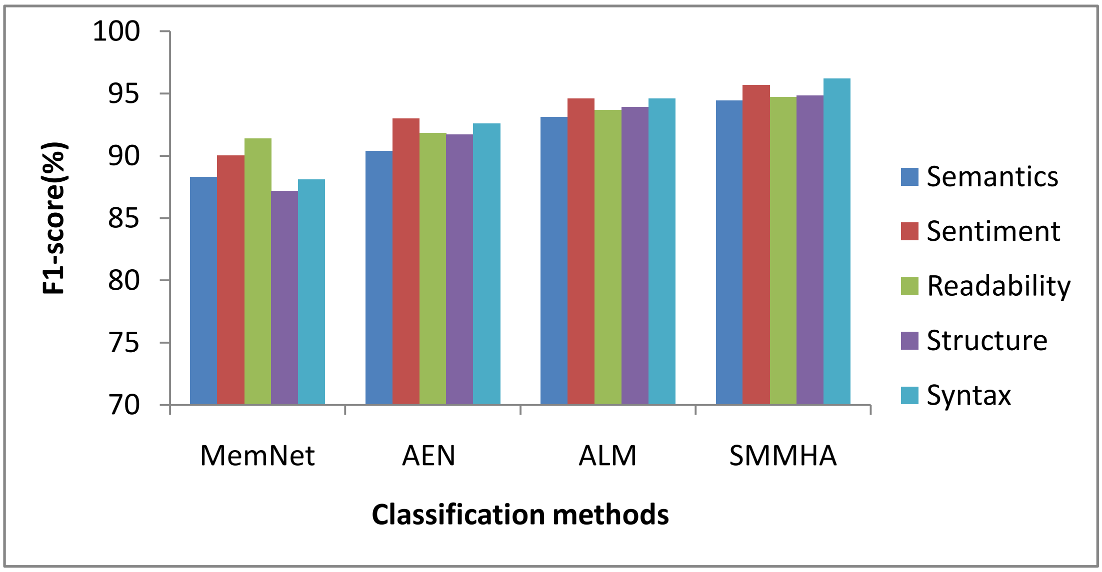

Figure 3. F1-Score Comparison Of Bert Classification Methods With Feature Set

The SMMHA model will give higher recall results of 93.24%, 95.46%, 94.48%, and 94.27% for semantics, sentiment, readability, and structure. BERT with classification methods among feature sets with respect to F1-score are illustrated in figure .3. Proposed system gives a higher F1-score of 95.62%, the other classifiers such as MemNet, AEN, and ALM give a lesser F1-score value of 89.63%, 93.22%, and 94.75% for syntax features. The SMMHA model will give higher F1-score results of 93.24%, 95.46%, 94.48%, and 94.27% for semantics, sentiment, readability, and structure.

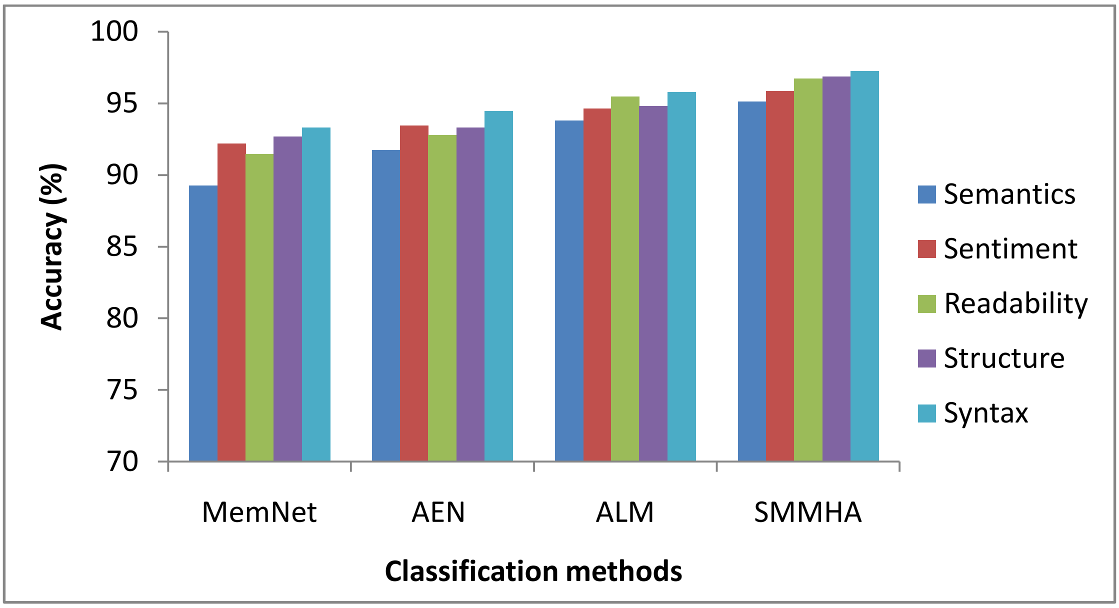

Figure 4. Accuracy Comparison Of Bert+ Classification Methods With Feature Set

Accuracy comparison of BERT with classification methods among feature sets are illustrated in figure 4. Present method provide better outcomes of accuracy is 97.24%, the another classifiers such as MemNet, ALM and AEN provide small values of precision 93.32%, 96.78% and 94.45%, for syntax features. The SMMHA model will give higher results of 95.11%, 95.84%, 96.71%, and 96.85% for semantics, sentiment, readability, and structure.

5.CONCLUSION AND FUTURE WORK

In this research, we offer a deep learning based technique to determine the sentiment polarity of opinion words on a given sentence component. For feature extraction, Serialized Multi-layer Multi-Head Attention-Bidirectional Encoder Representations from Transformers (SMMHA-BERT) is presented. Chaotic Cuckoo Search Optimization (CCSO) algorithm is introduced for feature selection. Feature construction process is created and feature candidates are identified by CCSO. CCSO algorithm aims to reduce the number of features to increase the accuracy of the classifier. CCSO algorithm highlights the feature combination patterns which lead to higher performance. CCSO algorithm incorporates the semantics features, sentiment, readability, structure and syntax features. This technique with a chaotic varied measure size α accelerates its occurrence. SMMHA method, serialized attention technique is carried out in a serialized manner, allowing the method to aggregate data with temporal context from deeper layers. The suggested approach provides supervisors with discriminative features by moving dependent and contextual information from opinion words to aspect words. Classification techniques yield better outcomes in terms of accuracy, F1-score, precision, and recall. According to research findings, the suggested method is now considered the state-of-the-art for aspect-based sentiment classification. Future research is aimed at examining the opportunities of developing a much better and efficient scheme for the generation of new membership functions thereby improving the classifier performance.

REFERENCES

[1] Musa, I.H., Zamit, I., Xu, K., Boutouhami, K. and Qi, G., 2023. A comprehensive bibliometric analysis on opinion mining and sentiment analysis global research output. Journal of Information Science, 49(6), pp.1506-1516.
[2] Zhang D, Zhu Z, Kang S, Zhang G, Liu P. Syntactic and semantic analysis network for aspect-level sentiment classification. Appl Intell. 2021;51:6136–47.
[3] Zhang, W., Li, X., Deng, Y., Bing, L. and Lam, W., 2022. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. IEEE Transactions on Knowledge and Data Engineering, 35(11), pp. 11019 - 11038.
[4] Nazir, A., Rao, Y., Wu, L. and Sun, L., 2020. Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. IEEE Transactions on Affective Computing, 13(2), pp.845-863.
[5] Cai Z. and Z. Xu, "A private and efficient mechanism for data uploading in smart cyberphysical systems," IEEE Transactions on Network Science and Engineering, vol. 7, no. 2, pp. 766–775, 2018.

[6] Nazir, A., Rao, Y., Wu, L. and Sun, L., 2020. Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. IEEE Transactions on Affective Computing, 13(2), pp.845-863.
[7] Lin Y., Z. Cai, X. Wang, F. Hao, L. Wang, and A. M. V. V. Sai, "Multi-round incentive mechanism for cold startenabled mobile crowdsensing," IEEE Transactions on Vehicular Technology, vol. 70, no. 1, pp. 993–1007, 2021.
[8] Devlin J., M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in NAACL-HLT, 2019, pp. 4171–4186.
[9] Liu Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "ROBERTA: A robustly optimized BERT pretraining approach," CoRR, vol. abs/1907.11692, 2019.
[10] Wu, Z., Li, Y., Liao, J., Li, D., Li, X. and Wang, S., 2020. Aspect-context interactive attention representation for aspect-level sentiment classification. IEEE Access, 8, pp.29238-29248.
[11] Uysal, A.K. and Murphey, Y.L., 2017, Sentiment classification: Feature selection based approaches versus deep learning. In 2017 IEEE International Conference on Computer and Information Technology (CIT), pp. 23-30.
[12] Zhang, D., Zhu, Z., Kang, S., Zhang, G. and Liu, P., 2021. Syntactic and semantic analysis network for aspect-level sentiment classification. Applied Intelligence, 51(8), pp.6136-6147.
[13] [13 Manek, A.S., Shenoy, P.D. and Mohan, M.C., 2017. Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World wide web, 20, pp.135-154.
[14] Akhtar, M.S., Gupta, D., Ekbal, A. and Bhattacharyya, P., 2017. Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. Knowledge-Based Systems, 125, pp.116-135.
[15] [15]Zhang, X., Xu, J., Cai, Y., Tan, X. and Zhu, C., 2021. Detecting dependency-related sentiment features for aspect-level sentiment classification. IEEE Transactions on Affective Computing, 14(1), pp.196-210.
[16] Tang, D., Qin, B. and Liu, T., 2016. Aspect level sentiment classification with deep memory network. arXiv preprint arXiv:1605.08900, pp.1-11.
[17] Song, Y., Wang, J., Jiang, T., Liu, Z. and Rao, Y., 2019, Targeted sentiment classification with attentional encoder network. In International Conference on Artificial Neural Networks, pp. 93-103.
[18] Pang, G., Lu, K., Zhu, X., He, J., Mo, Z., Peng, Z. and Pu, B., 2021. Aspect-Level Sentiment Analysis Approach via BERT and Aspect Feature Location Model. Wireless Communications and Mobile Computing, vol.2021, no. 5534615, pp.1-13.
[19] Gao, Z., Feng, A., Song, X. and Wu, X., 2019. Target-dependent sentiment classification with BERT. IEEE Access, 7, pp.154290-154299.
[20] Gan, C., Wang, L., Zhang, Z., and Wang, Z. (2020). Sparse attention based separable dilated convolutional neural network for targeted sentiment analysis. Knowledge Based Syst. 188, pp.1–10.
[21] Rida-E-Fatima, S., Javed, A., Banjar, A., Irtaza, A., Dawood, H., Dawood, H. and Alamri, A., 2019. A multi-layer dual attention deep learning model with refined word embeddings for aspect-based sentiment analysis. IEEE Access, 7, pp.114795-114807.
[22] Lin, P., Yang, M. and Lai, J., 2021. Deep selective memory network with selective attention and inter-aspect modeling for aspect level sentiment classification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, pp.1093-1106.
[23] Feng, A., Cai, J., Gao, Z. and Li, X., 2023. Aspect-level sentiment classification with fused local and global context. Journal of Big Data, 10(1), pp.1-17.
[24] Havrlant, L. and Kreinovich, V., 2017. A simple probabilistic explanation of term frequency-inverse document frequency (TF-IDF) heuristic (and variations motivated by this explanation). International Journal of General Systems, 46(1), pp.27-36.
[25] Riadi, I. and Widiandana, P., 2019. Mobile Forensics for Cyberbullying Detection using Term Frequency-Inverse Document Frequency (TF-IDF). vol, 5, pp.68-76.
[26] Kaji, N. and Kobayashi, H., 2017. Incremental skip-gram model with negative sampling. arXiv preprint arXiv:1704.03956, pp.1-16.
[27] Wang, G.G., Deb, S., Gandomi, A.H., Zhang, Z. and Alavi, A.H., 2016. Chaotic cuckoo search. Soft Computing, 20(9), pp.3349-3362.
[28] Huang, L., Ding, S., Yu, S., Wang, J. and Lu, K., 2016. Chaos-enhanced Cuckoo search optimization algorithms for global optimization. Applied Mathematical Modelling, 40(5-6), pp.3860-3875.
[29] Xiong, R., Yang, Y., He, D., Zheng, K., Zheng, S., Xing, C., Zhang, H., Lan, Y., Wang, L. and Liu, T., 2020, On layer normalization in the transformer architecture. In International Conference on Machine Learning ,pp. 10524-10533.

[30] Pang, G., Lu, K., Zhu, X., He, J., Mo, Z., Peng, Z. and Pu, B., 2021. Aspect-Level Sentiment Analysis Approach via BERT and Aspect Feature Location Model. Wireless Communications and Mobile Computing, vol.2021, no. 5534615, pp.1-13.
[31] Rida-E-Fatima S., A. Javed, A. Banjar et al., "A multi-layer dual attention deep learning model with refined word embeddings for aspect-based sentiment analysis," IEEE Access, vol. 7, pp. 114795–114807, 2019.
[32] Xiong, R., Yang, Y., He, D., Zheng, K., Zheng, S., Xing, C., Zhang, H., Lan, Y., Wang, L. and Liu, T., 2020, On layer normalization in the transformer architecture. International Conference on Machine Learning ,pp. 10524-10533.