Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6, 3313-3326
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i6.2711

A hybrid BiLSTM-CNN deep learning model for Chinese sentiment analysis of online car reviews

Dongmei Lee[1], Dingran Wang[2], Wen Zhang[3], Zhihong Song[4*]
[1]School of Economics and Management, Shanxi University, China; lidm@sxu.edu.cn (D.L.)
[2]School of Mathematical Sciences, Shanxi University, China; wdr7582@163.com (D.W.)
[3]Honghu No.1 Middle School, Honghu, China; 1154896478@qq.com (W.Z.)
[4]Shanxi University, China; songzhihong@sxu.edu.cn (Z.S.)

Abstract: Although sentiment analysis with datasets in English has achieved significant progress, there are still relatively few studies on sentiment analysis in the area of Chinesecar review texts. In addition, the existing Chinese text sentiment analysis methods cannot simultaneously extract the contextual features and the local semantic information from the review texts. To address the above issues, a hybrid deep learning model namely BiLSTM-CNN was designed and validated with Chinese car review texts. We trained word vectors with the CBOW model by combining the Chinese Wikipedia corpus with other open-source car-related corpora. Such word vectors include car-specific vocabulary, which improves sentimental classification accuracy. Experimental results show that the performance indices of the deep learning models (CNN, LSTM, BiLSTM) are much better than the KNN, SVM, Naïve Bayes, and RF model, which is attributed to deep learning's powerful feature extraction ability and nonlinear fitting ability. Furthermore, when compared to deep learning models such as CNN, LSTM, BiLSTM, and LSTM-CNN, the suggested hybrid BiLSTM-CNN model outperformed the others. The results may provide references for consumers to buy a car and for car companies to optimize their products.
Keywords: *BiLSTM-CNN, Chinese car reviews, Deep learning, Sentiment analysis, Text mining.*

1. Introduction

According to the report released by China Internet Network Information Center (CNNIC), the number of Chinese Internet users has reached 1.032 billion in December 2021, and the Internet penetration rate reached 73.0%, with Chinese Internet users spending 28.5 hours per week on average. Many Internet users post product reviews and express their personal thoughts on online platforms, such as microblogs, Zhihu, and Xiaohongshu, resulting in a massive amount of text data. Fast and accurate emotion extraction and analysis from the vast text data is extremely valuable for government opinion monitoring, business market research, and personal consumption decisions.

With the rapid growth of the mobile Internet, an increasing number of automobile manufacturers are turning to social media and other mobile platforms for marketing. At the same time, customers are eager to learn about other users' social media experiences and opinions in order to make purchasing decisions. As a high-value commodity, practically all buyers utilize search engines to find out other people's experiences with their favorite car models before purchasing [1]. There are already several professional platforms in China that provide customers with car reviews, such as PCauto, Xcar, and 12365auto. On these platforms, a vast number of users who have purchased cars share their own usage experiences. For example, PCauto, one of China's largest automobile platforms, offers both professionally produced and user-generated content, allowing internet users to acquire critical insights into the quality of an automobile before purchasing. An important issue that must be addressed is how to fully utilize text mining and other related technologies to efficiently extract users' emotional information contained in the review texts, thereby assisting consumers in making accurate purchase decisions and assisting automotive companies in improving their products and services.

* Correspondence: songzhihong@sxu.edu.cn

Sentiment analysis, often referred to as comment mining or opinion mining, is an important research stream in the field of natural language processing (NLP). The text feature analysis of sentiment analysis includes three kinds of approaches: lexicons-based approaches [2], machine learning-based approaches [3], and deep learning-based approaches [4-7]. Though the first two approaches are simple to construct, their generalization capability and accuracy are poor. In recent years, deep learning has been widely employed as an emerging method in the fields of computer vision, speech recognition, and NLP. Deep learning-based techniques can automatically extract the syntactic and semantic features from text without the need for feature engineering, which is labor-intensive and time-consuming [8]. Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are two popular deep learning algorithms [4,5]. The CNN model is a type of artificial neural network developed by LeCun et al. [9] that has demonstrated excellent prediction performance in the field of image recognition and NLP. The RNN is another type of artificial neural network that exhibits temporal dynamic behavior due to the connectivity between the neurons inside the hidden layers. These two neural network models, however, have disadvantages. Although CNNs may successfully extract local text elements, the contextual semantic information between words is omitted. Traditional RNNs, on the other hand, suffer the problems of vanishing and exploding gradients. Hochreiter and Schmidhuber [10] proposed the Long Short-term Memory (LSTM) network by adding a long-term memory into the network to address the issues. However, the RNN (e.g., LSTM) model is incapable of extracting the local features of texts effectively because the contextual information is not fully extracted. Thus, Graves and Schmidhuber [11] developed the bidirectional LSTM (BiLSTM) model, containing a forward and a backward LSTM that may utilize information from both directions.

To solve the issue that existing Chinese text sentiment analysis algorithms cannot simultaneously extract both contextual features and local semantic information from review texts, a hybrid deep learning model called BiLSTM-CNN is proposed. The hybrid model combines the benefits of BiLSTM and CNN, successfully overcoming the shortcomings of a single model and increasing the efficiency of learning. The primary contributions of this study are as follows:

(1) We trained word vectors with the CBOW model by combining the Chinese Wikipedia corpus with other open-source car-related corpora (e.g.,https://gitee.com/peixilong/car-corpus-lmodels). These word vectors include car-specific vocabulary, which improves sentimental classification accuracy, which supplement existing literature on machine learning applications in the area of car review texts [12].

(2) A hybrid deep learning model integrating BiLSTM and CNN was proposed. BiLSTM is utilized first to automatically extract contextual features from the review texts, which are then fed into CNN to obtain local semantic features. Thus, it is possible to take advantage of both BiLSTM and CNN to extract features and grasp the semantics of the review texts. Furthermore, the prediction performance of the hybrid BiLSTM-CNN model has been shown to outperform other different machine learning models (KNN, SVM, Naïve Bayes, and RF) and deep learning models (CNN, LSTM, BiLSTM, and lexicon-based techniques LSTM-CNN).

The rest of the paper is structured as follows. Section 2 provides a brief overview of the evolution of sentiment analysis and deep learning technologies. The BiLSTM-CNN model is developed in Section 3. The datasets and data pre-processing are described in Section 4. Section 5 compares the prediction performance of the hybrid BiLSTM-CNN model with that of the traditional machine learning model (Naïve Bayes) and other deep learning models (CNN, LSTM, BiLSTM, and LSTM-CNN). Section 6 concludes the paper.

2. Related Works

Sentiment analysis refers to the process of identifying people's emotions, attitudes, and opinions about subjects, products, individuals, organizational services, events, and other entities to extract and classify audio, video, and text features [13]. Traditionally, analyzing algorithms generate a numerical value representing the polarity of a single sentence in terms of sentiment, which is classified into three categories: positive, negative, or neutral polarity [14,15]. The object of sentiment analysis is a text corpus, which may be the whole documents, sentences, words, or phrases [16]. According to the

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

granularity of the object, sentiment analysis can be classified into document-level, sentence-level, and aspect-level. Automatic text classification algorithms have been developed based on lexicons [17], machine learning, either supervised or unsupervised [18,19], and deep learning [20].

2.1. Lexicons-Based Approaches

The earliest technique for sentiment analysis of texts is the lexicon-based approach, which includes two types: dictionary-based approach and corpus-based approach [12]. Dictionary-based sentiment classification makes use of predefined dictionaries like WordNet and SentiWordNet [21]. Corpus-based sentiment analysis, on the other hand, is based on a statistical study of the content of the documents, rather than predetermined dictionaries [22].

Taboada et al. [23] proposed the lexicon-based method with the creation of a new dictionary. They employed Mechanical Turk to test the consistency and reliability of the dictionaries as they labeled the semantic orientation with dictionaries of words comprising polarity, strength, and negation. Bravo-Marquez et al. [24] combined automatically tagged tweets with manually created opinion lexicons to extend the opinion lexicons using a supervised algorithm. They constructed two distinct techniques, PMI-SO and SGD-SO, and experiments demonstrate that the lexicons developed by the study produced notable improvements over SentiWordNet in classifying tweets into polarity classes.

The challenge of the lexicon-based method is to create sentiment dictionaries. Many researchers and organizations have now published numerous sentiment dictionaries. HowNet, NTUSD, and BosanNLP are a few of the most popular sentiment dictionaries. However, because of the rapid development of the internet and the ongoing updating of information, specific sentiment dictionaries cannot be generalized in multiple sectors or languages [12]. Furthermore, sentiment classification with lexicon-based techniques cannot utilize contextual semantic information, which makes it difficult to transfer them to other domains [25].

2.2. Machine Learning-Based Approaches

To increase the accuracy of sentiment classification, machine learning techniques are progressively gaining acceptance among academics. Several well-known algorithms for sentiment classification are discussed by Kawade and Oza [26], such as Random Forest (RF), Naïve Bayes (NB), and Support Vector Machine (SVM) (SVM). For example, Na et al. [27] explored the effectiveness of SVM using different text features to classify online product reviews as recommended (positive sentiment) and not recommended (negative sentiment). Shi et al. [28] proposed a multilayer Naïve Bayesian classifier for sentiment classification of movie reviews and showed that it outperformed other traditional approaches. Singh et al. [29] examined machine learning classifiers for sentiment analysis, including Naïve Bayes, J48, BFTree, and OneR, and found that Naïve Bayes had the fastest learning rate and the best classification accuracy. Al-Bakri et al. [30] employed conventional machine learning (CML) algorithms to evaluate Iraqi e-tourism firms based on extracting sentiments from Iraqi dialect reviews. on Facebook. Mikolov et al. [5] designed the Word2Vec model that laid the foundation for text processing using deep learning models. The Word2Vec consists of the Skip-gram model and the CBWO model, which acquires semantic information through a large amount of texts with an unsupervised learning method. The most important feature of the Word2Vec model is the vectorized representation of all words so that the connections between words can be analyzed quantitatively.

2.3. Deep Learning-Based Approaches

Deep learning has arisen in recent years, with increasing applications in sentiment analysis of NLP, because traditional machine learning approaches are labor-intensive in preprocessing the data as well as identifying and extracting important features [12,20]. Deep learning algorithms model the activity of neurons in the human brain when activated [20]). As compared to machine learning, deep learning can better deal with the issues of feature sparsity and feature extraction, as well as improve the efficiency of text analysis. Sentiment analysis based on deep learning techniques consists of two main steps. The first step is to transform the text corpus into a high-dimensional dense vector, i.e., text vectorization. The next stage is to create a deep-learning model for text sentiment analysis. Much of the work in deep

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

learning in the field of NLP has focused on word vector representations utilizing neural language models, with the Word2Vec methodology being the most common method used to construct word vectors [5].

Commonly used models for sentiment analysis in deep learning include CNN, RNN, and LSTM, as well as variants and hybrids of these models, such as bidirectional Long Short-Term Memory (BiLSTM), gated recurrent units (GRU), and bidirectional gated recurrent units (BiGRU) [20, 31]. A convolutional neural network (CNN) is a form of an artificial neural network created by LeCunet al. [9]. A CNN consists of a series of convolution and sub-sampling layers followed by a fully connected layer and a normalizing (e.g., softmax function) layer [31]. Liao et al. [32] used CNN in Twitter sentiment classification and demonstrated that it is more accurate than traditional machine learning-based approaches such as SVM. Socher et al. [33] built a sentiment tree library, introduced fine-grained sentiment labels, and used RNN for sentence-level sentiment prediction. Bengioet al. [34] pointed out that the original RNN had its own limitations, such as gradient disappearance and gradient explosion when dealing with a long series. To address these issues, Hochreiter and Schmidhuber [10] proposed Long Short-Term Memory (LSTM), which addressed the nonlinearity of information and tackled the vanishing and exploding gradient problems inherent in traditional RNNs by keeping essential information in memory cells and eliminating futile information.

In addition to using a single deep learning model, some literature has attempted different hybrid models for sentiment analysis (Xu et al., 2022). Ombabiet al. [35] proposed a novel deep learning model for Arabic language sentiment analysis based on one layer of CNN architecture for local feature extraction, and two layers of LSTM to maintain long-term dependencies. The feature maps learned by CNN and LSTM were passed to the SVM classifier, which demonstrated better classification performance and accuracy of 90.75%. Feng and Cheng [36] proposed a novel sentiment analysis model by integrating a multi-channel convolutional neural network with a multi-head attention mechanism (MCNN-MA). Their experimental results showed that the MCNN-MA model had a higher classification accuracy and a relatively low training time cost compared with other baseline models. Ramaswamy and Jayakumar [37] proposed a neural network structure based on LSTM and CNN, which combined the explicit knowledge from the external database (RecogNet) with the implicit information of the LSTM model to improve sentiment accuracy. Khan et al. [38] developed a hybrid architecture using CNN and LSTM, and experiments showed that its accuracy for text classification in Roman Urdu and English was improved by 5% on a relevant corpus. Sitaula et al. [39] proposed three different CNN models for the sentiment classification of tweets using fastText-based feature extraction, domain-specific probability-based feature extraction, and domain-agnostic probability-based feature extraction respectively.

In sum, previous research has shown that traditional approaches to developing sentiment classification models such as dictionary-based methods that rely on lexicons (e.g., WordNet) and traditional machine learning algorithms (e.g., Naïve Bayes and Random Forest) often fail to accurately predict the polarity of car review texts due to their poor feature extraction and nonlinear fitting capabilities.

3. BILSTM-CNN Model

In this paper, a hybrid deep learning model integrating BiLSTM and CNN was proposed for sentiment analysis on Chinese car review texts. The BiLSTM is designed to capture the semantics by containing a forward and a backward LSTM, which better compensates for the inability of LSTM to encode information from both directions [12]. However, using the BiLSTM model directly may result in too much computing cost owing to a large number of input dimensions. Thus, we first used BiLSTM to automatically extract contextual features from the review texts, which were then used as the input of CNN to extract local semantic features. It is possible to take advantage of both BiLSTM and CNN to extract features and understand the semantics of the review texts. The structure diagram of BiLSTM-CNN is shown in Figure 1.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

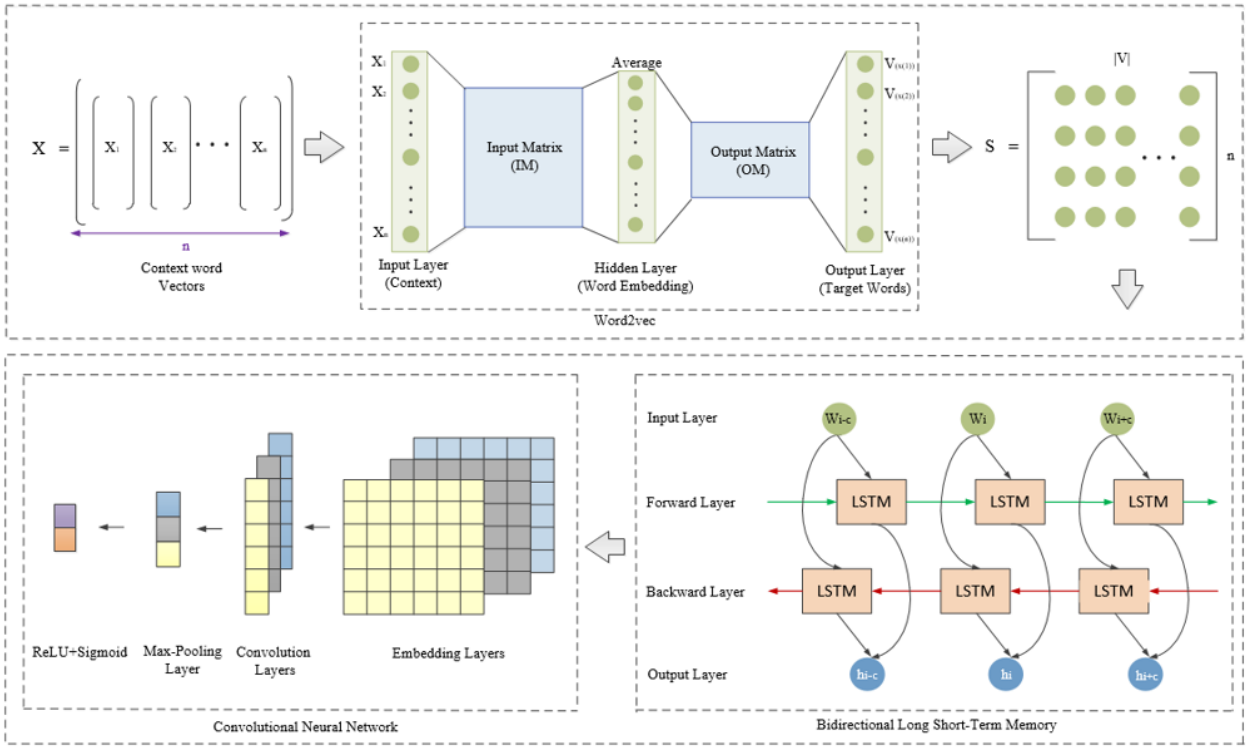

Figure 1.
The structure diagram of BiLSTM-CNN Model.

(1) To classify text data, it is necessary to convert the text data into a mathematical representation that can be processed by the model. One of the most common approaches is to use a distributed word vector to represent a word. Word2Vec is a classical method developed by Google in 2013 that creates word embeddings in the field of NLP. Word2Vec uses neural networks to create word embeddings such that embeddings with similar word meanings tend to point in a similar direction. The word2vec includes two training models: The Continuous Bag-of-Words (CBOW) model and the skip-gram model. The former uses context words to predict the target words. Conversely, the latter uses the target words to predict the context words.

The embedding layer transformed the input text $X = \{x_1, x_2, \cdots x_n\}$ into a lexicon of word vectors that can be recognized by the model through the CBOW or the skip-gram model. In this paper, the word vector was trained by the CBOW model with the dimensionality set as 300. In this way, it satisfied the model's demand for a large amount of corpus on the one hand, and on the other hand, the large amount of car terminology greatly improved the quality of the word vector. If no word in the input text appears in the dictionary, the values of all dimensions of the word are taken as 0. Then the word vectors of all words are stitched together to generate the word vector matrix S_{ij}.

$$S_{ij} = [V(x(1)), V(x(2)), \cdots, V(x(i))] \quad (1)$$

where $V(x(i))$ is the word vector of the ith word, S_{ij} is the j-dimensional word vector matrix consisting of i words merged together.

(2) The input of the BiLSTM layer comes from the word vectormatrix of the embedding layer. The hidden layer was set to 128 in this paper and the *sigmoid* function was used as the activation function. The sequences of the word vector matrix were used as input of the forward and backward LSTM, which used the contextual information features of the text from both directions and then took the data from both directions into the stacking operation as follows.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

$$h_{ijt} = BiLSTM(S_{ijt}) \quad (2)$$

where S_{ijt} is the word vector matrix in time t, h_{ijt} is the output of the BiLSTM layer in time t.

(3) The convolutional layer used the fine-tuning parameters for TextCNN to extract local features. The size of the convolutional kernels used for the combination was (3, 4, 5) and the number of each convolutional kernel was set as 128.

$$O_{ijt} = CNN(h_{ijt}) \quad (3)$$

where O_{ijt} is the output of CNN in time *t*. In the following experiment, the step size was set as 1. The ReLU activation function was used, and the maximum pooling was chosen to reduce the dimension of the matrix that had undergone convolutional operations, and was then connected by a fully connected layer.

(4) The output layer classified the sentiment of the data after a series of processing by a classifier. The experiment in this paper was a binary classification of the sentiment, so the sigmoid function was used.

$$y_i = sigmoid(w_i d_{ijt} + b_i) \quad (4)$$

where w_i is the weight matrix of the fully connected layer, b_i is the bias term, and d_{ijt} is the output of the fully connected layer.

4. Discussion

4.1. Datasets

To evaluate the performance of the proposed model for Chinese sentiment analysis, this paper used dataset from the following three websites: Pcauto (https://www.pcauto.com.cn/), Xcar (https://www.xcar.com.cn/), and 12365auto (http:// www. 12365auto. com/).

Pcauto: It is one of the largest automobile portals in China. It provides both comprehensive and the latest product information on various domestic and international automobiles through its local guides and directories serving the major cities in China.

Xcar: To date, Xcar is the largest car-themed community in China and the world, including 196 mainstream car clubs, and 43 personalized discussion forums. As an important portal for car lovers before and after purchasing a car, the car-themed community provides a more up-to-date, realistic, and professional user experience, car information, and an interactive communication platform.

12365auto: Based on its unique advantages in collecting a large amount of customer complaint information, producing professional evaluation content of auto products, and data analysis and research, 12365auto can not only help solve auto quality and service problems encountered by car owners and provide important information reference for auto consumers, but also provide authoritative data support and solutions for car companies and dealers in product development, quality control, and after-sales service improvement.

Since the review texts may involve different aspects related to car performance, the short text is not enough to describe it clearly. Therefore, it is necessary to improve the data quality by filtering the length of the review texts by setting the minimum text length to 80 characters. Two types of reviews with opposite polarities appear on car websites, ranging from "most satisfied" to "least satisfied" or "good" to "bad". In this paper, the positive comments were labeled with a value of 1, while the negative comments were labeled with 0.

The car review texts were obtained using a Python crawler from the above three websites on Nov. 22th, 2020. By de-duplication, 40209 valid comments were finally obtained, including 20465 positive comments and 19744 negative comments, and the data volume of both sentiment polarities was basically balanced. Examples of the car review texts are shown in Table 1.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

Table 1.
Examples of Chinese car review texts.

Label	Original review texts	Translation
1	电子助力转向非常轻盈，一个手指就可以自由得转动，然后就是它的空间真的是非常大，我身高 183 坐进去正常坐姿头部空间还有一拳左右非常的宽裕，这也是它的一大亮点。还有就是我们高速开起来稳定性还是很不错，我在高速上开到过 150 多也不会感觉车子轻飘飘的，再说说它得刹车性能，在日常用车过程中这套刹车系统还是可以给足我们足够的信心，整个刹车用起来还是很线性，基本上是你踩多少就给你多少。	Electronic power steering is very light, and you can manipulate it even with a finger. Then, its space is really large. I am 1.83m tall, and when I sit in the normal position, the space between my head and the car ceiling is about a fist or so, which is a major highlight. Besides, the stability of high-speed driving is very good. The car didn't seem to be particularly light when I drove over 150 mph on the highway. Finally, I would say that it has an excellent performance in braking in daily use. The brake system is still very smooth and you can brake as much as you want.
0	仪表盘没有水温表，只有高水温报警，这个问题其实很严重；雨刮器不好用，声音响还刮不干净。准备换无骨的用用看；驾驶侧玻璃在上升快到位时，明显的“疙瘩”一下，原因查不出来。在行驶时，有时听到驾驶侧门有异响；减震效果不好，低速通过不平路段时，都感觉车身抖动。3 档升 4 档有点困难，一定要到 2000 转以上才行。	The dashboard does not have a water temperature gauge, only equipped with a high water-temperature alarm, which is actually a very serious problem. The wipers do not work well, the sound is loud, and can't clean the front windshield. (I'm) ready to replace them with the boneless wipers; The lift of glass on the driver's side is also problematic and there is an obvious "lump", with the reason not being detected. When you are driving, you may sometimes hear the noise in the driver's side door; The shock absorption is poor and you may feel the car body shaking when passing through the uneven road at low speed. It is a bit of difficulty to move from 3rd up to 4th gear. (To do so,) the engine speed must reach 2000 rpm.

4.2. Data Preprocessing

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Preprocessing is a technique that is used to convert the raw data into a clean dataset.

First, we removed duplicate car review records. On the one hand, duplicate information will inevitably appear in the data crawling process, and on the other hand, the comments themselves may have duplication problems. In this paper, some information was even repeated up to 5 times in the original datasets. These increase the training time of the model during model training on the one hand, and seriously affect the reliability of the model training parameters on the other hand, which in turn leads to the low prediction performance of the sentiment classification. Therefore, it is very necessary to filter out duplicate information.

Second, we removed non-textual content from the data. In most cases, there is a lot of useless content in the fetched text data, such as HTML, codes, CSS tags, and unwanted punctuation, which need to be removed before sentiment analysis. Python's regularization works and Python's Beautiful Soup library were used as techniques for removing the non-text content.

Third, tokenization was used by splitting the sentences into words known as tokens. Western languages (e.g. English) naturally use spaces as word separators, so tokenization can be achieved by using only spaces or punctuation. In contrast, Since Chinese and some adherent languages (e.g. Japanese, Korean, etc.) are written without word segmentation markers, the meaning of a sentence may be completely opposite if it is split in different ways. For example, for the sentence "这辆汽车不大精致" (This car is not so exquisite), one way of tokenization is “这辆/汽车/不大精致”(This/car/is not so exquisite), which is a negative sentiment by evaluating the car from the appearance. Another way of tokenization is “这辆/汽车/不大/精致” (This/car/is not (big)/ so exquisite), which is evaluated from two aspects. One aspect indicates that the car is small and doesn't have enough room, which may be a negative or positive sentiment. The other aspect shows that the appearance of the car is exquisite, which is an obvious positive sentiment. Thus, it is necessary to accurately split the review texts. The result of the tokenization is a list of tokens separated without punctuation.

At present, there are more and more mature tools for Chinese text tokenization, such as jieba tool in python, ICTCLAS of CAS, and LTP-cloud of HUST, which are all open-source software. In this paper, we chose jieba for tokenization. Because all the code in this paper was done with python, jieba is easy to

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

install and widely used. Jieba word splitting has 3 modes: (1) the precise mode, which attempts to make the most accurate results after word splitting; (2) the full mode, which has the advantage of outputting all possible words of the sentence; (3) the search engine mode, which is based on the precise mode, and the long words are split and re-split again to improve the recall rate, which is suitable for search engine splitting. In this paper, we chose the precise mode, which is also the default mode of jieba. Tab. 2 shows the tokenization for Chinese review texts.

Table 2.
Examples of tokenization for Chinese care review texts.

Texts with tokenization	Translation
动力/吧/车子/也/不重/动力/应该/也/要点/表现/的/提速/快/动力/好/超车/随随便便/的/油门/就/过去/了	As for power, (since) the car is not heavy, the (engine) power ought to be strong with start and it is easy to overtake other cars by putting a little pressure on the gas.
反正/能/满足/我/的/要求/一人/一个/标准/吧/只要/给油/也有/推背感	Anyway, it can meet my requirements. everyone has their own rules. As long as you push the gas pedal, there is a feeling of being pushed into the seat (due to rapid acceleration).
市区/开/只有/多余/的/力/了/高速/两个/人/开起/刚好/满载/时/动力/就要/差些	(If you drive) in the city, (there may be) extra power; (but if you drive) on the highway with two people and the car is just fully loaded, the power will be worse.

Finally, the removal of stop words was applied. Stop words are words that are not necessary to be present in a sentence. In Chinese texts, stop words are words without real meaning, mainly including pronouns, conjunctions, prepositions, etc. For example, "我(I)", "不仅(not only)", "啊(oh)", "的(and the like)" and so on. These words have no effect on sentiment expressions but appear very frequently, which may slow down the training model. Thus, those stop words are usually removed before text mining.

In this paper, the 10-fold cross-validation technique was used to split the training and test dataset. The input of the model was the index of each word in the word vector matrix, and then the index was used to find the corresponding word vector in the embedding matrix. If the word is not in the dictionary, the index value is 0. Since the length of different text indexes was not the same, it was not conducive to the batch processing of data. Usually, a fixed length was set, and the text that was longer than the fixed length would be intercepted to the fixed length, while the text that was less than the fixed length would be made up with the value of 0. As for the choice of the fixed length, if the shortest text length was chosen as the fixed length, it would waste a lot of effective data. However, if the longest text length was taken as the fixed length, it would cause a waste of computational resources. Taking into account the influence of some extreme values, the final fixed length of 100 words was chosen in this paper.

4.3. Word Vector Training

After pre-processing, the text data need to be vectorized (Word Embedding) to be represented as the input of the text sentiment analysis model. In this paper, we used the Wikiipedia Chinese corpus as the base, and then added car-related review text corpus for the final word vector training, and used Google's Word2Vec tool for word vector training, and selected the CBOW neural network language model [5]. Mapping each word as a K-dimensional real vector is able to avoid the feature sparsity problem while reducing the dimensionality of the word vectors. The word vectors obtained from the training can be used to determine the semantic similarity between words based on the cosine distance or Euclidean distance between each other. The dimensionality of the word vectors is 300. We then fine-tune the word vectors along with other model parameters during training.

4.4. Evaluation Metrics

The confusion matrix is usually used to evaluate the goodness of fit of the model. There are four kinds of results: the sample is positive and still positive after the model classification, which is classified

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

as True Positive (TP); the sample is positive but negative after the model classification, which is classified as False Negative (FN); the sample is negative and positive after the model classification, which is classified as False Positive (FP); and the sample is classified as True Negative (TN) when it is still negative after the model classification.

Accuracy is the ratio of the sum of the number of all correct predictions to the number of all samples. It describes how frequently the sentiment rating predicted by the model is correct. Accuracy is calculated as:

$$accuracyy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Precision is the ratio of the number of positive evaluations predicted correctly to the number of all predictions that are positive. A larger value of precision indicates a better model.

$$precision = \frac{TP}{TP + FP} \quad (6)$$

Recall is the ratio of the number of positive evaluations predicted correctly to the number of positive evaluations for the entire sample. A larger value of recall rate indicates a better model.

$$recall = \frac{TP}{TP + FN} \quad (7)$$

Bothprecision and recall of test data are used to calculate the F1 score. It is calculated as follows.

$$F1 = \frac{2\,precision \times recall}{precision + recall} \quad (8)$$

4.4. Model Parameters and Comparison of Experiments Results

The experiment was based on the Keras deep learning framework and used ten-fold cross-validation. The training process used the dropout mechanism and the model accuracy was highest when the dropout rate was set to 0.5. The window size of the convolution kernels selected by the three channels are (3, 4, 5) and the batch size was set to 128. The Adam optimizer was used in the training process, and the initial learning rate was 0.001. Tab. 3 lists the parameters and Tab. 4 shows the prediction performance of the various models.

Table 3.
Parameters of the BiLSTM-CNN model.

Parameters	**Value/Category**
Dropout rate	0.5
Window size of the convolution kernels	(3,4,5)
Number of convolution kernel	128
Optimizer	Adam
Learning rate	0.001
Cross-entropy loss function	Binary
Classifier	Sigmoid

To validate the proposed hybrid model, the performance of the proposed model was compared to nine baseline models: KNN, SVM, Naïve Bayes, RF, CNN, LSTM, BiLSTM, and LSTM-CNN. As can be seen from Tab. 4, the performance indices of the deep learning models (CNN, LSTM, BiLSTM) are much better than the traditional machine learning model (e.g., Naïve Bayes model, RF, SVM, and KNN). Compared with the traditional machine learning model, the feature extraction of the neural network model uses word vectors trained by Word2vec, which are rich in semantic and contextual information. Therefore, the experiment results illustrated that deep learning models were more effective in sentiment analysis compared to traditional machine learning models (e.g., Naïve Bayes in this paper), which was attributed to the powerful feature extraction ability of deep learning and its nonlinear fitting ability.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

Table 4.
Summary of the prediction performance for different models.

Model	Accuracy	Precision	Recall	F1
(a) KNN	62.82%	60.71%	82.65%	70.18%
(b) SVM	72.26%	66.28%	88.64%	79.24%
(c) Naïve Bayes	73.37%	69.32%	95.34%	80.27%
(d) RF	78.61%	80.59%	90.71%	82.13%
(e) CNN	94.42%	91.67%	97.81%	94.64%
(f) LSTM	94.91%	93.34%	96.81%	95.05%
(g) BiLSTM	95.52%	94.64%	96.60%	95.60%
(h) LSTM-CNN	95.96%	95.51%	96.52%	96.01%
(i) BiLSTM-CNN	96.53%	97.70%	95.36%	96.52%

In addition, compared with the deep learning models, i.e., CNN, LSTM, BiLSTM, and LSTM-CNN, the proposed hybrid model obtained the best prediction performance in all performance indices except the recall rate. Compared with CNN, the accuracy rate, precision rate and F1 score of the BiLSTM-CNN model increased by 2.11%, 6.03%, and 1.88%, respectively. This is due to the temporal characteristics of the BiLSTM model, which is able to link the contextual relationships and thus has better predictive performance. Compared with BiLSTM, the accuracy rate, precision rate and F1 score of the BiLSTM-CNN model increased by 1.01%, 3.06%, and 0.92%, respectively. Since the proposed hybrid model integratedCNN to reduce the dimensionality of the features and thus could extract features more effectively and used them for sentiment analysis, which further improved the prediction accuracy and operational efficiency of the BiLSTM-CNN model.

Furthermore, as shown in Tab. 4, compared with the hybrid model of LSTM-CNN, the accuracy rate, precision rate and F1 score of the proposed hybrid model of BiLSTM-CNN increased by 0.57%, 2.19%, and 0.51%, respectively. Unlike standard LSTM, BiLSTM used the input flows in both directions, and it was capable of utilizing information from both sides. It was also a powerful tool for modeling the sequential dependencies between words and phrases in both directions of the sequence. Again, the only difference between LSTM-CNN and BiLSTM-CNN was the possibility for BiLSTM to leverage future context chunks to learn better representations of single words. As a result, the BiLSTM model is beneficial in some NLP tasks, such as sentence classification, translation, and entity recognition.

We further conducted the regular two-dimensional confusion matrixes to demonstrate the classification performances of different models based on testing data. As shown in Fig.2, the hybrid model BiLSTM-CNN demonstrated the best classification performance, followed by the LSTM-CNN, BiLSTM, LSTM, CNN, RF, NB, SVM, and KNN.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

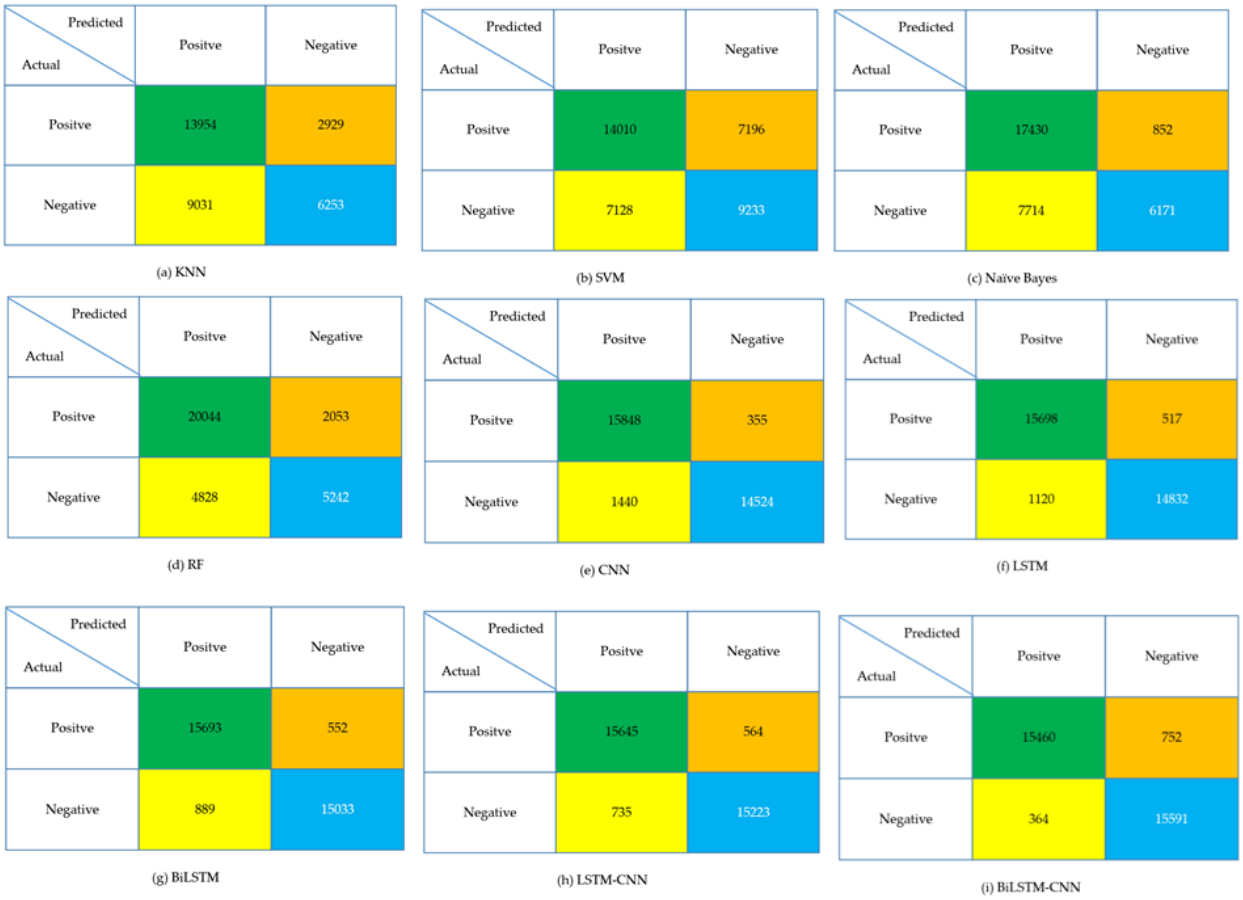

Figure 2.
Confusion matrixes for different models.

4.6. Word cloud of Chinese Car Reviews

In order to visualize the aspects that users prioritize, we created the word cloud of Chinese car reviews. As shown in Fig.3, the performance evaluations of the car are the most mentioned aspects in the reviewtext, such as "空间"(space), "内饰(interior system)", and"外观 (appearance design)". These are exactly in line with the theme classification of the car website, which indicates the high quality of the text selected in this paper. There are also some technical terms related to cars, such as "刹车(brake)","发动机(engine)"and "变速箱 (transmission box)". The correct classification of these words shows the reliability of tokenization withjieba. Some emotional words also appear with high frequency, such as "满意(satisfactory)", "好(good)","不好(bad)", which are exactly in line with the text requirements of the text for sentiment analysis.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

Figure 3.
Word cloud of car reviews.

5. Conclusions

With the rapid growth of the mobile Internet, car reviews by internet users have been constantly increasing. As a high-value commodity, almost all consumers use search engines to find out other users' comments about their favorite car models before they buy [1]. How to make full use of text mining and other related technologies to efficiently extract users' emotional information contained in the review texts, so as to help consumers in making accurate purchase decisions, and help automotive companies improve their products and services, is an important issue that needs to be addressed.

To address the problems that existing Chinese text sentiment analysis methods cannot simultaneously extract the contextual features and the local semantic information of review texts, a hybrid deep learning model namely BiLSTM-CNN was designed. The hybrid model incorporated the advantages of BiLSTM and CNN, successfully overcoming the drawbacks of a single model and making learning more efficient.

To evaluate the performance of the proposed model for Chinese sentiment analysis, this paper used care review datasets from three websites: PCauto, Xcar, and 12365auto. Experimental results based on Chinese car review texts showed that the performance indices of the deep learning models (CNN, LSTM, BiLSTM) were much better than the Naïve Bayes model, which was attributed to the powerful feature extraction ability of deep learning and its nonlinear fitting ability. In addition, compared with the deep learning models, i.e., CNN, LSTM, BiLSTM, and LSTM-CNN, the proposed hybrid BiLSTM-CNN model obtained the best prediction performance in all performance indices except the recall rate. In the e-commerce environment, car makers can infer consumers' repurchase intentions and behavior through the sentiment polarity of car review texts, which provides a reference basis for car makers to develop personalized marketing strategies.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

References

[1] Li, D.; Li, M.; Han, G.; Li, T. A combined deep learning method for internet car evaluation. Neural Comput. Appl. 2021, *33*, 4623-4637

[2] Wang,Y.; Wang, M.; Xu,W. A sentiment-enhanced hybrid recommender system for movie recommendation: A big data analytics framework. Wirel. Commun. Mob. Comput. 2018, 1, 1-9

[3] Revathy, G.; Alghamdi,S.A; Alahmari, S.M; Yonbawi, S.R.; Kumar, A.; Haq,M.A. Sentiment analysis using machine learning: Progress in the machine intelligence for data science. Sustain. Energy Techn. 2022, *53*,102557

[4] Kim,Y.Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar.2014, 1746-1751

[5] Mikolov, T.; Corrado,G.;Chen,K.;Dean,J.Efficient estimation of word representations in vector space. In *Int. Conf. Learn. Represent., ICLR - Workshop Track Proc.*, Scottsdale, AZ, USA. 2013, 1-12

[6] Zhou,C.; Sun,C.; Liu, Z.; Lau, F.C.M. A C-LSTM neural network for text classification. arXiv:1511.08630, 2015,https://doi.org/10.48550/arXiv.1511.08630

[7] David,M.S.; Renjith, D. Comparison of word embeddings in text classification based on RNN and CNN. In IOP Conference Series: Materials Science and Engineering. 2021, 1-9

[8] Chen, T.; Xu, R.; He, Y.; Wang,X.Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Syst. Appl.2017, 72, 221-230

[9] LeCun, Y.; Bottou,L.; Bengio, Y.; Haffner, P.Gradient-based learning applied to document recognition. Proc. IEEE.1998, 86, 2278-2324

[10] Hochreiter, S.; Schmidhuber, J.Long short-term memory. Neural Comput.1997,9, 1735-1780

[11] Graves,A.;Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks.2005, 18, 602610

[12] Xu, Q.; Chang,V.; Jayne,C.A systematic review of social media-based sentiment analysis: Emerging trends and challenges.Decis. Anal. J. 2022, *3*, 100073

[13] Serrano-Guerrero, J.; Olivas,J.A.; Romero,R.P.; Herrera-Viedma, E. Sentiment analysis: A review and comparative analysis of web services. Inform. Sciences. 2015, 311,18-38

[14] Choi,Y.; Cardie,C.Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In *EMNLP - Proc. Conf. Empir. Methods Nat. Lang. Process.: Meet. SIGDAT, Spec. Interest Group ACL, Held Conjunction ACL-IJCNLP*, Singapore. 2009,590-598

[15] Pai, P.-F.; Liu, C.-H.Predicting vehicle sales by sentiment analysis of Twitter data and stock market values. IEEE Access. 2018, 6, 57655-57662

[16] Wilson,T.;Wiebe,J.;Hoffmann,P. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Comput. Linguist. 2009, *35*, 399-433

[17] Li, H.; Chen, Q.; Zhong,Z.; Gong, R.; Han, G. E-word of mouth sentiment analysis for user behavior studies. Inform. Process. Manag. 2022, *59*, 102784

[18] Ahuja, R.; Sharma, S.C.Sentiment Analysis on Different Domains Using Machine Learning Algorithms. In *Advances in Data* and Information Sciences,1st. Trivedi,S. M. C.;Kolhe,M. L.;Mishra,K.K.; Singh ed.,B. K. Singapore. 2022, 143-153

[19] Pilar, G.; Isabel,S.; Diego,P.; Luis,G.J. A novel flexible feature extraction algorithm for Spanish tweet sentiment analysis based on the context of words. Expert Syst. Appl. 2023, 212, 118817

[20] Joseph, J.; Vineetha, S.G.; Sobhana, N.V.A survey on deep learning based sentiment analysis. Mater. Today Proc. 2022,58, 456-460

[21] Baccianella,S.; Esuli,A.; Sebastiani,F. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In*Proc. Int. Conf. Lang. Resourc. Eval., LREC*, Valletta, Malta. 2010, 2200-2204

[22] Dang,N.C.; Moreno-García,M.N.; De la Prieta,F. Sentiment analysis based on deep learning: A comparative study. Electronics-SWITZ.2020,9, 483

[23] Taboada,M.; Brooke,J.; Tofiloski,M.K.; Voll,K.;Stede,M. Lexicon-Based Methods for SentimentAnalysis.Comput. Linguist. 2011, 37, 267-307

[24] Bravo-Marquez, F.; Frank,E.; Pfahringer,B.Building a Twitter opinion lexicon from automatically-annotated tweets. Knowl-Based Syst. 2016,108, 65-78

[25] Fang,Z.; Zhang,Q.;Kok,S.; Li, L.;Wang,A.;Yang,S. Referent graph embedding model for name entity recognition of Chinese car reviews. Knowl. Based Syst. 2021,233, 107558

[26] Kawade,D.R.; Oza,K.S. Sentiment Analysis: Machine Learning Approach. Int. J. Eng. Technol. 2017,9, 2183-2186

[27] Na,J.C.; Khoo, C.; Wu,P. Use of negation phrases in automatic sentiment classification of product reviews.Libr. Collect. Acquis.2005,29, 180-191

[28] Shi, J.Z.; Guo,L.; Wei,S.M. A Multi-Level Naïve Bayes Classifier for Sentiment Classification.Adv. Electron. Forum. 2012,6-7, 553-560

[29] Singh,J.; Singh,G.; Singh,R.Optimization of sentiment analysis using machine learning classifiers.Hum.-Cent. Comput. Info.2017, 7, 1-12

[30] Al-Bakri,N.F.; Yonan,J.; Sadiq,A.T.; Abid,A.Tourism companies assessment via social media using sentiment analysis. Baghdad Sci. J. (2021). 2021, 19, 422-429

[31] K.Shrestha,A.; Mahmood,A.Review of deep learning algorithms and architectures. IEEE Access. 2019,7, 53040-53065

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711

[32] Liao, S.; Wang,J.; Yu, R.; Sato, K.; Cheng,Z. CNN for situations understanding based on sentiment analysis of twitter data. Proc. Comput. Sci.2017, 111, 376-381
[33] Socher,R.; Pennington,J.; Huang,E.H.; Ng,A.Y.; Manning,C.D. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *EMNLP - Conf. Empir. Methods Nat. Lang. Process., Proc. Conf.*, Edinburgh, Scotland, UK. 2011, 151-161
[34] Bengio,Y.; Simard,P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. IEEE T Neural Networ.1994,5, 157-166
[35] Ombabi, A.H.; Ouarda,W.; Alimi,A.M. Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks. Social Network Analysis and Mining.2020, 10, 1-13
[36] Feng,Y.;Cheng,Y. Short Text Sentiment Analysis Based on Multi-Channel CNN With Multi-Head Attention Mechanism. IEEE Access.2021,9, 19854-19863
[37] Ramaswamy,S.L.; Jayakumar,C.RecogNet-LSTM+CNN: a hybrid network with attention mechanism for aspect categorization and sentiment classification. J. Intell. Inf. Syst. 2022, 58, 379-404
[38] Khan,L.;Amjad,A.;Afaq,K.M.; Chang, H.Deep sentiment analysis using CNN-LSTM architecture of English and Roman Urdu text shared in social media. Appl. Sci. 2022, 12, 1-18
[39] Sitaula,C.; Basnet,A.; Mainali,A.; Shahi, TB.Deep Learning-Based Methods for Sentiment Analysis on Nepali COVID-19-Related Tweets. ComputIntellNeurosci. 2021, 1, 1-11
[40] Wikimedia Downloads[EB/OL]. https://dumps.wikimedia.org/zhwiki/

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 3313-3326, 2024
DOI: 10.55214/25768484.v8i6.2711