Sentiment Analysis and Emotion Detection using Text Mining in Natural Language Processing

Pramjit Rana[1], Sagar Choudhary [2]

[1] M-Tech Student, Department of CSE, Quantum University, Roorkee, India.

[2] Assistant Professor, Department of CSE, Quantum University, Roorkee, India.

Abstract

There are many studies being conducted in the field of natural language processing related to sentiment analysis due to the growth of computer science and social media. However, there are many applications of sentiment analysis, and it might be difficult for researchers to review and understand past studies in order to see the trends. Thus, this study proposes trends for sentiment analysis research to assist sentiment analysis researchers. The data sources we obtained from our collections are Google Scholar and Scopus. Therefore, reviews of studies that are using multiple types of modalities have examined change from a particular event and from language itself have been researched more frequently in the last few years. Sentiment analysis is a powerful driver of natural language processing with huge implications for business, social media, healthcare, and disaster response, among others. In this review, we explore the complex landscape of sentiment analysis, its implications, challenges, and recent trends. We cover several important topics, such as the best dataset to use, the best algorithm to use, whether to consider different languages, and various new sentiment tasks. We evaluate the applicability of existing datasets (e.g., IMDB Movie Reviews, Twitter Sentiment Dataset) and deep learning approaches (e.g., BERT) to sentiment analysis. There has been progress in sentiment analysis, however, emerging challenges remain (e.g., sarcasm and irony, ethical challenges, and new domains). We believe sentiment analysis is dynamic in nature, so future research is necessary to capture the complexity of sentiment expression exhibited by humans, as well as to ensure ethical and effective use of sentiment analysis across specializations and languages.

Keywords: - IMDB Movie Reviews, Twitter Sentiment Dataset, support vector machine (SVM), Naive Bayes, Logistic Regression, social media, healthcare, and disaster response.

1. INTRODUCTION

Text mining is a means of extracting valuable information from textual data. Although it is utilized as an analytical method in research in areas such as medical informatics, social science, and design, there is a limited amount of research related to text mining [1]. Text mining can be used for a variety of applications, text classification and question answering, among others. Sentiment analysis refers to the analysis of subjective information, such as opinions, sentiments, evaluations, and attitudes in texts, using a computer. Research in Natural Language Processing revolves around sentiment analysis, and research in data mining, text mining, and web mining covers this topic as well. Sentiment analysis is not only part of the computer science sub-domain but is also a growing awareness in social science and management science as it becomes more relevant in management and society. The importance of sentiment analysis is reinforced by the rapid expansion of social media, such as reviews and Twitter [2]. Research is even conducted in financial and pharmaceutical industries. Increasingly people are expressing their opinions, and finding information, in social media or comments because of the Internet. Sentiment analysis lends itself to these types of growing trends in research, and therefore it is highly likely that research will continue in this domain.

In recent years, sentiment analysis has become an essential part of the natural language processing (NLP) field, particularly given the explosive growth of digital data and the need for analysis of textual data. Social media platforms, online reviews, customer feedback, and many more textual data sources became available providing a need to understand the sentiment of what is contained in a text in order to derive value. Sentiment analysis has applications in business intelligence, social sciences, and beyond. Sentiment analysis (with opinion mining as the other name) is the computational process of identifying, extracting, and analyzing subjective information in a text to determine the overall sentiment (or attitude) expressed towards an entity, product, topic, or event [3–4]. This is commonly done by taking a text and classifying it into potentially predefined categories (positive, negative or neutral with slightly more nuanced emotions and opinions) [3–4]. Historically, sentiment analysis was performed with machine learning algorithms such as support vector machine (SVM), Naive Bayes, Logistic Regression, and also with supervised and unsupervised algorithms.

2. Literature Review

Prior studies fall into two categories. The first category includes studies which analyzed trends of research using text mining. Kim & Delen (2018) engaged in an investigation of the evolution of important domains in the field of health informatics. The Term-by-Document Matrix (TDM) supported by clustering analysis with interpretation was used in the study to find associations between words and documents. Six clusters of research domains were produced; biomedical, algorithmic, and statistical approaches, adoption of HIT, Internet-enabled research, and knowledge representation. This study allows for a holistic view of the advancements in the area of medical informatics during the 2002 to 2013 span for academic fields' maturity [5]. Minaee et al. (2021) evaluated the research trend for deep learning-based text categorization. The authors used text classification, which included sentiment analysis and news articles classification as categories. They describe over 150 deep learning models, ranging from LSTM to reinforcement learning, discussing their technical contributions, similiarities, and advantages. The authors also looked at datasets that they have cited the most often, along with deep learning models performance in each text classification category. Minaee et al. (2021) gave a general idea of the methodology and academics offered, but in the end, it is hard to get a feel for what the research is up with [6]. Kim et al. (2022) examined trends of research in domestic economics education using. In order to better appreciate the current works and find existing gaps of knowledge, this section goes over the existing academic literature relevant to sentiment analysis tasks, applications, and deep learning approaches. There were three survey papers we were able to find that tried to cover the same issues that we are. The first paper [7] discusses many deep learning architectures that can be applied to sentiment analysis. It underscores the use of deep learning approaches to sentiment analysis and forecasts future developments in deep learning to sentiment analysis. Deep learning architectures discussed by the paper were CNN, RNN, LSTM, attention mechanism RNN, memory network (Mem NN), and recursive neural network (Rec NN). Sentiment analysis tasks discussed in the survey paper were document-level sentiment classification, sentence-level sentiment classification, aspect-level sentiment classification, aspect extraction and categorization, opinion expression extraction, sentiment composition, opinion holder extraction, temporal opinion mining, sentiment analysis with word embedding, sarcasm analysis, emotion analysis, and multimodal data for sentiment analysis. Even though the paper discussed deep learning, it did not provide any experimental analysis that shows one deep learning architecture was better than all others, nor did it provide relevant examples of field applications. Next, paper [8]

3. Data

The sources of data gathering are Google Scholar and Scopus. Google Scholar is a scholarly search engine offered by Google. You can locate papers, publications, and academic journals with the aid of this tool. Additionally, an advantage of Google Scholar is that you can find early research publications on sentiment analysis that do not match those found on Scopus [9]. Scopus was developed as an academic database in 2004 by Elsevier. This database provides a thorough source of scholarly citations [10]. For this reason, we think it is the best place to search scholarly publications. We retained the titles of the publications found in Google Scholar and Scopus using "Sentiment Analysis." A consistent method was applied to extract relevant data from the papers that were reviewed to answer our research questions. These data included: • The sentiment analysis tasks being addressed • The real-world application focus • The deep learning algorithms or models used • The gaps and challenges described • The datasets and size of data • The language • The performance measures (e.g., F1 score, accuracy) The next sections provide a summary of the key findings related to each research question: Fig. 3 A simplified process for paper extraction 2.4.1 Major sentiment analysis tasks We looked at the literature review and identified many sentiment analysis tasks. These tasks include document-level sentiment analysis, sentence-level sentiment analysis, aspect-based sentiment analysis, aspect extraction, emotion detection, multi-domain sentiment classification, multilingual sentiment analysis, multimodal sentiment analysis, opinion summarization, opinion spam detection, opinion holder extraction and classification, time extraction and standardization, visual sentiment analysis, and graded sentiment analysis. The above-mentioned sentiment analysis tasks have appeared in several papers, where the majority of the research has focused on general sentiment analysis, and now research is being conducted extensively on the first six tasks mentioned [11-12, 13, 14]. From the literature review, we discovered that more sentiment analysis tasks could potentially be looked at in the near future, with no research currently on any of these tasks [11–12, 13, 14]. These are;

11. Research Model

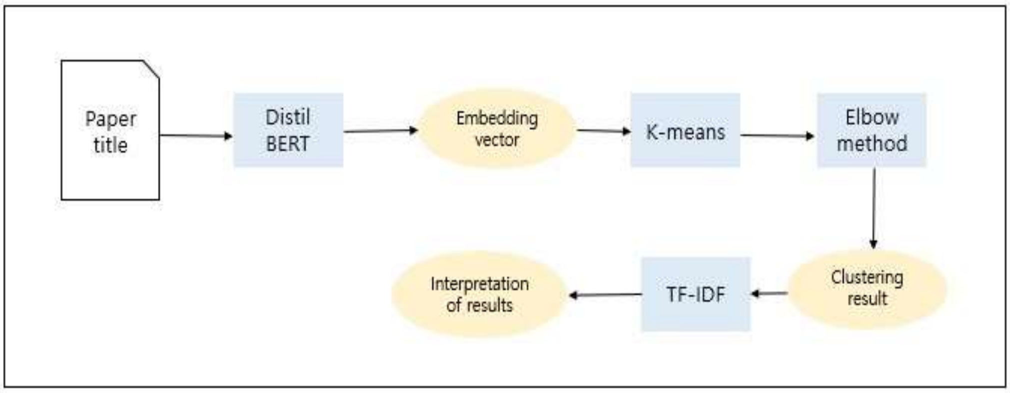

Figure 1- Search Model

The research model converts the gathered data to a vector within DistilBERT, determines clusters based on the elbow method, and clusters embedding vectors with K-means based on optimal clusters (see Figure 1). Distil BERT is a model proposed by HuggingFace in 2019. By using the distilling knowledge approach to the standard BERT model, the size and speed decreased, but performance is similar to the standard BERT. In the standard NLP area, researchers tended to increase model size in terms of their pre-trained models. While a large, pre-trained model benefits from better performance, the large model drawbacks are that these types of models are virtually impossible to use on small devices in real-time. To avoid these drawbacks, Distil BERT uses knowledge distilling technology [10]. Knowledge distilling is a compression mechanism in which a small model (the student) is trained to perform the same way as a large model (the teacher) [16, 17]. When applied to a BERT model, knowledge is transferred to a small BERT (the student) from the pretrained BERT-base (the teacher) model. While the performance was compared to the standard BERT, it was on the GLUE benchmark and had a comparable performance of around 81.0%.

4. Related Work

Existing work on sentiment analysis and mining mainly emphasizes three issues: text word embedding, text feature extraction, and model performance. First, Text Word Embedding. The processing of natural language with machine learning algorithms requires the mathematization of language, and word embedding is a way to mathematicize words of the language. Word2Vec is the word embedding model used in the earliest studies of Network opinion, Wei et al. [20] used the Word2Vec model to achieve semi-automatic construction of product reviews, and combined on this basis rule parsing and domain ontology to construct the feature-level sentiment analysis framework for product reviews. Li et al. [21] constructed a sentiment analysis model for Network restaurant reviews by combining Word2vec, Bi-GRU, and Attention techniques to solve the existing situation that consumers have difficulty in quickly finding information from the services of merchants. He et al. [22] deep learning-based analysis framework of university Network public opinion for the problem of university Network public opinion in real time, and MAP Word2Vec is used to embed words to text and LSTM-CFR method is used to classify Chinese words. The word vectors generated by Word2Vec have static characteristics and are not well suited to deal with the situation of multiple words, so to meet these situations that have been developed dynamic word vector models. In the past few years, the BERT model is the developed most frequently as a research method of Network opinion sentiment analysis, and Li et al. [23] for the situation that the BERT model cannot provide contextual information, a method called GBCN is proposed, which sends the text to the BERT and context-aware embedding layer, and then used a gating mechanism to control the sentiment characteristics of the BERT output. Liu et al. [24] proposed a Bert-BiGRU-Softmax deep learning model that effectively construct the input layer with affective Bert model to `more directly' understand consumers' emotional reaction as they evaluate goods.

Li et al. [25] used BERT to develop a sentiment analysis model based on Chinese stock market reviews to address the problem of lower than acceptable accuracy of stock market sentiment analysis, and used the pre, Concerning feature extraction from text. Most of the current feature extraction models used in Network opinion sentiment analysis make use of either single or multi-models in serial, which limits feature information extraction. For example, Ren et al. [26] constructed a lexicon-enhanced attention network (LEAN) based on bidirectional LSTM for the relationship between words and sentences. LEAN not only considers words that are characterized as sentiment words in sentences, it also attends to salient information that is in sentences. Wei et al. [27] developed a BiLSTM framework based on polled orthogonal attention structures to address the issue of a lack of differentiable explicit emotion words, and employed the solution for integration implicit emotion analysis to enable a corresponding word and emotion trends to be differentiated. Li et al. [28] also examined a bi-directional LSTM framework and self-attentive mechanism and multi-channel features as part of their investigation to enable the underlying relationships between polar components of target words to be explored, without the paradigm of establishing an artificial sentiment dictionary. Huang et al. [29] also put forward a lexicon-based contextual convolutional neural network for effective exploration Features and strength of sentiment based on words and across texts.

Gu et al. [30] provided insight concerning the situations faced within the current sentiment analysis discourse landscape that also warrants a new MBGCV method, based on Multi-granularity Sentiment features. This paper introduces a new MBGCV method that incorporates multi-granularity sentiment features based on the advantages of both the BiGRU model and the CNN model, to cope with the challenges that exist in current sentiment analysis research. Zhao et al. [31] used an aspect-dependent sentiment analysis model to understand different sentiment polarities in comments by combining convolutional neural networks with gated recurrent units in tandem, using the local feature extraction ability of the CNN and combine it with the long-term dependent learning ability of the GRU. Jain et al. [32] proposed a sentiment analysis method that sentential convolutional neural networks and long and short-term memory models by combining dropout, maximum pooling, and batch normalization to obtain the corresponding results. Lin et al. [33] proposed a sentiment analysis algorithm named FAST-BiLSTM to optimize sentiment inference, it used the FastText model to obtain word vectors, then trained th word vectors with a bi-directional long and short-term memory network (Bi-LSTM), then fusion it with FastText for comprehensive sentiment analysis. This paper utilizes a TextCNN and BiGRU parallel structure models as text feature extractor to address the weakness of the above studies to obtain deep text features under several dimensions, for better understand people's sentiment tendency on performance of the model.

There have been various studies aimed at increasing the performance of models making use of transfer learning or attention mechanisms. On one hand, relevant studies using attention mechanisms, Lv et al.[34] presented a contextual and aspectual memory network (CAMN) based on deep memory networks, bidirectional long- and short-term memory networks, along with different attention mechanisms to facilitate improving aspect-level sentiment analysis by more accurately describing the sentiment characteristics of short texts; Sweidan et al.[35] introduced a hybrid ontology-XLNet for sentiment analysis classification on a sentence level, where the XLNet network was used to extract neighboring context in the text and connect it to generate contextually clearer information for improved accuracy in feature extraction. Zhang et al.[36] developed a sentiment analysis model with BiTCN-Attention in order to better concentrate on sentiment words, which various attention mechanisms are presented in BiTCN to produce BiTCN-SA and BiTCN-MHSA to improve the weight of sentiment words, better improve feature extraction accuracy, and to enhance the impact of sentiment analysis. Yang[37] combined BERT, CNN, and BiLSTM to form the sentiment classification model BCBL; secondly these authors to address the issue that BCBL does not take into consideration the established a sentiment analysis model based on BiTCN-Attention, which gives greater focus to sentiment words, and developed BiTCN-SA and BiTCN-MHSA by incorporating different attention mechanisms in BiTCN, in order to increase the weight of sentiment words and the accuracy of feature extraction and improve the results of the sentiment analysis.

The above studies achieved improvements to the performance of the opinion analysis model using the attention mechanism but were also found to not utilize transfer learning; On the other hand, there are relevant studies, specifically using transfer learning, like Tao et al. [38] proposed a sentiment analysis method based on the ABSA model by integrating the ABSA model and transfer learning to take advantages of opportunities to compensate for the weaknesses of the ABSA sentiment analysis method. Sanagar et al. [39] studied an unsupervised sentiment dictionary that can transfer established seed vocabulary from category-level knowledge to the target domain and can be extended also to new domains in the same category. Cao et al. [40] . proposed a deep transfer learning mechanism (DTLM) for fine-grained cross-domain sentiment classification that transfers sentiment across domains more effectively, by incorporating BERT and KL divergence. Chandrasekaran et al.[41] were responsible to using different migratory learning models based on VGG-19, ResNet50V2, and DenseNet-121 models for image-based sentiment analysis, and they fine-tuned the model by freezing some layers while unfreezing others.

While the above studies did not have practically-focussed research on the integration of transfer learning and attention mechanisms, they did not adequately address the roles of data and different model synthesis factors. In addition to our extensive research on SOTA (state-of-the-art) and comparison with our approach literature, our approach literature [42] used the BERT model along with CNN and BiLSTM models to make an end-to-end model to attain state-of-the-art classification performance. Furthermore, literature [43] also used the BERT model to combine a BiLSTM model and attention mechanism. This model was used for two very different applications, pre-training and multi-task. Alternatively, this paper has used the ERNIE (an improved extension of BERT) model, which can improve pre-training efficiency, semantic representation, and semantic understanding. Whilst, this paper has used the joint model of TextCNN and Text BiGRU, where TextCNN is more specific than CNN for sentiment analysis.

5. Description of the model framework

Model framework overview To address the research question of the lack of labeled data in Network opinion sentiment analysis, this paper proposes a multi-modal fusion of transfer learning-based approach to Network opinion sentiment analysis (Transfer learning-ERNIE-TextCNNBiGRU-Attention, TETBA), as illustrate in Fig. 1. The proposed model leverages the capabilities of deep learning and transfer learning. The ERNIE model is used to generate the word embedding representation of the text of the dataset for model input and the jointly to use TextCNN and BiGRU model to extract the local features during input/text word vectorization and overall features with pre-training; using the established model for opinion sentiment analysis, we utilize annotated news analysis data (source domain data) with much rich annotated information model to train the opinion sentiment analysis model; all feature extraction model parameters from the trained opinion sentiment analysis model are transfer to the unrelated target domain task of opinion sentiment analysis and to promote textual sentiment extremes with attention; finally we fuse the local features with the overall features to conduct opinion sentiment analysis in the environment with little available labelling, or annotated data.

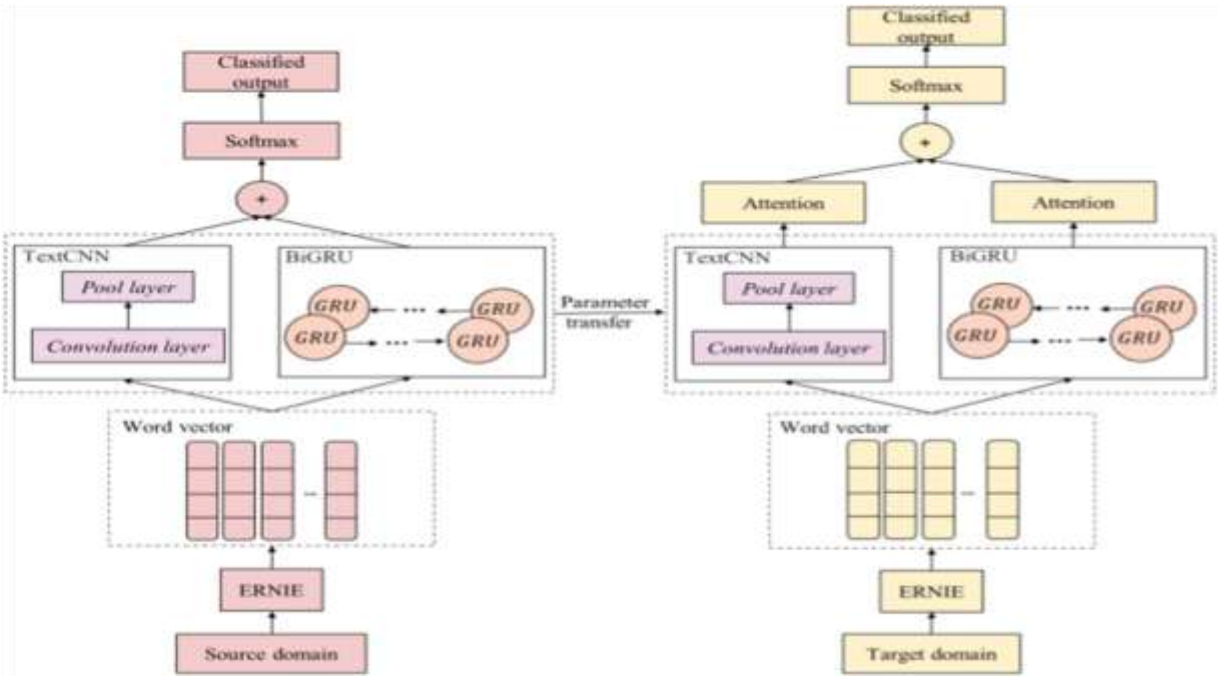

Fig. 2 TETBA model

The model consists of the following components:

1. Word vector representation of public opinion data: the ERNIE model is applied to realize the transformation of the text data and obtain the relevant word vectors to input into the model of public sentiment analytics.

2. Joint model feature extraction: the joint model of TextCNN and BiGRU to reach completeness in feature extraction; the local features are extracted using TextCNN, and the overall features are extracted using the BiGRU model and then merged to form the local features and overall features.

3. Parameter transferring: the parameters of the feature extraction layer are to be transferred using the transferring method from the pre-trained model to the target model, to serve as the initial parameters of the feature extraction layer of the target model; the target model is subsequently trained using the network opinion data.

4. Attention and classification predictions output: weight the matrix of the feature extraction layer prediction output, to draw out the weights of emotional keywords, then fuse the feature outputs of differing dimensions to reach classification of the Network opinion data; with the use of the Softmax function the emotional analysis of Network opinion is complete.

6. Word vector representation of public opinion data

The task of Network opinion sentiment analysis requires the language to be mathematical. Word vectors are a good way to realize the mapping of input text to vector form. The ERNIE model is employed in the paper to map text data to word vector.

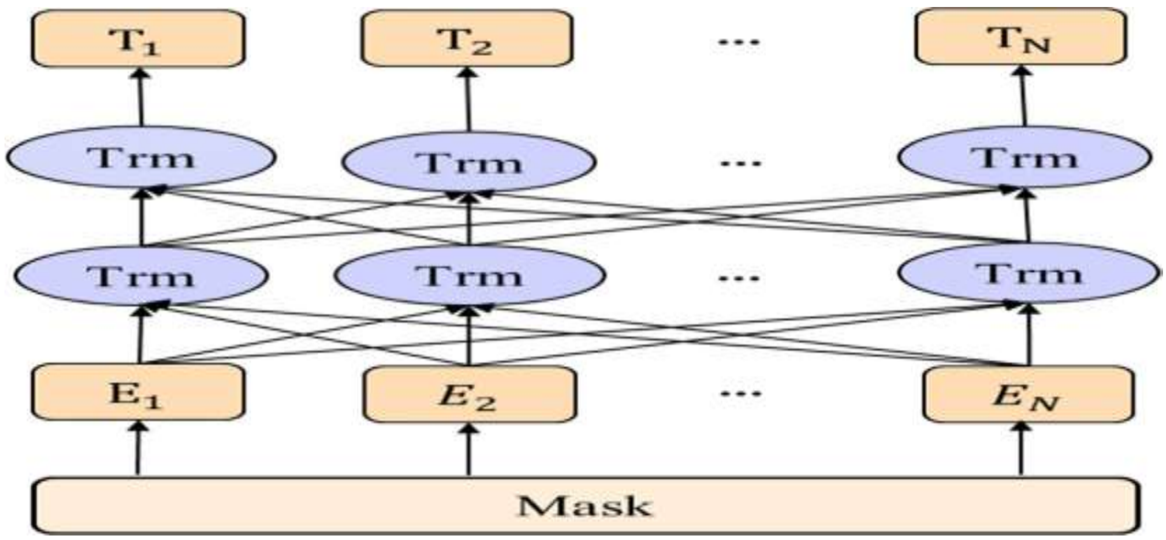

Fig. 4 ERNIE model framework diagram.

The best feature of ERNIE is it fuses a great deal of information sources, while at the same time, continually acquiring new knowledge, including vocabulary, structure, and semantics, within a great deal of abstract text data. The framework of the ERNIE model is visualized in Fig. 2, including knowledge integration and Transformer coding. Knowledge integration is coupling semantic knowledge of entities and phrases into linguistic expressions, and Transformer can extract contextual information of each sentence in a sentence by self-attentive mechanism to derive a series of contextual embedding details as follows

7. Knowledge integration.

The ERNIE model is an enhanced form of the BERT model and BERT is improved primarily in terms of the mask (masking) strategy. The mask strategy is to let the machine know the masked piece of input languages by masking it, and to achieve the training purpose in this way. On the contrary to the traditional BERT model which deploys only as a simply numeric word mask strategy, on this basis the ERNIE model substitutes a Chinese word mask for a consecutive entity word and phrase mask, thus greatly improving the extent to which it understands the semantic content being held within the Chinese text, the masking strategy of ERNIE model is shown in Fig. 3.

8. Dynamic representation of word vector

After using the masking method, there is the Transformer encoding part of the ERNIE model. There are twelve layers of Transformer encoder in ERNIE model, and while ERNIE model is, indeed based on technical enhancement of BERT, it builds a two-way self-attentive mechanism layer .

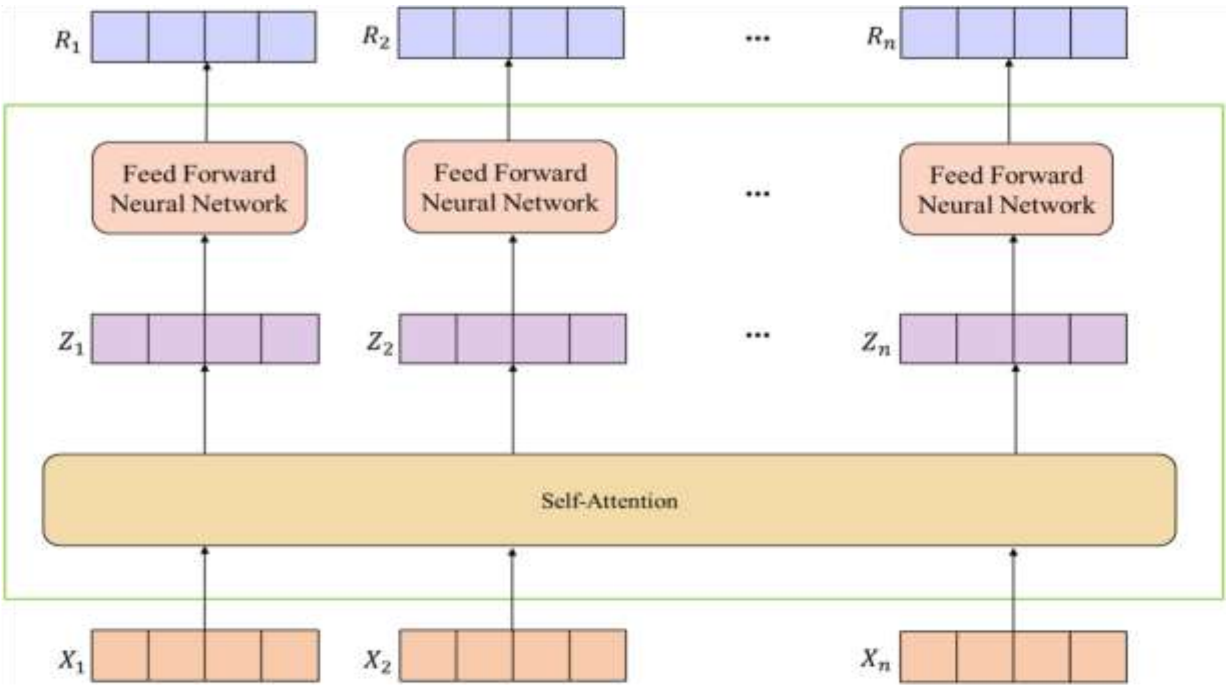

Fig.5 two-way self-attentive mechanism layer

encoder structure in the inner network structure, the encoding device consists of two layers with encoders. The encoder structure of the ERNIE model is shown in fig.4,

here Xi(i = 1, 2, ..., n) is the vector form of the text,Zi(i = 1, 2, ..., n) is the text obtained by the self-attentive mechanism from the text Xi(i = 1, 2, ..., n), and Ri(1, 2, ..., n) is the output of ERNIE model and input to the feature extraction stage model. Fig.4 Encoder structure The specific calculation process of the way that the ERNIE model calcultes by handling public opinion data is as follows: For a Network opinion text data, to which has a corresponding piece of text sequence X after masking training in the knowledge integration stage: X = [X1, X2, ..., Xn] (1) to obtain the text vector corresponding to each input input using the three weight matrices WQ, WK and WV multiplied by the query matrix Q, the key matrix K, and value matrix V based on that. Q = X · WQ (2) K = X · WK (3)

9. Joint Model Feature Extraction

In this experiment, we used the joint model of TextCNN and BiGRU to extract features from the word vectors produced by ERNIE, where the TextCNN model can extract local features within the text, while the BiGRU gets feature information for the text as a whole as illustrated in Fig. 5.

10. Text CNN

As a convolutional neural network, Text CNN can extract local features from a word vector matrix made up of multiple words and can automatically combine and filter multiple word vectors for multi-level semantic representation. The depth that comes from using multiple convolutional kernels allows the TextCNN model to focus on important feature pools and capture local relevance. Compared to CNN models, TextCNN models have a simpler network structure, less computation, and faster training. textCNN models extract important information in sentences at the same time, without regard to ordering by using multiple convolutional kernels of different sizes, so that the model can better understand local information

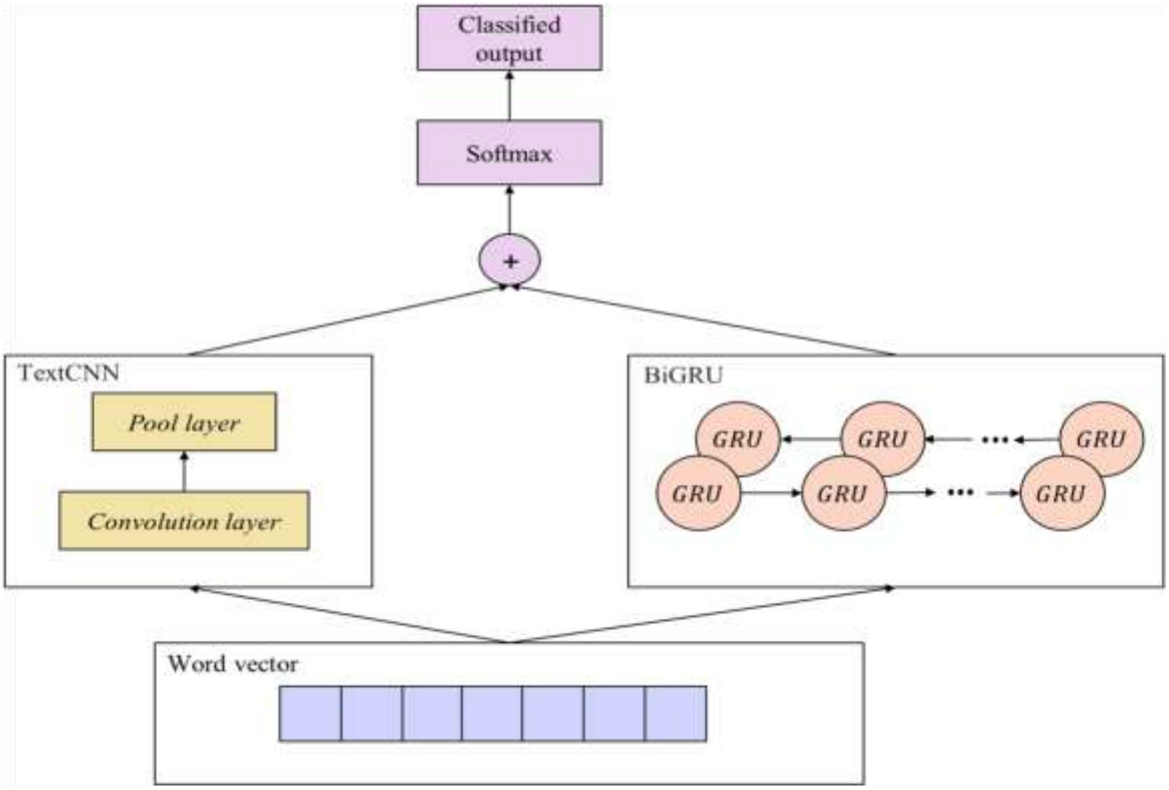

Fig.6 Text CNN

capture local relevance and thus increase the chances of engaging with the semantic meaning of the sentence more accurately. The model format of Text CNN employed in this paper is shown in Fig. 6: Fig 6 shows the feature vectors derived from the convolutional pooling operation, where the convolutional kernels are 2, 3, and 4 by size, and the feature vectors are composed of word vectors. A complex sequence of operations must be carried out by the Text CNN structure in order to carry out this local-level feature extraction, and a complex series of operations is needed to maintain the accuracy and reliability of the approach, and these are shown below.

(1) Word vector input.

For network public opinion sentiment analysis, the word vector matrix is used to generate model input data. In this chapter, the output from the ERNIE model is input into the Text CNN with a word vector dimension of 768. The input data obtained will be included in the TextCNN model as pre-trained word vectors, and different corpora can be used to acquire other based knowledge. During the training, the information will encapsulate things closely related to the textual content features of network opinion, where the sequence length of the input text is n. R = [R1, R2, ..., Rn] (7)

(2) Local Feature Extraction

Text CNN ingests the text data and uses a one-dimensional convolution method to convert it into one-dimensional data of TextCNN model vectors. The convolution conforms to the width and dimension of the word vector and can be selfdetermined in height. When the length of the input sentence has a certain length, a downward sliding trend will occur and then continue to slide down, which needs further analytical processing of the current output to achieve more accurate and complete information. For downward sliding, different windows are defined to extract different feature vectors. In this paper, the convolution kernels in the convolutional layer had three different sizes, and the width of the convolution kernels is the same as the dimension of the word vector is d , and the height is 2, 3, and 4 respectively. Thus there are several number of convolution kernels for each size. For a web opinion text of length n, the convolutional layer learns the properties of the text using a h sliding window of varying size to convolve the text input vector and obtain the convolutional feature values of each web opinion text at the location of the convolutional kernels. The notation for the feature computation can be expressed as : Ci = f (WRi:i+h−1 + b) (8) where W is the convolution kernel, which is a n×d dimensional weight matrix, b is the bias, and R.

(3) Max pooling

Pooling operations are typically made up of two components, either average pooling or max pooling. In this case, the TextCNN model uses the max pooling method which is able to keep the most representative features from multiple sliding windows and combine them to create a full vector representation. When analyzing Network opinion sentiment, each text has some worthless information; with maxpooling, it is able to keep the important keywords which may assist the model in finding the corresponding categories easier. In this paper, the maxpooling method is used and the full feature vector is replaced with the maximum value in the previous feature vector C and then maxpooling will also be used on the other same convolution kernels to convolve the obtained feature vector and will from a new feature vector S. Si = MAX{C} = MAX(C1, C2, ..., Cn−h+1) (10) S = [S1, S2, ..., Sm] (11) BiGRU A traditional neural network connects and transfers information uni-directionally and while this form is useful for learning tasks, it limits how the neural network models. In many real-world tasks, the outputs of the neural network are going to be equal to the input of the current moment plus the past outputs.

The GRU network's parameter update is given by: 1)

First, we will calculate the update gate zt where zt is the update gate state,σ is the sigmoid activation function, and Wz is the weight parameter for the update gate.

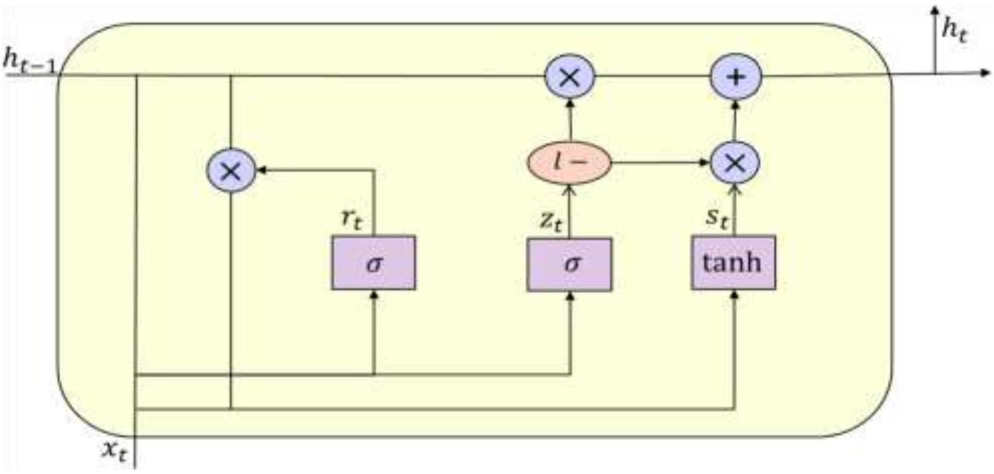

Fig.7 GRU model structure

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

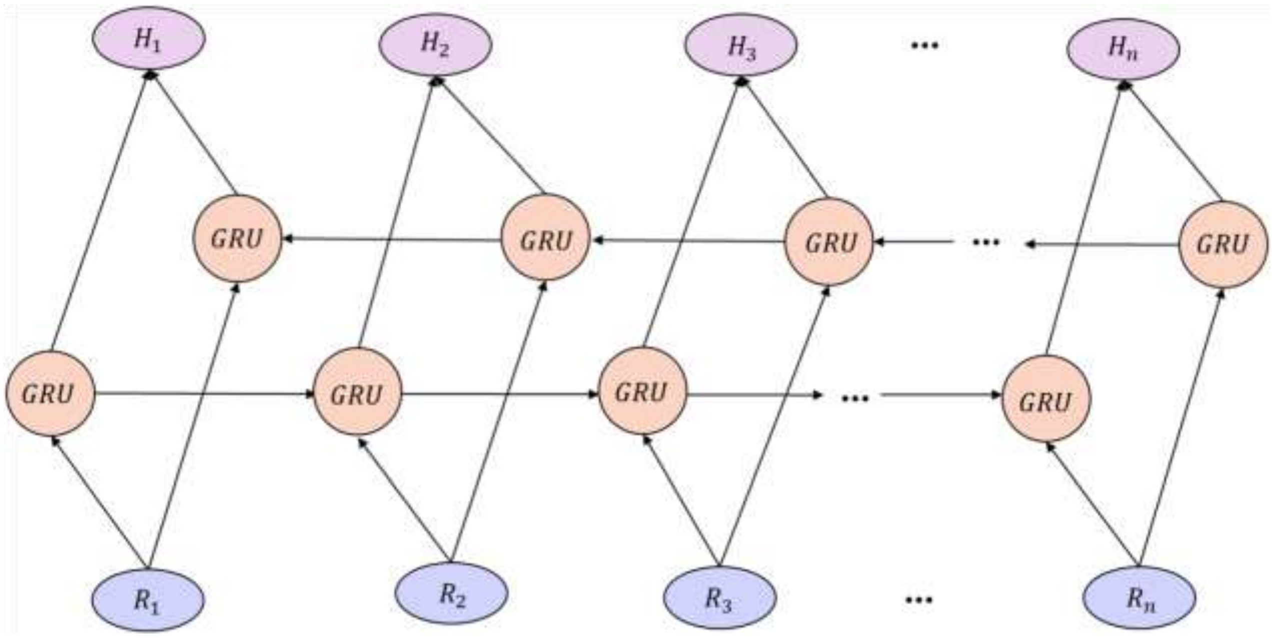

Fig. 8 BiGRU structure diagram

Parameter transfer

Transfer learning is a machine learning method that helps improve the target domain by transferring knowledge from some source domain to the target domain, The important part is to find some similarities of both domains and efficiently apply these similarities. There are four types of transfer learning depending on the difference of what should transfer: instance-based, feature-based, model-based (parameter-based), and relationship-based. In this paper, we will address parameter-based transfer learning, which can find parameters that could be shared between source and target models by creating a parameter-sharing model and effectively migrating and transforming these parameters, fit within because neural networks are structures that can initially transfer parameters. Parameter transfer is widely used in neural networks due to the neuronal structure that would directly transfer the parameters. In this paper, we divide it mostly into two parts: the source model and the target model, and we used the Text CNN and BiGRU dual-channel model to extract both features, specifically where the Text CNN model extracts the local features of text and the BiGRU model extracts the entire text

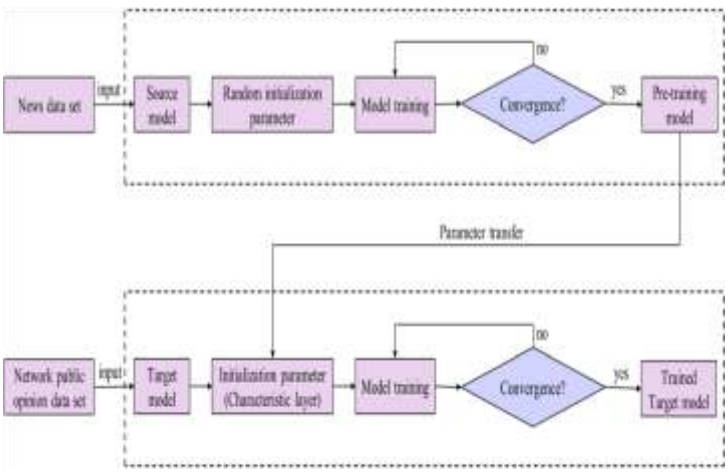

Fig. 9 Transfer learning flow chart

feature information. In the first stage, the label-rich news data are used to get the feature parameters using the Text CNN and BiGRU dual-channel models. The pre-training model can get the common language of the text, then the parameters of the feature extraction layer of the pre-training model are migrated to the target extraction layer of the target model.

These parameters will act as the initial parameters of the feature extraction layer of the target task and are fine-tuned with the public opinion data. The flow diagram of the transfer learning process is shown in Fig. 9.

To undertake the application of multiple pre-trained models to sentiment analysis of Network public opinion via transfer learning, follow these steps:

(1) Development of neural networks using the ERNIE model, and the joint model of Text

CNN and BiGRU, i.e. source models that have been pre-trained on a large news dataset, i.e. pre-trained models;

(2) Construct

a target model that has the same architecture as the pre-trained model, except the classification output layer, then add the attention mechanism in layer three of the network and classification layer in layer four, with an output size reflecting the number of categories in the web opinion dataset;

(3) Move parameters of the feature layer in the pre-trained model, to the feature layer at the same level in the target model, to be used as the original parameters of the feature layer in the target model, and both ERNIE parameters or parameters of the pre-trained model remain unchanged, and afterwards classification parameters of the target model will be initialized with a random generator to yield the best performance;

(4) Remove the labels on the pre-trained model and act on the acting labelled data in the target domain as training data and optimize the model.

Comparison of word embedding models.

To evaluate the performance of the ERNIE model selected in this study in terms of word embedding, an experimental comparative study between to-word embedding models, Word2Vec and BERT was undertaken with the ERNIE model. In the analysis, the word embedding ERNIE model in TETBA proposed in this study was substituted with the Word2Vec model and the BERT model respectively for experimentation purposes, the experimental results are shown in Table 5. From the experimental results presented in Table 5 is clear to see that the four evaluation indices using the ERNIE model as the word embedding model performed best and that the four evaluation indices using the WordVec model as the word embedding model performed worst. The accuracy of the ERNIE model was improved by 3.9% and 2.1% respectively compared to that of the Word2Vec model and the BERT model respectively. The main reasons were: first, the Word2Vec model represented transformed the text to a static word vector.

12. Analysis Result

The output from clustering with 7 clusters was transformed into a data frame that included titles for the papers and then, to visually assess if the clusters were well defined, they were reduced to the two-dimensional using t-distributed stochastic neighbor embedding (t-SNE) and then visualized which is an algorithm for reducing the dimensionality of high-dimensional data to two dimensions. t-SNE is one of several manifold learning techniques used to visualize complex data. After the cluster results were realized and reduced down with t-SNE, it was concluded that clusters 0, 2, 3, 4, and 5 stood out from the seven clusters. Because of the abundance of papers it was hard to conceptualize the themes of the papers in the cluster by looking at only the cluster results. So, we looked at the results of clustering and the results of TF-IDF for the frequently occurring keywords used for each cluster. Checking the keywords indicated that there was a preponderance of prepositions which made it difficult to assess the actual topic of the paper. So, instead, nltk was used to extract the noun-only segments. Excluding sentiment and analysis which occurred often since they were keywords that were evaluated, words like, multimodal, twitter, tweets and multi were in abundance in the search results. Checking the cluster results based on these words showed that.

13. Conclusions

This research collected a selection of paper titles from Google Scholar and Scopus. Using a combination of DistilBERT, K-means clustering, and TF-IDF we sought to provide other researchers who may be interested in this area of research in sentiment analysis with our data. The analysis revealed that, sentiment analysis studies were grouped in seven clusters and that those demonstrate the important research trends of studies based or are related to sentiment analysis. The following are some highlights and conclusions developed during this research. First, not only research topics such as stocks price prediction, but also twitter and other reviews of products seem to be occupying an important place in current sentiment analysis research. A search for new sentiment analysis target texts might be a useful research direction. It was also found that sentiment analysis research seems to be going from being limited to natural language processing research to multi-modal and multi-feature research that includes voice, video, image and has even more weight. Researchers categories see images and natural language not as different universes but rather as interconnected and the investigation of their interdependence. Based on this observation of change, sentiment analysis research should utilize natural language, image, voice, and video together. First, it seems that not only the predictable topics, such as stock price and Twitter or product review analysis, were important topics in sentiment analysis research. Researching new types of sentiment analysis target texts may be another good research line. Furthermore, noting that sentiment analysis was previously a natural language processing branch of research, see since the previous decade or so it is acquiring a research branch total research weight with various modalities and features like voice, video, and image. Rather than treating voice and image as separate sets of study, researchers have been studying their interplay due to their interrelationship. The trajectory of people's emotional place following an event or changing process (e.g., COVID-19) may also be a significant area of research. When people engage in specific events or changes, their thinking and emotional state become clearer. We should watch the research emerging from these states/reactions closely. In the current trend analysis in sentiment analysis research, various studies were evaluated based on methodology. In addition,

References

[1]. Jung, H., & Lee, B. G. (2020). Research trends in text mining: Semantic network and main path analysis of selected journals. Expert Systems with Applications, 162, 113851.

[2]. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

[3]. Zhang, W., Xu, M., Jiang, Q.: Opinion mining and sentiment analysis in social media: Challenges and applications. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 536–548. Springer, Berlin (2018) .

[4]. Taboada, M., Brooke, J., M., Voll, K., Stede, M.: Lexicon-Based Methods for Sentiment Analysis (2011). http:// direct.mit.edu/coli/article-pdf/37/2/267/1798865/coli_a_00049. pdf

[5]. Kim, Y. M., & Delen, D. (2018). Medical informatics research trend analysis: A text mining approach. Health informatics journal, 24(4), 432-452.

[6]. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning--based text classification: A comprehensive review. ACM Computing Surveys (CSUR), 54(3), 1-40.

[7]. Adriaanse, L. S., & Rensleigh, C. (2013). Web of Science, Scopus, and Google Scholar: A content comprehensiveness comparison. The Electronic Library.

[8]. Ravi, K., Ravi, V.: A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Know l. Based Syst. **89**, 14–46 (2015). https://doi.org/10.1016/j.knosys.2015.06.015

[9]. Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top-cited papers. Computer Science Review, 27, 16-32.

[10]. Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. Journal of Informetrics, 1(2), 161-169.

[11]. Zhang, L., Wang, S., Liu, B.: Deep learning for sentiment analysis: a survey. WIREs Data Min. Knowl. Discov. (2018). https://doi. org/10.1002/widm.1253

[12]. Liu, B., Zhang, L.: A survey of opinion mining and sentiment analysis. In: Mining Text Data. Springer US, Boston, pp. 415–463 (2012). https://doi.org/10.1007/978-1-4614-3223-4_13

[13]. Prabha, M.I., Umarani Srikanth, G.: Survey of sentiment analysis using deep learning techniques. In: 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), pp. 1–9 (2019). https://doi.org/10.1109/IC IICT1.2019.8741438.

[14] Yadav, A., Vishwakarma, D.K.: Sentiment analysis using deep learning architectures: a review. Artif. Intell. Rev.. Intell. Rev. **53**(6),4335–4385(2020).https://doi.org/10.1007/s10462-019-09 794-5.

[15] Buciluǎ, C., Caruana, R., & Niculescu-Mizil, A. (2006, August). Model compression. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 535-541).

[16] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, *2*(7).

[18] Liao W, Zeng B, Yin X, et al. An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. Appl Intell. 2021;51:3522–33.

[19] Naz H, Ahuja S, Kumar D. DT-FNN based effective hybrid classification scheme for twitter sentiment analysis. Multimed Tools Appl. 2021;80:11443–58.

[20] Wei W, Liu YP, Wei LR. Feature-level sentiment analysis based on rules and fine-grained domain ontology. KO Knowl Org. 2020;47(2):105–21.

[21] Li L, Yang L, Zeng Y. Improving sentiment classification of restaurant reviews with attention-based bi-GRU neural network. Symmetry. 2021;13(8):1517.

[22] He M, Ma C, Wang R. A data-driven approach for university public opinion analysis and its applications. Appl Sci. 2022;12(18):9136.

[23] Li X, Fu X, Xu G, et al. Enhancing BERT representation with context-aware embedding for aspect-based sentiment analysis. IEEE Access. 2020;8:46868–76.

[24] Liu Y, Lu J, Yang J, et al. Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRUSoftmax. Math Biosci Eng. 2020;17(6):7819–37.

[25] Li M, Chen L, Zhao J, et al. Sentiment analysis of Chinese stock reviews based on BERT model. Appl Intell. 2021;51:5016–24.

[26] Ren Z, Zeng G, Chen L, et al. A lexicon-enhanced attention network for aspect-level sentiment analysis. IEEE Access. 2020;8:93464–71.

[27] Wei J, Liao J, Yang Z, et al. BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis. Neurocomputing. 2020;383:165–73.

[28] Li W, Qi F, Tang M, et al. Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. Neurocomputing. 2020;387:63–77.

[29] Huang M, Xie H, Rao Y, et al. Sentiment strength detection with a context-dependent lexicon-based convolutional neural network. Inf Sci. 2020;520:389–99.

[30] Gu T, Xu G, Luo J. Sentiment analysis via deep multichannel neural networks with variational information bottleneck. IEEE Access. 2020;8:121014–21.

[31] Zhao N, Gao H, Wen X, et al. Combination of convolutional neural network and gated recurrent unit for aspectbased sentiment analysis. IEEE Access. 2021;9:15561–9.

[32] Jain PK, Saravanan V, Pamula R. A hybrid CNN-LSTM: a deep learning approach for consumer sentiment analysis using qualitative user-generated contents. Trans Asian Low-Resour Lang Inf Proc. 2021;20(5):1–15.

[33] Lin Z, Wang L, Cui X, et al. Fast sentiment analysis algorithm based on double model fusion. Comput Syst Sci Eng. 2021;36(1):175–88.

[34] Lv Y, Wei F, Cao L, et al. Aspect-level sentiment analysis using context and aspect memory network. Neurocomputing. 2021;428:195–205.

[35] Sweidan AH, El-Bendary N, Al-Feel H. Sentence-level aspect-based sentiment analysis for classifying adverse drug reactions (ADRs) using hybrid ontology-XLNet transfer learning. IEEE Access. 2021;9:90828–46.

[36] Zhang J, Zhang T, Chen J. Sentiment analysis of chinese reviews based on BiTCN-attention model. Int J Found Comput Sci. 2022;33(06n07):755–70.

[37] Yang H. Network public opinion risk prediction and judgment based on deep learning: a model of text sentiment analysis. Comput Intell Neurosci. 2022;2022:1.

[38] Tao J, Fang X. Toward multi-label sentiment analysis: a transfer learning based approach. J Big Data. 2020;7:1–26.

[39] Sanagar S, Gupta D. Unsupervised genre-based multidomain sentiment lexicon learning using corpus-generated polarity seed words. IEEE Access. 2020;8:118050–71.

[40] Cao Z, Zhou Y, Yang A, et al. Deep transfer learning mechanism for fine-grained cross-domain sentiment classification. Connect Sci. 2021;33(4):911–28.

[41] Chandrasekaran G, Antoanela N, Andrei G, et al. Visual sentiment analysis using deep learning models with social media data. Appl Sci. 2022;12(3):1030.

[42] Prottasha NJ, Sami AA, Kowsher M, et al. Transfer learning for sentiment analysis using BERT-based supervised fine-tuning. Sensors. 2022;22(11):4157.

[43] Xu G, Zhang Z, Zhang T, et al. Aspect-level sentiment classification based on attention-BiLSTM model and transfer learning. Knowl-Based Syst. 2022;245: 108586.