Journal of Information Systems Engineering and Management
2025, 10(12s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Learnable Conjunction with Adjective Enhanced Model for Mandarin Sentiment Analysis of Social Media Text

Zhang Jie [ab], Ruhaila Maskat [a,1] , Xu Zhaosheng [ab], Zhang Zhiping [ab], Li Shuliang [b]
[a] *College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam, Malaysia*
[b] *College of mathematics and computer, Xinyu University, Jiangxi, China*
Corresponding Author: Ruhaila Maskat, Email: ruhaila256@uitm.edu.my, College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA, Shah Alam, Malaysia.

ARTICLE INFO

Received: 18 Nov 2024

Revised: 02 Jan 2025

Accepted: 24 Jan 2025

ABSTRACT

The advent of research into neural networks has led to the widespread utilisation of deep learning methodologies in the domain of text sentiment analysis, owing to their formidable data processing and pattern recognition capabilities. In recent years, Transformer and its variant Bert have attracted considerable attention, and their performance has been demonstrated to be superior in practice. However, it is important to note that China's unique social media data set is characterised by a variety of languages, distinct cultural backgrounds, and intricate emotional expressions, which poses a significant challenge to the generalisability of models trained on more homogeneous data sets. Consequently, the conventional Transformer model's efficacy in feature extraction is constrained in certain scenarios, impeding the efficiency of its implementation. It is necessary to adapt and optimize the model further to account for the specific attributes of social media data in the task design. Nevertheless, within the framework of Chinese linguistics, the significance of word order and conjunction is inherently apparent. Conversely, the potential of adjectives in emotional expression is frequently disregarded. This study endeavours to investigate an innovative approach, underpinned by a Transformer encoder model, to construct a reinforcement model that can integrate and learn adjective features. The model will be evaluated through experimentation with three publicly available Chinese social media datasets. The experimental data demonstrate that the model can utilise the attention mechanism to not only identify the emotional tendency of keywords but also to effectively combine the positional characteristics of conjunctions and adjectives to obtain local details and contextual meanings of the text. This enhancement of the model's effectiveness in feature extraction is a significant contribution to the field.

Keywords: Sentiment Analysis, Social media content, Feature identification, Transformer model, Mandarin language.

1. Introduction

Sentiment analysis, otherwise referred to as opinion mining, is the process of identifying, extracting, and interpreting the emotions, attitudes, evaluations, opinions, and experiences expressed by individuals with regard to services, products, organisations, people, topics, events, and their attributes [1]. Sentiment analysis occupies a fundamental and important position in the field of natural language processing. Indeed, this topic has been the focus of considerable interest and in-depth study by researchers worldwide in recent years. The rapid development of the Internet and mobile networks has resulted in an increasing number of users sharing their personal opinions, consumption experiences, social perspectives, and emotional tendencies through various online platforms. In China, Internet users often express their emotions and opinions on social media platforms like Weibo and WeChat. Customers leave reviews for hotels and restaurants on websites such as Ctrip and Meituan, while shoppers share feedback on products through e-commerce websites like Taobao and Jingdong. It is also worth exploring how to mine important sentiment information from these contents. Emotion analysis can be categorised into two distinct classifications: document-level and aspect-level. The primary objective of document-level sentiment analysis is to evaluate the prevailing emotional tendency articulated by the entirety of a text, such as an article, commentary, or blog content. The assessment of document-level sentiment is typically classified into positive, negative, or neutral categories. When undertaking document-level sentiment analysis, it is imperative to take into account the lexical selection, grammatical characteristics, and the contextual milieu of the text. Aspect-Level Sentiment Analysis delves deeper by identifying specific topics or aspects mentioned within the text and analyzing the sentiment associated with each. For example, in a product review, the sentiment towards "price," "quality," and "customer service" might be assessed separately [2].

In the early stages, sentiment analysis methods primarily fell into two categories: those based on sentiment dictionaries and those relying on machine learning. These approaches required the construction of a sentiment dictionary, followed by manual annotation to assign polarity and intensity, and were then used to classify text sentiments. While effective for sentiment classification, this approach was labor-intensive and inefficient due to the need for manual dictionary creation and annotation. In the 1990s, machine learning methods began to gain traction in text sentiment analysis [3][4]. Despite the rudimentary functionality of these sentiment analysis models, they are heavily reliant on the sophisticated engineering of complex features. The efficacy of feature engineering is a pivotal factor in determining the accuracy of emotion classification outcomes. This process encompasses pivotal procedures such as feature selection, extraction, and subsequent optimisation. Additionally, machine learning models often struggle with generalization, limiting their applicability across diverse datasets. The advent of deep learning technology has precipitated a paradigm shift within the domain of sentiment analysis, effectively circumventing the constraints imposed by conventional sentiment dictionaries and traditional machine learning methodologies. In the field of natural language processing, particularly in the context of sentiment analysis, deep learning has demonstrated remarkable efficacy and generalisability.

CNN and RNN are two primary neural network models frequently employed in text sentiment analysis within the domain of deep learning. The utilisation of CNNs in text sentiment analysis enables their convolutional layer to precisely identify and extract local features and patterns in the input text, as well as the intricate connections between them. This facilitates the automatic acquisition of feature data which is indispensable for sentiment analysis. No need for manual feature engineering, the network learns features directly from data. Ideal for capturing sentiment-laden phrases or n-grams. Efficient for large datasets and can handle varying input lengths with appropriate preprocessing. By leveraging these feature extraction methods, CNN-based approaches excel in identifying local patterns crucial for sentiment classification, especially in structured or short texts. Textual sentiment analysis shares similarities with sequential modeling, as it involves learning contextual information from sequences. RNNs are commonly used for this purpose. In the context of text data, RNNs encounter limitations due to their sequential processing mode and gradient disappearance, which hinders the effective capture of long-distance dependencies. To address this challenge, research in the field of Chinese text sentiment analysis is progressively orienting towards integration of RNN, attention mechanism and Transformer model.The attention mechanism enhances the model's capacity to recognise long-distance dependencies by focusing on distinct regions of the input sequence. Concurrently, the Transformer model employs a self-attention mechanism to circumvent the limitations imposed by sequential processing in RNN, thereby demonstrating remarkable efficacy in the capture of long-distance dependencies. The Transformer model utilises a distinctive self-attention mechanism that enables it to evaluate the significance of each component within an input sequence in a concurrent manner. This capacity to process long sequences is significantly enhanced. The Transformer model's unique mechanism facilitates the capture of long-distance dependencies, thereby enabling it to concurrently assess the relationship between any two elements in a sequence and generate an attention-weight matrix that reflects the proximity between these elements. As a result, Transformers achieve better performance in tasks such as language modeling and translation.

In the context of Mandarin Chinese, word order plays an important role. It is not only the core element of expressing grammatical meaning, constructing sentences and transmitting information, but also the main way to distinguish grammatical structure and semantic differences in cases of relatively limited morphological changes in Chinese [5]. The integration of word order elements into Mandarin Chinese data sets is of significant importance when it comes to enhancing the efficacy of sentiment analysis. Word order remains constant in Chinese, and any adjustments made to it result in semantic alterations. The incorporation of word order can serve to enhance the model's capacity to recognise the logic of sentences, accurately interpret meaning, improve analysis accuracy, reduce misjudgement, and establish a theoretical and practical foundation for the application of sentiment analysis in Chinese. Therefore, effectively capturing and utilizing word order is crucial for accurate sentiment interpretation and overall model performance [6]. "especially good/特别好" refers to food that tastes particularly delicious, "very special/好特别" means that this food is different from others and has distinctive characteristics. Despite the absence of any overt emotional sentiment in these phrases, they do, nevertheless, reflect divergent levels of positive emotion. Research findings indicate that the observed variation in emotional inclination is predominantly attributable to the transformation of parts of speech subsequent to the alteration of word order. In the context of Mandarin Chinese, the arrangement of words has the capacity to modify the parts-of-speech properties of a phrase, thereby exerting an influence on its emotional tone or intrinsic meaning. Essentially, different word orders can shift how a word or phrase is understood, influencing whether the expression has a positive, neutral, or negative emotional impact. For example: "好人" (hǎo rén): This phrase is a noun in Chinese, means "good person" and has a positive connotation, describing someone with good character. "人好" (rén hǎo): This is a phrase with a subject-predicate structure in Chinese, where "ren" is a noun (subject) and "hao" is an adjective (predicate), meaning "this person is good" or "people are good" which is more neutral and simply states a fact about someone's character without the same focus or emphasis. Another example: "吃好" in Chinese is a phrase with a verb-object structure, which contains a verb and an adjective. "吃" is a verb, which means "eating" or "eating", "好" here is an adjective, which means "good" or "satisfied". Therefore, the overall meaning of "吃好" is "eating well" or "eating satisfactorily". This phrase itself is not an independent word, but a verb phrase, which cannot be classified into a certain part of speech. But in terms of composition, it is composed of a verb ("吃") and an adjective ("好"), expressing a state or result and does not have a strong emotional connotation. "好吃" is an adjective in Chinese, used to describe the taste of food, meaning "the food is delicious" or "delicious". It is composed of "好" (adjective) and "吃" (verb), but as a whole, it expresses the evaluation of food and is an adjective and this word clearly expresses a strong positive emotion. From the above analysis, we can find that adjectives frequently serve a crucial function in establishing sentiment polarity (positive, negative, neutral)[7][8][9].

Building on the aforementioned methods and challenges, this paper introduces learnable conjunction with an adjective-enhanced model for Mandarin Chinese sentiment analysis, leveraging a Transformer encoder framework. The integration of conjunction position attributes and adjective weight factors within the attention mechanism of the model facilitates the effective extraction of the global semantic content of the context, while concurrently maintaining a high degree of vigilance to key local information. This combination of attributes results in the model demonstrating excellent performance in the domains of sentiment analysis and semantic interpretation. The model incorporates enhanced modules, such as a weighted enhancement mechanism, to dynamically adjust the contribution of attention, thereby improving the attention module's capacity to handle complex contexts. The residual structure of the pre-trained language model has been enhanced, and a flexible design has been adopted for the purpose of optimising the network and facilitating the flow of information. An evaluation of three Chinese social media data sets has been conducted, and the model has been shown to possess strong text feature extraction and classification capabilities that surpass those of the baseline model, thereby demonstrating its superior performance.

The primary research outcomes of this paper concentrate on the following aspects:

a) It elucidates the fundamental role of word order in the sentiment analysis of Chinese social media texts and employs it as a pivotal feature.

b) A novel adjective-combined enhancement model is formulated for efficient extraction of Chinese affective features and precise prediction of affective categories.

c) Through experimental validation on three publicly available data sets, the proposed method attains substantial performance enhancement in comparison with the two baseline methods.

2. Related works

The objective of text sentiment analysis is to transform unstructured emotional text into a structured format that is readily interpretable and processable by computers. This transformation not only enhances the capacity of computers to discern the emotional content of text but also establishes a robust foundation for subsequent tasks, such as text classification and emotion recognition. It is necessary to identify and judge its meaningful information units, and then obtain the sentiment subject and evaluation opinion information. There are three primary methodologies employed to obtain evaluation information: the dictionary rule method, the general machine learning method, and the deep learning method. The dictionary rule method is characterised by simplicity, but its efficacy is constrained. General machine learning is predicated on feature engineering, while deep learning relies on neural networks with considerable learning capabilities. However, the training cost is high.

The method based on lexicon and rules generally uses existing knowledge resources, such as WordNet, to build a sentiment dictionary, and then builds rules based on the sentiment dictionary to judge emotions[10][11][12]. The approach utilizing machine learning was initially introduced by Pang et al. in 2002 [3]. In their method, the feature representation of the text is constructed using the sentiment dictionary, and then use NB, SVM, and ME models for positive and negative sentiment classification. After Pang, many people began to try to use machine learning methods for text sentiment analysis, and many new methods were proposed[13][14][15]. For machine learning methods, a big difficulty is the acquisition of training data. Training samples can be obtained by manual labelling, but this method is labour intensive and cannot obtain a large amount of labelling data. For texts such as Weibo, comments, etc., the emoticons in the text can be used to label the text[16], this labeling method might generate a degree of noise; nevertheless, it enables the effective gathering of significant training data while achieving satisfactory outcomes.

Convolutional Neural Networks—CNNs were originally proposed by LeCunY et al. and applied to handwriting recognition in the image domain[17], and LeNet-5 based on deep CNNs achieved good results. Collobert et al., 2011 proposed a novel application of CNN to NLP tasks, such as part-of-speech tagging. This development signified a significant expansion of the application scope of CNN [18]. In 2014, Kim suggested using CNNs for sentiment classification of English text and achieved notable results at the time [19]. In that same year, Kalchbrenner et.al developed an innovative wide convolution model that replaced the traditional max pooling layer in CNN [20]. This approach neither requires prior knowledge input nor the creation of complex handcrafted features. In 2016, Yin and Schutze pioneered the adoption of a multi-channel convolutional neural network architecture, incorporating convolutional cores of varying sizes for each channel. This innovation led to a substantial enhancement in the quality of feature extraction in sentence classification [21]. Building on this foundation, in 2018, Chen et.al advanced the state of the art by designing a multi-channel convolutional neural network model that enabled the comprehensive capture of sentence-level features through the utilisation of multiple CNN channels. This model yielded exceptional outcomes in the sentiment analysis task of Chinese microblogging, effectively surmounting the challenges posed by noisy data and the intricacies of language complexity [22]. However, a limitation of sentiment classification using CNNs is their inability to account for the contextual semantic information within sentences.

RNNs have a long history of development and are mainly used in speech processing-related problems. Rong et al., 2013 utilised a two-form RNN model with double closed-loop hidden layers to learn the vector representation of film review sentences and explore the emotion distribution. The model demonstrated the ability to capture timing and context information, generate accurate sentence vectors, support sentiment classification, and illustrate the potential of biformalised RNNs in sentiment analysis[23]. Kiros et al. and D. Tang et al. used a deep network based on gate units to model sentences[24][25]. Compared to CNNs, RNNs introduce memory units, enabling the network to retain certain information and capture long-distance dependencies in text. To address this, LSTM and GRU networks incorporate gating mechanisms, effectively mitigating the gradient vanishing issue in traditional RNNs. In 2015, D. Tang et al.[25] employed CNN and LSTM methods to obtain an effective representation of a single sentence. Subsequently, they utilised gated RNN technology to meticulously encode and process the correlations and semantic ties among sentences, thereby successfully constructing a high-precision text model at the text level. This approach effectively captures semantic information across sentences. In the same year, C. Zhou et al.[26] amalgamates the feature extraction advantages of CNN with the time series processing capability of LSTM, thereby facilitating the efficient operation of text classification tasks.

The hierarchical bidirectional LSTM model adopted by Ruder et al. [27] and the bidirectional LSTM method applied by Rao et al. [28] demonstrate excellent performance in both aspect- and document-level sentiment classification tasks. Furthermore, in the 2020 study, Sachin et al. utilised LSTM, GRU and their two-way forms to implement a detailed emotional analysis of Amazon user reviews, achieving notable research outcomes [29]. These pioneering research results not only promote the rapid development of text .

The attention mechanism represents a pivotal technique in the field of neural networks, with the capacity to enhance the efficacy of tasks such as NLP and computer vision by means of a targeted focus on critical components of the input data. Within the domain of NLP, the attention mechanism finds extensive application in text classification, machine translation, sentiment analysis, question-answering systems, and a range of other tasks [30]. Li et al. [31] proposed a new attention model, which optimises the performance of traditional RNNS and significantly improves the accuracy of Chinese emotion classification. The model accurately captures key text information and emotional trends, enhancing the ability of emotion recognition and providing technical support for NLP and emotion analysis.

In 2016, Z. Yang et al.[32] combined a bidirectional RNN with the attention mechanism, developing an attention model for chapter-level text classification. Later, in 2021, Gan et al. [33] have developed a multi-channel scalable design framework that integrates extended CNNs, BiLSTMs, and attention mechanisms to deeply explore the emotional content of Chinese texts. The framework utilises multi-channel design and can swiftly extract the original context and multi-level high-level features, thereby enhancing the accuracy and efficiency of sentiment analysis.

Transformers, a powerful class of models used widely in NLP, rely heavily on attention mechanisms. The Transformer architecture has been shown to achieve breakthroughs in NLP tasks by analysing input sequence element associations through a self-attention mechanism. Its encoder-decoder layout eschews RNNs and CNNs, instead relying on self-attention to capture sequence dependence for superior performance. Instead, the Transformer employs a self-attention mechanism, which captures the relationships between different words within the same sequence. For example, consider the sentence "Mary gave her friend a gift because she was grateful." As humans, we can easily understand that the word "she" refers to "Mary", based on the context of the sentence. However, for an algorithm, this association is not immediately clear. For the word "she", it calculates its relevance to all other words in the sentence. In this example, "she" has a stronger connection to "Mary" than to other words like "friend" or "gift", based on their semantic roles. By associating "she" with "Mary", the model understands they refer to the same person, allowing it to make a more accurate prediction or interpretation of the sentence. From studies, the Transformer model, which incorporates this Self-Attention Mechanism, has demonstrated exceptional performance across various tasks, including sentiment classification, surpassing the abilities of previous top models. Built upon the Transformer architecture, a range of groundbreaking pre-trained language models has been developed, demonstrating exceptional capability in learning general Chinese representations[34]. Z. Li et al., 2021 leveraged enhanced attention mechanisms to fully capture contextual information and encode the relative positions between words based on ELMo[35]. Li et al. utilised BERT to verify the efficacy of the Transformer model in the domain of Chinese sentiment analysis, particularly in its ability to comprehend text context and discern subtle emotional shifts, thereby surpassing the capabilities of RNNs[36]. Jing and Yang have devised a hybrid model that integrates RNNs, an attention mechanism, and Transformer, leading to a substantial enhancement in the accuracy of sentiment analysis[37]. Furthermore, Wang's team discovered that fine-tuning a pre-trained Transformer model leads to enhanced performance in Mandarin sentiment analysis when compared to traditional RNN models[38]. These findings underscore the efficacy of Transformer models in Chinese sentiment analysis and contribute to the advancement of the field.

RNNS frequently encounter challenges when processing sequence data due to signal attenuation, which can impede the establishment of long-term dependencies. Conversely, the Transformer architecture, with its self-attention mechanism, has the capacity to simultaneously consider the correlation of all elements in the input sequence. This capability enables the effective identification and interpretation of both short- and long-term dependency patterns, thereby demonstrating notable advantages in the context of dealing with this particular type of data. The Transformer architecture has been engineered to capture long-distance dependencies and parallel sequence processing, rendering it effective for tasks such as language modelling and sentiment analysis. However, contemporary pre-trained language models exhibit a substantial limitation in the domain of emotion analysis: they frequently fail to adequately integrate syntactic structure with semantic information and are unable to effectively extract and utilise key emotion-specific features.

This study introduces a learnable model enhanced with conjunction and adjective features for Mandarin Chinese sentiment analysis, built on the Transformer encoder architecture. The model has been developed to combine the position features of conjunctions and the weight features of adjectives into the attention mechanism. This has enabled it to successfully realise an accurate grasp of the global semantic context, while continuously focusing on key local information. The result is excellent performance and stability in natural language processing tasks.

3. System model

This study proposes a novel enhanced sentiment analysis model for Chinese social media texts based on a Transformer encoder. The model integrates an optimised multi-head attention mechanism, a trainable residual connection, and a residual design in the feedforward layer. The architecture of the model is shown in Figure 1. This architecture significantly improves the accuracy and processing efficiency of sentiment analysis.

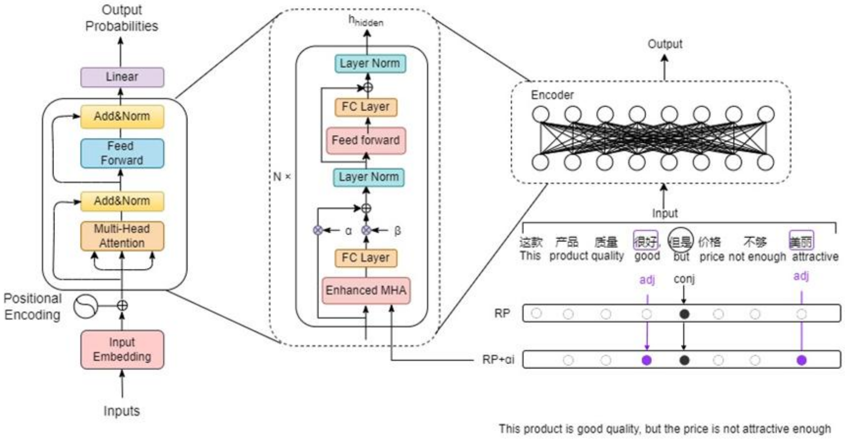


Figure 1. The overall system model architecture

3.1. Word embedding

In the field of sentiment analysis, the employment of models necessitates the provision of input data that is structured as vectors or tensors. Consequently, any text-based input must undergo transformation to align with these structural requirements. Within this paradigm, the advent of word embedding technology emerges as a pivotal development. This technological framework facilitates the mapping of words into compact, continuous vectors within a high-dimensional space, thereby ensuring that semantically analogous word vectors are proximate in this dimensionality. The integration of word embedding technology facilitates not only the conversion of text into a digital format but also the preservation of the semantic connections between words. This, in turn, ensures the provision of accurate and comprehensive data for sentiment analysis models. This allows text data, which is inherently discrete, to be transformed into a numerical format compatible with machine learning models like those used for sentiment analysis. By converting words into embeddings, the model can process text inputs as vectors or tensors, enabling advanced analysis like sentiment classification. This sentiment analysis strategy is characterised by a meticulous approach to the planning of the length of the input sequence, with the value of k being set to align with the broader text. In this strategy, the words in the sequence are transformed into vectors through the implementation of the word embedding technique. For words that are not present in the word dictionary, a random vector initialization strategy is adopted to ensure that the model is adequately equipped to handle diverse text inputs. Consequently, a $k \times d$ dimensional matrix is generated, where d indicates the dimensionality of the word representations. To standardize the length of the input sequence: If the text length is shorter than k, zero vectors of dimension d are added as padding until the sequence reaches length k. If the text length exceeds k, it is truncated to k. Following these rules, a text T can be expressed as:

$$T = w_1 \oplus w_2 \oplus \dots \oplus w_t \oplus \dots \oplus w_k, t = 1,2, \dots, k, \qquad (1)$$

In this study, w_t is the word embedding vector of the tth word in T sequence, which contains semantic information. $\oplus$ is a series operation of w_t vector to improve the model's ability to capture text semantics and context. This ensures that all input sequences have a consistent format for processing by the model.

In this study, word vectors that had been pre-trained on the Sogou corpus were used to initialise the baseline model. The dimension of the word vectors was set to 300. Following the training process, the model demonstrated effective mapping of Chinese words in vector space, and the distribution of similar semantic words was found to be close, thus establishing a solid foundation for sentiment analysis tasks.

3.2. Conjunction Enhanced Multi-Head Attention

The conjunction-enhanced multi-head attention mechanism incorporates relative location data to enhance the model's ability to recognise complex interactions between input text and conjunctive representations. In this research, conjunctions indicating transition, progression, selection, and coordination are identified using the Jieba tool, and it is imperative to consider that the distance between the sentence character and the initial character of the conjunction should be maintained at a positive value $d(d \geq 0)$. The relative position information of the conjunction is then normalised within the interval $(0,1)$, consequently resulting in the acquisition of the relative position feature RP:

$$RP = 1 - Sigmoid(d) = 1 - \frac{1}{1+e^{-d}} \tag{2}$$

In the absence of a conjunction, the relative position of each word in the sentence is characterised by calculating the distance between it and the starting point.

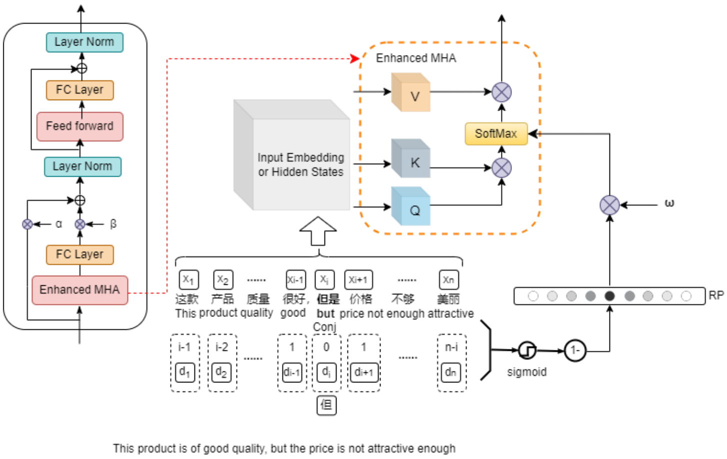

Layer Norm
FC Layer
Feed forward
Layer Norm
α
β
FC Layer
Enhanced MHA
Enhanced MHA
V
SoftMax
K
Q
Input Embedding or Hidden States
ω
X1
X2
......
Xi-1
Xi
Xi+1
......
Xn
这款 产品 质量 很好， 但是 价格 不够 美丽
This product quality good but price not enough attractive
Conj
RP
i-1
i-2
1
0
1
n-i
d1
d2
di-1
di
di+1
dn
但
sigmoid
1-
This product is of good quality, but the price is not attractive enough

Figure 2. Details of conjunction enhanced multi-head attention

As shown in Figure 2, the inclusion of RP significantly improves the model's ability to focus on conjunctive contexts. We optimized the calculation strategy for attention by adding the adjustable parameter ω and integrating it with the underlying input representation H to incorporate the relative position feature:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_K}} + \omega RP\right)V \tag{3}$$

$$where\ Q = HW^Q, K = HW^K, V = HW^V$$

3.3. Adjective Enhanced Multi-Head Attention

Adjectives have been shown to enhance multiple attention mechanisms, innovate traditional mechanisms, and integrate the syntactic attributes of adjectives. This change has been demonstrated to improve the semantic understanding of models and strengthen the analysis of complex structures. To focus more on adjectives, we introduced an adjective weight mask that amplifies the attention scores of adjectives. This can be done by adding a weight factor α_i to the softmax function, where α_i is greater than 1 for adjectives and 1 for other words. A_i is an indicator function that is 1 if the token at position i is an adjective, and 0 otherwise. λ is a hyperparameter that controls the amount of boost given to adjectives (e.g., $\lambda > 1$).

Then, the modified attention score for each token i becomes:

$$\alpha_i = 1 + (\lambda - 1) \cdot A_i \tag{4}$$

The enhanced attention weights are calculated as:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}} + \omega\alpha_i\right)V \tag{5}$$

Here we use the part-of-speech (POS) tagger jieba tool to identify which tokens in the sequence are adjectives (where $A_i = 1$)

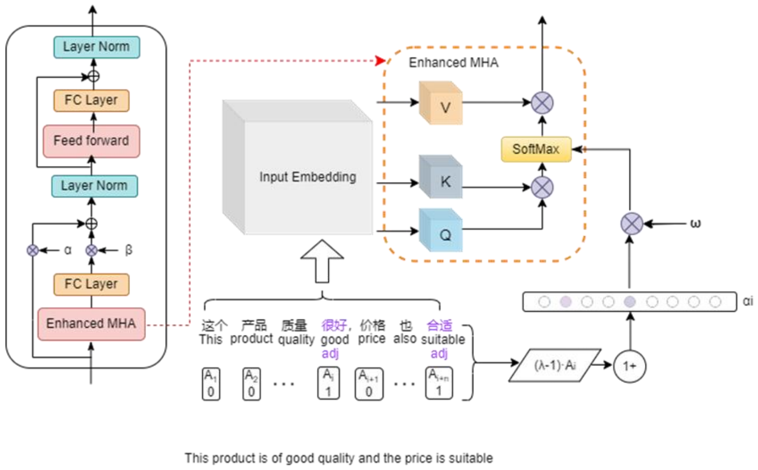


Figure 3. Details of adjective enhanced multi-head attention

As demonstrated in Figure 3, α_i accentuates the significance of the adjective weight, this is the adjective-boosting factor. For adjectives, $\alpha_i = \lambda$ $(where\ \lambda > 1$, e.g., $\lambda = 2$, boosting their contribution to the attention weights. For non-adjectives, $\alpha_i = 1$, leaving their weights unchanged. This adjustment ensures that adjectives, which play a critical role in sentiment polarity, are given higher attention weights, improving the model's focus on sentiment-bearing tokens in the sequence. Suppose the sentence is: "这个产品的质量非常好，价格也很合适("This product is of good quality and the price is suitable.") ", the POS tagging identifies "很好" (good) and "合适" (suitable) as adjectives. Adjectives often play a crucial role in sentiment analysis (e.g., words like "good" or "suitable" that directly reflect sentiment polarity), we set $\lambda > 1$ to amplify their attention weight. For example, if $\lambda = 2$, the weight of the adjective will be doubled, encouraging the model to focus more on these words, and the attention score for these adjectives will be amplified, giving them a stronger influence on the model's prediction.

3.4. Conjunction with Adjective Enhanced Multi-Head Attention

The overall model of this study adopts the architecture of Conjunction with Adjective Enhanced Multi-Head Attention (CAE-MHA), which aims to optimize the Transformer model's understanding of the semantics of sentences modified by conjunctions and adjectives. By introducing the information of adjective and conjunction annotations on the basis of the original attention mechanism and adjusting the attention weight, the MHA mechanism boasts a number of clear advantages when it comes to the processing of natural language. This is due to the mechanism's ability to take into account both local and contextual features, thereby enhancing the model's understanding.

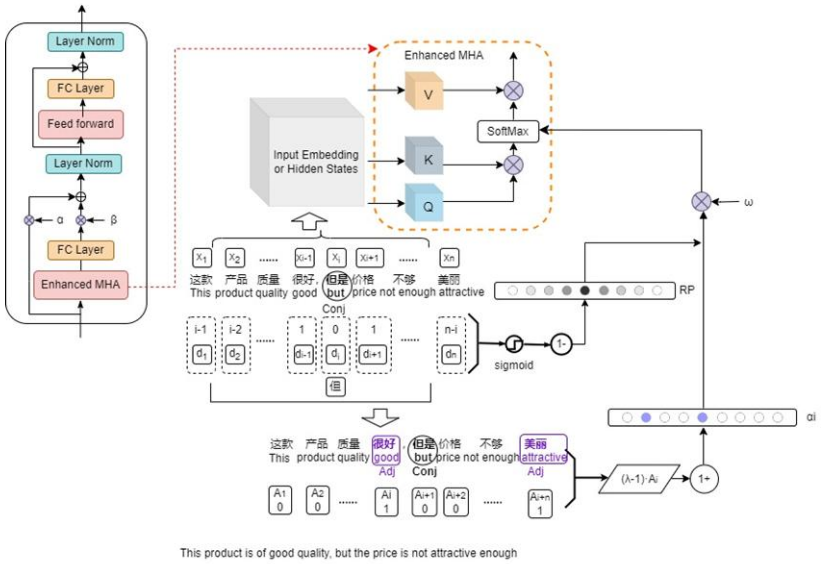

Layer Norm
FC Layer
Feed forward
Layer Norm
α
β
FC Layer
Enhanced MHA
Enhanced MHA
V
K
Q
SoftMax
Input Embedding or Hidden States
ω
X1 X2 ...... Xi-1 Xi Xi+1 ...... Xn
这款 产品 质量 很好，但是 价格 不够 美丽
This product quality good but price not enough attractive
Conj
RP
i-1 i-2 1 0 1 n-i
d1 d2 di-1 di di+1 dn
但
sigmoid
1-
αi
这款 产品 质量 很好，但是 价格 不够 美丽
This product quality good but price not enough attractive
Adj
Conj
Adj
A1 0
A2 0
Ai 1
Ai+1 0
Ai+2 0
Ai+n 1
(λ-1)·Ai
1+
This product is of good quality, but the price is not attractive enough

Figure 4. Details of conjunction with adjective enhanced multi-head attention

From the Figure 4, we can see that the input text "这款产品的质量非常好，但是价格不是很美丽。("This product is of great quality, but the price is not attractive enough")", is converted into a vector representation$x_1, x_2, \ldots, x_n$, with additional grammatical information (e.g., conjunctions d_i and adjectives A_i). The attention mechanism is a process which calculates weights based on queries and generates attention distributions with the objective of focusing on the Queries (Q), Keys (K), and Values (V).

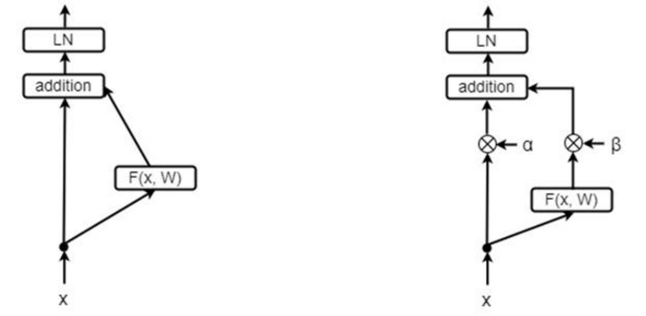
The model introduces conjunction POS d_i and adjective POS A_i and modifies the attention distribution through the weight adjustment function. The information of adjectives and conjunctions is weighted by weight factors α_i and RP, which reflect how much the model pays attention to words related to adjectives or conjunctions. If a word is an adjective or modified by an adjective, its attention value will be amplified by the adjective enhancement weight λ, if a word is a conjunction, then the local semantic features before and after the conjunction will be affected by the relative position of the conjunction. The final formula for conjunction with adjective enhanced attention becomes:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}} + \omega(\alpha_i + RP)\right)V \qquad (6)$$

Through attention distribution (Softmax) adjustment, the influence of conjunctions and adjectives on contextual information is weighted, ultimately forming a more semantically sensitive feature representation. The key semantic features of the optimised attention information are further refined by the subsequent processing of the feedforward neural network. Concurrently, residual connection technology can effectively alleviate the problem of gradient disappearance in deep networks and maintain the integrity of information. Moreover, the introduction of a layer normalisation module ensures data stability, improves the efficiency of training and the robustness of the model.

3.5. Learnable Residual Structure

The neural network demonstrates a high level of proficiency in feature expression, and the backpropagation algorithm enhances its architecture through gradient optimisation, thereby improving overall performance. However, during backpropagation, gradients may either diminish to near zero or grow exponentially, leading to ineffective parameter updates or gradient explosion. Additionally, deeper networks often face degradation issues. The residual learning technique employs the residual connection method to efficiently address the gradient problems that arise during the training of deep neural networks. These problems include gradient disappearance and explosion. In the domains of natural language processing and computer vision, this technology has been demonstrated to enhance the expressiveness of models and accelerate the development of related disciplines. Each component of the Transformer encoder is embedded with a hierarchical normalized residual design, which helps Transformer pre-trained variables optimize residual connections in a more efficient manner. This study proposes an innovative scheme that incorporates a configurable learning-based residual structure enhanced by conjunctions and adjective-induced self-attention features, while enabling flexible adaptive adjustments by assigning trainable parameters to each branch. The learnable residual structure is shown in Figure 5.



(a) Tranformer (b) Learnable Residual Structure in CAE-MHA

Figure 5. Residual Structure in Conjunction with Adjective Enhanced multi-Head Attention(CAE-MHA)

As posited by the residual design philosophy embedded in each layer of the Transformer architecture (Figure 5 (a)), the central role of layer normalisation in enhancing the overall performance of the model is widely acknowledged. It aids in optimizing nonlinear transformations to some extent. As posited by the concept of employing weight factors to dynamically adjust the branches of the residual network previously outlined, the core mechanism of the residual structure is to flexibly regulate the branch output with the assistance of weight factors. This is done in order to address the challenges posed by information flow barriers and gradient disappearance:

$$\mathcal{Y} = LN(\alpha x + \beta \mathcal{F}) \qquad (7)$$

As shown in Figure 5 (b), the trainable residual element underscores the residual architecture of the multi-head attention mechanism. Incorporating learnable parameters α and β enables the model to autonomously learn and promptly adjust the optimal proportional coefficient between the input branch x and the residual branch $\mathcal{F}$. This mechanism enhances the flexibility and adaptability of the model, optimises the flow of information and the presentation of features, and consequently contributes novel approaches and insights to the performance enhancement of deep neural networks. The enhanced attention mechanism facilitates the optimisation of scale setting during the dissemination of information. In this process, the scale factor successfully reduces the noise introduced by the positional characteristics of conjunctions and the syntactic attributes of adjectives, thus improving the accuracy of semantic representation and enhancing the efficiency of the model in dealing with complex language structures and deep semantic parsing. The scaling factors α and β jointly govern the distribution balance between x and $\mathcal{F}$. Additionally, layer normalization is applied to maintain consistency in the distribution across each layer, preventing issues like gradient vanishing or explosion that could arise due to changes in the learnable parameters.

3.6. Sentiment Classification and Optimization

The Transformer encoder extracts sequence features layer by layer, and finally maps the features to the category space through fc1. The result of the ultimate classification task is generated by the model's final fully connected layer (fc1), often referred to as the linear classifier. This process can be expressed using the following:

$$Y = W \cdot X^{\mathrm{T}} + b \qquad (8)$$

Y is the output matrix, representing the scores for each category for all input samples in the batch. The weight matrix, denoted by W, is responsible for the accurate mapping of the input data to the feature domain, which is defined by the number of categories required. Additionally, the weights are adjusted through the training process. b is the bias vector, providing a bias term for each output category. The output of the encoder provides a concise summary of the sequence's characteristics. The fully connected layer directly employs these features for classification, thereby eliminating the necessity for global pooling. This approach enhances the efficiency of the classification process while preserving its accuracy. The model uses the features at each position for classification instead of globally aggregating the sequence. This layer maps the output features of the encoder to the dimension of the category probability distribution through a linear transformation. This output is usually passed to a loss function to calculate the loss for classification tasks.

Adam optimizer is a widely used optimization method in deep neural networks, mainly because its adaptive learning rate and momentum mechanism can effectively deal with challenges in model training, such as gradient sparsity, gradient disappearance or explosion, and high-dimensional parameter optimization. The Adam optimisation algorithm is a machine learning algorithm that automatically adjusts the parameter learning rate according to the first-order and second-order statistics of the gradient. This enables it to adapt to changes in the gradient, thereby enhancing the efficiency and robustness of the optimisation process, and improving training efficiency. It performs well in models such as Transformer and BERT and is the default optimizer.

3.7. Loss Function

In the process of model training, a special loss function is utilised to enhance the expressiveness of the model. Specifically, for the execution of the standard classification task, the standard cross entropy function is utilised for the evaluation of the loss. A custom additional penalty term based on the model output is introduced to adjust the final loss. The final weighted total loss combining standard cross entropy and additional penalty terms:

$$L = L^{ce} + penalty \qquad (9)$$

Where cross-entropy loss is:

$$L^{ce} = -\frac{1}{N}\sum_{i=1}^{N} logP_{i,y_i} \qquad (10)$$

$O \in \mathbb{R}^{N \times C}$ is the model's output (logits) represents the prediction of N samples for C categories. $y \in \{1,2,\dots,C\}^N$ is the true labels. $P \in \mathbb{Z}^N$ is the replication count array corresponding to the label indicates the number of replications for each label. Where generate a copy tensor is: Using the number of replications P and the model output O, dynamically generate an expanded tensor R.

$$R = stack\left(O_{1,y1,\dots,}O_{1,y1,\dots,}O_{N,yN,\dots,}O_{N,yN}\right) \qquad (11)$$

Assume that the total number of samples after the expansion is $M = \sum_{i=1}^{N} p_i$, then the penalty can be calculated for the probability value of each sample in the expanded tensor R.

$$penalty = \beta \cdot \frac{1}{M} \sum_{j=1}^{M} \sum_{k=1}^{C} (1 - R_{j,k}) \qquad (12)$$

Where, $R_{j,k}$ indicates the projected probability of the kth category for the jth instance in the enhanced tensor R. Dynamic adjustment of the penalty term can strengthen the learning of certain categories or samples in a targeted manner. The hyperparameter β controls the effect of the penalty term on the total loss, balancing the main task (classification) with additional features (such as the prediction confidence of the model). The replication mechanism can handle complex sample weighting requirements, such as enhancing the focus on important categories or samples.

4. Experiment

4.1. Evaluation criteria

To further verify the model's superior performance, we use Accuracy, Precision, Recall, and F1 as evaluation metrics. For a classification problem with n categories, let TP_i/FP_i denote the True/False Positive of its class, and TN_i/FN_i represent the True/False Negative of the ith class, then some evaluation criteria to measure the model performance can be defined as follows.

In order to comprehensively evaluate the performance of the model, the following evaluation metrics were adopted: accuracy, precision, recall and F1 score. For the N-category classification task, the following basic statistics were set as the foundation for index calculation: TP_i/FP_i *and* TN_i/FN_i.

Acc: The ratio of accurately identified samples to the overall number of samples, indicates the overall accuracy of the model.

$$Acc = \frac{\sum_{i=1}^{n}(TP_i+TN_i)}{\sum_{i=1}^{n}(TP_i+TN_i+FP_i+FN_i)} \qquad (13)$$

P: The model's capacity to predict Type I positive samples is determined by comparing the proportion, with a high proportion indicating a low misjudgement, which is a crucial indicator for evaluation,

$$P_i = \frac{TP_i}{TP_i+FP_i} \qquad (14)$$

R: Within the subset of true positive samples, the model accurately identifies them as Class I positive samples,

$$R_i = \frac{TP_i}{TP_i+FN_i} \qquad (15)$$

F1: The harmonic mean of precision and recall, which combines the efficacy of both metrics, for the th class. When Precision and Recall are unbalanced, F1 is a better indicator,

$$F1_i = \frac{2 \times P_i \times R_i}{P_i+R_i} \qquad (16)$$

In this study, the model's effectiveness is evaluated using two key performance indicators: accuracy and F1 score. Accuracy is used to measure the overall accuracy of the model's predictions, while F1 score integrates accuracy and recall rate to provide a more comprehensive evaluation of the model's performance.

4.2. Datasets

The objective of this study is to undertake a comprehensive evaluation of the performance of the proposed method across a range of sample sizes and variations in text length. In accordance with this objective, the Ctrip dataset, the "merge" dataset and the takeout dataset have been selected as the experimental materials. Please refer to Table 1 for detailed descriptions of these datasets.

Table 1 Specific Information of Three Datasets

No	Datasets	Category	Number	Class
1	ctrip	Positive	24,662	3
		Negative	2,940	
		Neutral	881	
2	merge	Positive	6000	2
		Negative	6000	
3	waimai	Positive	4,000	2
		Negative	7,988	

The “ctrip” dataset: The Ctrip dataset comprised 28,483 hotel reviews that were sourced from Ctrip and annotated, but not segmented. Following the pre-processing of the text with Jieba segmentation technology and the removal of invalid data, the dataset consisted of 24,662 positive instances, 2,940 negative instances, and 881 neutral instances using "1", "0", and "2" as sentiment labels, respectively, as illustrated in Figure 6.

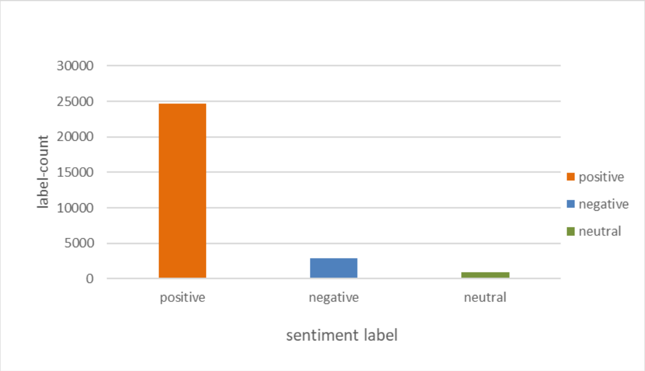
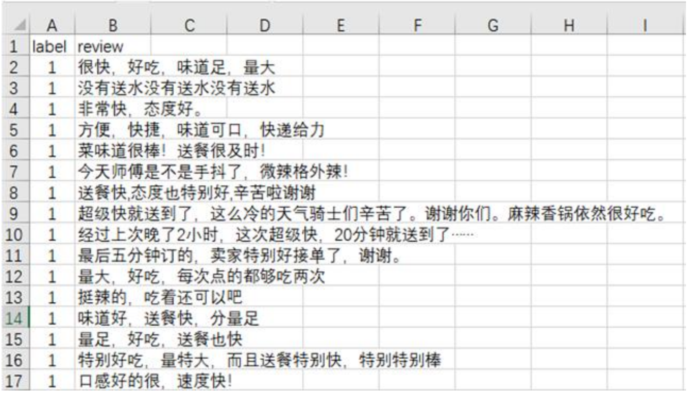


Figure 6. Sentiment label distribution of ctrip dataset

The “merge” dataset: The construction of this dataset entailed the integration of three distinct subsets derived from disparate social media platforms, namely tongcheng_hotel_4000, Dangdang_book_4000, and Jingdong_nb_4000. Each dataset has about 4,000 comments, and after merging, there are a total of 12,000 valid comment data. Among them, there are 6,000 positive comments and 6,000 negative comments. tongcheng_hotel_4000 is a hotel review dataset from ly.com, Dangdang_book_4000 is a book review dataset from Dangdang.com, and Jingdong_NB_4000 is a laptop review dataset from JD.com.

The “waimai” dataset: This dataset is a take-away food review corpus. After cleaning and removing invalid and sensitive data, it has 11,988 samples. The data comes from Baidu waimai, a Chinese food delivery platform (now acquired by Ele.me, China's largest food delivery platform). The present dataset encompasses all emotionally annotated comments (4,000 positive and 7,988 negative), derived from the CSDN platform and published for the purpose of natural language processing research in the domain of Chinese text sentiment analysis. In the construction of the dataset, the number "0" is assigned to negative emotions and the number "1" to positive emotions, as illustrated in Figure 7.



	A	B
1	label	review
2	1	很快，好吃，味道足，量大
3	1	没有送水没有送水没有送水
4	1	非常快，态度好。
5	1	方便，快捷，味道可口，快递给力
6	1	菜味道很棒！送餐很及时！
7	1	今天师傅是不是手抖了，微辣格外辣！
8	1	送餐快,态度也特别好,辛苦啦谢谢
9	1	超级快就送到了，这么冷的天气骑士们辛苦了。谢谢你们。麻辣香锅依然很好吃。
10	1	经过上次晚了2小时，这次超级快，20分钟就送到了……
11	1	最后五分钟订的，卖家特别好接单了，谢谢。
12	1	量大，好吃，每次点的都够吃两次
13	1	挺辣的，吃着还可以吧
14	1	味道好，送餐快，分量足
15	1	量足，好吃，送餐也快
16	1	特别好吃，量特大，而且送餐特别快，特别特别棒
17	1	口感好的很，速度快！

Figure 7. The sample content of waimai dataset

4.3. Text embedding and pre-training

In the text embedding part, the first step is to build a vocabulary and extract pre-trained word vector embeddings to provide basic word representations for subsequent text classification or Sentiment analysis, a branch of natural language processing, can be divided into several elements, firstly is vocabulary construction, the script processes a specified Excel dataset by extracting text data from a given column "text". A tokenizer, defaulting to character-level segmentation, splits the text into tokens. Token frequencies are then calculated, and only those with a frequency above min_freq and within the maximum vocabulary size (max_size, default 10,000) are retained. Two special tokens, <UNK> (unknown words) and <PAD> (sequence padding), are appended to handle unseen words and sequence alignment. The resulting vocabulary, mapping tokens to unique indices, is serialized into a .pkl file for future use. Secondly is loading pre-trained embeddings, the code integrates pre-trained word embeddings sgns.sogou.char, a character-level embedding file. For tokens in the vocabulary, it retrieves their corresponding embeddings from the pre-trained model. If a token's embedding is unavailable, its vector is initialized randomly. The embedding size (emb_dim) is set to 300 by default, ensuring consistency across tokens. The final embedding matrix, aligning each token index to a pre-trained or randomly initialized vector, is saved in a compressed .npz file. This ensures efficient storage and reuse during downstream model training.

By filtering tokens based on frequency and vocabulary size constraints, the script constructs a compact and task-relevant vocabulary. The inclusion of <UNK> and <PAD> ensures robustness against out-of-vocabulary tokens and supports variable-length sequence alignment, which is crucial for deep learning models. Leveraging pre-trained embeddings, such as character-level embeddings from sgns.sogou.char, provides rich semantic representations for known tokens, improving model performance and convergence speed. Random initialization for missing tokens ensures completeness of the embedding matrix, maintaining consistency. By serializing the vocabulary and compressing the embedding matrix, the reusability and storage efficiency of resources are optimized, providing standardized input for subsequent deep-learning models.

4.4. Training details

The embeddings are converted to a PyTorch tensor (torch.tensor) for model compatibility. In this study, the model configuration was set to utilise 300-dimensional word vectors to represent the semantic information of vocabulary. Furthermore, the learning rate was adjusted to 5e-4 with a view to enhancing the training efficiency. Concurrently, to mitigate the risk of overfitting, a dropout ratio strategy of 0.5 was adopted. Furthermore, an early stop training mechanism was introduced, whereby if the validation set performance does not improve significantly within 500 consecutive training batches, the training process is automatically terminated. The model trains for a maximum of 20 epochs, controlling the number of complete passes over the dataset. Mini-batch size is set to 128. If pre-trained embeddings are used, their dimension is derived automatically. Otherwise, it defaults to 300. The dimensionality of input and output embeddings passed to the transformer layers (300). In this study, the hierarchical structure of the model is meticulously delineated. The middle hidden layer is endowed with 1024 neurons, while the final hidden layer is furnished with 512 neurons. To enhance the model's capacity to discern intricate details, a multi-head attention mechanism is implemented, comprising five attention heads. Moreover, to optimise the trade-off between model performance and computational cost, a two-layer encoder design is adopted for the transformer architecture.

Table 2 Hyper Parameters of Experiment

Hyper Parameters Description	Number
Size of word vector	300
Learning rate	5e-4
Dropout rate	0.5
Require_improvement	500
Num of epochs	20
Mini-batch size	128
Pad size	32
Hidden size	1024
Number of head	5
Number of encoder	2

In accordance with the ratio of 6:2:2, the review text of each data set was randomly assigned to the training set, the verification set and the test set, with a view to maximising data utilisation. Concurrently, the adoption of unbalanced data partitioning methods served to enhance the generalisation capability of the model, thereby effectively reducing the likelihood of overfitting. Consequently, this ensured that the stability and accuracy of the evaluation results were significantly improved.

5. Result Comparisons

5.1. Accuracy and Loss comparison

In this study, the proposed model is compared with the transformer encoder and LCEM models on Ctrip, merged and outsourced datasets. The pre-trained model implements character-level processing, and the data sets are randomly divided and distributed in a 6:2:2 ratio. Table 3 shows the comparison of test accuracy to provide data support for performance evaluation.

Table 3. Comparison of Model Test Accuracies on Three Public Datasets

Datasets	ctrip	merge	waimai
	Acc(%)	Acc(%)	Acc(%)
Transformer	88.71%	76.75%	86.57%
LCEM	89.24%	77.17%	87.90%
Mymodel	90.19%	78.33%	88.61%

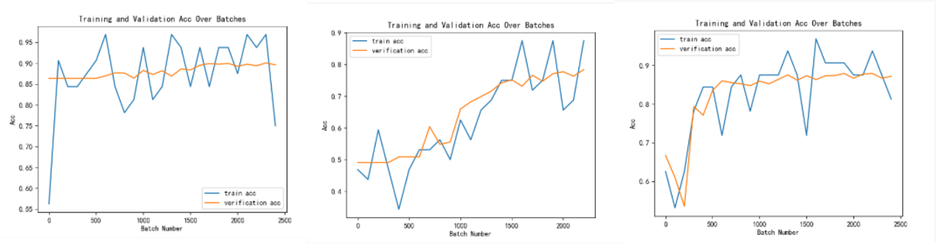
Table 3 shows the test accuracy comparison of three models (Transformer, LCEM and MyModel) on three public data sets. An analysis of Ctrip data sets reveals distinctive characteristics, with the text exhibiting a high degree of coherence and incorporating extensive classification details. These characteristics not only augment the complexity of data processing but also furnish valuable data resources and practical opportunities for the development of high-performance classification models. The transformer's accuracy is 88.71%, which is slightly weaker on this dataset. LCEM improves to 89.24%, capturing more semantic information by improving the model structure. MyModel achieved 90.19% and further improved the classification performance by introducing specific features such as the weight of adjectives and conjunctions. MyModel's position weighting and feature engineering work remarkably well on this dataset.

The characteristics of the merge data sets are that they may contain more complex or sparse feature distributions. The Transformer's accuracy is 76.75%, which is a weak performance and may not fully learn complex features. LCEM improved to 77.17%, and the improvement was limited, indicating that its feature-capturing ability has certain limitations. MyModel reaches 78.33%, which is 1.16% higher than LCEM, indicating that its targeted feature enhancement method is more suitable for processing complex text data. The advantages of MyModel are still significant in this dataset.

The characteristic of the waimai dataset is that it may involve more intuitive features, such as emotional tendencies in user comments. With an accuracy of 86.57%, Transformer performs relatively well on this dataset, but there is still room for improvement. LCEM improves to 87.90%, which is a large improvement, probably due to its enhanced feature modelling capabilities. MyModel reaches 88.61%, further improving accuracy, but the improvement (0.71% compared to LCEM) is not as significant as other datasets. MyModel's improvement on this dataset is mainly due to its model optimization and fine-grained feature extraction capabilities.

The above test accuracy shows the overall performance of the three models in three datasets. MyModel achieved the highest accuracy on all datasets, demonstrating the effectiveness and adaptability of its design. LCEM ranked second, indicating that its model structure has improved to a certain extent compared to Transformer. Transformer performed the weakest, indicating that the basic Transformer architecture is not sufficiently optimized for these specific tasks. The MyModel innovation has been shown to enhance text processing performance through two core technological solutions. Firstly, the integration of weight features of conjunctions and adjectives has been demonstrated to improve the model's capacity to identify and capture key text information. Secondly, the employment of a location-sensitive attention mechanism enables the model to focus on specific word groups in the text with greater precision. The improvement of LCEM may be reflected in more complex embedding layers or optimization algorithms, but the effect is limited. MyModel has significant improvements on the ctrip and “merge” datasets, indicating that it is suitable for processing complex text features and long-distance dependencies. The improvement on the waimai dataset is slightly smaller, probably because the features of this dataset are simple and the existing methods are close to the upper limit of performance.

As demonstrated in Figure 8, the training and verification accuracy of MyModel on three distinct data sets, namely Ctrip, Merge and Takeout, exhibit a clear trend of stability in its learning performance and strong potential for generalisation across diverse data sets. The experimental outcomes substantiate the model's adeptness in discerning the inherent characteristics of the data during the training phase, accompanied by a steady enhancement in verification accuracy. This validates its efficacy and broad adaptability across a range of data sets:



MyModel-ctrip-Acc MyModel-merge-Acc MyModel-waimai-Acc

Figure 8. Train and Validation comparison in accurate of MyModel

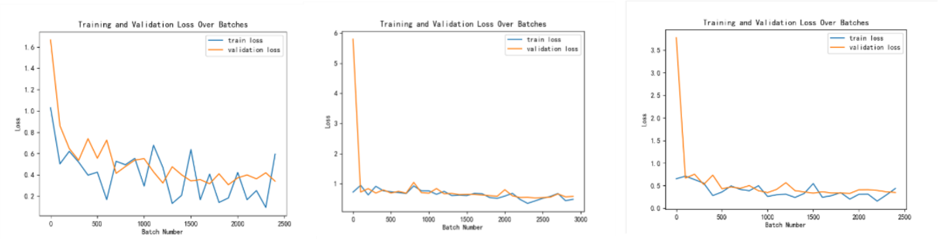
In the context of the Ctrip data set, Mymodel demonstrated a swift escalation in accuracy during the initial phase of training, readily attaining and surpassing the stringent precision criterion of 0.9. Despite the variability in accuracy during the subsequent stage, the model maintained a consistently high level of accuracy, thereby underscoring its effective learning capability and its capacity for precise data feature identification. The verification accuracy (verification acc) shows a slow growth trend and eventually stabilizes at around 0.85. The model quickly reaches a high accuracy on the training set, indicating that the learning effect of the training data is good. The verification accuracy is stable, but it is always lower than the training accuracy, which may be due to some overfitting.

The training accuracy (train acc) of MyModel on the combined dataset demonstrated considerable fluctuation in the early phases, but gradually increased and stabilized (about 0.85) as the training progressed. The verification accuracy (verification acc) gradually increased with less fluctuation, and finally approached the training accuracy (about 0.82). The training accuracy of the combined data set exhibits significant fluctuations, which may be indicative of the high complexity of the data set or the escalation of the difficulty of the model learning task. However, the verification accuracy is closely aligned with the training accuracy, indicating that the model effectively reduces overfitting and possesses excellent generalisation capabilities.

In the context of the Waimai dataset, MyModel exhibited optimal learning efficiency, demonstrating a rapid enhancement in accuracy at the commencement of the training process. It effectively surpassed the stringent precision standard of 0.9, thereby substantiating the efficacy and precision of the model. Although there were some fluctuations, it was generally stable. The verification accuracy (verification acc) showed a steady upward trend, eventually approaching 0.88, with small fluctuations. The training of the model on the waimai dataset was carried out smoothly, without any major complications, the verification accuracy performed well, and the gap between training and verification was small, indicating that the generalization ability was strong.

The verification accuracy of MyModel exhibited excellent stability across all the examined data sets, demonstrating a high degree of consistency with the training accuracy. This effectively substantiated the model's capacity to circumvent severe overfitting, its superlative generalisation capability, and its remarkable stability. These findings provide substantial validation for subsequent scientific research and practical applications. On the waimai dataset, MyModel performs best, with both training and validation accuracy reaching a high level (validation accuracy is close to 0.88). On the “merge” dataset, the training process of MyModel is still complicated, but the stability of validation accuracy is significantly improved. MyModel's performance on the merge data set is significantly smoother, and the verification accuracy curve oscillates less. The model demonstrates robust performance on the Waimaai dataset, exhibiting comparable efficacy to the reference model. Of particular note is its superior generalisation capability and heightened stability across all data sets, particularly in the context of merge and Waimaai data sets. Furthermore, the accuracy curve is found to be smooth, thereby establishing a reliable foundation for subsequent optimisation and extensive utilisation of the model.

As illustrated in Figure 9, the training and validation losses of MyModel on the Ctrip, merge and Waimai datasets demonstrate a rapid decrease over time, eventually stabilising, thereby substantiating the model's merits through its expeditious learning, robust convergence and substantial generalisation capabilities:



MyModel-ctrip-Loss MyModel-merge-Loss MyModel-waimai-Loss

Figure 9. Comparison of the accuracy of training and verifying MyModel

The training loss of the MyModel model on the ctrip dataset drops rapidly from an initial value of about 1.6, stabilizes after 500 batches, and finally converges to close to 0.1. The validation loss drops rapidly at the beginning, then approaches the training loss, and finally stabilizes at around 0.2. In the model training stage, an increase in the number of batches has been shown to result in a gradual decrease in both the training loss and the validation loss. An analysis of the early training stage reveals a sharp decrease in the training loss, suggesting that the model is capable of efficient learning. Subsequent analysis shows that, although the training loss experienced small fluctuations, the overall trend of decline remains clear, indicating that the model is undergoing a process of continuous self-adjustment to find a better solution. The validation loss fluctuates greatly, but generally shows a downward trend, and gradually approaches the training loss in the later stage. The experimental results show that the model has a certain degree of overfitting on this dataset. In some batches, the training loss is lower than the validation loss.

During the training stage of the combined data set, both the training loss and validation loss of MyModel underwent a significant decrease from a high to a low level, ultimately converging to a near-low level. Furthermore, the loss curves of both converged to a similar point. This finding serves to provide robust verification that MyModel possesses not only excellent fitting ability and the capacity to accurately capture the inherent characteristics of training data, but also exhibits excellent generalisation capability and the ability to maintain stable prediction accuracy on unseen data. The dataset is complex (the initial loss is as high as 5.0), but the model successfully overcame the data complexity through training and reached a stable convergence state. The model did not show overfitting or underfitting on this dataset, and the optimization process was effective.

The training loss of MyModel on the waimai dataset decreases swiftly from an initial value of approximately 3.5, and stabilizes after about 500 batches, eventually converging to close to 0. The validation loss also drops rapidly from the initial high value, then stabilizes, and finally remains in the range of 0.2-0.3. In the training and validation process of the waimai dataset, we noticed that the validation loss was slightly higher than the training loss, but the difference was slight and the trend was the same. This reflects the excellent performance of the model on the waimai dataset, with strong generalization ability, and no obvious overfitting phenomenon. The model can quickly cope with the initial high-complexity data set and promote the steady decline of the loss value, which fully proves the effectiveness and robustness of the model, and lays a solid foundation for the subsequent application on complex data sets.

On all three datasets, the model showed high training efficiency, and the training loss dropped rapidly. The “merge” dataset had the highest initial loss (5.0), but it dropped quickly, and the final convergence effect was consistent with the other datasets.

5.2. Ablation study

As illustrated in Table 4, the results of the Mymodel ablation experiments on three datasets are presented in a comprehensive manner, with particular emphasis placed on several core optimisation strategies. Specifically, the $+\alpha i$ symbol indicates the integration of adjective-weight masking in the self-attention mechanism of the baseline model; on the basis of $+\alpha i$, $+\omega\alpha i$ further incorporates weighted adjective weight masking; and $+\omega RP$ indicates that weighted relative position information is included in the baseline model. The symbol $+LRS$ denotes the implementation of a learnable residual design in the Transformer coding layer.It is noteworthy that these alphabetic optimisation elements are of paramount importance, collectively providing a robust foundation for a comprehensive evaluation of Mymodel performance and a solid basis for subsequent model enhancement efforts.

Table 4 Results of ablation experiment

Datasets	**ctrip**		**merge**		**waimai**	
	Acc(%)	**F1(%)**	**Acc(%)**	**F1(%)**	**Acc(%)**	**F1(%)**
(Baseline)Transformer	88.85	87.57	76.42	76.94	87.36	78.97
$+\alpha i$	87.38	81.49	78.17	78.92	86.86	79.79
$+\omega\alpha i$	87.13	82.95	78.12	77.14	86.65	79.70
$+LRS$	87.89	86.47	75.17	74.40	85.98	78.71
$+\ \alpha i\&LRS$	88.85	86.11	76.21	75.71	86.73	79.43
$+\ \omega RP\&LRS$	89.57	87.60	78.25	77.17	87.69	81.62
$+\ \omega\alpha i\&LRS$	89.73	87.88	78.96	78.35	86.98	80.88
$+\ \omega\alpha i\&\omega RP\&LRS$	90.26	88.18	79.79	80.72	88.57	81.97

First, A comprehensive and in-depth analysis of the performance of baseline model transformers (encoders) is provided on three distinct data sets: ctrip Accuracy = 88.85%, F1 = 87.57%; merge: Accuracy = 76.42%, F1 = 76.94%; waimai: Accuracy = 87.36%, F1 = 78.97%. Through observation and analysis, we can see that the baseline values of Accuracy and F1 of the ctrip dataset are both high, indicating that the ctrip dataset has a better adaptability to the Transformer model, and the feature distribution may be simpler or the data quality is higher. The “merge” dataset has the lowest Accuracy and F1, which are 76.42% and 76.94% respectively, indicating that the “merge” dataset may contain more complex features or greater noise.

In the ablation experiment, the performance of each single module after being added separately is as follows. The $+\alpha i$ module performance changes: ctrip Accuracy = 87.38%, F1 = 81.49%, F1 dropped significantly. merge Accuracy = 78.17%, F1 = 78.17%, slight improvement. waimai Accuracy = 86.86%, F1 = 79.79%, slight improvement. The $+\alpha i$ module has an unfavourable effect on the ctrip data set, with a decrease in both Accuracy and F1, which may indicate that the $+\alpha i$ module is not sensitive to specific characteristics of the ctrip data. On the “merge” and “waimai” data sets, the performance improvement brought by the $+\alpha i$ module is small, indicating that its independent effect on feature extraction is limited. $+\omega\alpha i$ module performance changes: ctrip Accuracy = 87.13%, F1 = 82.95%, slightly decreased. merge Accuracy = 78.12%, F1 = 77.12%, slightly improved. waimai Accuracy = 86.65%, F1 = 79.70%, slightly improved. The $+\omega\alpha i$ module performs better than ai, but the overall improvement is still limited. On the ctrip data set, although the F1 score drops slightly, its stability is slightly stronger than $+\alpha i$ module. $+LRS$ module performance changes ctrip Accuracy = 87.89%, F1 = 86.47%, close to the baseline. merge Accuracy = 75.17%, F1 = 75.17%, basically the same as the baseline. waimai Accuracy = 85.73%, F1 = 78.70%, close to the baseline. The $+LRS$

module has significant results on the ctrip data set, especially the improvement in F1 score, which shows that it has certain advantages in enhancing recall rate. On the “merge” and “waimai” data sets, the effect of LRS is slightly stable, and the performance improvement is limited.

Ablation experiments further verified the effect of multiple module combinations. $+\alpha i\&LRS$ module performance changes: ctrip Acc = 88.85%, F1 = 86.11%, slightly lower than the baseline. merge Acc = 78.12%, F1 = 78.12%, significant improvement. waimai Acc = 86.98%, F1 = 80.88%, the performance is significantly improved. It shows that the combination of $+\alpha i$ and $+LRS$ has a large performance improvement on the “merge” and “waimai” data sets. In particular, the F1 score of the “waimai” data set reaches 80.88%. The performance of the ctrip data set dropped slightly, possibly because these two modules have redundant modelling of ctrip features. $+\omega RP\&LRS$ module performance change: ctrip Acc = 89.57%, F1 = 87.60%, exceeding the baseline. merge Acc = 78.25%, F1 = 77.17%, significant improvement. waimai Accuracy = 87.69%, F1 = 81.62%, performance further improved. The combination of $+\omega RP$ and $+LRS$ has a relatively stable improvement effect on the three data sets, especially on the “waimai” data set, which shows that $+\omega RP$ can effectively capture feature relationships and optimize recall. $+\omega\alpha i\&LRS$ module performance changes: ctrip Accuracy = 89.73%, F1 = 87.88%, significant improvement. merge Accuracy = 78.96%, F1 = 78.35%, a huge improvement. waimai Accuracy = 86.98%, F1 = 81.62%, excellent performance. The combination of $+\omega\alpha i$ and $+LRS$ further enhances the feature extraction capability and outperforms the single module on the three datasets. $+\omega\alpha i\&\omega RP\&LRS$ module performance changes: ctrip Accuracy = 90.26%, F1 = 88.18%, reaching the highest. merge Accuracy = 79.79%, F1 = 80.72%, significant improvement. waimai Accuracy = 88.57%, F1 = 81.97%, the best performance. This is the best-performing module combination, and its performance on the three data sets has reached the highest level in the experiment. This shows that the synergistic effect of $+\omega\alpha i$, $+\omega RP$ and $+LRS$ can maximize the classification ability of the model.

The combination of $+\omega\alpha i\&\omega RP\&LRS$ brings Accuracy and F1 to 90.26% and 88.18%, which are the highest values in all experiments. The “merge” is the data set with the weakest performance, but the effect of module combination is significantly improved on this data set. The combination of $+\omega\alpha i\&\omega RP\&LRS$ increases Accuracy and F1 to 79.79% and 80.72% respectively, indicating that the synergy between modules is more effective in enhancing the ability to model complex features. The baseline performance on the “waimai” dataset is good, but the baseline value of F1 (78.97%) is slightly lower than ctrip. After multi-module combination, the F1 of $\omega\alpha i\&\omega RP\&LRS$ increased to 81.97%, showing high robustness.

In summary， introducing $+\alpha i$, $+\omega\alpha i$ or $+LRS$ modules alone has a small improvement in performance, indicating that their optimization of feature modeling is limited when acting alone. The combination of multiple modules can greatly improve the performance, among which $+\omega\alpha i\&\omega RP\&LRS$ is the optimal combination, which significantly enhances the classification ability of the three data sets. In terms of cross-dataset adaptability, the improvement effects of the ctrip and waimai data sets are more prominent, while the merge data set is limited by its characteristics, and the improvement is relatively small. This hierarchical, cross-module ablation experiment provides reliable evidence that the combination of $+\omega\alpha i$, $+\omega RP$, and $+LRS$ can effectively optimize model performance, especially with significant advantages when dealing with complex tasks.

5.3. Attention Visualization

In order to demonstrate the model effect more intuitively and effectively, in this research we use the seaborn library and matplotlib library to visualize the distribution of word attention weights in sentences in the experimental part. This paper selects one paragraph of positive and one paragraph of negative comments for visualization experiment.

文本 1：总体还是感觉很不错，房间十分干净、简洁、舒适。

Text 1: The overall ambiance is delightful. The environment is neat, minimalist, and welcoming.

文本 2：酒店前台小女孩的服务太不好了，态度十分差。

Text 2: The service of the little girl at the hotel front desk is too bad and her attitude is very bad.

For text 1 and text 2, We start by utilizing the Jieba word segmentation tool to segment the sentences, subsequently removing punctuation and stop words to isolate the individual terms "总体 overall", "感觉 feeling", "还 still", "不错 good", "房间 room", "很 very", "干净 clean", "简洁 simple", "舒适 comfortable" and "酒店 hotel", "前台 front desk", "

小 small", "女孩 girl", "服务 service", "太 too", "次 bad", "态度 attitude", "很差 very bad", and then draw the word-level attention weight heat map as shown in Figure 9 and Figure 10. As illustrated in the figure, a positive correlation exists between the colour depth and the grey value, which intuitively reflects the distribution of lexical weights.

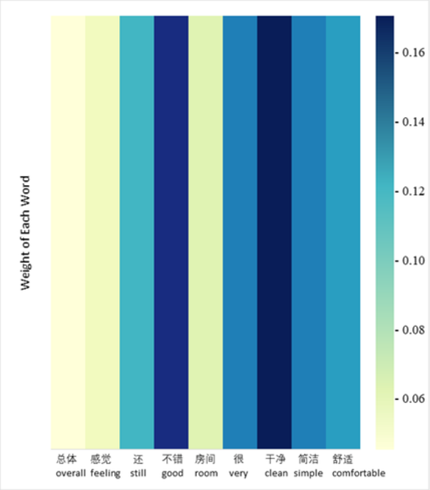


Figure 9 Text 1 word attention weight heatmap

After jieba word segmentation and the removal of stop words and punctuation, text 1 contains 9 words. From Figure 9, we can seethat the model allocates greater emphasis to the words "还 still", "不错 good", "很 very", "干净 clean", "简洁 simple", and "舒适 comfortable", which are all words related to positive comments, and 4 of these 6 words that express emotions are adjectives; text 2 also contains 9 words after jieba word segmentation and the removal of stop words. From Figure 10, it is clear that the model assigns a higher significance to specific words "太 too", "次 bad", "很差 very bad", which are all words related to negative comments, and 2 of these 3 words that express emotions are adjectives.

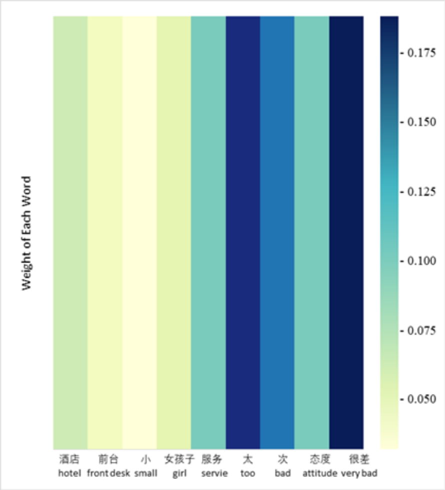


Figure 10 Text 2 word attention weight heatmap

In this paper, we utilise the attention weight visualization technique to examine the function of the attention mechanism in sentiment analysis. The findings demonstrate that the attention mechanism possesses the capacity to precisely identify and prioritize words that exert a substantial influence on the outcome of sentiment analysis. These words are adeptly captured in both positive words in positive comments and negative statements in negative comments. Going further, increasing the weight of adjectives in the attention mechanism is helpful for extracting sentiment features and enhancing the effect of sentiment classification.

6. Conclusion

In this paper, an innovative model is proposed, which combines adjective and conjunction reinforcement strategies to achieve significant improvements in the Chinese sentiment analysis task. This is achieved by integrating local features and contextual semantic information, adopting weighted relative location features and adjective-importance masks, and a flexible adaptive residual structure. The efficacy of this model is evidenced by its ability to reduce noise interference, deepen the learning of location-sensitive grammatical features, improve the performance of the self-attention mechanism, and enable the model to focus on core features more accurately and generate deeper semantic expressions. Consequently, this enhances the accuracy and stability of emotion classification. In the field of Chinese sentiment analysis, a novel approach involves the incorporation of an adjustable residual structure, predicated on a pre-trained language model. This innovation facilitates a nuanced adjustment of the interaction between the residual and input branches, thereby enhancing the model's adaptability. Empirical validation substantiates that by integrating relative location features with the adaptive residual structure, the model's capacity to discern salient text information is notably augmented, thus ensuring a more pronounced focus on the fundamental principles of sentiment analysis.

Despite the optimisation of local feature recognition and contextual semantic information extraction in Chinese text by this model, through an increased emphasis on conjunctions and adjectives, improvements in text emotion classification tasks remain inadequate. In comparison to conventional Chinese emotion classification techniques, which incorporate syntactic structure features, this model employs a novel approach, eschewing the use of such features. In the context of processing Chinese social media data sets, the existing Chinese text feature extraction techniques are inherently constrained. To address this limitation, we propose an integrated approach that combines deep learning methodologies with conventional syntactic dependency analysis techniques. This integration aims to enhance the utilisation of text features that are distinctive to Chinese social media, such as emoticons and Internet slang. This approach is expected to enhance the classification accuracy, thereby leading to substantial improvements in the performance of Chinese social media text emotion classification.

Acknowledgment

The authors would like to convey their appreciation for the support received from Universiti Teknologi MARA in Shah Alam, Selangor, Malaysia, and Xinyu University in Jiangxi Province, China.

References

[1] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press, 2020.

[2] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends® Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, 2008.

[3] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," *arXiv Prepr. cs/0205070*, 2002.

[4] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.

[5] B. Zhao, et al., "Learnable Conjunction Enhanced Model for Chinese Sentiment Analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 14232 LNAI, no. c, pp. 367–381, 2023, doi: 10.1007/978-981-99-6207-5_23.

[6] "word order," [Online]. Available: https://baike.baidu.com/item/词序/10715322.

[7] L. Shuxiang, *Eight Hundred Words in Modern Chinese*. Beijing: The Commercial Press., 1999.

[8] L. Yuming, *Contemporary Chinese Linguistic Studies (1949-2019)*. Beijing: China Social Sciences Press., 2019.

[9] Y. Lu, et al., "Exploring the sentiment strength of user reviews," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6184 LNCS, pp. 471–482, 2010, doi: 10.1007/978-3-642-14246-8_46.

[10] L. Y and B. Y, "Convolutional networks for images,speech,and time series," *Handb. brain theory neural networks*, vol. 10, p. 3361, 1995, doi: 10.1177/016555157900100111.

[11] S. Zhou, et al., "Fuzzy deep belief networks for semi-supervised sentiment classification," *Neurocomputing*, vol. 131, pp. 312–322, 2014, doi: 10.1016/j.neucom.2013.10.011.

[12] Y. Le Cun *et al.*, "Handwritten Digit Recognition: Applications of Neural Network Chips and Automatic Learning," *IEEE Commun. Mag.*, vol. 27, no. 11, pp. 41–46, 1989, doi: 10.1109/35.41400.

[13] L. Barbosa and J. Feng, "Robust sentiment detection on twitter from biased and noisy data," *Coling 2010 - 23rd Int. Conf. Comput. Linguist. Proc. Conf.*, vol. 2, no. August, pp. 36–44, 2010.

[14] X. Glorot, et al., "Domain adaptation for large-scale sentiment classification: A deep learning approach," *Proc. 28th Int. Conf. Mach. Learn. ICML 2011*, no. 1, pp. 513–520, 2011.

[15] S. Zhou, et al., "Active deep networks for semi-supervised sentiment classification," *Coling 2010 - 23rd Int. Conf. Comput. Linguist. Proc. Conf.*, vol. 2, no. August, pp. 1515–1523, 2010.

[16] D. E. Rumelhart, et al., "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986, doi: 10.1038/323533a0.

[17] Y. LeCun, et al., "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.

[18] R. Collobert, et al., "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, 2011.

[19] Y. Kim, "Convolutional Neural Network for Sentence Classification," *Conf. Empir. Methods Nat. Lang. Process.*, pp. 1746–1751, 2014, doi: 10.1109/ICEMI52946.2021.9679581.

[20] N. Kalchbrenner, et al., "A convolutional neural network for modelling sentences," *arXiv Prepr. arXiv1404.2188*, 2014.

[21] W. Yin and H. Schütze, "Multichannel variable-size convolution for sentence classification," *arXiv Prepr. arXiv1603.04513*, 2016.

[22] K. Chen, et al., "Chinese micro-blog sentiment analysis based on multi-channels convolutional neural networks," *J. Comput. Res. Dev*, vol. 55, no. 5, pp. 945–957, 2018.

[23] W. Rong, et al., "Semi-supervised dual recurrent neural network for sentiment analysis," *Proc. - 2013 IEEE 11th Int. Conf. Dependable, Auton. Secur. Comput. DASC 2013*, pp. 438–445, 2013, doi: 10.1109/DASC.2013.103.

[24] R. Kiros *et al.*, "NIPS-2015-skip-thought-vectors-Paper," no. 786, pp. 1–9.

[25] D. Tang, et al., "Document modeling with gated recurrent neural network for sentiment classification," *Conf. Proc. - EMNLP 2015 Conf. Empir. Methods Nat. Lang. Process.*, no. September, pp. 1422–1432, 2015, doi: 10.18653/v1/d15-1167.

[26] C. Zhou, et al., "A C-LSTM neural network for text classification," *arXiv Prepr. arXiv1511.08630*, 2015.

[27] S. Ruder, P. Ghaffari, and J. G. Breslin, "A hierarchical model of reviews for aspect-based sentiment analysis," *arXiv Prepr. arXiv1609.02745*, 2016.

[28] G. Rao, et al., "LSTM with sentence representations for document-level sentiment classification," *Neurocomputing*, vol. 308, pp. 49–57, 2018.

[29] S. Sachin, A. Tripathi, N. Mahajan, S. Aggarwal, and P. Nagrath, "Sentiment analysis using gated recurrent neural networks," *SN Comput. Sci.*, vol. 1, pp. 1–13, 2020.

[30] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.

[31] G. Li, Q. Zheng, L. Zhang, S. Guo, and L. Niu, "Sentiment infomation based model for chinese text sentiment analysis," in *2020 IEEE 3rd international conference on automation, electronics and electrical engineering (AUTEEE)*, 2020, pp. 366–371.

[32] Z. Yang, et al., "Hierarchical Attention Networks," *ArXiv*, pp. 1480–1489, 2016, [Online]. Available: http://arxiv.org/abs/1606.02393.

[33] C. Gan, et al., "Scalable multi-channel dilated CNN–BiLSTM model with attention mechanism for Chinese textual sentiment analysis," *Futur. Gener. Comput. Syst.*, vol. 118, pp. 297–309, 2021, doi: 10.1016/j.future.2021.01.024.

[34] S. Peter, et al., "Self-attention with relative position representations," *arXiv Prepr. arXiv1803.02155*, 2018.

[35] Z. Li, et al., "Chinese text sentiment analysis based on ELMo and Bi-SAN," *Appl. Res. Comput.*, vol. 38, no. 8, pp. 2301–2307, 2021.

[36] M. Li, et al., "Sentiment analysis of Chinese stock reviews based on BERT model," *Appl. Intell.*, vol. 51, pp. 5016–5024, 2021.

[37] H. Jing and C. Yang, "Chinese text sentiment analysis based on transformer model," in *2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI)*, 2022, pp. 185–189.

[38] J. Wang, et al., "Chinese Short-Text Sentiment Prediction: A Study of Progressive Prediction Techniques and Attentional Fine-Tuning," *Futur. Internet*, vol. 15, no. 5, p. 158, 2023.=