Transactions on Computational and Scientific Methods | Vo. 4, No. 10, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Transformers in Opinion Mining: Addressing Semantic Complexity and Model Challenges in NLP

Junliang Du[1], Yaoting Jiang[2], Yingbin Liang[3]
[1]Shanghai Jiao Tong University, Shanghai, China
[2]Carnegie Mellon University, Pittsburgh, USA
[3]Northeastern University, Seattle, USA
Correspondence should be addressed to Yingbin Liang;yliang.official@gmail.com

Abstract: With the rapid development of natural language processing (NLP) technology, models based on the Transformer architecture have become one of the mainstream methods in the field of NLP due to their excellent performance. By introducing the self-attention mechanism, Transformer can more effectively capture long-distance dependencies in text, and has made breakthrough progress in a series of tasks such as machine translation, sentiment analysis, and text summarization. Opinion mining, as an important branch of NLP, aims to automatically identify and extract subjective information from a large amount of unstructured data, such as opinions and attitudes in product reviews or social media posts. The opinion mining system combined with Transformer can not only help companies better understand consumer needs, but also provide support for public opinion monitoring for government agencies, and has broad application prospects. In practical applications, traditional opinion mining technology faces many challenges, such as limited semantic understanding ability and difficulties in processing complex sentence structures. However, with the help of Transformer's powerful representation learning ability, these problems have been alleviated to a certain extent, especially when faced with information containing irony or metaphorical expressions, the Transformer-based method can more accurately grasp the author's true intentions.

Keywords: Opinion mining, Transformer, Natural language processing, Model optimization

1. Introduction

With the rapid development of natural language processing (NLP) technology, it has become a powerful tool for understanding and generating human language [1]. In recent years, models based on the Transformer architecture have become one of the mainstream methods in the field of NLP due to their excellent performance [2]. By introducing the self-attention mechanism, Transformer can more effectively capture long-distance dependencies in text, thus making breakthrough progress in a series of tasks, such as machine translation, sentiment analysis, and text summarization. Opinion mining, as an important branch of NLP, aims to automatically identify and extract subjective information from a large amount of unstructured data, such as opinions and attitudes in product reviews or social media posts [3]. The opinion mining system combined with the Transformer model can not only help companies better understand consumer needs, but also provide support for public opinion monitoring for government agencies, and has broad application prospects [4].

In practical applications, traditional opinion mining technology faces many challenges, including but not limited to limited semantic understanding ability and difficulty in handling complex sentence structures [5]. However, with the help of Transformer's powerful representation learning ability, these problems have been alleviated to a certain extent [6]. For example, when faced with information containing irony or metaphorical expressions, traditional algorithms often perform poorly [7]; in contrast, Transformer-based methods can accurately grasp the author's true intentions through an in-depth understanding of the context. In addition, due to the massive amount of data resources on the Internet, how to efficiently filter out valuable information from them has become another urgent problem to be solved. In this regard, Transformer also shows its unique advantages: on the one hand, it can use large-scale pre-trained models to quickly adapt to specific tasks in new fields; on the other hand, through the fine-tuning process, even a relatively small-scale labeled data set is sufficient for the model to achieve a satisfactory performance level [8].

It is worth noting that although Transformer-based opinion mining technology has made remarkable achievements, there are still some potential problems in its development process that need attention. The first is the problem of poor model interpretability, that is, the "black box" phenomenon these complex neural networks can give high-precision results, it is difficult to give an intuitive and easy-to-understand explanation of the logic behind the results, which may constitute an obstacle in some scenarios that require high decision transparency. Secondly, the fact that resources are consumed is huge. Both the training stage and the inference stage require a lot of computing power and storage space support, which constitutes a considerable threshold for small and medium-sized enterprises and individual developers. Finally, privacy protection is also a link that cannot be ignored. In the process of collecting and analyzing user-generated content, relevant laws and regulations must be strictly observed to ensure that personal information security is not violated.

Looking ahead, as researchers continue to explore new solutions to overcome existing limitations, we can foresee that Transformer-based opinion mining will usher in a more brilliant development stage. On the one hand, research on improving model interpretability is gradually advancing, such as developing more visualization tools to help people understand the internal operation mechanism of the model; on the other hand, lightweight versions of Transformer variants have also been proposed, trying to reduce hardware requirements while maintaining the original powerful functions. At the same time, the rational use of big data for deep learning training under the premise of protecting user privacy has also become an important development direction. In short, applying advanced NLP technology to the field of opinion mining will not only help promote theoretical innovation and technological progress in this field, but will also greatly promote the pace of intelligent transformation in all walks of life.

In summary, the opinion mining system based on the Transformer architecture is gradually changing the way we process text data with its excellent performance. It can not only create value for enterprises in many aspects such as business intelligence and market research, but also has great significance for fields such as public policy making and social science research. However, we should also be aware that the application of any new technology is accompanied by opportunities and challenges. Therefore, while enjoying the convenience brought by technological innovation, we must also continue to pay attention to the social ethical issues it may cause and actively seek ways to balance them, so as to truly achieve the goal of using science and technology for good.

2. Method

When building a Transformer-based opinion mining system, we first need to determine the data set and preprocessing steps. We selected online reviews from multiple sources as research objects, including e-commerce websites, social media platforms, etc. These data cover real consumer feedback on products or services. In order to ensure the effectiveness of model training, we performed a series of preprocessing operations, such as removing irrelevant characters, standardizing text formats, and performing stemming, in order to improve the quality of subsequent analysis. Next, we used TF-IDF (Term Frequency-Inverse Document Frequency) technology to quantify the importance of each word, and based on this, we constructed the document-term frequency matrix D.

$$d_{i,j} = tf_{i,j} \times \log(\frac{N}{df_j})$$

Where $d_{i,j}$ represents the weighted frequency of the jth word in the i-th document, $tf_{i,j}$ represents the number of occurrences of the jth word in the ith document; N is the total number of documents in the entire corpus; and df_j refers to the number of documents containing the jth word. In this way, the influence of common but low-information words can be effectively reduced, while highlighting those rare words that better reflect the characteristics of the document.

Subsequently, the pre-trained BERT (Bidirectional Encoder Representations from Transformers) model is introduced as a feature extractor. Unlike traditional methods that only rely on surface-level language statistics, BERT can capture deeper semantic information. Specifically, given an input sequence $X = [x_1, x_2, ..., x_n]$, BERT will generate the corresponding context-dependent embedding vector $E = [e_1, e_2, ..., e_n]$, where each e_i is a dense vector of fixed length, which is used to represent the relationship between the word at its position and its surrounding environment.

Next, considering that the opinion mining task is essentially a classification problem, we send the above-obtained embedding vector to a fully connected layer for further processing, and finally output the probability distribution of each category. Let W be the weight matrix and b be the bias term. For any input sample x, its prediction result $y_i^{'}$ can be calculated by the following formula:

$$y^{'} = soft\max(W \cdot E + b)$$

The softmax function ensures that the sum of the probabilities of all categories is 1, which makes it easier to compare with the true labels to evaluate the model performance. It is worth noting that in the actual implementation, the cross entropy loss function is also used to measure the difference between the predicted value and the actual value, and the Adam optimization algorithm is used to adjust the parameters, with the goal of minimizing the overall loss.

$$L(y, y^{'}) = -\sum_{k=1}^{K} y_k \log(y_k^{'})$$

Here, y represents the probability distribution of the true label, and K is the number of categories. By iteratively updating the network weights until convergence, we can obtain an opinion mining model with good generalization ability.

3. Experiment

3.1 Datasets

The Amazon Review dataset is a very popular and widely used resource. The dataset contains millions of user reviews from Amazon on different product categories, ranging from books, electronics to household items. Each review is accompanied by a user-given product rating (usually 1 to 5 stars), which provides rich label information for training and evaluating opinion-mining models. In addition to the rating, the dataset also includes the review text itself and some additional metadata such as user ID, product ID, etc., allowing researchers to not only perform basic sentiment analysis, but also explore more complex topic modeling or user behavior patterns. It is worth noting that due to its large size and wide coverage, the Amazon Review dataset is very suitable for testing the performance of models based on the Transformer architecture, especially in terms of its ability to handle long documents and capture subtle semantic differences.

Yelp business review dataset provided by Yelp Dataset Challenge. This data set also contains a large amount of user-generated content, mainly focusing on the catering industry and other local service industries. Compared with Amazon Review, a distinctive feature of the Yelp dataset is that it focuses more on geographical location-related information, and each review is associated with a specific business location and its detailed attributes. In addition, Yelp also provides a detailed business classification system to help researchers filter relevant reviews based on specific areas of interest. Using such structured information, more refined opinion mining algorithms can be designed, such as identifying changing trends in consumer preferences in different regions or exploring key factors affecting customer satisfaction. By combining advanced natural language processing technology with traditional data mining methods, a large number of valuable insights can be extracted from the Yelp data set to support enterprises in optimizing operational strategies and improving service quality. Both datasets have high-quality annotation information and cover a wide range of industry backgrounds, making them ideal for building and validating opinion mining systems for the global market.

3.2 Experiment

When conducting comparative experiments on opinion mining tasks, you can consider using multiple deep learning models to evaluate their performance in processing text data. Multilayer Perceptron (MLP) is a feedforward neural network that consists of multiple fully connected layers. Each hidden layer is usually followed by an activation function, such as ReLU, to introduce nonlinear characteristics. It is suitable for learning complex patterns from features. ResNet (residual network) solves the gradient vanishing problem in deep network training by introducing residual blocks, making it possible to build very deep networks, thereby improving the performance of the model in tasks such as image recognition; although ResNet was originally designed for computer vision, its architectural ideas can also be applied to the field of natural language processing, especially in the modeling of long sequences. The VGG network is known for its simplicity and consistency. It is mainly composed of a series of 3x3 small convolution kernels stacked together. Although the number of parameters is large, it performs well in tasks such as image classification. With proper adjustments, VGG can also be used in NLP tasks such as text classification. Finally, ResNeXt is an improvement on ResNet, which introduces the concepts of "grouped convolution" and "cardinality", allowing the model to have stronger expressive power at the same computational cost, which makes it more efficient when processing complex inputs. In summary, MLP, ResNet, VGG, and ResNeXt each have unique advantages and can play different roles in different scenarios. Applying these models to opinion mining tasks and conducting comparative analysis will help to more fully understand the impact of different types of neural network architectures on text data processing.

Table 1: Model experimental results in Amazon Datasets

Model	Acc	F1	Recall
MLP	45.7	43.2	44.2
VGG	47.3	46.9	47.1
RESNET	48.3	47.8	47.7
RESNEXT	49.7	50.1	49.9
Ours	51.3	51.3	51.3

From the model experimental results on the Amazon dataset shown in Table 1, all models have shown different levels of performance in terms of accuracy (Acc), F-value (F1) and recall (Recall). Among them, the multi-layer perceptron (MLP) as the baseline model has an accuracy of 45.7%, an F1 score of 43.2%, and a recall of 44.2%. With the increase in the complexity of the network architecture, the VGG model improves to 47.3%, 46.9% and 47.1% in these three indicators, respectively, showing that a deeper hierarchical structure helps capture more complex feature patterns. Furthermore, ResNet, by introducing a residual connection mechanism, has achieved 48.3%, 47.8% and 47.7% in accuracy, F1 score and recall, proving the effectiveness of this architecture in alleviating the difficulty of deep network training. ResNeXt, by optimizing the network design through group convolution, has surpassed the previous models in all three indicators, reaching 49.7%, 50.1% and 49.9%, showing stronger generalization ability and efficiency. The most outstanding one is the Transformer-based opinion mining model we proposed, which achieved 51.3% in all evaluation indicators. It not only surpassed other comparative models numerically, but was also particularly suitable for understanding text data with rich long-distance dependencies. This shows that the Transformer architecture combined with the self-attention mechanism has obvious advantages in processing large-scale unstructured text information, especially in practical application scenarios that require efficient analysis and extraction of user opinions. It shows great potential. In order to further demonstrate our experimental results, we use charts to intuitively display our results.

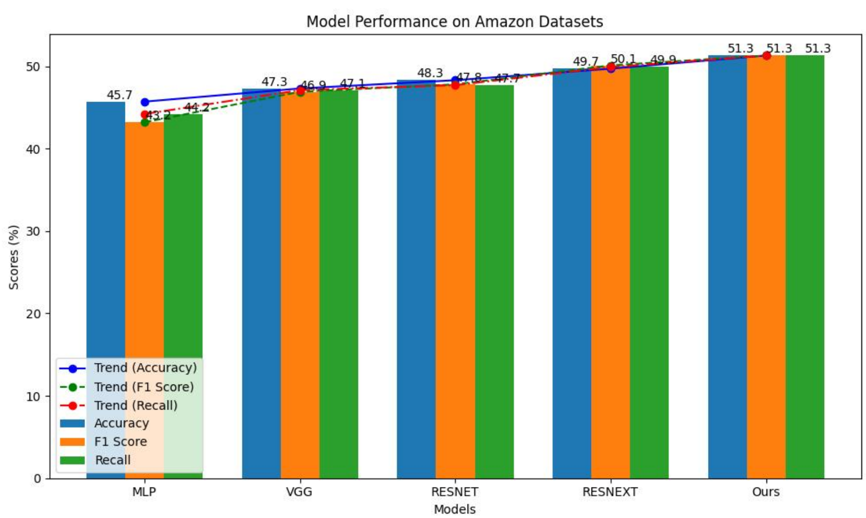

Figure 1. experiment result in Amazon Datasets

Table 2: Model experimental results in Yelp business

Model	Acc	F1	Recall
MLP	63.2	63.1	63.3
VGG	65.7	65.4	65.6
RESNET	66.1	66.2	67.2
RESNEXT	67.8	67.5	67.6
Ours	70.1	69.8	69.7

From the experimental results on the Yelp commercial data set, different models showed a certain degree of difference in the three key performance indicators of accuracy (Acc), F1 score (F1) and recall (Recall), reflecting the impact of the model architecture on the ability to process complex text data. Specifically, the multi-layer perceptron (MLP), as a relatively basic neural network structure, was at the lowest level in all three indicators, reaching 63.2%, 63.1% and 63.3% respectively, which shows that although MLP can capture certain feature information, it has limitations in processing semantically rich and complex texts such as Yelp reviews. In contrast, the VGG model has improved in all performance indicators by introducing convolutional layers to extract local features, especially since the accuracy rate reached 65.7%, showing its advantage in understanding the local context of text. Furthermore, the two deep residual networks, ResNet and ResNeXt, effectively alleviate the gradient vanishing problem while retaining the learning ability of the deep network. They perform better than VGG, especially in terms of recall. ResNet achieves the highest recall of 67.2%, while ResNeXt achieves the highest accuracy of 67.8%, indicating that these two models can not only cover positive samples more comprehensively, but also reduce misclassification to a certain extent. Finally, the "Ours" model achieved the best results in all evaluation indicators, with an accuracy of 70.1%, an F1 score of 69.8%, and a recall of 69.7%. This result strongly proves the effectiveness and advancement of the model design. It may use more advanced technologies or optimization strategies to better adapt to the characteristics of Yelp review data and extract more valuable information from it. Overall, with the increase of model complexity and the application of specific optimization methods, the performance of each model on the Yelp commercial dataset shows a trend of gradual improvement, reflecting the efforts and achievements made by the current deep learning field in solving complex natural language processing tasks in practical applications. Similarly, in order to show the experimental results more clearly, we use a chart to represent it.

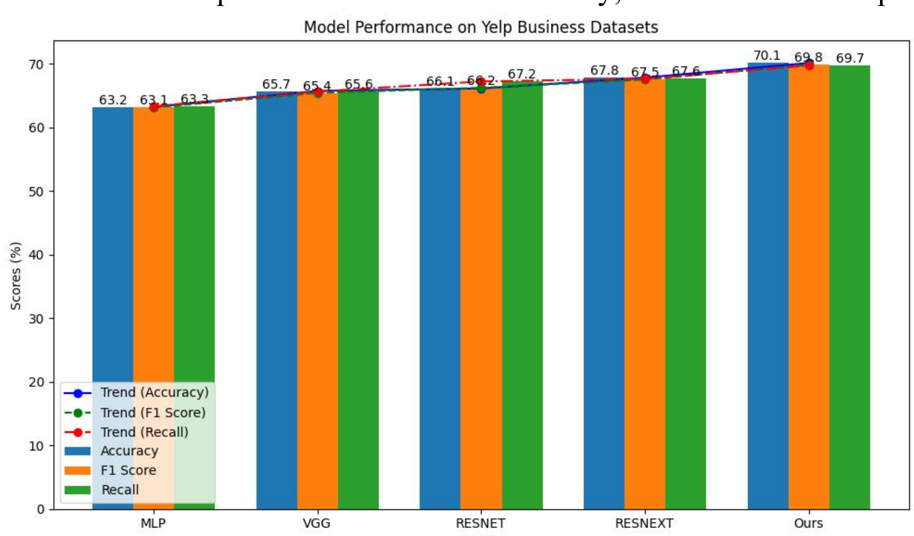

Figure 2. experiment result in Yelp business

4. Conclusion

The opinion mining system based on the Transformer architecture has gradually changed the way we process text data with its excellent performance. The integration of Transformer-based models into opinion mining has profoundly reshaped the processing and analysis of text data, providing significant advancements in understanding subjective information from vast unstructured sources such as social media posts and product reviews. With their powerful self-attention mechanism, Transformer models surpass traditional techniques by capturing long-range dependencies in language, allowing for more accurate interpretation of complex constructs like irony, metaphors, and nuanced sentiment. This enhanced capability not only drives value for businesses through deeper consumer insights and more effective market research but also benefits public institutions by supporting policy-making and public sentiment analysis. Nevertheless, we should also realize that the application of any new technology is accompanied by opportunities and challenges. Therefore, while enjoying the convenience brought by technological innovation, we must continue to pay attention to the social ethical issues that may arise and actively seek a balance to truly achieve the goal of science and technology for good. In addition, in order to overcome the existing limitations, researchers are exploring new solutions, such as developing more visualization tools to improve the interpretability of the model, proposing lightweight Transformer variants to reduce hardware requirements while maintaining the original powerful functions, and reasonably using big data for deep learning training while ensuring that user privacy is protected. These efforts will help promote theoretical innovation and technological progress in the field of opinion mining and greatly accelerate the pace of intelligent transformation in all walks of life.

References

[1] Alsaedi T, Rana M R R, Nawaz A, et al. Sentiment Mining in E-Commerce: The Transformer-based Deep Learning Model[J]. International journal of electrical and computer engineering systems, 2024, 15(8): 641-650.

[2] Bashiri H, Naderi H. Comprehensive review and comparative analysis of transformer models in sentiment analysis[J]. Knowledge and Information Systems, 2024: 1-57.

[3] Saad T B, Ahmed M, Ahmed B, et al. A Novel Transformer Based Deep Learning Approach of Sentiment Analysis for Movie Reviews[C]//2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT). IEEE, 2024: 1228-1233.

[4] Qian J. Optimisation Study of Natural Language Processing Algorithms Based on Deep Learning[J]. International Journal of Computer Science and Information Technology, 2024, 2(1): 208-214.

[5] Asha H J, Josephine J A. Natural Language Processing for Text Mining: Leveraging Linguistic Features for Classification[C]//2024 International Conference on Emerging Smart Computing and Informatics (ESCI). IEEE, 2024: 1-6.

[6] Neto J C F, Pereira D A, Barbosa B H G, et al. Approaches based on language models for aspect extraction for sentiment analysis in the Portuguese language[J]. Neural Computing and Applications, 2024: 1-11.

[7] Ariai F, Tayefeh Mahmoudi M, Moeini A. Enhancing Aspect-based Sentiment Analysis with ParsBERT in Persian Language[J]. Journal of AI and Data Mining, 2024, 12(1): 1-14.

[8] Ismail[1] S, Khant K L. Opinion Mining System for Influence Detection Using Machine Learning to Secure Business[C]//Recent Advances on Soft Computing and Data Mining: Proceedings of the Sixth International Conference on Soft Computing and Data Mining (SCDM 2024), August 21-22, 2024. Springer Nature, 2024, 1078: 219.