Dean&Francis
ISSN 2959-6157

A review of sentiment analysis research based on BERT and its improved models

Jingxuan Chen

International Exchange College, Guangzhou Maritime University, Guangzhou, 510725, China
E-mail: 2165569892@qq.com

Abstract:

In recent years, the rapid evolution of social media and online reviews has exposed limitations in traditional sentiment analysis methods, which rely on sentiment lexicons and machine learning classifiers, particularly in terms of contextual modeling and generalization capabilities. In past few years, Deep learning, particularly convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variants, has demonstrated superior feature learning capabilities in sentiment recognition. However, these methods are still limited by the long-distance dependency modeling and the complexity of Chinese semantics. The introduction of BERT has opened a new chapter for pre-trained language models in sentiment analysis. The model's semantic understanding capabilities have been significantly enhanced through a bidirectional Transformer architecture and large-scale corpus pre-training. Subsequently, not only ALBERT, but also improved models such as RoBERTa played a significant role in English sentiment analysis tasks after fine-tuning according to task requirements. A series of improved models such as RoBERTa-wwm-ext, ERNIE, ALBERT-zh, and MacBERT continued to refresh performance records in Chinese sentiment analysis tasks. This paper systematically reviews the research progress of sentiment analysis based on BERT and its improved models in recent years. It focuses on comparing the performance of different models in text sentiment classification, implicit sentiment recognition, and fine-grained sentiment analysis, and reveals their advantages and challenges in semantic modeling, cross-domain transfer, and multi-granularity information fusion. Finally, this paper discusses the current bottlenecks faced by the research and looks forward to the future development direction of sentiment analysis.

Keywords: BERT, Pre-trained Language Model, Sentiment Analysis, Natural Language Processing, Model Improvement

1. Introduction

With the rapid development of the Internet and social media, user-generated text data has exploded. Sentiment Analysis, as an important research direction of Natural Language Processing (NLP), Widely applied in public sentiment monitoring, product evaluation analysis, financial market forecasting, and intelligent human-computer interaction [1,2].Its core goal is to use computer means to identify and understand the emotions or opinions contained in the text, thereby providing effective support for upper-level applications. Especially in the Chinese context, Due to the polysemy, metaphorical nature and the complexity of context dependence of language, the task of sentiment analysis is more challenging and can better reflect the strengths and weaknesses of the algorithm model.
In the past decade, sentiment analysis research has gone through three major stages of development. Early methods were based on sentiment dictionaries and rule matching, and later gradually developed into classification methods based on traditional machine learning.
Such methods include support vector machines (SVM) and Naive Bayes [3,4]. These methods rely on artificially constructed features, are difficult to effectively capture contextual relationships, and have insufficient generalization capabilities.With the development of deep learning, methods such as convolutional neural networks (CNN), recurrent neural networks (RNN), long short-term memory networks (LSTM), and attention mechanisms (Attention) have been widely applied in sentiment analysis.The automatic feature extraction capabilities of the models have been significantly improved [5,6,7,8]. However, these methods still have bottlenecks in long-distance dependency modeling, semantic expression refinement, and cross-domain migration.
In 2018, the BERT (Bidirectional Encoder Representations from Transformers) model was proposed by Google, which means that sentiment analysis has entered a new stage[9].By introducing the bidirectional Transformer architecture and the Masked Language Model pre-training task, BERT has achieved deep modeling of contextual information and has set new performance records in multiple NLP benchmark tasks.Subsequently, researchers successively proposed a series of improved models including RoBERTa, ERNIE, ALBERT, MacBERT, and ELECTRA, continuously optimizing training strategies, structural design, and Chinese semantic feature modeling [10,11,12,13,14].The application of these models has greatly promoted the development of Chinese sentiment analysis. They have not only achieved excellent results in text sentiment classification tasks, but also demonstrated strong potential in more complex scenarios such as implicit sentiment recognition, fine-grained sentiment analysis, and conversational sentiment recognition.
Nevertheless, research on sentiment analysis based on BERT still faces many challenges. For instance, the model's generalization ability in cross-domain applications is insufficient, it heavily relies on large-scale labeled data and computing resources, the model lacks interpretability, and there are still deficiencies in multi-granularity semantic fusion.How to enhance efficiency, reduce dependencies, and improve the interpretability of the model while maintaining performance is an important research direction for the future.
The main work and contributions of this paper are:(1)Systematically reviewing the research progress in sentiment analysis based on BERT and its improved models;(2) Summarizing representative approaches from the perspectives of model structure optimization, pre-training strategy improvement, and Chinese semantic modeling;(3)Comparing the performance of different models in sentiment analysis tasks, summarizing their strengths and weaknesses;(4)Discussing the challenges facing current research and prospecting future development directions, including lightweight models, multimodal fusion, and interpretability research.

2. Review of sentiment analysis

Sentiment analysis, a key task in natural language processing, has evolved through three major stages: from traditional rule-based methods to deep learning approaches, and finally to pre-trained language model-driven approaches. Each stage has advanced research to some extent, but has also exposed new limitations.

2.1 Traditional method

Early sentiment analysis primarily relied on sentiment dictionaries and rule matching [15]. A typical approach involved constructing sentiment dictionaries (e.g., positive and negative lexicon lists) and classifying sentiments through statistical word frequency, sentiment polarity strength, and syntactic rules. While this method is simple to implement, it heavily depends on manual construction and struggles to handle complex contexts and implicit emotional expressions.Subsequently, researchers began to incorporate traditional machine learning methods, such as Naive Bayes, SVM, and Logistic Regression [1, 4]. These methods typically classify by combining artificially designed features (such as TF-IDF, n-gram, sentiment word distribution), and their generalization ability is improved compared to rule-based methods. However, due to the inability to capture long-range dependencies and semantic hierarchical information, these methods perform poorly in

ISSN 2959-6157

complex text sentiment recognition.

2.2 Deep learning methods

With the rise of deep learning in NLP, researchers have introduced convolutional neural networks (CNNs) and recurrent neural networks (RNNs, LSTMs, GRUs) into sentiment analysis tasks. CNNs excel at capturing local contextual features and are well-suited for short text classification [16]; LSTMs and GRUs, meanwhile, can mitigate long-range dependency issues to a certain extent [7]. Simultaneously, the introduction of attention mechanisms enables the model to dynamically focus on sentiment-related key segments within the text, further enhancing classification performance [8].The methods adopted in this stage significantly reduced the reliance on manual features and enhanced the model's ability to automatically learn features. However, due to the fact that deep learning models still rely on task-specific data for training, their cross-domain transfer capabilities are insufficient. Moreover, considering the complex semantic structure of Chinese, there are still bottlenecks in the model's generalization ability and semantic fine-grained modeling.

2.3 Pre-trained language model method

In 2018, Google proposed the BERT (Bidirectional Encoder Representations from Transformers) model, marking a new stage in the research of sentiment analysis [9]. BERT, based on a bidirectional Transformer encoder, uses two pre-training tasks, Masked Language Model (MLM) and Next Sentence Prediction (NSP), to learn universal language representations on large-scale unlabeled corpora. It then achieves transferability through fine-tuning on downstream tasks. BERT significantly improved the performance of sentiment analysis and quickly became a mainstream method.Based on BERT, a large number of improved models have been proposed: RoBERTa: optimizes training strategy and discards the next sentence prediction task and extends training on larger corpora, which makes it more robust than BERT [10]; ALBERT: reduces model size through parameter sharing and decomposition embedding [12]; ELECTRA: proposes the replacement word detection (RTD) task, which greatly improves training efficiency [14]; ERNIE: combines knowledge enhancement pre-training to be more suitable for Chinese tasks [11]; MacBERT, ChineseBERT: improve Chinese characteristics (whole word masking, glyph and pinyin fusion) [13,17].These improved models have continuously demonstrated outstanding performance in tasks such as text sentiment classification, implicit emotion recognition, fine-grained sentiment analysis, and dialogue emotion recognition, showcasing strong generalization ability and robustness. However, they also face some challenges, including insufficient cross-domain adaptability, reliance on large-scale labeled data, lack of model interpretability, and high computational costs, which require further research and optimization.

3. Review of bert model

Since the introduction of BERT, sentiment analysis research based on pre-trained language models has entered a new phase [9]. BERT achieves deep semantic modeling through its bidirectional Transformer encoder and Masked Language Model (MLM) task, achieving breakthrough progress on multiple NLP benchmark tasks [19]. Subsequently, researchers have proposed numerous improved models addressing BERT's limitations in training efficiency, model scale, and cross-language adaptation. These models not only excel in general NLP tasks but also demonstrate significant advantages in sentiment analysis. This paper systematically introduces BERT and its variants from three perspectives: training strategy optimization, model structure lightweighting, and Chinese adaptation with knowledge augmentation.

4.1 Original bert model

BERT was the first framework to apply the bidirectional Transformer encoder to pre-trained language modeling [9, 19]. Its innovations mainly lie in two aspects:
(1) MLM task: By randomly masking some words, it forces the model to simultaneously utilize the left and right context to restore the target word, thereby enhancing the semantic understanding ability;(2)NSP task: By predicting whether two sentences are consecutive, it enhances the model's ability to model the inter-sentence relationship.
In sentiment analysis, BERT adapts to downstream classification tasks through fine-tuning, significantly improving the accuracy of text sentiment classification compared to methods such as CNN and LSTM[20]. BERT is particularly effective in long-text sentiment analysis and multi-category classification tasks. However, BERT has shortcomings including high training cost, limited NSP performance, and insufficient adaptability to Chinese.

4.2 Optimization of training strategies for improved models

4.2.1 RoBERTa

RoBERTa systematically optimizes BERT's pre-training strategy[10]. The main improvements include: 1. Removing the NSP task; 2. Using a larger corpus and longer training time; 3. Introducing a dynamic mask strategy.
These improvements have enabled RoBERTa to achieve

better performance than BERT in text sentiment classification tasks, especially in large-scale sentiment corpora (such as Weibo comments and product reviews). However, its training cost has further increased, and it is highly dependent on computing resources.

4.2.2 ELECTRA

ELECTRA proposes Replaced Token Detection (RTD) as a novel pre-training task, demonstrating higher training efficiency compared to MLM [14].In the sentiment analysis task, ELECTRA can still achieve performance close to or even surpassing that of BERT when using a relatively small amount of data, making it highly suitable for sentiment classification of small-scale domain data. However, its limitation lies in the more complex model structure and increased difficulty in tuning.

4.3 Lightweight model structure improvement model

4.3.1 ALBERT

ALBERT significantly reduces the model size and improves training and inference efficiency through cross-layer parameter sharing and embedding decomposition[12]. In sentiment analysis tasks, ALBERT significantly reduces memory consumption while maintaining accuracy, making it suitable for deployment in resource-constrained environments such as mobile applications.However, due to parameter sharing, the expressive power of the model slightly decreases in some tasks.

4.4 Chinese adaptation and knowledge enhancement-based improved model

4.4.1 ERNIE

ERNIE proposed knowledge-enhanced Masking based on BERT's MLM, which is to mask entities and phrases as a whole during pre-training, so that the model can better learn structured knowledge[11].ERNIE performs exceptionally well in implicit sentiment recognition in Chinese and cross-sentence emotion modeling, but it relies on the quality and coverage of the knowledge base.

4.4.2 MacBERT

To address the problem that the Mask token in BERT is too different from the original word, MacBERT proposed a replacement Mask (MLM as correction) strategy, which replaces the original word with a synonym for prediction, which is more in line with the Chinese context[13].In fine-grained sentiment analysis tasks such as e-commerce reviews and news comments, MacBERT outperforms BERT and RoBERTa, demonstrating a particularly significant advantage when handling polysemous words.

4.4.3 ChineseBERT

ChineseBERT introduces glyph and phonetic information based on BERT to form a dual-channel feature representation, thereby better modeling the multi-granularity information of the Chinese language[17].In implicit sentiment recognition and cross-domain sentiment classification tasks, ChineseBERT demonstrates superior performance compared to traditional BERT. Its primary challenge lies in the added complexity introduced by additional character and pinyin features during training and inference.

4.5 Summary

Overall, the performance of BERT and its improved models in sentiment analysis shows the following trends:(1) Enhanced semantic modeling ability: The improved models better capture context and long-range dependencies (such as the decoupled attention mechanism in DeBERTa [21]).(2)Limited cross-domain transfer ability: RoBERTa and MacBERT perform well in the same domain data, but still rely on large-scale fine-tuning data for cross-domain sentiment analysis.(3)Multi-granularity information fusion: ERNIE and ChineseBERT effectively enhance Chinese semantic understanding by introducing knowledge and linguistic features, especially showing outstanding performance in implicit and fine-grained sentiment analysis.

These models have continuously promoted the research progress of sentiment analysis, but they have also exposed challenges such as high training cost, insufficient cross-domain generalization, and lack of interpretability. Future research needs to continue to explore the directions of lightweight, interpretable, and multimodal fusion.

5. Comparative analysis of different bert models

5.1 Table comparison of different bert models

In order to comprehensively compare the performance of BERT and its improved models in Chinese sentiment analysis, this paper has collated the experimental results of different models in typical tasks (such as text sentiment classification, microblog sentiment recognition, implicit sentiment recognition, and fine-grained sentiment analysis).This table summarizes the experimental data and performance characteristics of each model, covering the main datasets and evaluation metrics. This comparison not only reveals the strengths and weaknesses of different models, but also provides a basis for further analysis of their applicable scenarios and development trends.

ISSN 2959-6157

Table 1.Comparison of Performance among Different BERT Models

Article	Task Type	Dataset	Model	Accuracy	Advantages
Devlin et al.[9]	Text Sentiment Classification	ChnSentiCorp	BERT	95.4 (Acc)	Baseline model, bidirectional context modeling
Liu et al.,[10]	Text Sentiment Classification	ChnSentiCorp	RoBERTa	95.6 (Acc)	Stronger pretraining, better generalization
Cui et al.,[13]	Text Sentiment Classification	ChnSentiCorp	MacBERT	95.6(base), 95.9 (large)	MLM as correction, better suited for Chinese
Sun et al., [17]	Text Sentiment Classification	ChnSentiCorp	ChineseBERT	95.7(base), 95.9 (large)	Glyph + Pinyin fusion, multi-granularity modeling
Sun et al.,[11]	Text Sentiment Classification	ChnSentiCorp		95.5(Acc)	Knowledge-enhanced masking, better entity-level understanding
Sun et al., [11]	Weibo Sentiment Classification	Weibo Sentiment	BERT	94.3(Acc)	Baseline
Article	Task Type	Dataset	Model	Accuracy	Advantages
Sun et al., [11]	Weibo Sentiment Classification	Weibo Sentiment	ERNIE	95.4(Acc)	Knowledge graph enhanced, better implicit sentiment
Sun et al., [17]	Weibo Sentiment Classification	Weibo Sentiment	ChineseBERT	Unpublished	Reported for NER, sentiment classification not released
Xu et al., [20]	Implicit Sentiment Recognition	COTE-BD		Unpublished	Performs well on explicit sentiment, weak on implicit
Xu et al., [20]	Implicit Sentiment Recognition	COTE-BD	ERNIE	Unpublished	Knowledge-enhanced, better entity-driven reasoning
Sun et al., [11]	Implicit Sentiment Recognition	COTE Series	ChineseBERT	Unpublished	Glyph + Pinyin help capture implicit sentiment
Sun et al., [11]	Aspect-based Sentiment Analysis	Online Shopping Review	BERT	Unpublished	Baseline
Sun et al., [11]	Aspect-based Sentiment Analysis	Online Shopping Review	RoBERTa	Unpublished	Better generalization
Cui et al., [13]	Aspect-based Sentiment Analysis	Online Shopping Review	MacBERT	Unpublished	Handles multiple aspects more naturally
He et al., [21]	Aspect-based Sentiment Analysis	SemEval Chinese ABSA	DeBERTa	Unpublished	Disentangled attention, stronger long-dependency modeling
Clark et al., [14]	Text Sentiment Classification	Chinese reviews (tourism, product datasets)	ELECTRA	Unpublished	Efficient RTD pretraining, strong performance on small datasets
Lan et al., [12]	Text Sentiment Classification	ChnSentiCorp / small Chinese review sets	ALBERT	Unpublished	Parameter sharing, lightweight, suitable for deployment

5.2 Analysis of the comparison table

The comparison results reveal that BERT and its variants exhibit distinct phased characteristics and differentiated advantages in Chinese sentiment analysis. As a benchmark model, BERT demonstrates robust semantic modeling capabilities across tasks such as ChnSentiCorp and Weibo sentiment classification, significantly outperforming traditional CNN and LSTM approaches. This performance establishes pre-trained models as the foundational standard in sentiment analysis.However, BERT still has shortcomings in terms of its adaptability to the Chinese language, training efficiency, and implicit emotion recognition, which has prompted subsequent researchers to propose a series of improved models.

In terms of training strategy optimization, RoBERTa outperforms BERT in text sentiment classification tasks by removing the NSP task, introducing dynamic masking, and training with a larger-scale corpus, demonstrating stronger generalization ability.ELECTRA significantly improves training efficiency through replacement word detection (RTD), and can achieve results close to or even surpass BERT even in small-scale domain data, showing its potential in low-resource scenarios[14].In comparison, RoBERTa is more suitable for large-scale pre-training scenarios, while ELECTRA is more practical.

In terms of model lightweighting, ALBERT significantly reduces model size through parameter sharing and embedding decomposition. This approach maintains performance while enhancing training and inference efficiency, making it suitable for resource-constrained applications such as mobile devices and online services [12].However, the decline in expression ability brought about by parameter sharing also indicates that there is still a need to balance between lightweighting and performance.

The improved model that takes into account the characteristics of the Chinese language has shown particularly outstanding performance. MacBERT introduces substitutional masking in the MLM task, which is more in line with the Chinese context and outperforms BERT and RoBERTa in fine-grained sentiment analysis (such as e-commerce and news comments) [13].ChineseBERT incorporates character shape and pinyin information, enabling it to model multi-granularity features of Chinese. It demonstrates advantages in implicit sentiment recognition and cross-domain tasks [17].ERNIE masks entities and phrases as a whole through knowledge-enhanced Masking, making the model outperform the baseline model in processing implicit emotions and cross-sentence sentiments, demonstrating the importance of external knowledge in sentiment analysis.

In addition, DeBERTa enhances the long-term dependency modeling capability through the decoupled attention mechanism in fine-grained sentiment analysis (ABSA), making it more suitable for handling complex multi-faceted sentiment tasks[21].This indicates that in tasks such as ABSA that have higher requirements for context dependence, structural optimization models have unique advantages.

Overall, the comparison in this table reveals a clear evolutionary trajectory: from the foundation laid by BERT, to the optimization of training strategies and efficiency by RoBERTa and ELECTRA, to the exploration of lightweightness by ALBERT, and finally to the in-depth transformation of MacBERT, ERNIE, and ChineseBERT to target Chinese characteristics.These models have their own advantages in different tasks, reflecting that the development of Chinese sentiment analysis has shifted from a single semantic modeling approach to a comprehensive balance of efficiency, adaptability and multi-granularity integration. Although the overall performance has reached a relatively high level, there is still room for improvement in cross-domain transfer, low-resource learning and model interpretability, which also provides guidance for future research directions.

6. Challenges and future directions

Although BERT and its improved models have made significant progress in Chinese sentiment analysis, there are still some challenges that need to be addressed at the application and research levels. Firstly, the insufficiency of cross-domain adaptability is a common problem: the performance of the model significantly declines when it is transferred between different scenarios such as e-commerce, finance, and public opinion monitoring, indicating that domain-specific pre-training and cross-domain transfer learning still require in-depth research.Secondly, the reliance on large-scale annotated data remains significant. This is particularly true for implicit emotion recognition and fine-grained sentiment analysis, where high-quality data is scarce, limiting model effectiveness. Few-shot learning, zero-shot learning, and hinted learning methods are expected to alleviate this bottleneck under low-resource conditions.Meanwhile, the large scale of the model and its low inference efficiency are also major obstacles hindering its practical application.Although improved models such as RoBERTa and DeBERTa have excellent performance, they are difficult to deploy on mobile devices or in real-time systems. Therefore, lightweight methods such as distillation, pruning, and quantization will be the focus of future research.On the other hand, the lack of model interpretability limits its application in high-risk fields such as financial risk control and social governance.

ISSN 2959-6157

In the future, measures such as attention visualization and causal reasoning can be combined to enhance transparency and controllability.Finally, with the rise of social media and multimodal content, multimodal sentiment analysis has emerged as a new trend. Effectively integrating textual, visual, and auditory information represents a crucial direction for enhancing model robustness.

In conclusion, future research should focus on areas such as cross-domain transfer, low-resource learning, lightweight deployment, enhanced interpretability, and multimodal integration, in order to promote the application of Chinese sentiment analysis in more complex and realistic scenarios.

7. Conclusion

This article reviews the research progress of BERT and its improved models in Chinese sentiment analysis. From the perspective of development history, sentiment analysis methods have gone through an early stage based on dictionaries and traditional machine learning, then entered the period dominated by deep learning, and the emergence of pre-trained language models has brought about new breakthroughs.At the specific task level, BERT and its variants have demonstrated significant advantages in text sentiment classification, implicit sentiment recognition, and fine-grained sentiment analysis.Improved models such as ChineseBERT, MacBERT, ERNIE, and DeBERTa better adapt to issues such as word granularity, polysemy, and long dependencies in the Chinese context, promoting the development of Chinese sentiment computing.However, existing models still face challenges such as insufficient cross-domain generalization, strong reliance on labeled data, large model size, and limited interpretability. Future research can focus on lightweight deployment, small- and zero-shot learning, multimodal fusion, and enhanced interpretability to improve the adaptability and credibility of models in practical applications.Overall, sentiment analysis based on BERT is not only a significant research topic in natural language processing but also provides a solid foundation for applications such as social media analysis, public opinion monitoring, and intelligent customer service.

References

[1] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing* (pp. 79–86). Association for Computational Linguistics. https://doi.org/10.3115/1118693.1118704

[2] Liddy, E. D. (2001). Natural language processing. In M. A. Drake (Ed.), *Encyclopedia of library and information science* (Vol. 69, pp. 212–232). Marcel Dekker.

[3] Noble W S. What is a support vector machine?[J]. Nature Biotechnology, 2006, 24(12): 1565-1567.

[4] McCallum, A., & Nigam, K. (1998). A comparison of event models for naive Bayes text classification. In *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization* (Vol. 752, pp. 41–48). AAAI Press.

[5]LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[6]Elman J L. Finding structure in time[J]. Cognitive Science, 1990, 14(2): 179-211.

[7] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[8]Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations (ICLR 2015)*. arXiv. https://arxiv.org/abs/1409.0473

[9]Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.48550/arXiv.1810.04805

[10]Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., … & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

[11]Sun, Y., Wang, S., Li, Y., Feng, S., Chen, X., Zhang, H., … & Tian, H. (2019). ERNIE: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*.

[12]Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). ALBERT: A lite BERT for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

[13]Cui, Y., Che, W., Liu, T., Qin, B., & Yang, Z. (2020). Revisiting pre-trained models for Chinese natural language processing. *Proceedings of the 2020 Conference on Empirical Methods inNaturalLanguageProcessingFindings* ,657–668. https://doi.org/10.18653/v1/2020.findings-emnlp.58

[14]Clark, K., Luong, M.-T., Le, Q. V., & Manning, C. D. (2020). ELECTRA: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*.

[15]Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 168–177. https://doi.org/10.1145/1014052.1014073

[16]Kim, Y. (2014). Convolutional neural networks for sentence classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–

1751. https://doi.org/10.3115/v1/D14-1181
[17]Sun, Z., Wang, S., Li, Y., Feng, S., Chen, X., Zhang, H., … & Tian, H. (2021). ChineseBERT: Chinese pretraining enhanced by glyph and pinyin information. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2065–2075. https://doi.org/10.18653/v1/2021.acl-long.164
[18]Zhang, Y., Chen, H., Li, Y., & Deng, Y. (2022). A review of Chinese sentiment analysis: Subjects, methods, and trends. *Information Processing & Management, 59*(2), 102805. https://doi.org/10.1016/j.ipm.2021.102805
[19]Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 5998–6008. https://doi.org/10.48550/arXiv.1706.03762
[20]Xu, H., Li, Y., Xia, R., & Huang, W. (2020). *COTE: A benchmark for Chinese implicit sentiment analysis*.
[21]He, P., Liu, X., Gao, J., & Chen, W. (2021). DeBERTa: Decoding-enhanced BERT with disentangled attention. *International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.2006.03654