Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 13, No. 3, 2025

Deep Learning-Driven Text Sentiment Analysis: Research Progress, Challenges, and Future Trends in the Past Five Years

Sirui Song *

School of Telecommunication Engineering, Xidian University, Xi'an, Shaanxi, 710071, China
*** Corresponding author Email:** ssr263066659@163.com

Abstract: Text sentiment analysis has developed from a single text classification task into a complex multimodal fusion system that can achieve cross-scenario sentiment understanding with the help of deep learning technology. In this article, we will review the cutting-edge studies from 2020 to 2025 on domain adaptation, low-resource learning, complex semantic modeling, multimodal fusion, and ethical fairness based on review papers in leading journals (e. g., IEEE TPAMI, Nature) and at top conferences (e. g., ACL, NeurIPS). It has been found that the combination of domain knowledge graphs reduces the performance drop caused by cross-domains by 60%, causal reasoning increases the accuracy of mixed sentiment disentanglement by 12.4%, and multimodal dynamic alignment improves precision by 8% in conflicting scenarios. This article shows that technological development has a transversal and fused trend of "knowledge promotion-causal modelling-modal collaboration" and lightweight mechanism as well as fair mechanism are important in practical engineering implementation.

Keywords: Text Sentiment Analysis; Deep Learning; Domain Adaptation; Multimodal Fusion.

1. Introduction

Text sentiment analysis aims to reveal public opinion and gauge user feedback sentiment from the words they use and draws decision insights through analysing emotional trends; and the development of the relevant technologies depends on the tremendous advances of deep learning technologies. In the past few years, pre-trained language models such as BERT have reached more than 90 percent accuracy in sentiment classification [1]. However, three major difficulties remain as practical applications move forward: Firstly, variations in cross-domain training data distributions degrade the performance by more than 10%; secondly, the accuracy of distinguishing semantics such as irony and metaphors falls short of human-level capacity by 23%, and thirdly, divergences of multiple modes (text + image + audio) reduce their accuracy by 15-20% respectively [2-4].

Since 2020, with intensive exploration conducted by academia, the EAAF has quantified the reliability of domain knowledge, cutting the cross-domain performance gap between medical and e-commerce areas down from 10% to 4.2%, whereas the accuracy of irony recognition with structured sentiment analysis model exceeds 75% [5,6]. In an environment where 'text is happy and images cry', CAMF enhances accuracy up to 8%, but there remains room for further improving the domain specific semantic rules, tackling the limited data augmentation efficiency in low resource scenarios and simplifying the modelization of the multivariate relationship [7]. Based on previous achievements, this paper systematically analyses and sorts out existing research technology statuses and breakthrough ways to solve current problems, offering necessary theoretical evidence for academic research.

2. Research Status

Domain adaptation focuses on resolving model generalization bottlenecks by narrowing down semantic distribution differences between the source and target domains. Traditional methods (domain adversarial learning) are only concerned with feature distributions, but more recently researchers have started focusing on "knowledge enhancement + dynamic adaptation" way of thinking [8]. In the 2024 ACL work, legal term knowledge graphs were combined with pretrained models to find domain-specific phrases like "valid defence", using entity linking to enhance the identification of related concepts with improved legal text sentiment classification performances [9]. The EAAF framework introduced in IEEE TPAMI 2024 successfully quantifies "evidential uncertainty" that discriminates between the credibility of different sources of knowledge from the source domain, which improves performance by about 60% in multi-source unsupervised settings [5]. Diffusion models serve as an important data augmentation tool in low-resource scenario as well. The proposed diffusion-CLS at EMNLP 2024 generates domain-adapted texts through a conditional diffusion process, increasing the F1 score on the Arabic low-resource reviews dataset by 9.7% and achieving 89.2% emotional consistency [10]. The integration of meta-learning and prompting models improves performance in few-shot cross-domain sentiment analysis tasks [11]. Significant gains are shown in terms of differences in domain terminologies in big terms such as technologies and agro-based fields.

Breakthroughs on Complex Semantic Understanding Relies on Moving From "Surface Feature Matching" to "Deep Logical Modelling" Irony Recognition Latent Dependency Graph Parsing model captures "conflict between the value judgment and facts" (for example, the evaluation of "genius" and the negative context) from ACL 2024, improving the accuracy from 68.5% to 75.3%, but still needs to adopt manually annotated local structure [6].Mixed sentiment disentanglement took advantage of causal reasoning, as causal graphs could distinguish the "direct emotional causes," such as "the high price" inferring bad cost performance, from

"indirect associations," such as brand effect, for instance, leading to improved text separation between mixed sentiment like "expensive but worthwhile" and 12.4% higher disentanglement accuracy on text such as "expensive but worthwhile" [12]. The Hierarchical model (HierSenti) can achieve 81.6% on mixed sentiment recognition from long movie reviews using "word-sentence-discourse" three level attention, which is higher than traditional models by 7.2% [13]. Furthermore, a new finding was that big models with more than 100B parameters are worse performing on some complex semantic tasks. Such a result can be seen in irony recognition. A larger model with more than 100B parameters leads to lower accuracy in irony recognition (5.8%) compared to a small model of 7B parameters, which implies that the scale of semantic understanding is not always tied to the magnitude of parameters [3].

Multimodal fusion's key problem is handling and integrating different types of data heterogeneity and dynamic correlations across text, image and audio. In order to capture the temporal correlation such as "emotional enhancement of the text 'funny' two seconds after laughter appears" the CAMF (Cross-Attention Multi-layer Fusion Module) proposed in RCBEVDet at CVPR 2024 can attain the multimodal conflicting scenario improvement up to 8%. MJP-FRM-FDs in ECCV 2024 solves the problem of missing modality with the generative modeling by making up for the missing modality's information as text "angry" generate corresponding expression features, maintaining a high accuracy retention rate (92%) of 15% higher than the traditional method for missing modalities [14]. Besides, the light-weight technology in practical engineering solves this issue through Knowledge distillation-based lightweight methods reduce multimodal model parameters while maintaining acceptable accuracy, with inference speed improvements; while Causal structure learning explicitly models the relationships between sentiment factors (e.g., aspects and emotions), supporting more interpretable sentiment analysis.[15,16].

Ethics and interpretability requirements have been mandatory criteria for any type of new technology set up. According to NeurIPS 2024, there is an 18% higher negative prediction rate of women-related texts as "female programmer", as well as a 9% lower positive recognition rate towards the review of rural areas compared to cities [17]. In reply to this issue, using MaxMin-RLHF framework (ICML 2024), the overall performance has kept the same status, but the minority group sentiment recognition accuracy can be improved by 33%, which is accomplished by utilizing the diversified preference distribution to reach a good balance between different sentiments; concerning interpretability, Hierarchical attention mechanisms enhance the modeling of long-text sentiment by focusing on key words and sentences, improving the interpretability of emotion decision-making [13,18]. In the domain of privacy protection, through utilization of federated learning, Federated learning frameworks enable cross-device collaborative training for sentiment analysis while protecting data privacy, with reduced risks of information leakage [19].

Based on the trend in technology, we can see cross-fusion such as domain adaptation and knowledge graph (e. g., legal term association rules), complex semantic modeling introduces causal reasoning (e. g., mixed sentiment disentanglement), and multimodal fusion works well with generative modeling (e. g., modal completion); and lightweight techniques, for example, structured pruning, along with domain adaptation methods such as adversarial training, will help shift sentiment analysis from the labs into real-world applications [20,21].

3. Existing Challenges

The current technological bottlenecks in text sentiment analysis essentially reflect the mismatch between "model capabilities and real-world scenario requirements," specifically manifested in four sets of core contradictions:

The deep-seated obstacles in domain generalization lie not only in feature distribution differences but also in the difficulty of transferring implicit domain knowledge rules. Although the EAAF framework compresses cross-domain performance loss to 4.2% through evidential learning, in highly professional fields such as law and medicine, the emotional mapping of terms (e.g., "defense" has no direct emotional tendency in law and needs to be combined with contextual roles) still relies on manual intervention, and existing models lack the ability to autonomously mine implicit domain rules [5]. In low-resource scenarios, although augmented data generated by diffusion models can improve performance by 8%-10%, in semantically sparse fields (e.g., ancient Chinese reviews), the emotional consistency between generated texts and real corpora is still below 70%, exposing the shortcoming of semantic fidelity in data augmentation [10].

The cognitive gap in complex semantics is reflected in the model's defective modeling of the nonlinear relationship between "context and emotion." In irony recognition, even though the structured model achieves an accuracy of 75.3%, its performance drops by 15%-20% when facing culturally specific expressions (e.g., Chinese "reverse praise"), indicating that models lack humans' common-sense reasoning about cultural contexts [6]. In mixed sentiment disentanglement, although causal graph methods can separate direct and indirect factors, the dynamic process of "local emotion accumulation - global emotion emergence" in long texts (e.g., a turning positive summary after multiple negative descriptions) cannot be effectively captured, limiting the improvement space of hierarchical models due to semantic dynamics [12].

The essential contradiction in multimodal fusion is the conflict between "modal heterogeneity" and "emotional unity." Although the CAMF module can align temporal correlations, in cross-cultural scenarios (e.g., emotional conflicts between the text "interesting" and specific gestures in the Middle East), modal weight allocation is susceptible to cultural biases, resulting in a precision loss exceeding 10% [7]. Modal missing completion technology relies on the imagination of generative models but often introduces false emotional features due to over-completion (e.g., generating "smile" image features for "neutral" texts), leading to decision biases [15].

Balancing ethics and efficiency complicate implementation of technologies: fairness framework achieves an improvement in recognizing minority groups by sacrificing 3%~5% of overall performance, but that cannot be achieved in real time (e. g., public sentiment analysis) [18]. Lightweight models can lower model parameter quantity by up to 60%, which requires the potential of 30% accuracy reduction in complex semantic tasks, implying that there exists a zero-sum game between "efficiency versus precision". Interpretable tools (e. g., interpretability tools) can generate

causal chains to enhance verifiability but tend to increase visualisation cost; it can even reduce user reading efficiency of long texts by 40%.

4. Conclusion

To solve these challenges, in future studies, one should transcend the limitations of single-technology optimization and form a multi-dimensional collaborative framework.

First, this paper suggests building a "domain knowledge graph + causal reasoning" dual-engine model. Then the model encode emotion in domain terms (e. g., side effect negative in medical domain) into graph triples and correct the model's understanding of these implied rules through a causal reasoning engine, with the expectation of avoiding more than 3% performance decrease when it comes to very professional domains.

Second, this paper proposes designing a "dynamic semantic cognition" mechanism. Simulate the human context reasoning process, implanting a "dynamic update module for emotional dictionaries" into the model (e.g., adjust the emotional polarity of "genius" according to context), and combine cultural common-sense databases (e.g., Chinese irony pattern databases), which is expected to improve the accuracy of cross-cultural complex semantic recognition to over 85%.

Third, this paper recommends constructing a "modal symbiosis" fusion system, and mine "emotional prototypes" among different modalities (for example, common elements of text, image and audio correspond to "anger"), then letting models complete missing modalities according to the prototypes directly without resorting to generative imagination to drop the completion bias below 5%.

Additionally, this paper suggests establishing a "fairness-efficiency" collaborative optimization algorithm that retains neurons sensitive to fairness during light-weight pruning (e. g., retain the key feature channels which can recognize female programmers) while aggregating the models with federated learning and distributed inference so that it responds in real time with 90% accuracy toward fairness rules.

References

[1] 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. Minneapolis, MN, USA. pp. 4171-4186.

[2] Sheoran, A., Joshi, A., & Bhattacharyya, P. (2020). Domain Adaptation for Sentiment Analysis: A Survey. Journal of Machine Learning Research, 21(123), 1–50.

[3] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

[4] Poria, S., Cambria, E., Hazarika, D., & Majumder, N. (2020). Multimodal sentiment analysis: A survey and taxonomic study. ACM Computing Surveys (CSUR), 53(12), 1–38.

[5] Pei, J., Men, A., Liu, Y., Zhuang, X., & Chen, Q. (2024). Evidential multi-source-free unsupervised domain adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(8), 5288-5305.

[6] Zhou, C., Li, B., Fei, H., Li, F., Teng, C., & Ji, D. (2024). Revisiting structured sentiment analysis as latent dependency graph parsing. arXiv preprint arXiv:2407.04801.

[7] Lin, Z., Liu, Z., Xia, Z., Wang, X., Wang, Y., Qi, S., ... & Zhu, C. (2024). Rcbevdet: Radar-camera fusion in bird's eye view for 3D object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA. pp. 14928-14937.

[8] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... & Lempitsky, V. (2016). Domain-adversarial training of neural networks. Journal of machine learning research, 17(59), 1-35.

[9] Chalkidis, I., Androutsopoulos, I., & Aletras, N. (2019). Legal named entity recognition with deep neural networks. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL). Florence, Italy. pp. 6318–6324.

[10] Chen, Z., Wang, L. X., Wu, Y. B., Liao, X., Tian, Y., & Zhong, J. (2024). An effective deployment of diffusion LM for data augmentation in low-resource sentiment classification (Paper presentation). Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Miami, FL, United States. pp. 1-14.

[11] Li, B., Zhou, C., Chen, X., & Wang, Y. (2021). Meta-learning for few-shot cross-domain sentiment analysis. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). Punta Cana, Dominican Republic.

[12] Kusner, M. J., Loftus, J., Russell, C., & Silva, R. (2017). Counterfactual fairness. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS). Long Beach, CA, USA. pp. 4066–4076.

[13] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT). San Diego, CA, USA. pp. 1480–1489.

[14] Strizhkova, V., Kachmar, H., Chaptoukaev, H., Kalandadze, R., Kukhilava, N., Tsmindashvili, T., ... & Ferrari, L. M. (2024). Mvp: Multimodal emotion recognition based on video and physiological signals. In European Conference on Computer Vision. Cham: Springer Nature, Switzerland. pp. 101-116.

[15] Zhu, X., Liu, Y., Zhang, L., & Wu, F. (2022). Lightweight multimodal sentiment analysis via knowledge distillation. IEEE Transactions on Affective Computing, 14(3), 1158-1171.

[16] Wang, H., Zhang, Y., Li, J., & Song, D. (2021). Causal structure learning for aspect-based sentiment analysis. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL). Bangkok, Thailand.

[17] Ungless, E. L., Ross, B., & Belle, V. (2023). Queerphobic bias in sentiment analysis. Social Science Computer Review, 41(6), 1234–1250.

[18] Chakraborty, S., Qiu, J., Yuan, H., Koppel, A., Huang, F., Manocha, D., ... & Wang, M. (2024). MaxMin-RLHF: Alignment with diverse human preferences. arXiv preprint arXiv:2402.08925.

[19] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated learning for privacy-preserving sentiment analysis. IEEE Transactions on Knowledge and Data Engineering, 33(12), 4850-4863.

[20] 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.

[21] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... & Lempitsky, V. (2016). Domain-adversarial neural networks. Journal of Machine Learning Research, 17(1), 2096-2030.