様式8

<div align="center">論 文 内 容 要 旨</div>

報告番号	甲 先 第　　**４７０**　号	氏　　名	丁　飛

学位論文題目	Research on Interpretable Text Sentiment Analysis (解釈可能なテキスト感情分析に関する研究)

内容要旨

With the proliferation of Deep Learning (DL) methodologies and Large Language Models (LLMs) in sentiment analysis, significant advancements have been achieved in recent years. However, these models introduce a set of challenges that have not been adequately addressed. Foremost among them is the lack of model interpretability, which hinders the understanding of the mechanisms through which these models make decisions. Furthermore, the current state-of-the-art models necessitate expensive computational resources for training and require vast datasets with manual annotations, posing both financial and time constraints for researchers and practitioners.

Addressing these challenges, this research presents a unique, interpretable framework for text sentiment analysis that is not only cost-efficient but also high in precision. Our proposed framework synthesizes three principal methods:

(1) Enriched Semantic Layer: By merging the unsupervised topic clustering Latent Dirichlet allocation (LDA) matrix with the hidden expression matrix generated by Transformer-like models, we bolster the semantics of the hidden layer. This synthesis allows the model to capture deeper and more nuanced sentiments from the text, bridging the gap between raw data and interpretability.

(2) Symbolic Logic Integration: We incorporate symbolic logic systems, such as Real Logic and Logic Tensor Network (LTN), into our framework. By doing so, we translate the traditionally obscure operations of deep learning models into a more understandable and logical format. This layer of logic helps decode the complex operations, rendering the model interpretable to a significant extent.

(3) Large Language Model Supervision: Using a sophisticated language model, such as ChatGPT, as the teacher model, we generate target text. This text acts as a benchmark, evaluating the quality of the text produced by the student model. Through this teacher-student Causal Inference dynamic, our framework gains insights from state-of-the-art models without inheriting their inherent opacity. In this process, we also utilize technologies such as Prefix Instruc

tion Fine-tuning, within-Domain Further Fine-tuning, and Task-specific Decoding to further improve the efficiency and accuracy of the model.

Incorporating these strategies, our methodology prioritizes both model simplicity and transparency, while also leveraging domain knowledge. Initial results indicate that this hybrid approach melds interpretability with high performance, suggesting a compelling alternative to the prevailing deep learning-centric models. This research aspires to spearhead the development of more transparent, efficient, and accessible sentiment analysis tools in the future.

To further verify the effectiveness of the framework proposed in this research, we conducted extensive experiments on multiple sentiment analysis subtasks. These include Weibo emotion detection, emotion intensity analysis, financial argument analysis, human-machine customer service dialogue satisfaction evaluation, etc. The model outperforms state-of-the-art baselines on various subtasks and achieves first place in both NTCIR-16 DialEval-2 and NTCIR-17 FinArg-1 tasks.