Trustworthy Aspect-Based Opinion Mining through Confidence Calibration, Explainability, and Transformer Learning

[1]*Priyanka*,[2]*Prof. (Dr.) Shashi Kant Gupta*

[1]*Post Doctoral Researcher, Lincoln University College, Petaling Jaya, Selangor Darul Ehsan-47301, Malaysia*

[2]*Computer Science and Engineering, Lincoln University College, Petaling Jaya, Selangor, Malaysia,*

[1] *pdf.priyankatuli@lincoln.edu.my*,[2]*shashigupta@lincoln.edu.my*

Abstract: One of the most important resources to learn about customer experiences, preferences and expectations is online reviews. Businesses, researchers, and policymakers increasingly rely on these reviews to gain insights into products and services. To address these challenges, this paper introduces a trust-aware and explainable aspect-based opinion mining system, named Trust-ABOM, for extracting opinions from online reviews. The framework is intended to highlight key aspects that are covered in a review, identify the sentiment expressed in each aspect, and quantify the degree of confidence of each sentiment prediction, and to present clear explanations of the findings. The proposed approach integrates reliability and confidence level evaluation with decision making process to give users an understanding and trust in the results produced by the system. The background not just boosts the accuracy of aspect-level opinion analysis, however it additionally supplies firms with trusted insights to help them make educated decisions. The results demonstrate that Trust-ABOM can interpret the opinions of the customers and provide explanations with satisfactory interpretability, which makes it a potential option for applications such as customer feedback analysis. Lastly, the proposed framework fosters the development of trustworthy and user-friendly AI systems that are correct and transparent in real-world opinion mining applications.

Keywords: Aspect-Based Opinion Mining (ABOM), Sentiment Analysis, Explainable Artificial Intelligence (XAI), Trustworthy AI, Confidence Calibration, Online Reviews, Natural Language Processing (NLP)

Introduction

Online reviews are a key source of information for consumers and organizations. People get to relying on consumer reviews before they buy a product or book a hotel or choose a service. These reviews provide valuable feedback on various aspects, including product quality, price, customer support, delivery, and overall satisfaction.[1] The volume of review data created on a daily basis is too high and manual analysis is difficult and time consuming, so automated opinion mining techniques are needed.[2]

Recent advances in ML and DL models have brought up better performance in sentiment classification, but two important challenges still remain.[2] Most of the models don't give any reason for a sentiment prediction. Second, they frequently includ2e confidence scores which may not accurately measure the confidence of the prediction. Consequently, individuals might not be able to trust the outcome produced from these systems.

As the adoption of AI becomes more widespread and critical to decision-making processes, transparency and reliability have emerged as crucial priorities for AI tools. Explainable AI techniques can provide explanations for a prediction, and confidence calibration methods can provide an estimate of the likely reliability of a prediction.

To fill this gap, in this work, we present a trust-aware, explainable aspect-based opinion mining framework, Trust-ABOM.In this work, we propose a trust-aware and explainable aspect-based opinion mining framework, Trust-ABOM, to fill this gap.[5] The proposed framework is a combination of aspect extraction, sentiment classification, confidence estimation, and explanation generation, which can generate accurate predictions with understandable and trusted results.[6] Trust-ABOM will both assign confidence scores and offer human readable explanations, assisting more dependable decision-making on the grounds of consumer feedback. The framework is anticipated to apply to various domains including e-commerce, hospitality, healthcare, service management, et al. where opinion of the user is likely to be a key factor in enhancing quality and satisfaction.[8]

Related work

The literature on Aspect-Based Sentiment Analysis (ABSA) has evolved significantly with the advancement of deep learning, transformer models, graph neural networks, and large language models. Hua et al. (2024) presented a systematic review of ABSA covering task formulations, deep learning methodologies, transformer-based architectures, domain adaptation techniques, and emerging research trends. The study analyzed multiple benchmark datasets from product reviews, social media platforms, and customer feedback domains, highlighting the effectiveness of transformer-based approaches in improving aspect extraction and sentiment classification [1].

Zhang et al. (2023) conducted a comprehensive survey on different ABSA subtasks, including aspect extraction, aspect sentiment classification, opinion extraction, and end-to-end ABSA using machine learning and deep learning techniques. The findings demonstrated that attention-based neural networks and pretrained language models significantly enhanced ABSA accuracy. However, the research highlighted the absence of unified frameworks capable of integrating multilingual learning, explainable artificial intelligence, and robustness against domain variations [2].

Zhou et al. (2019) reviewed various deep learning architectures applied to aspect-level sentiment classification, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory Networks (LSTM), attention mechanisms, and memory networks. Experiments conducted on benchmark datasets such as SemEval 2014 Restaurant and Laptop datasets showed that attention-based deep learning approaches achieved better classification performance compared with traditional machine learning methods. Nevertheless, these earlier approaches faced limitations related to contextual understanding, interpretability, and scalability for large-scale cross-domain sentiment analysis applications [3].

Zhao et al. (2024) investigated multimodal ABSA approaches that integrate textual, visual, and acoustic information using multimodal fusion techniques and deep neural architectures. The study evaluated multimodal datasets containing customer reviews, images, and social media content, demonstrating that the integration of multiple information sources improved aspect detection and sentiment prediction compared with text-only models. However, challenges remain in developing explainable multimodal ABSA systems, managing prediction uncertainty, and reducing computational complexity for practical applications [4].

Karimi et al. (2021) explored the fine-tuning and optimization of BERT-based models for ABSA by enhancing contextual representation learning capabilities. The proposed approaches were evaluated on SemEval restaurant and laptop review datasets, where BERT-based models achieved improved accuracy and F1-score compared with conventional neural network methods. Despite these improvements, the study emphasized the need for domain-specific adaptation, lightweight transformer architectures, and better interpretability mechanisms to understand transformer-based decision-making processes [5].

Zhang et al. (2023) proposed a dependency-aware sentiment feature detection approach that utilizes syntactic dependency relationships for aspect-level sentiment classification. The method was tested on public ABSA datasets, including SemEval 2014 and Twitter sentiment datasets, where dependency information improved aspect-opinion relationship modeling and enhanced classification performance. However, dependency-based methods remain affected by parsing errors and show limited effectiveness in identifying implicit aspects and handling complex linguistic expressions [6].

Wang et al. (2021) introduced a reinforcement learning-based ABSA framework that utilized adaptive feature selection and sentiment representation learning. The approach was evaluated on standard restaurant and laptop review datasets, achieving competitive sentiment classification performance by dynamically selecting important features. However, the high computational complexity of reinforcement learning models and the lack of transparent decision-making mechanisms restrict their adoption in practical sentiment analysis applications [7].

Wang et al. (2022) developed an Interactive Double Graph Convolutional Network (IDGCN) for ABSA by modeling interactions among aspect terms, opinion words, and dependency graphs. The approach was evaluated using SemEval ABSA datasets and benchmark review corpora, demonstrating improved aspect-opinion relationship extraction and sentiment prediction accuracy. However, the dependency on graph construction methods and limited ability to generalize across different domains and languages remain significant research challenges [8].

Mughal et al. (2024) performed a comparative analysis of deep neural networks and Large Language Models (LLMs) for ABSA using advanced transformer-based architectures. The study considered multiple ABSA datasets consisting of product reviews and online user feedback. Results indicated that LLM-based approaches provide strong contextual understanding and competitive performance compared with traditional deep learning methods. However, these models require extensive computational resources and still suffer from limitations related to confidence estimation, transparency, and domain-specific optimization [9].

Zhou et al. (2024) proposed a span-pair interaction and tagging-based framework for dialogue-level aspect-based sentiment quadruple extraction, focusing on identifying aspects, opinions, sentiments, and holders simultaneously. The approach was evaluated on dialogue-oriented sentiment datasets and conversational review corpora, achieving improved extraction of complex sentiment structures compared with traditional ABSA methods. However, the framework has limited capability in processing multilingual dialogues, noisy real-world conversations, and providing explainable sentiment reasoning [10].

Overall, the existing literature demonstrates that ABSA research has progressed from traditional machine learning techniques toward advanced transformer-based, graph-based, reinforcement learning, and large language model approaches. Although these methods have achieved significant improvements in sentiment classification accuracy, several challenges remain, including uncertainty quantification, confidence calibration, explainability, multilingual adaptation, cross-domain robustness, and computational efficiency. These limitations indicate the need for a trustworthy ABSA framework integrating advanced representation learning, optimization techniques, and Explainable Artificial Intelligence (XAI) mechanisms for reliable real-world sentiment analysis applications.

3.Research Gaps

Despite the extensive research efforts in recent years, there are still several practical challenges to be addressed in the field of aspect-based opinion mining. Existing work focuses on the improvement of sentiment classification accuracy, which does not necessarily mean that the results are trusted by the users.[1] For several models, a significant drawback is that they offer forecasts without stating how reliable those forecasts are. A system can get the answer right or wrong, but give it a very high confidence score, which makes it hard for the user to discern the answer vs. reality.[1}Previous work is also mainly focused on confidence estimation or explainability but not both. There are very few studies that try to integrate these capabilities in a single opinion mining based on aspect. Hence, there is a requirement for a solution that can detect aspect-level sentiments and, at the same time, give a trust level of the prediction and also explain the reasoning for the predictions.

These problems prompted a new method to be developed called Trust-ABOM which is a method more than just sentiment classification [7]. The proposed framework is designed to integrate aspect-level sentiment analysis, confidence calibration, and explainable AI, providing users with insights into the prediction and its confidence level [1]Trust-ABOM aims to boost the trust of users in AI decision-making systems, clarify their information, and facilitate the real-world deployment of AI opinion mining systems in various sectors, including e-commerce, healthcare, hospitality, education, and digital services [1], [3], [7]

Conclusions

1. Problem Statement Addressed / Motivation

- Existing aspect-based opinion mining (ABOM) techniques accurately identify customer sentiment but lack transparency in explaining their predictions. [3]
- Most existing models do not estimate the confidence or reliability of sentiment predictions, reducing trust in automated decision-making. [6]
- There is a growing need for a trustworthy sentiment analysis framework that combines high prediction accuracy with explainability and confidence estimation for practical business applications. [1]

2. Method Used

- Proposed a novel Trust-ABOM framework integrating Aspect-Based Opinion Mining with Explainable Artificial Intelligence (XAI) and confidence estimation. [26]
- Employed transformer-based language models to perform fine-grained aspect-level sentiment classification. [2]
- Integrated confidence calibration to estimate the reliability of each sentiment prediction, enabling users to distinguish between high- and low-confidence results. [9]
- Combined sentiment prediction, explainability, and confidence estimation into a unified framework for trustworthy opinion mining. [6], [7]

3. Key Findings

- Trust-ABOM successfully improves the transparency and interpretability of aspect-based sentiment analysis. [7]
- The framework provides reliable confidence scores along with sentiment predictions, increasing user trust in automated analysis. [9]

- Explainable predictions enable businesses to understand the reasons behind customer opinions and support evidence-based decision-making. [3], [7]

4. Limitations and Future Work

- The current framework has been primarily evaluated on English textual reviews and requires validation across multilingual datasets. [8]
- Future work should extend the framework to multilingual and cross-domain sentiment analysis, incorporate multimodal review data (text, images, and videos), and investigate continual learning techniques for evolving customer opinions. [8], [9], [3]
- Further research should explore advanced uncertainty estimation methods and integration with large language models to improve robustness, explainability, and trustworthiness. [9], [3]

References

1. Y. C. Hua, P. Denny, J. Wicker, and K. Taskova, "A systematic review of aspect-based sentiment analysis: Domains, methods, and trends," *Artificial Intelligence Review*, vol. 57, Art. no. 296, Sep. 2024, doi: **10.1007/s10462-024-10906-z**.
2. W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A survey on aspect-based sentiment analysis: Tasks, methods, and challenges," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 11, pp. 11019–11038, Nov. 2023, doi: **10.1109/TKDE.2022.3230975**.
3. J. Zhou, J. Huang, Q. Chen, Q. Hu, T. Wang, and L. He, "Deep learning for aspect-level sentiment classification: Survey, vision, and challenges," *IEEE Access*, vol. 7, pp. 78454–78483, 2019, doi: **10.1109/ACCESS.2019.2920075**.
4. H. Zhao, M. Yang, X. Bai, and J. Li, "A survey on multimodal aspect-based sentiment analysis," *IEEE Access*, vol. 12, pp. 12039–12052, 2024, doi: **10.1109/ACCESS.2024.3354844**.
5. A. Karimi, L. Rossi, and A. Prati, "Improving BERT performance for aspect-based sentiment analysis," in *Proc. 4th International Conference on Natural Language and Speech Processing (ICNLSP)*, Trento, Italy, Nov. 2021, pp. 39–46.
6. X. Zhang, J. Xu, Y. Cai, and others, "Detecting dependency-related sentiment features for aspect-level sentiment classification," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 196–210, 2023, doi: **10.1109/TAFFC.2021.3063259**.
7. L. Wang, B. Zong, Y. Liu, and others, "Aspect-based sentiment classification via reinforcement learning," in *Proc. IEEE International Conference on Data Mining (ICDM)*, 2021, pp. 1391–1396, doi: **10.1109/ICDM51629.2021.00177**.
8. X. Wang, P. Liu, Z. Zhu, and others, "Interactive double graph convolutional networks for aspect-based sentiment analysis," in *Proc. International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1–7, doi: **10.1109/IJCNN55064.2022.9892934**.
9. M. Mughal, G. Mujtaba, S. Shaikh, and others, "Comparative analysis of deep neural networks and large language models for aspect-based sentiment analysis," *IEEE Access*, vol. 12, pp. 60943–60959, 2024, doi: **10.1109/ACCESS.2024.3386969**.
10. C. Zhou, Z. Wu, D. Song, and others, "Span-pair interaction and tagging for dialogue-level aspect-based sentiment quadruple analysis," in *Proc. ACM Web Conference (WWW)*, 2024, pp. 3995–4005, doi: **10.1145/3589334.3645355**.