Large Language Models in Accounting and Financial Research: A Review of Applications in Accounting-related Text Analysis

Yan Hao

School of Management, Hefei University of Technology, Hefei, China
1321498664@qq.com

Abstract: With the rapid advancement of large language models (LLM) technologies, new technical support and research paradigms have emerged for traditional accounting and finance studies. Despite these opportunities, scholars in the fields of accounting and finance often encounter challenges when navigating the extensive and complex domain knowledge of large language models, as well as the ever-evolving academic literature. To address this gap, this paper conducts a qualitative survey on the applications of large language models in early accounting and finance researches. This paper is structured into three main sections. First, it delves into the definition, underlying principles, and developmental trajectory of large language models. Second, it synthesizes the latest research on large language models applications in accounting and finance, categorizing these studies into three major domains, including sentiment analysis, report analysis and practical works. Finally, the paper highlights emerging trends and potential research directions, aiming to provide a comprehensive guide for future scholarly exploration in this dynamic field. Although it still has many shortcomings such as high cost and black box of process, the application of large language modeling provides a new technical support for text analysis in accounting and financial research, and also improves the actual efficiency of accounting and auditing.

Keywords: Large Language Model, Accounting Research, Financial Research, Text analysis

1. Introduction

In the wave of global digital transformation, the explosive growth of unstructured financial data in enterprises is giving rise to profound changes in the paradigm of accounting information processing. However, the utilization rate of traditional accounting research is less than 12%, forming a huge "data value depression". The breakthrough progress of generative artificial intelligence - especially the excellent performance of other large language models (LLMs) such as ChatGPT4 in the field of semantic understanding and pattern recognition - has provided a technical key to break this dilemma[1].

Although existing studies have provided a multi-dimensional overview of the application landscape of large language models in accounting, there is still a lack of systematic literature integration and theoretical construction of interdisciplinary research on large language models in the in-depth analysis of unstructured accounting texts. Of particular concern is that traditional literature reviews have not yet fully revealed the domain-specific challenges of large language models in

handling semantically complex texts such as management's discussion and analysis (MD&A) and note disclosure, and this theoretical blind spot directly restricts the innovative development of intelligent analysis paradigms for accounting texts. Therefore, this paper is dedicated to constructing a three-dimensional research framework of theory-technology-application of large language models in accounting researches: in the theoretical dimension, it explores the development of accounting text analysis , practical works and how the large language model can reconfigure the "text-data-knowledge" transformation mechanism of accounting information; In the technical dimension, it analyzes the special requirements of prompt engineering and domain adaptation on accounting-related text processing; in the application dimension, this paper summarizes the application of large language model analysis quality on sentiment analysis, report analysis and practical works.

By systematically combing through 25 core articles,, this paper combs through the current research status of large language models in accounting text analysis, and provides a new analytical paradigm for understanding how large language models can reshape the value discovery function of accounting and financial research.

2.　　Relevant concepts and research status

2.1.　Definition and type of large language models

A large language model (LLM) represents a cutting-edge form of generative artificial intelligence. As a natural language processing model grounded in deep neural network architectures [2] LLMs are designed to model and generate textual data. Their core characteristics are manifested across three key dimensions: technical architecture, training paradigm, and data scale.

In terms of architectural design, LLMs employ sequential deep neural networks, most notably the Transformer architecture, whose attention mechanism enables context-aware text modeling and captures complex linguistic relationships. Regarding the training mechanism, LLMs leverage a self-supervised pre-training paradigm. By processing extensive volumes of unlabeled textual data, including multilingual corpora and specialized literature, these models learn to represent language comprehensively. Common pre-training tasks, such as masked language modeling (MLM) and sequence generation prediction, facilitate the acquisition of semantic and syntactic knowledge [1,3,4]).

Distinguishing itself from traditional language models, the defining innovation of LLMs lies in its "pre-training & fine-tuning" two-stage learning framework. Starting from a foundation of general language understanding, LLMs can be adapted to specialized domains—such as accounting—by integrating domain-specific lexicons and fine-tuning for task-oriented objectives, like financial report generation. This dual-stage approach empowers LLMs to handle both open-ended language generation tasks, such as management discussion and analysis, and structured semantic parsing. As a result, LLMs have redefined the boundaries of intelligent text processing, offering new possibilities for diverse applications.

2.2.　The current development status of large language models

The development of large language models includes four steps:

(1) Statistical language models (SLM)

The evolutionary trajectory of language modeling commenced with Statistical Language Models (SLMs)[5] , foundational systems grounded in probabilistic frameworks for linguistic pattern recognition. It predicts subsequent word probabilities by counting historical n-gram frequencies (e.g., binary models [6].

(2) Neural language models (NLM)

Neural language models (NLMs) are a class of machine learning algorithms that utilize artificial neural networks to process and interpret language data. These models typically employ multilayer

perceptrons (MLPs) and recurrent neural networks (RNNs) to analysis the preceding word in a sequence, which can predict the probability distribution of the subsequent word.

(3) Pre-trained language models (PLM)

With the introduction of the highly parallelizable Transformer architecture based on a self-attentive mechanism[7]. Models based on the Transformer architecture such as BERT model, has achieved learning of contextual content from both the left and right directions of the text, and adjusted the prediction results according to the contextual content [8]. Besides, significant advancements in pre-training achieved by pre-trained language models (PLMs), which allow for prediction-based adjustment.

(4) Large Language Models (LLMs)

According to scaling laws for neural language model[9], models with larger sizes are significantly more sample efficient. Large language models usually contain billions or even trillions of parameters, which improves their accuracy in predicting text sentiment higher than that of common text sentiment analysis methods such as the BERT model[10].

3. Application of large language models in accounting research

3.1. Sentiment analysis

The large amount of unorganised text on social media and investor communication platforms (like user comments, Q&A sessions, and executive statements) contains a lot of information about feelings, and these positive or negative emotions often give hints about how well a company will do in the future to people like management, creditors, and investors through the way information spreads in the market. With the technology iteration of Large Language Model (LLM), its deep semantic understanding in unstructured text sentiment analysis has been significantly improved, providing a technological breakthrough for mining hidden market sentiment.

In the field of news text sentiment analysis, studies have confirmed the significant advantages of LLM: ChatGPT is able to accurately capture market sentiment fluctuations in financial news by building a multi-dimensional sentiment analysis framework [11]. Further expanding to investor interaction scenarios, a study utilized ChatGPT 4 to semantically parse investor communication data from Oriental Wealth Net, and found that there is a significant dynamic interaction effect between investor sentiment index and stock return.[12]. Investor sentiment can even be used to make predictions about stock prices.[13]. User-generated content on social media platforms has also become an important object of analysis. A study of corporate executives' Twitter texts shows that the intensity of negative emotions such as fear and anger in their work-related tweets is negatively correlated with the company's market value[14], revealing the potential impact of executives' non-public channel expressions on the capital market. With the deepening of the global ESG investment concept, big language modeling has been applied to the sentiment analysis of carbon disclosure: by parsing Interactive Carbon Disclosure (ICD) texts in investors' social networks, related studies have found that positive carbon disclosure sentiments can significantly enhance firm value through paths such as lowering the enhancement of corporate reputation [15].

3.2. Report analysis

In addition to sentiment analysis, big language models show unique advantages in the parsing of standardized texts of companies, and their in-depth semantic understanding capability provides technical support for cracking information asymmetry and restoring the real operational picture of enterprises.

First, in the field of financial fraud identification,[16] constructs a financial report text analysis framework based on big language models, which assists financial investigators in identifying

accounting fraud signals by mining abnormal semantic features in unstructured texts such as management discussion and analysis (MD&A) and notes. This technology path breaks through the limitations of traditional quantitative indicators and realizes the intelligent capture of implicit risks in text.Secondly, in the supply chain risk management scenario, in the study of [17] ChatGPT is applied to the semantic parsing of firms' site visit records to construct a semantic assessment model of supply chain risk by extracting the key information in the text about qualitative descriptions such as supplier stability and logistics efficiency. This method provides a qualitative research dimension to the traditional quantitative assessment system and effectively fills the analytical blind spot in the dynamic monitoring of supply chain risk. Finally, in the face of such difficult-to-quantify assessment objects as ESG performance, big language modeling technology promotes the structured conversion of unstructured data:[18] propose the ESGReveal method to accurately extract key elements such as environmental compliance indexes, social contribution data, and governance structure characteristics from lengthy ESG reports through custom-trained LLM models, and transform unstructured text into quantifiable standardized datasets. Unstructured text is transformed into quantifiable and standardized data sets. This technological innovation significantly improves the transparency of ESG disclosure and provides high-precision analytical tools for regulators to conduct compliance reviews, investors to construct ESG portfolios, and researchers to conduct cross-company comparative studies. It also helps to solve the long-standing problems of fragmented ESG data and inconsistent evaluation standards.

3.3. Accounting practical works

LLMs also play an important role in actual accounting work[19]. Recent research has focused on the transformative impact of large language models (LLMs) on accounting practice. It systematically demonstrated the technological suitability of ChatGPT for scenarios such as financial reporting, risk assessment, and so on[20] proposed that certified public accountants can optimize their workflow by using LLMs to complete text generation-type tasks to optimize workflows, and constructed an operational framework for secure deployment. In terms of domain-specific research[21] quantitatively analyze the boundaries and ethical risks of ChatGPT's application in manuscript generation and anomaly detection for external auditing scenarios, while[22] specialize in forensic accounting, and develop a cue-engineering-based fraud investigation aid. Academic consensus suggests that the core value of LLMs lies in the automation of standardized processes - freeing up practitioners' higher-order cognitive resources by taking over repetitive paperwork (e.g., audit communication letter drafting, tax memo writing) [23]. This feature of the technology is particularly beneficial for resource-constrained small and medium-sized firms, providing them with a sustainable capacity expansion solution in the context of the industry's talent shortage.

4. Discussion

Although the development of large language models has made significant progress and demonstrated a wide range of application potential in the academic research fields of accounting and finance[24], there are still multidimensional limitations in their technical systems. To be more specific, firstly, the practical application of the large language model is highly dependent on large-scale human and arithmetic costs for customized fine-tuning, and this high development threshold leads to the concentration of resources in a few institutions, which objectively exacerbates the fairness dilemma of information access. Secondly, the black-box characteristics of algorithms and model illusion problems have not yet been effectively solved[25], which directly affects the reliability of the output results and the reference value of decision-making. In addition, the current technical framework is

still limited to text processing, and there are obvious research gaps in the integration and analysis of multimodal information such as images and audio.

In this context, future academic research needs to make breakthroughs in three aspects: First, while focusing on the integration of text analysis capability of large language models and financial decision-making scenarios, more attention should be paid to the new type of information asymmetry that may be triggered by differences in technological accessibility -- even in the face of homogenized information inputs, the gap in technological resource endowment among different subjects may still lead to a differentiation of decision-making advantages. Even in the face of homogenized information input, different subjects may still have differentiated decision-making advantages due to the gap in technical resource endowment. Secondly, improving the interpretability of the model operation mechanism and constructing an illusion detection and correction system will become the core research direction to consolidate the credibility of the technology. Third, the research scope needs to break through the traditional text boundary, explore the technical path of multimodal information processing, and promote the evolution of the big language model from a single text analysis to the in-depth fusion analysis of image, audio and other multi-media information, so as to provide a richer data dimension and methodological support for academic research.

5. Conclusion

This paper reviews the evolution of large language modeling from SLM to LLM and its recent applications in accounting and finance. The paper identifies areas that have been well-studied as well as areas that require further research. On the one hand, well-researched areas include sentiment analysis of textual content on social media and news media, analysis of the content of various types of company reports, and how large language modeling can help accountants and auditors to improve the efficiency of their work as well as the quality of their work in the real world of accounting and auditing. On the other hand, there are still a lot of difficulties in the application of big language modeling, such as high cost, insufficient modeling itself, and still a single analysis content. Solving these challenges will require more time. In addition, this paper analyzes the future research trends of big language models in accounting and finance, including the new information asymmetry brought by technological accessibility, the performance enhancement of big language models themselves, and the expansion of the types of content analyzed. However, there are still some limitations about this paper. First of all, this study only focuses on the current state of research on the application of big language modeling in text analysis, and lacks the excavation of other aspects, and the breadth of research needs to be improved. Second, due to the limited research time, it is difficult to realize the in-depth collection and analysis of the development process of related literature in the time dimension, which may lead to the short-term nature of the research results.

References

[1] Chang Y, Wang, X, Wang, J, et al. A survey on evaluation of large language models[J], 2024, 15(3): 1-45.
[2] Feuerriegel S, Hartmann, J, Janiesch, C, et al. Generative ai[J], 2024, 66(1): 111-126.
[3] Naveed H, Khan, A U, Qiu, S, et al. A comprehensive overview of large language models[J]. arXiv preprint 2023: arXiv:.06435.
[4] Teubner T, Flath, C M, Weinhardt, C, et al. Welcome to the era of chatgpt et al. the prospects of large language models[J], 2023, 65(2): 95-101.
[5] Rosenfeld R. Two decades of statistical language modeling: Where do we go from here?[J]. Proceedings of the IEEE, 2000, 88(8): 1270-1278.
[6] Horowitz J L, Savin, N J J o e p. Binary response models: Logits, probits and semiparametrics[J], 2001, 15(4): 43-56.
[7] Vaswani A, Shazeer, N, Parmar, N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[8] Devlin J, Chang, M-W, Lee, K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[A],Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)[C], 2019.

[9] Kaplan J, McCandlish, S, Henighan, T, et al. Scaling laws for neural language models[J]. arXiv preprint, 2020: arXiv:.08361.

[10] Wen Y, Liang, Y, Zhu, X. Sentiment analysis of hotel online reviews using the BERT model and ERNIE model-Data from China[J]. PLoS One, 2023, 18(3): e0275382.

[11] Fatouros G, Soldatos, J, Kouroumali, K, et al. Transforming sentiment analysis in the financial domain with ChatGPT[J]. Machine Learning with Applications, 2023, 14.

[12] Zhuang Y, Wang, F, Chiu, D K W, et al. Leveraging large language models to examine the interaction between investor sentiment and stock performance[J]. Engineering Applications of Artificial Intelligence, 2025, 150.

[13] Zhen K, Xie, D, Hu, X. A multi-feature selection fused with investor sentiment for stock price prediction[J]. Expert Systems with Applications, 2025, 278.

[14] Wang Q, Yiu Keung Lau, R, Xie, H, et al. Social Executives' emotions and firm value: An empirical study enhanced by cognitive analytics[J]. Journal of Business Research, 2024, 175.

[15] Zhu J, Zhang, C, Sun, J, et al. The impact mechanism of interactive carbon disclosure on firm value moderated by investors' online social networks[J]. Research in International Business and Finance, 2025, 75: 102771.

[16] Bhattacharya I, Mickovic, A. Accounting fraud detection using contextual language learning[J]. International Journal of Accounting Information Systems, 2024, 53.

[17] Fan S, Wu, Y, Yang, R. Measuring firm-level supply chain risk using a generative large language model[J]. Finance Research Letters, 2025, 77.

[18] Zou Y, Shi, M, Chen, Z, et al. ESGReveal: An LLM-based approach for extracting structured data from ESG reports[J]. Journal of Cleaner Production, 2025, 489.

[19] Zhao J, Wang, X J J o C A, Finance. Unleashing efficiency and insights: Exploring the potential applications and challenges of ChatGPT in accounting[J], 2024, 35(1): 269-276.

[20] Street D, Wilck, J, Chism, Z. Six principles for the effective use of ChatGPT and other large language models in accounting[J]. CPA Journal, 2023.

[21] Fotoh L, Mugwira, T. Exploring Large Language Models (ChatGPT) in External Audits: Implications and Ethical Considerations[J]. Available at SSRN 4453835, 2023.

[22] Street D, Wilck, J. 'Let's Have a Chat': Principles for the Effective Application of ChatGPT and Large Language Models in the Practice of Forensic Accounting[J]. Journal of Forensic Investigative Accounting, 2023.

[23] Boritz J E, Stratopoulos, T C. AI and the accounting profession: Views from industry and academia[J]. Journal of Information Systems, 2023, 37(3): 1-9.

[24] Dong M M, Stratopoulos, T C, Wang, V X. A scoping review of ChatGPT research in accounting and finance[J]. International Journal of Accounting Information Systems, 2024, 55: 100715.

[25] Yi Z, Cao, X, Chen, Z, et al. Artificial intelligence in accounting and finance: Challenges and opportunities[J]. IEEE Access, 2023, 11: 129100-129123.