

Research on Intelligent Business Data Analysis Methods Driven by Large Models

Bowen Ma

Tencent Technology (Shenzhen) Co. Ltd, Shenzhen, Guangdong, 518000, China

Keywords: Large model; Intelligence; Business data analysis

Abstract: In the current digital era, business data analysis has become a key basis for enterprise decision-making. With the continuous expansion of data scale and the increasing complexity of business requirements, traditional data analysis methods are gradually revealing their limitations. The emergence of large models has brought new opportunities for intelligent business data analysis. Their powerful computing capabilities and intelligent algorithms can deeply mine the potential value in the data and provide more accurate and efficient decision support for enterprises. Systematically exploring the methods, advantages, challenges and coping strategies of large model-driven intelligent business data analysis can provide useful references for enterprises to enhance their data analysis capabilities during digital transformation.

1. Introduction

In the era of information explosion, data has become an important asset for enterprises. Business data analysis, as an important means of mining data value and assisting in decision-making, is playing an increasingly crucial role. With the diversification of enterprise businesses and the exponential growth of data volume, traditional data analysis methods are facing huge challenges in handling complex data structures, real-time requirements and in-depth insights. Large models, with their powerful processing capabilities and intelligent algorithms, provide new ideas and methods to solve these problems. Tencent Data Governance Platform and Tianqiong ChatBI, as excellent practical cases in the industry, offer valuable experience for us to study the application of large models in business data analysis.

2. Relevant Theories and Technical Foundations of Large Models

A large model is a machine learning model with massive parameters, complex structures and super computing capabilities. It is based on deep learning technology and uses multi-layer architectures to automatically extract data features like neural networks. Whether dealing with large-scale data or complex relationship patterns, large models can handle them with ease and show extremely high prediction accuracy. Its training methods cover various categories such as supervised learning, unsupervised learning and reinforcement learning. By training with a large amount of labeled or unlabeled data, large models can continuously adjust their parameters to enhance the performance and generalization ability of the model. This major technological

breakthrough enables large models to understand and process massive amounts of information more accurately.

Looking forward, large models will continue to evolve towards more efficient, more accurate, more universal and more interpretable directions. It will integrate more closely with other emerging technologies and give birth to more innovative and practical applications. These applications will strongly support us in solving various complex practical problems. Whether in the fields of business decision-making, medical diagnosis or autonomous driving, large models will exert their huge potential. The development of large models will undoubtedly bring more possibilities and opportunities for our future.

3. Connotation and Characteristics of Large Models

Large models are not merely marked by their large scale. Their essence lies in having advanced and unique architectures, numerous and rich parameters, as well as outstanding learning capabilities. Such models have the powerful ability to handle massive data and can keenly detect and reveal the patterns and rules hidden deep within the seemingly complex data ocean. Large models have the excellent ability to extract deep-level and abstract features from the original data automatically. Relying on their advanced algorithms and powerful computing capabilities, they can accurately identify and capture the key information and core elements in the data without much manual intervention, laying a solid foundation for subsequent analysis and processing.

When faced with new data that has never been encountered before, large models can fully utilize the rich learning experience accumulated in the past to conduct rapid and accurate predictions and analyses. It shows extremely strong adaptability and flexibility, can quickly understand and deal with various data characteristics and application scenarios, and always maintain a high level of analysis and prediction accuracy. When dealing with large-scale and diverse data, large models demonstrate extraordinary advantages. No matter how large the data scale is or how diverse the types are, it can handle them with ease and deeply mine the potential valuable information. Through comprehensive analysis and in-depth understanding of large-scale data, large models can provide accurate and effective strategies and solutions for solving various complex problems [1].

4. Advantages of Intelligent Business Data Analysis Driven by Large Models

Large models have strong parallel computing capabilities and efficient algorithms, which can process massive amounts of data in a short period, significantly shortening the data analysis cycle. They can capture complex relationships and subtle features in the data, and the accuracy of the analysis results has also been significantly improved. Traditional data analysis methods often can only reveal surface phenomena, while large models can deeply mine hidden patterns and potential laws in the data, providing enterprises with new perspectives and innovative ideas. Through in-depth analysis of customer behavior data, new customer needs and market opportunities can be discovered, promoting product innovation and business expansion. In a highly competitive business environment, enterprises need to be able to respond to market changes in a timely manner and adjust business strategies. Large models can monitor the dynamic changes of data in real-time, quickly update the analysis model, and provide real-time decision support for enterprises. Through precise data analysis, enterprises can allocate resources more reasonably, avoiding waste and idleness of resources. The automated processing capabilities of large models can reduce manual intervention and lower labor and operational costs[2].

5. Insights from Tencent ola Data Governance Platform

Tencent ola Data Governance Platform is an intelligent platform based on big data technology. Through data governance, data mining, and data analysis methods, it provides comprehensive data support for enterprises. This platform adopts a large model-driven intelligent analysis method to extract valuable information from massive data and provide strong support for enterprise decision-making.

5.1 The Key Significance of Data Governance

Tencent ola Data Governance Platform clearly demonstrates the core essence of data governance for business data analysis. In today's digital business environment, data has become a key asset for enterprises, and high-quality data governance is the cornerstone for giving full play to its value. Through effective data governance, strict control over the quality of data can be achieved to ensure it is accurate, complete, and consistent. This process can eliminate errors and biases in the data and maintain its consistency and reliability in different application scenarios. With such a solid foundation, subsequent data analysis work can be carried out smoothly, providing strong support for enterprise decision-making and promoting the continuous development and innovation of the business[3].

5.2 The Importance of Data Integration and Sharing

This platform has achieved remarkable success in data integration and sharing, successfully breaking the predicament of isolated and scattered data. In traditional enterprise architectures, different departments and business systems often operate independently, making data difficult to circulate and integrate, resulting in numerous "data silos". Tencent ola Data Governance Platform, through advanced technologies and reasonable architectural design, realizes seamless data integration and sharing. This enables different departments within the enterprise to freely exchange and utilize data, promoting cross-departmental collaboration and information sharing. The marketing department can conduct more precise market analysis and predictions by combining data from the sales department, and the R&D department can improve product designs based on user feedback data. This collaborative analysis model greatly improves the operational efficiency and decision-making scientificity of the enterprise, accelerating business innovation and optimization.

5.3 The Necessity of Data Security and Privacy Protection

In the context of business data analysis, the importance of data security and privacy protection is self-evident. With the continuous increase in data value, the risks of data leakage and privacy infringement are also increasing. Tencent ola Data Governance Platform attaches great importance to this and has adopted a series of rigorous and effective security protection measures. Through encryption technology, access control, and identity authentication means, it ensures that only authorized personnel can access and use sensitive data. A complete data privacy management mechanism is established to strictly regulate and supervise the collection, storage, use, and destruction of data to safeguard the legitimate rights and interests of data subjects. Such powerful safeguard measures maintain the enterprise's reputation and credibility, enhance users' trust in the enterprise, and create favorable conditions for the sustainable development of the enterprise.

6. Characteristics and Application Value of Tianqiong ChatBI

Tianqiong ChatBI is an AI-based business intelligence analysis tool that realizes intelligent data analysis through technologies such as natural language processing. This tool adopts a large model-driven intelligent analysis method and can provide data support for enterprises quickly and accurately.

6.1 Advantages of Intelligent Interaction and Visualization

Tianqiong ChatBI demonstrates excellent intelligent interaction and visualization features, bringing a brand-new experience and significant value to users. In terms of intelligent interaction, it breaks the operation mode of traditional data analysis tools, enabling users to communicate with the system smoothly and conveniently in natural language. Without the need for complex instructions and professional operation skills, users only need to express their needs in daily language, and the system can quickly understand and give accurate responses. This natural and intuitive interaction method greatly lowers the user's usage threshold, allowing personnel of different technical backgrounds and professional levels to easily handle the data analysis process. In terms of visualization, Tianqiong ChatBI performs outstandingly. It can present complex data in an intuitive and clear chart form, making the information and trends behind the data readily apparent. Whether it is a bar chart, line chart, pie chart, or more complex data visualization forms, it can precisely present the characteristics and relationships of the data. This visualization presentation method significantly improves the efficiency of data analysis, enables users to quickly obtain key information in a short period of time, and significantly enhances the readability and understandability of the data[4].

6.2 ChatBI Problems need to be solved

The challenges faced in the interactive experience of finding numbers, analyzing data, and generating Business Intelligence (BI) charts through natural language in Tencent Cloud BI are multifaceted. One significant issue arises from colloquial user input, where users often phrase their requests in a conversational manner. This colloquial language can lead to unclear or ambiguous intentions, which hampers the model's ability to accurately interpret and process the request. As a result, subsequent BI analysis may fail, or the output may be incorrect, creating a poor user experience. Another key challenge is the lack of domain-specific knowledge. When the model lacks understanding of the relevant business context, it struggles to grasp the user's real intention. In such cases, the model may generate hallucinations or incorrect analyses because it doesn't have sufficient background knowledge to provide the right insights. Additionally, limitations in the model's capabilities, especially in relation to understanding and solving complex problems, pose significant barriers. Despite the ongoing advancements in AI and the large model ecosystem, issues such as limited sequence support length, tendency to hallucinate, difficulty in following complex instructions, and struggling with intricate problem-solving remain persistent. These limitations hinder the model's performance, especially when users request intricate analyses or have very specific needs that the model is not adequately trained to handle.

7. Detailed Explanation of Intelligent Business Data Analysis Methods Driven by Large Models

The ChatData framework consists of three main modules: the Agent layer, the foundational model layer, and the atomic tool layer. The foundational model layer provides the core capabilities

of large language models, including knowledge, linguistic, comprehension, and reasoning abilities, among others. These capabilities serve as the bedrock for the framework's overall functionality, ensuring that the system can understand and process complex language inputs. The BigData Agent layer is designed to interact with users in natural language, enabling the system to interpret user intent, perform task planning, and execute complex tasks. In BI scenarios, this layer can quickly form a specialized BI Agent capable of handling specific user queries and requests, making the system adaptable and flexible to different business intelligence needs. The atomic tool layer is composed of the essential components required for data extraction, analysis, insights generation, and visualization, among others. Each individual tool within this layer is responsible for completing a specific atomic function in the BI process, ensuring that tasks such as data visualization, statistical analysis, and insights generation are carried out effectively. Based on these three modules, users can interact with the system through natural language conversations with the intelligent Agent to complete tasks like data analysis and visualization. As illustrated in the framework, this approach can replace traditional BI development and delivery processes driven by data engineers, significantly improving development efficiency. The system can provide analytical results to users with very low latency, allowing for faster decision-making based on these insights and supporting quicker, data-driven business decisions.

8. Applications of Large Models in Business Data Analysis

The role of the data analysis module is to identify the user's intent for data analysis based on the instructions provided and automatically generate executable SQL queries. The implementation of this module consists of four key components: intent optimization, metadata enhancement, Text2SQL, and SQL optimization. Intent optimization focuses on addressing the fact that user queries are often conversational in nature and are frequently accompanied by specific business context. The intent optimization component is responsible for generating clear and precise query instructions, providing more understandable inputs for the Text2SQL model. Metadata enhancement comes into play in certain business scenarios where the length of database table schemas may exceed the maximum input limit that the large model can handle. Additionally, the model's ability to generate SQL diminishes as the context length increases. The metadata enhancement component automatically adjusts the input sequence length to overcome the limitation of the model's supported input length. Text2SQL is the component responsible for reading the user's instructions along with the database table schema, then generating accurate and executable SQL statements. By fine-tuning a large model built on a vast dataset of both open-source and domain-specific SQL data, the model gains excellent generalization capabilities as well as the ability to generate complex SQL queries. SQL optimization is necessary because the majority of training data consists of generic SQL statements, so the Text2SQL component is primarily capable of generating generic SQL queries. However, certain business environments may require support for different SQL dialects, such as ClickHouse, SuperSQL, and others. Moreover, considerations regarding the executability and performance efficiency of the generated SQL queries must also be taken into account. To address this, a unified SQL optimization component is designed to support SQL syntax conversion, SQL correction, and SQL performance optimization, ensuring that the generated SQL queries are not only syntactically correct but also optimized for execution efficiency[5].

9. Challenges and Countermeasures

With the increase in data volume and the rise in data value, data privacy and security have become the primary concerns. When large models handle a large amount of sensitive business data,

it is necessary to ensure that the collection, storage, transmission, and use of data comply with laws, regulations, and ethical norms to prevent data leakage and abuse. For enterprise decision-makers, they need to know how the analysis results are derived in order to make reasonable judgments and decisions. Improving the interpretability of large models is an urgent problem to be solved. The development and application of large models require profound technical skills and rich practical experience. Currently, there is a shortage of compound talents who understand both business and technology in the market, which to a certain extent restricts the wide application of large models in commercial data analysis. Constructing and running large models require a large amount of computing and storage resources, and a significant amount of time and funds are needed for model training and optimization, which may be a heavy burden for small and medium-sized enterprises.

Enterprises should establish a complete data governance system, strengthen the training of employees' data security awareness, and formulate strict data usage norms and processes. Researchers should be committed to developing more transparent and interpretable large model algorithms and architectures. Through visualization tools and feature importance analysis methods, they can help users understand the decision logic of the model. In the process of model development, developers should pay attention to communication and cooperation with business experts to ensure that the model results are in line with business common sense and logic. Enterprises increase investment in talent cultivation and, through internal training, external recruitment, and industry-university-research cooperation methods, cultivate and introduce professional talents with large model application capabilities. Enterprises should establish cross-departmental data analysis teams to promote the deep integration of technology and business.

10. Conclusion

The intelligent commercial data analysis method driven by large models has brought rare opportunities and challenges to enterprises. Large models have prominent advantages in commercial data analysis, can deeply explore data value, provide accurate predictions, and offer strong support for decision-making. Judging from the application practices of cases such as Tencent Data Governance Platform and Tianqiong ChatBI, large models have achieved remarkable results in data governance, user profiling, risk assessment, and natural language interaction. If large models are to be widely used and continuously developed in commercial data analysis, numerous technical, talent, and management issues must be solved. Looking forward to the future, with the continuous innovation of technology and the continuous expansion of application scenarios, the role of large models in the field of commercial data analysis will become increasingly crucial and create more considerable commercial value for enterprises. It is believed that in the near future, large models will become a powerful driving force for the development of enterprises.

References

- [1] Li Xiaoming. Application Exploration of Big Data-based Business Intelligence System in E-commerce Data Analysis [J]. *Microcomputer*, 2024, (04): 40-42.
- [2] Ji Hongyi. Application and Impact Analysis of Artificial Intelligence and Big Data in News Dissemination, Copyright Protection, and Media Asset Management Commercial Innovation [J]. *Journalism & Communication*, 2023, (24): 33-35.
- [3] An Ran, Chu Jihua, Hong Xianfeng. Research on the Framework of Intelligence Analysis Method System for Unstructured Data [J]. *Information Studies: Theory & Application*, 2024, 47(02): 143-150.
- [4] Gao, Chao, Lin, Hong-Lei, Hu, Hai-Qing, et al. A review of models of forest fire occurrence prediction in China [J]. *The journal of applied ecology*, 2020, 31(09): 3227-3240.
- [5] Mattia Bazzoni. *The Geographical Expansion of Management Consulting Firms* [D]. Shanghai: Fudan University, 2012.