

# Big Data Intelligent Analysis and Data Mining

Guanchen Zhuang

Sonengda Shanghai, Shanghai 200235, China

**Abstract:** *The advent of the era of big data has brought great changes to people's lives, but also provided unprecedented opportunities for enterprises and organizations. However, in the face of massive data, how to extract valuable information has become an urgent problem to be solved. In this issue, big data intelligent analysis and data mining technology play an important role. This paper starts with the intelligent analysis technology of big data and analyzes the big data information mining technology for its reference.*

**Keywords:** Big data; Intelligence; Technology; Information mining.

## 1. INTRODUCTION

Big data intelligent analysis refers to the use of various artificial intelligence and machine learning algorithms to process and analyze big data, discovering hidden patterns, trends, and correlations to assist decision-making and predict the future. It can not only significantly improve the efficiency of data analysis, but also discover information that is difficult for humans to discover, providing powerful reference for enterprises and organizations. Data mining refers to the process of extracting useful information and knowledge from big data to aid decision-making and problem-solving. Data mining techniques can use clustering, classification, association rules, and other methods to identify relationships and patterns between data, and apply them to different fields. For example, in the financial field, data mining techniques can help detect fraudulent behavior; In the field of marketing, data mining can help businesses identify potential customers and accurately locate the market. Yang et al. [1] developed innovative dynamic hedging strategies for derivatives markets using LLM-driven sentiment analysis. Healthcare applications show particular promise, with Ming et al. [2] demonstrating the feasibility of telemedicine for pediatric care and Yuan et al. [3] advancing tumor classification through self-supervised multimodal learning. Legal technology innovations include Wang et al. [4]'s work on automated regulatory compliance using explainable LLMs. Sustainable development has benefited from several AI applications. He et al. [5] explored material selection for carbon reduction in construction, while Xu et al. [6] optimized healthcare facility design using graph convolutional networks. Feng et al. [7] further contributed to this field by enhancing energy efficiency in green buildings through AI. In healthcare monitoring, Lin et al. [8] investigated intelligent exercise tracking for children with ADHD, complemented by Peng et al. [9]'s research on exercise impacts on cognitive function. Logistics automation has progressed through Luo et al. [10]'s transformer-GCN hybrid algorithm for robot path planning. Consumer behavior studies in electric vehicles have yielded important insights, with Xu et al. [11] examining user-perceived value and Xu et al. [12] developing experience management tools. Clinical AI applications continue to expand, as shown by Shen et al. [13]'s LSTM-based anesthetic dosing system and Xu et al. [14]'s analysis of adversarial attacks in cybersecurity.

## 2. BIG DATA INTELLIGENT ANALYSIS TECHNOLOGY

### 2.1 Big Data Technology

Big data technology refers to a series of techniques and tools used to process and analyze large-scale, high-speed, and diverse data. It involves the collection, storage, processing, and analysis of data, aiming to extract valuable information and insights from massive amounts of data. The importance of big data technology will be explained from three aspects: data collection and storage, data processing and analysis, and data security and privacy protection.

#### 2.1.1 Data Collection and Storage

With the rapid development of the Internet and the popularization of Internet of Things technology, massive data sources continue to emerge. Big data technology provides efficient data collection and storage solutions, capable of receiving various types and sources of data, and integrating and storing them. This includes structured data, semi-structured data, and unstructured data. Through big data technology, high availability, scalability, and redundancy of data can be ensured to meet the needs of large-scale data storage and access.

### 2.1.2 Data Processing and Analysis

Big data technology provides efficient data processing and analysis capabilities, making it feasible and efficient to process and analyze large-scale data. Traditional data processing and analysis methods are unable to handle such tasks, while big data technology, based on distributed computing, parallel processing, and high-performance hardware, can quickly process massive amounts of data. Through big data technology, real-time analysis, batch processing, and streaming processing can be achieved, supporting applications such as data mining, machine learning, and artificial intelligence. These analysis results can help enterprises carry out refined marketing, intelligent decision-making, and business optimization, enhancing competitiveness and efficiency.

### 2.1.3 Data Security and Privacy Protection

With the generation and use of large-scale data, data security and privacy protection have become important concerns. Big data technology provides various mechanisms and tools to protect the security and privacy of data. This includes technologies such as data encryption, access control, identity authentication, and data anonymization to prevent data leakage, unauthorized access, and improper use. At the same time, big data technology can also respond to data risks and threats, promptly identifying and addressing potential security issues. Data security and privacy protection are important considerations in big data technology, providing guarantees for the use of data and enhancing the trust of users and enterprises in big data.

Big data technology plays an important role in modern data processing and analysis. Through efficient data collection and storage, big data technology can process and store large-scale, high-speed, and diverse data. Through efficient data processing and analysis, big data technology can extract valuable information and insights from massive amounts of data. Through strict data security and privacy protection measures, big data technology can ensure the security and credibility of data.

## 2.2 Artificial Intelligence Technology

Artificial intelligence technology is a branch of computer science aimed at developing and building systems with intelligent behavior. Artificial intelligence technology imitates human intelligence and behavior, enabling computers to perceive, learn, reason, and make decisions like humans. The importance of artificial intelligence technology will be explained from four aspects: machine learning, natural language processing, computer vision, and expert systems.

### 2.2.1 Machine Learning

Machine learning is the core of artificial intelligence technology, which improves its performance and accuracy by allowing computers to learn and discover patterns from data. Machine learning algorithms can automatically extract features from large amounts of data and make predictions, classifications, and decisions through data analysis and pattern recognition. Machine learning enables computers to further improve algorithms and models through self-learning and optimization, enhancing the efficiency and accuracy of task execution. Machine learning has a wide range of applications in various fields, such as personalized recommendation, fraud detection, medical diagnosis, etc.

### 2.2.2 Natural Language Processing

Natural language processing is an artificial intelligence technology that studies the interaction and understanding between computers and human language. It involves speech recognition, semantic understanding, machine translation, sentiment analysis, and other aspects. Through natural language processing technology, computers can recognize and understand human natural language, process and analyze text and speech, and perform tasks such as text generation and question answering systems. The application fields of natural language processing are very extensive, such as intelligent assistants, intelligent customer service, intelligent translation, etc., greatly improving the efficiency and naturalness of human-computer interaction.

### 2.2.3 Computer Vision

Computer vision is an artificial intelligence technology that studies the visual understanding and analysis of computers through images and videos. It involves image recognition, object detection, image generation, and other aspects. Through computer vision technology, computers can simulate the human visual system, recognize and understand the content in images, and perform tasks such as object recognition, scene understanding, and face recognition. Computer vision technology has a wide range of applications in security monitoring, autonomous driving, medical imaging, and other fields, providing basic support for automation and intelligence.

#### 2.2.4 Expert System

Expert system is a computer system constructed based on artificial intelligence technology and domain expert knowledge. It enables the system to simulate the decision-making process and problem-solving ability of experts by programming their knowledge into the system. Expert systems solve domain specific problems through rule engines and inference mechanisms, providing decision support and professional advice. The application of expert systems includes fields such as medical diagnosis, fault diagnosis, and financial risk assessment, playing an important role in improving intelligent decision-making and knowledge management.

### 2.3 Natural language processing capabilities

#### 2.3.1 Text Processing

Natural language processing can process and analyze textual data. Through text processing technology, computers can recognize and understand the grammar, semantics, and structure of text data, and perform tasks such as part of speech tagging, word segmentation, and stem extraction. In addition, text processing can also be used for text classification, sentiment analysis, entity recognition, etc., to extract useful information and insights from text data. In the business field, text processing can be applied to tasks such as public opinion analysis, user comment sentiment analysis, news summarization, etc., helping companies better understand user needs and market dynamics.

#### 2.3.2 Speech Recognition

Speech recognition is one of the important functions in natural language processing, which enables computers to recognize and understand human speech input. Through speech recognition technology, computers can convert users' voice input into text form and perform subsequent text processing and analysis. Speech recognition technology has been widely applied in fields such as voice assistants (such as Siri, Alexa), voice interaction systems, and telephone customer service, improving the convenience and naturalness of human-computer interaction.

#### 2.3.3 Semantic Understanding

Semantic understanding is an important technology in natural language processing, aimed at enabling computers to understand the meaning and context of human language. Through semantic understanding technology, computers can extract key information from text and understand its meaning and context. This involves tasks such as ambiguity resolution of word meanings, analysis of grammatical structures, and annotation of syntactic roles. Semantic understanding plays an important role in applications such as information retrieval, question answering systems, and machine translation, improving the system's accurate understanding of user intent and the precision of information retrieval.

#### 2.3.4 Machine Translation

Machine translation is an important application of natural language processing, aimed at automatically converting text from one language to another. Through machine translation technology, computers can understand text in the source language and convert it into text in the target language, achieving cross lingual communication and understanding. Machine translation technology has been widely applied in business communication of multinational corporations, online translation platforms, and the tourism industry. In recent years, with the development of deep learning and neural network technology, machine translation based on neural networks has made breakthrough progress, improving translation quality and efficiency.

### 3. BIG DATA INFORMATION MINING TECHNOLOGY

#### 3.1 Data Information Mining Technology

Data preprocessing is the first step in data information mining, used to clean, organize, and prepare data for subsequent analysis. Data preprocessing includes tasks such as data cleaning, missing value processing, outlier detection, and data integration. Through data preprocessing, the quality and accuracy of data can be improved, and the interference of data noise and errors on the analysis results can be reduced. Data preprocessing is crucial for the reliability and effectiveness of data information mining. Association analysis is a commonly used technique in data information mining, used to discover association rules and patterns in data. Association analysis can reveal hidden relationships and co-occurrence patterns in data, such as in shopping basket analysis where "users who purchase product A also frequently purchase product B. Through correlation analysis, enterprises can understand the correlation between products or services, optimize recommendation systems, improve marketing strategies, increase sales volume and user satisfaction. Classification and prediction are important tasks in data information mining, aiming to establish models based on the characteristics and categories of known data to predict the category of unknown data. Classification and forecasting can help businesses with market segmentation, customer segmentation, risk assessment, and more. For example, by establishing a user purchase model, companies can predict products that users may be interested in based on their historical purchase records, thereby providing personalized recommendations and targeted marketing [5]. Clustering is another important technique in data information mining, used to group similar data points in a dataset into different categories. Clustering can help companies discover potential patterns and group structures in data, conduct market segmentation, user analysis, and product positioning. For example, in customer relationship management, by clustering user behavior data, enterprises can identify user groups with similar purchasing behaviors and needs, and provide them with more personalized services and recommendations.

#### 3.2 Big Data Information Mining Technology

##### 3.2.1 Specific operational methods for big data information mining

Big information mining is the process of extracting valuable information and insights from large-scale datasets by analyzing and mining patterns, relationships, trends, and knowledge. Here are some specific methods for big data information mining: Firstly, it is necessary to collect large-scale data and clean and prepare it. This includes deleting duplicate data, handling missing values, handling outliers, etc., to ensure the quality and accuracy of the data. This step also includes selecting appropriate datasets and features for analysis. Before conducting data mining, exploratory analysis of the data is necessary to understand its basic characteristics and structure. This includes statistical description, data visualization, correlation analysis, etc. Exploratory analysis of data can help select appropriate mining methods and discover potential correlations and trends. Text mining refers to extracting useful information and knowledge from a large amount of textual data. It includes tasks such as text classification, sentiment analysis, and entity recognition.

##### 3.2.2 Types of Big Data Information Mining

###### (1) Text mining

Text mining is another important field of big data information mining, focusing on extracting useful information and knowledge from large amounts of textual data. Text data refers to textual data expressed in natural language form, such as news articles, social media posts, emails, etc. Text mining tasks include text classification, sentiment analysis, topic modeling, named entity recognition, etc. These tasks involve technologies such as natural language processing, machine learning, and statistical analysis, which can help businesses analyze user reviews, public opinion analysis, market research, and more.

###### (2) Graph mining

Graph mining refers to the technique of extracting valuable information and patterns from large graph data. In the era of big data, network data such as social networks, knowledge maps, Internet page links, etc. can be represented as graph structures. Graph mining includes tasks such as graph analysis, graph clustering, and graph classification. For example, social network analysis can help businesses make social recommendations, influence marketing, and more by mining communities, influential nodes, and important relationships in social networks.

### (3) Social Network Analysis

Social network analysis is a technique used to explore and analyze interpersonal relationships, social influence, and information dissemination from large-scale social network data. With the rapid development of social media, the data on social networks is growing rapidly. Social network analysis can help businesses understand user social relationships, identify opinion leaders in social networks, and analyze the process of information dissemination. The technologies include node centrality measurement, community detection, network influence analysis, etc. These analysis results can be used for applications such as recommendation systems, social advertising, and public opinion monitoring.

### (4) Abnormal detection

Anomaly detection is an important task in big data information mining, aimed at identifying anomalies or abnormal patterns in data. By analyzing a large amount of data, data points or behaviors that do not conform to the normal pattern can be discovered. Anomaly detection has important applications in financial fraud detection, network security, equipment failure warning, and other fields. For example, in the financial sector, anomaly detection can help banks detect suspicious transactions and fraudulent behavior in a timely manner, protecting customers and businesses from losses.

## 4. CONCLUSION

In short, the arrival of the big data era has made intelligent analysis and data mining technologies particularly important. They can improve the accuracy and efficiency of decision-making, identify potential business opportunities and risks, and enhance the competitiveness of enterprises. However, we also need to face challenges in terms of data quality and security, technology and talent, law and ethics. Only by overcoming these challenges can we better utilize the power of big data intelligent analysis and data mining, bringing more opportunities and benefits to our society and economic development.

## REFERENCES

- [1] Yang, Jie, et al. "Dynamic Hedging Strategies in Derivatives Markets with LLM-Driven Sentiment and News Analytics." arXiv preprint arXiv:2504.04295 (2025).
- [2] Ming DY, Li T, Ross MH, et al. Feasibility of post-hospitalization telemedicine video visits for children with medical complexity. *J Pediatr Health Care*. 2022;36(2):e22–e35
- [3] Yuan, J. (2025). Self-Supervised Multimodal Learning for Tumor Classification in Chest Radiography. Authorea Preprints.
- [4] Wang, J., Yuan, J., Liu, J., & Evans, L. (2025). Simple Legal Compliance: Automating Regulatory Audits with Explainable LLMs.
- [5] He, J., Meng, Q., Xu, H., & Liu, Y. (2025, January). Exploring Material Selection and Applications for Embedded Carbon Reduction in the Built Environment. In The 1st International scientific and practical conference "Technologies for improving old methods, theories and hypotheses"(January 07–10, 2025) Sofia, Bulgaria. International Science Group. 2025. 405 p. (p. 62).
- [6] Xu, Haoran. "Sustainability Enhancement in Healthcare Facility Design: Structural and Functional Optimization Based on GCN." (2025).
- [7] Feng, Zhang, Minyue Ge, and Qian Meng. "Enhancing energy efficiency in green buildings through artificial intelligence." *Applied Science and Engineering Journal for Advanced Research* 3.5 (2024): 10-17.
- [8] Lin, L., Li, N., & Zhao, S. (2025). The effect of intelligent monitoring of physical exercise on executive function in children with ADHD. *Alexandria Engineering Journal*, 122, 355-363.
- [9] Peng, Y., Zhang, G., & Pang, H. (2025). Impact of Short-Duration Aerobic Exercise Intensity on Executive Function and Sleep. arXiv preprint arXiv:2503.09077.
- [10] Luo, H., Wei, J., Zhao, S., Liang, A., Xu, Z., & Jiang, R. (2024). Intelligent logistics management robot path planning algorithm integrating transformer and gcnn network. *IECE Transactions on Internet of Things*, 2(4), 95-112.
- [11] Xu, Y., & Lin, Y. (2024, November). Exploring the Influence of User-Perceived Value on NEV-Enterprises Using an Empirical Computer Model. In 3rd International Conference on Financial Innovation, FinTech and Information Technology (FFIT 2024) (pp. 4-10). Atlantis Press.

- 
- [12] Xu, Y., Shan, X., Guo, M., Gao, W., & Lin, Y. S. (2024). Design and application of experience management tools from the perspective of customer perceived value: A study on the electric vehicle market. *World Electric Vehicle Journal*, 15(8), 378.
  - [13] Shen, Z., Wang, Y., Hu, K., Wang, Z., & Lin, S. (2025). Exploration of Clinical Application of AI System Incorporating LSTM Algorithm for Management of Anesthetic Dose in Cancer Surgery. *Journal of Theory and Practice in Clinical Sciences*, 2, 17-28.
  - [14] Xu, J., Wang, Y., Chen, H., & Shen, Z. (2025). Adversarial Machine Learning in Cybersecurity: Attacks and Defenses. *International Journal of Management Science Research*, 8(2), 26-33.