Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/67/20240471

Applications of knowledge graph in medical and financial fields: Data integration and intelligent decision-making from an interdisciplinary perspective

Ruichen Xi

Huazhong Agricultural University, Wuhan, Hubei, 430070, China

chen77687768@outlook.com

Abstract. With the rapid advancement of AI and deep learning technologies, knowledge graphs have emerged as a key technology for improving the performance of intelligent decision-making systems and driving interdisciplinary innovation. This article outlines the core principles and structure of knowledge graphs, including how they construct knowledge networks to support complex queries and intelligent reasoning. It reviews their innovative applications in the healthcare and financial industries, emphasizing their significant roles in data integration, decision support, and risk assessment. In the healthcare domain, knowledge graphs contribute to improving the accuracy of medical diagnoses, accelerating drug discovery, and enabling intelligent semantic searches. In the financial sector, they optimize risk management and aid in fraud prevention. The article also looks ahead to the future potential of knowledge graphs, stressing the importance of interdisciplinary collaboration and technological innovation in their development. It aims to provide valuable references for further research and application of knowledge graphs.

Keywords: Knowledge Graphs, Artificial Intelligence, Machine Learning, Medical Applications, Financial Risk Management.

1. Introduction

In the era of big data, knowledge graphs have gradually become a key tool for data management and analysis. Knowledge graphs were officially introduced by Google on May 17, 2012 [1], originating from the semantic web [2] and ontology, representing an evolution and refinement of semantic web standards and technologies[3]. They provide a structured framework for understanding complex and dynamic information environments, aiding in the excavation and utilization of vast data resources across various industries.

In recent years, with the rapid development of machine learning and deep learning technologies, knowledge graphs have gradually become a foundational infrastructure for knowledge-based intelligent services on the Internet [4], finding applications across various industries. This paper will focus on the medical and financial sectors, where knowledge graphs have been extensively applied, to review the research and applications of knowledge graph technology within these areas. By comparing and analyzing the methods by which knowledge graphs process and analyze information in specific

DOI: 10.54254/2755-2721/67/20240471

application domains, this paper demonstrates how they realize the true value of knowledge representation across different industrial fields.

2. Fundamental Principles of Knowledge Graph Technology

The core components of knowledge graphs include entities, relationships, and attributes. Entities are the basic units of the graph, typically representing objects or concepts from the real world. Relationships define the connections between entities, while attributes provide descriptive details for the entities. Additionally, the concept of ontology is closely related to these three elements, serving as an explicit specification of a shared conceptual model [5]. It outlines a set of terms and concepts along with their interrelationships, offering a framework to organize and interpret entities, relationships, and attributes within knowledge graphs, thus ensuring their structural integrity and consistency.

The construction of knowledge graphs involves extracting and integrating knowledge from various data sources. This process is usually carried out in a bottom-up manner, including importing knowledge from structured sources or extracting entities, relationships, and attributes from unstructured documents (such as news articles and research papers). After acquiring new knowledge, it is integrated to resolve contradictions and ambiguities. Finally, qualified knowledge is selectively incorporated into the repository through manual screening or quality assessment [6].

Table 1. Overview of Knowledge Graph Construction Technologies

Technology	Description	Core Feature
Knowledge Extraction	Extracts key information from unstructured data.	Handles large-scale text data.
Entity Recognition	Identifies specific entities in text.	Improves data structuring.
Relationship Extraction	Discovers semantic relationships between entities.	Enhances graph semantic richness.
Attribute Extraction	Extracts descriptive attributes of entities.	Enriches entity information.
Knowledge Fusion	Merges knowledge from different sources, resolves ambiguities.	Increases graph accuracy and consistency.

Additionally, the implementation of knowledge graphs cannot be separated from updates and maintenance to adapt to the dynamic changes in information, data, and knowledge. Currently, updates to knowledge graphs are primarily categorized into manual updates and automatic updates, which utilize the timestamps or geolocation information retained within the knowledge graph [7]. Effective updates and maintenance involve continuous updates and quality control to ensure that the knowledge graph reflects the most recent and accurate information.

3. Application of Knowledge Graphs in the Medical Field

3.1. Characteristics and Processing Requirements of Medical Data

Medical data is diverse and complex, including but not limited to Electronic Health Records (EHRs), medical imaging, laboratory test results, genetic information, and more. These data types involve structured, semi-structured, and unstructured data [8], each with its unique storage, management, and analysis requirements, adding to the complexity of medical data processing. Moreover, the quality and accuracy of data are crucial for clinical decision-making, but medical data often has issues with missing, erroneous, or inconsistent information [9], which are distributed across various information systems without a unified standard.

DOI: 10.54254/2755-2721/67/20240471

Table 2. Characteristics of Medical Data Requirements

Data Type	Data Type
Electronic Health Records (EHR)	Requires systematic storage and management for easy retrieval and analysis
Medical Imaging	Requires large-capacity storage and advanced image analysis tools
Laboratory Test Results	Standardized storage formats to ensure accuracy and comparability
Genetic Information	Requires complex data processing and analysis capabilities, storage of large amounts of data
Others (such as case reports, patient feedback)	Requires text analysis and natural language processing techniques to transform into analyzable information

3.2. Specific Application Studies

Knowledge graph technology has had a profound impact on the medical field, whether it's improving diagnostic accuracy, accelerating drug discovery, or optimizing intelligent question-answering systems. Its capability for deep integration and analysis of data has introduced new possibilities to the healthcare industry.

Knowledge graphs can be applied to disease diagnosis and clinical decision support. Medical professionals can use the structured data and relationships within the graph to enhance diagnostic accuracy and provide treatment decision recommendations. Cai Xi [10] proposed a disease diagnosis method that combines medical knowledge graphs with deep learning, named CKGDL, which obtains structured disease knowledge from medical knowledge graphs through entity linking disambiguation and knowledge graph embedding and extraction. Qiu Yongjian and others [11] proposed a visualization algorithm model based on deep learning, which uses disease feature word vectors from disease description texts and corresponding knowledge entity vectors as multi-channel inputs to convolutional neural networks through knowledge graph embedding and extraction. Yin Yating and others [12] used knowledge graphs as structured data sources to provide high-quality knowledge information for a medical question-answering system focused on hepatitis B. The system, containing a total of 8,563 entities, 96,896 relationships, 32 types of entities, and 40 types of relationships, integrates infectious disease diagnosis guidelines with literature and real medical record knowledge graphs to generate a rule library for infectious disease monitoring and early warning. Using interrupted time series analysis, the system reduced the misdiagnosis rate by an average of 4.4037%.

Knowledge graph technology can be applied to drug discovery and medical research. With knowledge graphs, we can identify potential therapeutic compounds and understand their mechanisms of action, mapping interactions between drugs and biological entities to simplify the drug development process. Lou Pei and others [13] discovered potential targets and candidate drugs for reanalyzing coronaviruses using a knowledge graph-based method. By semantically mapping literature knowledge with existing drugs and genetic knowledge, the coronavirus knowledge graph (CovKG) was constructed, demonstrating that learning effective molecular feature knowledge representations to promote molecular property prediction is significant for drug discovery. Zhu Zhaocheng and others [14] developed TorchDrug, a powerful and flexible drug discovery machine learning platform built on PyTorch. TorchDrug leverages knowledge graph technology for important tasks in drug discovery, such as molecular property prediction, pre-trained molecular representation, retrosynthesis prediction, biomedical knowledge graph reasoning, and priority ranking of target genes for diseases, optimizing the drug development process.

Knowledge graphs can be applied to medical intelligent semantic search and question-answering. Knowledge graphs can organize a vast amount of medical data in a way that is efficiently searchable and retrievable, facilitating professionals and patients to access relevant information and improving medical information retrieval and complex query answering. Liang Min [15] used web crawlers to extract knowledge from authoritative medical websites and Baidu Encyclopedia, integrating it with cardiovascular disease medical textbooks using the Neo4j graph database for knowledge storage, constructing a cardiovascular disease knowledge graph. Faced with the massive amount of information generated daily by medical Q&A websites, Li Yaliang and others [16] proposed a Medical Knowledge Extraction (MKE) system that can automatically provide high-quality knowledge triplets extracted from noisy Q&A pairs. Quantitative evaluations and case studies have shown that the MKE system can successfully provide effective and accurate medical professional knowledge.

4. Application of Knowledge Graphs in the Financial Sector

4.1. Complexity of Financial Data and the Value of Knowledge Graphs

The complexity of financial data stems from its vast volume, diverse data types, and rapidly changing market conditions. Knowledge graph technology can visualize these complex data relationships, breaking the limitations of traditional data storage, and transforming originally discrete multi-source heterogeneous data into a unified structured form. This provides strong data support and technological improvements for the financial sector [17].

4.2. Specific Application Studies

In the financial sector, the application of knowledge graphs is becoming increasingly diverse and profound, not only playing a core role in traditional risk assessment and investment management but also showing great potential in emerging areas such as financial fraud prevention and intelligent decision support.

Table 2. Main Applications of Knowledge Graphs in the Financial Sector

Application Area	Purpose	Technologies/Methods	Technologies/Methods
Risk Management	Identify and assess financial risks	Knowledge graph construction,data visualization, association analysis	Improves the accuracy and efficiency of risk assessment, better monitoring and prevention of systemic risks
Investment Management	Analyze asset correlations and market conditions	Graph databases, semantic analysis, machine learning	Supports decision-making, optimizes investment portfolios, enhances the robustness of investment strategies
Financial Fraud Prevention	Detect and prevent financial fraud activities	Anomaly detection algorithms, graph analysis techniques	Increases the speed and accuracy of fraud detection, reduces financial losses

Financial institutions use knowledge graphs to identify, assess, and monitor risks. In investment management, knowledge graphs are employed to analyze the correlations between assets and market conditions, aiding in the formulation of more robust investment strategies. Liu Fang and others [18] utilized knowledge graphs for a visual analysis of systemic financial risk research in China from 2010 to 2020, employing Citespace V to mine research hotspots, delineate its evolutionary paths, and analyze its research trends. On this basis, a systemic integration framework was constructed, propelling the further development of systemic financial risk research. Yerashenia and others [19] proposed a novel intelligent approach to constructing bankruptcy prediction computational models, comprising three layers: a bankruptcy prediction ontology, a semantic search engine, and a semantic analysis graph database system. The results indicate that this method, leveraging advanced semantic data management mechanisms, can process data and perform relevant calculations more effectively than methods using traditional relational databases.

In the prevention of financial fraud, knowledge graphs identify abnormal activities by analyzing transaction patterns, customer behavior, and associated networks. For example, by analyzing transaction frequency, amounts, locations, and similarities to known fraud cases, knowledge graphs can discover potential fraudulent behavior and trigger alerts. Zhao Gang and others [20] discussed an ontology-based knowledge engineering approach to combat financial fraud in information systems. The forensic ontology, developed based on laws, regulations, and cases related to the illegal solicitation of financial products online, employs fuzzy clustering and expert reasoning to identify companies affected by fraudulent financial reporting. Shen Yuming and others [21] combined traditional features with knowledge graphs and explored enriched representations through feature embedding across various financial categories for financial statement fraud detection. Experiments show that financial feature representations enriched with relevant information significantly enhance the classification performance of SVM and K-NN, slightly outperforming decision trees and logistic regression.

5. Conclusion

This article provides a brief introduction to knowledge graph technology and interdisciplinary reviews of its applications in the medical and financial sectors. Through comparative analysis, it reveals the roles of knowledge graphs in data integration, decision support, and risk assessment, along with sector-specific application strategies. We find that knowledge graphs can build a cross-domain, structured knowledge base, offering a flexible and powerful way to represent, integrate, and share knowledge across domains. This not only promotes data integration and knowledge sharing across different fields but also opens new pathways for scientific research, technological innovation, and knowledge-driven applications. It provides a powerful tool for data management, analysis, decision support, and risk assessment across various fields, helping to address many disciplinary bottlenecks [22] and showing a broad range of application prospects and significant benefits.

At the same time, we must recognize that knowledge graph technology still faces many challenges, including ensuring data quality, handling data heterogeneity, managing large-scale data, achieving dynamic updates and maintenance, maintaining semantic consistency, and ontology management. To address these, we need to adopt comprehensive solution strategies, improving the system's performance and reliability through a series of complementary technical means. For example, ensuring the accuracy of input data through data cleaning and quality control, solving data heterogeneity issues using middleware and ontology mapping techniques, and effectively managing massive data using distributed storage and big data technologies. The integrated application of these strategies allows knowledge graphs to maintain their value and utility in a constantly changing information environment.

Future research needs to continue exploring more efficient knowledge extraction methods, more accurate data fusion technologies, and more flexible ontology modeling methods to meet the evolving application demands. Additionally, with the advancement of artificial intelligence and machine learning technologies, this article anticipates more research on using knowledge graphs to enhance the semantic understanding and reasoning capabilities of AI models. The deep integration of knowledge graphs with these technologies will open new research directions and application scenarios, enabling knowledge graphs to play a greater role in intelligent decision support systems. This article looks forward to the continuous development of knowledge graph technology, bringing revolutionary and profound changes to more fields.

References

[1] Amit Singhal. Introducing the knowledge graph. Official Blog of Google, America, 2012.
[2] Tim Berners-Lee, James Hendler, Ora Lassila. The Semantic Web. Scientific American Magazine, 2008, 23(1): 1-4.

[3] Zenglin Xu, Yongpan Sheng, Lirong He, et al. A Survey of Knowledge Graph Techniques. Journal of University of Electronic Science and Technology of China, 2016, 45(04): 589-606.

[4] Juanzi Li, Lei Hou. A Survey on Knowledge Graph Research. Journal of Shanxi University (Natural Science Edition), 2017, 40(03): 454-459. DOI: 10.13451/j.cnki.shanxi.univ(nat.sci.).2017.03.008.

[5] Thomas R. Gruber. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 1993, 5(2): 199–220.

[6] Qiao Liu, Yang Li, Hong Duan, et al. A Survey of Knowledge Graph Construction Techniques. Computer Research and Development, 2016, 53(03): 582-600.

[7] Tao Li, Cichen Wang, Huakang Li. The Development and Construction of Knowledge Graphs. Journal of Nanjing University of Science and Technology, 2017, 41(01): 22-34. DOI: 10.14177/j.cnki.32-1397n.2017.41.01.004.

[8] Mengwei Hou, Rong Wei, Liang Lu, Xin Lan, Hongwei Cai. A Survey of Knowledge Graph Research and Its Application in the Medical Field. Computer Research and Development, 2018, 55(12): 2585-2599. https://doi.org/10.7544/issn1000-1239.2018.20180623

[9] Chuanyu Jiang, Xiangyu Han, et al. A Survey of Medical Knowledge Graph Research and Applications. Computer Science, 2021, 50(3), 83-93. https://doi.org/10.11896/jsjkx.220700241

[10] Xi Cai. A Novel Disease Diagnosis Method Using Combining Knowledge Graph and Deep Learning. Journal of Medical Imaging and Health Informatics, 2021.

[11] Yongjian Qiu, Jing Lu. A Visualization Algorithm for Medical Big Data Based on Deep Learning. Measurement, 2021.

[12] Yating Yin, Lei Zhang, Yiguo Wang, Mingqiang Wang, Qiming Zhang, Guo-Zheng Li. Question Answering System Based on Knowledge Graph in Traditional Chinese Medicine Diagnosis and Treatment of Viral Hepatitis B. Biomed Research International, 2022. (IF: 3)

[13] Pei Lou, An Fang, Wanqing Zhao, Kuanda Yao, Yusheng Yang, Jiahui Hu. Potential Target Discovery and Drug Repurposing for Coronaviruses: Study Involving A Knowledge Graph-Based Approach. Journal of Medical Internet Research, 2023.

[14] Zhaocheng Zhu, Chence Shi, Zuobai Zhang, Shengchao Liu, Minghao Xu, Xinyu Yuan, Yangtian Zhang, Junkun Chen, Huiyu Cai, Jiarui Lu, Chang Ma, Runcheng Liu, Louis-Pascal Xhonneux, Meng Qu, Jian Tang. TorchDrug: A Powerful and Flexible Machine Learning Platform for Drug Discovery. ARXIV-CS.LG, 2022. (IF: 3)

[15] Min Liang. Research on The Construction Technology of Cardiovascular Disease Q&A System Based on Knowledge Graph. International Conference on Intelligent Computing, ..., 2023.

[16] Yaliang Li, Chaochun Liu, Nan Du, Wei Fan, Qi Li, Jing Gao, Chenwei Zhang, Hao Wu. Extracting Medical Knowledge from Crowdsourced Question Answering Website. IEEE Transactions on Big Data, 2020. (IF: 3)

[17] Shuangshuang Qian. Construction and Application of Knowledge Graphs in the Financial Field. Zhejiang University of Technology, 2021. DOI: 10.27463/d.cnki.gzgyu.2020.001189.

[18] Fang Liu, Yi Zhang, Li Li. Review of Systematic Financial Risk Research Based on Knowledge Map. Procedia Computer Science, 2022, 199, 315-322. https://doi.org/10.1016/j.procs.2022.01.039

[19] N. Yerashenia, A. Bolotov. Computational Modelling for Bankruptcy Prediction: Semantic Data Analysis Integrating Graph Database and Financial Ontology. 2019 IEEE 21st Conference on Business Informatics (CBI), Moscow, Russia, 2019, pp. 84-93, doi: 10.1109/CBI.2019.00017.

[20] Gang Zhao, John Kingston, Koen Kerremans, Frederik Coppens, Ruben Verlinden, Rita Temmerman, Robert Meersman. Engineering An Ontology of Financial Securities Fraud, 2004. (IF: 3)

DOI: 10.54254/2755-2721/67/20240471

[21] Yuming Shen, Caichan Guo, Huan Li, Junjie Chen, Yunchuan Guo, Xinying Qiu. Financial Feature Embedding with Knowledge Representation Learning for Financial Statement Fraud Detection. Procedia Computer Science, 2021. (IF: 3)