

人工智能在企业经济数据管理中的应用

——基于知识图谱的智能检索模型

丁雅芳, 魏晓光

河北金融学院河北省金融科技应用重点实验室, 河北 保定

收稿日期: 2025年10月22日; 录用日期: 2025年11月19日; 发布日期: 2025年11月27日

摘要

知识图谱融合人工智能技术, 可实现复杂经济数据的语义关联与智能分析, 在企业数据管理中具有重要价值。本文分析了企业经济数据管理中多源数据整合难、异构信息清洗效率低、查询精度不足及风险预警滞后的现状, 研究了基于知识图谱的智能检索模型构建思路, 提出以知识图谱为语义支撑、融合自然语言理解与语义向量匹配的智能检索架构, 旨在围绕多源异构经济数据的语义融合方法展开研究, 构建统一本体模型与嵌入式表示机制, 实现跨系统数据的标准化与智能可检索性。

关键词

人工智能, 企业经济数据管理, 知识图谱, 智能检索模型

Application of Artificial Intelligence in Enterprise Economic Data Management

—An Intelligent Retrieval Model Based on Knowledge Graph

Yafang Ding, Xiaoguang Wei

Hebei Provincial Key Laboratory of Financial Technology Application, Hebei Finance University, Baoding Hebei

Received: October 22, 2025; accepted: November 19, 2025; published: November 27, 2025

Abstract

The integration of knowledge graph and artificial intelligence technology can achieve semantic correlation and intelligent analysis of complex economic data, which has important value in enterprise data management. This article analyzes the current situation of difficult integration of multi-source data, low efficiency of heterogeneous information cleaning, insufficient query accuracy, and lagging risk

warning in enterprise economic data management. It studies the construction idea of an intelligent retrieval model based on knowledge graph and proposes an intelligent retrieval architecture that integrates natural language understanding and semantic vector matching with knowledge graph as semantic support. The aim is to conduct research on semantic fusion methods for multi-source heterogeneous economic data, construct a unified ontology model and embedded representation mechanism, and achieve standardization and intelligent retrievability of cross system data.

Keywords

Artificial Intelligence, Enterprise Economic Data Management, Knowledge Graph, Intelligent Retrieval Model

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在数字化经济背景下,企业经济数据呈现“海量、异构、碎片化”特征。当前已有研究在知识图谱构建、语义检索优化和风险识别方面取得初步进展,但多数侧重于结构化数据管理,尚缺乏对非结构化与半结构化经济数据的语义处理机制,且普遍缺乏统一本体支撑与风险推理结构的集成方案。基于此,本文提出基于知识图谱的智能检索模型思路,将知识图谱的语义关联能力与人工智能的智能推理技术结合,聚焦多源异构经济数据的语义整合方法,构建本体驱动的结构映射机制与嵌入表达模型,提升数据融合的一致性与语义可计算性,该研究在提升企业经济数据管理效率与决策智能化水平方面具有重要学术与应用价值。

2. 知识图谱相关概述

2.1. 知识图谱的定义

知识图谱是一种以图结构组织语义知识的信息表达形式,核心结构由实体、关系、属性构成。知识图谱基于语义解析的问答技术,通过对用户查询问句进行语义解析获取查询对象、对象约束与用户意图,再根据关联关系形成查询图,最后与知识图谱匹配推理以获取并推送所需知识,其中涉及命名实体识别、实体链接、语义解析等关键技术[1]。在企业经济数据管理中,实体表现为企业主体、合同、财务指标等关键要素,关系体现要素间的逻辑连接,属性描述各要素的特征信息。知识图谱具备可解释性、可扩展性、可视化性、可推理性、可检索性五项能力。如图1所示,可解释性强化数据间逻辑关联展示,可扩展性支持多维信息融合更新,可视化性实现结构关联的直观呈现,可推理性基于语义关系发现潜在规律,可检索性依托语义匹配实现跨系统精准定位。该结构模型以语义网络方式支撑经济数据管理的全流程认知与分析。

2.2. 知识图谱构建流程

知识图谱构建过程包含数据采集、知识抽取、融合处理与语义加工四个阶段。如图2所示,数据采集面向结构化、半结构化与非结构化数据源,通过统一接入工具完成格式解析与初步归集。实体抽取采用深度学习模型识别核心经济要素,关系抽取借助有监督方法提取要素之间的语义关联。融合阶段引入实体对齐与消歧策略解决同源异名现象,提升图谱一致性与准确性。知识加工环节构建本体模型,规范

图谱结构, 辅以规则推理实现语义扩展。最终构建结构完备、语义清晰、可推理的企业经济知识图谱, 为后续智能检索模型提供数据基础与语义支撑。

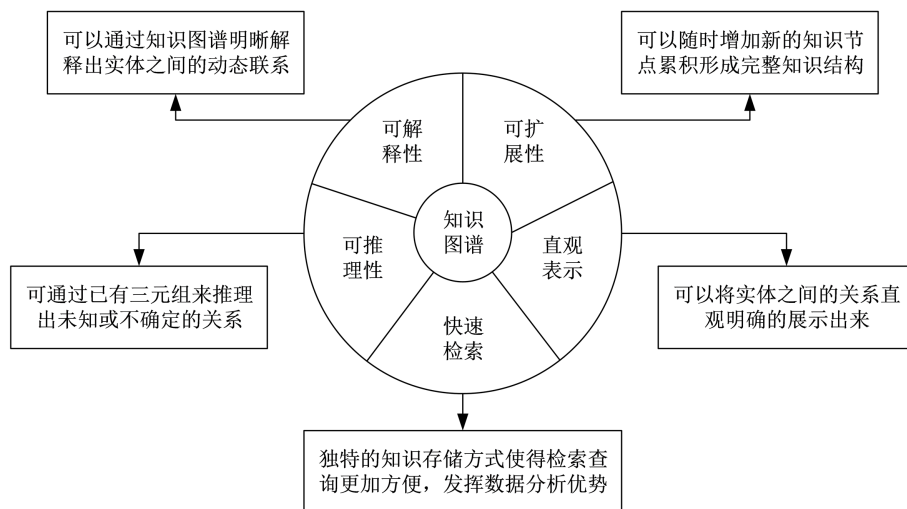


Figure 1. Core feature structure diagram of knowledge graph

图 1. 知识图谱的核心特征结构图

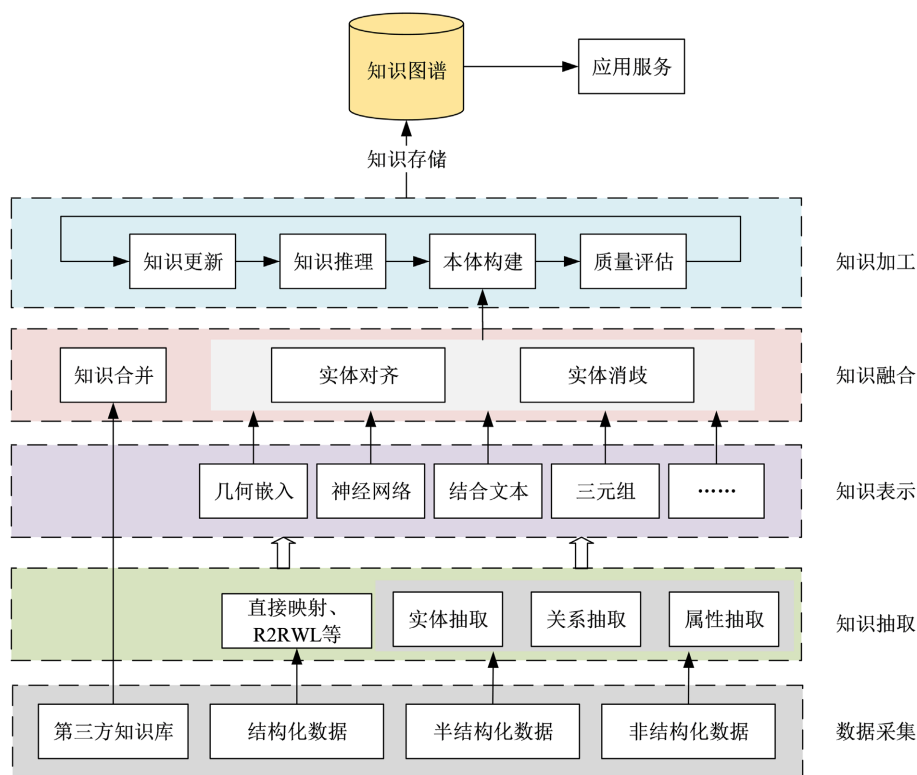


Figure 2. Flow chart of knowledge graph construction

图 2. 知识图谱构建流程图

3. 智能检索模型架构设计

智能检索模型架构由查询、检索、评估、输出四个层面组成, 构建完整的语义驱动数据管理体系。

如图 3 所示, 查询层以自然语言理解模型为核心, 对用户输入的经济数据查询进行语义解析与意图识别, 将自然语言转化为可被知识图谱识别的结构化指令。模型在这一阶段通过语义消歧与上下文分析形成语义向量, 使系统能够准确理解查询目标。

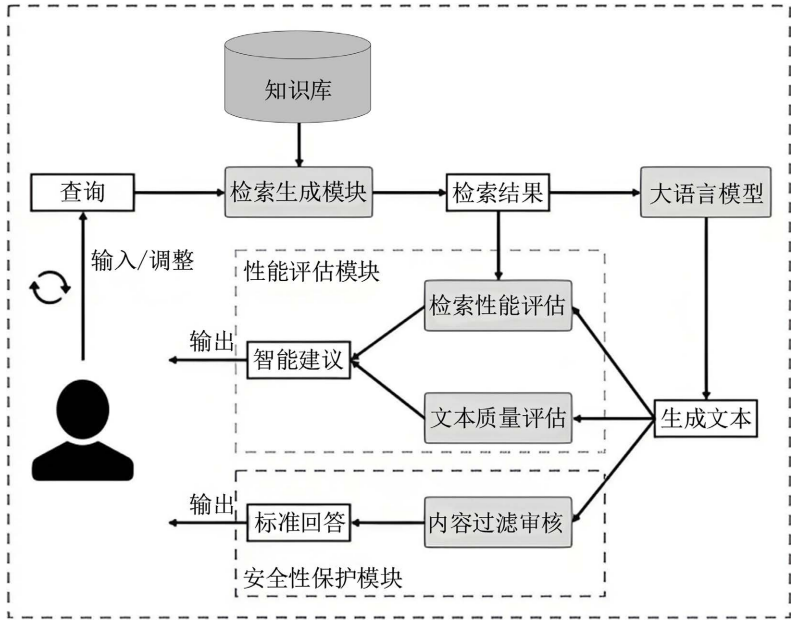


Figure 3. Architecture design of intelligent retrieval model
图 3. 智能检索模型架构设计图

检索层基于知识图谱语义网络执行数据匹配任务, 以嵌入式向量计算识别实体与关系间的深层关联, 并利用图神经网络进行结构推理, 挖掘经济指标与业务要素之间的潜在逻辑联系, 从而实现高精度检索。

评估层承担结果验证职能, 通过性能评估与文本质量评估双模块运作。性能评估部分基于人工智能算法计算准确率、召回率及 F1 值, 对结果进行综合评价; 文本质量评估部分利用语言模型分析逻辑一致性与表达规范性, 排除冗余与冲突信息, 保持输出内容的专业与连贯。

输出层负责生成反馈结果, 集成智能建议模块与安全过滤模块, 根据用户历史查询习惯与知识图谱关联关系动态推荐相关内容, 并运用命名实体识别技术屏蔽敏感数据。采用 Neo4j 数据库存储产业链知识图谱数据, 借助 Python 的 Py2neo 库将结构化数据转化为节点和关系, 同时利用 Echarts 实现知识图谱可视化, 支持平移、缩放、延展等操作, 直观展示数据关联[2]。最终, 模型以可视化图谱与语义化文本双形式输出, 实现企业经济数据检索的智能化与语义化联动, 形成查询、推理、评估、输出的闭环架构, 支撑复杂经济数据的智能管理与语义决策。

4. 人工智能在企业经济数据管理中的应用

4.1. 数据标准整合

企业经济数据存在来源系统多、字段结构异、单位表达混杂等现象, 影响数据的有效整合。数据标准整合需在格式转换、字段匹配与权重优化三方面建立统一模型结构。格式归一化阶段, 采用序列标注模型识别日期、数值、文本等字段类型, 将日期格式统一映射为标准时间戳, 数值字段统一以基础单位计量。单位转换由模型自动识别单位标记并生成转换系数矩阵, 用于动态调整输入值的数量级, 消除格式冲突。

字段对齐阶段, 构建经济数据字段知识库, 使用深度语义匹配模型计算非标准字段与标准字段之间的语义向量相似度。嵌入式语言模型生成的向量经由相似度函数判定归属关系, 实现多系统字段的自动匹配与融合。该方法适用于跨语义层表达场景中字段含义一致但术语各异的情况。

在多源数据整合后, 为解决不同数据源对决策影响权重不等的问题, 引入特征权重分配机制。采用基于梯度提升树的加权模型, 学习各数据源在历史决策中的表现, 构建特征权重优化函数:

$$w_i = \frac{G_i}{\sum_{j=1}^n G_j} \quad (1)$$

式中, w_i 表示第 i 个数据源的特征权重, G_i 表示其在样本中对目标变量提升效果的增益值, $\sum_{j=1}^n G_j$ 为全部数据源总增益和。该函数用于衡量每个数据源在目标预测或判断中的边际贡献, 权重越高代表整合优先级越高, 影响后续指标的融合排序。模型以此结果动态调整数据融合顺序与可信度参数, 提升整合精度与决策支撑能力。

4.2. 异构信息清洗

企业经济数据存在数据类型多样、来源系统分散、格式表达不一等问题, 导致脏数据频发, 影响后续分析处理。如图 4 所示, 清洗过程需覆盖结构化、半结构化与非结构化三类信息, 建立统一的识别、去重、纠错机制。结构化数据中异常值问题突出, 孤立森林算法构建多棵随机切分树, 度量每个样本孤立难度, 孤立路径越短则越异常。该算法可识别逻辑矛盾、极端偏离的财务数据样本, 有效剔除影响建模质量的值。

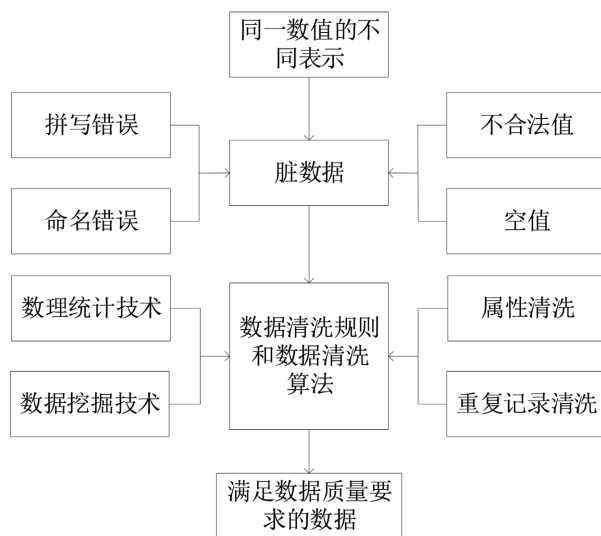


Figure 4. Schematic diagram of heterogeneous information cleaning
图 4. 异构信息清洗原理图

半结构化数据存在格式松散、重复嵌套问题, 基于文本向量模型构建表示空间, 提取合同编号、金额、时间等关键信息生成向量, 采用余弦相似度指标判定重复数据对。相似度高于设定阈值即视为重复, 保留一条并删除冗余记录, 适用于文档、表单、邮件类数据统一归并处理。

非结构化数据易出现命名混乱、语义歧义等问题, 结合 BERT 语言模型与编辑距离算法构建语义纠错模块, 对文本与语音转写数据进行标准化修正。采用 RoBERTa + BiLSTM + CRF 模型对非结构化数据进行命名实体识别和关系抽取, 能准确提取关键信息并转化为结构化数据, 有效解决非结构化数据语义

处理难题[3]。模型利用上下文理解能力识别表达错误,再依据最小编辑操作路径判断最可能的标准表达,并输出修正结果,提升原始文本数据的语义准确性与逻辑一致性。数据清洗流程整体以异常识别、重复消解、语义修复为主线,构建多源经济数据的自动化清洗机制。

4.3. 语义查询解析

企业经济数据查询中存在语言表达自然、系统识别依赖字段的问题,形成查询语义与数据结构不匹配的常见障碍。语义消歧环节引入上下文感知模型,结合用户角色与历史行为识别查询词义。模型构建动态语义偏移向量,通过对查询上下文与身份标签的双重建模,区分同词异义的字段指向,提升语义解析准确度。

意图识别模块采用多层神经网络结构,提取语句特征后进行意图分类[4]。输入语句经卷积网络提取局部关键词,再由双向循环网络捕捉上下文依赖,完成对查询目的的精细分类。分类标签控制后续知识图谱检索策略选择,使系统根据意图调用相应规则库,提高语义响应的精度与效率。

语义结构生成模块基于序列到序列神经网络实现自然语言向结构化查询语句的映射。模型输入编码为嵌套向量,输出阶段根据实体识别结果与关系抽取路径构建三元组映射,自动生成可执行的 SPARQL 结构语句。该过程建立自然语言表达与知识图谱语义之间的映射桥梁,支撑非结构化查询向结构化检索的转化。

4.4. 风险决策预警

本研究聚焦于基于图神经网络的风险路径推理方法,在企业经济数据知识图谱中构建多实体链式关联结构,挖掘复杂条件下的潜在风险传导逻辑,结合历史事件归纳与推理算法,建立多类风险触发条件与路径推导模式。提出融合关系特征的动态图注意力网络(RDGAT)模型,采用动态注意力机制计算邻居实体间注意力系数,融合实体间关系特征,有效预测实体间关系,提高知识图谱完整性[5]。

模型采用图神经网络 GCN (Graph Convolutional Network)结构,输入为实体节点特征矩阵 X 与邻接矩阵 A ,模型计算公式如下:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^{(l)} W^{(l)}\right) \quad (2)$$

式中, $\tilde{A} = A + I$ 表示加自环邻接矩阵, $\tilde{D}$ 为度矩阵, $W^{(l)}$ 为第 l 层权重, σ 为激活函数。该结构实现风险因子间的图卷积传播与潜在路径挖掘。

图神经网络用于发现延迟、违约、流动性压力间的链式传导结构,扩展风险识别维度。风险量化阶段采用逻辑回归模型构建评分函数,融合关键指标差值与时间因素,统一衡量各类风险暴露程度。以项目成本为核心的风险评分函数设为:

$$R_s = \frac{C_a - C_b}{C_b} \times 100 + D_p \times \gamma \quad (3)$$

式中, R_s 为风险评分, C_a 为实际成本, C_b 为预算成本, D_p 为延期天数, γ 为延期影响系数。评分值结合阈值划分风险等级,形成多级预警响应机制,用于精准分类与干预策略分派。

实时监测环节部署流处理引擎,在数据更新瞬间同步计算评分结果,并比对预设阈值。如触发规则条件,则自动生成预警信息包,包含风险类型、关键指标、触发时间、风险路径、干预建议等内容。系统按角色映射路径推送至具体责任岗位,并在风险看板中动态刷新,支撑企业运作的即时风险控制反馈闭环。

5. 结论

本文以人工智能与知识图谱融合为核心, 构建面向企业经济数据管理的智能检索模型, 从语义层面实现数据标准整合、异构信息清洗、语义查询解析与风险决策预警的系统化管理。知识图谱在多源数据关联与语义逻辑推理中发挥关键作用, 人工智能算法在自然语言解析、模式识别与风险量化中展现高效性能, 形成查询、推理、评估、输出的闭环体系。未来研究将围绕跨领域知识图谱融合、模型可解释性优化及多模态经济数据处理展开, 推动人工智能在企业经济治理中的纵深应用, 为智能化决策与数据价值再造提供更坚实的技术支撑。

基金项目

2025 年度河北省金融科技应用重点实验室课题《基于大语言模型的用户画像与金融营销场景应用技术研究》(课题编号: 2025002)。

参考文献

- [1] 王辉, 邢伟, 曹帅, 等. 工业企业数据挖掘和知识图谱系统建设研究[J]. 中国标准化, 2023(20): 75-80.
- [2] 万轶轩, 张福安, 章永龙. 基于知识图谱的产业链大数据平台设计与实现[J]. 现代计算机, 2024, 30(15): 97-103+110.
- [3] 张涛, 周卫. 广西非遗文化知识图谱构建与数据处理研究[J]. 智能计算机与应用, 2025, 15(3): 72-78.
- [4] 李少华. 深度学习的自然语言处理技术进展分析[J]. 产业科技创新, 2025, 7(4): 69-72.
- [5] 江树涛. 企业数据资产知识图谱融合与补全算法研究[D]: [硕士学位论文]. 大庆: 东北石油大学, 2023.