

基于金融知识图谱的会计欺诈风险识别方法

陈强¹, 代仕娅²

1. 兴业银行信息科技部, 上海 201201; 2. 蚂蚁科技国际事业群数据算法技术部, 上海 200120

摘要

针对商业银行会计案件日益复杂且频发的问题, 将会计案防领域的行业知识与金融知识图谱技术结合, 以更精准地识别与防范商业银行会计风险。采用图分析、图挖掘等技术, 提取深层关联风险特征, 并与行业经验知识相结合, 构建了249条单点规则及425条组合规则, 形成了丰富、可灵活配置的反欺诈策略体系。将该智能化反欺诈方法应用于银行活期账户的风险排查, 与传统规则策略相比, 识别精准度大幅提升, 且对于筛选出的高度可疑账户, 识别精准度达到85%左右, 极大提升了会计案件核查的效率。

关键词

会计案防; 金融知识图谱; 反欺诈; 关联交易

中图分类号: TP311, F832

文献标识码: A

doi: 10.11959/.issn.2096-0271.2021029

Recognition method of accounting fraud risk based on financial knowledge graph

CHEN Qiang¹, DAI Shiya²

1. Information and Technology Department, Industrial Bank Co., Ltd., Shanghai 201201, China

2. Data and Algorithm Department, Ant Technology International Business Group, Shanghai 200120, China

Abstract

Since the accounting risk events exhibit complexity increasingly and occur frequently, a method merged by industrial knowledge and financial knowledge graph was proposed to recognize and prevent commercial bank's accounting risk more precisely. Based on the financial knowledge graph of account transaction, deep graph connected risk features were extracted via various graph analysis and mining technologies. Combining the graph features with industrial knowledge, 249 single rules and 425 assembled rules were constructed to form a more affluent and flexibly configurable anti-fraud strategy system, which was then applied to verify commercial bank's current accounts and select the high suspicious ones. The experimental results show that the risk recognition accuracy rate of the intelligent strategy is much higher than the traditional one and reaches up to 85% above, which significantly promotes the efficiency of the accounting risk verification.

Key words

accounting risk event management, financial knowledge graph, anti-fraud, connected transaction

1 引言

商业银行会计风险案件形式多样,涵盖反洗钱、资金诈骗、非法集资、虚假交易、恶意刷单、套现挪用等多种作案手段。随着商业银行业务范围不断拓展,经济形势复杂性加剧,会计风险案件日趋高发,给商业银行及社会经济带来严重的损失,也给金融系统的稳定性造成不良影响。除了明确对反洗钱等金融犯罪行为的打击防范,监管对银行资金的使用规范也日益严格,如《商业银行互联网贷款管理暂行办法》明确表示,贷款资金不得用于购买房产、股票、金融衍生品等。加强对会计风险案件的防控是新形势下银行风险管理的重要内容,具有较大的经济效益和社会效益,一是能更好地保障客户资金的安全,切实做好维护客户利益的本职工作;二是有利于监测并规范信贷资金的用途,使其切实用于企业的经营发展,支持与服务实体经济;三是能增强反洗钱能力,防范金融犯罪,助力金融系统的稳定与社会经济健康发展。而现实中各类风险事件的作案手法越发复杂,内外部勾结、团伙合作等形式较难被及时察觉,传统事后核查方式的防范效率较低。这些都对银行传统的风险管理模式提出更严峻的挑战,需要借助新技术、新方法来更有效地应对。

近年来,金融科技蓬勃发展,驱动金融业务经营与管理模式的创新变革。中国人民银行印发的《金融科技(FinTech)发展规划(2019—2021年)》明确指出,金融科技已成为防范化解金融风险的新利器,是新形势下金融风险管理的内在需求和重要选择。金融科技在会计非现场监控工作中的应用能实现对会计异常数据的自动监控、自动预警、自动通知,有效提升会计监

控的效率^[1];基于大数据、人工智能技术的反欺诈智能平台能对海量金融交易进行监测,在降低异常交易预警的同时,极大地提高了识别欺诈交易的命中率^[2]。为此,商业银行积极探索大数据、人工智能、知识图谱等核心技术在会计案防领域的高效应用就显得格外必要。将传统的风险业务知识与大数据、人工智能技术结合,充分发挥数据模型在风险特征挖掘上的能力,构建更完善、更精准的智能会计案件风险防控体系,能有效地提升银行的风险管理水平,推动银行的高质量发展。

本文引入知识图谱相关技术,提出了结合关联风险特征的会计案防智能化反欺诈策略体系构建方案。首先,构建了银行账户金融知识图谱,并采用多种图分析技术挖掘提炼账户之间的关联风险特征;其次,将关联风险特征与传统会计案防规则结合,形成完整的会计案防智能化反欺诈策略体系;最后,基于银行真实交易账户对反欺诈策略进行检验,检验结果表明,该智能化反欺诈策略体系能有效识别出高风险账户。

2 相关工作

作为人工智能的重要领域之一,知识图谱相关技术在金融反欺诈领域的应用日益广泛,对金融风险防范起到显著的促进作用。姜增明等人^[3]认为商业银行反欺诈的关键在于建立以大数据为支撑的风控体系,采用知识图谱、社交网络分析等技术,更有效地防范复杂模式下的欺诈风险。柴洪峰等人^[4]指出依托知识图谱强大的知识推理和逻辑判断能力,在提升风险监管决策准确性的同时,也能增强其可解释性。陶睿等人^[5]通过构建上市公司知识图谱,对企业风险进行智能化监控,提升了监管穿透性,缓解了监管时滞性;陶士贵等人^[6]基

于股权关系图谱,采用复杂股权关系路径算法,有效识别企业的最终受益人,实现反洗钱风险的预警提示。上述文章主要侧重于从企业控制人角度出发,采用知识图谱挖掘股权关系中的风险,但这些文章均仅描述了研究思路,并未就具体图计算过程和实验结果进行详细阐述。金磐石等人^[7]通过构建企业关联图谱,将企业画像特征与从知识图谱中提取的企业关系特征结合,并将其作为分类模型的变量输入,预测小微企业欺诈的概率;王成等人^[8]采用知识图谱的网络嵌入方法预测互联网借贷的欺诈行为,将网络中的节点嵌入低维的向量空间,以自动学习网络中隐含的关联关系,增强了欺诈预测的性能。这两篇文章主要将知识图谱与AI算法结合,通过算法模型预测信贷业务中的欺诈风险,但对于会计案防业务来说,这种方法不仅欺诈样本比例较低,影响模型训练效果,而且对案件发生逻辑的分析要求高,完全采用机器学习模型预测欺诈缺乏相应的可解释性。凡友荣等人^[9]基于电信通话知识图谱,通过最短路径算法、PageRank算法、聚类算法等识别电信通话中的关联路径、核心人物及社会关系,为电信资金诈骗的识别提供重要线索;魏瑾等人^[10]搭建了基于知识图谱的欺诈电话智能决策平台,提升了电信诈骗识别的准确率。这两篇文章采用知识图谱技术提取个人之间的交互网络特征,能获得更深层次的风险信息,但主要为个人关系层面,未涉及账户资金交易,场景应用上存在较大差异。

为此,本文基于会计案防领域的业务特点,构建了一种将行业知识与知识图谱技术结合的智能化反欺诈策略方案,在交易账户金融知识图谱的基础上,采用多种图分析技术,从网络结构、账户资金流向、资金交易闭环、账户关联强度等多维度进行深度挖掘,提炼出账户关联风险特征;

将账户关联风险特征与账户基础特征结合起来,基于行业知识构建反欺诈规则策略集,以实现会计案件欺诈风险更精准的识别。

3 会计案防智能化反欺诈策略构建技术

在商业银行会计案防工作中,基于长期业务经验积累形成的风险规则逻辑是识别欺诈案件的重要行业知识,但是对于资金交易闭环、账户穿透转移等关联交易特征,由于涉及主体较多、交易路径复杂,人工难以及时识别,这也是会计案防工作的痛点与难点。知识图谱通过将蕴含业务规律的、跨系统跨主体的数据信息以图的形式进行关联、展示与分析,能够挖掘出隐含的欺诈风险信息。对于银行账户交易来说,不同账户之间通过资金转移形成了一张交易关联网络,采用知识图谱技术能更及时、更精准地提取出账户深层的关联交易风险特征,并通过图可视化技术进行展示,这有助于提升会计案防风险识别的质效。

由于会计案防欺诈样本相对较少,且对案件发生逻辑及可解释性要求很高,会计案防核查最终需要基于行业知识的风险规则策略对案防事件进行识别与确认。传统案防风险策略一般以账户的基础特征为主进行构建,缺乏账户之间的关联网络特征。而在本文的智能化反欺诈方案中,将行业知识与知识图谱结合起来构建出更前瞻、更精准的风险策略集。一是通过账户交易知识图谱提炼、挖掘更加丰富的交易网络特征,全面获得账户的基础特征(也可以将其看成账户节点的属性)与账户间隐藏的关联特征,形成更完整的特征规则集;二是基于会计案防行业知识,对融合了网络关联特征的规则集进行组合配置,形

成更加精准有效的会计案件反欺诈策略体系。在整个过程中,如何通过知识图谱技术挖掘出账户之间隐藏的关联特征,是本文研究的重点。智能化反欺诈策略方案的整体架构如图1所示,主要包含指标库、金融知识图谱分析、规则策略等部分,其中指标库是指对会计案防规则进行解耦,并加工出账户的基本特征,是实现风险策略可灵活配置应用的重要工程化基础;金融知识图谱分析部分主要是对账户交易知识图谱的构建以及关联网络特征的提取;规则策略部分主要是对融合了账户基础特征和关联网络特征的规则进行组合配置,形成不同场景下的会计案防反欺诈策略。

一旦规则逻辑发生任何细微的改动,都需要重新编写SQL规则语句,这种应用方式一方面导致规则配置及更新迭代的效率较低,另一方面也难以与知识图谱等技术结合进行挖掘分析。为了实现会计案件的智能分析预警,需要对规则文本进行进一步的拆解,提炼出影响案件性质的关键要素,形成指标库,以便通过对指标的灵活组合配置,形成适应不同时间、不同场景的业务规则,对会计风险案件做出更快、更精准的响应。指标拆解示例见表1。

同时,对于一些难以从结构化数据中直接提取的规则指标,将采用知识图谱、

3.1 业务规则指标化

会计案防业务中原始的规则逻辑大多以文本的形式描述,在应用过程中将对每一条规则逻辑进行独立的SQL语句编写,

表1 指标拆解示例

规则逻辑描述	提取指标	指标化规则配置
本行对公活期存款账户发生借方出账金额达单笔≥800万元,且交易走款未验印或验印手工通过	主账户类型 借贷标志 交易金额 验印结果	主账户类型=“对公定期” 且:借贷标志=“借方” 且:交易金额≥“800万元” 且:验印结果=“未验印”or“手工通过”

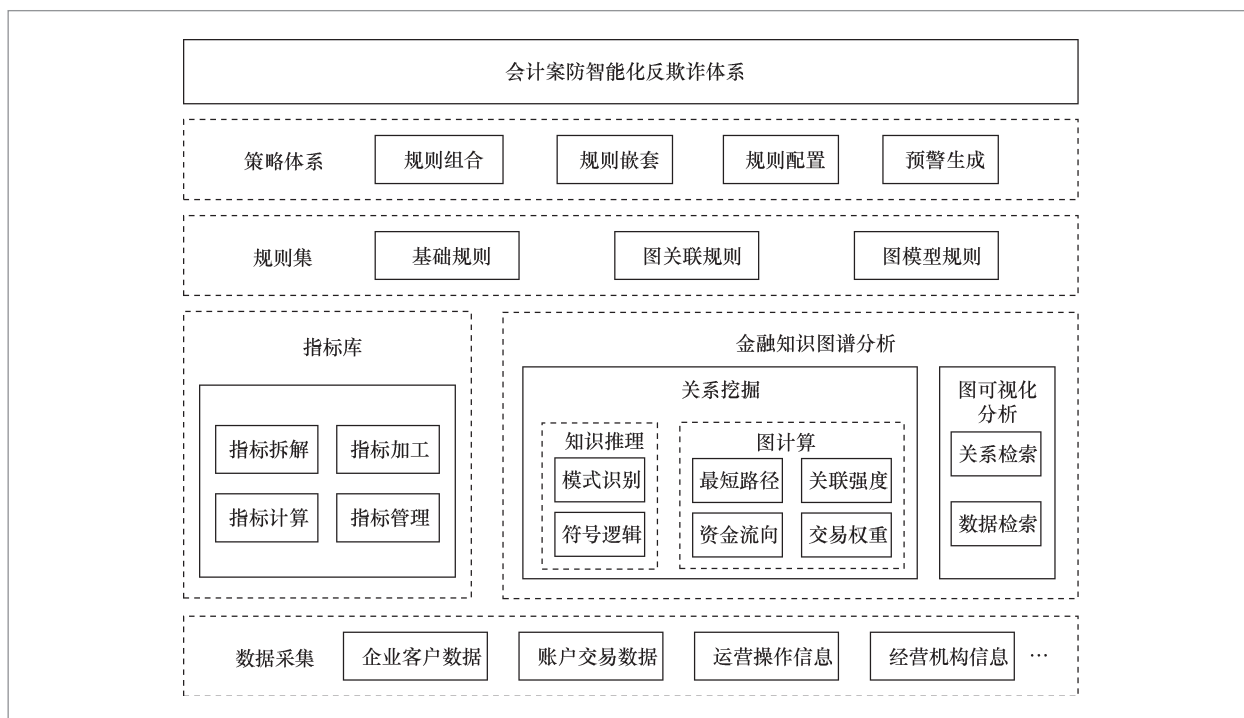


图1 会计案防智能化反欺诈体系架构

机器学习等技术进行相应的处理,以更充分地实现指标自动化、智能化生成。如在“票据审查真实”这条规则中,传统票据审查方式只能靠人工在线下进行,在此则依托光学字符识别(optical character recognition, OCR)技术对承兑人、票面金额等票据要素进行自动提取与真实性判断;又如,对于“是否是股东”这条规则,可以采用自然语言处理(natural language processing, NLP)技术从股权变更公告中自动提取股权相关信息,图2为指标智能化生成方法示例。

3.2 金融知识图谱构建与分析

知识图谱具有节点和边的概念,节点表示信息实体或者实体的属性值,边表示两个被连接实体的关系或者一个实体的某个属性^[11]。金融知识图谱主要针对金融行业的业务场景,通过大规模语义网络,将金融业务领域中的结构化、半结构化、非结构化等不同类型数据进行整合,将事物或者概念作为实体,基于实体之间的语义关系相互连接,构建出网状的知识链接,形成结构化的语义知识库,以更清晰的形式描

述金融业务中复杂的关联逻辑^[12]。金融知识图谱的内容和结构根据具体金融业务的不同而不同,对于会计案防领域,防范重点在于资金的违规流动,故该领域金融知识图谱的重点在于对账户之间资金交易关系的刻画,与传统知识图谱相比,账户实体之间的边增加了表示资金流向的箭头,边的粗细也会根据交易量进行区分。

知识图谱的构建首先是语义框架的设计,即确定会计案防业务领域中的实体,以及实体之间的语义关系结构;其次是知识内容的构建,即根据语义框架中定义好的实体及关系类型,从多源异构数据中提取出实体,将实体、关系、属性以三元组的形式存储在图数据库中^[13],形成完整的关联图谱。

采用各类知识图谱分析技术能够发掘出图结构中隐含的业务深层联系,实现智慧搜索与智能交互,进而增强金融机构的业务智能化水平。本方案主要采用图可视化分析、知识推理、图计算等图技术,提炼金融知识图谱的浅层关联特征、深层关联特征等多层次结构信息,赋能复杂业务经验的智能化应用,提高欺诈风险识别的覆盖率、准确率。主要的图分析及特征提取内容如图3所示。

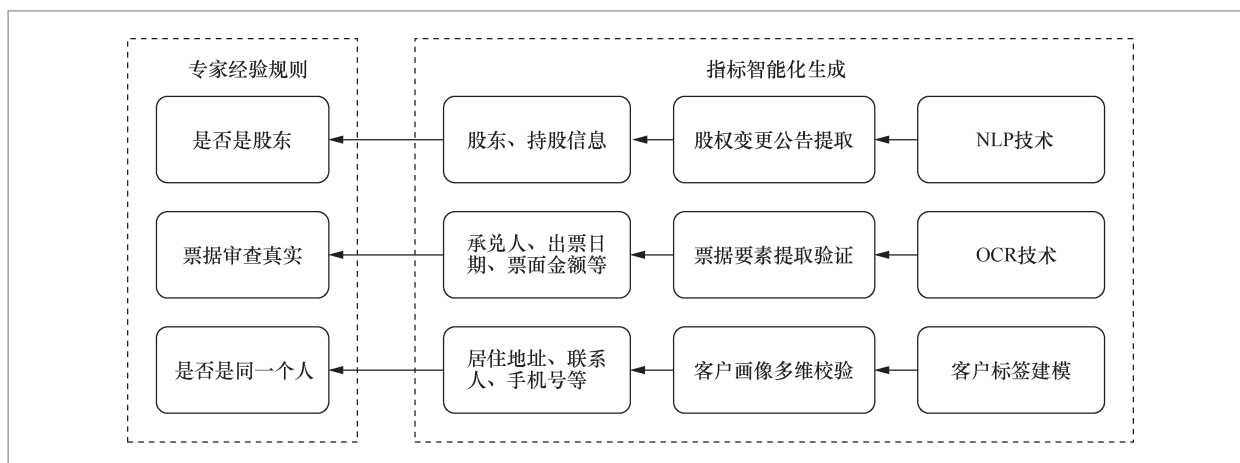


图2 指标智能化生成方法示例

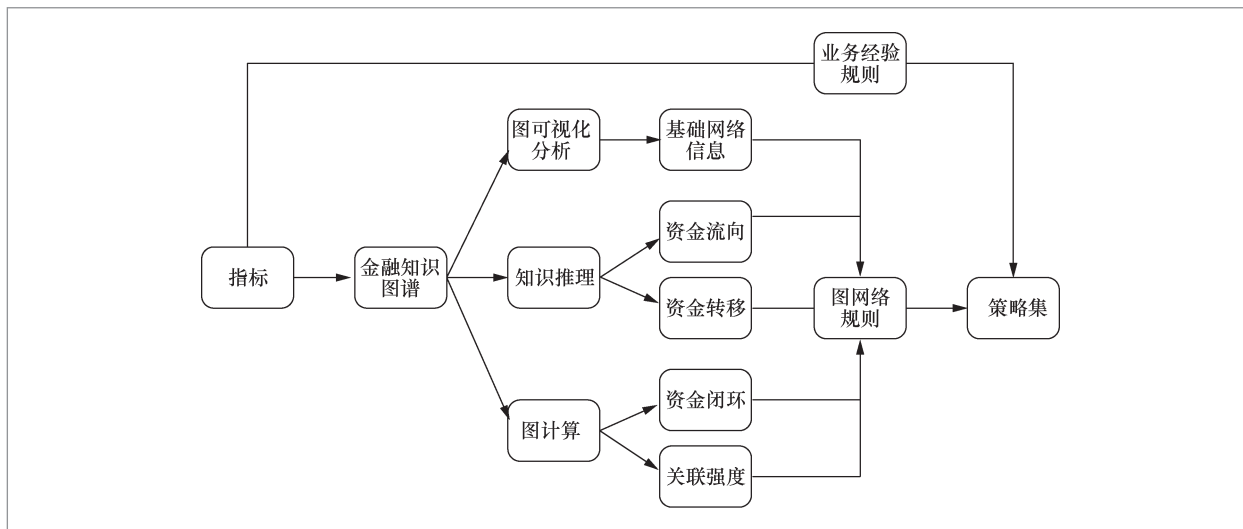


图3 图分析和特征提取

3.2.1 图可视化分析

图可视化分析主要是将图结构的数据存储在Neo4j、JanusGraph等图数据库中，并结合Elasticsearch搜索分析引擎，实现高灵活性、高准确性、低时延及大规模并行化的检索查询^[14]，形成对图结构数据的可视化关联展示，以提取实体之间的关系类型、实体属性等基础的网络特征。

3.2.2 图关系挖掘

通过图计算、知识推理等分析挖掘技术，提取出图数据结构中的深层关联特征，揭示隐含的欺诈风险。本方案主要采用的图关系挖掘方法见表2。

(1) 资金交易闭环分析

找出资金从某一账户转出，与其他账户交易后又转入原始账户的交易闭环，有利于排查虚假交易、洗钱等风险行为。账户交易图谱是有向图，可采用Tarjan算法识别交易网络中的强连通图，即图中任意两个顶点可以互相通达，从而挖掘出闭环交易。Tarjan算法基于深度优先搜索进行

表2 图关系挖掘方法

方法	识别目标
图计算	资金交易闭环分析 账户关联强度分析
知识推理	账户分层穿透匹配 资金转移模式匹配

节点访问，定义 $dfn[u]$ 为深度优先搜索时节点 u 的时间戳（即次序编号）， $low[u]$ 为 u 或 u 的子树能够追溯到的最早的栈中节点的序号^[15]。对于当前节点 u 和关联节点 v ：若 (u,v) 为搜索边， $low[u]=\min\{low[u],low[v]\}$ ；若 (u,v) 为回边， $low[u]=\min\{low[u],dfn[v]\}$ 。最终选取 $low[u]=dfn[u]$ 时，该节点往上的所有节点构成的强联通分量。

(2) 账户关联强度分析

将两个账户之间的交易路径长度、交易时间、交易金额、交易次数等多个维度综合起来进行评估，挖掘出有直接或间接交易关系的账户之间的紧密程度。对于一条交易路径，如图4所示，A1、A2两个账户之间并无直接交易，但通过账户B2、B3建立起间接的交易关系。

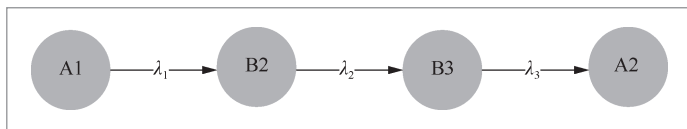


图4 交易路径示例

用 λ 表示两个直接交易账户之间的交易权重, λ_i 为资金从账户A1到A2的整个交易链路中第 i 段交易的权重:

$$\lambda_i = \frac{1}{T_i} \sum_{t=1}^{T_i} \left(\frac{N_{it}}{d_{it}} + \frac{\lg M_{it}}{N_{it}} \right) \quad (1)$$

其中, T_i 表示第 i 段交易中两个账户在特定时间内的时间切片个数, d_{it} 表示第 t 个时间切片, N 表示交易次数, M 表示交易金额。这里交易权重主要由单位时间段的交易次数以及单次交易的金额这两个因素决定。整个交易链条中A1和A2的关联强度 R 为各

$$\text{段交易权重的平均值} \frac{\sum_{i=1}^n \lambda_i}{n}。$$

(3) 账户分层穿透匹配

分层穿透是指通过层层追溯, 找到最终的间接利益关联方^[16]。此处主要是对资金在转出初始账户后的交易路径进行层层追踪, 找到资金的最终流向, 能有效挖掘出多层复杂场景下资金通过中间账户转移到非法账户的风险。推理示例如下:

$$\begin{aligned} & \text{edge}(a \rightarrow b), \text{edge}(b \rightarrow c) \\ & \Rightarrow \text{path}(a \rightarrow b), \text{path}(b \rightarrow c) \quad (2) \\ & \Rightarrow \text{path}(a \rightarrow c) \end{aligned}$$

其中, a 、 b 表示两个节点, edge表示存在边关系, path表示存在关联路径, 箭头表示关系及路径的方向。

(4) 资金转移模式匹配

资金转移模式匹配主要用来判断“分散转入、集中转出”以及“分散转出、集中转入”这两种可疑的账户交易模式, 前者通常隐含非法集资的风险, 后者则具有典型的洗钱特征。为了更加精准地评估某一资金转移与风险交易模式的匹配程度, 将账户

交易的网络图谱转换为账户余额随时间变化的交易波形图, 用波形的相似程度表示账户交易模式的相似程度, 从而定量评估资金转移模式是否具有较高的风险特征。

在波形相似度的衡量上, 相关系数一般用来衡量波形总体相似程度, 容易忽视局部特征的差异^[17]。本文在计算波形相关系数的基础上, 选取波形的主波峰值、主波起始点至顶峰的时间这两个主要的波形局部特征, 将两者结合起来, 得到最终相似度, 其计算式为:

$$S = \beta_1 \left(\frac{|H_p - H_w|}{H_w} + \frac{|T_p - T_w|}{T_w} \right) + \beta_2 \frac{\sqrt{\sum (W_t - \bar{W})(P_t - \bar{P})}}{\sqrt{\sum (W_t - \bar{W})^2} \sqrt{\sum (P_t - \bar{P})^2}} \quad (3)$$

其中, W 为基础波形, P 为对比波形, W_t 、 P_t 分别表示在 t 时刻两个账户的金额; \bar{W} 、 \bar{P} 分别表示两个账户在该时间段内的平均金额; H 表示主波峰值; T 表示主波起始点至顶峰的时间; β 表示权重。

4 实验设计与结果分析

在实验中, 将实际金融活动中真实发生的15起会计风险案件作为规则设计的参考蓝本, 基于行业经验知识从中提炼出可疑风险规则, 并对规则进行拆解, 形成基础指标库, 作为账户的基础特征; 选取某银行2019年6月至2020年1月的所有活跃账户数2 900万个, 将其中有历史管控和确认欺诈的账户作为欺诈样本, 共6 571个, 将账户交易情况、企业及个人基本信息等作为数据基础, 构建账户交易知识图谱, 提取出账户的关联风险特征。

4.1 业务规则指标库构建

与会计案防相关的完整的指标体系涉

及的内容广泛。从指标对象来说,涵盖了客户、账户、交易、员工、营业机构等不同层面的主体,如客户注册地、账户开户地址、是否人工验印、员工操作笔数、机构开户数量等。从指标加工的技术来说,涵盖了基础指标、统计指标等,基础指标是指客户或账户的基本属性,统计指标是指经数据运算后的指标。**表3**展示了一些业务规则指标,将这些指标作为账户的基础特征。

4.2 金融知识图谱关联特征提取

本文图网络结构标识的实体主要包含两类,一类是账户实体,包含企业、个人开立的可发生资金交易行为的银行账户;另一类是非账户实体,如经营机构、操作设备等。对应的实体之间的关系也主要分为两类,一是账户与账户之间,主要是交易关系,由是否有资金往来确定,若发生过转账、支付等交易行为,则有交易关系;二是账户与非账户之间,具体关系需要根据非账户实体确定,如账户与经营机构之间是开户关系,账户与操作设备之间是使用关系。在本文账户关联图谱中,实体用节点表示,关系用边表示,其中交易关系具有箭头和粗细区分,边的箭头由转出账户指向转入账户,边的粗细则主要由交易金额决定;而对于账户与非账户之间,未涉及资金流向,其关系大多是固定的,也是可逆的,如某一账户和开户证件号,可以从账户关联到证件号,也可以从证件号关联到账户,可逆关系无须箭头和粗细的区分^[18]。对于与其他实体无任何关联的孤立实体节点,其表示该账户无资金交易行为,也无任何基本信息关联,将该孤立节点删除。最终构建形成的金融知识图谱由诸多非连通的子图组成,各个子图之间未构成关联,子图内部的账户之间以多种方式进行连接。知识图谱中涉及的主要实体及关系

表3 业务规则指标示例

指标类别	指标名称	指标含义
客户类	对账地址	企业在银行对公账户对账单上填写的地址
	联系电话	企业在银行开立账户的联系电话
	是否开启通知	企业是否开启银行短信、精灵信使等通知
交易类	主账户类型	发生交易的企业账户主账户类型
	借贷标志	账户交易的资金转移方向
	交易账户类型	主账户发生交易的对手账户类型
运营类	账户验印结果	企业提交材料上的印章与开户时的预留印章是否一致
	对账是否成功	交易账目核对是否一致
机构类	近3月账户开立数	营业机构近3个月内开立的对公账户数量
	近1月借方累计金额	营业机构近1个月内所有对公主账户发生借方交易的总金额

类型见表4。

图5为账户交易知识图谱的子图示例。以子图1为例,A1~A5代表资金账户实体,A1账户的资金主要流向A2、A3、A4 3个账户,其中A1→A2的资金交易量最大,连线边也相应较粗,A1→A3的资金交易量最小,连线边则较细;IP地址、证件等为非账户实体。从**图5**可知,账户A2、A3、A5关联同一个证件号码,非账户实体与账户实体之间无交易行为,故连接边为无箭头的实线。

4.2.1 浅层关联特征

浅层关联特征主要指可从图结构中直接查询得到,无须进一步加工计算的特征,也可以通过图谱可视化直接读取,如**图5**中的子图1,A4、A5账户对应同一个对账地址,当多个非同名账户都对应同一个对账地址时,则存在虚假开户的风险。**表5**给出了浅层关联特征示例。

4.2.2 深层关联特征

与浅层关联特征的获取有所不同,深

表4 知识图谱实体与关系

实体	关系类型	关系表示	关系表示
账户实体	交易关系	“账户”_(交易)_“账户”	⊙A1 ⊙A2
非账户实体(证件、手机号、设备IP、经营机构等)	非交易关系	“账户”_(拥有)_“证件”	⊙A1 ⊙I
		“账户”_(联系方式)_“手机号”	⊙A1 ⊙P
		“账户”_(使用)_“设备IP”	⊙A1 ⊙D
		“账户”_(开户)_“经营机构”	⊙A1 ⊙O

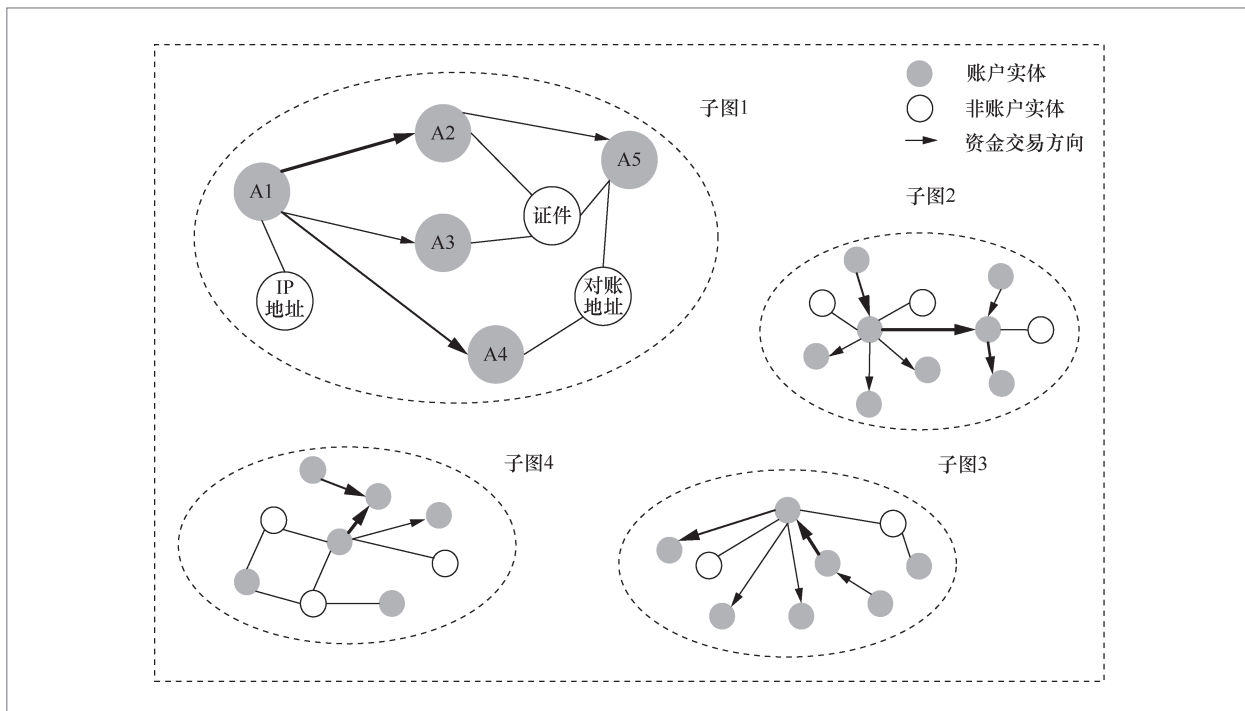


图5 账户交易知识图谱的子图示例

表5 浅层关联特征示例

类别	浅层关联特征
图谱类	多个账户的联系人电话相同
图谱类	账户在多个设备登录
图谱类	不同客户对账信息中共享同一个对账IP
图谱类	经营机构存在多个异常账户

层关联特征无法从图结构中直接获取,需要基于图计算等方式深入挖掘提取得到。

(1) 资金交易闭环分析

如图6所示,基于前述Tarjan算法挖

掘强联通分量,可以发现资金从账户A1转出,经过账户A2、A3、A5路径,以及账户A2、A3、A4、A5路径,回流至账户A1,形成两条资金闭环,提示存在洗钱或虚假交易等疑似欺诈的行为。

(2) 账户关联强度分析

当两个账户之间存在多条交易路径时,对不同交易路径下的关联强度进行计算,选取最大值作为最终的关联强度值。如图7所示,A和B代表账户实体(为了进一步区分,以A表示起始和终点账户,B

表示中间账户)；V表示非账户实体，与账户间无资金交易关系，也不影响账户关联强度的计算。A1与A2之间存在两条关联路径，分别是A1→B2→B3→A2，以及A1→B1→A2。计算两条交易路径下A1、A2的账户关联强度，得到在较长路径下，两者的关联强度反而更大，有助于对可疑账户的高关联账户及核心交易路径进行排查。

(3) 账户分层穿透匹配

如图8所示，对账户资金流向进行层层穿透，可以发现账户A与账户D并无直接的资金交易，但经过知识图谱对交易层级的拓展，显示账户A中的资金经过B1到Bn，以及C1到Cn这两条路径，最终流向了账户D这类投资股票、房产的高风险账户，属于违规使用信贷资金的行为。

(4) 资金转移模式匹配

如图9所示，通过限定资金转入转出特点，得到两种交易模式图，一种模式是分散转入、集中转出，账户X1、X2、X3、X4中的资金先后流向账户X中，呈现陆续分散转入的现象，然后账户X再将资金集中转出到账户Y中，这有非法集资的风险；另一种模式是集中转入、分散转出，与前一种模式正好相反，资金先由账户X集中转入账户Y中，然后再由账户Y陆续转出到账户Y1、Y2、Y3、Y4等账户中，这有较严重的洗钱嫌疑。

进一步分析不同账户的交易波形与欺诈样本交易波形的相似程度。图10表示“分散转入、集中转出”模式下某一欺诈账户的关联交易模式，图11为对应的风险交易波形图；波形一（如图12所示）和波形二（如图13所示）分别代表两类账户的交易波形。根据前述计算方式，得到波形一、波形二与风险波形的相似度分别为0.423和0.825，表明波形二对应的账户交易模式存在与风险波形类似的欺诈特点，需要重点排查；波形一与风险波形的相似度较小，

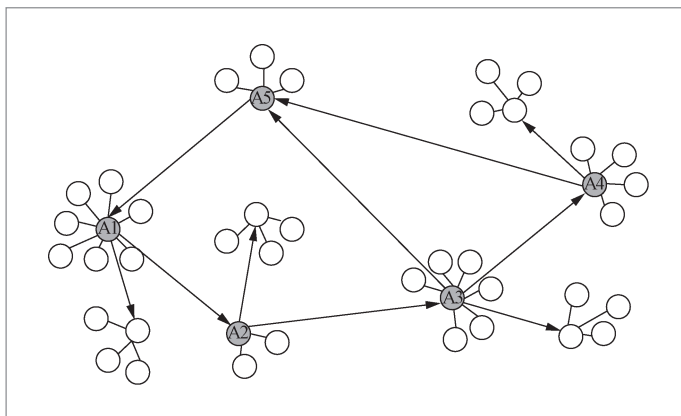


图6 资金交易闭环示例

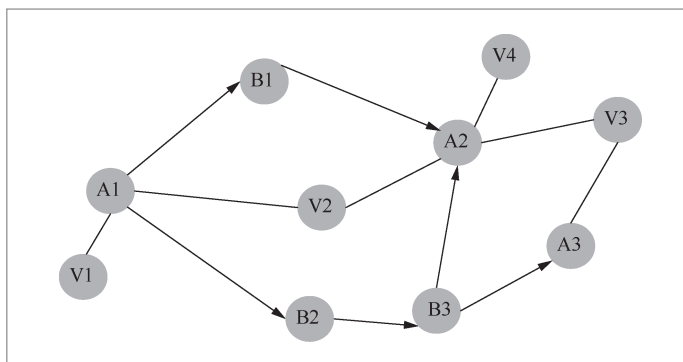


图7 账户关联强度分析

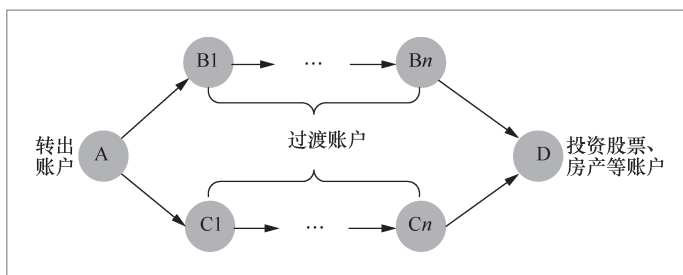


图8 账户分层穿透

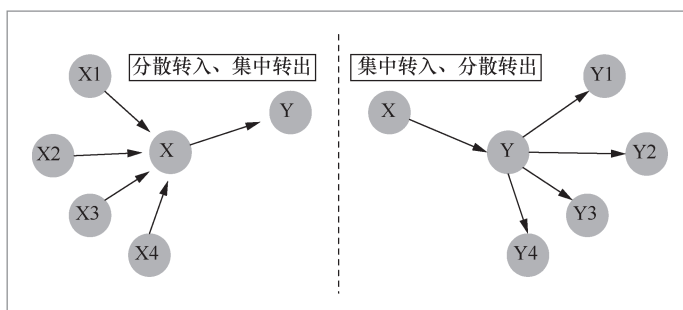


图9 资金转移模式

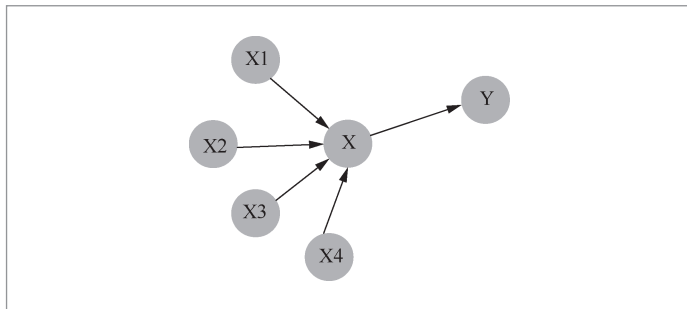


图 10 风险交易模式

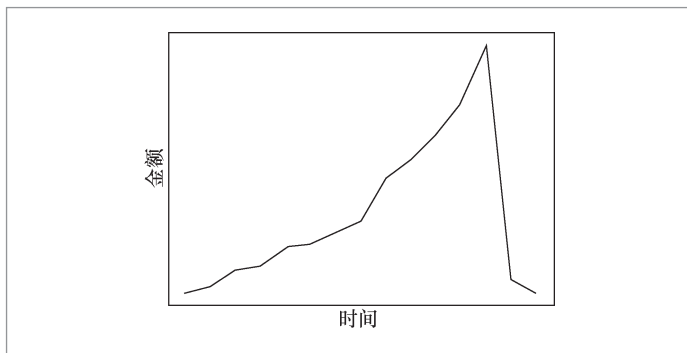


图 11 风险交易波形

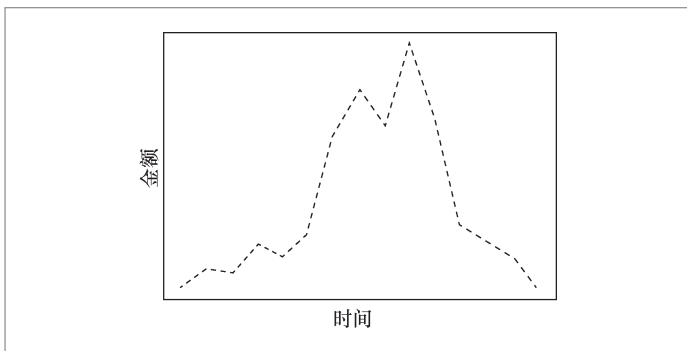


图 12 波形一

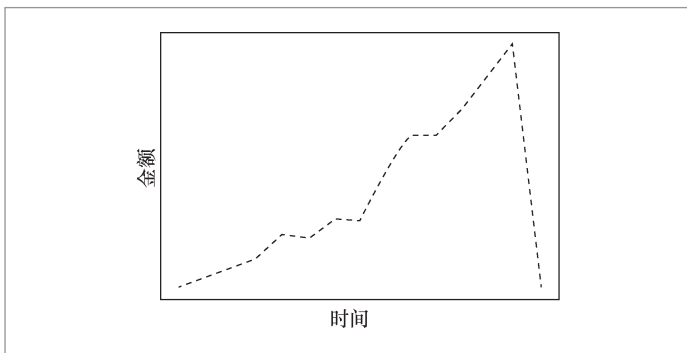


图 13 波形二

表明波形一对应的账户交易模式属于前述欺诈行为的可能性较小。

基于以上分析，本文进一步提取出一些具有较高欺诈风险的深层关联特征，见表6，这些深层关联特征和浅层关联特征一起形成了对账户基础特征的有力补充。

4.3 规则策略体系

将账户基础特征与从知识图谱中提取出的关联风险特征相结合，搭建了较完整、丰富的风险指标集。对每一指标设置阈值，即相应的单点规则；由于单点业务规则不足以解释风险现象，需对其进行灵活的搭配形成规则组合，并辅以指标阈值的调整。最终构建出会计案防领域反欺诈规则策略体系，该体系共涵盖了249条单点规则，以及425条组合规则，图14为规则配置的示例。

4.4 实验结果

将以上规则策略集应用于选定历史时间段的2 900万个银行活跃账户，将其中有历史管控或确认过欺诈的账户标记为欺诈样本，共6 751个。将该数据集分为两组进行对比试验，一组采用账户基础特征构建的规则策略进行欺诈识别，另一组采用融合网络关联特征后的规则策略，通过对比两组策略的风险识别结果，验证知识图谱应用的效果，实验结果见表7。

策略1表示仅使用账户基础特征规则进行欺诈识别，策略2表示将账户基础特征与关联特征结合起来形成反欺诈策略。将样本近乎平均地分配到两组实验中，最终策略1筛选出2.1万个疑似风险账户，其中确认为欺诈的共4 230个，确认率为20.1%；策略2提示有风险账户数为4 067个，其中属于欺诈的有2 621个，确认率为64.4%；进一步从策略2预警提示的风险账户中挑选

出规则命中最多的前100个账户,其中85个被确认为欺诈账户,头部欺诈确认率达到85%。实验结果表明,融合了知识图谱关联特征后的规则策略能够显著提升会计案防欺诈识别的准确性,缩小异常账户核查的范围,使风险管理措施集中到最可疑的账户中,从而提升会计案防工作的质效。

5 结束语

本文将会计案防领域的行业知识与金融知识图谱技术相结合,构建了基于图谱关联特征的智能化反欺诈策略方案。在实验中,以行业真实会计风险案件和银行内部账户数据为分析基础,对欺诈经验规则进行指标化拆解、加工,形成反欺诈基础规则指标库;采用图可视化分析、知识推理、图计算等知识图谱技术,对基础网络结构、账户资金流向、资金交易闭环、账户关联强度等进行深度挖掘,提取出关联风险特征,最终形成了丰富、灵活的会计案防智能化反欺诈规则策略体系,共包含了249条单点规则,以及425条组合规则。将智能化反欺诈规则策略用于银行内部账户的风险核查,挑选其中疑似欺诈程度较高的账户进一步人工核验,最终欺诈确认率高达85%,这充分表明了智能化反欺诈策略在风险识别上的精确性,能显著提升会计案防工作效率,增强银行的风险管理能力,也有助于维护金融系统的稳定性。

未来智能化反欺诈策略中将融入更多的机器学习、深度学习等AI算法,将算法模型的分析预测结果进一步纳入规则集中,使反欺

表 6 深层关联特征示例

类别	深层关联特征
图谱类	资金通过内部账户转向非同名账户
图谱类	账户转出资金高度集中于某一个账户
图谱类	指定购买理财资金转入其他类型账户
图谱类	贷款资金经多层周转后用于禁止性领域
图谱类	多对一账户归集
图谱类	与异常账户关联强度较高
图谱类	资金以固定时间间隔转出相同金额
图谱类	大额资金经多条路径转入关联方账户

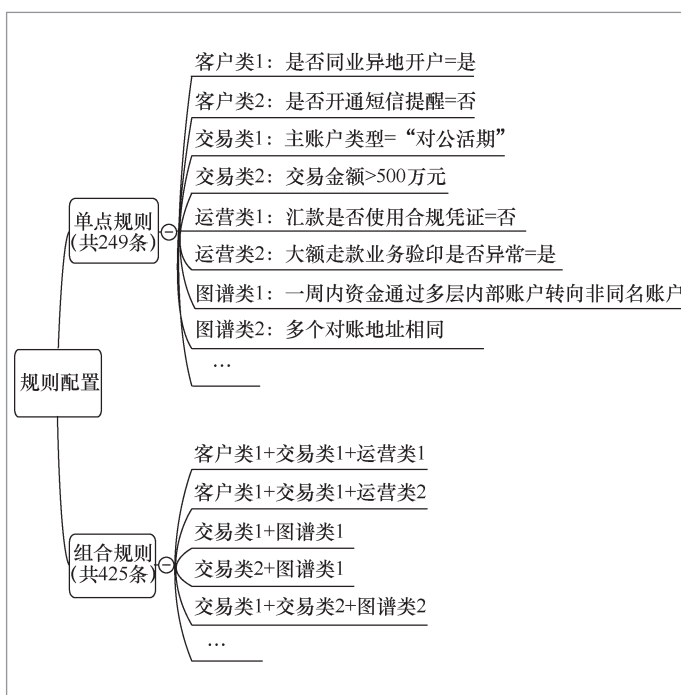


图 14 规则配置示例

诈策略体系更加完善与精细;也将依托大数据相关技术搭建智能化会计案防平台,以平台化运作的方式实现规则策略的自动化加工计算、组合配置及提示预警,最终在真实会计案防场景中高效运行,助力银行风险管理的数字化、智能化转型升级。

表 7 实验效果对比

实验组	样本数/个	应用策略	预警数/个	确认数/个	确认率
策略1	1 500万	账户基础特征规则	2.1万	4 230	20.1%
策略2	1 400万	账户基础特征规则+账户关联特征规则	4 067	2 621	64.4%

参考文献:

- [1] 吴建光. 金融科技与会计管理数字化转型[J]. 金融会计, 2020(1): 36-38.
WU J G. Financial technology and digital transformation of accounting management[J]. Financial Accounting, 2020(1): 36-38.
- [2] 刘瑞霞. 打造基于大数据的智能化风控体系[J]. 金融电子化, 2018(8): 57-58.
LIU R X. Configuration of intelligent risk management system based on big data[J]. Financial Computerizing, 2018(8): 57-58.
- [3] 姜增明, 陈剑锋, 张超. 金融科技赋能商业银行风险管理转型[J]. 当代经济管理, 2019, 41(1): 85-90.
JIANG Z M, CHEN J F, ZHANG C. The empowerment of financial technology on the risk management transformation of commercial banks[J]. Contemporary Economy & Management, 2019, 41(1): 85-90.
- [4] 柴洪峰, 王帅, 涂晓军, 等. 智能化金融科技创新监管工具: 理念、平台框架和展望[J]. 智能科学与技术学报, 2020, 2(3): 214-226.
CHAI H F, WANG S, TU X J, et al. Intelligent innovative regulatory tools on financial technology: concept, platform framework, and prospects[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(3): 214-226.
- [5] 陶睿, 吴继春, 谢胜强, 等. 深度学习和知识图谱在智能监管中的应用研究[J]. 金融纵横, 2019(8): 56-66.
TAO R, WU J C, XIE S Q, et al. Applied research on deep learning and knowledge graph in intelligent regulation[J]. Financial Perspectives Journal, 2019(8): 56-66.
- [6] 陶士贵, 相瑞. 基于大数据技术的商业银行反洗钱风险识别“穿透”研究[J]. 金融发展研究, 2020(7): 73-78.
TAO S G, XIANG R. Research on penetration of anti-money laundering risk identification of commercial banks based on big data technology[J]. Journal of Financial Development Research, 2020(7): 73-78.
- [7] 金磐石, 万光明, 沈丽忠. 基于知识图谱的小微企业贷款申请反欺诈方案[J]. 大数据, 2019, 5(4): 100-112.
JIN P S, WAN G M, SHEN L Z. Knowledge graph-based fraud detection for small and micro enterprise loans[J]. Big Data Research, 2019, 5(4): 100-112.
- [8] 王成, 舒鹏飞. WEB: 一种基于网络嵌入的互联网借贷欺诈预测方法[J]. 大数据, 2019, 5(6): 85-100.
WANG C, SHU P F. WEB: a fraud prediction method of Internet lending using network embedding[J]. Big Data Research, 2019, 5(6): 85-100.
- [9] 凡友荣, 杨涛, 孔华锋, 等. 基于知识图谱的电信欺诈通联特征挖掘方法[J]. 计算机应用与软件, 2019, 36(11): 182-187.
FAN Y R, YANG T, KONG H F, et al. Calling features mining method of telecom fraud based on knowledge graph[J]. Computer Applications and Software, 2019, 36(11): 182-187.
- [10] 魏瑾, 李伟华, 潘炜. 基于知识图谱的智能决策支持技术及应用研究[J]. 计算机技术与发展, 2020, 30(1): 1-6.
WEI J, LI W H, PAN W. Research on intelligent decision support technology and application based on knowledge graph[J]. Computer Technology and Development, 2020, 30(1): 1-6.
- [11] 陈成, 陈跃国, 刘宸, 等. 意图知识图谱的构建与应用[J]. 大数据, 2020, 6(2): 57-68.
CHEN C, CHEN Y G, LIU C, et al. Constructing and analyzing intention knowledge graphs[J]. Big Data Research, 2020, 6(2): 57-68.
- [12] 王昊奋, 丁军, 胡芳槐, 等. 大规模企业级知识图谱实践综述[J]. 计算机工程, 2020, 46(7): 1-13.
WANG H F, DING J, HU F H, et al. Survey on large scale enterprise-level knowledge graph practices[J]. Computer Engineering, 2020, 46(7): 1-13.
- [13] 胡扬, 闫宏飞, 陈翀. 面向金融知识图谱的实

- 体和关系联合抽取算法[J]. 重庆理工大学学报(自然科学), 2020, 34(5): 139-149.
HU Y, YAN H F, CHEN C. Joint entity and relation extraction for constructing financial knowledge graph[J]. Journal of Chongqing University of Technology(Natural Science), 2020, 34(5): 139-149.
- [14] 李任之, 李博杰, 张国栋, 等. 基于CouchDB和Elastic Search的高性能化学结构搜索引擎与数据库的构建[J]. 化学物理学报, 2018, 31(3): 341-349.
LI R Z, LI B J, ZHANG G Z, et al. A high-performance and flexible chemical structure & data search engine built on CouchDB & Elastic Search[J]. Chinese Journal of Chemical Physics, 2018, 31(3): 341-349.
- [15] 吴金全. 有向图的强连通分量及应用[J]. 软件, 2014, 35(3): 72-75.
WU J Q. The SCC of digraph and application[J]. Computer Engineering & Software, 2014, 35(3): 72-75.
- [16] 吕华揆, 洪亮, 马费成. 金融股权知识图谱构建与应用[J]. 数据分析与知识发现, 2020, 4(5): 27-37.
LYU H K, HONG L, MA F C. Constructing knowledge graph for financial equities[J]. Data Analysis and Knowledge Discovery, 2020, 4(5): 27-37.
- [17] 吴结凤, 黄学进, 李晋川, 等. 脉动流模拟系统血压波形相似度算法研究[J]. 医用生物力学, 2020, 27(5): 515-520.
WU J F, HUANG X J, LI J C, et al. Similarity measurement of blood pressure waveforms for pulsatile flow simulation system[J]. Journal of Medical Biomechanics, 2020, 27(5): 515-520.
- [18] 藏根林, 王亚强, 吴庆蓉, 等. 知识图谱中的关系方向与强度研究[J]. 大数据, 2019, 5(3): 96-103.
ZANG G L, WANG Y Q, WU Q R, et al. Study on direction and strength of relation based on knowledge graph[J]. Big Data Research, 2019, 5(3): 96-103.

作者简介



陈强 (1976-), 男, 博士, 兴业银行信息科技部创新技术总监、高级工程师、高级经济师, 厦门大学统计学与数据科学系兼职教授, 中国计算机学会 (CCF) 会员, 北京金融科技产业联盟人工智能专业委员会委员, 兴业银行集团科技架构专家委员会常任委员。主要研究方向为数据科学, 人工智能算法在金融风险控制、财富管理、理财投资等业务领域的应用研发及系统落地。



代仕娅 (1990-), 女, 蚂蚁科技国际事业群数据算法技术部数据产品专家, 主要研究方向为数据科学、人工智能相关产品在金融领域的研发及落地应用。

收稿日期: 2021-01-13

通信作者: 陈强, aoyu_77@qq.com

基金项目: 2018年上海市人工智能创新发展专项资金资助项目 (No.XX-RGZN-01-18-9814); 2020年中国计算机用户协会金融信息科技风险管理及审计领域最佳实践入库项目; 2019年国家互联网数据中心产业技术创新战略联盟 (NIISA) 技术创新二等奖项目

Foundation Items: 2018 Shanghai Artificial Intelligence Innovation and Development Project Supported by Special Fundation(No. XX-RGZN-01-18-9814), 2020 Best Practice Project in Financial Information Technology Risk Management and Audit Field Sponsored by China Computer Users Association, 2019 Second Prize of Technology Innovation Sponsored by National Internet Data Center Industrial Technology Innovation Strategic Alliance (NIISA)