

doi: 10.19920/j.cnki.jmsc.2024.03.004

考虑审计要素多重语义关联的财务欺诈识别^①

李建平^{1,4}, 孙 灏^{2,3}, 常闫芃^{2,3}, 朱晓谦^{1,4*}

(1. 中国科学院大学经济与管理学院, 北京 100190; 2. 中国科学院科技战略咨询研究院, 北京 100190;
3. 中国科学院大学公共政策与管理学院, 北京 100049; 4. 中国科学院大学数字经济监测预测预警与
政策仿真教育部哲学社会科学实验室(培育), 北京 100190)

摘要: 现有的财务欺诈识别研究大多基于公司、审计师、会计师事务所等审计要素中较为简单的关系特征, 罕有研究能够系统刻画各类审计要素之间错综复杂的关联关系. 本文创新性地引入知识图谱(Knowledge Graph)技术, 构建出包含公司、审计师和会计师事务所的多重语义关联网络, 并利用图神经网络(Graph Neural Networks)方法捕捉知识图谱中审计要素之间复杂隐秘的关联关系以提高财务欺诈识别效果. 基于我国 2018 年—2019 年的上市公司样本, 构建出包含 12 373 个审计要素和 111 194 条关系的审计知识图谱. 实证研究发现引入审计要素关联关系能够提升财务欺诈识别准确率; 在多种审计要素关联关系中, 考虑审计师对公司出具的审计意见对欺诈识别更为重要; 对比不同历史时长的审计要素, 使用公司历史 5 年的审计要素识别财务欺诈的效果更好. 本研究可以为投资者、分析师以及监管机构在大数据时代下的财务欺诈识别提供科学参考.

关键词: 财务欺诈; 审计要素关系; 语义关联; 知识图谱; 图神经网络

中图分类号: TP181; F830 文献标识码: A 文章编号: 1007-9807(2024)03-0058-13

0 引言

近年来康美药业、瑞幸咖啡等国内外上市公司的重大财务欺诈事件频发, 严重扰乱资本市场正常秩序, 阻碍金融经济的健康发展^[1,2]. 财务欺诈是指在财务报表中蓄意错报、漏报或泄露以欺骗使用者的行为^[3]. 根据国泰安经济金融研究数据库(China Securities Market and Accounting Research, CSMAR)的统计, 2008 年—2021 年我国涉嫌财务欺诈的公司达到 991 家, 涵盖 2 681 个公司财政年.

独立审计作为连接公司内部与外部投资者的重要纽带, 包含与财务欺诈相关的多种信息, 使用审计数据识别财务欺诈受到越来越多的关注^[4].

目前利用审计信息识别财务欺诈的研究主要基于较为简单的定量审计数据, 大致可以分为两类: 一类是分析审计费用、审计用时等反映审计工作具体内容的结构化数据与财务欺诈的关系, 如 Cao 等^[5]发现审计师在审计欺诈公司的财务报表时通常需要更多的审计用时. 还有一类关注审计师、会计师事务所等审计主体的特点在财务欺诈识别中的作用, 如刘文军等^[6]、Hobson 等^[7]认为具有较长任期、较高专业胜任能力的审计师能够更好地识别出公司的欺诈信号.

在实际审计业务中, 公司、审计师、会计师事务所等审计要素之间存在多重语义关联, 包括审计师与会计师事务所之间存在“任职”关系, 会计

① 收稿日期: 2022-10-26; 修订日期: 2022-10-28.

基金项目: 国家自然科学基金资助重点项目(92046023); 国家自然科学基金资助项目(72371236; 71971207); 中央高校基本科研业务费专项资金; 中国科学院大学数字经济监测预测预警与政策仿真教育部哲学社会科学实验室(培育)基金.

通讯作者: 朱晓谦(1989—), 女, 云南昆明人, 博士, 副教授. Email: zxq@ucas.ac.cn

师事务所与公司之间存在“委托”关系等,忽略这些关联关系可能会影响欺诈识别的结果^[8]。已有部分研究开始考虑审计要素之间关联关系,但大多关注单一类型的关联规则,如涉嫌欺诈的公司倾向频繁变更审计师^[9],欺诈公司可能与审计师存在合谋关系^[10,11];公司共用不良会计师事务所进行审计更易被查处涉嫌欺诈等^[12]。

不同审计要素之间的语义关联错综复杂,通过简单的关联规则难以有效刻画。例如当公司变更会计师事务所,可能是正常遵守证监会的审计师定期轮换规定,也可能是因为审计师跳槽至新的事务所时将其负责的部分客户一并带到新的事务所,公司委任的会计师事务所变更但审计师实际并未改变,即“换所不换师”现象^[9];还有研究认为涉嫌欺诈的公司更换事务所可能是为了寻找“听话”的事务所以默许其掩盖信息的行为^[13]。公司、审计师、会计师事务所之间语义关联的变化中也蕴含着与财务欺诈有关的信息,但现有的方法难以系统地整合这些存在复杂关联的审计要素,刻画其中隐藏的相关性^[14]。

本文以2018年—2019年中国上市公司为研究对象,引入知识图谱(Knowledge Graph)技术,基于公司、审计师、会计师事务所等审计要素之间的语义关联构建出审计知识图谱,系统刻画审计要素之间的多重语义网络,并进一步运用图神经网络(Graph Neural Networks, GNNs)中的Heterogeneous Graph Transformer(HGT)方法捕捉知识图谱中不同审计要素之间隐含的关联关系,提高财务欺诈识别效果。研究发现,与仅使用传统财务特征相比,进一步考虑不同审计要素之间基于语义的复杂关联关系可以提升财务欺诈识别的准确率。在多种类型的关联关系中,审计师对公司出具的审计意见对识别欺诈更为重要,并且对比不同时长的历史信息,使用上市公司历史5年审计信息的欺诈识别效果最好。

本文的创新和贡献主要体现在三个方面:第一,罕有研究实现对多种审计要素及其之间复杂关联关系的系统刻画,本文创新性地引入知识图谱技术,系统刻画出公司、审计师、会计师事务所等审计要素之间的语义关联;第二,主流机器学习方法难以处理知识图谱中以图结构表示的复杂关联关系,本文引入新兴的图神经网络方法捕捉知

识图谱中不同审计要素之间的关联特征,为识别欺诈提供更多有效信息;第三,分析了审计知识图谱中不同关联关系对识别公司财务欺诈的重要性,为复杂语义关联中关键关系的捕捉提供理论依据。

1 文献综述

审计作为上市公司预警系统的主要组成部分,包含着与财务欺诈相关的多种信息,现有文献使用审计信息对财务欺诈识别进行了大量研究^[5],大致可以分为两类。第一类是关注审计意见、审计费用、审计用时等反映审计工作具体内容的信息与财务欺诈的关系。例如,洪荭等^[15]发现审计意见在一定程度上可以揭露上市公司财务舞弊,并将其纳入财务欺诈识别指标体系中。还有研究指出审计费用中包含可以衡量公司会计质量的信息^[16,17]。Chakrabarty等^[18]提出一种基于Benford定律的新方法来评估审计费用和财务不当行为之间的关系。Lobo和Zhao^[19]度量了审计费用和公司随后进行财务重述之间的关系,发现较高的审计费用可能反映审计师评估公司欺诈风险的增加,并试图阻止欺诈行为。Cao等^[5]、Knapp等^[20]也指出,发现财务报表异常通常需要更多的审计精力,这表现为更高的审计费用和更长的审计用时。

还有一类研究通过分析公司聘请的审计师、会计师事务所等审计主体的特点,探究其在财务欺诈识别中的作用。审计师作为保障上市公司财务信息披露质量的屏障之一,其行业经验、任期都可能在一定程度上影响欺诈识别效果。例如,刘文军等^[6]、Hobson等^[7]分别通过分析审计师专长与审计失败的关系、审计师对公司电话会议内容的判断,发现具有较长任期、较高专业胜任能力的审计师可以帮助提高审计质量,更准确地识别出公司的欺诈信号。Patterson等^[21]、Brazel和Schmidt^[22]也指出,经验不足或任期较短的审计师可能更不容易识别出涉嫌欺诈的财务报表。此外,Perols^[23]、钱苹等^[24]认为会计师事务所的规模也可以作为识别财务欺诈的指标。

虽然现有研究采用了多种类型的审计信息进

行财务欺诈识别,但是这些信息相对独立,罕有研究考虑审计要素之间错综复杂的关联关系。在实际审计业务中,公司、审计师、会计师事务所等要素之间的关联关系例如“委任”、“任职”等,大多以语义关系的形式呈现,忽略这些关联信息也可能会影响财务欺诈的识别结果。例如,孙亮等^[25]、陈武朝等^[9]发现变更审计师在一定程度上会影响后任审计师的独立性,使欺诈公司达到收买审计意见的目的;Sun等^[12]发现共用同一网络会计师事务所的子公司和母公司更有可能因为有误导性或欺诈性的财务报告而受到监管机构的制裁。这些研究为探究审计要素之间的语义关联奠定了基础,但是目前在识别财务欺诈时使用的关联关系还较为简单。

综上所述,当前研究大多利用审计业务中较为简单直观的信息检测财务欺诈,少部分研究虽然开始关注不同审计要素之间的相关性,但实际审计业务中,审计要素之间的关联关系错综复杂,现有的方法难以对其进行系统描述并挖掘其中蕴含的复杂关联信息。利用知识图谱技术可以从语义的角度将公司、审计师、会计师事务所以关系图的形式连接成审计知识图谱,更清晰地理解审计业务活动。此外,与传统的机器学习方法不同,图神经网络技术可以处理图结构类数据中的复杂关联信息,捕捉知识图谱中审计要素之间隐秘的关联关系以提高财务欺诈识别效果^[26]。

2 研究方法

2.1 知识图谱构建方法

在实际的审计业务中,公司、审计师和会计师事务所等审计要素通过“委托”、“任职”等关联关系连接成错综复杂的多重语义网络,而已有的数值型关联关系刻画方法难以直接刻画这些语义关联,因此本文引入知识图谱方法刻画审计要素之间错综复杂的语义关联。知识图谱方法由Google在2012年正式提出,主要目的是刻画和解释现实世界中实体之间复杂的关联关系^[2]。

知识图谱是由实体节点和关系边构建的多关系图,包含实体、关系及属性三种知识要素。其最小组成单元为三元组(头实体,关系,尾实体),表

示两个实体通过特定关系连接在一起,同时可以通过三元组(实体,属性,属性值)对实体的属性信息进行描述^[27]。本文拟构建的中国上市公司审计知识图谱的示例如图1所示,包括三种类型的实体(公司、审计师、审计事务所)、三种类型的关系(任职、委托、审计意见)以及公司财务指标属性。基于公司的审计信息以及财务信息,实体、关系及属性的抽取过程如下:

1) 实体抽取:中国上市公司会在年度报告中披露聘请的会计师事务所的信息,其中“签字会计师姓名”及“会计师事务所名称”两项字段信息可以分别用于构建“审计师”及“会计师事务所”实体节点。考虑到“公司”实体发生欺诈行为具有时间信息,故根据被审计公司财务报表的会计年份的不同建立不同的公司实体。此外,部分公司实体名称会出现变化(如康美药业变为ST康美),及公司委托的会计师事务所可能为各省份的分所(如中证天通会计师事务所江苏分所),本文根据分别根据证券代码及会计师事务所证书编号对实体名称进行统一。

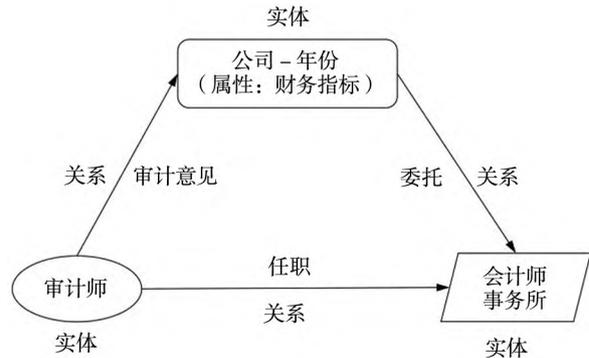


图1 审计知识图谱中的实体、关系和属性示例

Fig. 1 Example of entities, relations and properties in the audit knowledge graph

2) 关系抽取:审计要素的关联关系来源于实际审计过程,即公司委托会计师事务所审计其财务报表后,由在会计师事务所任职的审计师执行审计任务,并给公司出具审计意见。本文构建三种语义关联关系连接不同实体节点构成知识图谱,包括审计师和公司之间的“审计意见”关系,公司和会计师事务所之间的“委托”关系,审计师和会计师事务所之间的“任职”关系。考虑审计师根据财务报表的会计质量会出具不同审计意见,故根据审计意见类型建立不同类型的“审计意见”

关系。

3) 属性抽取: 考虑现有文献大多利用公司的财务信息识别欺诈行为^[3,24], 为了探究审计要素关联关系能否提升欺诈识别效果, 本文将公司财务报表中的相关财务指标作为公司实体的属性。

当公司在年度报告中披露委托会计师事务所对其年度财务报表进行审计, 在会计师事务所任职的审计师对该公司出具了审计意见, 则可如图 1 所示, 根据此审计信息建立 3 个三元组将“公司-年份”、“会计师事务所”、“审计师”实体节点连接在一起: (公司-年份, 委托关系, 会计师事务所) (审计师, 任职关系, 会计师事务所) (审计师, 审计意见关系, 公司-年份)。为了与基准指标的欺诈效果对比, 抽取公司的财务指标作为公司实体的属性信息, 并通过三元组表示为(公司-年份, 属性, 财务指标)。最终形成的中国上市公司审计知识图谱, 能够系统刻画多种审计要

素及其之间复杂的关联关系。下面将进一步引入图神经网络方法对审计知识图谱进行分析和知识推理, 以实现审计知识图谱中欺诈公司节点的识别。

2.2 图神经网络模型

图神经网络(GNNs)是一种基于图结构的深度学习方法, 用于处理包含复杂关联关系的图结构形式数据, 通过聚合节点自身以及邻居节点的特征信息, 完成关系信息的融合, 从而发现图中隐藏的信息^[2,28]。尽管主流机器学习方法可以处理欧几里得空间中的结构化数据, 但此类方法的一个核心假设是样本之间是相互独立的。而知识图谱属于非欧几里得空间中的图结构类数据, 个体之间并不是相互独立的, 而是通过关联关系连接在一起, 难以通过现有主流机器学习方法处理^[28]。因此, 本文引入 GNNs 方法处理中国上市公司审计知识图谱中的关联关系, 实现对公司实体的推理和分类, 识别财务欺诈公司。

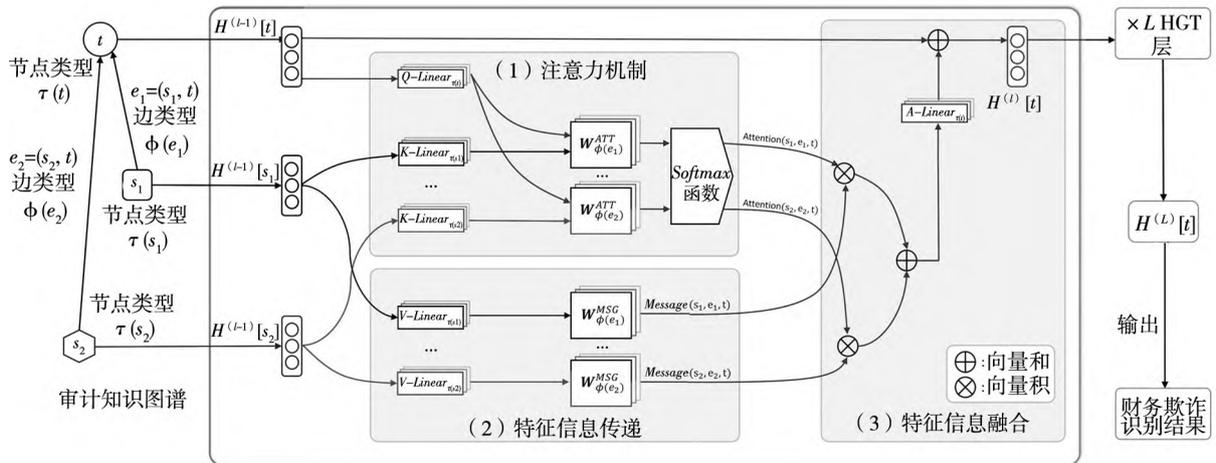


图 2 Heterogenous graph transformer (HGT) 模型结构示意图

Fig. 2 The framework of heterogeneous graph transformer (HGT)

针对节点和关系都只有一种类型的同构图, 常用的图神经网络模型包括 Graph Attention Network(GAT) 和 Graph Sample and Aggregate(Graph-SAGE) 等^[28]。本文构建的中国上市公司审计知识图谱中包含多种类型的实体节点和关系, 属于典型的异构图, 通常采用 Heterogeneous Graph Attention Network(HAN) 和 Heterogeneous Graph Transformer (HGT) 等模型进行处理^[29]。相比于其它适用于异构图的模型, Hu 等^[29]提出的 HGT 模型从元关系三元组(meta relation triplet) 的角度同时考

虑节点及关系的类型并为之构建相应的异构注意力机制, 能够更有效地捕捉异构图中的特征信息。因此本文主要采用 HGT 模型对审计信息知识图谱进行处理, 实现对欺诈公司的识别。

HGT 模型通过对图中邻近节点的特征信息进行融合, 利用多次迭代融合后的特征信息, 实现对图中实体节点的分类^[29]。HGT 的模型结构如图 2 所示, 首先保留知识图谱的结构, 将每个节点参数化为向量, 以矩阵的形式进行结构化表示, 输入到 HGT 模型的注意力层, 通过注意力机制衡量

目标节点的各邻居节点对自身的重要性,同时每一个邻居节点将自身的特征经过处理后传递到目标节点,然后目标节点根据其邻居节点的重要性及传递的特征信息进行信息融合,三个阶段的具体工作原理如下:

1) 注意力机制阶段: 衡量三元组 (s, e, t) 中源(邻居)节点 s 对于目标节点 t 的重要性,其计算方法如下

$$Attention_{(s, e, t)} = \text{Softmax}_{\forall s \in N(t)} (\|_{i \in [1, h]} ATT - head^i(s, e, t)) \quad (1)$$

$$ATT - head^i(s, e, t) = (K^i(s) W_{\phi(e)}^{ATT} Q^i(t)^T) \times \frac{\mu < \tau(s), \phi(e), \pi(t) >}{\sqrt{d}} \quad (2)$$

$$K^i(s) = K - \text{Linear}_{\tau(s)}^i(H^{(l-1)}[s]) \quad (3)$$

$$Q^i(t) = Q - \text{Linear}_{\tau(t)}^i(H^{(l-1)}[t]) \quad (4)$$

针对 h 个注意力机制中第 i 个,首先将类别为 $\tau(s)$ 的源节点 s 的特征 $H^{(l-1)}[s]$ 通过 $K - \text{Linear}_{\tau(s)}^i$ 线性层变换为键向量 $K^i(s)$,将类别 $\tau(t)$ 的目标节点 t 的特征 $H^{(l-1)}[t]$ 通过 $Q - \text{Linear}_{\tau(t)}^i$ 线性层变换为值向量 $Q^i(t)$;然后通过键向量 $K^i(s)$ 和值向量 $Q^i(t)$ 之间的相似度来衡量源节点 s 对于目标节点 t 的重要性,其中根据异构图中边 e 的类型 $\phi(e)$ 参数化权重矩阵 $W_{\phi(e)}^{ATT}$;而后根据三元组的类型 $\langle \tau(s), \phi(e), \pi(t) \rangle$, HGT 给三元组设定权重 $\mu_{\langle \tau(s), \phi(e), \pi(t) \rangle}$. 最终通过整合 h 个注意力机制并经过 Softmax 函数得到三元组 (s, e, t) 中源节点 s 对于目标节点 t 的重要性 $Attention(s, e, t)$.

2) 特征信息传递阶段: 对三元组 (s, e, t) 中需要传递至目标节点 t 的源节点 s 的特征信息进行变换,其计算方法为

$$Message(s, e, t) = \|_{i \in [1, h]} MSG - head^i(s, e, t) \quad (5)$$

$$MSG - head^i(s, e, t) = M - \text{Linear}_{\tau(s)}^i(H^{(l-1)}[s]) W_{\phi(e)}^{MSG} \quad (6)$$

对于 h 个特征变换器中的第 i 个,首先将类别 $\tau(s)$ 的源节点 s 的特征信息 $H^{(l-1)}[s]$ 经过 $M - \text{Linear}_{\tau(s)}^i$ 线性层变换为消息特征,然后通过对类型 $\phi(e)$ 的边 e 设立权重矩阵 $W_{\phi(e)}^{MSG}$ 得到第 i 个特征变换器的消息 $MSG - head^i(s, e, t)$. 最终通过整合 h 个消息头的消息得到源节点 s 向目标节点

t 传递的特征消息 $Message(s, e, t)$.

3) 特征信息融合阶段: 根据第一阶段各三元组 (s, e, t) 中源节点 s 对目标节点 t 的重要性 $Attention(s, e, t)$,对传递至目标节点的特征消息 $Message(s, e, t)$ 进行信息融合

$$\tilde{H}^{(l)}[t] = \bigoplus_{\forall s \in N(t)} (Attention(s, e, t) \times Message(s, e, t)) \quad (7)$$

$$H^{(l)}[t] = A - \text{Linear}_{\tau(t)}(\sigma(\tilde{H}^{(l)}[t])) + H^{(l-1)}[t] \quad (8)$$

根据重要性 $Attention(s, e, t)$,对目标节点 t 邻居节点集 $N(t)$ 的特征消息 $Message(s, e, t)$ 加权平均得到特征信息 $\tilde{H}^{(l)}[t]$,然后将目标节点 t 自身的特征信息 $H^{(l-1)}[t]$ 考虑在内经过 $A - \text{Linear}_{\tau(t)}$ 线性层得到融合后的新特征 $H^{(l)}[t]$.

记 HGT 中第 l 层神经网络输出的所有节点特征为 $H^{(l)}$,其可以作为第 $(l+1)$ 层 HGT 神经网络的输入,并通过上述步骤再一次进行信息融合以更新节点特征.通过构建 L 层的 HGT 神经网络,最终得到模型输出的图中所有节点特征 $H^{(L)}$,其中每个节点特征不仅包含自身的特征信息,而且包含了多次迭代更新后的多阶邻居节点特征. HGT 进一步利用此节点特征 $H^{(L)}$ 可以实现图中节点分类任务^[29]. 具体而言, HGT 对审计知识图谱中审计要素关联关系进行处理,进一步根据信息融合后的特征,得到公司节点的欺诈概率值,从而实现欺诈识别.

2.3 评价指标

由于公司财务欺诈识别可以看成二分类任务,本文选用了常用的分类评价指标对欺诈识别模型的效果进行评估,包括总准确率、一型准确率、二型准确率及 AUC(Area Under Curve)^[30,31]. 其中总准确率、一型准确率、二型准确率分别表示欺诈识别模型将所有公司、欺诈公司、非欺诈公司正确分类的百分比,具体定义如下

$$\text{总准确率} = \frac{TP + TN}{TP + FN + FP + TN} \quad (9)$$

$$\text{一型准确率} = \frac{TP}{TP + FN} \quad (10)$$

$$\text{二型准确率} = \frac{TN}{FP + TN} \quad (11)$$

其中 TP (True Positive) 为欺诈公司被正确识别为欺诈的数量, FN (False Negative) 为欺诈公司被错

误识别为非欺诈的数量, TN (True Negative) 为非欺诈公司为正确识别为非欺诈的数量, FP (False Positive) 为非欺诈公司被错误识别为欺诈的数量. AUC 指标通过 ROC (Receiver Operating Characteristic) 曲线下面积计算, 其取值范围在 0 和 1 之间, AUC 越接近 1, 分类效果越好.

3 实证分析

基于我国上市公司的审计信息, 构建包含审计要素复杂关联关系的审计知识图谱, 利用图神经网络中的 HGT 方法处理知识图谱数据以实现财务欺诈识别, 并进一步分析了审计要素之间各类关联关系的重要性.

3.1 实证数据

实证分析以中国 A 股上市公司为研究对象. 由于公司从发生财务欺诈至欺诈行为被披露一般需要 14 个月^[32], 为了保证样本中的公司能被准确界定为是否财务欺诈, 样本区间确定为 2018 年 1 月至 2019 年 12 月. 样本数据包括用于界定公司是否发生财务欺诈的财务违规数据、审计信息数据、以及财务指标数据, 所有数据均来源于国泰安经济金融研究数据库 (CSMAR).

为了分析引入审计要素关联关系对财务欺诈的提升效果, 实证研究中将大多数研究中采用的基于公司财务指标数据的欺诈识别结果作为比较基准. 由于 Bao 等^[31]构建的财务指标的欺诈识别效果优于常用的财务指标^[33], 本文参考该研究的财务指标作为实证研究的基准对比指标. Bao 等^[31]构建了 28 个美国公司的财务指标体系, 考虑到中国和美国会计指标的差异, 以及数据的可获得性, 最终确定了 10 个财务指标作为本文的基准指标, 分别为货币资金、存货净额、流动资产、非流动资产、应付账款、实收资本、留存收益、收盘

价、股票交易数量及个股交易金额.

在界定财务欺诈样本时, 参考 Liao 等^[34]的研究, 将存在虚构利润、虚列资产、虚假记载、重大遗漏、披露不实五类财务违规情况的公司界定为财务欺诈样本, 其他公司定义为非财务欺诈样本. 进一步剔除财务指标存在缺失值的样本后, 最终得到公司样本共计 5 469 个, 其中欺诈样本所占比例为 2.16%. 表 1 展示了样本的统计分布情况.

表 1 样本公司数量统计

Table 1 Statistics of company samples

年份	欺诈样本	非欺诈样本	样本总计	欺诈样本占比
2018	71	2 631	2 702	2.63%
2019	47	2 720	2 767	1.70%
总计	118	5 351	5 469	2.16%

如前所述, 公司的审计信息通常会在年报中披露, CSMAR 数据库根据各个公司的年报将审计信息整理为结构化的数据格式. 如表 2 展示了深大通公司 2014 年—2018 年间的五条审计信息, 包括会计年份、审计意见类型、审计师姓名、会计师事务所名称等. 因此本文直接从 CSMAR 数据库中提取公司、审计师及会计师事务所等审计要素, 以及各审计要素之间的关联关系.

3.2 中国上市公司审计知识图谱

根据 2.1 节中介绍的知识图谱构建方法, 构建中国上市公司审计知识图谱. 例如从表 2 可以看出深大通公司于 2018 年委托中汇会计师事务所对其年度报告进行审计, 审计师王甫荣和章归鸿出具了保留意见, 则可根据此审计信息提取“深大通 - 2018”、“中汇会计师事务所”、“王甫荣”、“章归鸿”实体及“委托”、“任职”、“保留意见”关系, 并通过三元组表示为 (深大通 - 2018, 委托, 中汇会计师事务所)、(王甫荣, 任职, 中汇会计师事务所)、(章归鸿, 任职, 中汇会计师事务所)、(王甫荣, 保留意见, 深大通 - 2018) 及 (章归鸿, 保留意见,

表 2 上市公司审计信息示例

Table 2 Examples of listed company audit information

序号	上市公司	年份	审计意见类型	审计师	会计师事务所
1	深大通	2018	保留意见	王甫荣, 章归鸿	中汇会计师事务所
2	深大通	2017	标准无保留意见	王甫荣, 章归鸿	中汇会计师事务所
3	深大通	2016	标准无保留意见	王甫荣, 章归鸿	中汇会计师事务所
4	深大通	2015	标准无保留意见	陈松波, 李东安	瑞华会计师事务所
5	深大通	2014	标准无保留意见	李泽浩, 李东安	瑞华会计师事务所

深大通-2018). 为了与基准财务指标的欺诈效果对比, 抽取公司的财务指标作为公司实体的属性信息. 例如深大通 2018 年的货币资金为 19 亿元、流动资产为 26 亿元, 通过三元组表示为(深大通-2018, 货币资金, 19 亿元)、(深大通-2018, 流动资产, 26 亿元). 从而根据基本单元三元组构建了上市公司审计知识图谱, 对审计要素中公司、审计师即会计师事务所的信息, 及各实体之间的关联关系进行了刻画, 以知识图谱的形式系统呈现.

考虑到连续多年的审计信息可能比公司当年

包含更丰富的信息, 例如可以体现审计师变动、会计师事务所变动等情况, 因此实证研究利用公司历史 5 年的审计信息来识别公司的财务欺诈. 例如基于 2014 年—2018 年的审计信息判别 2018 年公司是否存在欺诈, 基于 2015 年—2019 年的审计信息判别 2019 年公司是否存在欺诈, 利用公司历史 5 年的审计信息构建知识图谱. 表 3 展示了审计知识图谱中各类实体及关系的数量统计结果, 其中实体节点数量共有 12 373 个, 关系数量共有 111 194 条, 显示本文构建的中国上市公司审计知识图谱较为庞大, 关联关系也十分复杂.

表 3 审计知识图谱中各类实体和关系的数量统计

Table 3 Statistics of entities and relationships in the audit knowledge graph

类别	数量	
实体 (12 373)	公司	5 469
	审计师	6 857
	会计师事务所	47
关系 (111 194)	审计师-保留意见-公司	555
	审计师-保留意见加事项段-公司	104
	审计师-无保留意见加事项段-公司	1 414
	审计师-无法发表意见-公司	178
	审计师-标准无保留意见-公司	47 933
	审计师-任职-会计师事务所	36 332
	公司-委托-会计师事务所	24 678

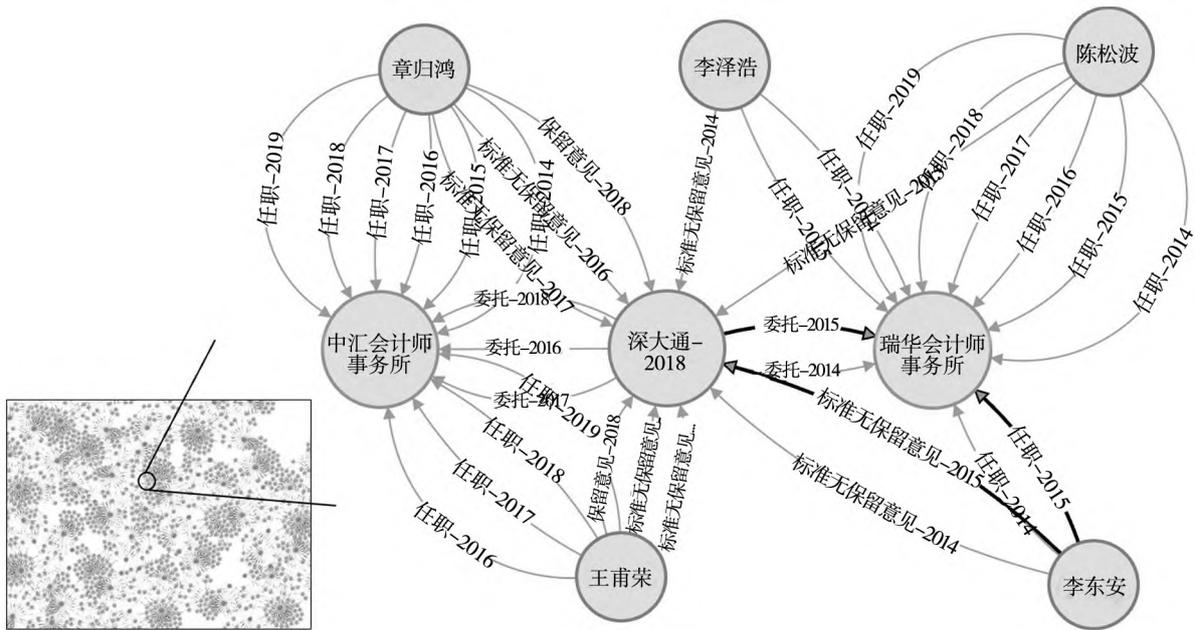


图 3 中国上市公司审计知识图谱及其局部示例

Fig. 3 Audit knowledge graph of listed companies in China and partial examples

为了更为直观地展示构建出的审计知识图谱, 本文利用常用的 Neo4j 图数据库对知识图谱

中的数据进行存储及可视化。图3展示了本文基于公司历史5年审计信息构建的中国上市公司审计知识图谱的可视化结果,由于其整体较为庞大且复杂,只展示了局部示例。从图3的示例中可以看出,审计知识图谱能够系统并且清晰地刻画出公司、审计师、会计师事务所之间错综复杂的关联关系。如加粗线条所示,对于2018年的深大通公司,其曾在2015年委托瑞华会计师事务所审计其财务报表,2015年在瑞华会计师事务所任职的陈松波接任此审计工作,最终出具了标准无保留意见。此外,图3还清楚展示了深大通公司的会计师事务所变动情况,即在2016年由瑞华会计师事务所变动为中汇会计师事务所。

3.3 财务欺诈识别结果的对比分析

本节利用图神经网络方法HGT对中国上市公司审计知识图谱进行分析,以识别财务欺诈。为了验证本文提出的财务欺诈识别方法的有效性,如3.1节中的叙述,将Bao等^[31]构建的财务指标的欺诈识别结果作为基准,与引入审计要素关联关系进行欺诈识别的效果进行对比。

在利用财务指标识别财务欺诈时,本文选用了五种主流的机器学习方法,包括逻辑回归、支持向量机、标准神经网络、随机森林和XGBoost^[2,31,35]。在各种模型的构建过程中,逻辑回归的正则化参数为5,支持向量机的核函数为高斯核函数、正则化参数为5,标准神经网络共4层、隐藏层维度为10,随机森林及XGBoost中基决策树数量为100。

在利用审计知识图谱识别财务欺诈时,首先将知识图谱转化为能够输入图神经网络模型的图结构数据,即知识图谱中的实体节点及关系——对应至图结构中的节点及边,公司实体的财务指标对应至节点的属性,进而利用图神经网络模型对图中的公司节点进行分类。如第2.2节中的描述,本文采用的图神经网络模型包括常见的GAT、GraphSAGE、HAN和HGT,通过Python图神经网络框架库PyTorch Geometric(PyG)实现,其中同构图神经网络GAT、GraphSAGE模型在处理本文构建的异构图时需要通过PyG库中的异构卷积包装器进行处理。在图神经网络构建过程中,各网络层数为4、隐藏层维度为10,带有注意力机

制的模型使用2个注意力头。

在图神经网络及主流机器学习模型训练过程中,采用现有研究中常用的样本数据集划分方式^[31],按照训练集80%、测试集20%,得到4375个训练样本及1094个测试样本。由于模型需要调整相应的超参数以提升模型的性能,本文从训练集中划分20%样本作为验证集,并采用网格搜索方法对各模型的超参数进行寻优。此外,由于财务欺诈识别问题中欺诈样本的数量通常显著少于非欺诈样本,具有样本不均衡的特点^[2]。采用代价敏感(cost sensitive)学习方法,对不同类别样本的分类错误赋以不同的惩罚权重,提高模型对欺诈类样本的重视程度,并通过网络搜索方法确定最优的惩罚权重^[36]。各模型共重复训练5次,从而得到在测试集上评估效果的平均值。

表4展示了分别基于财务指标、财务指标和简单审计要素关系指标、以及包含审计要素之间较为全面的关联关系的审计知识图谱的欺诈识别结果。本文参考Bao等^[31]构建的财务特征,基于主流机器学习方法,得出的欺诈识别结果显示XGBoost的AUC值最高,为78.37%,对应的总准确率为71.86%,一型准确率为70.83%,二型准确率为71.89%,表明通过机器学习分析财务指标的方法确实能够在一定程度上识别出公司财务欺诈。本文财务指标选取的参考研究,即Bao等^[31]基于美国公司数据的欺诈识别AUC值为72.50%,而本文基于中国公司数据的欺诈识别AUC值更高,验证了财务指标选取的合理性。

已有研究开始关注审计要素的关联关系信息,并从中挖掘出简单关系特征进行财务欺诈识别,因此本文在实证当中进一步引入现有研究中的审计要素关系指标^[9,12,24],包括更换审计师的频率、更换会计师事务所的频率、是否与其它欺诈公司共用审计师、是否与其它欺诈公司共用会计师事务所。表4展示了在财务指标的基础上引入这4类简单审计要素关系指标的识别效果,可以看出各类准确率相较于仅仅使用财务特征都有提升,其中XGBoost方法的AUC指标仍然最高,为79.20%,在仅使用财务特征的基础上提升了0.83%。

表4 财务欺诈识别结果对比

Table 4 Comparison results of company financial statement fraud identification

欺诈识别方法		特征	总准确率/%	一型 准确率/%	二型 准确率/%	AUC/%	训练 时间/s
主流机器学习方法	逻辑回归	财务指标	67.86	67.50	67.87	76.77	0.71
	支持向量机		66.36	66.67	66.36	76.13	2.40
	随机森林		72.65	68.33	72.75	77.81	10.52
	标准神经网络		70.66	70.83	70.65	77.34	14.51
	XGBoost		71.86	70.83	71.89	78.37	39.34
	逻辑回归	财务指标 + 简单 审计要素关系指标	72.63	68.33	72.73	77.94	1.57
	支持向量机		72.74	68.33	72.84	78.00	3.56
	随机森林		72.71	71.67	72.73	78.55	11.93
	标准神经网络		71.02	72.50	70.99	78.22	17.87
	XGBoost		72.36	70.83	72.39	79.20	46.14
图神经网络方法	GAT	审计知识图谱 (财务指标 + 全面的 审计要素关系)	73.25	73.46	73.24	81.45	157.03
	GraphSAGE		74.46	74.88	74.87	82.02	67.08
	HAN		76.64	75.83	76.65	83.20	197.72
	HGT		77.81	76.67	77.83	85.22	251.92

注: AUC 列加粗表示同类方法对比中欺诈效果 AUC 指标最优. 模型的运行环境为 Python 3.7.9, 设备为 Mac Mini(Apple M1 芯片), CPU 8 核, 内存 16G.

本文基于图神经网络方法,分析包含财务指标以及全面的审计要素关系的审计知识图谱,进行财务欺诈识别.从表4中可以看出,利用四种图神经网络方法分析审计知识图谱的欺诈识别效果进一步显著提升,GAT、GraphSAGE、HAN和HGT四种方法的AUC值分别为81.45%、82.02%、83.20%和85.22%,均显著高于利用财务指标及简单审计要素关系指标的最高AUC值(XGboost方法得到的79.20%).这表明除了4类简单的审计要素关系特征,本文构建的审计知识图谱中的关联关系包含更丰富的潜在特征.相比于用于处理结构化数据(财务指标及简单审计关系指标)的主流机器学习方法,图神经网络方法虽然需要更多的模型训练时间,但优势在于能够处理非结构化数据(复杂的审计要素关联关系)并从中挖掘潜在的有效信息,不过需要花费更多的模型训练时间作为代价.值得注意的是,在采用的四种图神经网络方法中,适用于处理异构图的HAN和HGT方法得到的各类准确率指标值均大于适用于处理同构图的GAT和GraphSAGE方法,其中HGT方法的欺诈识别效果最优,AUC达到85.22%.相比于使用财务指标和简单审计要素关系的XGBoost方法提高了6.02%,同时总准确率

提高5.45%,一型准确率提高5.84%,二型准确率提高5.34%.实证结果验证了HGT方法确实更适用于处理本研究中包含多种类型实体和关系的审计知识图谱,后续研究将继续采用HGT模型进行分析.

3.4 审计要素关联关系的重要性分析

本文构建的上市公司审计知识图谱中包含三种类型的关联关系,即公司与会计师事务所之间的委托关系、会计师事务所与审计师之间的任职关系、审计师和公司之间的审计意见关系.本节进一步通过消融实验探究了这三类关系在财务欺诈识别中的重要性.在利用HGT模型对各类关联关系进行分析时,分别删除审计信息知识图谱中任职关系、委托关系、审计意见关系,通过对比删除各类关系前后欺诈识别效果的变化程度,分析各类关系的重要性.

图4展示了包含完整关系的审计知识图谱,及删除任职关系、委托关系、审计意见关系后的欺诈识别结果的ROC曲线,其横轴、纵轴分别代表分类模型的评估指标特异性和灵敏性,ROC曲线下的面积即为AUC指标值,该面积越大表示欺诈识别效果越好.

从图4中可以看出,相较于审计信息知识图

谱中包含完整的三种关系时,分别删除三类关系后,HGT模型的欺诈识别效果都有所下降。进一步对比删除每一种关系后欺诈识别效果的变化程度,发现对比于删除任职关系(下降4.18%)和删除委托关系(下降1.6%),删除审计意见关系后AUC指标值下降最大,为8.41%。这表明在识别公司财务欺诈时,相比于公司委托会计师事务所、审计师任职于会计事务所的信息,审计师给公司出具的审计意见对识别欺诈更为重要。

为了验证欺诈和非欺诈公司之间各类审计意见是否存在显著差别,本文首先统计了欺诈和非欺诈公司各类审计意见的数量和比例,如表5所

示,发现各类审计意见占比在欺诈和非欺诈公司之间存在一定差异。为了检验差异是否显著,进一步将公司的审计意见转为虚拟变量,即为每个公司样本构建5个审计意见指标,根据审计意见类型将对应指标设为1,其他为0。表5展示了各类审计意见变量的统计性描述,通过独立样本t检验方法,发现各类审计意见指标在0.01水平下显著。对比欺诈和非欺诈公司各类审计意见指标的均值,可以看出非欺诈公司更倾向于收到“标准无保留意见”,而欺诈类公司更倾向于收到“保留意见”等其它审计意见。进一步验证了审计师出具的审计意见能相对有效地识别财务欺诈。

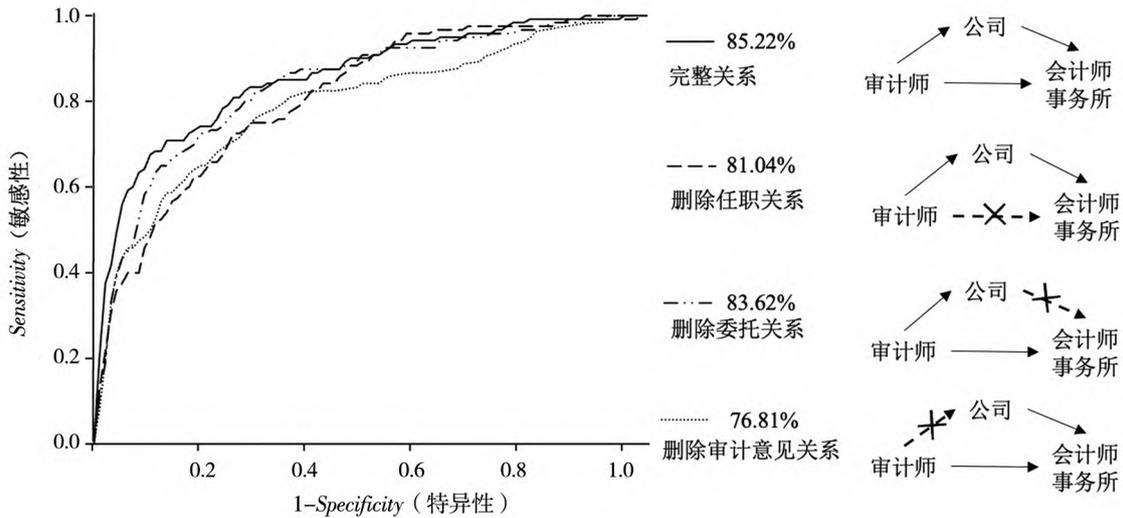


图4 审计知识图谱中分别去除各种要素关联关系后的欺诈识别效果对比

Fig. 4 Comparison results of fraud identification after removing the relationships between elements in the audit knowledge graph

3.5 审计信息历史时长对欺诈识别效果的影响分析

考虑到审计要素之间的关联关系变化需要通过一定时间段内的历史审计信息体现,本节进一步分析公司不同历史时长的审计信息对欺诈识别

效果影响。分别采用公司历史1年—7年的审计要素及其关联关系构建审计知识图谱,并利用图神经网络HGT模型进行财务欺诈识别。图5展示了审计知识图谱中分别包含公司1年—7年的审计要素关联关系时的欺诈识别结果。

表5 各类审计意见统计性描述和检验结果

Table 5 Statistical description and test results of audit opinions

审计意见	非欺诈				欺诈				t 检验
	数量	比例	均值	标准差	数量	比例	均值	标准差	
保留意见	194	1.78%	0.02	0.13	40	16.88%	0.17	0.38	-11.43***
保留意见加事项段	70	0.64%	0.01	0.08	12	5.06%	0.05	0.22	-5.53***
无保留意见加事项段	314	2.88%	0.03	0.17	22	9.28%	0.09	0.29	-4.01***
无法发表意见	56	0.51%	0.01	0.07	18	7.59%	0.08	0.27	-9.54***
标准无保留意见	10 279	94.19%	0.94	0.24	145	61.18%	0.61	0.49	14.62***

注: 样本中未出现“否定意见”故无此特征。***表示在0.01水平下显著。

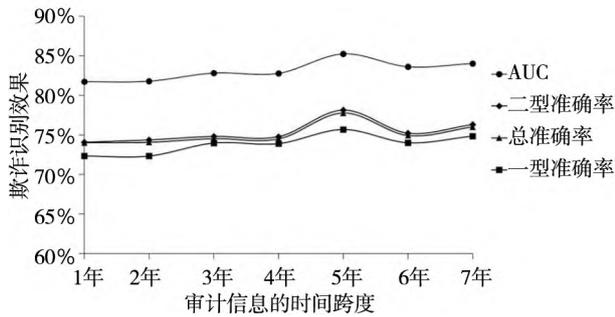


图5 不同时间跨度审计信息的欺诈识别效果

Fig. 5 The fraud identification results with audit information over different accounting periods

从图5中可以发现仅使用1年的审计要素关联关系的欺诈识别效果最差.此外,基于5年审计要素关联关系的识别效果最佳,验证了本文在构建审计信息知识图谱时选择5年审计信息是合理的.当审计信息的历史时长进一步增加,例如使用公司历史6年或7年的审计信息,欺诈识别效果相较于5年有所下降,可能是因为公司过于久远的审计关联信息难以反映公司当前的实际运营情况.总之,在判断公司的财务欺诈时,不仅需要考察公司当前审计信息,还要合理考虑公司历史审计信息.基于本文结果,选择公司历史5年的审计信息进行财务欺诈识别较为合适.

4 结束语

本文通过引入知识图谱技术,实现对审计业务中公司、审计师、会计师事务所等审计要素之间

复杂关联关系的系统刻画,并采用图神经网络方法中的HGT模型处理知识图谱中的关联信息,以更为有效地识别财务欺诈.

基于2018年—2019年中国上市公司样本,实证研究发现:1)在现有研究基于公司财务特征指标识别欺诈的基础上,引入审计要素相关性可以提升财务欺诈识别的准确率.在防范欺诈风险时,监管机构除了关注公司的财务信息,还应当审查公司审计业务中不同审计要素之间的关联关系;2)相较于审计业务活动中公司与审计事务所之间的委托关系、审计事务所与审计师之间的聘用关系,审计师给公司出具的审计意见对欺诈识别的影响最大,在识别财务欺诈时应当重点关注;3)对比不同时长的历史审计信息对财务欺诈的识别效果,使用历史5年的审计信息识别财务欺诈的准确率最高.因此在判断财务欺诈时不仅应当考虑当年的审计信息,还应当合理考虑公司历史审计信息中审计要素关联关系的变化.

本文的研究结果可以为投资者、分析师以及监管机构识别上市公司财务欺诈提供了考虑审计要素相关性的新视角.在未来的研究中,可以将更多类型的审计要素纳入研究范围,例如现有研究中常用的审计费用、审计用时,充分利用审计业务中产生的多类型信息.此外,虽然图神经网络方法可以较为有效地处理知识图谱中审计要素之间的语义信息,但是难以推理出清晰的关联规则,模型的可解释性需要进一步提升.

参考文献:

- [1] 马长峰, 陈志娟, 张顺明. 基于文本大数据分析的会计和金融研究综述[J]. 管理科学学报, 2020, 23(9): 19-30. Ma Changfeng, Chen Zhijuan, Zhang Shunming. A survey on accounting and finance research based on textual big data analysis[J]. Journal of Management Sciences in China, 2020, 23(9): 19-30. (in Chinese)
- [2] Zhu X, Ao X, Qin Z, et al. Intelligent financial fraud detection practices in post-pandemic era[J]. The Innovation, 2021, 2(4): 100176.
- [3] 袁先智, 周云鹏, 严诚幸, 等. 财务欺诈风险特征筛选框架的建立和应用[J]. 中国管理科学, 2022, 30(3): 43-54. Yuan Xianzhi, Zhou Yunpeng, Yan Chengxing, et al. A feature extraction method on corporate financial fraud[J]. Chinese Journal of Management Science, 2022, 30(3): 43-54. (in Chinese)
- [4] 王鲁平, 陈羿. 管理舞弊的形成机理及治理对策研究[J]. 管理工程学报, 2018, 32(1): 107-116. Wang Luping, Chen Yi. Research on formation mechanism and countermeasure of management fraud[J]. Journal of Industrial Engineering and Engineering Management, 2018, 32(1): 107-116. (in Chinese)
- [5] Cao J, Luo X, Zhang W. Corporate employment, red flags, and audit effort[J]. Journal of Accounting and Public Policy, 2020, 39(1): 106710.

- [6]刘文军,米莉,傅惊轩. 审计师行业专长与审计质量——来自财务舞弊公司的经验证据[J]. 审计研究,2010,(1): 47-54.
Liu Wenjun, Mi Li, Fu Jingxuan. Auditor industry specialization and audit quality [J]. Auditing Research, 2010, (1): 47-54. (in Chinese)
- [7]Hobson J L, Mayew W J, Peecher M E, et al. Improving experienced auditors' detection of deception in CEO narratives [J]. Journal of Accounting Research, 2017, 55(5): 1137-1166.
- [8]董沛武,程璐,乔凯. 客户关系是否影响审计收费与审计质量[J]. 管理世界,2018,34(8): 143-153.
Dong Peiwu, Cheng Lu, Qiao Kai. Does client-auditor relationship affect audit fees and audit quality? [J]. Management World, 2018, 34(8): 143-153. (in Chinese)
- [9]陈武朝,张海燕,范语辰. 审计师变更与审计质量——基于签字注册会计师未变更的经验证据[J]. 管理评论,2020,32(12): 204-220.
Chen Wuzhao, Zhang Haiyan, Fan Yuchen. Auditor change and audit quality: Empirical evidence from the perspective of signatory CPAs unchanged [J]. Management Review, 2020, 32(12): 204-220. (in Chinese)
- [10]余玉苗,田娟,朱业明. 审计合谋的一个博弈均衡分析框架[J]. 管理科学学报,2007,10(4): 32-37.
Yu Yumiao, Tian Juan, Zhu Yeming. Game equilibrium analysis framework for auditing collusion [J]. Journal of Management Sciences in China, 2007, 10(4): 32-37. (in Chinese)
- [11]Markelevich A, Rosner R L. Auditor fees and fraud firms [J]. Contemporary Accounting Research, 2013, 30(4): 1590-1625.
- [12]Sun J, Wang J, Kent P, et al. Does sharing the same network auditor in group affiliated firms affect audit quality? [J]. Journal of Accounting and Public Policy, 2020, 39(1): 106711.
- [13]Chen F, Peng S, Xue S, et al. Do audit clients successfully engage in opinion shopping? Partner-level evidence [J]. Journal of Accounting Research, 2016, 54(1): 79-112.
- [14]洪亮,马费成. 面向大数据管理决策的知识关联分析与知识大图构建[J]. 管理世界,2022,38(1): 207-219.
Hong Liang, Ma Feicheng. Knowledge association analysis and big knowledge graph construction for big data management and decision-making [J]. Journal of Management World, 2022, 38(1): 207-219. (in Chinese)
- [15]洪荃,胡华夏,郭春飞. 基于 GONE 理论的上市公司财务报告舞弊识别研究[J]. 会计研究,2012,(8): 84-90,97.
Hong Hong, Hu Huaxia, Guo Chunfei. Research on the identification of list companies' financial reporting fraud: Based on GONE theory [J]. Accounting Research, 2012, (8): 84-90, 97. (in Chinese)
- [16]赵子夜,杨庆,杨楠. 言多必失? 管理层报告的样板化及其经济后果[J]. 管理科学学报,2019,22(3): 53-70.
Zhao Ziye, Yang Qing, Yang Nan. The less said the better? Economic consequences of textual similarity in management discussion and analysis [J]. Journal of Management Sciences in China, 2019, 22(3): 53-70. (in Chinese)
- [17]梁日新,李英. 年报文本语调与审计费用——来自我国 A 股上市公司的经验数据[J]. 审计研究,2021,(5): 109-119.
Liang Rixin, Li Ying. Annual report text ton and audit fees: Empirical data from Chinese A-share listed companies [J]. Auditing Research, 2021, (5): 109-119. (in Chinese)
- [18]Chakrabarty B, Duellman S, Hyman M A. A new approach to estimating the relation between audit fees and financial misconduct [J]. Accounting Horizons, 2019, 34(2): 41-61.
- [19]Lobo G J, Zhao Y. Relation between audit effort and financial report misstatements: Evidence from quarterly and annual restatements [J]. The Accounting Review, 2013, 88(4): 1385-1412.
- [20]Knapp C A, Knapp M C. The effects of experience and explicit fraud risk assessment in detecting fraud with analytical procedures [J]. Accounting, Organizations and Society, 2001, 26(1): 25-37.
- [21]Patterson E R, Smith J R, Tiras S L. The effects of auditor tenure on fraud and its detection [J]. The Accounting Review, 2019, 94(5): 297-318.
- [22]Brazel J F, Schmidt J J. Do auditors and audit committees lower fraud risk by constraining inconsistencies between financial and nonfinancial measures? [J]. Auditing: A Journal of Practice & Theory, 2018, 38(1): 103-122.
- [23]Perols J. Financial statement fraud detection: An analysis of statistical and machine learning algorithms [J]. Auditing: A Journal of Practice & Theory, 2011, 30(2): 19-50.
- [24]钱苹,罗玫. 中国上市公司财务造假预测模型[J]. 会计研究,2015,(7): 18-25,96.
Qian Ping, Luo Mei. Predicting accounting fraud in China [J]. Accounting Research, 2015, (7): 18-25, 96. (in Chinese)
- [25]孙亮,刘春,柳建华. 御用会计师: 合作抑或合谋[J]. 管理科学学报,2016,19(2): 109-126.
Sun Liang, Liu Chun, Liu Jianhua. Underwriter-accountants: Collusion or cooperation [J]. Journal of Management Sci-

- ences in China, 2016, 19(2): 109–126. (in Chinese)
- [26]肖斌卿, 杨 旻, 李心丹, 等. 基于模糊神经网络的小微企业信用评级研究[J]. 管理科学学报, 2016, 19(11): 114–126.
Xiao Binqing, Yang Yang, Li Xindan, et al. Research on the credit rating of small and micro enterprises based on fuzzy neural network[J]. Journal of Management Sciences in China, 2016, 19(11): 114–126. (in Chinese)
- [27]Ji S, Pan S, Cambria E, et al. A survey on knowledge graphs: Representation, acquisition, and applications[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(2): 494–514.
- [28]Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4–24.
- [29]Hu Z, Dong Y, Wang K, et al. Heterogeneous Graph Transformer[C]// Proceedings of The Web Conference, 2020: 2704–2710.
- [30]Bertomeu J, Cheynel E, Floyd E, et al. Using machine learning to detect misstatements[J]. Review of Accounting Studies, 2021, 26(2): 468–519.
- [31]Bao Y, Ke B, Li B, et al. Detecting accounting fraud in publicly traded U. S. firms using a machine learning approach[J]. Journal of Accounting Research, 2020, 58(1): 199–235.
- [32]Correlation of Certified Fraud Examiners. 2020 Global Study on Occupational Fraud and Abuse[R]. Austin, USA, 2020.
- [33]Dechow P M, Ge W, Larson C R, et al. Predicting material accounting misstatements[J]. Contemporary Accounting Research, 2011, 28(1): 17–82.
- [34]Liao L, Chen G, Zheng D. Corporate social responsibility and financial fraud: Evidence from China[J]. Accounting & Finance, 2019, 59(5): 3133–3169.
- [35]Al-Hashedi K G, Magalingam P. Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019[J]. Computer Science Review, 2021, (40): 100402.
- [36]Hoppner S, Baesens B, Verbeke W, et al. Instance-dependent cost-sensitive learning for detecting transfer fraud[J]. European Journal of Operational Research, 2022, 297(1): 291–300.

Financial statement fraud identification considering the multiple-dimensional semantic associations of auditing elements

LI Jian-ping^{1,4}, SUN Hao^{2,3}, CHANG Yan-peng^{2,3}, ZHU Xiao-qian^{1,4*}

1. School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China;
2. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China;
3. School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing 100049, China;
4. MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation at UCAS, Beijing 100190, China

Abstract: Existing studies on financial fraud identification mainly focus on analyzing relatively simple features of audit elements such as companies, auditors, and audit firms. Few studies can systematically describe the intricate relationships between various auditing elements. To improve the accuracy of financial fraud identification, this paper innovatively introduces the knowledge graph technology to construct a multi-dimensional semantic network including companies, auditors, and audit firms, and applies graph neural network models to analyze the complex relationships among the auditing elements in the knowledge graph to better detect financial fraud. Based on the auditing information of listed companies in China from 2018 to 2019, an auditing knowledge graph containing 12 373 auditing elements and their 111 194 relationships is constructed. The empirical results demonstrate that the relationships among auditing elements can indeed help improve the accuracy of financial fraud identification. Among the various relationships, the auditing opinion of auditors to companies is more important for fraud identification. Using the auditing element information of a company in the past 5 years to identify fraud is better than in other historical periods. This study can provide a scientific reference for investors, analysts, and regulators to identify financial fraud in the era of big data.

Key words: financial statement fraud; audit element correlation; semantic association; knowledge graph; graph neural network