山西票号知识表示模型构建与知识发现研究

毋江波　朱鼎弘
（山西大学经济与管理学院，太原 030031）

摘要：山西票号的创立是中国近代金融业的开端，山西票号的经营模式和管理制度对现代金融业具有重要启示价值。整合票号领域分散的知识资源，构建系统化的知识表示模型，有利于实现票号文化遗产的数字化保护、深度挖掘和跨领域应用。本研究首先通过本体构建七步法将金融本体框架与山西票号知识深度融合，构建由山西票号静态本体、社会本体及动态本体构成的三维票号知识表示模型，覆盖票号组织机构、业务逻辑和制度体系等核心内容；其次借助Protégé本体编辑与Neo4j图数据库工具构建知识图谱；最后实现知识可视化和知识发现。本研究为票号文化的知识关联和系统阐释、数字化保护和有效开发利用提供了较为完整的知识框架，同时，从业务流程、制度佐证和风险识别等不同角度进行的知识推理也充分展示了本模型在票号价值发现方面的潜力。

关键词：金融本体；山西票号；知识表示；知识发现
中图分类号：G254；F832　DOI：10.3772/j.issn.1673-2286.2026.04.006
引文格式：毋江波，朱鼎弘. 山西票号知识表示模型构建与知识发现研究[J]. 数字图书馆论坛，2026，22（4）：56-67.

票号，亦称票庄、汇票庄或汇兑庄，是一种专门经营汇兑业务的金融机构。山西票号传承百年的经营理念、管理制度与风控机制等，不仅为中国金融业发展留下了宝贵遗产，也是传统金融智慧的集中体现。在金融市场瞬息万变的当下，现代金融业发展面临新的机遇与风险，对票号知识进行系统梳理，不仅可以促进对票号历史和文化价值的有效发现，也可以鉴古知今，为现代金融业的发展提供丰富的实践参照、宝贵经验和重要启示。

票号研究多聚焦史料抢救与汇编[1-3]、内部文书保护[4-6]、传统建筑修复[7-8]、票号经营管理[9-10]、人力资源管理[11-12]、经营模式[13-14]、风险控制[15-16]等问题，主要采用传统史学方法或定性分析方法展开，缺乏运用数字技术进行的知识组织与挖掘研究，难以有效揭示票号知识之间的关联和丰富内涵。本体是从客观世界抽象出的一组概念集合，提供了对给定领域的一种共识，为相关领域提供了一套标准化的术语体系[17]，以属性关系为基础支持知识的关联发现。现代金融领域已建有相对成熟的金融本体，这些金融本体在文化遗产[18-20]、文献资源[21-23]、金融知识[24]等与票号资源高度关联的领域已有广泛应用。在金融知识表示方面，已有研究构建了多元时序金融超图[25]、多层金融股权本体立方体[26]；在金融知识发现方面，已有研究探索了金融本体在风险识别[24]、事件推理[25,27]中的应用。上述研究为本研究提供了基本的金融知识框架和方法论参考，但现有研究尚未涉及以票号知识为代表的传统金融领域知识，难以支持对金融业发展历程的全面描述及其文化内涵的深度挖掘。

为了促进票号知识的系统化组织和语义关联发现，弥补现有金融本体缺乏传统金融领域知识表示的

收稿日期：2026-03-10

不足，本研究采用Protégé本体编辑和Neo4j图数据库工具，通过本体构建七步法将金融本体框架与山西票号知识深度融合，构建覆盖票号组织机构、业务逻辑和制度体系等核心内容的山西票号知识表示模型，以促进票号价值的有效发现与利用。

1　数据来源

本研究以票号相关史料为主要语料，并通过文献研究和实地调研补充语料。

1.1　史料汇编

本研究主要以内容翔实、系统权威的经典史料汇编为核心语料基础。2002年出版的《山西票号史料》[1]系统整合了票号经营过程中形成的原始文献，涵盖账簿、总分号往来公信、商业合约、号规制度、人物事略等多元数据，全面覆盖票号的机构设置、分号网络、业务流程、激励机制、风险管控措施等核心内容，同时包含票号兴衰历程中的关键历史事件。卫聚贤[3]的著作《山西票号史》内容涵盖经济背景、地域帮派、组织管理、人事制度和业务经营细节，整合大量原始信件、账簿及人物事略，可以提供晋商金融实践的必要补充。因此，这两本著作构成了本研究的核心语料。

1.2　文献研究与实地调研

鉴于上述著作完成时间较早，搜集史料范围受到当时研究方法、史料挖掘技术及学术视野的限制，可能仍有疏漏，且无法反映新出现的资料。因此，本研究通过文献研究和实地调研进行语料补充。

科研论文可以涵盖票号研究的最新成果。本研究以“票号”为检索词，以主题、关键词、篇名为检索入口，在中国知网开展检索，经人工筛选剔除信息不全、与研究主题不符的文献，最终得到有效论文549篇。经过LDA（Latent Dirichlet Allocation）主题分析发现，现有研究主要集中在金融行业、管理制度、分号分布、外部影响、历史文化五大领域，为了使语料更加集中，根据文献的研究深度和内容全面性，进一步精选出36篇论文作为补充语料。

同时，研究团队于2024年7月先后走访山西平遥、祁县等票号发源地的博物馆、图书馆及传统票号旧址，不仅搜集到《山西票号史料》中未收录的账本实物、票号家族族谱等资料，还通过与当地票号史专家、票号讲解员进行交流，获取了有关票号内部管理等方面的口述历史资料，这些语料与经典史料记载相互印证、互为补充，有效弥补了单一文献史料在具象化场景还原、隐性规则呈现等方面的不足，提升了数据来源的全面性、丰富性与立体性。

本研究通过扫描、图文转换、视频音频文字提取等方式将上述语料统一转换为文本语料，经过人工编码标注，完成基础语料库构建。山西票号相关语料及核心内容如表1所示。

表1　山西票号相关语料及核心内容

语料来源		核心内容
史料汇编	《山西票号史料》	票号股本与组成手续
		票号内部组织人员设置
		票号汇兑制度及发展
		票号账务与报告制度
		票号利润来源与红利分配
		票号的经营方针与盈利手段
		票号资本家及其他投资人类型
		票号经理与财东的关系
		票号与官吏的关系
		票号与清政府的关系
		票号与近代工业交通的关系
		票号在国内外商业、对外贸易中的作用
		票号存放款业务、汇兑业务方式及其他
		票号发行与代办捐项
		票号汇兑公款成为省关财政支柱
	《山西票号史》	票号的帮别、各路平码、分号设立
		票号营业报告与来往信件具体史料
学术文献	中国知网	票号经营、管理制度、人事、社会关系、业务、事件、风险相关论文36篇
实地调研	雷履泰故居	票号产权制度：股份制、两权分离制度
		票号管理制度：铺保制、学徒制在内共8项
		日昇昌分号网络分布图
	日昇昌票号旧址	日昇昌票号财东、经理相关史料
	蔚泰厚票号旧址	“蔚”字五联号财东、经理相关史料
		蔚泰厚票号经营年谱（1826—1921年）
	协同庆钱庄博物馆	协同庆票号年谱
		协同庆票号财东、经理相关史料
	平遥一城两寺研究院	平遥22家票号相关业务事件史料
	乔家大院	大德通票号经营管理制度
	渠家大院	百川通、三晋源、永春原票号介绍
	日昇昌讲解员	票号各项制度讲解

2 山西票号知识表示模型构建

2.1 构建方法

斯坦福大学本体工程小组开发的七步法[28]是构建知识表示模型的通用方法。因此，本研究采用七步法，并参照唐旭丽等[24]对金融本体的分类，从静态本体、社会本体和动态本体三个维度构建山西票号知识表示模型。

第一步：明确本体领域范围。本研究聚焦山西票号机构、业务、制度、社会关系与风险管控机制等内容，时间上覆盖1823年我国第一家票号日昇昌创立至20世纪20年代票号衰亡的完整历史周期，在空间上以平遥、祁县、太谷这3个山西票号发源地为核心。

第二步：选取参考本体。FIBO（Financial Industry Business Ontology）是最具代表性的金融本体，通过众包形式，基于层级关联完成金融基本概念描述，可以支持金融静态本体的构建[24]。本研究选取FIBO以及人文领域适用性较高的通用本体模型和词表为参照，包括CIDOC-CRM模型、FOAF词表、Schema.org词表。相关本体与词表为山西票号知识表示模型构建提供了多维度、体系化支撑，能有效提升本体模型的科学性、完整性。其中，FIBO中成熟的标准化金融术语，能为票号知识表示模型提供金融术语定义参考；FOAF词表对定义票号人物、梳理人物社会关系等特定场景有重要的借鉴意义；CIDOC-CRM模型主张以事件为纽带，将人、物、时、空进行关联适配票号以金融活动为核心、多方参与、时空交织的业务特点；Schema.org词表可适配票号组织、机构设立、分号地点等多元实体的定义需求。

第三步：自下而上梳理概念术语。本研究以前期搜集的语料为基础，围绕专著目录、核心内容、文献标题、关键词、摘要、重点段落等进行梳理，以广泛收录、适度筛选为原则，尽可能多地覆盖具有代表性的专门概念术语，并进一步提炼山西票号的核心概念范畴，部分山西票号核心概念术语如表2所示。

表2 山西票号核心概念术语（部分）

概念范畴	概念定义	术语示例
组织机构	与票号经营发展直接相关的各类组织、机构实体，涵盖票号本身、关联金融机构、业务关联机构	票号、银号、钱庄、银行、政府、工商企业
票号人员	参与票号运营、管理、业务执行、后勤保障等各个环节的各类工作人员	财东、经理、协理、襄理、跑街、学徒
关联人员	不隶属于山西票号内部，但与票号经营发展、业务开展存在密切关联的各类外部人员	官员、皇室成员、实业家
划分部门	票号内部依据业务职能、管理需求划分的专业化组织单元	账房、柜房、信房、中厅
经营业务	票号开展的主要金融服务及相关业务类型	异地汇兑、存放款、代办捐项、公款汇兑
业务凭证	票号明确权利义务、验证业务真实性的各类凭证	汇票、兑条、即票、迟票
制度规范	票号内部运营、管理及业务开展所遵循的规则体系	两权分离制、身股制、总分号制、联号制
金融文件	票号业务过程中产生的各类账目、信件等书面材料	营业报告、往来信件、账簿
资本类型	票号运营过程中涉及的各类资本形态及相关权益	正本、副本、银股、身股

第四步至第六步：定义顶层概念类、结构层次及类的属性。因本研究遵循静态本体、社会本体、动态本体的顺序进行模型构建，静态本体的顶层概念类、结构层次及类的属性定义是基础，为便于后续社会本体、动态本体的描述，将第四步至第六步合并阐述。

第七步：创建实例，构建山西票号知识图谱。以前文构建的基础语料库文本为基础，以大语言模型辅助和人工核查相结合的方式进行实体、属性和关系抽取，构建以三元组存储、支持查询与推理的图数据库。

2.2 本体构建

2.2.1 山西票号静态本体

静态本体用于描述金融市场的基本概念、属性及概念间的关系，主要涵盖六大顶层概念：自主实体、机构、时间、空间、事件和金融合约[24]。票号静态本体是静态本体在票号知识领域的体现，是票号及相关金融活动的静态知识描述体系，重在定义票号这一传统金融领域内的基本概念、本质属性、分类层级与概念间静

态关联。首先，基于前期术语概念的梳理并结合静态本体六大顶层概念，本文选取FIBO以及票号领域适用性较高的通用本体和词表作为参照。人物类使用FOAF词表，机构类使用Schema.org词表，时间类、地点类、事件类使用CIDOC-CRM模型，文件类使用FIBO，完成票号静态本体初步构建。其次，由于现有本体中尚无可复用的制度知识参照，无法反映“制度规范”这一核心概念范畴的相关知识，因此，本研究根据表2提炼的概念范畴定义和相关术语自建词表Shanxi Piaohao Resources（SPR），以支持对制度类（spr:Regime）等概念的描述。至此，本研究完成7个票号静态本体顶层概念的定义：人物类指从资本供给到日常运营的全链条票号内部角色，机构类指票号本身及经营演化过程中的衍生机构，时间类指票号发展过程中的重要时间，地点类指票号所在地理位置，事件类指票号内部人员及票号自身的活动，制度类指支撑票号运营及人员管理的制度规范，文件类指业务记录与权益凭证等。最后，自上而下与自下而上相结合，根据顶层概念的基本含义向下划分子类，选择相关概念术语向上归纳，完成山西票号静态本体顶层概念类及子类定义（见表3）。

表3　山西票号静态本体顶层概念类及子类定义

顶层概念类	子类	子类定义
人物类（foaf:Person）	财东	票号出资者与所有者
	经理	票号经营管理者
	信房工作人员	负责处理票号文书信件的工作人员
	柜房工作人员	负责驻守柜台的业务执行人员
	账房工作人员	负责票号账务的工作人员
	中厅工作人员	负责接待、业务接洽、部门协调的工作人员
	学徒	通过选拔进入票号学习的年轻从业者
机构类（org:Organization）	票号	中国近代晋商创办的专业金融机构
	票号前身	票号创办前晋商经营的商业机构
	银行	综合性金融机构
时间类（crm:TimeSpan）	时间点	明确的具体时刻
	时间段	具有特定特征的时间区间
地点类（crm:Place）	具体地址	具有明确指向的具体位置
	分布地区	一定规模的地域空间
事件类（crm:Event）	历史事件	票号发展过程中的重要事件，如开设分号等
	业务事件	票号经营过程中所承揽的各类金融业务活动
文件类（fibo:Document）	金融合约	票号业务过程中开具的汇票、兑条等凭证
	营业报告	分号对于总号报告全月及通年营业状况的账目
	往来信件	总号和分号之间互通消息的信件
制度类（spr:Regime）	体系制度	票号支撑核心运营的基础性制度
	管理制度	票号针对内部人员管理制定的规章制度

山西票号静态本体关系定义如表4所示。通过对上述顶层概念类及子类关系的梳理，结合对语料中票号组织机构、业务流程、制度体系等核心知识内容的逻辑分析，本研究归纳并定义了7种顶层概念关系。定义域是源类，值域是目标类，其中人物类之间可能存在亲属关系，如李大全是日昇昌票号财东，之后其子李箴视接任。进一步根据票号知识的核心内容表达要求，完成对顶层概念类的数据属性定义及属性约束（见表5）。数据属性中的xsd:string、xsd:dateTime、rdfs:Literal分别表示字符串型、时间日期型、长文本类型；xsd:decimal、xsd:int分别表示任意精度的十进制数（可含小数）、固定范围的整数。

2.2.2　山西票号社会本体

社会本体主要关注金融机构在业务和管理活动中产生的各类社会关系[24]。在票号领域，社会本体主要关注票号及内部人员之间以及与外部其他社会主体之间的金融社会关系，其核心概念类主要涵盖人物类和机构类。票号静态本体的顶层概念中，人物类主要包括从

表4　山西票号静态本体关系定义

顶层概念关系	定义域	关系描述	值域
人际关联关系	人物类	父母、子女、配偶、同事	人物类
事件参与关系	事件类	涉及	人物类
	人物类	负责、参与	事件类
	机构类	启动、结束	事件类
隶属关系	人物类	任职	机构类
	机构类	上级、包含	机构类
	机构类	在……期间	时间类
	机构类	隶属于	地点类
实践关系	人物类	创办	机构类
	机构类	执行	制度类
	机构类	发布、接收	文件类
演化关系	机构类	前身	机构类
	事件类	发生	机构类
地理关系	人物类	出生、去世	地点类
	机构类	所属、设分号	地点类
	事件类	发生	地点类
时间关系	人物类	出生于、去世于	时间类
	机构类	成立、歇业	时间类
	事件类	开始、结束、发生	时间类

资本供给到日常运营的全链条票号内部角色。为了充分揭示票号各类社会关系，本研究进一步细分顶层人物类，设置外部人员、内部人员两个子类，内部人员子类，纳入静态本体中的财东、经理、信房工作人员、柜房工作人员、账房工作人员、中厅工作人员和学徒；外部人员子类，纳入关联人员，即票号与外界业务往来所涉及的社会主体包括官员、皇室成员、实业家等；同理，在票号机构类中增设除票号及票号演化机构以外的同级子类，将外部机构类社会主体（包括钱庄、商号、近代工商企业、政府等）纳入本体框架。

票号社会本体的对象属性重点关注票号各类社会主体在票号经营、业务往来过程中形成的各类社会关系。本研究参考已有社会关系定义[24]并以票号领域示例进行解释，山西票号社会关系定义与示例如表6所示。

2.2.3　山西票号动态本体

动态本体用于刻画金融知识的动态特征，包括状态、状态转换和过程[24]。山西票号动态本体是在金融动态本体基础上，结合票号自身特征构建的动态知识描述体系，聚焦票号业务运营、市场参与随时空演化的动态过程，用于精准刻画票号各类业务的运作流程、事件发展的演化轨迹，以及各类事件之间的逻辑关系。

票号业务运作往往伴随金融风险。风险是在金融活动中，由于各种不确定性因素导致金融资产价值波动或经济损失的可能性，这些风险潜藏在各类业务事件中，同时与特定票号制度相关。故对于票号动态本体的定义，除复用静态本体中与业务密切相关的人物类（foaf:Person）、机构类（org:Organization）、时间类（crm:TimeSpan）、地点类（crm:Place）、事件类（crm:Event）之外，本研究进一步定义了风险类（spr:Risk），对应设置风险名称、来源、类型等数据属性。参考金融学专业书籍中关于金融风险的分类定

表5　山西票号静态本体顶层概念类的属性定义及属性约束

顶层概念类	属性名称	属性约束	顶层概念类	属性名称	属性约束
人物类	姓名	xsd:string	机构类	机构名称	xsd:string
	性别	xsd:string		机构地址	xsd:string
	生年	xsd:string		机构创立日期	xsd:dateTime
	卒年	xsd:string		机构介绍	rdfs:Literal
	籍贯	xsd:string		原始资本	xsd:decimal
	官职	xsd:string		后期资本	xsd:decimal
地点类	省	xsd:string		分号数量	xsd:int
	市	xsd:string	时间类	年	xsd:string
	县	xsd:string		朝代	xsd:string
制度类	修订原因	rdfs:Literal		时间跨度	xsd:string
	适用范围	xsd:string	事件类	事件名称	xsd:string
	制度内容	rdfs:Literal		原因	xsd:string
文件类	所属机构	xsd:string		结果	xsd:string
	签署时间	xsd:dateTime		影响	xsd:string
	文件内容	rdfs:Literal		发生地	xsd:string

表6　山西票号社会本体关系定义与示例

社会关系	定义	子关系	示例
直接金融关系	以资本市场为基础，资金提供者与资金需求者之间直接进行金融交易的关系	投资	财东出资成立票号
		捐输	票号财东以票号或个人名义向清政府捐输银两，换取官衔功名
		银股	财东等出资者持有银股
		身股	经理、伙计等出力者持有身股
间接金融关系	资金提供者通过金融中介机构将资金提供给资金需求者	同业对交	票号与钱庄之间的银两互通
		汇兑结算	商号通过票号完成异地货款汇兑
		存款	商号将闲置资金存入票号
		资金借贷	票号向信誉良好的近代工商企业发放贷款
调控关系	政府及金融监管机构通过货币政策、财政政策等手段对金融市场和金融机构进行宏观调控	直接控制	清政府指定票号为汇兑机构，管控公款汇兑
		货币政策	清末银钱比价调整，票号汇兑业务随政策调整计价
监管关系	政府及金融监管机构对金融机构和金融市场进行监督管理	账目核对	清政府对票号汇兑公款的账目进行定期核查
		注册管理	清末清政府令官商银钱行号注册
雇用关系	机构内部的人事关系	员工雇用	票号通过学徒制招录年轻从业者并雇用为伙计
		经理聘用	财东通过考查聘用有经验的经理执掌票号
		铺保	票号雇用新员工需有同业商号或乡绅出具铺保
信用关系	对投资者而言，如果两个金融机构营业方式雷同，则存在着信用关系，即使此类机构本身并无其他关联	无	一个票号因信用问题产生的影响波及其他票号

义[29]，将风险类（spr:Risk）分为6个子类，定义如下。①市场风险：利率、汇率、股价等波动所导致的风险。如票号需跨地区进行银两兑换，若兑付地的银价暴跌，将导致汇兑实际价值减损。②信用风险：债务人或交易对手未能履行合同规定的义务或信用质量发生变化，影响金融产品的价值，从而给债权人或金融工具持有人带来损失的风险。如票号放贷后，因债务人破产而形成坏账。③操作风险：人为错误、技术缺陷或不利的外部事件造成损失的风险。如票号密押规则泄露，导致汇票被伪造。④流动性风险：因市场成交量不足或缺乏愿意交易的对手而导致未能按合理价格交易资产的风险。如清末大清银行的成立导致客户偏向于选择官方银行机构进行存贷款。⑤法律和政策风险：法律风险主要关注签署的各类合同等法律文件的有效性问题所带来的风险；政策风险则主要体现为政府宏观金融政策、货币制度调整带来的冲击。⑥突发事件风险：战争、自然灾害等突发事件带来的风险。如太平天国运动、义和团运动给票号分号经营带来风险。

动态本体重在揭示业务逻辑和事件演化，因此基于事理知识图谱对事件关系的定义[30]，对票号动态本体相关的事理关系进行定义（见表7）。票号动态本体可通过事理关系分析事件的演化与发展，如针对票号汇兑业务事件，可将业务流程分为几个子事件，各子事件之间又包含因果、顺承、并发等关系。

表7　基于事理关系的山西票号动态本体事件关系定义

事理关系	定义	票号相关事理关系描述示例
因果事理	某一事件导致某一事件发生	密押规则泄露导致汇票被伪造
条件事理	某一事件条件下另一事件发生	如果银价暴跌，那么汇兑价值缩水
反转事理	某一事件与另一事件形成对立	客户原本选择票号存款，但因大清银行成立转向官方机构
顺承事理	某一事件紧接着另一事件发生	贷款发放后即启动贷款回收
组成事理	某一事件是另一事件的组成部分	汇兑发起、汇票流通、兑付汇款，组成完整汇兑流程
并发事理	某一事件与另一事件同时发生	汇款人付款同时票号收到款项

笔者进一步对山西票号知识表示模型的本体层逻辑进行梳理。首先，票号静态本体是票号知识框架的基础，通过定义人物类、机构类、事件类、地点类、时间类、文件类、制度类建立类间基础关联，初步搭建了票号基本知识体系。其次，从社会本体角度，将票号外部相关主体纳入知识体系，构建了人物之间（如财东与经理）、机构之间（如票号与钱庄）、人物与机构之间（如客户与票号）的金融社会关系网络，丰富了票号本体对票号复杂社会关系的知识表达能力。最后，将动态本体的事理逻辑嵌入票号基本知识体系中，增强了票号本体的动态刻画能力：通过将票号业务流程与其他票号关联事件拆分为子事件，以事理关系为纽带，同时增设风险类，可以支持对事件关系的知

识挖掘与深层阐释。静态本体、社会本体、动态本体三者融合，实现了对票号知识的全局描述。山西票号知识表示本体层关联展示如图1所示。

2.3 山西票号知识表示模型可视化

本研究选择Protégé作为本体编辑工具，将前述定义的各实体类、子类及其相应的数据属性、对象属性录入，并进一步定义关系属性的必要约束，如票号的创办人、票号的成立时间等都具有唯一性；人物关系中的同事关系、票号之间的信用关系等具有对称性；地点的隶属关系等具有传递性等，以增强本体的语义表达能力，更好地支撑后续的知识发现任务。最后通过OntoGraf插件对构建的山西票号知识表示模型进行可视化呈现。

2.4 山西票号知识图谱构建

上述模型构建过程已完成传统金融领域的抽象概念关系梳理，提供了山西票号知识框架体系和推理逻辑。为了支持后续知识发现任务，山西票号知识表示还需要纳入不同本体层面的具体知识，构建以三元组组织、可规模存储、查询、推理的图数据库，即知识图谱。首先，以前述构建的基础语料库文本为基础，以大语言模型辅助和人工核查相结合的方式进行实体抽取。通过抽取实践比较，本研究选择结构化输出稳定、对强制输出约束响应好、支持长文本抽取的大语言模型豆包为工具，基于Few-shot少样本提示示例与“实体识别→关系判定→属性抽取→三元组生成”分步骤提示链提升字段识别的准确性。其次，将语料文本根据内容划分，以大模型抽取结果为参照，人工逐一核查，并对可能遗漏的低频实体、领域专属实体进行人工抽取补充，确保无关键信息缺失。最后，对同义不同表述的实体进行合并归一化，统一实体命名规范，明确实体唯一标识，如地点名称统一使用现代地名、人物统一使用姓名等，形成标准化、高准确率的最终实体库。按照上述步骤，本研究抽取得到137个人物、55个机构、76个地点、100个时间、14个制度、110个事件、34个文件、8个风险，合计534个实例。将实例录入Protégé，并采用Protégé自带的推理机进行自动化推理，结果表明本体整体逻辑一致，未发现明显的语义冲突。

Onto QA是定量评估本体质量的常用工具。本研究选择Onto QA本体评估工具对构建的山西票号知识表示模型进行评价：类的丰富度为0.943，表明山西票号知识表示模型中的类得到了有效利用；关系丰富度为0.571，表明在山西票号本体中父子类关系以外的关系数占比高于父子类关系数，本体类间关系较为丰富；类实例分布为7.628个，说明该本体中每个类平均包含的实例数目较多，包含了较为丰富的票号知识。知识表示模型抽取的实体覆盖核心概念类，囊括核心代表性票号。尤其是地处平遥的22家票号实例最为丰富，其发展跨越票号发展史各个阶段，业务领域全面，包含票号

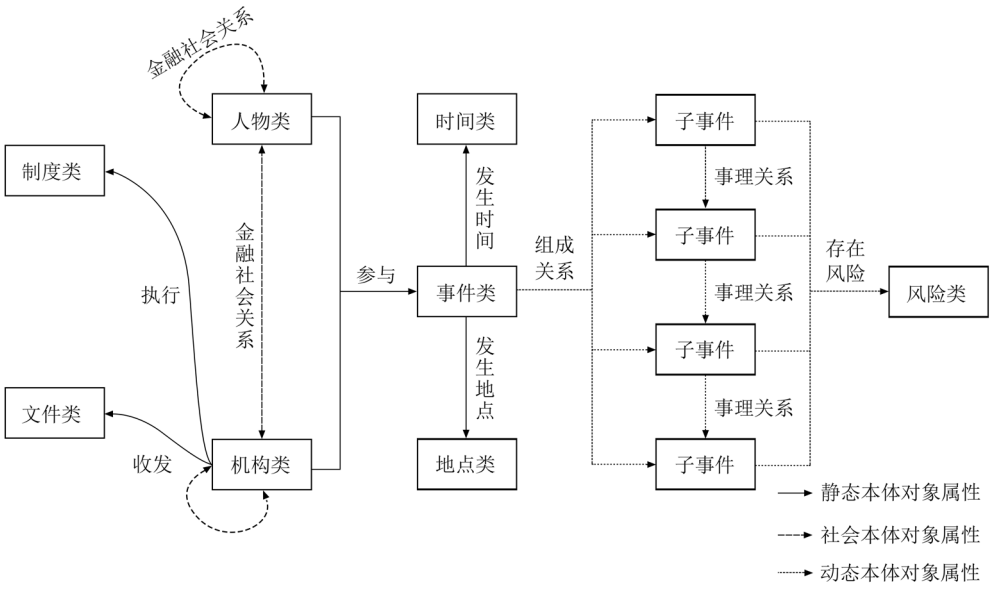

图1 山西票号知识表示本体层关联展示

创立、经营、衰亡全生命周期的关键场景，能充分体现本研究票号知识表示模型中票号组织机构、业务逻辑和制度体系等核心内容。

本研究选用Neo4j图数据库，将构建的山西票号知识表示模型及实例映射为图数据库中的节点、关系和属性。其中，Neo4j中的节点主要对应本体中的类及实例；关系（边）对应本体中的对象属性；属性对应本体中的数据属性，完成山西票号知识图谱构建。

3　山西票号知识发现

3.1　基于静态本体的票号知识全关联

本研究基于山西票号知识图谱的静态本体逻辑，使用Neo4j提供的Cypher语言进行可视化检索，实现票号基本知识的体系化。以检索式“MATCH (n:`ns0_总号`{ns0_机构名称:'日昇昌票号'})-[m]-(l) RETURN n,m,l”查询，得到较为完整的日昇昌票号知识体系（见图2）：日昇昌票号成立于1823年，歇业于1923年；核心人物有经理雷履泰、协理毛鸿翔、襄理程大培，以及财东李大全、李箴视、李五典等，相关人物关系清晰，如雷履泰、毛鸿翔与程大培是同事关系，李大全、李箴视、李五典是亲属关系；且直观展示了日昇昌遍布北京、重庆、香港、上海、天津的庞大分号网络，以及票号独创的铺保制、学徒制、股份制等制度体系，点击具体节点可以同步了解相关实体对象的知识描述。可见，该图谱较全面呈现了日昇昌票号的运营网络、制度逻辑和历史脉络，可以有效支撑对票号这一近代金融机构的全面了解和认识。

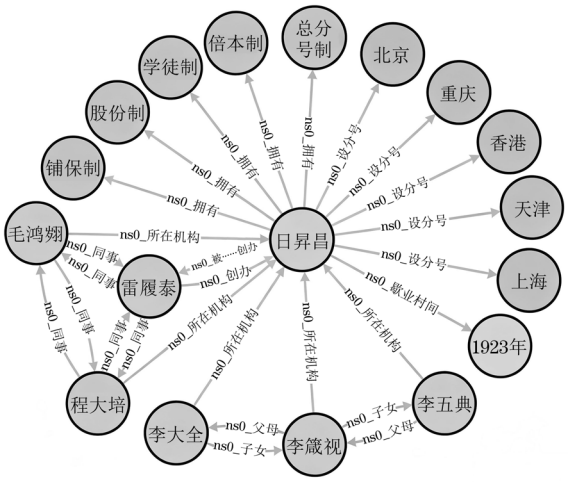

图2　日昇昌票号知识体系

3.2　基于社会本体的票号治理制度梳理

本研究基于社会本体，以山西票号独特的身股制度为例进行票号治理制度挖掘。身股制是山西票号以人身入股的特别制度，指在总股本中，财东为主要出资人，以银股享有分红权，并对票号亏损、债务负无限连带责任；掌柜、伙计虽然没有出资，但可以凭借自身的资历以及在山西票号中的功劳作为身股入股，与财东一起参与票号的分红；已顶身股的掌柜、伙计去世后，票号还会视他们生前贡献大小及所顶身股数量，在一定时间内给其家属发放其生前享受的分红，称故身股。

语义网规则语言（Semantic Web Rule Language，SWRL）可用于表示和执行语义知识推理。根据身股制的定义，设计SWRL语言推理规则：财东只持有银股，掌柜、伙计只持有身股，持有银股者和身股者均参与分红，已故掌柜、已故伙计的家属持有故身股者亦可参与分红。并输入HermiT本体推理机执行推理，可以实现对身股制关联实体的知识发现。

以大德通票号经营后期银股、身股持有情况为例，在山西票号知识图谱中进行相关制度查询，可以剖析票号控股的复杂关系。使用Cypher语句检索式“MATCH o=()-[r:`ns0_银股`]->(m:`ns0_总号`{机构名称:'大德通票号'})MATCH p=()-[s:`ns0_身股`]->(n:`ns0_总号`{机构名称:'大德通票号'})MATCH q=()-[t:`ns0_故身股`]->(l:`ns0_总号`{机构名称:'大德通票号'})RETURN o,p,q”查询，可以得到大德通票号经营后期银股、身股、故身股持有情况。从图3可以发现，大德通票号的大量人员持有身股。结合身股制定义可知，票号的盈亏主要由财东负责，身股持有者不承担赔偿义务但参与分红。随着票号业务发展，财东银股不变，而新的身股不断加入，身股不断累积，如果身股占比超过银股，会导致票号责任利益关系严重失衡，印证了原始史料中“亏折之祸独归财东，利害不均，号基渐危”[1]的论断。

3.3　基于动态本体的票号业务流程梳理与风险识别

山西票号的核心业务流程蕴含复杂的内在逻辑和动态演化特征，梳理票号业务流程，对于全面了解票号经营管理的理念与方法至关重要。知识图谱的路径检索功能可通过逐层遍历节点间的关联链路，深入挖掘并呈现节点之间的潜在联系。

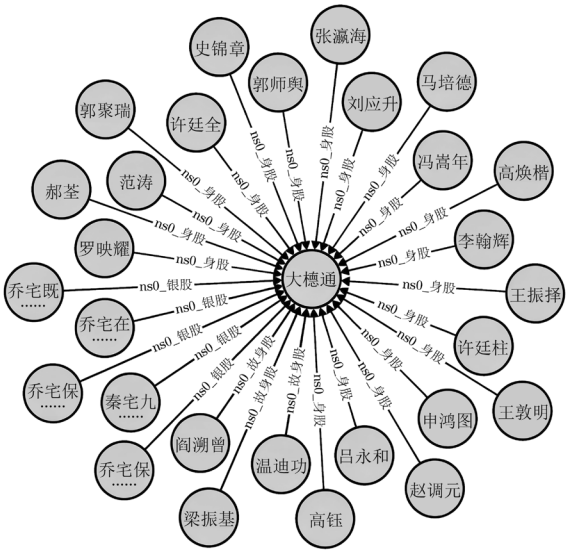

图3　大德通票号银股、身股持有情况

为系统呈现票号业务的运作机理、演化路径及风险传导机制，以道光二十四年姚宅与蔚泰厚票号之间进行的汇兑业务为例，使用Cypher语言检索式“MATCH path=(start:ns0_业务受理机构{机构名称:'蔚泰厚北京分号'})-[*]-(end:ns0_操作风险{风险名称:'失落票券'}) RETURN path”，以业务受理机构为出发点，以业务风险为终点，进行路径查询，可以直观呈现姚宅与蔚泰厚票号汇兑业务流程（见图4）。

本研究围绕道光二十四年的姚宅与蔚泰厚票号汇兑业务进行知识链路挖掘（见图5），该业务中包含层层递进的4个子事件[3]：首先“蔚泰厚北京分号收关批足纹银四百两又伊足纹银一百二十两”，再顺承至“两部分银两各立会票一张”，接着推进至“送票人收会票后前往汇兑”，最后顺承至“据年貌姓氏单盘问送票人后照票交付款项”。

在票号风险管理中，针对失落票券、仿制汇票等导致款项冒领的操作性风险事件，票号采取票据秘押制措施，包括专用统一印制的汇票、专人书写、字体备案、加盖多项印鉴、汇票暗号、汇票金额、取款人外貌清单等，确保款项交付妥当。本研究以失落票券为例，围绕该操作性风险进行挖掘，制定SWRL规则语言进行推理，建立制度与风险事件的应对关联，得到失落票券风险的防控措施为“蔚泰厚苏州分号据年貌姓氏单盘问送票人后照票交付款项”。

以检索式“MATCH p=()-[r:`ns1_存在风险`]->() RETURN p”查询票号图谱，检索其他风险事件，可以发现票号经营业务中存在的主要风险（见图6）。

依据史料内容解读，结合检索结果，可以发现：针对流动性风险，票号采取倍本制，在山西票号账期分红时，财东将一部分红利留在票号内用于周转，以增强流动资本的韧性，这与现代金融业构建的多元流动性储备体系异曲同工；针对信用风险，山西票号通过担保制抵御信用风险，对于信誉情况不明的商号，要求其寻找信誉达标的商号出面介绍，到期不履约则由介绍商号承担经济损失，使票号在无法收回贷额时，减少了不良资产的持有，为现代金融担保制度的建立提供了宝贵的历史经验；针对市场风险，票号对各地银两平码的掌握等

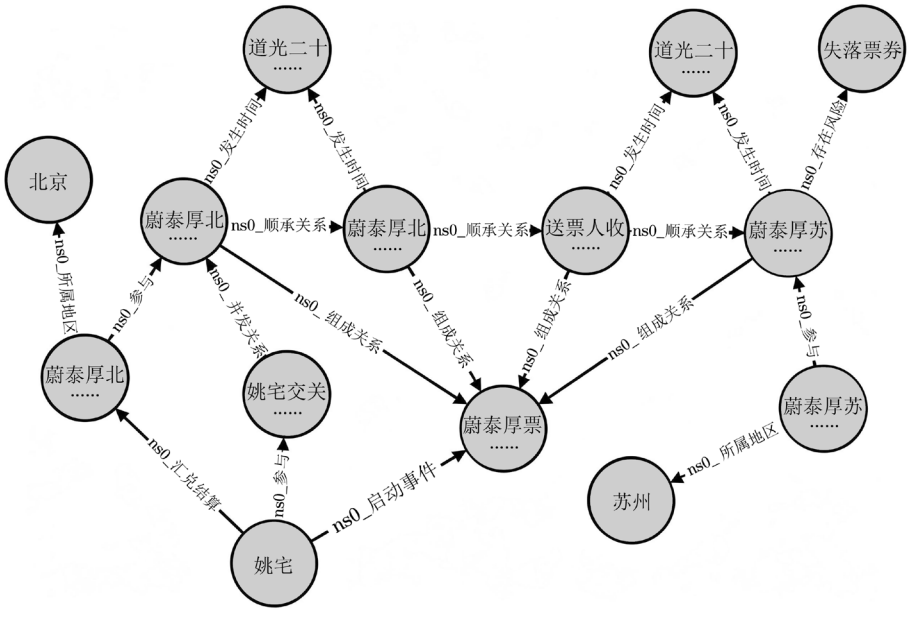

图4　姚宅与蔚泰厚票号汇兑业务流程

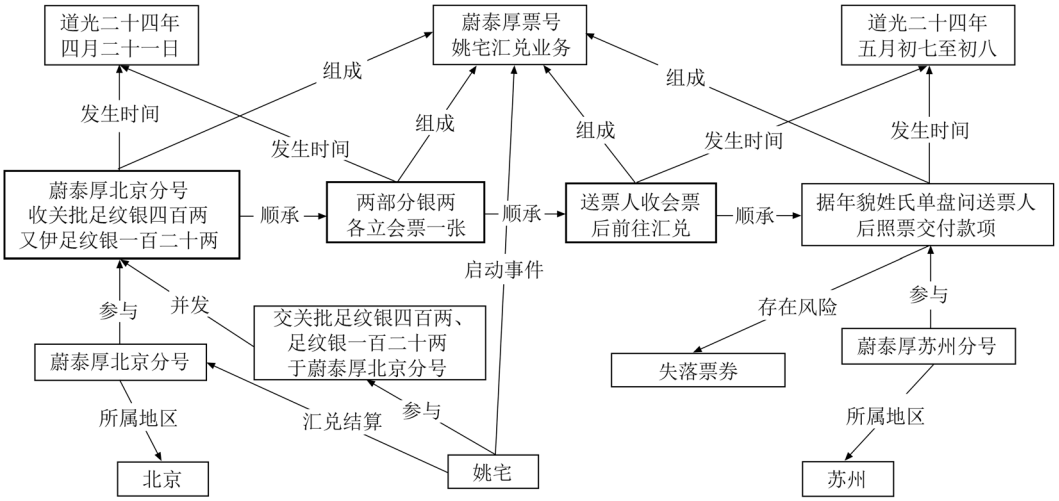

图5　票号汇兑业务知识链路

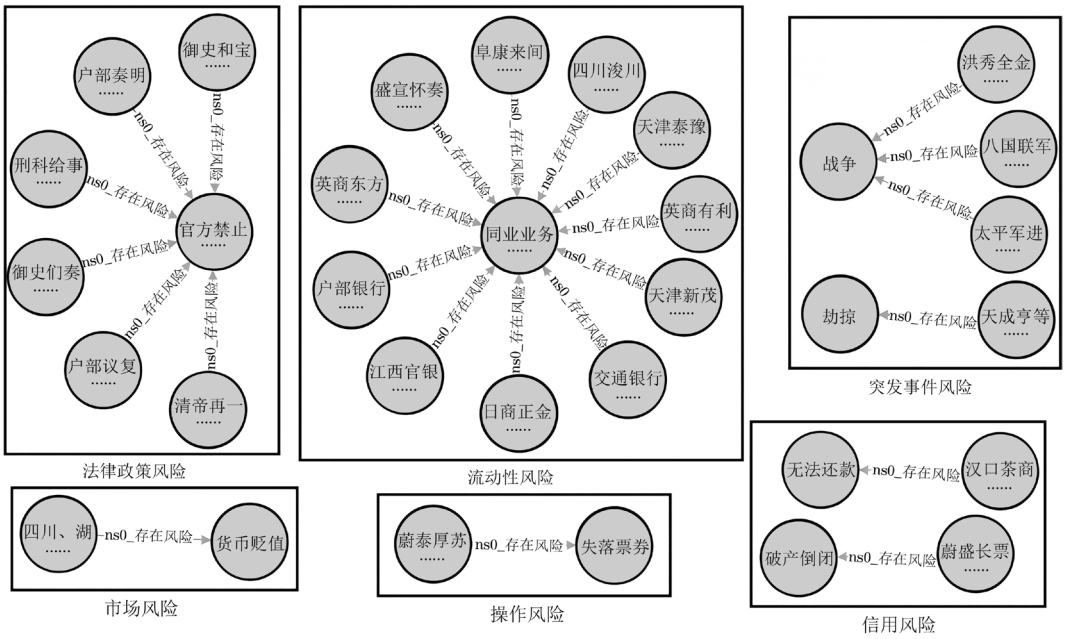

图6　票号经营风险识别

市场风险防控逻辑为“抽疲转快，酌盈剂虚”[1,3]，为现代金融业构建动态风险应对机制、强化市场预判能力提供了重要参照；针对法律政策风险和突发事件风险，票号并没有构建良好的应对机制。

4　结语

本研究融合金融本体理论与知识表示技术，针对当前票号研究以定性分析和文献整理为主，缺乏应用数字技术进行知识关联挖掘和体系构建的问题，构建了覆盖静态概念、社会关系、动态演化的山西票号知识表示模型，为票号文化的知识关联和系统阐释、数字化保护和有效开发利用提供了较为完整的知识框架，研究从业务流程、制度佐证和风险识别等不同角度进行的知识推理，展示了本模型对票号价值发现的能力和潜力。

本研究还存在一定不足：研究重在呈现票号这一传统金融领域的关键知识，以挖掘票号作为近代金融机构的核心价值为主，并未纳入政治、经济、文旅等相关领域知识，因而限制了本研究所提出模型的跨领域应用；在知识发现部分，仅依据史料内容逻辑进行推理规则设计，在全面性、可靠性方面仍有待进一步检验。

未来，本研究将进一步扩大史料收集范围，强化关联知识建模，并深入探索本体与机器学习的融合应用，

提升知识图谱构建、知识发现的自动化水平和推理能力，并尝试将票号知识体系与晋商文化、近代经济史等领域关联，以拓展本模型的应用领域。

参考文献

[1] 黄鉴晖. 山西票号史料[M]. 太原：山西经济出版社，2002.

[2] 李锦彰. 晋商老账[M]. 北京：中华书局，2001.

[3] 卫聚贤. 山西票号史[M]. 太原：三晋出版社，2017.

[4] 何庄. 晋商票号文书档案及其管理初探[J]. 档案学通讯，2019（6）：94–99.

[5] 周亚，李善靖. 论票号账簿的归户、特征与价值：以恒隆光票号账簿身份考证为例[J]. 中国经济史研究，2022（4）：174–189.

[6] 赵晶. 文书档案的文体特征与文化意蕴[J]. 山西档案，2025（1）：106–109.

[7] 程明洋，丁传标，陶伟. 平遥票号传统建筑空间研究：基于空间句法的分析[J]. 地域研究与开发，2017，36（2）：175–180.

[8] 朱向东，刘霞. 晋商票号金库设计理念分析：以山西平遥协同庆为例[J]. 山西档案，2016（4）：33–35.

[9] 陈凌，王萌，楼静波. 家族企业委托代理关系研究：山西票号与中世纪意大利家族企业的分析与比较[J]. 山东大学学报（哲学社会科学版），2009（6）：37–46.

[10] 吕长全. 晋商票号东掌关系再考察[J]. 会计之友，2023（21）：150–155.

[11] 丁丽芳，刘建军. 晋商票号人力资源管理特色分析[J]. 人力资源管理，2012（2）：72–74.

[12] 成瑾，郑仪. “尊贤型”高绩效人力资源系统研究：以山西票号为例[J]. 管理案例研究与评论，2013，6（5）：341–356.

[13] 郝平，李善靖. 清中后期山西票号合伙经营的多元模式及其影响[J]. 清史研究，2024（5）：110–119.

[14] 王治胜，周亚. 分号经营，总号核算：山西票号资产负债表的逻辑结构与呈现[J]. 河北经贸大学学报，2025，46（3）：98–108.

[15] 孔祥毅，张亚兰. 山西票号的风险控制及其现实意义[J]. 金融研究，2005（4）：1–11.

[16] 张宏彦，寇建轶. 山西票号兴衰与风险管控[J]. 中国金融，2017（22）：98–99.

[17] 付远方，方朝阳. 基于本体的革命历史事件叙事化开发模型构建与实践研究[J]. 图书情报工作，2025，69（24）：65–80.

[18] 李旭晖，吴燕秋，王晓光. 基于角色关联的叙事型文化遗产知识表示方法[J]. 图书情报工作，2017，61（9）：116–122.

[19] 闻心玥，许浩，吴丹. 线性文化遗产双语本体构建研究：以嘉兴运河文化为例[J]. 图书情报工作，2023，67（7）：107–120.

[20] WEI T，CHEN Y. A methodology for building domain ontology of cultural heritage[J]. Digital Scholarship in the Humanities，2023，38（4）：1710–1719.

[21] 秦春秀，杨智娟，赵捧未，等. 面向科技文献知识表示的知识元本体模型[J]. 图书情报工作，2018，62（3）：94–103.

[22] 何琳，吕星月，蒋畅畅，等. 基于LRMoo模型的红色书信资源共享本体构建及应用研究[J]. 图书情报工作，2026，70（2）：15–26.

[23] BARROS T H B，DE SOUSA R T B. Archival science and knowledge organization: mapping methodological relationships[J]. Knowledge Organization，2019，46（7）：493–501.

[24] 唐旭丽，马费成，傅维刚，等. 知识关联视角下的金融知识表示及风险识别[J]. 情报学报，2019，38（3）：286–298.

[25] 高峰，郑丽丽，顾进广. 面向多元时序关系的金融知识图谱表示与构建[J]. 山西大学学报（自然科学版），2022，45（4）：873–883.

[26] 刘政昊. 基于知识关联的多层本体立方体设计与实现：以金融证券领域为例[J]. 现代情报，2022，42（1）：72–86.

[27] 强韶华，罗云鹿，李玉鹏，等. 基于RBR和CBR的金融事件本体推理研究[J]. 数据分析与知识发现，2019，3（8）：94–104.

[28] 李景，孟连生. 构建知识本体方法体系的比较研究[J]. 现代图书情报技术，2004（7）：17–22.

[29] 徐望. 金融风险管理知识与应用：基于MATLAB编程[M]. 上海：上海财经大学出版社，2019.

[30] 刘政昊，曾曦，张志剑. 面向应急管理的金融突发事件事理知识图谱构建与分析研究[J]. 信息资源管理学报，2022，12（3）：137–151.

作者简介

毋江波，女，博士，副教授，研究方向：知识组织与数字人文。

朱鼎弘，男，硕士研究生，通信作者，研究方向：知识组织与数字人文，E-mail：1352978998@qq.com。

Research on Knowledge Representation Model Construction and Knowledge Discovery of Shanxi Piaohao

WU JiangBo ZHU DingHong
(School of Economics and Management of Shanxi University, Taiyuan 030031, P. R. China)

Abstract: The establishment of Shanxi Piaohao marked the beginning of modern Chinese financial industry. Its business models and management systems still offer valuable insights for the contemporary financial sector. Integrating scattered knowledge resources of Piaohao and constructing a systematic knowledge representation model facilitates the digital preservation, in-depth exploration and cross-domain application of Piaohao cultural heritage. This study first deeply integrates the financial ontology framework with the traditional financial knowledge of Shanxi Piaohao through the seven-step ontology construction method and builds a three-dimensional knowledge representation model of Shanxi Piaohao composed of a static ontology, a social ontology, and a dynamic ontology, covering core contents such as the organizational structure, business logic, and institutional system. Secondly, using the Protégé ontology editor and Neo4j graph database tools, builds a Piaohao knowledge graph. Finally, knowledge visualization and knowledge discovery have been achieved. This study provides a relatively comprehensive knowledge framework for the knowledge association, systematic interpretation, digital preservation, and effective development and utilization of Piaohao culture, the knowledge reasoning conducted from different perspectives, such as business processes, institutional evidence, and risk identification, demonstrates the potential of this model in discovering the value of Piaohao.

Keywords: Financial Ontology; Shanxi Piaohao; Knowledge Representation; Knowledge Discovery