

·业务研究·

基于司法判决书的知识图谱构建与知识服务应用分析

黄茜茜^{1,2}, 杨建林^{1,2}

(1. 南京大学 信息管理学院, 江苏 南京 210023; 2. 江苏省数据工程与知识服务重点实验室, 江苏 南京 210023)

摘要:【目的/意义】构建基于司法判决书的案件知识图谱是对司法数字资源的有效利用,有助于提升司法智能化水平,积极响应国家“智慧法院”建设发展战略。【方法/过程】以“网络诈骗”领域为例,用“自顶向下”的方式构建知识图谱。首先,结合文书内容与专家意见构建案件领域本体;接着,通过知识抽取、知识表示、知识融合等环节获取实体、属性及关系;再利用Neo4j生成案件知识图谱。最后,提出了基于知识图谱的智慧司法知识服务框架。【结果/结论】基于2015年-2020年的“网络诈骗”领域司法判决书,构建了含有约3万个实体和18万条关系的案件知识图谱,并详细阐述了具备基础资源层、知识图谱层、服务应用层的智慧司法知识服务框架设计。【创新/局限】实现了案件知识图谱的实体类型扩充,以丰富图谱应用场景,并将知识图谱技术与智慧司法知识服务框架进行融合;局限于仅使用网络诈骗领域判决书数据进行实证研究。

关键词: 知识图谱; 知识服务; 司法判决书; Bert-BiLSTM-CRF; Neo4j

中图分类号: G250.2; D926.2

DOI: 10.13833/j.issn.1007-7634.2022.02.018

1 引言

法院信息化建设是我国贯彻实施网络强国战略的重大举措之一^[1]。2016年7月,由中共中央办公厅、国务院办公厅印发的《国家信息化发展战略纲要》首次将“智慧法院”建设列入国家信息化发展战略^[2],旨在推动人工智能、大数据等新兴技术同司法工作的深度融合,提升司法活动各环节信息化水平,促进司法信息公开与公平正义^[3]。近年,国家已颁布多项法规政策助力智慧法院建设,强调对案件数字资源的有效利用。目前,已有中国裁判文书网、无讼案例网等平台向公众开放司法判决书,文书包含案情描述和审理结果等信息,可辅助司法工作人员全面了解案件情况。而冗长复杂的文书内容增加了批量处理和分析案件的难度,因此如何自动从判决书中提取、整合和存储所需信息是全面实现法院信息化需首要攻克难点。

当前,以图模型描述对象及关系的知识图谱技术正在多个行业领域内得以广泛应用。知识图谱是一种智能、高效的知识组织方式,能从海量数据中抽取结构化知识,具备广阔的应用情景^[4]。因此,本文选择利用深度学习等技术,从司法判决书中提取案件实体、属性及实体间关系,构建面向特定犯罪领域的案件知识图谱,以助力智能司法应用发展。

2 相关研究

知识图谱根据其覆盖范围可划分为通用知识图谱和领

域知识图谱两大类,通用知识图谱主要强调知识的广度,包含大量常识性知识;而领域知识图谱面向不同的领域,其数据模式和应用需求均各不相同,因此并未有一套通用的构建规范标准,常需通过特定行业工程师与业务专家来协作完成^[5]。基于司法判决书的案件知识图谱属于一类关注犯罪案件信息和关系的领域知识图谱,是有效提取和组织海量案件构成要素的利器,有助于切实提升司法服务智能性。目前,国内外学者围绕案件知识图谱的构建和应用已展开不少研究。

案件知识图谱除了包括涉案的人、物、地、事等基本实体外,学者们会根据各自研究的目的和侧重点,选取不同的数据源针对不同方向开展实体类型及实体间关系的扩充:

(1)有些学者针对某一特定犯罪类型,构建具有案件特色的知识图谱,强调实体及关系的全面性和多样性。例如杨阳^[6]融合公安、工商、银行等多库情报资源信息,搭建涉众型经济犯罪知识图谱,着重描绘案件、罪犯、账户三者之间的社交、资金流向等各类关系,为警方打击此类犯罪提供决策支持。Pandey R^[7]针对凶杀案调查年表展开研究,从中提取证人、嫌疑人、侦探、证据四大类型的实体,构建凶杀案件知识图谱,可辅助调查人员对是否能侦破案件做出合理判断。洪文兴^[8]等基于司法判决书构建了针对交通事故的案件知识图谱,包含罪犯、车型、责任认定等实体,及驾驶、搭乘等实体关系,旨在为交通事故领域的司法人工智能应用提供语义支撑。

(2)有些学者则针对节点的数据类型进行拓展,强调多模态数据的语义融合。例如Srinivasa K^[9]等基于在线报刊网

收稿日期: 2021-03-23

作者简介: 黄茜茜(1998-),女,浙江温州人,硕士研究生,主要从事数据挖掘、信息检索研究;杨建林(1970-),男,江苏泰州人,博士,教授,博士生导师,主要从事数据挖掘、信息检索、学术评价研究。

站上的案件报导构建案件知识图谱,除了包含常见的时间、地点、人物等实体节点外,还纳入了图片类型实体,丰富案件信息的表现形式。Wu H^[10]利用物联网技术融合案件电子数据和纸质报告、光盘等非电子数据,再依托自然语言处理、图像识别等人工智能技术构建含有文本、音频、视频等多模态数据的案件知识图谱。

目前,案件知识图谱已成为诸多智能司法应用场景实现的重要基础,较为典型的有罪犯分析和案件分析两大用途:

(1)应用于罪犯分析,着重关注如何最大化利用案件知识图谱中对罪犯信息提炼后的精简描述。例如王杰^[11]等基于案件知识图谱构建用户画像以全面描绘罪犯信息,协助智能警情分析,方便精准查人找物。Qazi N^[12]则开发了基于多层次关联分析方案的罪犯匹配系统,该系统可利用案件知识图谱从已结案件的罪犯中匹配可能参与某未破案件的嫌疑人,以辅助警方侦查破案。

(2)应用于案件分析,要点在于如何利用案件知识图谱的结构特点做更深层次的分析工作。例如赵洪^[13]等对案件知识图谱进行了扩展延伸,构建了更细粒度、更专业的应用图谱,如重大案件关联图谱、违法行为相似地域图谱等,旨在提升办案过程的智能化水平和工作效能。张蓝^[14]开发了一个嵌入案件知识图谱的案件可视化分析系统,并引入相关算法实现对图谱内节点的影响力分析、社区发现分类等,以期为公安机关定位潜在的侦查方向。Elezaj O^[15]则设计了一个基于案件知识图谱的智能查案分析框架,以协助犯罪调查人员开展作案特征挖掘、案件自动分类、犯罪网络可视化等工作。

综上所述,目前国内外关于案件知识图谱的自动构建研究还较少融合案件审判部分所含的法规法条、量刑调整因素、判定结果等类型实体,而这些实体的加入可丰富图谱的使用场景,深入赋能智慧司法工作。并且,目前也较少有学者对案件知识图谱的应用场景整合展开讨论。因此,本文以司法判决书为数据源,完善扩充可提取的与案件信息相关的实体,构建面向特定犯罪领域的案件知识图谱,并提出基于知识图谱的智慧司法知识服务框架,以期如何利用知识图谱技术挖掘司法大数据价值、提升司法智能化水平提供方法基础和参考依据。

3 案件知识图谱构建

本文以“网络诈骗”类型案件为例,构建基于司法判决书的案件知识图谱。通过Python编写网页爬虫工具,自动采集无讼案例网上2015年至2020年的“网络诈骗”类型案件一审判决书。对所获取的网页内容进行数据清洗,去除正文内容为空或结构不完整的判决书实例,再将去重和规范化处理后的数据存入MySQL数据库中,最终共计有4493份判决书。

知识图谱主要有自顶向下和自底向上两种构建方式,其中自顶向下的方式适用于具备固定知识体系或可定义固定模式数据的特定行业^[16]。由于判决书具备规范的格式和组

成要素,故本文选择采用自顶向下的方式构建案件知识图谱,构建流程如图1所示:首先,构建案件领域本体,以为后续实体加入知识库做准备;接着,基于本体从司法判决书中抽取知识,包括实体抽取、属性抽取和关系抽取;然后,再以JSON格式表示所抽取知识,并通过引入第三方知识库完成知识融合工作,提升知识质量;最后,利用Neo4j构建知识图谱。

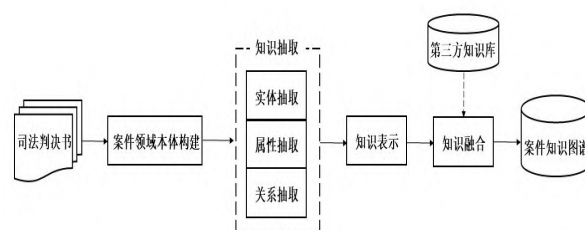


图1 基于司法判决书的案件知识图谱构建流程

Figure 1 The construction process of case knowledge graph based on judicial decision documents

3.1 案件领域本体构建

案件领域本体旨在结构化表述案件的作案信息和审理过程。本文基于“七步法”^[17],结合现有案件领域本体与司法判决书基本结构,再辅以高校、企业内的司法领域专家指导,构建案件领域本体知识模型,为后文面向网络诈骗领域的案件知识图谱构建提供逻辑支撑。图2展示了该本体的RDF描述,详述如下:

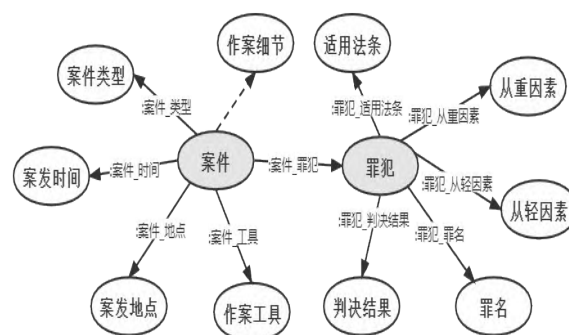


图2 案件本体RDF图

Figure 2 The RDF diagram of case ontology

(1)实体类型。案件是案件领域本体的核心,其设有作案细节这一数据属性,此外还包含案件类型、案发时间、案发地点、作案工具、罪犯五类案情描述实体;适用法条、从重因素、从轻因素、罪名、判决结果五类判定信息实体。

(2)实体关系。除定义上述实体外,本文还定义了10条对象属性来描述网络诈骗案件领域本体中实体间的关系。其中,以案件实体为出发点的关系主要描述案件基本信息,以罪犯实体为出发点的关系主要描述某罪犯在某案件中所获判定信息。

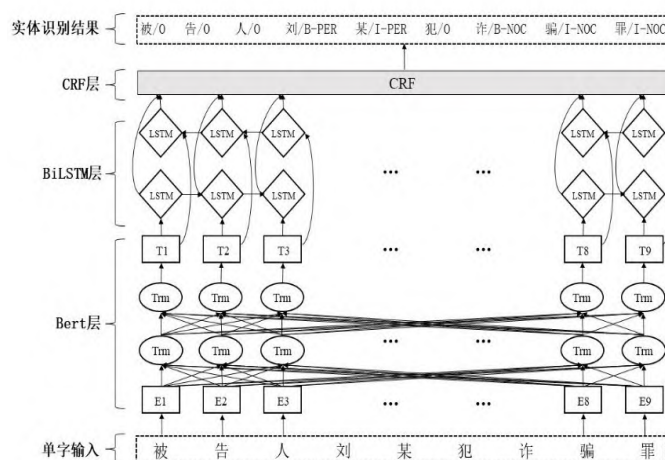


图3 基于Bert-BiLSTM-CRF模型的实体识别流程

Figure 3 The entity recognition process based on Bert-Bilstm-CRF model

3.2 知识抽取

3.2.1 实体抽取

经过预处理后的司法判决书分为半结构化文本和非结构化文本两部分。针对半结构化文本部分,本文采用基于规则的方法进行实体抽取,由此获取案发时间、案发地点两类实体;针对非结构化文本,本文通过 Bert-BiLSTM-CRF 模型进行命名实体识别,实现过程如下所述:

首先,采用 BIO 标注法对实体进行人工标注,其中作案细节触发词、作案工具、适用法条、罪犯、从重因素、从轻因素、罪名、判决结果等实体分别用{MTD,IST,LAW,PER,HVP,LTP,NOC,ROJ}进行标记;各实体文本的开始单字、其余单字分别用{B-,I-}作为标记前缀;文本中其余无关单字则用{O}标记。

标注完成后,以 6:1:1 的比例将所有单句划分成训练集、验证集、测试集,分别共计有 111610 条、19688 条、19018 条句子。接着,引入 Bert-BiLSTM-CRF 模型完成命名实体识别任务,已知该模型是中文命名实体识别领域目前较为主流的训练模型^[18],图 3 展示了该模型的基本实现原理:首先,将待标注语句输入 Bert 模型中,得到输入向量 $E_i (i = 1, 2, \dots, n)$,然后通过多层 Transformer 对其进行特征提取,得到特征向量 $T_i (i = 1, 2, \dots, n)$;接着,将上层得到的特征向量输入 BiLSTM 模型中进一步训练,得到每个单字可能为各序列标签的概率值;最后,根据 CRF 层的约束规则获得每个单字的最大概率化序列标注,即为最终的实体识别结果。

实验模型中部分重要的超参数设置如下:将输入句子的最大长度 maxlength 设为 512;每轮实验的样本批量 batchsize 设为 50;最大迭代次数 maxepoch 设为 100;为避免过拟合,将 dropout 设为 0.5。实验性能的评价指标包括精确度 Precision、召回率 Recall 及 F1 值。

将上文得出的训练集和验证集投入训练模型,在迭代

74 轮次后,模型在测试集上取得了最佳表现。表 1 展示了各类实体的识别情况,其中适用法条、罪犯、罪名、判决结果四类实体的 F1 值表现较好,因为这些实体的词汇表达或在语句内的位置较为固定,如法条一般会以书名号作为词汇首尾;而作案细节、作案工具、从重因素、从轻因素四类实体的 F1 值表现稍微差一些,归因于这些实体的词汇表达较多样,如作案细节中表示诈骗理由的就有“以……为由”“用……的名目”等多种表达,进而提升了模型语义学习的困难度。

基于以上识别的“作案细节触发词”实体,提取并合并含有这类触发词的单句,将合并的结果作为该案件的作案细节,即案件这一实体的数据属性。

表 1 实体识别实验测试集最佳结果(%)

Table 1 Best result of entity identification experimental test set (%)

实体名称	Precision	Recall	F1
作案细节触发词	89.73	91.76	90.55
作案工具	85.84	86.20	86.02
适用法条	96.53	98.98	97.74
罪犯	98.27	99.45	98.85
从重因素	83.16	97.50	89.76
从轻因素	76.85	87.27	81.73
罪名	99.00	99.60	99.30
判决结果	98.47	98.69	98.58

3.2.2 “案件类型”实体生成

至此,除“案件类型”外的所有实体均已成功获取,而“案件类型”并不能直接在司法判决书中获取,需根据作案细节进行二次判断,为此本文构建了案件自动分类模型,具体实验流程如下:首先,随机挑选 3000 个案件的作案细节部分作为实验数据集,并以 8:2 的比例划分训练集与测试集;接着,结合作案细节描述和专业领域研究报告^[19-20],人工为每个案件标注犯罪类型,共分为冒充身份诈骗、金融诈骗、交友诈骗、兼职诈骗、网购诈骗、发布虚假信息诈骗、贩卖个人信息

诈骗、网络赌博诈骗、虚拟物品诈骗、其他诈骗等十种类型；然后，再采用 jieba 完成语句分词，通过 TF-IDF 算法进行特征提取，并分别投入 LinearSVC、MultinomialNB、RandomForestClassifier、LogisticRegression 四种分类器中自我学习，再将测试集依次投入训练好的模型中，得到各分类器在测试集上的表现（表 2），评价指标为加权平均精确度 Weighted-Precision、加权平均召回率 Weighted-Recall、F1 值；最终，选用表现最好的 LinearSVC 分类器完成其余案件的自动分类。

表2 自动分类实验测试集结果(%)

Table 2 Automatic classification experimental test set results (%)

分类器名称	Weighted-Precision	Weighted-Recall	F1
LinearSVC	92.53	90.95	91.74
MultinomialNB	89.43	77.06	82.78
RandomForestClassifier	84.04	76.37	80.02
LogisticRegression	88.71	87.16	87.93

3.2.3 关系抽取

由上文构建的本体可得,实体间关系主要可分为以“案件”为出发点和以“罪犯”为出发点两大类,其中:前一类关系提取较为简单,只要该实体出现在该案件的判决书中,则关系成立;而后一类关系采用规则匹配的方式从文本语句中提

取,以图4为示例,首先提取“诈骗罪”“判处有期徒刑四年”这两个实体所在的单句,再通过匹配该句中的罪犯类型实体“刘某”,则关系(刘某,罪犯_罪名,诈骗罪)和(刘某,罪犯_判决结果,判处有期徒刑四年)成立。

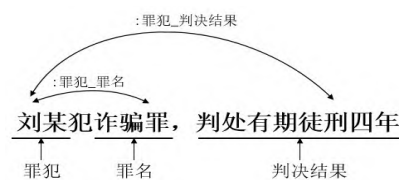


图4 以“罪犯”为出发点的关系提取示例

Figure 4 An example of relationship extraction with ‘criminals’ as the starting point

3.3 知识表示与融合

为了方便数据的读取,本文选择采用JSON语言来表示实体三元关系。进一步地,因实体识别结果仍存在歧义和冗余信息,故需展开知识融合工作,提升知识质量^[21]。以“案发地点”实体对齐为例,在司法判决书中,关于同一地点的语言表述常存在差异,所涉及的区域粒度也不尽相同。譬如“浙江省”,存在“浙江”“浙江省”等表达,或仅具备“温州市”“文成县”等下属地区名称。为了实现地点实体名称统一,本

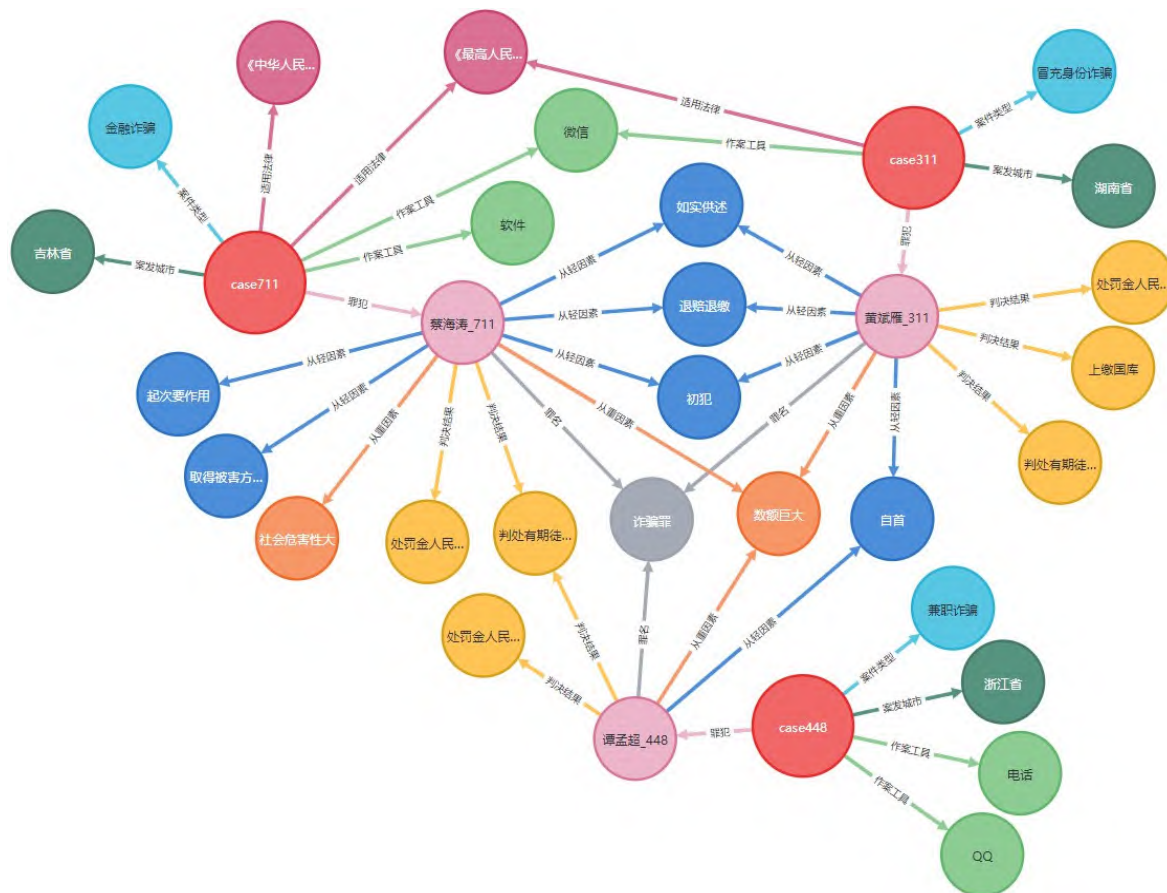


图5 基于司法判决书的网络诈骗案件知识图谱(部分)

Figure 5 Part of knowledge graph of internet fraud cases based on judicial decision documents

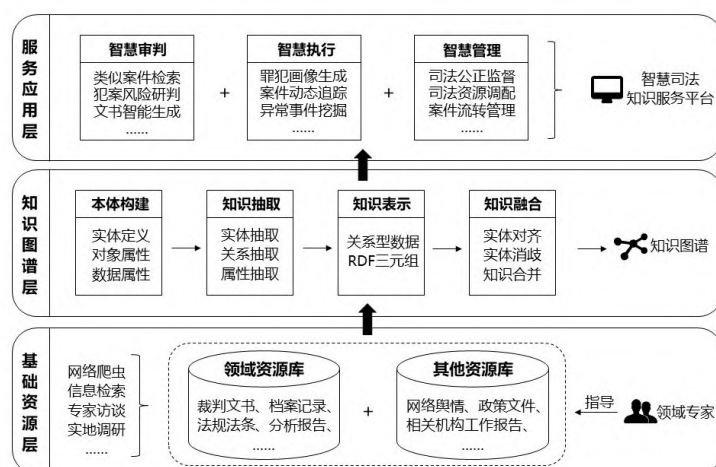


图6 基于知识图谱的智慧司法知识服务框架

Figure 6 Knowledge graph-based intelligent judicial knowledge service framework

文选取了中国省市县区域划分公开数据集^[22-23]作为第三方数据,再将所抽取实体与第三方数据库的实体链接,进而实现实体对齐。

3.4 知识图谱生成

本文利用Python的Py2neo模块,将网络诈骗案件相关实体及其关系从本地JSON文件存储到Neo4j数据库中,最终构建了一个含有约3万个实体和18万条关系的基于司法判决书的网络诈骗案件知识图谱。图5为部分案件知识图谱,展示了三起网络诈骗案件的相关信息。此外,还可在Neo4j可视化界面使用Cypher查询语言实现知识检索。

4 智慧司法知识服务框架设计

4.1 知识服务框架概述

传统的司法信息服务平台更关注如何为用户的标准信息需求提供服务,例如司法判决书、法律法规等专业文献资源的检索和获取,而针对更深层次的个性化、专业化智能司法知识需求,其满足程度仍存在较大提升空间。为此,本文以司法工作人员为服务对象,提出了基于知识图谱的智慧司法知识服务框架,主要包括基础资源层、知识图谱层、服务应用层三个层次(见图6)。

(1)基础资源层。基础资源层的主要功能为构建知识服务平台底层资源库,包括数据资源采集、整合与存储环节,旨在为司法领域知识图谱构建提供夯实的数据资源保障。为了准备更完善的数据对象,提供更全面的知识服务功能,可利用网络爬虫、专家访谈、实地调研等多种方式收集各类型数据,再在领域专家的指导下,构建存储案件、法规等司法领域专业数据的领域资源库和存储网络舆情、政策报告等补充数据的其他资源库。

(2)知识图谱层。知识图谱层可视作对基础资源层内各

类数据的二次整合,是知识服务框架中的关键层次。基础资源层中存储了各类结构化、半结构化、非结构化数据,数据类型的不统一为今后各种复杂分析和决策支持功能的实现带来不少难度。利用知识图谱技术,可以三元组的形式表示与存储从海量数据中提取的实体及关系,实现多源数据的相互融合与支持,为后续智慧司法知识服务应用提供基础,核心构建方法细节如以上章节所述。

(3)服务应用层。服务应用层是基于知识图谱层搭建的知识服务功能层,展现形式为集成多个功能板块的智慧司法知识服务平台。近年,最高人民法院连续通过《人民法院信息化五年发展规划(2019-2023)》《全面推进司法人工智能建设的总体思路》等多项指导文件,文中指出就“司法工作人员”这一服务对象而言,智慧法院建设应加快推进“智慧审判、智慧执行、智慧管理”三环节的实现。依照文件精神,本文分别为各环节设计了基于知识图谱的智慧司法知识服务应用模块,各模块的作用功能、应用场景示例及知识图谱技术嵌入使用方法将于下一小节展开详细阐述。

4.2 服务应用层模块介绍

(1)智慧审判模块。“智慧审判”模块致力于提升审判工作的自动化程度,可包含类似案件检索、犯案风险研判、文书智能生成等功能,本小节将以类似案件检索这一应用场景为例展开详细介绍。2017年8月,最高人民法院首次提出了类似案件检索机制,至今已密集出台多项政策法规推进实施,强调法院应秉持“类案类判”原则开展审判工作。司法审判人员可借助类案检索系统,输入待查案件相关的检索式,获取类似案件。依据类案检索结果,法院可对比同类生效案件审判尺度来评估待查案件审判结果合理性,增强审判公信力;此外,若遇见棘手难办案件,也可通过检索以往是否存在相关类似案件,为审判意见的合理给定寻求参考经验。

背靠司法大数据环境,类案检索系统的构建可依托知识图谱、自然语言处理等技术。首先,开发人员需利用自然语

言处理技术提取司法判决书中的各类情节要素及其携带属性,如作案情节、量刑情节、审判情节等案件要素,并判断各要素间的逻辑关系,从而将半结构化、非结构的文本内容转为结构化标签,再基于此结果构建案件知识图谱,以作为类案检索系统的资源库。接着,需要确定检索式的输入形式,例如用户可以选择手动输入待查案件的要素关键词,也可以选择将整篇判决书传入系统,由系统自动提取案件要素及关系。最后,引入知识相关性算法度量待查案件与知识图谱中各案件的语义相似程度,再以相似度大小对检索结果进行排序,实现对类似案件的精准匹配。应注意在设计案件要素实体相似度测量算法时,既要考虑概念自身的语义,又要结合概念的属性和上下文内容,如同样是自首认罪,行为发生时间的不同可能会导致量刑调整力度的差异,进而影响最终审判结果。

(2)智慧执行模块。“智慧执行”模块聚焦于为司法执行工作减负增效,可包含罪犯画像生成、案件动态追踪、异常事件挖掘等功能。其中,罪犯画像是智慧执行业务下较为典型的辅助工具,本节将以此为例展开介绍。被执行人规避执行一直是困扰法院执行工作的一大难题,在很大程度上阻碍了司法的高效发展。为了解决这类问题,法院常采取对被执行人在社会上的各类行动实施联合限制的方法,即若被执行人不按时执行审判结果,则将“寸步难行”。罪犯画像有助于司法工作人员准确把控罪犯的各项社会信息,进而实现精准限制罪犯各项行动。

若采用传统的用户画像构建技术来制作罪犯画像,大致过程为从司法判决书中识别提取罪犯的身份信息、履历信息、行踪信息等,再进行罪犯标签的识别、判断和归类。而由于数据源单一和文本挖掘技术的限制,常导致所构建的罪犯画像产生数据稀疏、粒度太粗等问题。于此,本文提出一种基于知识图谱的罪犯画像构建技术。首先,利用上述传统方法构建一个基于司法判决书的基础罪犯画像。接着,再从多方渠道搜集罪犯的相关信息作为画像补充数据,例如出行、社交、购物等平台的行为数据。然后,基于知识图谱技术将上步所得数据转化为罪犯信息知识图谱,图谱中的实体均可视作画像的备选标签。最后,可通过概念网络对最开始的基础罪犯画像进行标签扩展、标签联想以解决标签稀疏的不足。其中,关于标签扩展方面,可将罪犯信息知识图谱中的实体补充至每个罪犯相应的标签体系中,也可基于罪犯信息知识图谱计算两罪犯子图的相似性,若两罪犯的关联度较高,可考虑二者互相补充标签;关于标签联想方面,可利用知识图谱的推理联想机制延展个人标签,使得罪犯画像更为饱满,例如可根据罪犯消费记录推导出个人偏好等。

(3)智慧管理模块。“智慧管理”模块旨在为司法机关内外部管理提供新的解决方案,可包含司法公正监督、司法资源调配、案件流转管理等功能。以下将阐述司法公正监督这一功能的实现,作为知识图谱技术在智慧管理方向的运用示例。随着网络电子技术的不断发展,案件审理过程越来越公开透明,广大人民群众可通过各大信息流平台对司法办案各

环节进行监督。若对办案过程中某步骤有不明晰或不满意的地方,群众可在平台上发表意见或展开讨论,而这些讨论话题中往往蕴含着可能出现工作漏洞的信号,可辅助法院内部管理人员定位问题根源。因此,通过对案件相关的网络舆情文本展开深入挖掘,并与法院审判流程节点实时联动分析,寻找判决存在潜在争议的案件,是推进司法公正建设的有力助手。

为了支持案件审判态势和舆情状态的实时更新,本文提出一种以动态时序知识图谱为核心的司法公正监督系统构建方法设想。动态时序知识图谱是一种含有时间维度的知识图谱,图谱中的实体及实体间关系可以随着时间推移不断发生变化。具体构建流程如下:首先,利用自然语言处理技术完成多源数据融合工作,将记录案件审判过程的各类司法内部文书和与案件相关的各信息流平台舆情文本有效结合。接着,基于上述语料构建含有案件审判环节详情、舆情主题、观点态度等实体的知识图谱,并设有相关管理人员定期运维知识图谱的动态实时更新。最后,可嵌入时序分析技术、复杂网络分析技术等挖掘蕴藏在图谱中的异常点,再就问题寻根溯源,实现法院工作与众人意见的快速对齐,借助全民监督的力量倒逼法院规范行为,推动司法公正建设,达到智慧管理目的。

5 结 语

本文以“网络诈骗”领域案件知识图谱构建为实例,阐述了基于司法判决书的知识图谱构建过程,包括案件领域本体构建、知识抽取、知识表示、知识融合、知识图谱生成等环节,并提出了基于知识图谱的智慧司法知识服务框架,可视为对案件知识图谱应用场景的一种整合,以期提升司法工作智能化水平。

相比前人研究,本文在案件知识图谱的实体类型上做了完善扩充,纳入了除基本案情描述外的案件类型、适用法条、量刑调整因素、判决结果等实体,可丰富图谱的应用场景。同时,本文所设计的知识服务框架为如何利用知识图谱技术推动司法工作信息化建设提供了新思路。

此外,本文工作仍存在以下局限:首先,仅选择了“网络诈骗”这一领域的司法判决书作为研究数据,可适当扩充案件类型;其次,仅选择了司法判决书作为数据源,可进一步考虑融入舆情、政策等其他数据源。未来研究将进一步致力于智慧司法知识服务平台的实现与应用。

参考文献

- 1 陈甦,田禾,吕艳滨,胡昌明.中国法院信息化发展报告 No.4(2020)[M].北京:社会科学文献出版社,2020:2.
- 2 中共中央办公厅国务院办公厅印发《国家信息化发展战略纲要》[EB/OL]. [2021-03-18]. http://www.gov.cn/xinwen/2016-07/27/content_5095336.htm.
- 3 习近平出席中央政法工作会议并发表重要讲话[EB/OL].

- [2021-03-18].http://www.gov.cn/xinwen/2019-01/16/content_5358414.htm.
- 4 闫树,魏凯,洪万福.知识图谱技术与应用[M].北京:人民邮电出版社,2020:1-3.
- 5 王昊奋,漆桂林,陈华钧.知识图谱:方法、实践与应用[M].南京:电子工业出版社,2019:420.
- 6 杨阳.基于知识图谱的涉众型经济犯罪知识建模与情报分析[J].中国人民公安大学学报(自然科学版),2020,26(2):87-95.
- 7 Pandey R.Text Mining for Social Harm and Criminal Justice Applications[D].West Lafayette,United States:Purdue University,2020.
- 8 洪文兴,胡志强,翁洋,张恒,王竹,郭志新.面向司法案件的案情知识图谱自动构建[J].中文信息学报,2020,34(1):34-44.
- 9 SrinivasaK, Thilagam P S. Crime base: Towards building a knowledge base for crime entities and their relationships from online news papers[J].Information Processing & Management, 2019,55(6):102059.1-102059.19.
- 10 Wu H.Research on Electronic Evidence Management System Based on Knowledge Graph[C]//Communications in Computer and Information Science,2019:416-425.
- 11 王杰.基于公安知识图谱的用户画像研究[D].武汉:武汉理工大学,2019.
- 12 Qazi N, Wong B L W. Behavioural & Tempo-Spatial Knowledge Graph for Crime Matching through Graph Theory [C]//2017 European Intelligence and Security Informatics Conference (EISIC).IEEE Computer Society,2017.
- 13 赵洪,王芳,王晓宇,张维冲,杨京.基于大规模政府公文智能处理的知识发现及应用研究[J].情报学报,2018,37(8):805-812.
- 14 张蓝.基于知识图谱的网络传销案件可视化研究[D].北京:中国人民公安大学,2020.
- 15 Elezaj O,Yayilgan S Y,Kalemi E,et al.Towards Designing a Knowledge Graph-Based Framework for Investigating and Preventing Crime on Online Social Networks[C]//International Conference on e-Democracy. Cham, Switzerland: Springer,2019:181-195.
- 16 王昊奋,漆桂林,陈华钧.知识图谱:方法、实践与应用[M].南京:电子工业出版社,2019:425.
- 17 邓诗琦,洪亮.面向智能应用的领域本体构建研究——以反电信诈骗领域为例[J].数据分析与知识发现,2019,3(7):73-84.
- 18 吴俊,程焱,郝瀚,艾力亚尔·艾则孜,刘菲雪,苏亦坡.基于BERT嵌入BiLSTM-CRF模型的中文专业术语抽取研究[J].情报学报,2020,39(4):409-418.
- 19 司法大数据专题报告之网络犯罪特点和趋势[EB/OL]. [2021-03-18]. <http://www.court.gov.cn/fabu-xiangqing-202061.html>.
- 20 2019年网络诈骗趋势研究报告[EB/OL].[2021-03-18]. <http://zt.360.cn/1101061855.php?dtid=1101062366&did=610412125>.
- 21 杨波,廖怡茗.面向企业动态风险的知识图谱构建与应用研究[J].现代情报,2021,41(3):110-120.
- 22 中国行政区划及行政编码[EB/OL].[2021-03-18]. <https://github.com/luckydesigner/china>.
- 23 China Province City Json Data[EB/OL].[2021-03-18].<https://github.com/wu-jianyong/ChinaProvinceCityJsonData>.

(责任编辑:赵红颖)

Construction of Knowledge Graph and Analysis of Knowledge Service Application Based On Judicial Decision Documents

HUANG Xi-xi^{1,2}, YANG Jian-lin^{1,2}

(1. School of Information Management, Nanjing University, Nanjing 210023, China;

2. Jiangsu Key Laboratory of Data Engineering and Knowledge Service, Nanjing 210023, China)

Abstract:【Purpose/significance】The construction of case knowledge graph based on judicial decision documents is an effective use of judicial digital resources, which helps to improve the level of judicial intelligence and actively respond to the national "smart court" construction and development strategy.【Method/process】Take the field of cyber fraud as an example to build the knowledge graph in a top-down way. Firstly, the case domain ontology is constructed by combining the contents of documents and expert opinions. Then, entities, attributes and relationships are acquired through knowledge extraction, knowledge representation and knowledge fusion. Then Neo4j was used to generate knowledge graph of crime domain. Finally, an intelligent judicial knowledge service framework in crime domain based on knowledge graph is proposed.【Result/conclusion】Based on the judicial decision documents in the field of cyber fraud from 2015 to 2020, a knowledge graph containing about 30,000 entities and 180,000 relationships is constructed, and the framework design of intelligent judicial knowledge service with basic resource layer, knowledge graph layer and service application layer is elaborated in detail.【Innovation/limitation】The entity types of the case knowledge graph were expanded to enrich the application scenarios of the case knowledge graph, and the knowledge graph technology was integrated with the intelligent judicial knowledge service framework. The limitation lies in the empirical research only using the data of judgment in the field of network fraud.

Keywords: knowledge graph; knowledge service; judicial decision documents; Bert-BiLSTM-CRF; Neo4j

(上接第 132 页)

Studies on Influencing Factors Related to Students' Reading Literacy Based on HLM

——Evidence from PISA 2018 in China

HU Jie, WANG Mei-shu

(Department of Linguistics, Zhejiang University, Hangzhou 310030, China)

Abstract:【Purpose/significance】Reading literacy is the most essential skills in human development, therefore, to investigate what factors influence students' reading literacy will enhance students' reading ability and comprehensive development and provide an effective way to excellent reading education.【Method/process】This present study based on psychological theory of educational productivity, applying HLM to further explore the factors related to students' reading literacy. A total of 11,993 participants from 361 schools in China were extracted from PISA 2018 database.【Result/conclusion】The results indicated that students' aptitude, teachers' instruction as well as social-psychological environment all have a significant impact on students' reading literacy, and different factors have different effects on students' reading literacy.【Innovation/limitation】This current study provided a multidimensional perspective by probing into the dynamic development factors of students' reading literacy from student- and school-level, restoring the whole picture of the system of reading education, which is conducive to the quality and governance of reading education.

Keywords: reading literacy; affecting factors; psychological theory of educational productivity; hierarchical linear modeling; PISA2018