基于知识图谱的文学作品智能服务研究
——以《人世间》小说为例

谢政霖，钱　清，陈清容[通信作者]
（贵州财经大学信息学院，贵州　贵阳　550000）

摘要：构建当代文学知识图谱旨在深化对文学作品的理解和分析，同时为文学研究者提供一个高效的工具，帮助其从多维度和多层面解读和研究当代文学作品。该文以小说《人世间》为对象，采用自顶向下的方式构建了《人世间》小说的知识图谱，构建了包括关系检索系统和智能问答系统的《人世间》小说智能服务系统。本系统通过前端页面直观地展示《人世间》小说人物的各类关系，通过智能问答搜索小说的相关内容，提升了面向用户的交互知识服务效果，为推动当代文学作品传播提供有效的信息服务。

关键词：中国当代文学；知识图谱；Neo4j；智能服务
doi：10.3969/J.ISSN.1672-7274.2024.09.083
中图分类号：G　353.1；TP　3　　文献标志码：A　　文章编码：1672-7274（2024）09-0248-05

Research on Intelligent Services for Literary Works Based on Knowledge Graph
--Taking the Novel "The World" as an Example

XIE Zhenglin, QIAN Qing,CHEN Qingrong
(School of Information, Guizhou University of Finance and Economics, Guiyang 550000, Guizhou, China)

Abstract: The construction of a knowledge graph of contemporary literature aims to deepen the understanding and analysis of literary works; At the same time, it provides an efficient tool for literary researchers to interpret and study contemporary literary works from multiple dimensions and levels. This article takes the novel "The World" as the object and constructs a knowledge graph of the novel from top to bottom. It also constructs an intelligent service system for the novel, including a relationship retrieval system and an intelligent question answering system. This system intuitively displays various relationships between characters in the novel "The World" through front-end pages, and can search for relevant content of the novel through intelligent Q&A, improving the effectiveness of user oriented interactive knowledge services and providing effective information services to promote the dissemination of contemporary literary works.

Keywords: contemporary chinese literature; knowledge graph; Neo4j; intelligent services

文学作品是精神文化产品的重要组成部分，其中蕴含着世界各地不同历史时期人类的经验、情感和智慧。近年来，随着自然语言处理（NLP）和知识图谱技术的发展，学者们已开始利用这些工具挖掘和分析文学作品，以期从新的角度解读和传播这些作品[1]。

然而，首先文学作品中的语言表达具有丰富的修辞手法和意象，这需要深入的语义理解和分析[2]。这种多样性给实体识别和关系抽取带来了挑战。其次，文学作品中的实体和关系是动态和多层次的，这需要复杂的模型来捕捉和表示[3]。最后，由于文学作品的数量庞大和类型多样，构建大规模的文学作品知识图谱需要高效的方法和技术[4]。尽管存在以上挑战，构建文学作品知识图谱仍具有重要的理论和实践价值。知识图谱可以帮助读者从新的角度解读文学作品，揭示作品中的人物、事件、主题和动态变化等关键要素，为文学研究提供新的视角和方法[5]。同时，可以为文学作品推荐系统、在线学习平台、虚拟故事生成等应用提供基础数据和工具[6]。

本文以当代中国文学作品《人世间》为研究对象，探索构建文学知识图谱的技术与方法，分析其在文学研究和应用上的重要价值。希望通过对文学作品知识图谱的构建，推动文学作品知识图谱研究的进展，为文学研究和应用提供新的工具和资源。

1　相关研究

近年来，针对中文语料的知识图谱创建的研究逐渐受到关注。徐彤阳等[7]使用骨架法与七步法相结合的方式构建了晚明戏曲家的本体模型，实现了晚明戏

作者简介：谢政霖（1998—），男，汉族，江西赣州人，硕士研究生，研究方向为信息服务。
钱　清（1986—），女，汉族，贵州贵阳人，副教授，博士研究生，研究方向为信息资源管理。
通信作者：陈清容（1985—），女，汉族，贵州贵阳人，讲师，硕士研究生，研究方向为信息资源管理。

曲家知识图谱的创建。张强等[8]以皖籍开国将军为研究对象，运用了自顶向下的方法构建知识图谱，利用GIS技术描绘了人物活动的轨迹，并创建了一个以智能问答为核心的红色历史人物智能服务系统。张云中等[9]采用CBDB、上图人名规范库、上图古籍资源、上海地方志、古诗文网、历史人物年谱等作为数据来源，在CBDB数据库框架的基础上提炼和完善了历史文化名人游学足迹关系数据模型。欧阳剑等[10]通过知识图谱技术对中国历代典籍进行了知识组织，构建了一个涵盖需求层、模型层、应用层三个部分的典籍知识图谱框架模型。

在文学作品研究领域，知识图谱技术的应用正在悄然兴起，但目前的研究对象主要集中于古代文学作品和历史人物，对当代文学作品的知识图谱构建研究较少，同时在知识服务方面的研究成果也相对匮乏。针对这一问题，本文采用自顶向下的方法构建《人世间》的知识图谱，并基于知识图谱和图数据库设计文学作品知识服务系统。这既为未来传统小说的知识图谱构建研究提供了参考，也满足了对数字人文研究中小说资源的再利用需求，具有重要的实践意义。

2 实证研究

2.1 数据来源

本文以梁晓生先生所著的《人世间》作为主要研究素材。《人世间》分为上、中、下三部曲，通过讲述周氏家族等十几位平民子弟跌宕起伏的人生展示了改革开放给中国社会带来的巨变。该作品于2019年8月16日荣获第十届茅盾文学奖。笔者通过熊猫搜书平台下载《人世间》三部曲电子书资源，作为相关研究的原始数据。

2.2 本体构建

结合《人世间》小说原始数据的特点以及人物关系的描述，本研究以七步法为基础，借鉴已有的文学类本体构建框架并进行调整，构建《人世间》小说本体模型（如图1所示）。

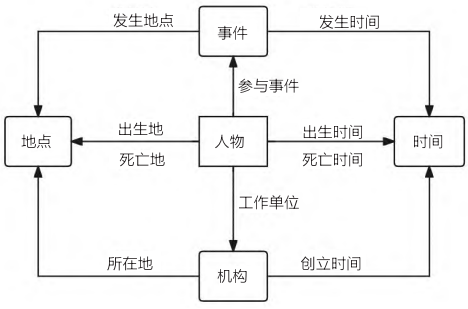

图1 《人世间》小说本体模型

（1）明确本体的应用范围。小说本体应准确描述小说人物的基本信息、具体时间、机构名称、发生事件，才能厘清小说故事发展的脉络，以确保小说要素分类符合《人世间》原著中的叙事与描述。

（2）本体构建。本文采用FOFA本体模型、Clinga本体模型，以实现《人世间》小说本体的初步构建。

（3）重要术语提取。本研究从《人世间》小说中提取出相应的重要术语，包括人物、地点、时间、机构、时间、上下级、下属单位、兄弟姐妹、出生地、出生时间、创建时间、单位改编、性别、地名、政府部门等。

（4）定义类及其等级体系。本文将提取的《人世间》术语归纳为五大核心类别，分别为人物、地点、时间、机构、时间，将其他术语归纳为核心类别的子类别。

（5）定义类的属性及约束。本文进一步对文中其他类别进行归纳整理，将其作为属性划分给对应的类，并确立类别之间的关联。类的属性包括对象属性与数据属性。其中对象属性包括上下级、下属单位、兄弟姐妹、出生地、出生时间、去世时间等22个，数据属性包括别称、事件内容、性别、地名、机构名、姓名等12个。

（6）本体表示。为了对本体进行表示，本文对小说本体进行了表示，部分OWL文件内容表示如下。

①类的定义。

<owl:Class rdf:about="http://www.semanticweb.org/administrator/ontologies/2023/4/untitled-ontology-10#人物"/>

②对象属性的定义。

<owl:ObjectProperty rdf:about="http://www.semanticweb.org/administrator/ontologies/2023/4/untitled-ontology-10#参与事件">

<rdfs:subPropertyOf rdf:resource="http://www.w3.org/2002/07/owl#topObjectProperty"/>

<rdfs:domain rdf:resource="http://www.semanticweb.org/administrator/ontologies/2023/4/untitled-ontology-10#人物"/>

<rdfs:range rdf:resource="http://www.semanticweb.org/administrator/ontologies/2023/4/untitled-ontology-10#事件"/>

</owl:ObjectProperty>

③数据属性的定义。

<owl:DatatypeProperty rdf:about="http://www.

semanticweb.org/administrator/ontologies/2023/4/untitled-ontology-10#事件内容">

</owl:DatatypeProperty>

（7）实例创建。本文在Protégé中填充实例以便判断类与类之间的关系是否明确，本体结构是否符合应用需求。图2是《人世间》本体模型周秉义实例化的成功运用，从图中可以看出，本体模型的可用性较强，且能够准确表达小说中主要术语、对象属性和数据属性之间的组织关系。

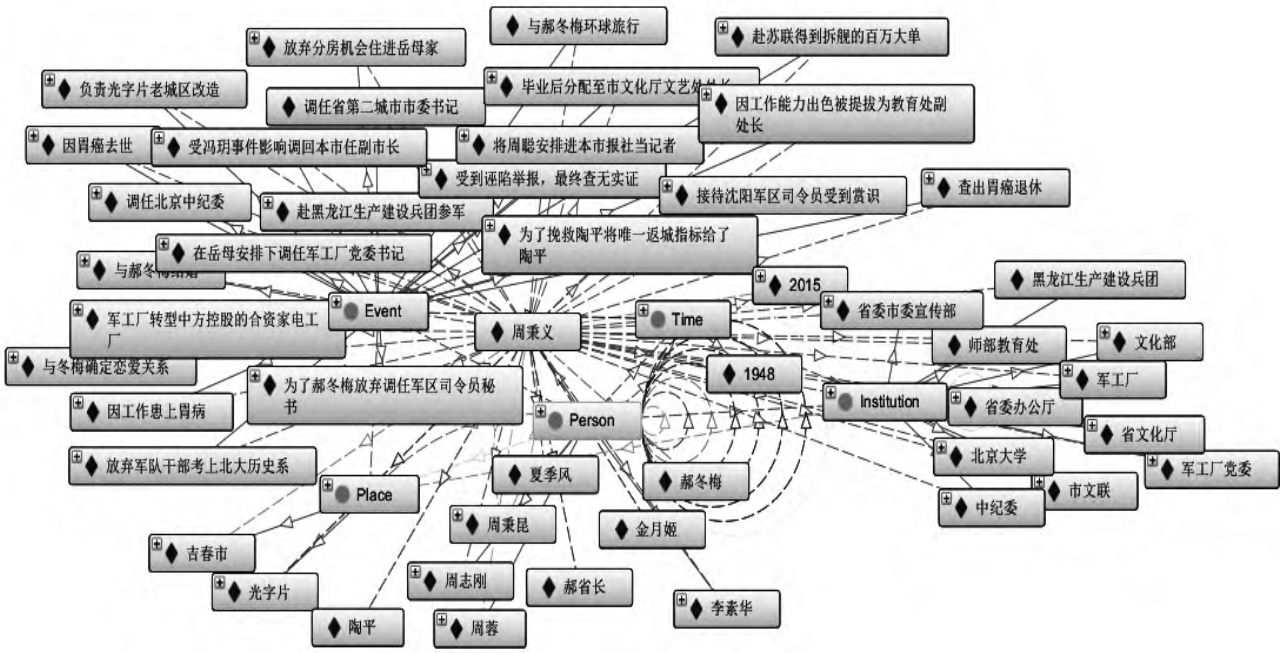

图2 《人世间》本体模型周秉义实例化演示

2.3 知识抽取

实体识别是知识抽取的第一步，目标是识别文本中的实体，通常使用命名实体识别（Named Entity Recognition，NER）技术实现。本研究选用结合双向神经网络和softmax函数的模型对《人世间》三部曲原始数据数据进行知识抽取，使用深度学习方法从非结构化数据中提取结构化数据，以构建高质量的知识三元组。例如，“周秉昆在酱油厂工作”这句话经过三元组抽取得到三元组“周秉昆-工作-酱油厂”。

2.4 知识融合

经过知识抽取之后的部分实体和关系可能存在歧义现象，如人物姓名这一数据属性可能存在多种称呼，以“周秉昆”为例，其母亲对其的称呼为“老疙瘩”，亲朋好友称呼其为“秉昆”，这些称呼都指代同一人物实体“周秉昆”。对于存在上述问题的实体和关系，本研究采用计算余弦相似度的方法进行同义实体的合并，同时结合人工方法消除部分有歧义的实体。余弦相似度的计算公式如式（1）所示。在研究应用领域，0.8的相似度已经足够高，可以认为两个对象是相似的。

$$\text{similarity}(A,B)=\frac{A*B}{\|AB\|} \quad (1)$$

2.5 知识存储

Neo4j图数据库由标签、节点、关系及节点属性4类要素组成，本研究将类与标签、实例与节点、对象属性与关系、数据属性与节点属性一一对应，从而实现小说本体模式层到图谱数据层的匹配映射。如人物类与地点类组成的对象属性“周秉昆，出生地，光字片”。标签分别为人物类与地点类，实例分别为“周秉昆”与“光字片”，关系为“出生地”。

本研究采用了Neo4j4.0.7图数据库4.0.7版本，JDK的依赖版本为11.0.18。将protege中的三元组数据导出为RDF/XML文件，再通过Neosemantics插件和Cypher语言将RDF/XML格式的三元组数据导入图数据库Neo4j图数据库中。最终，本研究构建了由1 194个节点和2 706条边组成的知识图谱，图3为部分内容构成的知识图谱。其中，本体模型中五大类的实例映射在图数据库中为不同颜色的节点。数据属性映射为每个节点的属性。对象属性映射为节点之间的连线。通过该图，可以较好地展示小说中人物、时间、事件、地点等属性之间的关系，由此为开发智能服务系统奠定基础。

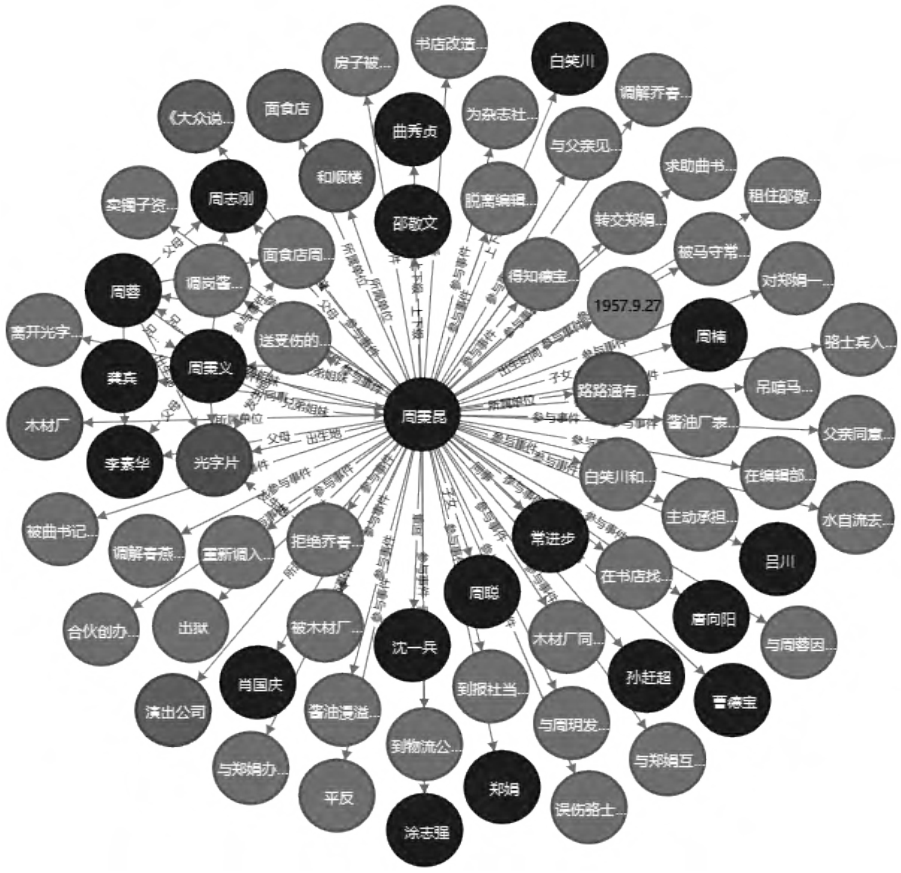

图3 《人世间》小说部分知识图谱

2.6 知识服务

2.6.1 关系检索系统

关系检索是指在前端页面直接输入某个历史人物，实现对该人物相关信息和事件的检索。“前端通过POST方法发送query和category数据。后端接收这些数据，构建Cypher查询语句，查询Neo4j数据库，并返回相关的图谱数据（节点、边和描述）。

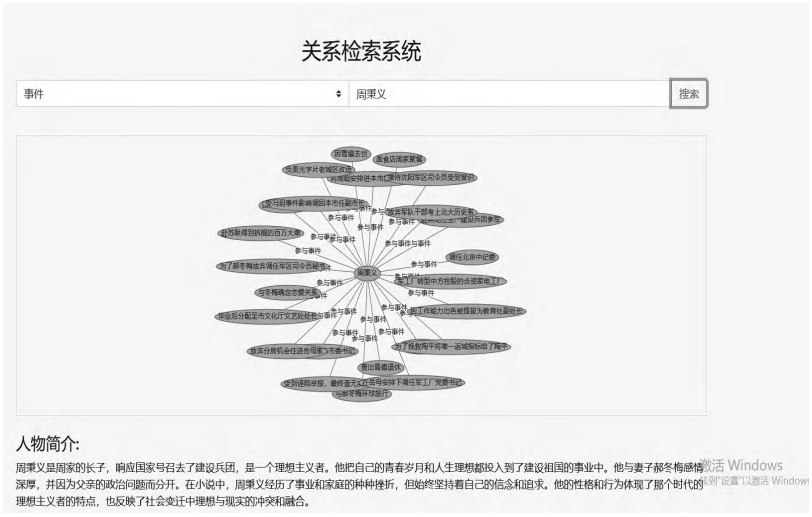

图4 检索系统实例图

图4为检索系统，展示《人世间》小说中周秉义的人物与事件关系检索结果，从图中可以看到，通过检索可以有效地展示其人物相关的知识图谱和人物简介，让使用者能够快速了解小说中人物的信息。

2.6.2 问答系统

本文还构建了一个智能问答系统，后端服务通过结合Flask和Neo4j技术，为前端提供了一个高效、准确的关系检索功能。当处理搜索请求时，系统首先从前端接收到的数据中提取查询文本。接着，采用字符串操作方法解析实体和关系。例如，问题“周秉昆的妻子是谁”，系统将其解析为实体“周秉昆”和关系“配偶”。随后，构建了一个Cypher查询语句，用于在Neo4j数据库中检索与该实体和关系相关的节点和边。查询结果被处理并转换为前端可以使用的格式，包括节点、边和目标实体的简介。最后，这些数据以JSON格式返回给前端并进行信息显示。

在前端展示中，该系统采用HTML5标准进行构建，引入了Bootstrap和vis.js两个主要的前端库。Bootstrap为系统提供了响应式的用户界面设计，使得该系统能够在不同的设备和屏幕尺寸上均保持良好的用户体验。vis.js则为系统提供了图形可视化功能，使得用户可以直观地查看和理解检索到的关系数据。系统的主界面包含一个下拉选择框，在查询关系全貌的同时，还允许用户选择查询不同的类别，如“人物”“事件”等关系，以及一个输入框，用于输入查询的实例名称。用户在输入查询后，可以点击“搜索”按钮，触发search函数。该函数通过jQuery发送POST请求至后端，并将返回的数据用于更新图形和描述部分。此外，系统还引入了其他的JavaScript库，如jQuery和Popper.js，以支持前端的交互功能和Bootstrap的组件功能。

在系统的主界面，用户通过输入框输入问题。输入完成后，点击“搜索”按钮，触发search函数。此函数利用jQuery发送POST请求至后端，并将返回的数据用于更新图形和简介部分。整体布局采用了Bootstrap栅格系统，将页面分为两个主要部分：左侧展示关系图谱，而右侧则显示目标实体的简介。

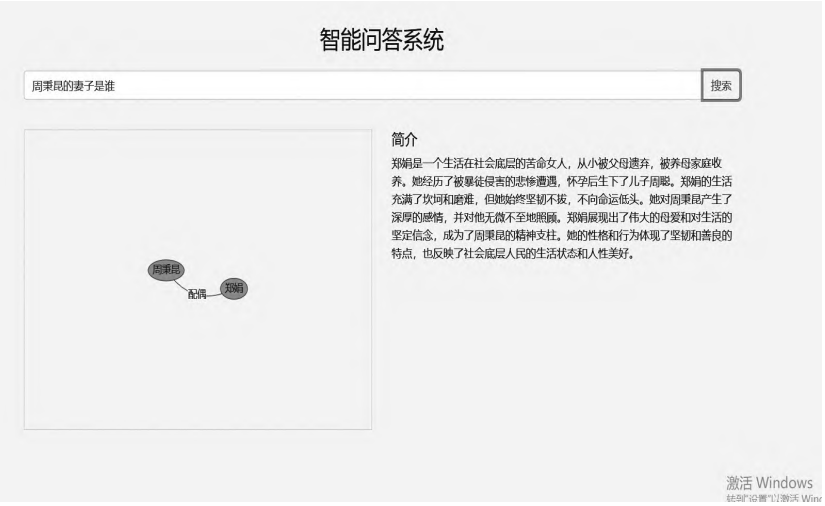

图5 智能问答系统实例图

图5为智能问答系统，以问题“周秉昆的妻子是谁”为例，结果展示区显示出周秉昆和郑娟的配偶关系，并在右侧提供关于郑娟的简介，达到信息提取的目的。

3 结束语

随着数字人文技术的发展，传统人文知识的组织和服务方式得到了革新。本文构建了小说《人世间》的智能服务原型系统，为当代小说信息服务带来了新的方法和视角。通过知识图谱的应用，文中实现了《人世间》小说的深度知识组织、关联的直观展示。此外，该系统不仅可为图书馆、博物馆等机构提供了参考，助力文学研究和教育活动的推进，还通过直接检索和智能问答的方式，极大地提高了用户的人机交互体验。值得一提的是，本研究所采用的方法不仅具有高度的通用性，还可以根据不同的文学资源特征进行灵活调整和复用。未来将进一步拓展当代文学作品的样本集和数据集，同时探索和实施更多元化的智能服务方式，如微信小程序和智能服务App，以期将研究成果更好地应用于实践中。■

参考文献

[1] Manolis Koubarakis, G. Stoilos, Ian Horrocks, Phokion G. Kolaitis. An Introduction to Ontology-Based Query Answering with Existential Rules [J]. Reasoning Web, 2014(8):245-278.
[2] Paulheim H. Knowledge graph refinement: A survey of approaches and evaluation methods[J]. Semantic web, 2018(3):489-508.
[3]Harispe S, Ranwez S, Janaqi S,Montmain J. Semantic similarity from natural language and ontology analysis[J]. Synthesis Lectures on Human Language Technologies, 20188(1):1-254.
[4] 奥德玛，杨云飞，穗志方．中文医学知识图谱CMe KG构建初探[J]．中文信息学报，2019，33（10）：1-7.
[5] 林峰，赵广平，林娜，等．《红楼梦》文本的社会网络结构分析[J]．石家庄铁道大学学报，2018，12（1）：58-63.
[6] Gangemi A, Presutti V, Reforgiato Recupero D, Nuzzolese, A. G., Draicchio F, Mongiovì M.Semantic web machine reading with FRED. [J].Semantic Web,2017(8): 873-893.
[7] 徐彤阳，黄映思．名人年谱资源的知识图谱构建——以徐朔方《晚明曲家年谱》为例[J]．数字图书馆论坛，2022（12）：36-45.
[8] 张强，高颖，刘飞，等．基于知识重组的红色历史人物智能服务研究[J]．现代情报，2023（7）：96-108.
[9] 张云中、孙平．历史文化名人游学足迹知识图谱的构建与可视化[J]．图书馆杂志．2021，40（9）：81-87.
[10] 欧阳剑，梁珠芳，任树怀．大规模中国历代存世典籍知识图谱构建研究[J]．图书情报工作．2021，65（5）：162－173.

（上接第244页）

参考文献

[1] 王帆．烟草物流的数字化赋能影响因素研究[J]．中国储运，2023（11）：8-9.
[2] 陈浩，杨明，黄丹．我国烟草物流协同发展现状与趋势[J]．中国市场，2023（07）：11-12.
[3] 周杰，王申，田敏．基于无线传感器网络的远程医疗监护系统设计[J]．科技与创新，2019（13）：59-60.
[4] 胡伟，李博为，孟建明，尹志良．物联网技术在烟草机械设备中的应用[J]．内燃机与配件，2020（03）：2-3.
[5] 尚晏莹．“互联网+”商业模式创新影响因素及路径研究-以制造企业为例[D]．西安：西安理工大学，2019（05）：21-22.
[6] 潘博．“双碳”目标下烟草物流碳排放管理机制研究与探索[J]．物流技术与应用，2022（05）：33-34