

# 学科发展状态的知识图谱构建

陆菁宇,张绍阳,黄文旒  
(长安大学 信息工程学院,陕西 西安 710064)

**摘要:**使用科学知识图谱的方法,对交通信息工程及控制学科发展状态进行探索。对学科发展状态进行分析可帮助各方深入认识学科的历史、现状和趋势。以2009–2018年间交通信息工程及控制学科评估等级B+以上的高校硕博学位论文为分析对象,绘制科学知识图谱。分析知识图谱发现该学科科研技术种类多样,并分别从公路交通、水路交通和铁路交通三个领域进行对比分析,发现该学科三个领域中的研究技术热点以及不同领域之间的科研技术共性。另外,通过构建交通信息工程及控制学科的技术演进知识图谱,呈现出该学科科研的技术发展脉络和知识流动趋势,并得出机器学习以及车联网技术为交通信息工程及控制学科未来科研的研究趋势。为该学科的科研工作者提供深入研究的参考。

**关键词:**交通信息工程及控制;学科发展状态;研究热点;知识图谱;citespace

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2020)06-0145-06

doi:10.3969/j.issn.1673-629X.2020.06.028

## Analysis of Development Status of Discipline Based on Knowledge Graph

LU Jing-yu, ZHANG Shao-yang, HUANG Wen-ni

(School of Information Engineering, Chang'an University, Xi'an 710064, China)

**Abstract:** By using the scientific knowledge graph, the development status of traffic information engineering and control discipline is explored. The analysis of the development status of disciplines can help all parties to have a better understanding of the history, current situation and trend of the discipline. The dissertations of master's degree and doctor's degree from universities with B+ or above in traffic information engineering and control disciplines from 2009 to 2018 to be chosen as object to draw the scientific knowledge graph. Discovering a wide variety of research techniques in the discipline, it also makes a comparative analysis of highway traffic, waterway traffic and railway traffic respectively, which shows the research focuses in three fields and technical commonalities among them. In addition, by analyzing the evolution of technology in the three fields, it presents the technological development thread and knowledge flow trend of scientific research in this discipline. And draw the research trend of machine learning and vehicle networking technology for future scientific research of traffic information engineering and control discipline, which provides a reference for further research for scientific researchers in this discipline.

**Key words:** traffic information engineering and control; subject development status; research hotspot; knowledge graph; citespace

## 0 引言

学科是高等院校教学和科学研究的基本构成单元,凝聚了一批高水平、高层次的科技人才,是大学进行人才培养、科学研究的重要依托。学科的科研水准反映的是一所高校的核心竞争能力<sup>[1]</sup>。对学科状态进行分析,不仅可以让科研机构明确认识到当前学科的长处和短板,帮助提升硕博研究生的培养质量;同时也为外界了解高等院校以及科研机构学科内涵建设提供渠道。

教育部学位与研究生教育发展中心自2002年起对全国有授予博士或硕士学位资格的一级学科开展整体水平评估<sup>[2]</sup>。到2017年为止已经进行了四次,第四轮学科评估以“公共数据和单位填报相结合”的方式获取数据,并且设置了“师资队伍与资源”“人才培养质量”“科学研究水平”“社会服务与学科声誉”四个一级指标及对应的二级指标。学科评估全面反映了学科的总体状况,但由于学科评估是针对所有学科的,不会针对每个学科的内涵进行分析。

收稿日期:2019-08-31

修回日期:2019-12-31

基金项目:中央高校基本科研业务费专项资金项目(300102249202);陕西省技术创新引导专项(2018XNCG-G-16)

作者简介:陆菁宇(1996-),女,硕士研究生,CCF会员(D4252G),研究方向为自然语言处理;张绍阳,博士(后),教授,研究方向为交通信息大数据及标准化、软件技术及理论。

知识图谱可以直观清晰地展现科学知识的发展进程和知识结构关系<sup>[3]</sup>,因此很多研究者将它应用在自己的研究领域。一方面可以利用知识图谱挖掘学科领域的基础知识,展现学科知识架构。例如,傅居正、喻国明使用知识图谱对数据新闻学科进行研究<sup>[4]</sup>。廉同辉、余菜花、宗乾进使用知识图谱对旅游学科研究领域、主要研究机构、学术代表人物进行探究<sup>[5]</sup>。吴远仁研究2001–2017年间人口、资源与环境学科的核心研究领域<sup>[6]</sup>。另一方面可以通过知识图谱找到科学知识的内在连接,并利用“图谱”特征直观呈现学科的发展规律和演化路径<sup>[7]</sup>。例如,孙艳、田丽梅使用中国知网收录的“中文核心期刊”和“CSSCI”相关研究文献对舆情的研究现状进行分析,总结文献内在的联系和科学结构<sup>[8]</sup>。张学召使用共词知识图谱探究材料学和工程学两门学科的研究热点、趋势<sup>[9]</sup>。饶武元、刘浩使用共现知识图谱对一流学科的科研热点以及科研前沿进行分析<sup>[10]</sup>。当前采用知识图谱对学科展开的研究主要使用的数据源是Web of Science和CNKI中收录的期刊文献,并且研究内容大多是对研究领域、发文量、领域经典文献、研究热点进行分析,而对于学科的研究技术、技术热点和发展有所忽略<sup>[11]</sup>。科研过程中使用的技术直接反映了研究人员的科研水平和学科发展状况,因此文中从研究技术的角度对学科的发展状态进行探析。

交通信息工程及控制学科是“交通运输工程”一级学科下的二级学科,也是现代交通运输工程发展的核心学科,为国家输送交通科技人才,对社会和国民经济建设有重要的意义。当前尚没有对交通信息工程及控制学科发展状态进行分析的文献。硕博士是高校科研的主力军,其研究水平能够从一个层面反映一个学科的水平,硕博士的主要产出即学位论文。学位论文能够体现作者的基础理论知识、研究领域、手段和成果,综合起来就是一个学科的整体状态。为此,文中以硕博士论文为分析对象,对交通信息工程及控制学科的发展状态进行分析。探究该学科的科研技术热点和科研现状,探索该学科不同领域的科研技术热点共性、科研趋势等,为科研机构了解自身的科研水平以及确定该学科未来科研方向提供参考和借鉴。

## 1 数据来源与处理

学科评估是对一个学科的整体水平进行评估,对于评价学科建设的成效是一个重要的评价依据。在本研究中,选择在《全国第四轮学科评估》中评估等级“B+”及以上的高校作为代表对该学科进行分析<sup>[12]</sup>。分别有:东南大学、西南交通大学、北京交通大学、北京航空航天大学、同济大学、大连海事大学、哈尔滨工业

大学、武汉理工大学、中南大学和长安大学。这几所高校在该学科科研研究基础深、具有较高影响力,可以反映出该学科真实的科研现状。

文中以CNKI中该学科的硕博学位论文文献为数据。在高级检索中的硕博论文中设置检索条件:学科专业名称=交通信息工程及控制;学位年度=2009–2018;学位单位=上述高校其中之一。在检索过程中,没有检索到同济大学和北京航空航天大学相关的学位论文,因此经过筛选采用剩下八所高校检索出的2393篇学位论文作为研究样本。具体发表数量分布如图1所示。

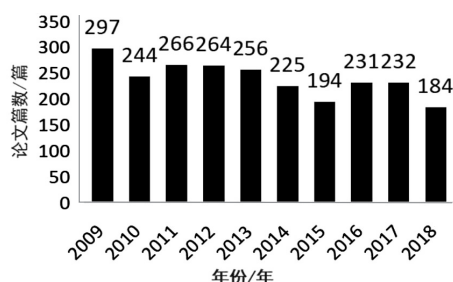


图1 2009–2018 学位论文年度分布

如图1所示,该学科的硕博论文的发文量平均波动程度较小,并且整体呈缓慢下降的趋势。总的来说,交通信息工程及控制学科硕博论文一直维持着一定程度的发文量,反映出该学科的研究已经迈入一个比较成熟的研究领域。

将数据进行导出并完成数据格式转换,如图2和图3所示。



图2 数据格式转换



图3 数据处理前与处理后(部分)

## 2 学科发展状态图谱构建

知识图谱区别于传统的文字、表格,而是以网状结构描述世界上的实体、概念以及实体与概念之间的关

系,并且采用图形化的方式使人们更加直观地发现、理解这些信息<sup>[13-14]</sup>。知识图谱可以针对某一学科专业领域的科学技术发展态势及其相关知识结构进行深入研究,进而发现科技活动中潜在的一般和特殊规律,也使得管理人员可以有效监测学科发展动态,从而推动学科整体的发展。因为交通信息工程及控制学科相关高校的主要研究领域可分为公路交通、水路交通和铁路交通三个领域。为了对学科有更加全面、深入的了解,按照相关研究领域构建图谱。

知识图谱的构建主要包括确定研究领域、数据检索与清洗、图谱构建与图谱分析等步骤。具体如图4所示。

## 2.1 参数设定

### (1) 时间切片划分。

将时间分区设定为1,则为每一年为一个时间分区,利用熵作为各时间段的宏观指标来衡量网络的有序性。

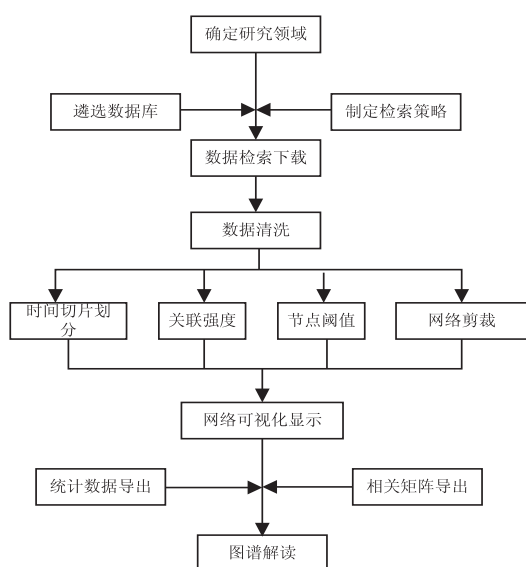


图4 图谱构建流程

### (2) 关联强度计算。

图谱中的节点显示的是分析的对象,连线表示着两个词之间的共现联系,连线的粗细则是表示着这种共现联系的强度。知识连接强度的计算主要有三种方法:

#### (1) 夹角余弦距离计算。

$$\text{Cosine}(x, y) = \frac{C_x C_y}{\sqrt{(\sum_{i=1} C_{xi})^2 (\sum_{i=1} C_{yi})}} \quad (1)$$

#### (2) Jaccard 距离计算。

$$\text{Jaccard}(x, y) = \frac{CC_{xy}}{C_x + C_y - C_{xy}} \quad (2)$$

#### (3) Dice 距离计算。

$$\text{Dice}(x, y) = \frac{2CC_{xy}}{C_x + C_y} \quad (3)$$

$x, y$ :任意的两个节点;

$C_x$ :与 $x$ 相连的节点构成的向量;

$C_y$ :与 $y$ 相连的节点构成的向量;

$C_{xi}$ : $x$ 与 $i$ 节点共现次数;

$C_{yi}$ : $y$ 与 $i$ 节点共现次数;

$C_{xy}$ : $x$ 与 $y$ 节点共现次数。

文中主要使用夹角余弦距离算法计算知识的连接强度。

### (3) 节点阈值。

数据量过多会造成图谱过于庞大和杂乱,因此需要对节点设定阈值,对数据进行筛选,去除冗余信息使图谱更加清晰。文中将阈值设定为 TOP30,即将每个时间切片中频次最高的 30 个节点数据作为绘制数据。

### (4) 网络裁剪。

在数据量庞大的情况下,图谱杂乱不宜解读,需要进行网络的裁剪使图谱更加清晰简洁。文中采用最小生成树算法来对网络进行简化,即在网络中,构建一个能够包含所有顶点,无圈且权值最小的子网络。

## 2.2 图谱绘制

点的中心性是一个用以量化点在网络中地位重要性的图论概念。节点中介中心性的大小代表着这个节点与图谱中其他节点的联系密切程度,中心性高的点往往位于连接两个不同聚类的路径上<sup>[15]</sup>。中心性超过 0.1 的节点称为关键节点。在 citespace 中对图谱进行构建。

### (1) 公路交通领域研究技术热点。

该学科的公路交通领域研究技术热点知识图谱如图5所示。图5中高中心性词见表1。

表1 公路交通领域研究技术热点高中心性词

序列	关键词	频次	中心性
1	神经网络	43	0.29
2	嵌入式系统	36	0.42
3	仿真	33	0.26
4	dsp	25	0.14
5	图像处理	23	0.23
6	模糊控制	23	0.17
7	支持向量机	20	0.05



图5 公路交通领域技术研究热点知识图谱



由图5和表1可知,在该学科研究公路交通领域中常用的技术热点有神经网络、嵌入式系统、仿真、DSP、图像处理、模糊控制、支持向量机。

## (2) 水路交通领域研究技术热点。

该学科的水路交通领域研究技术热点知识图谱如图6所示。图6中高中心性词见表2。

表2 水路交通领域研究技术热点高中心性词

序列	关键词	频次	中心性
1	航海模拟器	61	0.18
2	仿真	49	0.21
3	建模	43	0.23
4	模糊综合评价	37	0.19
5	神经网络	37	0.01
6	层次分析法	33	0.14

on Criteria: Top 50 per slice, LRF=3, LBY=8, e=2.0  
k: 10-115, E=122 (Density=0.0186)  
tCC: 72 (63%)  
Labeled: 5.0%  
p: None

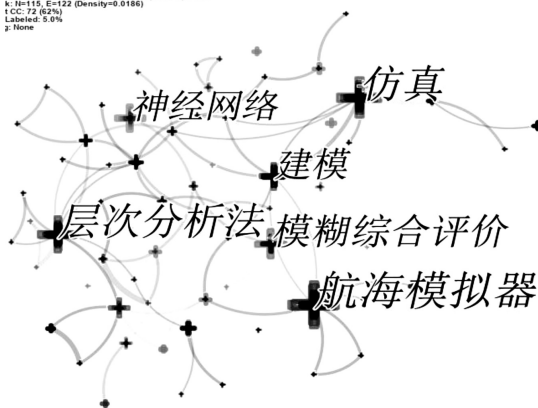


图6 水路交通领域技术研究热点知识图谱

由图6和表2可知,在该学科水路交通领域研究的技术热点主要为航海模拟器、仿真、建模、模糊综合评价、神经网络、层次分析法。

## (3) 铁路交通领域研究技术热点。

该学科的铁路交通领域研究技术热点知识图谱如图7所示。图7中高中心性词见表3。

图7所示。图7中高中心性词见表3。



图7 铁路交通领域技术研究热点知识图谱

表3 铁路交通领域研究技术热点高中心性词

序列	关键词	频次	中心性
1	仿真	89	0.34
2	建模	76	0.21
3	遗传算法	63	0.23
4	神经网络	59	0.11
5	自适应控制	53	0.06
6	时间自动机	48	0.09
7	贝叶斯网络	18	0.04
8	模糊控制	17	0.01
9	uppaal	16	0.06

自表3和图7可知,该学科在铁路交通领域主要的研究技术热点有仿真、建模、遗传算法、神经网络、自适应控制、时间自动机、贝叶斯网络、模糊控制、uppaal。

## (4) 知识演化路径。

为了更好地总结交通信息工程及控制学科的研究演化路径,文中对相关文献进行时区分析。通过时间的演变来展示知识的变化的图谱<sup>[16]</sup>。如图8所示,以一年为间隔,根据各个阶段内的连线,可以看出知识之间的演化关系。

Source: x=year-1  
/50 per slice, LRF=3, LBY=8, e=2.0  
19 (Density=0.017)

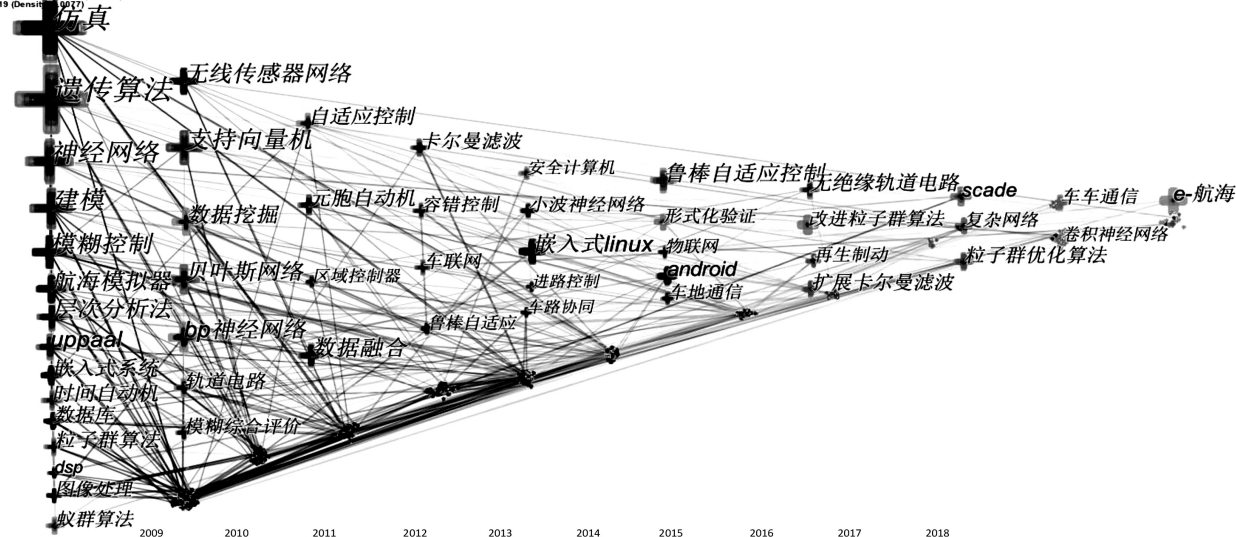


图8 交通信息工程及控制学科研究技术时区图

### 3 图谱分析

图谱中节点表示当前学科热点研究知识。点击节点可以显示与之相连的节点,展现知识之间的关系互联,帮助研究者对知识进行分析。

#### (1) 研究技术热点分析。

由图5和表1可知,在该学科研究公路交通领域中常用的技术热点。通过图谱可以发现,神经网络技术主要包括BP神经网络和卷积神经网络。主要被用于交通流量预测、路面裂缝检测、安全目标检测等。嵌入式系统技术主要包括嵌入式系统和嵌入式Linux。主要被用于智能家居、车辆定位、移动目标智能检测等。

仿真技术主要包括交通仿真和其他仿真。被用于隧道交通诱导仿真、语音识别仿真平台等。DSP技术主要被用于车速检测、智能车载抓拍系统等。图像处理技术主要包括图像处理、数字图像处理等,主要被用于疲劳驾驶检测、车牌定位、路面裂缝识别、干涉条纹检测等。模糊控制技术主要被用于高速公路智能控制、监控系统控制等。支持向量机技术主要被用于交通流预测、交通事件检测等。

由图6和表2可知,在该学科水路交通领域研究的技术热点。其中航海模拟器技术主要包括航海模拟器、拖轮模拟器、船舶操纵模拟器和海上搜救模拟器等主要被用于模拟真实场景以及模拟器性能的优化。仿真技术主要包括运动仿真、船舶交通仿真、三维仿真等。主要被用于船间效应仿真、雷达图像仿真、船舶交通流演化、船只避碰仿真、海上立体搜寻态势仿真。

建模技术主要包括建模、交通流模型、船舶运动数学模型等。主要被用于构建大规模地形场景模型、海图要素建模、航向保持、动态路径规划等。模糊综合评价主要包括模糊综合评价、模糊综合评判、粗糙集等。主要被用于智能避碰、溢油风险评价、安全评估、船员适任性评估等。神经网络主要被用于船舶自动靠泊、船舶直线航迹控制等。层次分析法主要被应用于通航安全评价、船舶碰撞责任划分辅助决策、船舶定线制优化等。

由图7和表3可知,交通信息工程及控制学科在铁路交通领域主要的研究技术热点。其中仿真技术主要包括仿真测试、仿真平台 视景仿真等。主要被用于CBTC系统测试、列车调度、列车自动驾驶、三维视景、等。建模技术主要包括uml、交通流模型、统一建模语言、模型转换等。主要被用于车-地通信、车-车通信、车载设等。遗传算法主要被用于节能运行、自动驾驶系统优化、交通流预测控制、故障诊断、交通事件检测、交叉口信号控制、交通流预测等。

神经网络主要包括BP神经网络、小波神经网络,

主要被用于故障诊断、列车定位、信号智能控制等。自适应控制技术主要包括无模型自适应控制、鲁棒自适应控制。主要被用于多列车协同控制、防滑牵引与制动数据驱动控制等。时间自动机技术主要被用于RBC控车、异常诊断等。模糊控制主要被用于车路协同、列车自动驾驶优化、交通信号协调控制、交通诱导、电机传动控制系统的节能运行控制等。

#### (2) 热点共性分析。

对交通信息工程及控制学科不同领域的技术热点进行研究分析,可以发现在不同的研究领域中,使用的研究技术是有一定的共性的。

仿真技术、神经网络是三个领域的共性技术。BP神经网络则是最高频使用的神经网络技术,其次是卷积神经网络。建模技术为水路交通和铁路交通领域的共性技术,建模经常与仿真技术一起使用,用来对列车或者船舶的设备、运行、通信等进行研究。模糊控制为公路交通和铁路交通领域的热点共性技术。模糊控制技术被用于信号协调控制等。

#### (3) 发展演进分析。

图8显示了该学科研究技术发展的历程。可以看出,该学科的博硕士生在科研过程中使用的技术种类多样化。

图8整体呈三角形,随时间的变化,技术由浅入深。如使用频次最高的仿真技术,从最开始的仿真到三维仿真、视景仿真。神经网络技术,从一开始的神经网络到BP神经网络、小波神经网络和卷积神经网络。神经网络有能力学习和构建非线性的复杂关系模型、高速寻找最优解。通常被应用在流量预测、目标识别、信号控制等方面。

嵌入式系统从嵌入式、Linux到嵌入式Linux,并且出现了专业的嵌入式软件scade。各类机器学习算法,如蚁群算法、支持向量机、贝叶斯网络等也随着机器学习的发展在研究中使用。控制技术也在不断发展,随着研究的深入以及技术的发展,也不断出现在相关研究中。

随着智能交通越来越热门,智能交通有关的技术也一直在深入,出现了车联网技术以及相关的车-地通信技术和车-车通信技术等实现车路协同、人车交互。

整体来说,该学科的研究技术类别多样化,并且出现的新技术呈逐年深入的状态,说明该学科的研究已经步入一个较为成熟的阶段。

### 4 结束语

以CNKI数据库中2009-2018年间交通信息工程及控制学科评估等级B+以上的高校硕博学位论文文献

为数据样本,构建该学科知识图谱,并对图谱进行探析,结论如下:

(1)当前学科科研过程中使用的研究技术种类比较多样化,主要是计算机技术、控制技术、通信技术。并且三个领域在科研过程中使用的技术有一定的共性,都使用了仿真技术和神经网络。

(2)因为学科的发展以及科研的深入,相关技术也在不断的发展。由时区图可知,各类机器学习算法不断改进,深度神经网络的不断深入,机器学习成为该学科的科研趋势。在发展“智能交通”的大背景下,各类通信技术在进步,云计算技术也在升级,使得车联网技术也成为未来的研究趋势。

#### 参考文献:

- [1] 郑海英. 地方高校办学特色的形成机制与发展战略研究——以三峡大学为例[D]. 宜昌:三峡大学,2012.
  - [2] 王晨,刘虹. 第四轮学科评估指标体系的变化、问题及启示[J]. 上海教育评估研究,2018,7(6):6-10.
  - [3] CHEN Chaomei. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377.
  - [4] 傅居正,喻国明. 数据新闻的学科建构:演进逻辑、知识图谱与前沿热点——基于美国核心期刊数据库 Web of Science (1992-2018)的文献分析[J]. 新闻记者,2018(10): 57-69.
  - [5] 廉同辉,余菜花,宗乾进,等. 基于 CSSCI 的 2000~2010 年旅游学科研究知识图谱分析[J]. 旅游学刊,2013,28(3): 114-119.
  - [6] 吴远仁. 基于知识图谱的人口、资源与环境学科前沿动态追踪[J]. 武汉商学院学报,2018,32(4):30-37.
  - [7] CHEN Chaomei. Science mapping: a systematic review of the literature[J]. Journal of Data and Information Science, 2017, 2(2): 1-40.
  - [8] 孙艳,田丽梅. 基于多维尺度分析的舆情研究主题词知识图谱[J]. 计算机技术与发展,2016,26(4):187-190.
  - [9] 张学召. 基于共词知识图谱的材料工程学科研究热点及研究趋势分析[J]. 创新科技,2015(8):80-85.
  - [10] 饶武元,刘浩. 一流学科研究的主题、前沿及问题探讨——基于关键词共现知识图谱的可视化分析[J]. 现代教育管理,2017(4):11-17.
  - [11] MODAK N M, MERIGÓ J M, WEBER R, et al. Fifty years of transportation research journals: a bibliometric overview[J]. Transportation Research Part A, 2019, 120: 188-223.
  - [12] 汤强. 以学科评估促进高校学科建设——基于教育部第四轮学科评估指标体系的分析[J]. 中国高等教育评估, 2018, 29(1): 9-11.
  - [13] 廖凤灵. 科学知识图谱构建与应用方法研究[D]. 武汉:武汉大学,2018.
  - [14] 戈其平,钟艳如. 基于数学教学的知识图谱构建[J]. 计算机技术与发展,2019,29(3):187-189.
  - [15] SMALL H. Co-citation in the scientific literature: a new measure of the relationship between two documents[J]. Journal of the American Society for Information Science & Technology, 2014, 24(4): 265-269.
  - [16] 朱记伟,蒋雅丽,翟翌,等. 基于知识图谱的国内外 BIM 领域研究对比[J]. 土木工程学报,2018,51(2):113-120.
- +++++
- (上接第 134 页)
- 交通科技与经济,2017,19(6):9-11.
  - [8] GUO J, XIE Z, QIN Y, et al. Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM[J]. IEEE Access, 2019, 7(99): 42946-42955.
  - [9] LI B. Urban rail transit normal and abnormal short-term passenger flow forecasting method[J]. Journal of Transportation Systems Engineering & Information Technology, 2017, 15(7): 43-48.
  - [10] DOU F, JIA L, WANG L, et al. Fuzzy temporal logic based railway passenger flow forecast model[J]. Computational Intelligence and Neuroscience, 2014, 2014: 1-9.
  - [11] 江天河. ARIMA 模型与 BP 神经网络模型在铁路春运客流量预测中的应用——以广东省为例[J]. 科技经济市场, 2017(7): 4-6.
  - [12] 冯冰玉,鲍学英,王起才,等. 新型组合模型在铁路客运量预测中的应用[J]. 铁道标准设计,2015,59(12):6-9.
  - [13] 黄召杰,冯硕. 灰色预测模型在铁路客流预测中的应用[J]. 交通科技与经济,2014,16(1):57-60.
  - [14] XIE M, LI X, ZHOU W, et al. Forecasting the short-term passenger flow on high-speed railway with neural networks[J]. Computational Intelligence and Neuroscience, 2014, 2014: 375487.
  - [15] FENG X, ZHANG H, GAN T, et al. Random coefficient modeling research on short-term forecast of passenger flow into an urban rail transit station[J]. Transport, 2016, 31(1): 94-99.
  - [16] NI M, HE Q, GAO J. Forecasting the subway passenger flow under event occurrences with social media[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(6): 1623-1632.