

Analysis on Research Status and Hot Spots of Rail Transit Based on Data Mining

Chunming Li Yifan Yang

Traffic Control Technology Co., Ltd., Beijing, 100070, China

Abstract

With the progress of social economy and technology and the acceleration of urbanization, China's rail transit industry has developed rapidly since 2001. According to the "Outline for the Construction of a Powerful Transportation Country" and the "Outline of the National Comprehensive Three-dimensional Transportation Network Planning" issued by the Central Committee of the Communist Party of China and the State Council, in recent years, the construction of China's rail transit industry will play an important role in the construction of a strong transportation country by the middle of this century. In order to clarify the current research status and hotspots of the rail transit industry, and to understand the development trend of the industry, this paper adopts method of data mining to research and analyze, taking the Chinese core literature with the subject word "rail transit" in CNKI and the top 3000 relevant documents as the target. The conclusion shows that from 2002 to the present, scholars' research on "rail transit" has shown an overall growth trend; among the top 3000 Chinese core literature data related to "rail transit" "Urban Rail Transit Research" and "Urban Rail Transit" are the most publications of all; in recent years, the most researched topics are traffic engineering, land use, signal systems, etc. and the most frequently used methods of related research are genetic algorithms, statistical analysis methods, etc.

Keywords

rail transit; data mining; research status; research hotspot

基于数据挖掘的轨道交通研究现状及热点分析

李春明 杨逸凡

交控科技股份有限公司, 中国·北京 100070

摘要

随着社会经济技术的进步、城市化进程的加快,中国轨道交通行业自2001年来,迅猛发展。据近年中共中央国务院发布的《交通强国建设纲要》及《国家综合立体交通网规划纲要》显示,至21世纪中叶,中国轨道交通行业的建设将在交通强国建设中发挥不容小觑的作用。为明确轨道交通行业的研究现状及研究热点,进而了解行业的发展趋势,论文采用数据挖掘方法,以中国知网(CNKI)中主题为“轨道交通”、相关度前3000的中文核心文献为研究对象进行目标的研究与分析。结论显示,2002年至今,学者们对于“轨道交通”的研究整体呈增长态势;在与“轨道交通”相关度前3000篇的中文核心文献数据中,《城市轨道交通研究》与《都市快轨交通》是发文量最多的期刊;近年,研究热度最高的话题是交通工程、土地利用、信号系统等;相关研究使用频率最高的方法是遗传算法、统计分析方法等。

关键词

轨道交通;数据挖掘;研究现状;研究热点

1 研究背景

2019年9月14日,中共中央国务院印发的《交通强国建设纲要》提出^[1],到2020年,完成决胜全面建成小康社会交通建设任务和“十三五”现代综合交通运输体系发展规划各项任务,为交通强国建设奠定坚实基础;到2035年,基本建成交通强国;到21世纪中叶,全面建成人民满意、保障有力、世界前列的交通强国。2021年发布的《国家综合立体交通网规划纲要》中多次提到都市圈、城市群中轨道交通间的衔接协调,城市内部充分利用城市轨道交通

通地下空间、优化客流疏散的公共交通系统运营。轨道交通建设也已成为中国交通强国建设中不可或缺的一环,更是成为了新型城镇化和城市化地区高质量发展的命脉。

轨道交通是运营车辆在特定轨道上行驶的一类交通工具或运输系统,随着城市化进程的加快,近年来,中国轨道交通的发展步入快车道。根据城市轨道交通协会数据显示^[2],2020年末,中国城轨交通运营线路长度为7969.7km。截至2021年9月30日,运行线路长度增长至8553.4km。2021年前三年季度共新增线路21条,新增运营线路长度为588.7km。行业的迅猛发展,不仅激发了企业家的投资热情,同时也为行业内的学者专家提供“产学研政”相结合的研究依据。

目前,学术界对于轨道交通的研究,主要集中于轨道

【作者简介】李春明(1983-),男,中国黑龙江齐齐哈尔人,硕士,工程师,从事轨道交通信号系统设计与研究。

交通规划、安全、信号等多个方面。现有对轨道交通规划方面的研究,主要从技术^[3]、政策^[4]、运营管理^[5]、发展理念^[6]等多角度出发研究阐述规划对于城市、国家发展的重要性。

为明确轨道交通行业的研究现状及研究热点,进而了解行业的发展趋势,论文采用数据挖掘的方法,对目标问题进行分析。论文首先在中国知网采集与“轨道交通”主题相关度前3000篇的核心期刊中文文献的数据,数据包括文章的题名、摘要、发表期刊及发表年份。然后,分别研究分析轨道交通行业的研究现状及研究热点。在研究现状部分,论文从年度、期刊两个角度出发,运用EXCEL进行相应的统计分析;在研究热点部分,论文运用ROST软件对采集得到的数据进行分词统计,进而依据词频得出热点关键词。最后,总结得到轨道交通行业的研究现状及研究热点相关的结论。

2 数据来源

2.1 数据爬虫

数据爬虫是一种自动化的程序,通过根据某些特定的设定,实现循环打开网页、选取数据、收集数据,完成自动采集所需数据的目标。

论文在中国知网中,以“轨道交通”为主题,“核心期刊”“排序:相关度”为约束条件,采集得到的相关度前3000篇中文期刊文献为数据研究基础。每条数据包括文章题名、摘要、发表期刊及发表年份。经分析统计,该相关度前3000篇的中文核心期刊文献,时间跨度为2002年至2021年,共来源于304个期刊。

2.2 数据预处理

在采集得到的相关度前3000篇中文核心期刊文献中,部分文献属于期刊发布的新闻,缺少关键词,如《都市轨道交通》中题为《南京全国首个“5G+智慧轨道交通项目”实验成功》的文章等。经统计,此类文献数据共计123条,不属于论文的研究范围,剔除该部分数据。筛选后,共计得到论文相关有效分析数据2877条。建立在该2877条数据的基础上,论文展开轨道交通研究现状及热点分析的研究。

3 研究现状分析

论文在进行研究现状分析时,从年度文献数据分析、期刊文献数据分析这两个角度出发。年度文献数据分析时,将2877条有效数据按年份归类统计,根据每年的文献数量进行年度分析;期刊文献数据分析时,将所有有效数据按期刊进行分类,从不同期刊发表的文章角度分析“轨道交通”类文章的现状。

3.1 年度文献数据分析

论文将采集获得的所有数据导入Excel进行数据处理。经分类、计算、统计,得到如表1所示的文献年份分布。2002年至今,学者专家对于“轨道交通”的研究整体呈增长态势。其中,学者们在2002年至2007年期间的研究处于

平稳发展阶段,2007年至2013年为逐渐发展阶段,2013年至今则发展为迅猛阶段。

表1 相关度前3000有效文献数据年份—频次—占比表

年份	频次	占比	年份	频次	占比
2002	6	0.21%	2012	81	2.82%
2003	6	0.21%	2013	100	3.48%
2004	8	0.28%	2014	157	5.46%
2005	9	0.31%	2015	211	7.33%
2006	6	0.21%	2016	190	6.60%
2007	14	0.49%	2017	216	7.51%
2008	27	0.94%	2018	249	8.65%
2009	49	1.70%	2019	296	10.29%
2010	52	1.81%	2020	516	17.94%
2011	67	2.33%	2021	617	21.45%

随着经济、技术的进步,轨道交通飞速发展,更多的学者专家将精力投入于“轨道交通”的研究。技术的飞速进步,促使轨道交通的发展日新月异,进一步使得轨道交通相关的技术越发成熟、先进。“产学研”紧密结合,形成良性发展循环。这将促进轨道交通行业更好更快的发展,为中国实现全面建成人民满意、保障有力、世界前列的交通强国提供坚实的支撑。

3.2 期刊文献数据分析

论文将2877条相关有效数据,按期刊合并同类项,得到相关度前3000条有效文献数据中,频次最高的前6个期刊,该6个期刊统称为“高度相关期刊”。经统计比较,高度相关期刊分别是《城市轨道交通研究》《都市轨道交通》《铁道运输与经济》《铁道设计标准》《交通运输系统工程与信息》及《铁道学报》。详细分布数据如表2所示。

表2 高度相关期刊频次及总文献数占比

期刊名称	频次	占比
城市轨道交通研究	763	26.52%
都市轨道交通	389	13.52%
铁道运输与经济	127	4.41%
铁道标准设计	123	4.28%
交通运输系统工程与信息	98	3.41%
铁道学报	67	2.33%

如表2所示,相关度前3000的有效文献数据频次高于60的期刊仅有6个,分别占比26.52%、13.52%、4.41%、4.28%、3.41%、2.33%。该6个期刊有效文献数据在所有高度相关期刊的占比中,《城市轨道交通研究》发表数量高达49%,其他依次为《都市轨道交通》《铁道运输与经济》《铁道设计标准》《交通运输系统工程与信息》及《铁道学报》,占比分别为25%、8%、8%、6%及4%。由此可见,相关度前3000篇的有效文献数据中,发文章最多的期刊是《城市轨道交通研究》,其次是《都市轨道交通》。

4 研究热点分析

论文在研究热点分析时,以分词词频数据作为辅助依

据,即词频高,则其被研究的热度就高;相反,词频越低,则意味着该方面的研究热度越低。

在进行词频分析前,将采集得到数据的文献题名及关键词作为研究对象,运用ROST软件进行数据算法的分词及词频统计。数据处理完毕后,进行20年数据的分词总词频分析。在分词总词频统计分析时,论文剔除“城市轨道交通”“轨道交通”“城市交通”“地铁”等与主题词“轨道交通”相同或相近的词汇,并选取词频频次前20的分词进行排序。

如表3所示,与“轨道交通”相关度前3000篇的中文核心文献中,分词词频最高的是交通工程,其次是土地利用、信号系统等。常被研究的问题主要集中于交通工程、土地利用、信号系统、复杂网络等;研究轨道交通相关问题的常用方法为遗传算法、统计分析、仿真、特征价格模型、大数据及层次分析法。

表3 词频前20分词表

分词	频次	分词	频次
交通工程	46	安全性	21
土地利用	39	客流特征	21
信号系统	37	仿真	21
复杂网络	36	客流	20
客流预测	31	大数据	20
遗传算法	26	特征价格模型	19
全自动运行	26	运营管理	19
互联互通	24	层次分析法	18
统计分析	22	可靠性	18
网络化运营	21	规划	17

在互联网时代,实体经济依旧是不容小觑的一环,交通工程、土地利用等实体的规划、运营管理,无不彰显着一个行业的能力。与时代共同进步,也是行业能力的一大衡量标准。信号系统、复杂网络、全自动运行、互联互通、网络化运营、大数据等,这些被贴上时代标签的词汇,也正是“轨道交通”这个行业与时俱进的诠释。遗传算法、统计分析、

层次分析法,用科学量化的数学方法研究问题亦是行业未来发展强有力的支撑。从热点看未来趋势,轨道交通行业的发展也必将安全、可靠。

5 结语

随着社会经济的发展、科学技术的不断进步,中国轨道交通行业也迎来了高速发展。论文通过对“轨道交通”相关度前3000篇的有效中文核心期刊文献进行年度、期刊两个角度的统计分析,并依据计算机分词技术得到词频统计数据,进而识别出轨道交通行业内的研究现状与研究热点。近20年,学者们对于“轨道交通”的研究整体呈增长态势;研究热点问题主要集中于交通工程、土地利用、信号系统等方面。此外,该领域内的研究方法多采用遗传算法、统计分析、仿真等方法。论文对轨道交通行业的研究现状与热点进行深入探讨,有利于学者们更好地明确学术界对轨道交通行业的认识,并为后续开展更深入的研究提供参考。

参考文献

- [1] 中华人民共和国中央人民政府.中共中央国务院印发《交通强国建设纲要》[EB/OL].(2019-09-19) http://www.gov.cn/zhengce/2019-09/19/content_5431432.htm.
- [2] 前瞻经济.2022年中国城市轨道交通行业市场现状及建设规划分析“十四五”计划新增运营里程3000公里[EB/OL].(2021-10-28) <https://baijiahao.baidu.com/s?id=1714850619564738335&wfr=spider&for=pc>.
- [3] 陈锋.基于手机大数据的城市轨道交通规划策略研究——以石家庄市为例[J].都市快轨交通,2021,34(4):8-15.
- [4] 任兵杰.国家政策对都市圈轨道交通发展的影响研究[J].城市轨道交通,2021(9):50-52.
- [5] 房庆恒,田鑫,罗晨伟.广州都市圈视角下的穗莞城市轨道交通衔接规划研究[J].交通与港航,2021,8(4):31-37.
- [6] 刘超,丁波.城市轨道交通运营管理的规范化策略探究[J].中国储运,2021(9):78-79.