



开放科学
(资源服务)
标识码
(OSID)

离群专利与论文视角下新兴技术主题识别研究

邱悦文

江苏大学科技信息研究所 镇江 212013

摘要: [目的/意义] 针对当前新兴技术识别主要围绕热点专利与论文开展,对低热度专利与论文不够关注,然而由于其类别较多、总体规模较大,极易产生长尾效应。因此,以低热度专利与论文为研究对象,设计新兴技术主题识别方法,以期为国家、企业、机构提供情报支持。[方法/过程] 构建专利、论文相似矩阵,利用社会网络分析工具聚类筛选出离群专利分类号与论文关键词;通过 LDA 模型获取技术主题;短语相似度测算,归并技术主题;围绕新兴技术特征构建识别指标体系对技术主题进行综合评价,识别出新兴技术。[结果/结论] 以车联网为实验对象,识别出无人驾驶技术是车联网领域新兴技术,识别结果与业界分析一致,说明方法有效性。

关键词: 主题识别;新兴技术;识别方法;车联网

中图分类号: G353.1; U463.6

Research on Emerging Technology Topic Identification based on Outlier Patent and Paper

QIU Yuewen

Institute of Science and Technology Information, Jiangsu University, Zhenjiang 212013, China

Abstract: [Objective/ Significance] The topic identification of current emerging technologies is mainly carried out around hot patents and papers, and there is not enough attention to outlier patents and papers. However, these patents and papers are easy to produce long tail effect. Therefore, this research uses outlier patents and papers to design a new method of identifying emerging technology topics, for providing intelligence support for countries, enterprises, and institutions. [Methods/Process]

基金项目 国家重点研发计划项目“基于人工智能技术创新服务技术集成研发与应用示范”(2019YFB1405200)。

作者简介 邱悦文(1996-),硕士研究生,研究方向为专利情报分析, E-mail: jsuqiuyuewen@126.com。

引用格式 邱悦文. 离群专利与论文视角下新兴技术主题识别研究[J]. 情报工程, 2021, 7(1): 81-93.

First, construct similarity matrix, use social network analysis to cluster and filter out the outlier patent classification numbers and paper keywords; then use the LDA model to obtain the topic of outlier patents and papers. Next, through the phrase similarity measurement, the technical topics are merged. Finally, the identification system is constructed from characteristics of emerging technology to evaluate the technical topics and identify emerging technologies. [Results/Conclusions] Taking the Internet of Vehicles as the experimental object, it has successfully identified that unmanned driving technology is an emerging technology in the field. The identification result is consistent with the industry analysis, indicating the effectiveness of the method.

Keywords: Topic identification; emerging technology; identification Method; IOV

引言

随着社会进步,尤其新一代通信技术发展,新兴技术如雨后春笋般,层出不穷。新兴技术并非只带来旧技术变革,更重要改变人们思想观念与生产方式,新兴技术对人类社会影响深远。识别新兴技术并长期监测全球技术前沿,不仅更早捕捉和发现技术动态,更能尽早把握住发展机会,无论国家还是地区,关注新兴技术都至关重要。不仅如此,新兴技术识别对政策制定者、企业、研究所与科研人员,都能及时而准确提供科技研究趋势与领先技术领域动态变化情况,帮助决策者选择技术领域和对象开展政策支持工作;为企业自身定位和确定优势领域提供参考,挖掘潜在合作机遇;使得研究所与科研人员,尽早发现技术领域发展新趋势^[1]。

1 文献综述

围绕专利与论文开展新兴技术识别工作,首先,构建相似矩阵,通过聚类筛选出离群专利与论文;其次,利用 LDA 模型将离群专利与

论文记载内容进行技术主题识别,分别生成专利主题短语和论文主题短语;再次,短语相似度测算,主题归并;最后,根据新兴技术特征设计识别指标体系,综合评价确定最终的新兴技术。因此,文献综述主要围绕新兴技术识别方法和新兴技术特征及判别指标体系展开。

在数据统计中与其他群组在数值上相隔较远的观察值被定义为孤立点,集群中弱关系值容易被忽视,这些弱关系值也可能会蕴含潜在开发价值,仅考虑孤立点可能会造成识别结果不准确。因此,本文借鉴罗素平等^[2]思路,将研究对象范围定在特定领域关系网络中的弱关系与孤立点的集合,避免遗漏结果。利用共现矩阵转换相似矩阵,对相似矩阵聚类后选出远离聚类簇的专利与论文作为对象开展研究。

1.1 新兴技术识别方法

新兴技术识别方法可以分为三类:第一类是以专家头脑风暴法为主的定性方法;第二类为以文献计量、文本挖掘发为主的定量方法;第三类为混合型方法。F.M. Tseng 等^[3]采用情景归纳、德尔菲法、技术替代模型来识别新兴技术;谈毅等^[4]则用技术路线图、实物期权来

识别新兴技术。随着新一代通信技术发展和信息不断增长,越来越多学者开展论文、专利等规范化数据定量研究,减少或者避免定性方法主观性较强的缺陷,增强识别方法客观性与可操作性。

1.2 新兴技术特征与识别指标

新兴技术特征:新颖性、根本创新性、相对增长性、社会影响性^[5-6]。新颖性指技术在发展初期,相关文献出现时间越晚越新颖^[7];根本创新性指旧技术转变为新技术发生本质性变化,该变化是判断新旧技术转折点;相对增长性:与同一领域其他技术相比,新兴技术增速明显更快;社会影响性:新兴技术可能会对未来社会经济产生影响。在识别指标方面, Lee C 等^[8]通过专利家族成员数、权利要求项数等多项识别指标来识别新兴技术; Kyebambe M N 等^[9]则加入技术成长周期指标来预测新兴技术;黄鲁成等^[10]利用高关注度、高成长潜力度以及高关联度指标对精密单位定点技术有效识别;宋欣娜等^[11]利用新颖性、持久性、社区性、增长性来跟踪新兴技术。

目前关于新兴技术主题识别研究,仍有不足:第一,定性研究结果主要取决于专家知识,专家不同,结果不同,预测准确性下降。第二,新兴技术指标体系往往只将单个指标计算出来,并未考虑各指标权重大小,导致预测结果可能出现偏差。第三,目前识别研究主要围绕热点专利、论文高频词展开,对与热点相隔较远离群专利、关键词关注度不够,然而这些离群专利与离群词,因其特点与当下主流技术的技术特点相异,内部存在巨大潜力成为新兴技术,

忽视它们存在可能会导致结果不全面。

针对不足,在深入理解新兴技术特征基础上,为进一步提升新兴技术识别的实际效果,完善新兴技术识别方法体系,对专利与论文数据,进行深度挖掘,实现对新兴技术主题有效识别。

2 新兴技术主题识别方法

利用社会网络分析法对相似矩阵进行聚类,筛选出离群专利与论文后,借助 LDA 主题模型处理离群论文与专利文本内容,挖掘技术主题,然后利用短语相似度计算技术将专利与论文技术短语进行归并,避免造成最终识别结果重复。接着根据新兴技术特征构建新兴技术识别指标体系评价归并后的技术主题,完成新兴技术识别。具体识别流程如图 1 所示。

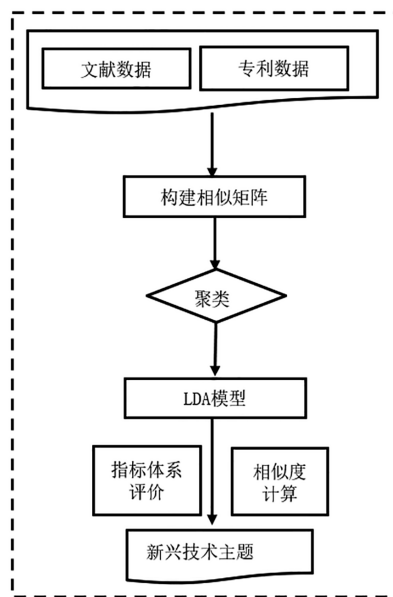


图 1 新兴技术主题识别流程

2.1 新兴技术主题识别模型

(1) 构建相似矩阵

相似性可以通过耦合分析、共被引分析、共分类分析来测度,但由于被引关系需要一定时间存在滞后性问题,耦合分析经常会造成数据缺失问题,而共分类分析方法体系的IPC分类号已被绝大多数国家和地区专利局使用并作为技术分类一种方式,具有权威性、准确性、易获取等优点,因此本文采用共分类分析来构建专利相似矩阵。利用分类号,借助文献计量分析软件构建共现矩阵,再通过系数转换成相似矩阵。同样地,虽然论文关键词由作者自行选定,其表达方式会有差异,但根据其含义也可以进行词形归并,例如“LDA”“LDA模型”就可归并为LDA。通过此办法,利用excel对关键词进行预处理之后,再进行相似矩阵构建。

(2) 筛选离群专利与论文

社会网络里每个人都是一个节点,节点之间连线代表二者关系,该方法最初研究人之间关系,后扩展到研究技术间联系^[12]。利用Ucinet^[13]分别对专利分类号与论文关键词进行聚类,调节阈值,当关系弱于某个阈值时,因为与其他节点关系不紧密而被排出在网络之外,离群分类号与关键词形成。专利文件中分类号有主、次分类号,其中次分类号仅包含该项专利次要技术。因此本文选用最能代表该发明信息IPC主分类号为研究对象,关键词则根据人工判读,剔除与研究无关的关键词,例如“方法”等,形成关键词集合。考虑到分类号与关键词有时会脱离语义环境,因此有必要将离群分类号与关键词对应专利与论文查找出来,将摘要与说明书部分摘取出来,全文无意义的词语太多,而摘要和说明书往往能够准确地代表论文

与专利内容,且篇幅更加精炼,能够提取出较完备的语义内容,即主题短语。

(3) 提取主题短语

通过LDA主题模型处理离群论文摘要部分与专利说明书部分,生成专利主题短语与论文主题短语。LDA常被用于大规模文档集合建模。运用LDA模型开展文本挖掘提取文本主题词是一种较为惯常和科学的做法^[14]。

(4) 相似度计算

为避免识别结果重复,有必要将专利主题短语与论文主题短语进行归并。利用余弦相似度^[15]将论文主题短语与专利主题短语进行相似度计算。相似度在0.95以上,将主题短语进行归并,形成一个主题短语;如果低于0.95,则为不同的两个主题短语。

2.2 新兴技术识别指标体系

根据新兴技术根本创新性、相对增长性、影响性、新颖性等特征,构建识别指标对相关技术主题进行评估。

新兴技术主题新颖性,采用论文平均发表年、专利平均授权年,二者进行比较,选取更早的年份来表征新兴技术主题新颖性。原因是一项新兴技术起点的可能是从理论研究开始或先从应用研究出发,而论文是主要理论研究成果,专利是主要应用研究成果,因此需要考虑论文与专利出现时间早晚,并进行比较。为了方便计算,将其标准化,公式为: $1 / (\text{计算年} - \text{研究主题出现年} + 1)$,其中计算年为2020年,研究出现年为论文平均发表年/专利平均授权年。

新兴技术主题根本创新性由其主题下的专利权利要求平均数来表征^[8]。新兴技术主题影

响性可以参考 Lee C 等人做法^[7]通过专利家族成员数来表征。

此外还应该考虑新兴技术关注人数情况，关注人数越来越多情况下，其规模也会越来越

壮大，该技术越有可能成长为真正新兴技术。
综上所述，新兴技术判别指标体系包括：新颖性指标、相对增长性指标、根本创新性与影响性指标、关注度指标。具体内容见表 1。

表 1 新兴技术主题识别指标

指标	计算规则	含义
新颖性	$1 / (\text{计算年} - \text{研究主题出现年} + 1)$	年份越晚技术，更可能是新兴技术
相对增长性	专利与文献数据分别曲线拟合，得到斜率 A、B，写作 (A,B)	研究在逐年增长技术，更可能是新兴技术
根本创新性	权利要求总数 T/件数 N	创新性越强技术，更可能是新兴技术
影响性	专利家族成员总数 P/件数 N	保护布局范围越广，反映出影响力越大，更可能是新兴技术
关注度	$(\text{论文平均关注人数} Q + \text{专利平均关注人数} M) * 0.5$	关注度越高技术，更可能是新兴技术

3 实验数据及结果分析

车联网是我国战略性新兴产业之一^[16]。本文以该领域为例，验证方法的有效性，同时识别出该领域新兴技术主题，为政府决策、企业战略、科学研究等提供参考。

3.1 数据采集与描述性统计

(1) 数据采集
专利数据以 Incopat 专利数据库作为数据来源，以车联网领域中国专利为研究对象，检索时间为 2020 年 11 月 01 日。检索词为车联网、车辆物联网、车辆网联、汽车通信；检索方式为标题检索；申请人国别选择中国；专利申请日选择 2008 年 1 月 1 日 -2019 年 12 月 31 日。考虑发明专利一般比实用新型与外观设计的实质审核周期长，其质量也更高，更能代表中国车联网当前专利情况。因此选择已授权发明专利或者申请发明专利，去除与领域不相干专利，

共得 2422 条有效数据。
论文数据以中国期刊全文数据库（中国知网）作为数据来源；以车联网领域核心期刊、学位论文、会议论文为研究对象。检索时间为 2020 年 11 月 01 日，检索词为车联网、车辆物联网、车辆网联、汽车通信，检索方式为主题检索，时间范围限定 2008 年 1 月 1 日至 2019 年 12 月 31 日，检索时间为 2020 年 11 月 01 日。精确检索，去除会议通知、消息、序言等记录，共获取 387 篇核心期刊论文，1412 篇硕博论文与 146 篇会议论文。

(2) 描述性统计
通过统计可看出该领域存在大量低热度关键词与专利分类号，值得深入研究。
论文关键词分布区间见表 2，50 以上数量的关键词有 4 枚，10~50 数量之间关键词有 59 枚，构成论文领域关键词“头部”；数量在 1~3 之间关键词达 4584 枚，这部分关键词内容丰富，占比较高，构成论文领域关键词“长尾”，是可以发掘新兴技术主题的重要数据源。

表 2 论文关键词分布情况

词频数量区间	关键词数	关键词展示（部分）
1000（不含）~	1	智能网联
50（不含）~	3	物联网；智能交通；车载自组织网络
40（不含）~10（含）	59	智能交通；隐私保护；大数据；路由协议等
10（不含）~06（含）	12	MAC协议；路径规划；信息中心网络；模糊逻辑；D2D通信等
06（不含）~04（含）	66	无人驾驶；移动云计算；嵌入式系统；粒子群算法等
00（不含）~03（含）	4584	疲劳驾驶；环签名；深度神经网络等

专利分类号分布区间见表 3，数量达 500 以上的分类号仅 1 枚，数量 50 以上的分类号有 12 枚，数量 10~50 之间的分类号有 37 枚，数量在 08~10 之间的分类号有 104 枚，构成专利

领域关键词“头部”；数量在 1~3 之间的分类号达到 984 枚，这部分分类号的内容丰富，占比比较高，构成专利领域分类号“长尾”，是可以发掘新兴技术主题的重要数据源。

表 3 专利分类号分布情况

数量区间	分类号数	分类号展示（部分）
500（不含）-	1	H04L29/08
300（不含）-50（含）	12	H04L29/08;H04W4/44;H04W4/40;H04L29/06;H04W4/40等
50（不含）-10（含）	37	H04W12/06;G08G1/0967;G08G1/00;H04W4/029等
10（不含）-08（含）	104	G08G1/0962;G06K9/62;G06Q10/02;G06F17/50等
08（不含）-04（含）	64	H04W72/12;H04W12/02;H04W76/02;G08B21/06等
04（不含）-01（含）	984	H04W72/00;H04W40/24;G07F15/00;G06Q50/30等

综上所述，本文设计新兴技术主题识别方法对车联网领域可行。

再进行聚类可视化。阈值为 0.4，结果较为清楚，共筛选出 97 个关系较弱节点和孤立节点。

3.2 实验结果及分析

- (1) 专利分类号相似矩阵
- 借助 Bib Excel 软件^[17]构建分类号共现矩阵，通过 Ochiai^[18]再转换成相似矩阵，结果如表 4。
- (2) 离群专利识别
- 将表 4 相似矩阵（660×660）导入 Ucinet，
- 结果如图 2 所示，其中有三个大聚类簇，包含多个分类号，簇 1 主分类号为 H01Q5/00，含义为使天线同时工作在两个或两个以上不同波段装置，如双频或多频装置；簇 2 主分类号为 F03D9/25，含义为驱动装置为电能产生装置；簇 3 主分类号为 G06F3/01，含义为用于用户和计算机之间交互输入装置或输入和输出组合装置。

表 4 相似矩阵结果（专利部分）

	H04L29/08	H04W4/44	H04W4/40	H04L29/06	H04W4/46	G08G1/01	...	H04W72/04
H04L29/08	1	0.1361	0.121	0.1776	0.1071	0.0733	...	0.0444
H04W4/44	0.1361	1	0.0824	0.0696	0.301	0.0366	...	0.0416
H04W4/40	0.121	0.0824	1	0.0555	0.1113	0.075	...	0.0852
H04L29/06	0.1776	0.069	0.0555	1	0.0385	0.0105	...	0
H04W4/46	0.1071	0.301	0.1113	0.0385	1	0.0342	...	0.0777
G08G1/01	0.0733	0.0366	0.075	0.0105	0.0342	1	...	0
...	...	...	...	...	...	...	...	...
H04W72/04	0.0444	0.0416	0.0852	0	0.0777	0	...	1

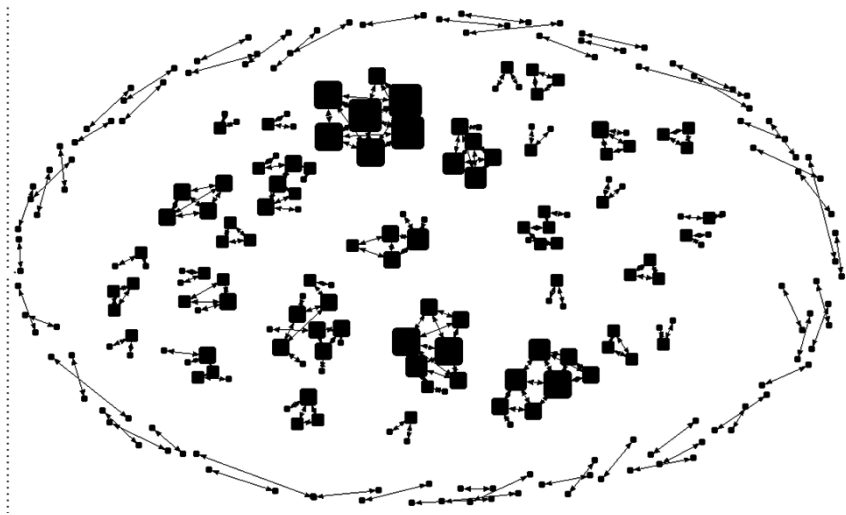


图 2 聚类簇和分离群分类号

(3) 论文相似矩阵

关键词，构建相似矩阵，见表 5。
 与专利分类号处理过程相同，利用论文关

表 5 相似矩阵结果（论文部分）

	智能交通	隐私保护	大数据	车载自组织网络	...	资源分配
智能交通	1	0	0	0.023	...	0
隐私保护	0	1	0	0.046	...	0
大数据	0	0	1	0.024	...	0
车载自组织网络	0.023	0.046	0.024	1	...	0.0267
...	...	...	...	...	...	...
资源分配	0	0	0	0.0267	...	1

注：短语相似度计算结果，小数点后有长达 10 位数之多，这里只保留前四位。

(4) 离群论文识别

将表5相似矩阵(1041×1041)导入Ucinet,进行聚类可视化。阈值调为0.4,结果较为清楚,共筛选出107个关系较弱节点和孤立节点。

图3展示了阈值为0.4时所有聚类簇和部分离群关键词,其中有六个大聚类簇:簇1关键词为机动车保险、差异化市场费率、保险公司、车联网,含义为车联网环境下车险定价策

略;簇2关键词为车道时分复用、实时优化、快速公交,含义为车联网信道传输优化技术;簇3关键词为数字签名、可信计算、远程证明,含义为车联网安全路由技术;簇4关键词为数字交换、区块链技术、去中心化,含义为车联网数据交换系统设计;簇5关键词为动态网组、组播通信模型、三角网,含义为车联网通信模型;簇6关键词为交叉群口、仿真、速度引导,含义为车联网仿真系统设计。

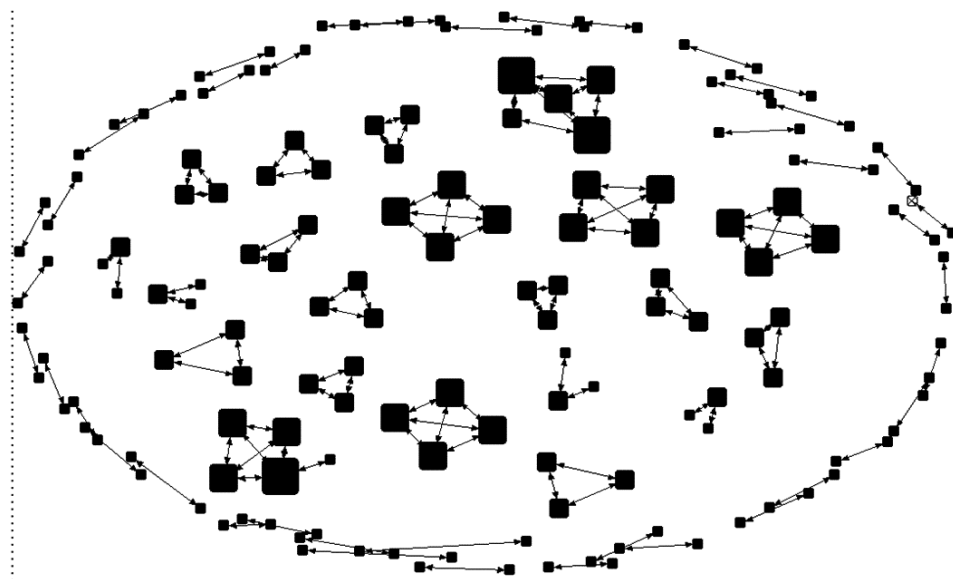


图3 聚类簇和部分离群关键词

(5) 提取主题短语

利用97个关系较弱或者孤立专利分类号节点和对应专利文献,利用LDA主题抽取,生成专利主题短语;将107个关系较弱或者孤立论文关键词含义和论文文献记载内容进行LDA主题抽取,生成论文主题短语。部分结果见表6。

(6) 相似度计算

利用余弦相似度^[14]计算论文主题短语与专利主题短语相似度,见表7。

表6 部分专利论文主题短语

专利主题短语	论文主题短语
车联网频谱感知方法	车联网入侵检测技术
车联网连通性方法	区块链车联网安全通信技术
直连通信方法	V2V通信技术
车联网匿名认证技术	D2D通信技术
车联网信息安全方法	DOA估计方法
车联网拥塞控制方法	车联网移动云系统
...	...
雾计算	车联网大数据处理系统

表 7 相似度计算（部分）

	边缘计算技术	驾驶行为分析技术	车联网定位技术	车联网导航技术	...	车联网云平台系统
边缘计算技术	1	0.106	0.2619	0.2619	...	0
驾驶行为分析技术	0.106	1	0.4805	0.4805	...	0.3238
车联网定位技术	0.2619	0.4805	1	0.7142	...	0.4017
车联网导航技术	0.2619	0.4805	0.7142	1	...	0.4017
...	...	...	...	...	...	...
车联网云平台系统	0	0.3238	0.4017	0.4017	...	1

注：短语相似度计算结果，小数点后有长达 10 位数之多，这里只保留前四位。

(7) 备选主题短语集合

为避免相同主题在计算指标时重复，节省时间与精力，在对主题短语综合评价之前，将相似度水平为 0.95 以上的主题词进行归并，形成一个主题短语；反之为两个主题短语。结果见表 8。

表 8 备选主题短语集合

主题短语名称	主题短语名称
无人驾驶	车联网隐私保护方法
车联网边缘计算技术	V2V通信技术
雾计算	群智感知车联网
中继节点选择方法	自动驾驶
车联网信息管理系统	云计算车联网
车联网协同控制方法	车联网连通性方法

3.3 新兴技术主题短语评价

为进一步确定上述主题短语是否为真正的新兴技术主题，还需结合新兴技术指标体系来确定。

新兴技术主题短语新颖性通过比较论文、专利文献出现平均时间，取更早年份来确定。为了方便后续综合计算，将新颖性标准化，参

考范少萍等人^[19]做法，公式为： $1/(\text{计算年} - \text{研究主题出现年} + 1)$ ，例如：无人驾驶主题出现年为 2016 年，计算年为 2020 年，那么标准化后无人驾驶主题新颖性为： $1/(2020 - 2016 + 1) = 0.2$ 。数值越接近 1，新颖性越强。

相对增长性则按照新兴技术主题短语历年专利数量与论文数量变化情况进行曲线拟合，若曲线预测斜率 k 为正值，则比较大小，若为负值，则直接从备选集合中踢出，例如，备选主题短语车联网连通性方法，专利部分预测增长斜率为 0.1，论文部分预测增长斜率为 0，都为正值，保留结果，写作 (0.1,0)，见图 4。

新兴技术主题短语影响性、根本创新性、关注度，按照表 1 计算规则分别算出结果。最终各项指标的测度结果见表 9。

本文采用 CRITIC 法确定指标权重，CRITIC 法是客观赋权法^[20]。指标权重结果见表 10。需要说明的是，相对增长性有 (A,B) 两个值，为方便计算，将 A 与 B 值相加，取平均数，作为计算综合指数里面的相对增长性值。CRITIC 法得到各指标权重，再与对应指标值相乘，接着相加，得到综合指数。

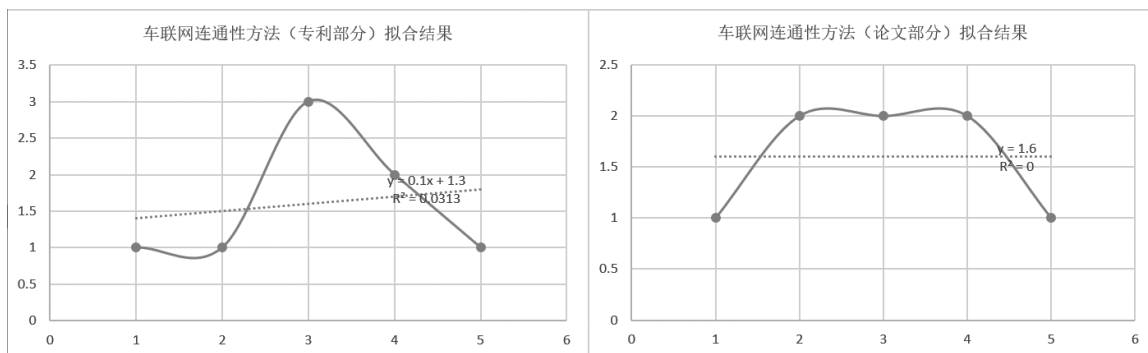


图4 相对增长性曲线拟合结果示例

表9 主题短语测各项指标测度结果

主题短语名称	根本创新性	影响性	关注度	新颖性	相对增长性
无人驾驶	8.76	1.23	2.43	0.20	1.50
群智感知车联网	8	1	2.90	0.17	0.30
车联网边缘计算技术	6.2	1.06	2.71	0.25	6.25
雾计算	6	1.2	2.46	0.20	1.40
云计算车联网	6	1.6	1.93	0.11	0.20
中继节点选择方法	7	1.5	3.50	0.13	0.25
车联网信息管理系统	8	1.5	2.63	0.11	0.00
车联网隐私保护方法	5.55	1.5	3.33	0.11	1.48
V2V通信技术	6.37	1.12	2.77	0.17	1.50
自动驾驶	7.54	1.18	2.73	0.14	1.50
车联网连通性方法	3.13	1.62	5.44	0.11	0.05
车联网协同控制方法	8	1.33	3.40	0.11	0.05

如无人驾驶主题各指标值如下：根本创新性（8.76）、影响性（1.23）、关注度（2.43）、取平均数后相对增长性（1.50）、标准化后新

颖性（0.2），综合指数为 $8.76 \times 0.43 + 1.23 \times 0.05 + 2.43 \times 0.20 + 1.5 \times 0.31 + 0.2 \times 0.01 = 4.769829$ 。按照综合指数对备选主题短语进行排序，见表11。

表10 CRITIC 法权重计算结果

指标名称	指标变异性	指标冲突性	信息量	权重
根本创新性	1.702	5.166	8.792	0.43
影响性	0.205	5.263	1.081	0.05
关注度	0.849	4.734	4.018	0.20
相对增长性	1.611	3.994	6.433	0.31
新颖性	0.043	4.174	0.179	0.01

注：为方便计算，权重值保留两位小数。

表 11 车联网领域相关指标值排名情况

主题短语名称	综合指数	排名	主题短语名称	综合指数	排名
车联网边缘计算技术	5.208027	1	中继节点选择方法	3.8461875	7
无人驾驶	4.769829	2	V2V通信技术	3.80555	8
自动驾驶	4.301380857	3	车联网隐私保护方法	3.57696067	9
车联网协同控制方法	4.182567667	4	雾计算	3.55828	10
群智感知车联网	4.14709	5	云计算车联网	3.09814667	11
车联网信息管理系统	4.02491667	6	车联网连通性方法	2.50943467	12

由表 10 可知，CRITIC 法客观赋权结果中，根本创新性占综合指数 43%，相对增长性占综合指数 31%，关注度占综合指数 20%，影响性和新颖性占比为 5% 与 1%。因此，判断该技术是否为新兴技术关键指标为根本创新性、相对增长性与关注度，即关注度越高、相对增长性越高、根本创新性越强技术更可能成为新兴技术。

将综合指数全部相加取平均数得 3.919030892，考虑到篇幅所限，本文仅进一步研究综合指数在平均水平之上的技术主题。综合指数在平均水平之上的技术主题有：车联网边缘计算技术、无人驾驶、自动驾驶、车联网协同控制方法、群智感知车联网、车联网信息管理系统。按照综合指数排名顺序，依次对这些技术主题做进一步分析。

（1）边缘计算技术

由表 9 可知，边缘计算技术相对增长性高达 6.25，排在第 1 位，关注度排在中等位置，但其根本创新性仅为 6.2，位于待分析的技术主题的末位，不符合新兴技术本质特征：根本创新性。因此排除。

（2）无人驾驶技术

由表 9 可知，无人驾驶技术根本创新性高

达 8.76，排在第 1 位，其相对增长性为 1.5，排在第 2 位，且其综合指数也排在第 2 位，符合新兴技术根本创新性、相对增长性特征，虽然关注度排名不够靠前，随着时间推移，无人驾驶技术逐渐兴起会吸引更多人员，关注度也会随之提高。因此无人驾驶技术为车联网识别领域新兴技术。

（3）自动驾驶技术

无人驾驶处于驾驶最高级别，无人驾驶等于完全自动驾驶，美国汽车工程师协会将自动驾驶技术进行了分级，L0 为属于传统驾驶，L1 和 L2 属于驾驶辅助，L3-L5 属于自动驾驶，L5 自动驾驶技术等级也称为无人驾驶。前文已经论证了无人驾驶技术为车联网识别领域新兴技术，为了更加准确描述新兴技术，本文认为自动驾驶技术不是车联网识别领域新兴技术，自动驾驶技术高级阶段即无人驾驶技术为车联网识别领域新兴技术。

（4）车辆协同控制方法

由表 9 可知，车辆协同控制方法根本创新性高达 8，排在第 2 位，其关注度也排在第 3 位，名次靠前，但其相对增长性仅为 0.05，远低于同领域其他技术相对增长速度，不符合新兴技术相对增长性特征。所以排除。

(5) 群智感知技术

由表9可知,群智感知技术根本创新性为8,排在第2位,关注度排在中等位置,但其相对增长性仅为0.3,不符合新兴技术相对增长性特征。因此排除。

(6) 车联网信息管理系统

由表9可知,车联网信息管理系统根本创新性为8,排在第2位,关注度也排在前列,但其相对增长性为0,完全不符合新兴技术相对增长性特征。因此排除。

综上所述,本文确定无人驾驶技术为车联网领域新兴技术。

3.4 新兴技术主题解读

无人驾驶能够有效避免因驾驶技能、心理变化、疲劳程度等人为因素而导致交通事故,有助于合理管控道路交通流量以改善道路通行能力,具有广阔应用前景以及潜在社会效益。2020年8月召开中国车联网大会暨第十五届中国卫星导航运营商会,业内专家认为无人驾驶技术是未来很有前景的技术,但还存在诸多问题需要攻克。例如无人驾驶对感知传感器精度依赖性较强,且在非常态环境下,安全性无法100%保证,因此需要攻克其核心技术激光雷达传感器等。业界分析与识别结果保持一致,说明新兴技术识别方法有效性。

4 结论与不足

为避免单纯用专家评估而带来过强的主观性和目前新兴技术主题识别研究对离群数据源关注度不够的问题,以及新兴技术主题识别指

标体系的不足之处,提出一种离群专利与论文新兴技术主题识别方法。通过构建相似矩阵,聚类筛选出专利与论文中存在孤立节点与弱关系节点,再通过LDA主题模型抽取技术主题,最后围绕新兴技术本质特征,设计指标体系,利用CRITIC法计算各指标权重之后,计算综合指数,增强了评估方法的客观性。选取车联网这一新兴领域为实证对象,为国家、企业以及相关科研人员与科研机构,提供了情报支撑。

本研究只针对新兴技术根本创新性、影响性、相对增长性等特征来综合构建评价指标体系,未考虑对新兴技术主题不确定性进行量化。如何测度新兴技术主题不确定性将成为下一步重点研究内容。

参考文献

- [1] 刘小玲,谭宗颖.新兴技术主题识别方法研究进展[J].图书情报工作,2020,64(11):145-152.
- [2] 罗素平,寇翠翠,金金,等.基于离群专利颠覆性技术预测——以中药专利为例[J].情报理论与实践,2019,42(7):165-170.
- [3] Tseng F M, Cheng A C, Peng Y N. Assessing market penetration combining scenario analysis, Delphi, and the technological substitution model: the case of the OLED TV market[J]. Technological forecasting and social change, 2009, 76(7): 897-909.
- [4] 谈毅,黄燕丽.基于过程新兴技术规划与选择模型研究[J].科技管理研究,2007(8):5-8.
- [5] 周萌,朱相丽.新兴技术概念辨析及其识别方法研究进展[J].情报理论与实践,2019,42(10):162-169.
- [6] 徐建国,李孟军,游翰霖.新兴技术识别研究进展[J].情报杂志,2018,37(12):8-12+7.
- [7] Wang Q. A bibliometric model for identifying emerging research topics[J]. Journal of the Association for Information Science and Technology, 2018,

- 69(2):290-304.
- [8] Lee C, Kwon O, Kim M, et al. Early identification of emerging technologies: A machine learning approach using multiple patent indicators[J]. *Technological Forecasting & Social Change*, 2018(127):291-303.
- [9] Kyebambe M N, Cheng G, Huang Y, et al. Forecasting emerging technologies: A supervised learning approach through patent analysis[J]. *Technological Forecasting & Social Change*, 2017(125): 236-244.
- [10] 黄鲁成, 唐月强, 吴菲菲, 等. 基于文献多属性测度新兴主题识别方法研究 [J]. *科学学与科学技术管理*, 2015, 36(2): 34-43.
- [11] 宋欣娜, 郭颖, 席笑文. 基于专利文献多指标新兴技术识别研究 [J]. *情报杂志*, 2020, 39(6):76-81+88.
- [12] 张克群, 魏晓辉, 郝娟, 等. 基于社会网络分析方法专利价值影响因素研究 [J]. *科学学与科学技术管理*, 2016, 37(5):67-74.
- [13] 刘光阳, 李根. Ucinet 国内学术贡献计量与图谱分析——基于 CNKI 中文期刊 2004-2017 年数据 [J]. *图书情报工作*, 2017, 61(S2):88-98.
- [14] 胡吉明, 陈果. 基于动态 LDA 主题模型内容主题挖掘与演化 [J]. *图书情报工作*, 2014, 58(2):138-142.
- [15] 王春柳, 杨永辉, 邓霏, 等. 文本相似度计算方法研究综述 [J]. *情报科学*, 2019, 37(3): 158-168.
- [16] 田波, 刘鹏远, 郑羽莎, 等. 基于 ANP 方法车联网信息平台风险识别与评价研究 [J]. *情报科学*, 2018, 36(6):100-104+155.
- [17] Persson O, Danell R, Schneider J W. How to use Bibexcel for Various Types of Bibliometric Analysis. In *Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at his 60th birthday* [M]. Leuven: International Society for Scientometrics and Informetrics, 2009.
- [18] 李海林, 邬先利. 基于时间序列聚类主题发现与演化分析研究 [J]. *情报学报*, 2019, 38(10): 1041-1050.
- [19] 范少萍, 安新颖, 倪萍. 基于文献研究前沿特征与识别方法研究 [J]. *情报工程*, 2017, 3(1):56-65.
- [20] 王昆, 宋海洲. 三种客观权重赋权法比较分析 [J]. *技术经济与管理研究*, 2003(6):48-49.