

# 基于主题突变检测的颠覆性技术识别 ——以无人机技术领域为例

刘忠宝, 康嘉琦, 张静

中北大学软件学院, 太原 030051

**摘要** 颠覆性技术对实现中国科技创新跨越式发展具有重要战略意义。以无人机技术领域为例, 获取 Web of Science 论文数据库和德温特(Derwent)专利数据库 2005—2019 年收录的 2812 篇论文与专利数据, 基于 LDA-LSTM 文本分类算法得到技术主题, 采用 CiteSpace 构建共现网络, 从突变权重排序及突变时间段和突变共词聚类知识图谱 2 个角度进行主题突变检测, 进而识别出无人机交互技术中的脑机接口技术及手势控制技术为该领域内的颠覆性技术, 通过 2020 年无人机技术领域相关论文专利验证了识别框架的有效性。

**关键词** 颠覆性技术识别; 文本分类; 主题突变检测

2008 年国际金融危机以来, 世界主要国家对技术预见高度重视, 颠覆性技术成为各国战略布局的重点<sup>[1]</sup>。2016 年, “颠覆性技术”被写入中国《国家创新驱动发展战略纲要》和《“十三五”国家科技创新规划》。2017 年, 党的十九大报告提出, 要“突出颠覆性技术创新”。2018 年, 习近平总书记在两院院士大会上指出以颠覆性技术创新为突破口。在新形势新起点下, 中国力争在更多战略性基础科学领域实现率先突破, 在关键核心技术领域取得颠覆性创新, 力求使中国在新一轮产业变革中赢得竞争优势。但由于颠覆性技术本身发展具有高度不

确定性, 任何一项技术在成为颠覆性技术之前很长一段时间内不被大众所关注, 因此颠覆性技术的识别和预测一直是学界的重点和难点。鉴于此, 探索颠覆性技术识别方法, 识别该领域内可能存在的颠覆性技术, 具有至关重要的意义。

## 1 相关研究

颠覆性技术(disruptive technology)的概念由 Bower 和 Christensen 在 20 世纪 90 年代提出, 它是一种另辟蹊径、会对已有传统或主流技术产生颠覆性

收稿日期: 2020-04-28; 修回日期: 2020-05-27

基金项目: 国家社会科学基金一般项目(19BTQ012)

作者简介: 刘忠宝, 教授, 研究方向为数据挖掘、机器学习、信息资源推荐, 电子邮箱: liuzb@nuc.edu.cn; 康嘉琦(共同第一作者), 硕士研究生, 研究方向为数据挖掘、智能信息处理

引用格式: 刘忠宝, 康嘉琦, 张静. 基于主题突变检测的颠覆性技术识别——以无人机技术领域为例[J]. 科技导报, 2020, 38(20): 97-105; doi: 10.3981/j.issn.1000-7857.2020.20.015

效果的技术,它能重新配置价值体系,并引领全新的产品和服务<sup>[2]</sup>。颠覆性技术识别目的是有效地对处于潜伏期的技术进行预测。系统梳理国内外关于颠覆性技术识别与预测方法的研究文献,发现颠覆性技术预测识别的主要方法有以下4种。

第1种是专家评议法。该方法通常基于专家经验的德尔菲法与技术路线图,例如 Vojak 等<sup>[3]</sup>分析技术创新的规律,然后绘制出技术路线图,帮助决策者确定可能的颠覆性技术;也有学者利用德尔菲法识别具有潜在性的颠覆性技术<sup>[4-6]</sup>。但这些方法在预测过程中需要大量的人力物力,并且很大程度上依赖于专家的个人经验,具有较强的局限性。

第2种为技术演化法。例如,许泽浩<sup>[7]</sup>以新能源汽车为例,运用发明问题与解决理论(TRIZ)的技术成熟度分析工具,采用4个尺度变量数据对汽车产业进行技术成熟度分析,从而判别作为汽车行业颠覆性创新的电动汽车培育的最佳时期。该方法利用进化法则和产品进化S曲线,可以分析确认技术系统的发展阶段,预测未来发展趋势,但不能准确地描述技术演化的完整过程,从而难以识别技术突变。

第3种为模型分析法。Nagy 等<sup>[8]</sup>构建基于功能、技术标准、所有权的“3步走”模型,以3D打印技术为例,验证了模型的有效性;黄鲁成等<sup>[9]</sup>运用物种入侵模型及属性文本集进行分析,对评估技术方案出现前后属性集的相似性和颠覆性进行测度,该方法是在颠覆性技术成长期进行识别。基于模型的分析法相较于专家评议法简单易操作,预测周期短,但其用于模型预测的数据获取与处理较为复杂,影响结果准确性。

第4种是基于专利、论文的分析法。近些年有许多学者尝试以专利或科学论文作为数据源,按照一定的评判思想和指标,在尽可能早的时期从客观角度识别颠覆性技术。例如, Buchanan 等<sup>[10]</sup>、Momeni 等<sup>[11]</sup>以专利分析为基础,借助数据挖掘方法来识别颠覆性技术;罗素平等<sup>[12]</sup>以中药专利为例,构建专利相似矩阵,经聚类后筛选出离群专利,根据离群专利主分类号划分技术组合,从知识关联性、

技术潜力和市场潜力3个维度评估技术组合,最后运用孔多塞投票法预测颠覆性技术。苏敬勤等<sup>[13]</sup>从专利数量、引用量、引用率3个指标绘制技术演变趋势曲线图,并以智能手机为例,对专利影响因子指标预测早期颠覆性技术。

目前基于专利、论文的分析识别方法主要集中于热点技术的研究,在某种程度上具有一定优势。论文、专利等数据承载了相关技术领域科学研究和技术创新活动的大量信息,其中专利包含丰富的技术信息,且专利数据规范、可靠,从专利信息中可以把握技术特征及演化趋势<sup>[14]</sup>。不同的文献类型可以表示技术发展的不同阶段,而且可以反映该领域学术研究的理论水平和发展速度。因而对论文、专利等数据进行分析,可以尽可能早地识别颠覆性技术。因此,本研究以论文、专利为数据源,使用主题模型分类算法得到各技术类别主题及所对应的关键词,基于 CiteSpace 的突变检测功能,运用突变检测算法识别无人机技术领域内的颠覆性技术。

## 2 研究思路

### 2.1 数据来源

以 Web of Science(WoS)论文数据库与德温特(Derwent)专利数据库为数据来源,检索时间范围是2005年1月1日—2019年12月31日,检索方式为(主题=“Disruptive drone technology”OR 主题=“Disruptive innovation of drones”),得到与颠覆性技术研究相关的文献2935篇,除去通知、要目、指南等,共计得到有效文献专利2812篇。转换 Wos 和 Derwent 数据库下载格式并以相应的形式保存,以便在 CiteSpace 软件中处理。

### 2.2 研究框架

基于现有研究的基础上,提出颠覆性技术识别框架:(1)确定颠覆性技术研究领域,对该技术领域的论文及专利摘要数据进行检索、获取与特征提取,并划分训练集和测试集;(2)使用主题模型分类算法对获取的数据摘要分类,选择分类准确率较高的分类器,进而得到各技术类别主题及所对应的关键词;(3)选取技术主题构建共现网络实现主题

突变监测与分析,进而得到某一领域内颠覆性技术。最终构成主题分类—突变检测—技术识别的

流程框架(图1)。

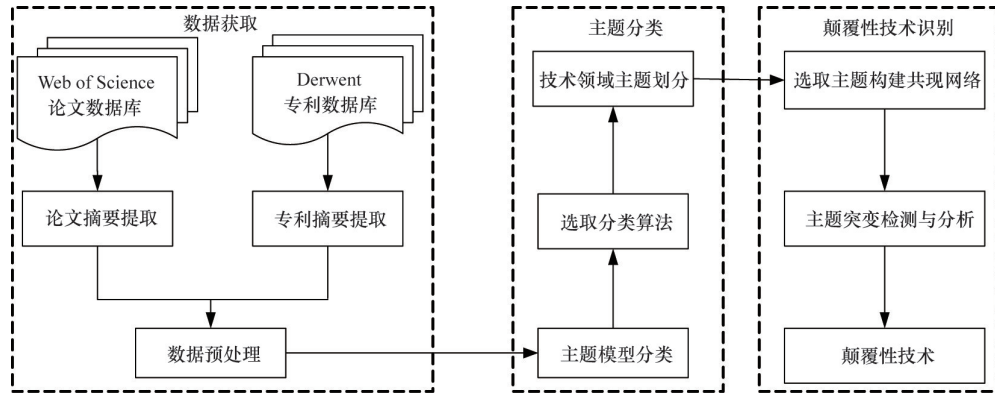


图1 颠覆性技术识别框架

### 2.2.1 主题模型分类算法

采用机器学习分类算法,对论文专利摘要进行分类。文本分类即依照一定的分类标准或体系对文本进行自动分类,即通过对已分类数据的分析,总结出划分各类别的规则,并以此为基础识别其他更多数据所属的类别,常用的有传统机器学习分类算法如 K 近邻算法<sup>[15]</sup>(K-nearest neighbor, KNN)、支持向量机<sup>[16]</sup>(support vector machine, SVM)、深度学习算法如卷积神经网络<sup>[17]</sup>(convolutional neural

networks, CNN)、长短时记忆神经网络<sup>[18]</sup>(long short-term memory network, LSTM)等。以上分类算法均采用基于 LDA(latent dirichlet allocation)主题模型的特征提取方法构建文本分类器,形成 LDA-KNN、LDA-SVM、LDA-CNN、LDA-LSTM 分类器,对论文专利摘要数据集进行分类,分类框架如图2所示。最终表明 LDA-LSTM 分类性能最好,对于待分类数据分类流程图如图3所示。

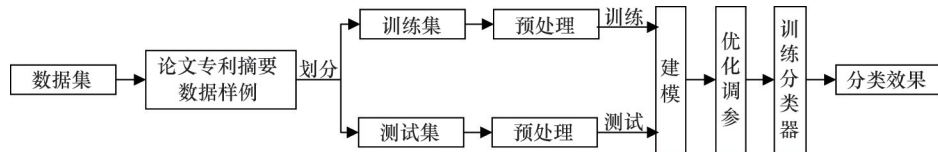


图2 机器学习分类主题模型流程图

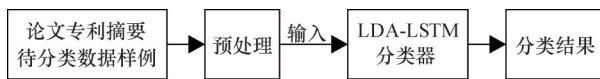


图3 LDA-LSTM 分类流程图

1) LDA 主题模型。隐含狄利克雷分布(latent Dirichlet allocation, LDA)是 Blei 等<sup>[19]</sup>于 2003 年提出的基于概率模型的主题模型算法,是一种无监督的文本聚类方法。LDA 主题模型是包含文档-主题-词的 3 层贝叶斯结构。其基本思想是把文档

看成其隐含主题的混合,而每个主题则表现为跟该主题相关的词项的概率分布。该方法通过对语料库的建模训练,从而发现文档中潜在隐藏的主题信息,LDA 主题模型如图4所示。

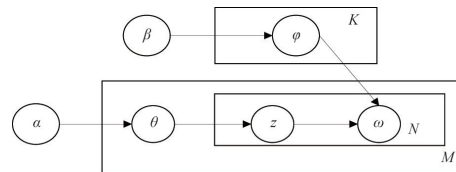


图4 LDA 主题模型

其中,  $M$  为文章数量;  $N$  是文章中词语总数;  $K$  为主题个数;  $\varphi$  是主题上的词分布;  $\theta$  是文章的主题分布, 表示文档中每个隐含主题所占的比重。  $z$  是每次生成文档词  $w$  时被选择的主题;  $w$  是一个可观测变量。  $\alpha$  和  $\beta$  是 2 个超参数, 作为 Dirichlet 分布的先验因子,  $\alpha$  表示每篇文档的潜在隐含主题在文档中的 Dirichlet 分布;  $\beta$  表示潜在隐含主题在每篇文档的关键词中的 Dirichlet 分布。

2) LDA-LSTM 模型。长短时记忆神经网络是循环神经网络模型的一种改进结构, 能够充分利用上下文信息并抓取全局特征, 学习长短依赖关系, 缓解梯度消失的问题。选择分类性能最高的

LDA-LSTM 模型对所有数据进行分类, 其结构如图 5 所示。包括文本输入层、特征提取层、特征融合层、文本分类层以及最后的输出层。(1) 将输入文本序列转化为向量化表示, 得到的文本表示向量作为下一层的输入; (2) 特征提取层中使用 LSTM 神经网络和 LDA 主题模型对输入文本进行局部特征提取和全局特征提取, 提取不同层次、具有不同含义的特征向量; (3) 将全局特征信息和局部特征信息进行融合, 进而实现多层次提取文本信息; (4) 经特征融合后的向量信息输入文本分类器中, 进行分类操作; (5) 输出模型分类结果。

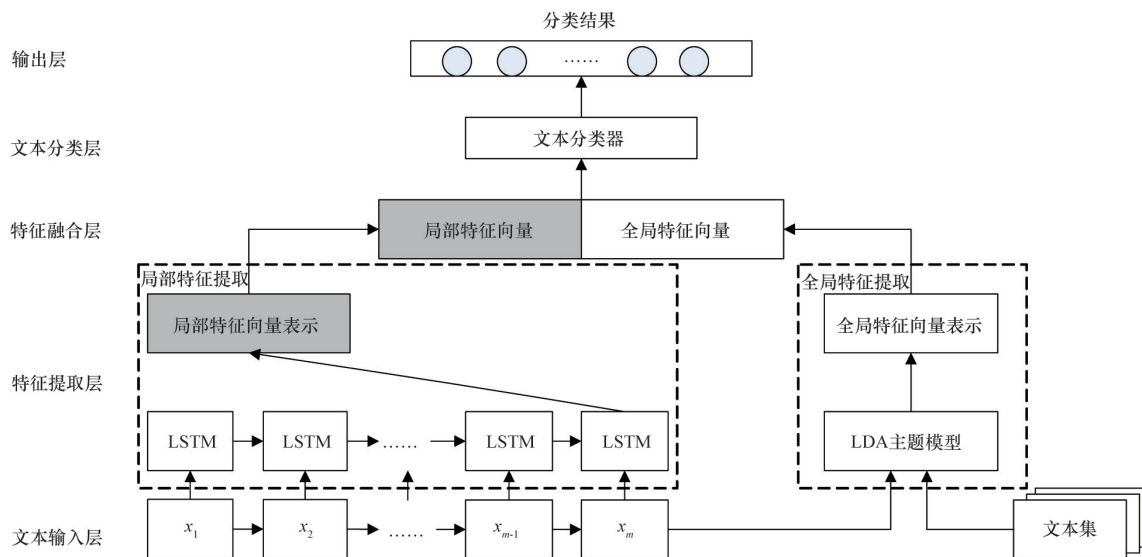


图 5 LDA-LSTM 分类主题模型

LDA-LSTM 模型中的参数设置为: LDA 主题模型,  $\alpha$  反应文本中隐含主题的先验分布为  $50/K$ ,  $\beta$  反应隐含主题下词的先验分布为 0.01, Nites Gibbs 为模型迭代次数设为 1000,  $K$  为主题维数 200。LSTM 神经网络中, LSTM 隐藏层单元数目设置为 128, 损失函数采用 binary\_crossentropy, 迭代次数 epoch 设置为 10, 激活函数采用 Sigmoid。

### 2.2.2 主题突变检测法

Kleinberg 于 2002 提出基于词频增长率的突变检测法, 该法可用于检测某一学科内研究兴趣突变性的增长, 根据词语突发状态进行词语排列, 并确定突发状态的持续时间, 从而揭示词语现在是否处

于突发状态, 对词的发展分析具有动态性和历史性<sup>[20]</sup>。突变检测是根据词在时间顺序的阶段性发展, 挖掘突发词突变率的变化, 进而发现那些低频且具有意义的词<sup>[21]</sup>。该方法可以对所有的词进行突发分析, 从关注词自身的发展变化出发, 关注单个词发展的阶段性。突变检测旨在关注低频但变化率较高的词, 即相对增长率突然增长的词, 这类增长势头不断加强的词在揭示研究领域的发展上更具及时性和情报价值。

突变词是以文献关键词为分析对象, 根据词频年度分布情况, 利用时间序列中突变点识别方法检测出发生突变的时间点, 那么在发生突变的当年该

关键词就称为突变关键词。突发词会根据词频阈值监测热点出发点不同,通过词汇突发性,从而掌握领域发展及时性,并且突变词所代表的主题有可能是研究领域未来主题演化的方向所在。在对某一领域文献的研究中,突变词的突变权重大小和突变发生时间段可以帮助分析该领域研究热点的兴起、发展及消退过程。找到处于上升阶段的新型突变术语有助于发现推动学科发展中的微观因素,可以揭示学科领域的新趋势。

Chen<sup>[22]</sup>认为趋势上升的突变词能够帮助了解研究领域的新趋势和动态发展,遂将 Kleinberg 的突变检测算法思想加入科学图谱软件 CiteSpace。本文利用 CiteSpace 的突变检测,从突变权重排序及突变时间段和突变共词聚类知识图谱进行分析,进而得到领域内的颠覆性技术。

### 3 实验过程

#### 3.1 主题分类实验

以收集 2005—2019 年论文专利数量共计 2812 篇为实验数据集,再随机不重复地选取分类数据与待分类数据。得到分类数据 704 篇,按照 3:1 分配训练集与测试集,其中训练集 528 篇,测试集 176 篇,用于验证分类模型的性能;待分类数据 2108 篇。本次试验采用精确率  $P_m$ , 召回率  $R_m$  和  $F_1$  值(模型精确率和召回率的一种调和平均)对所构建的文本分类器进行评价,计算公式如下:

$$P_m = \frac{N_{tm}}{N_{rm}} \quad R_m = \frac{N_{tm}}{N_{bm}} \quad F_{1m} = \frac{2P_m \cdot R_m}{P_m + R_m}$$

其中,  $N_{tm}$  表示第  $m$  类分类正确的论文数,  $N_{rm}$  表示实际分类类别为  $m$  的论文数,  $N_{bm}$  表示标准分类为  $m$  类的论文数。  $P_m$  表示第  $m$  类精确率,  $R_m$  表示第  $m$  类召回率,  $F_{1m}$  表示第  $m$  类的  $F_1$  值。

对于分类数据分别采用 LDA-KNN、LDA-SVM、LDA-CNN、LDA-LSTM 进行分类,各分类算法的准确率  $P$ 、召回率  $R$ 、 $F_1$  值见表 1。

由表 1 得, LDA-KNN 算法分类的精确率比 LDA-SVM 算法的精确率的低 18%, 是由于该算法对训练样本数量敏感, 训练集较少, 导致结果精确

表 1 各分类主题算法精确率  $P$ 、召回率  $R$ 、 $F_1$  值

算法类别	精确率 $P/\%$	召回率 $R$	$F_1$ 值
LDA-KNN	58	0.57	0.57
LDA-SVM	76	0.81	0.78
LDA-CNN	63	0.75	0.68
LDA-LSTM	85	0.82	0.83

率较低。LDA-SVM 该分类算法中采用若干个二分类分类器的组合解决一个多分类问题,每次对一个目标类别和剩余其他类别的集合进行二分类,通过循环多次,最终得到分类结果。虽然这一方法时间复杂度高且运算时间长,但较 LDA-KNN 效果有明显提升。LDA-CNN 算法中,提取文本语义特征的 LDA 模型忽略了词义特征的相互关系,将该表示特征为 CNN 输入矩阵,导致该算法的精确率只有 63%。基于 LDA-LSTM 分类模型中, LSTM 神经网络模型获取的是文本的局部特征表示,而 LDA 主题模型获取的是文本的全局特征表示,通过融合深度神经网络获取的文本局部特征和主题模型获取的文本全局特征,使得分类精确率达到 85%。最终结果表明, LDA-LSTM 模型分类的精确率、召回率及  $F_1$  值均优于其他模型,可见该分类模型在保证分类质量的同时具有较好的稳定性。

采用 LDA-LSTM 模型对待分类数据进行分类,得到的分类结果包括技术主题类别及技术主题相对应的特征词(表 2)。经过分类得到 5 大类无人机技术主题类别: 无人机动力技术、无人机导航技术、无人机交互技术、无人机通信技术和无人机集群技术。

#### 3.2 主题突变检测

基于 CiteSpace 的突变检测方法是通过可视化图谱呈现出论文专利数据中所隐含的相关领域研究动态及研究趋势<sup>[23]</sup>。针对无人机技术领域主题突变检测,将构成的基本数据库导入 CiteSpace 软件,从突变权重排序及突变时间段和突变共词聚类知识图谱角度进行分析。

##### 3.2.1 突变权重排序及突变时间段

突变词是指可能成为研究前沿或热点的概念<sup>[24]</sup>,为了验证突变词是否真的能反映研究前沿和

表2 基于主题模型识别的关键技术

序号	技术主题类别	技术主题特征词
1	无人机动力技术	new battery, hybrid, ground power, wireless charging
2	无人机导航技术	positioning technology, speed measurement, obstacle avoidance, tracking technology
3	无人机交互技术	gesture control technology, brain-computer interface technology
4	无人机通信技术	4G/5G communication technology, Wifi communication technology
5	无人机集群技术	intelligent cluster control, intelligent command dispatch, UAV collaboration and cluster control

趋势,将2005—2019年的突变词按突变权重由大到小排序,对其频次的增长趋势状况进行分析,检测它是否具有持续的增长速度。由于无人机技术更新周期一般为3~5年,在此设定每3年为1个时间段,列出突变权重排在前10位的高权重突变词,如表3所示。

从表3中可以看出,突变权重最高的是跟踪技术,突变频次最高的是通信技术。根据突变性质、突变频次及持续时间,将其分为:(1)上升型,例如跟踪技术、通信技术、避障技术、无线充电,在2005

—2019年呈现上升趋势,它们频次较高,突变权重也较大,属于高频焦点词,预示了也是当前热点研究主题;(2)平稳型,例如混合动力及速度测量,其波动较小,属于持续性研究热点主题;(3)下降型,例如传统集群技术从2010年以后开始下降,表示其已是成熟热点或属于逐渐过时的主题,研究已经向新的方向转移;(4)突现型,例如脑机接口技术及手势控制技术,在2017—2019年突现,频次虽然不高,但是突变权重相对较大,具有重大意义,预示这将是领域内具有颠覆性意义的研究主题。

表3 无人机技术主题词突变

序号	突变权重	突变主题	突变时间	突变时间段突变词频次					趋势
				2005—2007	2008—2010	2011—2013	2014—2016	2017—2019	
1	7.88	跟踪技术	2005	9	12	19	27	32	上升
2	7.24	通信技术	2005	4	15	22	36	41	上升
3	6.67	脑机接口技术	2017	—	—	—	—	12	突现
4	5.74	避障技术	2008	—	9	20	22	38	上升
5	4.92	控制技术	2018	—	—	10	—	—	突现
6	4.77	混合动力	2013	—	—	—	15	13	平稳
7	3.24	传统集群技术	2007	43	31	10	8	1	下降
8	2.23	新型电池	2015	—	—	7	12	—	上升
9	1.56	速度测量	2005	13	12	17	14	15	平稳
10	1.32	无线充电	2015	—	—	—	5	17	上升

### 3.2.2 突变共词聚类知识图谱

使用突现词共现网络,能够将数据全集清晰地展示出来。节点越大,说明该关键词词频越大,与主题的相关性越大<sup>[25]</sup>。为了更清晰更细致地识别某一时间段内研究主题突变,设定时间跨度为2005—2019年,top值为30,如图6、图7所示。通过对突变主题词进行共现分析以后建立的突变共词聚类知识图谱,能够全面直观地展示无人机技术领域主题在这一时间段内的研究热点、突变词之间的

关系及重要程度<sup>[26]</sup>。

由图6和图7可以看出,“brain-computer interface technology(脑机接口技术)”“gesture control technology(手势控制技术)”2个节点较大,说明这2个主题突变词的词频高,且与主题的相关性更大。脑机接口指在人脑与计算机或其他电子设备之间建立直接交流和控制通道,通过这种通道人就可以直接通过脑来表达想法或操纵设备。目前,无人机主要通过操纵杆或者手机操控,意味着飞行



30日无人机技术领域的262篇论文专利,采用LDA-LSTM模型进行分类,得到与表2相对应的类无人机技术主题,分析汇总每一类所对应的论文专利数量如图8所示。

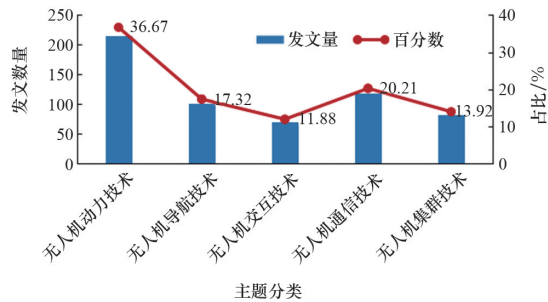


图8 2019年论文、专利的发文章数及所占百分比

由图8可得,基于2019年论文专利数量,将颠覆性技术的未来发展分为3种情况:(1)论文专利数量多,代表可能是无人机技术领域内持续性的研究热点,例如无人机动力技术;(2)论文专利数量较多,发展较为平稳,具有一定的研究热度,例如无人机导航技术、无人机通讯技术、无人机集群技术;(3)论文专利数量较少,但是该技术领域具有较大的研究空间,且具有一定程度的颠覆性,可作为中国研究发展的重点,集中资源形成技术突破,取得技术上的领先地位,例如无人机交互技术。经过验证,对每一类主题的论文专利数量进行时间尺度上的纵向比较(图9),2019—2020年无人机交互技术相关论文专利数量少,结合表2、表3,图6、图7,证明了无人机交互技术为无人机技术领域内的颠覆性技术,同时也证明了本文颠覆性技术识别框架的有效性和准确性。

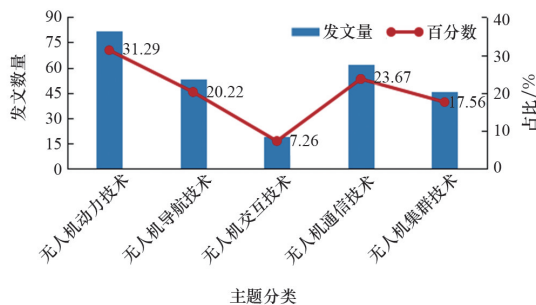


图9 2020年论文、专利的发文章数及所占百分比

## 5 结论

根据LDA-LSTM算法,对输入文本进行局部特征和全局特征提取,进而实现多层次文本信息的提取,有效地对无人机技术进行主题分类,利用CiteSpace的突变检测功能,尝试了从突变词的视角探测无人机技术领域的研究热点,直观地揭示了无人机技术领域中的颠覆性技术。结合相关文献分析,发现以往的研究多集中在理论方面的策略研究及应对措施,缺乏对理论的实践论证。作为实证研究,以2020年无人机技术领域相关论文专利数量分析,证明了本研究框架与方法的有效性和适应性,并得出该技术领域的颠覆性技术主题——无人机交互技术中脑机接口技术及手势控制技术。无人机交互技术颠覆了人类对于技术发展的认知,也颠覆了传统遥控器进行飞行控制,用“意念”或手势控制飞行器。因此,基于文本分类与主题突变检测的颠覆性技术识别,可在一定程度上为研究人员探索颠覆性技术提供有价值的参考。

## 参考文献 (References)

- [1] 李政, 刘春平, 罗晖. 浅析颠覆性技术的内涵与培育——重视颠覆性技术背后的基础科学研究[J]. 全球科技经济瞭望, 2017, 31(10): 53-61.
- [2] Bower J L, Christensen C M. Disruptive technologies: Catching the wave[J]. Harvard Business Review, 1995, 73(1): 43-53.
- [3] Vojak B A, Chambers F A. Roadmapping disruptive technical threats and opportunities in complex, technology-based subsystems: The SAILS methodology[J]. Technological Forecasting and Social Change, 2004, 71(1): 121-139.
- [4] Borup M, Brown N, Konrad K, et al. The sociology of expectations in science and technology[J]. Technology Analysis and Strategic Management, 2006, 18(3): 285-298.
- [5] Shen Y C, Chang S H, Lin G T R, et al. A hybrid selection model for emerging technology[J]. Technological Forecasting and Social Change, 2010, 77(1): 151-166.
- [6] Carlsen H, Dreborg K H, Godman M, et al. Assessing socially disruptive technological change[J]. Technology in Society, 2010, 32(3): 209-218.
- [7] 许泽浩. 基于TRIZ理论的颠覆性技术选择环境研究[J]. 工业工程, 2016, 19(4): 43-47.
- [8] Nagy D, Schuessler J, Dubinsky A. Defining and identify-

- ing disruptive innovations[J]. *Industrial Marketing Management*, 2016, 57(8): 119–126.
- [9] 黄鲁成, 成雨, 吴菲菲, 等. 关于颠覆性技术识别框架的探索[J]. *科学学研究*, 2015, 33(5): 654–664.
- [10] Buchanan B, Corken R. A toolkit for the systematic analysis of patent data to assess a potentially disruptive technology[R]. London: Intellectual Property Office (United Kingdom), 2010.
- [11] Momeni A, Rost K. Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling[J]. *Technological Forecasting and Social Change*, 2016, 104(5): 16–29.
- [12] 罗素平, 寇翠翠, 金金, 等. 基于离群专利的颠覆性技术预测——以中药专利为例[J]. *情报理论与实践*, 2019, 42(7): 165–170.
- [13] 苏敬勤, 刘建华, 王智琦, 等. 颠覆性技术的演化轨迹及早期识别——以智能手机等技术为例[J]. *科研管理*, 2016, 37(3): 13–20.
- [14] 孙轶楠, 杜建, 唐小利. 基于科学论文和专利信息的基因测序技术领域创新态势分析[J]. *中华医学图书情报杂志*, 2017, 26(5): 31–37.
- [15] 张著英, 黄玉龙, 王翰虎. 一个高效的KNN分类算法[J]. *计算机科学*, 2008, 35(3): 170–172.
- [16] 萧嵘, 王继成, 张福炎. 支持向量机理论综述[J]. *计算机科学*, 2000, 27(3): 1–3.
- [17] 陈文, 张恩阳, 赵勇, 等. 基于多分类器协同学习的卷积神经网络训练算法[J]. *计算机科学*, 2016, 43(9): 223–226.
- [18] 任勉, 甘刚. 基于双向LSTM模型的文本情感分类[J]. *计算机工程与设计*, 2018, 39(7): 272–276.
- [19] Blei D M, Ng A Y, Jordan M I, et al. Latent dirichlet allocation[J]. *Journal of Machine Learning Research*, 2003, 3(8): 993–1022.
- [20] Kleinberg J. Bursty and hierarchical structure in streams[J]. *Data Mining & Knowledge Discovery*, 2003, 4: 373–397.
- [21] 杨选辉, 蔡志强. 基于突变检测与共词分析的关联数据新兴趋势探测[J]. *情报科学*, 2018, 36(11): 166–170.
- [22] Chen C M. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. *Journal of the American Society for Information Science & Technology*, 2006, 57(3): 359–377.
- [23] 罗明英, 李东丽, 岳丕昌, 等. 基于突变词检测的世界核桃研究前沿热点分析[J]. *农业图书情报学刊*, 2016, 28(2): 60–63.
- [24] 奉国和, 李媚婵. 基于Citespace的档案学研究可视化分析[J]. *档案学研究*, 2014, 3(5): 18–23.
- [25] 陆泉, 赵琴. 文档内可视化分析工具的比较研究[J]. *信息资源管理学报*, 2016, 6(2): 63–71.
- [26] 成小娟. 我国电子书包的研究热点与发展趋势——基于关键词频和共词网络图的分析[J]. *中国教育信息化*, 2015, 8(11): 22–26.

## The disruptive technology of recognition based on topic mutation detection: With the drone technology as an example

LIU Zhongbao, KANG Jiaqi, ZHANG Jing

School of Software, North University of China, Taiyuan 030051, China

**Abstract** The disruptive technology is of great strategic significance for the leap-forward development of China's technological innovation. In order to get rid of the situation of the lack of original innovation capabilities, and to deal with the varied key core technologies, the subversive technology identification is of great significance. This paper takes the UAV technology field as an example, reviews 2812 papers and patent data collected in the Web of Science (Wos) paper database and Derwent patent database from 2005 to 2019, with the technology based on the LDA-LSTM text classification algorithm Theme, and the CiteSpace to build a co-occurrence network, to realize the theme mutation detection from the perspective of mutation weight ranking and mutation time period and mutation co-word clustering knowledge map, and to identify the brain-computer interface technology and the gesture control in the UAV interaction The technology is a disruptive technology in this field. Finally, the validity of the recognition framework is verified through patents in the related papers in the field of the drone technology in 2020.

**Keywords** disruptive technology identification; text classification; subject mutation detection ●



(责任编辑 王志敏)