doi:10.3772/j.issn.2095-915x.2024.03.006

开放科学
（资源服务）
标识码
（OSID）

基于文本挖掘的颠覆性技术识别方法综述

孙心怡　梁洪振　刘桂锋

江苏大学科技信息研究所　镇江 212013

摘要：[目的 / 意义] 对颠覆性技术的基础理论与识别方法进行梳理归纳有助于把握技术发展态势，推动数智时代面向科技创新的智能化情报服务体系构建。[方法 / 过程] 明确颠覆性技术与文本挖掘方法的基础内涵，通过对该研究领域强关联的核心论文进行定性解读，将基于文本挖掘的颠覆性技术识别方法进行分类比较与组合分析。[局限] 当前颠覆性技术识别领域应用的文本挖掘技术尚不成熟，需要进一步改进并与其他方法结合使用来优化识别效果。[结果 / 结论] 基于文本挖掘的颠覆性技术识别方法包括突变词检测、术语词分析、共词分析、文本聚类、主题模型、语义分析。文本挖掘侧重从技术属性角度识别颠覆性技术，深度挖掘科技文本数据隐藏的潜在技术主题。

关键词：颠覆性技术；技术识别；文本挖掘；技术范式创新；突破性技术

中图分类号：G350

A Review of Research on Identification of Disruptive Technologies Based on Text Mining

SUN Xinyi　LIANG Hongzhen　LIU Guifeng

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

Abstract: [Objective/Significance] The summarization of basic theories and identification methods of disruptive technologies can help grasp the technology development trend and promote the construction of intelligent intelligence service system towards scientific and technological innovation in the era of digital intelligence. [Methods/Processes] This paper firstly clarifies the connotation of disruptive technology and its related concepts, secondly introduces the concept of text mining method and the difference and connection with bibliometrics, and finally summarizes the identification methods of disruptive technologies based on text mining and carries out comparative analysis. [Limitations] Current text mining techniques applied in the field of disruptive technologies recognition are immature and need to be further improved and used in combination with other methods to optimise recognition. [Results/Conclusions] The identification methods of disruptive technologies based on text mining include mutation word detection, terminology word analysis, co-word analysis, text clustering, topic modelling, and semantic analysis.

基金项目　国家社科基金一般项目“科学数据融合模式设计与体系建构研究”（21BTQ080）；江苏省教育科学规划课题高校重点项目“数智化背景下江苏省高校纪检监察工作运行模式研究”（B/2023/01/194）。

作者简介　孙心怡（1999-），硕士研究生，主要研究方向为文本挖掘与主题识别，Email：1014905154@qq.com；梁洪振（1976-），博士，副研究员，主要研究方向为信息资源管理与高等教育；刘桂锋（1980-），博士，教授，主要研究方向为科学数据管理。

引用格式　孙心怡 , 梁洪振 , 刘桂锋 . 基于文本挖掘的颠覆性技术识别方法综述 [J]. 情报工程 , 2024, 10(3): 77-87.

doi:10.3772/j.issn.2095-915x.2024.03.006

Text mining focuses on identifying disruptive technologies from the perspective of technical attributes, and deeply explores the potential technical themes hidden in scientific and technological text data.
Keywords: Disruptive Technologies; Technology Identification; Text Mining; Technological Paradigm Innovation; Breakthrough Technologies

引言

当前全球科技创新已经进入密集活跃期，以颠覆性技术（Disruptive Technologies）为代表的技术革新为全球政治、社会、经济格局带来深刻影响。对于这些具有颠覆性潜力的技术进行识别具有重要的现实意义，一方面有助于发现创新机会，推动科技进步和经济发展；另一方面及时识别出颠覆性技术可以对潜在风险进行预警从而制定出应对措施，保障社会的稳定和安全；此外，对颠覆性技术识别进行深入研究可以加强学术研究与产业创新的融合，促进资源的优化配置。

目前主流的颠覆性技术的识别方法可以大致概括为以下两种：一是主观识别方法，其主要依据是领域专家经验和学术背景，包括专家意见法、技术路线图法、情景分析法等。这类方法主要通过行业或领域专家的学科知识背景来评估技术属性及潜在价值，从而判断某项技术是否具有颠覆性效应，是使用最早、最广泛成熟的方法。这类方法的操作流程简单、使用成熟，能够综合考虑技术要素和市场要素，分析维度全面，但大多从宏观层面评估技术要素，存在主观性强、过程复杂耗时、实施难度大等问题。二是客观识别方法，主要依赖科技文献信息的相关定量特征，包括文献计量法、技术指标评估法、模型分析法等。这类方法大多将科技文献作为数据来源，挖掘能够表征颠覆性技术的关键信息并进行定量分析，优势在于数据来源科学广泛、数据量大，识别结果更加全面可观，是当前科技情报学界较为推崇的方法。文献计量法是对专利、论文等科技文献中能够反映技术特性信息的计量，如共词分析、词频统计、引文分析等 [1]；技术指标评估法主要针对技术特征、市场变化、外部环境等方面将技术量化表示，通过构建综合指标、分配权重从而进行统计识别 [2]；模型分析法是以理论准则为基础的技术量化模型，代表性的理论包括 TRIZ 理论、QFD 理论、种间竞争理论、传染病模型等 [3-5]。这类方法虽然能够有效避免主观性，但大多是对外部特征的描述，缺少针对多源数据的内容特征分析、关联关系分析、多特征项交叉共现分析等。

近年来兴起并日渐流行的文本挖掘方法，不仅能够有效规避主观识别方法的一系列问题，还能处理非结构化数据，弥补客观识别方法中文本内部特征分析不全面的缺陷。基于文本挖掘的方法可以从科技文本数据中抽取技术术语，通过共词分析、突发词检测、文本聚类、SAO 结构分析、主题模型等方法及相应的软件工具提取科技文本中的关键词、主题词、语义结构等来识别符合某些特征的技术，从而能够更好地挖掘深层次的技术隐性知识。

1　理论与方法基础

1.1　颠覆性技术的内涵

颠覆性技术的概念最早由 Christensen 等 [6-7] 在 1995 年提出，他们认为这类技术“早期以简单、方便、便宜的特征由低端或边缘市场切入，在性能的不断升级中逐渐取代主流技术，从而形成新的市场格局与价值体系”。此后国内外学者也从不同视角提出了各自的见解，然而这一概念在不断地发展演变中尚未统一，本文将其中代表性的观点按不同视角进行归纳，详见表 1。

表 1　不同视角下国内外学者对颠覆性技术的概念界定

视角	侧重点	定义	来源
市场环境	低端市场切入	技术特性及功能与主流市场相似，但初期性能较差，后期性能跃升并取代主流技术。	[8]
	高端市场切入	初期因技术成本过高而导致受众范围较小，后期技术性能显著跃升并取代主流技术。	[9]
技术属性	替代性改进技术	基于现有技术的改进突破，通过不断升级迭代占领主流市场，并形成新的价值体系。	[10]
	原始性创新技术	以全新的概念或原理为基础的原始创新，被接纳后快速传播并给社会带来根本性变化。	[11]
行业角度	企业竞争	一类改变企业竞争基础与能力的绩效指标。	[12]
	产品与服务	现有技术范式、产品与服务功能的变革，并对产业的市场与竞争格局造成破坏性影响。	[13]
其它		一是基于新原理、新发现的原始创新；二是基于现有技术的集成创新与应用；三是科学原理与成熟技术的转移与创新应用。	[14]

从市场环境来看，可根据市场颠覆方式分为高端颠覆性技术和低端颠覆性技术，具体的技术轨迹如图 1 所示。在将市场分为高端、主流、低端三类的基础上，进一步将高端市场和低端市场分别细分为高端分离、高端边缘市场以及低端分离、低端边缘市场，由此可以得到四条技术轨道：高端颠覆性技术从高端分离市场切入，经过高端边缘市场进入并颠覆主流市场；Rafii 等 [8] 认为新市场颠覆性技术轨道从低端分离市场切入，经由低端边缘市场进入并颠覆主流市场；低端颠覆性技术从低端边缘市场切入并随时间推移进入并颠覆主流市场；持续性技术则始终位于主流市场，是对主流技术的持续性创新。

从技术属性来看，一种是替代性改进技术，是对现有技术性能的突破性改进，在经历多次迭代升级后能够提供完全替代甚至超越主流技术性能的产品，最终彻底颠覆占领主流市场。另一种是原始性创新技术，Christensen 等 [11] 将其定义为以全新的技术理念为前提所产生的原始性创新，在技术渗透的过程中逐渐满足主流需求后还能进一步开发新的技术需求与市场需求，最后形成新的价值体系。还有卢光松等 [13] 从行业角度对颠覆性技术概念进行界定，一是从企业竞争角度考虑，关注技术的颠覆性效应对企业竞争能力以及绩效水平的改进；二是基于产品与服务视角，侧重技术创新过程中对产品性能、服务水平、产业变革、社会效益等方面所带来的重大影响和突破。

doi:10.3772/j.issn.2095-915x.2024.03.006

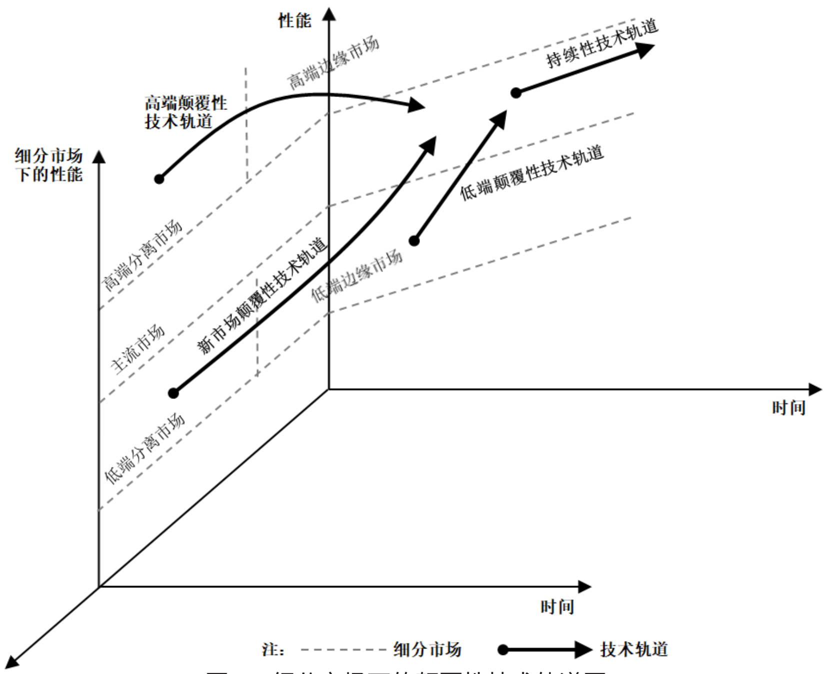

图 1　细分市场下的颠覆性技术轨道图

综合来看，当前研究一致认为颠覆性技术是由现有技术产生替代改进式突破或原始性创新技术动态演变而来的，此类技术初期由边缘市场切入，因技术性能较低而无法满足主流需求而容易被忽视，后期通过性能跃升入侵、威胁、最终彻底破坏主流技术轨道，颠覆主流市场格局，最终塑造全新的价值体系。

通过梳理相关文献可以发现，有些概念与颠覆性技术之间存在密切关联及差异，因而清晰梳理并辨别它们之间的异同有助于为后续相关研究的开展提供理论依据。本文在已有研究基础上对这 5 个相关技术的定义、技术特征、要求进行梳理归纳，如表 2 所示。

表 2　颠覆性技术的关联概念辨析

名称	定义	技术特征及要求	来源
新兴技术	新出现的，具有创造新产业或改造现有产业的潜力的技术。	新颖性、明显发展潜力、持续增长性	[15]
前沿技术	高技术领域中具有前瞻性、先导性和探索性的重大技术。	技术的前瞻性、新兴性、探索性	[16]
突破性技术	通过新知识淘汰现有技术，具有变革性和不连续性的技术。	对已有技术的根本性创新	[17]
破坏性技术	借助低端市场及成本优势对已有技术进行融合创新从而颠覆主流市场。	初始性能较差、结构简单、低成本、高易用性	[18]
“卡脖子”技术	具有高技术价值并同时存在技术垄断的关键核心技术或高技术产品。	垄断性、潜伏性、高技术特征，是国际竞争的威胁性工具，关键核心技术缺乏自主权	[19]

1.2　文本挖掘方法

颠覆性技术具有潜在变革性的影响，对其进行识别与分析对社会、经济、科技等领域的意义重大，近年来文本挖掘技术凭借高效灵活、大数据适应性、信息揭示性、可扩展性等优势被更多地应用到颠覆性技术识别的研究中。文本挖掘指的是从大规模的文本形式的数据集合中提取以往未知的、隐性的、可理解的、实用

的信息的技术，能够处理非结构化数据并常应用于文本表示模型、特征项抽取、模式发现（如关联规则抽取、文本分类、文本聚类）等领域[20-21]。现有的文本挖掘方法主要分为三种：一是基于数理统计的方法，如词频统计、TF-IDF、文本分类等；二是基于人工设定规则的方法，如语法分析、句法分析等；三是基于机器学习的方法，如 LDA、SVM、朴素贝叶斯、深度学习等。文本挖掘方法在技术主题识别的研究中已得到广泛成熟的应用，包括引文聚类、技术术语词频分析、主题模型、异常监测等[22-24]。而数据的增量重组与学科的交叉融合促使文本挖掘方法被更多地应用到计算机科学、图书馆学情报学、教育学、社会学、心理学、技术管理等众多学科领域中，研究对象也逐渐从学术文献扩展到社交网站，如新闻评论、博文帖子、客户评价、语音文本（STT）等，当前的研究趋势聚焦于大数据分析、社交媒体分析、情绪分析、话题建模等方面。随着基础理论与算法的创新优化，文本挖掘方法的应用场景在不断扩张，但在实际应用中也存在诸如文本数据噪声、算法的普适性等问题，因而需要设计更加准确高效的算法加以优化，文本挖掘在不同应用场景下与其他方法的优势组合将成为重要的研究趋势。

2 基于文本挖掘的颠覆性技术识别方法

文本挖掘因其良好的数据挖掘分析能力，被越来越多地运用到技术主题分析中，本文确定检索式：“TI=(颠覆性技术 OR 颠覆性创新 OR 破坏性技术 OR 破坏性创新 OR 突破性技术 OR 突破性创新 OR 不连续技术 OR 技术不连续性) AND TI= 识别”，在 CNKI 中获取到与研究主题强关联的核心期刊论文 40 篇，学位论文 7 篇，并对这 47 篇文献进行解读，将基于文本挖掘的颠覆性技术识别总结为如下六种方法，并对这六种方法进行比较分析。

2.1 方法类型

2.1.1 突变词检测

突变词检测（Burst Items Detection）依据词频增长率的突变情况判断某一领域的发展趋势和研究动向[25-26]，颠覆性技术的发展路径符合突变词的变化规律，颠覆性技术发展之初潜伏于边缘市场，容易被主流市场所忽略，后经过技术性能的更迭升级迅速取代、颠覆主流市场，而突变词则是捕捉一定时期内词频增长率发生异常骤变的词语，并将其定义为突变词。同颠覆性技术一样，突变词的增长率在时段初期不易被关注到，但却能够在某一时间节点发生突变。突变性是颠覆性技术的重要特征，通过观察科技文本数据中发生突变的词语，可以有效发现新兴技术主题和热门话题，结合观察突变词的动态变化情况，有助于挖掘具有颠覆性潜力的词项，为后续的技术识别奠定基础。现有研究中大多基于主题词的突变情况进行异常检测，通过结合不同的机器学习算法来优化识别的精度，如 Miao 等[23]利用 LDA 主题模型和主题变异模型识别技术演化路径中的关键节点；王秀红等[27]基于 CBLOF 算法进行关键技术主题的异常监测来表征颠覆性技术的突变属性，从而识别出颠覆性技术。突变词检测常用于新兴前沿技术主题的探测，以此观察颠覆性

doi:10.3772/j.issn.2095-915x.2024.03.006

技术的突变性特征。

2.1.2 术语词分析

文本挖掘的首要任务是命名实体的识别，延伸到颠覆性技术识别的研究中则表现为颠覆性技术术语词的抽取，首先需要借助自然语言处理（NLP）对论文专利等科技文本中的标题、摘要等内容进行分词、去停用词、保留词等操作，进而得到能够体现技术属性或特征的技术术语，并用以反映技术主题，在此基础上进行分析。术语词分析在颠覆性技术的识别研究中表现为对高频词的时序分析，通过术语词词频的年代分布观察技术主题的发展情况，能够发现技术动态演变中的新兴技术术语及其最新动向，观察不同时期的术语词分布有助于发现技术主题的演变规律，从而更好地识别出具有颠覆性潜力的技术，如邢晓昭等 [28] 根据专利标题和摘要中抽取到的术语关键词构建时间窗下的术语关键词网络并识别出类脑智能领域的颠覆性技术主题。术语词分析主要用于技术实体的抽取，并且更多情况下仅作为一种手段或环节，要想更加准确高效地识别颠覆性技术还需要与其他方法组合搭配进行。

2.1.3 共词分析

共词分析本是文献计量的传统方法，随着方法应用的深度融合逐渐被改进并融入文本挖掘中。共词分析在文献计量学中更多地被用于分析同一篇期刊论文中关键词对的共现情况，从而观察某研究领域的学科结构与主题变化。然而，识别颠覆性技术的数据来源中除了论文，都缺少规范的关键词，文本挖掘能够自动抽取非结构化文本数据中的术语关键词，利用共词分析发现技术主题。目前共词网络分析法使用较成熟，根据技术术语词在同一文本中的共现情况绘制共词网络图，从而更直观地反映出术语词之间的共现关系，揭示技术主题的分布关联情况。如 Dotsika 等 [29] 使用关键词共现网络揭示颠覆性技术领域的主题内容、结构及时间趋势。此外，共词聚类分析和战略图分析也是共词分析中使用较多的方法，共词聚类常用 K-means 聚类、层次聚类等方法将复杂的术语词汇聚成不同的技术主题，而战略图分析则是在共词聚类的基础上计算主题的中心度和密度来绘制战略图 [25]。共词分析法常与文本聚类法相结合，用于技术主题的识别与分析。

2.1.4 文本聚类

文本聚类在颠覆性技术识别的研究中主要表现为技术主题聚类。通过测量已抽取出的术语关键词、突现词等词项各自的语义相似度，将不同文本数据集合中相似度较高的文档聚为一类，并探测相应的技术主题，对科技文本内容进行深度挖掘，从而发现潜在的知识结构，识别出更能够表征颠覆性技术属性的技术主题，进一步优化识别效果。文本聚类一般分为两个阶段：第一个阶段是将术语词根据语义关联或共词关系进行聚类，从而形成多个技术主题团簇；第二阶段是将不同的主题团簇进行命名，生成相应的主题词或构建主题词表。常用的文本聚类方法包括 TF-IDF、K-means、层次聚类、DBSCAN 等，Li 等 [30] 将文本聚类算法用于定位与技术特征相对应的文本内容来识别颠覆性技术，Liu 等 [31] 将 K-means 聚类与表示学习相结合从多源文本中识别潜在的颠覆性技术及其特征。综合来看，文本聚类方法可以从宏观上快速直观地把握技术主题的分布，多数情况下

用于技术特征的识别与抽取，但由于目前的文本挖掘技术并不足以深入、精准地捕捉潜在的语义信息，因而聚类效果往往无法切合预期，这就不可避免地需要人为干预辅助分析。要想充分解决这一问题，首要任务在于攻克文本挖掘的技术难关。

2.1.5 主题模型

主题模型是现阶段使用最广泛而普遍的颠覆性技术识别方法，是对大量多源异构的科技文档集合中所识别出的术语词进行统计，从而反映出文本内容中隐藏的潜在技术主题分布、主题结构以及主题演化趋势等。现阶段在识别颠覆性技术的研究中使用最多的主题模型是隐含狄利克雷分布（Latent Dirichlet Allocation，LDA）模型，它的本质是一种概率图模型，根据文档集中的词项分布生成隐藏的技术主题分布从而对文本数据进行降维，发现技术隐性知识。单晓红等 [32] 利用 LDA 主题模型识别出颠覆性技术，但由于传统的 LDA 模型存在无法把握词项间的语义关系、容易混淆歧义较大的特征词等问题，从而造成识别结果存在误差。因而有学者尝试将 LDA 与其他方法结合使用来优化识别效果。如将 LDA 模型与专利开发路径、K-core 分析相结合，识别与检测光伏行业潜在的颠覆性技术 [33]。还有一些常用于识别颠覆性技术的主题模型，如 LDAVec、DTM、STM 等。主题模型法最大的优势在于能够处理大量的文本数据，并对文本内容进行降维处理从而揭示潜在的技术趋势和创新点，但由于在模型的适用性、稳定性、文本数据的质量与处理等方面仍面临一些挑战，因而需要不断进行模型调试与改进来优化识别效果。

2.1.6 语义分析

由于颠覆性技术发展路径存在高度不确定性，仅仅对关键词或主题词进行词频统计无法精准把握颠覆性技术的内在特征，越来越多的学者开始尝试利用语义分析来识别颠覆性技术。语义分析法是一种基于统计的文本数据挖掘技术，常用的方法是 SAO（Subject-Action-Object）结构语义分析，“S”“A”“O”三个字母分别表示的是技术组件的主语、宾语以及表示实体间关联关系的谓语，Li 等 [30] 使用 TF-IDF 与专利化的 SAO 结构提取技术特征，并结合深度分类模型识别颠覆性技术。除此之外，国外的一些研究机构和政府部门也开始将语义分析用于识别和跟踪颠覆性技术，如美国国防部高级研究计划局（DARPA）通过对科学论文、新闻报道等其他公共数据进行文本主题的语义分析来识别具有颠覆性潜力的技术创新。借助语义分析可以较好地挖掘科技文本中所包含的技术实体及其关联要素，从而更精准地进行技术识别。语义分析主要用于颠覆性技术识别的科技文本分类及技术特征提取等方面。

2.2 方法分析

2.2.1 基于文本挖掘的颠覆性技术识别方法比较

基于文本挖掘的颠覆性技术识别方法类型上以定量为主，主要表现为对文本内容数据的量化统计或结构分析，方法特征上大多从技术特征视角切入，利用文本挖掘技术从科技文档中抽取能够表征颠覆性技术属性的实体，进而识别出颠覆性技术。为了更直观清晰地了解上述六种文本挖掘方法的区别与联系，本文从作用思路、方法内核、优势、局限性、应用范围几方面进行进一步的比较分析，如表 3 所示：

doi:10.3772/j.issn.2095-915x.2024.03.006

表 3　基于文本挖掘的颠覆性技术识别方法比较

方法名称	作用思路	优势	局限性	应用范围
突变词检测	将词频增长率的突变情况作为技术领域研究态势的判断标准	切合技术路径	仅依赖词频增长率判断难以掌握技术本质特征	技术突变性特征计量
术语词分析	借助 NLP 技术从科技文本中抽取技术术语来反映技术主题	抽取技术特征	无法反映术语词或技术主题之间的关系	技术实体抽取
共词分析	对相同文本中术语词的共现情况进行分析	把握技术热点分布及其关联	存在时间滞后性	技术主题分析；技术特征识别
文本聚类	根据计算出的语义相似度将高关联的文本聚为一类探测技术主题与知识结构		难以进行技术主题的横向分析，算法准确度有待提高	
主题模型	从大量科技文档集合中识别出隐藏的潜在技术主题并进行主题分析	处理大量复杂的文本数据	模型准确性和稳定性有待长时间监测	科技文本降维
语义分析	从科技文本中自动抽取语言知识并进行语义计算	更精准理解科技文本含义	需要大量语料库和语言知识的支持，准确性受语言复杂程度影响	技术特征识别；科技文本分类

2.2.2　基于文本挖掘的颠覆性技术识别方法组合

上述六种文本挖掘方法在颠覆性技术识别过程中具体的应用位置如图 2 所示。从整体的识别过程来看，现阶段的学者更倾向于多种方法的组合使用，从而将各个方法的优势发挥到最大。

（1）文本数据来源：识别颠覆性技术的数据大多来源于与技术本质密切相关的科技文本，包括科技论文、专利、科技新闻、政策文本、科技报告等，从数据源上看，基于科技论文或专利的单一数据源进行颠覆性技术识别的研究占比最多，其次是基于论文和专利两种数据源，而随着文本挖掘技术的不断发展，以多种文本数据类型作为数据来源的研究也在不断增多，未来如何基于多源数据来识别颠覆性技术将成为研究热点之一。

（2）数据预处理：在确定数据源之后，接下来则需要对获取到的数据进行预处理，这一步需要借助自然语言处理技术将大量的文本内容进行分词、去停用词、保留词、自定义技术术语词典等操作来完成文本内容的向量化、结构化处理。

（3）技术实体抽取：从处理完的结构化数据中抽取技术实体，这一步可以用到术语词分析的方法，依据词频进行高频术语词分布情况或时间序列上的分析，部分研究会按照具体需要进一步划分时间周期或进行时间切片的操作。

（4）科技文本降维、分类与分析：在识别出技术实体后，则需要对科技文本进行降维，大多数研究会使用主题模型从技术实体中进一步凝练技术主题。在抽取出相应的技术主题后，多数研究会进行技术主题分析，通过使用共词分析和文本聚类观察主题分布情况、演化趋势以及主题间的关联情况等，也有研究利用语义分析进行科技文本的分类后再进行技术主题分析或直接进入到技术特征识别这一步骤。

（5）颠覆性技术特征识别：技术特征识别是整个识别过程的核心，需要将文本内容特性与颠覆性技术特征相匹配，通常将颠覆性技术特征归纳为成长性、融合性、突变性、创新性、突破性等，而突变词检测的方法通常用于识别颠覆性技术的突变性特征，此外，共词分析、文本聚类和语义分析也常被用于颠覆性技术特

征的识别中。

综上，文本挖掘方法或与其他技术识别方法相结合可以提取出符合颠覆性技术属性的技术名称或主题，从而完成某技术领域的颠覆性识别。需要明确的是，现阶段的文本挖掘方法尚不够成熟，且由于文本内容存在语义结构复杂等问题致使语义理解上存在困难，因而无法实现全过程的自动化抽取与分析，在研究过程中往往还是需要依靠技术领域专家进行辅助，从而达到理想的识别效果。

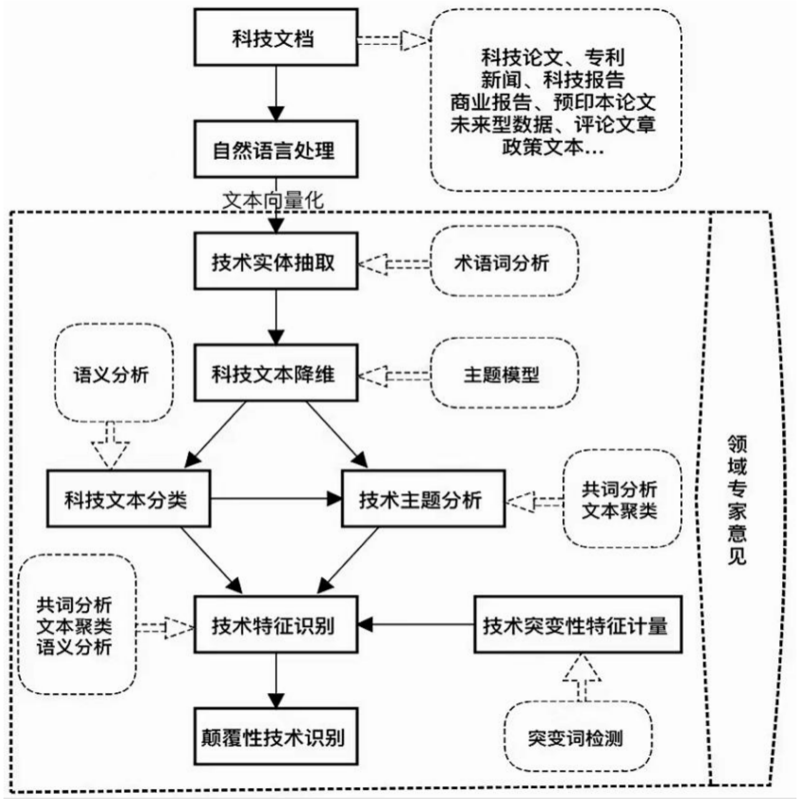

图 2　文本挖掘法在颠覆性技术识别过程中的应用

综合来看，文本挖掘法能够从微观层面探测到颠覆性技术的相关信息，相较于传统识别方法能够更好地发现潜在技术主题，随着机器学习、深度学习等新兴技术的快速发展，多种方法的组合使用成为必然趋势。但由于文本挖掘法在颠覆性技术识别领域的发展时间较短，可直接获取利用的工具有限，自动处理效率不高，需要结合专家意见进行干预。尽管已有研究验证了识别结果的可靠性，但未来如何进行技术优化还有待探索。

3　结语

颠覆性技术对社会和经济发展具有深远的影响，因此其准确识别具有重要意义。识别和研究颠覆性技术，有助于为科研、产业政策等领域提供有力的支持和指导。文本挖掘法能够从大量文本数据中自动提取关键信息，实现对颠覆性技术的准确识别，具有高效、自动化、可扩展性强等优势，可以有效应对大规模数据和多样化的颠覆性技术类型。本文通过文献梳

doi:10.3772/j.issn.2095-915x.2024.03.006

理将基于文本挖掘的颠覆性技术识别方法总结为六种：突变词检测、术语词分析、共词分析、文本聚类、主题模型、语义分析，并对六种方法进行了比较分析。发现文本挖掘法更注重从技术特征角度进行颠覆性技术识别，不少研究也证明了该方法的科学可靠性，然而缺乏市场环境、行业角度等综合视角的分析。其次，虽然文本挖掘方法在颠覆性技术识别中取得了一定的成果，但仍然存在一些挑战。例如，文本数据的质量和多样性，颠覆性技术的时效性和隐蔽性等都会对识别结果产生影响。此外，模型的可解释性、标注数据的获取等问题也需要解决。因而未来研究中需要更加关注改进文本挖掘方法来应对数据特点的变化、结合其他技术手段提高颠覆性技术识别的精度和可靠性、建立更加完善的数据集等方面的问题。此外，跨学科合作、开展实证研究等也是未来发展的重要方向。

参考文献

[1] JIA W F, XIE Y P, ZHAO Y N, et al. Research on Disruptive Technology Recognition of China's Electronic Information and Communication Industry Based on Patent Influence[J]. Journal of Global Information Management, 2021, 29(2): 148-165.

[2] GUO J F, PAN J F, GUO J X, et al. Measurement Framework for Assessing Disruptive Innovations[J]. Technological Forecasting and Social Change, 2019, 139: 250-265.

[3] SUN J, GAO J, YANG B, et al. Achieving disruptive innovation forecasting potential technologies based upon technical system evolution by TRIZ[C]//International conference on management of innovation & technology. Bangkok: IEEE, 2008: 18-22.

[4] NILCHIANI R, EDWARDS C M, GANGULY A. Introducing a tipping point measure in explaining disruptive technology[C]//2019 International Symposium on Systems Engineering. Edinburgh: IEEE, 2019: 1-5.

[5] MULLER J M, KUNDERE R R. Ex-ante prediction of disruptive innovation: the case of battery technologies[J]. Sustainability, 2019(11): 5229.

[6] CHRISTENSEN C M, BOHMER R, KENAGY J. Will disruptive innovations cure health care?[J]. Harvard business review, 2000, 78(5): 102-112.

[7] BOWER J L, CHRISTENSEN C M. Disruptive Technologies: Catching the Wave [J]. Harvard Business Review, 1995, 73(1): 43-53.

[8] RAFII F, KAMPAS P J. How to Identify Your Enemies Before They Destroy You[J]. Harvard Business Review, 2002, 80(11): 115-123, 134.

[9] PARRY M E, KAMAKAMI T. The encroachment speed of potentially disruptive innovations with indirect network externalities: The case of e-readers[J]. Journal of Product Innovation Management, 2017, 34(2): 141-158.

[10] GANGULY A, NILCHIANI R, FARR J V. Defining a set of metrics to evaluate the potential disruptiveness of a technology[J]. Engineering Management Journal, 2010, 22(1): 34-44.

[11] CHRISTENSEN C M, MCDONALD R, ALTMAN E J, et al. Disruptive innovation: An intellectual history and directions for future research[J]. Journal of Management Studies, 2018, 55(7): 1043-1078.

[12] DANNEELS E. Disruptive technology reconsidered: A critique and research agenda[J]. Journal of Product Innovation Management, 2004, 21(4): 246-258.

[13] 卢光松，卢平．企业颠覆性技术路线图制定研究[J]. 科技进步与对策，2011, 28(11): 81-86.

[14] 栾恩杰，孙棕檀，李辉，等．国防颠覆性技术在航天领域的发展应用研究 [J]. 中国工程科学，2017, 19(5): 74-78.

[15] PORTER A L, ROESSNER J D, JIN X Y, et al. Measuring national 'emerging technology' capabilities[J]. Science & Public Policy, 2022, 29(3): 189-200.

[16] 中国政府网．国家中长期科学和技术发展规划纲　要 [EB/OL]. (2006-02-09)[2023-06-08]. http://www.gov.cn/jrzg/2006-02/09/content_183787_5.htm,

2021-07-12.

[17] TUSHMAN M L. Technological discontinuities and organizational environments[J]. Administrative Science Quarterly, 1986, 31(3): 439-465.

[18] 李平 , 臧树伟 . 基于破坏性创新的后发企业竞争优势构建路径分析 [J]. 科学学研究 , 2015, 33(2): 295-303.

[19] 李昱璇 , 方卫华 . " 卡脖子 " 技术概念辨析 ——内生性矛盾、国家主体与外部限制的共同建构 [J]. 科学学研究 , 2024, 42 (1): 31-37, 135.

[20] ZHANG Y, CHEN M, LIU L. A Review on Text Mining[C] // IEEE Beijing Section. Proceedings of 2015 IEEE 6th International Conference on Software Engineering and Service Science (ICSESS 2015). Institute of Electrical and Electronics Engineers, 2015: 5.

[21] RONEN FELDMAN, IDO DAGAN. Knowledge Discovery in Textual Databases (KDT) [C]//In: Proceedings of the First International Conference on Knowledge Discovery and Data Mining. Montreal: AAAI Press, 1995: 112-117.

[22] LI X, FAN M, ZHOU Y, et al. Monitoring and forecasting the development trends of nanogenerator technology using citation analysis and text mining[J]. Nano Energy, 2020, 71: 104636.

[23] MIAO Z, DU J, DONG F, et al. Identifying technology evolution pathways using topic variation detection based on patent data: A case study of 3D printing[J]. Futures, 2020, 118(2): 102530.

[24] WANG Y, FENG L, WANG J, et al. Technology Trend Forecasting and Technology Opportunity Discovery Based on Text Mining: The Case of Refrigerated Container Technology[J]. Processes, 2022, 10(3): 551.

[25] KLEINBE RG J. Bursty and hierarchical structure in streams[J]. Data Mining and Knowledge Discovery, 2003, 7(4): 373-397.

[26] CHEN C. 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.

[27] 王秀红 , 王欣 , 王少凡 , 等 . 基于 SimCSE-LDA 和异常检测的颠覆性技术识别方法 —— 以农业机器人为例 [J]. 情报理论与实践 , 2023, 46(5): 135-143.

[28] 邢晓昭 , 任亮 , 雷孝平 , 等 . 基于专利主题演化的颠覆性技术识别研究 —— 以类脑智能领域为例 [J]. 情报科学 , 2023, 41(3): 81-88.

[29] DOTSIKA F, WATKINS A. Identifying potentially disruptive trends by means of keyword network analysis[J].Technological Forecasting Social Change, 2017, 119(6): 114-127.

[30] LI R, YU W, HUANG Q, et al. A new identify disruptive technologies algorithm based on technology develop network[J]. Mathematical Problems in Engineering, 2022, 2022(1): 7354535.

[31] LIU X, WANG X, LYU L, et al. Identifying disruptive technologies by integrating multi-source data[J]. Scientometrics, 2022, 127(9): 5325-5351.

[32] 单晓红 , 韩晟熙 , 刘晓燕 . 基于技术主题演化的颠覆性技术识别研究 [J]. 情报理论与实践 , 2023, 46(8): 113-123.

[33] MOMENI A, ROST K. Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling[J]. Technological Forecasting Social Change, 2016, 104: 16-29.