

基于技术成熟度曲线的大数据发展趋势分析

摘 要

引入技术成熟度曲线的概念，在分析 Gartner 公司发布的 2014 年新兴技术成熟度曲线基础上，深入解读其在 2012—2014 年发布的大数据技术成熟度曲线的特点与变化，分析大数据近年来的总体发展趋势和热点技术，为大数据研究提供一种新的工具、方法和视角。

关键词

技术成熟度曲线 大数据 Gartner

1 技术成熟度曲线概述

技术成熟度曲线（Hype Cycle），又称炒作周期曲线，是指新技术、新概念在媒体上的曝光度（即可见度）随时间而变化的曲线，描述了一项技术从诞生到成熟的过程，是评估技术发展周期的一种预测模型。自 1995 年起，全球权威 IT 研究与顾问咨询公司高德纳（Gartner）每年发布新兴技术成熟度曲线报告，分析预测与推论各种新科技的成熟演变速度及达到成熟所需的时间，利用一致性评价确定技术在成熟度曲线中的位置。如图 1 所示，技术成熟度曲线的横坐标轴代表技术所处的阶段，包括技术萌芽期、期望膨胀期、泡沫谷底期、稳步复苏期和生产高峰期五个阶段（见表 1），纵坐标轴代表技术可见度。曲线上的技术运用不同符号标注预期成熟时间，分为少

于2年、2—5年、5—10年、多于10年、成熟期前放弃。

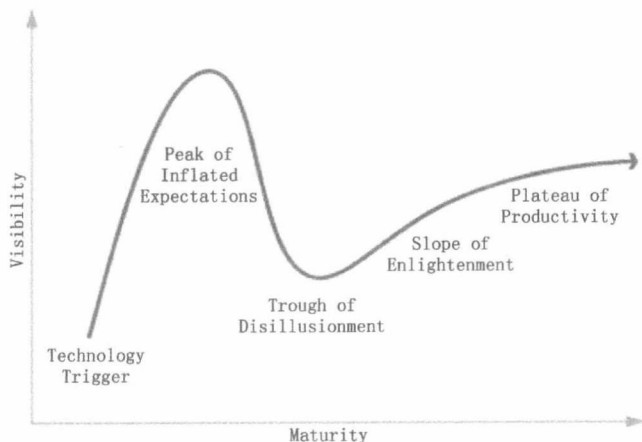


图1 技术成熟度曲线示意图

表1

技术成熟度曲线的五个阶段

序号	阶段名称	定义
1	技术萌芽期 (Technology Trigger)	新技术出现的初期。早期的概念验证与媒体的兴趣引发了公众的关注。在此阶段尚未产生可用的产品,也没有证明其商业可行性
2	期望膨胀期 (Peak of Inflated Expectations)	早期宣传产生了大量的成功案例,同时也伴随着大量的失败案例。只有少部分公司采用该项新技术
3	泡沫谷底期 (Trough of Disillusionment)	随着实验和实施的失败,人们对新技术的兴趣开始减弱。技术生产商草草收尾或以失败告终,幸存的供应商只有提高产品质量并达到早期采用者的满意才能获得更多投资
4	稳步复苏期 (Slop of Enlightenment)	新技术使企业受益的案例增多,并被广泛接受。第二代及第三代技术开始出现,更多企业开始试用,保守的公司依然持谨慎态度
5	生产高峰期 (Plateau of Productivity)	新技术已经作为主流技术被广泛采用,并制定出评价新技术生产可行性的标准,新技术具备了广泛的市场适用性和相关性

2 2014 年新兴技术成熟度曲线

2014 年, Gartner 公司从市场成熟度、商业利益与未来方向等角度评估了 119 个领域的 2000 多项技术, 得到 45 项新兴技术的成熟度曲线, 从图 2 中可以看出: 处于萌芽期和膨胀期的技术比较密集, 但是大部分需 5—10 年才能达到生产高峰, 包括物联网、可穿戴技术、加密货币、个人助手、联网家庭、智能机器人、自动驾驶车辆等; 而处于谷底期之后的技术, 大多需 2—5 年达到成熟, 如语音识别、云计算、内容分析、企业级 3D 打印等, 详见表 2。



图 2 2014 年新兴技术成熟度曲线

资料来源: Gartner (2014 年 8 月)。

2014 年新兴技术成熟度曲线报告以数字化企业的发展蓝图为主题, 总结出发展成为数字化企业的六个阶段: 模拟、网络、电子商务、数字化营销、数字化商业和自动化。由于新兴技术成熟度曲线把焦点放在新近崛起的技术上, 因此技术多集中在后三个阶段:



表 2

2014 年新兴技术成熟阶段分布

阶段时间	技术萌芽期	期望膨胀期	泡沫谷底期	稳步复苏期	生产高峰期
<2 年				内存分析	语音识别
2—5 年		数据科学	内存数据库管理系统 内容分析 混合云计算 云计算 近距离无线通讯	3D 扫描仪 企业级 3D 打印 活动流 手势控制	
5—10 年	数字安全 虚拟个人助手 智能工作场所 联网家庭 量化自我 软件定义一切 3D 生物打印系统 智能机器人 情感计算 生物芯片 规范分析	智能顾问 自动驾驶车辆 语音翻译 物联网 自然语言问答 可穿戴用户界面 消费级 3D 打印 加密货币 复杂事件处理	大数据 游戏化 增强现实 机器间交互通讯 移动健康监测 虚拟现实	消费级车联网	
>10 年	生物声学传感 脑机接口 人类机能增进 量子计算 立体与全息显示 神经商业				

(1) 数字化营销（第四阶段）：此阶段强调多方力量的结合（如移动、社交、云）。聚焦于以新的、更先进的方式抵达消费者，寻找愿意参与营销活动、获取产品及服务的消费者，此类消费者青睐移动设备和社交网络带来的生活便利，同时他们也带动了数字化营销。此阶段的相关技术包括：软件定义一切、立体与全息显示、神经商业、数据科学、规范分析、复杂事件处理、大数据、内存中数据库管理系统、内容分析、混合云计算、游戏化、增强现实、云计算、近距离无线通讯、虚拟实境、手势控制、内存中分析、活



动流及语音识别。

(2) 数字化商业（第五阶段）：这个阶段焦点转移到人员、企业与事物的融合上。物联网以及现实与虚拟世界的模糊化成为这个阶段的强烈概念。数字化实物资产在业务价值链中的重要性等同于数字化实体。如 3D 打印使实物数字化更进一步，并有可能在供应链与制造层面引发颠覆性变革；将与人相关的各种属性（如健康生命体征）数字化也是这个阶段的特色之一；连货币也早已数字化转型为比特币等加密货币。此阶段相关技术包括：生物声学传感、数字安全、智能工作场所、联网家庭、3D 生物打印系统、情感计算、语音翻译、物联网、加密货币、可穿戴用户界面、消费级 3D 打印、机器对机器通信服务、移动健康监测、企业级 3D 打印、3D 扫描仪及消费级车联网。

(3) 自动化（第六阶段）：该阶段的企业将有能力提供各种类人或替代人类的技术。例如以自动驾驶交通工具来运载人员或物品，或使用认知系统撰写文案或回答顾客问题。此阶段相关技术包括：虚拟个人助手、人类机能增进、脑机接口、量子计算、智能机器人、生物芯片、智能顾问、自动驾驶车辆及自然语言问答。

3 大数据技术成熟度曲线及发展趋势

Gartner 公司认为“大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。可见，大数据指的是无法使用传统流程或工具处理分析的信息，具有日益增长的数据量（volume）、高增长速度（velocity）、多样性（variety）的特点。从某种程度上说，大数据是数据分析的前沿技术。简言之，从各种各样类型的数据中，快速获得有价值信息的能力，就是大数据技术。

3.1 大数据发展趋势

2011 年，大数据（Big Data）的概念首次进入 Gartner 技术成熟度曲线，经历了 2012 年的期望膨胀期，于 2013 年达到期望膨胀顶峰，于 2014 年进入泡沫谷底期（见图 3）。Gartner 公司在《2014 年技术成熟度曲线特别报告》中指出：“虽然对大数据的兴趣依然不减，但它已经离开高峰期，因为

该市场已经安定下来，有了一整套合理的方法，新的技术和实践被添加进现有方案。”但 Gartner 认为，大数据还有 5—10 年才会达到稳定期。近几年企业级市场围绕大数据的炒作不断加温，很多企业也的确面临数据量激增的现实困境。大数据绝非仅仅是数据量大的挑战，其核心问题还是取决于数据挖掘背后所能产生的价值，在经历了一段热潮之后，大数据的概念转向谷底期，开始切实为企业解决问题，大数据相关技术的演进在未来一段时间内仍将展现出强大的生命力。

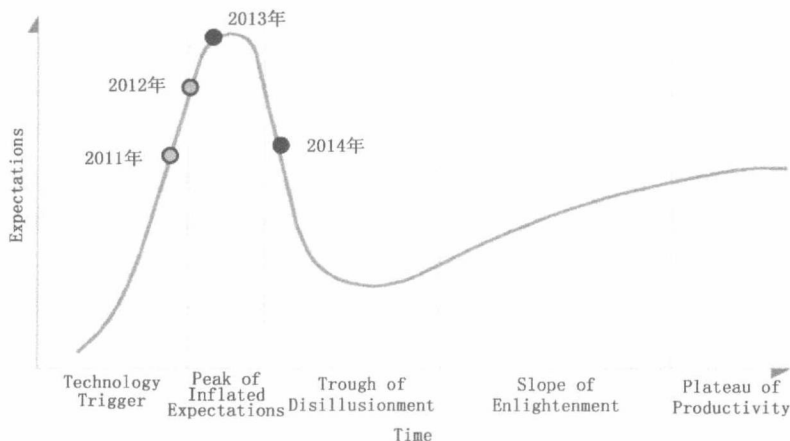


图 3 2011—2014 年大数据技术在技术成熟度曲线上的位置

3.2 大数据技术成熟度曲线

3.2.1 2012 年大数据技术成熟度曲线

2012 年大数据的技术成熟度曲线共列出了近 50 项技术（见图 4），其中大数据革命技术之一 MapReduce 出现在成熟度曲线的谷底期，预期将在 2—5 年内达到生产高峰期。MapReduce 是由 Google 提出的一种并行分布式编程模型，该模型通过简单的分布式并行运算有效处理海量数据，不断从中挖掘出有价值的信息，成为大数据发展的重要支撑技术。根据技术在大数据中的用途可以分为三类：大数据的使能技术（Entries that describe enabling technologies for big data）、典型应用（Entries that describe typical use cases for big data）、信息类型和来源（Entries that describe new information types, sources and roles）（见表 3）。

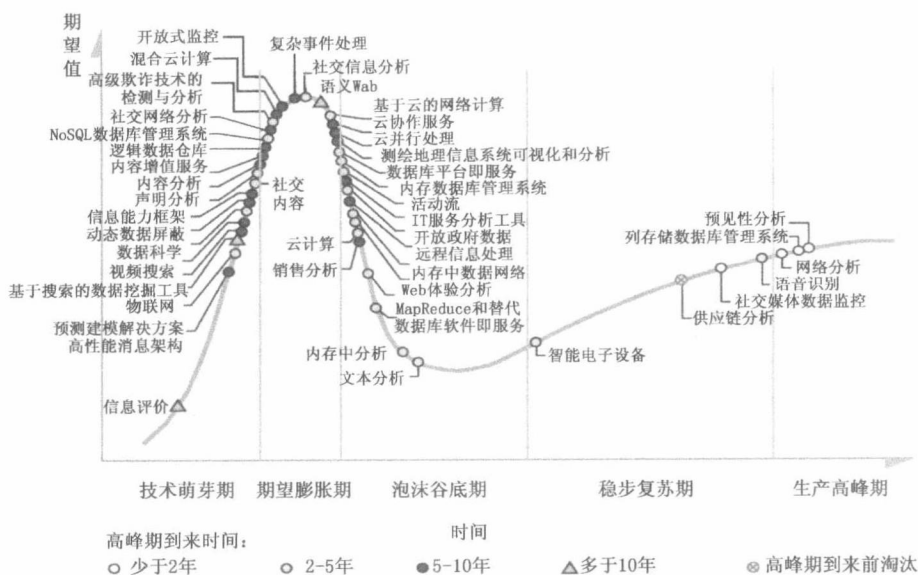


图 4 2012 年大数据技术成熟度曲线

资料来源：Gartner（2012 年 8 月）。

表 3 2012 年大数据相关技术分类

类 别	技术条目
使能技术	高性能消息架构、物联网、基于搜索的数据挖掘工具、视频搜索、动态数据屏蔽、信息能力框架、内容分析、逻辑数据仓库、NoSQL 数据库管理系统、混合云计算、复杂事件处理、基于云的网格计算、云协作服务、云并行处理、数据库平台即服务、内存数据库管理系统、内存中数据网络、MapReduce 和替代、数据库软件即服务、内存中分析、文本分析、智能电子设备、语音识别、存储数据库管理系统、预见性分析
典型应用	信息评价、预测建模解决方案、声明分析、内容增值服务、社交网络分析、高级欺诈技术的检测与分析、社交信息分析、IT 服务分析工具、远程信息处理、Web 体验分析、供应链分析、社交媒体数据监控、网络分析
信息类型 和来源	数据科学、社交内容、开放式监控 SCADA、语义 Web、测绘地理信息系统可视化和分析、活动流、开放政府数据

3.2.2 2013 年大数据技术成熟度曲线

2013 年大数据的技术成熟度曲线共列出了 40 项技术（见图 5），新进入成熟度曲线的技术有：量化自我（Quantified Self）、图形数据库（Graph Database）、基于电子商务的大数据分析（Big Data for Electronic Commerce）、意向联结的客户系统、表式数据库管理服务（Table - Style Database Management Services）、信息语义服务（Information Semantic Services）、HadoopSQL 界面（Hadoop SQL Interfaces）、文件存储数据库管理系统（Document Store Database Management Systems）、关键值数据库管理系统（Key - Value Database Management Systems）、车载信息服务（Vehicle Information Service）和 Hadoop 发布工具（Hadoop Distributions）。

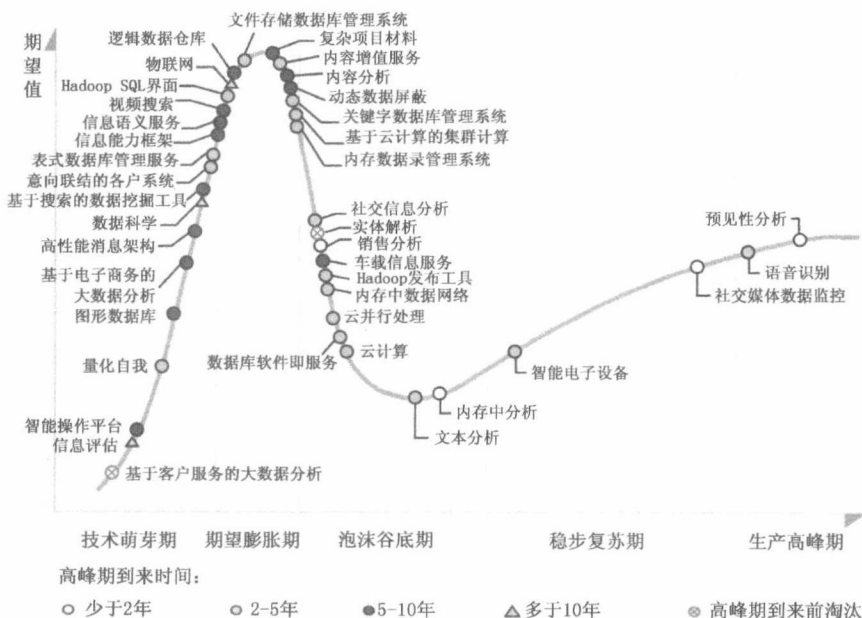


图 5 2013 年大数据技术成熟度曲线

资料来源：Gartner（2013 年 8 月）。

2013 年 MapReduce 技术没有出现在技术成熟度曲线上，而 Hadoop 技术成为大数据关注热点，成熟度曲线上出现了两项与 Hadoop 相关的技术，即 Hadoop SQL 界面和 Hadoop 发布工具。Hadoop 是一个由 Apache 基金会所开发的分布式系统基础架构，用户可以在不了解分布式底层细节的情况下开发

分布式程序，充分利用集群的威力进行高速运算和存储，Hadoop 实现了一个分布式文件系统（Hadoop Distributed File System，HDFS）。Hadoop 框架最核心的设计就是：HDFS 和 MapReduce。HDFS 为海量的数据提供了存储，MapReduce 为海量的数据提供了计算。

3.2.3 2014 年大数据技术成熟度曲线

2014 年大数据的技术成熟度曲线共列出了 47 项技术（见表 4）。新进入成熟度曲线的技术有：语境经纪人（Context Brokers）、虚拟个人助手（Virtual Personal Assistants）、个人分析（Personal Analytics）、数据即服务（Data as a Service）、数据湖（Data Lakes）、公开数据（Open Data）、移动设

表 4 2014 年大数据技术成熟度曲线

阶段	技术萌芽期	期望膨胀期	泡沫谷底期	稳步复苏期	生产高峰期
技术条目	语境经纪人 虚拟个人助手 个人分析 信息价值和信息学 数据即服务 智能操作平台 供应链大数据分析 量化自我 数据湖 基于 Hadoop 的数据挖掘 图形分析 公开数据 基于搜索的数据挖掘工具 图形数据库管理服务 MDM 和社交数据 规范分析	基于客户智能的 大数据分析 数据科学 高性能消息架构 预测分析 智能顾问 信息能力框架 防止欺诈和保护安全的 大数据分析 信息语义服务 表式数据库管理服务 企业信息架构 Hadoop SQL 界面 物联网 逻辑数据仓库 自然语言问答 可穿戴用户界面 文件存储数据库 管理系统 复杂事件处理	关键值数据库 管理系统 内存中数据库 管理系统 内容分析 动态数据屏蔽 远程信息管理 Hadoop 发布工具 实体解析 内存中数据网格 数据库平台即服务 社交信息分析 云计算 关联数据	内存中分析 数据集成/ 虚拟化	

资料来源：Gartner（2014 年 8 月）。

备管理和社交数据 (MDM and Social Data)、规范分析 (Prescriptive Analytics)、基于客户智能的大数据分析 (Big Data Analytics for Customer Intelligence)、智能顾问 (Smart Advisors)、Big Data Analytics for Fraud and Security、企业信息架构 (Enterprise Information Architecture)、自然语言问答 (Natural - Language Question Answering)、可穿戴用户界面 (Wearable User Interfaces)、关联数据 (Linked Data)、数据集成/虚拟化 (Data Federation/Virtualization)。

2014 年新进入的技术涉及信息类型和来源、典型应用的技术较多,在信息类型和来源方面增加了数据湖 (Data Lakes)、公开数据 (Open Data)、移动设备管理和社交数据 (MDM and Social Data)、关联数据 (Linked Data) 等,在典型应用方面增加了基于客户智能的大数据分析 (Big Data Analytics for Customer Intelligence)、为防止欺诈和保护安全的大数据分析 (Big Data Analytics for Fraud and Security) 等。

4 结 语

综上所述,大数据经过 4 年的发展正由膨胀期转向谷底期,内涵逐渐清晰,市场趋向稳定,在未来 5—10 年将达到生产高峰期。综合近 3 年大数据技术成熟度曲线可以勾勒出大数据的整体轮廓,即大数据的应用依托于社交内容、开放式监控 SCADA、语义 Web、测绘地理信息系统可视化和分析、活动流、开放政府数据、公开数据、关联数据等信息数据,综合利用物联网、数据挖掘、视频搜索、动态数据屏蔽、内容分析、逻辑数据仓库、NoSQL 数据库管理系统、混合云计算、复杂事件处理、基于云的网格计算、云协作服务、云并行处理、内存中数据网络、MapReduce、Hadoop、内存中分析、文本分析、数据集成、语音识别、预见性分析等技术工具,实现信息评价、预测建模、声明分析、内容增值服务、社交网络分析、高级欺诈技术的检测与分析、社交信息分析、IT 服务分析工具、远程信息处理、Web 体验分析、供应链分析、社交媒体数据监控、网络分析等技术。



参考文献

- [1] Gartner 发布 2014 年新兴技术成熟度曲线报告勾勒数字化商业发展路径. 现代电信科技, 2014, 44 (9): 76.
- [2] Gartner Inc. Gartner Hype Cycle [EB/OL]. <http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp>, 2014/12/02.
- [3] Gartner Inc. Hype Cycle for Big Data, 2012 [EB/OL]. <http://www.gartner.com/doc/2100215>, 2012/07/31.
- [4] Gartner Inc. Hype Cycle for Big Data, 2013 [EB/OL]. <http://www.gartner.com/doc/2574616>, 2013/07/31.
- [5] Gartner Inc. Hype Cycle for Big Data, 2014 [EB/OL]. <https://www.gartner.com/doc/2814517>, 2014/08/04.

(作者: 王志玲、燕光谱、蓝洁、管泉, 青岛市科学技术信息研究所)

基于专利的跨领域技术特征测度与分析

摘 要

跨领域技术作为整合多领域知识的新兴技术,在统筹资源解决共性问题方面具有绝对优势,是推动技术创新发展的重要力量。以知识从主要 IPC 流向追加 IPC 作为构建技术影响网络的基础,依照世界知识产权组织公布的技术领域分类原则。首先,提出了基于专利的跨领域技术特征测度方法,并根据跨领域特征对技术领域进行分类;其次,利用社会网络分析指标描述不同类型技术领域的重要性及作用,该方法与利用引文信息分析技术跨领域特征相比,构建网络的限制条件更少,与利用文献的学科信息分类相比,更能反映出技术内容;最后,以 3D 生物打印技术为例进行实证分析,验证了该方法的可行性,为进一步研究跨领域技术的演化提供了方法依据。

关键词

跨领域技术 专利数据 跨领域特征 多样性 关联性

1 引 言

随着科学技术的不断发展,学科间的交叉融合与技术的汇聚发展已成为主流,而技术的跨领域研究也以其改变传统规程、实现资源共享等特点得到更多关注。跨领域技术是指为了实现同一个目标,通过不同技术领域深度协

作、有机组合而形成的一类新技术。不同于多学科或跨学科技术，它并非将知识简单地叠加，而是根据技术领域的功能将其有机地整合起来，是一个技术再造和技术创新的过程。正因为跨领域技术整合了多个领域的知识和方法，所以在统筹资源方面具有先天优势，可以解决领域间共性问题，推进技术的创新发展。随着 2014 年年初首都创新大联盟的正式成立，跨领域的研究方法被率先应用于推动产业融合和模式创新中，这是一种全新的尝试。在这种新形势下，政府和研发机构如何根据自身优势及特点高效地利用资源实现产业联动，进而推进跨领域技术的快速发展是亟待解决的问题。

跨领域技术具有四个基本特征：多样性——以多种技术知识结构为基础；功能新颖性——具有单一技术无法实现的新技术功能；构成复杂性——需要不同技术领域专家协同攻破；实施高效性——能有效解决社会、经济科技发展面临的重大问题。目前，关于跨领域技术特征评价的研究多以学科分类为主，研究的内容主要集中于评价原则、评价指标及评价方法三个方面。在评价原则方面，国内学者刘凡丰、余诗诗等指出评价跨学科研究时需要充分考虑跨学科研究的属性，同时要保证学术性贡献和应用性贡献、个体贡献和团队贡献、针对性和灵活性之间的平衡。在评价指标方面，以 Porter 为代表的研究团队提出了跨领域研究程度评价内容；Leydesdorff 根据科技文献的引文关系绘制了学科领域的分布地图，为使用引文网络进行跨领域研究打下了基础。此后，Ismael 与 Meyer 提出了跨领域特征的评价可以从多样性（diversity）和关联性（coherence）两方面把握，其中多样性包括相关知识主体的种类（variety）、均衡性（balance）以及差异性（disparity）三个指标；关联性则是指在研究过程中使用的具体主题、工具、数据等的相关程度。在此基础上，Stirling 根据跨领域多样性评价的内容提出了基于引文网络映射出学科种类、均衡性和差异性的度量方程，Rafols 和 Meyer 完善了该度量方程，提高了跨领域多样性测算的准确性。

专利同类分析（Patent Co-classification Analysis）是指使用专利分类代码来识别技术知识在不同技术间转移的方法。在美国专利分类（United States Patent Classification, USPC）体系中，每件专利被赋予了一个或多个专利代码，这些分类代码被归为两类：主要分类和追加分类。一件专利可以拥有一个主要分类和多个追加分类，前者被认为是专利发明关注的技术领域，而后者则是指专利发明涉及却并不关注的附加技术领域。目前，专利同类分析法多用于测量产业间的技术溢出情况，不仅可以用于识别产业部门间的技

术媒介,还可以识别核心技术。我国学者雷滔、陈向东尝试以专利的主分类和追加分类号作为切入点研究信息和通信技术 (Information and Communication Technology, ICT) 领域和非 ICT 领域间的跨学科技术关联;张宪义则使用专利同类分析研究了技术领域间的联系及变化趋势,通过引入时间变量分析发现技术领域间的关系随时间的推移会出现融合和分化的现象。外国学者 Yoon 和 Kim 分别验证了专利同类分析方法在应对专利数据较新、缺乏引文数据情况时具有的优势,这种优势使专利同类分析作为一个替代引文网络测量知识在不同技术间流动的方法越来越受到关注。

决策实验室分析法 (Decision - making and Trial Evaluation Laboratory, DEMATEL) 由 Battelle Memorial 研究所于 1971 年提出,是在多因素决策问题中协助决策者判断决策因素重要性的矩阵计算方法。它既可以识别因素间的直接和间接影响,又可以分别测量各因素间的影响强度。国内学者徐国东、郭鹏利用 DEMATEL 方法对问卷数据进行处理,便于在知识联盟中识别知识转移的关键影响因素;王莉使用 DEMATEL 方法辨识制造业产业技术创新的影响因素。然而这些研究在使用 DEMATEL 方法时无法避免评价中带有主观因素,直到 2014 年韩国 KO 团队提出了一个克服传统 DEMATEL 方法中主观因素干扰的办法,将技术类别的直接关系矩阵作为 DEMATEL 方法的初始输入矩阵,使得技术影响力的判断更加客观而准确。

社会网络分析 (Social Network Analysis, SNA) 是一种映射参与者间关系、并测量其整合度的常用方法。在技术分析方面,该分析法在直观地显示节点间联系和可视化结果输出方面具有优势,被广泛地应用在节点影响力的判断和网络动力学检验中。外国学者 Kim 使用中心度指标从各个角度衡量一个节点的重要性;Yan 和 Ding 利用引文关系构建技术影响力网络,合成相关指标测量节点间的相似性和关联性。

综上所述,目前国内外关于跨领域技术特征的研究多数是通过科技文献和专利文献的引文网络开展,以学科分类替代技术领域分类。然而,基于引文网络的分析方法前提是要有引文信息,通常论文的参考文献著录完整规范,但专利的引文信息因为不是必须著录项,很多专利数据库中不能提供所有专利的完整引用信息,这也就使得利用专利引用信息构建技术影响网络测量跨领域特征的方法受到较多的限制;在技术影响力的评价方面,多数研究不能在判断技术领域影响力的同时考虑技术领域间的直接和间接影响。鉴于使用传统的引文网络及社会网络分析法开展跨领域特征评价时所存在的问

题,本文依据知识是从专利的主要国际专利分类(International Patent Classification, IPC)流向追加 IPC 这一前提,提出了使用专利数据测算跨领域特征的方法,在此基础上根据技术领域的跨领域特征对其进行分类,并利用社会网络指标深入分析不同类型技术领域在网络中的重要性及作用。

2 研究方法

本文从技术知识在专利主要 IPC、追加 IPC 间的流动出发,利用专利同类分析和 DEMATEL 方法构建技术领域间的总体影响网络,提出基于专利数据的跨领域特征测度方法,并根据测算结果对技术领域进行分类。在此基础上使用社会网络分析指标进一步探讨跨领域技术中各类技术领域的重要性及作用。

研究步骤包括:

第一,将专利的主要 IPC、追加 IPC 分离,依据世界知识产权组织(World Intellectual Property Organization, WIPO)公布的技术领域分类准则,把分离得到的主要、追加 IPC 归到 35 个技术领域中,构建直接关系矩阵,再使用专利同类分析法与决策实验室分析法用于生成技术领域的总体影响网络;

第二,基于技术领域间的总体影响网络,使用专利数据测算各技术领域的跨领域特征指标(多样性和关联性);

第三,依据技术领域的跨领域特征对各技术领域进行分类(见图 2),再使用社会网络指标进一步分析各类技术领域的特点;

第四,以 3D 生物打印技术为例进行实证分析,验证了该方法的可行性,为进一步研究跨领域技术的演化提供依据。

本文的技术路线见图 1。

2.1 构建总体关系矩阵及技术领域影响网络

IPC 是 WIPO 制定的便于查找和管理专利技术分类的系统,它由部、大/小类和部/小组五级构成。本文为了便于将主要和追加 IPC 限定到具体的技术领域中,采用 IPC 的前四位进行分析(即小类,如 A05B)。由于每件专利包含了若干个 IPC,需要利用 C 语言编程对得到的专利数据进行主要和追

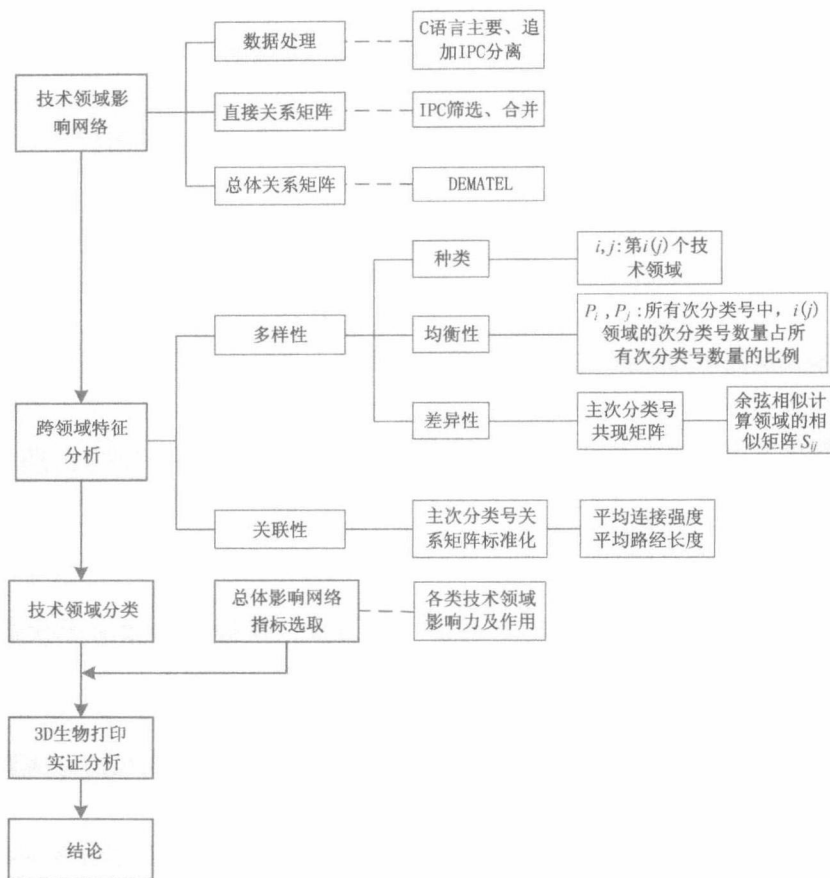


图1 技术路线

加 IPC 的分离, 然后根据 WIPO 的技术领域划分准则把 IPC 划分到 35 个技术领域, 从而构建技术领域间的直接关系矩阵 Z , 矩阵的行与列分别为主要和追加 IPC 所在的技术领域 (以下分别简称“主要技术领域”和“追加技术领域”), 矩阵内的元素表示专利中知识从主要技术领域 i 流向追加技术领域 j 的次数。再把矩阵 Z 作为 DEMATEL 的初始矩阵进行变换: 首先对 Z 进行标准化处理得到标准化矩阵 M , 然后使用 (1) 式计算求得技术领域间的总体关系矩阵 T , 这样矩阵 T 包含了技术领域间的直接和间接影响。

$$T = M(I - M)^{-1} \quad (1)$$

最后, 利用得到的总体关系矩阵 T 生成跨领域技术各技术领域间的总体影响网络。网络可视化以及指标的计算过程借助 Ucinet 和 Netdraw 两个软

件完成。

2.2 基于专利数据的跨领域特征测算

Stirling 的度量方程指出,跨领域多样性的特征评价可以映射出某技术整合学科领域的情况。“多样性”是指技术涉及的知识主体“数量”“均衡性”与“差异性”,这一指标与跨领域程度呈正比。其中“均衡性”显示了各技术领域的比重,不同技术领域的占比越接近,均衡程度也就越高,跨领域程度也就越强;“差异性”是指不同技术领域的差别,可以体现在理论、研究方法、研究工具等方面,它反映了技术领域间的相似程度,相似程度越低,差异性越大,也就意味着跨领域程度越高。跨领域技术的“关联性”体现的是技术领域间的紧密程度,紧密度越高,跨领域程度也就越高。而在指标的计算方面,多样性通常是借助引用报告的学科分类及学科间的引用关系计算,将科技文献与学科领域关联起来进行测量;关联性的评价指标则通常使用学科间的平均连接强度或平均路径长度来表示。本文借鉴了以上研究方法,提出了基于专利的主要和追加 IPC 测度跨领域“多样性”“集成性”的方法,具体测算指标及内容如下:

首先,本文明确了以专利主要 IPC 作为技术领域界定的前提条件,虽然专利拥有多个 IPC,每个 IPC 又可能归属不同的技术领域,但只根据专利唯一的主要 IPC 来认定专利所属的技术领域。在多样性测算方面:知识主体“数量”的测算以专利 IPC 涉及的技术领域数量表示;均衡性为某一技术领域内所有专利的追加 IPC 在各技术领域内分布的均匀程度;差异性通过计算主要和追加技术领域之间的距离来表示,为了便于计算,可以使用 SPSS 软件计算技术领域之间的余弦相似性,再用 1 减去包含相似性的计算结果间接求得多样性得分 I ,见公式(2):

$$I = 1 - \sum_{i,j} d_i s_{ij} d_j \quad (2)$$

公式(2)中, i 、 j 表示第 $i(j)$ 个技术领域; $d_i(d_j)$ 表示属于领域 $i(j)$ 的追加 IPC 数量占有追加 IPC 数量的比例; s_{ij} 表示技术领域间的相似性矩阵,而相似性矩阵可从技术领域的直接关系矩阵 Z 中获得,再利用 SPSS 软件计算求得各技术领域间的余弦相关系数矩阵即可。

在关联性测算方面,本文沿用了 2013 年国内学者郑锐萍提出的以接近中心度(Closeness Centrality)表示技术领域间紧密程度的方法。首先创建