

第一部分 研究方法

全球工程前沿研究在以数据为支撑、专家为核心的原则下，确定了数据与专家多轮交互、迭代决策的研究方法，实践了定量分析与定性分析的深度融合，共遴选获得 2018 年度 94 个工程研究前沿和 96 个工程开发前沿，9 个领域的前沿数量分布如表 1.1.1 所示。

在定量分析上，研究依托科睿唯安基于共被引聚类方法获取了 9 个领域 702 个文献聚类主题和 9 个领域 53 个专业学科组的专利地图。

在定性分析上，9 个领域分别在数据源确定之前和数据挖掘之后两次征集专家提名前沿，以期对定量分析查漏补缺。各领域专家经过多轮研讨以及问卷调查，最终遴选出 2018 年度本领域 10 个左右工程研究前沿和 10 个左右工程开发前沿，并从中选出 3 个研究前沿和 3 个开发前沿进行重点解读。

1 工程研究前沿的遴选

本报告中，工程研究前沿的基础素材主要来自

以下两种途径，一是科睿唯安基于 Web of Science 核心合集的 SCI 期刊论文和会议论文数据，通过共被引聚类方法获得文献聚类主题；二是专家提名备选工程研究前沿。两种途径获得的前沿经过专家论证、提炼得到 482 个备选工程研究前沿，再经过问卷调查和多轮专家研讨，遴选得出每个领域 10 个左右工程研究前沿。

1.1 文献聚类主题的获取和筛选

科睿唯安将 Web of Science 学科与工程院 9 个学部领域建立映射关系，获得每个领域对应的期刊和会议列表，经领域专家补充确认，确定 9 个领域分析用数据源共计 12 882 种期刊和 28 626 个会议。此外，对于《Nature》等 67 种综合学科的期刊，采用单篇文章归类的方法，即根据期刊内单篇文章的参考文献主要归属的学科来定义这篇文章的领域学科。在此基础上，检索得到 2012—2017 年上述期刊和会议收录的论文，数据采集时间为 2018 年 2 月。

对于每个领域，根据每篇文章所在的 Web of

表 1.1.1 9 个领域前沿数量分布

领域	工程研究前沿 / 个	工程开发前沿 / 个
机械与运载工程	10	10
信息与电子工程	10	10
化工、冶金与材料工程	12	12
能源与矿业工程	13	14
土木、水利和建筑工程	10	10
环境与轻纺工程	10	10
农业	10	10
医药卫生	9	10
工程管理	10	10
合计	94	96

Science 学科，筛选在所属学科中被引频次位于前 10% 的高影响力论文。并且考虑到期刊与会议的差别以及被引频次受出版年影响等因素，将期刊与会议论文分开筛选，不同年度发表的论文分开筛选，最终获得该领域的数据集。对上述数据集进行共被引聚类分析，得到每个领域的全部文献聚类主题，这是第一步工作。

第二步是从每个领域的文献聚类主题中遴选 50 个左右不相似¹的备选工程研究前沿。为体现前沿的新兴性，对平均出版年在 2016 年之后的聚类主题给予更多的关注。对于平均出版年在 2016—2017 年的聚类主题，按照核心论文的数量、总被引频次、常被引论文占比依次筛选，获得 20 个不相似的文献聚类主题；对于平均出版年在 2016 年之前的聚类主题，按照核心论文的数量、总被引频次、平均出版年、常被引论文占比依次筛选，获得 30 个不相似的文献聚类主题。此外，针对没有聚类主题覆盖的学科按关键词进行定制检索。最终得到 9 个领域 702 个文献聚类主题。

1.2 专家提名研究前沿

专家提名研究前沿是数据挖掘方法的重要补充。

在数据挖掘之前，9 个领域的院士专家提出研究前沿问题，文献情报专家将前沿问题转化为检索式，这是初始数据源的重要组成部分。在数据挖掘之后，针对没有文献聚类主题覆盖的学科，领域专家提供关键词、代表性论文或代表性期刊，用于支撑科睿唯安进行定制检索。

在专家研判阶段，领域专家对照科睿唯安提供的文献聚类结果进行查漏补缺，对于未出现在数据挖掘结果中而专家认为重要的前沿进行第二轮提名，文献情报专家提供数据支撑。

最终，领域专家对数据挖掘和专家提名的工程研究前沿素材进行归并、修订和提炼，获得 482 个

备选工程研究前沿，而后经过问卷调查和会议研讨，遴选出 94 个工程研究前沿。

2 工程开发前沿的遴选

工程开发前沿的基础素材也来自两种途径，一是科睿唯安基于 Derwent Innovation 专利检索平台，对 9 个领域 53 个学科组中被引频次位于各学科组前 5000 个的高影响力专利家族进行文本聚类，获得 53 张专利地图，领域专家从专利地图中解读出备选工程开发前沿；二是专家提名备选工程开发前沿。两种方式获得的 415 个备选开发前沿通过多轮专家研讨、问卷调查，获得每个领域 10 个左右工程开发前沿。

2.1 专利地图的获取和解读

科睿唯安建立德温特手工代码与工程院学部专业划分标准体系的匹配关系，初步确定 9 个领域的专利数据检索范围及检索策略。领域专家对德温特手工代码删减、增补和完善，确定 53 个学科组的专利检索式。专利检索时间范围为 2012—2017 年，专利引用时间截至 2018 年 2 月。

通过 Derwent Innovation 专利平台检索，综合考虑年均被引频次和技术覆盖宽度指标，筛选获得每个学科组对应的前 5000 个高被引的专利家族。利用专利文本间的语义相似度，获得 53 张能快速直观呈现工程开发技术分布的 ThemeScape 专利地图。

各领域专家在文献情报专家辅助下，从专利地图提炼技术开发前沿、归并相似前沿、确定开发前沿名称，得到每个学科组的备选工程开发前沿。

2.2 专家提名开发前沿

在数据挖掘之前，专家根据对领域技术发展

¹ 不相似主题：如果两个聚类主题的核心论文中有 25% 以上不相同，称两个聚类主题不相似。

的认识和见解，提名工程开发前沿的关键问题，文献情报专家将其转化为专利检索式，作为原始数据源的一部分。在数据挖掘之后，专家进行第二轮前沿提名，对照专利地图解读结果进行查漏补缺。

专利地图解读结果与专家提名前沿共同进入专家研判环节，得到 415 个备选工程开发前沿，而后通过问卷调查或专题研讨，获得 96 个工程开发前沿。

在 94 个工程研究前沿和 96 个工程开发前沿中，领域专家从发展前景、受关注程度等角度遴选出 27 个关键的工程研究前沿和 27 个关键的工程开发前沿进行重点解读。

3 术语解释

被引频次：指论文被科睿唯安 Web of Science 核心合集收录的论文引用的次数。

文献（论文）：包括 Web of Science 中经过同行评议的公开发布的研究型期刊论文、综述和会议论文。

高影响力论文：指被引频次在同出版年、同学科论文中排名前 10% 的论文。

文献聚类主题：对高影响力论文进行共被引聚类分析获得的一系列主题和关键词的组合。

核心论文：指与 702 个备选工程研究前沿相关

的高影响力论文。

施引核心论文：指引用核心论文的文献的集合。

平均出版年：指对该文献聚类主题中所有文献的出版年取平均数。

引文速度：引文速度是一定时间内衡量累计被引频次增长速度的指标。在本研究中，每一篇文献的引文速度是从发表的月份开始，记录每个月的累计被引频次。

常被引论文：指引文速度排名前 10% 的论文。

专利地图：通过分析专利文献中的语义相似度，将相关技术的专利聚集在一起，并以地图形式可视化展现，是形象地反映某一行业或技术领域整体面貌的主题全景图。

技术覆盖宽度：指每一篇专利被分入的不同德温特手工代码的数量。该指标可以体现每件专利的技术覆盖广度。

论文比例 / 公开量比例：对于核心论文（专利）中不同国家（地区）或不同机构作者合著的情况，在统计国家（地区）或机构的产出量时，每个国家（地区）或机构都算一次产出，因此论文比例或公开量比例合计超过 100%。

中国工程院学部专业划分标准体系：包含中国工程科学技术（含农、医）的 9 个学部所涵盖的 53 个专业学科，按照《中国工程院院士增选学部专业划分标准（试行）》确定。