

第一章 研究方法

2019 年度全球工程前沿在以专家为核心、数据为支撑的原则下，采用专家与数据多轮交互、迭代遴选研判的方法，实现了专家主观研判与数据客观分析的深度融合，共遴选出 2019 年度 93 个全球工程研究前沿和 94 个全球工程开发前沿，并按发展前景、受关注程度等原则筛选出重点解读的 28 个工程研究前沿和 28 个工程开发前沿。9 个领域组的前沿数量分布如表 1.1 所示。

在具体研究方法流程上分为数据对接、数据分析和专家研判三个阶段。数据对接阶段主要通过领域专家和图书情报专家的交互，明确数据挖掘的范围；数据分析阶段主要通过聚类方法获得基于数据的研究热点和专利地图，并通过专家研读获得工程热点；专家研判主要通过专家研讨、问卷调查等方法确定最终前沿。同时，为弥补数据挖掘中因算法局限性或数据滞后所导致的研究前沿性不足的问题，鼓励领域专家对照数据分析结果查漏补缺，提名前沿。具体实施流程如图 1.1 所示，其中绿色部分以数据分析为主，紫色部分以专家研判为主。

1 工程研究前沿的产生

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 论文数据的获取与预处理

科睿唯安将 Web of Science 学科与中国工程院 9 个学部领域进行匹配映射，获得每个领域对应的期刊和会议列表，经过领域专家修正与补充，确定 9 个领域数据分析的数据源为 10 817 本期刊和 24 330 个会议列表。此外，对于《Science》等 70 本综合学科的期刊，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分配这篇文章的学科领域。在此基础上，检索得

表 1.1 9 个领域的前沿数量分布

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	11	12
能源与矿业工程	12	12
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	10	10
工程管理	10	10
合计	93	94

到发表于 2013—2018 年的上述期刊论文和会议论文，被引用次数截至 2019 年 2 月。

对于各个领域，科睿唯安综合考虑期刊与会议差别、出版年等因素，对上述文献列表进行检索和数据挖掘，将期刊与会议论文分开筛选，筛选出其中被引频次位于前 10% 的高影响力论文，作为研究热点分析的原始数据集。如表 1.1.1 所示。

1.2 文献聚类主题的获取和筛选

通过对上述的数据集中 9 个领域前 10% 的高

被引论文进行共被引聚类分析，得到 9 个领域的全部文献聚类主题。对于出版年在 2017—2018 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 25 个不相似的文献聚类主题；对于出版年在 2017 年之前的聚类主题，按照核心论文数量、总被引频次、核心出版物平均出版年份、常被引论文占比依次进行筛选，提取出 35 个不相似的文献聚类主题。其中各领域聚类主题如有交叉，则递补不交叉的聚类主题。此外，对于没有聚类主题覆盖的学科按关键词进



图 1.1 全球工程前沿研究实施流程图

表 1.1.1 各领域对应的期刊、会议数量及其前 10% 高被引的论文数

序号	领域	期刊数 / 本	会议数 / 个	高被引论文数 / 篇
1	机械与运载工程	457	1768	38 676
2	信息与电子工程	986	9632	109 507
3	化工、冶金与材料工程	1128	2313	219 081
4	能源与矿业工程	226	785	440 641
5	土木、水利与建筑工程	359	512	28 384
6	环境与轻纺工程	1003	605	93 524
7	农业	1575	975	105 523
8	医药卫生	4328	7059	392 142
9	工程管理	755	681	32 927

行定制挖掘。最终得到 9 个领域 806 个文献聚类主题（见表 1.2.1）。

1.3 专家研判

专家提名研究前沿是数据挖掘方法的重要补充。在数据对接阶段，9 个领域的专家提出研究前沿问题，图书情报专家将前沿问题转化为数据挖掘的检索式，这是作为分析数据源的重要组成部分。在数据分析阶段，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，图书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，而后经过网络问卷调查、院士问卷调查和多轮会议研讨，遴选出 93 个工程研究前沿。

2 工程开发前沿的产生

2.1 工程开发热点的遴选

工程开发前沿的基础素材主要来自以下两种途径，一是主要以科睿唯安公司的 Derwent Innovation

专利数据库为原始数据，通过建立德温特专利分类号与中国工程院学部专业划分体系的映射关系，获得分析的基础数据，而后对 9 个领域 53 个学科组被引频次位于前 10 000 的高影响力专利进行主题聚类，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，其中工程管理领域作为单独的学科组进行分析；二是专家提名或小同行专利分析备选工程开发前沿。两种方式获得的备选开发前沿通过问卷调查和多场研讨，最终获得每个领域 10 个左右工程开发前沿。

2.2 专利地图的获取和解读

科睿唯安建立德温特手工代码与中国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。领域专家通过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 2013—2018 年，专利引用时间截至 2019 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 10 000 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。

表 1.2.1 各领域共被引聚类结果统计

序号	领域	聚类主题数 / 个	前 10% 高被引论文数 / 篇	备选工程研究热点数 / 个
1	机械与运载工程	6720	29 960	138
2	信息与电子工程	16 816	76 015	67
3	化工、冶金与材料工程	26 563	116 361	68
4	能源与矿业工程	10 624	47 860	100
5	土木、水利与建筑工程	5 594	25 867	115
6	环境与轻纺工程	18 486	80 850	90
7	农业	7663	33 909	81
8	医药卫生	46 264	203 487	63
9	工程管理	4240	18 321	84

各领域专家在图书情报专家辅助下，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。同时，为避免专利数据挖掘中单纯以数字指标衡量前沿性，领域组专家尤其注重专利地图中低频次、相关性较差的专利解读力度。

2.3 专家提名与研判

为弥补数据挖掘中因算法局限性或数据滞后所导致的开发前沿性不足问题，鼓励领域专家查漏补缺，提名开发前沿。在数据对接阶段，领域专家提出开发前沿关键词与描述，图书情报专家将前沿转化为专利检索式，并作为数据源的重要组成部分。在数据分析阶段，挖掘被统计数据埋没的冷门、偏门、小同行前沿，纠正专利地图偏差。在专家研判阶段，领域专家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，图书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开发前沿素材进行归并、修订和提炼，共获取候选开发前沿 351 个，而后经过网络问卷调查、院士问卷调查和多轮会议研讨，遴选出 94 个工程开发前沿。

3 术语解释

前沿：全球工程前沿研究中的“前沿”是指具有前瞻性、先导性和探索性，对工程科技未来发展有重大影响和引领作用的主要方向，是培育工程科技创新能力的重要指南。

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：工程研究前沿相关联的高影响力论文。

施引核心论文：指引用核心论文文献的集合。

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，每一篇文章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。

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。