

# 基于SAO结构的关键技术要素识别研究<sup>\*</sup>

## ——以高速芯片为例

孔令凯<sup>1</sup> 杨朝均<sup>1</sup> 齐浩<sup>1</sup> 马永红<sup>2</sup> 杨晓萌<sup>3</sup>

(1. 昆明理工大学管理与经济学院, 昆明 650504; 2. 哈尔滨工程大学经济管理学院, 哈尔滨 150001; 3. 江苏科技大学经济与管理学院, 镇江 212003)

**摘要:** 面对日益严峻的技术封锁及遏制风险, 深入技术本身, 掌握技术系统内部关键组件、工艺以及方法等技术要素是突破技术瓶颈的关键环节。归纳关键技术要素的3个特征——关键性、创新性以及基础性。以高速芯片为例, 利用专利文献数据, 分析高速芯片专业术语。引入SAO (Subject-Action-Object) 语义结构, 提取高速芯片技术要素, 以技术要素为节点、技术要素关联强度为边, 构建高速芯片SAO语义网络。借鉴社会网络分析方法, 量化技术要素权利、地位以及创新力, 识别高速芯片系统内关键技术要素, 探究高速芯片技术结构, 寻找技术卡点。实证结果显示, 高速芯片“卡脖子”技术中存在“高速数据”“单片机”“fpga芯片”“晶圆”4个关键技术要素。

**关键词:** 关键技术要素; 高速芯片; SAO结构; 中心性; 结构洞; 覆盖度

**中图分类号:** G250.7 **DOI:** 10.3772/j.issn.1673-2286.2023.10.005

**引文格式:** 孔令凯, 杨朝均, 齐浩, 等. 基于SAO结构的关键技术要素识别研究: 以高速芯片为例[J]. 数字图书馆论坛, 2023 (10): 40-51.

中美技术摩擦加剧, 针对新兴产业以及高端产业的技术封锁政策频出, 关键技术、共性技术随时面临“卡脖子”风险, 直接威胁技术需求企业的生存及国家经济的安全发展, 突破技术封锁是构建新发展格局的首要任务之一。党的二十大报告强调“坚决打赢关键核心技术攻坚战, 加快实施一批具有战略性全局性前瞻性的国家重大科技项目, 增强自主创新能力”, 而关键技术、共性技术是由基础工艺、核心元部件等技术要素构成的复杂技术簇<sup>[1]</sup>, 技术突破难以一蹴而就, 唯有通过技术学习和整合, 掌握技术系统内部关键技术要素, 摆脱技术依赖, 才能拥有持久的创新能力。技术系统中关键技术要素不同于普通技术要素: 关键技术要素是

技术系统的重心, 在技术系统中具有主导性作用, 是突破技术瓶颈的关键。

当前学术界大部分研究将关键技术、共性技术作为技术簇展开研究, 这不利于后发国家理解和攻克技术瓶颈。只有将技术系统“黑箱”打开, 挖掘技术系统内部要素, 深入理解技术要素间关联关系, 识别关键技术要素, 掌握技术系统的“卡点”<sup>[2]</sup>, 获取高质量技术情报, 才能找到技术创新的路径和机制, 补齐技术短板, 突破国外技术封锁。因此, 打开技术系统“黑箱”, 识别技术系统内关键技术要素, 理解关键技术要素关联关系, 对于提升国家及企业自主创新能力, 实现科技自立自强意义重大。

收稿日期: 2023-08-25

<sup>\*</sup>本研究得到国家自然科学基金“区域创新系统网络结构与产业关键共性技术创新涌现的构效关系与作用机理研究”(编号: 71874040)、“基于专利数据挖掘的我国芯片产业关键技术研发路径选择研究”(编号: 72104064)、“基于多源文本挖掘的我国高端装备制造制造业关键共性技术突破路径研究”(编号: 23XJC630006)资助。

## 1 文献回顾

为深入揭示技术系统整体特性,考虑细粒度信息,学者从微观层面提出技术要素是技术系统的最小单元<sup>[3]</sup>,并指出技术要素包含科学效应、组件、方法、材料、材料产品、部件、功能、应用领域等<sup>[4]</sup>,而处于关键地位的技术要素对整个技术系统发展具有制约作用<sup>[5]</sup>,是技术成熟度评估以及实现技术突破的关键,对于技术系统发展意义重大。

对于识别关键技术要素,学者利用不同数据源及方法展开研究。专家经验是关键技术要素识别最早的数据源。早期学者利用专家知识制定关键技术要素候选清单,结合专家打分确定关键技术要素<sup>[6]</sup>,也有学者将数理统计分析方法与专家经验相结合,利用粗糙集和信息熵<sup>[5]</sup>、目标规划方法<sup>[7]</sup>、模糊层次分析方法<sup>[8]</sup>确定关键技术要素。此外,部分学者基于“模块论”以及“链接论”分析技术内在构成<sup>[9-10]</sup>,打开技术系统“黑箱”,例如:武建龙等<sup>[11]</sup>根据技术水平以及功能重要性,将技术功能性模块划分为关键模块以及外围模块,指出突破性技术创新就是突破关键模块技术;洪勇等<sup>[12-13]</sup>基于案例分析方法探讨核心技术链包含的核心元件技术。借助专家经验可准确识别关键技术要素,但实践中存在专家资源稀缺、主观性强、不确定性大,且识别结果粒度较粗等缺陷。

为了弥补利用专家经验识别关键技术要素的缺陷,提升关键技术要素识别精确性,学者利用承载技术信息的专利文献,采用SAO (Subject-Action-Object) 结构<sup>[3, 14]</sup>、LDA (Latent Dirichlet Allocation) 模型<sup>[15]</sup>确定微观层面的技术要素。此后,随着命名实体识别研究的不断丰富,机器学习<sup>[16]</sup>、深度学习<sup>[17]</sup>、迁移学习<sup>[18]</sup>等模型被用于提取技术要素,但该类研究停留于技术要素抽取阶段,缺乏对不同技术要素属性及特征的分析,尚未探讨不同技术要素之间的关联关系,难以确定技术系统内的关键技术要素。不过,引用关系<sup>[19]</sup>、社会网络分析<sup>[20-21]</sup>、专利技术共现<sup>[22]</sup>、专利文献聚类<sup>[23]</sup>、技术路线图分析<sup>[24]</sup>、文本挖掘<sup>[25]</sup>等方法已经应用于关键技术识别。例如:马永红等<sup>[25]</sup>采用LDA模型以及社会网络分析方法识别新材料领域关键共性技术;毛荐其等<sup>[26]</sup>在关键技术概念界定基础上,利用专利共类方法识别关键技术。这些技术识别方法能够在抽取不同技术信息的基础上,探讨不同技术信息之间的共现、相似以及引用等关系,明晰技

术信息的作用与被作用关系,进而为细粒度技术要素抽取以及不同技术要素之间关联关系分析提供了方法指导。

基于以上文献梳理,借鉴现有理论及方法,以专利文献为基础,抽取专利文献中的SAO结构,获取领域内部更细粒度的技术要素,在此基础上构建技术要素SAO网络。结合关键技术要素关键性、创新性以及基础性特征,以度数中心性、中介中心性、接近中心性判断技术要素的关键性,以有效规模、效率、限制度、等级度判断技术要素的创新性,以覆盖度判断技术要素的基础性,并构建关键技术要素识别指标体系。融合技术要素SAO网络以及关键技术要素识别指标体系,探讨不同技术要素在技术系统中的权利、地位以及创新力,选取地位及权利较高的技术要素作为关键技术要素,以此完成关键技术要素识别,掌握技术系统卡点,为突破技术瓶颈提供参考。

## 2 关键技术要素概念及特征

技术要素是构成技术系统并维持技术发展的最小单元,微观层面的技术要素可分为技术、产品、方法、部件、属性等,该层面的技术要素可从专利文献中直接提取,表示为专利文献中的具体词组或短语<sup>[4]</sup>。研究基于微观视角,界定技术要素概念,利用专利文献数据,获取技术簇内不同类型的技术要素。其中,关键技术要素在技术系统中占据关键位置,对其他类型的技术要素具有制约作用,主导整个技术的发展走向,并具有较强竞争力,为其他技术要素发展提供支撑,技术要素覆盖面较广。基于前人的研究,归纳出3个关键技术要素基本特征。

### 2.1 关键性

关键技术要素与不同类型的技术要素相互作用共同构成有机技术整体,而关键技术要素能影响技术整体发展速度及成熟度<sup>[5]</sup>,对整体技术系统具有控制能力,占据技术系统的关键位置,权利较高,决定技术系统能否突破技术瓶颈,是技术发展的关键环节<sup>[27]</sup>,同时决定当前及未来时期的技术发展方向及产品效用,对产业的整体发展以及技术的继承与发展具有深刻意义,关键性特征明显<sup>[28]</sup>。因此,关键性是关键技术要素的重要特征之一。为了定量分析技术要素关键性特征,借鉴栾春娟<sup>[21]</sup>以及马永红等<sup>[25]</sup>对技术地位的研究成

果,采用度数中心性、中介中心性、接近中心性分析不同技术要素的地位,量化技术要素权利。其中:度数中心性反映与技术要素连接的其他技术要素数量总和,较高的度数中心性说明技术要素的凝聚力较高;中介中心性以经过技术要素的最短路径数目进行测度,中介中心性越高,技术要素地位越高;接近中心性表示某技术要素与其他技术要素之间的接近程度,较高的接近中心性表示该技术要素距离其他技术要素较近。当度数中心性、中介中心性以及接近中心性较高时,技术要素居于关键地位。

## 2.2 创新性

关键技术要素往往是高质量技术要素,与关键技术均能表现出较强的竞争力<sup>[29]</sup>,体现技术领域最新的研究方向,提供新的功能、技术标准,并通过扩散应用对产业工艺及其他领域发展进程产生影响<sup>[30]</sup>,带来持久创新力<sup>[31]</sup>。关键技术要素影响国家或地区的发展,国家或地区往往高度重视关键技术要素,旨在促进整体技术发展,提升技术竞争力<sup>[32]</sup>。因此,创新性是关键技术要素的重要特征之一。为分析技术要素是否具备创新性特征,采用结构洞指数进行判断。结构洞占据者可获得更多更新的非重复异质性信息,拥有更多接触知识的机会,具有保持信息和控制信息的优势,具备更强的竞争力和创新力<sup>[33]</sup>。因此,依据不同技术要素的结构洞指数能够判断其创新性。

## 2.3 基础性

关键技术要素作为更细粒度的技术,是整体技术

系统生存的基础,与多种技术要素相互联系<sup>[34]</sup>,为不同类型的技术要素发展提供支撑。若关键技术要素消失,其他技术要素的功能将会受到影响,这波及多个技术子系统,乃至整个技术系统,制约整体技术系统发展。因此,将基础性定义为关键技术要素的基本特征之一。为了定量分析技术要素的基础性特征,借鉴黄鲁成等<sup>[35]</sup>的研究,采用技术覆盖度指标测度技术要素基础性。技术覆盖度最早由Geum等<sup>[36]</sup>提出,该指标能够衡量技术系统中某技术要素辐射其他技术要素的范围:技术覆盖度越高,技术要素基础性越强。

## 3 关键技术要素识别模型

### 3.1 关键技术要素识别流程

根据上文总结的关键技术要素特征以及文本挖掘方法,构建研究框架,具体流程如图1所示。

(1) 数据收集。以智慧芽数据库(PatSnap)为数据源,结合IPC分类号以及关键词,制定专利检索表达式,收集专利文献数据。

(2) 构建核心词集,筛选SAO结构。利用词频-逆文档频率(TF-IDF)算法,构建核心词集;实施专利摘要词性标注以及句法依存度分析,提取SAO结构,结合核心词集以及SAO结构相似度,筛选和整合SAO结构。

(3) 构建SAO网络。筛选和整合SAO结构,以技术要素为网络节点,技术要素关联强度为边,构建技术要素SAO网络。

(4) 关键技术要素识别。根据SAO网络,采用中心性、结构洞指数以及技术覆盖度分析技术要素特征,识别关键技术要素。

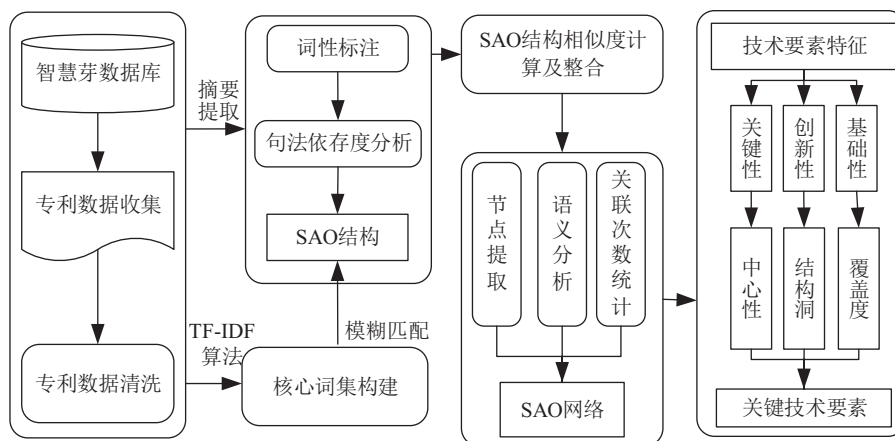


图1 关键技术要素识别流程



## 3.2 数据收集及预处理

核心技术不一定会申请专利,但核心专利往往承载着核心技术<sup>[37]</sup>,蕴含丰富的技术信息。因此,将处理后的专利文献作为关键技术要素识别的数据源。智慧芽数据库收集世界范围内多个国家或地区的专利数据,数据库涵盖50多个高科技产业,每条专利文献记录包含专利题目、专利摘要、专利引用文献、专利价值、专利申请人、专利申请国家等信息,并且数据库处于不断更新状态,为识别关键技术要素提供了可能,因此,智慧芽数据库被用于收集专利文献数据。专利摘要是专利全文的总结,信息表达凝练,因此,将从专利摘要中提取SAO结构。在专利文献收集基础上,Python软件被用于剔除摘要空白以及重复的专利文献;此外,NLTK术语库为多种语言提供了使用界面,将其应用于专利摘要词干提取及词性还原,实现专利文献预处理;利用TF-IDF算法测度不同专利文献中词语的重要性,构建核心词集。

## 3.3 技术要素提取及SAO网络构建

### 3.3.1 技术要素提取

SAO结构强调“关键概念”,在抽取技术要素的同时,保留语义信息,描述文本中语义之间的关联关系,

较好获取专利技术文献技术要素信息;与实体命名技术相比,考虑不同技术要素之间语义关系,涵盖丰富的技术信息。SAO结构由Subject、Action、Object 3个部分构成<sup>[38]</sup>,如图2所示:Subject和Object表示主语和宾语,表现为不同类型的技术关键词或短语,而技术关键词或短语可表示微观层面的技术要素;Action表示谓语,描述技术系统内技术要素的关联关系及其技术要素功能的实现方式。现有研究利用Stanford parser、Goldfire、Alcheml API、ClausIE等软件提取文献SAO结构,综合考虑各软件优缺点,Stanford parser被用于实现专利文献数据词性标注,分析句子结构及依存度,结合团队制定的SAO抽取规则,实现专利文献SAO结构提取。

关联关系包含直接关系和间接关系两大类,其中直接关系包含共现关系和语义关系,而语义关系能够依据SAO结构中的Action识别,根据Action可将语义关系划分为包含类、功效类、用途类以及相互作用类4种类型。借鉴李晓曼等<sup>[3]</sup>研究梳理的包含类、功效类、用途类以及相互作用类包含的核心动词,结果如表1所示。



图2 SAO结构中技术要素关系

表1 关联关系类型及核心动词

关联关系类型	核心动词	关联关系类型	核心动词
包含类	comprise; have; include; contain; involve	用途类	be used for; (be) used in
功效类	improve; increase; reduce; promote; degrade; achieve; realize	相互作用类	prepare; provide; develop; form; produce

### 3.3.2 SAO结构预处理

基于Python软件以及团队制定的SAO提取规则,得到若干SAO结构,为降低无效SAO结构对于后续研究的影响,实施SAO结构预处理。首先,通过TF-IDF算法抽取核心词,制定核心词集,并依据核心词集剔除无关SAO结构。其次,为保障筛选SAO结构的有效性,选择失效率指标评价SAO结构的有效性,其中失效率指标是指抽取的SAO结构中无效的、不可接受的SAO结

构的占比,失效率越高,SAO结构有效性越差。最后,为整合相似度较高的SAO结构,结合词义相似度计算SAO结构相似度,合并同一篇专利文献中的SAO结构。TF-IDF算法、失效率指标计算公式如式(1)、式(2)所示。借鉴已有研究<sup>[39]</sup>,按照式(3)和式(4)展开SAO结构相似度测算,通过SAO结构整合,得到语义区别明显的SAO结构。

$$F = \frac{f_{A,k}}{W_k} \times \ln \frac{D_a}{D} \quad (1)$$

$$R_{\text{invalid}} = \frac{n_{\text{invalid}}}{n_{\text{actual}}} \times 100\% \quad (2)$$

$$s(c_a, c_b) = \frac{2 \times d(c_a, c_b)}{d(c_a) + d(c_b)} \quad (3)$$

$$s_{\text{SAO}} = (s_{\text{SO}}, s_{\text{A}}) / 2 \quad (4)$$

式中:  $F$  表示词频-逆文档频率,  $f_{A, k}$  表示 Action 在文档  $k$  中的出现频次,  $W_k$  表示文档  $k$  中的单词总数,  $D_a$  表示出现单词  $a$  的文档数量,  $D$  表示文档总数;  $n_{\text{invalid}}$ 、 $n_{\text{actual}}$  表示无效的、不可接受的 SAO 结构数量和 SAO 结构总数;  $s(c_a, c_b)$  表示两个词语的相似度,  $d(c_a, c_b)$  表示词语在 WordNet 概念层次树中的共有最小深度,  $d(c)$  则表示单个单词的深度<sup>[40]</sup>;  $s_{\text{SO}}$ 、 $s_{\text{A}}$  分别表示 SAO 结构各部分的相似度。

### 3.3.3 SAO网络构建

为构建技术要素 SAO 网络, 以由 SAO 结构提取的技术要素为网络节点, 节点间的关联强度为边, 构建 SAO 网络<sup>[41]</sup>。节点之间关联强度由 Action 语义重要度及关联次数决定, 其中 Action 语义重要度使用 TF-IDF 算法测度, 关联次数则通过统计获取, 计算二者的乘积, 得到特定 Action 下技术要素关联强度, 并对所有 Action 下特定技术要素关联强度求和, 得到 SAO 网络中技术要素间的关联强度, 如式 (5) 所示。

$$S_{\text{S-O}} = \sum r_{\text{S-A-O}} \times F \quad (5)$$

式中:  $S_{\text{S-O}}$  为技术要素关联强度,  $r_{\text{S-A-O}}$  为特定 Action 下技术要素关联次数。

## 3.4 关键技术要素识别指标选择

### 3.4.1 技术要素关键性测度指标

根据上文关键技术要素关键性特征分析, 采用度数中心性、中介中心性、接近中心性量化不同技术要素的权利, 探讨不同类型的技术要素在技术系统中的地位。中介中心性、接近中心性测度方法如式 (6) 和式 (7) 所示, 而度数中心性则是与技术要素连接的其他技术要素数量总和。

$$B_i = \frac{1}{(N-1)(N-2)/2} \sum_{s \neq i \neq t} \frac{g_i}{g_{st}} \quad (6)$$

式中:  $B_i$  表示技术要素  $i$  的中介中心性,  $N$  表示技术

要素总数,  $g_i$  表示技术要素  $i$  上的最短路径数量,  $g_{st}$  表示技术要素  $s$  和  $t$  的最短路径数量。

$$C_i = \frac{N-1}{\sum_{j=1}^N d_{ij}} \quad (7)$$

式中:  $C_i$  表示技术要素  $i$  的接近中心性,  $d_{ij}$  表示技术要素  $i$  和  $j$  的最短距离。

### 3.4.2 技术要素创新性测度指标

采用结构洞指数衡量技术要素的创新性特征, 而结构洞包含有效规模、效率、限制度以及等级度 4 个典型指标。

(1) 有效规模。技术要素有效规模等于该技术要素的个体网络规模减去网络冗余度, 有效规模越大可能结构洞越多, 如式 (8) 所示。

$$\delta_{\text{ES}} = \sum_j (1 - \sum_q p_{iq} m_{jq}) \quad (8)$$

式中:  $j$  与  $i$  表示连接的技术要素,  $q$  是除  $i$ 、 $j$  之外的第三者;  $p_{iq}$  代表  $i$  投入到  $q$  的关系所占的比例,  $m_{jq}$  代表  $j$  投入到  $q$  的关系所占的比例,  $p_{iq} m_{jq}$  代表  $i$  和  $j$  之间的冗余度。

(2) 效率。效率等于技术要素的有效规模与其个体网络实际规模之比。

(3) 限制度。限制度描述网络中技术要素对其他技术要素的依赖程度, 限制度越低, 技术要素网络能力越强, 如式 (9) 所示。

$$\delta_{\text{RD}} = (p_{ij} + \sum_q p_{iq} p_{pj})^2 \quad (9)$$

式中:  $p_{ij}$  代表  $i$  直接投入到  $j$  的关系占总关系的比例,  $\sum_q p_{iq} p_{pj}$  代表间接投入。

(4) 等级度。等级度表示约束性在多大程度上集中在某个技术要素上, 如式 (10) 所示。

$$\delta_{\text{H}} = \frac{\sum_j \left( \frac{\gamma_{ij}}{\gamma/T} \right) \ln \left( \frac{\gamma_{ij}}{\gamma/T} \right)}{T \ln T} \quad (10)$$

式中:  $T$  表示  $i$  的网络规模,  $g$  表示技术要素约束性。

### 3.4.3 技术要素基础性测度指标

基础性是关键技术要素的另一基本特征, 采用覆盖度指标进行测度, 覆盖度计算公式如式 (11) 所示。

$$\delta_{\text{CI}} = \sum_{j=1}^N V_j \quad (11)$$

式中:  $\delta_{\text{CI}}$  表示技术要素  $i$  的覆盖度,  $V$  取值为 0 或 1。

当技术要素 $i$ 与技术要素 $j$ 存在于同一SAO结构中时,若SAO结构为主动语态,技术要素 $i$ 位于Subject位置,而技术要素 $j$ 位于Object位置,则 $V=1$ ,否则取0;若SAO结构为被动语态,技术要素 $i$ 位于Object位置,而技术要素 $j$ 位于Subject位置,则 $V=1$ ,否则取0。

## 4 高速芯片关键技术要素识别案例

随着物联网和5G时代到来,人们对电子产品的功能、体积及操作速度的要求提高,芯片作为电子设备的核心,其操作速度和集成度也不断提升,市场对于高速芯片的需求随着时间推移持续上升<sup>[42]</sup>。高速芯片是一种将高性能、低耗能的功能器件和系统通过大规模半导体集成的方式制作在单一衬底上而形成的系统级芯片,相较于普通芯片可以进行更高速的数据处理和通信<sup>[43]</sup>。然而,知名分析机构IC Insights的相关数据统计显示,2021年中国芯片进口额高达4 397亿美元,占全球芯片销售额的77.8%,国内芯片自给率仅为16.7%,其中仅低速光芯片和电芯片已经实现国产,而需求较高的高速芯片仍需依赖进口。芯片制造原材料、关键设备等供应受制于人,关键核心技术陷入“卡脖子”的困境<sup>[44]</sup>。对此,选择高速芯片作为研究对象,探究高速芯片技术结构,识别高速芯片关键技术要素,分析高速芯片技术要素间的关联关系,获取高速芯片技术信息,为突破高速芯片技术提供参考。

### 4.1 高速芯片专利数据收集与处理

高速芯片技术领域复杂程度高、更新速度快,通过单一的关键词或IPC分类号难以全面检索高速芯片专利数据。查阅高速芯片文献,结合专家意见,确定高速芯片关键词。此外,高速芯片专利分布在多个IPC分类号下,包含A61B、A61K、B01L、C12Q、C07K、G06B、G06F、G06K、G01R、G01N、G09G、G09Q、H01L、H04L、H04B、H02J、H05B、H04W、H04N。结合关键词及IPC分类号制定检索表达式TACD\_ALL: (high speed OR high velocity OR high-speed OR express OR high rate OR clipping OR fast) AND (microchip OR chip OR microarray OR microfluidics OR MEMS) AND IPC: (A61B OR A61K OR B01L OR C12Q OR C07K OR G06B OR G06F OR G06K OR G01R OR G01N OR G09G OR G09Q OR H01L OR

H04L OR H04B OR H02J OR H05B OR H04W OR H04N),利用智慧芽数据库进行数据收集。截至2021年7月8日,共收集1965—2021年有关高速芯片的专利文献记录22 933条,剔除专利记录缺失数据,剩余高速芯片专利文献记录共19 074条。较早研究高速芯片的国家有美国、德国、日本以及韩国,其保持较强的研发态势,专利申请量稳居前列;此外,新加坡、中国的高速芯片专利申请量不断增加,尤其中国的高速芯片专利申请量自2010年起显著增加。

## 4.2 高速芯片专利SAO结构整合及网络构建

### 4.2.1 高速芯片核心词集构建

为保证提取的专利SAO结构与高速芯片相关,需构建高速芯片核心词集,根据核心词集删除无关SAO结构。核心词集构建包含两步:首先,采用Python软件的NLTK术语库实现高速芯片专利文献词性还原,并利用TF-IDF算法计算各词语重要度;其次,筛选权重较高的词或词组,反复与电子信息科学与技术专业博士生进行商讨,剔除与高速芯片领域无关的词或词组,最终确定930个高速芯片核心词或词组。可视化重要度排名前100的词语,确定核心词集,结果如图3所示。

### 4.2.2 高速芯片专利SAO结构提取及有效性分析

为提取专利SAO结构,利用Python软件,结合Stanford parser模块,实现专利文献摘要词性标注,完成摘要句法依存度分析,基于SAO提取规则,提取专利文献SAO结构。从19 074条专利文献记录摘要中提取673 258个SAO结构,为检验高速芯片专利SAO结构提取的准确性,结合构建的高速芯片核心词集,采用模糊匹配原则:当且仅当高速芯片核心词位于SAO结构中时,认为匹配成功,保留该条SAO结构。经过筛选,保留有效的SAO结构共45 631个。经过模糊匹配过滤大部分关联性不强的SAO结构,为进一步评估SAO结构的有效性,随机抽取10%的SAO结构共4 563个,进行人工研判,其中无效SAO结构共20个,失效率达到0.4%,失效率比较低,表明抽取的高速芯片专利SAO结构有效,为后续关键技术要素提取奠定基础。

高速芯片SAO结构有效性评估后,结合Python软件测算SAO结构间的语义相似度,整合相似度较高的



SAO结构。当两个SAO结构相似度高于0.7时,整合两个SAO结构,提高后续研究测算速度。经过语义相似度

整合,共得到25 943个SAO结构。高速芯片技术领域部分SAO结构及频次如表2所示。

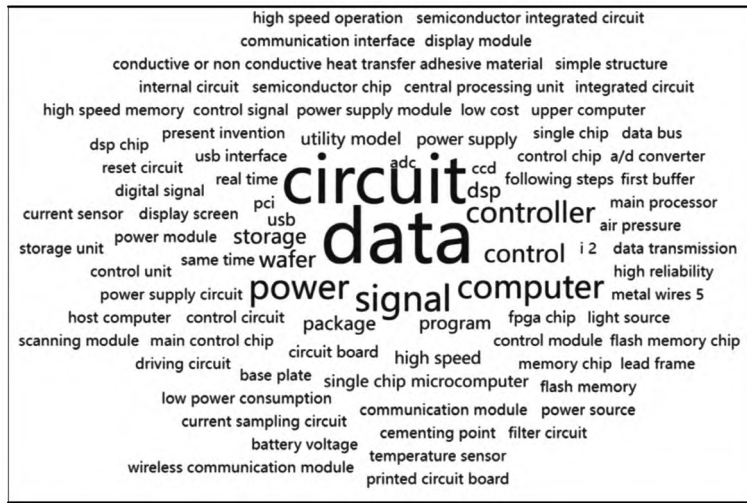


图3 高速芯片核心词集可视化结果

表2 高速芯片专利文献SAO结构(部分)

Subject	Action	Object	SAO结构频次/次
analog digit convert	comprise	4g commune module	7
dsp chip	connect	reset circuit	6
liquid phase chip	comprise	asp primer	12
liquid phase chip	comprise	microsphere	12
microsphere	coat	asp primer	8
single chip microcomputer	connect	sequence	7
single chip microprocessor	connect	keyboard	11
system	comprise	single chip microcomputer	16
usb interface	connect	compute	6

#### 4.2.3 高速芯片SAO网络构建

基于筛选整合的SAO结构,抽取SAO结构中的Subject、Action、Object部分,以Subject和Object为网络节点,计算各节点间语义重要度及关联次数,根据式(5)测算不同技术要素之间的关联强度。同时,因为同一个SAO结构多次出现可能会提升某些技术要素之间的关联强度,为减弱这种效应,在构建SAO网络过程中,对来自同一篇专利文献的相似SAO结构进行了合并处理。具体来说,如果两个SAO结构的Subject和Object词义高度相似,则将SAO结构进行整合,避免同一文献中重复的SAO结构造成关联强度放大效应。需要说明的是,来自不同文献的相似SAO结构仍然保留,以反映各篇文献对技术要素关系的判断。通过以上处理,提高SAO网络中技术要素关联强度的可靠性,进而构建高速芯片技术要素SAO网络。由于提取的技术要

素众多,过滤度数中心性较低的节点,仅展示高速芯片领域部分技术要素节点及关联关系,利用Gephi软件导出可视化结果,如图4所示。图4中连线代表技术要素之间的关联,连线越粗表示技术要素的关联关系越强。技术要素single chip microcomputer(单片微机)、high speed data(高速数据)、wafer(晶圆)、main control chip(主控芯片)、fpga chip(fpga芯片)等出现在多个SAO结构中,与多个技术要素相互作用,表现出较高的活跃度,在SAO网络中具有较高地位。

### 4.3 高速芯片关键技术要素识别

#### 4.3.1 高速芯片技术要素关键性测度

综合考虑技术要素度数中心性、中介中心性、接近中心性,准确量化高速芯片技术要素地位,评估高速芯

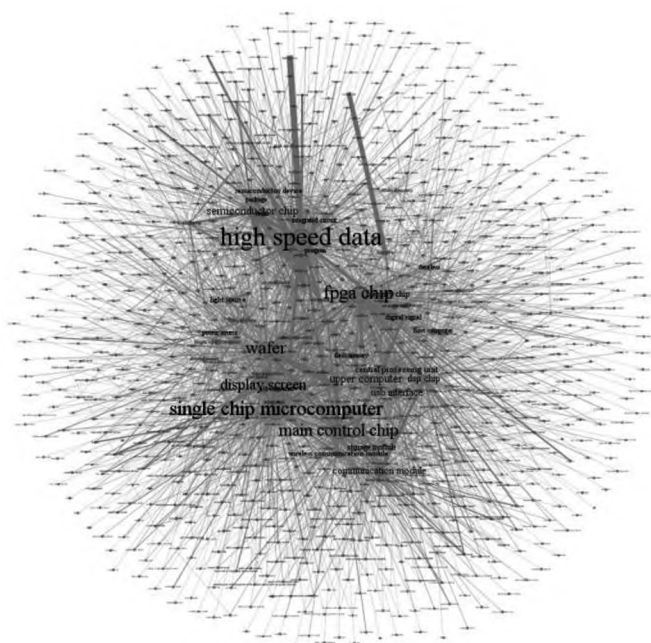


图4 高速芯片SAO网络

片技术系统中地位较高、凝聚力较强的技术要素。结合高速芯片SAO网络结构,筛选度数中心性排名前20的高速芯片技术要素进行展示,结果如表3所示。

由高速芯片领域技术要素关键性测算结果可知,技术要素“高速数据”度数中心性、中介中心性以及接近中心性排名均为第1,说明“高速数据”在高速芯片SAO网络中拥有较高权利,位于网络中心,其与不同类型的技术要素相互作用,影响技术系统发展,旨在提升芯片运算与数据传输速度。技术要素“单片微机”度数中心性、接近中心性排名第2,而中介中心性排名第3,说明该技术要素在网络中处于核心地位。单片微机由于具有体积小、高度集成、数据运算速度快等优点,目前已经被广泛应用于工业控制领域。此外,“fpga芯片”度数中心性排名第3,而中介中心性以及接近中心性排名第4,排名靠前。相对于其他类型的芯片,fpga芯片能够针对特定领域或者产品进行优化设计,技术应用广泛。此外,高速芯片领域关键性较高的技术要素还包含“主控芯片”“显示屏”“晶圆”等。

### 4.3.2 高速芯片技术要素创新性测度

采用结构洞指数衡量高速芯片技术要素创新性,包括有效规模、效率、限制度、等级度4个指标。通过综合运用多个指标能更好地量化高速芯片技术要素创新性,准确把握各技术要素在SAO网络中的地位及作用,

挖掘技术要素信息。根据构建的SAO网络,筛选有效规模排名前20的技术要素进行展示,结果如表4所示。

根据高速芯片技术要素创新性测算可知,高速芯片领域技术要素有效规模均值为6.518,限制度均值为0.798,效率均值为0.735,等级度均值为0.323。技术要素有效规模越大,限制度越低,则技术要素发挥的技术信息传递作用越重要。高速芯片领域中,“单片微机”“高速数据”“fpga芯片”“晶圆”“主控芯片”“显示屏”的有效规模均高于150,显著高于平均有效规模,区别于其他技术要素,说明上述技术要素能够获取网络非冗余信息,存在结构洞的可能性比较大。此外,上述技术要素的限制度低于0.13,显著低于平均限制度0.798,说明上述技术要素与其他技术要素相互联系,技术要素应用范围广泛,产生更多的结构洞,具备信息优势和控制能力,创新性较强。

### 4.3.3 高速芯片技术要素基础性测度

基础性技术要素是技术系统发展的基础,采用覆盖度衡量特定技术要素的基础性:覆盖度越高,技术要素基础性越强。综合考虑SAO结构中各部分之间的关联关系,利用SAO结构测度不同技术要素的基础性。鉴于技术要素众多,仅展示排名前20的技术要素,结果如图5所示。技术要素“高速数据”“单片微机”“fpga芯片”“晶圆”“主控芯片”“半导体芯片”存在于多个



表3 高速芯片技术要素关键性指标 (部分)

技术要素	度数中心性		中介中心性		接近中心性	
	排 名	数 值	排 名	数 值	排 名	数 值
高速数据	1	0.305 9	1	0.215 9	1	0.548 3
单片微机	2	0.237 6	3	0.150 3	2	0.538 6
fpga芯片	3	0.220 8	4	0.106 8	4	0.513 6
晶圆	4	0.205 0	5	0.103 8	5	0.511 3
主控芯片	5	0.192 1	6	0.092 8	7	0.507 4
显示屏	6	0.171 3	7	0.079 7	6	0.510 8
上位机	7	0.122 8	8	0.047 2	8	0.486 9
半导体晶片	8	0.121 8	11	0.034 4	37	0.438 3
串行总线接口	9	0.117 8	10	0.042 0	9	0.475 1
通信转化单元	10	0.111 9	9	0.042 1	18	0.455 3
dsp芯片	11	0.105 9	12	0.033 4	20	0.452 2
线路板	12	0.086 1	13	0.027 4	13	0.460 4
内存芯片	13	0.078 2	18	0.014 6	78	0.416 5
存储模块	14	0.077 2	16	0.018 0	33	0.441 0
封装	15	0.074 3	17	0.016 8	48	0.428 3
半导体原件	16	0.070 3	22	0.011 1	115	0.406 4
无线通信模块	17	0.070 3	14	0.018 9	29	0.442 4
集成电路	18	0.068 3	20	0.013 5	92	0.412 5
数字信号	19	0.060 4	19	0.014 5	54	0.425 4
模拟转化器	20	0.060 4	23	0.010 7	21	0.450 9

表4 高速芯片技术要素结构洞指标 (部分)

技术要素	有效规模	限制度	效 率	等级度
高速数据	299.071	0.072	0.968	0.643
单片微机	214.444	0.111	0.897	0.722
fpga芯片	198.277	0.124	0.893	0.781
晶圆	186.874	0.104	0.907	0.734
主控芯片	180.444	0.114	0.935	0.672
显示屏	162.902	0.101	0.947	0.626
半导体芯片	113.929	0.157	0.934	0.608
上位机	113.612	0.150	0.924	0.623
usb接口	106.948	0.128	0.906	0.589
通信模块	103.161	0.135	0.921	0.618
dsp 芯片	96.301	0.138	0.908	0.572
中央处理单元	76.288	0.173	0.887	0.659
存储芯片	71.897	0.189	0.922	0.652
光源	70.937	0.169	0.959	0.598
存储模块	70.519	0.135	0.916	0.486
集成电路	65.247	0.175	0.960	0.570
快闪存储	65.151	0.136	0.931	0.467
半导体元件	63.502	0.192	0.907	0.527
主机	60.857	0.162	0.922	0.533
能源模块	59.427	0.192	0.900	0.680
均值	6.518	0.798	0.735	0.323

SAO结构中,对不同技术要素产生影响,波及多个技术子系统,制约高速芯片整体技术发展,属于高速芯片领域基础性技术要素。其中,“高速数据”基础性显著强于其他技术要素,对芯片高速数据处理的要求使得该技术要素影响力较强。

4.3.4 高新芯片关键技术要素识别结果及分析

(1) 高速芯片关键技术要素识别结果。通过对高速芯片技术要素的关键性、创新性以及基础性的测度可知,“高速数据”“单片微机”“fpga芯片”“晶圆”4

个技术要素显著区别于其他技术要素: 地位突出, 关键性显著; 占据网络结构洞, 能够获取网络非冗余信息, 创新性较强; 影响其他技术要素, 基础性显著。为避免研究内容泛化, 将“高速数据”“单片微机”“fpga芯片”“晶圆”界定为高速芯片关键技术要素。

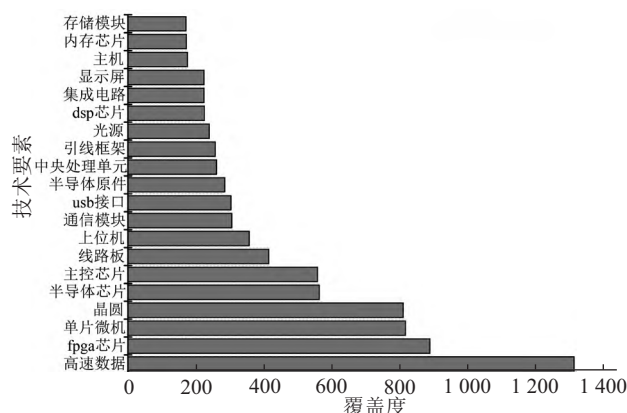


图5 技术要素基础性测度结果 (部分)

(2) 高速芯片关键技术要素专利申请分析。采集的高速芯片专利文献中, 涉及高速芯片关键技术要素“高速数据”“单片微机”“fpga芯片”“晶圆”的专利文献分别有5 618、579、469、577篇, 专利申请国家涉及中国、日本、美国、德国等, 其中大部分专利为我国申请, 但我国在高速芯片领域仍然存在“卡脖子”问题, 原因在于: 部分创新主体片面追求专利数量, 创新程度不足, 且部分专利处于未开放状态, 被束之高阁, 不能转化成生产力。我国从2016年开始集中申请“高速数据”“单片微机”“fpga芯片”“晶圆”专利, 专利申请年份比较近, 关键技术要素发展不成熟。相关专利文献被引频次较低, 说明高速芯片领域关键技术要素核心专利被其他国家掌握, 技术突破难度较高。

(3) 高速芯片关键技术要素关联关系分析。根据高速芯片关键技术要素识别结果以及涵盖关键技术要素的SAO结构, 统计高速芯片领域不同关键技术要素关联关系的类型。“高速数据”“单片微机”“晶圆”3个技术要素的关联关系中, 相互作用类占比分别为45.57%、49.63%、56.52%, 显著高于其余3种类型的关联关系。高速数据技术与缓存器、闪存芯片等相互作用, 旨在提升芯片数据传输和运算速度; 单片微机与不同设备和器件相互作用, 技术应用范围不断拓展; 晶圆制造是半导体产业中的关键环节, 需要高精度的设备和流程以保证芯片的质量和性能, 因此, 在SAO结构中晶圆与不同的工艺及流程相互连接。此外, 关键技

术要素“fpga芯片”关联关系中包含类占比较高, 达到51.12%。fpga芯片复杂程度比较高, 包含数字模块、内存模块等, 旨在提升芯片运行速度, 改善芯片性能, 是未来芯片技术发展的重要方向。

## 5 结论

本研究基于专利文献, 融合SAO结构以及社会网络分析方法, 构建关键技术要素识别模型, 尝试打开技术系统“黑箱”, 挖掘高质量技术信息, 为后发国家或企业深入理解复杂技术簇提供参考, 并以高速芯片关键技术要素识别为例, 综合考虑技术要素的关键性、创新性以及基础性, 得到“高速数据”“单片微机”“fpga芯片”“晶圆”4个高速芯片领域关键技术要素, 为高速芯片领域技术突破提供参考, 也为该领域不同创新主体的创新资源投入指明方向。同时, 高速芯片关键技术要素的关联关系中, 相互作用类、包含类占比较高, 功效类次之, 占比最低的关联关系类型为用途类, 这充分说明关键技术要素是整个技术系统的中枢, 连接不同类型的技术要素。本研究运用SAO结构以及社会网络分析, 结合关键性、创新性以及基础性特征识别关键技术要素的方法有效, 能够为其他领域的关键技术要素识别提供参考, 以获取高质量技术情报, 掌握技术系统的“卡点”, 实现技术突破。

本研究也存在不足之处: 部分技术要素尚未申请专利, 进而制约关键技术要素的识别; 数据仅来源于智慧芽数据库, 缺乏不同类型的数据。未来研究将结合论文及相关报告, 综合运用多源异构数据, 更全面地识别技术体系内部的关键技术要素, 深层次打开技术“黑箱”。

## 参考文献

- [1] 李红建. 创新: 瞄准“卡脖子”技术[N]. 学习时报, 2020-03-05 (1).
- [2] 肖广岭. 以颠覆性技术和“卡脖子”技术驱动创新发展[J]. 学术前沿, 2019 (13): 55-61.
- [3] 李晓曼, 张学福, 宋红燕, 等. 专利文献技术要素识别方法研究: 以纳米肥料领域为例[J]. 图书情报工作, 2020, 64 (6): 59-68.
- [4] 李晓曼. 基于专利要素特征的技术演化分析[D]. 北京: 中国农业科学院, 2020.

- [5] 朱永国, 陶斌斌, 宋利康, 等. 基于粗糙集和信息熵的技术成熟度关键技术要素识别方法[J]. 现代制造工程, 2018 (1): 1-5.
- [6] 张新国. 国防装备系统工程中的成熟度理论与应用[M]. 北京: 国防工业出版社, 2013.
- [7] 卜广志. 武器装备体系的技术成熟度评估方法[J]. 系统工程理论与实践, 2011, 31 (10): 1994-2000.
- [8] 孙凌宇, 朱丽莉, 张小俊, 等. 基于模糊层次分析法的关键技术元素识别[J]. 科学技术与工程, 2020, 20 (24): 9816-9821.
- [9] KIAMEHR M, HOBDAI M, HAMED M. Latecomer firm strategies in complex product systems (CoPS): the case of Iran's thermal electricity generation systems[J]. Research Policy, 2015, 44 (6): 1240-1251.
- [10] LEE J J, YOON H. A comparative study of technological learning and organizational capability development in complex products systems: distinctive paths of three latecomers in military aircraft industry[J]. Research Policy, 2015, 44 (7): 1296-1313.
- [11] 武建龙, 王宏起. 战略性新兴产业突破性技术创新路径研究: 基于模块化视角[J]. 科学学研究, 2014, 32 (4): 508-518.
- [12] 洪勇, 苏敬勤. 发展中国家核心产业链与核心技术链的协同发展研究[J]. 中国工业经济, 2007 (6): 38-45.
- [13] 洪勇, 苏敬勤. 我国复杂产品系统自主创新研究[J]. 公共管理学报, 2008, 5 (1): 76-83, 124.
- [14] CHOI S, PARK H, KANG D, et al. An SAO-based text mining approach to building a technology tree for technology planning[J]. Expert Systems with Applications, 2012, 39 (13): 11443-11455.
- [15] 冯立杰, 马亚坤, 王金凤, 等. 多维技术创新地图融合TO-RFM模型的技术机会识别与评价路径研究[J]. 情报理论与实践, 2023, 46 (2): 145-155, 108.
- [16] LUO W C, YANG F. An empirical study of automatic Chinese word segmentation for spoken language understanding and named entity recognition[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016: 238-248.
- [17] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition[EB/OL]. [2023-05-06]. <https://arxiv.org/abs/1603.01360>.
- [18] 臧凌玉, 张应中, 罗晓芳. 基于双重深度迁移学习的机械领域命名实体识别[J]. 计算机应用与软件, 2022, 39 (9): 219-224.
- [19] 杨武, 陈培, DAVID G. 专利引证视角下技术轨道演化与技术锁定识别: 以光刻技术为例[J]. 科学学研究, 2022, 40 (2): 209-219.
- [20] 栾春娟. 战略性新兴产业共性技术测度指标实证研究[J]. 中国科技论坛, 2012 (6): 73-77.
- [21] 栾春娟. 战略性新兴产业共性技术测度指标研究[J]. 科学学与科学技术管理, 2012, 33 (2): 11-16.
- [22] 李瑞茜, 陈向东. 基于专利共类的关键技术识别及技术发展模式研究[J]. 情报学报, 2018, 37 (5): 495-502.
- [23] LEE Y G, SONG Y I. Selecting the key research areas in nano-technology field using technology cluster analysis: a case study based on National R&D Programs in South Korea[J]. Technovation, 2007, 27 (1/2): 57-64.
- [24] YASUNAGA Y, WATANABE M, KORENAGA M. Application of technology roadmaps to governmental innovation policy for promoting technology convergence[J]. Technological Forecasting and Social Change, 2009, 76 (1): 61-79.
- [25] 马永红, 孔令凯, 林超然, 等. 基于专利挖掘的关键共性技术识别研究[J]. 情报学报, 2020, 39 (10): 1093-1103.
- [26] 毛荐其, 杜艳婷, 苗成林, 等. 基于专利共类的关键核心技术识别模型构建及应用: 以光刻技术为例[J]. 情报杂志, 2022, 41 (11): 48-54.
- [27] MIROWSKI P. An evolutionary theory of economics change: a review article[J]. Journal of Economic Issues, 1983, 17 (3): 757-768.
- [28] MARIANI M S, MEDO M, LAFOND F. Early identification of important patents: design and validation of citation network metrics[J]. Technological Forecasting and Social Change, 2019, 146: 644-654.
- [29] 胡登峰, 黄紫微, 冯楠, 等. 关键核心技术突破与国产替代路径及机制: 科大讯飞智能语音技术纵向案例研究[J]. 管理世界, 2022, 38 (5): 188-209.
- [30] 杨艳萍, 董瑜, 韩涛. 基于专利共被引聚类和分析的产业关键技术识别方法研究: 以作物育种技术为例[J]. 图书情报工作, 2016, 60 (19): 143-148, 124.
- [31] 陈旭, 江瑶, 熊焰, 等. 关键核心技术“卡脖子”问题的识别及应用: 以AI芯片为例[J]. 中国科技论坛, 2023 (9): 17-27.
- [32] 汤勇力, 李剑敏, 李从东, 等. 面向制造企业价值链升级的产业关键技术识别研究[J]. 科学学研究, 2016, 34 (12): 1777-1788.
- [33] 马骥, 汤小银. 产业集群网络、结构演化与协同发展: 以叶集木竹产业为例[J]. 安徽师范大学学报(人文社会科学版), 2019, 47 (4): 111-121.
- [34] 任海英, 李真. 基于输入输出型SAO网络的核心技术链识别



- 方法研究: 以量子计算领域为例[J]. 图书情报工作, 2021, 65 (19): 117-129.
- [35] 黄鲁成, 刘春文, 吴菲菲, 等. 基于NPCIA的核心技术识别模型及应用研究[J]. 科学学研究, 2020, 38 (11): 1998-2007.
- [36] GEUM Y, KIM C, LEE S, et al. Technological convergence of IT and BT: evidence from patent analysis[J]. ETRI Journal, 2012, 34 (3): 439-449.
- [37] 贾军, 魏洁云. 新兴产业核心技术早期识别方法与应用研究[J]. 科学学研究, 2018, 36 (7): 1206-1214.
- [38] CASCINI G, FANTECHI A, SPINICCI E. Natural language processing of patents and technical documentation[M]//MARI-NAI S, DENGEL A. Document Analysis Systems. Heidelberg: Springer, 2004: 508-520.
- [39] 李欣, 谢前前, 黄鲁成, 等. 基于SAO结构语义挖掘的新兴技术演化轨迹研究[J]. 科学学与科学技术管理, 2018, 39 (1): 17-31.
- [40] WU Z, PALMER M. Verb semantics and lexical selection[EB/OL]. [2023-05-06]. <https://arxiv.org/abs/cmp-lg/9406033>.
- [41] 郭俊芳, 汪雪锋, 李乾瑞, 等. 一种新型的技术形态识别方法: 基于SAO语义挖掘方法[J]. 科学学研究, 2016, 34 (1): 13-21.
- [42] 杨晓莉. 高速芯片封装技术的EMI辐射研究[D]. 杭州: 浙江大学, 2018.
- [43] 杨迎. 适用于高速芯片的老化测试系统研究[D]. 武汉: 武汉邮电科学研究院, 2020.
- [44] 谢非, 罗安. 非对称性贸易依赖关系对中国芯片进口的影响探究[J]. 改革, 2022 (6): 116-130.

## 作者简介

孔令凯, 男, 博士, 讲师, 研究方向: 技术创新与关键共性技术, E-mail: 861586018@qq.com。

杨朝均, 男, 博士, 教授, 研究方向: 低碳创新网络。

齐浩, 男, 硕士研究生, 研究方向: 高铁网络。

马永红, 女, 博士, 教授, 研究方向: 科技创新与管理创新。

杨晓萌, 女, 博士, 讲师, 研究方向: 关键共性技术与技术合作研发。

## Identification of Key Technology Elements Based on SAO Structure: Taking High-Speed Chip as an Example

KONG LingKai<sup>1</sup> YANG ChaoJun<sup>1</sup> QI Hao<sup>1</sup> MA YongHong<sup>2</sup> YANG XiaoMeng<sup>3</sup>

(1. Faculty of Management and Economics, Kunming University of Science and Technology, Kunming 650504, P. R. China; 2. School of Economics and Management, Harbin Engineering University, Harbin 150001, P. R. China; 3. School of Economics and Management, Jiangsu University of Science and Technology, Zhenjiang 212003, P. R. China)

**Abstract:** Facing the growing risk of technological embargoes and containment, digging deeper into the key core technology information and mastering the key technology elements within the technology system, such as key components, processes, and methods, is the premise of breaking through the technology bottleneck. The three characteristics of key technology elements are summarized as critical, innovative, and fundamental. This research takes the high-speed chip as an example, adopts patent literature data, and analyzes terminologies for high-speed chips. We introduce SAO (Subject-Action-Object) semantic structure, extract high-speed chip technology elements, and take technology elements as nodes and technology element association strength as edges to construct high-speed chip SAO semantic network. Afterwards, we draw on social network analysis methods, quantify technology element rights, status, and innovation capability, identify key technology elements inside high-speed chip system, explore high-speed chip technology structure, and find technology chokepoints. The research results show that there are four key technology elements of stranglehold on technology: "high-speed data", "single-chip microcomputer", "fpga chip", and "wafer".

**Keywords:** Key Technology Element; High-Speed Chip; SAO Structure; Centrality; Structural Hole; Degree of Coverage

(责任编辑: 王玮)