

DNA 计算研究主题的演化路径分析

张子婷, 郑彦宁, 袁 芳

(中国科学技术信息研究所, 北京 100038)

摘要: 本文基于 WoS 核心合集检索得到 DNA 计算领域的 2 373 篇论文, 以关键词为分析单元, 通过划分研究时间区间, 使用共词分析法和聚类分析法, 利用 SciMAT 软件绘制关键词覆盖图、主题战略图、主题路径演化图, 识别不同研究时期的研究主题及其演化路径。DNA 计算领域的研究始于 1994 年, 早期发展较为迟缓, 后期关键词增多, 研究热度高。各时期研究主题变化不大, 演化关系较为简单, 在 2 个研究方向上呈现 4 条演化路径。分子计算和逻辑门的构建是主要研究主题, 将持续成为研究热点。

关键词: 演化路径; 主题战略图; DNA 计算; SciMAT

中图分类号: G354.4 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2019.11-12.013

1994 年, 美国加利福尼亚大学的 Adleman 博士^[1]首次通过生化方法求解了 7 个顶点的哈密顿回路问题, 显示了用 DNA 进行特定目的计算的可行性。DNA 计算机模型克服了电子计算机存储量小和运算速度慢这两个严重的不足, 具有并行性高、运算速度快、贮存容量大、耗能低、DNA 分子资源丰富等优点。此外, 基于生化机理的 DNA 计算在解决图与组合优化这类电子计算机几乎无法解决的问题上有一种“天然”的优势^[2]。关于 DNA 计算的研究目前已经取得了不少成果, 越来越多的学者和专家投入到该领域的研究中来。未来, DNA 计算机将应用于逻辑研究、密码破译、基因编程、疾病防治及航空航天等领域。通过对 DNA 计算领域的研究主题演化路径进行梳理, 可以厘清各主题的发展脉络, 便于相关研究学者全面了解领域发展现状并科学预测未来研究方向。

最早人们使用 TRIZ 法^[3]、德尔菲法^[4]等定性方法进行知识演化状况的研究, 此类方法主要依赖专家知识与经验判断, 具有较强的主观性。但随后科技快速发展, 且信息传播快捷, 研究人员难以全局、系统地把握学科领域的宏观发展状况, 很难对

某个学科领域的知识结构进行全面客观的描述^[5]。随着科学计量学的发展, 学者们利用词频分析法、共词分析法、引文分析法等方法描述研究主题的演化路径, 实现了定性与定量的结合。共词聚类分析是利用词与词之间的关联程度对关键词进行聚类, 得到多个主题, 展示某一学科或领域的研究结构。目前已有多种成熟的分析工具被用于共词聚类分析中, 例如 Ucinet、Pajek、CiteSpace 等^[6]。张敏等^[7]采用 Ucinet 的社会化网络分析功能得到国际大数据领域的研究热点分布, 结合 CiteSpace 提取高被引文献分析其演化路径。陈建龙等^[8]利用 Pajek 等分析工具对现代大学制度领域进行聚类分析、分类分析及历时分析, 并总结了该领域的研究热点与发展态势。任晓松等^[9]利用 CiteSpace 软件绘制中国碳排放研究的知识基础和热点演化知识图谱。但利用 CiteSpace 软件得到的演化图只能大致看出不同研究时期领域研究热点的演化, 关于某一领域研究热点在各个时期如何演化无法得知。同时也出现了新的分析与可视化工具, 如 ThemeRiver^[10], TextFlow^[11], SciMAT^[12]等。其中 SciMAT 软件在展示学科演化方面有特有的优势: 具备数据预处理、

第一作者简介: 张子婷 (1996—), 女, 在读硕士研究生, 主要研究方向为产业竞争情报。

项目来源: 2017 年博士后基金第 61 批“颠覆性技术的识别方法研究” (2017M610970)。

收稿日期: 2019-10-07

分析方法选取、计量指标选取等功能, 可根据分析需求进行选择与组合; 利用关键词覆盖图、战略坐标图、主题演化图, 由点及面进行全方位分析, 准确掌握各时间段主题的发展状况与全时期的演变情况。鉴于此, 本文利用 SciMAT 软件, 以 DNA 计算领域论文产出为对象, 以关键词为分析单元, 描述不同时期 DNA 计算领域的研究主题与演化路径, 揭示该领域的主题演化关系、发展状况及发展脉络, 以期制定科学研究政策和开展研究活动提供参 考依据。

1 数据来源及分析方法

1.1 数据来源

本文以 Web of Science 核心合集数据库 (以下简称 WoS 数据库) 中收录的关于 DNA 计算机的论文为数据来源进行分析。由于 2019 年数据不完整, 故将检索年限设置为 1994 年至 2018 年, 通过检索得到 2 373 篇论文。

1.2 分析方法

本文采用 SciMAT 软件对数据进行清洗、处理与分析。以目标文献集的关键词集合——WoS 数据库中的关键词字段 (Keywords & Keyword Plus) ——为分析单元, 首先利用关键词覆盖图直观表示关键词变化; 接着通过对目标文献关键词集构建共现矩阵并进行聚类分析, 绘制研究领域的聚类主题战略图, 以把握各研究主题在该研究区间内的发展状况; 最后根据各主题之间的关联关系, 制作主题演化路径图^[13], 以识别研究主题演化路径。

1.2.1 关键词覆盖图

关键词覆盖图从数量上直观表示关键词新生、继承与消亡的情况。如图 1 所示, 每个圆圈代表设定的时期, 圈内数字代表该研究区间出现的关键词数量; 向上的箭头代表消亡关键词的数量, 向下的箭头代表新生关键词的数量, 水平箭头表示从上个时期延续到下个时期的关键词数量及稳定性指数 (研究区间 A 延续到研究区间 B 的关键词数量占研究区间 A 全部关键词数量的比例)^[13]。如示意图中, 研究区间一的关键词数量为 60, 研究区间二的关键词数量为 90。研究区间一向研究区间二的演化过程中, 有 20 个消亡关键词, 有 50 个新生

关键词, 有 40 个关键词保留至研究区间二, 稳定指数为 0.67 (40/60)。

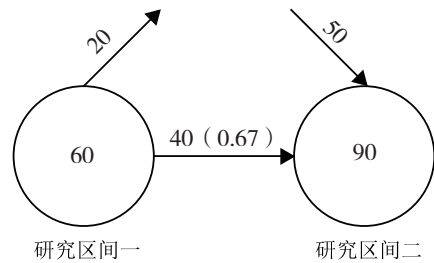


图 1 关键词覆盖图

1.2.2 主题战略图

主题战略坐标图在共词和聚类的基础上, 描述特定研究领域内研究主题的自身发展情况和主题间相互影响情况^[14]。通过战略坐标图, 可以把握各主题在各研究区间的发展状况, 同时结合上下时期聚类所处位置的变化判断其演变方向。如图 2 所示, 主题坐标战略图从中心度 (Centrality) 和密度 (Density) 两个维度表示研究主题在该时期的发展状况。中心度表示聚类与其他主题的互相影响程度, 代表聚类的重要程度; 密度表示聚类内部关键词联系强度, 代表聚类在研究领域中的发展程度^[15]。节点表示聚类, 其大小表示分析时选择的聚类的影响指标, 本文中的战略图聚类大小表示 h 指数。不同象限代表不同的发展状况: 右上象限所包含的聚类为引擎类聚类, 中心性高且密度大, 代表发展较好且重要的主题; 右下象限所包含的聚类为基本的横向的聚类, 与其他主题联系较强但本身发展不足或已经到达趋于平稳的水平; 左下象限所包含的聚

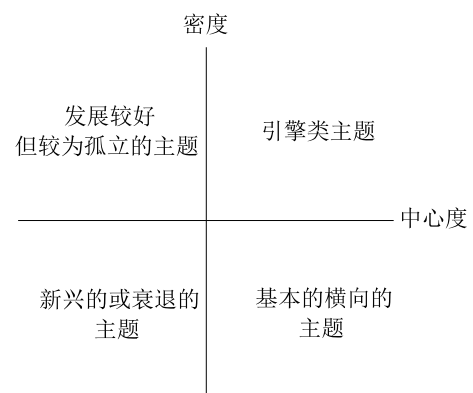


图 2 战略图^[3]

类为新兴的或衰退的聚类，中心性低且密度小；左上象限所包含的聚类为发展较好但较孤立的聚类，主题本身发展较好但与其他主题联系不紧密。

1.2.3 主题演化路径图

主题演化路径图在主题战略图的基础上，采用相邻区间聚类间的主题相似度进行关联，通过主题的出现、分化、合并等关系可以大致确定研究主题演化路径，同时依据各时期各主题在主题战略图中的位置变化确定主题演化方向。如图3所示，每个圆圈代表该列所在的研究区间内的关键词聚类，与各时期主题战略图一一对应。连线代表研究的连续性，线条粗细及颜色深浅与主题关联度成正比，主题关联度越大，线条越粗且颜色越深，相邻聚类越相似，研究连续性越强。其中实线表示两个聚类之间共享的是核心关键词，代表主流的演化方向，虚线表示两个聚类之间共享的是次要关键词，代表支流的演化方向。孤立点表示在其所在研究时期单独出现的研究主题，与前后时期关联性不强^[6]。

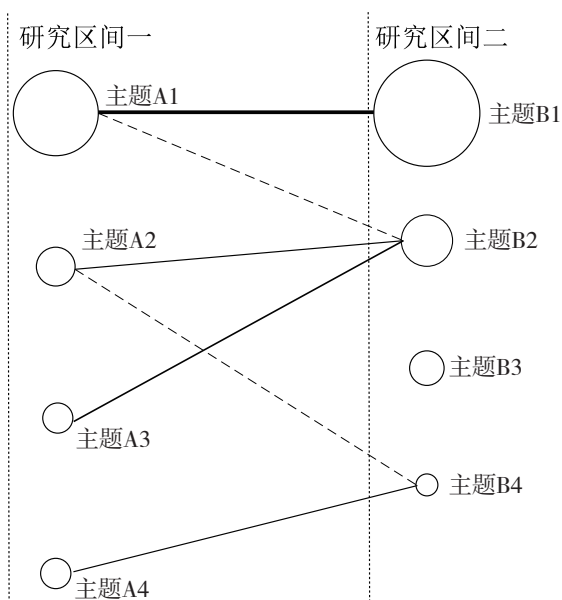


图3 主题演化图

1.3 参数设置

首先将25年数据以5年为周期分为5个研究区间：1994—1998年、1999—2003年、2004—2008年、2009—2013年、2014—2018年。5个研究区间的发文量依次为103篇、311篇、635篇、611篇、713篇。

在分析参数选择上，以关键词为分析单元，考虑到各研究区间的论文数量，5个时间段的数据精简阈值分别为1、2、3、3、3。构建关键词共现网络，分别设置网络精简阈值1、1、2、2、2，网络标准化的相似性度量指标为等价指数（Equivalence Index）。聚类算法为简单中心算法（Simple Centers Algorithm），经过多次尝试后设置网络最大值为20，最小值为3，聚类的影响指标为文献总数和h指数。演化图的相似度指标选择Jaccard系数，关键词覆盖图的相似度指标选择包容指数（Inclusion Index）。

2 结果与分析

2.1 发文量统计分析

论文的逐年产出情况如图4所示。1994年，Adleman首次提出利用DNA分子算法求解特定的NP完全问题，从而开创了DNA计算研究的新纪元^[1]。在此之后，DNA计算机领域文献发表量呈现缓慢上升趋势。2002年之后发展迅猛，保持在一定的发展水平，于2017年达到最高发文量，有171篇文献。通过各研究区间的发文总量可以看出，从第3个研究区间开始，DNA计算领域的文献保持在比较稳定的水平。

2.2 关键词数量变化分析

图5为1994—2018年25年间5个研究区间关键词新生与消亡的情况。由于1994—1998年处于研究的起步阶段，关键词数量较少。每个时期关键词增长量都保持在一定水平（92、84、90、56），各时期新生关键词大于都消亡关键词，说明学科正处于蓬勃发展阶段。从上一阶段延续至下一时期的关键词数量远大于消亡的关键词数量，说明学科继承性较强。后3个研究时期的发文量差别不大（635、611、713），但关键词数量均增加较多，说明这期间研究人员从不同方向对DNA计算领域开展探索性研究。稳定指数持续上升，说明研究人员对DNA领域的多数研究方向进行了深入性研究。

2.3 主题发展分析

（1）1994—1998年期间生成7个聚类，见图6，其各项指标见表1：算法（Algorithm）、基因（Gene）、序列（Sequences）、DNA计算机（DNA-Computer）、

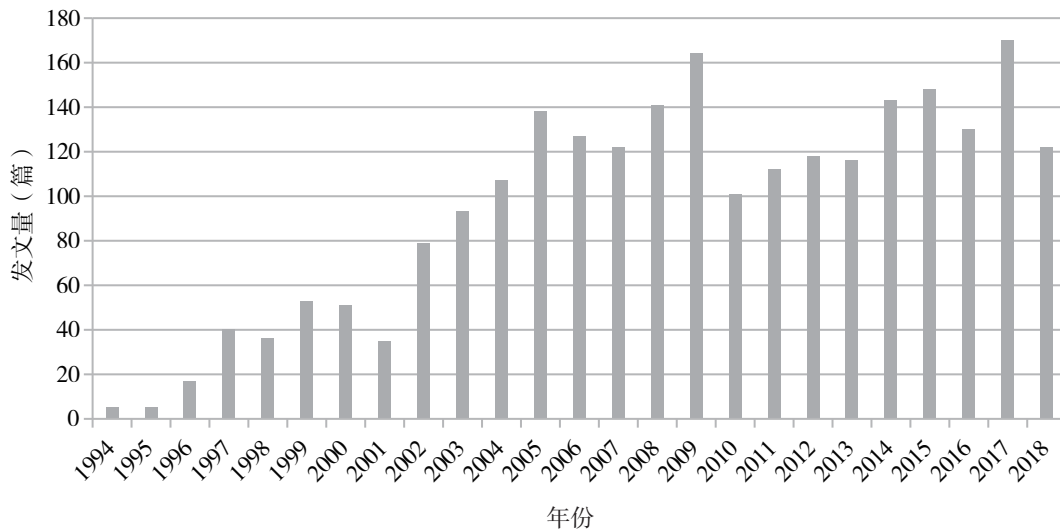


图 4 DNA 计算领域 1994 年至 2018 年历年发文量统计

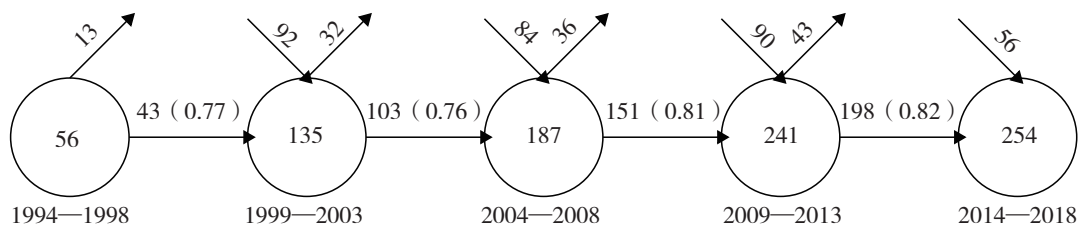


图 5 DNA 计算领域各研究区间关键词覆盖图

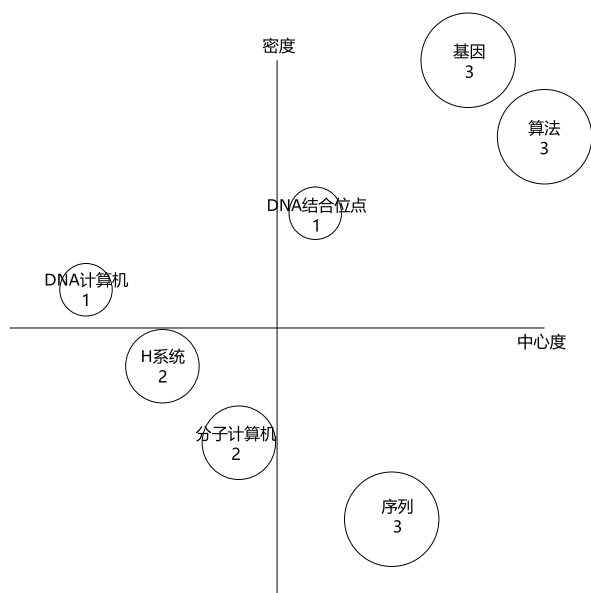


图 6 DNA 计算领域 1994—1998 年主题战略图

H 系统 (H-Systems)、分子计算机 (Molecular-Computer)、DNA 结合位点 (DNA-Binding-Site)。

此阶段正处于起步期, 各聚类核心文献数量及关键词均较少。位于右上象限的聚类有算法、基因和 DNA 结合位点, 受关注程度较高, 是发展较好且重要的主题。位于右下象限的聚类为序列, 是基本的主题。位于左下象限的聚类有 H 系统、分子计算机, 是新兴研究主题。位于左上象限的聚类为 DNA 计算机, 是发展较好但较为孤立的主题。

(2) 1999—2003 年期间生成 5 个聚类, 见图 7, 其各项指标见表 2: 分子计算 (Molecular-Computation)、算法 (Algorithm)、系统 (Systems)、核酸 (Nucleic-Acids)、NP 完全问题 (NP-Complete-problem)。在此期间, 文献数量急剧增加, 但研究主题更为集中。位于右上象限的聚类有分子计算、算法、系统, 受关注程度较高, 是发展较好且重要的主题。分子计算聚类继承了上一时期 DNA 计算机的部分关键词, 同时出现了许多新的关键词, 代表该主题出现较多新的研究方向, 例如为了克服试管型的 DNA 计算的缺点提出了表面与

表 1 DNA 计算领域 1994—1998 年聚类各项指标

聚类名称	中心度	中心度 标准化	密度	密度标准化	核心文献 数量 (篇)	核心文献 h 指数	关键词数量 (个)
算法	10.9	1	100.69	0.86	3	3	9
基因	6.65	0.86	116.25	1	3	3	8
序列	3.75	0.71	61.61	0.14	3	3	7
DNA 计算机	0	0.14	100	0.57	2	1	5
H 系统	0	0.29	95	0.43	2	2	5
分子计算机	0	0.43	62.5	0.29	2	2	4
DNA 结合位点	0	0.57	100	0.71	1	1	3

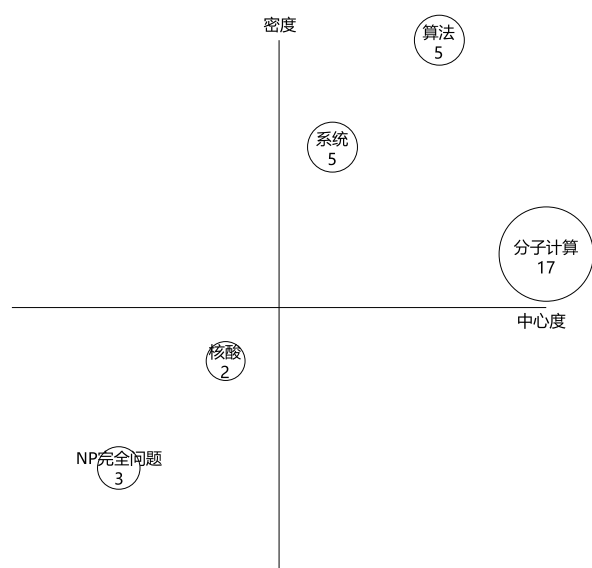


图 7 DNA 计算领域 1999—2003 年主题战略图

芯片 DNA 计算模型。在分子计算聚类中, 有两篇关于逻辑门的文章^[17, 18]被引量分别为 336 次

和 209 次, 对后期研究影响很大。但由于关键词标引的原因, 两篇文章皆没有作者关键词, 只有 WoS 数据库根据施引文献标题自动生成的扩展关键词, 对结果造成了一定影响。算法聚类继承了上一时期算法聚类的主要关键词和基因聚类的次要关键词, 密度有所提升, 说明研究人员对该领域的研究深入。系统聚类继承了上一时期 DNA 计算机聚类分子计算机聚类和 H 系统的部分关键词。位于左下象限的聚类有核酸、NP 完全问题, 是新兴研究主题。DNA 计算的兴起为解决 NP 完全问题提供了可能。

(3) 2004—2008 年期间生成 6 个聚类, 见图 8, 其各项指标见表 3: 分子计算 (Molecular-Computation)、算法 (Algorithm)、逻辑门 (Logic-Gates)、纳米结构 (Nanostructures)、剪接系统 (Splicing-Systems)、二级结构 (Secondary-Structure)。位于右上象限的聚类有纳米结构, 是发展较好且重要的主题。DNA 纳米技术因其可编

表 2 DNA 计算领域 1999—2003 年聚类各项指标

聚类名称	中心度	中心度 标准化	密度	密度标准化	核心文献 数量 (篇)	核心文献 h 指数	关键词数量 (个)
分子计算	40.14	1	9.92	0.6	36	17	20
算法	22.85	0.8	21.28	1	9	5	10
系统	10.49	0.6	11.86	0.8	10	5	8
核酸	9.56	0.4	8.33	0.4	3	2	4
NP 完全问题	1.68	0.2	3.61	0.2	3	3	3

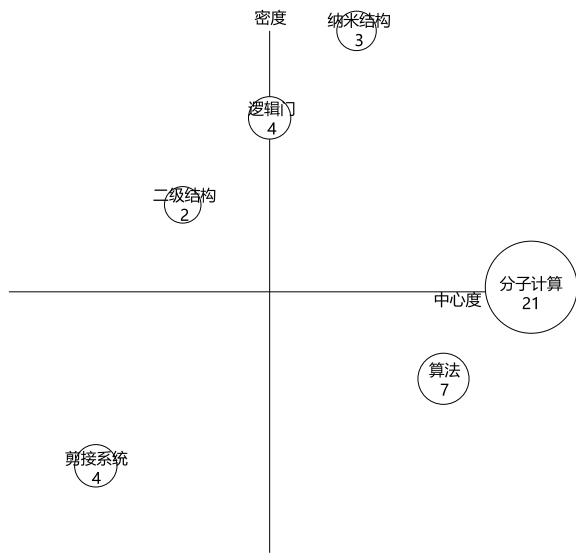


图 8 DNA 计算领域 2004—2008 年主题战略图

程的优势对 DNA 计算机产生了巨大的影响。分子计算聚类位于中心度轴上,较上一时期稍向下移。分子计算聚类继承了上一时期分子算法聚类的主要关键词,同时也增加了一些关键词,结合相关文献发现 DNA 计算在本研究期间主要有建立在链置换基础上的、建立在瓦片基础上的以及建立在纳米颗粒基础上的 DNA 计算模型。位于右下象限的聚类是算法,是本时期内基本的研究主题。位于密度轴上的逻辑门首次作为一个聚类出现,说明这段时间研究人员对逻辑门的研究力度加大且取得了一定的研究成果。位于左下象限的聚类为剪接系统,是上一时期中系统聚类的关键词,在本时期作为一个聚类出现,故认为其是新兴研究主题。位于左上象限的聚类为二级结构,是发展较好但较为孤立的主题。

表 3 DNA 计算领域 2004—2008 年聚类各项指标

聚类名称	中心度	中心度标准化	密度	密度标准化	核心文献数量(篇)	核心文献 h 指数	关键词数量(个)
分子计算	15.15	1	6.78	0.5	58	21	20
算法	6.8	0.83	5.31	0.33	12	7	6
逻辑门	3.41	0.5	11.19	0.83	4	4	4
纳米结构	3.91	0.67	15.11	1	3	3	4
剪接系统	1	0.17	5.28	0.17	8	4	4
二级结构	2.5	0.33	7.41	0.67	3	2	3

(4) 2009—2013 年期间生成 6 个聚类,见图 9,其各项指标见表 4: 逻辑门 (Logic-Gates)、分子计算 (Molecular-Computation)、算法 (Algorithm)、纳米结构 (Nanostructures)、DNA 词汇 (DNA-Words)、程序 (Program)。位于右上象限的聚类是逻辑门,具有高中心度和高密度,核心文献数量与 h 指数都显著增加,是发展较好且重要的主题。分子计算聚类仍旧位于向心度轴上,继承了上一时期分子计算聚类和剪接系统聚类的主要关键词。该聚类在战略图中所处的位置虽与上时期相差无几,但中心度和密度均有所提高,核心文献数量、h 指数及关键词数量均明显增加,是领域中自身发展良好且与其他主题联系十分密切的主题。位于右下象限的聚类是纳米结构,是本研究时期内基

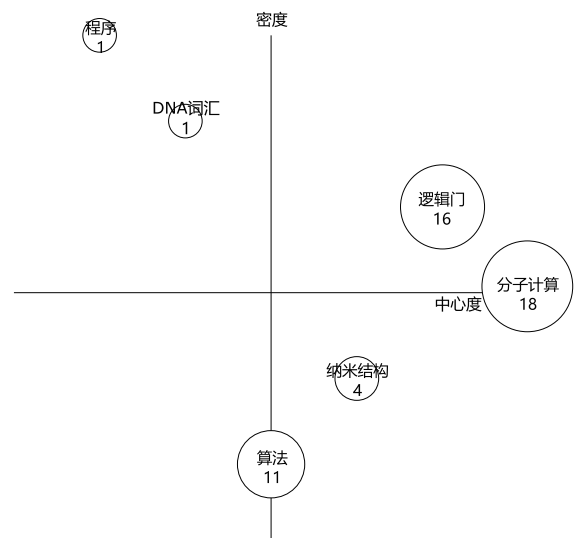


图 9 DNA 计算领域 2009—2013 年主题战略图

本的研究主题。算法聚类位于密度轴上,密度较低,中心度一般。算法聚类从第一个时期开始一直持续到此阶段,从右上象限往左下转移,且下一阶

段不再以单独聚类出现,而是作为分子计算聚类的关键词出现。位于左上象限的聚类有DNA词汇、程序,是发展较好但较为孤立的主题。

表4 DNA计算领域2009—2013年聚类各项指标

聚类名称	中心度	中心度标准化	密度	密度标准化	核心文献数量	核心文献h指数	关键词数量
分子计算	26.45	1	10.29	0.5	51	18	20
逻辑门	24.45	0.83	10.4	0.67	28	16	15
算法	6.39	0.5	2.64	0.17	27	11	12
程序	1.15	0.17	17.78	1	2	1	3
DNA词汇	2	0.33	16.79	0.83	2	1	3
纳米结构	6.56	0.67	7.08	0.33	5	4	3

(5) 2014—2018年期间生成6个聚类,见图10,其各项指标见表5:逻辑门(Logic-Gates)、分子计算(Molecular-Computation)、纳米颗粒(Nanoparticles)、DNA序列(DNA-Sequences)、DNA计算(DNA-Computation)、可能性(Probability)。位于中心度轴上的逻辑门聚类,其h指数首次超过分子计算聚类,是领域内的重要研究领域。位于右下象限的聚类有分子计算、纳米颗粒,是本研究时期内基本的研究主题。分子计算聚类吸收了上一时期分子计算聚类和算法聚类的主要关键词,并出现了许多新的关键词,说明关于分子计算的研究呈现多样化。纳米颗粒聚类吸收了逻辑门聚类的主要关键词和分子计算聚类的次要关键词。DNA计算聚类位于密度轴上,主要关注图与优化问题。位于左上象限的聚类有DNA序列、可能性,是发展较好但较为孤立的主题。

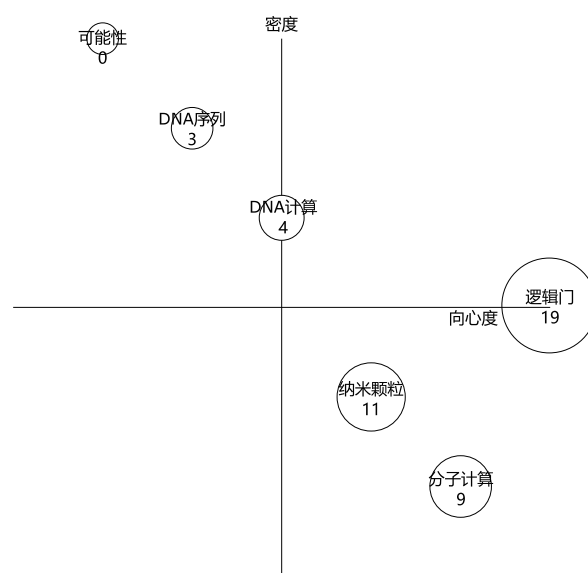


图10 DNA计算领域2014—2018年主题战略图

表5 DNA计算领域2014—2018年聚类各项指标

聚类名称	中心度	中心度标准化	密度	密度标准化	核心文献数量(篇)	核心文献h指数	关键词数量(个)
逻辑门	25.76	1	7.12	0.5	69	19	20
分子计算	13.67	0.83	5.24	0.17	68	9	20
纳米颗粒	9.62	0.67	6.12	0.33	29	11	11
DNA序列	1.57	0.33	13.54	0.83	6	3	4
DNA计算	1.74	0.5	9.82	0.67	16	4	4
可能性	0	0.17	30.16	1	3	0	3

3 DNA计算领域研究主题演化路径分析

通过历年发文量统计分析、各时期关键词覆盖图分析、各时期主题战略图分析, 大致了解 DNA 计算领域的发展情况。在此基础上, 计算相邻区间研究主题的关联度, 绘制主题演化路径图, 识别 DNA 计算领域研究主题演化路径。同时, 利用主题在各时期主题战略图中的位置变化, 判断演化方向。本文利用颜色对各主题在战略图中的象限位置进行区分。

图 11 为 1994—2018 年间 DNA 计算领域研究主题的演化路径图。从主题动态演化路径图可以看出, 研究主题随着研究的深入呈现多样化的情况, 研究内容不断丰富, 节点增大, 文献数量持续增长。1994—1998 年处于 DNA 计算领域的萌芽阶段,

受到 Adleman 的影响与启发, 研究主题分布广泛。1999—2003 年的研究主题聚类数量减少, 研究主题更为集中, 出现了一个孤立节点“NP 完全问题”。2004—2008 年间的研究主题增多, 出现了新的聚类“逻辑门”, 分子计算和算法持续成为研究热点。2009—2013 年该领域研究热度升高, 上一时期新出现的研究主题成为新的主流研究方向, 原来的主流研究方向热度持续不减。2014—2018 年各研究主题继承性强, 主流演化方向稳定发展且研究更为深入。总体来说, DNA 计算领域的研究正处于发展阶段, 取得了阶段性的研究成果, 但并未成熟, 各时期的主流研究主题继承性强, 演化关系较为简单。研究发展至今, 大致在分子计算、逻辑门 2 个方向上呈现 4 条演化路径。

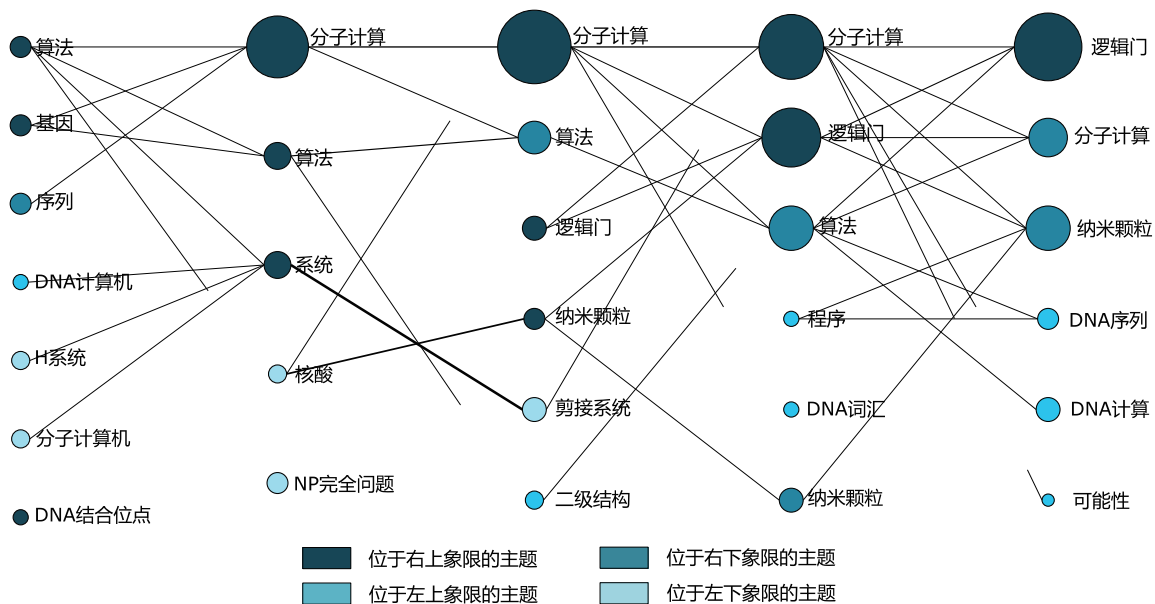


图 11 DNA 计算领域 1994—2018 年主题演化路径图

3.1 分子计算

(1) 分子计算: 算法、基因、序列→分子计算→分子计算→分子计算、序列→分子计算。这条演化路径由早先的序列主题的主流和算法主题的支流演化而来, 继承性强, 关注分子计算主题。这条演化路径上的主题一直处于右上和右下象限, 具有高向心度, 是领域中的核心研究领域。通过分析各阶段聚类关键词可以发现, 不断有新的研究内容涌现,

大致可归纳为以下几个方面: 一是关于 DNA 计算模型构建的研究。DNA 计算模型一直是 DNA 计算领域研究的热点问题, 结合各阶段关键词和相关文献可以发现剪接系统模型、粘贴系统模型、双链 DNA 计算模型、单链 DNA 或 RNA 模型、表面 DNA 计算模型、芯片 DNA 计算模型、DNA Tile 自组装模型等。二是关于编码问题的研究。在 DNA 计算中, 信息总是通过特定的 DNA 序列来表示。DNA 计算中序列的

编码问题为解决图与组合优化的某些问题提供了可能,图论问题、NP 完全问题、NP 难题等关键词反映了研究人员在此方向的努力。三是微流控制系统在 DNA 计算机中的应用研究。微流控制系统可以将 DNA 计算机中的基本操作单位集合在一张芯片上,为实现真正意义上的 DNA 计算机提供可能。

(2) DNA 计算模型的构建:算法、DNA 计算机、分子计算机、H 系统→系统→剪接系统、支持→分子计算→分子计算、DNA 计算、可能性。这条演化路径由早些算法的主流、DNA 计算机与分子计算机的主流及 H 系统的支流演化而来,是分子计算演化路径的重要分支,主要关注 DNA 计算模型和 DNA 计算机模型的构建及其应用于基因分析与疾病诊断的研究。

(3) DNA 计算序列编码问题:算法→算法、核酸→算法、二级结构→基因比对、序列、工具→基因。这条演化路径由早期的算法的支流、序列的支流演化而来,继承性很强。在前三个研究区间内,算法聚类从右上象限转移到右下象限,成为领域中的基础研究主题。在 2004—2008 年到 2009—2013 演化时,算法聚类演化为分子计算聚类、算法聚类和序列聚类,算法聚类消失,原先算法聚类中的基因比对、DNA 序列等关键词演化成为新的聚类。主要关注图与组合优化问题的解决。

3.2 逻辑门

序列→分子计算、核酸→逻辑门、纳米结构、分子计算→逻辑门、机器→逻辑门、纳米颗粒、分子计算。这条演化路径由早期的序列、分子计算、核酸演化而来,早期继承性一般,后期继承性强。在 1999—2003 年期间虽然没有出现逻辑门的聚类,但是在分子计算聚类中出现了关于逻辑门的文献,被引频次分别为 336 和 209,说明此时逻辑门已经受到人们的关注并取得了具有一定影响力的研究成果。从第三个研究区间(2004—2008 年)以来,逻辑门聚类作为独立节点出现,均处于右上象限。2014—2018 年期间出现在纵轴右部分,仍具有高中心度,与其他主题联系紧密,存在从右上象限向右下象限移动的趋势,说明逻辑门研究方向呈现多元化趋势。随着研究数量的增多和研究内容的深入,逻辑门迅速演化为研究的重点领域。布尔逻辑电路的模拟是 DNA 计算研究中非常具有实际应用意义

的研究主题。传统电子计算机的硬件电路是由不同类型的逻辑门,诸如 OR、AND、XOR、NAND 等构成。同样地,DNA 计算机体系结构的产生基础和 DNA 计算机实现技术的硬件基础也是由基于 DNA 的逻辑电路构成的^[19]。逻辑门研究的深入为 DNA 计算机的发展与实现奠定了基础。

4 结语

总体来说,DNA 计算领域的早期研究较为迟缓,研究主题少,作为一个新领域具备一定热度;随后主题关键词数量逐渐增多,新的研究内容不断涌现,研究深度不断拓展;目前领域发展未到达成熟期,各时期的研究主题变化不大,主题演化关系较为简单。演化过程稳定,聚类间继承性强。分子计算和逻辑门的构建是 DNA 计算领域的主要研究脉络,预计将持续成为研究热点。DNA 计算机的具体实现是未来需要突破的研究方向。

本研究通过划分连续的研究区间,对各时期文献的关键词进行聚类,对其进行主题战略分析和演化路径展示,在一定程度上实现了 DNA 计算领域的研究主题演化路径识别,但是仍存在粗粒度的缺点,关于主题演化的更多细节问题需要深入挖掘,这也是今后进一步研究的方向。另外,由于本文是基于关键词进行聚类及关联计算,故关键词对于结果的影响比较大。而部分论文没有作者标引的关键词,只有数据库根据其施引文献标题进行自动标引的关键词,此类关键词存在泛化冗余、专指性不强的缺点,尤其当一个主题首次出现时无法快速识别。在今后的研究中,将进一步克服关键词标引不准确的问题,提升主题演化的准确度与深度。■

参考文献:

- [1] Adleman L M. Molecular computation of solutions to combinatorial problems[J]. Science, 1994, 266(5 187): 1 021-1 024.
- [2] 许进,张雷. DNA 计算机原理、进展及难点(I): 生物计算系统及其在图论中的应用[J]. 计算机学报, 2003, 26(1): 1-11.
- [3] Ilevbare I M, Provert D, Phaal R. A review of TRIZ, and its benefits and challenges in practice[J]. Technovation, 2013, 33(2-3): 30-37.

- [4] Powell C. The Delphi technique: myths and realities[J]. *Journal of Advanced Nursing*, 2003, 41(4): 376-382.
- [5] 郝心宁, 孙巍, 张学福. 1995—2012 年生物育种领域知识演化分析 [J]. *中国农业科技导报*, 2014, 16 (2) : 174-181.
- [6] 胡阿沛, 张静, 雷孝平, 等. 基于文本挖掘的专利技术主题分析研究综述 [J]. *情报杂志*, 2013 (12) : 88-92, 61.
- [7] 张敏, 刘玉佩, 朱明星. 国际大数据领域研究热点及其演化路径分析 [J]. *情报科学*, 2016, 34 (4) : 158-163.
- [8] 陈建龙, 韩秋明. 近十年现代大学制度研究热点领域及发展脉络分析——基于共词分析 [J]. *西藏大学学报 (社会科学版)*, 2012, 27 (4) : 165-171.
- [9] 任晓松, 孙天美, 赵国浩. 中国碳排放研究热点演化知识图谱分析 [J]. *科技管理研究*, 2018, 38 (10) : 235-243.
- [10] Havre S, Hetzler E, Whitney P, et al. ThemeRiver: Visualizing thematic changes in large document collections [J]. *Ieee Transactions on Visualization and Computer Graphics*, 2002, 8(1): 9-20.
- [11] Cui W, Liu S, Tan L, et al. TextFlow: Towards Better Understanding of Evolving Topics in Text[J]. *Ieee Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2 412-2 421.
- [12] Cobo M J, Lopez-Herrera A G, Herrera-Viedma E, et al. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field[J]. *J Informetr*, 2011, 5(1): 146-166.
- [13] Cobo M J, Lopez-Herrera A G, Herrera-Viedma E, et al. SciMAT: A new science mapping analysis software tool[J]. *Journal of the American Society for Information Science and Technology*, 2012, 63(8): 1 609-1 630.
- [14] 杨颖, 崔雷, 等. 基于共词分析的学科结构可视化表达方法的探讨 [J]. *现代情报*, 2011, 31 (1) : 91-96.
- [15] 张云, 华薇娜, 袁顺波, 等. WoS 数据库中专利分析论文的主题动态演进研究 [J]. *现代图书情报技术*, 2015 (1) : 17-23.
- [16] 刘艳华, 华薇娜, 钱爱兵. 国际健康信息研究主题的演化路径分析 [J]. *现代情报*, 2018, 38 (7) : 135-141.
- [17] Stojanovic M N, Mitchell T E, Stefanovic D. Deoxyribozyme-based logic gates[J]. *Journal of the American Chemical Society*, 2002, 124(14): 3 555-3 561.
- [18] Saghatelian A, Volcker N H, Guckian K M, et al. DNA-based photonic logic gates: AND, NAND, and INHIBIT[J]. *Journal of the American Chemical Society*, 2003, 125(2): 346-347.
- [19] 唐静静. DNA 计算在逻辑门中的应用 [D]. 淮南: 安徽理工大学, 2014.

Analysis of Theme Evolutions of DNA Computing

ZHANG Zi-ting, ZHENG Yan-ning, YUAN Fang

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: Based on the Web of Science core collection, the paper retrieves 2 432 papers related to DNA computing. By selecting keywords as the unit of analysis, co-word analysis and cluster analysis as method, SciMAT software as tool, this paper draws keywords overlapping map, topic strategic diagrams and theme evolutions map to identify research topics and their revolution paths. The research on DNA computing began in 1994. In the early stage, development was slow. With the more researchers focusing on this field, research topics changed diversely. The differences of research topics between each period were not highly obvious, and the evolution path were relatively simple. The research demonstrated four evolution paths in two research directions. Molecular computation and logic gates were the main research topics which are supposed to be research hotpots in the future.

Key words: evolution path; strategic diagram; DNA computing; SciMAT