doi:10.3772/j.issn.2095-915x.2018.05.005

基于单层引文网络的技术演进路径研究
——以机器学习领域为例

张爽　刘非凡　夏昊翔

大连理工大学系统工程研究所　大连　116024

摘 要 科学地把握技术发展动态和了解技术发展脉络对企业和政府制定创新战略具有重要意义。本文提出了一种基于专利单层引文网络社区划分识别技术演进路径的方法，并以机器学习领域为例，构建五个时间段下的专利单层引文网络，通过分析网络结构，发现机器学习领域存在一定的技术壁垒；接着使用社区发现算法识别专利主题，利用德温特手工代码映射主题内容，依据相邻时间段引文网络社区之间的引用强度构建主题演化冲积图，最终识别出四条主要技术演进路径，验证了方法的有效性，表明本文提出的方法有望对现有技术演进路径的分析方法起到一定的补充作用。

关键词：机器学习；技术演进路径；专利；引文网络

中图分类号：G35；G306

开放科学（资源服务）标识码（OSID）

Researching on Technological Evolution Path Based on the Direct Patent Citation Network
——A Case Study in the Field of the Machine Learning

ZHANG Shuang　LIU Feifan　XIA Haoxiang

Institute of Systems Engineering, Dalian University of Technology, Dalian 116024, China

Abstract It is of great significance for enterprises and governments to scientifically grasp the development trend of the technology. A method is proposed in this paper to track technological evolution paths by community detection in original direct citation network. After constructing five successive citation

基金项目：国家自然科学基金面上项目（71371040），国家自然科学基金重点项目（71533001）。

作者简介：张爽（1994–），硕士研究生，研究方向：复杂科研网络；刘非凡（1991–），博士研究生，研究方向：群体行为、复杂科研网络；夏昊翔（1972–），教授，博士，博士生导师，研究方向：复杂社会系统、知识管理与知识组织，E-mail: hxxia@dlut.edu.cn。

doi:10.3772/j.issn.2095-915x.2018.05.005

networks in the patent field of Machine Learning, the key topics in every period are identified through community detection algorithm and the topic content are represented by Derwent Manual Code. Then it is found that there are certain technical barriers in the field of machine learning through the analysis of network structure. Finally, four main technological evolution paths are tracked by detecting strong connections between communities in adjacent sliced time periods, which show that the method proposed in this paper is expected to supplement the existing approaches of tracking technological evolution paths.

Keywords：Machine learning; technological evolution path; patent; citation network

1 引言

进入 21 世纪，科学技术迅猛发展，新兴技术不断涌现，技术发展呈现出领域内不断深入、领域间交叉融合的态势 [1]。然而技术体系的日益复杂，使得技术人员越来越难以全面、准确地获悉特定领域的发展状态，捕捉技术发展的新兴方向。如何科学地把握技术发展动态，识别技术发展脉络，理解技术演进模式，对于企业和政府预测技术变化趋势，识别新兴技术的出现，合理配置科技资源，制定创新战略具有重要意义。因此，探究技术演进路径成为广大学者重要的研究课题。自 1982 年意大利经济学家 Dosi[2] 提出“技术轨道”的概念以来，技术演进路径研究的理论和方法不断丰富。目前关于技术演进路径的研究方法主要包括共现分析，主题模型分析，文本挖掘分析和引文分析 [3]。

共现分析是指利用技术内容间相互关联的信息，构建共现网络，描述技术领域的结构特征。Suzuki 等（2004）建立专利信息中 IPC 分类号共现网络，追溯了日本不同公司不同领域的产品发展的技术轨迹 [4]。Wu（2016）提出了一种基于关键字的专利网络方法，结合了德尔菲技术、层次分析法和网络分析，发现了生物燃料的技术演化轨迹 [5]。主题模型分析是指通过对专利文本数据建模，提取特征词汇表征专利主题。陈亮等（2017）采用 hLDA 从专利文本数据中提取主题，并以硬盘驱动器磁头领域为例进行实例分析 [6]。Wang（2014）结合 LDA 模型和词性标注方法揭示了 LTE 技术的新兴热点 [7]。文本挖掘是指利用数据挖掘的原理和方法直接对专利的文本信息的进行关键词提取，来分析技术主题和技术演化轨迹。李欣（2018）等基于不同专利技术主题的 SAO 语义结构相似性，揭示了新兴技术的演化轨迹中的规律性特征 [8]。然而，共现方法、主题模型分析和文本挖掘对技术水平要求高，分析结果的好坏依赖于语料库的有效性、关键词的质量、算法参数的设置和算法本身的准确性 [9-11]。由于上述三种方法的局限性，越来越多的学者开始着眼于利用专利的引文关系分析技术演进路径。

专利间的引用体现了技术的流动，通过引用关系构成的专利引文网络反映了技术内容的传递、积累、交叉和渗透。在文献引文网络中关键路径反映了知识流动的主路径 [12]，而专利引文网络中的关键路径可以被视为技术进步的主干。Verspagen（2005）开创性地对燃料电池领域进行了基于专利引文网络的关键路径识别

研究[13]。随后，Martinelli（2012）利用最优主路径演化网络算法（NETP）分析了电信交换行业的关键技术路径[14]。国内的学者围绕技术关键路径的搜索也展开了广泛的研究，其中最为广泛引用的算法是搜路径连接数目法（SPLC）和搜寻路径节点对法（SPNP）。SPLC 是通过计算两节点间的路径占所有路径的频次搜寻关键路径，而 SPNP 则是通过计算节点对连接节点的数量确定关键路径。范维熙等（2014）以 SPNP 算法为基础，提出了太阳能电池技术发展的四个阶段[15]。杨中楷（2016）利用 Pajek 软件计算了专利引文网络中每个节点 SPLC 值，识别出了混合动力汽车驱动系统的技术发展路径[16]。黄鲁成等（2013）提出了一种基于 SPLC 算法的技术轨道动态分析方法[17]。然而关键路径识别算法，虽然能得到技术领域演化的基本骨架，但是常常忽略次级路径包含的主要信息，且很难反映出领域的产生、分裂、融合和消亡等演化过程[18]。

专利之间的引用关系在一定程度上可反应所属技术主题间的相似性。据此，也有学者基于耦合和同被引关系，结合聚类算法，分析专利聚簇的关系和其随时间变化的情况识别技术演进路径。Huang（2003）针对台湾高科技电子公司技术研发状况绘制了专利耦合引文图，通过聚类将公司分为 6 类，结果发现半导体行业的企业之间存在着更高的相似性[19]。王健美（2014）通过构建专利同被引矩阵，并对其进行聚类和多维尺度分析，揭示了不同时期纯电动汽车产业技术的演进规律[20]。技术主题相似的专利之间引用关系较为密切，而属于不同技术主题的专利间引用相对较少，这使得专利引文网络存在一定的社区结构，即社区内连边稠密而社区间连边稀疏，而且这种社区结构可能和专利主题的结构相对应[21]。然而利用共施引关系建立的耦合网络，由于高被引专利的影响，往往连边十分稠密；利用共被引关系建立共被引网络，由于被引关系的滞后性，往往无法准确反映该阶段的研究内容[22]。这就使得以这两种方式构建的引文网络都不能很好的体现出社区结构。相比之下，利用直接引用关系建立的单层引文网络既能直接反应专利之间的联系，又能一定程度上弥补耦合和同被引网络的不足。因此，本文将基于复杂网络理论，利用直接引文关系建立专利引文网络，从网络社区划分的角度出发识别技术演进路径。

2 研究方法

本文研究方法的主要思路是：构建不同时间段下的专利单层引文网络，通过社区发现算法识别专利主题，利用德温特手工代码映射主题内容，依据相邻时间段专利网络社区之间的引用强度构建主题演化冲积图，追溯专利主题演化路径。具体分析流程如下图 1 所示：

图 1 基于单层引文网络的技术路径分析流程图

doi:10.3772/j.issn.2095-915x.2018.05.005

构建专利单层引文网络。专利单层网络是指利用专利间的直接引用关系构建的网络，即利用授权时间在所分析时间段中的专利和其所有直接引用的领域内专利构成的网络。对于专利主题分析，使用单层引文网络可以将专利间的相关性清晰且完全地呈现。在分析一个时间段的引用关系时，若只分析授权时间在这一阶段的专利和其之间的引用关系，网络则丢失了由共同引用更早时期的专利而产生的专利间的耦合关系；若分析授权时间在这一阶段及之前的专利和其引用关系，那么大量早期专利的存在将会影响对该时段专利内容的分析。而专利单层引文网络是由所分析阶段产生的专利以及对此阶段产生重要影响的已有专利和其间的引用关系构成的，这样既避免了大量早期专利的影响，又包含了由高被引专利产生的此阶段专利间的耦合关系，清晰且完全的呈现出某阶段的专利之间的相关性。在此基础上进行社区划分和聚类分析，更有利于识别专利主题内容。

识别技术主题。专利单层引文网络是有向图，其中网络中的节点表示专利，边代表专利之间的直接引用关系，其方向由施引专利指向被引专利。专利间引用关系的密切性可以体现专利主题的相似性，在网络中表现为主题相同或相似的专利之间连边稠密，而属于不同主题的专利间连边稀疏。这使得专利引文网络存在一定的社区结构，且这种社区结构很可能是专利之间主题聚集性的映射结果。因此依据专利引文网络结构进行社区划分，一定程度上能实现专利主题的聚类。对此，本文采用 Blondel 等 [23] 提出的模块度优化启发式算法进行社区划分。该算法是将模块度作为目标函数，通过迭代使目标函数最大化得到最终的社区划分结果。模块度是度量社区划分质量的指标，其值越接近于 1，表明划分出的社区结构强度越强，质量越好。其计算公式为：

$$Q=\frac{1}{2m}\sum_{i,j}\left[A_{i,j}-\frac{k_ik_j}{2m}\right]\delta\left(c_i,c_j\right) \quad \text{（公式 1）}$$

公式（1）中 $A_{i,j}$ 代表节点 i，j 间连接的权重，$k_i=\sum A_{i,j}$ 即与节点 i 相连的所有边的权重之和，c_i，c_j 表示节点 i，j 所在的社区，若 i，j 在同一个社区，则 $\delta(c_i,c_j)$ 等于 1，否则为 0。$m=\frac{1}{2}\sum_{i,j}A_{i,j}$ 代表整个网络的权重连接总和。

利用关键词映射主题。得到不同阶段的社区划分结构后，利用每个社区内的专利对应的关键词反应社区内容。本文选取德温特数据库作为数据源，由于德温特数据库中专利标题和摘要是为了方便读者理解重新编写过的，所以专利的标题和摘要都是描述性的、通俗的词语，缺乏能够体现技术特征的专业性、概括性词组。而在德温特数据库中，每个专利都有其对应三个或三个以上的德温特手工代码，这些手工代码（MC）是根据专利原始文献的摘要和全文对发明的重要特点和应用进行独家索引，能够反应某项专利的技术特征，技术创新方面和应用领域。基于以上考虑，本文将德温特手工代码作为专利关键词，将专利社区作为整体统计 MC 的频率，利用 MC 对应的含义和频率绘制词云图，以反映社区内容。

识别技术演进路径。对特定时间段的专利引文网络社区结构和主题词进行考察，可以把握该阶段技术的主题内容和主题之间的关系。专利引用关系反映了技术的传递、积累、继承和发展，依据相邻时间段社区之间的引用关系

可以建立不同时间段社区间的联系，进而可以根据联系的强度识别技术演进路径。本文采用回溯的方法寻找经典技术演进路径，对于某一特定社区，其包含的施引专利引用前一阶段某个社区的数量越多、引用关系越强，则可认为这两社区间存越强的继承性。相邻阶段的引用强度计算公式：

$$W_{i,j}^{t}=\frac{N_{i,j}^{t}}{\sum_{j}^{m}N_{i,j}^{t}} \qquad \text{（公式 2）}$$

公式 (2) 中，$W_{i,j}^{t}$为归一化后的第 t 阶段的 i 社区对第 t-1 阶段的 j 社区的引用强度；$N_{i,j}^{t}$为施引专利属于第 t 阶段的 i 社区且被引专利属于第 t-1 阶段的 j 社区的引文数量；$\sum_{j}^{m}N_{i,j}^{t}$为施引专利属于第 t 阶段的 i 社区且被引专利属于第 t-1 阶段的所有社区的引文数量，m 为第 t-1 阶段的社区数量。依据相邻时间段社区引用关系绘制专利主题演化冲积图，自后向前追溯最强的社区间引用关系，进而识别出技术演进路径。

3 案例分析

本文选取“机器学习”（Machine learning）领域作为分析对象。从上世纪 80 年代末，机器学习的研究开始复兴。如今作为推动人工智能发展的核心动力，机器学习几乎占据了当今技术话题的半壁江山。AlphaGo、索菲亚、智能仓储和无人驾驶汽车等无一不在彰显机器学习对科技发展的影响。2017 年，图灵奖的获奖者 John E. Hopcroft 表示机器学习将推动下一次的工业革命。同年，国务院发布的《新一代人工智能发展规划》，更是将人工智能的发展上升到国家战略。因此，了解机器学习相关技术领域应用的发展现状，把握发展趋势，对技术研发人员、企业和政府都有参考价值和借鉴意义。

3.1 数据

以 Derwent Innovation Index(DII) 数据库为来源数据库，本文根据“机器学习”领域的核心文献和最新综述对“机器学习”领域的概括确定检索式为：TS=(‘machine learning’) OR TS=(‘support vector machine’) OR TS=(‘neural network’) OR TS=(‘decision tree’) OR TS=(‘genetic algorithm’) OR TS=(‘bayesian network’)。该检索式包含了“机器学习”领域各个学术流派的代表性方法，能够较为精确且全面搜集到“机器学习”领域的专利，设置检索时间跨度为 1963-2017，最终检索到 30496 条专利。

对检索到的专利数据进行预处理，即抽取领域内专利之间的引用关系、专利对应的德温特手工代码以及合并同族专利。同族专利是指在不同国家申请的相同专利及其衍生专利的集合，其中最早申请的专利是专利族中的基础专利。故本文在构建专利单层引文网络时，将同族专利集合视为一个专利，同族专利集合的引文视为基础专利的引文，以基础专利的授权时间作为同族专利集合的时间。

3.2 基础数据分析

为了把握机器学习领域技术发展的基本状况，本文首先统计了每年授权的专利数量，如图 2。机器学习领域最早的专利出现于 1977 年，该专利是一个使用决策树方法发明的实时条码分解器，此后机器学习领域的技术开始缓慢发展。直到 2000 年，每年的专利

doi:10.3772/j.issn.2095-915x.2018.05.005

授权数量开始快速增长，尤其是 2017 年专利授权数量剧增，机器学习领域的技术进入了高速发展的阶段。

通常认为专利的被引频次可以衡量某项专利技术对领域发展的贡献和作用，因为往往是提出了创造性的技术或方法的专利会得到后续专利的大量引用，为技术的发展奠定基础。统计发现机器学习领域的专利被引频次分布近似呈现幂律分布，如图 3 所示。大量的专利仅有 1-4 次的被引次数，只有少量的专利的被引次数较高，最高被引次数为 65。专利的被引次数整体偏低，这一特点可能是由专利申请者的引用动机造成的。这里需要说明的是，本文中的被引频率仅是指被数据集内专利引用的次数。高被引专利往往代表了技术的突破，会对技术的发展起到很大的驱动作用，是分析技术演化过程的重要切入点 [24]。表 1 列出了被引频次最高的 5 篇机器学习专利的有关信息。这五项专利的授权时间在 1987-1991 年，处在机器学习技术起步阶段。这五项均与神经网络技术有关，其中 2520 号专利将反馈神经网络应用到关光学字符识别系统，28539 号专利将混合模拟 - 数字联想神经网络方法用于存储器的信号处理过程，16875 号专利和图像处理系统有关，28500 号专利使用反馈神经网络进行信号处理，28379 号专利将神经网络应用于生产过程的控制中。可以看出，28539 号和 28500 号专利的主题相同，均和信号处理有关，2520 号和 16875 号专利的主题相似，这一结论也在后续社区划分中得到了印证。

专利的手工代码表征了专利的技术特征和应用领域，可以认为高频手工代码能够反映技术领域常用的研究方法和重点研究方向，手工代码的频率占比的变化可以反映某个方法或研究方向生长或衰退的发展趋势。因此，统计手工代码频率的历时变化情况，能够发现技术重点的演变以及和新兴技术的涌现。机器学习领域总共涉及 6312 种手工代码，限于篇幅，表 2 展示了前 20 个高频手工代码。根据高频手工代码可知，神经网络、遗传算法、模糊逻辑和专家系统等是机器学习领域专利常用的研究方法。专利通常涉及医疗、图像分析与识别、语言文字识别、工业控制、搜索引擎和汽车微处理器系统等应用领域。这里需简要说明，考虑到“Neural networks”的词频很高，会对主题内容可视化效果造成干扰，故在后续分析时，我们去除了该手工码；且经查看专利标题和摘要，我们发现“On-line Education”，“Claimed software products”和“Educational aids”这三个手工代码和其对应专利的内容关联度很小，故在此节以及后文分析时，我们去除了这三个手工代码。通过对手工代码频率占比统计发现，手工代码频率占比的变化趋势有波动不变，波动上升，下波动降和急剧上升四种情况。这里以图 4 显示的部分高频手工代码频率占比变化趋势为例，进行简要描述。在机器学习技术领域，专家系统和模糊逻辑系统一直常用的方法，人们对“图像分析”，“数据库应用和存储”方面研究的关注度越来越高，而与“医疗”相关的频次呈下降趋势，很可能是当前机器学习与“医疗”相结合的技术领域更加细分所致。“数据转换”方面的技术主题在近十年来呈现快速发展的态势。

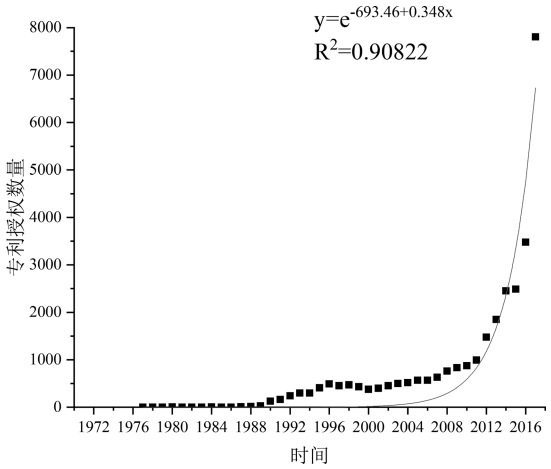


图 2 专利授权数量变化时序图

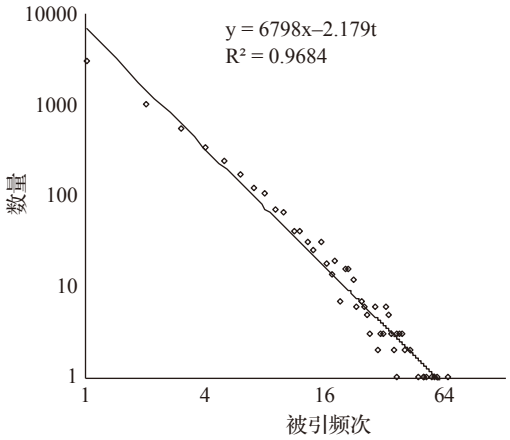


图 3 专利被引次数分布

表 1 被引频次最高的 5 篇机器学习专利的有关信息

编号	被引频次	专利授权时间	专利标题
2520	65	1991	Neural network for optical character recognition system - has back propagation and adjustment of neural weight an bias values controlled through output confidence measure
28539	58	1987	Hybrid analog-digital associative neural network - distributes feedback from successive rows of synaptic information to sample-and-hold circuit array until all data in RAM have been stored
16875	57	1990	Neural network image processing systems is multilayer - has layers of neurons receiving signals from input stimulus and also output of other layers
28500	55	1990	Signal processing system using neural network - has processing section which learns according to rule of back propagation to provide efficient learning
28379	52	1991	Process control using neural network - predicts values of indirectly controlled process variables to change directly values and optimise system

注：此处编号是本文定义的清洗后专利数据集中的专利编号

表 2 机器学习领域高频手工代码

排序	手工代码	手工代码的意义	频率
0	T01-J16C1	Neural networks	10770
1	T01-J10B2	Image analysis	3859
2	T01-J16C4	Genetic algorithms	3327
3	T01-J05B4P	Database applications	2805
4	T01-J05B2	Storage	2580
5	T04-D04	Recognition	2400
6	T01-J16C2	Learning	1831
7	T01-J07B	Computer control of manufacturing/industrial machines and Quality Control(QC)	1803
8	T01-J04B2	Correlation function	1513
9	T01-N01D	Data transfer	1486
10	W04-W05A	Educational equipment in general	1438
11	T01-J10B2A	For recognition	1432
12	T01-J05B3	Search and retrieval	1408
13	T01-N03A2	Search engines and searching	1209
14	T01-J03	For evaluating statistical data	1172
15	T01-E01A	Sorting	1172
16	T01-J04E	Mathematical modelling	1156
17	T01-J07D1	Vehicle microprocessor system	1121
18	T01-J16A	Expert systems	1064
19	B11-C11	General computing methods and apparatus	1059

doi:10.3772/j.issn.2095-915x.2018.05.005

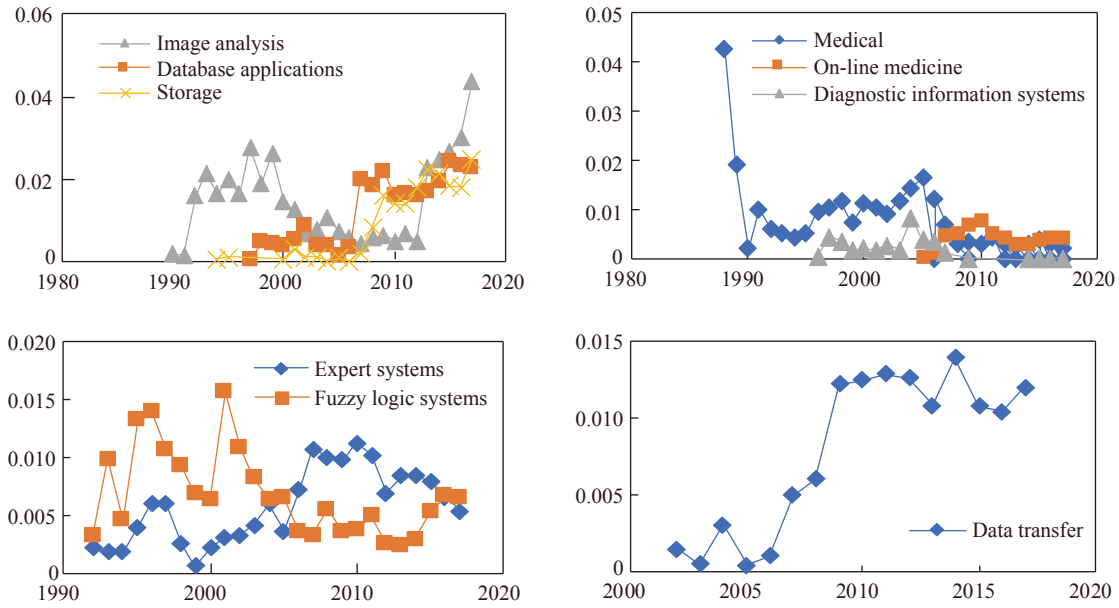


图 4 部分手工代码频率占比变化图

3.3 专利单层引文网络结构分析

本文统计发现，机器学习领域内施引专利和其被引专利之间的授权时间间隔集中在 3-6 年，如图 5 所示。基于此结果，本文将获得的数据集划分为 5 个阶段。考虑到 1997 年之前专利数量较少，第一个阶段的时间跨度较大，为 1963~1997。后四个阶段的时间跨度均为 5 年，分别为 1998~2002，2003~2007，2008~2012，2013~2017。基于前面的研究方法，构建了 5 个阶段的专利单层引文网络，网络结构如图 6 所示。为了进一步验证相对于耦合网络和同被引网络，单层引文网络的社区结构更加凸显，有利于社区划分，本节比较了三种方式下网络的基本特征。表 3 展示了单层引文网络、同被引网络和耦合网络的各项基本网络特征指标。

从表 3 看，单层引文网络各个阶段的网络模块度较高，且显著高于同规模随机网络的模块度，这说明网络中存在明显的社区结构，可以对其进行社区划分。通过对比可以看出，单层引文网络的节点和边的数量都比较适中，网络的密度也明显小于耦合网络和同被引网络，这表明单层网络的结构相对清晰，其相对高的模块度值也印证了这一点。

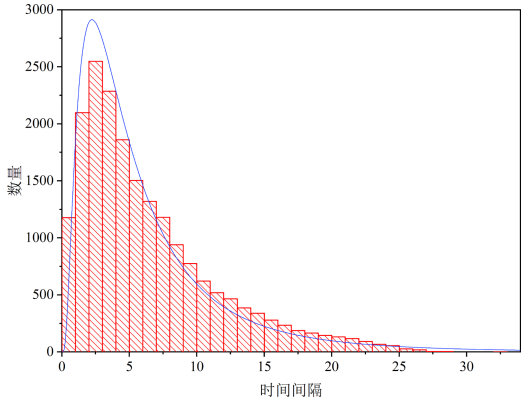


图 5 专利的时间间隔分布

图 6 是各阶段单层引文网络结构图，图中每个节点代表一个专利，连边代表专利之间的引用关系，不同节点颜色代表不同的技术主题社区。由图 6 可以看出，机器学习领域的专利引文网络呈现辐条状结构，这种结构通常为由一个节点或若干节点向外呈链状延展，也使得

总体的网络直径相较于同阶段的耦合和同被引网络要更长。社区内部存在更小的子团，这表明该领域存在一定的技术壁垒，即某一专利只会引用具有与其十分相似或相同的技术主题的专利。表 4 包含了第五阶段三号社区中的两个“辐条”（图 7）的标题信息。这两个辐条对应的专利内容均与动力系统有关，辐条 1 聚焦于使用 SVM 的方法解决动力系统的预测和控制问题，辐条 2 则着重于利用贝叶斯神经网络解决动力系统的故障诊断问题。可以看出两个辐条对应的专利从关注的子问题到使用的方法均有一定的差别。

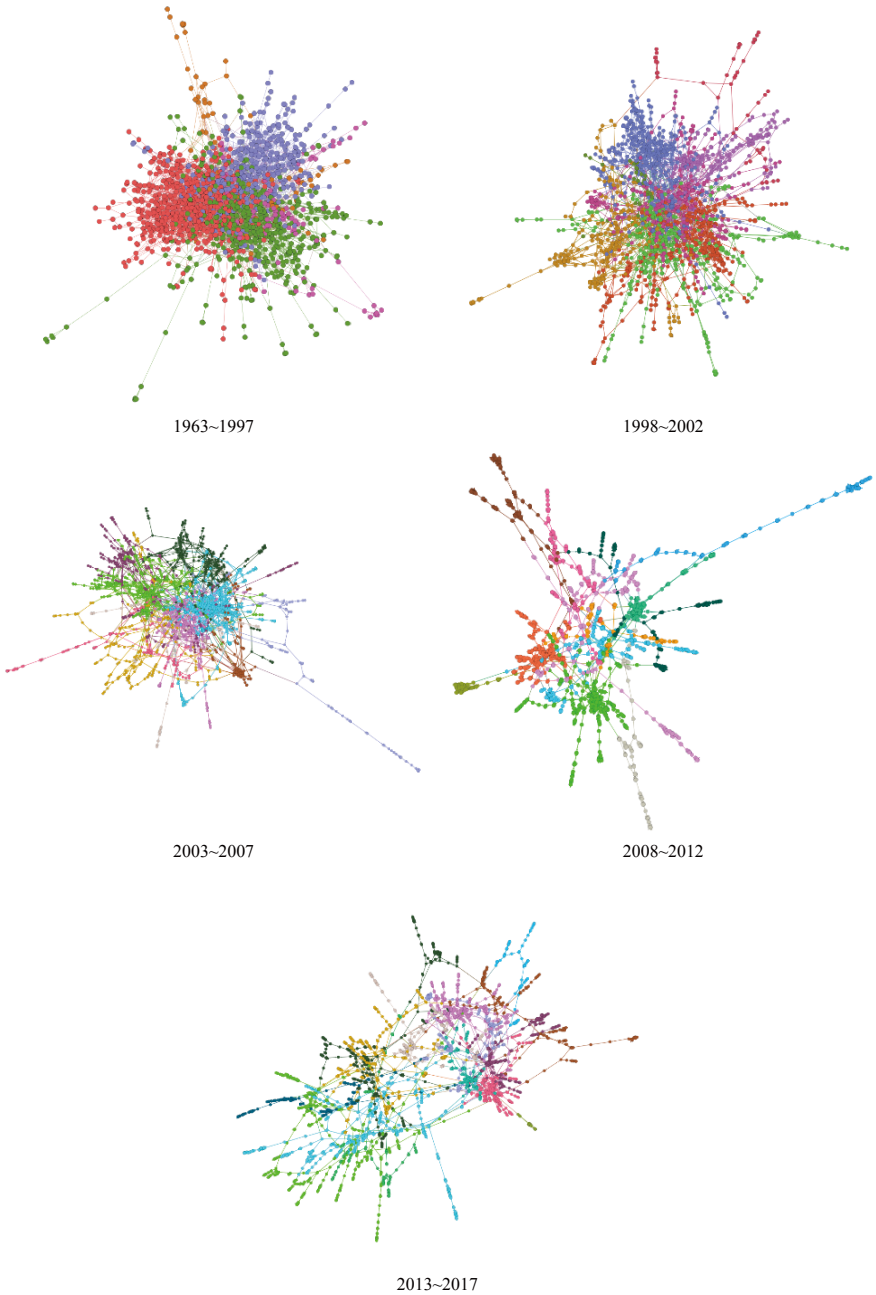


图 6 各阶段专利单层引文网络

doi:10.3772/j.issn.2095-915x.2018.05.005

表 3 三种引文网络基本特征指标

	阶段	最大连通子图节点数量（占比）	最大连通子图边数量（占比）	网络密度	模块度	社区数
同被引	1963~1997	766(94.68%)	12470(99.74%)	0.043	0.349	7
	1998~2002	1092(90.47%)	7715(98.8%)	0.013	0.549	8
	2003~2007	1133(84.81%)	7271(97.24%)	0.011	0.67	12
	2008~2012	1037(65.26%)	5316(90.29%)	0.01	0.805	14
	2013~2017	1583(58.41%)	10357(90.38%)	0.008	0.635	16
耦合	1963~1997	1004(98.62%)	22088(989.95%)	0.044	0.351	8
	1998~2002	696(93.67%)	3873(99.28%)	0.016	0.478	9
	2003~2007	714(88.92%)	3953(97.58%)	0.016	0.648	14
	2008~2012	598(60.47%)	1893(82.74%)	0.011	0.754	10
	2013~2017	1134(60.38%)	3351(83.94%)	0.005	0.796	15
单层引文	1963~1997	1265(94.19%)	4110(98.89%)	0.005	0.422	5
	1998~2002	1607(84.67%)	2889(93.71%)	0.002	0.672	8
	2003~2007	1800(78.47%)	2967(90.05%)	0.002	0.765	10
	2008~2012	1768(53.05%)	2336(67.97%)	0.001	0.854	12
	2013~2017	3021(50.83%)	3933(66.46%)	0.001	0.878	15

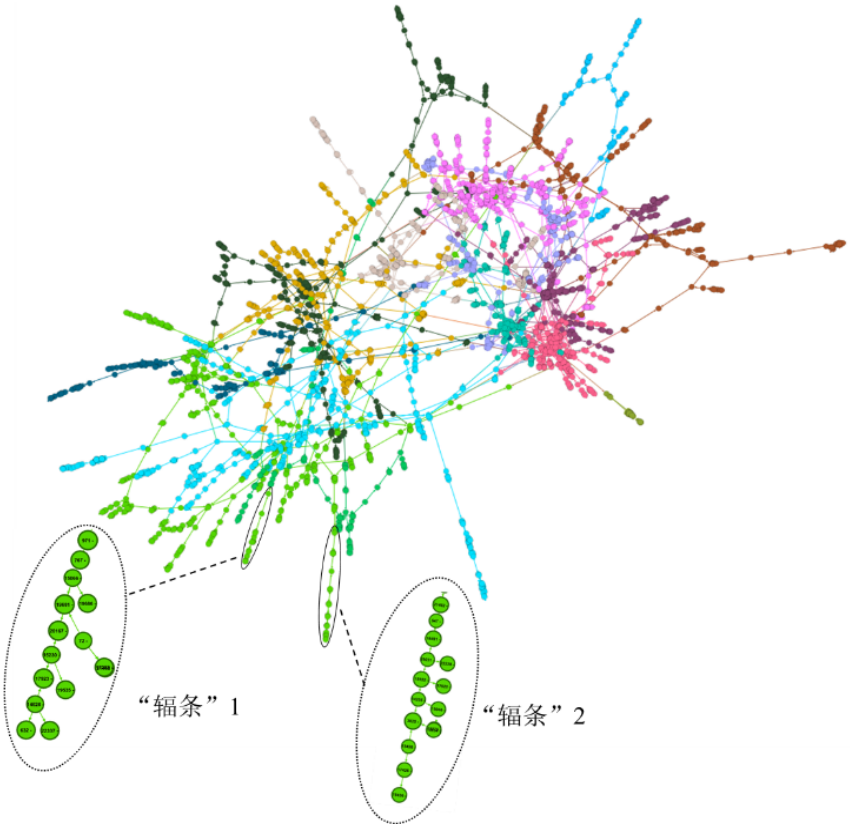


图 7 第五阶段三号社区中的两个“辐条”结构

表 4 第五阶段第三社区中的两个“辐条”的标题信息

编号	标题
	“辐条”1
971	**Wind power prediction method** for e.g. electric power system, involves obtaining minimum value of square regression model parameter and transferring optimized model of least square **support vector machine** based on minimum value
767	Wind electric field short-term **wind speed prediction method**, involves extracting and recording data of wind power field data collecting and monitoring control system, and obtaining wind speed prediction result
15066	Optimal small wave conversion and auto **regressive-support vector machine** bag based stable wind **speed forecasting method**, involves forecasting forecast data of low and high frequency signals to obtain forecast result
19681	Method for predicting organic mixture explosion characteristics, involves determining explosion characteristics of organic mixture using support vector machine model
20167	Online **mixed prediction method** for short-term wind speed of wind power plant, involves reconfiguring prediction results given by time sequence prediction and **support vector machine** prediction so as to obtain wind speed predicted value
15230	Maximum power tracking control method of **mini-type permanent-magnetic direct drive wind power generation system**, involves determining fan rotating speed based on observation using maximum power to preset step length tracking
17923	Device for tracking large power of **mini-type permanent-magnet direct-drive wind generating system**, has rectifier connected with tri-phase output end of fan, and load whose two ends are respectively connected with capacitor and ground
14626	Gravitational neural network based **wind power generating system MPPT controlling method**, involves configuring power rotating speed-sample of wind speed, and providing disturbance observation for tracking maximum power of fan rotating speed
	“辐条”2
21462	**Power system fault diagnosing** method for topological transformation, involves forming fault possible index set formula of fault candidate element, and determining fault element according to fault possible index set
547	**Bayesian network** transformer based comprehensive malfunction diagnosing method, involves collecting characteristic of transformer operation monitoring part, and establishing process of **dynamic malfunction diagnosis** by Bayesian network model
18491	**Bayesian network based transformer fault diagnosis analysis method**, involves determining property variables and fault type variables of transformer oil chromatography using characteristic gas of transformer oil chromatography
29231	Optimization method of alarm system based on **Bayesian network**, involves using Bayesian network to adjust alarm network, and making rapid positioning of root cause of alarm to reduce alarm flood
15152	Semi physical simulation technology based electric power communication network risk evaluating method, involves judging whether **failure degree of network** is determined, and determining sequence probability distribution value of root node
14255	Method for analyzing dynamic fault tree in repairing time limitation system, involves establishing **dynamic Bayesian network**, fixing repairing box to sub-system, establishing dynamic fault tree model, and obtaining system reliability
3825	Polymorphic malfunction tree-based **Bayesian network** malfunction forecasting method, involves obtaining component malfunction detection event, and performing multi-predicted malfunction mode state probability distribution function
13426	**Pipeline fault prediction method**, involves obtaining current pipeline status data, and determining current line happen fault probability, pipeline generating history fault probability and outlet pipeline failure prediction probability
17426	Multi-index ground **Bayesian network** working status data evaluating method, involves constructing Bayesian network based on satellite working status, obtaining history data, and performing Bayesian network working status evaluating process
18436	Operator status assessment reliability analysis method for power plant, involves establishing causal relation of performance shaping factors, and performing **Bayesian network** causal analysis based on probability distribution

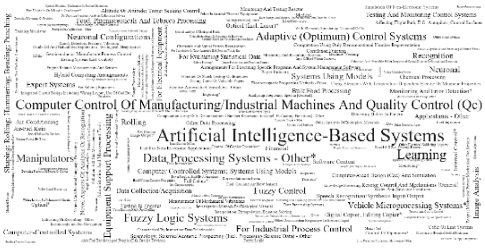
3.4 主题识别与技术演进路径分析

本节将划分出的社区看作整体，利用手工代码作为主题词表征社区内容，并制作词云图进行可视化。从可视化结果看，同社区内的主题词的相似性和关联性明显高于社区间的，且每个社区中的主题词既包含了技术主题涉及的方法又包含了技术的用途，能够明确全面的概括主题内容。限于篇幅，这里仅展示第一阶段的词云图（图 8），并对其进行简要说明。第一阶段划分为 5 个社区，1 号社区主题词主要

doi:10.3772/j.issn.2095-915x.2018.05.005

有“语音识别”、“伴音系统”、“自然语言和图像语言处理”和“字符信号的模式识别”等，该主题可概括为“数据分析与识别”。2 号社区为“工业控制”主题，高频主题词包括“人工智能系统”、“机器和质量控制”、“适应控制系统”、“模糊逻辑系统”和“模糊控制”。3 号社区为“图像处理”，“颜色系统”、“彩印”、“全彩”、“影像增强”表明了第一阶段该主题的重点。4 号社区的词云图展示了“医疗主题”发展初始阶段的内容。“数据处理系统”、“其他数据处理”、“图像处理”等主题词表明 5 号社区的主题为“数据处理”，该主题主要涉及“光电子设备”、“存储硬件”等方面、且使用“混合计算”、“神经网络”等方法。

1号社区



2号社区

3号社区

4号社区

5号社区

图 8 第一阶段的词云图

在得到的各个阶段的专利主题后，根据相邻阶段社区间的引用关系绘制主题演化冲积图，结果如图 9 所示。图中每一列方格代表一个阶段，每个方格代指一个社区，方格的大小代表社区被引量在所处阶段占比的相对大小。社区间河流的宽度是由引用强度决定的，其中社区的流出量衡量了该社区专利对下一阶段各个社区影响力的大小，而社区的流入量表明了该社区受上一阶段各个社区影响的大小。冲积图能够较为完整地展示技术主题间的交叉融合的关系，在此基础上依据最大引用强度，自后向前追溯各个社区的前驱社区，即可获得主要的技术演进路径。图 9 中 4 条深色线条表示了 4 条主要技术演进路径。

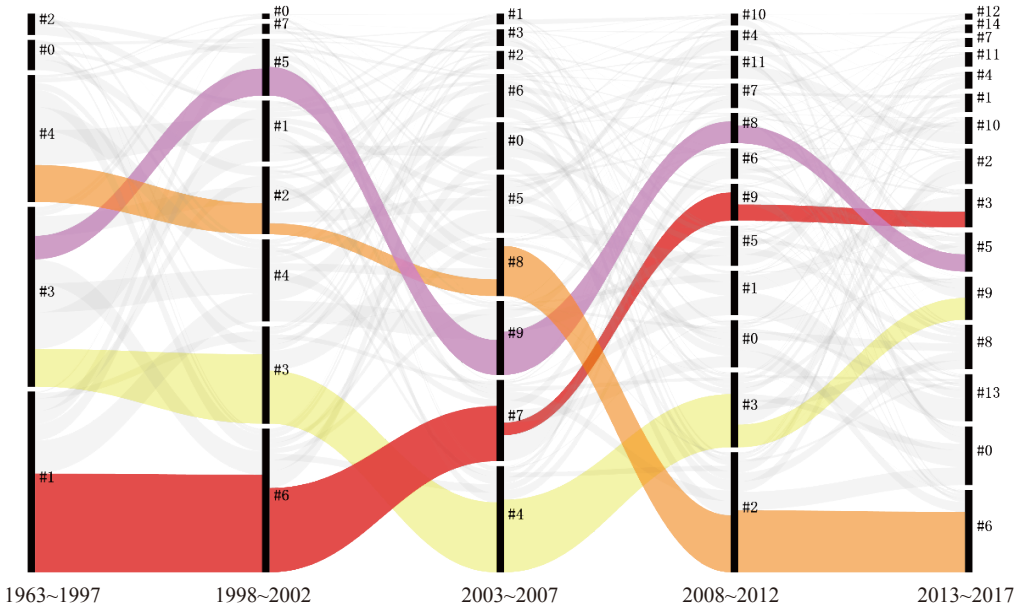


图 9 主题演化冲积图

图 10，11，12 和 13 分别展示了四条核心技术主路径及其对应的词云图。从词云图可以看出，处在同一主分支的社区，其词云内容无论是在主题词种类还是频次上都显示了高度的相似性，而不同分支上的词云内容呈现很大的差异性，这表明这 4 条主分支代表了 4 条不同的技术演进路径。而每个主分支的词云内容在不同阶段又呈现一定的差异性，正是这些差异性反映出了技术的演进。本文将结合主分支对应的词云图和专利文本内容对四条演化路径进行解读。

a) 第一条技术演进路线。

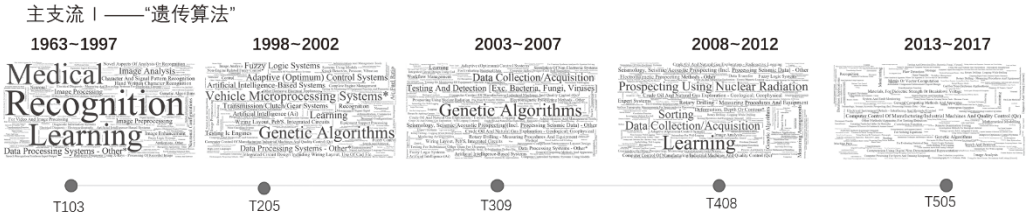


图 10 主分支 I

由图 10 知，该条路径主要关于遗传算法的技术应用。遗传算法主要有两种技术应用方式，一是直接用于技术设计解决优化问题，如确定设备参数，解决调度问题，选择最优方案等；二是优化技术设计中神经网络的结构。起初，遗传算法的直接应用比较多，比如评估制造设备生产线配置方法，计算机集成电路的功耗优化设计，设计半导体参数调节器，优化数字信号编码方案等，优化油井循环蒸汽的调度等。随后，遗传算法和神经网络结合使用的技术增加。例如，利用矿井参数确定开采方案，开发企业任务调度系统等。然后，遗传算法的直接应用开始减少，

doi:10.3772/j.issn.2095-915x.2018.05.005

而以神经网络为主，遗传算法为辅的石油和天然气的勘测技术的研究较多，如预测石油产量，确定开采位置等。近期，遗传算法应用频率相对降低，支持向量机和神经网络成为工业控制、故障诊断等技术研发常用的方法。

b) 第二条技术演进路线。

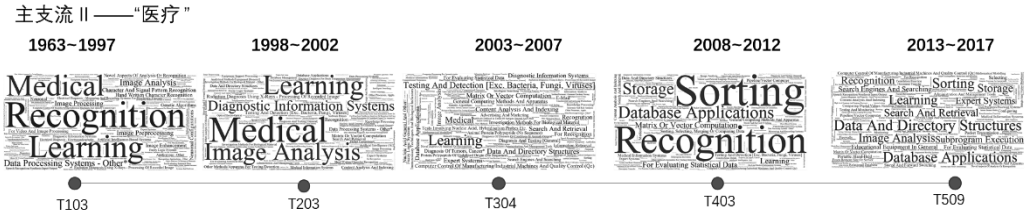


图 11 主分支 II

由图 11 可知，这是关于机器学习技术应用于医疗领域的演化发展路径。起初，人工神经网络技术开始被用于医疗诊断方面，且大多用于癌症和心脏的检测。比如，利用反馈神经网络监测心率信号识别心律失常；对心电图分类为患者提供心脏复建和除颤治疗方案；将细胞图像信息输入神经网络从而对细胞标本，血液样本进行分类；根据超声图像的像素强度特征，诊断乳腺癌；利用神经网络处理生理信息对疾病，如脑梗、心梗进行预测等。神经网络还用于放射影像的分析，比如根据图像的轮廓识别微钙化点检测癌症，对胸透图像进行分析判断间质性肺病等。随后，神经网络的应用更加广泛，如用于呼吸机监测系统，血管相关疾病的非入侵式诊断。所处理的信息也不只是图像信息，还有声音信号，如提取声信号特征检测心脏疾病。支持向量机（SVM）方法和决策树也开始被应用，医学诊断系统的功能进一步完善，如处理组织样本信息分析病理，根据患者数据确定药物剂量。接着，医疗领域开始利用机器学习方法对生物信息，如基因组，蛋白质进行研究，并利于这些信息辅助诊断。这时，支持向量机方法的使用开始增多，如 SVM 被用于基因探针的构建，对生物组织感兴趣区域进行分类等。然后，机器学习方法开始用于医疗方案推荐系统。例如使用 SVM 预测糖尿病患者对罗格列酶的治疗反应，决策树方法确定神经刺激疗法的参数配置等。近期，医疗领域的通用基础技术，比如图像识别，数据库系统也开始不断的发展和完善。

c) 第三条演进路径

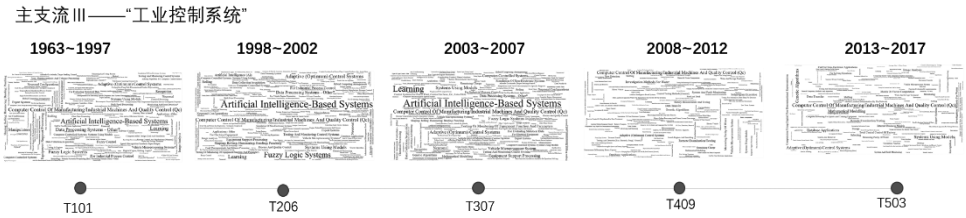


图 12 主分支 III

由图 12 可以看出，该路径上五阶段的词云图内容相差不大，这表明工业流程控制主题的内容没有发生较大的改变。起初，该领域主要通过使用神经网络和模糊逻辑算法实现工业生

产过程的自适应控制。比如利用神经网络实时调节注塑成型机加工参数，利用模糊神经网络控制器调节家用电器运行状态，利用神经网络和模糊逻辑构建自适应振荡控制系统减少飞机汽车的振动等。随后，遗传算法逐渐开始被应用起来。比如利用遗传算法确定输入特征，或对控制器中的模糊规则进行处理，或者优化控制器中神经网络权重值等。然后，该领域逐渐使用神经网络和遗传算法对生产诊断和质量预测方面的技术进行研究。比如利用神经网络构建轴承拉伸损伤状态识别系统，利用反馈神经网络设计电力异常故障数据分析装置，利用贝叶斯网络对变压器故障进行分类，基于遗传算法和神经网络算法对水泥原料力度进行测量等。

d) 第四条演进路径

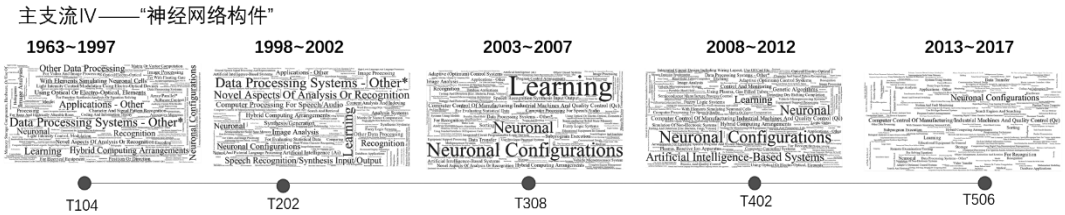


图 13 主分支Ⅳ

由图 13 可知，该条路径反映了有关神经元构件技术的演化过程。神经元构件主要用于构建物理神经网络。物理神经网络是指模拟真实的神经网络设计电路或芯片等硬件中元件和元件之间的结构，比如用可调电阻材料模拟神经突触的功能。起初，物理神经网络技术还处于起步阶段，大量专利研究集中于如何构造物理神经网络。随着纳米技术的发展，人们对物理神经网络中突触间的连接进行了改进，物理神经网络技术也逐渐应用于超大集成电路，半导体器件电路，微电子电路中。近期，人们又针对突触元件，突触间连接权重的设置等进行了改进。如在物理神经网络中相连的模块中间添加一个信号识别模块，构成基于记忆的图像识别系统。

4 结论

分析技术演进路径，把握技术的发展趋势，对企业发展和政府决策具有重要的参考意义。针对当前技术演进路径研究尚存的不足，本文提出了一种基于单层引文网络社区划分的技术路径分析方法，以期对目前的研究起到一定的补充作用。

本文以复杂网络理论和方法为基础，提供了一种从单层引文网络社区划分的角度出发，利用主题词映射主题内容，通过社区间引用强度识别技术演进路径的方法。并以机器学习领域为例，对该领域的技术演化进行分析，识别出了四条技术演进路径，证明了方法的有效性。而且，通过与耦合网络和同被引网络的网络特征对比发现，相较于前两者，单层网络在社区划分方面具有一定的优势。但这并不意味着单层网络分析可以取代耦合网络和同被引网络分析，因为不同的数据集，会呈现不同的领域特征，从而适合不同的网络构建方法。

本文所提方法虽然在机器学习领域的技术

doi:10.3772/j.issn.2095-915x.2018.05.005

主题识别和演化分析结果中得到了验证，但仍有需改进的地方。一方面，德温特手工代码不能详细的反映专利所用的具体方法，如从根据德温特手工代码我们不能看出某项专利究竟使用了哪种神经网络。另一方面，在四条技术演进路径中，每条路径的前四阶段的专利研究内容差异性较小，而最后一阶段的词云内容和前四阶段的具有较大的差异性，这种差异性的不同在一定程度上可以从技术的两种演化方式的角度解释。技术的进步主要有两种方式，一种是渐进的、积累的、连续性的线性演化过程，另一种表现为突变的，爆发式的非线性演化过程[25]。从图 2 和表 3 可知，第五阶段的专利数量和手工代码种类的指数级增长是导致最后一阶段与前四阶段在词云内容上差异较大的原因。未来将利用专利标题和摘要信息，结合德温特手工代码表征主题内容，并进一步探究造成领域爆发式增长的深层原因。除此之外，利用论文和专利的引用关系研究科学和技术的相互转化也是也是后续值得探究的课题。

参考文献

[1] 郭颖, 朱东华, 汪雪峰,等. 科学技术可视化[J]. 科学学与科学技术管理, 2011, 32(12):36-44.

[2] Dosi G. Sources, Procedures, and Microeconomic Effects of Innovation[J]. Journal of Economic Literature, 1988, 26(3):1120-1171.

[3] 赵阳, 文庭孝. 专利技术信息挖掘研究进展[J]. 图书馆, 2018(4):28-36.

[4] Suzuki J, Kodama F. Technological diversity of persistent innovators in Japan[J]. Research Policy, 2004, 33(3):531-549.

[5] Wu C. Constructing a weighted keyword-based patent network approach to identify technological trends and evolution in a field of green energy: a case of biofuels[J]. Quality & Quantity, 2016, 50(1):213-235.

[6] 陈亮, 张静, 张海超,等. 层次主题模型在技术演化分析上的应用研究[J]. 图书情报工作, 2017(5):103-108.

[7] Wang B, Liu S, Ding K, et al. Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: a case study in LTE technology[J]. Scientometrics, 2014, 101(1):685-704.

[8] 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(1):17-31.

[9] 王燕鹏. 国内基于主题模型的科技文献主题发现及演化研究进展[J]. 图书情报工作, 2016(3):130-137.

[10] 陈亮, 张志强. 技术演化研究方法进展分析[J]. 图书情报工作, 2012(17):59-66.

[11] 许海云, 王振蒙, 胡正银,等. 利用专利文本分析识别技术主题的关键技术研究综述[J]. 情报理论与实践, 2016(11):131-137.

[12] Hummon N P, Dorelan P. Connectivity in a citation network - The development of DNA theory [J]. Social Networks, 1989,11(1):39-63.

[13] Verspagen B. Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research[J]. Advances in Complex Systems, 2007, 10(1):93-115.

[14] Martinelli A. An emerging paradigm or just another trajectory? Understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry[J]. Research Policy, 2012, 41(2):414-429.

[15] 范维熙, 费钟琳. 基于德温特专利引文网络的技术演进路径研究——以太阳能电池技术为例[J]. 情报杂志, 2014(11):62-66.

[16] 杨中楷. 混合动力汽车驱动系统技术轨道识别研究[J]. 情报工程, 2016(2):8-17.

[17] 黄鲁成, 石媛嫄, 吴菲菲. 基于专利引用的技术轨道动态分析——以太阳能电池为例[J]. 科学学研究, 2013, 31(3):358-367.

[18] 张娴, 方曙. 专利引用网络主路径方法研究述评与

展望[J]. 图书情报工作, 2016(20):140-148.

[19] Huang M, Chiang L, Chen D. Constructing a patent citation map using bibliographic coupling: A study of Taiwan's high-tech companies[J]. Scientometrics, 2003, 58(3):489-506.

[20] 王健美, 刘志芳, 戴爱兵. 纯电动汽车产业关键技术演进分析——专利引文分析视角[J]. 图书情报工作, 2014(14):21-27.

[21] 张文琪. 综合主题分析的交叉学科合著网络演化研究[D]. 大连: 大连理工大学, 2016.

[22] 罗双玲, 张文琪, 夏昊翔. 基于半积累引文网络社区发现的学科领域主题演化分析——以"合作演化"领域为例[J]. 情报学报, 2017(1):100-110.

[23] Blondel V D, Guillaume J, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics-Theory and Experiment, 2008(10):155-168.

[24] Breitzman A F, Mogee M E. The many applications of patent analysis[J]. Journal of Information Science, 2002, 28(3):187-205.

[25] Adner R, Levinthal D. Demand heterogeneity and technology evolution: Implications for product and process innovation[J]. Management Science, 2001, 47(5):611-628.