
基于专利交易数据的高校技术创新团队 识别和跟踪方法研究 ——以浙江大学为例

赵显基¹ 张云^{1, 2} 袁顺波¹ 周勇¹¹

(1. 嘉兴学院 商学院, 浙江 嘉兴 314001;

2. 南京大学 信息管理学院, 江苏 南京 210093)

【摘要】: 技术创新团队的发现、培育工作,对于促进高校产学研联系具有重要的战略意义。利用 IncoPat 中高校专利交易数据,通过共现网络、聚类、战略坐标图识别技术创新团队,依据历时网络的关联、变化进行演变分析,结合发明人的科学研究背景信息获悉其科学技术关联,并以浙江大学为例,进行了实例研究。结果表明,自然语言处理技术与可视化知识图谱相结合的方法,能促进技术创新团队识别、跟踪工作的深入开展,并能提高中文专利数据的分析深度。

【关键词】: 技术创新团队 SciMAT 高校专利申请 技术转移 浙江大学

【中图分类号】:G306. 3 **【文献标志码】:**A **【文章编号】:**1671-3079(2020)03-0132-08

近年来,高校专利申请、技术转移工作逐渐引起众多学者的关注。一方面,随着现代科技、经济全球化的快速发展,越来越多的呼声希望高校能直接为经济社会发展作出贡献,高校的目标,正从传统的以教育和基础研究为主,转向进一步包括第三任务,即面向产业的广泛的技术转移活动,如研究合作、初创企业、专利许可等。^[1]另一方面,高校的专利申请、商业化工作取得显著成果并产生极大影响。如美国国会的《专利和商标法修正案》在一定程度上有效驱动了美国高校专利工作的发展,每年专利授权数量不断增长,向企业的专利许可数量及由此产生的收益呈大幅增长趋势。^[2]我国高校知识产权创造、运用和管理水平也在不断提高,专利申请量、授权量大幅提升。

专利发明人是对发明创造实质性特点做出创造性贡献的自然人,是技术进步最基本的要素和财富创造的原动力。^[3]而技术创新团队具备先进的技术水平,符合地区、国家经济和社会发展需要,能带来较大经济效益和社会效益,故其发现、培育工作,具有重要的战略意义。本文提出结合自然语言处理技术和开源知识图谱生成工具的发现方法,对高校技术创新团队进行识别和跟踪,旨在解决如何发现高校重要的技术创新团队、如何从历时角度揭示团队及其研究主题的动态演变、如何进一步了解这些研究团队的学术背景等问题,并以浙江大学为例,进行具体说明。

基金项目:2019年浙江省大学生科技创新活动计划暨新苗人才计划项目(2019R417002)

作者简介:赵显基(1999-),男,湖南娄底人,嘉兴学院商学院学生;张云(1974-),女,湖北黄冈人,嘉兴学院商学院副教授,研究方向为信息计量学。

一、研究方法

从技术创新团队识别和跟踪方法看,目前多数研究主要借助于频数统计、社会网络分析方法或数据挖掘算法,如张敏、沈雪乐利用普赖斯定律初选出核心作者,再利用综合指数法对每位作者进行精确的测评,以识别最终的核心作者。^[4]李红、陈少龙借助于社会网络分析法对智能手机专利发明人合作网络进行演化分析,以分析定位智能手机核心发明人和核心发明人团队。^[5]DUYong-ping 等人利用排序算法从基于专利数据的异构网络中确定核心发明人,并采用 LDA 模型对发明人的研究主题进行详细分析。^[6]传统的基于频数统计、社会网络分析方法研究视角较为单一,如仅能获得专利发明人或发明人合作程度的信息,难以有效结合其他线索进行深入分析。基于数据挖掘的算法虽能发现隐藏的模式,但结果展示较为简单,实现难度较大。

本文中的识别方法将自然语言处理技术与开源知识图谱生成工具相结合,能有效解决上述问题,具体流程如下:首先是领域选择,其范围可大可小,如某行业、特定技术领域、特定机构或企业等,以服务于不同的决策目标;其次是数据集下载,考虑到近年来我国高校专利申请数量激增,但质量水平普遍偏低,如高校申请的专利中高被引专利较少、推向国际市场的专利较少、实际进行科技成果转移、转化的专利较少。因此,本研究将申请人定义为高校,并使用授权发明专利中产生专利交易的记录用于技术创新团队的揭示和跟踪,主要考虑这部分数据反映的技术人员往往具有较为密切的产学研联系,是高校与企业重要的联系纽带,也是深入研究高校科研如何促进地方经济、行业发展的重要线索。接着进行数据清理工作,包括清除无效数据、数据格式转换、提取专利交易中的信息等。

在以上工作基础上,对技术创新团队进行识别:先基于清理后的专利交易数据生成专利发明人共现网络,利用聚类方法识别出重要的发明人团队,采用中心度、密度指标对发明人团队的战略位置进行布局分析,并结合团队的专利交易记录进行技术创新团队研究主题的识别;接着进行时间段划分,利用不同时间段聚类的演进、关联分析进行技术创新团队的跟踪;最后利用采集程序,进行重要发明人科学研究背景信息的采集和下载,并从专利数据和科学研究背景数据结合的角度,探究技术创新团队中的科学与技术关联关系。整个流程如图 1 所示。

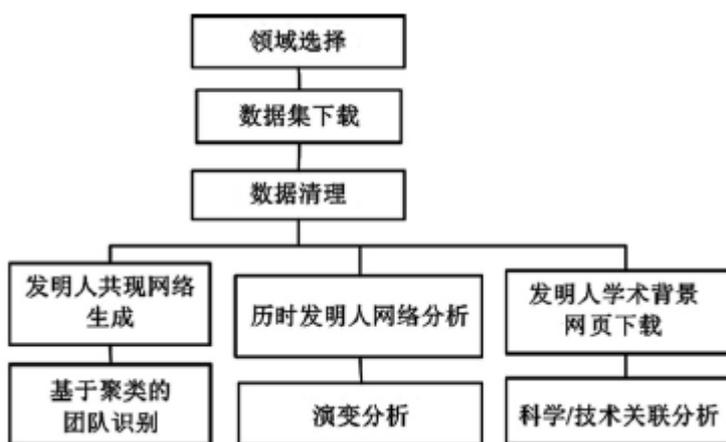


图 1 研究流程图

在分析过程中,本文选择 SciMAT 作为知识图谱生成工具。该软件分析单元主要包括作者、参考文献、词等,可进行发明人或发明人团队分析;可对分析单元进行多视角联合分析,如将聚类分析、共现分析、历时分析、各种计量指标,如 h 指数、发文量、被引次数等有效结合起来,便于发明人团队的多角度分析;上下文语义联系较为丰富,可将战略坐标图、演进图、聚类网络图有效联系起来,输出报告中还含有上下文语义信息,便于对发明人团队的动态跟踪和深入分析。^[7-8]此外,该软件是开源软件,易获得性较好,能获取源代码,便于二次改造。

二、数据处理

(一)数据来源

本文以浙江大学技术创新团队发现、跟踪为例示,选择 IncoPat 数据库为数据源。IncoPat 是将全球顶尖发明智慧深度整合,并将数据翻译为中文,面向中国企业决策者、研发人员、知识产权管理人员的专利信息平台。目前,IncoPat 收录了全球 102 个国家、地区或组织超过 1 亿件的专利文献,其中还包括中国大陆、美国、日本和中国台湾地区的诉讼数据,中国大陆和美国的转让数据以及中国大陆的许可、质押、复审、无效和海关备案数据。^[9]

为下载数据集,确定数据范围为“中国发明授权”,经 2020 年 1 月间的多次试检和试分析,最终确定检索式为“AP=(浙江大学)ANDAD=[20000101to20191231]”,其中 AP 为申请人,AD 为申请日期,初步检索后得到 21325 条发明授权记录。所得结果先按法律事件“诉讼、质押、转让、许可、海关备案、变更”进行筛选,共得到 1610 条记录,再进行手工排查,如去掉申请人仅为大学士工程管理有限公司、杭州华泰电液成套技术有限公司、杭州浙达精益机电技术股份有限公司等的无关记录,共得到 1588 条有效记录。

(二)数据预处理

从 IncoPat 下载的数据集并不符合知识图谱生成工具 SciMAT 的制作要求,如格式不符合导入规格,中文数据导入时会出现乱码问题。此外,记录集中部分数据还需进行规范化处理。为此,采用自编程序对源数据进行预处理,如进行格式转换、增加文件首尾标志、增加记录首尾标志、在记录末尾增加“ER”标记、在文件末尾增加“EF”标记等,为突出不同时段专利交易的主体,本文中的时间主要以转让或许可日期信息中提取的“年”为准。

(三)时间阶段划分及参数设置

SciMAT 是一个嵌入了方法、算法、度量指标的开源知识图谱制作工具,能够为图谱制作中的不同步骤提供支持。^[7]为识别和跟踪技术创新团队,需利用 SciMAT 进行发明人数据预处理,划分时间阶段,设置网络生成、聚类生成所需各种阈值,比较不同图谱的可解释性和有效性。以浙江大学为例,提取记录中“许可/转让年”后,按“年”进行许可/转让数量统计可得到图 2。

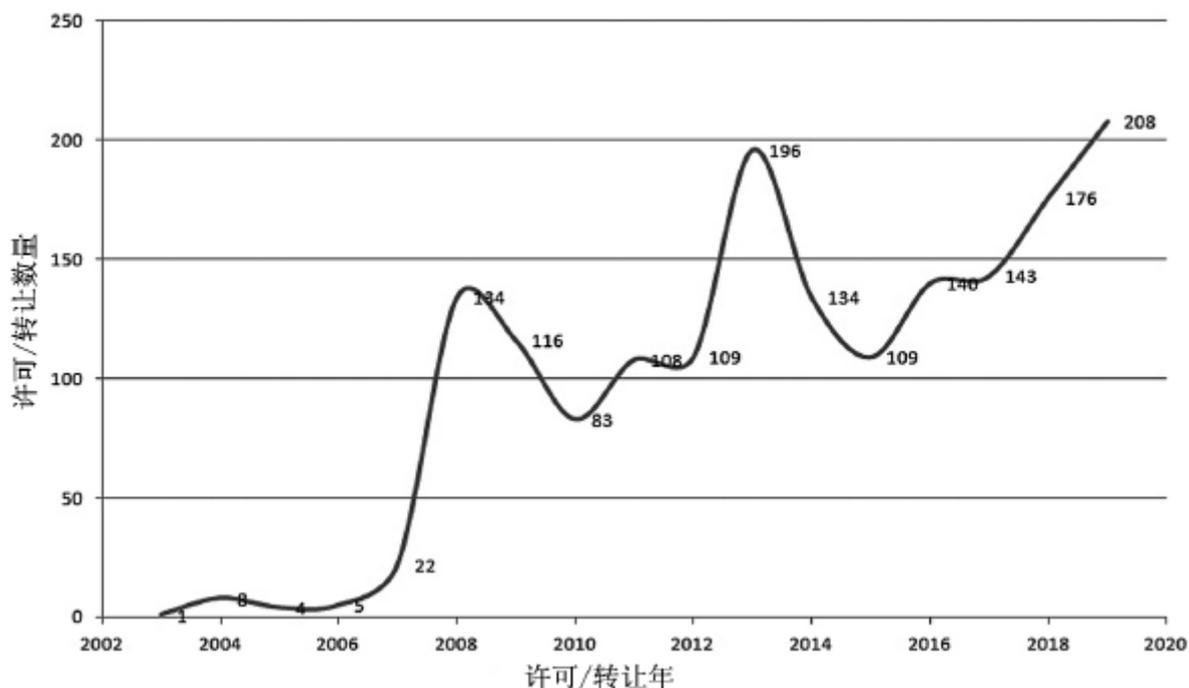


图 2 浙江大学授权发明专利交易数量统计图

由图 2 可知,浙江大学授权发明专利的许可/转让数量呈波浪式上升趋势。据此可划分出 3 个时间阶段,分别是 2003-2008 年、2009-2013 年、2014-2019 年。

经不同图谱分析对比,最终确定 SciMAT 中相关参数设置如下,分析单元为发明人,网络类型为共现网络,3 个时间段的最低频次分别为 3、5、5,最低共现频次分别为 1、1、1,聚类算法为 SimpleCenters,其中聚类成员数在 3~18 之间等,可得到 2003-2019 年浙江大学授权发明专利交易中发明人团队演进图、各时段战略坐标图、聚类网络图及相关的输出报告。

(四) 基于 Python 的数据采集

高校专利发明人具有特殊地位,尤其在科学与技术关联方面。来自于高校的专利发明人往往也从事科学研究工作,在特定领域进行论文发表、项目研究,有相对固定的校外合作关系,较有影响力的学者还通过指导博士生、硕士生等促进科学技术知识的扩散和转移。

为进一步获悉发明人、发明人团队的科学研究背景,特别是了解发明人技术发明与科学研究的关联,本文利用网络信息采集程序,从 CNKI 的作者知网节中进行发明人科学研究信息的搜索和下载,如发明人所属研究机构、研究领域、总发文量、总下载量、关键词、同机构主要合作者、其他机构主要合作者及合作机构、参与的基金项目类型及数量等,并在此基础上进行科学技术关联分析。

三、结果分析

(一) 技术创新团队的识别

1. 2003-2008 年主要技术创新团队分析

技术创新团队识别主要利用发明人聚类及聚类指标进行。其中,2003-2008 年浙江大学授权发明专利交易中主要发明人团队聚类的战略坐标图如图 3 所示。图中圆圈代表发明人聚类,圆圈大小被设置为交易专利数量。图中横坐标是中心度,表明不同聚类的重要程度;纵坐标为密度,表明不同聚类的发展程度。由此得到 4 个不同象限,不同象限的聚类具有不同性质。右上象限的聚类发展较好,且对领域的构成较为重要,通常被称为引擎性聚类,这些聚类内部紧密相关,且具有较好的外部相关性;左上象限的聚类内部联系较好,但不具有重要的外部联系,往往非常专业,但有一定外围性。左下象限的聚类既薄弱又边缘化,主要表现为新兴或消失的聚类;右下象限的聚类对研究领域很重要,但尚未开发,往往是横向、一般的或基本的聚类。^[10]

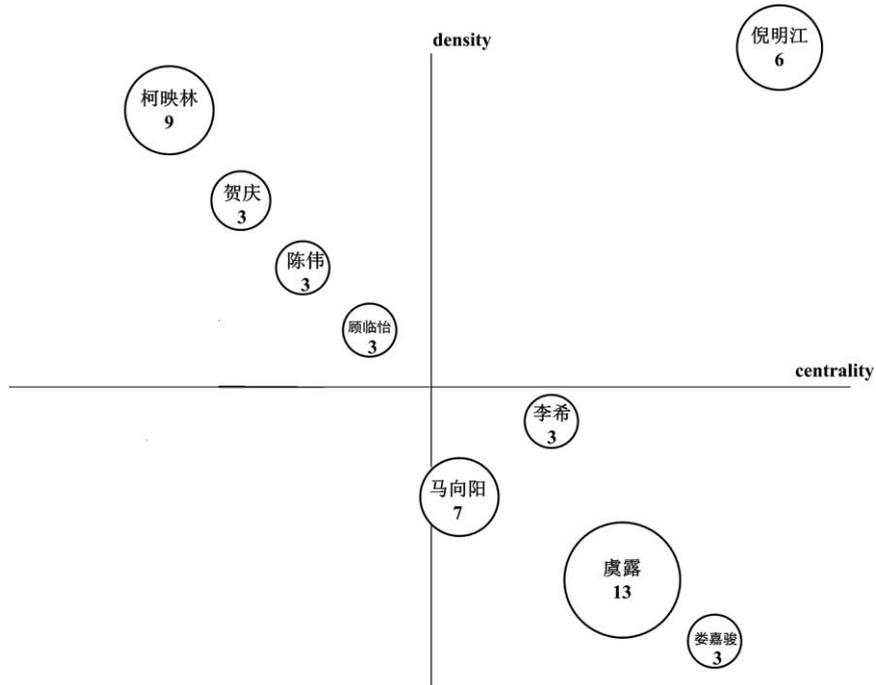


图 3 2003-2008 年浙江大学授权发明专利交易中主要发明人团队战略坐标图

图 3 中位于右上象限发展较好且较为重要的发明人聚类是倪明江团队。结合 SciMAT 输出报告中的聚类网络图、上下文信息可以发现,倪明江团队共 18 人,团队成员间联系较为紧密,主要成员包括岑可法、骆仲泐、高翔、施正伦、余春江等人,其专利交易内容多与烟气治理、燃烧工艺有关。

位于左上象限的相对专门性、偏外围性发明人聚类包括柯映林、贺庆、陈伟、顾临怡团队。柯映林团队共 18 人,团队成员间联系密切,主要成员包括贾叔仕、杨卫东、秦龙刚等人,其专利交易内容多与飞机部件位姿调整有关。贺庆团队共 7 人,主要成员包括程冀宇、李云飞等人,其专利交易内容多与丹皮制药有关。陈伟团队共 6 人,其专利交易内容多与电动助力转向控制器有关。顾临怡团队共 4 人,其专利交易内容多与液压马达有关。

位于右下象限的横向的、广义的、基本的发明人聚类包括娄嘉骏、虞露、李希、马向阳团队。娄嘉骏团队共 3 人,专利交易内容涉及气动阀门、家庭控制网络系统等。虞露团队共 3 人,专利交易内容主要涉及视频、图像编解码处理等。李希团队共 3 人,专利交易内容多与芳香羧酸生产有关。马向阳团队共 4 人,专利交易内容多涉及直拉硅片内吸杂工艺、制备等。

2. 2009-2013 年主要技术创新团队分析

2009-2013 年间浙江大学授权发明专利交易中主要发明人团队聚类的战略坐标图如图 4 所示。其中发展较好、且较为重要的

发明人聚类为叶兴乾、王龙虎团队。叶兴乾团队共 3 人,其专利交易内容主要涉及液态奶、乳品检测检验方法。王龙虎团队共 3 人,其专利交易内容主要涉及原花青素提取、复方茶含片制备等。

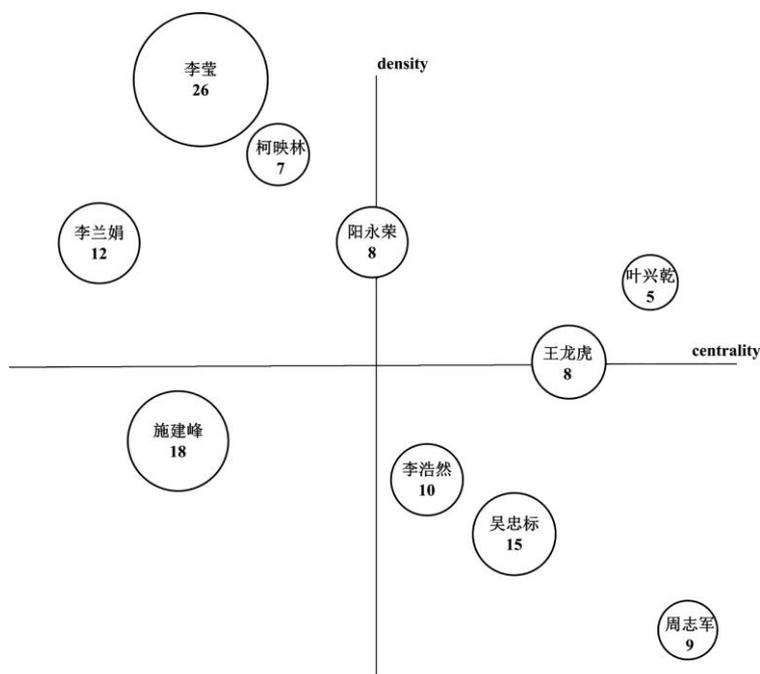


图 4 2009-2013 年浙江大学授权发明专利交易中主要发明人团队战略坐标图

相对专门性、偏外围性发明人聚类是阳永荣、柯映林、李莹、李兰娟团队。阳永荣团队共 5 人,主要成员包括王靖岱、吴文清等人,专利交易内容主要涉及烯烃聚合反应的装置、制备等。柯映林团队 5 人,主要成员包括蒋君侠、李江雄等人,专利交易内容主要涉及大飞机中机身数字化装配布局方法、飞机机身柔性化、自动化调姿方法等。李莹团队共 5 人,包括吴朝晖、尹建伟等人,专利交易的主要内容与 Web 服务有关。李兰娟团队共 7 人,团队成员包括喻成波、潘小平等人,专利交易的主要内容多与人工肝支持系统有关。

横向的、广义的、基本的发明人聚类是周志军、吴忠标和李浩然团队。周志军团队共 4 人,包括岑可法、张林等人,专利交易内容主要涉及渗透汽化膜组件制备等。吴忠标团队共 4 人,包括官宝红、赵伟荣等人,专利交易内容主要涉及脱硫脱硝工艺。李浩然团队共 4 人,包括陈志荣、胡柏剡等人,专利交易内容主要涉及化工产品、化学物质制备等。

新兴或衰退的发明人聚类是施建峰团队,该团队共 6 人,包括郑津洋、郭伟灿等人,专利交易内容主要与电熔焊机有关。

3. 2014-2019 年主要技术创新团队分析

2014-2019 年,浙江大学授权发明专利交易中主要发明人团队聚类的战略坐标图如图 5 所示。其中发展较好且较为重要的发明人聚类为陈丰秋、杨灿军、甘春标团队。陈丰秋团队共 12 人,包含陈新志、钱超等人,专利交易内容主要涉及裂解气分离。杨灿军团队共 4 人,包括金波、李德骏等人,专利交易内容主要涉及海底观测网络。甘春标团队共 4 人,专利交易内容涉及永磁涡流联轴器。

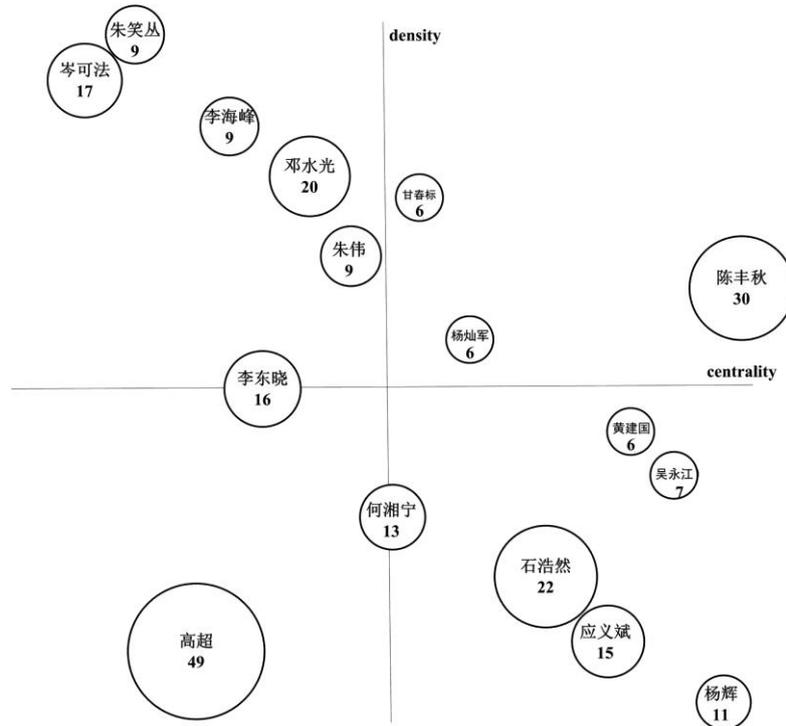


图 5 2014-2019 年浙江大学授权发明专利交易中主要发明人团队战略坐标图

相对专门性、偏外围性发明人聚类是朱伟、邓水光、李东晓、李海峰、朱笑丛、岑可法团队。朱伟团队共 5 人,包括武建伟、许方富等人,专利交易内容主要涉及卡簧自动装配机等。邓水光团队共 5 人,主要成员包括吴朝晖、尹建伟等人,专利交易内容主要涉及数据密集型服务协同系统组件服务部署等。李东晓团队共 6 人,主要成员包括张明、王梁昊等人,专利交易内容主要涉及深度图像处理。李海峰团队共 7 人,主要成员包括刘旭、郑臻荣等人,专利交易内容主要涉及显微成像系统、方法等。朱笑丛团队共 15 人,主要成员包括裴翔、张学群等人,专利交易内容主要涉及可穿戴式助力外骨骼控制方法等。岑可法团队共 18 人,主要成员包括倪明江、高翔、骆仲泱等人,专利交易内容涉及烟气脱硝等。

横向的、广义的、基本的发明人聚类是杨辉、吴永江、黄建国、应义斌、石浩然、何湘宁团队。杨辉团队共 3 人,专利交易内容主要涉及二氧化钛掺银抗菌剂等材料的制备方法。吴永江团队共 3 人,专利交易内容主要涉及中药材多指标成分含量测定等。黄建国团队共 3 人,专利交易内容主要涉及碳纤维材料的制备方法、产品及应用等。应义斌团队共 4 人,主要成员包括饶秀勤、谢丽娟等人,专利交易内容主要涉及农产品中图形图像处理。石浩然团队共 4 人,主要成员包括王维锐、葛正等人,专利交易内容主要涉及电动助力制动器及其控制等。何湘宁团队共 5 人,主要成员包括李武华、赵一等人,专利交易内容主要涉及单相五电平逆变器等。

新兴或衰退的发明人聚类是高超团队。高超团队共 9 人,包括彭鑫、许震等人,交易内容主要涉及石墨烯-锰复合纤维及其制备方法等。

(二) 2003-2019 年主要技术创新团队演变分析

为揭示技术创新团队演变情况,本文主要对 3 个时段中发明人团队、研究主题等的发展进行了关联分析。2003-2008 年,浙江大学授权发明专利交易总量为 172 条记录,交易量虽不大,但形成了两个比较重要的人数较多、成员间联系紧密的发明人团队:一是以柯映林、杨卫东等人为主的研究团队,研究内容涉及机用位姿调整装置或相关方法;二是以倪明江、岑可法等人为主的研究

团队,交易内容涉及烟气治理、燃烧工艺等。此外,该阶段专利交易主要分布在化工领域。

2009-2013年,浙江大学授权发明专利交易总量为587条记录,交易量、团队数量有所增长,专利交易内容明显丰富。其中,前一阶段的两个团队继续相关的研究,如何映林、杨卫东等人的机位姿调整装置,岑可法、周志军等人的烟气治理、燃烧工艺等。该阶段专利交易的主要领域仍旧分布在化工领域,但一些新的研究团队、研究主题不断涌现出来,如李兰娟、潘小平团队的人工肝支持系统的专利交易,李莹、吴朝晖团队的Web服务中一些管理与技术实现方法的专利交易。其中,李莹团队表现较为突出,团队成员人数不多,但专利交易数量较多。

2014-2019年,浙江大学授权发明专利交易总量为829条记录,专利交易活动进一步活跃,发明人团队不断壮大,交易内容进一步丰富,且呈现较强的新技术特征。其中岑可法、骆仲泐团队是浙江大学较为重要的发明人团队,并在3个时间阶段连续出现,从事着烟气治理、烟气脱硝等方面的研究和专利交易。吴忠标、陈新志团队在后两个时间阶段表现更为突出,其专利交易内容多与脱硫脱硝、裂解气分离等有关。值得关注的是,该阶段专利交易的领域不断扩大,除化工领域外,计算机技术、医学、新材料等领域的专利交易都表现出明显的增长势头。如2014-2019年,计算机技术领域中主题更为多样化,包括应义斌、饶秀勤团队的农产品图形图像处理,杨灿军、李德骏团队的海底观测网络,李东晓团队的深度图像处理等。医学领域中,朱笑丛团队的可穿戴式助力外骨骼控制方法相关专利、李海峰等的实时静脉定位显示装置和方法等都具有高科技特色。此外,高超团队的石墨烯-锰复合纤维及其制备方法、黄建国团队的碳纤维材料制备方法等,表征着新材料领域的专利交易正成为新的专利交易增长点。

(三)重要发明人科学研究背景分析

对授权发明专利交易数量排名前60的发明人,从CNKI的作者知网节中进行科学背景信息的搜索和下载,经分析后发现:一、重要发明人往往具有较高的科学影响力,如浙江大学前60的发明人中,文章下载次数超过1万次的达49人,发文量在100篇以上的有40人;二、高校技术发明团队具有一定的学科聚集性,并在这些领域形成了相对较大或较有影响力的发明团队,如专利交易数量前60位的发明人主要科学研究领域涉及有机化工、无机化工、化学、计算机软件及计算机应用、动力工程、电信技术等,而在浙大授权发明专利交易中,有机化工、无机化工、计算机软件及计算机应用的专利交易相对较多,且在这些领域都出现了一些较为显著的研发团队,如有机化工中吴忠标、陈新志等人的团队,环境科学与资源利用、化工领域的岑可法、骆仲泐、倪明江、高翔等人的团队,计算机软件与计算机应用中吴朝晖、邓水光等人的团队;三、重要发明人在科学研究过程中,主要的合作机构是我国高校、科研院所,但也有相当部分的发明人与企业有一定的科研合作关系,如前60位发明人中有38位与企业具有科学研究合作关系,其中16位发明人主要是与浙江企业具有科学研究合作关系;四、高校重要发明人往往主持或参与多项国家级、省部级重大科学基金项目,如专利交易总量排名前60的发明人参与最多的项目类型为国家自然科学基金、浙江省自然科学基金、国家高技术研究发展计划(863计划)、国家科技支撑计划、浙江省科技计划、国家重点基础研究发展计划(973计划)、高等学校博士学科点专项科研基金、中国博士后科学基金、国家科技攻关计划、教育部新世纪优秀人才支持计划、教育部留学回国人员科研启动基金、浙江省教育厅科研计划等。由此可以发现,自然科学基金项目、高水平人才培养发展计划基金项目等有力地推动了高校科学研究、技术研发工作的发展和壮大。除此之外,重要发明人的技术发明内容与科学研究方向具有较为紧密的关联,且技术研发团队和科学研究团队中同机构成员的构成具有相似性,以吴朝晖为例,其专利交易记录共51条,交易时间分布在2009-2019年,专利交易内容多与Web服务有关,其学科研究背景为计算机软件及计算机应用、电信技术及计算机硬件技术,关注的科学领域关键词主要是Web服务、产生式规划、传感器网络等,科学研究和技术发明内容之间表现出较强的相关性,此外,专利交易形成的团队与科学研究中同机构的合作者表现出较高的相似性,都包括邓水光、吴健、李莹等人。

四、结语

随着创新型国家建设战略的进一步推进,高校技术成果转移、转化工作日益重要,深入分析高校专利数据,不仅有利于把握高校技术优势、资源优势,也有利于从深层次对我国高校科学研究、技术研发进行宏观调控和指导,促进高校科研与企业、地方经济的有效结合,使高校能更好地为我国经济建设做出贡献。

利用自然语言处理技术与开源知识图谱生成工具相结合的认识方法,能简便、高效地实现对重要发明团队的识别和跟踪。但这一方法的使用,需要注意两个问题:一是发明人、学者身份识别问题。虽然近年来关于研究者身份识别的 ORCID 在不断推进,但近期进行作者、发明人分析时,人名的规范化和同名区分仍然是大规模研究中较为困难的课题;二是工具支持问题。虽然存在不少面向科学研究的可视化知识图谱工具,但如何借助于自然语言处理技术,加深对源数据的深入解析和提取,最大程度上发挥自然语言处理技术和可视化知识图谱工具的长处,还有待进一步研究。

参考文献:

[1] FISCH C O, HASSEL T M, SANDNER P G, et al. University Patenting: A Comparison of 300 Leading Universities Worldwide[J]. *The Journal of Technology Transfer*, 2015, 40:318-345.

[2] VINIG T, LIPS D. Measuring the Performance of University Technology Transfer Using Meta Data Approach: The Case of Dutch Universities[J]. *The Journal of Technology Transfer*, 2015, 40:1034-1049.

[3] 肖沪卫. 专利地图方法与应用[M]. 上海:上海交通大学出版社, 2010:97.

[4] 张敏, 沈雪乐. 国际知识发现研究领域核心作者群成熟度分析[J]. *情报杂志*, 2014, 33(8):111-116, 122.

[5] 李红, 陈少龙. 基于社会网络分析的智能手机专利发明人合作网络演化研究[J]. *科技管理研究*, 2013(20):157-160, 165.

[6] DU Y, YAO C, LI N. Using Heterogeneous Patent Network Features to Rank and Discover Influential Inventors[J]. *Frontiers of Information Technology & Electronic Engineering*, 2015, 16(7):568-578.

[7] COBO M J, L PEZ-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):1609-1630.

[8] SciMAT. SciMAT Version 1.0 User Guide [EB/OL]. [2017-11-22]. <http://sci2s.ugr.es/scimat/download.html>.

[9] 专利数据库[EB/OL]. [2017-12-11]. <https://baike.baidu.com/item/%E4%B8%93%E5%88%A9%E6%95%B0%E6%8D%AE%E5%BA%93>.

[10] L PEZ-HERRERA A G, COBO M J, HERRERA-VIEDMA E, et al. A Bibliometric Study about the Research Based on Hybridating the Fuzzy Logic Field and the Other Computational Intelligent Techniques: A Visual Approach [J]. *International Journal of Hybrid Intelligent Systems*, 2010, 7(7):17-32.