



开放科学
(资源服务)
标识码
(OSID)

基于投融资和专利融合的投资方向发现模型

黄葆春

国家知识产权局专利局自动化部 北京 100088

摘要: [目的/意义] 针对通过对新兴技术领域专利数据以及投融资数据的分析从而发现优先投资领域这一需求, 构建了以融资企业为基础融合专利的数据整合策略, 以及相应的数据分析, 入库体系。[方法/过程] 本文详细阐述了面向投融资数据和专利数据整合加工理论及方法, 包括采取以融资企业为基础的数据抽取方法, 以投资赛道聚合为目标对抽取数据进行清理和加工, 最终形成投融资和专利整合的数据库过程。[结果/结论] 该方法能对投融资和专利两种异构数据进行有效整合, 从而为投融资领域寻找提供决策支持。

关键词: 投融资数据; 专利数据; 文本挖掘; 知识发现

中图分类号: G35

Research of Analysis and Mining Method Based on Integration of Investment Data and Patents

HUANG Baochun

Automation Department Patent Office, China National Intellectual Property Administration, Beijing 100088, China

Abstract: [Objective/Significance] In order to find new technology investment opportunities through the analysis of patent data and investment and financing data, a data integration strategy based on financing enterprises, and related data analysis and storage system are constructed. [Methods/Processes] This paper describes the theories and methods for the integration and processing of investment and financing data and patent data, including data extraction and cleaning and processing procedures based on financing enterprises and investment tracks. [Results/Conclusions] This method can effectively integrate the heterogeneous data of investment and financing and patents, so as to provide decision support for the search in the new opportunity of investment.

Keyword: Investment data; patents data; text mining; knowledge discovery

基金项目 国家重点研发计划“知识产权数据智能分析技术及知识产权服务评价、模式标准体系建设”(2017YFB1401905)。

作者简介 黄葆春(1973-), 工学学士, 研究方向为信息化项目建设管理, E-mail: huangbaochun@cnipa.gov.cn。

引用格式 黄葆春. 基于投融资和专利融合的投资方向发现模型[J]. 情报工程, 2022, 8(2): 109-118.

引言

硬科技概念正在被人们所熟知, 具有较高技术门槛和技术壁垒的硬科技越来越受到资本的关注。硬科技类型企业具有高技术壁垒、高研发投入、高学历的管理层以及低营销投入的特点。硬科技行业将是未来一段时间内的投资热点。

随着科创投资时代的来临, 科学技术对投融资事件的影响越来越大, 所有投资机构都在努力发现具有高新技术的公司和研发团队。对科技资源进行检测, 分析评价科学技术发展的状态和趋势, 选择合理优先投资的科技领域、合理配置投资资金、有效开展科创投资的方法和途径已经成为国内投资机构普遍关注的热点。投融资数据是资本的表现形式, 专利文献是技术创新成果的表现形式。随着科技、资本与产业融合发展, 投融资数据与专利文献之间相互参考与相互作用的价值日益凸显, 两者之间的有效链接除了有助于用户便捷、高效的获取科技文献信息外, 还能加强技术创新成果和资本之间的联系, 有助于实现科创投资的目的。然而, 当前的文献服务体系中, 投融资数据与专利文献资源表现出明显的局部有序但整体无序的孤岛特征, 目前, 研究人员也进行了一些关于海量信息分析及机会发现的研究^[1]。

传统的研究主要是基于情景分析法^[2]、德尔菲法^[3]、AHP(层次分析法)^[4]等方法进行信息分析和结论演绎。上述方法基于非系统过程, 并依靠专家的主观意见。为了克服上述方法的缺陷, 研究者提出了一些系统性的客观的方法, 例如 FUSE(Foresight and Understanding from

Scientific Exposition)^[5], CUBIST(Combining and Uniting Business Intelligence with Semantic Technology)^[6], 用于技术管理的文本挖掘软件(VantagePoint)^[7], KISTI的信息分析和科学技术领域机会发现的 InSciTe^[8,9], 任智军等探索了基于数据挖掘的技术机会发现^[10]等。但国内学者大都局限于将投融资数据或者专利数据进行单独研究, 将结论进行综合分析, 这样以单一视角和不全数据进行的分析得到的结果相对比较片面。

因此, 投融资数据和专利数据的融合分析对于理解资本和技术之间的联系、提高实现科创投资和技术创新方向有着重要的意义, 而且大量的实证研究也表明这两种资源的集成揭示分析有助于理解技术发展趋势、产业资本技术关系等。因此, 研究如何进行投融资数据和专利数据整合, 同时改变现有数据孤岛式的深加工的加工方式, 形成投融资数据和专利数据整合理论和方法是具有重要实际意义的。本文提出一种新的研究方法, 将投融资数据和专利两种异构化的数据进行融合对比, 然后分析与挖掘可以得到更加全面多元化的分析结果, 文章最后使用了最近10年投融资数据和专利数据对本方法进行了实证研究, 结果也相对比较理想。

1 投融资数据与专利数据整合

目前投融资数据与专利数据整合实践研究的很少, 更多的是进行论文和专利的整合, 任智军等研究了论文与专利整合数据的研究, 通过两类数据中都共同存在的人名数据字段将两

类资源进行整合^[11]，赖院根提出了通过分类体系之间的映射关系实现期刊论文与专利文献在领域层面的对接，利用主题词表，建立期刊论文与专利文献在主题层面的链接^[12]。目前尚无对投融资数据与专利数据整合研究，因此本文提出一种以融资企业为基础融合专利的整合策略，从信息资源整合的角度入手，通过企业名称为核心将两类资源进行整合，使得投融资数据与专利数据变得集中有序储存。

首先，基于大规模投融资数据和专利文献数据收集和整理。在数据收集整理阶段要对各种类型（中国专利著录项目及全文数据、DOCDB 专利著录项目、投融资数据、工商数据）数据进行归并整理。

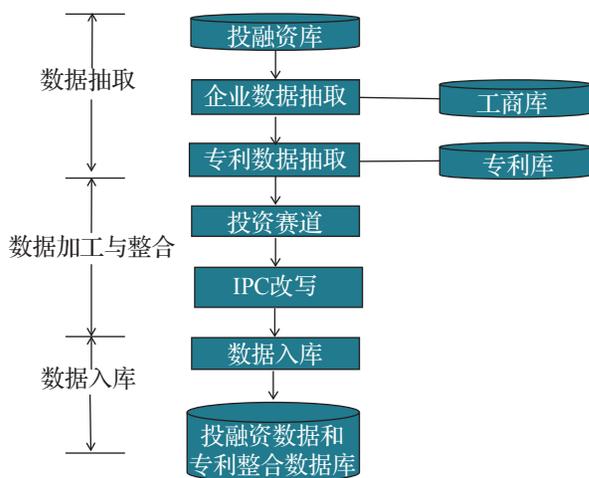


图1 投融资数据与专利数据整合

其次进行数据整合，数据整合分成三个部分：

(1) 数据抽取：先将投融资数据中的被投资企业从简称得到企业的全称，然后将企业全称获取相关的工商数据，同时根据企业全称利用申请人获取专利数据；

(2) 数据清理与加工：基于投融资数据自底向上构建投资赛道，把所有在融企业标引到投资赛道中，同时采用前期研究积累中已有的针对专利文献的IPC抽取工具对IPC进行提取，并将IPC数据进行中文标引和改写。

(3) 数据导入：将加工好的数据导入到投融资和专利整合数据库。

最后，本文选取最近10年投融资数据和专利数据进行实证研究，进行投融资数据和专利的整合，从信息资源整合的角度入手，通过挖掘其理论深度，把无序分散的资源集中起来，把无序的资源变为有序，使之有效实现投融资数据和专利数据知识化组织、方便后继的分析与挖掘。

2 投资方向识别及标注

2.1 投资方向识别概述

在投融资数据和专利数据整合基础上，对新融资企业进行标注和度量，然后利用分类算法对新融资企业进行投资赛道分类，对投融资数据和专利数据融合程度进行分析，研究投融资和专利两种数据之间存在的相互影响和排斥关系，从而探索资本和技术之间的发展规律，确定科创投资方向。基于投融资数据和专利数据挖掘科创投资方向的首要问题是识别在融企业所在投资赛道，本文采用的方法则是利用在融企业特征做分类，在分类的基础上，进一步进行企业聚类并标注聚类名称，企业聚类的结果就是科创投资方向，具体如图2所示。

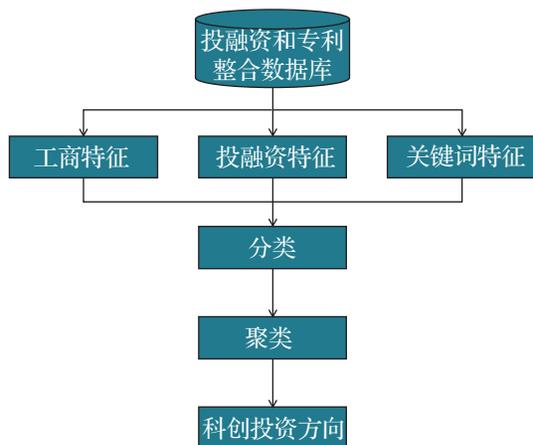


图2 科创投数据和专利数据投资方向识别模型图

分类和聚类的基础是企业技术和产品特征，企业技术特征是全面刻画企业技术信息的系统性方法，基于工商、投融资和专利等多维数据可以更加全面了解企业的真实情况，也为后续科创投资方向识别算法提供有力支持，通过多样化的产品标签，更加清晰地标识了企业技术和产品的多种属性等。其企业技术特征如下表1所示：

表1 企业技术特征

特征编号	特征名称	特征内容说明
1	工商特征	经营范围中技术关键词
2	融资特征	投融资新闻中技术关键词
3	专利特征	专利IPC、专利关键词

2.2 投资方向识别

有了企业技术特征我们就可以使用分类算法进行产业的分类，分类算法使用 fasttext^[13]，在分类之后，我们可以对产业下的赛道进行聚类，聚类赛道作为投资方向，基于聚类的投资方向识别模型说明如下：

(1) 从投融资数据和专利整合数据库中抽

取投融资特征和专利关键词。

(2) 利用投融资指标和关键词作为特征进行分类。

(3) 在每个分类里面，采用 KMeans 聚类算法^[14] (sklearn kmeans^[15]) 进行融资企业聚类。

(4) 根据聚类结果将每个类别作为投资方向。

2.3 投资方向标注

在得到研究方向后，需要对研究方向进行标注，研究方向的标注主要通过科技术语。互信息是表示两个变量之间关联程度有用的度量，因此可以通过关键词与研究方向的互信息来进行研究方向的标注。在对共同研究方向投融资数据和专利聚类后，具有相同研究方向的投融资数据和专利聚为一类，这些投融资数据和专利在关键词上具有共性，提取这个共性词汇就是对研究方向的标注。由于与同一个研究方向上互信息最大的关键词不只有一个，因此共同研究方向标注有可能由一组词汇组成。基于互信息的研究方向描述的算法如下：

(1) 统计聚类结果中总的文献数 n 与各类别文献数 $n_i, i=1,2,\dots,k$ (k 为聚类数目)。

(2) 根据聚类结果，利用互信息公式计算每个关键词与所有类别的互信息，MI 公式如下所示

$$MI(t_k, c_i) = P(t_k, c_i) \log \frac{P(t_k, c_i)}{P(t_k)P(c_i)} \quad (1)$$

(3) 对 MI 进行规划化处理。

(4) 分类别统计最大互信息值及对应的主关键词。

3 投资趋势走向分析模型

投资趋势走向分析是根据最新的投资方向

和技术发展趋势进行分析和挖掘的结果，它们可能创立一个新投资赛道或改变某个老投资赛道。投资趋势走向代表了一个新的投资赛道或者投资方向，被用来找到未来科创投资的技术领域和投资趋势。投资趋势走向分析对于投资方向研究意义重大：一方面，投融资与专利包括了资本和技术，在最新被投资企业所在的新技术赛道热点，最有可能涌现新兴投资趋势。另外，新投资方向也会推动资本投资趋势和技术发展，吸引投资机构和商业公司转向研发，从而提高投资机构投资走向与专利的产出和提高技术创新水平有着重要的意义。由于资本是技术创新的催化剂，因此在投资趋势走向是非常必要的，本文研究的投资趋势走向分析是指在研究最近刚刚出现有新的投资方向或者由冷变热的投资方向，投资趋势走向分析模型采用的是基于聚类的离群点检测算法^[16]。基于聚类的离群点检测算法如下所示：

(1) 计算现有投资分类或者聚类的质心；

(2) 计算被投资企业到最近质心的距离；

(3) 计算各对象到它的最近质心的相对距离；

(4) 与给定的阈值作比较，超过阈值即认为是离群点；

(5) 对离群点企业进行聚类，找到新的投资方向。

4 实验及分析

为验证基于投融资数据与专利数据整合数据的投资方向发现模型的可行性，本文选择了最近10年投融资数据和专利数据进行实例研究，在互联网上采集投融资数据和专利局专利下载获取了数据，利用投融资数据与专利数据整合之后构建了投融资专利整合数据库，然后对投资赛道进行投资方向分析和新兴投资方向分析。

4.1 投融资与专利数据整合

投融资数据与专利整合数据中抽取10家融资企业的投融资和专利整合数据见表2。

表2 投融资和专利整合

序号	企业简称	企业全称	公司简介	专利数量	专利IPC	专利关键词
1	蚂蚁集团	蚂蚁科技集团股份有限公司	蚂蚁集团是中国最大的移动支付平台支付宝的母公司，也是全球领先的金融科技开放平台，致力于以科技推动包括金融服务业在内的全球现代服务业的数字化升级，携手合作伙伴为消费者和小微企业提供普惠、绿色、可持续的服务，为世界带来微小而美好的改变	3883 (支付宝(杭州)信息技术有限公司)	B60, 10-, B61, G06, G03, G10	区块链、电子设备、交易、合约、智能、网络、风险、图像、数据处理
2	快手科技	北京快手科技有限公司	北京快手科技有限公司是国内知名互联网公司。公司围绕着“发现真实有趣的世界”理念，打造了全球最大的短视频分享社区——快手。与其他社交类软件公司不同的是，快手是一家技术驱动型公司，倡导工程师文化，希望通过技术的力量使人与人之间的联系更加紧密。在这里，用户可免费使用快手平台上传7-57秒的视频，记录和分享生活点滴，增进人与人之间的联系和了解。公司致力于为用户打造“简单，好用”的使用体验，提供“真实，有趣”的短视频展现。这是一个充满朝气、求实坦诚的公司，梦想通过短视频，给普通人生活带来更多的欢乐	6 (广州弹幕网络科技有限公司)	H04, 11-, G03	弹幕、信息、屏幕、影院、终端、输入界面、服务器资源、用户数量、观众、观影

(续表 2)

3	北京京邦 达贸易有 限公司	<p>随着无界零售时代的到来,京东物流作为无界物流的引领者和实践者,以降低社会物流成本为使命,致力于成为全球供应链基础设施服务商。将基于短链供应,打造高效、精准、敏捷的物流服务;通过技术创新,实现全面智能化的物流体系;与合作伙伴、行业、社会协同发展,构建共生物流生态。通过智能化布局的仓配物流网络,京东物流为商家提供仓储、运输、配送、客服、售后的正逆向一体化供应链解决方案、快递、快运、大件、冷链、跨境、客服、售后等全方位的物流产品和服务以及物流云、物流科技、物流数据、云仓等物流科技产品。目前,京东物流是全球唯一拥有中小件、大件、冷链、B2B、跨境和众包(通达)六大物流网络的企业</p>	50	G06, G01, B60, G05, A47, B64, B65, 09- B07, B23, G07, H04	存储介质、仓 库、订单、准 确性、图像、 物品、仓储物 流、数量、货物、 发货
4	广州橙行 智动汽车 科技有限 公司	<p>小鹏互联网汽车团队成立于 2014 年中,于 2015 年 1 月正式注册,是一家创新型的互联网汽车科技公司。公司的主要目标是在物联网、大数据和 O2O 的大环境下,研发下一代智能化电动汽车。使用新的材料、制造工艺和营销模式,大胆创新、勇于探索,创造出全新的智能出行工具,用心做以产品体验为核心的互联网汽车。公司由国内顶尖高校毕业,并在知名汽车、互联网企业拥有多年研发、生产、经营经验的工程师创立。是一个扎实开展技术创新为核心的工程师团队</p>	226	B60, G06, H04, G01, H01, H02, G05, G10, G07, E05, H05, B62, F16, B22, G08, B24, B64, E04, G03	装置、车辆、 存储介质、电 动汽车、电池、 数据处理、车 载、驾驶员、 电池包、动力 电池
5	菜鸟网络 科技有限 公司	<p>菜鸟网络科技有限公司成立于 2013 年 5 月 28 日,由阿里巴巴集团、银泰集团联合复星集团、富春控股、三通一达(申通、圆通、中通、韵达)等共同组建。菜鸟是一家互联网科技公司,专注于物流网络的平台服务。通过大数据、智能技术和高效协同,菜鸟与合作伙伴一起搭建全球性物流网络,提高物流效率,加快商家库存周转,降低社会物流成本,提升消费者的物流体验。菜鸟的使命是与物流合作伙伴一道,致力于实现中国范围内 24 小时送货必达、全球范围内 72 小时送货必达。数据显示,2017 年中国已经进入日均包裹 1 亿个的超级繁忙时代。物流行业引入智能、开放的互联网协同模式,而非传统自建模式,才能更好地适应未来的物流需要。菜鸟的商业逻辑是搭建平台,让物流供应链条上不同服务商、商家和消费者可以实现高效连接,从而提升物流效率和服务品质,降低物流成本。通过菜鸟与合作伙伴的努力,全球智慧物流网络已经覆盖 224 个国家和地区,并且深入到了中国 2900 多个区县,其中 1000 多个区县的消费者可以体验到当日达和次日达的极致配送。以历年天猫双 11 为例,菜鸟网络成立以来,通过智慧物流的提升,虽然单日物流订单量从 1.52 亿攀升到 8.12 亿,但是配送 1 亿个包裹的时间却从 9 天下降到了 2.8 天,创造了世界物流业的奇迹。2022 年前,阿里巴巴和菜鸟网络还将投入 1000 亿元升级全球智慧物流网络,加快实现“全国 24 小时、全球 72 小时必达”</p>	3(浙江 菜鸟供应 链管理有 限公司)	G06	信号收发、天 线、射频、物 流、电子标签、 电磁波、胶层、 货物、交换平 台、仓储管理

(续表 2)

6	Lu.com, 陆金所	上海陆家嘴国际金融资产交易市场股份有限公司	<p>陆金所, 全称上海陆家嘴国际金融资产交易市场股份有限公司, 是全球领先的互联网财富管理平台, 平安集团旗下成员, 2011 年 9 月在上海注册成立, 注册资本金 8.37 亿元, 位于国际金融中心上海陆家嘴。陆金所致力于结合金融全球化发展与信息技术创新, 以健全的风险管控体系为基础, 为广大机构、企业与合格投资者等提供专业、高效、安全的综合性金融资产交易信息及咨询相关服务。陆金所旗下 lu.com 网络投融资平台 (www.lu.com, 原域名 www.lufax.com) 2012 年 3 月正式上线运营。作为中国平安集团倾力打造的平台, lu.com 结合全球金融发展与互联网技术创新, 在健全的风险管控体系基础上, 为中小企业及个人客户提供专业、可信赖的投融资服务, 帮助他们实现便捷高效的低成本融资和财富增值。除了通过陆金所平台交易的资产规模保持高速增长以外, 陆金所还推出了业内独树一帜的投资者适当性管理体系—KYC 2.0 系统。该体系主要包括“投资者评估 (Know Your Customer, KYC)、产品风险评估 (Know Your Product, KYP), 投资者与产品风险的适配、信息披露、投资者教育”五方面内容, 最大的特点就是更多地利用了大数据技术、机器学习等在资金端对投资者进行“精准画像”, 并提供智能推荐服务, 能实现投资者风险承受能力与产品风险的精准匹配, “将合适的产品卖给合适的人”。</p>	57	G06, H04, 14-,	<p>存储介质、标识、终端发送、计算机可读、产品标识、用户标识、镜像、数据安全、服务器发送</p>
7	中航锂电	中航锂电科技有限公司	<p>中航锂电是专业从事锂离子动力电池、电池管理系统、储能电池及相关集成产品和锂电池材料的研制、生产、销售和市场应用开发的高科技企业, 致力于为全球客户提供完整的产品解决方案和完善的全生命周期服务。公司成立于 2007 年, 产品先后通过 IATF16949 等权威质量体系认证, 获得 CE、UL、TUV、RoHS 等国际认证。公司现已设立常州、洛阳、厦门、成都、武汉五大产业基地; 同时, 作为国家认定企业技术中心, 公司设立专业的独立研发机构——中航锂电研究院, 致力于动力电池关键技术的研发与应用, 打造具有全球影响力的锂电科技创新平台。十四五期间, 公司规划产能 300GWh+, 并根据市场及客户需求实时扩产, 全力打造全球领先的动力电池制造商和全球优秀企业最信赖的合作伙伴! 中航锂电以“持续创新, 造福人类”为使命, 坚持创新发展战略, 在材料体系创新、系统结构创新、智能制造创新和生态发展创新上不断发力, 持续打造产品力领先, 努力创建伟大公司, 成就伟大事业! 为全球新能源产业的健康快速发展, 为中国“碳达峰、碳中和”目标的实现, 为创造世界更加美好的未来不懈奋斗!</p>	5	<p>H01, G01, B08, B23, B05, B65, H05, 13-, A62, B01, B21, B32, B62, C01, G08,</p>	<p>电池技术、极片、电芯、电池模组、注液孔、底板、极柱、浆料、盖板组件、锂电池</p>

(续表 2)

8	京东数科 京东科技控股股份有限公司	<p>京东数字科技集团创立于京东集团内部，前身为京东金融，于 2013 年 10 月开始独立运营。公司以大数据、人工智能、物联网、区块链等时代前沿技术为基础，建立起核心的数字化风险管理能力、用户运营能力、产业理解能力和 B2B2C 模式的企业服务能力。公司经营宗旨是从数据中来，到实体中去，通过数字化手段连接金融与实体产业，助力金融与实体产业实现互联网化、数字化和智能化，进而降低成本、提高效率、提升用户体验和模式升级，在实体经济不断数字化的过程中创造公平与普惠的社会价值。截至目前，公司完成了在数字金融、智能城市、数字农牧、数字营销、数字校园等领域的布局，旗下品牌包括京东金融、京东城市、京东农牧、京东钜媒、京东少东家等，在客户群体上实现了个人端、企业端、政府端的三端合一。2018 年，公司完成 B 轮融资，估值超过 1300 亿人民币。</p>	69	G06, H04, G08	存储介质、电子设备、数据、介质、模型、数据库、人工、图像、数据处理
9	威马汽车 威马智慧出行科技(上海)股份有限公司	<p>威马汽车是一家智能电动汽车研发商，专注于从事新能源汽车的研发与设计，基于 Teke 核心架构将会延伸出 STD 和 PL 两大整车平台，未来将推出至少 8 款新车。同时还会提供车联网及售后等配套服务。</p>	36	B65, H05, H01, B62, G01, H02, E05, F21, F16, A61, B60	车辆、支架、汽车、线束、外板、稳定性、车身、底板、侧板
10	紫光展锐 紫光展锐(上海)科技有限公司	<p>作为紫光集团旗下核心企业，紫光展锐是我国集成电路设计业的标杆企业，也是整个集成电路产业的领军企业，其致力于移动通信和 AIoT 领域核心芯片的自主研发及设计，产品涵盖 2G/3G/4G/5G 移动通信基带芯片、AIoT 芯片、射频前端芯片、无线连接芯片、电视芯片，是全球全面掌握 2G/3G/4G/5G 移动通信技术以及 IoT 等全场景通信技术的少数企业之一。目前，紫光展锐拥有近 4500 名员工，其中 90% 以上是研发人员，在全球拥有 15 个技术研发中心及 7 个客户支持中心。已发展成为全球第三大面向公开市场的手机芯片设计企业、全球领先的 5G 通信芯片企业和中国最大的泛连接芯片供应商之一。</p>	105 (紫光展锐(重庆)科技有限公司)	G01, H03, H04, G10, G06	存储介质、信号、信息、数据、电子设备、终端设备、语音、时域、芯片

4.2 投资方向识别与标注

利用标注算法投资方向识别及标注方法对 2020 年投融资数据与专利整合数据进行识别之后，得到融资企业 6043 家，并通过第 2.3 节中公式 (1) 对化学药的类别进行了标注，结果如表 3 所示。从表 3 可以看出，投资方向为“色谱柱、药品包装，化学药药理学理论，崩解剂，化学药

制备设备”等 10 个方向。

4.3 新兴投资方向分析

根据第 3.3 节中新兴技术术语的识别方法，2020 年投融资数据对 2020 年新兴投资方向识别，识别出植物肉、射频滤波器、液态储氢、内窥镜和手术机器人等 5 个新兴投资方向，具体如表 4 所示。

表 3 投资方向标注

类别编号	互信息	标注结果	类别编号	互信息	标注结果
1	色谱柱, 0.0007; 溶液: 0.0006; 收率, 0.0006	色谱柱	6	编码, 0.0018; 抗体, 0.0017; 宿主细胞, 0.0016	基因编码
2	名称, 0.0036; 照片, 0.0035; 图案, 0.0031, 包装, 0.0021;	药品包装	7	疫苗, 0.0013; 序列, 0.0009; 样本, 0.0008	疫苗
3	化合物, 0.0046; 药学, 0.0043; 疾病: 0.0035	化学药药理学理论	8	供试品溶液, 0.0012; 高效液相, 0.001; 对照品溶液, 0.001	供试品溶液
4	出度, 0.001; 室温, 0.0009; 崩解剂, 0.0007	崩解剂	9	抗体, 0.0019; 重链, 0.0013; 肿瘤细胞, 0.0012	肿瘤抗体
5	制药, 0.0013; 化学药, 0.0011; 设备, 0.009	化学药制备设备	10	中药材, 0.0008; 功效, 0.0008; 冰片, 0.0008	中药材

表 4 新兴投资方向识别表

序号	投资方向	相关企业	技术关键词	关键 IPC
1	植物肉	植物肉(杭州)健康科技有限公司; 植物肉(杭州)健康科技有限公司; 深圳市星期零食品科技有限公司	A23J3, A23L	抽提物, 植物蛋白, 半胱氨酸, 丝氨酸, 人造肉, 大豆, 植物
2	射频滤波器	珠海晶讯聚震科技有限公司; 开元通信技术(厦门)有限公司; 杭州左蓝微电子技术有限公司; 锐石创芯(深圳)科技有限公司; 北京超材信息科技有限公司	H03H9, H03H3, H03F3, H04B1	谐振器, 射频, 射频前端, 滤波器, 膜体声波, 压电薄膜, 功率放大器, 基板
3	液态储氢	深圳国氢新能源科技有限公司; 上海重塑能源集团股份有限公司; 北京中科富海低温科技有限公司; 国家电投集团氢能科技发展有限公司	F17C, F25J1/02, F17D3/01	氢气, 储氢装置, 低温储罐, 阀体, 液氢, 气瓶, 管道, 管路, 制冷机, 转换器, 液氢储罐, 压缩机
4	内窥镜	沈阳沈大内窥镜有限公司; 深圳市资福医疗技术有限公司; 湖南省华芯医疗器械有限公司; 新光维医疗科技(苏州)股份有限公司; 苏州阿酷育医疗科技有限公司	A61B1, A61B17, A61B5/07, A61B19/00	内窥镜, 胶囊内窥镜, 磁控, 组件, 装置, 导管, 手柄, 磁控
5	手术机器人	深圳市精锋医疗科技有限公司; 江苏一影医疗设备有限公司; 元化智能科技(深圳)有限公司; 杭州健嘉机器人有限公司; 杭州三坛医疗科技有限公司	A61B	手术, 机器人, 操作臂, 手术器械, 连接组件, 动力机构, 机械臂, 关节组件, 末端执行器, 驱动装置, 连杆

5 总结与展望

本文提出了一种基于投融资数据与专利数据整合的分析与挖掘的研究方法, 构建了基于投融资数据与专利数据整合的分析与挖掘知识挖掘模型, 从投融资与专利共同研究方向和新兴技术发现两个方面来揭示投融资与专利关系, 最后, 我们对最近 10 年投融资数据和专利数据

进行了实例验证, 加强技术创新成果和资本之间的联系, 为科创投资等提供了一个新的研究思路。下一步的工作将在本文提出的模型基础上, 进行基于投融资和专利整合的企业融资预测研究。

参考文献

[1] Richard, S. Patent Trends as a technological

- Forecasting Tool[J]. World Patent Information, 1983, 5(3): 137-143.
- [2] Wright G, Goodwin P. Decision making and planning under low levels of predictability: Enhancing the scenario method[J]. International Journal of Forecasting, 2009, 25(4): 813-825.
- [3] Okoli C, Pawlowski S. The Delphi method as a research tool: an example, design considerations and applications[J]. Information and Management, 2001, 42(1):15-29.
- [4] Liu G, Jin Y, Li F. The application of AHP method in well control risk evaluation by controllable factor analysis[J]. Journal of Southwest Petroleum University, 33(2):137-141.
- [5] IAPRA. Foresight and understanding from scientific exposition[EB/OL]. [2010-11-06]. <https://www.iarpa.gov/research-programs/fuse>.
- [6] CUBIST. Project description[EB/OL]. [2019-08-01]. <https://cordis.europa.eu/project/id/257403>.
- [7] Search Technology, Inc. Introduction to VantagePoint. <http://thevantagepoint.com>.
- [8] Kim J, Lee S, Lee J, et al. Design of TOD Model for Information Analysis and Future Prediction. Communications in Computer and Information Science, 264(1):301-305.
- [9] Kim J Y, Hwang M G, Jeong D H, et al. Technology trends analysis and forecasting application based on decision tree and statistical feature analysis [J]. Expert Systems with Applications, 2012:39.
- [10] 任智军, 乔晓东, 徐硕, 等. 基于数据挖掘的技术机会发现模型研究 [J]. 情报杂志, 2015(6): 174-177.
- [11] 任智军, 詹淑琳, 范婷婷. 基于论文与专利整合数据的研究方向发现模型研究 [J]. 情报工程, 2015, 1(4):88-94.
- [12] 赖院根, 曾建勋. 期刊论文与专利文献的整合框架研究 [J]. 图书情报工作, 2010, 54(4):109-112
- [13] Joulin A, Grave E, Bojanowski P, et al. FastText.zip: Compressing text classification models[J]. 2016.
- [14] 周丽娟, 王慧, 王文伯, 等. 面向海量数据的并行 KMeans 算法 [J]. 华中科技大学学报(自然科学版): 2012(S1):150-152.
- [15] 黄永昌. scikit-learn 机器学习: 常用算法原理及编程实战 [M]. 北京: 机械工业出版社, 2018.
- [16] 胡婷婷. 数据挖掘中的离群点检测算法研究 [D]. 厦门: 厦门大学, 2014.