



开放科学
(资源服务)
标识码
(OSID)

基于语义聚类的行业及产业技术研发方向分析

台一鸿¹ 高宇强² 张娇²

1. 国家知识产权局专利局专利审查协作江苏中心 江苏 215163;
2. 汉唐信通(北京)咨询股份有限公司 北京 100012

摘要: 基于专利 IPC 与国民经济分类映射技术, 可以获取行业专利数据。在此基础上, 本文研究了基于语义聚类的行业及产业技术研发方向分析, 定量进行了文化及相关产业的行业融合分析, 在融合方向上寻找创新发展方向。最后, 我们对文化及相关产业进行了实例验证, 实验效果表明基于文化及相关产业内部融合可以找到重点融合数据, 在融合数据中使用语义聚类 and MI 算法可以发现文化及相关产业的技术创新方向。

关键词: 文化产业; 专利分析; 语义聚类; 研发方向

中图分类号: H087

Analysis of Industry and Industry Technology R&D Direction Based on Semantic Clustering

TAI Yihong¹ GAO Yuqiang² ZHANG Jiao²

1. Patent Examination Cooperation Jiangsu Center of the Patent Office, CNIPA, Jiangsu 215163, China;
2. Sino-tong (Beijing) Consulting Co., Ltd, Beijing 100012, China

Abstract: Based on the mapping of patent IPC and national economic classification, industry patent data can be obtained. On this basis, this paper conducted the industry and industry technology research and development direction analysis based on semantic clustering, quantitatively analyzed culture and its related industries to seek innovative development in the direction of fusion.

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

作者简介: 台一鸿(1982-), 硕士, 研究方向: 专利审查、专利分析, E-mail: tlm666@126.com; 高宇强(1982-), 本科, 高级工程师, 研究方向: 专利数据分析、自然语言处理; 张娇(1993-), 研究生, 助理工程师, 研究方向: 文本挖掘、社会网络分析、科技情报分析。

The experimental results showed that the key integration data can be found based on the internal integration of culture and related industries. The semantic clustering and MI algorithm can be used in the fusion data to discover the technological innovation direction of culture and related industries.

Keywords: Cultural industry; patent analysis; semantic clustering; research and development direction

引言

以数据驱动把握行业创新态势,充分发挥科技创新对行业发展的重要引擎作用。通过对相关技术领域的专利数据进行统计和分析,可以定量衡量行业的创新状况,寻找发展中存在的问题与薄弱环节^[1]。通过对行业中的专利数据进行分析,能进一步明晰行业的需求与发展方向,发挥专利数据对于行业创新的监控和推动作用,从而有的放矢,促使行业创新获得成功。从宏观上明确哪些技术能够直接或者间接推动行业的发展,可以引导创新主体有效运用行业专利状况信息及技术创新信息,推动创新成果的转移转化,引导民间资本投融资等。从而为制定相关制度促进文化产业的科技创新、优化文化产业结构等提供数据支撑和决策依据。

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^[2]。传统的研究主要是基于情景分析法^[3]、德尔菲法^[4]、AHP(层次分析法)^[5]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)^[6], CUBIST (Combining and

Uniting Business Intelligence with Semantic Technology)^[7],用于技术管理的文本挖掘软件(VantagePoint)^[8],KISTI的信息分析和科学技术领域机会发现的InSciTe^[9,10],任智军等^[11]探索了基于数据挖掘的技术机会发现等。但是在行业中找到新技术发展方向的研究很少,因此以专利数据为驱动研究行业的技术研发方向具有实际意义。找到技术创新可能发生点进行分析,发现和识别技术的研发方向具有推动行业变革和产业升级的重要意义。因此本文第二部分主要研究了以行业融合分析,试图在行业融合点上找到可能的融合专利,第三部分利用语义聚类的方法找到技术研发方向。最后以文化及相关产业为例进行了实证研究。

1 研究行业融合分析

1.1 行业产业专利数据获取

以专利数据为基础的专利分析反映了技术创新的程度,基于国民经济分类和IPC的映射关系可以得到行业和产业的专利数据,国家知识产权局制定了《国际专利分类与国民经济行业分类参照关系表(2018)》(以下简称《2018年参照关系表》)^[12]。本参照关系表建立了专

利与国民经济行业的映射关系，为专利的行业分类提供直接对照，有助于从产业角度出发结合科技、经济数据开展相关统计分析。由映射

关系我们可以得到国民经济行业分类（相关门类）与国际专利分类（8个部）的对应关系，如图1所示：

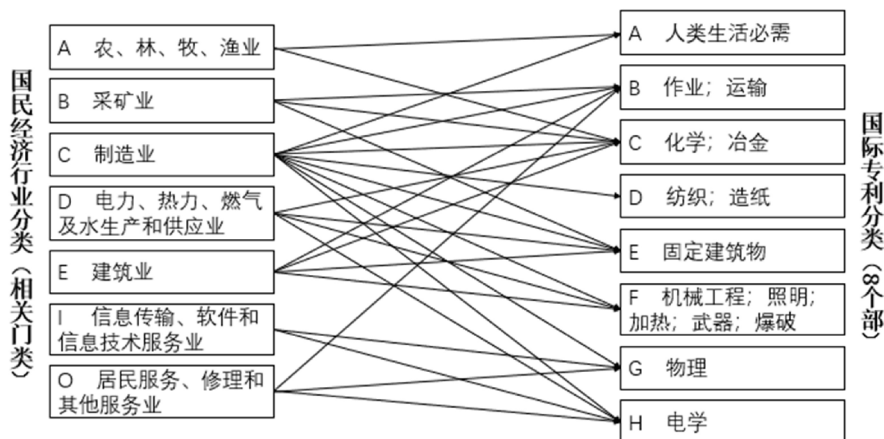


图1 国民经济行业分类（相关门类）与国际专利分类（8个部）的对应关系

由以上进行映射关系，检索专利数据库，可以得到行业或者产业的专利数据。

1.2 产业融合度计算

产业融合是指不同产业或同一产业内的不同行业相互渗透、相互交叉，最终融为一体，逐步形成新产业的动态发展过程^[13]。产业融合通常涉及两个或两个以上行业之间的交叉融合、互相渗透的演进过程。这种两个或两个以上行业的交叉融合程度或知识整合程度，称之为产业融合度，简称融合度。本文的产业融合度是对专利进行测度，专利经归属到行业类别后，可以做出行业的共现矩阵。对于一个行业的共现矩阵，本文要测度的目标是，在此矩阵中，非对角线上的数值越接近于对角线上的数值时，该领域的行业融合度越大，在设定行业融合度测度公式时，必须符合这样的测度目标。构建产业之间的相关度矩阵，通常用于测度共现矩

阵距离的，主要有皮尔逊相关系数^[15]、欧氏距离^[16]和夹角余弦函数^[17](Cosine Similarity Measure)等。经过细致认真比较它们的优缺点，本文认为夹角余弦函数（简称余弦函数或 Cosine 值）很符合产业融合度的测度目标。本文利用 Cosine 值来测度行业融合度，测度公式是相似的，Cosine 值的公式为^[18]。

$$\cos(XY) = \frac{\sum_i x_i y_i}{\sqrt{(\sum_i x_i^2) \cdot (\sum_i y_i^2)}} \quad (1)$$

2 技术研发方向发现

(1) 在产业融合的专利数据上，首先基于 BERT 得到专利的特征，共有 768 维度的特征。BERT 模型用 Transformer 的双向编码器表示。与最近的其他语言表示模型不同，BERT 旨在通过联合调节所有层中的上下文来预先训练深度双向表示。因此，预训练的 BERT 表示可以

通过一个额外的输出层进行微调,适用于广泛任务的最先进模型的构建,比如问答任务和语言推理,无需针对具体任务做大幅架构修改。针对不同的任务,模型能够在一个 token 序列中明确地表示单个文本句子或一对文本句子

(比如[问题,答案])。对于每一个 token,其输入表示通过其对应的 token embedding, 段表征(segment embedding)和位置表征(position embedding)相加产生。图2是输入表示的直观表示^[19]:

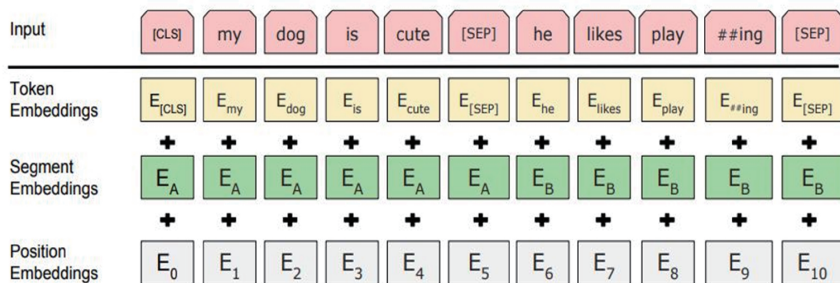


图2 BERT 输入表示。输入嵌入是 token embeddings, segmentation embeddings 和 position embeddings 之和

然后基于 768 维度特征计算专利之间的文本相似度,在文本相似度基础上进行聚类。聚合完成之后,把聚类结果作为技术研发方向。

在聚类文献上利用互信息的方法识别方向上的核心关键词,用关键词定义融合发展的新技术研发方向。

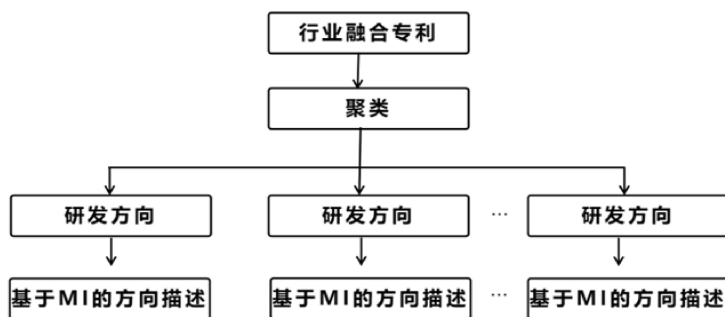


图3 研究方向示意图

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。

(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^[14]

$$MI(t_k, c_i) = P(t_k, c_i) \log \frac{P(t_k, c_i)}{P(t_k)P(c_i)} \quad (2)$$

(3) 分类别统计最大互信息值及对应的主关键词。

3 案例分析

3.1 文化及相关产业概述

新时代下“文化+”成为文化产业发展的

重要议题和创新思路。十九大报告指出“健全现代文化产业体系 and 市场体系，创新生产经营机制，完善文化经济政策，培育新型文化业态。”如何能够实现文化产业与相关产业的跨界融合、创新突破，一直以来都是文化产业领域的主要课题。在经济发展新常态与新一轮科技革命交汇的大背景下，新技术、新业态不断涌现，创新成为文化产业发展的基本动力，文化与大数据的结合推动了文化产业与众多产业的跨界融合，使文化产业呈现繁荣的发展态势。互联网时代下，数字网络技术也给世界创造了新的巨大资源。人工智能赋能文化产业创新发展初现端倪。截至 2018 年底，北京已有规模以上文化产业法人单位 4000 余家，文化产业产值达到 10703 亿元，年增长率达到 11.9%。近年来，人工智能带动产业发展的头雁效应已经在文化领域初现端倪。如电影《流浪地球》的制作中使用了大量的人工智能技术，展现出前所未有的视听效果，令观众叹为观止。人工智能技术正从根本上改变文化内容生产、分发和变现过程，提供了全新的文化展陈模式，彰显出人工智能

技术赋能文化传播的“无边界想象”。

文化科技创新融合背景下，知识产权战略作用日益凸显。文化产业是以创新为核心，以知识产权运作为手段，向大众提供文化、艺术、娱乐等精神产品的新兴产业集群。当代文化产业在全球化的消费社会背景中迅猛发展起来，其中“创新”或“创意”是文化产业的核心动力。知识产权是文化企业获取经营收益的核心资源，文化产业知识产权保护问题日益突出。文化产业的创新水平直接关系到文化产业的整体实力和竞争力。而专利是科技创新成果的典型代表，而且是内容最明确、边界最清晰、表现形式最规范的科技成果。因此如何判断文化产业创新趋势，以及如何推动新的科学技术对于文化发展的促进作用，需要对文化产业的创新态势进行统计分析，从而为进一步完善产业政策提供决策依据。本文在梳理文化及相关产业分类与国际专利分类关系的基础上，基于《文化及相关产业分类（2018）》^[20]与国际专利分类对照表，《文化及相关产业分类（2018）》进行数据收集和整理，文化及相关产业与国际专利分类对照表示例表 1 所示。

表 1 文化及相关产业与国际专利分类对照表示例

文化及相关产业分类	GB/T 4754-2017代码	IPC分类
文化相关领域		
02 内容创作生产		
026 工艺美术品制造		
0261 雕塑工艺品制造	2431	B44B、B44C
0262 金属工艺品制造	2432	A44C3、A47G1、B44B、B44C
0263 漆器工艺品制造	2433	A47G1、B44B、B44C
0264 花画工艺品制造	2434	A41G1、B44B、B44C、B44D、B44F
0265 天然植物纤维编织工艺品制造	2435	B27J1、B44B、B44C
0266 抽纱刺绣工艺品制造	2436	B44B、B44C、D04C1、D05C17、D06P、D06Q
0267 地毯、挂毯制造	2437	A47G27、D03D27、D04G3、D05C17
0268 珠宝首饰及有关物品制造	2438	A44C1、A44C11、A44C13.....
0269 其他工艺美术及礼仪用品制造	2439	A41G11、A41G3、A41G5.....

3.2 数据获取与数据预处理

为了验证基于专利挖掘的文化及相关产业技术研发方向发现模型的可行性,本文选择了文化产业及相关产业 2013-2018 年申请的发明和实用新型专利进行实例研究。在专利检索与分析系统-专眼查中获取数据,共 1854690 篇。并在其数据基础上,构建文化及相关产业第二

层产业中的文具制造及销售和印刷设备制造两个产业融合的专利数据库,共 26357 件。

3.3 基于知识挖掘的文化及相关产业融合分析

使用 $\cos(XY)$ 算法,计算得到 2013-2018 年文化及相关产业各第二层产业之间融合度,其结果如表 3 所示。

表 2 2013-2018 年文化及相关产业相同专利在不同第二层产业之间申请情况

	乐器制造及销售	互联网信息服务	互联网文化娱乐平台	信息服务终端制造及销售	印刷复制服务	印刷设备制造	工艺美术品制造	广播电视电影设备制造及销售	广播电视节目传输	摄录设备制造及销售	数字内容服务	文具制造及销售	文化辅助用品制造	游乐游艺设备制造	演艺设备制造及销售	玩具制造	笔墨制造	艺术陶瓷制造	节庆用品制造
乐器制造及销售	/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
互联网信息服务	0	/	0	0	0	0	0	0	0	0	58461	0	0	0	0	0	0	0	0
互联网文化娱乐平台	0	0	/	378433	72320	2089	0	0	0	0	0	0	0	0	0	0	0	0	0
信息服务终端制造及销售	0	0	378433	/	0	0	0	0	0	61188	0	0	0	0	17923	0	0	0	0
印刷复制服务	0	0	72320	0	/	11707	0	0	0	0	0	4641	0	0	0	0	0	0	0
印刷设备制造	0	0	2089	0	11707	/	0	18932	0	18953	0	26357	18932	0	0	0	0	0	0
工艺美术品制造	0	0	0	0	0	0	/	0	0	0	0	0	0	359	539	2004	0	12925	539
广播电视电影设备制造及销售	0	0	0	0	0	18932	0	/	1121	27384	0	0	18932	0	0	0	0	0	0
广播电视节目传输	0	0	0	0	0	0	0	1121	/	0	0	0	0	0	0	0	0	0	0
摄录设备制造及销售	0	0	0	61188	0	18953	0	27384	0	/	0	0	18932	0	0	0	0	0	0
数字内容服务	0	58461	0	0	0	0	0	0	0	0	/	0	0	0	0	0	0	0	0
文具制造及销售	0	0	0	0	4641	26357	0	0	0	0	0	/	0	0	0	0	0	0	0
文化辅助用品制造	0	0	0	0	0	18932	0	18932	0	18932	0	0	/	0	0	0	85	0	0
游乐游艺设备制造	0	0	0	0	0	0	359	0	0	0	0	0	0	/	0	0	0	0	0
演艺设备制造及销售	0	0	0	17923	0	0	539	0	0	0	0	0	0	0	/	0	0	0	539
玩具制造	0	0	0	0	0	0	2004	0	0	0	0	0	0	0	0	/	0	0	0
笔墨制造	0	0	0	0	0	0	0	0	0	0	0	0	85	0	0	0	/	0	0
艺术陶瓷制造	0	0	0	0	0	0	12925	0	0	0	0	0	0	0	0	0	0	/	0
节庆用品制造	0	0	0	0	0	0	539	0	0	0	0	0	0	0	539	0	0	0	/

表 3 2013-2018 年文化及相关产业各第二层产业之间融合度

	乐器制造及销售	互联网信息服务	互联网文化娱乐平台	信息服务终端制造及销售	印刷复制服务	印刷设备制造	工艺美术品制造	广播电视电影设备制造及销售	广播电视节目传输	摄录设备制造及销售	数字内容服务	文具制造及销售	文化辅助用品制造	游乐游艺设备制造	演艺设备制造及销售	玩具制造	笔墨制造	艺术陶瓷制造	节庆用品制造
乐器制造及销售	/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
互联网信息服务	0	/	0	0	0	0	0	0	0	0	0.528	0	0	0	0	0	0	0	0
互联网文化娱乐平台	0	0	/	0.547	0.312	0.007	0	0	0	0	0	0	0	0	0	0	0	0	0
信息服务终端制造及销售	0	0	0.547	/	0	0	0	0	0	0.187	0	0	0	0	0.105	0	0	0	0
印刷复制服务	0	0	0.312	0	/	0.092	0	0	0	0	0	0.063	0	0	0	0	0	0	0
印刷设备制造	0	0	0.007	0	0.092	/	0	0.138	0	0.127	0	0.262	0.153	0	0	0	0	0	0
工艺美术品制造	0	0	0	0	0	0	/	0	0	0	0	0	0	0.012	0.013	0.031	0	0.513	0.089
广播电视电影设备制造及销售	0	0	0	0	0	0.138	0	/	0.025	0.232	0	0	0.193	0	0	0	0	0	0
广播电视节目传输	0	0	0	0	0	0	0	0.025	/	0	0	0	0	0	0	0	0	0	0
摄录设备制造及销售	0	0	0	0.187	0	0.127	0	0.232	0	/	0	0	0.178	0	0	0	0	0	0
数字内容服务	0	0.528	0	0	0	0	0	0	0	0	/	0	0	0	0	0	0	0	0
文具制造及销售	0	0	0	0	0.063	0.262	0	0	0	0	0	/	0	0	0	0	0	0	0
文化辅助用品制造	0	0	0	0	0	0.153	0	0.193	0	0.178	0	0	/	0	0	0	0.003	0	0
游乐游艺设备制造	0	0	0	0	0	0	0.012	0	0	0	0	0	0	/	0	0	0	0	0
演艺设备制造及销售	0	0	0	0.105	0	0	0.013	0	0	0	0	0	0	0	/	0	0	0	0.105
玩具制造	0	0	0	0	0	0	0.031	0	0	0	0	0	0	0	0	/	0	0	0
笔墨制造	0	0	0	0	0	0	0	0	0	0	0	0	0.003	0	0	0	/	0	0
艺术陶瓷制造	0	0	0	0	0	0	0.513	0	0	0	0	0	0	0	0	0	0	/	0
节庆用品制造	0	0	0	0	0	0	0.089	0	0	0	0	0	0	0	0.105	0	0	0	/

通过 2013-2018 年文化及相关产业各第二层产业的专利申请数量和相同专利在不同产业之间申请情况的统计及分析,可以看出近 6 年的各第二层产业之间融合度较高,其中专利总申请量排名第 4 的印刷设备制造业是 19 个第二层产业中与其他行业融合数量最多的行业,该行业与互联网文化娱乐平台、印刷复制服务、广播电视电影设备制造及销售、摄录设备制造及销售、文化制造及销售、文化辅助用品制造

行业均有融合,印刷设备制造业与文化制造及销售行业融合度最高为 0.262。由 4.1 可知,信息服务终端制造及销售和互联网文化娱乐平台之间的融合度最高,达到 0.547,其次是数字内容服务和互联网信息服务,融合度达到 0.528,而其他 15 个行业中互联网信息服务行业仅与数字内容服务行业融合,融合度为 0.528;印刷复制服务和互联网文化娱乐平台行业融合度最高为 0.312;印刷设备制造业与文化制造和销售

行业融合度最高为 0.262；工艺美术品制造行业与艺术陶瓷制造行业之间融合度为 0.513；广播电视电影设备制造及销售行业与摄录设备制造及销售行业融合度最高为 0.232；广播电视节目传输行业仅有广播电视电影设备制造及销售行业有所融合，融合度为 0.025；摄录设备制造及销售行业与广播电视电影制造及销售行业融合度最高为 0.232；传统上我们认为没有关系的文具制造及销售行业与印刷设备制造行业融合度达到了 0.262（需要进行分析）；文化辅助用品制造行业与广播电视电影设备制造及销售行业融合度最高为 0.193；游乐游艺设备制造行业仅与工艺美术品制造行业有融合且融合度为 0.012；演艺设备制造及销售行业与节庆用品制造行业和信息终端制造及销售行业融合度一致，均为 0.105；玩具制造行业仅与工艺美术品制造行业融合，融合度为 0.031；笔墨制造行业仅与文化辅助用品制造行业融合，融合度是 19 个第二层产业中融合度最低的为 0.003；艺术陶瓷制造行业仅与工艺美术制造行业有融合，

融合度为 0.513；节庆用品制造行业与演艺设备制造及销售行业融合度最高为 0.105。值得注意的是，2013-2018 年互联网信息服务专利申请的数和该产业与数字内容服务专利一致，数量为 58461 件，且均为相同专利。

3.4 文化及相关产业融合的技术研发方向发现

传统上，我们认为文具制造及销售和印刷设备制造是完全不同的两个行业，但是它们在相似度达到了 0.262，我们希望在传统上完全不相关的行业间找到技术研发方向。分析如下所示：

文具制造及销售和印刷设备制造两个文化产业融合的专利数据上，基于 bert_as_service^[21]进行专利 768 维特征提取，提取的结果如下表 4 所示。

基于相似度进行文具制造及销售和印刷设备制造 768 维特征，用 kmean 算法^[22]进行聚类，聚类类别 10 个，根据聚类结果绘制趋势曲线，X 轴是相似时间，Y 轴是数量。

表 4 特征列表

专利号	特征1	特征2	特征3		特征768
201621031311	-0.3370079	0.25185606	-0.08963802		0.039689276
201520773517	0.020356992	0.2674941	-0.07812791		0.06516011

表 5 10 个类别 2103-2018 每年专利申请量

	2013	2014	2015	2016	2017	2018	合计
1	440	502	611	627	676	450	3306
2	348	292	323	294	280	204	1741
3	356	432	712	727	889	691	3807
4	248	287	341	442	524	525	2367
5	156	222	244	191	256	169	1238
6	497	423	489	467	497	522	2895
7	244	259	288	261	253	233	1538
8	152	110	251	320	592	696	2121
9	168	138	222	241	349	349	1467
10	519	615	616	1021	1673	1433	5877

由以上数据，可以得到如下研发方向发展趋势图。

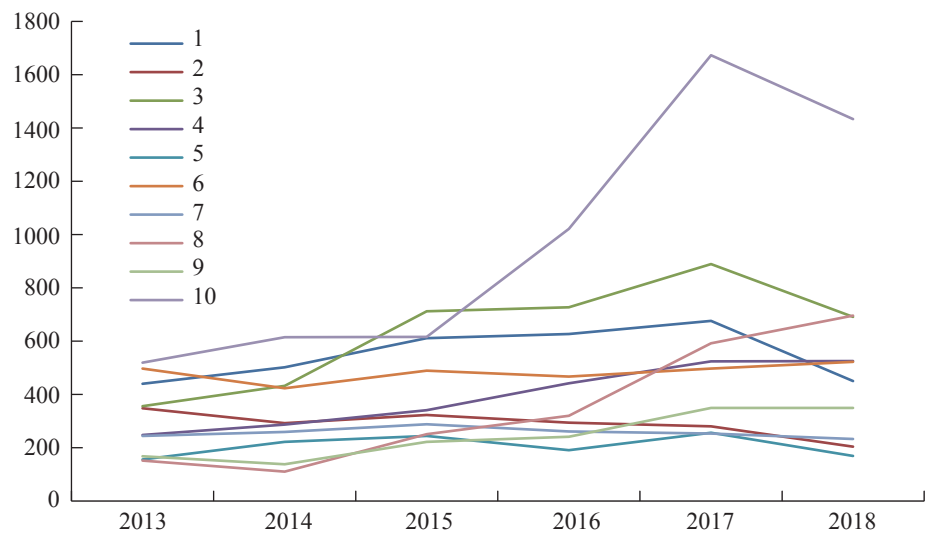


图 4 研究方向发展趋势

从图 4 中曲线可以看出，研发方向 10(印刷装置) 的申请趋势增长最快，说明该类别在近 6 年的申请中增长较快，属于新兴产业。研发方向 8 (档案、票据存储装置) 的申请趋势呈稳步增长时态。研发方向 2 (记录装置) 的

申请趋势为逐年递减，说明该研发方向在近 6 年的申请中逐渐减少，说明技术已经成熟。

在聚类文献上，利用互信息的方法识别方向上的核心关键词，用关键词定义融合发展的新技术研发方向。

研发方向	关键词	方向定义
1	出墨口、喷墨打印、驱动机构、输送方向、打印喷头、打印平台、输送装置、安装部、标签打印机	喷墨打印
2	打印介质、出墨口、容纳部、压力室、印刷头、喷出液体、注入口、印刷介质、检测部、输送路径	记录装置
3	打印介质、出纸口、出墨口、驱动机构、输送方向、输送装置、供墨系统、步进电机、调节装置、升降装置	打印介质
4	打印设备、印刷装置、喷墨头、打印介质、激光打标机、打印喷头、打印平台、打印装置	打印设备
5	记录介质、激光打标机、出墨口、喷墨打印、输送方向、打印系统、热敏打印机、打印机构、打印方法	记录介质
6	热敏打印机、记录头、压力室、打印过程、喷墨印刷、压电元件、喷墨打印头、打印单元、喷墨打印装置	热敏打印机
7	记录介质、输送方向、下游侧、注入口、支承部、喷嘴列、收容部、喷出口、电路基板、液体容纳容器	记录介质
8	固定杆、连接块、放置板、活动板、加持装置、分隔板、文件盒、放置盒、弹簧夹、存放盒	档案、票据存储装置
9	长度方向、下表面、上端面、固定杆、中间位置、安装槽、伸缩杆、对称装置、限位板、弹簧夹	打印机机构
10	印刷装置、打印系统、热敏打印机、打印机构、打印平台、标签打印机、打印机主体、打印过程	印刷装置

综上,我们发现了虽然文具制造及销售和印刷设备制造是完全不同的两个行业,但是有产业融合相互渗透、相互交叉的研发方向。特别是4(打印设备)、8(档案、票据存储装置)、10(印刷装置)四个研发方向上,这几年有逐年上升的趋势。说明文具制造企业开始从事设计、制作、印刷各种宣传资料、招贴画等业务,同时文具制造的研发方面,文具的设计制作开始更加精美,需要特殊印刷装置进行印制。这是一个未来可能大力发展的新兴领域。

4 小结

以专利数据为基础的专利分析反映了技术创新的程度,基于国民经济分类和IPC的映射关系可以得到行业和产业的专利数据,基于语义聚类的行业及产业技术研发方向分析,定量进行了行业和产业的融合分析,在融合方向上寻找创新发展方向。通过对文化及相关产业的实例分析,发现了文具制造及销售和印刷设备制造两个文化产业融合的专利数据上的融合分析的新研发方向。

本文在数据分析上还有待加强,第一是没有预测新技术的未来发展趋势,下一步将结合趋势分析和数据挖掘可以实现。三是融合其他数据对文化产业进行验证,新闻数据、期刊数据和专利数据的融合分析将能够帮助我们进一步得到更加准确的研发方向和研发主体。

参考文献

- [1] 彭茂祥,徐勇.基于专利数据的我国文化及相关产业创新态势分析[J].中国发明与专利,2016(7):36-40.
- [2] Campbell R S. Patent trends as a technological forecasting tool[J]. World Patent Information, 1983, 5(3):137-143.
- [3] 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.
- [4] Okoli C, Pawlowski S D. The Delphi method as a research tool: an example, design considerations and applications[J]. 2004, 42(1):15-29.
- [5] 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, 2011, 33(2):137-141.
- [6] Foresight and Understanding for Scientific Exposition (FUSE)[EB/OL]. [2019-01-13]. http://www.iarpa.gov/solicitations_fuse.html
- [7] Combining and Uniting Business Intelligence with Semantic Technology (CUBIST) [EB/OL]. [2019-03-22]. <http://www.cubist-project.eu>.
- [8] VantagePoint[EB/OL]. [2019-03-11]. <http://www.thevantagepoint.com0>
- [9] Kim J, Lee S, Lee J, et al. Design of TOD Model for Information Analysis and Future Prediction[M]. U- and E-Service, Science and Technology. Springer Berlin Heidelberg, 2011.
- [10] Kim I H, 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.
- [11] 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.
- [12] 国家知识产权局.国际专利分类与国民经济行业分类参照关系表(2018)[M].北京:国家知识产权局办公室,2018.
- [13] 罗勇,曹丽莉.产业融合的创新点[J].中国发展观察,2004(10):40-42.
- [14] 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.
- [15] McCain K W. Mapping authors in intellectual space: A technical overview[J]. Journal of the Association for Information Science & Technology, 2010, 41(6):433-443.
- [16] 张文彤. SPSS11 统计分析教程 (高级篇) [M]. 北京: 希望电子出版社, 2002
- [17] Leydesdorff L, Vaughan L. Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment[M]. John Wiley & Sons, Inc, 2006.
- [18] 杨良斌, 金碧辉, 等. 跨学科测度指标体系的构建研究 [J]. 情报杂志, 2009, 28(7):65-69.
- [19] Devlin J, Chang M W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. <https://arxiv.org/abs/1810.04805>
- [20] 国家统计局社会科技和文化产业统计司, 中宣部文化体制改革和发展办公室 编. 2018 中国文化及相关产业统计年鉴 [M]. 北京: 中国统计出版社, 2018.
- [21] Han X. xiao2018bertservice[EB/OL]. [2018-09-13]. <https://github.com/hanxiao/bert-as-service>.
- [22] 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.