

大数据视域下阳明心学主题挖掘及作者发现研究

杨秀璋¹,武帅¹,夏换^{2,3},窦悦琪¹,赵紫如¹,朱涤尘¹,张亚成¹

(1. 贵州财经大学信息学院,贵阳 550025; 2. 贵州财经大学,贵州省经济系统仿真重点实验室,贵阳 550025;
3. 贵州财经大学,贵州省电子商务大数据营销工程研究中心,贵阳 550025)

摘要:

随着互联网和大数据的迅速发展,如何从海量信息中快速获取所需知识,精确识别学术成果的核心主题及作者关系变得越来越重要。阳明心学作为中国传统思想文化重要组成部分,传承与弘扬的重要性不言而喻。提出一种基于大数据分析技术的阳明心学研究方法,包括文献计量、社交网络、LDA 模型、词云可视化,旨在挖掘阳明心学文献的热门主题和核心作者,并结合可视化技术展现阳明心学的研究热点及作者关系图谱。实验数据采集了中国期刊全文数据库(CNKI)收录的 2463 篇阳明心学期刊文献,最终发现以左东岭、杨国荣、张新民、王路平、陆永胜等为代表的阳明心学核心研究者,挖掘的热门主题包括“王阳明”“阳明心学”“知行合一”“致良知”“传习录”等,LDA 模型识别出三个核心主题。所提出的方法能有效地挖掘出中国传统文化的主题关键词及作者关系图谱,使研究的主题脉络更加清晰,为推动阳明心学的弘扬与发展提供帮助,具有一定的应用前景和实用价值。

关键词:

阳明心学; 大数据; 社交网络; LDA 模型; 主题挖掘

基金项目:

贵州省科技计划项目:多源地理数据融合知识图谱构建方法在舆情分析中的应用——以贵州省为例(黔科合基础[2019]1041);贵州省科技计划项目:圆形地下连续墙结构时变性仿真研究(黔科合基础[2019]1403);贵州省科技计划项目:基于大数据及图像识别的水族文献及濒危水书抢救性整理研究(黔科合基础[2020]1Y279);贵州财经大学 2019 年度校级项目:基于多源异构数据融合的建设项目管理知识图谱研究与实现(2019XQN01)

0 引言

阳明心学是明代著名思想家王阳明的心学思想,是中国传统文化的精华,其精神内涵包括“知行合一”“致良知”“身之主宰便是心”等^[1]。阳明文学强调“心即是理”,即最高的道理不需外求,而从自己本心得到,其中的四句教“无善之恶心之体,有善有恶意之动,知善知恶是良知,为善去恶是格物”广为流传。目前共分为七个学派:右派(江右学派)、南中王门学派、闽粤王门学派、北方王门学派、楚中王门学派、左派(浙中王门学派)、泰州学派^[2-3]。阳明心学文献作为其交流与传播的重要载体,有效地推动阳明心学的发展。当前阳明心学的研究主要采用期刊阅读、书籍查阅、学派交流,核心作者仅仅依据发文量来判断,缺乏利用大数据技术

来进行分析和研究。

随着互联网和大数据技术的迅速发展,传统文化相关的学术成果和在线博客呈爆炸式增长,如何从海量文献中识别出核心科研作者、挖掘出热门主题变得越来越重要。核心作者是学科研究的坚实基础,决定着学科研究方向以及学术成果的质量^[4]。传统识别方法是通过发文量来认定,却忽略了论文质量,缺乏利用社交网络或知识图谱技术构建核心作者间的关系,识别结果往往存在片面性,针对这些问题,本文通过社交网络方法挖掘阳明心学的核心作者深层次合作关系。同时,利用 LDA 主题模型挖掘阳明心学文献的核心主题关键词,并结合可视化技术直观地反映阳明心学的研究热点。

近年来,国内外学者致力于学术文献研究并提出各自的分析方法。侯建华^[5]通过 PathFinder 算法,基于美国科学情报研究所的 Web of Science 三个检索数据库(SCI、SSCI、A&HCI)绘制和分析工商管理学科主干理论演进的关键路径图谱。杨秀璋等^[6]基于 LDA 主题模型挖掘水族文献知识,揭示学科领域作者合作关系。何超^[7]借助知识图谱理论与方法研究我国工商管理学科,并系统绘制管理科学学科知识图谱。严红等人^[8]基于 CSSCI 引文索引数据库为数据来源,采用 CiteSpace 引文空间可视化分析方法,绘制国内协同创新研究热点知识图谱。黄海瑛^[9]通过知识图谱和可视化技术系统分析外国文学文献及热点变迁。

综上,结合国内外学者的研究现状,本文提出一种基于大数据分析技术的阳明心学研究方法,包括社交网络、LDA 模型、词云可视化,旨在挖掘阳明心学文献的热门主题和核心作者,并结合可视化技术展现阳明心学的研究热点及作者关系图谱。

1 研究方法

1.1 基本思路与框架

本文旨在利用大数据分析技术研究阳明心学文献,通过文献计量、社交网络、LDA 模型、综合指数、可视化分析方法挖掘阳明心学文献的热门主题和核心作者。该算法的框架图如图 1 所示。

(1)采用 Python 和 Selenium 技术自定义爬虫抓取 KI 阳明心学文献数据。

(2)数据预处理包括中文分词、停用词过滤和特征提取,提取阳明心学相关的特征并存储至数据库,包括文献标题、文献作者、文献摘要、文献来源、发表时间、被引量 and 下载量。

(3)阳明心学大数据分析包括文献计量分析、核心作者计算、科研团体发现和研究主题挖掘,所涉及的技术包括文献计量、普赖斯定律、综合指数、社交网络、LDA 模型和可视化技术。

1.2 数据采集及预处理

本文旨在挖掘中国知网阳明心学文献的核心作者,分析阳明心学核心期刊杂志及研究主题,共抓取中国知网 1974 年至 2019 年 4 月期间 2463 篇阳明心学相关的学术成果,并对抓取的文献进行数据预处理,从而得到更准确、更完整、质量更高的文献信息。实验所

采用的数据预处理主要包括异常值处理、数据清洗和中分分词。

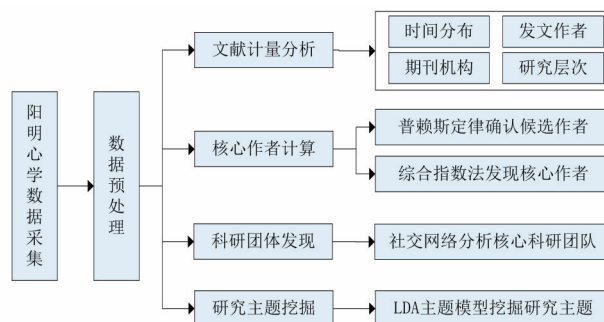


图1 阳明心学研究框架图

1.3 LDA 模型

LDA (Latent Dirichlet Allocation) 是一种文档主题生成模型,由 Blei 等人^[10]在 2003 年首次提出,是一种基于主题(T)、文档(D)和主题词(W)的三层贝叶斯结构,其中文档到主题(D-T)以及主题到主题词(T-W)层面均服从多层分布。LDA 模型将一篇文本的每个词都按照一定概率分布到某个主题上,并从这个主题中选择相关的词语集,如图 1 所示,将 d 篇文档映射到 k 个主题中,每个主题包括一定量的主题词。

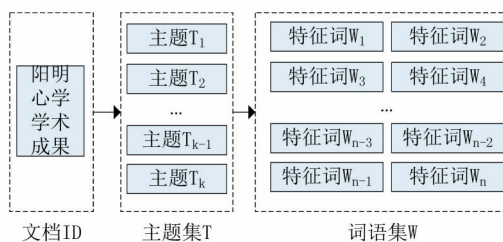


图2 文档-主题-词映射模型

本文针对阳明心学文献完成主题挖掘研究,经过 LDA 主题分布后,得到各个文档的不同主题所占比例,实现阳明心学各主题关键词的挖掘。

2 阳明心学文献计量分析

本文采用 Python 自定义爬虫抓取中国知网阳明心学相关的学术成果,共获取 2463 篇学术成果,包括学术期刊论文 1825 篇、博士论文 134 篇、硕士论文 354 篇、国内外会议论文 63 篇、报纸文献 87 篇,详细信息

如表 1 所示。

表 1 中国知网 1974 年至 2019 年 3 月的阳明心学学术成果汇总表

学术成果类型	刊载量	下载量	引用量	单篇最高下载量	单篇最高引用量	篇均下载量	篇均引用量
学术期刊	1825	398637	4623	3008	71	218.43	2.53
博士论文	134	172675	879	6780	49	1288.62	6.56
硕士论文	354	133733	802	2611	21	377.78	2.27
会议论文	63	4137	6	210	2	65.67	0.10
报纸	87	3609	9	200	4	41.48	0.10
总计	2463	712791	6319	15560	159	2127.56	12.34

2.1 近十年发展综述分析

本文首先利用文献计量和可视化技术分析阳明心学近十年的发展概况。通过 PyEcharts 库绘制如图 3 所示的趋势图,图中横轴为 2009 至 2018 年这十年的时间轴,纵轴为近十年中国知网阳明心学相关的学术成果的发文量(紫色折线图)、最高引用量(红色柱状图)和最高下载量(蓝色柱状图)。整体而言,发文量呈平缓增长趋势,可见越来越多的学者关注阳明心学研究,对阳明心学的传承和弘扬起到一定的推广作用。其中,2014 年陈琪^[1]《王阳明“致良知”思想研究》博士论文下载量最高,共被下载 6780 次,被引用 29 次。

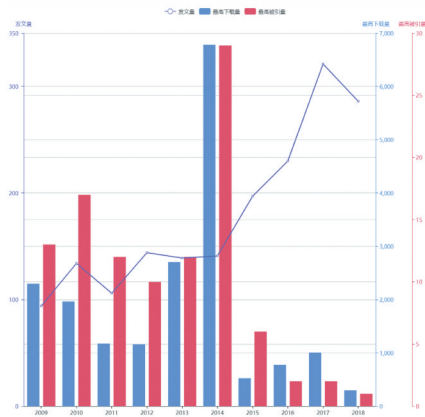


图 3 近十年学术成果发展趋势图

2.2 发文作者分析

对阳明心学学术成果相关作者发文量进行汇总排名,得出如表 2 所示的结果。其中,贵州大学中国文化书院的张新民、贵阳学院阳明学与黔学研究所陆永胜均发表 26 篇以阳明心学为主题的文献。

表 2 作者发文量前十统计表

排名	作者	机构	发文量
1	张新民	贵州大学	26
2	陆永胜	贵州学院	26
3	王路平	贵州省社会科学院	22
4	张明	贵州大学	18
5	何静	宁波大学	14
6	王胜军	贵州大学	14
7	姚才刚	湖北大学	14
8	吴震	复旦大学	13
9	潘立勇	浙江大学	13
10	李承贵	南昌大学	13

2.3 学科类别和研究类型分析

针对中国知网阳明心学学术成果的学科类别和研究类型进行计量分析,得出如图 4 和图 5 所示的结果。其中,图 4 是阳明心学学术成果学科类别分析,以哲学与文学为主,其中哲学占 59.07%、文学占 10.11%,阳明心学主要归于哲学和文学范畴。

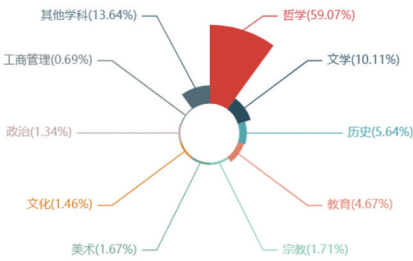


图 4 学科类别统计分析

图 5 是阳明心学研究类型分析,排名最高的分别是基础研究(社科)、政策研究(社科)、行业研究(社科),分别占 80.07%、4.38%、3.69%,该结果间接反映出我国大多数研究阳明心学的为基础社科研究。

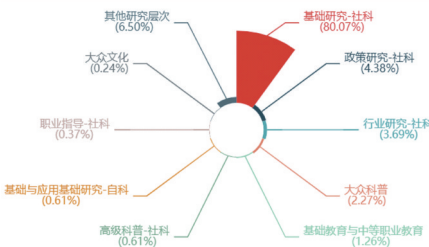


图 5 研究类型统计分析

2.4 学术期刊分析

学术期刊分析能有效挖掘研究阳明心学的中坚力量,它们有效推动了中国传统文化的研究。本文采用

阳明心学文献核心作者的候选人。

(4)计算阳明心学核心作者候选人的平均发文量和平均被引用量。其计算公式如下:

$$\bar{y} = \frac{Y}{n} = \frac{1794}{45} = 39.867 \quad (4)$$

(5)依据发文量与被引用量构建综合指数,从阳明心学在中国知网收录文献的数量和质量两个角度评估核心作者候选人。综合指数计算公式如下:

$$score_i = \frac{x_i}{\bar{x}} \times 0.5 + \frac{y_i}{\bar{y}} \times 0.5 \quad (5)$$

运用此综合指数方法对 45 位核心作者候选人进行计算,得出如表 3 所示的前 20 位核心作者。其中,左东岭发表阳明心学相关文献 9 篇,被引用量为 180 次,综合指数为 2.77;杨国荣发表阳明心学相关文献 11 篇,被引用次数为 159 次,综合指数为 2.62;张新民发表阳明心学相关文献 26 篇,被引用量为 46 次,综合指数为 2.06;王路平发表阳明心学相关文献 22 篇,被引用次数为 58 次,综合指数为 1.98;陆永胜发表阳明心学相关文献 26 篇,被引用次数为 29 次,综合指数为 1.85。该方法有效挖掘出阳明心学研究的核心科研工作者。

表3 阳明心学学术成果核心作者统计表

排名	作者	发文量	被引用量	综合指数	排名	作者	发文量	被引用量	综合指数
1	左东岭	9	180	2.77	11	姚才刚	14	51	1.44
2	杨国荣	11	159	2.62	12	丁为祥	9	72	1.42
3	张新民	26	46	2.06	13	宋克夫	9	63	1.30
4	王路平	22	58	1.98	14	吴光	10	54	1.25
5	陆水胜	26	29	1.85	15	方国根	7	64	1.20
6	潘立勇	13	84	1.80	16	陈宝良	4	73	1.14
7	吴震	13	77	1.71	17	钱明	10	40	1.07
8	何静	14	68	1.65	18	李承贵	13	13	0.91
9	张明	18	47	1.62	19	华建新	10	22	0.85
10	刘宗贤	11	71	1.52	20	袁琼	4	49	0.84

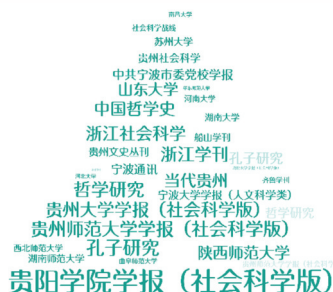


图6 期刊文献分析

3 阳明心学核心作者计算

(1)依据普赖斯定律统计核心作者候选人刊发文章的累计最低发文量。其计算公式如式(1)所示:

$$M_p = 0.749 \times \sqrt{N_p \max} = 0.749 \times \sqrt{26} = 3.819 \quad (1)$$

(2) 计算核心作者候选人刊发文章的累计最低被引用量。实验发现, 阳明心学文献中作者发文被引用量最高为 71 次, 依据普赖斯定律确定核心作者候选人的最低被引用量, 其计算公式(2)如下:

$$M_c = 0.749 \times \sqrt{N_{c \max}} = 0.749 \times \sqrt{71} = 6.311 \quad (2)$$

其中, Mc 为普赖斯定律统计阳明心学文献的最低被引用量, N_{max} 为作者发文单篇被引用最高量。依据普赖斯定律, 选择被引用量在七次以上的作者作为

4 阳明心学作者关系发现

针对阳明文化核心作者关系发现,本文提出一种基于社交网络和共现矩阵的分析方法,利用 Gephi 构建阳明心学文献作者间的关系图谱,进一步挖掘对阳明心学做出重要贡献的科研群体。具体操作如下:

(1)首先计算中国知网 2463 篇阳明心学学术成果的所有作者共现矩阵,当两位作者出现在同一篇文章中时,则认为共现并构建一条相关关系边,否则没有。

(2)采用 Gephi 构建阳明心学作者关系图谱,绘制如图 7 所示的效果图。由图可知,圆圈越大表示合作关系越密切,反之越少。

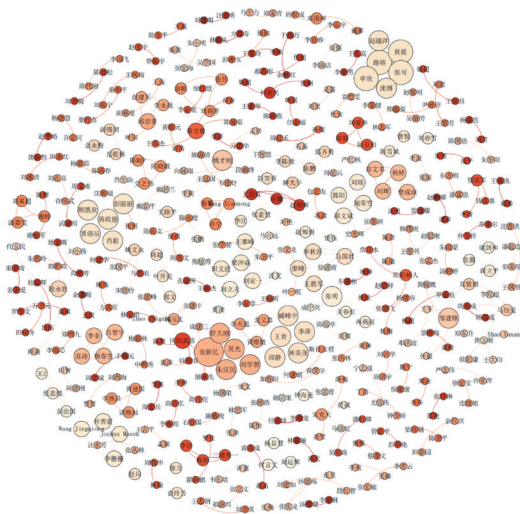


图7 阳明心学核心作者合作关系图谱

该知识图谱共挖掘具有合作关系的核心作者 435 名,合作次数 321 次,平均聚类系数为 0.844,平均路径长度为 1.137。阳明心学核心科研群体包括以张新民老师为代表的贵州大学中国文化书院科研团队,其研究方向是中国古代史、哲学和阳明文化;以邹建锋老师为代表的湖州师范学院团队,他们的研究方向是明代思想史、明代儒学和哲学;以陆永胜老师为代表的贵阳学院科研团队,其研究方向是中国文学、哲学和王阳明心学美学。

综上,这些科研团队对中国传统文学及阳明文化研究均有一定的贡献,同时,更多阳明心学研究者采用独自科研并发表相关学术成果的方式。本文也鼓励科研工作合作研究、共同探讨,为中华传统文化的保护

和传承提供帮助。

5 阳明心学研究主题挖掘

本文采用WordCloud和LDA模型挖掘阳明心学的主关键词。阳明心学学术成果的热门主题词主要包括“王阳明”“阳明心学”“王阳明心学”“致良知”“传习录”“知行合一”“龙场悟道”“程朱理学”等,它们被研究的频率较高,直接反映了阳明心学的核心价值及学术研究方向。

在基于 LDA 模型的主题挖掘实验中,本文通过实验发现最佳主题数为 3,迭代次数为 500,最终实验结果如图 8 所示,三个圆圈分别代表阳明心学学术成果的三个核心主题,即阳明心学和知行合一(Topic 1)、哲学思想和美学启示(Topic 2)、儒学思潮和明代文学(Topic 3),它们明显区别且联系密切。图 8 展示了主题 3 对应的词频大小,LDA 主题模型有效凝聚了阳明心学的主题,可视化直观地呈现各主题对应的关键词分布差异及重要程度。

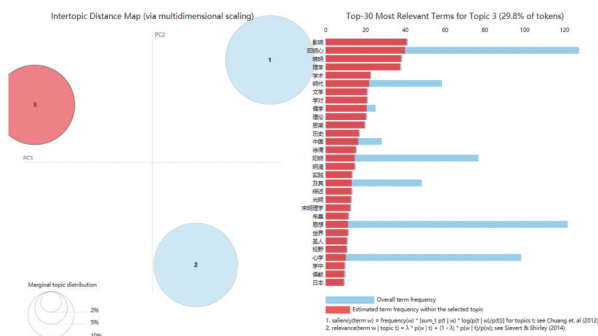


图8 阳明心学LDA主题挖掘

同时,为进一步挖掘各主题的核心关键词,通过 LDA 主题模型提取三大主题的前 20 个主题词,如表 4 所示。

表 4 LDA 模型主题-关键词识别结果

主题	主题词
Topic1 阳明心学 知行合一	心学 王阳明 明代 知行合一 发展 阳明心学 阳明 关系 书院 研究 探析 贵州 学派 龙场悟道 传习录 学说 思想 心性 解读 中国
Topic2 哲学思想 美学启示	王阳明 哲学 良知 阳明心学 教育 哲学思想 思想 心学 美学 道德 精神启示 绘画 内涵 传承 人格 智慧 伦理思想 审美 本体 诗歌
Topic3 儒学思潮 明代文学	儒家 浅析 文化 影响 阳明心学 明代 文学 儒学理论 思潮 历史 中国 理学 学术 宋明理学 朱熹 心学 佛教 渊源 晚明 文学探讨

由表可知,每个主题的差异明显,主题 1 中的特征词主要是阳明心学和知行合一,包括“心学”“王阳明”“知行合一”“龙场悟道”“研究”等主题关键词;主题 2 中的特征词主要是哲学思想和美学启示,包括“王阳明”“哲学”“良知”“哲学思想”“伦理”“精神启示”等主

题关键词;主题 3 中的特征词主要是儒学思潮和明代文学,包括“儒家”“文化”“阳明心学”“影响”“明代”“文学”“儒学理论”等主题关键词。

6 结语

针对中国传统文学和哲学缺乏利用大数据技术分析的特点,本文提出一种基于大数据分析技术的阳明心学研究方法,主要采用文献计量、社交网络、LDA 模型、词云可视化对中国知网所收录的阳明心学相关的 2463 篇学术成果进行研究。

实验结果表明,本文方法发现了以左东岭、杨国荣、张新民、王路平、陆永胜为代表的阳明心学核心研究者,挖掘的热门主题包括“王阳明”“阳明心学”“知行合一”“致良知”“传习录”等,LDA 模型识别出三大核心主题,分别是阳明心学和知行合一、哲学思想和美学启示、儒学思潮和明代文学。本文方法能有效地挖掘出中国传统文化的主题关键词及作者关系图谱,使研究的主题脉络更加清晰,为推动阳明心学的弘扬与发展提供帮助,具有一定的应用前景和实用价值。

参考文献:

[1]王守仁. 王阳明集[M]. 北京:中华书局,2016.
[2]刘宗贤. 从“四句教”看王阳明心学的哲学意蕴[J]. 东岳论丛,2004(06):155-160.
[3]郦波. 五百年来王阳明[M]. 上海:上海人民出版社,2017.
[4]廉清. 《图书情报工作》核心作者群分析研究[J]. 现代情报,2004(11):55-59.
[5]侯剑华. 工商管理学科演进与前沿热点的可视化分析[D]. 辽宁:大连理工大学,2009.
[6]杨秀璋,夏换,于小民,等. 基于综合指数和知识图谱的水族文献核心作者群分析研究[J]. 计算机时代,2019(4):13-17.
[7]何超. 我国管理科学学科演进的知识图谱研究[D]. 湖南:湖南大学,2012.
[8]严红,尹继东,许水平. “2011 计划”实施以来协同创新研究热点与前沿分析——基于 Citespace 引文空间可视化分析方法的实证研究[J]. 科学管理研究,2019(01):13-15.
[9]黄海璇. 基于知识图谱的外国文学研究的可视化分析[J]. 信息资源管理学报,2014(1):14-20.
[10]BLEI D M,NG A Y,JORDAN M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research,2003(3):993-1022.
[11]陈琦. 王阳明“致良知”思想研究[D]. 吉林大学,2014.
[12]丁海东. 文献计量学基础[M]. 北京:北京大学出版社,1992:204-209+220-232.

作者简介:

杨秀璋(1991-),男,贵州凯里人,硕士,研究方向为 Web 数据挖掘、知识图谱、文献分析
通信作者:夏换(1981-),男,湖南永州人,博士,教授,研究方向为计算机仿真、大数据分析
收稿日期:2020-09-15 修稿日期:2020-10-18

Research on Yangming's Psychology Theme Mining and Author Discovery Based on Big Data

YANG Xiu-zhang¹, WU Shuai¹, XIA Huan^{2,3}, DOU Yue-qi¹, ZHAO Zi-ru¹, ZHU Di-chen¹,
ZHANG Ya-cheng¹

(1. School of Information, Guizhou University of Finance and Economics, Guiyang 550025;

2. Guizhou Key Laboratory of Economics System Simulation, Guizhou University of Finance and Economics, Guiyang 550025;

3. Guizhou Province E-Commerce Big Data Marketing Engineering Research Center, Guizhou University of Finance and Economics, Guiyang 550025)

Abstract:

With the rapid development of the Internet and big data, how to quickly obtain the required knowledge from massive amounts of information and accurately identify the core themes of academic achievements and author relationships has become more and more important. As an important part of Chinese traditional ideology and culture, Yangming's psychology is of great importance for inheritance and promotion. This paper proposes a research method of Yangming's psychology based on big data analysis technology, including bibliometrics, social networks, LDA model, wordcloud visualization, and aims to mine the popular topics and core authors of Yangming's psychology literature, combined with visualization technology Show the research hotspots of Yangming's psychology and the author's relationship group. The experimental data collected 2463 articles of Yangming's Psychology journals included in the CNKI, and finally found that the core researchers of Yangming's Psychology represented by Zuo Dongling, Yang Guorong, Zhang Xinmin, Wang Luping, Lu Yongsheng, etc. Popular themes include "Wang Yangming", "Yangming's Psychology", "Unity of Knowledge and Action", "To Conscience", and "Records of Practice". The LDA model identifies three core themes. The method in this paper can effectively identify the subject keywords of traditional Chinese culture and the author relationship map, make the research subject line clearer, and provide help to promote the promotion and development of Yangming's Psychology. It has certain application prospects and practical value.

Keywords:

Yangming's Psychology; Big Data; Social Network; LDA Model; Topic Mining

(上接第 7 页)

Design of Double-Plate Composite Machine

GUO Peng^{1,3}, MEI Xue-chuan^{1,3}, CHEN Jie², GUO Ru-feng^{1,3}

(1. SINOMACH Intelligence (Suzhou) Co., Ltd., Suzhou 215134;

2. School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640;

3. SINOMACH Intelligence Technology Co., Ltd., Guangzhou 510700)

Abstract:

The double-plate composite technology is widely used in the fields of metal materials, polymer materials and inorganic materials. Especially in recent years, China has accelerated the modernization of the housing industry, and the demand for composite panels in the construction and decoration industry is growing. A fully automated double-plate composite machine was designed for the needs of double-plate composite automation production. The double-plate composite machine is mainly composed of a bottom plate positioning device, a panel positioning device, a gantry robot arm, a panel suction tooling and a PLC control system. Through the PLC system to control each device to achieve the composite of two plates, the control principle and work flow of the automatic double-plate composite is expounded.

Keywords:

Plate Positioning; Double-Plate Composite; PLC Control; Automation