国际传播与区域发展视角下的陕西外媒报道热点研究——基于 LDA 模型

刘洁琳1*. 周钰颖2

¹西安文理学院外国语学院 陕西西安 ²西安外国语大学商学院 陕西西安

【摘要】随着中华文化"走出去"战略的不断推进,区域文化的对外传播已成为国家形象构建与国际文化话语权竞争的重要组成部分。本研究基于潜在狄利克雷分配(Latent Dirichlet Allocation, LDA)主题模型,对近年来境外媒体关于"陕西"的英文新闻报道语料进行数据挖掘与热点分析,考察境外媒体对陕西新闻热点话题的内容分布、演变逻辑及发展趋势。借鉴舆情研究中新闻与公众评论的主题关联建模思路,本研究进一步探讨境外新闻话语中传播内容与区域经济发展之间的关系,从而揭示国际传播中区域形象构建与媒体关注热点的异同。本研究将 LDA 主题建模与新闻数据挖掘方法应用于陕西新闻热点分析,有助于提升新闻话语分析的科学性与前瞻性,并为陕西乃至西部地区优化国际传播路径与热点形象构建提供实证支持与策略参考。

【关键词】陕西;境外媒体英文新闻报道; LDA 主题模型; 热点分析

【基金项目】陕西省哲学社会科学研究专项"陕西省公共空间汉字文化英译调查与研究"(项目编号: 2025HZ0719);陕西省哲学社会科学研究专项"服务国际传播的高校外语教师发展创新示范中心建设研究"(项目编号: 2025HZ0850)

【收稿日期】2025 年 9 月 17 日 【出刊日期】2025 年 10 月 16 日 【DOI】10.12208/j.ssr.20250407

Shaanxi hotspots in foreign media coverage: An LDA-based study from the perspective of international communication and regional development

Jielin Liu^{1*}, Yuving Zhou²

¹School of Foreign Studies, Xi'an University, Xi'an, Shaanxi ²School of Business, Xi'an International Studies University, Xi'an, Shaanxi

【Abstract】With the continuous advancement of China's "Going Global" strategy, the international dissemination of regional culture has become a key component in shaping national image and competing for cultural discourse power. This study employs the Latent Dirichlet Allocation (LDA) topic model to conduct data mining and hotspot analysis on English news reports about Shaanxi published by overseas media in recent years. It examines the content distribution, evolutionary patterns, and development trends of news hotspots concerning Shaanxi. Drawing on approaches from public opinion research that associate news with audience commentary, the study further explores the relationship between the dissemination of overseas news discourse and regional economic development, thereby revealing the similarities and differences in regional image construction and media attention in international communication. By applying LDA topic modeling and news data mining to the analysis of Shaanxi-related news hotspots, this study contributes to enhancing the scientific rigor and foresight of news discourse analysis and provides empirical support and strategic insights for optimizing international communication strategies and hotspot image construction in Shaanxi and the broader western region of China.

Keywords Shaanxi; Overseas media coverage; Latent Dirichlet Allocation (LDA); Hotspot analysis

1 引言

近年来,陕西在对外形象传播中,力图呈现立体、

真实、全面的区域形象。2023年,陕西省人民政府新闻办公室在新闻发布会上强调"文化陕西"品牌的构建

与传播路径。作为中国西部的重要省份,陕西不仅拥有丰富的历史文化资源,西安作为丝绸之路的起点城市,也在经济发展中发挥了重要作用。然而,区域形象的呈现不仅取决于省份及城市自身的传播策略,也受到外媒报道选择与关注焦点的影响。尤其在中国推进高水平对外开放、构建"双循环"发展格局的背景下,境外媒体报道是否与地方宣传策略相匹配,以及其关注焦点与区域发展趋势之间的关系,成为亟待深入分析的问题。

随着大数据和自然语言处理技术的发展,基于新闻文本的主题识别与分析成为可能。潜在狄利克雷分配(LDA)主题模型已被广泛应用于新闻、评论和政务文本的自动主题识别与分类分析,为理解区域外宣热点和传播机制提供了技术手段[1,2]。基于此,本研究以陕西为典型案例,构建涵盖 2005 至 2025 年境外媒体报道的英文语料库,通过 LDA 主题模型识别核心传播主题,并探讨报道主题与官方对外宣传及区域发展之间的关联。本文研究拟在两个层面展开:一是境外媒体报道热点分析,结合国际媒体报道数据,识别外媒关注的高频主题与潜在话题,并分析其与陕西官方对外宣传策略的匹配程度;二是区域形象与经济发展分析,结合陕西经济与产业特征,探讨外媒报道与区域发展趋势、投资环境及产业结构调整的关联,为内陆地区国际传播与经济发展研究提供参考。

2 文献综述

2.1 陕西区域形象的构建与海外呈现

2023 年 9 月 25 日,陕西省人民政府新闻办公室在新闻发布会上提出,要强化新的文化使命担当,扎实开展"三个年"活动,并奋力推进陕西文化旅游高质量发展。在发布会上,明确提出"文化陕西"的品牌构建和传播路径,既强调秦岭文化、丝绸之路起点、兵马俑之乡等传统品牌的长期化,也提出挖掘更多陕西特色品牌,如"根脉陕西""臻美陕西"^[3]。在对外形象构建中,陕西长期依托深厚的文化底蕴和丰富的旅游资源,向世界展示千年文化的省份形象。近年来,通过沉浸式演出、博物馆建设以及非物质文化遗产与旅游的结合,陕西形成了独特的传播品牌和路径。

在海外社交媒体中,例如 YouTube, 西安的城市形象主要集中于餐饮美食、历史人文与经济发展, 其中餐饮美食和历史人文是境外视频媒体平台的关注重点^[4]。已有研究指出,西安作为世界文化名城的形象深入人心,其宣传仍以旅游、文化、美食和宗教为主,而科技和经济发展相关内容关注度较低^[5]。语料库分析显示,

外媒对西安的形象塑造较为片面,集中于古都与旅游资源,由此得出外媒对陕西整体形象塑造单一的结论^[6]。自"一带一路"建设推进以来,尤其是 2020 年 11 月中国与东盟签署《区域全面经济伙伴关系协定》(RCEP)后,中国与东盟的商业往来进一步加强。研究发现,东盟媒体对西安的报道主题涵盖国际经济合作、酒店与地产业、贸易往来以及旅游和历史文化产业发展等方面,体现了区域形象在文化、旅游与经济产业间的多元呈现^[7]。

总体来看,陕西对外宣传以文化旅游为核心,外媒 报道则在关注文化与旅游的基础上,逐步聚焦经济与 产业议题。这为理解官方传播策略与外媒实际关注点 的匹配度提供了研究空间,也显示了区域形象塑造与 国际投资环境之间的潜在联系。

2.2 境外媒体舆论与区域经济发展

研究表明,境外媒体对陕西经济发展的关注日益增加,尤其依托"一带一路"建设与区域发展热点。新闻文本数据作为高频、实时信息源,可用于分析外媒报道对区域经济与投资行为的潜在影响。已有研究表明,国际媒体报道的数量与基调,与外商直接投资(FDI)流入及企业决策行为存在显著关联^[8-10]。此外,移动互联网与社交媒体的发展,进一步扩大了舆论影响力,使外媒报道能够间接影响投资选择与产业布局。跨文化交流活动则有助于改善偏向、提升外资认知水平与区域形象^[11]。

然而,现有研究仍存在局限:第一,多数关注国家 或省级宏观层面,缺乏典型内陆省份如陕西的微观动 态分析与热点识别;第二,外媒报道的关注焦点与潜在 信号分析不足;第三,方法上仍以传统统计为主,缺乏 对新闻文本的深度挖掘和结构化建模。

2.3 本研究的突破与研究重点

为弥补上述不足,本文以陕西为典型案例,结合海量国际媒体报道数据,利用 LDA 主题模型识别外媒关注热点与潜在话题,同时结合区域经济与产业结构,分析报道主题与官方宣传及区域发展趋势的关联。研究旨在: (1)提炼境外媒体对陕西的高频报道主题及潜在关注点; (2)分析外媒报道主题与官方对外宣传策略的匹配度,以及报道如何反映区域热点; (3)探讨外媒报道主题与陕西经济发展和产业结构调整之间的关联,为理解内陆地区国际传播与投资环境提供参考。

通过整合传播学与区域经济视角,本研究不仅丰富了LDA在外宣文本分析中的应用实践,也为地方政府和媒体优化国际传播策略、提升区域全球形象和投

资吸引力提供了实证参考。

3 LDA 主题模型分析

本研究基于专业新闻数据库 Factiva,以"Shaanxi"为检索关键词,系统收集了近二十年(2005-2025年)欧美、亚洲等境外主要媒体对陕西的相关报道。初步检索共获得9185条新闻记录,剔除重复内容后,得到有效新闻文本8175篇。

在数据预处理阶段,首先对原始文本进行去噪处理,清除 HTML 标签、特殊字符及广告链接等无关信息。随后,利用自然语言处理工具对文本进行分词,将连续文本拆分为可识别的词语单位。在此基础上,结合 NLTK 标准英文停用词表与自定义停用词表,去除高频虚词,并采用 WordNet Lemmatizer 对动词时态、名词复数等进行词形还原,以提高主题建模的准确性和一致性。

完成预处理后,整理好的语料输入 LDA (Latent

Dirichlet Allocation) 主题模型进行训练。训练前需预设主题数 K,为确定最优 K 值,研究团队通过多轮实验,综合评估困惑度(Perplexity)和一致性得分(Coherence Score)。其中,困惑度用于衡量模型对未见数据的拟合能力,值越低表示模型泛化能力越强;一致性得分反映主题内高频词的语义一致性,得分越高说明主题内聚性越好。

如图 1 所示,通过对 8175 篇报道的分析,最终确定最优主题数为 K=10,表明外媒对陕西的报道大致聚焦于 10 个主题。每个主题的关键词存在差异,需要进一步对关键词内容进行深度分析,以明确主题内涵。经反复训练与参数优化,本研究所构建的 LDA 模型在困惑度与一致性得分方面均表现良好,能够有效识别并提取外媒报道中的主题结构,为理解陕西形象的国际传播特征提供了坚实的数据基础。

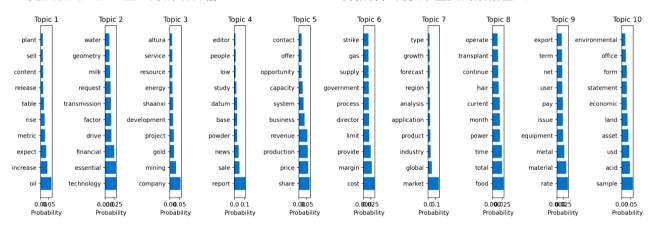


图 1 LDA 聚类主题分布图

表 1 LDA 结果及编码结果

主题序号	主题命名	一级类别	核心关键词	关键词(英文)
1	资源开发与工业根基	工业生产与市场行为	工厂、销售、发布、油、指标	Oil, increase, metric, sell, plant
3		矿业资源开发	资源、黄金、采矿、能源、 陕西	Company, mining, gold, project, development, energy, resource, service
2	陕西地域特色产业	陕西本地资源	水、奶、运输、工厂	Technology, financial, factor, transmission, request, milk, geometry, water
4		产业研究与行业调研	评估、研究、销售、交易、 信息	Report, sale, news, powder, base, study, editor
5	一带一路与区域企业国际化战略	企业运营与商业交易	份额、价格、供应、产能、 收入、合同	Share, price, production, revenue, business, system, capacity, opportunity, offer, contact
6		行业监管与外部影响	成本、利润、气、政府、 供应、限制、冲击	Cost, margin, limit, process, government, supply, gas, strike
7		产业趋势与国际市场	市场、全球、产业、地域、 预测、分析、类型	Market, global, industry, product, application, analysis, region, forecast, growth, type
8		产业运营与转型	运营、转型、持续、能源	Food, total, power, current, continue, transplant, operate
9		国际贸易与金属产业	出口、条款、金属、设备	Rate, material, metal, equipment, issue, pay, term, export
10		环境与产业经济	环境、机构、声明、经济、 土地	Sample, acid, asset, land, economic, statement, form, office, environmental

基于 LDA 模型分析结果,本研究识别出 10 个主 题,并通过人工编码总结出外媒关于"Shaanxi"报道 的三个核心主题:资源开发与工业基础、陕西地域特色 产业以及一带一路背景下的区域企业国际化战略。而 涉及旅游、文化、兵马俑、大雁塔等关键词的报道仅有 271 篇, 占总样本的 3.3%。这一结果表明, 境外媒体 关注的主题与陕西官方主导的文化导向对外官传策略 存在明显偏离。造成这一差异的原因可能包括:一方面, 外媒关注更多的是经济与产业相关内容,以满足其受 众对投资、贸易和产业发展的信息需求;另一方面,文 化和旅游主题尽管是官方重点,但在国际新闻中相对 碎片化,缺乏连续报道的新闻价值或受众兴趣,从而导 致在报道数量上占比偏低。因此,在LDA聚类分析中, 本研究未将文化相关主题纳入主要聚类范围,但这一 差异本身也反映了境外媒体报道与地方外宣策略之间 的结构性偏差, 为后续研究提供了值得关注的对比视 角。

3.1 聚类 1: 资源开发与工业根基

在近二十年的境外媒体报道中,陕西的资源开发和工业发展一直是关注的重点之一。作为中国内陆典型的资源富集型地区,陕西的能源与矿产资源构成了区域重工业体系的基础,形成以"资源开采一工业生产一价值转化"为核心的纵向产业链条。这一链条不仅涵盖矿产与能源资源的初级开发(Topic 3),还延伸至深度转化及工业产品制造(Topic 1),实现了从原材料输入到市场销售的全流程管控,体现出资源型经济向制造业主导型经济的战略转型。

与中东、澳大利亚等主要依赖资源出口的发展路径不同,陕西展现出较高的产业内嵌性与产业链自主性,形成资源一制造闭环的战略特征。这一模式不仅强化了国家能源安全,还通过高效整合上下游资源、制造与市场体系,实现对产业价值链关键环节的掌控。与此同时,陕西在传统矿产开采中积极引入科技创新、生态修复及文旅融合理念,逐步摆脱"资源一开采一废弃"的线性路径,探索"多元价值释放一绿色可持续发展"的转型模式。例如,通过"绿色矿山"建设、废弃矿区生态重构及文化资源再利用等方式,推动矿业经济向"生态一经济一社会"多重协同发展。

在"一带一路"建设背景下,陕西获得了更多国际 媒体关注。尽管矿产资源并非官方宣传的主要主题,但 外媒在关注经济和投资信息时,更侧重于基础资源建 设与产业发展。在国家"双碳"战略背景下,陕西依托 传统能源基础,积极推进能源结构绿色转型,加快清洁 能源消费占比,努力构建以煤为基础、多元互补、安全 高效的新型能源体系。这种"立足传统、融入创新"的 发展路径,体现了中国式现代化进程中产业政策"先立 后破、以新促转"的系统性谋划。

综上所述,陕西的资源开发与工业基础建设不仅体现了产业双轨驱动的发展策略:一方面通过资源一制造闭环保障国家能源与工业安全,另一方面以资源多元化开发推动可持续发展;同时,在新发展格局下,展现了资源型地区在"新旧动能转换""绿色低碳转型"和"高质量发展"中的实践价值与投资潜力。因此,境外媒体对陕西资源开发和工业基础的聚焦,既反映了其新闻关注热点,也为全球资源密集型地区提供了中国路径的参考,同时吸引更多境外企业对陕西相关产业的投资与合作。

3.2 聚类 2: 陕西地域特色产业

核心主题 2 是陕西的地域特色产业。陕西作为中 国西部的重要省份,在资源禀赋、产业基础与区位优势 上具有鲜明特色。境外媒体所关注的乳制品产业(特别 是羊乳)、水资源调控及跨境运输通道建设等领域不仅 构成了陕西产业体系的重要支撑, 也在一定程度上决 定了其未来产业升级与开放格局。陕西奶业最显著的 特征在于羊乳优势。关中平原以富平为核心的奶山羊 养殖区,已形成较为完整的"种羊繁育—规模化养殖— 乳品加工—市场品牌"产业链。陕西奶山羊存栏量和鲜 羊乳产量居全国首位,被誉为"中国羊乳之都"。企业 层面,银桥、百跃、雅泰等龙头企业带动了乳品加工与 品牌建设,推动羊乳粉、功能性乳制品走向高端化与差 异化。尽管羊乳在陕西乳业中占据主导,但牛乳产业也 在稳步发展。陕西统计年鉴数据显示,2023年全省生 牛奶产量同比增长超过1%,部分地市增速更快。另外, 陕西原奶品质与加工标准与国际标准接轨的问题,品 牌效应和地理标志保护不足的问题也限制了陕西乳品 在全国和国际市场的竞争力。外媒关于陕西的报道中, 也部分提到了这些亟需解决的问题。与此同时,随着消 费升级与跨境市场拓展, 羊乳粉和功能化乳制品被认 为是陕西乳业实现差异化竞争的突破口。除了乳制品 以外,陕西水资源产业发展也是外媒关注的核心。虽然 陕西总体水资源紧缺, 人均占有量远低于全国平均水 平,且呈现显著的时空分布差异,但陕南相对丰富。"引 汉济渭"工程被认为是缓解关中城市群及产业用水压 力的关键性举措,同时该工程已对西安、咸阳等地的工 业与民生供水产生显著效益,并间接保障了乳业加工 与饲草种植的水源安全。同时,调水工程与节水灌溉、

再生水利用相结合,能够为陕西奶业提供长期的水资源支撑。部分研究尝试构建"水—草—畜—乳"链条模型,强调通过节水灌溉技术、非常规水资源利用与水价机制改革,提升牧草单产与质量,进而稳定奶源供给。这为陕西奶业的可持续发展提供了政策与技术层面的启示。

此外,伴随"一带一路"建设的发展,中欧班列的建成通车视为陕西产业外向化的重要依托。2024年西安中欧班列开行量居全国首位,累计突破3万列,占全国总量的四分之一。西安—杜伊斯堡全程时刻表班列的开行,将运输时效稳定在12天左右,大幅提升了陕西对欧亚市场的连通度。铁路运输相较航空和公路在成本与碳排放上更具优势,而奶粉等产品对冷链依赖度较低,更适合通过班列规模化输出。目前,西安已逐步形成"国际班列+航空快线+公铁联运"的多式联运格局,为乳品及其上下游产业(如包装、装备)提供了高效物流网络。外媒也指出,陕西应进一步建设境外节点的乳品保税仓与检疫通道,以降低通关与资金周转成本。

综上所述,陕西地域特色产业的发展是"资源禀赋—产业结构—区位优势"综合作用的结果。奶业以羊乳为核心,水资源短缺通过跨流域调水与节水技术得到缓解,运输优势则依托中欧班列得以放大。三者相互耦合,为陕西在全国乳业版图与"一带一路"市场中塑造差异化竞争优势提供了条件。文献的整体趋势显示,未来陕西应在保障水资源、提升产业标准化与拓展国际市场之间寻求平衡,实现特色产业的高质量发展。这也让更多的境外媒体看到了陕西经济产业发展的突破点。

3.3 聚类 3: "一带一路"与区域企业国际化战略在"一带一路"倡议背景下,外媒报道中陕西的最后一个聚焦主题涉及数字化驱动、供应链韧性、全球拓展、政策调适与绿色转型。当前,陕西产业生态呈现高度组织化与多维协同的发展格局,成为中国内陆地区引资潜力最强、产业体系最具韧性的区域之一。

在企业运营机制层面,陕西表现出明显的数据驱动导向与系统性策略规划能力。LDA模型中 Topic4 所揭示的 editor(评估)一study(研究)一sale(销售)链条,反映能源、矿业等传统行业积极融入数字化技术,通过精准市场调研与数据分析优化资源配置,指导销售策略,实现从"产什么"到"为谁而产"的结构性转变。这种以信息为基础的决策机制,被外媒视为"一带一路"背景下传统产业主动拥抱数字化、提升决策效率与市场响应速度的典型案例。

综合 LDA 主题分析结果可见,陕西企业通过"数据驱动一合同稳定一绿色转型一全球布局"的多维策略,构建了高效、稳健且具国际拓展性的产业生态体系。这一体系不仅保障区域经济韧性与政策适应力,还为外资提供了从传统能源到绿色科技、从资源输出到价值共创的投资路径。不同于欧美"去工业化+高金融依赖"模式,也不同于部分发展中地区"资源依赖+低附加值"格局,陕西正以中国特色的产业逻辑探索"产业升级+全球治理"双轨战略:以 Topic4、5、6 为内核,强化产业运营的精细化与风险管理能力;以 Topic7、8、9 为前沿,开拓全球市场空间并融入国际规则体系;以 Topic10 为支点,推动绿色转型与产业可持续发展,从而形成三重战略协同支撑。

4 研究结论

本研究基于 2005-2025 年境外媒体对陕西的报道,通过关键词提取与 LDA 主题分析,揭示了外媒关注的主要内容及其与区域发展实践之间的关系。分析结果显示,外媒报道的高频主题主要集中在资源开发、工业制造、企业运营与全球化、绿色转型等领域。相比之下,官方宣传中主打的文化旅游、历史文化和旅游资源形象在外媒报道中占比较低,仅有少量涉及文化类主题的报道,未进入 LDA 主题聚类。这表明外媒的关注重点更多体现产业发展逻辑和投资价值判断,而非文化标签。

外媒报道的主题与陕西官方对外宣传存在一定差异,但在产业与投资议题上与区域经济发展趋势表现出明显契合。例如,陕西的资源型经济基础与现代产业体系建设得到了持续关注,包括能源与矿产开发、工业转化、市场销售链条的纵向整合,以及数字化驱动的企业运营模式和绿色转型实践。此外,"一带一路"建设下的全球化企业运营和国际投资布局,也成为外媒报道的核心内容,反映出陕西作为区域枢纽和产业高地的战略价值。

从区域发展视角来看,外媒报道与陕西实际发展路径高度一致。陕西通过"资源主导—制造牵引—创新驱动—绿色转型"的发展模式,构建了高韧性、多维协同的产业生态体系。企业在数据驱动、供应链稳控、绿色转型及全球市场拓展等方面的实践,不仅保障了区域经济韧性与政策适应力,也为外资提供了从传统能源向绿色科技、从资源输出向价值共创的投资路径。伴随"一带一路"建设,陕西逐步从中国西部资源腹地迈向全球绿色产业价值链的重要节点,形成可复制的内陆发展模式。

综上,外媒报道不仅反映了陕西在资源开发、产业建设和企业运营方面的特色实践,也呈现出与区域发展趋势高度契合的内容结构。虽然文化旅游类形象已深入人心,但经济与产业议题正成为外媒报道关注的主要方向。这提示陕西未来在国际传播中,应在保持文化品牌优势的同时,更加注重经济与产业层面的对外宣传,提升全球经济地位与投资吸引力,实现多元化宣传。进一步的政策方向包括打造绿色标杆、强化自主创新能力、培育高附加值产业链条,并以国际标准对接本地实践,推动资源高地向价值高地、产业洼地向投资高地的转型飞跃。

参考文献

- [1] 林墨丞,赵会军.基于语言模型的地区文化外宣新闻热点 趋势分析——以福建为例[J].外语电化教学,2024,(05): 35-42+110.
- [2] 王曰芬,许杜娟,杨振怡,等.與情评论与新闻报道的话题 识别及其主题关联分析[J].现代情报,2018,38(06):3-10.
- [3] 陕西省人民政府. 陕西省政府政策例行吹风会 (第九场) [EB/OL].(2023-09-25) [2024-06-11]. https://www.shaanxi.gov.cn/szf/xwfbh/202309/t20230925 2301719.html.
- [4] 王瑞媛,刘燕梅.海外社交媒体中西安城市形象研究——基于优兔平台的考察[J].新闻研究导刊,2023,14(08):52-

55.

- [5] 梁琳琳,王宏俐.城市形象跨文化传播策略研究——以西安为例[J].对外传播,2021,(11):57-60.
- [6] 刘丹.语料库视域下西安城市形象对比研究[J].新楚文化,2024,(22):78-82.
- [7] 厉文芳,汪顺玉.东盟媒介中的西安城市形象——基于 LexisNexis 新闻数据库的文本挖掘[J].新闻知识,2021, (10):25-30.
- [8] Zhu Yicheng. Determining economic news about China in global news feed: Evidence from Global Database of Events, Language and Tone [J]. Global Media and China, 2019, 4(2): 1-14.
- [9] 程盈莹,成东申,逯建.国际舆论对中国引进外资的影响: 基于 Gdelt 新闻大数据的实证研究[J].世界经济研究, 2021,(07):19-33+135.
- [10] 成东申.国际舆论对中国引进外资的影响研究[D].西华大学,2022.
- [11] 晏艳阳,汤会登.东道国媒体情绪对中国企业跨境并购的 影响研究[J].国际贸易问题,2023,(01):158-174.

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC) 所有。本文章按照知识共享署名许可条款发表。

http://creativecommons.org/licenses/by/4.0/

