

基于LDA主题模型的乡村人才政策演化研究

袁婷, 徐馨雨, 曾宇, 张思忆, 黄婷

成都信息工程大学统计学院, 四川 成都

收稿日期: 2026年3月8日; 录用日期: 2026年4月20日; 发布日期: 2026年4月28日

摘要

在城乡融合加快推进背景下, 乡村人才流失与结构失衡日益突出, 乡村人才是乡村振兴的核心资源。为探究我国乡村人才政策的重点内容与演化规律, 文章基于2016~2025年公开发布的37份乡村人才政策文本, 采用LDA主题模型进行主题分析, 识别出“培养提升、政府管理、基础设施、人才就业、社会保障”五大核心主题, 并分析其时序演化规律。研究发现: 1) 政策主题结构多元, 呈现全链条特征; 2) 主题热度演化表现出动态互补与阶段性转换; 3) 部分主题热度波动剧烈, 存在“重培养轻服务”的政策执行短板。文章提出政策协同机制、执行优化路径及动态监测体系建设等建议, 以提升政策的系统性与精准性。

关键词

人才吸引力, 政策要素, LDA主题模型, 文本挖掘

Research on the Evolution of Rural Talent Policies Based on the LDA Topic Model

Ting Yuan, Xinyu Xu, Yu Zeng, Siyi Zhang, Ting Huang

School of Statistics, Chengdu University of Information Technology, Chengdu Sichuan

Received: March 8, 2026; accepted: April 20, 2026; published: April 28, 2026

Abstract

Amid the accelerated integration of urban and rural areas, the loss of rural talent and structural imbalances have become increasingly prominent, with rural talent serving as a core resource for rural revitalization. To explore the key content and evolutionary patterns of China's rural talent policies, this study analyzes 37 publicly available rural talent policy documents issued between 2016 and 2025. Using the LDA topic model for thematic analysis, it identifies five core themes—“Training and Development, Government Management, Infrastructure, Talent Employment, and

Social Security” —and examines their temporal evolution. The findings reveal that: 1) the policy theme structure is diverse, exhibiting a full-chain characteristic; 2) the evolution of thematic prominence demonstrates dynamic complementarity and phased transitions; and 3) some themes experience significant fluctuations in prominence, revealing a policy implementation shortfall characterized by “emphasizing training while neglecting services.” Accordingly, this paper proposes recommendations in three areas—policy coordination mechanisms, implementation optimization pathways, and dynamic monitoring system construction—to enhance the systemic nature and precision of policies.

Keywords

Talent Attractiveness, Policy Elements, LDA Topic Model, Text Mining

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

在全球知识经济竞争日益激烈的背景下,人才资源已成为推动区域创新与高质量发展的核心。自2010年《国家中长期人才发展规划纲要(2010~2020年)》发布以来,各省份相继将人才政策纳入地方发展战略,力图吸引和留住高层次人才[1]。然而,现有研究指出,我国人才政策存在政策工具结构不完善、子系统不健全等问题[2]。亟需通过系统化政策文本分析,探索人才吸引力的长效提升路径。

近年来,人才政策研究渐成热点,涌现出多种定性定量方法。马一鸣等发现住房补贴、税收优惠等经济手段短期有效,但长期效果受限于公共服务配套[3]。宋月萍等指出,低收入城市中“宽松落户+高生活补助”组合能有效吸引高校毕业生[4]。李晓萍等证实区域人才政策可提升创新绩效,政策效度越高效果越好[5]。以东北三省为例,苗宏慧等发现其科技人才政策虽已形成引进、培养、使用、评价、激励、服务主线体系,但在政策供给、需求牵引、针对性等方面仍存不足[6]。

现有研究存在三方面不足:缺乏全国政策文本的系统性量化分析、政策主题动态演化规律研究不足、传统分析方法难以深入挖掘大规模非结构化文本。文本挖掘技术为政策分析提供了新路径,LDA主题模型作为无监督机器学习方法,能从大规模文本中提取隐含主题,已广泛应用于多领域[7]。王慧英等运用LDA模型指出,应推动科技创新人才政策与多元化创新政策耦合,构建系统性人才发展框架[8]。基于此,本研究基于2016~2025年公开发布的37份乡村人才政策文本,采用LDA主题模型进行主题分析,提取文本隐含主题。

乡村人才被视为乡村振兴的核心路径,对实现农业强、乡村美、农民富的目标至关重要。当前乡村人才外流、结构失衡、老龄化问题严峻[9],存在吸引力不足、专业人才缺乏、激励保障不到位、培养水平不高等问题[10]。以湖南为例,人才流失严重、激励机制不健全等问题突出[11]。曹中秋提出“培引结合”双轨策略,通过本土培育与外来引进协同发力[12]。此外,优化农村发展环境、完善教育培训体系、构建科学化管理机制成为普遍共识[13]。

为促进农村实用人才队伍建设,需建立新型乡土人才认定体系,科学认可人才贡献[14]。肖国俊指出农村干部人才利用不足[15],王潇仪运用史密斯模型分析政策执行问题[16],钱再见等倡导多元化“人才下乡”渠道吸引“新乡贤”回流[17],黄梅呼吁打破城乡分割、增强政策协同[18],田书芹等强调制度供给与整体治理的核心驱动力[19]。

综上，乡村人才吸引政策研究已成为学界和实践的热点。研究成果为政策制定提供了丰富经验，但仍需进一步深化区域差异研究，完善政策执行与评估体系。结合大数据与人工智能技术，运用 LDA 等文本挖掘方法对政策文本进行系统分析，将为优化人才政策设计、促进城乡融合发展提供科学支撑与理论依据。

2. 研究思路及方法

2.1. 研究思路

本文主要研究步骤如下：(1) 收集有关乡村人才的政策报告；(2) 对数据进行分词、去停用词、文本向量化等预处理；(3) 根据一致性得分与困惑度确定主题数，计算各主题热度值；(4) 对主题进行分析和阐述，总结启示和应对措施(图 1)。

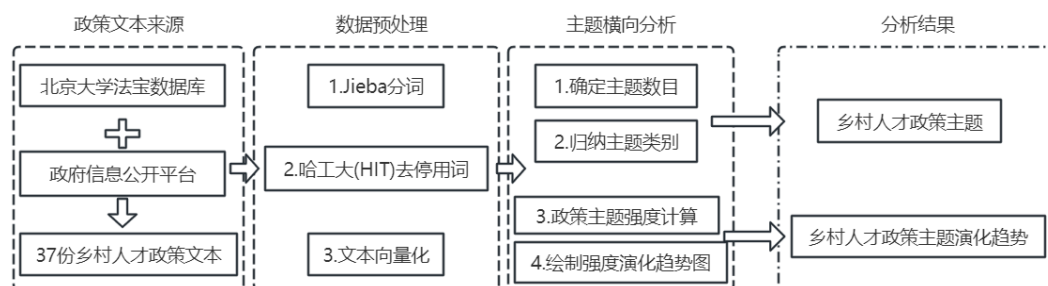


Figure 1. Research framework

图 1. 研究思路

2.2. 研究方法

2.2.1. LDA 主题模型

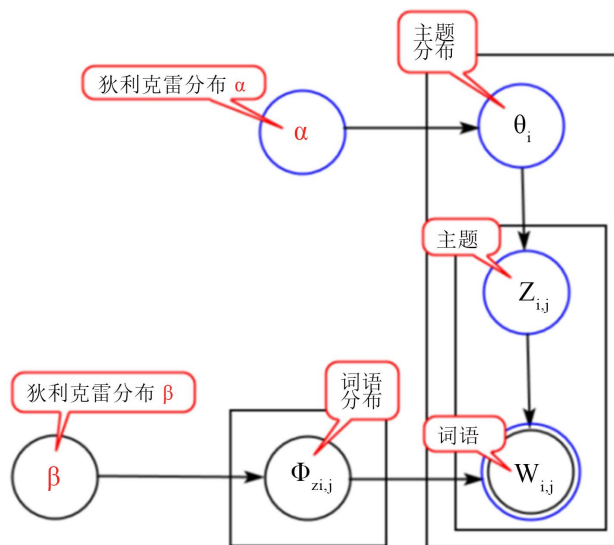


Figure 2. Framework diagram of the LDA Topic model

图 2. LDA 主题模型框架图

LDA (Latent Dirichlet Allocation)是一种基于贝叶斯网络的概率主题模型，广泛应用于非监督式语义特征提取，其模型框架如图 2 所示。

2.2.2. 模型参数

为评估不同主题数量下 LDA 模型的性能, 确定模型的最优主题数量, 本文选择一致性得分和困惑度指标衡量主题间相似度, 一致性得分的最大值和困惑度的最小值对应的主题是最佳主题数目。其中, 困惑度公式如下所示:

$$\text{Perplexity}(D) = \exp \left\{ \frac{-\sum_{d=1}^N \log p(w_d)}{\sum_{d=1}^N N_d} \right\} \quad (1)$$

注: D 为语料库中的训练集, N 为文档总数, K 为主题数量, θ_d 为 d 的主题分布, N_d 为文档 d 中单词数, 为任意文档 d 中的词。

主题显著性(Topic Significance), 是衡量特定概念在时间窗口内受学术界/社会舆论关注的量化指标, 在一定时间内包含该主题的文本数量越多, 意味着主题热度越大, 该主题则有概率成为热点主题[20]。主题热度的计算公式为:

$$P_k = \frac{\sum_{i=1}^N \theta_{ki}}{N} \quad (2)$$

注: P_k 为第 k 个主题热度, N 为文本数, θ_{ki} 为第 k 个主题在第 i 篇文本中概率。

3. 数据收集及预处理

3.1. 数据收集

本研究基于政策文本分析法, 采集 2016~2025 年间 “乡村人才” 相关的政策文件。数据来源包括北京大学法宝数据库和政府信息公开平台, 通过 “乡村”、“人才”、“引进” 等核心关键词组合检索, 确保政策文本与研究主题的高度契合。文献筛选遵循三重标准: 内容聚焦乡村人才引进措施及保障; 效力层级优先选择中共中央、国务院纲领性文件及农业农村部、人社部等部委专项政策; 时效性截取 2025 年 2 月 1 日前有效文件, 剔除废止或重复政策。最终纳入 37 份乡村人才政策文本, 如附录所示。

3.2. 数据预处理

本文首先, 采用基于深度学习的 Jieba 分词工具对政策文本进行分词处理, 将连续的字符序列切分为具有语言学意义的词序列。其次, 依据哈工大(HIT)中文停用词表, 剔除高频虚词、标点符号等无实际语义贡献的停用词, 以降低噪声对关键词频统计的干扰。最后, 通过 LDA 主题模型构建语料库并生成词典, 借助词袋模型将非结构化的文本数据转换为数值形式的向量表示。由此, 原始文本被转化为可供机器学习算法直接处理的结构化数值特征。

3.3. 最优主题数确定

为防止主题数过多造成过拟合, 本文在 1 至 14 的范围内遍历不同主题数, 并绘制主题数与一致性得分(左)以及主题数与困惑度(右)的曲线图, 如图 3 所示。

由图 3, 随着主题数的增大, 一致性得分逐渐减小, 困惑度逐渐增大。当 $K=3$ 时, 存在一个显著的拐点, 一致性得分取得极大值, 困惑度取最小值, 但是主题数过少, 可能存在欠拟合情况; 当主题数为 5 时, 一致性得分与困惑度都表现较为优异。绘制 $K=5$ 的主题-词分布和文档-主题分布, 如图 4 所示, 五个主题在图中分布均衡, 未出现重叠现象, 且各圆心间距离较大, 表明这些主题之间, 具有较

好的区分度。因此，本文选择 $K = 5$ 作为最优主题数。通过主题可视化图，可知主题 1 聚类效果较好。

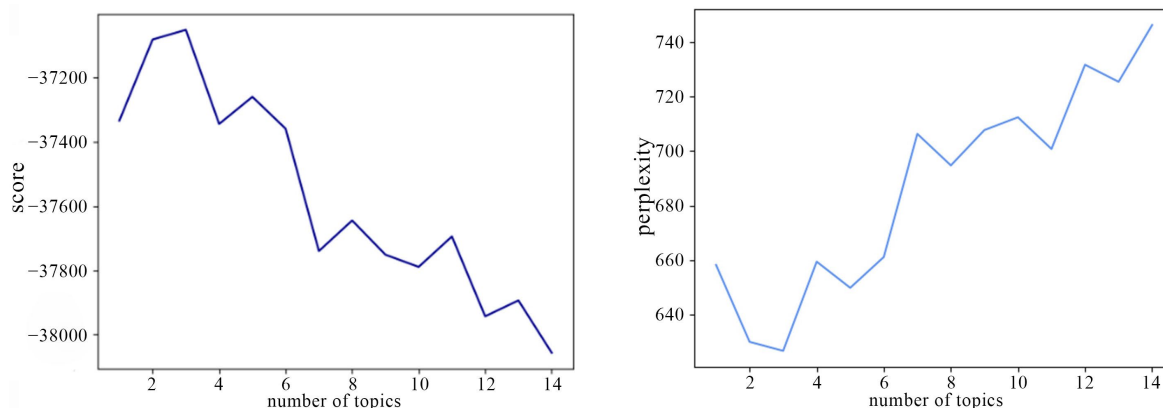


Figure 3. Coherence score (Left) and perplexity (Right) curves
图 3. 一致性得分(左)和困惑度(右)曲线

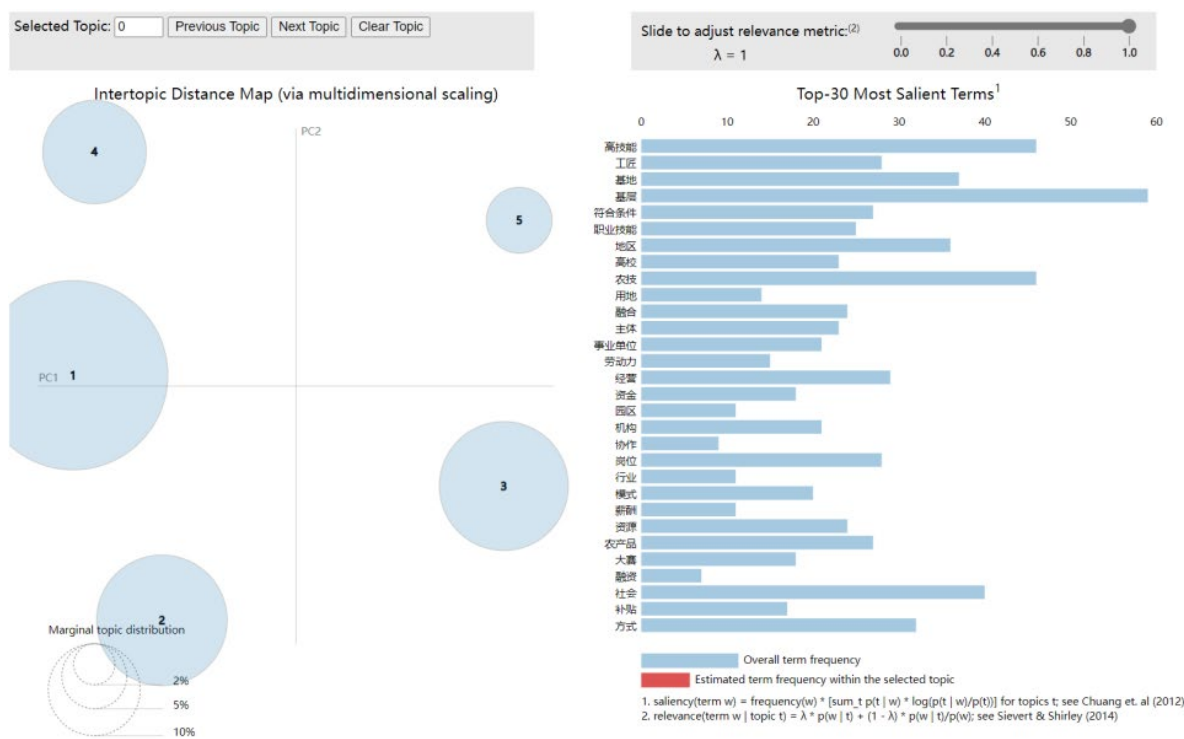


Figure 4. Policy topic bag-of-words (number of topics = 4)
图 4. 主题数为 4 的政策主题词袋

4. LDA 主题建模结果

4.1. 主题横向分析

基于前文确定的最优主题数量 $K = 5$ 构建 LDA 主题模型，获得“主题 - 特征词”以及“文档 - 主题”的概率分布情况。针对“主题 - 特征词”分布，从每个主题里挑选出词频位于前 30 位的词语，将这些特征词作为该主题的关键词。通过对各个主题的特征词进行分析，判断主题的具体内容，归纳出每个主题

的具体类别。经过识别与分析，最终确定 5 个潜在主题，具体内容如下表 3 所示。

Table 3. LDA topic model and keywords

表 3. LDA 主题模型及关键词

主题	建模结果	主题名称
Topic 1 关键词	基地、人才队伍、农技、管理、科技、能力、平台、力度、培训班、融合、重点、农产品、社会、合作、全国、质量、改革、指导、企业、经营、制度、模式、信息、方式、大力	培养提升
Topic 2 关键词	高技能、技能、地区、教师、社会、重点、企业、力度、制度、高校、岗位、评价、学校、基层、工匠、符合条件、能力、劳动力、管理、工资、部门、科研人员、事业单位、比例、奖励	政府管理
Topic 3 关键词	科技、基层、乡镇、体育、人才队伍、医生、大学生、企业、行动、制度、专业人才、特派员、农技、经营、主体、医疗卫生、方式、队伍、符合条件、文化、体育教师、法律、社会、机构、优化	基础设施
Topic4 关键词	制度、事业单位、改革、大力、职业技能、科技、技能、岗位、高技能、人才队伍、行动、人力资源、国家、乡镇、大赛、战略、领域、企业、精神、基层、头雁、着力、体制、工作人员	人才就业
Topic 5 关键词	工匠、企业、补贴、技能、队伍、符合条件、贷款、基地、农民工、管理、人口、城乡、机构、体系、主体、实训、制度、优化、载体、条件、情况、落户、科技人员、名录、用工	社会保障

4.2. 主题热度分析

为探究新时代我国乡村人才主题热度演变历程，本文针对基于 LDA 模型生成的主题展开演化分析。在研究过程中，采用按时间后离散式的方法进行主题热度演变分析，最终得到我国乡村人才政策逐年主题热度分布情况，如表 4 所示。

Table 4. Temporal distribution of theme intensity allocation in rural talent policies

表 4. 乡村人才政策主题热度配置时序分布

	Topic1	Topic2	Topic3	Topic4	Topic5
2016	0.494627	0.499824	0.005549	0.000000	0.000000
2017	0.000000	0.000000	0.499965	0.000000	0.500036
2018	0.009514	0.000000	0.000000	0.340850	0.647793
2019	0.250365	0.250172	0.000000	0.244838	0.252303
2020	0.500793	0.000000	0.000000	0.000000	0.499208
2021	0.143925	0.559594	0.146011	0.145591	0.003576
2022	0.735724	0.000000	0.255527	0.007906	0.000000
2023	0.398579	0.000000	0.196693	0.003065	0.397538
2024	0.000000	0.000000	0.299679	0.694188	0.006133
2025	0.319351	0.014243	0.659258	0.000000	0.000000

依据注意力理论可知，决策者在特定时段内的注意力属于稀缺资源，这就导致其无法对政策的所有

主题同时予以同等程度的关注[21]。以时间切片作为横轴、主题热度当作纵轴，绘制主题热度折线图，如图5所示。

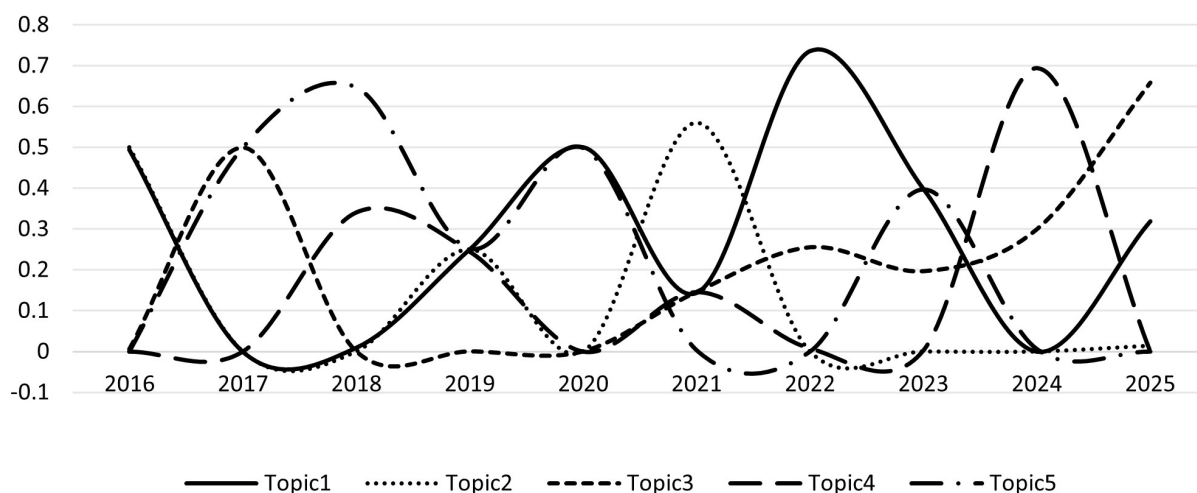


Figure 5. Evolution Chart of Theme Intensity in Rural Talent Policies
图5. 乡村人才政策主题热度演进图

4.2.1. 基础设施主题前期呈现大幅波动，中后期缓慢上升

从图5可知，基础设施主题2016年热度较低(0.0055)，2017年因《“十三五”推进基本公共服务均等化规划》发布迅速升至0.5，成为政策焦点，通过改善校舍、医疗设备、实训基地等为人才引进提供硬件支撑。此后政策重心向人才培养转移，主题热度回落。2021年《关于加快推进乡村人才振兴的意见》推动议题升温，2022年回升至0.26。2023~2025年，随着《关于推进体育助力乡村振兴工作的指导意见》《乡村全面振兴规划(2024—2027年)》等政策出台，2025年达峰值0.66，政策从“硬件建设”向“设施+服务”协同发展。

4.2.2. 政府管理主题前期具有较高的注意力热度，后期热度较低

2016年政府管理主题热度接近0.50，得益于《关于激发重点群体活力带动城乡居民增收的实施意见》等政策对治理机制的强化，推动教育、农业、人社等部门协同，提升资源配置效率。2017~2020年主题热度在0.25左右波动，持续关注人才评价、岗位管理等议题。2021年《关于加强国家乡村振兴重点帮扶县人力资源社会保障帮扶工作的意见》发布，主题短暂回升至0.56峰值，此后持续下降，2023年降至0.00，2025年仅回升至0.01，暴露出“重培养轻管理”的结构性问题，政策注意力向“产出端”倾斜，治理机制逐渐边缘化。

4.2.3. 培养提升主题与人才就业主题变化趋势形成“此消彼长”的动态平衡

培养提升主题长期处于高位，与人才就业主题呈现明显“此消彼长”关系。2016~2017年，《关于支持返乡下乡人员创新创业促进农村一二三产业融合发展的意见》等政策聚焦人才技能提升。2018年培养主题热度仅0.01，2019年回升至0.25，2020年达0.50。2022年，随着《关于做好2022年农村综合性改革试点试验申报工作的通知》等出台，培养主题升至0.74峰值，人才就业主题仅0.01，反映政策向人才储备倾斜。2022~2024年，随着培养成果显现，政策重心向就业安置转移。2023年培养主题回落至0.40，2024年降至0.00，人才就业主题从2022年的0.01持续升至2024年的0.69，呈现“储备-输出”阶段性转换。2025年，随着《农业农村部落实乡村振兴工作部署的实施意见》发布，培养主题回升至0.32，新

一轮培养周期开启，体现“培引结合”策略的动态调整。

4.2.4. 社会保障主题的演化轨迹呈现双重特征：剧烈震荡以及阶梯式下行

社会保障主题呈现“前期波动剧烈、后期阶梯下行”特征。2016~2017年热度接近于0，2018年因《财政部贯彻落实实施乡村振兴战略的意见》升至0.65峰值。2019~2021年，《关于促进乡村产业振兴的指导意见》等政策持续规划社会保障事务，2020年达0.50次高峰。此后剧烈波动：2021年骤降至0.004，2022年为0.00，2023年小幅回升至0.40，2024年再度降至0.01，2025年为0.00。这种“脉冲式”波动反映城乡权益衔接机制不健全，政策延续性不足，可能削弱人才长期留任意愿，亟需建立长效机制。

5. 结论与建议

本研究基于2016~2025年中国大陆37份乡村人才政策文本，通过LDA主题模型与文本挖掘方法，确定最优主题数 $K=5$ ，系统揭示政策主题结构及时序演化特征。研究发现：政策内容聚焦于“培养提升-政府管理-基础设施-人才就业-社会保障”五大主题。主题间呈现动态互补特征，培养提升与人才就业主题“此消彼长”；基础设施主题后期热度提升，凸显“硬件+服务”协同导向；政策执行层面存在结构性短板，政府管理主题后期热度趋零；社会保障主题热度震荡，折射城乡权益衔接机制与政策延续性不足。基于以上发现，提出如下政策建议：

第一，强化政策协同机制，构建“培养-就业-保障”一体化政策链条。针对政策主题间“此消彼长”的互补特征，需打破部门壁垒，统筹人才培养、岗位开发与社会保障。推动产教融合，支持高校与乡村产业基地联合开设“订单班”，实现精准育才。建设省级乡村人才信息平台，提升人岗匹配效率。探索社保双轨衔接机制，允许返乡人才保留原城市社保权益，同时在属地享受基本公共服务，降低跨区域流动成本。

第二，优化执行机制，增强政策落实效能。针对“重培养轻服务”现象，应改善制度供给与基层治理能力。在乡镇层面设立“人才服务窗口”，整合多部门职能，提供政策咨询、待遇兑现等一站式服务。完善财政激励机制，对企业吸纳乡村人才给予税收减免，鼓励社会组织参与公共服务。设立乡村人才专项基金，聚焦医疗、住房等民生领域，建立补贴动态调整机制，确保保障稳定可持续。

第三，建立动态监测体系，推动政策从“经验驱动”向“数据驱动”转型。针对主题热度波动与创新议题识别不足问题，应推动政策监测由“经验驱动”向“数据驱动”转型。依托自然语言处理技术，搭建政策主题演化监测平台，实时绘制主题热度曲线，动态评估政策效果。推动分区域差异化施策，根据东中西部发展阶段与人才需求差异，分类定制“政策工具包”。持续追踪“科技成果入股”“城乡权益衔接”等隐性议题，推动制度从被动应对向主动布局演进。

综上，乡村人才政策亟需在主题协同、执行优化与数据驱动之间形成闭环机制，通过全链条设计与多主体参与，推动政策从“规模扩张”向“结构优化”与“质量提升”转型，实现“引得进、留得住、用得好”的目标，为乡村振兴战略提供持续的人才支撑。

基金项目

成都信息工程大学大学生创新创业训练计划项目资助，成都信息工程大学大学生创新创业训练计划项目(202410621030)。

参考文献

- [1] 人民出版社，编. 国家中长期人才发展规划纲要 2010-2020 年[M]. 北京：人民出版社，2010.
- [2] 谭雨婷，岑晓潼，刘艳，等. 人才政策分析[J]. 合作经济与科技，2024(10): 78-79.

- [3] 马一鸣, 祁勇. 中国人才政策研究的现状与趋势——基于 Cite Space 的可视化分析[J]. 北京城市学院学报, 2024(3): 87-92.
- [4] 宋月萍, 王若霖. 城市吸引高校应届毕业生的政策组合及效果——基于模糊集定性比较分析[J]. 西北人口, 2023, 44(3): 27-37
- [5] 李晓萍, 赵浩云, 江飞涛, 等. 中国区域人才政策与区域创新绩效——基于 3171 份政策文本的创新测度与实证检验[J]. 财经研究, 2024, 50(9): 4-18.
- [6] 苗宏慧, 全情爽, 舒心. 基于 BERTopic 的科技人才政策文本主题识别与量化分析——以东北三省为例[J]. 现代情报, 2025, 45(4): 110-121.
- [7] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*, **3**, 993-1022.
- [8] 王慧英, 向蓓珊. 中国科技创新人才政策注意力演变与整合转型研究[J/OL]. 科技进步与对策: 1-11. <http://kns.cnki.net/kcms/detail/42.1224.G3.20250123.1539.019.html>, 2025-04-11.
- [9] 韩克勇, 孟维福, 汪小愉. 日本乡村振兴发展模式创新的经验与启示[J]. 江西师范大学学报(哲学社会科学版), 2022, 55(4): 74-82.
- [10] 张静宜, 陈洁. 强化乡村人才支撑有效供给实现脱贫攻坚乡村振兴有效衔接[J]. 宏观经济管理, 2021(8): 54-60.
- [11] 余剑平. 乡村振兴背景下农村乡土人才队伍建设刍议——以湖南省 HD 县为例[J]. 南方农机, 2020, 51(17): 24-25, 28.
- [12] 曹中秋. 打造人才引擎助力乡村振兴[J]. 人民论坛, 2019(23): 70-71.
- [13] 张亚玉. 乡村振兴视阈下农村实用人才队伍建设研究[J]. 科学大众(科学教育), 2018(7): 150-151.
- [14] 冯丹丹. 农村实用人才队伍建设思考[J]. 理论月刊, 2010(12): 172-175.
- [15] 肖国俊. 乡村振兴战略背景下农村干部队伍建设研究[D]: [硕士学位论文]. 广州: 华南理工大学, 2018.
- [16] 王潇仪. S 市农村实用人才政策的执行问题研究[D]: [硕士学位论文]. 沈阳: 辽宁大学, 2019.
- [17] 钱再见, 汪家焰. “人才下乡”: 新乡贤助力乡村振兴的人才流入机制研究——基于江苏省 L 市 G 区的调研分析[J]. 中国行政管理, 2019(2): 92-97
- [18] 黄梅. 乡村振兴战略中人才资源配置机制与策略研究——基于职业结构变迁的视角[J]. 中国人事科学, 2020(2): 52-61.
- [19] 田书芹, 王东强. 乡村人才振兴的核心驱动模型与政策启示——基于扎根理论的政策文本实证研究[J]. 江淮论坛, 2020(1): 10-17.
- [20] 孙瑞英, 陈宜泓. 基于 LDA 主题模型的国内智慧阅读研究热点及发展导向研判[J]. 图书馆建设, 2023(3): 82-93, 103.
- [21] 杨剑, 石孝宇, 郭正茂, 等. 新时代我国青少年体育政策注意力配置及演进——基于 LDA 主题模型的政策文本分析[J]. 体育学研究, 2024, 38(3): 77-88.