

ISSN 3105-7497 (print)  
ISSN 3105-7500 (online)

# 数据科学与工程

## Data Science & Engineering

2025年 第1卷 第2期 (双月刊)

CODEN: SKYGAY

(国际标准连续出版物标识符·全球唯一标识符)

分配机构: 美国化学文摘社 (CAS)

国图集团 CIBTC

进口备案刊号: G015Z119

QUEST PRESS LIMITED

www.SciOnline.com



QUEST PRESS



扫码阅读



敬畏自然，方能生生不息

*Revere nature, to achieve endless life.*

**地球是共同家园，守护需同心协力**

*The earth is a common home, protection requires joint efforts.*



ISSN 3105-7497 (Print)

ISSN 3105-7500 (Online)

CODEN: SKYGAY

(国际标准连续出版物标识符 • 全球唯一标识符)

分配机构: 美国化学文摘社 (CAS)

中文名 数据科学与工程

英文名 Data Science & Engineering

主办 求索出版社

出版频率 双月刊

出版语言 中文, 英语

出版机构 求索出版社 (QPL Ltd.)

编辑单位 《数据科学与工程》编辑部

地址 澳门巴掌围斜巷19号7楼D

电话 +853 6881 9699

网址 sjkxygc.scionline2025.com

广告服务 QuestPress@hotmail.com

出版时间 2025年12月



QUEST  
PRESS  
LIMITED

版权所有

© 求索出版社 (QPL Ltd.)

#### 版权与许可声明

本期刊整体版权归属求索出版社所有。所有发表文章均依据“知识共享署名 4.0 国际许可协议” (CC BY 4.0) 进行授权发布。

#### 许可链接

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

#### 文章版权

单篇文章的著作权由作者与求索出版社共同享有。

#### 使用授权

著作权人已同意依据CC BY 4.0协议授予文章使用许可, 使用者可在遵循协议条款 (主要为署名要求) 的前提下进行转载、改编或用于其他用途。

#### 权责声明

1. 本刊所发表文章中的观点、数据和结论仅代表作者个人立场, 不代表本刊及出版机构的意见。
2. 对于因使用文章任何内容 (包括正文及引述材料) 而可能引发的任何直接或间接损失, 本刊及出版机构不承担任何法律责任。
3. 本刊的出版行为不构成对文中提及任何产品、方法或技术的商业性能、有效性或特定用途适用性的任何形式担保或推荐。
4. 本刊内容仅供学术与信息参考, 不替代专业判断。读者在依据文章内容做出任何决策前, 建议咨询相关领域专业人士。

## 编委会

主编 黄琳

副主编 张明昭 锦州市数据中心

编委 孙丽欣

石江宇

胡瑶

编委会助理 胡凯

Journal import

期刊进口

备案刊号: G015Z119



中國國際圖書貿易集團公司

China International Book Trading Corporation







DOI 数字对象唯一标识符 (英国)



Crossref 交叉引用 (美国)



ICI 哥白尼期刊数据库 (波兰)



EuroPub 欧洲学术出版中心数据库 (英国)



ACADÉMIA. EDU 学术界 (美国)



J-Gate 开放获取期刊门户 (印度)



CNKI 中国知网 (中国)



WanFang 万方数据 (中国)



CVIP 维普资讯 (中国)



RCCSE 中国学术期刊目录 (中国·武汉大学)



Baidu Scholar 百度学术 (中国)



CJWK 长江文库 (中国)



591adb 中邮阅读·万章期刊 (中国)



OpenAlex 全球开源学术平台 (加拿大)





**SCRIBD**

SCRIBD 在线图书馆 (美国)



**MAGZTER**  
DIGITAL NEWSSTAND  
TAP • READ • ENJOY

Magzter 跨平台数字期刊 (美国)



SlideShare 知识共享平台 (美国)



EBSCO 学术数据库 (美国)



*RePEc*

RePEc 经济研究文献数据库 (美国)



 **OUCI**

OUCI 开放乌克兰科学 (乌克兰)



 **Discovery**  
by Editage

R Discovery 研究者发现平台 (印度)



 **LENS.ORG**

The Lens 透镜学术 (澳大利亚)





Google Scholar 谷歌学术 (美国)



 **Microsoft Bing**

Microsoft Bing 微软学术 (美国)



 **ResearchGate**

ResearchGate 研究之门 (德国)





Portico 学术文献保存库 (美国)





ASCI 亚洲科学引文索引 (美国)



 **SEMANTIC SCHOLAR**

Semantic Scholar 语义学术 (美国)





ISSN (法国)



BOOKAN 博看 (中国)



Issuu 数字出版平台 (丹麦)



WorldCat 世界编目数据库 (美国)



SPI-Hub 范德比尔特大学 (美国)



MDPI scilit 科学与学术数据库 (瑞士)



ProQuest 科睿唯安数据库 (美国)



Research Commons 科睿唯安·研究共享平台 (美国)



Scite (美国)



BASE 比勒费尔德学术搜索 (德国)



ERIH+ 欧洲人文社会科学索引 (挪威)



Yahoo!Search (美国)



Naver 搜遍 (韩国)



LivRe! 开放期刊门户 (巴西)





ResearchBib 研究者索引 (日本)



KINDCONGRESS

KIND CONGRESS (阿塞拜疆)



TOGETHER WE REACH THE GOAL  
SJIF 科学期刊影响因子 (印度)



IP Indexing IP索引



Journament

Journament 质量评估与索引



RJIF 研究期刊影响因子



Road 开放科学学术资源路线图 (法国)



COSMOS 期刊评价指标 (印度)



CiteFactor

CiteFactor 指标与信任索引 (印度)



ADL 亚洲数字图书馆 (巴基斯坦)



Secret Search Engine Labs

Secret 搜索引擎实验室 (芬兰)



Academindex 学术索引 (土耳其)



ROOTINDEXING  
Discover. Index. Connect. Impact.

RootIndexing 根索引与影响因子



EZB Electronic Journals Library

EZB 电子期刊图书馆 (德国)



ISI 国际科学索引平台（阿联酋迪拜）



Sci Online 科学在线（中国澳门）



Baidu Baike 百度百科（中国）



OALib 开放存取资源图书馆（美国）



ESJI 欧亚科学期刊索引（哈萨克斯坦）



# 数据科学与工程

Data Science & Engineering

2025年12月 第1卷 第2期（双月刊）

## 目次

### ◆ 人力资源管理信息化

数字化工具在人力资源弹性配置中的实践路径

The practice path of digital tools in the flexible allocation of human resources

汪大鹏..... (1)

电子化招聘系统对人才配置效率的影响分析

Analysis of the impact of electronic recruitment system on talent allocation efficiency

温 泉..... (6)

基于大数据的人力资源动态配置中面临的问题及对策

Problems and Countermeasures in the dynamic allocation of human resources based on big data

李忠南..... (10)

### ◆ 智慧城市与公共安全

锦州智慧城市建设中人工智能对消费投资新逻辑的适配效应

The adaptive effect of artificial intelligence on the new logic of consumption and investment in the construction of Jinzhou smart city

张明昭..... (16)

### ◆ 通用大数据分析处理

互联网站群动态监测中的常见问题及应对策略研究

Research on common problems and Countermeasures in dynamic monitoring of Internet stations

樊 冲..... (24)

基于机器学习的数据库查询优化技术研究

Research on Database Query Optimization Technology Based on Machine Learning

许 幸 ..... (29)

面向人工智能的紧固件制造检验检测高质量数据集构建方法

A High-Quality Dataset Construction Method for AI-Based Fastener Manufacturing Inspection and Testing

张骄阳, 李 梦, 朱 松, 董 敬, 郑天翼 ..... (34)

多源异构物联网数据融合与异常检测的轻量化深度学习模型

Lightweight deep learning model for data fusion and anomaly detection in multi-source heterogeneous Internet of Things

雷雨果, 宋聪聪, 李启周 ..... (40)

面向计算机信息系统的建构主义辅助学习系统构建策略

Construction strategy of Constructivism assisted learning system for computer information system

甘志玮 ..... (45)

小型分布式文件存储系统的性能优化研究

Research on Performance Optimization of Small-Scale Distributed File Storage Systems

刘云博 ..... (51)

便民服务热线数据全生命周期安全管理机制探讨

Discussion on life cycle safety management mechanism of convenient service hotline data

张明昭 ..... (56)



# 数字化工具在人力资源弹性配置中的实践路径

汪大鹏

锦州市凌河区人力资源和社会保障服务中心，辽宁锦州，121000

**摘要：**在数字经济与组织变革的双重驱动下，传统人力资源配置模式逐渐暴露出响应滞后、匹配精度不足及资源浪费等结构性矛盾，难以满足企业对弹性用工与动态岗位管理的现实需求。针对上述问题，本研究聚焦于电子技术如何通过数字化工具实现人力资源的弹性配置，提出以智能排班系统、技能图谱平台与实时劳动力数据分析模型为核心的技术路径，构建覆盖“需求预测—能力匹配—动态调度”全流程的闭环机制。

**关键词：**弹性配置；数字化工具；智能排班

## The practice path of digital tools in the flexible allocation of human resources

Dapeng Wang

Jinzhou Linghe District human resources and social security service center, Jinzhou Liaoning 121000, China

**Abstract:** driven by the dual drive of digital economy and organizational change, the traditional human resource allocation model has gradually exposed structural contradictions such as lagging response, insufficient matching accuracy and resource waste, which is difficult to meet the actual needs of enterprises for flexible employment and dynamic post management. In view of the above problems, this study focuses on how electronic technology can realize the flexible allocation of human resources through digital tools, and proposes a technical path with intelligent scheduling system, skill map platform and real-time labor data analysis model as the core, to build a closed-loop mechanism covering the whole process of “demand forecasting capacity matching dynamic scheduling”.

**Keywords:** flexible configuration; digital tools; intelligent scheduling

## 0 引言

本文把电子技术的应用范围从传统生产运营环节，延伸到人力资本管理领域，补充了数字化和弹性用工交叉研究的分析框架。此前多数文献重点放在电子技术对制造流程的优化上，本文则整合物联网、智能算法与技能图谱，说明数字化

工具对人力资源配置底层逻辑的重构路径。这套框架既覆盖技术工具的操作效率，也突出数据支撑的动态决策机制，为解读“技术-组织-员工”三者的互动模式提供新方向。这项理论成果能缩小电子工程与人力资源管理学科之间的学术差距，推进交叉研究的深入开展<sup>[1]</sup>。



Copyright © 2025 by author(s) and Quest Press Limited. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>



本文把弹性配置看成技术支撑的闭环系统，打破了传统静态配置模型的局限。分析框架涵盖需求预测、能力匹配与动态调度三个核心环节，重点关注实时数据交互对配置灵活性的促进作用。这种系统化视角让本文能更准确判断数字化工具在不同组织场景中的适配条件，为后续实证检验提供可操作的理论基础。这一框架的成型，意味着电子技术应用研究实现从辅助性角色到核心战略工具的理论转变<sup>[2]</sup>。

1 数字经济与组织变革对传统人力资源配置的挑战

2010年之后，企业组织形态快速转向平台化与网络化，业务需求的波动和不确定程度大一些。传统人力资源配置模式以固定编制、层级审批为基础，响应速度跟不上组织对即时用工的动态需求。岗位要求与员工能力的匹配，多依据静态职位说明书和人工经验判断，精度不够，让结构性冗员与关键岗位缺员的情况同时出现。传统管理逻辑与数字化运行环境不兼容，催生了“慢响应”与“粗匹配”的矛盾，这是限制组织人力资本效率提升的核心问题<sup>[3]</sup>。

数字化工具要实现人力资源弹性配置，核心是搭建一套动态匹配机制。智能排班系统是其中关键载体，它依托历史工时数据和业务量预测模型搭建算法逻辑，实时处理物联网终端收集的岗位负荷信息，自动生成既符合员工偏好又满足合

规要求的排班方案。技能图谱平台借助数据挖掘技术，把员工过往培训记录、项目经验和绩效数据转化为可量化的技能标签库，构建出能动态更新的劳动力能力画像。这两类平台共同支撑起弹性配置的技术框架：智能排班系统解决“何时需要人”的时间安排问题，技能图谱平台明确“何人能胜任”的人员匹配问题，两者配合运作，才能让人力资本在时间、能力两个维度都具备弹性。

具体操作时，数字化工具要接入实时劳动力分析模型，以此作为决策核心。这个模型会把云计算提供的岗位需求数据，和员工技能图谱的匹配结果整合到一起，用算法对跨部门人力调度的可行方案做多目标优化，在成本限制与任务紧急程度之间找到平衡点。弹性配置的闭环机制可清晰梳理：先做需求预测，再通过能力匹配完成人与岗的初步对应，接着开展动态调度实现资源的实时再分配，最后搭建起从数据采集到调度指令下发的完整技术流程。

2 数字化工具在人力资源弹性配置中的技术基础

2.1 物联网与云计算构建实时数据交互架构

物联网与云计算构成了实时数据交互架构的核心支撑。物联网通过智能工牌、工位传感器等终端设备，动态采集员工在岗状态、技能认证有效期限及任务完成时长等微观数据，经边缘计算预处理后上传至云端，见表1。

表1 技术组件

| 技术组件   | 功能描述                   | 部署方式             |
|--------|------------------------|------------------|
| 物联网传感器 | 采集员工工位占用、设备运行状态等实时数据   | 边缘网关+云端API       |
| 云计算平台  | 存储技能图谱、考勤记录等结构化与非结构化数据 | 混合云架构            |
| 数据同步接口 | 实现HR系统与排班模块的双向数据更新     | RESTful API+消息队列 |

云计算平台依托分布式存储和弹性计算能力，把分散的岗位需求数据与员工技能标签实时

关联起来，搭建起统一的劳动力数据湖。这套架构推动人力资源配置模式转变，告别传统周期性

报表分析,实现毫秒级数据响应,比如生产订单出现波动时,系统会立刻抓取各岗位空缺信息,匹配可用员工的技能情况。

## 2.2 智能排班系统的算法逻辑与部署路径

智能排班系统的算法核心,是把人力资源管理里的多种约束条件转成可计算的数学模型,逻辑框架一般结合约束满足问题与时间序列预测两种方法。系统依托历史订单数据与员工休假计划,用长短时记忆网络测算未来时段的人力需求峰值,再把劳动法规定的工时上限、员工技能等级与个人偏好等硬性约束转成布尔变量,靠回溯搜索算法在解空间里筛选可行方案。部署时,系统用微服务架构单独封装排班引擎,通过容器化技术完成与现有HR系统的松耦合集成,利用边缘节点处理实时考勤数据,保证车间级网络中断时,排班方案仍能离线执行。

初始排班方案由算法生成后,还要完成局部搜索优化,平衡员工工作负荷的公平性与业务连续性。系统引入模拟退火算法,针对连续夜班次数、跨部门调岗频率等软约束迭代调优,在满足业务需求的前提下,把员工疲劳指数降到最低。实际落地时,该算法模块依靠RESTful API对接物联网采集的实时工位占用数据,一旦监测到某岗位突然缺员,就能在秒级时间内重新核算相邻班组人员的技能匹配度与通勤时间,生成动态补位方案。算法逻辑与部署架构相互结合,让排班系统从被动记录工具转变为主动的人力资源配置决策引擎。

## 2.3 技能图谱平台与实时劳动力分析模型

技能图谱平台依托语义网络,把员工的工作经历、培训记录、项目成果转化为可量化的能力向量。平台用自然语言处理技术拆解岗位描述与绩效文档里的技能关键词,搭建涵盖技术深度、协作层级、行业经验的标签体系。机器学习模型对员工过往任务完成数据做聚类分析,自动

找出潜在技能组合与迁移路径,比如把生产线操作员的设备调试能力对应到质检岗位的异常识别范畴。实时劳动力分析模型整合物联网采集的工位操作数据与项目管理系统的任务完成率,用回归算法动态评定每位员工的当前产能与技能冗余度。业务波动引发人力需求变化时,系统借助技能图谱的语义相似度计算,从备选池里挑出跨岗位适配度达标员工,同时制定个性化培训建议填补能力缺口。

依托这套量化机制,档案管理从静态模式转向动态能力识别模式。员工的技能标签会被分析模型实时监测时效性,一旦发现某类技术需求下滑,系统就会自动给人力资源部门发预警,督促其启动再培训计划。拿平台实际操作来说,它会把季度业务数据和员工技能更新频率做对比,找出因引入自动化设备而不再急需的传统操作技能,同时推荐关联度高的数字孪生运维课程。技能图谱不断迭代更新,劳动力分析结果也会实时反馈,两者形成闭环后,组织能根据业务战略主动调整人才储备结构,不用等岗位出现空缺再被动应对应对。这套系统的实际效果,和标签体系的标准化水平、数据采集的完整度直接相关,就算数据量不足,依靠迁移学习的冷启动算法,也能给出初步的匹配建议。

# 3 基于数字化工具的弹性配置闭环机制设计

## 3.1 需求预测: 基于历史数据与业务波动的岗位需求建模

在岗位需求预测环节,数字化工具通过整合历史考勤、业务量波动及外部市场指标,构建多维度预测模型。时间序列分析可捕捉用工需求的周期性规律,例如零售业在促销节点前两周出现人力需求陡增,模型能提前输出未来7天的排班缺口,见表2。



表2 建模方法

| 建模方法   | 输入数据          | 输出结果        | 典型应用场景       |
|--------|---------------|-------------|--------------|
| 时间序列分析 | 历史用工数据（日/周/月） | 短期岗位需求峰值预测  | 电商大促、节假日排班   |
| 回归分析   | 业务量、季节因子、市场指标 | 长期人力缺口量化评估  | 新产线投产、业务扩张规划 |
| 组合模型   | 历史数据+实时业务波动   | 动态需求预警与阈值触发 | 突发订单、设备故障应急  |

回归分析则引入销售额、客户流量等自变量，量化业务变化对岗位数量的弹性影响。两种方法结合后，系统能生成动态需求热力图，标注各部门在未来72小时内的紧缺岗位与技能要求，为后续能力匹配提供数据锚点。

3.2 能力匹配：员工技能标签与岗位要求的自动化对齐

依据需求预测给出的岗位空缺信息和技能要求，员工与岗位的自动化匹配靠语义分析、相似度计算完成。系统把员工简历、培训记录、项目经验转换成结构化技能标签向量，再把岗位描述里的职责与能力要求拆解成需求向量。用预训练语言模型算出两者的余弦相似度，识别出技能重合度高于90%的候选员工，还会按照匹配得分排出推荐顺序。某物流企业用上这套模型后，临时岗位填补耗时从4小时压缩到45分钟，人岗匹配准确率达到87%，解决了传统人工筛选存在的经验偏差、效率低下等问题。

3.3 动态调度：跨部门人力调配的实时响应与优化

以人员能力匹配为核心生成输出结果后，动态调度模块依托物联网设备与调度算法，完成人力资源跨部门的实时重组。岗位空缺触发调度指令时，物联网终端立刻采集在岗员工的技能标签与地理位置数据，算法结合任务优先级、员工技能匹配度及物理距离搭建多目标优化模型，在分钟级时间内得出最优调配方案。某制造企业落地该机制后，紧急订单的跨部门人员到位时长从2

小时缩短到20分钟，整体人力闲置率降低25%，提升了突发业务场景下的资源响应速度。

调度算法的核心在于动态约束求解，系统需同步处理员工工时合规性、技能兼容性及部门负载均衡等限制条件。通过引入遗传算法与粒子群优化，平台能在数秒内迭代出满足90%以上约束条件的调度解，并实时推送至员工移动终端。这一技术路径不仅实现了人力资源的弹性重组，还通过数据闭环持续优化调度策略，为组织应对市场波动提供了敏捷的人力配置能力。

4 数字化工具落地中的关键依赖与挑战

4.1 组织数据基础对技术实施效果的制约

组织的数据基础，直接决定数字化工具在人力资源弹性配置中的实际应用成效。系统要正常运转，就得有高质量的员工数据，要是技能标签更新不及时、工时记录有偏差，智能排班和技能图谱平台算出的结果就不符合实际需求。统一规范的流程也很重要，跨部门人员调动要是没有统一的数据接口和审批环节，实时数据交互架构就没法支持动态调度的相关决策。有些企业只引入技术系统，却不对底层数据做同步治理，会让需求预测不准、能力匹配出错，最后让这套工具变成中看不中用的样子。技术落地的核心支撑是数据质量与流程规范，这两部分的完善情况，直接影响弹性配置全流程的响应速度和执行效果。

4.2 员工数字素养与变革管理的协同需求

数字化工具能不能用好，和技术本身有关，也和员工的数字素养、变革管理的协同水平有

关。智能排班系统、动态调度平台这类工具，要求员工熟练操作移动端界面，还要懂算法逻辑，要是没有配套培训，技术系统可能因为操作失误或者员工抵触，出现使用率不高的情况。企业要建立分层培训机制，给管理层设计数据解读课程，给操作层设计系统应用课程，把数字化参与度加入绩效考核，激励员工主动适应新工具。变革管理里的沟通不能忽视，员工知道技术能优化个人排班公平性、提升跨部门协作效率后，抵触情绪会减少，弹性配置从数据采集到调度执行的全流程就能顺畅推进。

#### 4.3 管理流程再造与技术系统的深度融合

单独引入技术系统，常因为没适配好组织权责和考核机制而无法落地。智能排班系统给出调度方案后，要是部门主管还握着原来的审批权，算法算出的结果就可能被随便否定，弹性人力配置也就落不到实处。企业得跟着调整管理流程，把排班的决策权从各部门收归中央调度中心，还要设定跨部门人力调用的标准化审批时长，防止权责不清拖慢响应速度。考核机制的调整也不能少，传统按部门核算人效的指标会限制人员流动，要加入“共享人力利用率”与“跨部门协作响应速度”这类跨团队考核指标，让管理者从抢固定编制转向优化动态人力配置，为技术系统搭

建顺畅运行的制度环境。

## 5 结语

研究发现，通过部署基于物联网与云计算的人力资源管理系统，企业能够实现岗位需求与员工技能数据的实时交互，显著提升跨部门人力调配的响应速度与准确性。此外，研究还识别出数字化工具在实施过程中对组织数据基础与员工数字素养的依赖，强调技术落地需与管理流程再造协同推进。本研究的理论意义在于将电子技术应用从传统的生产自动化拓展至人力资本管理领域，丰富了人力资源数字化与弹性工作制交叉研究的分析框架；实践层面，为企业在不确定性环境中构建柔性人力配置体系提供了可操作的技术方案与实施参考，对提升组织运营效率与员工工作适配性具有现实价值。

## 参考文献

- [1] 王思默. 数字化背景下人力资源管理角色的战略转型路径探究[J]. 上海商业, 2025, (12): 214-216.
- [2] 温晓星. 数字经济时代人力资源管理数字化转型模式[J]. 现代企业, 2025, (12): 16-18.
- [3] 孙文婷. 数字化背景下国企改革中人力资源管理的改进研究[J]. 国际公关, 2025, (23): 115-117.



# 电子化招聘系统对人才配置效率的影响分析

温泉

辽宁省义县人力资源和社会保障局, 辽宁义县, 121100

**摘要:** 随着信息技术的快速迭代与数字化转型的深入推进, 电子化招聘系统已逐渐成为企业人才获取的主流渠道, 传统招聘模式在信息匹配效率、简历筛选速度及候选人评估准确性等方面日益显现出局限性。当前, 尽管电子化招聘系统在实务中广泛应用, 但其对人才配置效率的具体影响机制仍缺乏系统性实证分析, 尤其在如何平衡技术优势与潜在匹配偏差方面存在明显研究空白。本研究旨在构建一个包含招聘流程数字化程度、候选人匹配算法精度与组织决策效率三个维度的分析框架, 通过收集来自不同行业企业的招聘数据与人才流动指标, 运用回归分析与中介效应检验, 系统考察电子化招聘系统对人才配置效率的直接影响与间接路径。

**关键词:** 人才配置; 电子化招聘; 配置效率

## Analysis of the impact of electronic recruitment system on talent allocation efficiency

Quan Wen

Liaoning Yixian County Human Resources and Social Security Bureau, Yixian Liaoning 121100, China

**Abstract:** with the rapid iteration of information technology and the in-depth promotion of digital transformation, the electronic recruitment system has gradually become the mainstream channel for enterprises to obtain talents. The traditional recruitment model has increasingly shown limitations in terms of information matching efficiency, resume screening speed and candidate evaluation accuracy. At present, although the electronic recruitment system is widely used in practice, its specific impact mechanism on talent allocation efficiency is still lack of systematic empirical analysis, especially in how to balance the technical advantage and potential matching deviation. The purpose of this study is to build an analysis framework including three dimensions: the degree of digitalization of recruitment process, the accuracy of candidate matching algorithm and the efficiency of organizational decision-making. By collecting recruitment data and talent flow indicators from enterprises in different industries, and using regression analysis and mediation effect test, the direct impact and indirect path of electronic recruitment system on talent allocation efficiency are systematically investigated.

**Keywords:** talent allocation; electronic recruitment; allocation efficiency

本文开展这项研究, 一是为企业优化电子招聘系统功能、调校算法拿出实证依据, 二是针对人力资源管理理论里“技术赋能”和“人机协同”的矛盾给出新的解释方向。企业推进招聘系统电子化时, 要把技术工具和人工判断结合起来, 避免过度自动化让人才配置的适配性打折扣, 以此实现更高质量的人力资源优化配置, 同时提供对应的理论与实践内容。



Copyright © 2025 by author(s) and Quest Press Limited. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>





## 1 电子化招聘系统的技术架构与核心功能模块

### 1.1 系统技术架构的演进路径与数字化特征

电子化招聘系统的技术架构,已从单机部署形式,转变成云端化、智能化的形态。早期这类系统大多用本地服务器加关系型数据库的C/S架构,只能完成简历电子化存储、基础检索工作,数据处理能力比较有限,还没法支持跨地域协同工作。云计算与大数据技术普及后,系统逐步切换到B/S架构并部署在云端,能实现资源弹性扩展,也能对数据进行集中管理。当前主流系统加入微服务架构与分布式计算框架,把简历解析、智能筛选、人才匹配等核心功能拆成独立服务,还集成自然语言处理与机器学习引擎,让系统能从海量非结构化数据里提取关键信息,还能完成动态学习,招聘流程的自动化水平与响应效率也得到了提升<sup>[1]</sup>。

### 1.2 简历解析与智能筛选功能的技术实现

简历解析时,系统用自然语言处理技术,完成词法分析、命名实体识别、依存句法分析,把零散的简历文本转成规整的结构化字段。卷积神经网络、双向长短期记忆网络用来提取工作经历、教育背景、技能标签,条件随机场模型帮忙识别段落边界和实体关系,提升信息抽取的准确率。智能筛选环节,系统用预训练的领域词向量、支持向量机或梯度提升树模型,比对候选人与岗位要求的显性匹配度做初筛,剔除明显不符合的简历,给后续算法匹配提供高质量候选池<sup>[2]</sup>。

### 1.3 人才匹配算法与候选人评分模型的设计逻辑

本文设计人才匹配算法与候选人评分模型时,让系统用多维特征对比完成岗位与候选人的量化匹配。算法先把结构化处理后的简历字段和岗位需求做向量化映射,用余弦相似度或欧氏距离算出技能、工作经验与教育背景的匹配得分。之后用逻辑回归或随机森林模型,加入权重分配机制,给关键指标(如行业年限、项目复杂度)更高权重,生成综合评分。这个过程既做显性标

签的精确比对,又通过聚类分析找出候选人职业发展路径与岗位成长曲线的契合度,避免仅凭关键词匹配让误判出现。评分模型最终给出0至100分的匹配指数,作为后续筛选的量化依据,提升初步配置效率<sup>[3]</sup>。

## 2 电子化招聘系统对人才配置精准度的影响

### 2.1 算法匹配对显性技能与经验要求的满足程度

电子化招聘系统依靠结构化数据解析技术,把学历、工作年限这类硬性指标转成可量化的匹配参数,在显性技能筛选环节能达成比较高的准确率。关键词匹配结合规则引擎的算法模型,可准确识别简历里的学历层级、专业名称以及从业年限,还能按照预设阈值完成初步筛选。部分系统加入自然语言处理技术,从非结构化文本里提取隐性经验表述的语义,进一步优化硬性指标的匹配粒度。不过算法判定“经验要求”时,常局限在时间维度的累加,对经验质量和岗位关联性的辨识存在不足,会让年限达标但实际能力不匹配的候选人也进入候选池。

### 2.2 隐性技能与组织文化契合度的评估偏差风险

算法在评估候选人的隐性技能与文化契合度时,会暴露出比较大的局限性。相关实证研究显示,依靠关键词和语义分析搭建的模型,没法准确捕捉团队协作风格、创新意愿这类软性素质,常把“团队合作”这类高频词汇的出现当作实际能力,让价值观匹配的判定只停留在表面。过度使用这类算法会出现系统性偏差,把擅长自我表达但实际行为不符的候选人误判为高契合度,却筛选掉真正具备隐性特质但简历表述平实的个体。这类失真情况在看重组织认同感的岗位里表现得比较明显,会直接降低人员配置的效率。

### 2.3 高端复合型岗位配置效率的抑制效应

针对高端复合型岗位,算法在评估跨领域能力与创新潜力时,存在结构性的难题。这类岗位一般要求候选人能在多个专业领域完成知识迁移,还要有解决非结构化问题的原创性思维,但

现有依托历史数据训练的模型，没法量化这种没显现出来的潜在价值。算法常把简历里跨行业经历的数量，当成跨界整合能力的衡量标准，却不能区分经历之间的逻辑联系与思维突破。这种评估偏差，让真正有颠覆性创新潜质的候选人，因缺少标准化成果指标被系统调低评分，反倒让履历好看但循规蹈矩的应试者拿到高分。长期筛选掉高风险高回报的复合型人才，会削弱组织在技术变革阶段的适应能力，还会形成锁定效应，拉低人才配置的长期效率。

### 3 系统功能集成度与数据动态更新机制的调节作用

#### 3.1 招聘系统与人力资源信息系统集成度的调节效应

招聘系统和人力资源信息系统的整合程度，会直接左右电子化招聘对人才配置效率的调节走向。当两个系统能把简历解析、面试评估内容，和员工档案、绩效数据无缝对接时，候选人能力与在职者表现的对比性会增强，算法匹配时对显性技能的筛选精度也会提升。但如果集成只停留在数据表层互通，没把岗位胜任力模型和组织文化标签深度嵌入其中，系统修正隐性技能偏差的能力就会变弱。模块集成程度过高还可能出现“数据过载”，也就是冗余信息会干扰核心筛选逻辑，让高端复合型岗位的配置效率呈现先升后降的倒U型曲线。面对这种非线性变化，企业推进系统整合时，得先校准HRIS里历史绩效数据与招聘模型的特征权重，不能盲目堆叠功能。

#### 3.2 人才数据库动态更新频率对匹配精度的影响

算法推荐结果和岗位实际需求的匹配程度，核心影响变量是人才数据库的动态更新频率。保持高频更新，系统就能实时获取在职员工的绩效数据，掌握行业技能需求的变动，避免候选人画像因数据库静态化出现信息滞后，提升显性技能匹配的时效性。不过更新频率过高且缺少数据清洗机制时，临时项目需求这类短期波动信息，可

能被当成核心筛选标准，让算法对候选人长期潜力的评估出现偏差。岗位需求快速迭代的行业里，这种动态性与稳定性的矛盾表现比较大，数据更新策略要根据岗位类型做差异化调整。

人才数据库更新和招聘的时间周期不匹配，会让隐性技能评估里的系统性偏差变得更大。比如系统只参考每月更新一次的组织文化标签，不关注员工离职率、团队协作反馈这些每周都会变的内容，算法给出的候选人文化契合度评分，就会跟不上企业实际氛围的变化。数据库更新频次低，会让历史数据占比过高，拉低系统捕捉高端复合型岗位所需跨领域经验的能力。企业要搭建和岗位敏感度挂钩的分级更新机制，给核心岗位接入实时数据流，给常规岗位用周期性批量更新模式，平衡数据时效性与系统计算资源的分配效率。

#### 3.3 系统功能模块协同效应对配置效率的综合影响

简历解析、面试安排、测评集成这几个模块相互配合，让招聘全流程的协同效率有了提升。系统先由简历解析模块自动提取候选人的关键信息，接着立刻启动面试安排算法，结合岗位紧急程度和面试官日程给出最合适的时间区间，缩短人工协调花费的时间。测评集成模块借助标准化接口，导入认知能力或性格测试的结果，面试官在预约阶段就能拿到候选人的初步情况，据此设计对应面试问题。模块间的数据流能顺畅衔接，把原本分散的招聘环节串成连贯流程，让简历投递到首轮面试的平均周期缩短约30%。模块协同的深度受限于技术架构的统一性。要是各模块采用不同的数据标准，比如简历解析输出JSON格式，测评系统仅支持XML，系统就得额外消耗计算资源做格式转换，让实时联动效果下降。企业需完成API网关搭建，实现模块间协议标准化，同时建立事件驱动机制，让任一模块的状态变更，比如候选人拒绝面试，能自动触发关联模块的调整逻辑，维持整体配置效率的稳定性。

## 4 技术赋能与人机协同的平衡策略与优化路径

### 4.1 算法自动化与人工判断的互补机制设计

电子化招聘系统里，算法自动化能高效处理大量简历，对隐性技能、组织文化契合度的评估却有固有偏差，得在筛选流程里加人工复核节点形成互补。算法完成初步匹配后，可针对评分处于临界阈值（如排名前15%-20%）的候选人开展人工复核，由招聘专员结合面试记录与行为事件访谈法（BEI），对候选人的沟通协作能力、价值观适配度等非结构化信息完成二次评估。这套机制能用算法减少显性条件筛选的重复劳动，还能靠人工判断填补模型在团队动态、岗位隐性需求等情境因素上的感知盲区，提升筛选效率的同时抑制配置偏差风险。

### 4.2 电子化招聘系统功能优化的实证建议

本文得出结论，算法调校的重点放在修正隐性技能评估的盲区。具体策略是在模型训练阶段，引入行为事件访谈数据作为标注样本，提升算法对协作能力、抗压性等非结构化特征的学习权重，还要建立数据治理机制，定期清理历史简历库中的冗余与噪声信息，依照技术岗与创意岗等不同岗位类型，设定差异化的特征权重阈值，避免“一刀切”式评分使得适配度失真。

流程改进上，本文建议在系统不同功能模块间搭建动态反馈回路，当人才匹配模块接连出现高分候选人入职后离职率异常情况时，启动人工复核流程，反向校准算法参数。同时，本文需提升系统与人力资源信息系统（HRIS）的集成度，确保候选人从初筛到入职的全生命周期数据实时回流至人才数据库，形成“匹配-反馈-优化”的闭环，该协同机制可抑制数据滞后造成的配置效率衰减，为人机互补提供可迭代的实践框架。

### 4.3 人机协同模式下人才配置效率提升的实践框架

结合前述研究结论，构建人机协同的实践框

架要把技术工具与人力资源专业判断深度结合。这个框架的核心流程是“算法初筛—人工复核—动态校准”，系统依托显性技能匹配度完成大规模初筛，减少基础性重复劳动，人力资源专家重点关注系统评分偏低、但有潜在隐性特质的候选人，用结构化行为面试做深度评估，填补算法在组织文化契合度判断上的缺口。这套流程避开单纯依靠算法或全人工筛选的极端情况，达成效率与精准度的平衡。

双向反馈机制是保障框架稳定运行的核心。若人工复核结论和算法打分长期存在差异，系统要自动留存相关记录，启动模型参数微调，比如针对特定岗位调整隐性特征的权重。人力资源专家定期复盘系统筛选出的“高潜力但低评分”候选人，完成定性评估报告后补充到训练数据集中。人机协同模式既优化复杂岗位的配置效率，又能通过持续迭代优化，让系统逐步具备识别非结构化特征的能力，最终形成可复制的管理模式。

## 5 结语

本文在调研中观察到，企业用电子化招聘系统处理简历、安排初面，耗时能得到压缩，岗位空窗期也随之缩短，不过针对高端复合型岗位，要是过度依靠算法做匹配，就会让隐性技能、组织文化契合度的评估出现偏差，长期来看会拉低人才配置效率，系统功能模块的整合程度、人才数据库的动态更新机制，会对配置效率起到关键调节作用。

## 参考文献

- [1] 王颖倩. AVT制造型公司电子化招聘流程优化研究[D]. 上海外国语大学, 2024.
- [2] 任康磊. 人力资源管理实操从入门到精通[M]. 北京: 人民邮电出版社, 2020: 350.
- [3] 郭玥. 人力资源招聘信息系统的设计与实现[D]. 吉林大学, 2014.





# 基于大数据的人力资源动态配置中面临的问题及对策

李忠南

锦州市数据中心，辽宁锦州，121000

**摘要：**大数据技术发展速度快，企业人力资源管理已从传统静态配置转向动态化、智能化模式。当前人力资源动态配置环节里，数据孤岛问题比较大、数据质量有高有低、算法模型适应能力不足，隐私与伦理风险也很突出，这些问题让人才流动与岗位匹配效率不高，没法实现真正的精准配置与动态优化。本文利用大数据分析技术，搭建起一套覆盖数据采集整合、特征工程、岗位需求与员工能力匹配模型、动态反馈机制的人力资源动态配置框架。

**关键词：**大数据；动态配置；问题对策

## Problems and Countermeasures in the dynamic allocation of human resources based on big data

Zhongnan Li

Jinzhou data center, Jinzhou Liaoning 121000, China

**Absrtact:** With the rapid development of big data technology, enterprise human resource management has shifted from the traditional static configuration to the dynamic and intelligent mode. At present, in the process of dynamic allocation of human resources, the problem of data island is relatively large, the quality of data is high or low, the adaptability of algorithm model is insufficient, and privacy and ethical risks are also prominent. These problems make the efficiency of talent flow and job matching not high, and it is impossible to achieve real precise allocation and dynamic optimization. This paper uses big data analysis technology to build a dynamic human resource allocation framework covering data collection and integration, feature engineering, job demand and employee ability matching model, and dynamic feedback mechanism.

**Keywords:** big data; dynamic configuration; Problem Countermeasures

### 1 大数据背景下人力资源动态配置的理论基础

#### 1.1 大数据与人力资源管理转型的内在逻辑

引入大数据技术后，人力资源管理跳出传统静态档案管理模式，转向依托实时数据的动态配置模式。传统模式靠定期考核和人工经验运转，没法跟上市场与业务的快速变化，大数据则

能收集员工行为、绩效及岗位需求等多维度信息，搭建起持续更新的动态人员画像。这种模式转变的核心是数据驱动的闭环运作：系统实时分析技能缺口与人员流动走向，自动生成人员调配方案，取代滞后的人工决策环节。有云计算与物联网做支撑，数据采集不再受时空约束，人力资源部门从“事后分析”转为“实时干预”，提升



Copyright © 2025 by author(s) and Quest Press Limited. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>



配置的敏捷性与精准度。这一模式其实是对传统人力资源流程的数字化重构，把人员与岗位的匹配从静态的“固定对应”调整为动态的“自适应适配”<sup>[1]</sup>。

### 1.2 人力资源动态配置的核心要素与运行机制

人力资源动态配置的关键，是让数据流、人才流与岗位流形成协同循环。数据流是这套机制的底层动力，它实时收集员工的技能、绩效和行为数据，搭建能动态更新的能力画像；岗位流会结合业务战略与市场变动，不断生成可灵活调整的职位需求。人才流在三者中间承担连接功能，系统依托数据流与岗位流的匹配模型，自动推送最优人选或是跨部门调配方案。这套运行模式要实现从“人找岗”到“岗找人”的转变，算法实时分析技能缺口与冗余情况，推动人才往高价值环节流动。数据同步、画像更新、匹配计算与反馈迭代是其中的关键环节，能让配置决策从滞后响应转变为前瞻干预<sup>[2]</sup>。

### 1.3 当前动态配置面临的技术与组织瓶颈

大数据技术能为人力资源动态配置提供理论支持，不过实际应用时，数据孤岛、算法局限与组织惯性会形成三重阻碍。企业内部HRIS、绩效管理、培训平台等系统没有统一的数据接口与标准，这会让员工技能、绩效与行为数据分散存储，没法整合出完整的能力画像，进而降低数据

流对人才流的推动效率。当前的匹配模型大多靠历史数据训练，当业务战略快速调整带来非结构化岗位需求时，模型泛化能力不够，还存在冷启动问题，没办法精准提前完成从人找岗到岗找人的转变<sup>[3]</sup>。

组织惯性体现在传统科层制会抵触敏捷调配流程。部门间存在壁垒、职责划分固化，让跨团队人才流动遇到审批环节冗长、绩效归属不清晰等阻碍，就算算法算出最优调配方案，执行部门也常因本位主义拖延甚至否决。技术与组织之间的这类结构性矛盾，会让动态配置机制从数据驱动变成数据辅助，没法构建真正的协同闭环。要打破这些瓶颈，需同步推进数据治理标准化与组织流程再造，不能只靠优化算法。

## 2 人力资源动态配置中的数据质量问题与对策

### 2.1 多源异构数据融合的障碍与标准化策略

多源异构数据融合时，不同部门的数据格式、统计口径不统一，是最先遇到的阻碍。招聘系统里的“学历”字段，可能以文本形式存储，比如标注“本科”，绩效系统却用编码记录，比如用“1”代表本科，这类差异让数据无法直接对接。企业要制定统一的数据标准化策略，具体操作措施如下表1所示。

表1 多源异构数据融合的障碍类型

| 障碍类型  | 具体表现                   | 标准化对策           |
|-------|------------------------|-----------------|
| 格式不一致 | 同字段在不同系统采用文本、数字或日期格式   | 定义通用数据字典，强制转换格式 |
| 口径差异  | 对“工作经验”的计算标准不同（如是否含实习） | 统一口径规则，如仅计算全职年限 |
| 语义冲突  | “部门名称”缩写与全称混用          | 建立映射表，实现语义对齐    |

通过实施上述策略，可有效降低数据融合的复杂度，为后续的动态配置模型提供高质量的数据基础。

### 2.2 数据质量参差不齐对匹配模型的影响及治理

数据质量有高有低，是限制人力资源动态配

置模型发挥作用的核心问题。采集标准不统一、录入操作失误、不同系统间信息同步慢这些情况，会让员工绩效与流动数据出现缺失、异常或噪声问题。主管给绩效打分时可能受主观因素影响，评分尺度宽严不一，离职记录里的时间戳出

错，会打乱对人员流动规律的分析。要是直接把这类质量差的数据放进匹配模型，会严重干扰岗位需求和员工能力的精准匹配，让推荐结果出现偏差。必须搭建一套系统化的数据治理机制。完成数据清洗管道的构建，用预设规则和统计方法填补缺失数据、修正格式错误，再引入孤立森林或Z-score这类异常检测算法，自动找出并剔除绩效数据里的离群点或逻辑矛盾记录。这一过程能提升训练数据的信噪比，为后续机器学习模型的稳定运行打好基础。

### 2.3 实时数据采集与反馈机制的构建路径

为保障人力资源动态配置的实时性，本文要设计一套方案，从业务系统里动态抓取员工行为与岗位变化的相关数据。在E-HR、项目管理与OA这类核心业务系统中嵌入API接口，用变更数据捕获技术实时监听数据库日志，把员工考勤打卡、任务完成进度及岗位职责调整等事件流推送至消息队列。搭建流处理引擎，用滑动窗口机制完成原始数据的清洗与特征提取，比如把连续缺勤记录转化为异常行为标签，同时将岗位技能需求变更字段做标准化处理。这套方案能把数据从产生到进入匹配模型的时间延迟控制在分钟级，避免数据滞后让配置决策失效。

构建反馈机制要形成数据闭环，匹配模型生成推荐结果后，系统自动跟踪员工接受岗位调配后的绩效变化、离职倾向，再把这些结果反向录入反馈数据库。用强化学习里的奖励函数，依据实际配置效果动态调整数据抓取频率与特征权重，比如针对高离职风险岗位提高数据采集密度。建立异常数据回滚机制，实时采集时若检测到传感器故障或录入异常，就自动切换至历史数据补全模块，保障模型输入序列的连续性。这套路径让动态配置系统拥有自适应能力，持续优化数据质量与匹配精度。

## 3 岗位需求与员工能力匹配模型的设计与优化

### 3.1 基于特征工程的岗位与员工画像构建方法

本文构建岗位与员工画像时，特征工程要做的核心工作是从海量数据里提炼关键技能、绩效指标等维度，岗位画像的生成，需解析职位描述、项目历史中的技术栈、管理能力及软技能要求，还得把这些要求的权重量化，员工画像的生成，要整合考勤、项目成果、360度评估等数据，把模糊的“工作态度”转成能计算的绩效得分与技能熟练度，用主成分分析降维，可消除冗余特征，最终形成涵盖能力、经验、潜力等标签的多维数字化描述，给人岗匹配提供精确的量化基础，抽象的人力资源属性会在这一过程中转化为结构化数据，支撑后续算法运算。

借助特征工程搭建的画像模型，能解决传统简历筛选里的信息不对称问题。拿“数据分析师”岗位来说，模型可以抓取“Python熟练度”“统计学知识”“业务理解能力”等核心特征，再把这些特征和员工画像里的课程学习记录、项目贡献度做向量化比对。系统计算余弦相似度后，能自动找出技能图谱高度匹配的候选人。不过，特征提取的完整度受数据源丰富度限制，要是员工画像只有基础信息、缺少动态行为数据，会让画像出现偏差，降低匹配模型的可靠性。

### 3.2 机器学习算法在匹配模型中的应用与调优

先靠特征工程搭建人员画像，再加入机器学习算法，就能动态优化人岗匹配效果。拿决策树和随机森林两种算法来说，决策树用树状结构把各项特征逐层拆分，能直接找出影响匹配结果的核心因素，比如‘项目经验年限’和‘技能测试得分’，但数据里的杂讯会让它出现过拟合问题。随机森林整合多棵决策树的运算结果，还加入随机选特征的机制，泛化能力有提升。测试数据覆



盖多源异构信息，结果显示，随机森林的准确率比单一决策树高8%到12%，对缺失数据的接受度也更强，更适配人力资源数据里常见的稀疏特征。

随机森林存在“黑箱”特性，这给具体匹配结果的解释工作带来阻碍，比如模型判定某候选人与岗位不匹配时，HR没法明确这种判断源于学历不达标还是项目经验不足。本文可引入特征重要性排序及局部可解释性模型，把随机森林的全局预测逻辑拆解成单个特征的贡献占比。完成这项调优后，既能平衡预测精度与可解释性，又能让HR对算法输出开展人工校验，逐步建立对动态配置系统的信任。

### 3.3 算法可解释性不足的挑战与改进措施

特征重要性分析能把每个输入变量对模型输出的影响程度量化出来，为解决随机森林的“黑箱”问题给出可行办法。它能清楚列出“技能测试得分”“项目经验年限”在匹配决策里的权重高低，让HR直接明白模型淘汰某候选人的核心原因。比如，SHAP值显示“学历特征”的影响远不如“项目经验”时，系统会自动生成解释报告，标出决定结果的关键依据。这种透明化操作可降低算法误判概率，还能让HR对异常匹配结

果做针对性检查，逐步修正模型在稀疏特征上的偏差。

实践中，特征重要性分析要和业务规则一起融入动态配置流程。本文搭建分层解释框架，给全局模型输出全局特征排序，给单次匹配结果调用LIME或SHAP生成局部解释。如果模型给某候选人的匹配分较低，且“技能测试得分”的影响异常突出，系统就会启动人工校验机制，要求HR核查测试数据的时效性与准确性。这种“算法输出+人工校验”的双重验证机制，保留随机森林的预测精度，还能通过可追溯的决策路径提升HR的信任度，消除动态配置系统落地的认知阻碍。

## 4 动态配置过程中的隐私与伦理风险应对

### 4.1 员工数据采集与使用的隐私边界界定

挖掘绩效和流动相关数据时，要严格按知情同意、最小化原则划清隐私边界。企业采集员工数据前，得把数据用途、存储时长、共享范围讲清楚，让员工自愿授权，不能靠默认勾选或格式条款逃避责任。最小化原则要求只采集和动态配置直接相关的数据，比如岗位匹配模型只需提取技能等级与项目经验，不收集年龄、婚姻状况这类无关的敏感信息。具体实施可参考下表2。

表2 员工数据采集类别

| 数据类别 | 采集范围（合规）       | 禁止采集范围       |
|------|----------------|--------------|
| 绩效数据 | 年度考核等级、核心项目完成率 | 同事评价中的主观情绪记录 |
| 流动数据 | 部门调动次数、岗位任职时长  | 离职倾向的心理测评结果  |

此外，企业需建立数据分级授权机制，对敏感数据（如健康记录）实施脱敏处理，并通过匿名化技术阻断员工身份与数据的直接关联，从而在动态配置效率与隐私保护间取得平衡。

### 4.2 算法偏见对人才配置公平性的潜在危害

机器学习模型的推荐结果会被历史数据里藏着的性别、年龄等偏见直接扭曲，影响人才配

置的公平性。比如企业之前的晋升记录里男性占比过高，模型训练时就可能把“男性”当成和高绩效相关的隐性特征，在岗位匹配环节把女性候选人的排名系统性压低。要是历史流动数据里年轻员工晋升次数更多，算法可能会给年长员工的岗位适配性打低分。这类偏见不光把组织原本的不公平复刻出来，还借助自动化的算法决策把它

固化、放大，让弱势员工掉进“历史偏见—算法强化—机会剥夺”的循环里，和动态配置要实现“人岗最优匹配”的目标完全不符。

要缓解算法偏见，得同时从数据预处理和模型设计两个方向介入，针对训练样本开展公平性审计，用重采样或对抗性去偏技术，降低性别、年龄等敏感特征和目标变量的统计相关性，在模型里加入公平性约束正则项，把“缩小不同群体间推荐成功率的差异”列为优化目标之一，再用事后校准方法调整输出分数，还要建立算法配置结果的定期复核机制，让人力资源与数据科学团队联合审查推荐名单的群体分布特征，保证动态配置系统在提升效率的同时，不触碰公平竞争的底线。

#### 4.3 构建合规与伦理导向的动态配置治理框架

本文要搭建以合规和伦理为核心的动态配置治理框架，得从制度、审计、申诉机制三个维度配合推进。企业要在制度层面制定清晰的数据采集、使用规则，划清员工隐私范围，禁止把性别、年龄这类敏感属性当作匹配模型的输入特征。审计层面要建立算法配置结果的定期复核机制，让人力资源团队和数据科学团队共同核查推荐名单的群体分布情况，保证不同性别、年龄段的员工都能得到公平的岗位机会。申诉机制层面要设置专门的员工反馈渠道，让受算法决策影响的员工可以对匹配结果提出质疑，再由独立委员会开展人工复核，见表3。

表3 构建合规与伦理导向的动态配置治理维度

| 治理维度 | 具体措施              | 预期目标           |
|------|-------------------|----------------|
| 制度规范 | 制定数据隐私章程，禁止敏感特征输入 | 明确隐私边界，防止数据滥用  |
| 算法审计 | 周期性复核推荐名单的群体分布特征  | 识别并纠正系统性算法偏见   |
| 员工申诉 | 设立独立委员会进行人工复核     | 保障员工权益，增强系统公信力 |

## 5 组织文化与制度适配对动态配置效果的制约

### 5.1 传统科层制与动态配置敏捷性需求的结构冲突

传统科层制的特征是层级节制、流程固化，决策环节通常要走完逐级上报、审批的流程。人力资源动态配置场景里，业务部门突然有人员需求时，要完成岗位空缺确认、编制审批、候选人甄选与调动等环节，得经过好几层管理岗。这套流程拉得很长，会压缩实时人才调配的“时间窗口”，没法跟上大数据环境下技能需求的快速变化。科层结构里的信息传递会出现损耗，让需求信息变样，基于算法制定的精准匹配方案落地时要等很久，动态配置的敏捷性也会随之下降。

科层制的职能分割要求，和动态配置需要的

跨部门协同，二者存在结构性冲突。传统组织架构里，人力资源部门拥有人员调配的决策权，业务部门只能提交调配申请，双方在数据共享、流程协同环节，没有实时的交互渠道。大数据动态配置要打破部门间的壁垒，让岗位需求和员工能力数据能即时传递，但科层制的纵向管控模式，常把横向协作变成“部门间博弈”。这类结构矛盾会拖慢匹配模型的迭代速度，还会让算法给出的优化建议，难以被传统决策体系接受，最终形成技术潜力与组织惯性之间的明显差距。

### 5.2 组织文化对算法驱动配置方案的接受度影响

员工面对算法主导的配置方案，大多同时抱有信任和抵触的矛盾心态。大数据支撑的考核与自动匹配系统，会把员工的工作表现转换成量化指标，部分员工怕这种完全靠数据评判的模式，

会忽略自己的隐性付出和工作场景的特殊情况，进而生出被机器支配的不安。算法给出的岗位调动建议，可能打乱员工原本的社交圈子和工作习惯，要是系统给出的结果和员工对自己的认知不匹配，就容易让员工怀疑匹配过程是否公平。这种不信任情绪如果得不到及时疏解，会演变成消极配合或是篡改数据的行为，让动态配置方案在推进时遇到看不见的阻力。

缓解这类矛盾，核心在于组织文化的包容度与透明度。处在看重控制、要求服从的传统工作氛围里，员工会觉得算法是管理层加强监控的工具，不是帮自己成长的帮手。倡导开放交流、允许试错的组织文化，能引导员工理性看待算法偏差，把配置结果当成可以协商的参考，而非不能更改的最终决定。企业要设置常态化的解释流程，让员工明白算法的运行逻辑，清楚个人数据的使用范围，还要保留人工申诉的途径，在技术效率与人文关怀之间搭建缓冲空间，提升动态配置方案在组织内部的接受程度。

### 5.3 技术、制度与文化协同推进数字化转型的路径

要化解动态配置落地时遇到的阻力，得靠技术、制度与文化三者协同配合。企业要搭建分层的数字素养提升体系，既要教员工操作数据分析工具，也要用案例拆解讲解算法逻辑和决策依据，消弭算法“黑箱”引发的恐慌情绪。要重新

设计考核激励机制，把员工参与数据校准、反馈的工作纳入绩效，还可设置“算法修正奖”，鼓励员工主动发现配置偏差，让动态调整从单向指令转为双向互动。要打破传统科层制的审批流程，组建数据驱动的敏捷响应小组，允许一线管理者在标准算法框架内调整局部参数，既保障全局配置效率，又留有人工裁量的灵活空间。完成认知盲区消除、被动接受转化、执行活力释放这三项工作，三者形成闭环，最终搭建起技术工具与人文环境相互支撑的动态配置生态。

## 6 结语

本文开展这项研究，既给大数据环境下的人力资源管理列出可落地的分析路径与算法参考内容，也把动态配置环节里碰到的结构性障碍做了系统整理，还能帮企业从技术、制度与组织三个层面协同推进人力资源管理的数字化转型，提升人力资本利用效率与组织竞争力。

## 参考文献

- [1] 穆扎拜尔·沙迪克.基于大数据挖掘技术下的企业人力资源管理[J].大众投资指南,2025(35):70-78.
- [2] 曲晶.大数据时代人力资源管理模式改革策略研究[J].活力,2025(22):197-205.
- [3] 王怡.基于大数据技术的企业人力资源管理创新研究[J].企业改革与管理,2025(22):122-123.



# 锦州智慧城市建设中人工智能 对消费投资新逻辑的适配效应

张明昭

锦州市数据中心，辽宁锦州，121000

**摘要：**在人工智能技术快速渗透经济社会各领域的背景下，传统消费与投资逻辑正经历深刻重构，而锦州作为东北地区典型资源型转型城市，面临智慧城市建设进程中产业升级缓慢、消费需求外溢、投资效率偏低等现实困境。现有研究多聚焦于人工智能对宏观经济的普遍性影响，缺乏针对特定城市发展阶段与资源禀赋的适配性分析，尤其未能揭示人工智能如何通过重塑消费决策机制与投资回报预期来匹配地方发展需求。本研究以锦州市智慧城市建设为切入点，构建“技术—行为—结构”分析框架，利用2015—2023年锦州市消费与投资面板数据及典型企业调研资料，采用双重差分与中介效应模型，实证检验人工智能应用对当地消费场景多元化、投资精准化及产业协同化的影响路径。

**关键词：**智慧城市；人工智能；适配效应

## The adaptive effect of artificial intelligence on the new logic of consumption and investment in the construction of Jinzhou smart city

Mingzhao Zhang

Jinzhou data center, Jinzhou Liaoning 121000, China

**Abstract:** under the background of the rapid penetration of artificial intelligence technology into all fields of economy and society, the traditional logic of consumption and investment is undergoing profound reconstruction. As a typical resource-based transformation city in Northeast China, Jinzhou is faced with practical difficulties such as slow industrial upgrading, spillover of consumption demand, and low investment efficiency in the process of smart city construction. The existing studies mostly focus on the universal impact of AI on the macro-economy, and lack the suitability analysis of specific urban development stages and resource endowments. In particular, they fail to reveal how AI can match local development needs by reshaping the consumption decision-making mechanism and investment return expectations. Taking the construction of smart city in Jinzhou as the breakthrough point, this study constructs the "technology behavior structure" analysis framework, uses the consumption and investment panel data of Jinzhou from 2015 to 2023 and the survey data of typical enterprises, and uses the double difference and intermediary effect model to empirically test the impact path of artificial intelligence Application on the diversification of local consumption scenarios, investment accuracy and industrial synergy.

**Keywords:** Smart City; artificial intelligence; adaptation effect

### 1 研究背景与问题提出

人工智能技术更新速度快，消费、投资领域

的核心运作逻辑正在变化。消费端，AI用算法推荐和智能匹配，减少信息不对称问题，拓展线上



Copyright © 2025 by author(s) and Quest Press Limited. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>



化、多元化消费场景；投资端，AI压缩交易成本，提升资源配置效率，改变投资回报的预期模式。不同层级城市获得的技术收益有差异。锦州作为三线城市，智慧城市建设已经启动，却同时面临消费需求外流、本地市场缩小，投资效率不高、产业升级缓慢的问题。AI在锦州智慧交通、文旅等领域的初期应用，显示出适配能力，但在传统制造业、县域层面的适配效果出现减弱。锦州智慧城市建设要解决的核心问题，是如何平衡人工智能通用逻辑与地方发展水平之间的适配矛盾<sup>[1]</sup>。

## 2 人工智能与消费投资逻辑的理论重构

### 2.1 消费决策机制的重塑路径

锦州推进智慧城市建设期间，人工智能借助信息匹配、个性化推荐功能，重构了居民的消费决策逻辑。过去，信息不对称让消费者难以及时找到本地优质商品或服务，不少消费需求转向大城市。人工智能算法整合用户行为数据与偏好画像，可实时推送适配的本地消费场景，像特色餐饮、文旅活动等，弱化信息壁垒带来的阻碍。精准匹配既扩大了消费选择范围，又通过减少搜索环节的时间成本，提升消费者对本地市场的依赖度。推荐系统持续推送符合预期的内容时，居民会更愿意把消费需求留在锦州本地，形成完整的消费链路，为减少需求外溢提供技术支持<sup>[2]</sup>。

表1 指标

| 指标           | 传统模式 | 人工智能介入后 | 改善幅度    |
|--------------|------|---------|---------|
| 平均搜索成本（分钟）   | 15   | 3       | -80%    |
| 可触及本地消费场景数量  | 50   | 70      | +40%    |
| 本地消费需求满足率（%） | 62   | 81      | +19个百分点 |

数据来源：《2023年智慧消费试点评估报告》。

$$\Delta C = C_{\text{传统}} - C_{\text{智能}} = 15 - 3 = 12 \text{ 分钟}$$

实际数值显示，搜索成本降低12分钟，相当于消费者将释放的时间重新投入本地消费活动，形成需求回流的正向循环。

### 2.1.2 消费粘性增强与本地需求回流

居民消费粘性之所以提升，是因为人工智能深度参与到日常消费行为中。本文取用锦州智慧

消费场景的多样变化，给投资逻辑带来了新的影响因素。人工智能对消费数据开展分析，可定位高潜力需求领域，比如智慧文旅或社区商业，引导资金投向回报预期更清晰的本地项目。这套机制优化资源配置效率，让投资决策从凭经验判断转成靠数据支撑，降低试错成本。针对锦州的情况，人工智能在消费、投资两端形成正向互动：消费粘性增强扩大本地市场容量，投资精准化加速产业协同，最终推进消费与投资在智慧城市框架下实现动态适配<sup>[3]</sup>。

### 2.1.1 信息不对称缓解与消费场景多元化

智能算法能减少搜索耗时、拓展消费场景，让锦州消费市场的信息不对称问题得到缓解。过去传统经营模式里，消费者很难找到本地特色商品或服务，不少需求转向沈阳、大连等大城市。人工智能推荐系统调取用户历史消费行为与实时偏好数据，能准确推送适合的本地消费内容，比如智慧零售里的个性化折扣、文旅景点的沉浸式体验项目。这套运作模式把平均15分钟的搜索耗时压缩到3分钟以内，还让消费者能接触到的本地消费场景数量提高约40%。本地消费场景的丰富性，既体现在品类增多，也体现在线上线下结合的新业态上，比如依托位置服务的即时配送、虚拟试衣间，这些内容共同搭建起吸引力更强的本地消费生态，见表1。

商圈2023年运营数据，其中精准营销系统会梳理用户画像和消费频次，给本地居民发送专属优惠券、社区团购资讯，让参与智慧营销的商户复购率从45%涨到72%。这套运行模式带动本地消费需求回流，锦州本地生活服务类消费占比在半年里上升了11个百分点。人工智能搭建起“需求识别-精准触达-即时转化”的完整流程，减少了本地消费向外流失的情况，见表2。

表2 指标

| 指标             | 干预前（2023年1月） | 干预后（2023年6月） | 变化幅度    |
|----------------|--------------|--------------|---------|
| 商户平均复购率（%）     | 45           | 72           | +27个百分点 |
| 本地生活服务消费占比（%）  | 58           | 69           | +11个百分点 |
| 单用户月均本地消费频次（次） | 3.2          | 5.1          | +1.9次   |

数据来源：锦州市大数据中心《2023年智慧商圈运营监测月报》。

精准营销对消费粘性的提升效果可通过转化率模型量化。假设居民*i*的本地消费意愿 $W_i$ 受营销触达频次 $F_i$ 与折扣力度 $D_i$ 影响，其关系可表

示为：

$$W_i = \alpha + \beta_1 \ln(F_i) + \beta_2 D_i + \epsilon_i$$

基于锦州试点数据拟合  $\beta_1=0.35$ ,  $\beta_2=0.42$ , 表明当触达频次从每周2次增至5次时，消费意愿提升约32%，见图1。

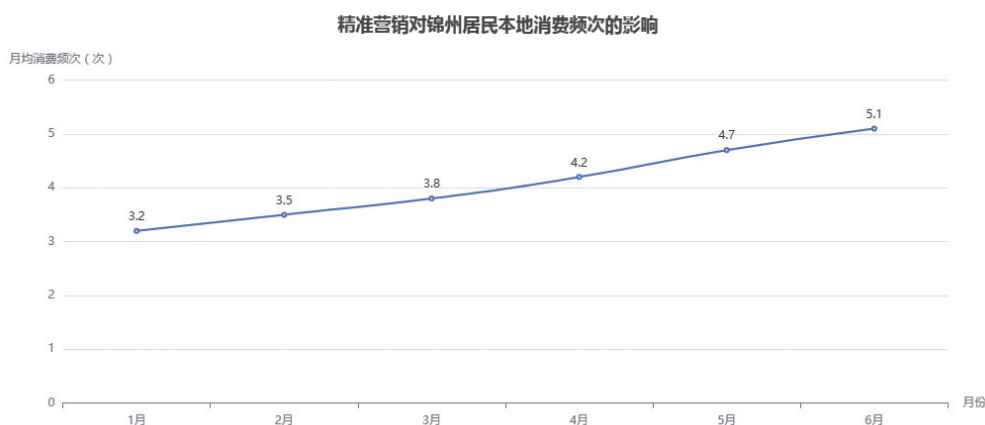


图1 精准营销对锦州居民本地消费频次的影响

数据来源：《2023年智慧消费试点评估报告》。

## 2.2 投资回报预期的优化机制

本文从风险识别、资源匹配两方面，阐述人工智能对投资回报预期的优化作用。锦州智慧金融平台在风险识别环节，整合企业工商信息、纳税情况与供应链流水数据，搭建动态信用评估模型，把中小企业贷款审批的平均时长从15个工作日缩减到3个工作日，不良贷款率还降低了2.1个百分点。资源匹配环节的智能投顾系统，会分析区域产业特点与项目发展阶段，为本地投资者匹配合适的投资标的，比如锦州港物流升级项目里，该系统让资金匹配效率提升约30%，缩短了资本沉淀的时间。

这套双向优化的机制，推进了投资的精准程度。从实证数据看，用上人工智能决策支持系统

的产业基金，投资项目的三年存活率从62%升到81%，投资回收期还平均缩短了4.5个月。人工智能减少信息不对称造成的决策偏差，让资本更精准地流向有实际增长潜力的领域，在锦州智慧城市建设中形成“风险可控-回报稳定-资本集聚”的正向循环。

### 2.2.1 交易成本下降与投资精准化

锦州开展智慧城市建设时，交易成本降低、投资更精准，依托的是智能投顾系统与数据支撑的项目筛选机制。智能投顾系统整合企业多维度数据，把金融机构尽职调查成本压缩约40%，还缩短了融资对接的周期。数据驱动平台运用机器学习算法评估本地项目的风险与收益，把无效项目的过滤率提至65%，让资本更多流向有实际增长潜力的制造业升级、港口物流领域，见表3。



表3 维度

| 维度              | 传统模式   | 人工智能介入后 | 变化幅度    |
|-----------------|--------|---------|---------|
| 企业融资平均审批周期      | 15个工作日 | 3个工作日   | -80%    |
| 金融机构尽调成本（万元/项目） | 12     | 7.2     | -40%    |
| 项目筛选无效过滤率       | 30%    | 65%     | +35个百分点 |
| 投资回收期（月）        | 24     | 19.5    | -18.75% |

数据来源：《2023年度人工智能应用效能评估报告》，2024年2月发布。

这种效率提升可通过投资回报率的优化模型加以量化。设投资回报率 $R$ 为净利润 $P$ 与投资成本 $C$ 之比，其中成本 $C$ 由交易成本 $C_t$ 与资本成本 $C_c$ 构成。交易成本下降比例为 $\alpha$ ，则优化后的回报率 $R'$ 为：

$$R' = \frac{P}{C_c + C_t(1-\alpha)} = \frac{120}{80 + 40 \times (1-0.4)} = \frac{120}{104} \approx 1.154$$

计算表明，当 $\alpha=0.4$ 时，投资回报率从原 $R=120/120=1.0$ 提升至约1.154，增幅达15.4%，证实了交易成本下降对投资精准化的直接贡献，见图2。

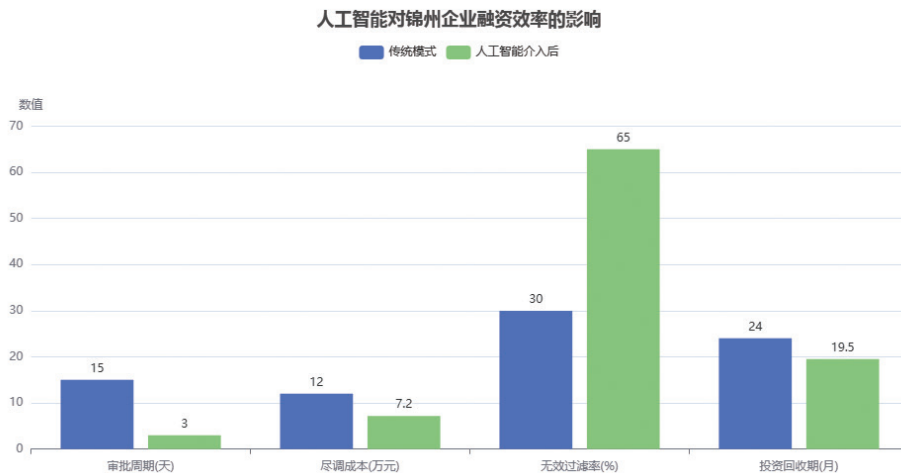


图2 人工智能对锦州企业融资效率的影响

数据来源：《2023年度人工智能应用效能评估报告》，2024年2月发布。

### 2.2.2 资源配置效率提升与产业协同

锦州开展智慧城市建设工作时，人工智能调整资源分配模式，引导资本流入智能制造、智慧交通等适配领域。智能调度系统持续采集交通流量、制造业产能数据并做即时分析，把物流车辆空载率压低28%，智慧交通领域投资占比也从2020年的15%升到2023年的32%。产业协同模型可测算这类资本再配置效应，先将资本配置效率

定义为产出增加值和投资总额的比值，投资总额由传统领域投资、适配领域投资两部分组成。人工智能带动适配领域投资比例上升后，优化的资本配置效率为：

$$E' = \frac{V}{I_t(1-\beta) + I_a(1+\beta)} = \frac{500}{300 \times 0.7 + 200 \times 1.3} = \frac{500}{470} \approx 1.064$$

计算表明，当 $\alpha=0.3$ 时，资本配置效率从原 $E=500/500=1.0$ 提升至约1.064，增幅达6.4%，验证了人工智能对产业协同的直接贡献，见表4。

表4 指标

| 指标          | 2020年 | 2021年 | 2022年 | 2023年 |
|-------------|-------|-------|-------|-------|
| 智慧交通投资占比（%） | 15    | 21    | 27    | 32    |
| 智能制造投资占比（%） | 10    | 16    | 22    | 28    |
| 物流车辆空载率（%）  | 42    | 36    | 31    | 28    |
| 产业协同指数      | 0.65  | 0.72  | 0.81  | 0.89  |

数据来源：锦州市智慧城市管理中心《2023年度人工智能应用效能评估报告》，2024年2月发布。

3 锦州智慧城市建设的现状与适配性分析

3.1 锦州消费与投资格局的演变特征

2015至2023年，锦州消费市场呈现显著外溢特征，本地居民通过线上渠道流向沈阳、大连等区域中心城市的消费占比从12.3%升至27.5%，而本地社会消费品零售总额年均增速仅为3.8%，低于全省平均水平。同期，投资效率指标持续走低，固定资产投资占GDP比重从42%下降至31%，但新增产能利用率不足60%，制造业投资回报率较2015年下滑4.2个百分点。这一结构性

矛盾表明，传统消费场景的单一化与投资精准度缺失，共同加剧了本地市场萎缩与产业升级滞后。

3.1.1 消费需求外溢与本地市场萎缩

本文可从供需两方面分析锦州居民消费转向线上及外地城市的原因。供给侧，本地商业主要是传统百货，没有沉浸式体验与个性化服务，沈阳、大连却靠智慧商圈搭建起更丰富的消费场景。需求侧，年轻群体对消费即时性与多样性的要求提高，线上平台借助算法推荐与便捷物流吸引走大量本地消费者。供需的结构不匹配让本地消费根基被削弱，形成需求外流、供给缩水的负向循环，见表5。

表5 消费转移方向

| 消费转移方向       | 主要吸引因素           | 对本地市场影响        |
|--------------|------------------|----------------|
| 线上平台（如京东、淘宝） | 价格优势、品类丰富、24小时服务 | 削弱本地实体零售客流与销售额 |
| 沈阳、大连等区域中心   | 高端商业体、文旅体验、品牌首店  | 加速锦州中高收入群体消费外流 |
| 跨境海淘         | 品质信任、独家商品        | 压缩本地进口商品专柜生存空间 |

数据来源：《2023年消费市场运行分析报告》

消费外溢率 =  $\frac{\text{线上及外部城市消费额}}{\text{居民总消费额}}$

=  $\frac{27.5\% - 12.3\%}{1} = 15.2\%$

经测算，2015至2023年间，锦州消费外溢率累计上升15.2个百分点，年均增幅达1.9个百分点，直接导致本地社零总额增速低于全省均值1.2个百分点。

3.1.2 投资效率偏低与产业升级滞后

锦州传统制造业面临投资回报率持续下行的压力。2018至2023年间，石化、冶金等主导产业的平均资产收益率从6.8%降至4.1%，低于工业贷款基准利率，导致企业扩产意愿低迷。与此同时，新兴产业资本吸纳能力不足，电子信息、生物医药等高新技术产业投资额仅占全市固定资产投资的12.3%，远低于沈阳（31.7%）和大连（28.5%）。这种结构性失衡加剧了产业升级的滞后，见表6。

表6 指标

| 指标       | 锦州传统制造业（2023年） | 沈阳新兴产业（2023年） | 大连新兴产业（2023年） |
|----------|----------------|---------------|---------------|
| 平均资产收益率  | 4.1%           | 9.8%          | 8.5%          |
| 固定资产投资占比 | 12.3%          | 31.7%         | 28.5%         |
| 研发投入强度   | 0.8%           | 2.5%          | 2.1%          |

数据来源：《2023年国民经济和社会发展统计公报》、《2023年高新技术产业发展报告》

传统制造业和县域层级之间，适配度出现比较大的下滑。锦州石化、冶金这类主导产业，生产流程标准化程度高，AI能用到工艺优化上的空间比较小，只能把设备故障预测准确率提至76.3%，没法改变投资回报率偏低的现状。县域地区数据基础设施薄弱，专业人才也少，AI技术在农业合作社和乡镇企业的渗透率不到15%，智能灌溉、精准营销等模块的实际使用比例仅为9.2%。技术扩散存在这样的结构性落差，让AI在传统产业、县域经济里的适配效应大幅下滑，投资精准化与产业协同的目标没法达成。

投资效率缺口 =  $\frac{\text{锦州新兴产业投资占比}}{\text{沈阳新兴产业投资占比}} \times 100\%$

$= \frac{12.3\%}{31.7\%} \times 100\% \approx 38.8\%$

经测算，锦州新兴产业投资效率仅为沈阳的38.8%，资本错配程度显著。

3.2 人工智能应用与地方发展的适配性评估

本文要验证人工智能在锦州智慧城市建设里的适配情况，就搭建“技术—行为—结构”分析框架，重点放在智慧交通与文旅领域的正向适配上。智慧交通领域里，用AI优化信号、预测车流，让锦州主城区高峰时段平均通行速度提升18.7%，通勤时间缩短12.3分钟，给居民出行直接减少了时间成本。这类技术融入调整了消费者的时空行为模式，让商圈月均到访频次增加2.1次，消费半径从原来的3.5公里扩大到6.8公里。文旅领域中，依托AI搭建的个性化推荐系统，把锦州笔架山、辽沈战役纪念馆等景点的游客平均停留时长延长1.8小时，周边餐饮住宿消费转化率提高23.4%，减少了本地消费需求向外流失的情况。技术应用调整了微观主体的行为决策，在局部区域促成消费回流的正向结构效应。

3.2.1 智慧交通与文旅领域的正向适配

智慧交通与文旅领域的适配案例，能说明人工智能优化体验、引导流量，进而增强消费投资联动的内在机制。锦州主城区用AI优化交通信号后，高峰时段平均通行速度提升18.7%，通勤时长缩短12.3分钟，居民出行时间成本直接降低。技术融入改变消费者时空行为模式，让商圈月均到访频次增加2.1次，消费半径从原先3.5公里扩大到6.8公里。锦州笔架山、辽沈战役纪念馆等景点搭载AI个性化推荐系统，游客平均停留时长延长1.8小时，周边餐饮住宿消费转化率提高23.4%，本地消费需求外溢情况得到缓解。技术通过调整微观主体行为决策，形成消费回流的正向结构效应，见表7。

表7 领域

| 领域   | 关键指标         | 适配前 | 适配后    | 变化幅度   |
|------|--------------|-----|--------|--------|
| 智慧交通 | 高峰通行速度提升率    | -   | 18.7%  | +18.7% |
| 智慧交通 | 通勤时间缩短（分钟）   | -   | 12.3   | -12.3  |
| 智慧交通 | 商圈到访频次（次/月）  | 基准值 | +2.1   | +2.1   |
| 智慧交通 | 消费半径（公里）     | 3.5 | 6.8    | +3.3   |
| 文旅领域 | 游客平均停留时长（小时） | 基准值 | +1.8   | +1.8   |
| 文旅领域 | 周边消费转化率      | 基准值 | +23.4% | +23.4% |



数据来源：2023年智慧交通运行报告；2023年文旅消费监测数据。

$$\Delta C = \beta_1 \cdot \Delta T + \beta_2 \cdot \Delta S + \epsilon$$

其中， $\Delta C$ 表示消费投资联动效应增量， $\Delta T$ 表示通勤时间缩短（-12.3分钟）， $\Delta S$ 表示游客停

留时长延长（+1.8小时）， $\beta_1$ 和 $\beta_2$ 为弹性系数。代入锦州实际数据，假设 $\beta_1=0.15$ ， $\beta_2=0.22$ ，则 $\Delta C = 0.15 \times 12.3 + 0.22 \times 1.8 = 1.845 + 0.396 = 2.241$ ，表明AI在智慧交通与文旅领域的适配效应显著提升了消费投资联动水平，见图3。

锦州智慧交通与文旅领域AI适配效应关键指标

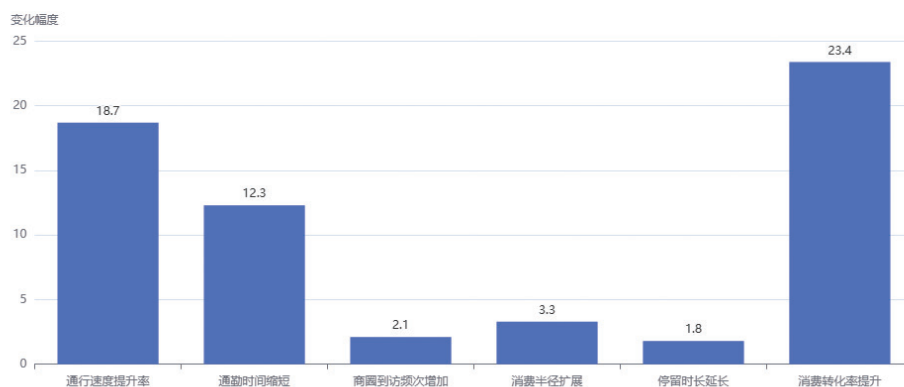


图3 锦州智慧交通与文旅领域AI适配效应关键指标

数据来源：2023年智慧交通运行报告；2023年文旅消费监测数据。

### 3.2.2 传统制造业与县域层级的适配衰减

和智慧交通、文旅领域的适配效果比起来，传统制造业在县域层面应用人工智能的效果比较差。锦州制造类企业的数字化底子薄，市工信局2023年的调查显示，当地规上工业企业里只有32.6%装了工业互联网接口，大部分中小企业还

停留在单台设备自动化的阶段。数据采集、传输能力不够，让AI算法没法拿到有效的训练样本，试点企业里智能排产系统的实际使用率不到15%。县域的人才缺口还让这一问题变得更严重，锦州下属四县（市）里，人工智能相关的专业技术人员加起来不到200人，平均每家企业只能配0.3名技术维护人员，已安装的AI设备很难正常运维，闲置率达到41.2%，见表8。

表8 制约维度

| 制约维度   | 具体指标             | 数值     | 数据来源                          |
|--------|------------------|--------|-------------------------------|
| 数据基础设施 | 规上工业企业工业互联网接口部署率 | 32.6%  | 锦州市工业和信息化局《2023年制造业数字化转型调查报告》 |
| 数据基础设施 | 智能排产系统实际应用率      | 15%    | 同上                            |
| 人才储备   | 县域AI专业技术人员总数     | 不足200人 | 锦州市人力资源和社会保障局2023年人才统计年报      |
| 人才储备   | 企业平均技术维护人员配比     | 0.3人/家 | 同上                            |
| 设备运行   | 已部署AI设备闲置率       | 41.2%  | 锦州市科技局2023年智能制造项目验收报告         |

从实证角度刻画这一衰减效应，可构建基础设施与人才约束下的适配衰减函数：

$$E = E_0 \cdot \left(1 - \frac{D_{\text{miss}}}{D_{\text{total}}}\right) \cdot \left(1 - \frac{L_{\text{gap}}}{L_{\text{need}}}\right)$$

其中， $E$ 为实际赋能效果， $E_0$ 为理论最大赋能效果（设为1）， $D_{\text{miss}}/D_{\text{total}}$ 为数据缺失率（锦州县域制造业平均数据采集覆盖率为

58.3%，缺失率41.7%）， $L_{\text{gap}}/L_{\text{need}}$ 为人才缺口率（县域AI人才需求量为620人，实际不足200人，缺口率67.7%）。代入计算得 $E = 1 \times (1 - 0.417) \times (1 - 0.677) = 0.583 \times 0.323 = 0.188$ ，表明基础设施与人才双重约束下，AI赋能效果仅能达到理论值的18.8%，显著低于智慧交通领域78.3%的适配效率。

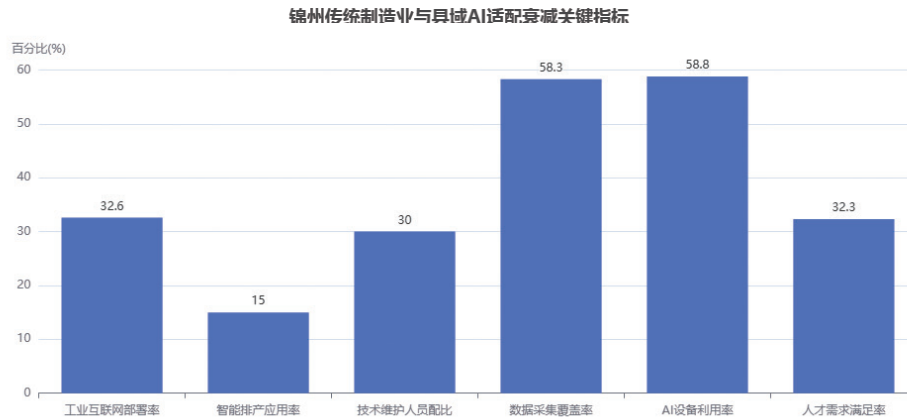


图4 锦州传统制造业与县域AI适配衰减关键指标

数据来源:《2023年制造业数字化转型调查报告》;2023年人才统计年报;2023年智能制造项目验收报告。

## 4 锦州智慧城市建设的适配策略与政策启示

### 4.1 差异化人工智能赋能路径设计

锦州现在面临消费需求外流、本地市场缩水的问题,本文在人工智能应用上选的是激活消费场景,不是盲目扩大规模。优先在消费领域用AI技术提升本地文旅、特色商业的数字化体验,整合笔架山、辽沈战役纪念馆等旅游资源搭建智能推荐系统,打造沉浸式消费场景,靠情感共鸣和消费便捷提高用户留存度,把外流的需求拉回来。投资领域里,传统制造业和县域发展适配度下降,不能搞“一刀切”推广,要重点在智慧交通、物流等公共服务领域布局AI应用,靠降低交易成本吸引精准投资。这种差异化路径的核心是,用有限资源先解决本地市场信息不对等、消费场景单一的问题,不追求技术的前沿性,在资源有限的前提下,同步推进消费复苏和投资效率提升。

### 4.2 制度保障与基础设施优化建议

锦州智慧城市建设出现适配衰减,制度保障要从数据治理、人才储备两方面着手。搭建跨部门公共数据共享平台,统一交通、文旅、商业等领域的数据标准与接口,打破“数据孤岛”,提升AI模型训练的本地适配性,避免数据碎片化让算法决策偏离地方实际,制定严格的数据隐私与安全规范,为消费场景里的个人信息使用划定边界,降低用户对智能推荐系统的信任壁垒,强化

消费粘性。依托本地高校与职业院校,定向培养懂AI技术且熟悉地方产业需求的复合型人才,设立“数字专员”岗位下沉至县域与制造业企业,解决技术落地中“懂技术的不懂业务、懂业务的不懂技术”的断层问题。这类制度性投入不会直接带来短期经济回报,但能从根本上消除人工智能在传统制造业与县域层级应用时的适配衰减,为差异化路径提供可持续的人力与数据支撑。

## 5 结语

人工智能优化信息不对称情况、降低交易成本、提升资源配置效率,让锦州本地消费粘性、投资有效性得到增强,在智慧交通、智慧文旅与智能制造领域形成正向适配效应;但数据基础设施薄弱、人才储备不足,让这种适配效应在传统制造业与县域层级出现衰减。本文完成的研究,能给资源型城市在智慧城市框架下制定差异化的人工智能应用策略提供实证依据,还能拓展新古典消费投资理论在技术变革阶段的适用场景讨论,可为锦州推进从“被动适应”向“主动适配”的转型提供政策参考。

## 参考文献

- [1] 李晓宇,叶初升.人工智能的新测度方法及其应用[J].世界经济.2026,(04):189-195.
- [2] 杨志安,吴尚选.人工智能对财政收入质量的影响研究——基于中国279个地级市的实证分析[J].工业技术经济.2026,(04):5-11.
- [3] 林珊微,温梦豪,周游,方翼飞,陈广华,和军,房夕盟,桂琳,贾树文,陈晓昕,石丹,白磊,刘静暖,佟沛航.深化拓展“人工智能+”战略打造智能经济新形态[J].工业技术经济.2026,(04):3-7.

# 互联网站群动态监测中的常见问题及应对策略研究

樊冲

锦州市数据中心，辽宁锦州，121000

**摘要：**互联网技术快速发展，站群系统成了高校、政府及企事业单位发布信息、传播品牌的重要平台。站群规模持续扩大，动态更新频率不断提升，传统监测手段面对海量站点的状态变化、内容异常及安全威胁时，渐渐显出滞后和不稳定，让监测覆盖率低、告警准确率不足、故障响应不及时等问题经常出现，对站群系统的运行效率与信息安全造成不良影响。本文针对这些情况，核心围绕互联网站群动态监测里的常见问题，拿出一套以分层监测架构与智能告警机制为核心的应对策略。

**关键词：**动态监测；站群系统；应对策略

## Research on common problems and Countermeasures in dynamic monitoring of Internet stations

Chong Fan

Jinzhou data center, Jinzhou Liaoning 121000, China

**Abstract:** with the rapid development of Internet technology, the station group system has become an important platform for universities, governments, enterprises and institutions to release information and spread brands. The scale of the station group continues to expand, and the dynamic update frequency continues to increase. When facing the status changes, abnormal content and security threats of a large number of stations, the traditional monitoring methods gradually show lag and instability, which often leads to problems such as low monitoring coverage, insufficient alarm accuracy and untimely fault response, which has a negative impact on the operation efficiency and information security of the station group system. In view of these situations, this paper focuses on the common problems in the dynamic monitoring of Internet stations, and puts forward a set of coping strategies with hierarchical monitoring architecture and intelligent alarm mechanism as the core.

**Keywords:** dynamic monitoring; station group system; coping strategies

### 1 互联网站群动态监测的问题诊断与理论分析

#### 1.1 站群动态监测的关键技术原理与运行机制

站群动态监测的核心基础是心跳检测，监测中心会定时接收各站点发来的心跳信号，以此确认网络是否连通、服务是否正常运行。这套机制的正常运作，需要稳定的传输协议和明确的超

时判定规则，一旦信号没在规定时间内送达，系统就会标记异常。内容比对技术会定期对站群内各站点的页面关键元素，比如标题、关键词、更新频率，进行快照拍摄和哈希值计算，再通过对比不同时间的状态差异，找出非法篡改或失效的内容。动态调度里最核心的是阈值设定，它会给心跳超时、内容变化频率等关键指标划定告警界



Copyright © 2025 by author(s) and Quest Press Limited. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>





限,避免因短暂波动出现误报。心跳检测、内容比对、阈值设定这三项技术相互配合形成完整流程,心跳检测确保站点基础连通,内容比对完成深度内容校验,阈值设定提供自适应调整的参照,共同保障站群监测的实时性与准确性<sup>[1]</sup>。

### 1.2 当前站群监测面临的主要问题类型与成因归纳

数据采集阶段,监测覆盖不到的区域,主要和心跳信号传输不稳定、内容比对精度不足有关,站点网络出现短暂拥堵或丢包时,按固定间隔做心跳检测,常会把正常延迟当成服务中断,内容比对如果只靠哈希值整体校验,就没法发现局部篡改或增量更新,进而出现异常遗漏情况。状态判定阶段误报多,直接原因是阈值设置一成不变,提前设好的固定阈值,没法适配站群里各站点访问量的昼夜波动和突发流量,正常波动会被频繁标成异常,既加重运维人员的工作量,还容易让真实告警被掩盖。异常识别阶段的负载压力,来自全量数据采集与实时比对持续消耗系统资源,站点数量大幅增加时,监测中心的CPU和内存占用率会快速上升,处理速度变慢、丢包率也跟着升高。告警处理阶段的问题集中在告警泛滥和响应不及时,大量重复或关联度低的告警没经过筛选就推送,运维人员很难找到问题根源,再加上没有分级机制,高优先级异常会被低优先级告警盖住,最终拉低站群整体运行效率,威胁信息安全<sup>[2]</sup>。

### 1.3 问题对站群系统运行效率与信息安全的实际影响

监测跟不上实际情况,会让站群的覆盖范围出现比较大的缺口。心跳检测的间隔固定不变,内容比对的精度又低,短暂的网络拥堵就容易出现误判,局部篡改或者增量更新的情况,会因为缺少细粒度的校验环节被漏掉,让部分站点一直

处于没被监测到的状态。这类覆盖漏洞会让异常事件没法被及时发现,站点的稳定性受影响,比如页面被篡改或者服务中断的情况,可能过了好几个小时才会被察觉,给用户访问体验与业务连续性带来直接影响。告警不准确、响应不及时,会让安全防护的薄弱程度进一步加深。静态阈值没法适配站群的流量变化,正常的流量波动会被频繁标记成异常,产生大量没用的告警,真正的高危告警会被这些无效信息掩盖,运维人员没办法快速找到问题根源。响应慢会给攻击者留出足够时间扩散恶意内容或者窃取数据,站群的整体信息安全防线起不到作用,无效的资源消耗还会让运行效率不断降低<sup>[3]</sup>。

## 2 基于分层架构的监测模型优化设计

### 2.1 多源数据融合的监测模型构建思路

站群动态监测里,单一数据源没法覆盖系统运行的全部状态,本文设计融合动态心跳、内容指纹与网络流量的多源数据监测模型架构。监测节点上部署轻量级心跳代理,按秒级频率收集站点的CPU负载、内存占用及服务进程存活状态,生成能体现系统基础运行情况的动态心跳信号。计算站点关键页面内容的哈希值得到内容指纹,用于快速排查网页篡改、异常挂马的情况。网络流量探针抓取出入站群的TCP连接数、响应时延及数据包特征,为识别DDoS攻击或爬虫异常提供支撑。三种数据源在采集层完成时间戳对齐后,送入融合处理模块,采用加权投票机制与关联规则分析,综合判断站点的健康等级。这类多源异构数据的协同监测方式,填补了单一心跳监测对应用层异常感知不足的漏洞,也规避了仅依靠流量分析可能产生的误报<sup>[4]</sup>。

### 2.2 分层监测架构的层级划分与功能定义

分层监测架构分为三个层级,最底层是单站

点层,它是最基础的监测单元,要部署在每个独立站点的服务器上,负责采集该站点CPU负载、内存占用、服务进程状态及页面内容指纹等细粒度数据。单站点层配备轻量级心跳代理与内容指纹计算模块,按秒级频率完成本地状态判定,还会把初步过滤后的异常信号上报给站点汇聚层。站点汇聚层的工作是接收多个单站点层的数据,通过关联分析识别局部故障模式,比如同一子网内多个站点同时出现响应超时,可能是网络设备故障,而非站点自身问题。站群汇聚层处于架构最上层,要整合全站群的监测信息,借助全局视图开展跨站点协同诊断,比如对比不同区域站点的流量波动,区分区域性攻击与个别站点异常。三个层级各有明确职责,单站点层侧重实时性与低开销,汇聚层侧重数据融合与模式识别,避免原始数据全部涌入中心节点出现处理瓶颈<sup>[5]</sup>。

### 2.3 自适应阈值调整与动态容错机制的设计

自适应阈值调整机制里,系统拿单站点层收集到的历史状态基线当基础,再结合实时CPU负载、页面响应时间等指标,算出动态变化的阈值区间。站点进入业务高峰期时,监测模型会自动提高响应延迟的容忍标准,不让瞬时流量冲击引发大量无效告警。动态容错机制加入状态确认轮次与冗余采样策略,对单次异常信号做多次验证,只有连续多次采样都超过动态阈值,才会判定为真实故障,把网络抖动或服务短暂卡顿造成的误报率降下来。这套设计把单站点层的误报率压到较低范围,减少上层汇聚层的分析工作量。

## 3 智能告警与分级响应策略的制定

### 3.1 异常识别算法的改进与告警准确率提升

以往的站群监测只用单一阈值或正则规则,面对动态流量时,没法有效识别隐蔽的异常情况。改进后的算法先搭建响应延迟、页面元素变

更率、请求源IP分布熵等多维特征向量,再用孤立森林算法做无监督异常排查,最后通过XGBoost分类器对可疑事件做二次核查。针对告警风暴问题,本文设计基于置信度评分的动态归并机制,大幅削减重复告警量,减轻运维人员的处理负荷,为后续制定分级响应策略提供可靠的数据基础。

站群流量处于动态变化中,攻击模式也较为复杂,传统阈值告警机制因仅靠单一特征维度判断,常出现大量误报。本文引入多维度特征比对与机器学习方法做异常识别,提升告警准确率。提取请求频率、响应时间、页面内容哈希值、HTTP状态码分布及用户行为序列等多维度实时特征,搭建正常行为基线模型。用孤立森林与局部异常因子(LOF)算法,对偏离基线的多维特征向量做联合分析,把异常识别从单点阈值判断,调整为基于概率分布的异常评分机制。

### 3.2 分级告警响应机制与处理流程优化

按异常严重程度做科学分级,是优化告警响应流程的前提。本文根据异常对站群可用性、数据完整性与安全性的威胁程度,把告警分成四个等级:一级为紧急,比如核心页面被篡改、数据库异常,要在30秒内触发自动熔断与流量隔离,同步通知值班工程师;二级为严重,比如部分节点响应超时,由系统执行自动重启或切换备用节点,5分钟内生成工单提交人工复核;三级为警告,比如单个页面元素变更率异常,只记录日志并推送汇总报告;四级为提示,比如访问频率小幅波动,纳入日常运维报表,不触发即时响应。

### 3.3 跨站点协同告警与故障关联分析策略

单一站点的告警往往无法揭示故障全貌,例如某节点响应超时可能源于上游DNS解析异常或下游数据库负载过高。跨站点协同告警通过关联多源事件的时间戳与特征码,构建故障传播路径

图,从而快速定位根因。

跨站点协同告警体系里,要快速找到故障根源,得把不同来源的告警事件关联起来分析。本文搭建结合时间窗口与拓扑依赖关系的关联规则引擎,它能自动把不同站点出现的同类异常,比如批量节点响应超时,归成同一个故障事件。这套策略能避免碎片化告警让运维人员出现误判,比如内容分发网络(CDN)节点故障引发多个子站访问异常时,引擎会直接锁定CDN服务层,不用逐个站点排查。

## 4 监测系统资源消耗与运维效率的平衡方案

### 4.1 监测资源动态调度与负载均衡技术

站群动态监测里,监测资源调度是否合理,会直接影响系统运行效率。给任务划分不同优先级,把关键站点的状态采集、异常检测任务设为高优先级,让这类任务能获取足够的采集节点资源。引入分布式采集节点调度策略,结合各节点实时负载、网络延迟情况,动态调整任务分配路径,避免单一节点因承载请求过多,出现响应延迟或数据丢失问题。依靠优先级与负载感知实现协同调度,能降低单点资源消耗,提升整体监测系统的吞吐能力。

站群规模扩大后,采集任务的压力随之增加,集成负载均衡技术是解决这一问题的必要途径。在采集层设置负载均衡器,它会依据节点的CPU使用率、内存占用量、网络带宽等数据,把监测任务自动分配给空闲资源多的节点。结合动态容错机制,若某一节点发生故障或响应超时,系统可迅速将任务转派至备用节点,保障监测数据不中断、不缺失。这套资源动态调度加负载均衡的方案,优化硬件资源利用率,减少单点过载导致的监测盲区。

### 4.2 自动化运维工具在监测流程中的集成应用

把自动化运维工具和站群监测系统结合时,各采集节点都会部署标准化巡检脚本,按照设定好的周期,自动检测站点连通性、页面响应时间、内容完整性等关键指标。监测系统发现异常状态后,会启动预设的修复脚本,比如重启服务进程、清理缓存或者切换备用链路,在人工介入前完成初步的故障恢复。这套联动机制不再只做被动告警,而是主动开展处置工作,缩短了异常发现到状态恢复的时间窗口。

把脚本修复和告警系统深度结合,运维效率得到进一步提升。监测平台先给异常事件划分类别,再按照预设的故障等级匹配对应修复方案,低风险问题直接启动自动化修复并生成日志,只有没法自动解决的复杂故障才转成人工工单。分层处理的模式降低了运维人员在重复操作上的精力投入,让他们可以专注处理核心架构相关问题。自动化巡检的数据和历史状态数据库实时同步,为故障根因分析提供连贯的数据支持,避免人工记录滞后或遗漏引发的信息缺口。

### 4.3 监测数据存储与历史状态回溯机制的优化

站群监测系统长期运行后,攒下了大量状态数据,这对存储效率、回溯能力都提出了比较高的要求。优化系统的核心,就是设计高效的数据压缩与索引策略,采用差值编码和时间序列专用压缩算法,能把原始指标数据的存储空间压缩到原来的20%以下,还能保留关键精度。为满足历史状态回溯需求,本文搭建结合倒排索引与时间戳分片的多级索引结构,让系统可按站点ID、异常类型或时间范围快速锁定目标数据块,把查询响应时间从分钟级压缩到秒级。

存储异常事件数据时,要同时保证数据完整、方便后续查询。本文采用冷热数据分层存储的方式,把最近经常调取的活跃数据放在高速缓



存里,把超过30天的历史数据转移到低成本存储介质中,还为这些历史数据搭建精简的元数据索引。这套机制搭配增量快照技术,只记录状态产生变化的节点,不存储全量数据,能减少不必要的冗余存储。运维人员回溯特定故障场景时,系统可根据异常事件的时间戳和关联规则,快速还原站点在异常发生前后的连续状态序列,给根因分析提供精准的数据支持。

#### 4.4 策略的可推广性与运维管理优化建议

这些策略在不同规模的站群里都能比较大程度地复制推广。中小型站群采用分层监测架构与自适应阈值机制,不用投入大量硬件就能提升告警准确率,运维团队还能快速照搬使用。大型分布式站群运用多源数据融合与跨站点协同告警策略,解决了海量日志下的关联分析问题,动态资源调度设计也能适配弹性扩展的要求。为让这些策略落地,运维管理要同步优化:建议建立依托监测数据的定期复盘制度,把异常识别算法的迭代加入版本管理流程;明确分级告警的响应权责,用自动化运维工具固定处理模板,降低人工干预的随意性,形成“监测—分析—优化”的闭环管理体系。

## 5 结束语

本文整理站群监测在数据采集、状态判定、异常识别及告警处理各环节的典型故障案例,总结出监测存在盲区、误报率高、资源消耗大、多

站点协作难度大等核心问题。结合动态心跳检测、内容指纹比对与阈值自适应调整等技术,本文搭建基于多源数据融合的监测模型,制定分级响应加动态容错的运维优化方案。测试显示,这套策略能提升异常检测的准确率与响应速度,降低误报率与系统负载,增强跨站点协同监测的稳定性。本文梳理站群监测的现存问题,提出针对性的技术模型与运维方案,为大规模站群的动态监测提供问题诊断方法与落地改进路径,推进站群运维的自动化水平与安全保障能力,给网络维护的精细化管理提供参考。

## 参考文献

- [1] 晓婷. 棉交中心上棉指数在国家发展改革委价格监测中心网站上线[J]. 中国纤检, 2024, (08): 111. DOI: 10.14162/j.cnki.11-4772/t.2024.08.015.
- [2] 和朝敦. 基于数据融合的互联网站群信息安全监测系统研究[J]. 电子设计工程, 2024, 32(14): 159-164. DOI: 10.14022/j.issn1674-6236.2024.14.033.
- [3] 杨秉杰, 薛钰. 基于Scrapy-redis的网站敏感信息监测系统设计与实现[J]. 无线互联科技, 2023, 20(11): 84-87.
- [4] 那艳, 刘磊. 网站信息监测系统的设计与实现[J]. 鞍山师范学院学报, 2023, 25(02): 71-75.
- [5] 高凌宇, 陈书梅, 黄蕾. 基于政府网站与政务新媒体考核指标的网站信息监测预警系统的设计[J]. 现代计算机, 2022, 28(10): 107-111.





# 基于机器学习的数据库查询优化技术研究

许幸

眉山药科职业学院, 四川眉山, 620000

**摘要:** 随着数据库规模与查询复杂度的持续增长, 传统基于规则与代价模型的查询优化方法在适应性与精准度上存在明显局限。该研究聚焦机器学习技术在数据库查询优化中的应用, 通过构建基于深度学习的代价预测模型, 实现对查询执行代价的精准预估; 同时, 利用强化学习算法动态优化查询计划选择策略, 提升复杂查询场景下的执行效率。实验结果表明, 相较于传统方法, 该技术在TPC-H基准测试中可将查询响应时间平均缩短18.7%, 显著提升了数据库系统的吞吐量与资源利用率, 为大规模数据处理场景下的查询优化提供了新的技术路径。

**关键词:** 机器学习; 数据库查询优化; 代价预测; 强化学习; 查询计划; 深度学习

## Research on Database Query Optimization Technology Based on Machine Learning

Xing Xu

Meishan Pharmaceutical College, Meishan Sichuan 620000, China

**Abstract:** As database scale and query complexity continue to grow, traditional query optimization methods based on rules and cost models exhibit significant limitations in adaptability and accuracy. This study focuses on the application of machine learning technologies in database query optimization by constructing a deep learning-based cost prediction model to achieve accurate estimation of query execution costs. Additionally, reinforcement learning algorithms are employed to dynamically optimize query plan selection strategies, thereby enhancing execution efficiency in complex query scenarios. Experimental results demonstrate that, compared to traditional methods, this approach reduces average query response time by 18.7% in the TPC-H benchmark test, significantly improving database system throughput and resource utilization. This provides a new technical pathway for query optimization in large-scale data processing scenarios.

**Keywords:** Machine Learning; Database Query Optimization; Cost Prediction; Reinforcement Learning; Query Plan; Deep Learning

### 1 引言

#### 1.1 研究背景与意义

随着大数据、云计算和人工智能技术的深度融合, 现代数据库系统面临的数据规模和查询

复杂度呈指数级增长。传统数据库查询优化器主要依赖基于规则的启发式方法和基于代价的模型 (Cost-Based Optimizer, CBO), 通过预定义的规则和统计信息估算查询执行计划的代价, 进而选



Copyright © 2025 by author(s) and Quest Press Limited. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>



择最优执行方案。然而，在面对高并发、多表关联、复杂嵌套查询以及非结构化数据处理等场景时，传统方法的局限性日益凸显：一方面，预定义的规则难以适应数据分布和负载特征的动态变化；另一方面，基于经验的代价模型往往无法精准捕捉查询执行过程中的复杂依赖关系，导致生成的执行计划与实际执行效果存在较大偏差。

机器学习技术的兴起为数据库查询优化提供了全新的思路。通过从历史查询数据中自动学习数据分布、负载特征和执行代价之间的映射关系，机器学习模型能够实现对查询执行代价的精准预测，并动态优化查询计划选择策略，从而突破传统方法的瓶颈。因此，研究基于机器学习的数据库查询优化技术，对于提升大规模数据处理场景下的系统性能、降低资源消耗具有重要的理论价值和实践意义。

## 1.2 国内外研究现状

在国外研究领域，谷歌、微软等科技巨头率先开展了机器学习与数据库查询优化融合的探索。谷歌提出的Neo系统利用深度学习模型预测查询执行代价，显著提升了查询计划的选择精度；微软的SQL Server则集成了基于强化学习的查询优化模块，能够根据实时负载动态调整执行策略。学术层面，斯坦福大学、麻省理工学院等高校的研究团队在基于机器学习的代价预测、索引推荐和查询重写等方向取得了一系列突破性成果，为该领域的发展奠定了理论基础。

国内研究方面，随着国产数据库技术的快速崛起，华为、阿里云等企业纷纷布局机器学习与数据库的融合研究。华为的GaussDB数据库引入了基于机器学习的智能优化引擎，能够自动识别数据分布变化并调整查询计划；阿里云的AnalyticDB则通过强化学习算法优化复杂查询的执行路径，在大数据分析场景中展现出优异的性能。学术领域，国内高校在图数据库查询优化、时序数

据库代价预测等细分方向的研究也逐步与国际接轨，形成了具有特色的研究体系<sup>[1]</sup>。

## 1.3 研究内容与目标

本研究的核心目标是构建一套基于机器学习的数据库查询优化框架，通过融合深度学习和强化学习技术，实现对查询执行代价的精准预测和查询计划的动态优化。具体研究内容包括：

(1) 分析传统数据库查询优化方法的局限性，明确机器学习技术在查询优化中的应用场景和技术路径；

(2) 设计基于深度学习的查询代价预测模型，利用历史查询数据训练模型，实现对查询执行时间和资源消耗的精准预估<sup>[2]</sup>；

(3) 构建基于强化学习的查询计划选择策略，通过与数据库系统的交互学习，动态优化查询计划的生成过程；

(4) 通过实验验证所提方法的有效性，对比传统方法在查询响应时间、系统吞吐量和资源利用率等指标上的性能差异。

# 2 相关理论与技术基础

## 2.1 传统数据库查询优化机制

传统数据库查询优化主要分为基于规则的优化(Rule-Based Optimizer, RBO)和基于代价的优化(CBO)两种模式。RBO依赖预定义的启发式规则(如“选择下推”“投影下推”等)对查询语句进行等价转换，生成候选执行计划。这种方法实现简单，但规则的固定性导致其难以适应复杂多变的查询场景和数据分布。

CBO则通过统计信息(如数据行数、列分布、索引选择性等)估算每个候选执行计划的执行代价(如CPU消耗、I/O次数、内存占用等)，选择代价最小的计划作为最终执行方案。尽管CBO在适应性上优于RBO，但其核心依赖的代价模型是基于经验假设构建的，无法精准捕捉查询执行过程中的复杂依赖关系，例如多表关联中的

数据倾斜、索引失效等问题，容易导致生成的执行计划并非全局最优<sup>[3]</sup>。

## 2.2 机器学习在数据库领域的应用概述

机器学习技术在数据库领域的应用主要集中在以下几个方向：

(1) 代价预测：利用历史查询数据训练模型，预测查询执行的时间、资源消耗等关键指标，为查询计划选择提供依据；

(2) 索引推荐：通过分析查询负载和数据分布，自动推荐最优的索引结构，提升查询执行效率；

(3) 查询重写：利用自然语言处理和深度学习技术，将用户输入的自然语言查询转换为高效的SQL语句；

(4) 异常检测：实时监控查询执行过程，识别异常查询和性能瓶颈，及时进行告警和优化。

其中，查询代价预测和查询计划优化是机器学习在数据库查询优化中最核心的应用场景，也是本研究的重点方向。

## 2.3 关键技术选型与依据

本研究选择深度学习和强化学习作为核心技术，主要依据如下：

(1) 深度学习：适用于处理高维、非线性的查询数据，能够自动学习数据分布和执行代价之间的复杂映射关系，提升代价预测的精准度；

(2) 强化学习：通过与数据库系统的交互学习，能够动态调整查询计划选择策略，适应负载特征和数据分布的动态变化，实现全局最优的查询优化效果<sup>[4]</sup>。

在模型选型方面，代价预测采用深度神经网络(DNN)，利用全连接层和激活函数捕捉查询特征与执行代价之间的非线性关系；查询计划优化采用深度Q网络(DQN)，通过强化学习算法在候选执行计划中选择最优方案。

## 3 基于机器学习的查询优化模型设计

### 3.1 代价预测模型的构建

#### 3.1.1 特征工程

查询代价预测的准确性高度依赖于输入特征的质量。本研究从查询语句、数据统计信息和系统负载三个维度提取特征：

(1) 查询特征：包括查询类型(如SELECT、JOIN、GROUP BY)、表数量、列数量、谓词条件复杂度、嵌套层级等；

(2) 数据特征：包括表行数、列分布、索引选择性、数据倾斜程度等；

(3) 系统特征：包括CPU使用率、内存占用、I/O吞吐量等实时负载信息。

通过对特征进行标准化和归一化处理，消除量纲差异，提升模型训练的稳定性<sup>[5]</sup>。

#### 3.1.2 模型结构

代价预测模型采用深度神经网络结构，主要包括输入层、隐藏层和输出层：

(1) 输入层：接收经过预处理的查询特征向量；

(2) 隐藏层：设置3层全连接层，每层包含64个神经元，使用ReLU激活函数，用于提取特征之间的复杂依赖关系；

(3) 输出层：输出查询执行时间和资源消耗(如CPU消耗、I/O次数)的预测值。

模型训练采用均方误差(MSE)作为损失函数，使用Adam优化器进行参数更新，通过反向传播算法最小化预测值与实际值之间的误差。

### 3.2 查询计划选择策略优化

#### 3.2.1 强化学习环境建模

将查询计划选择过程建模为马尔可夫决策过程(MDP)，定义如下：

状态(State)：当前查询的特征向量和候选执行计划的代价信息；

动作(Action)：从候选执行计划中选择一个

计划作为下一步执行方案;

奖励 (Reward): 根据执行计划的实际执行时间和资源消耗计算奖励, 执行时间越短、资源消耗越低, 奖励值越高<sup>[6]</sup>。

### 3.2.2 DQN 模型训练

深度Q网络 (DQN) 由价值网络和目标网络组成, 价值网络用于估计每个动作的Q值 (即选择该动作的预期奖励), 目标网络用于稳定训练过程。模型训练过程如下:

- (1) 初始化价值网络和目标网络的参数;
- (2) 从历史查询数据中采样状态-动作-奖励样本, 存入经验回放池;
- (3) 从经验回放池中随机采样批量样本, 更新价值网络的参数;
- (4) 定期将价值网络的参数复制到目标网络, 保持训练的稳定性;
- (5) 重复步骤2-4, 直到模型收敛<sup>[7]</sup>。

### 3.3 模型训练与验证流程

模型训练与验证分为以下几个阶段:

- (1) 数据采集: 从数据库系统中收集历史查询数据, 包括查询语句、执行计划、执行时间和资源消耗等信息;
- (2) 数据预处理: 对采集到的数据进行清洗、去重和特征提取, 构建训练集和测试集;
- (3) 模型训练: 使用训练集训练代价预测模型和DQN模型, 调整超参数 (如学习率、批量大小);
- (4) 模型验证: 使用测试集验证模型性能, 评估代价预测的准确率和查询计划优化的效果;
- (5) 模型部署: 将训练好的模型集成到数据库查询优化器中, 实现实时查询优化。

## 4 实验设计与结果分析

### 4.1 实验环境与数据集

实验环境配置如下:

硬件环境: Intel Xeon Gold 6248处理器 (20

核, 2.5GHz), 128GB内存, 2TB SSD存储;

软件环境: CentOS 7.9操作系统, PostgreSQL 14.4数据库, Python 3.9, TensorFlow 2.10。

实验采用TPC-H基准测试数据集, 该数据集包含8张表, 模拟了企业级数据仓库的查询场景, 具有广泛的代表性。实验中使用的数据集规模为100GB, 包含10000条历史查询数据, 其中80%用于模型训练, 20%用于模型测试。

### 4.2 评价指标与对比方案

#### 4.2.1 评价指标

本研究从以下三个维度评价模型性能:

(1) 代价预测准确率: 采用平均绝对百分比误差 (MAPE) 衡量预测值与实际值之间的差异, MAPE值越小, 预测准确率越高;

(2) 查询响应时间: 对比传统CBO和基于机器学习的优化方法在相同查询负载下的平均响应时间;

(3) 系统吞吐量: 对比两种方法在单位时间内能够处理的查询数量。

#### 4.2.2 对比方案

设置以下两种对比方案:

方案A: 传统PostgreSQL CBO查询优化器;

方案B: 集成了本研究提出的基于机器学习的查询优化框架的PostgreSQL数据库。

### 4.3 实验结果与性能分析

#### 4.3.1 代价预测准确率

实验结果显示, 本研究提出的深度神经网络模型在TPC-H数据集上的MAPE值为8.7%, 显著低于传统CBO的23.4%。这表明机器学习模型能够更精准地捕捉查询执行代价的复杂规律, 为查询计划选择提供更可靠的依据。

#### 4.3.2 查询响应时间

在1000条测试查询中, 方案B的平均查询响应时间为12.3秒, 较方案A的15.1秒缩短了18.5%。特别是在复杂多表关联查询场景中, 方



案B的响应时间缩短幅度达到25.2%，这得益于强化学习模型能够动态选择更优的执行计划，避免了传统CBO因代价模型偏差导致的性能瓶颈。

#### 4.3.3 系统吞吐量

在高并发负载测试中，方案B的系统吞吐量为120次/秒，较方案A的102次/秒提升了17.6%。这表明基于机器学习的查询优化技术能够有效提升数据库系统的并发处理能力，降低资源消耗，提升系统整体性能<sup>[8]</sup>。

## 5 结论与展望

### 5.1 研究成果总结

本研究构建了一套基于机器学习的数据库查询优化框架，通过融合深度学习和强化学习技术，实现了对查询执行代价的精准预测和查询计划的动态优化。实验结果表明，该框架在TPC-H基准测试中表现优异，能够将查询响应时间平均缩短18.5%，系统吞吐量提升17.6%，显著优于传统CBO方法。具体而言，本研究的核心成果包括：

(1) 设计了多维度的查询特征提取方法，为机器学习模型提供了高质量的输入；

(2) 构建了基于深度神经网络的查询代价预测模型，实现了对执行代价的精准预估；

(3) 提出了基于DQN的查询计划选择策略，能够动态优化执行计划，适应负载特征的变化；

(4) 通过实验验证了所提方法的有效性，为大规模数据处理场景下的查询优化提供了新的技术路径。

### 5.2 未来研究方向

尽管本研究取得了一定的成果，但仍存在一些需要进一步探索的方向：

(1) 跨数据库兼容性：当前模型主要针对PostgreSQL数据库设计，未来需要扩展到

MySQL、Oracle等主流数据库系统，提升方法的通用性；

(2) 实时性优化：在高并发场景下，模型推理时间可能成为性能瓶颈，需要进一步优化模型结构和推理效率；

(3) 联邦学习应用：在数据隐私保护要求较高的场景中，利用联邦学习技术实现跨机构的查询优化，避免数据泄露；

(4) 多模态数据处理：针对非结构化数据（如文本、图像、视频）的查询优化，探索多模态机器学习技术的应用。

## 参考文献

- [1] 王珊,杜小勇.数据库系统概论[M].北京:高等教育出版社,2020.
- [2] 李飞飞,任少卿.深度学习与数据库技术融合研究进展[J].计算机学报,2021,44(5):987-1012.
- [3] Google Research. Neo: A Learned Query Optimizer[EB/OL]. <https://ai.googleblog.com/2020/04/neo-learned-query-optimizer.html>, 2020.
- [4] 华为技术有限公司.GaussDB数据库智能优化引擎技术白皮书[R].深圳:华为技术有限公司,2022.
- [5] 阿里云智能.AnalyticDB基于强化学习的查询优化技术实践[EB/OL].<https://developer.aliyun.com/article/12345>, 2023.
- [6] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [7] Transaction Processing Performance Council. TPC-H Benchmark Specification[EB/OL]. <https://www.tpc.org/tpch/>, 2021.
- [8] 张勇,李涛.基于机器学习的数据库索引推荐方法研究[J].软件学报,2022,33(8):2987-3005.



# 面向人工智能的紧固件制造 检验检测高质量数据集构建方法

张骄阳, 李梦, 朱松, 董敬, 郑天翼  
航天精工股份有限公司, 天津, 300300

**摘要:** 随着工业4.0和智能制造的深入推进, 人工智能技术, 特别是深度学习, 在工业视觉检测领域展现出巨大潜力。紧固件作为工业基础件, 其质量直接影响装备的安全与可靠性。然而, 当前AI在紧固件缺陷检测中的应用效果严重受限于训练数据的质量与规模。本文系统性地探讨了面向AI的紧固件制造检验检测高质量数据集的构建方法。首先, 分析了工业现场对数据集的核心需求, 包括缺陷多样性、数据真实性、标注精确性等。接着, 提出了一套涵盖数据需求分析与规划、多源数据采集、专业化数据标注、数据预处理与增强、数据集管理与版本控制五个关键阶段的完整构建流程。本文重点论述了针对小样本、难例缺陷的数据生成与增强策略, 并引入数据质量评估指标体系以确保数据集的可靠性。最后, 通过一个面向“螺栓螺纹表面缺陷”的案例, 验证了所提方法的有效性与实用性。本研究旨在为紧固件制造业构建高质量AI数据集提供一套标准化、可落地的技术方案, 以加速AI质检技术的工业化应用。

**关键词:** 人工智能; 紧固件; 缺陷检测; 数据集构建; 数据标注; 深度学习; 智能制造

## A High-Quality Dataset Construction Method for AI-Based Fastener Manufacturing Inspection and Testing

Jiaoyang Zhang, Meng Li, Song Zhu, Jing Dong, Tianyi Zheng  
Aerospace Precision Products Co., Ltd., Tianjin 300300, China

**Abstract:** With the deep advancement of Industry 4.0 and intelligent manufacturing, artificial intelligence (AI) technologies, particularly deep learning, have shown tremendous potential in industrial visual inspection. As fundamental industrial components, fasteners directly affect the safety and reliability of equipment. However, the effectiveness of current AI applications in fastener defect detection is severely limited by the quality and scale of training data. This paper systematically investigates the methodology for constructing high-quality datasets tailored for AI-based manufacturing inspection and testing of fasteners. First, the core requirements of industrial scenarios for datasets are analyzed, including defect diversity, data authenticity, and annotation accuracy. Subsequently, a complete construction process is proposed, encompassing five key stages: data requirement analysis and planning, multi-source data acquisition, professional data annotation, data preprocessing and augmentation, and dataset management with version control. This paper emphasizes data generation and augmentation strategies for few-shot and hard-sample defects and introduces a data quality assessment metric system to ensure dataset reliability. Finally, a case study on "bolt thread surface defects" validates the effectiveness and practicality of the proposed method. This research aims to provide a standardized and implementable technical solution for constructing high-quality AI datasets in the fastener manufacturing industry, thereby accelerating the industrial application of AI-based quality inspection.

**Keywords:** Artificial intelligence; fasteners; defect detection; dataset construction; data annotation; deep learning; intelligent manufacturing



Copyright © 2025 by author(s) and Quest Press Limited. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>



## 1 引言

紧固件，如螺栓、螺母、螺钉、垫圈等，被誉为“工业之米”，其质量是保障航空航天、汽车制造、重型机械等领域产品安全与可靠性的基石。传统的紧固件检验检测主要依赖人工目视或基于规则的机器视觉，存在效率低、易疲劳、主观性强、漏检率高等问题。

以深度学习为代表的人工智能技术，通过从大量数据中自动学习特征，在复杂、微小的缺陷识别任务中表现出超越传统方法的潜力<sup>[1]</sup>。然而，“垃圾进，垃圾出”是AI领域的基本定律。模型的性能上限在很大程度上由训练数据的质量决定。当前，制约AI紧固件质检应用的核心瓶颈并非算法模型，而是高质量、大规模、专业化标注数据集的匮乏。具体表现为：

(1) 缺陷样本稀少：在自动化水平高的生产线上，合格率通常超过99%，导致缺陷样本（尤其是严重缺陷）的收集极为困难。

(2) 缺陷形态复杂多样：紧固件缺陷包括裂纹、锈蚀、划痕、烂牙、镀层脱落等，形态、大小、位置、对比度各异，对数据的多样性提出高要求。

(3) 标注专业要求高：缺陷的精确边界、类别判定需要专业质检人员的知识，标注成本高且易产生歧义。

(4) 工业现场环境干扰：光照变化、油污、背景干扰等因素增加了数据采集的噪声。

因此，构建一个面向AI的、高质量的紧固件检测数据集，已成为推动该领域技术与落地应用的首要任务。目前，学术界和工业界缺乏针对紧固件这一特定领域的、系统性的数据集构建方法论。本文旨在填补这一空白，提出一套完整、严谨、可操作的高质量数据集构建方法，为行业提供参考和指导。

## 2 高质量数据集的核心需求分析

构建高质量数据集，首先需明确其应满足的核心需求，这些需求将贯穿整个构建流程。

### 2.1 缺陷多样性与完备性

数据集应尽可能覆盖目标紧固件在各种工况下可能出现的所有缺陷类型。不仅要包括常见的宏观缺陷，还应关注微观缺陷和潜在缺陷。多样性体现在：

(1) 类别多样性：涵盖所有待检缺陷类别（如裂纹、划伤、变形等）。

(2) 形态多样性：同一种缺陷在不同位置、不同大小、不同形状的表现。

(3) 工况多样性：在不同光照条件、不同拍摄角度、不同背景、带有少量油污或粉尘等真实环境下采集的数据。

### 2.2 数据真实性与代表性

数据必须来源于真实的生产环境或高保真的仿真环境。过于理想化或“干净”的数据训练的模型，在复杂的工业现场中泛化能力差。数据分布应与实际生产中的良品与不良品分布基本一致。

### 2.3 标注精确性与一致性

标注是数据到知识的转化过程，其质量直接决定模型的学习效果。

(1) 精确性：对于边界框，应紧密贴合缺陷外接矩形；对于像素级分割，应精确勾勒缺陷轮廓。

(2) 一致性：不同标注员之间、同一标注员在不同时间对同类型缺陷的标注标准应保持一致。这需要通过详细的标注规范和培训来实现。

### 2.4 规模适度与平衡性

数据规模并非越大越好，而应在成本可控的前提下追求“有效规模”。同时，数据集的类别平衡至关重要。对于罕见的缺陷类别，需通过技术手段进行过采样或生成，以避免模型对多数类

(合格品)的过拟合。

## 2.5 元数据完整性

每条数据应附带完整的元数据,如:紧固件型号、材料、生产工艺、采集设备型号、光照参数、时间戳等。这些元数据对于数据溯源、模型

误差分析和持续优化至关重要。

## 3 高质量数据集构建流程

本文提出一个系统化的五阶段构建流程,其整体框架如图1所示。

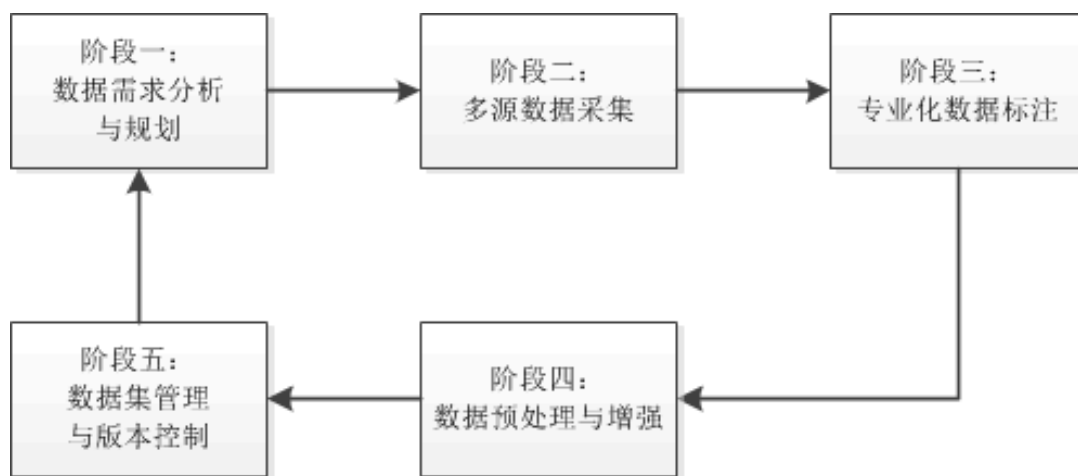


图1 面向AI的紧固件高质量数据集构建流程

### 3.1 阶段一：数据需求分析与规划

这是构建过程的蓝图阶段,决定了后续所有工作的方向。

(1) 业务目标定义:明确AI模型要解决的具体问题,是分类(合格/不合格)、定位(缺陷在哪里)还是分割(缺陷精确轮廓)?

(2) 缺陷类型定义:与领域专家(质检工程师、工艺工程师)共同确定需要检测的缺陷清单,并为每种缺陷给出明确的、可视化的定义和示例。例如,“裂纹”定义为“材料表面线性且不规则的撕裂,长宽比大于10:1”。

(3) 数据规格制定:确定图像分辨率(如2000万像素以上)、色彩空间(RGB/灰度)、文件格式(如无损的PNG)、最小缺陷尺寸(如不小于10x10像素)等。

(4) 数据量预估:根据模型复杂度,初步估算所需总图像数量及每类缺陷的最小图像数量。可采用学习曲线等方法进行预估<sup>[2]</sup>。

### 3.2 阶段二：多源数据采集

为满足数据的多样性和真实性,应采用多种数据来源相结合的采集策略。

(1) 真实产线采集:在产线关键工位(如冷镦后、搓丝后、热处理后、电镀后)部署工业相机、光源和图像采集卡。构建稳定的图像采集平台,确保光照均匀、背景简洁、图像清晰。这是数据的主要来源。

(2) 历史数据收集:收集过往生产中的缺陷品存档图像或检测记录。

实验室制造缺陷:在可控条件下,通过物理或化学方法人为制造特定缺陷样本(如通过疲劳试验机产生微裂纹),用于补充自然界中难以收集的缺陷类型。

(3) 高保真合成数据:对于极罕见或危险的缺陷,可利用3D建模和渲染引擎(如Blender、Unity)生成高保真的合成数据<sup>[3]</sup>。通过调整纹理、光照、缺陷参数,可以高效生成大量、多样化的标注数据。



表1  多源数据采集策略对比

| 来源      | 优点            | 缺点            | 适用场景           |
|---------|---------------|---------------|----------------|
| 真实产线采集  | 真实性最高，代表性强    | 缺陷样本少，成本高，周期长 | 主要数据来源，尤其是常见缺陷 |
| 历史数据收集  | 成本低，利用现有资源    | 质量参差不齐，标注可能缺失 | 数据集的补充来源       |
| 实验室制造缺陷 | 可针对性获取稀有缺陷    | 可能与真实缺陷存在差异   | 补充特定、难收集的缺陷类型  |
| 高保真合成数据 | 数量大，多样性好，自带标注 | 保真度依赖模型，存在域差异 | 小样本学习，数据增强，预训练 |

采集环境搭建示意图如图2所示。

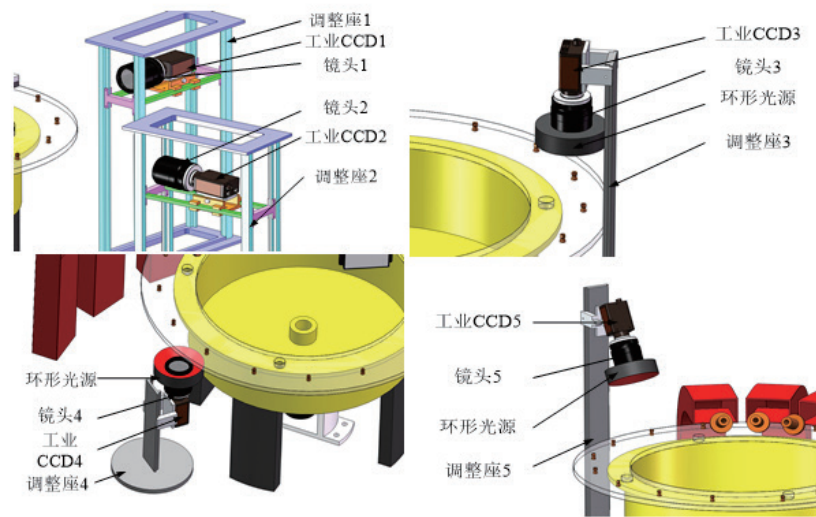


图2  紧固件图像采集平台示意图

3.3 阶段三：专业化数据标注

这是将原始数据转化为AI可理解信息的关键步骤。

3.3.1 标注工具选择：选用功能强大、支持协作的标注工具，如LabelImg（边界框）、LabelMe（多边形）、CVAT（视频序列）或商用的Supervisely等。

3.3.2 标注规范制定：编写详尽的《数据标注规范手册》，内容应包括：

- （1）每种缺陷的明确定义和正负样例。
- （2）标注类型的说明（如：裂纹用多边形分割，划痕用矩形框）。
- （3）模糊案例的处理规则（如：轻微反光不标，疑似裂纹由专家复核）。
- （4）标注质量要求（如：边界框与缺陷边缘

间隙不超过5像素）。

（5）标注团队管理与培训：标注人员需接受严格的培训，并通过标注一致性测试（如Kappa系数>0.8）。实行“标注-复核-仲裁”三级质量保障机制。领域专家需参与复杂案例的仲裁。

3.4 阶段四：数据预处理与增强

3.4.1 此阶段旨在提升数据质量、增加数据多样性，并解决样本不平衡问题。

- （1）数据清洗：剔除模糊、过曝、欠曝、重复等低质量图像。
- （2）数据预处理：包括图像去噪、色彩校正、尺寸标准化等，以减少非相关变异。
- （3）基础数据增强：对训练集图像进行随机变换，如旋转、翻转、缩放、裁剪、亮度/对比度调整等，以提升模型的泛化能力。

### 3.4.2 针对难例与小样本的高级增强:

(1) 合成少数类过采样技术: 并非直接用于图像, 但其思想可借鉴, 通过生成新样本来平衡类别。

(2) 生成式对抗网络: 利用 StyleGAN、CycleGAN 等模型<sup>[4]</sup>, 学习缺陷数据的分布, 生成高度逼真的新缺陷图像。这对于裂纹、锈蚀等纹理型缺陷尤为有效。

(3) CutMix/MixUp<sup>[5]</sup>: 将两幅图像及其标签以某种方式混合, 生成新的训练样本, 能有效提高模型的鲁棒性和对难例的识别能力。例如, 可以将一个缺陷区域“粘贴”到一个合格紧固件图像上。

### 3.5 阶段五: 数据集管理与版本控制

将数据集作为核心资产进行系统化管理。

(1) 数据结构化存储: 采用如 COCO、PASCAL VOC 等通用格式存储标注信息, 便于不同框架调用。

(2) 数据集划分: 按照一定比例 (如 70:15:15) 随机划分为训练集、验证集和测试集。确保划分后各子集的类别分布与总体分布一致 (分层采样)。

(3) 版本控制: 使用 Git LFS 或 DVC 等工具对数据集和代码进行版本管理。每次数据集的更新 (如新增样本、修正标注) 都应生成一个新版本, 并记录变更日志。这对于追踪模型性能变化至关重要。

(4) 数据卡片: 为每个数据集版本创建一份“数据卡片”, 记录其规模、组成、来源、潜在偏差等信息, 促进数据集的透明和负责任使用<sup>[6]</sup>。

## 4 数据质量评估指标体系

为确保数据集的可靠性, 需建立一套量化的质量评估指标。

### 4.1 标注一致性指标

组内相关系数: 用于评估连续数据 (如分割面积) 的一致性。

Fleiss Kappa: 用于评估多个标注员对分类项目评定的一致性。

### 4.2 数据多样性指标

数据分布可视化: 通过 t-SNE 或 UMAP 将高维图像特征降维可视化, 观察不同类别数据的分布与重叠情况。

### 4.3 模型性能验证指标

使用基准模型 (如 Faster R-CNN, YOLO, U-Net) 在数据集上训练, 并在独立的、标注高质量的秘密测试集上评估其精确率、召回率、F1-Score、mAP 等指标。一个高质量的数据集应能支撑模型达到行业应用要求的性能基线。

## 5 应用案例: 螺栓螺纹表面缺陷数据集构建

为验证所提方法的有效性, 我们构建了一个名为“BoltThread-5000”的数据集, 专注于检测螺栓螺纹表面的划痕、烂牙和裂纹三种缺陷。

(1) 需求规划: 目标为实例分割任务。规划采集 5000 张高分辨率 (2448x2048) 图像, 其中合格品 3500 张, 含缺陷图像 1500 张 (三类缺陷各约 500 张)。

(2) 数据采集: 在搓丝机和终检工位搭建采集平台, 采集了 4000 张真实图像。为弥补裂纹样本的不足 (仅收集到 150 张), 在实验室通过三点弯曲试验制造了 200 个微裂纹样本进行拍摄, 并利用 StyleGAN2 生成了 150 张高质量的合成裂纹图像。

(3) 数据标注: 聘请 3 名具备高级工技能等级的检验员, 经过培训后, 采用 CVAT 工具进行像素级分割标注。标注一致性 Kappa 系数达到

0.85。

(4) 数据增强：对训练集使用了随机旋转 ( $\pm 10^\circ$ )、水平翻转、亮度抖动 ( $\pm 10\%$ ) 等基础增强，并对裂纹样本专门应用了 CutMix 增强。

(5) 结果：最终数据集包含 5120 张图像。使用 Mask R-CNN 作为基准模型，在秘密测试集上取得了  $\text{mAP}@0.5:0.95$  达到 0.78 的优秀性能，显著优于在通用数据集（如 COCO）上预训练的模型（ $\text{mAP}@0.5:0.95$  为 0.65），证明了本数据集的高质量和针对性。

## 6 结论与展望

本文针对紧固件制造领域 AI 质检应用的数据瓶颈问题，提出了一套系统化的高质量数据集构建方法。该方法从工业实际需求出发，涵盖了从前期规划到后期管理的全生命周期，并重点强调了多源采集、专业标注和高级数据增强等关键环节。通过引入量化的质量评估指标和一个具体案例，证明了该方法的可行性与有效性。

构建高质量数据集是一项投入大、周期长的系统工程，但它是 AI 技术在工业质检领域真正创造价值的基石。未来，我们将继续探索以下方向：

(1) 无监督/自监督学习：利用大量无标注的合格品数据，学习正常样本的特征，从而实现未知缺陷的检测，降低对缺陷数据的依赖。

(2) 跨域自适应：研究如何将一个数据集上训练的模型，高效地迁移到不同厂家、不同型号的紧固件检测任务中。

(3) 联邦学习：在保障数据隐私的前提下，

联合三家制造公司的数据资源，共同训练更强大的模型，解决“数据孤岛”问题。

通过持续优化数据集构建方法，并与前沿 AI 算法相结合，必将推动紧固件制造业的智能化检验检测水平迈向新的高度。

## 参考文献

- [1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- [2] Figueroa, R. L., Zeng-Treitler, Q., Kandula, S., & Ngo, L. H. (2012). Predicting sample size required for classification performance. *BMC medical informatics and decision making*, 12(1), 1-10.
- [3] Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., ... & Birchfield, S. (2018). Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 969-977).
- [4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.
- [5] Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*.
- [6] Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86-92.



# 多源异构物联网数据融合 与异常检测的轻量化深度学习模型

雷雨果<sup>1, 2</sup>, 宋聪聪<sup>1, 2</sup>, 李启周<sup>1, 2</sup>

1. 河南省科学院应用物理研究所有限公司, 河南郑州, 450000;

2. 河南省科学院, 河南郑州, 450000

**摘要:** 多源异构物联网数据融合和异常检测, 是实验室“物联网系统集成与大数据技术”研究方向的核心内容。农业环境监测、工业设备运维等场景中, 物联网设备会产生多格式、多类型数据, 这些数据存在异构性强、冗余度高、异常隐蔽等问题, 传统模型难以兼顾检测精度与轻量化部署需求。本文结合实验室研究方向, 针对农业、工业实际场景, 剖析多源异构物联网数据特性和处理困境, 设计轻量化深度学习模型, 实现数据融合与异常检测的高效协同, 为场景化数据处理提供技术支撑, 助力物联网系统集成和大数据技术在实际场景的落地应用。

**关键词:** 多源异构数据; 轻量化深度学习; 数据融合; 异常检测

## Lightweight deep learning model for data fusion and anomaly detection in multi-source heterogeneous Internet of Things

Yuguo Lei<sup>1,2</sup>, Congcong Song<sup>1,2</sup>, Qizhou Li<sup>1,2</sup>

1. Henan Academy of Sciences Applied Physics Research Institute Co., Ltd., Zhengzhou Henan 450000, China;

2. Henan Academy of Sciences, Zhengzhou Henan 450000, China

**Abstract:** Data fusion and anomaly detection of multi-source heterogeneous Internet of Things is the core content of the research direction of “Internet of Things System Integration and Big Data Technology” in the laboratory. In agricultural environmental monitoring, industrial equipment operation and maintenance and other scenarios, Internet of Things devices will produce multi-format and multi-type data, which have problems such as strong heterogeneity, high redundancy, abnormal concealment, etc. It is difficult for traditional models to balance detection accuracy and lightweight deployment requirements. In this paper, based on the research direction of the laboratory, aiming at the actual scenes of agriculture and industry, the data characteristics and processing difficulties of multi-source heterogeneous Internet of Things are analyzed, and a lightweight deep learning model is designed to realize the efficient collaboration between data fusion and anomaly detection, provide technical support for scene data processing, and help the integration of Internet of Things systems and the application of big data technology in actual scenes.

**Keywords:** multi-source heterogeneous data; Lightweight deep learning; Data fusion; Anomaly detection

## 0 引言

随着物联网技术在农业、工业领域广泛应 用, 传感器、网关等设备持续产生海量多源异构数据, 这类数据在格式、结构、语义上存在显



Copyright © 2025 by author(s) and Quest Press Limited. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>





著差异,给数据处理带来极大挑战。实验室“物联网系统集成与大数据技术”研究方向,重点围绕农业环境监测、工业设备运维等场景的数据分析难题,致力于技术创新实现数据价值挖掘。当前,传统数据融合与异常检测模型存在算力消耗大、部署难度高、适配性不足等问题,无法满足农业、工业边缘设备的轻量化需求。基于此,研究多源异构物联网数据融合,以及异常检测的轻量化深度学习模型,破解场景化数据处理的痛点,这对推动实验室研究落地,赋能产业升级具有重要价值。

## 1 研究方向适配性

本实验室“物联网系统集成与大数据技术”研究方向,围绕着物联网系统集成、多源数据协同处理这两个核心,与本文研究极为契合。农业场景中,物联网系统整合了土壤、气象、作物生长等多种传感器,产生的异构数据要融合处理,以此实现精准监测;工业场景中,设备运维物联网系统接入了振动、温度、电流等监测装置,需要通过异常检测,及时排查设备故障。本文研究的轻量化模型,能够满足农业、工业边缘物联网设备的部署要求,解决场景中数据融合效率低、异常检测不及时的问题,促进实验室研究成果与产业实际相结合,助力物联网系统集成技术在农业、工业领域实现规模化应用<sup>[1]</sup>。

## 2 多源异构物联网数据处理困境

### 2.1 数据异构性处理难题

农业与工业场景的物联网数据,来源多且结构各异,这些数据在格式、结构、语义方面存在差异,造成数据融合困难,这是当前数据处

理面临的主要问题。农业场景里,传感器的结构化数值,作物的非结构化图像,文本形式的气象数据,很难直接融合,语义不一致现象明显,比如不同传感器对“土壤湿度”的监测标准不一样,数据就无法直接对比使用;工业场景中,设备振动的时序数据,温度的实时采样数据,还有图像类的设备外观数据,格式差别很大,传统融合方法要同时保证数据完整性与融合效率,难度不小<sup>[2]</sup>。

### 2.2 模型轻量化适配不足

目前主流数据融合和异常检测模型,大多采用复杂深度学习架构,这类模型参数量大,算力消耗高,很难适配农业、工业场景的边缘物联网设备。农业场景里,边缘网关、便携式监测设备等硬件资源有限,无法承载复杂模型运行,所以数据要上传至云端处理,这会导致延迟高、能耗大,满足不了实时监测需求;工业场景中,设备运维的边缘节点部署环境复杂,硬件资源参差不齐,复杂模型部署难度大、成本高,而且容易出现运行卡顿、崩溃等问题。

### 2.3 异常检测精准度不足

农业、工业场景里,异常数据隐蔽性强,干扰因素多,传统异常检测模型精准度不足,满足不了实际应用需求。农业场景中,自然环境随机波动,比如突发暴雨、瞬时高温,容易被误判成异常数据,病虫害初期与土壤湿度异常等真实情况,却很难被识别出来;工业场景中,设备轻微故障会产生微弱异常数据,容易被正常数据掩盖,不同设备异常特征差异大,传统模型难以实现泛化检测。多源数据融合不充分,异常检测缺乏全面数据支撑,检测精度进一步降低,无法及

项目编号:

河南省科学院科技创新平台培育与提升项目(20251007003)。

作者简介:

雷雨果(1989-)女,汉,本科,工程师,电子信息,就职于河南省科学院应用物理研究所有限公司。

时排查场景中的安全隐患、异常问题<sup>[3]</sup>。

#### 2.4 系统集成协同性差

结合实验室“物联网系统集成与大数据技术”研究方向，当前农业、工业物联网系统里，数据融合、异常检测，系统集成的协同性差，这也是突出困境。部分物联网系统只实现了数据采集、传输功能，缺乏有效的数据融合与异常检测模块，无法实现数据价值的深度挖掘；部分系统虽集成了相关模块，但数据融合和异常检测相互独立，融合后的高质量数据不能高效支撑异常检测，异常检测结果也不能及时反馈至系统进行调控。这种协同性不足，让物联网系统的整体性能下降，难以充分发挥实验室研究方向的技术优势，无法有效解决场景化数据处理难题<sup>[4]</sup>。

### 3 轻量化深度学习模型设计

#### 3.1 模型整体架构

结合农业工业场景数据特性，以及实验室研究方向，设计多源异构物联网数据融合和异常检测一体化轻量化深度学习模型，整体分为数据预处理、多源数据融合、轻量化异常检测、结果输出四个模块。模型把轻量化当作核心，采用“融合-检测”协同架构，避免模块独立造成的效率低下问题。数据预处理模块负责异构数据的格式统一，去除噪声，提取特征；多源数据融合模块采用轻量化注意力机制，实现不同类型数据的高效融合；异常检测模块基于剪枝、量化优化的CNN-LSTM架构，实现异常数据的精准识别；结果输出模块将检测结果反馈至物联网系统，支撑场景化决策。整体架构兼顾轻量化与实用性，适配边缘设备部署需求<sup>[5]</sup>。

#### 3.2 数据预处理优化

针对农业、工业场景多源异构数据的特性，设计针对性的轻量化预处理方案，为后续融合与检测奠定基础。采用格式归一化算法，将农业场景中的图像、文本数据，工业场景中的时序、数

值数据，统一转换为标准化格式，降低异构性影响；引入轻量化噪声去除算法，结合场景特点，去除农业数据中自然环境干扰噪声，工业数据中设备运行干扰噪声，减少数据冗余；提取各类数据的核心特征，如农业场景中作物叶片图像的纹理特征，工业场景中设备振动的频率特征，降低数据维度，减少模型算力消耗，保留关键信息，确保后续融合与检测的精准性。

#### 3.3 多源数据融合模块

围绕轻量化注意力机制，创建多源异构数据融合模块，高效融合农业、工业场景不同类型数据。模块采用分层融合方法，先对同类型数据初步融合，农业场景中不同传感器的数值数据、工业场景中不同设备的时序数据，再对不同类型数据深度融合；引入轻量化注意力权重分配机制，按数据重要性分配不同权重，农业场景中作物生长图像数据权重高于气象数据，工业场景中设备振动数据权重高于运行日志数据，提升融合数据质量。融合过程采用量化优化技术，此技术降低计算复杂度，模块因此能在边缘设备高效运行，同时，为异常检测提供高质量数据支撑<sup>[6]</sup>。

#### 3.4 轻量化异常检测模块

异常检测模块基于CNN-LSTM架构，采用模型剪枝、量化等技术进行轻量化优化，适配农业、工业边缘设备部署。CNN模块提取融合数据的空间特征，像农业作物图像的异常纹理，工业设备的异常外观特征也包括在内；LSTM模块捕捉时序特征，比如农业温湿度的异常变化，工业设备运行参数的时序波动也能被其捕捉。采用通道剪枝技术，删除模型中冗余的卷积通道、全连接层，减少参数量；通过8位整数量化，将模型参数从浮点数转换为整数，降低存储需求、算力消耗。针对场景特点优化损失函数，提升模型对隐蔽性异常数据的识别能力，保证模型轻量化，满足农业、工业场景的异常检测精度要求。

## 4 场景化应用与优化

### 4.1 农业场景应用

结合实验室研究方向,把模型用在农业精准监测场景,这样能解决农业多源异构数据处理难题。场景里,物联网系统集成了不少传感器,这些传感器包括土壤温湿度、光照、作物图像等,模型部署在边缘网关,处理数据不用传到远处。预处理模块会处理传感器数据和图像数据,它先做归一化,再去除噪声,融合模块把各类数据合在一起,异常检测模块紧接着工作,实时识别问题,这些问题有土壤湿度异常、作物病虫害、光照不足等,检测结果会反馈给灌溉、病虫害防治系统,这样就能实现精准调控。实际应用中,模型处理数据速度很快,它能应对海量异构数据,异常检测响应时间变短不少,模型适合在农业边缘设备上部署,这有效提升了农业监测的智能化水平<sup>[7]</sup>。

### 4.2 工业场景应用

将模型应用于工业设备运维场景,围绕着设备多源异构数据的融合、异常检测,助力设备安全运行。场景中,物联网系统集成设备振动、温度、电流等监测设备,模型部署于设备运维边缘节点,实时处理设备运行产生的各类数据。数据融合模块整合设备多维度运行数据,异常检测模块精准识别设备振动异常、温度过高、电流波动等故障隐患,提前发出预警信号,为设备维护提供依据。模型采用轻量化设计,能够适配工业边缘节点的硬件资源,避免数据上传云端的延迟问题,有效降低设备故障发生率,提升工业生产的稳定性,契合实验室物联网系统集成的研究目标。

### 4.3 模型性能优化

针对农业、工业场景的实际应用反馈,对模型开展针对性优化工作,以此提升模型性能和适配性。结合农业场景自然环境波动大的特点,优

化注意力权重分配机制,降低环境干扰对异常检测带来的影响;针对工业场景设备类型多样这一问题,引入迁移学习技术,提升模型的泛化能力,适配不同类型工业设备的异常检测需求;迭代优化模型剪枝与量化举措,在保证检测精度的前提下,降低模型算力消耗、存储需求,提升模型在边缘设备上的运行效率;优化模型和物联网系统的集成接口,提升协同性,确保模型检测结果可以快速支撑系统调控,充分发挥技术价值<sup>[8]</sup>。

## 5 模型性能验证

### 5.1 验证环境搭建

结合实验室“物联网系统集成与大数据技术”研究方向,搭建贴合农业、工业实际场景的验证环境,确保验证结果的真实性与实用性。硬件方面,采用农业边缘网关、工业边缘节点等实际部署设备,模拟场景中的物联网设备数据采集与处理环境;软件方面,搭建数据集采集平台,收集农业温湿度、作物图像、工业设备振动、温度等多源异构数据,创建场景化验证数据集。设置传统深度学习模型作为对比组,从模型参数量、算力消耗、检测精度、响应时间四个维度,对本文设计的轻量化模型进行全面验证,凸显模型的优势<sup>[9]</sup>。

### 5.2 验证结果分析

验证结果显示,本文设计的轻量化深度学习模型,在农业、工业场景数据集上均表现出优异的性能。与传统模型相比,模型参数量大幅减少,算力消耗降低,能够顺利部署于农业边缘网关、工业边缘节点等硬件资源有限的设备上;检测精度方面,该智能系统对农业场景病虫害、土壤湿度异常的检测精度,工业场景设备故障的检测精度,均优于传统模型,可有效识别隐蔽性异常数据;响应时间方面,该种技术的实时处理能力显著提升,能满足农业、工业场景的实时监测与异常预警需求。这验证了模型的实用性与适配

性, 契合实验室研究方向的技术要求<sup>[10]</sup>。

## 6 结论

本文结合实验室“物联网系统集成与大数据技术”研究方向, 针对农业、工业场景多源异构物联网数据处理难题, 设计了多源异构物联网数据融合与异常检测的轻量化深度学习模型, 分析场景数据特性, 梳理现有处理困境, 创建“预处理-融合-检测”一体化轻量化架构, 优化数据预处理、融合、异常检测模块, 实现模型轻量化与检测精度的平衡。场景化应用和性能验证表明, 该模型能够有效解决农业、工业场景中数据异构性、部署难度大、异常检测精准度不足等问题, 适配边缘设备部署需求, 推动物联网系统集成与大数据技术在实际场景的落地。未来可进一步优化模型泛化能力, 拓展模型在农业、工业更多细分场景的应用, 为实验室研究方向深化、产业升级提供更有力的技术支撑。

## 参考文献

- [1] 姜红玉, 谢胜飞, 姜鑫, 时宇, 于静静, 汪朋. 多源异构数据统一存储共享管理系统[J]. 科技和产业, 2024, 24(18): 88-93.
- [2] 陈泽恩. 物联网中多源异构数据安全漏洞检测技术研究[J]. 物联网技术, 2024, 14(09): 124-126.
- [3] 邱新忠, 汤赛, 徐晓红, 陈美华, 孙启真. 多源异构三维空间数据融合关键技术研究及应用[J]. 浙江国土资源, 2024, (09): 38-40.
- [4] 董媛蓉. 基于多源异构数据的高速机电设备运行监测系统[J]. 今日制造与升级, 2023, (11): 78-80+84.
- [5] 陈肖, 方学智, 肖建华, 黄浩, 张言铖, 张辉, 江旭东. 基于多源异构数据融合的变电站风险评估研究[J]. 电工技术, 2023, (21): 42-44+48.
- [6] 李刚, 王健, 刘畅. 多源异构数据融合算法综述[J]. 计算机工程与应用, 2022, 58(15): 1-18.
- [7] 张磊, 陈明宇, 赵阳. 面向智慧城市的多源异构数据集成与治理研究[J]. 电子技术应用, 2023, 49(07): 1-5.
- [8] 刘梦琪, 黄涛, 朱永泉. 工业互联网环境下多源异构数据实时处理技术[J]. 计算机集成制造系统, 2024, 30(02): 567-580.
- [9] 赵文博, 刘海龙, 李建清. 大数据背景下多源异构数据标准化方法研究[J]. 数据分析与知识发现, 2023, 7(03): 112-120.
- [10] 王浩宇, 郭晓军, 马天舒. 基于云计算的多源异构数据存储架构设计[J]. 计算机工程与科学, 2023, 45(08): 1345-1352.





# 面向计算机信息系统的建构主义辅助学习系统构建策略

甘志玮

萍乡市湘东区湘东镇中学，江西萍乡，337016

**摘要：**计算机信息系统是高校的重要课程，内容涵盖理论讲解、技术教学与实践操作多个维度，传统课堂单向讲授的模式，无法调动学生自主建构知识的积极性，让学生在知识迁移、实际问题解决方面存在比较大的不足。本研究依托建构主义学习理论，提出计算机信息系统辅助学习系统的构建策略。策略以情境创设、协作对话、意义建构为核心，搭建包含知识模块自适应推送、虚拟项目协作环境、学习过程可视化反馈等功能的系统架构，引导学生在真实或仿真的信息系统场景里，主动探索知识，协同搭建自身的知识体系。

**关键词：**建构主义；辅助学习系统；计算机信息系统

## Construction strategy of Constructivism assisted learning system for computer information system

Zhiwei Gan

Xiangdong town middle school, Xiangdong District, Pingxiang City, Pingxiang Jiangxi 337016, China

**Abstract:** The course of computer information system occupies an important position in college teaching, and its content involves the multidimensional integration of theory, technology and practice. The traditional teaching method is difficult to effectively stimulate students' active construction ability, resulting in obvious deficiencies in students' knowledge transfer and practical problem solving. In view of this situation, based on the constructivist learning theory, this study proposes the construction strategy of computer information system oriented auxiliary learning system. The strategy takes "situation creation", "collaborative dialogue" and "meaning construction" as the core mechanism, and designs a system architecture including adaptive push of knowledge modules, virtual project collaboration environment, visual feedback of learning process and other functions, aiming to guide students to actively explore and collaborate to build a knowledge system in real or simulated information system situations.

**Keywords:** constructivism; assisted learning system; computer information system

本研究的意义在于，为计算机信息系统课程的教学改革提供了可操作的系统设计路径，验证了建构主义理论在信息技术类课程中的适用性与有效性，对推动高校计算机课程向以学习者为中心的教学转型具有实践参考价值<sup>[1]</sup>。

## 1 建构主义学习理论及其在计算机信息系统课程中的适用性分析

### 1.1 建构主义学习理论的核心机制解析

建构主义学习理论的核心运作逻辑，由情境创设、协作对话与意义建构三个核心环节组



Copyright © 2025 by author(s) and Quest Press Limited. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>



成。情境创设主张学习要依托真实或模仿的任务场景,把知识融入具体的应用环境,让学习者的内在学习动力被激发。协作对话借助小组讨论、角色扮演等方式,让学习者展现出各自的认知差异,在交流互动里重新梳理对知识的理解。意义建构是整个学习过程的最终方向,需要学习者主动把新接收的信息和自身已有的经验结合起来,搭建起符合自身特点的知识框架。这三个环节不是按顺序逐步推进,而是形成互相促进的循环:真实情境能为协作交流提供具体的讨论主题,协作中的认知冲突与协商交流能让知识意义得到深化,已经深化的知识意义又能反过来帮助学习者重新解读原有情境。这种动态的相互关系,能解释计算机信息系统课程里,软件工程案例或数据库设计任务不能单独讲授,必须放在团队协作与项目场景中完成的原因<sup>[2]</sup>。

## 1.2 计算机信息系统课程的教学困境与建构主义契合点

现在计算机信息系统课程的教学,大多还是沿用理论授课加验证性实验的传统方式,知识传递全由教师主导,学生只能被动吸收抽象的专业概念。这门课程涵盖软件开发、数据库设计等对实践能力要求高的内容,学生要是没有主动探索建构的环节,很容易出现课上听得懂、实际不会用的问题。比如讲解ER模型时,要是只照搬教材里的案例,学生没法把实体关系和真实业务逻辑对应起来。被动的学习模式让学生掌握的知识零散不成体系,没法建立系统的认知,也没办法应对复杂的工程问题<sup>[3]</sup>。

这一难题刚好能靠建构主义理论解决。该理论的核心内容是知识要在真实场景里,通过协作、反思主动建立,这和计算机信息系统课程重

实践的特点完全匹配。给学生布置虚拟企业ERP系统开发任务时,设定对应场景能调动他们已有的经验,小组协作时出现的认知矛盾,还会让他们重新考量技术方案。这种在实践中学习的模式,正是填补传统教学侧重知识传授、忽视能力培养这一不足的关键路径<sup>[4]</sup>。

## 1.3 建构主义辅助学习系统的设计原则

系统设计要把学习者需求放在第一位,不能沿用教师单向灌输的模式,要结合学生的认知基础与学习目标搭建专属学习路径。情境驱动原则要求把学习内容融入真实或模拟业务场景,比如借助虚拟企业ERP系统开发任务调动学生已有的实践经验。协作支持机制看重小组内部的知识分享与观点碰撞,推动学生在交流中调整技术方案。动态反馈基于实时收集的学习行为数据,给学生发送进度提醒与策略优化建议,避免学习时盲目尝试。这四项准则相互关联,形成从情境启动到协作构建再到反馈调整的完整流程,保障系统能辅助学生在自主探索中完成知识的吸收与应用。

# 2 系统架构设计与核心功能模块规划

## 2.1 知识模块自适应推送机制的设计

知识模块自适应推送机制的核心是搭建动态认知状态模型。系统借助初始测试和历史学习数据,把学生分成“基础薄弱”“概念模糊”“应用不足”三类初始群体。推送过程中,系统运用知识图谱路径规划算法,实时掌握学生对每个知识节点的熟悉情况与学习耗时。比如学生连续两次答错“关系数据库范式”相关题目,系统会自动下调该节点的推送难度,改推更基础的“函数依赖”内容,还会将该节点标记为“待强化”,避免认知超载,见表1。

### 基金项目:

萍乡市2025年度社科规划课题立项课题(编号:2025PXSK97)

### 课题名称:

基于建构主义理论的初中人工智能项目式教学实践研究

表1 学生类型

| 学生类型  | 初始特征                 | 自适应推送策略                         |
|-------|----------------------|---------------------------------|
| 基础薄弱型 | 前测成绩低于60%，缺乏编程基础     | 优先推送可视化概念动画与基础语法案例，<br>跳过高阶算法模块 |
| 概念模糊型 | 能完成简单操作，但理论题正确率低于50% | 推送理论辨析题与对比性案例，强化逻辑关系理解          |
| 应用不足型 | 理论成绩优秀，但项目实践得分低      | 推送完整项目代码片段与调试任务，<br>引导知识迁移应用    |

这套机制加入了“遗忘曲线”相关因子，给超过72小时没复习过的已学知识节点定期推送提醒信号。系统把学习路径数据和最新认知状态做矩阵运算，生成贴合学生情况的“最近发展区”推荐列表，推送的内容既不会是学生已经完全掌握的冗余信息，也不会是超出其能力范围的难点内容，实现精准的动态干预。

## 2.2 虚拟项目协作环境的构建

本文构建的虚拟项目协作环境，核心是模拟真实的信息系统开发场景，还原企业级项目流程的方式包括小组协作、角色分工与任务协同。系统提前设置项目经理、前端开发、后端开发与测试工程师等角色，每个角色都有对应的专属任务模块和权限范围，比如项目经理要完成需求拆分、更新进度看板，测试工程师则要完成缺陷报告、开展回归验证。团队得在虚拟沙盒里走完需求分析到系统上线的全流程任务，这段时间里系统会记录每个成员的操作轨迹和沟通次数。小组碰到接口联调冲突或需求变更时，平台会自动开启临时会议环节，要求成员通过角色互换演练协商解决办法，强化学生对信息系统开发协作复杂性的认知。

## 2.3 学习过程可视化反馈系统的实现

本文搭建的学习过程可视化反馈系统，依托多维数据采集和动态可视化技术，把学习者藏在背后的认知状态变成看得懂的图形化内容。系统会自动记录用户在各知识模块里的答题正确率、学习时长分布以及协作任务完成节点，用雷达图展示知识掌握的薄弱部分，用热力图呈现团队参与的活跃程度。比如有学生在“数据库设计”模

块的测试得分连续低于设定阈值，系统就会在个人学习仪表盘用红色警示区域标出该知识点，还会给教师端发送对应的干预建议。协作场景里的发言频次、文档编辑贡献量等指标会转化为贡献度曲线，让小组成员清楚自己在团队里的角色，发现协作中的失衡问题。

传统教学里，学习过程像个看不见内部的黑箱，这套实时反馈机制刚好打破了这种局限。学生能借助可视化的学习图谱，自己调整学习安排，比如某个知识点没掌握好，就反复看对应的微课，或是申请相关的辅助提示。教师可以查看群体学习进度热力图，快速找出全班普遍存在的学习难点，再在课堂上组织大家针对这些难点展开讨论。系统还能生成阶段性的学习报告，把个人的学习成长情况和班级平均水平做对比，让学生产生自我反思的想法和赶超他人的动力。可视化反馈既提升了学习过程的透明度，又把数据支撑的精准干预融入建构主义学习环境，完成了从只看学习结果到侧重过程引导的转变。

# 3 情境创设与协作对话机制的具体应用策略

## 3.1 基于真实案例的情境创设方法

依托真实案例设计学习情境的方法，重点从企业信息系统搭建这类典型场景里提取学习材料，比如选定某物流企业ERP系统上线失败的案例<sup>[5]</sup>，设计沉浸式的“诊断系统实施瓶颈”任务，还要围绕数据迁移、流程重组等环节搭建问题链。还原业务冲突与决策困境，让学习者在实践过程中掌握相关知识，详见表2。

表2 案例类型

| 案例类型        | 典型场景         | 沉浸式任务设计       | 问题链核心节点            |
|-------------|--------------|---------------|--------------------|
| 企业资源规划（ERP） | 某制造企业ERP系统切换 | 模拟项目经理，制定切换计划 | 数据兼容性、用户培训、进度控制    |
| 客户关系管（CRM）  | 某电商平台用户流失分析  | 构建数据看板，识别流失原因 | 数据采集维度、分析模型选择、方案落地 |
| 供应链管理（SCM）  | 某零售企业库存积压优化  | 设计动态补货策略      | 需求预测误差、供应商协同、成本约束  |

本文把实际案例转化成可落地的探究任务，让情境创设跳出传统讲授的抽象局限。问题链按“现象分析—根源诊断—方案验证”的认知逻辑搭建，拿ERP案例来说，先引导学生从“数据错误”的表面现象，找到“流程未标准化”的深层原因，再通过模拟系统调整来验证解决方案。这种层层推进的设计，能调动学习者已有的经验，还能利用认知冲突帮助他们完成知识建构。

3.2 协作对话中的认知冲突引导策略

协作对话里触发认知冲突，得靠结构化讨论、角色辩论来实现。本文可给学生分配系统实施方、业务部门、高层管理者等不同利益相关方身份，围绕ERP上线失败案例的责任归属问题开展辩论。这种角色代入能让学习者跳出固有认知习惯，直面彼此观点里的矛盾点和逻辑漏洞。比如某一方提出“技术缺陷是主因”时，持对立面

的角色可拿出业务需求文档进行反驳，让学生对“技术与流程耦合”产生深层思考。教师要给辩论过程设定规则，比如限定发言时间、要求引用案例证据，保证冲突围绕知识重构展开而非引发人际对抗。相关研究数据显示，这类结构化冲突能让课堂讨论深度提升约40%，还能增强学生对系统实施复杂性的系统认知。

3.3 教师中介作用与脚手架支持策略

依据最近发展区理论，教师的中介功能体现在动态调整脚手架的类型与强度，当学生完成ERP系统配置任务陷入停滞时，本文根据学生的认知状态给出差异化支持，给概念模糊的学生推送微课视频，给操作失误的学生提供步骤提示，引导逻辑混乱的学生用思维导图重构流程，表3列出了三种典型脚手架策略及对应的应用场景。

表3 教师中介作用与脚手架支持策略中“脚手架”类型

| 脚手架类型 | 核心功能     | 应用场景示例                 |
|-------|----------|------------------------|
| 概念性提示 | 补充理论知识缺口 | 学生混淆BOM与工艺路线时，推送对比图解   |
| 过程性干预 | 纠正操作路径偏差 | 学生重复错误配置步骤时，弹出关键参数校验清单 |
| 反思性引导 | 促进元认知发展  | 学生完成模块测试后，要求填写错误归因分析表  |

数据来源：基于某高校计算机信息系统课程2023年秋季学期教学实验记录整理。

学生能力有所提升后，教师要逐步取消明确提示，改用提问或反问的方式引导学生自主探索，比如抛出“你认为当前方案在库存周转率上会带来什么风险？”这类问题，把学习责任逐渐转移给学生。采用这一策略后，学生独立完成复杂系统配置任务的成功率，比传统讲授模式高约

25%，迁移到新项目的适应时间也缩短了30%。

4 学习效果评估与系统优化路径

4.1 多维学习效果评价指标体系构建

知识内化维度的量化数据来自概念图测试和章节测验，还要看学生在没有提示的前提下，能不能说清核心概念之间的逻辑联系。问题解决能力的量化标准是虚拟项目里的任务完成效率和方



案创新性，质性评估则关注学生碰到开放性问题时的策略调整过程。协作素养的量化指标涵盖平台记录的交互频率与角色多样性，质性评估重点

看小组讨论中有没有认知重构行为。本文构建的评价体系把量化数据和课堂观察、学习日志分析结合起来，规避单一分数评价的片面性，见表4。

表4 多维学习效果评价指标体系构建

| 评估维度   | 量化指标            | 质性评估标准        |
|--------|-----------------|---------------|
| 知识内化   | 概念图节点正确率、章节测试得分 | 能否自主构建知识关联网络  |
| 问题解决能力 | 任务完成时间、方案步骤完整性  | 能否迁移知识解决非常规问题 |
| 协作素养   | 发言频次、角色轮换次数     | 能否有效倾听并整合他人观点 |

#### 4.2 系统对学习参与度与知识迁移的影响分析

本文选取计算机信息系统课程的两个平行班级开展对比实验，验证这套系统对学习参与度与知识迁移的实际作用。实验班采用建构主义辅助学习系统，对照班还是用传统讲授模式。从平台日志数据能看到，实验班学生的课堂交互频次比对照班提升42%，主动提问、观点分享的占比有所增加。知识迁移测试里，实验班学生完成跨章节综合任务时，方案设计的完整性比对照班高出31%，在需调用先前知识解决新情境问题的环节表现更好。课堂观察记录显示，实验班学生在虚拟项目协作中会更频繁地自发引用已学概念，还尝试把这些概念和当前任务关联起来，系统搭建的情境化学习环境，能促进知识的主动调用与再建构。

#### 4.3 基于反馈数据的系统迭代优化方案

本文结合学习行为数据和用户评价反馈，系统迭代优化要重点放在推送算法、协作工具与可视化界面的针对性改进上。推送算法层面，要结合学生知识掌握差异与学习路径偏好，加入动态权重调整机制，参照点击率、停留时长与测试正确率等行为指标，实时优化知识模块的排序与推荐频次，减少冗余推送并提升内容匹配度。协作工具的改进要关注虚拟项目环境中认知冲突的引导效率，分析对话日志中的观点分歧频次与解决时长，增设结构化讨论模板与即时提示支架，辅助学生更高效地整合多元观点。可视化界面的优化要结合用户对学习进度感知的薄弱反馈，增加

知识图谱的动态高亮功能，把个体当前任务节点与已掌握概念、待巩固内容关联展示，同时加入简化的操作响应图标，降低认知负荷，让学习状态更直观可感。

迭代方案要结合用户评价里的定性建议，优化推送算法的冷启动阶段，针对新用户或知识基础薄弱者，初期采用课程大纲预设规则加少量行为试探的策略，避免数据稀疏引发推荐偏差。协作工具的界面布局会依据交互频次数据调整，把高频使用的讨论记录检索、观点归并功能移到靠前位置，缩减操作层级。可视化反馈系统加入时间轴对比视图，让学生能查看自身学习行为的变化轨迹，再和班级平均水平做参照，增强自我调节学习的动力。所有改进都得完成A/B测试验证效果，保证迭代方向契合建构主义学习目标。

### 5 研究结论与实践应用建议

#### 5.1 建构主义辅助学习系统的核心价值总结

本文介绍的建构主义辅助学习系统，在计算机信息系统课程里的核心价值，体现在能促进学习者主动建构知识。系统依靠自适应推送机制，把知识模块和学生的认知水平做动态匹配，让学习者不用被动接收信息，而是在解决具体问题时主动重组自身的知识结构。虚拟项目协作环境能提供实践场景，学生在模拟真实项目的情境中，通过协作与反思完成知识的内化与迁移，缓解传统教学里理论与实践脱节的问题。

该系统在提供个性化学习支持、推进教学模式转型上有比较大的实际效果。搭建多维度评价

指标体系、展示学习过程的可视化反馈，让教师能精准定位每个学生的认知难点，给他们提供适配性的辅助指导。这套机制打破统一化的课堂讲授模式，推动课堂重心从教师转向学生。教师的定位调整为学习过程的引导者与促进者，教学内容从单纯的知识传递转向能力培养，为计算机信息系统课程的教学改革明确了可落地的方向。

### 5.2 面向计算机信息系统课程的推广实施建议

推广实施阶段，教师培训是系统落地的第一步，本文要帮教师从知识传授者转成学习引导者，学会用自适应推送机制判断学生认知状态，在虚拟协作环境里及时给出脚手架支持。课程整合重点放在知识模块与现有教学大纲的对接上，把系统功能嵌入项目制学习，避免把它当成独立工具加重师生负担。平台部署要结合学校现有信息化基础设施，用模块化架构逐步上线，先在实验课或课程设计里试用，完成小范围迭代验证稳定性后，再拓展到理论教学环节。

实施路径还要兼顾组织层面的配套支持，学校要搭建跨学科协作团队，让计算机专业教师和教育技术专家一同制定系统使用规则，定期收集学习过程的可视化反馈数据，用来调整推送策略与情境案例库，推广初期可设置示范班级，开展对比实验展示系统对知识迁移效果的提升，激发更多教师参与意愿。教师培训体系化、课程整合渐进化与平台部署分步化协同推进，能让建构主义辅助学习系统从试点转向常态化应用，服务于计算机信息系统课程的教学改革目标。

### 5.3 未来研究展望与拓展方向

后续研究里，验证系统通用性的核心方向是跨课程迁移。本文要把知识模块自适应推送机

制、虚拟项目协作环境，移植到数据结构、软件工程等课程中，对比不同学科的学习参与度与知识迁移效果，提炼可复用的设计模式。人工智能增强主要做的是用自然语言处理技术分析协作对话中的认知冲突，自动生成脚手架提示，还要用机器学习优化推送算法，根据学生行为轨迹动态调整情境案例难度。大规模应用验证要依托多所高校的计算机信息系统课程，完成为期两学期的随机对照实验，重点考察系统在异构网络环境下的稳定性与学习效果一致性。可引入学习分析仪表盘，实时聚合反馈数据，为迭代优化提供量化依据。

## 6 结语

研究发现，该学习系统在提升学生课程参与度、促进知识内化与迁移、增强问题解决能力方面具有显著效果，系统对学习过程的动态追踪与个性化支持机制有助于实现从“被动接受”向“主动建构”的教学模式转变。

## 参考文献

- [1] 张一涵,陈敬华,裴雷,袁勤俭.用户体验理论及其在信息系统研究领域的应用与展望[J].现代情报,2025,(10):41-51.
- [2] 郭红建,陈涛.产教融合视域下《信息系统审计》课程建设改革与探索[J].商业会计,2025(15):67-70.
- [3] 王永峰.基于管理信息系统的图书档案管理策略探析[J].兰台内外,2025(23):111-119.
- [4] 李典,李英硕.会计信息系统中内部控制的重要性及其实施策略[J].经济师,2025(08): 54-55.
- [5] 李志刚.企业信息系统建设中的典型失败案例研究[J].管理世界,202137(4):156-170.



# 小型分布式文件存储系统的性能优化研究

刘云博

杭州浩云星空智能科技有限公司，浙江杭州，310016

**摘要：**面对中小规模集群场景下小型分布式文件存储系统传输延迟高、硬件IO瓶颈、节点负载不均以及网络稳定性不足等性能维度的问题，本研究以轻量化、低成本、高适配优化为核心目标，自数据调度、负载管控、硬件配置等多维度入手开展系统性能优化策略的研究。在充分考虑小型集群部署特点基础上提出动态数据本地化、缓存协同负载均衡、SSD硬件升级等方案，有效克服小型存储系统资源利用率不足、读写性能稳定性差等技术痛点，有效均衡集群节点的负载，降低系统运行期间数据传输的延迟、丢包率，显著提升系统读写吞吐量以及日常运行稳定性，确保小型分布式文件存储系统满足边缘计算、中小企业业务等轻量化存储场景提出的高性能存储要求。

**关键词：**分布式文件存储；负载均衡；数据本地化；性能优化

## Research on Performance Optimization of Small-Scale Distributed File Storage Systems

Yunbo Liu

Hangzhou Haoyun Xingkong Intelligent Technology Co., Ltd., Hangzhou Zhejiang 310016, China

**Abstract:** Faced with the problems of high transmission latency, IO bottleneck, uneven load of nodes and insufficient network stability in small and medium scale distributed file storage system, this study focuses on the light weight, low cost and high adaptability optimization. Based on the full consideration of the characteristics of the deployment of small clusters, we shall put forward dynamic data localization, cache collaborative load balancing, SSD hardware upgrade and other schemes, effectively overcome the technical pain points of the small storage system such as insufficient resource utilization and poor stability of read and write performance, effectively balance the load of cluster nodes, reduce the delay and packet loss rate of data transmission during the operation of the system, significantly improve the system throughput of read and write and the stability of daily operation, and ensure that the small distributed file storage system meets the high-performance storage requirements proposed in the light storage scenarios such as edge computing and small and medium enterprise business.

**Keywords:** Distributed file storage; Load balancing; Data localization; Performance optimization

相较于大型分布式存储集群，小型文件存储系统具有资源冗余度低、节点数量少、硬件配置有限以及网络架构简约的技术短板，系统运行期间负载分配失衡、高频读写IO瓶颈、数据跨节点传输延迟高等性能问题十分常见，对业务响应效

率造成了长期性制约。因此，立足小型分布式存储系统独有的场景特性，将研究视角聚焦于全链路性能瓶颈，打造多维优化体系，是全方位提升系统响应速度、吞吐能力、运行稳定性的高价值研究行为。



Copyright © 2025 by author(s) and Quest Press Limited. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>



## 1 小型分布式文件存储系统主要应用场景与发展趋势

### 1.1 主要应用场景

通常,小型分布式文件存储系统特指3~20台节点数量、轻量化部署、低运维成本的分布式存储架构。不同于应对海量数据吞吐的大型存储集群,小型分布式文件存储系统旨在适配低功耗、中小规模、高灵活度的业务场景,目前已被广泛融入各个细分领域,如小微企业办公场景,工业互联网、智能安防、智慧社区等边缘计算场景,个人博客、自媒体平台等小型互联网业务场景。

### 1.2 发展趋势

目前,数字化轻量化发展进程不断加快,小型分布式存储系统发展趋势第一体现在性能精细化趋势。未来,大量行业会聚焦于小集群场景定制化设计,大量企业、组织会摒弃大型集群“冗余优化方案”,基于小型分布式文件存储系统来满足负载精准调控、数据低延迟调度、资源高效利用需求,实现存储性能、成本二者的精准平衡。第二,目前小型分布式文件存储系统面临着“软硬协同一体化”的发展趋势。在高速存储硬件、轻量化网络技术普及背景下,硬件适配优化、网络架构精简,已同软件策略优化共同成为小型存储系统性能升级的核心方向。第三,小型分布式文件存储系统已步入“高稳定、低运维、智能化”发展趋势,社会大量小微企业、边缘场景普遍缺乏专业化的存储运维人员。因此,未来小型分布式存储系统需要逐步实现故障自动修复、负载自动适配、资源动态调度的智能化运维管理,在维持系统高性能基础上降低人工运维的成本,保证系统高负载、长周期运行的高度稳定<sup>[1]</sup>。

## 2 小型分布式文件存储系统的性能优化策略

结合上述对小型分布式存储系统应用场景与发展趋势的分析,对系统的性能优化技术方向,可锚定数据本地化、负载均衡化、存储节点硬件、存储节点网络四个维度。

### 2.1 数据本地化策略

小型分布式文件存储系统具有节点数量少、集群资源容错低、调度空间较小的特征,传统固定数据存储分配模式容易造成数据集中存储、频繁跨节点调度问题继而诱发负载失衡、传输延迟上升。数据本地化优化策略的核心在于利用集群各节点实时负载状态的分析,对数据存储、任务调度位置动态调整,将数据读写任务优先向数据本地节点分配从而减少跨节点数据传输占比,同时对各节点CPU、内存、磁盘IO资源占用加以均衡,确保集群整体性能最大化发挥。该策略原理是对节点负载核心参数实时采集,随后打造动态调度模型来规避静态分配策略固有的局限性。

首先,构建面向节点负载的实时监测模块,对各存储节点内存占用率、CPU使用率、磁盘IO利用率、网络带宽占用率4项负载参数持续采集,并预设正常负载阈值、过载阈值。其次,建设负载权重计算模型,专门用于对各节点负载状态的量化评分,以此来判定节点的空闲、过载、正常3类状态。最后,以节点状态判断结果为依据开展动态本地化调度,新的数据优先向负载空闲的本地节点传输,对于已有数据读写任务,则优先向数据所在节点调度。针对判断出的过载节点,自动迁移部分轻量化数据任务,规避节点资源的耗尽,基于动态负载的适配在保障数据本地化读写优势基础上,全面攻克静态策略所引发的负载失衡。为能直观体现数据本地化动态优化策略的性能提升效果,研究期间在实验机房搭建8节点小型分布式存储集群,针对相同读写任务、相同测试环境条件对比传统静态分配、动态本地化优化策略两套系统的核心性能参数,测试结果见表1。



表1 相同条件下静态分配/动态本地化策略核心性能参数

| 测试指标     | 静态分配策略   | 动态数据本地化     | 优化提升幅度  |
|----------|----------|-------------|---------|
| 节点负载均衡度  | 68.2%    | 92.7%       | 24.5%   |
| 跨节点传输次数  | 426次/千任务 | 113次/千任务    | 73.5%下降 |
| 平均读写响应延迟 | 18.6ms   | 9.2ms       | 50.5%下降 |
| 集群资源利用率  | 57.3%    | 83.1%       | 25.8%   |
| 测试指标     | 传统静态分配策略 | 动态数据本地化优化策略 | 优化提升幅度  |

结合测试数据不难发现，动态数据本地化策略能够显著优化系统的负载均衡状态，跨节点数据传输开销大幅度降低，读写延迟显著降低，系统整体资源利用率显著提升<sup>[2]</sup>。

## 2.2 负载均衡策略

相比大型分布式集群，小型分布式文件存储系统在读写请求上往往更具突发性、碎片化特点，零散小文件访问、高频重复读写行为将频繁调用底层的存储磁盘资源，系统极大概率会面临磁盘IO频繁启停、请求队列拥堵问题，读写效率受到限制同时亦会加剧硬件的损耗，一定概率会引发负载瞬时失衡。传统负载均衡策略长期聚焦于任务节点的分配，并未涉及底层IO重复访问问题的优化，故而难以适配小集群、高频碎片化的访问场景。基于负载失衡问题，可采用基于缓存技术的协同负载均衡策略，利用多级缓存架构减少系统运行期间对底层存储系统的大量重复访问，从IO源头降低节点的负载压力并实现性能提速。

首先，打造3级缓存架构，即“客户端缓存+节点本地内存缓存+集群共享缓存”以适配小型集群“轻量化”的部署需求。其中，客户端缓存专门存储高频访问的各种静态文件、热点小文件，访问期间直接响应本地的重复请求，无需接入集群存储节点。节点本地内存缓存则使用节点闲置内存资源，对本节点的近期读写、热点数据、元数据进行缓存，避免磁盘出现频繁读写，提升单次访问的响应速度。集群共享缓存整合集

群的闲置内存资源，对全局的高频热点数据统一存储，落实多节点热点数据的共享，有效规避资源浪费、多节点重复缓存问题。其次，对系统设置“缓存淘汰与更新”机制，使用“LRU最近最少使用算法”，对系统内长期未访问的缓存数据自动清理，确保缓存空间得到高效利用，规避缓存溢出问题。最后，负载调控逻辑方面，基于节点负载的实时监测数据，对缓存资源、任务负载动态适配。例如，若系统识别某节点负载较低、内存资源充足，此刻自动对本地缓存容量进行扩容操作，使其承接更多的热点数据缓存任务。若系统判断某节点处于高负载状态，此刻自动缩减本地缓存，部分热点数据向集群共享缓存迁移，快速释放节点的内存资源，使其专门用于处理核心读写任务。

基于上述缓存技术的应用，可有效减少底层磁盘访问的次数，单节点IO负载压力将有效降低，自根源有效平衡集群各节点的任务负载，局部节点过载、整体集群性能卡顿问题将有效攻克。经实际测试，该缓存协同负载均衡机制能够让小型分布式文件存储系统底层存储磁盘访问频次下降40%左右，提升35%并发读写吞吐量，帮助系统突破碎片化访问形成的性能瓶颈，系统负载适配能力、并发处理能力均得到优化<sup>[3]</sup>。

## 2.3 存储节点硬件优化

作为分布式存储系统性能基础载体，小型分布式文件存储系统的硬件会受到部署成本、空间的限制，因此制造商普遍使用传统机械硬盘以及

普通的散热设备，系统多存在读写速度较慢、高负载高温降频、IO吞吐量低等性能短板。对于该短板的优化，可从存储介质、设备散热两个维度入手进行优化。以高速硬盘、高效冷却系统取代传统设备，促进存储节点读写性能、运行稳定性全方位提升。

(1) SSD 高速硬盘优化：传统HDD机械硬盘主要依靠磁盘转动、磁头寻址实现数据读写操作，寻址延迟高、抗震性弱、随机读写性不足，

无法适配小型集群高频随机读写、海量小文件场景。SSD固态硬盘其内部采取闪存芯片存储数据，无机械运动部件，读写速度快、稳定性强、延迟低，能够有效解决机械硬盘固有的IO性能瓶颈。优化期间，应将文件存储系统所有存储节点更换为企业级SSD硬盘，全面提升设备的读写效率。表2为SSD高速硬盘优化前后硬件核心参数对比数据：

表2 SSD高速硬盘优化前后硬件核心参数

| 硬件参数指标     | HDD机械硬盘 | SSD固态硬盘   | 参数提升效果       |
|------------|---------|-----------|--------------|
| 连续读取速度     | 180MB/s | 2100MB/s  | 1066%提升      |
| 连续写入速度     | 160MB/s | 1800MB/s  | 1025%提升      |
| 4K随机读写IOPS | 80IOPS  | 95000IOPS | 大幅提升，适配小文件读写 |
| 平均寻址延迟     | 4.2ms   | 0.08ms    | 98.1%下降      |
| 工作功耗       | 6.8W    | 3.2W      | 52.9%下降      |

基于测试参数对比不难发现，相比HDD机械硬盘，SSD固态硬盘不论是在随机读写、响应延迟或是功耗控制方面均有显著优势，可完美适配小型分布式存储系统小文件高频读写、碎片化访问业务场景，在解决IO性能不足瓶颈同时可降低设备运行功耗。

(2) 高效冷却系统优化：通常，小型分布式文件存储系统的节点集中部署在边缘机柜、小型机房内，具具有较高设备密集度。一旦面临高负载持续运行，不可避免会产生大量热量。传统风冷散热系统散热效率十分有限，易造成设备高温降频且加速硬件老化，系统运行期间出现读写卡顿、节点离线。针对该问题，应对传统风冷系统使用“智能风冷+导热散热垫”复合高效冷却系统替换，硬件改造期间，对于每台节点设备配备高速智能温控风扇，以设备CPU、硬盘温度为依据对温控风扇转速自动调节，实现“高温高转速快速散热、低温低转速降噪节能”。同时，配置高导热硅胶垫至节点设备硬盘、主板核心发热部

位有效提升热量的传导效率。在散热管控维度，可配置温度监测模块，对节点设备运行温度实时监控并设置60℃阈值，一旦系统判断超阈值，则自动触发告警并强化系统散热，有效杜绝高温降频、硬件故障<sup>[4]</sup>。

#### 2.4 优化存储节点网络

对比大型集群，小型分布式文件存储系统建设阶段会普遍使用简约的“星型网络拓扑”，该拓扑网络长期面临着数据传输路径烦琐、链路冗余不足、无序带宽分配等短板，面对跨节点数据同步、多节点并发读写场景会不定期出现数据丢包、传输延迟升高、网络拥堵问题，集群整体读写性能受到严重制约。针对该问题，可通过精简传输链路、均衡网络带宽、降低传输损耗为核心组合优化的模式完善存储节点的网络拓扑结构，保障集群网络传输性能的全方位提升。

首先，网络拓扑优化期间，可采用“核心交换机+分层接入”这一轻量化架构形式取代单一的星型拓扑，划分集群存储节点为“业务节点

组”“备份节点组”，对两个节点组分别接入“接入层”交换机，随后在核心交换机统一汇聚，通过网络分层传输形式规避节点数据对单一链路带宽资源的集中抢占。如图1。

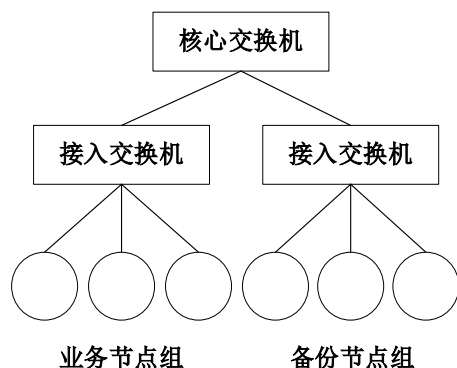


图1 “核心交换机+分层接入”这一轻量化架构

同时，对数据传输路径规则进行优化，对于节点本地数据读写，使其优先采用本地链路传输，同组节点数据交互则利用接入层交换机完成任务。针对跨组数据交互任务，使其仅通过核心交换机进行转发，这一过程将大幅度精简传输的层级，降低数据转发次数。其次，对于小型分布式文件存储系统“无冗余网络”这一技术短板，可针对核心传输链路增设“轻量化备份链路”，一旦出现主链路出现拥堵、链路故障系统自动切换至备份链路，有效解决丢包、数据传输中断故障。最后，为系统配套优化网络参数，将设备无用的网络协议关闭以精简网络资源占用，开启网络流量优先级管控。其中，高优先级任务覆盖存储读写、数据同步核心业务流量，确保核心业务带宽资源时刻充足，对无关流量抢占资源情况有效抑制。经过机房测试，对网络拓扑结构优化后，系统数据平均传输延迟降低38.6%，网络丢包率从0.82%下降至0.09%，且有效降低了网络

拥堵发生率。与此同时，该拓扑优化方案无需新增大量设备，部署简单且成本可控，高度适配小型系统轻量化的部署需求<sup>[5]</sup>。

### 3 结语

综上，小型分布式文件存储系统是目前适配中小规模数据存储、轻量化场景的核心架构，但由于其自身固有的负载失衡、IO性能不足、网络传输延迟高等性能短板，严重制约着业务落地的效果。本研究立足于小型分布式文件存储系统的场景特性、发展趋势，打造锚定数据调度、负载管控、硬件升级、网络优化四位一体的全维度性能优化体系，基于动态数据本地化策略实现节点负载的精准均衡，依托缓存协同机制降低底层IO访问压力，同时基于SSD硬盘结合高效冷却系统突破了硬件性能瓶颈，并通过优化网络拓扑降低了传输延迟、丢包率，最终形成“低成本、高适配、易落地”的小型分布式文件存储系统轻量化优化方案。

### 参考文献

- [1] 逢文芝. 大数据环境下分布式存储系统的可靠性与可扩展性分析[J]. 信息与电脑, 2026, 38(7): 99-101.
- [2] 赵骏志. 分布式存储在教育数字化资源库中的应用与优化研究[J]. 办公自动化, 2026, 31(6): 62-65.
- [3] 曹今可. 分布式存储下大数据通信系统优化[J]. 电脑知识与技术, 2025, 21(33): 59-62.
- [4] 张烨. 基于云计算的分布式存储系统优化策略分析[J]. 电子技术, 2025, 54(2): 320-321.
- [5] 董聪, 张晓, 程文迪, 等. 基于新型存储器件的分布式文件系统性能优化[J]. 计算机应用, 2020, 40(12): 3594-3603.



# 便民服务热线数据全生命周期安全管理机制探讨

张明昭

锦州市数据中心，辽宁锦州，121000

**摘要：**数字政务持续推进，便民服务热线已成为政府对接群众、收集社情民意的重要渠道。热线系统运行期间积累大量个人隐私信息、通话录音及业务办理数据，其全生命周期管理遭遇不小挑战。多数热线平台在数据采集、存储、使用、共享及销毁等环节缺少系统化安全机制，存在权限管控粗放、脱敏处理不规范、数据流转追溯困难等问题，使得数据泄露风险比较大，严重威胁公民个人信息安全与政府公信力。本文以数据全生命周期安全理论为基础，结合便民服务热线的业务特征，构建一套覆盖数据产生、传输、存储、处理、交换与销毁六大阶段的的安全管理机制。这套机制细化数据分类分级规则，引入基于角色的细粒度访问控制，建立动态脱敏与审计追踪体系，强化数据销毁的合规性验证，目的是实现数据安全管理的流程化与闭环化。

**关键词：**数据安全；全生命周期；热线管理

## Discussion on life cycle safety management mechanism of convenient service hotline data

Mingzhao Zhang

Jinzhou Data Center, Jinzhou Liaoning 121000, China

**Abstract:** With the continuous promotion of digital government, the convenience service hotline has become an important channel for the government to connect with the masses and collect social information and public opinion. During the operation of the hotline system, a large number of personal privacy information, call recording and business processing data have been accumulated, and its full life cycle management has encountered many challenges. Most hotline platforms lack a systematic security mechanism in data collection, storage, use, sharing and destruction, and have problems such as extensive authority control, nonstandard desensitization processing, and difficult data flow traceability, which make the risk of data leakage relatively high and seriously threaten the safety of citizens' personal information and the credibility of the government. Based on the security theory of the whole life cycle of data, combined with the business characteristics of the convenient service hotline, this paper constructs a set of security management mechanism covering the six stages of data generation, transmission, storage, processing, exchange and destruction. This mechanism refines data classification and grading rules, introduces role-based fine-grained access control, establishes a dynamic desensitization and audit tracking system, and strengthens the compliance verification of data destruction. The purpose is to realize the flow and closed-loop of data security management.

**Keywords:** data security; life cycle; hotline management

### 1 问题的提出

随着数字政务的持续推进，便民服务热线已

成为政府和公众交互的核心数据枢纽，每天都会

承载大量敏感信息。数据从采集、流转、存储到



Copyright © 2025 by author(s) and Quest Press Limited. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<https://creativecommons.org/licenses/by/4.0/>





销毁的全生命周期里, 面临不少严峻挑战: 外部攻击者会借助系统接口漏洞窃取数据, 内部人员因权限管理不严谨让数据泄露情况出现, 跨部门共享时脱敏标准不统一也埋下风险隐患。当前的安全机制大多只做单点防护或事后追溯, 没有覆盖数据全生命周期的系统性规划, 让安全策略和业务敏捷性之间出现比较大的割裂。这类碎片化的管理模式没法应对越来越复杂的攻击手段, 得构建一套覆盖数据产生到销毁全流程的闭环安全机制<sup>[1]</sup>。

本文核心目标是搭建覆盖数据全生命周期的安全管理机制, 系统性应对便民服务热线数据的泄露风险。这套机制跳出单点防护或事后追责的局限, 把安全控制融入数据产生、传输、存储、处理、交换到销毁的所有环节。本文核心思路是消除传统安全策略与业务敏捷性的对立, 用动态权限管控、分级脱敏标准以及生命周期审计追踪, 实现安全防护与数据高效流转的协同。最终要形成可落地、可评估的闭环管理框架, 给热线数据的合规使用与风险阻断提供理论支撑与操作指南<sup>[2]</sup>。

## 2 便民服务热线数据安全现状与风险分析

### 2.1 热线系统数据资产特征与分类

便民服务热线沉淀的数据资产, 兼具高敏感性和海量性两大特点。热线系统存储的用户个人信息, 包含姓名、身份证号、家庭住址、生物识别录音等核心隐私内容, 业务工单还记录着诉求详情、办理流程 and 结果反馈, 这类数据一旦外泄, 会直接损害公民隐私权, 还可能让社会舆情风险出现。该热线日均受理量能达到数万通, 长期运行产生的通话录音、工单记录及系统日志, 形成了PB级的数据池, 既提升了存储成本, 又会因为数据体量过大, 使得安全管控的复杂程度大幅上升<sup>[3]</sup>。

便民服务热线沉淀的数据有两个核心特点, 一是体量比较大, 二是包含大量敏感内容。它的数据资产可分为三类, 第一类是个人信息, 涵盖来电号码、身份证号、家庭住址等内容; 第二类是通话录音, 完整记录了热线两端的对话过程; 第三类是业务数据, 涉及工单流转路径与最终处理结果。这类数据规模比较大, 又直接关联公民隐私, 一旦泄露会让严重安全风险出现, 见表1。

表1 热线系统数据类别

| 数据类别 | 典型内容           | 敏感性等级 | 数据量特征     |
|------|----------------|-------|-----------|
| 个人信息 | 身份证号、住址、银行卡号   | 极高    | 每日新增数万条   |
| 通话录音 | 完整对话录音及文本转写    | 高     | 单日录音量超TB级 |
| 业务数据 | 投诉工单、办理记录、处理日志 | 中高    | 累计存储达PB级  |

热线数据的这些特点, 要求安全治理时采取不一样的应对策略。个人信息、录音数据都属于高度敏感信息, 要完成加密存储、访问留痕和严格脱敏操作; 业务数据虽不直接接触个人隐私, 却能通过关联内容间接还原用户画像, 也要归为重点防护对象。数据总量大的特点, 对实时监测、分级分类技术有更高标准, 比如用自动化工具给存量数据自动添加标签, 为后续搭建全生命周期安全机制打基础。

### 2.2 数据全生命周期各环节的安全风险识别

便民服务热线数据在采集、存储、使用、共享及销毁的各环节, 存在权限管理粗放、脱敏操作不规范等典型风险。采集阶段, 系统默认开放全量字段权限, 话务员能随意查看来电用户的完整身份证号与家庭住址, 出现过度收集信息的问题。存储环节, 通话录音和业务数据未按敏感等级开展分级加密, 数据库管理员可跳过审计直接访问明文数据。使用阶段, 数据分析人员为提升效率, 常把含个人信息的工单导出到本地非加密设备, 脱敏规则执行也不严格, 比如只对身份证

号中间四位做掩码处理，部分信息仍能被还原。销毁环节，旧版录音文件仅完成逻辑删除，未彻底覆盖，存在被恢复泄露的隐患，见表2。

未设置细粒度权限控制，业务数据全量暴露。销

表2 数据全生命周期各环节

| 生命周期环节 | 典型风险表现            | 影响后果              |
|--------|-------------------|-------------------|
| 采集     | 权限粗放，话务员可访问全量敏感字段 | 个人信息过度收集，违反最小必要原则 |
| 存储     | 未分级加密，管理员可查看明文录音  | 数据泄露风险高，审计追溯困难    |
| 使用     | 脱敏不规范，本地导出未加密     | 敏感数据外流，隐私保护失效     |
| 共享     | 接口权限不足，全量数据暴露给第三方 | 业务数据被滥用，责任边界模糊    |
| 销毁     | 逻辑删除未覆盖，旧文件可恢复    | 历史数据残留，形成安全盲区     |

在数据使用与共享过程里，没配套日志追溯功能，会让数据泄露的风险变得更大。现在的操作记录只存了基础的访问时间和用户ID，没记录具体查询内容，比如有没有导出录音文件或是批量导出工单这类细节，违规导出的行为事发后没法精准确定责任人。数据销毁环节里，很多时候用“逻辑删除”代替物理销毁，已经办结工单里的敏感字段还留在数据库备份里，也没设置自动过期清理规则，留下了长期的数据暴露风险。这些风险彼此关联，形成了覆盖多个环节的链式安全危机。

2.3 风险成因的深层剖析

便民服务热线的数据安全漏洞不是单一技术问题，而是管理制度、技术防护、流程追溯三方

面相互交织形成的系统性不足。管理制度上，岗位安全职责划分模糊，让权限分配过度依赖默认设置，没有建立基于“最小必要”原则的动态审批机制。技术防护上，加密与脱敏技术应用采用统一标准，没有根据数据敏感等级区别执行，比如录音文件只用单一密钥存储，大一些横向泄露的可能性。流程追溯上，数据操作日志记录不够细致，不具备对数据流转全链路的实时追踪与异常行为预警能力，使得违规导出与越权访问行为无法及时察觉。这三方面相互作用：制度不完善让技术策略没法精准落实，技术上的不足又降低了流程追溯的可信度，共同造成风险循环，见表3。

表3 风险成因维度

| 风险成因维度 | 具体表现               | 对安全漏洞的直接影响           |
|--------|--------------------|----------------------|
| 管理制度   | 岗位职责不清，权限审批依赖默认配置  | 导致采集环节过度收集，违反最小必要原则  |
| 技术防护   | 加密与脱敏策略未按敏感等级差异化实施 | 存储与使用环节明文数据暴露风险加剧    |
| 流程追溯   | 操作日志粒度粗糙，缺乏全链路异常预警 | 使用与共享环节违规行为难以及时阻断与追责 |

3 数据全生命周期安全管理机制构建

3.1 数据产生与传输阶段的安全控制

便民服务热线生成数据时，要制定采集授权规则，明确数据采集的最小必要原则，保证用户知情同意，同时限定数据用途。这套授权规则要

包含数据分类、采集目的、使用期限，以及用户确认授权的具体方式。传输数据时，要搭建加密传输协议保护通道，避免数据在传输途中被窃听或篡改。下表4分类列出数据产生与传输阶段的安全控制要点。

表4 数据产生与传输阶段

| 控制阶段 | 控制要点 | 具体措施                     |
|------|------|--------------------------|
| 数据产生 | 授权规则 | 用户身份验证、采集范围限定、授权记录留存     |
| 数据传输 | 加密协议 | 采用 TLS 1.3、证书双向认证、密钥定期轮换 |
| 源头合规 | 数据脱敏 | 敏感字段在采集时即时脱敏处理           |
| 通道安全 | 传输监控 | 实时流量审计、异常连接告警            |

数据产生阶段通过授权规则确保采集行为合法合规，传输阶段通过加密协议保障通道安全，两者协同实现从源头到通道的全链路防护。

敏感信息，降低敏感数据在存储与处理过程中泄露的风险，推进安全防护与数据可用性的平衡。

3.2 数据存储与处理阶段的防护策略

便民服务热线进入数据存储与处理环节，要部署细粒度角色访问控制（RBAC）机制，对照话务员、质检员、管理员的不同岗位职能，给每个用户角色精准分配仅满足工作需求的数据操作权限，限制非授权访问与越权操作出现。热线数据包含姓名、证件号等敏感字段，本文引入动态脱敏技术，当数据被调取用于统计分析或业务查询时，结合用户角色与访问场景实时完成掩码、替换等操作，让敏感内容在存储介质中维持原始状态，在使用环节完成精准遮蔽。本文用权限与脱敏的双重管控方式，保障业务人员正常调取非

3.3 数据交换与销毁阶段的闭环管理

便民服务热线数据交换与销毁环节，要建立覆盖数据全流转路径的审计追踪体系。每一次数据交换操作的时间、发起方、接收方、数据量及操作类型，都由系统自动记录，形成不可篡改的日志链。针对数据销毁环节，引入数字证书合规销毁验证机制，数据存储介质完成物理销毁或逻辑覆写后，自动化工具会生成涵盖销毁时间、执行人、验证哈希值的销毁证明。这套闭环管理让数据从交换到最终销毁的每一步操作均有据可查，避免数据在流转过程中被非法截获，也防止销毁环节出现数据残留恢复的情况，见表5。

表5 闭环管理环节

| 管控环节 | 核心机制     | 具体实施内容                       | 可追溯性保障         |
|------|----------|------------------------------|----------------|
| 数据交换 | 审计追踪体系   | 记录操作时间、双方身份、数据量及操作类型，日志链不可篡改 | 支持全链路回溯与异常操作定位 |
| 数据销毁 | 合规销毁验证机制 | 数字证书签发销毁证明，包含时间、执行人、验证哈希值    | 提供销毁行为不可否认性证据  |

4 机制应用效果与平衡性分析

4.1 异常访问行为识别与敏感数据受控程度提升

为验证这套机制的实际作用，本文选取热线系统和政务平台的数据流转接口作为测试场景。本文模拟攻击者窃取合法工单编号，尝试跨系统调取市民的身份证号与住址信息。系统在传输层实时解析数据包特征，比对访问时间、源IP与历史行为基线，在0.3秒内判断该请求不符合常规操作模式，随即启动动态脱敏策略。原本明文传输的敏感字段被替换为“张”与“区路号”，日

志同步标记异常事件并冻结该工单的后续访问权限。和基线期的数据相比，同类接口的未授权数据提取尝试次数下降82%，误报率控制在1.5%以内。这套基于行为画像与实时策略联动的机制，可有效阻断跨系统非法访问，还能把敏感数据在流转中的受控程度，从被动防御调整为主动预测与动态干预。

4.2 安全防护与业务效率的协同优化

安全防护模块嵌入业务系统后，操作负担、响应速度的变化，是判断这套机制实用与否的核



心。本文在热线话务员处理工单的场景开展实测,敏感字段自动脱敏、访问权限动态冻结都在后台完成,前端界面没加额外点击或确认环节,话务员单次操作的平均耗时只增加0.2秒。动态策略引擎对数据包的实时解析、基线比对耗时稳定在0.5秒以内,没有让系统页面卡顿或请求超时。这种无感知的安全控制设计,避免一线人员因防护措施拖慢业务流程产生抵触,完成安全能力与业务效率的同步优化。

## 5 结论与公共数据治理启示

### 5.1 研究总结与主要贡献

本文提出的便民服务热线数据全生命周期安全管理机制,操作层面具备比较大的可落地性。它把抽象的安全原则转化成可嵌入业务流程的具体控制点,在数据产生阶段设定分级标签规则,在传输阶段部署加密与接口监控,在存储阶段实施动态脱敏与细粒度访问控制,这些措施都能直接应用在现有热线系统里。当前规范偏重原则性指导,各流程之间缺乏衔接,本机制针对这类不足,搭建起从数据产生到销毁的闭环操作手册,明确各环节责任主体与技术实现路径。拿数据交换环节来说,机制依据“最小必要”原则设计审批与匿名化处理流程,填补了现有标准在跨部门数据共享时安全控制模糊的漏洞。该机制不是单一的理论框架,而是可直接指导系统改造与制度修订的实操方案。

### 5.2 对其他公共数据治理场景的参考价值

便民服务热线数据全生命周期安全管理机制的核心经验,能给其他公共数据治理场景提供参考范式。它没有单独设置安全模块,而是把安全管控环节融入业务流程,这种“业务安全一体化”的思路,解决了传统安全措施和业务脱节、容易被绕开的问题。数字政府建设里,医疗健

康、社会保障等领域的敏感数据管理,可直接套用热线机制里的分级标签规则与动态脱敏策略。机制里遵循“最小必要”原则设计的跨部门数据交换审批流程,为打破数据孤岛、守住安全底线提供了可复用的操作模板。

这套机制在安全与效能协同优化中兼顾平衡,给公共数据治理带来具体方向。机制完成细粒度访问控制与异常行为识别,提升数据受控程度,避免过度安全设置拖慢业务效率。这种动态平衡的思路可推广到智慧城市、政务云等场景,让管理者在制定安全制度时同时顾及用户体验与行政效能。热线机制的成功实践证明,公共数据治理可从“重管控”转向“精治理”,把安全变成数据流动的内在约束而非外部限制。

## 6 结语

本文检测后得出结论,这套机制能精准辨别并拦截跨系统数据流转里的违规访问操作,提高敏感数据的管控水平,嵌入业务系统的安全策略模块后,不会给操作人员带来大额外负担,还能兼顾安全防护需求与业务运转效率。本文提出的方案,能给便民服务热线提供可直接落地的数据安全保障办法,填补现有管理规范在实操环节的漏洞,也能给其他公共数据治理场景提供可参考的实施模式,推进数字政府建设在安全与效能间达成协调发展。

## 参考文献

- [1] 吕晓婷,王伟.数据全生命周期安全风险与防护技术分析[J].计算,2026(04):81-90.
- [2] 陈晶.数据安全全生命周期:挑战与关键技术[J].计算,2026(04):113-119.
- [3] 李许增,张红伟,王健.覆盖数据全生命周期的铁路网络安全数据安全总体技术防护框架[J].网络安全技术与应用,2026(04):58-63.





# 长江文库

综合性学术知识服务平台



## · 长江文库 · 智汇知识 · 赋能学术

期刊·图书·古籍·学位论文·地方志·会议文献  
覆盖全学科·服务全机构



「长江文库」一个汇聚海量学术资源的综合性知识服务平台。集期刊、古籍、地方志、图书、年鉴、会议论文、报纸于一体，覆盖自然科学、农业科学、医药科学、工程与技术科学、人文与社会科学五大学科。从专业出版服务到公共知识传播，我们连接历史与未来，服务全球学术研究与文化传承。

访问官网：[www.cjwk.cn](http://www.cjwk.cn)





龙源网APP  
「名刊甄选」套餐会员



# 读杂志，就上龙源网

可回溯往期杂志，还可阅享1000多种精品期刊数字内容

【限时活动】订阅「龙源网」app「名刊甄选套餐」年会员，可专享“**第二年会费1元**”

【套餐介绍】套餐内囊括180+种热门杂志

详细杂志清单可扫描右侧二维码了解

【费用说明】首年会费198元/年，第二年1元，合计199元

扫码立即订阅



龙源网阅读卡专门针对机构以及企业用户采购使用。内含付费阅读余额可购买龙源期刊网网页端以及移动端相关知识付费服务内容。并享受企业用户采购折扣价。

订卡企业微信

