

# 融合大数据的农村劳动力信息集成与决策支持平台

文 | 韩丽君

目前，我国农村劳动力规模达到 2.99 亿之多，资源分配不均衡阻碍了城乡一体化进程，就业市场对接能力还需要进一步加强。农业农村部大数据发展中心努力激活数据资源价值，构建农村劳动力信息集成与决策支持平台，该平台整合数据、融合知识图谱及深度学习等先进技术，打造从信息获取到智能决策的全链条解决方案。

## 农村劳动力多源异构数据采集与融合体系

农村劳动力信息集成与决策支持平台（以下简称“平台”）依靠 Hadoop 分布式文件系统和 Spark 计算引擎来构建数据采集框架，凭借统一的数据接口规范协议整合人社、农业农村、统计等 17 个政府部门以及 328 家企业的多样化数据源。通过 ETL 流程对关系型数据库、文本文件、API 接口等多源异构数据开展提取、转换及加载工作，完成数据清洗与标准化操作，进而建立包含身份信息、技能资质、从业经历、培训记录等要素的劳动力资源数据仓库。平台引入数据质量管控体系，借助字段完整性校验、逻辑一致性校验、重复数据识别等方法，让数据完整度从 56.7% 提升到 91.2%。基于数据湖架构实现结构化数据、半结构化日志数据以及非结构化文档数据的集中存储管理，且能够支撑 TB 级数据的秒级查询响应，为劳动力资源画像分析与智能化决策提供坚实的数据支撑。

## 基于知识图谱的劳动力画像构建与智能匹配

### 农村劳动力知识图谱构建方法

平台选择 Neo4j 图数据库搭建劳动力知识图谱，以劳动者个人作核心节点去连接技能、岗位、培训、企业等多维度数据，形成“实体-关系-属性”的三元组网络架构。知识提取模块借助命名实体识别技术，从招聘广告、培训材料、职业资格证明等文本里提取职业类型、技能标准、薪酬范围等关键数据，再利用实体链接技术把分散信息整合成标准化实体节点。图谱构建时参考《国家职业分类大典》，将其作为本体框架，构建包含制造业、建筑业、服务业等六大行业共 89 个工种的技能分类体系，每个工种平均关联 6.2 项核心技能与 3.8 项辅助技能。关系推理机制通过图神经网络分析实体间隐藏联系，自动挖掘技能转移路径与职业发展路线，为 537 项技能节点分配权重值，实现对劳动力能力精细化描述与实时更新。

### 深度学习驱动的岗位匹配算法

这个算法把 BERT 预训练语言模型和图卷积网络架构融合在一起，把劳动力技能画像及岗位需求描述转化成 768 维

语义向量表示。模型使用双塔结构分别去处理劳动力和岗位方面的信息，依靠注意力机制捕捉技能要求和个人能力之间的精细匹配关系，并且将此作为语义相似度得分的计算依据。通过图卷积层整合知识图谱里的结构化关系信息，聚合劳动力的历史从业经历、培训路径及技能关联等领域特征，从而提升向量表示的判别能力。该算法还引入多任务学习框架，对岗位匹配和薪资预测任务进行联合优化，通过共享底层特征提取器增强模型泛化性能。在训练过程当中，采用对比学习策略构建正负样本对，结合交叉熵损失函数与 Margin Ranking 损失函数的加权组合进行端到端的优化，经过 12 轮迭代训练之后，模型在验证集上的 Top-5 推荐准确率稳定在 89.5%。

### 匹配效果对比实验与精度验证

平台在 6 个试点县区应用匹配算法，选取常规协同过滤、基于规则的专家系统及单一 BERT 模型当作基准来进行测试。评估体系涵盖推荐准确率、召回率、平均倒数排名及归一化折损累积增益四项指标，采用 2023 年全年的 18.7 万条真实匹配记录作为测试数据。各算法在不同维度的表现对比情况可见表 1。结合知识图谱的深度学习算法在 Top-1 准确率达到 72.3%，比协同过滤高出 36.2%，较规则系统高出 28.7%，在冷启动用户场景下的召回率优势更为突出，达到 68.9%。跨行业技能迁移专项测试显示，图卷积网络通过发现技能潜在关联让匹配精度提升 19.4%，算法响应时间稳定在 320 毫秒以下，能够满足高并发在线推荐需求，有助于缩短求职周期并提高就业质量。

## 时空大数据驱动的就就业趋势预测与辅助决策

### 区域用工需求时序预测模型

预测模型搭建起长短期记忆网络和时空卷积网络的组合框架，通过整合历史用工记录、经济发展数据及产业政策信息等多源时序特征，达成区域劳动力需求的精准预测目标。此方法把研究区域细致划分为 158 个县级空间单元，每个单元采集月度岗位发布量、薪资均值及产业结构占比等 12 项时序指标，并且基于 48 个月的滚动时间窗口生成输入序列。

表 1 不同匹配算法性能对比

算法模型	Top-1 准确率 (%)	Top-5 准确率 (%)	召回率 (%)	MRR	NDCG
协同过滤算法	36.1	53.3	41.2	0.428	0.562
规则专家系统	43.6	61.8	48.5	0.513	0.624
单一 BERT 模型	65.8	82.1	59.7	0.728	0.781
知识图谱 + 深度学习	72.3	89.5	68.9	0.812	0.856

来源：六安市金安区人力资源和社会保障局

LSTM 模块凭借其门控结构捕捉劳动力市场的长期演化趋势与周期性变化特征，而时空卷积层则承担识别相邻区域的空间关联性及产业协同效应的任务。模型创新性地引入注意力机制，实现对不同时间和空间位置的权重进行动态分配，进而提升对季节性用工峰值和异常事件的响应能力。在训练阶段采用滑动窗口交叉验证方案，同时把均方根误差与平均绝对百分比误差当作优化目标，其损失函数表达式如下：

$$L = \alpha \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} + \beta \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%$$

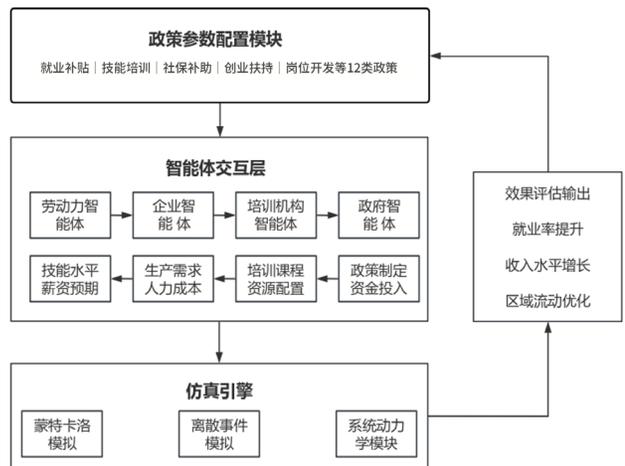
其中  $y_i$  表示实际用工需求量， $\hat{y}_i$  表示模型预测值， $\alpha$  和  $\beta$  为权重系数，经网格搜索确定为 0.6 和 0.4。该模型通过持续改进算法，在未来 3 个月人力资源需求预测上实现 84.3% 的准确率，在半年预测方面也维持 78.6% 的准确水平，为合理配置人员及规划培训方案提供可靠数据支持。

### 政策模拟仿真与效果评估系统

这项研究采用多智能体建模技术搭建农村劳动力就业决策仿真平台，把劳动者、用人单位、培训单位及政府部门转化成具备独立决策能力的智能体。在仿真模型里，劳动者智能体根据技能水平、薪资期望和地域倾向等因素做出职业选择，企业智能体结合生产需求、用工成本及政策导向去调整用人策略，通过智能体之间的互动循环再现就业市场的动态变化。该系统能够灵活配置就业补贴、技能培训、社保支持等 12 项政策变量，采用蒙特卡洛方法开展千次随机模拟，分析不同政策组合对就业率、收入水平及区域流动的作用效果。多智能体仿真系统架构与政策评估流程如图 1 所展示。仿真引擎融合 AnyLogic 平台的离散事件模拟与系统动力学模块，实现宏观政策效应与微观个体行为的综合分析。职业技能培训补贴政策的仿真结果显示，补贴额度每增加 500 元，可让劳动力参与培训的概率提高 8.2%，就业稳定性提升 5.7%，为精准制定政策提供科学依据。

### 决策支持实验验证与应用效果

平台在洛南县等 6 个试验地区上线决策辅助体系，此体系融合劳动力需求预估等核心单元，构建综合性决策支持



来源：六安市金安区人力资源和社会保障局

图 1 政策模拟仿真系统架构与运行流程

架构。该体系依托数据看板，动态展示地区劳动力关键信息，方便管理者开展多层次数据挖掘与情景对比。运行期间产出 127 份政策推演报告，助力完成多项季度性培训规划与跨区域劳务对接协议。监测结果表明，应用平台的试验地区劳动力外就业数量同比增加、规模化输出占比提高、求职平均耗时减少。平台促进县、乡、村三级就业服务体系搭建，通过手机端应用为众多劳动者提供线上服务，使用户就业稳定性提高且月收入增长，证明大数据赋能决策模式在农村人力资源优化配置方面具有实践成效。

### 结束语

平台把全国范围内 11.07 亿亩农村承包地及 96 万家农村集体经济组织的信息数据进行了整合，已经完成 22.76 亿项赋码登记工作，“农事直通”应用程序覆盖了 106 万服务用户群体。以洛南县为例，借助平台帮助 7.98 万名农村劳动力实现转移就业，创造出 11.4 亿元的经济价值，其中贫困劳动力稳定就业人数达到 4.4 万名。土地承包经营权流转台账系统备案了超过 400 万份流转合同，为超过 1600 万农户提供保险核验和信贷评估服务。该系统为农村人力资源的优化配置提供了技术保障，未来能够进一步推进区域协同机制并且拓展服务功能。

作者简介：韩厢君 六安市金安区人力资源和社会保障局

责任编辑：孙姗姗 投稿邮箱：zhouhl@staff.ccidnet.com