

基于 Apriori 算法的铁路行业就业人才画像构建与特征挖掘研究——以某二级学院 2024-2026 届铁路局录用毕业生为例

胡驰

陕西铁路工程职业技术学院 陕西省渭南市 714000

DOI: 10.32629/ems.v8i8.21558

[摘要] 在铁路系统招聘高标准与缩招的背景下，精准刻画铁路行业录用人才特征，服务高职院校就业指导精准滴灌。本文以某二级学院 2024-2026 届入职铁路局的 503 名毕业生为研究对象，利用 Apriori 关联规则算法，挖掘学业成绩、政治面貌等 10 项核心指标。研究发现，铁路局招聘选拔呈现综合素质导向特征，专业排名前 30% 与共青团员构成人才画像的基础，其次干部经历与奖学金荣誉表现出极强的关联规则，最后无违纪、无挂科、无色弱色盲及无纹身构成不可逾越的刚性门槛。本文构建了铁路行业就业人才画像模型，并提出“学业—实践—品德—学业”闭环培养策略，为提升高职毕业生人岗匹配度提供了数据支撑。

[关键词] Apriori 算法；关联规则；人才画像；铁路就业；数据挖掘

Research on the Construction of Employment Talent Portrait and Feature Mining in the Railway Industry Based on Apriori Algorithm

——A Case Study of Graduates from the College (Classes of 2024 - 2026) Employed by Railway Bureaus

Hu Chi

Shaanxi Railway Institute Weinan Shaanxi 714000

[Abstract] Against the backdrop of high recruitment standards and downsizing in the railway system, accurately characterizing the traits of hired talents in the railway industry is crucial for implementing precise career guidance in vocational colleges. Taking 503 graduates from the College (classes of 2024 - 2026) who were successfully employed by railway bureaus as the research subjects, this study utilizes the Apriori association rule algorithm to deeply mine ten core indicators, including academic performance and political status. The findings reveal that the recruitment and selection process of railway bureaus demonstrates a significant “comprehensive quality orientation.” Specifically, ranking within the top 30% of the major and holding the status of a Communist Youth League member constitute the foundation of the talent profile. Furthermore, student cadre experience and scholarship honors exhibit extremely strong association rules. Finally, clean records (no disciplinary violations, no failed courses), along with the absence of color weakness and tattoos, form insurmountable rigid thresholds. Based on these findings, this paper constructs a talent portrait model for employment in the railway industry and proposes a closed-loop cultivation strategy of “Academics-Practice-Morality-Academics,” providing data support for improving the person-job fit of vocational college graduates.

[Key words] Apriori algorithm; association rules; talent portrait; railway employment; data mining

1. 引言

在“交通强国”的背景下，目前，高速铁路与重载铁路以及保持持续发展，铁路运维行业已进入高质量发展的关键时期。据行业数据显示，2024 年铁路客运量已恢复至 2019 年的 120%，加之货运增量行动与智能化转型的双重驱动，铁路行业对高素质技能人才的需求依旧旺盛。但随着铁路局招聘形势呈现出“结构性缩招”与“标准精细化”的新趋势，铁路局在总体招聘计划上有所收紧。以西安铁路局为例，2024 年与 2025 年公示招聘录取专科生数量分别为 2001 人和 1558 人，但 2026 年专科生招聘计划缩减至 936 人。这对高职毕业生群体的综合素质与核心胜任力提出了更严苛的要求。因此，如何精准刻画铁路局青睐的“人才画像”，已成为铁路类高

职院校就业指导工作亟待解决的关键问题。

当前传统的经验式就业指导已经难以满足铁路局精准化服务的需求，在“铁路类专业背景”的前提下，毕业生的技能证书、有无国奖等多维度的综合表现，成为铁路局选拔人才的重要标准。为挖掘此类数据背后复杂的关联规则，笔者通过引入 apriori 算法，对某二级学院 2024-2026 届成功入职铁路局的毕业生数据进行深度分析。通过分析支持度、置信度和提升度，识别频繁共现的特征组合^[1]，量化其对录用结果的影响强度，旨在构建一套基于可视化数据的铁路行业就业人才画像模型^[2]。为铁路类高职院校就业指导提供科学依据，实现毕业生精准就业。

2. 数据来源与预处理

本研究所用数据来源于某二级学院 2024–2026 届毕业生就业数据库，筛选出最终成功被铁路局录用的学生样本，共计 503 份有效数据。

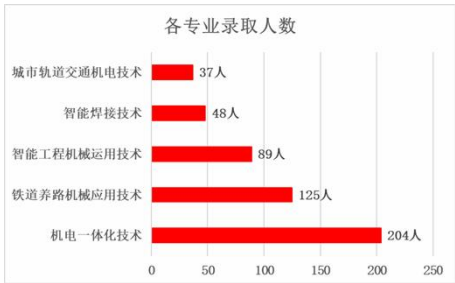


图 1 各专业录取人数

为构建人才画像，本文选取了与铁路局招聘标准密切相关的 10 个核心特征指标，具体包括：

表 1 毕业生特征核心指标

变量类别	特征名称	变量符号	含义说明
专业排名	专业排名前 30%	X1	是=1 否=0
奖项荣誉	国奖、省奖	X2	是=1 否=0
政治面貌	共青团员等	X3	
是否学生干部	学生干部	X4	是=1 否=0
是否获奖学金	奖学金	X5	是=1 否=0
是否纹身	无纹身	X6	是=1 否=0
是否色盲色弱	无色盲色弱	X7	是=1 否=0
是否违纪	无违纪	X8	是=1 否=0
是否有技能证书	技能证书	X9	是=1 否=0
是否挂科	挂科	X10	是=1 否=0

由于原始数据为结构化表格数据且 Aprior 算法要求输入数据为二值化的事务型数据。在此基础上，笔者对原始数据进行预处理，具体步骤如下：

(1) 数据清洗。对原始数据进行去重处理，注重检查缺失值。针对部分学生某些信息缺失的字段，通过咨询其辅导员等方式进行补充，确保数据完整真实。

(2) 二值化编码。对原始数据特征进行二值化转换。符合条件特征标记为 1，不满足条件特征标记为 0。

3. Apriori 算法与模型构建

3.1 Apriori 算法原理

目前，aprior 算法主要应用于关联规则挖掘方面。其核心包括频繁项集、支持度和置信度。频繁项集通常可以挖掘出数据的高频信息，支持度可以衡量项集在数据集中的出现

频率，置信度可以评估关联规则的可靠性^[3]。笔者在研究者通过设定合适的置信度和置信度阈值，挖掘铁路局录取学生特征的关联规则，从而在铁路行业就业人才画像构建中发挥重要作用。

3.2 算法流程

笔者在研究中主要分为：挖掘频繁项集和生成关联规则两大步骤。

(1) 挖掘频繁项集

①扫描数据集寻找所有单个项目的支持度，筛选满足最小支持度阈值的频繁 1-项集；

②利用迭代以及剪枝方法，逐步生成更高阶的频繁项集，直至不能继续生成最新的频繁项集为止^[4, 5, 6]。

(2) 生成关联规则

在频繁项集的基础上，计算置信度并生成满足设定的最小置信度阈值的强关联规则^[4, 7]。

3.3 模型构建

基于预处理后的数据，笔者构建了适用于铁路行业就业人才特征挖掘的 aprior 算法模型。通过初始化频繁 1-项集开始构建模型，经过多次迭代生成更高阶的频繁项集，最终形成完整的频繁项集集合。研究中笔者利用哈希树作为数据存储的主要结构，其可以将事务数据映射到不同的哈希桶中，显著降低候选项集的比较次数，提升频繁项集的发现速度。同时，笔者在模型中还加入了生成关联规则模块，通过计算频繁项集之间的置信度，确定满足最小置信度阈值的强关联规则^[4]，其表明铁路行业就业人才在专业技能和综合素质等方面的关联关系，为后续铁路行业录取人才画像构建了丰富的特征信息。

经过多次测试，本研究将最小支持度设定为 0.2，最小置信度设定为 0.8。此外，笔者通过设定合理的剪枝策略和先验约束，有限减少了无效候选项集的数量，提升算法的运行效率。以上不仅增强模型的实用性，还为后续在更大范围内应用 Apriori 算法进行铁路行业人才特征挖掘提供了技术支持。

4. 基于 Apriori 算法的铁路局录用特征挖掘与分析

笔者通过 aprior 算法，挖掘某二级学院 2024–2026 届铁路局录取毕业生特征关联规则，旨在揭示影响铁路局录取结果的关键组合特征及其内在逻辑，为学生进行高效的就业指导教育，实现学生高质量就业。

在设定最小支持度和最小置信度阈值后，aprior 算法挖掘出以下多条高置信度和提升度的强关联规则：

表 2 强关联规则

序号	关联规则	置信度	提升度
1	专业排名前 30%，共青团员，无违纪=>学生干部，奖学金	0.91	1.04
2	专业排名前 30%，共青团员=>学生干部，奖学金	0.91	1.04
3	专业排名前 30%，共青团员=>无违纪，学生干部，奖学金	0.91	1.04
4	专业排名前 30%，共青团员=>学生干部，奖学金，无挂科	0.88	1.03
5	专业排名前 30%，共青团员，无违纪=>学生干部，奖学金，无挂科	0.88	1.03
6	专业排名前 30%，共青团员=>无违纪，学生干部，奖学金，无挂科	0.88	1.03

7	专业排名前 30%, 共青团员, 无挂科=>学生干部, 奖学金	0.91	1.03
8	专业排名前 30%, 共青团员, 无挂科=>无违纪, 学生干部, 奖学金	0.91	1.03
9	专业排名前 30%, 共青团员, 无挂科, 无违纪=>学生干部, 奖学金	0.91	1.03
10	专业排名前 30%, 共青团员, 奖学金=>无违纪, 学生干部	0.91	1.03

通过挖掘样本数据之间的关联规则,结合 aprior 算法分析结果,可以清晰揭示铁路局在招聘过程中的核心特征与逻辑。

(1) 学业成绩与政治面貌是双重基准

学业成绩在专业排名前 30%是铁路局录用的基础门槛,录取的学生中有 90%达到了这个标准,这也验证了模型的预测结果,说明学业优秀是能否被铁路局录取的基础条件。

政治面貌虽以团员为主,是与其在样本数据中团员基数较大有关。然而,在铁路局的招聘过程中,中共党员或中共预备党员往往更具优势,在同等条件会优先录取,这说明铁路系统当中对于政治可靠性有更高的要求。

(2) 综合素质的显性标志

铁路局比较看重干部经历和奖学金荣誉。如规则 2 显示,当毕业生具备专业排名前 30%和共青团员的身份时,其有 91%的概率同时担任学生干部,并获得奖学金。结果中的提升度均大于 1,表明前后事项之间存在正相关关系,且这种关联并非属于偶然现象。以规则 1 为例,当毕业生具备良好的学业基础(专业排名前 30%)、合格的政治素养(共青团员)以及端正的品行记录(无违纪)时,有 91%的概率同时具备“学生干部”经历和“奖学金”荣誉。这揭示了铁路局录用人才的一个逻辑:学习优秀的学生通常综合素质和组织能力较强。在铁路系统高标准、严要求的管理背景下,具备干部经历的毕业生常被认为具有更优秀的纪律性、沟通协调和团队协作能力,更加契合于铁路基层岗位对复合型人才的需求。

(3) 品行与纪律的硬性要求

无违纪和无挂科是铁路局录取的刚性门槛。研究结果表明:无违纪和无挂科频繁作为后项出现,且置信度在 0.88-0.91 区间,说明任何品行不端或学业瑕疵都可能会成为最终能否录用的潜在风险。

无纹身和无色盲色弱虽然没有在关联规则中直接体现,是因其属于铁路局招聘的硬性要求,503 组样本数据均符合两项条件,进一步验证了在安全意识和规则意识方面,铁路行业对从业人员有着极高的标准要求。

(4) 品行与学业的稳定性要求

铁路局在招聘阶段,不仅关注学生在校期间的表现,更关注学生未来的可持续性发展。当列入“是否挂科”这一特征时,置信度虽然有所波动,但依旧超过 0.8,说明铁路行业精益求精的职业特质,这种选拔模式,高度契合于铁路行业技术岗位的要求,确保铁路系统从业人员具备可持续发展的专业能力。

(5) 规则特征的内在逻辑

笔者将其概括为:以学业要求为基准(专业排名前 30%),以共青团员为政治底色,以干部经历和奖项荣誉为综合素质标准,以无违纪、无挂科为辅助。这种招聘要求不仅反映铁

路系统“德才兼备、以德为先”的录用标准,也为铁路类院校提供了精准方向,强化专业教学与实践锻炼,形成“学业—实践—品德—学业”的闭环培养模式。

5. 结论

笔者基于 aprior 算法,深度挖掘某二级学院 2024—2026 届 503 名铁路局录用毕业生特征数据,构建了铁路行业就业人才画像。研究结果表明:

(1) 以学业要求为基准(专业排名前 30%),以共青团员为政治底色,以干部经历和奖项荣誉为综合素质标准,以无违纪、无挂科、无纹身、无色盲色弱为辅助,这是铁路局选拔用人标准的核心指标。

(2) 高校需强化专业教育和实践锻炼联合培养,鼓励学生参加学生工作管理或勤工俭学锻炼,以提升自身综合素质。同时严格要求学业与纪律管理,培养附合各个行业的德能双优人才。

(3) 铁路系统在招聘时,也可进一步优化招聘体系,点考察学业排名、实践经历及长期稳定性,提升人岗匹配效率。

[参考文献]

[1]李海杰,苗蕾,聂磊,等.基于关联规则和主成分分析的高铁旅客购票行为特征研究[J].铁道科学与工程学报,2023,20(06):2013-2025.DOI:10.19713/j.cnki.43-1423/u.t20221575.

[2]兰洁,董挺锴,朱梦琳,等.基于 Aprior 算法的学生就业画像研究[J].现代信息科技,2024,8(03):107-110.DOI:10.19850/j.cnki.2096-4706.2024.03.023.

[3]王佳嘉.煤矿冲击地压事故致因耦合分析及风险评价研究[D].华北科技学院,2025.DOI:10.27861/d.cnki.gnckj.2025.000087.

[4]大数据时代数据挖掘与分析技术应用研究[EB/OL].(2025-05-21).http://www.renrendoc.com/paper/421936066.html.

[5]朱恒蓓.基于 Aprior 算法的校园网络安全分析模型研究[J].枣庄学院学报,2018,35(02):92-100.

[6]马蓉.Aprior 算法的改进及其在高职学生就业中的应用[J].价值工程,2013,32(30):182-185.DOI:10.14018/j.cnki.cn13-1085/n.2013.30.132.

[7]王彩强,赵宪中,刘涌,等.大数据环境下改进的 Aprior 算法研究[J].科技通报,2019,35(07):182-185.DOI:10.13774/j.cnki.kjtb.2019.07.036.

项目来源:陕西铁路工程职业技术学院;

课题名称:陕西铁路工程职业技术学院辅导员工作精品项目;

课题编号:2025fd-03。