

# 基于 K-means 就业数据分析与人才培养教学改革 实践路径探索

王霞

(海南软件职业技术学院,海南东方 571400)

**[摘要]**针对当前高职院校人才培养与市场需求对接不精准的问题,本研究以我校毕业生真实就业数据为基础,构建数据仓库,并引入 K-means 聚类算法,进行深度数据挖掘。通过设定多维特征向量(涵盖专业类别、就业岗位、薪资水平及地域分布等),运用 K-means 算法对毕业生群体进行自动分群,精准识别出“高匹配度就业群”“跨专业转型群”及“低质量就业群”等不同特征的簇。研究不仅揭示了专业设置与区域产业结构、岗位能力需求之间的内在关联规则与潜在矛盾,还分析了导致就业质量差异的关键因素。基于聚类结果提炼出规律,提出管理者修订人才培养方案、优化专业布局及重构课程体系的具体实践路径,旨在建立“数据挖掘—规律提炼—决策优化”的闭环机制,为高职院校实现精准育人和提升就业质量提供科学依据与决策支持。

**[关键词]** K-means 聚类;就业数据分析;教学改革

**[作者简介]**王霞(1982—),女,海南东方人,海南软件职业技术学院讲师,硕士,研究方向:数据挖掘和数据库应用。

**[基金项目]**本文系 2024 年海南省教育厅关于下达 2024 年度海南省高等学校教育教学改革研究的项目(项目编号:Hnjg2024-209)的阶段性研究成果。

**[DOI]** <https://doi.org/10.62662/kwxz0305026>

**[中图分类号]** G642

**[本刊网址]** [www.oacj.net](http://www.oacj.net)

**[投稿邮箱]** [jkw1966@163.com](mailto:jkw1966@163.com)

## 一、研究背景和意义

近年来,国家相继出台《国家职业教育改革实施方案》等一系列政策,明确要求高职院校深化产教融合、校企合作,推动人才培养与市场需求精准对接。一方面,学校积累了海量的毕业生就业数据,但这些数据都停留在简单的统计报表层面,缺乏深度的数据挖掘与价值提炼,未能有效转化为指导教学的“智慧资产”。在此背景下,数据挖掘技术,特别是无监督学习中的 K-means 聚类算法能够通过高维就业数据进行自动分群,揭示出隐藏在数据背后的就业规律与关联规则。通过构建就业数据驱动的教学改革框架,不仅能够识别不同专业群的就业特征与短板,还能为优化专业布局、重构课程体系及开展个性化就业指导提供量化支撑。

因此,本研究聚焦于如何运用 K-means 算法深度挖掘高职院校就业数据,探索数据背后的“专业—岗位—地域”关联模式,旨在打通“数据”与“教改”之间的壁垒,构建“数据驱动”新范式,使研究成

果为专业动态调整、人才培养方案修订及课程体系重构提供精准决策支持,有效破解人才培养与市场需求脱节的痛点,为提升高职院校人才培养的针对性与适应性提供科学的实践路径。

## 二、研究内容与方法

研究聚焦于我校毕业生就业数据与人才培养质量的关联性分析。研究内容包括:构建就业数据仓库,选取专业、岗位、薪资、地域等多维指标,运用 K-means 聚类算法对毕业生进行群体特征划分,挖掘专业设置与市场需求的内在规律,并据此提出教学改革路径。研究方法主要利用数据挖掘技术(K-means 聚类)进行实证分析,对比分析验证不同专业群的就业差异,最终通过逻辑推演提出教改策略,实现“数据驱动决策”。

## 三、就业数据驱动高校教改的系统框架构建

本项目以 SQL Server Data Tools 为平台,构建“数据层—技术层—分析层—应用层”四层系统框架,各层相互关联、层层递进,形成闭环赋能体系,

实现就业数据从采集处理到教改落地的全流程转化。

### (一)数据层:数据构建与预处理

#### 1. 数据仓库

数据仓库是一个集成的、面向主题的数据集合,主要用于实现决策支持功能,其中每个数据单元都和某个时刻相关。

#### (1)概念模型

概念模型是对客观事实的主观表达的结果,是主观意识的工具。本项目的数据库需要能从专业、籍贯、岗位、区域和单位多方面进行分析。数据仓库概念模型如下:

数据需要包括:单位、专业、岗位、工作地区等数据。然后再进行描述,有如下内容:键码、联系、属性。维度要有:学生情况、单位情况、专业情况、城市情况。定义维度:学生情况有姓名、性别、籍贯、联系方式、学号、学院代码;专业情况设定专业代码、名称、学院代码;岗位设定岗位类别、岗位名称;学院设定学院代码、学院名称。城市情况设定城市名称、区域代码、区域名称;单位情况设定名称、性质、薪水。

#### (2)逻辑模型

逻辑模型是概念模型的细分,还要能在物理实施上有实际的指导意义,是把不同目标和维映射到数据仓库中的事实表。星型模型是典型的结构之一,包括事实表、维表和主键,维表围绕事实表显示到线上。星型模型包含三个实体:维度、指标和类别。工作主要有主题分析、粒度划分、数据表以及数据提炼。

#### (3)物理模型

物理模型遵循数据库设计方法。星型模型中事实表与维表会转变成对应的数据表,事实表有关注的信息,数据量大;维表数据量小,信息要稳定些。物理模型如表1、表2、表3、表4、表5所示:

表1 就业情况表

名称	类型	键	说明
XH	nvarchar(12)	primary key	学生学号
XM	nvarchar(30)		学生姓名
ZYDM	nvarchar(4)	Foreign key	专业代码
DWMC	nvarchar(50)	Foreign key	学生工作单位名称
GWLB	nvarchar(30)	Foreign key	学生从事工作的岗位
CSMC	nvarchar(30)	Foreign key	学生工作所在的城市

表2 学生维表

名称	类型	键	说明
XH	nvarchar(12)	primary key	学生学号
XM	nvarchar(30)		学生姓名
XB	nvarchar(2)		性别
JG	nvarchar(20)		籍贯
PHONE	nvarchar(11)		电话号码
XYMC	nvarchar(30)		学院名称

表3 单位维表

名称	类型	键	说明
DWMC	nvarchar(50)	primary key	工作公司或者单位名称
DWXZ	nvarchar(20)		工作公司或者单位的性质

表4 城市维表

名称	类型	键	说明
CSMC	nvarchar(20)	primary key	工作城市的名称
QYDM	nvarchar(4)		工作城市在哪个区域
QYMC	nvarchar(30)		工作城市区域名称

表5 岗位维表

名称	类型	键	说明
GWLB	nvarchar(20)	primary key	工作岗位属于哪类
GWMC	nvarchar(30)		工作岗位名称
Salary	nvarchar(10)		工作得到的报酬

#### 2. 就业数据预处理

原始就业数据收集时会存在缺失情况、格式不一致情况或噪声情况,需通过数据预处理提升数据质量,为后续挖掘分析奠定基础,具体步骤包括:

**数据清理:**剔除异常数据、补充空值数据等填充缺失值,避免影响挖掘结果准确性;

**数据集成:**将不同来源、不同格式的数据整合为统一格式,消除数据冗余与冲突;

**数据变换:**对数据进行标准化、将非结构化数据转化为结构化数据,适配数据挖掘算法需求

**数据规约:**简化数据维度,保留核心特征,提升算法运行效率。

#### 3. 就业数据体系构建

结合我校就业数据,构建多维度就业数据体

系,确保数据的全面性与实用性,具体包括三大类数据:毕业生基础数据:涵盖学号、姓名、籍贯、专业、学院等;就业核心数据:包括工作单位类别、就业岗位名称、就业薪资、就业区域等,反映就业质量与分布特征;市场需求数据:涵盖行业发展趋势、岗位要求标准、区域产业布局、企业招聘偏好等,反映市场对人才的需求动态。数据来源包括我校的就业管理系统、毕业生跟踪调查、企业招聘平台、地方产业发展报告等,通过多源数据整合,形成完整的就业数据仓库,如图1所示。

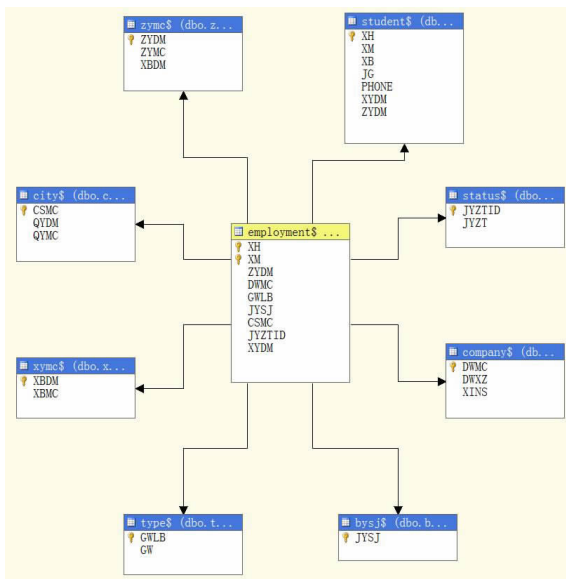


图1 就业数据仓库多维数据集的存储架构图

(二)技术与分析层:数据挖掘算法筛选与适配应用

筛选适配高校就业数据分析的核心算法,明确各算法的应用场景与优势,构建多算法融合的挖掘技术体系。

1. 聚类算法基本描述

聚类分析是依据样本间关联的度量标准将样本自动分成几个群组,使同一群组内的样本相似,不同群组样本相异。聚类分析的输出是数据集的几个组(聚类),附加结果是对每个类的概括描述,这个结果对深入分析数据集特征尤为重要。

2. K-means(K-平均值)算法

K-means(K-平均值)算法是一种基于形心的技术。先是指定所需寻找的聚类个数,这是参数K,而后随机选出K个点作为聚类的中心,根据普通的欧几里得距离变量,将所有的实例分配到各自最靠近的聚类中心,下一步是计算出实例所在的每个聚类的质心,或者均值,这些质心将成为各个聚类的新的中心值。最后用新的聚类中心重复整个过程。迭代过程不断继续直到在连续的几轮里,每个聚类上分到的点与上一轮分到的点相同,这时聚类的中心才算固定,并且会永远保持。本项目是从三个方向进行挖掘,用“城市”“岗位”“专业”为输入项,挖掘就业情况、专业需求、岗位方向,如图2、图3所示。

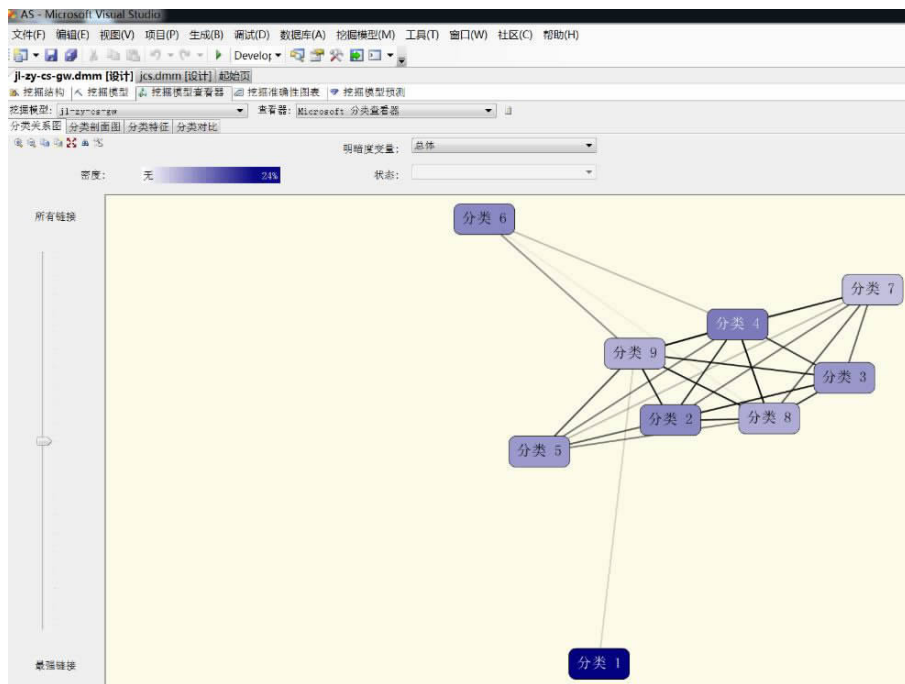


图2 聚类挖掘分类关系图

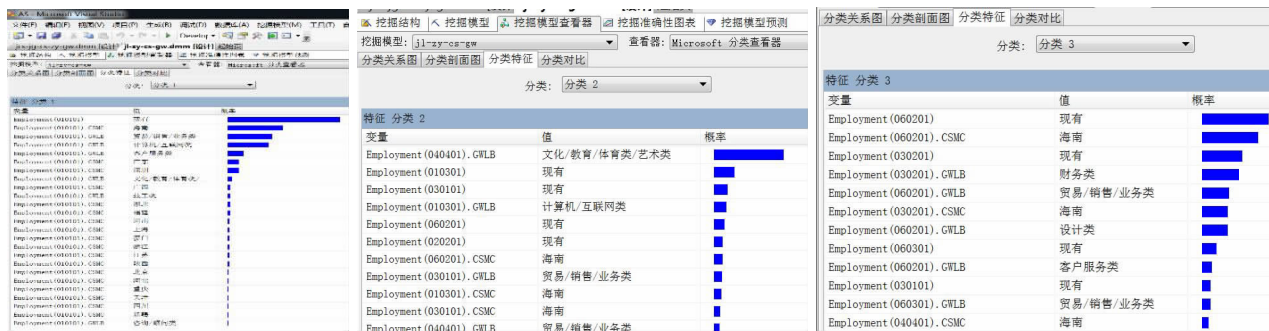


图3 聚类挖掘分类特征图

分类1 分类特征:典型群体画像为以海南本地就业为核心去向,主要从事贸易/销售/业务类、计算机/互联网类岗位,少数流向珠三角(广州/深圳)、从事客户服务类工作的已就业毕业生群体。

分类2 核心特征:音乐教育(040401)专业就业毕业生是本聚类的核心就业群体,在海南本地就业率高,核心岗位:文化/教育/体育类/艺术类。其次是会计与人工智能专业。

分类3 核心特征:汽车制造与试验技术(060201)、电子商务(030201)专业以海南本地就业为绝对核心去向,两个核心专业的海南就业概率均处于高位,但是汽车制造与试验技术专业以贸易/销售/业务类、设计类、客户服务类(多岗位并行,销售为核心)岗位为主。

由结果可以得出:聚类分析将岗位分成具有共同特征的几类,在共性中发现有价值的信息。比如:绝大多数专业的毕业生首选在海南工作;音乐教育、学前教育的专业设置相似度较高,自动聚合为同一类;涉及教育性质的专业大多数从事教育类的工作,专业与工作岗位的匹配度较高,而电子技术与就业岗位的相关性较低,从事销售/业务类岗位的较多;软件技术专业分别在分类1、5、6、7、8中都有影子且出现的概率较高,工作岗位涉及销售/业务类、计算机/互联网类,可以认为软件技术专业课程设置的覆盖面广。

#### 四、实证分析——以海南软件职业技术学院就业数据为例

##### (一) 实证对象与数据来源

选取我校2019—2023届毕业生就业数据为研究样本,涵盖6个二级学院、30个专业,共收集有效数据4898条。数据来源于我校就业管理系统、毕业生跟踪调查与企业反馈问卷,经预处理后,保留核心数据维度8个,形成实证分析数据集。

##### (二) 数据挖掘过程与结果

##### 1. 分类1 特征为海南地域就业的集中性特征,

特征为“生源地—就业地—专业—岗位”高度集中,聚类样本量占比高,主要由海南籍生源构成,就业区域集中在海南本地。专业以财经类、文教体艺类为主,其中会计、财务管理为核心专业,就业岗位集中在海南本地,而且专业与岗位对口率高。

2. 分类2 特征为专业与就业岗位存在显著错位,聚类样本主要包含两类专业群体:一是软件技术专业毕业生,二是学前教育专业毕业生。两类专业毕业生的就业岗位均大量流向贸易销售类岗位,与自身专业培养方向偏差较大,其中软件技术专业作为技术类专业,未进入计算机/互联网相关岗位,学前教育专业作为教育类专业,未进入教育相关岗位,形成了明显的“专业与岗位错配”聚类特征。

3. 分类3 特征为技术类专业向沿海发达地区集聚,聚类样本主要为软件技术、信息安全技术应用等技术类专业毕业生,就业区域集中在广东、浙江等沿海地区。其中,浙江地区的软件技术专业毕业生专业对口率较高(90.9%),主要进入计算机/互联网类岗位;广东地区的信息安全技术应用专业毕业生,75%进入互联网行业,形成了“技术专业—沿海地域—互联网岗位”的集聚聚类。

4. 分类4 特征为“地域产业—专业—岗位”高度适配,包含三个细分聚类分支,分别对应不同地域的特色产业:一是广东/深圳分支,聚类样本以各类专业毕业生为主,就业岗位集中在贸易销售、互联网类,体现广东/深圳贸易与互联网产业的集聚效应;二是浙江分支,聚类样本以计算机相关专业为主,就业岗位集中在计算机/互联网类,凸显浙江互联网产业的主导地位;三是北京分支,聚类样本以广告设计相关专业为主,就业岗位集中在广告设计与互联网行业。

##### (三) 实证分析

基于以上5个聚类集群的特征,结合就业实际场景,开展专业与就业匹配、岗位需求与技能缺口、区域就业与产业对接三大维度的实证分析,结论

如下:

### 1. 专业与就业匹配度分析

**高匹配群体:**海南财经类专业、特色产业适配分支、设计类的专业与岗位匹配度较高,其中室内设计专业、会计专业、财务管理专业、广告设计相关专业对口率突出。

**低匹配群体:**“专业—岗位”错配集群、部分技术专业明显匹配不足,其中软件技术跨行业就业比例高,仅浙江地区对口率较高;大数据毕业生未进入互联网核心岗位集群,反而分散在其他岗位,体现出专业与岗位的弱连接,专业培养方案未对接行业需求、技术专业技能可迁移性强、海南本地互联网岗位供给不足,导致专业与就业匹配度偏低。

### 2. 岗位需求与技能缺口分析

**高需求岗位:**聚类5中的贸易/销售/业务类岗位,跨专业吸纳能力强,覆盖10多个专业,形成高需求岗位集群;聚类3、聚类4中的计算机/互联网类岗位,集中吸纳计算机相关专业毕业生,地域分布广泛,需求旺盛。

**技能缺口识别:**结合聚类特征,聚类5中贸易销售岗位对“具备专业背景的销售人才”需求旺盛(跨专业吸纳但需专业基础);聚类4中互联网、广告设计等岗位,对“既懂技术又懂业务的复合型人才”需求稀缺;特定行业(如汽车电子)相关专业未形成独立聚类,说明该领域专业人才培养存在偏差,人才供给与岗位需求脱节;大数据专业未进入核心互联网岗位聚类,反映出该专业毕业生未掌握数据分析核心技能,无法满足岗位需求。

### 3. 区域就业与产业对接分析

**对接良好的产业:**海南本地集群体现海南本地产业(文教、贸易)与财经类、文教体艺类专业的深度对接,形成本地化就业内向循环;特色产业适配集群中,浙江互联网产业与软件技术专业、广东贸易产业与各类专业销售人才、北京互联网+广告行业与广告设计专业,均实现了产业与专业、岗位的精准对接,形成良性互动。

**对接不足的聚类与产业:**聚类3中技术人才向沿海流动,反映出海南本地产业无法完全吸纳技术人才,导致技术专业毕业生外流;大数据专业未形成与数据分析岗位对应的聚类,体现出大数据专业与数据分析岗位的对接不足,专业培养与产业需求存在脱节,未形成有效的“专业—产业—岗位”关联。

## 五、就业数据驱动高校教改的优化策略

### (一) 调控与疏导错位明显专业

对于“软件技术”流向销售的现象,优化该专业

课程,增加产品经理、技术销售等“技术+业务”融合课程模块,主动培养复合型人才。同时对于志在核心技术的学生,加强与浙江、广东等地互联网企业的实习实训合作,疏通技术就业渠道。对于“学前教育”进入销售岗:评估是意愿选择还是就业挤压。若是后者,需严格控制招生规模或转型为“儿童发展与教育”(拓宽至早教、教培、儿童产品策划等),增加沟通、策划等通用能力课程。针对“复合型销售”“技术+业务融合人才”缺口:在传统专业基础上,普遍开设“辅修专业”,如“工程技术+市场营销”“数据分析+金融业务”,打破专业壁垒。针对“汽车电子”等专精人才偏差,建立“产业教授”制度,邀请企业专家深度参与人才培养方案制定与核心课程教学,确保教学内容与行业技术发展同步。

(二) 重构人才培养标准,从“知识传授”到“能力集成”

制定“核心+模块”式能力标准,各专业均强化数据分析基础、设计思维、沟通协作等跨领域通用能力。根据专业对应的主流岗位簇(如技术开发、技术销售、数据分析、创意设计等),定义清晰的、可衡量的能力模块,学生可根据职业规划组合选修。推行“认证嵌入式”培养,将行业权威认证(如大数据分析、设计软件认证等)的知识体系与技能要求融入课程,鼓励学生获取证书,提升岗位适配度。

### (三) 课程体系改革

分析专业就业流向,及时调整课程,引入行业认证课程,针对高需求岗位(如互联网、销售)引入行业认证。为技术专业增设商务沟通、销售技巧课程,为财经专业增设数据分析课程。优化实践教学,区域化实习基地建设,海南重点建设文教、贸易类实习基地;广东/浙江建设互联网、跨境电商实习基地。项目制学习:围绕区域产业需求设计跨专业实践项目。与龙头企业合作开发“行业定制课程包”,如“汽车电子软硬件协同设计”“跨境电商供应链金融”等。更新现有课程内容,融入新兴技术与行业标准。例如,在计算机专业课程中增设人工智能、大数据分析内容补齐技能缺口。将核心技能证书考核融入课程体系,实现课程学习与证书获取、岗位需求的联动。

### (四) 建立数据反馈与持续改进闭环

每年收集毕业生就业公司反馈、产业趋势数据,作为专业设置、培养方案的核心依据。与高质量就业的毕业生及其公司建立长期合作关系。鼓励教师赴企业实践、开发产教融合课程。通过系统性的改革,能够将数据分析发现的“集中特征、错位

现象、产业特色、技能缺口”等关键信息,有效转化为“专业优化、标准重构、课程更新、服务精准”的具体行动,最终实现人才培养与市场需求的高度适配,提升毕业生的就业质量与发展潜力。

#### (五)优化就业指导体系

从一开始向学生展示各专业真实的就业地图(地域、岗位、薪资、发展路径),引导学生构建“学生能力—岗位需求”画像匹配系统,通过记录学生的课程项目、技能证书、实习经历等,生成个人能力画像,与企业岗位画像进行智能匹配与缺口提示,提供个性化提升建议。搭建精准就业信息平台,同时,实时跟踪行业发展与岗位需求变化,及时更新就业指导内容,帮助学生把握就业趋势。

#### (六)培养兼具教育教学经验与数据挖掘能力的复合型人才

定期开展培训,提升就业工作人员与教师的数据应用能力。同时与企业、科研机构合作,借助外部技术力量,提升数据挖掘与教改应用的专业性。深化校企合作,建立校企协同育人机制,邀请企业参与专业设置、课程体系构建与人才培养标准制定,确保人才培养符合企业需求。同时,通过企业提供的岗位需求数据、技术标准,优化数据挖掘维度与教改方向,实现高校人才培养与企业需求的同频共振。

## 六、结论

本项目研究构建了数据层、技术层、分析层、应用层的就业数据驱动高校教改系统框架,研究表明数据挖掘中聚类分析算法可有效挖掘就业数据中专业与就业、岗位、区域的内在关联,识别影响就业的核心因素;基于挖掘结果提出的就业指导优化、专业定位调整、人才培养重构、课程体系完善,可有效打通就业数据与教改实践的壁垒,实现人才培养与市场需求的精准对接。

## 参考文献:

- [1]宋丽萍,韦建国.关联规则挖掘在学生就业数据处理中的应用[J].湖南理工学院学报(自然科学版),2016,29(4):43-48.
- [2]王北阳.基于数据挖掘技术的就业数据实证分析——以宿州学院为例[J].赤峰学院学报(自然科学版),2015,31(3):273-275.
- [3]邱瑾.决策树在高职学生就业数据分析中的应用研究[J].柳州职业技术学院学报,2012,12(2):54-56,64.
- [4]李俊,李延平,赵玲令,高梦薇.基于数据挖掘的精准就业数据研究[J].信息技术与信息化,2020(11):37-39.
- [5]周燕,李凤,黄蔚.基于毕业生就业数据挖掘的高校教学管理探究[J].西部素质教育,2019,5(7):101-102.
- [6]王霞.数据仓库与数据挖掘技术在学生就业信息分析中的应用研究[D].广州:中山大学,2013.

## Exploration on Practical Paths of Teaching Reform for Talent Cultivation Based on K-means Analysis of Employment Data

WANG Xia

(Hainan Software Vocational College, Dongfang Hainan 571400, China)

**Abstract:** Aiming at the inadequate alignment between talent cultivation in higher vocational colleges and market demands, this study establishes a data warehouse based on real employment data of graduates from our university and adopts the K-means clustering algorithm for in-depth data mining. By constructing multi-dimensional feature vectors covering major categories, employment positions, salary levels and regional distribution, the K-means algorithm is applied to automatically group graduate cohorts and accurately identify clusters with distinct characteristics, including the high employment matching group, cross-major transition group and low-quality employment group. This research not only reveals the inherent correlation rules and potential contradictions between major setup, regional industrial structure and job competency requirements, but also analyzes the key factors leading to differences in employment quality. The laws summarized from the clustering results provide practical approaches for administrators to revise talent training programs, optimize major distribution and restructure curriculum systems. This study intends to build a closed-loop mechanism of “data mining—rule summarization—decision optimization”, so as to offer scientific reference and decision support for higher vocational colleges to conduct targeted talent training and improve employment quality.

**Key words:** K-means clustering; employment data analysis; teaching reform