

基于类平均聚类法的普通高等专科学校办学规模区域差异研究*

陈江丽

(滇西科技师范学院 信息科学与技术系,云南 临沧 677000)

【摘要】在市场经济的背景下,全国各地的普通高等专科学校呈现出区域发展的趋势。在学校招生规模、培养规模和毕业生规模等指标的基础上,采用类平均聚类法对各地区普通高等专科学校的办学规模进行分析,利用Matlab平台进行仿真,生成不同区类的划分,有利于进一步研究不同区域间学校发展的战略和因地制宜的人才培养方向。

【关键词】类平均聚类法;高等专科学校;办学规模;系统聚类树;不一致系数

【中图分类号】TP301.6;G648.2 **【文献标志码】**A **【文章编号】**1673-1891(2015)03-0046-03

DOI:10.16104/j.cnki.xccxb.2015.03.014

《中国教育大辞典》中指出:高等专科学校的主要宗旨是实施全日制高等职业技术教育,培养专门技术人员,注重实际知识和理论的掌握及实际技能的训练,以培养适应经济和社会发展的技能型人才为目标^[1]。因此,普通高等专科学校的毕业生主要面向和服务社会的技术和应用型岗位,辐射较广泛,社会需求较大,在经济发展和社会进步中起到巨大的促进作用。

但由于经济和文化条件不同,各地普通高等专科学校办学规模的差异较大,区域差异在总体上呈扩大趋势。区域教育发展不协调,带来了许多经济问题和社会问题。它不仅严重制约了我国经济的可持续发展和社会的全面进步,而且造成了教育机会和教育过程的不公平,对社会和谐、政治稳定构成了潜在威胁,同时也进一步加剧了技能型人才的流失。

因此,准确分析和全面把握高等专科学校办学规模区域发展的差异性,有利于国家立足于本地区的经济现状、社会需求和自身实际情况,准确定位学校的人才培养目标、服务面向和办学特色,更好地推动区域经济和社会的发展。

1 类平均聚类法

类平均聚类法是将类与类间的距离定义为样品间平方距离的平均值,将距离最近的两类合并为一类,最终实现聚类的目的^[2]。假设存在类 G_k 、 G_l , 则 G_k 和 G_l 之间的平方距离的计算如公式(1)所示^[3]。

$$D_{kl}^2 = \frac{1}{n_k n_l} \sum_{x_i \in G_k, x_j \in G_l} d_{ij}^2 \quad (1)$$

若某一步类 G_k 和 G_l 合并成新类 G_M 时, G_M 与另一类 G_j 间的平方距离可以推广为公式(2)。

$$D_{Mj}^2 = \frac{n_k}{n_M} D_{kj}^2 + \frac{n_l}{n_M} D_{lj}^2 \quad (2)$$

从公式(1)、(2)中可以看出,类平均法充分地利用了所有样品之间的信息,同一类中的平均距离小,表示样本间的相似度高;而不同类间的平均距离大,表示样本间的相似度低。因此,类平均法是一种较合理有效的系统聚类法。

2 利用类平均聚类法分析普通高等专科学校办学规模的区域差异

普通高等专科学校办学规模的评价主要参照学校招生规模、培养规模和毕业生规模等指标,如表1列出2013年全国31个省、市、自治区和直辖市高等专科学校的各项数据。利用类平均聚类法进行聚类分析,将地区按办学规模划分为不同的类,根据同一类地区普通高等专科学校办学规模的相似性,不同类地区间的差异性,为区域间普通高等专科学校办学规模的差异性研究提供重要依据,对全面提高普通高等专科学校办学能力、促进区域协调发展提供科学的指导。

表1 2013年全国各地区普通高等专科学校办学规模指标数据

地区	办学规模指标				办学规模指标				
	招生数	在校生数	毕业生数	预计毕业生数	招生数	在校生数	毕业生数	预计毕业生数	
北京	34531	107128	36780	35525	湖北	183634	563945	181294	195890
天津	56705	166749	53836	51683	湖南	149714	448981	152325	147211
河北	165135	528989	188077	193390	广东	265182	760296	211824	236808
山西	96907	286163	96884	89594	广西	110139	325393	104460	106287
内蒙古	53992	169161	57997	59730	海南	23641	69150	23818	21680
辽宁	96235	292215	94362	96707	重庆	79746	239271	66194	74978
吉林	50543	145169	42953	47215	四川	174962	535377	153501	168067
黑龙江	71990	207962	67541	66671	贵州	52930	161831	44969	47869
上海	46971	142029	49158	49179	云南	63735	189663	62982	68357

收稿日期:2015-03-25

*基金项目:云南省教育厅科学研究基金资助(项目编号:2013C037);临沧师范高等专科学校自然科学、基础应用研究基金资助项目(项目编号:LCSZL2013009)。

作者简介:陈江丽(1984-),女,云南大理人,讲师,硕士,研究方向:数据仓库和数据挖掘。

江苏	195005	683635	237480	240621	西藏	3584	12174	4321	4184
浙江	121414	372219	116674	122743	陕西	126086	390258	117761	127705
安徽	147545	469034	158141	167906	甘肃	50666	162892	48344	54761
福建	91927	273269	92780	90668	青海	5957	17367	5634	5817
江西	124731	380638	139581	130826	宁夏	12476	37082	9647	11357
山东	254094	763065	264197	253465	新疆	43058	127104	37963	35717
河南	230764	708164	266016	237413					

数据来源:中华人民共和国国家统计局网站, 2014年《中国统计年鉴》

2.1 类平均聚类分析各地区普通高等专科学校办学规模的Matlab实现

利用 MATLAB 对各地区普通高等专科学校的办学规模进行聚类分析。需要先读取表 1 中的招生数、在校生数、毕业生数和预计毕业生数 4 个变量数据放入矩阵 X 中, X 就是一个 31 × 4 的矩阵。再读取表 1 中的地区名称放入矩阵 area 中, area 是一个包含 31 个元素的列向量。完成数据的准备工作后, 接着按以下四个步骤进行分析处理。

2.1.1 样本的预处理

为了保证分析结果的准确性, 需要对样本数据进行一些预处理, 如平滑处理、标准化变换和极差归一化变换等。在实际应用中, 可根据数据特征选择相应的处理。

针对学校办学规模指标数据各变量的量纲和数量级不一致的情况, 采用 zscore 函数对原始指标数据进行标准化处理, 如式(3)所示。

$$X = zscore(X); \quad (3)$$

2.1.2 计算样本间距离

聚类开始时, 先将 31 个地区样本各自作为一类, 接着计算各个样本间的距离, 为接下来对距离最近的两类进行合并提供依据。

样本间距离的计算方法包括明氏距离、兰氏距离、马哈拉诺比斯距离和斜交空间距离等。其中最常用的是明氏距离中的欧式距离。第 i 个样本 和第 j 个样本 之间的欧式距离计算如公式(4)所示^[4]。

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (4)$$

MATLAB 中使用 pdist 函数计算样本间的欧式距离, 对于各地区的指标数据 X, 调用函数格式为 y = pdist(X), 输出 y 是 31 个样本间的距离向量, 分别对应第 i 个(其中 i=2,3,...,31)和第 j 个(其中 j=1,2,...,30)样本间的距离。

2.1.3 利用类平均聚类法创建聚类树

创建聚类树是聚类分析中的关键环节, 可以完整地反映聚类的过程。MATLAB 工具箱中提供了

linkage 函数创建系统聚类树, 对 31 个地区指标样本的距离向量 y, 调用函数格式为 Z=linkage(y, 'average')。其中 'average' 表示使用类平均法进行聚类。输出 Z 是创建的各地区专科学校办学规模聚类树, 包含 31 个元素的矩阵, 它的每一行对应一次聚类。

为了清晰直观地反映创建的聚类树, 可利用 MATLAB 工具箱中的 dendrogram 函数生成聚类树形图, 对聚类树 Z, 地区名称 area, 调用函数格式为 H = dendrogram(Z,0,'orientation','right','labels',area), 生成的聚类树形图如图 1 所示。图中的倒 U 形线用来连接聚类对象, 线的高度表示聚类距离, 距离的远近反映聚类的先后顺序。聚类树形图的方向从右至左, 地区名作为叶节点标签, 显示在图的左侧。通过聚类树形图, 可以完整和清晰展现聚类的过程。

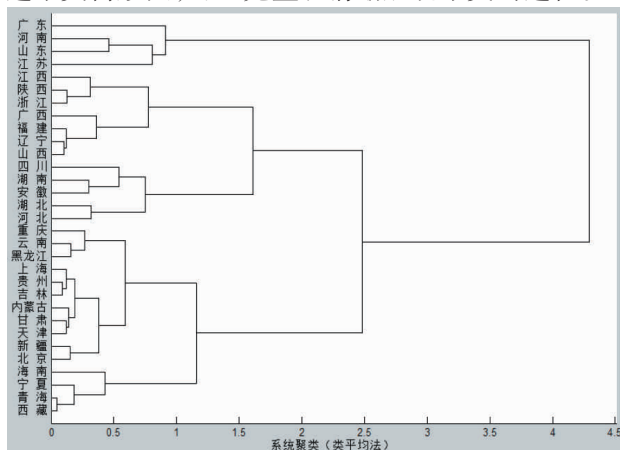


图1 聚类树形图

2.1.4 计算聚类树的不一致系数

系统聚类树的不一致系数用来确定最终的分类型数, 可使用 MATLAB 工具箱中的 inconsistent 函数进行计算。对上一步创建的聚类树 Z, 调用函数格式为 T=inconsistent(Z), 输出参数 T 是一个包含 30 × 4 个元素的矩阵, 其中的第 4 列为每一次聚类的不一致系数, T(:,4) = [0, 0, 0, 0, 0.7071, 0, 0.7071, 0, 0.7071, 0, 0, 0.7071, 1.0928, 0.7071, 0, 0.7071, 0, 0.7071, 1.1418, 0.7071, 0, 0.7071, 1.0843, 0.9893, 1.1488, 0.7071, 0.7071, 1.1293, 1.1544, 1.0867, 1.0233]'。

若某一次聚类的不一致系数较上一次有大幅增加, 则说明这次聚类的效果并不好, 而上一次的聚类效果比较好。通过观察和比较不一致系数 T(:, 4)。考虑倒数第 4、5 次聚类的不一致系数的变化, 不一致系数大幅增加 0.4222, 说明倒数第 5 次的聚类效果是比较好的, 对照图 1 的系统聚类树形图可看出, 此时样本被划分为 5 类。因此, 可明确将各地区专科学校的招生规模指标数据聚为 5 类是最合适的。

2.1.5 生成聚类结果

最终根据创建好的系统聚类树和确定的聚类个数生成聚类结果并输出,以完成整个聚类过程。

聚类结果的生成可利用MATLAB工具箱中的cluster函数实现,由创建的聚类树Z和聚类个数5,调用函数格式为C=cluster(Z,'maxclust',5),输出参数C为每一个样本的所属的类序号。每个聚类所包含的地区如下所示:

(1)第1类地区:海南、西藏、青海、宁夏;

(2)第2类地区:北京、天津、内蒙古、吉林、黑龙江、上海、重庆、贵州、云南、甘肃、新疆;

(3)第3类地区:河北、安徽、湖北、湖南、四川;

(4)第4类地区:山西、辽宁、浙江、福建、江西、广西、陕西;

(5)第5类地区:江苏、山东、河南、广东。

2.2 分析结果

利用类平均聚类法对全国31个地区的普通高等专科学校办学规模进行聚类分析,最终划分为5类地区。通过分析每类地区间的差异,可以发现普通高等专科学校的办学规模与地理位置、经济发展水平、人口数量、教育发展水平等客观因素有关。

第1类地区的招生规模最小。首先四个省份或自治区的地理位置属偏远地区,人口较少,教育水平相对较差。尤其西藏、青海和宁夏地区的经济发展水平较落后。这些因素都客观地影响着普通高等专科学校的办学规模。

第2类地区的招生规模处于较低水平。其中,

北京、上海、天津作为国际化的大都市,是中国政治、文化、科教的中心,对学历层次的要求较高,因此在一定程度上也制约了专科学校的发展。重庆、贵州、云南、甘肃、新疆、内蒙古等省份作为西部地区城市,吉林和黑龙江位于祖国的东北,地理位置和经济发展水平等因素都制约了普通高等专科学校的办学规模的发展。

第3类地区的招生规模处于中等水平。河北、安徽、湖北、湖南、四川均处于中部地区,地理条件较好,教育发展水平在全国处于前列,这些客观因素都促进了普通高等专科学校办学规模的扩大。

第4类地区的招生规模处于中上水平。这些地区大都处于东部或西部工业较发达的城市,对技能型和应用型人才的需求量较大,在很大程度上促进了专科学校毕业生的就业。因此,在一定程度上推动了普通高等专科学校办学规模的发展。

第5类地区的招生规模最大。其中的江苏和广东都是全国工业发展水平最高,经济实力最强的城市,对专科型应用技能人才的需求量最大。河南和山东分别是中国人口的第一和第二大省,人力资源丰富,对专科学校的办学需求量较大。

3 结束语

利用类平均聚类法对全国各地普通高等专科学校的办学规模进行聚类分析,划分为不同的区类,进而研究区域间差异的因素,有利于国家立足于本地区的实际,促进专科学校区域办学的协调发展,更好地推动经济和社会的可持续发展。

注释及参考文献:

- [1]李晓娟.东西部高等教育协调发展研究[D].石河子:石河子大学,2010.
- [2]Jiawei Han, Micheeline Kamber著,范明,孟小峰等译.数据挖掘概念与技术[M].北京:机械工业出版社,2010.
- [3]于秀林,任雪松.多元统计分析[M].北京:中国统计出版社,2008.
- [4]郑红英.数据挖掘聚类算法的分析和应用研究[D].沙坪坝:重庆大学,2002.
- [5]赵光龙.中国高等教育资源配置空间分布研究[D].上海:华东师范大学,2014.
- [6]袁静.区域高等教育与经济的协调发展研究[D].苏州:苏州大学,2008.

Study on the Regional Differences between Higher Specialty Institutions Based on the Class Average Clustering Method

CHEN Jiang-li

(Department of Information Science and Technology, Dianxi Science and Technology Normal University, Lincang, Yunan 677000)

Abstract: The development of higher specialty institutions in our country shows regional tendency under the market economy background. On the basis of the scale of enrollment, training and graduates and other indicators, we

(下转第72页)

注释及参考文献:

- [1]吴铨叙.军事训练学[M].北京:军事科学出版社,2003.
- [2]查有梁.教育模式[M].北京:教育科学出版社,1997.
- [3]朱如珂.现代军校教学论[M].北京:军事科学出版社,1997.
- [4]吴庆麟.教育心理学[M].上海:华东师范大学出版社,2003.
- [5]俞文钊.管理心理学[M].上海:东方出版社,2002.

Some Thoughts on the Training of the Students in the Police College

YI Jian-qu

(Basic Part, Public Security Border Forces College, Guangzhou, Guangdong 510663)

Abstract: The police college student is an important source of cadres, they infused the new blood for the troops construction.. Team cadres as the first "teacher" of the police college students' career, directly affect their future work style, behavior habits .So student team cadres in the daily management of the college students, military training, learning, life and other aspects should be refined management. This article carries on the discussion from the perspective of people oriented, combined with the characteristics of the group, how to do a good job in the training of college students' training work

Key words: student team cadres; college students; training; two changes

(上接第 48 页)

use the class average clustering method to analyze school running scale of higher specialty institutions in every region, use MATLAB to simulation, and then we get different class divisions. The research is conducive to the further study of the school development strategy and local talent training direction between different regions.

Key words: class average clustering method; higher specialty institutions; school running scale; system clustering tree; inconsistent coefficient

(上接第 59 页)

the enterprise, which is an important means of supervision and restraint to leading cadres. Economic responsibility audit can implement effectively prevent corruption on the mechanism and the source, which has become a necessary link in the development of state-owned enterprise leading cadres career. In the current situation, based on the governance in accordance with the enterprise, risk prevention and control and the requirement of advancing the fine management of grid enterprises, it has significant practical value and significance to comprehensively improve the term economic responsibility audit quality, which can conducive to enhance the level of power supply enterprise management and guard against operational risks.

Key words: term economic responsibility; power supply enterprise; audit