

基于 SPSS 的学生能力倾向聚类分析研究

罗家国 罗浩 仲佳嘉

【摘要】 本文主要介绍了一种基于 SPSS 聚类分析功能的学生就业潜能评价方法,以求从教育统计学的视角探索毕业生就业咨询的新途径,为教学管理改革提供参考,进而达到改进大学生职业生涯教育、寻求增强毕业生成长后劲的最佳择业取向之目的。实验结果表明此方法快速、直观、全面且针对性较强。

【关键词】 就业咨询 SPSS 聚类分析

【收稿日期】 2012 年 1 月

【作者简介】 罗家国,江西理工大学教务处教授;罗浩,江西理工大学应用科学学院讲师;仲佳嘉,江西赣州供电公司工程师。

择业是人生的重要转折点,它在相当的程度上决定着人一生所从事的职业及今后事业的发展,因此择业是一项非常重要和谨慎的工作。一个人在工作上能否有所作为,最主要取决于两点:首先是兴趣、爱好和性格特点;其次是其能力倾向,不同人的能力倾向是有差异的^[1]。

怎样根据学生的学业成绩指导学生就业?笔者通过多年的研究发现,通过聚类分析对学生潜能进行评价,是提高学生就业能力和就业质量的一种定量分析方法。所谓“聚类分析”,它是一种教育统计分析方法,是数据挖掘技术的重要组成部分,它能够在不同的潜在数据中发现数据分布模式,从而找出修正这一模式的方法^[2]。换言之,我们可以通过采集学生的学习成绩数据,对学生的能力倾向进行“聚类分析”,从而有的放矢地对学生进行就业指导。具体操作是:首先,对学生的学业成绩进行“聚类分析”,把学生按某种能力属性分成若干小组(类);再根据各门课程的特点分析每类学生的能力倾向;最后,根据每类学生的能力倾向和不同职业特点进行科学的就业指导。

一、数学模型及 SPSS 简介

本课题研究所采用的聚类分析数学模型:

欧氏距离平方(Squared Euclidean Distance)

$$SEUCLID = \sum_{i=1}^k (x_i - y_i)^2 \quad [2][3]$$

其中,k 表示每个样本有 k 个变量; x_i 表示第一个样本在第 i 个变量上的取值; y_i 表示第二个样本在第 i 个变量上的取值。

SPSS 的全称是 Statistical Program for So-

cial Sciences,即社会科学统计程序。SPSS 的基本功能包括数据管理、统计分析、图表分析等,几乎无所不包。它名为社会科学统计软件包,这是为了强调其社会科学应用的一面,而实际上它在社会科学、自然科学的各个领域都能发挥巨大的作用,并已经应用于经济学、生物学、教育学、心理学、医学以及体育、工业、农业、林业、商业和金融等各个领域。

二、用 SPSS 对学生学习状况进行差异分析^[4]

用 SPSS 对学生学习状况进行差异分析的步骤是:

第一步,建立学生成绩 EXCEL 样本文档(如表 1)。

第二步,启动并进入 SPSS 系统,将上述表中数据导入系统。

第三步,进入系统分析功能,在“Analyze”菜单“Classify”中选择 Hierarchical cluster 命令。

第四步,在弹出的 Hierarchical cluster analysis 对话框中,从对话框左侧的变量列表中选择“高等数学、工业催化、金属学、英语、有色冶金原理……”等变量,使之添加到右边的 Variable(s) 框中。

第五步,确定变量的 Q 型聚类,选择复选框: Variables。

第六步,单击 Method 按钮,选择聚类方法: Between-groups linkage 类间平均法,即两类距离为两类元素两两之间平均平方距离。在 Inter

表 1 学生成绩表

姓名	中国近 现代史 纲要	高等 数学	工业 催化	思想道德 修养与 法律基础	计算机 文化 基础	金属 学	英语	马克思主 义基本 原理	有色 冶金 原理	冶金 物理 化学	冶金实 验研究 方法	冶金 化工 设备	现代 企业 管理	线性 代数	冶金反 应工程 学
1	65	55	78	67	77	79	67	64	75	83	89	83	70	74	78
2	77	60	63	78	80	80	63	78	70	31	79	60	70	60	69
3	81	81	75	75	76	64	76	73	70	68	81	60	65	63	89
4	63	60	75	74	76	62	64	69	85	72	86	65	61	68	92
5	66	76	98	75	78	72	82	64	84	65	73	80	68	75	92
6	70	76	73	69	72	69	65	73	71	68	78	75	62	76	69
7	66	60	96	65	71	71	60	71	82	72	95	65	65	40	88
8	60	60	85	64	62	72	70	64	81	60	70	80	70	60	73
9	77	78	75	68	80	68	65	69	74	68	80	72	65	86	78
10	75	73	60	73	76	88	66	74	82	70	72	50	65	72	68
11	81	69	84	74	80	62	58	78	83	71	78	74	72	71	90
12	81	69	90	67	67	82	61	73	86	60	96	65	70	66	68
13	63	84	88	60	64	70	63	69	85	86	89	60	80	87	83
14	75	82	73	76	64	88	65	72	91	60	72	60	66	60	68
15	76	73	76	72	73	87	74	80	69	60	95	62	67	64	94
16	67	88	90	78	80	76	76	70	90	60	80	60	62	81	71
17	84	77	88	71	66	84	60	67	83	60	82	60	61	69	93
18	77	60	63	75	79	67	66	80	79	64	79	65	78	19	62
19	66	63	70	72	90	50	71	54	73	67	84	86	62	60	87
20	81	80	76	68	66	60	60	74	70	85	65	62	64	60	74
21	65	62	83	65	64	68	60	78	60	86	89	73	77	63	87
22	77	71	86	77	72	73	60	76	70	70	99	69	73	63	81
23	79	73	84	79	87	96	74	84	90	60	85	60	80	70	85
24	81	77	80	68	77	84	64	67	82	69	85	85	75	60	78
25	67	75	84	70	65	63	72	72	62	60	89	60	65	70	87
26	84	60	75	76	70	79	74	74	78	60	73	61	61	97	73
27	67	74	77	75	77	74	48.5	78	67	65	70	61	61	82	77
28	66	82	83	60	61	88	63	60	92	64	85	60	62	84	78
29	75	73	60	73	76	88	66	74	82	70	72	50	65	72	68
30	81	69	84	74	80	62	58	78	83	71	78	74	72	71	90

注:为方便运算分析,作者将分学期的外语、高数成绩进行了合并。

val 栏目中选皮尔逊相关系数(R):Pearson correlation。

第七步,单击 Plots 按钮,选中 Dendrogram 项,并选择纵向 (Vertical) 输出聚类全过程 (All Clustres) 的冰柱图。

第八步,显示凝聚状态表:单击 Statistics 按钮,选中 Agglomeration Schedule 和 Proximity matrix 项,选中 Cluster membership 中的 Single solution 项,并在其后文本框中输入 3,显示将变量分成 3 类时各个变量的归属情况。

第九步,单击 OK 按钮,SPSS 自动完成分析过程。

三、分析结果和建议

1. 分析结果。

(1) 第一部分输出的是层次 Q 型聚类的分析结果(见表 2),从结果中可以看出 30 个样本都进入了聚类分析。

表 2 Case Processing Summary(a)

		Cases				
		Valid		Missing		Total
N	Percent	N	Percent	N	Percent	
30	100.0%	0	.0%	30	100.0%	

a Correlation between Vectors of Values used

(2) 输出 SPSS 层次聚类分析各变量的距离矩阵(见表 3)。从中可以看出各个变量之间的距离,并按“距离”的长短进行分类。

(3) 凝聚状态表(表 4)。表格的第一列表示聚类分析的第几步;第二列、第三列表示聚类中哪

表 3 Proximity Matrix

Case	Matrix File Input														
	中国近现代史纲要	高等数学	工业催化	思想道德修养与法律基础	计算机文化基础	金属学	英语	马克思主义基本原理	有色冶金原理	冶金物理化学	冶金实验研究方法	冶金化工设备	现代企业管理	线性代数	冶金反应工程学
中国近现代史纲要	10.	— 0.	0.	0.	0.2242	— 0.	0.	0.	— 0.	— 0.	— 0.	0.	— 0.	— 0.	— 0.
14442	21437	419287	177098	0.09017	466591	043944	23304	1151	22848	072441	05988	0962			
高等数学	0.	1	0.	— 0.	— 0.	0.	0.	— 0.	0.	— 0.	— 0.	— 0.	— 0.	— 0.	— 0.
14442	1	182042	0264	17615	191957	106588	04714	238426	059223	15851	33969	14969	372153	0053	
工业催化	— 0.	0.	1	— 0.	— 0.	— 0.	0.	— 0.	0.	0.	0.	0.	0.	0.	0.
21437	182042	1	28231	274	07579	047708	18881	229011	175068	388227	286484	139913	154184	509945	
思想道德修养与法律基础	0.	— 0.	— 0.	1	0.	0.	0.	0.	— 0.	— 0.	— 0.	— 0.	— 0.	— 0.	— 0.
419287	0264	28231	1	606054	084771	24613	458202	00747	47978	24572	21585	08891	07219	04856	
计算机文化基础	0.	— 0.	— 0.	0.	1	— 0.	0.	0.	— 0.	— 0.	0.	0.	— 0.	0.	
177098	17615	274	606054	1	14999	204215	108549	029446	17773	06329	23367	073003	08269	097417	
金属学	0.2242	0.	— 0.	0.	— 0.	1	0.	0.	— 0.	— 0.	— 0.	— 0.	0.	— 0.	
191957	07579	084771	14999	1	096161	203981	425077	33843	040259	43129	094319	173089	29994		
英语	— 0.	0.	0.	0.	0.	0.	1	— 0.	0.	— 0.	— 0.	— 0.	0.	0.	
09017	106588	047708	24613	204215	096161	1	22093	121479	23046	02601	054585	05243	089146	095956	
马克思主义基本原理	0.	— 0.	— 0.	0.	0.	0.	— 0.	— 0.	— 0.	— 0.	— 0.	— 0.	— 0.	— 0.	
466591	04714	18881	458202	108549	20398	1	220931	21558	12672	0044	44456	381055	17677	10659	
有色冶金原理	0.	0.	0.	0.	0.	0.	— 0.	1	— 0.	— 0.	— 0.	— 0.	0.	— 0.	
043944	238426	229011	00747	029446	425077	121479	21558	09922	10136	13525	053127	097682	13394		
冶金物理化学	— 0.	0.	— 0.	— 0.	— 0.	— 0.	— 0.	— 0.	0.	0.	0.	0.	0.		
23304	059223	175068	47978	17773	33843	23046	12672	09922	1	08935	219066	189642	065486	226231	
冶金实验研究方法	— 0.	— 0.	0.	— 0.	— 0.	0.	— 0.	— 0.	— 0.	0.	1	0.	0.	— 0.	
1151	15851	388227	24572	06329	040259	02601	0044	10136	08935	1	127193	301409	15527	38234	
冶金化工设备	— 0.	— 0.	0.	— 0.	— 0.	0.	— 0.	— 0.	0.	0.	0.	1	— 0.	— 0.	
22848	33969	286484	21585	23367	43129	054585	44456	13525	219066	127193	1	207093	11395	24493	
现代企业管理	0.	— 0.	0.	— 0.	0.	0.	— 0.	0.	0.	0.	0.	0.	1	— 0.	
072441	14969	139913	08891	073003	094319	05243	381055	053127	189642	301409	207093	1	31653	005613	
线性代数	— 0.	0.	0.	— 0.	— 0.	0.	— 0.	0.	— 0.	— 0.	— 0.	— 0.	— 0.	0.	
05988	372153	154184	07219	08269	173089	089146	17677	097682	065486	15527	11395	31653	1	10244	
冶金反应工程学	— 0.	— 0.	0.	— 0.	0.	— 0.	0.	— 0.	— 0.	0.	0.	0.	0.	0.	
0962	0053	509945	04856	097417	29994	095956	10659	13394	226231	38234	24493	005613	10244	1	

表 4 Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4	5	.606	0	0	7
2	3	15	.510	0	0	5
3	1	8	.467	0	0	7
4	6	9	.425	0	0	10
5	3	11	.385	2	0	8
6	2	14	.372	0	0	10
7	1	4	.291	3	1	13
8	3	12	.220	5	0	11
9	10	13	.190	0	0	11
10	2	6	.175	6	4	12
11	3	10	.170	8	9	14
12	2	7	.103	10	0	13
13	1	2	-.006	7	12	14
14	1	3	-.088	13	11	0

两个样本或小类聚成一类；第四列是相应的样本距离或小类距离；第五列、第六列表示本步聚类中，参与聚类的是样本还是小类。0 表示样本，数据 n(非 0)表示由第几步聚类产生的小类参与本步聚类；第七列表示本步聚类的结果将在下面聚

类的第几步中用到。

(4) 表 5 是变量层次聚类分析聚成 3 个类时变量的类归属情况。

从表 5 中我们可以推出表 6 所示的分类，并逐一定义。

表 5 Cluster Membership

Case	3 Clusters
中国近现代史纲要	1
高等数学	2
工业催化	3
思想道德修养与法律基础	1
计算机文化基础	1
金属学	2
英语	2
马克思主义基本原理	1
有色冶金原理	2
冶金物理化学	3
冶金实验研究方法	3
冶金化工设备	3
现代企业管理	3
线性代数	2
冶金反应工程学	3

表 6 专业课程分类

课程名称	课程分类
中国近现代史纲要、思想道德修养与法律基础、计算机文化基础、马克思主义基本原理	第 1 类:我们定义为,属于偏重行政类
高等数学、金属学、英语、有色冶金原理、线性代数	第 2 类:我们定义为,属于偏重研究类
工业催化、冶金物理化学、冶金实验研究方法、冶金化工设备、现代企业管理、冶金反应工程学	第 3 类:我们定义为,属于偏重工程类

表 7

Vertical Icicle

Number of clusters	Case														
	现代企业管理	冶金物理化学	冶金化工设备	冶金实验研究方法	冶金反应工程学	工业催化	英语	有色冶金原理	金属学	线性代数	高等数学	计算机文化基础	思想道德修养与法律基础	马克思主义基本原理	中国近现代史纲要
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
5	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
6	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
7	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
8	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
9	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
10	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
11	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
12	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
13	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
14	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

根据上述分析所得到的 3 个分类以及我们分别对每个类的定义,可以看到,一类学生在“高等数学”、“外语”、“金属学”等方面学习能力比较强,具有较强的抽象思维、逻辑推理和外语基础,适合于从事科学研究工作,因此可以建议这些学生继续深造,有条件报考研究生的应动员其报考研究生,暂时不能考研的也要抓紧学习以争取进一步提高;一类学生行政管理和政治敏锐力较强,建议他们可以去考公务员,将来可能会在政界取得比较大的成就;再一类学生,形象思维能力和动手能力较强,他们比较适合从事本专业的工程技术工作,建议他们进一步打好基础,深入掌握实际中的

(5) 由层次聚类分析的冰柱图(见表 7),可以非常清楚地看到各变量按 3 类归属的情况。

(6) 图 1 为聚类分析的树形图(Dendrogram),表明每一步中被合并的类及其系数值,把各类之间的距离转换成 1~25 间的数值。

2. 基于能力倾向的就业指导。

一些技巧,以便将来能成为本专业的技术骨干和专家。^[5]

四、结语

本研究采用教育统计原理对学生的专业课程成绩进行数据采样,并借助 SPSS 软件平台的因子、聚类分析方法对学生就业潜能进行客观评价,取得了明显的效果^[6]。尤其是通过本研究所建立的与学生就业倾向相关的数据库,为更好实施大学生职业生涯教育、改进大学生就业管理积累了丰富的实验数据和素材,并从数据挖掘的新视角探索出一种学生思想工作的新方法。

(下转第 135 页)

- [2] 阳仁宇:《高校课程改革与发展趋势》,《高等理科教育》2003年第2期。
 [3] 李慧仙:《高校课程负责人制的改革建议》,《高等工程教育研究》2005年第1期。

- [4] <http://zsxx.e21.edu.cn/sqlimg/school-h.html-jt/2005-06-22/article-22307.htm>
 [5] 叶辉:《应用型高校的生命线——访宁波高等专科学校校长高浩其》,《光明日报》2003年11月11日。

Construction of University Core Curriculum and Its System

Qin Yanfen

At present in our country, basic trends of curriculum reform in universities are as follows: to establish a curriculum system which adapts to popular higher education; to respond to the new requirements the gradually improved socialist market economy system put forward; to educate talents of high qualities and core competency to meet the social economic development needs. In view of these trends and missions, core curriculum construction has to be valued in curriculum reform of universities or colleges. The internal and external relationships in the curriculum system predominated by the core curriculums should be well handled. At the same time, a set of system frame including curriculum director system should be established in order to make concrete measures into effect.

(上接第 104 页)

CASE 0 5 10 15 20 25
 Label Num +-----+-----+-----+-----+-----+

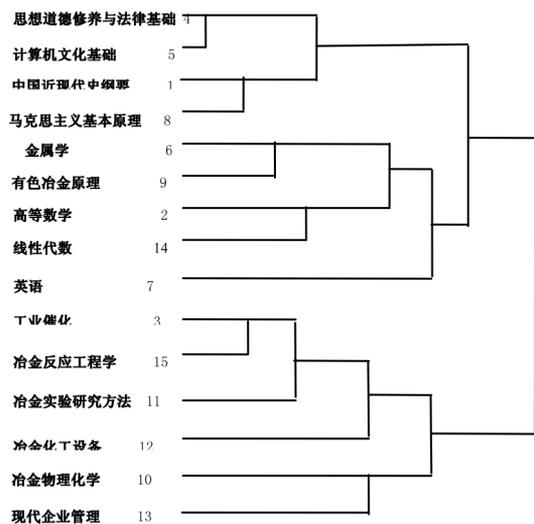


图 1 聚类分析的树形图

参 考 文 献

- [1] 佟庆伟、胡迎宾、孙青:《教育科研中的量化方法》,中国科技出版社 1997 年版,第 153~160 页。
 [2] 林治:《聚类分析在学生成绩管理中的应用探析》,《福建电脑》2009 年第 12 期。
 [3] 李志辉、罗平:《SPSS for Windows 统计分析教程(第 2 版)》,电子工业出版社 2005 年版,第 335~365 页。
 [4] 余建英、何旭宏:《数据统计分析与 SPSS 应用》,人民邮电出版社 2005 年版,第 251~289 页。
 [5] 张敏强:《教育与心理统计学》,人民教育出版社 2003 年版,第 332~369 页。
 [6] 罗家国:《基于 SPSS 的课程因子分析研究》,《江西理工大学学报》2011 年第 5 期。

Cluster Analysis of Students' Aptitude Based on SPSS

Luo Jiaguo, Luo Hao, Zhong Jiajia

This paper mainly introduces a method to evaluate the students' employment potentials based on SPSS clustering analysis. From the perspective of education statistics, the method is introduced to explore new ways of graduates' employment counseling as well as offering references for teaching management reform, thus reaching the goal of improving university students' career education and seeking the best way to strengthen student's employment potential orientation. The experimental results show that this method is rapid, visual, comprehensive and highly targeted.