

基于皮尔逊系数的管制仿真训练数据独立化与因子分析下的数据可视化研究

骆菁菁¹ 唐卫贞² 丁继婷³

1 中国民用航空飞行学院空中交通管理学院 四川 广汉 618307

2 中国民用航空飞行学院民航监察员培训学院 四川 广汉 618307

3 中国民用航空华东地区空中交通管理局 上海 200335

摘要 为解决管制仿真训练的指标重复性评分问题,文中基于皮尔逊相关系数进行指标关系研究,利用显著水平进行检验,并对评分数值算法进行修正。文中提取云端小程序评分数据库 4162 项,以因子分析理论为研究方法,采用主成分分析法进行因子载荷阵求解,并使用正交旋转法扩大载荷值,合理解释公共因子,建立管制学员素质能力模型,实现数据可视化。研究结果显示,去关联性后的指标评分统一呈下降态势,起伏变化与原评价分基本一致,可依照数据相关性与数值独立性进行部分指标描述更改;因子分析后可通过雷达图清晰地反映出学员能力,显示学员素质能力的优劣情况,也便于科学地进行管制岗位分配,是数据可视化的有效运用。

关键词: 数据可视化;皮尔逊系数;因子分析;管制仿真训练;独立化

中图法分类号 V335

Research of ATC Simulator Training Values Independence Based on Pearson Correlation Coefficient and Study of Data Visualization Based on Factor Analysis

LUO Jing-jing¹, TANG Wei-zhen² and DING Ji-ting³

1 College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan, Sichuan 618307, China

2 CAAC Academy, Civil Aviation Flight University of China, Guanghan, Sichuan 618307, China

3 East China Regional Air Traffic Management Bureau, Shanghai 200335, China

Abstract In order to solve the problem of repeated scoring on training indicators of ATC simulation, the indicators' relationship is studied based on Pearson correlation coefficient, and the significance level is used to test and the scoring algorithm is modified. This paper extracts 4162 items from the cloud small program scoring database, takes factor analysis theory as the research method, uses principal component analysis method to solve the factor load matrix, and uses orthogonal rotation method to expand the load value, reasonably explains the common factors, establishes the quality and ability model of the control students, and realizes data visualization. The results show that the score of the indicators after de-correlation show a downward trend, and the fluctuation is basically the same as that of the original evaluation score. Some indicator descriptions can be changed according to the correlation and independent values. The ability value after factor analysis can clearly reflect the ability of students through radar chart, showing the quality and ability of students, and is also convenient for scientific control of post allocation, which is an effective application of data visualization.

Keywords Data visualization, Pearson correlation coefficient, Factor analysis, Air traffic control simulator training, Retrieval

1 引言

管制模拟机训练是基于仿真模拟平台,对初级或复训管制员进行场景复现的一种学习模式,是管制员上岗训练和能力提升的重要环节。当前管制模拟机训练质量的反馈,主要以评分表纸质记载为主,根据多个指标将管制模拟机训练分类,逐一打分汇总,形成总分成绩以对学员进行评价。

在地方局培训部调研座谈中,研讨出当前评价突出的 3 个问题:1)数据记载均采用纸质记录,存在大量数据丢失与处理难度;2)各评分指标评分内容出现直接或潜在关联,导致了重复评分的现象;3)学员能力表述不清,除学员均总分以外,无其余数据反馈,导致分岗和进阶提升困难。

在 21 世纪这个信息时代,网络云端技术在大数据需求下应运而生。云服务通过互联网,使用网络服务器来构建云应

基金项目:民航安全能力项目:基于保障复杂度的民航中小机场空管人员培训体系建设(14002600100020J124);民航发展基金:民航监察员培训课程体系建设(2146999)

This work was supported by the Construction of Training System for Air Traffic Controllers in Medium and Small Airports of Civil Aviation Based on Complexity of Security(14002600100020J124) and Construction of Training Curriculum System for CAAC Supervisors(2146999).

通信作者:骆菁菁(acool@cafuc.edu.cn)

用程序,免去了服务器的应用安装,减少了维护和支持使用云服务开发的应用程序的工作。云服务具有不同层,包括软件即服务(SAAS)、基础设施即服务(IAAS)、平台即服务(PAAS)^[1]。云显示出了极好的运用价值,能够更高效地利用资源。

研究前,我们对于学员信息存储采用了阿里云服务^[2],购买云服务器 ECS(Elastic Compute Service),减小机房和物理组建建设成本,以微信小程序端为输入端,在云中存储结构化数据,实现了无纸化训练记载。

由于采用结构化数据存储模式,并一直实时接收手机端上的训练数据,数据逐渐演化为大数据,并具备 5V 特性,即数据量大(Volume)、速度快(Velocity)、类型多(Variety)、价值(Value)、真实性(Veracity),使数据分析成为了可能^[3]。管制训练指标独立化可基于大数据进行规律性探索,合理处理相应的数据。

与此同时,Analytic Visualizations(可视化分析)主要面向管理人员与普通用户,属于数据分析中的热点内容,数据可视化是数据分析工具最基本的主要体现。可视化可以直观地展示数据,让数据自己说话,让用户看到结果^[4]。这样就解决了上述 3 个问题,可以直观体现素质能力,便于分岗决策。

2 管制仿真训练数值独立化研究

评分指标由局方文件统筹下发标准,由各培训单位具体制定实施。但在具体制定时,底层指标或多或少会出现重复内容覆盖。以考核评估表指标为例,其不同二级指标下有底层考评指标内容重叠,不同的类别的底层指标之间的关系是错综复杂的。这样的关系会导致在某一技能素质中重复给分,进而使指标失衡。因此,考虑采用皮尔逊相关系数,建立数值独立化计算模型,对原始权重进行调节,可以解决重复性评分问题,客观反映学员训练的真实情况。

2.1 皮尔逊相关系数及矩阵构造

皮尔逊相关系数(Pearson Correlation Coefficient)需要在数据源组 $n > 2$,协方差是皮尔逊相关系数计算的前提,表示两个变量(x, y)间相互关系的数值特征^[5],如式(1)所示:

$$COV(x, y) = \frac{1}{n-1} \sum_1^n (x_i - \bar{x})(y_i - \bar{y}) \quad (1)$$

其中, x_i, y_i 为实际数据值, \bar{x}, \bar{y} 为相应模型的对值。指标发展趋势相互比较时,以当前指标为模型观察 x, y 的变化情况,可以看到当 x, y 变化趋势相同时,协方差为正, x, y 变化趋势相反时,协方差为负。其值区间大,关系明确,但是其值范围不一,受量纲影响。pearson 相关系数利用协方差除以标准差,如式(2)所示:

$$COR(x, y) = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_1^n (x_i - \bar{x})^2 \sum_1^n (y_i - \bar{y})^2}} \quad (2)$$

相关系数很好地反映了相关程度,其取值范围为 $[-1, 1]$,接近 1 表示相关性好,0 表示不相关,接近 -1 表示负相关^[6]。相关系数能够很好地实现去中心化处理,使得其值不再受量纲影响。为了清晰展示出指标间的系数关系,采用相关系数矩阵展示每两个指标的关联性^[7]。在大量的样本数据中,需要对整体关联性做出评估,以保证相关系数具备可靠性,因此采用置信区间做出检验。

置信区间计算样本数据落在模型周围的概率,来确定该模型的可靠性,如式(3)所示:

$$Pr(c_1 \leq v \leq c_2) = 1 - \delta \quad (3)$$

其中, δ 是显著性水平,常用 0.05 作为检验。由于发生概率满足正态分布,可利用标准差和正态分布表对区间极限进行计算,创建均值为中心,左右为阈值区间,数据落在正态分布概率率区间的发生率满足置信率,置信区间如式(4)所示:

$$\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}}, \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}} \right) \quad (4)$$

其中, \bar{X} 是“样本均值”,指单项所有数据均值; α 是置信度,显著性水平是 $\delta = 1 - \alpha$; $Z_{\frac{\alpha}{2}}$ 为 Z 值,指标准正态分布表对应值; σ 是总体标准差, N 是学员样本个体数, $\frac{\sigma}{\sqrt{N}}$ 是样本的标准误差。

在显著性检验中,存在单尾与双尾。单尾为单一方向检查,验证是否偏高或者偏低,以模型一侧为标准,即 90% 的置信区间;双尾检验抽样的样本统计量和假设参数的差是否过大,以模型双侧为标准,风险分摊到两侧,即 95% 的置信区间。表 1 列出了部分指标特性 SPSS 分析结果。

表 1 部分指标相关性与显著性表

Table 1 Partial index correlations

		陆空 通话	复诵 发现	口误	语速 时机	显示 处理
陆空通话	Pearson Correlation	1	0.324**	0.323**	0.302**	0.155**
	Sig. (1-tailed)		0	0	0	0
复诵发现	Pearson Correlation	0.324**	1	0.401**	0.286**	0.420**
	Sig. (1-tailed)	0		0	0	0
口误	Pearson Correlation	0.323**	0.401**	1	0.344**	0.281**
	Sig. (1-tailed)	0	0		0	0
语速时机	Pearson Correlation	0.302**	0.286**	0.344**	1	0.203**
	Sig. (1-tailed)	0	0	0		0
错误显示 处理	Pearson Correlation	0.155**	0.420**	0.281**	0.203**	1
	Sig. (1-tailed)	0	0	0	0	

** Correlation is significant at the 0.01 level (1-tailed).

* Correlation is significant at the 0.05 level (1-tailed).

表 1 中,单星表满足单尾检验,双星表满足双尾检验,在数据表中对应的相关系数值右上方标识。若出现相应指标相关系数未满足双尾显著性水平检验,则认为相关两项指标不具备相关性。

2.2 关联性去除与独立值的确定

相关系数矩阵表现了指标间的关联性,目的是为了去除指标的多余关联部分,使评分数值独立化。为此,我们建立独立值计算模型,以 1 记为完全独立量,计算每个指标的独立值,如式(5)所示:

$$\text{round} \left(1 - \frac{n-1}{2} \sum_1^n p_i \right), i=1, 2, 3, \dots, n \quad (5)$$

其中, P_i 表示指标与另一指标的单独相关关系,round 表示控制小数点位数。所求的独立值是以所有相关指标总关联量分析得出的。将独立值转化为指标权值,建立评分修正模型,计算修正后的学员评分与总均分,并与原均评分进行比较。图 1 给出了学员修正前后的分数比较。从图 1 可以观察到,独立修正后的分数明显低于原分数,双分数变化起伏基本一致。分数按百分比扩展之后能够恢复到原分数状态,部分略微有小幅度缩减,这与实际重复扣分情况是相符的。

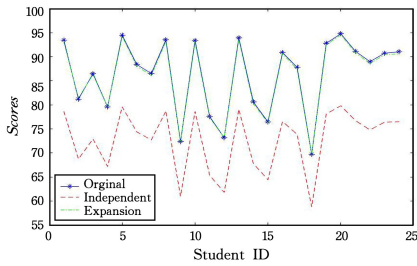


图1 独立化分与原分对比图

Fig. 1 Line chart of independent and original scores

由于变化规律基本一致,修正分数后学员排名与原排名基本一致,可不做变化;分数减少较多,可做修正处理,对独立性较差指标进行重新描述。参考表2所列的高关联度指标,来减少教员的重复评分与评分误差。

表2 高关联度指标列表

Table 2 List of high correlation indicators

Index Abbreviation	Pearson Correlation	Index Content	Associated Repetive Index
复诵发现	8.489	对机组的错误复诵能够进行监听和发现	对机长位的错误操作进行及时发现并处理
语言能力	7.776	具有良好的语言表达能力	管制指令中发生口误
应变能力	8.116	具有较好的应变能力	对雷达上的不正常显示进行及时处理
心理承受	8.397	具有较好的心理承受能力	冲突解脱及时有效
情绪控制	8.216	是否具有良好情绪控制能力	发话时机及语音语速的合理控制

3 学员技能素质可视化研究

数据去关联性反映了客观的学员训练分数,但学员的训练反馈仅此一个分数,毕竟以偏概全,删去了很多可评判内容。为了使呈现的数据更丰富、技能反映更直观、使数据说话,对训练数据进行可视化研究。根据云端数据库中的指标评分,使用因子分析法进行大数据分析,提取出指标数据中的共性因子,并加以合理解释,构造新的能力模块,形成数据可视化图形^[8]。

3.1 R型因子分析理论及运用流程

因子分析类型主要有两种,分别为R型与Q型。R型主要对变量做因子分析,Q型因子分析对样品做分析^[9]。由于学员能力素质为单个变量,R型更具有适应性。

1)R型因子分析基本模型。因子分析将变量解释为公共因子与特殊因子的线性之和,如式(6)所示:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{im}F_m + \epsilon_i, i = 1, 2, \dots, p \quad (6)$$

其中, F_m 是公共因子,学员的一种共性存在的能力, ϵ_i 是特殊因子,是学员独有的个体能力,模型用矩阵表示,如式(7)所示:

$$\mathbf{X} = \mathbf{A}\mathbf{F} + \boldsymbol{\epsilon}$$

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pm} \end{bmatrix},$$

$$\mathbf{F} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{bmatrix} \quad (7)$$

因子分析法主要利用降维思维,其公共因子数量小于变量数;除此以外,因子间互不相关,需体现很好的独立性,因此

相关系数可以进行整合。当其为0时表示线性无关,当其不为0但有相关关系时,就将因子整合到公共因子中。

2)KMO与Bartlett球形度检验。若变量间的相关程度变化差异大,则表明不适合做因子分析;显著性过高,也说明因子分析可能存在不合理的情况。为此进行了KMO与Bartlett球形度检验,结果如表3所列。KMO统计量值为0.853,位于0.8~0.9之间,变量相关程度差异小,适合做因子分析^[10];巴特利特球形检验结果小于0.05,球形假设被拒绝,原始变量存在相关性,适合做因子分析^[11]。

表3 KMO与Bartlett球形度检验表

Table 3 KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.853
Approx. Chi-Square		27677.162
Bartlett's Test of Sphericity	df	171
	Sig.	0

3)因子载荷阵求解。在式(7)中, A 称为因子载荷矩阵, a_{ij} 称为因子载荷,是第*i*个变量在第*j*个因子上的负荷。由于管制评分记载量纲相同,因此协方差标准化也比较简单,以相关矩阵为计算基底。因子载荷阵求法采用主成分分析法,设随机向量 $\mathbf{X} = (x_1, x_2, \dots, x_p)'$,均值为 μ ,协方差为 Σ , $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ 是协方差的特征根, $\mu_1, \mu_2, \dots, \mu_p$ 为对应标准化特征向量,则可推导出:

$$\Sigma = \mu \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix} \mu' \quad (8)$$

$$[\mu_1 \ \mu_2 \ \dots \ \mu_p] \begin{bmatrix} \lambda_1 & & & 0 \\ & \ddots & & \\ & & & \lambda_p \end{bmatrix} \begin{bmatrix} \mu_1' \\ \mu_2' \\ \vdots \\ \mu_p' \end{bmatrix} \quad (9)$$

$$= \lambda_1 \mu_1 \mu_1' + \lambda_2 \mu_2 \mu_2' + \dots + \lambda_m \mu_m \mu_m' + \lambda_{m+1} \mu_{m+1} \mu_{m+1}' + \dots + \lambda_p \mu_p \mu_p'$$

$$= [\sqrt{\lambda_1} \mu_1 \ \sqrt{\lambda_2} \mu_2 \ \dots \ \sqrt{\lambda_p} \mu_p] \begin{bmatrix} \sqrt{\lambda_1} \mu_1' \\ \sqrt{\lambda_2} \mu_2' \\ \vdots \\ \sqrt{\lambda_p} \mu_p' \end{bmatrix} \quad (10)$$

由于公共因子数量小于变量数,因此 $m+1 \sim p$ 无意义^[12],则优化如式(11)所示:

$$\Sigma \approx \lambda_1 \mu_1 \mu_1' + \lambda_2 \mu_2 \mu_2' + \dots + \lambda_m \mu_m \mu_m'$$

$$= \sqrt{\lambda_1} \mu_1 \ \sqrt{\lambda_2} \mu_2 \ \dots \ \sqrt{\lambda_m} \mu_m \begin{bmatrix} \sqrt{\lambda_1} \mu_1' \\ \sqrt{\lambda_2} \mu_2' \\ \vdots \\ \sqrt{\lambda_m} \mu_m' \end{bmatrix} \quad (11)$$

需要注意的是,因子载荷矩阵并不是唯一的,可通过因子正交变化的方法得到新的因子载荷阵,使得公共因子差别更清晰,线性关系更清晰^[13]。

4)因子命名和因子旋转。使用最大方差旋转法,其原理是使得旋转后因子载荷阵 A 的每一列元素的方差之和达到最大,从而实现使同一列上的载荷尽可能地靠近1和靠近0两极分离^[14]。这样突出了载荷较大变量与公共因子之间联系,便于对公共因子进行合理解释^[15]。

5)因子得分。因子得分可根据因子载荷阵得出,其结果

类似主成分分析,将用因子得分来构建模型表达公共因子。因子模型 $X=AF+\epsilon$,当其除去特殊因子影响,且矩阵 A 可逆, $m=p$,在每个指标因子 F 上相应取值, $F=A^{-1}X$,即因子得分。采用回归法进行, $F=A'\Sigma^{-1}X$, Σ^{-1} 为样本相关阵的逆。

3.2 公共因子提取与能力素质模型构建

公共因子可以被视为管制学员素质能力的体现,其公共因子需要数量合适,且具有代表性。其代表性的表达在因子旋转后更容易体现,也便于我们去解释公共因子对应的具体含义。在提取公共因子时,以特征值为提取要素。图 2 为因子载荷阵特征值碎石图。

我们从碎石图中观察到,前六项特征值差距较大,选取特征值 λ 大于 0.73 作为提取标准,构建因子模型。需要特别注意的是,管制学员由于各训练差异性,在数据整合时具备一定量的特殊因子,因此分析时不再多考虑贡献率,特征值反映良好即可。根据提取标准,我们得到了因子得分系数矩阵(见表 4),数据源为局方 4162 项云端电子评分记录。

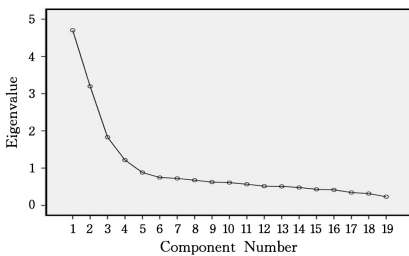


图 2 因子载荷阵碎石图

Fig. 2 Scree plot of factor load array

样本基数大,因子分析效果明显。根据局方评估需求,对公共因子进行合理命名^[16]。考虑载荷占比较大的指标并进行归纳总结,参考载荷占比最大指标确定因子名称。同时,在确定每个公共因子的对应指标时,还需要对因子得分进行归一化处理,使得最终各能力值不会产生较大差异,这是标准化手段^[17]。因子命名及模型系数如表 5 所列。

表 4 因子得分系数矩阵

Table 4 Component score coefficient matrix

	Component					
	1	2	3	4	5	6
陆空通话	-0.082	-0.041	0	0.368	-0.15	0.073
复诵发现	-0.027	-0.033	0.082	-0.137	0.479	-0.008
口误	0.223	-0.019	-0.033	-0.001	0.147	0.085
语速时机	0.278	-0.048	-0.035	0.197	-0.225	0.302
关键信息	0.225	0.041	0.112	-0.01	0.04	0.019
计划外活动	-0.025	-0.043	0.48	-0.136	0.216	-0.058
冲突意识	-0.011	0.045	0.022	0.491	0.054	-0.411
冲突解脱	-0.005	-0.108	-0.054	0.051	0.547	-0.128
交通意识	0.023	-0.039	0.465	0.032	-0.141	0.017
管制节奏	0.091	-0.125	-0.11	0.502	-0.081	0.075
顺序优化	-0.216	-0.037	-0.105	0.079	0.105	0.071
未放行数量	-0.185	-0.014	0.005	0.006	0.071	0.174
满足协议	-0.235	0.037	0.259	0.005	-0.067	0.023
语言能力	0.04	0.129	-0.027	-0.025	-0.222	0.428
沟通能力	-0.012	0.316	-0.012	0.014	-0.381	0.178
理解反应	0.061	-0.195	-0.004	-0.161	0.14	0.688
应变能力	-0.002	0.357	-0.087	0.058	-0.04	-0.248
心理承受	-0.002	0.376	0.033	-0.135	0.035	-0.17
情绪控制	0.002	0.374	-0.015	-0.099	0.022	-0.167

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

表 5 因子命名及模型系数

Table 5 Factor named & normalizing proportion

	Larger load	Factor named	Normalizing proportion
K1	X3, X4, X5	信息获取能力	0.307, 0.383, 0.31
K2	X14, X15, X17, X18, X19	心理沟通能力	0.083, 0.204, 0.23, 0.242, 0.241
K3	X5, X6, X9, X13	协调观察能力	0.085, 0.365, 0.353, 0.197
K4	X1, X4, X7, X10	节奏控制能力	0.236, 0.126, 0.315, 0.322
K5	X2, X3, X6, X8, X11	应变解脱能力	0.321, 0.098, 0.145, 0.366, 0.07
K6	X4, X12, X14, X15, X16	逻辑思维能力	0.171, 0.098, 0.242, 0.101, 0.389

3.3 数据可视化图形表达

学员模拟机中能力反馈并不是单一分数,不能单纯加减,是一种多维度的反馈。这种反馈通过因子分析方法进行总结归纳,建立一个科学化的素质能力模型,来计算单个学员的多个分值。为了表达这种多维度信息,雷达图成为了一种绝佳选择。雷达图可以在一张图表绘制多个信息,既能够突出单个信息的影响,也能通过图形面积反映总体情况,这与管制学员能力强弱展示和综合评估的需求是相匹配的。根据上文的能力模型以及学员总均分的计算方法,参照对比雷达图设定的标准数据,可计算出当前学员的能力值,具体如式(12)所示:

$$A = \begin{bmatrix} a_{1k1} & a_{1k2} & \cdots & a_{1kn} \\ a_{2k1} & a_{2k2} & \cdots & a_{2kn} \\ \vdots & \vdots & & \vdots \\ a_{nk1} & a_{nk2} & \cdots & a_{nkn} \end{bmatrix}$$

$$X' = \frac{V_{\max} \cdot X_i \cdot A}{X_{\max} \cdot A} \cdot \frac{V_e}{X_i \cdot A} \quad (12)$$

其中, A 矩阵为筛选归一化后的因子得分矩阵, V_{\max} 为雷达图表达中的标准满分, V_e 为雷达图中均值的表达值, X 为学员评分总均值, X 为评分最大均值。式(12)中, $\frac{V_e}{X_i \cdot A}$ 若不计算,则为最值雷达图,若计算,则为均值雷达图。

(1)最值雷达图,也称传统雷达图,是将各维度数据按照设定的满值换算而来,显示的是每种能力与满值之间的关系,可以看出学员能力与要求的差距。这种显示方式能清晰洞悉能力强弱,但不能与组员间形成对比,会形成部分认识误差。如图 3 所示,该学员单纯观察红框,逻辑思维能力强比节奏控制能力强,但在组员对比中,该学员的节奏控制能力会更占优势,这恰恰是管制分岗所关注的,因此需要优化雷达图。

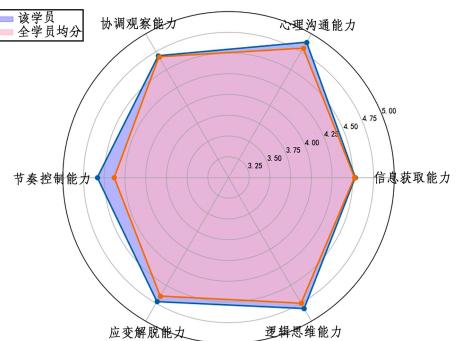


图 3 最值雷达示例图

Fig. 3 Maximum value radar example chart

(2)均值雷达图,是优化后的雷达图,如图 4 所示,基本思路是将所有组员能力均分作为参考对比,计算出学员的新能力值并表达出来。这样符合视觉习惯,均分形成了正六边形,更便于进行组员对比,获知当前学员是高于平均水平还是低于平均水平,与平均水平相差多少。

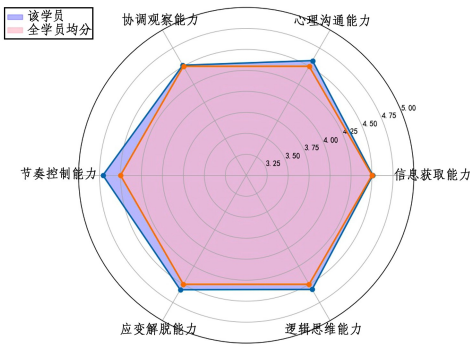


图4 均值雷达示例图

Fig. 4 Mean value radar example chart

参考最值雷达图,可以为学员今后的能力训练方向做出

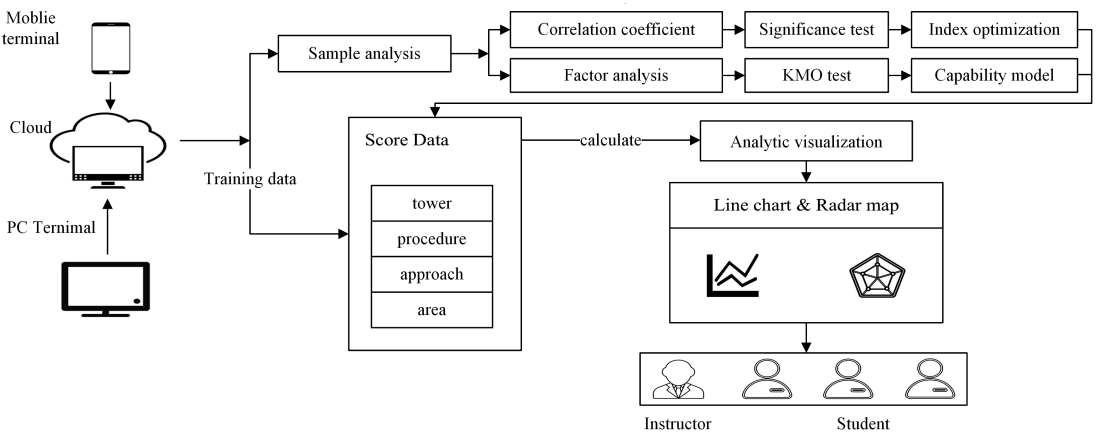


图5 模型的主要结构和实现流程

Fig. 5 Main structure of proposed model & realization process

教员用户通过移动端(pc, mobile terminal)将信息输入,由云端存储。接着将完整期班数据作为样本以定制数据端口dict形式, json格式输入进行算法分析,调用其numpy&pandas库修正分数计算模型和能力模型代码。然后,通过用户请求,将对应类别人员的数据输入模型进行计算,输出数字型信息。最后,使用matplotlib库,将信息数据可视化,返回至用户端,实现实时查看与更新。

结束语 本文参考一线管制培训部需求,对管制评价指标的独立性和评价数据的可视化进行了深入研究,提出使用相关系数与因子分析方法进行问题解剖,研究内容包含以下方面:

(1)评价数值可通过皮尔逊相关系数来去除关联性,其分数去关联后明显低于原分数,但与原分数起伏化基本一致。基于此研究结果建议修改指标描述,可不修改学员分数排名。

(2)因子分析下能力模型建立有一定的科学性和实用性。通过两种雷达图既可以表达单个学员的多维度能力值,又能够使学员进行均值对比,能够在不同需求下反映出各个能力的强弱,对学员后续的训练重点和岗位分配有一定的指导意义。

学员模拟机分数评价不仅包含指标数据,还包括文字信息。在今后的研究中,可推进词频分析(word frequency analysis)和基于BP nerve network的定量分析,进一步完善评价系统。

参考文献

[1] BELLO S A, OYEDELE L O, AKINADE O O, et al. Cloud com-

调整;参考均值雷达图,可以获知学员相较于平均水平能力的强项。因此,参考该两种雷达图,掌握学员能力长短,快捷对应岗位,得到学员岗位分配建议。

能力与岗位之间也存在着一定的分配关系,其中典型的包括具备较好协调观察能力的更适合区域协调席,应变解脱能力好的更适合到塔台席,逻辑思维能力突出的适合到进近管制席。均值雷达图能够直观表现该学员在该组学员中能力较强的项,从而给出了岗位分配建议。

4 算法融合及实现流程

对数据进行实时的二次分析,并以可视化数据形式表达出来,提供给管理教员与培训学员,这都归功于定制的云服务器。图5给出了整体模型结构与实现流程。

puting in construction industry: Use cases benefits and challenges[J]. Automation in Construction, 2021, 122:103441.

[2] LI X G. The development of designated driving application based on Android platform and Ali cloud sever[C]// 2016 2nd Workshop on Advanced Research and Technology in Industry Applications. 2016:633-636.

[3] KUMAR N, KUMAR G, KUMAR S R. Big data analytics application for sustainable manufacturing operations: analysis of strategic factors[J]. Clean Technologies and Environmental Policy, 2021(23):1-25.

[4] CHOI D, LEE H, BOK K, et al. Design and implementation of an academic expert system through big data analysis[J]. The Journal of Supercomputing, 2021(78):1-25.

[5] LIANG J Y, FENG C J, SONG P. A Survey on Correlation Analysis of Big Data[J]. Chinese Journal of Computers, 2016, 39(1):1-18.

[6] JIN L, LI Y. Discrimination of Several Correlation Coefficients and Their Implementation in R Software[J]. Statistics & Information Forum, 2019, 34(4):3-11.

[7] SHI S, LIU W Y. Factor Analysis Method Research of Evaluation Index Connection[J]. Mathematics in Economics, 2010, 27(2):23-27.

[8] XU W P. Design and Implementation of Data Visualization System Architecture[D]. Beijing: Beijing Jiaotong University, 2015.

[9] LEE A, KIM S E, SUH K D. Estimation of Stability Number of Rock Armor Using Artificial Neural Network Combined with

- Principal Component Analysis[J]. *Procedia Engineering*, 2015, 116:149-154.
- [10] LIU Y, YANG Q W, YAO Y H, et al. Current Situation and Existing Problems of Factor Analysis Method[J]. *Journal of Finance and Accounting*, 2019, 7(6):192-199.
- [11] WANG Y, DENG X D. Empirical Study on the Performance Evaluation of Agricultural Products Supply Chain Based on Factor Analysis[J]. *China Business and Market*, 2015, 3:10-16.
- [12] MOSLEM I, DIANAT I, OLFAT I, et al. Selection of elite F3 lines in cucumber using factor analysis method[J]. *International Journal of Vegetable Science*, 2020, 26(5):498-505.
- [13] WANG M M. Research on PTC Cable Materials Based on Principal Component Analysis and Quantitative Correspondence Factor Analysis Method in Big Data Technology[C]//IOP Conference Series: Materials Science and Engineering, 2019, 493(1):012048.
- [14] WEN R Y, WANG H Y. Evaluation of Air Traffic Complexity Based on Factor Analysis and K-means Clustering[J]. *Journal of Taiyuan University of Technology*, 2016, 47(3):384-388.
- [15] WEIDE A C. A Beauducel. Varimax Rotation Based on Gradient Projection is a Feasible Alternative to SPSS[J]. *Frontiers in Psychology*, 2019, 00645.
- [16] ALAVI S M, EBRAHIMI A, POURMOURDININ E, et al. Identifying and Classifying the Factors Affecting Fast Food Store Selection Using Factor Analysis Method[J]. *Journal of Business Management*, 2018, 9(4):827-854.
- [17] YE Y. Study on the Evaluation Index System of Guangdong's All-for-One Tourism Destination Competitiveness Based on Factor and Cluster Analysis[J]. *Open Journal of Social sciences*, 2019, 7(3):418-428.



LUO Jing-jing, born in 1990, postgraduate, probationer. Her main research interests include air traffic control training data visualization and flight procedure design.

(上接第 618 页)

参 考 文 献

- [1] YANG W X, FAN K L, LEUNG L P. The practice and safety profile of endotracheal intubation in an emergency department: A single-center study in Shenzhen, China [J]. *HongKong Journal of Emergency Medicine*, 2020, 27(6):428-433.
- [2] ANTONIO C, SARA N, DEBORA S, et al. Impact of COVID-19 on the Pattern of Access to a Pediatric Emergency Department in the Lombardy Region, Italy [J]. *Pediatric Emergency Care*, 2020, 36(10):213-225.
- [3] WANG X M. Research on mobile model and emergency communication system of post-earthquake emergency rescue [D]. Donghua University, 2020.
- [4] TIAN Y. Design and implementation of Emergency Mobile Communication Command Platform [D]. Tianjin: Tianjin University, 2018.
- [5] SHAO J W. An emergency command system based on scattering communication [D]. Xidian University, 2015.
- [6] JAMIE M, JANETTE T, RICHARD J, et al. Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review[J]. *Diagnostic and Prognostic Research*, 2020(10):725-734.
- [7] GE Y, WANG T, XU Y B. Assessment of emergency management capability based on gray analytic hierarchy process [J]. *Journal of Safety Science and Technology*, 2014, 10(12):80-86.
- [8] SHEN H Z. Emergency sequential group decision-making and support system for sudden critical events [M]. Beijing: Science Press, 2011:79-85.
- [9] DU Y Y, JIANG C J, ZHOU M C. A Petri net-based model for verification of obligations and accountability in cooperative systems [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 2009, 39(2):299-308.
- [10] HSIEH M C, WANG M Y, LEE W C, et al. Application of HFACS, fuzzy TOPSIS, and AHP for identifying important human error factors in emergency departments in Taiwan[J]. *International Journal of Industrial Ergonomics*, 2018, 67:171-179.
- [11] OZDEMIRA, TUYSUZ F. An Integrated Fuzzy DEMATEL and Fuzzy ANP Based Balanced Scorecard Approach: Application in Turkish Higher Education Institutions[J]. *Journal of Multiple-Valued Logic & Soft Computing*, 2017, 28:251-287.
- [12] CHEN E Y, YUAN J H, LI Z W. Research on JSON-based sensor source trust evaluation model for Environmental monitoring [J]. *Journal of the Xi'an University of Technology*, 2020(7).
- [13] DU Y Y, NING Y H, LIANG Q. Reachability analysis of Logic Petri Nets using incidence matrix [J]. *Enterprise Information Systems*, 2014, 8(6):630-647.
- [14] DU Y Y, NING Y H. Property analysis of Logic Petri Nets using reachable marking graphs [J]. *Frontiers of Computer Science in China*, 2014, 8(4):684-692.
- [15] ZHEN S Z, DENG Y F. Evaluation System of Urban Public Emergency Response Capability and its application [J]. *Journal of the Liaoning Technical University*, 2006, 25(6):943-946.
- [16] WANG W, CAO Y. Evaluation of Government Emergency Management Capability based on BP neural network [J]. *Research on Science and Technology Management*, 2018(10):117-125.
- [18] YUAN S J, LIANG R W. PSO algorithm based emergency logistics vehicle scheduling [J]. *Logistics Technology*, 2019(7):339-346.
- [19] WANG F. Research on crowd evacuation method and simulation based on BAS algorithm [D]. Jinan: Shandong Normal University, 2019.



NING Yu-hui, born in 1981, Ph.D., senior engineer. His main research interests include design and application of computer software, command communication and emergency command equipment.