

摘要

城市的发展对人才的吸引力两者之间是相互联系与制约的,因此需要政府决策与措施影响,使城市增强对人才的吸引力。本文结合武汉的发展特点,利用爬取后的数据较全面的选择合适的指标进,通过因子分析的方法为武汉市不同年份人才吸引力打分;再对人才进行产业类别分类,通过灰色关联分析筛选出合适的因子后再横向比较各城市综合得分,并结合当地人才政策评价对人才吸引力的水平提出合理化建议。

针对问题一,通过因子分析模型分析武汉 2013—2017 年相关数据,对武汉各年的人才吸引力进行量化打分。本组首先根据题目要求和以往人才吸引要素研究,确立要素框架并爬取合适的人才吸引力因素。在标准化数据通过 KMO 检验后,利用 SPSS 对各个因素进行主成分分析以确定公因子数目,求解旋转后因子荷载矩阵并对公因子命名,计算武汉各年人才吸引力综合得分。综合得分结果为-77.99、-30.08、-11.90、42.35、77.63。综合得分表明武汉 2013—2017 人才吸引力逐年上升,且 2015—2016 年上升幅度最大,与城市实际发展情况相符合。

针对问题二,本文在问题一模型的基础上进行改进,将人才分为第一、第二和第三产业人才并用灰色关联分析筛选出每类人才相应的人才吸引力因素。然后将筛选后的数据用 SPSS 进行因子分析得到成都、天津、西安、南京和武汉 2013—2017 年的三类人才吸引力因子分析模型,并计算每个城市各年针对三类人才的人才吸引力综合得分。横向比较各个城市的综合得分,相较与其他城市,武汉市近几年对第二、三产业人才的吸引力有较高水平,相对于西安和成都,武汉市具有较强的人才吸引力,符合地区的经济发展特点。

针对问题三,基于问题一中的武汉市人才吸引力评价模型以及人才政策对人才吸引力的量化评价,并结合问题二中武汉相较于其他四个同类城市在人才吸引力上的优势与不足,给武汉市人力资源管理部门的领导写一篇建议报告。

本文中所提到的模型优点主要有两点:一、选取的指标从多维度、全方面考虑,收集的数据真实可靠;二、利用灰色关联度筛选出三类人才的人才吸引力要素,相较于人工选取指标具有客观性、严谨性。

关键词: 因子分析模型 主成分分析 灰色关联分析 数据库爬虫

目录

一、 问题重述	4
1.1 问题背景	4
1.2 问题提出	4
二、 模型假设	4
三、 符号说明	5
四、 问题一模型的建立与求解	6
4.1 问题的描述与分析	6
4.2 预备工作	6
4.2.1	6
4.2.2 同向化处理	7
4.2.3	7
4.2.4	7
4.3 模型的建立与求解	8
4.3.1 公因子的确定	8
4.3.2 因子命名与载荷矩阵计算	10
4.3.3 求得因子得分和综合绩效得分	12
4.4 结果分析	14
五、 问题二模型的建立与求解	14
5.1 问题的描述与分析	14
5.2 模型的建立与求解	15
5.2.1 人才及其吸引力影响因素的分类	15
5.2.2 分类型人才吸引力分析	17
5.3 结果分析	19
六、 写给武汉市人力资源部门的建议报告	20
七、 模型的评价	22
7.1 模型的优点	22
7.2 模型的缺点	22
附录 A 代码	24
A.1 数据预处理—python 源代码	24

A.2 数据可视化—python 源代码	24
A.3 灰色关联度—matlab 源代码	25
A.4 相关系数矩阵	28
A.5 初始因子载荷矩阵	28
A.6 旋转后的因子载荷矩阵	30
A.7 成分得分系数矩阵	32

一、问题重述

1.1 问题背景

植物的种类繁多,要了解和掌握如此多的植物,必须进行一个科学的分类。人们常常根据植物的用途,或根据植物的一个或几个明显的形态进行分类,植物的识别与分类对于区分植物种类,探索植物间的亲缘关系,阐明植物系统的进化规律具有重要的意义。因此植物分类学是植物科学甚至整个生命科学的基础学科。目前对于树叶识别与分类主要由人完成,但是树叶种类庞大,依赖人工地进行树叶识别与分类是不现实的。所以树叶的研究对于植物总体的研究能提供很大的帮助。

从树叶的各个方面,纹理,硬度,离心率等方面都可作为主要方向研究,现对树叶的研究主要通过采集树叶图形,利用数字图像处理来对树叶进行分类识别,这种方法只停留在处理形态特征,有很大的局限性,忽略了树叶的生理特征和其他特征,所以研究方法来综合处理树叶平面图像特征,形态特征和生理特征很有必要性。

1.2 问题提出

围绕植物分类进行树叶识别与分类,以树叶二值化图片为依据,依次提出以下问题:

- (1) 结合附件中的二值化的图片数据,建立合适的图片数据提取方案,量化处理图片数据,并具体分析说明所提取数据信息的量化指标体系。
- (2) 基于问题一中提取的数据信息,建立合适的数学模型由数据出发判断叶子的种类,研究判别分类的核心指标,并估计出模型的性能以及核心指标对模型判别性能的影响。
- (3) 基于二值化图片数据,结合附件中叶子纹理的数据信息,对原有模型进行改进,并对新旧模型进行比较分析。

二、模型假设

- (1) 假设相同的行业在不同的城市里吸引力影响因素相同,每个行业在不同城市里的发展模式近似相同。
- (2) 人才会考虑未来一段时间内自身对于发展前景、收入、环境的需求的变化。
- (3) 政策对人才吸引力的影响转换成三个影响因素——地方财政支出与收入和固定资产投资总额。
- (4) 人才的迁移是在追求效用最大化,人才的行为仅受到城市因子的影响,忽略人的非理性行为。

三、符号说明

符号	说明
F	城市各年人才吸引力综合得分
F_1	工业发展与薪酬因子
F_2	医疗卫生环境因子
F_3	经济贸易因子
F_4	拥挤程度因子
X	原始指标
\bar{X}	指标平均值
\tilde{X}	同向化指标
δ_X	指标标准差
Z	标准化指标
R	相关系数矩阵
λ_p	相关系数矩阵特征值
η_p	标准正交化特征向量
Λ	因子载荷矩阵
σ_i	方差
α_{ij}	载荷因子
k	两极最小差
K	两极最大差
$\Delta_i(t)$	特征序列与因素序列的序列差

四、问题一模型的建立与求解

4.1 问题的描述与分析

针对问题一，本题要求对武汉的人才吸引力做出量化评价。对于人才吸引力水平模型，现在主要存在的问题是定性研究较多、量化评价较少，对变量之间的关系讨论更是不足。本文采用因子分析法建立模型，将通过不同的因子对变量的影响程度进行分析，建立客观评价模型。首先根据题目要求和以往人才吸引要素研究，寻找合适的人才吸引力因素，确立要素框架并爬取各个因素相对应的数据。在将所得数据进行无量纲化处理后，进行 KMO 检验和巴特利特检验以验证所选变量是否适合做因子分析。通过主成分分析法确定公因子数目，求解旋转后因子荷载矩阵并对公因子命名，最后利用因子得分与其奉献率计算武汉各年人才吸引力综合得分，其流程图如图 ?? 所示：

4.2 预备工作

4.2.1

树叶图片的属性分为两种，一种是按树叶的形态特征，另一种是按树叶的纹理特征。

I 表示目标树叶所对应的图像， ∂I 表示图像边界， $D(I)$ 表示外接圆直径， d 表示内切圆半径， $A(I)$ 表示其范围面积， B 表示界坐标， $L(\partial I)$ 表示研究对象的轮廓线长度， $A(H(I))$ 表示研究对象的凸包域， C 为凸型区域中心坐标，运算符 $d(\cdot)$ 代表欧式距离。以下是对部分属性的具体的数学描述。

- (1) 离心率 (Eccentricity): 是标量，目标树叶对应椭圆的离心率，变化范围为 $(0,1)$ 。
- (2) 密实度 (Solidity): 是标量，是一个仿射特征，反映出树叶图片区域的固靠性程度，计算公式为 $\frac{A}{A(H(I))}$ 。
- (3) 等周因子 (Isoperimetric Factor): 计算公式为 $\frac{4\pi \cdot A}{L(\partial I)^2}$ ，叶子边缘越规则，越接近于最大值 1。
- (4) 伸长率 (Elongation): 计算最大距离 $d_{max} = \max_{x \in I} d(X, \partial I)$ ，伸长率即为 $1 - \frac{2d_{max}}{D(I)}$ ，范围为 $(0,1)$ ，树叶越趋于圆形，相应的伸长率越小，比率 $\frac{2d_{max}}{D(I)}$ 为最大的内切圆的直径和最小的外接圆的直径的比值。□
- (5) 最大压痕深度 (Maximal IndentationDepth): 记 $C_{H(I)}$ 表示几何中心， $L(\partial I)$ 表示为 $H(I)$ 的轮廓线长度， $\forall X \in H(I)$ 和 $\forall Y \in \partial I$ ，计算距离 $d(X, C_{H(I)})$ 和 $d(Y, C_{H(I)})$ 。定义函数 $\frac{d(X, C_{H(I)}) - d(Y, C_{H(I)})}{L(H(I))}$ ，该函数的最大值即最大压痕深度。
- (6) 长宽比 (Aspect Ratio): 记 X_1 为图片二值化后的 $0-1$ 矩阵出现 1 的最小行数， X_2 为出现 1 的最大行数， Y_1 为出现 1 的最小列数， Y_2 为出现 1 的最大列数，计算公式为 $\frac{X_1 - X_2}{Y_1 - Y_2}$ 。

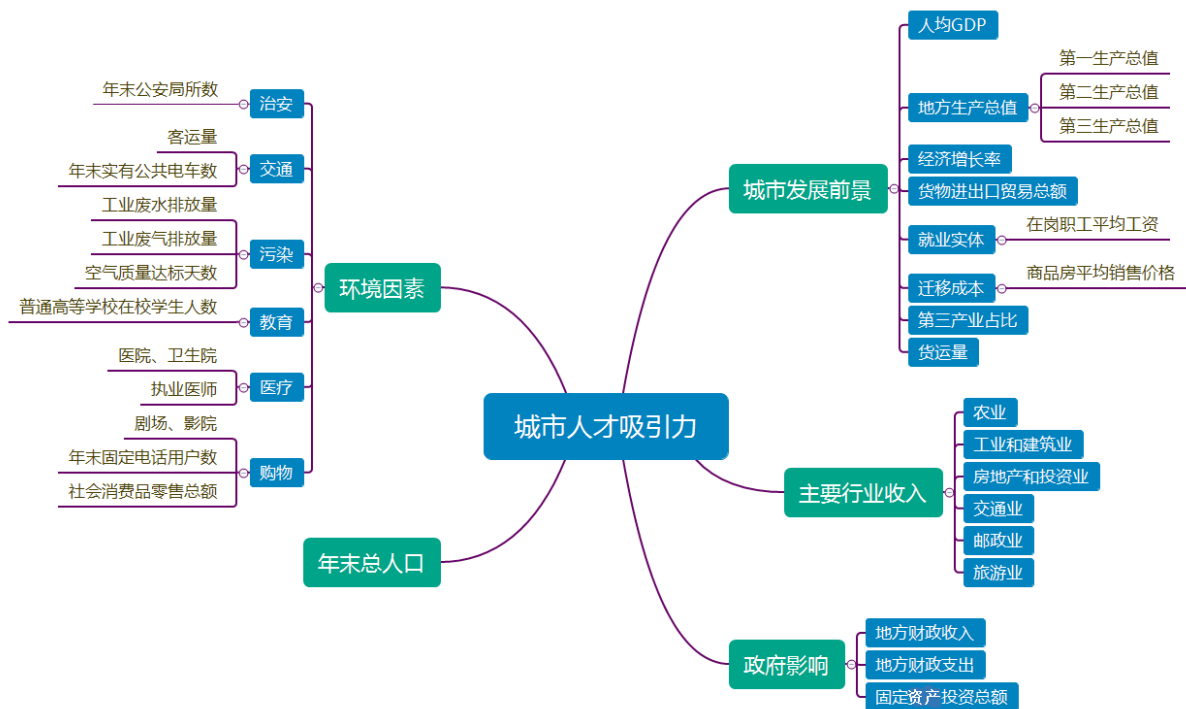


图1 城市人才吸引力指标体系

4.2.2 同向化处理

在评价城市人才的指标中，工业废水废气排放量和商品房平均销售价格不是越高越好，为方便比较，对这三个指标进行同向化处理 [1]，其公式如下：

$$\tilde{X} = -(X - \bar{X}) \quad (1)$$

式中：\$X\$ 原始指标，\$\bar{X}\$ 为原始指标 \$X\$ 的平均值，\$\tilde{X}\$ 为同向化指标。

4.2.3

为了对变量进行比较并消除由于观测量纲的差异及数量级差异所造成的影响，将样本观测数据进行标准化处理，使标准化的变量均值为 0，方差为 1。其处理方法如下：

$$Z = \frac{X - \bar{X}}{\delta_X} \quad (2)$$

式中：\$\bar{X}\$ 为 \$X\$ 的均值，\$\delta_X\$ 是 \$X\$ 的标准差。

4.2.4

根据标准化处理后的指标数据，利用 SPSS23.0 统计软件进行 KMO 和巴特利特检验，以确认所选变量是否适合做因子分析，结果如表所示：

表 1 KMO 和 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量		0.859
Bartlett 的球形度检验	近似卡方	1183.981
	df	78
	Sig.	.000

由上表可知，巴特利特球度检验统计量的观测值为 1183.981，相应的概率 P 值接近 0。若显著性水平 为 0.5，概率 P 小于显著性水平，应拒绝零假设，认为相关系数矩阵与单位阵有显著差异，即因子协方差矩阵不是单位阵。同时， KMO 值为 0.859， $KMO > 0.8$ ，根据 KMO 度量标准可知原有变量适合进行因子分析。

4.3 模型的建立与求解

4.3.1 公因子的确定

根据上述指标标准化后，其对应相关系数矩阵部分数据见表 2 所示（详见附录）：

表 2 相关系数矩阵部分数据

	X_1	X_2	X_3	X_4	X_{33}
X_1	1.000	0.727	-0.423	0.967	-0.928
X_2		1.000	-0.796	0.822	-0.814
X_3			1.000	-0.620	0.682
X_4				1.000	-0.955
.....				
X_{33}					1.000

设 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ 为样本相关系数矩阵 R 的特征值， $\eta_1, \eta_2, \cdots, \eta_p$ 为相应的标准正交化特征向量。设 $m < p$ ，则因子载荷矩阵 Λ 为：

$$\Lambda = \left[\sqrt{\lambda_1} \eta_1, \sqrt{\lambda_2} \eta_2, \cdots, \sqrt{\lambda_m} \eta_m \right] \quad (3)$$

用 $\mathbf{R} - \mathbf{\Lambda}\mathbf{\Lambda}^T$ 对角元来估计特殊因子的方差:

$$\sigma_i^2 = 1 - \sum_{j=1}^m \alpha_{ij}^2 \quad (4)$$

式中: σ_i 为方差, α_{ij} 表示载荷因子。由上述分析, 得总方差解释部分数据如表 3 所示:

表 3 总方差解释

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积	总计	方差百分比	累积
1	25.330	76.758	76.758	25.330	76.758	76.758
2	4.025	12.197	88.956	4.025	12.197	88.956
3	2.373	7.191	96.147	2.373	7.191	96.147
4	1.271	3.852	99.920	1.271	3.852	99.920
5	5.129e-15	1.554e-14	100.0			
6	1.330e-15	4.031e-15	100.0			
7	6.062e-16	1.837e-15	100.0			
8	5.067e-16	1.535e-15	100.0			
9	4.532e-16	1.373e-15	100.0			
10	3.335e-16	1.011e-15	100.0			
11	3.137e-16	9.506e-16	100.0			
12	2.585e-16	7.834e-16	100.0			
13	1.990e-16	6.031e-16	100.0			
.....	100.0			
.....	100.0			
33	-2.279e-15	-6.908e-15	100.0			

由表 3 特征根知, 因子 F_1 的特征值 $\lambda_1 = 25.330$, 占方差的 76.75%。由碎石图 2 可知, 当提取 1,2 个公因子时, 特征值变化非常明显, 当提取第 5 个以后的公因子时, 特

征值变化比较小，基本趋于平缓。由此说明，提取 4 个公因子对原变量信息的刻画有显著作用。因此，在这里我们提取 4 个公共因子，这 4 个公因子的累计方差达到 99.92%，即这 4 个公因子可以反映原来 33 个指标的 99.92% 的信息量，可见采用前 4 个公因子对武汉 2013—2017 年人才吸引力进行评价是比较合适的。



图 2 碎石图

4.3.2 因子命名与载荷矩阵计算

表 4 是初始因子载荷矩阵（详见附录），由此可写出因子分析模型的如下：

$$\begin{aligned}
 X_1 &= 0.989F_1 + 0.123F_2 - 0.074F_3 - 0.031F_4 \\
 X_2 &= 0.795F_1 - 0.556F_2 - 0.193F_3 - 0.151F_4 \\
 &\dots\dots\dots \\
 X_{33} &= -0.958F_1 + 0.176F_2 - 0.060F_3 + 0.220F_4
 \end{aligned}
 \tag{5}$$

表 4 初始因子载荷矩阵

指标	F_1	F_2	F_3	F_4
地方财政预算支出	0.989	0.123	-0.074	-0.031
固定资产总额	0.795	-0.556	-0.193	0.151
年末邮政局数	-0.548	0.768	-0.333	0.008
执业医师人数	0.994	-0.027	-0.245	0.002
.....
年末固定电话数	-0.958	-0.176	-0.060	0.220

表 4 中的每个数据表示了相应因子变量对相应原变量的相对重要程度。由于得到的公共因子与各指标的载荷分布归类比较困难，需要对因子载荷矩阵进行正交旋转，这里运用方差最大正交旋转法 [2]，得到旋转后的因子载荷矩阵表 5 所示（详见附录）：

表 5 旋转后的因子载荷矩阵

指标	F_1	F_2	F_3	F_4
地方财政预算支出	0.870	0.356	0.076	0.331
固定资产总额	0.469	0.311	0.678	0.473
年末邮政局数	-0.291	-0.390	-0.867	0.105
执业医师人数	0.787	0.496	0.259	0.259
.....
年末固定电话数	-0.872	-0.277	-0.384	-0.124

根据表 5 发现，旋转在 9 次迭代后已收敛。旋转后的因子系数已经明显向两极分化，有了更鲜明的实际意义。因子载荷的绝对值越大，则表明该因子与变量的重叠性越高，在解释因子的时候就越重要。

第一因子 F_1 主要包含地区生产总值、第一第二产业生产总值、在岗职工平均工资和人均 GDP 等，这些指标包含了地区工业发展水平以及人均薪酬水平故将其命名为工业发展与薪酬因子。第二因子 F_2 主要包含持证医师人数，医院卫生院个数以及一年内

空气质量达到及好于二级的天数等，这些指标包含了地区医疗卫生水平和居民生活环境，故将其命名为**医疗卫生环境因子**。第三因子 F_3 主要包含固定资产投资总额与货物进出口总额，该指标反映了城市贸易与经济发展水平，故将其命名为**经济贸易因子**。第四因子 F_4 包含年末总人口与旅游生产总值，反映了城市人口与拥挤程度故将其命名人口与拥挤程度因子。

4.3.3 求得因子得分和综合绩效得分

采用回归法估计因子得分系数，并输出因子得分系数矩阵，其结果见表 6 所示（详见附录）：

表 6 成分得分系数矩阵

指标	F_1	F_2	F_3	F_4
地方财政预算支出	0.052	-0.017	-0.021	0.024
固定资产总额	-0.063	0.045	0.127	0.137
年末邮政局数	0.022	-0.079	-0.195	0.109
执业医师人数	0.007	0.064	0.021	0.009
.....
年末固定电话数	-0.109	0.093	-0.068	0.094

则根据表 6 可以计算出四个主因子得分，表达式分别为：

$$\begin{aligned}
 F_1 &= 0.052x_1 - 0.063x_2 + 0.022x_3 + 0.007x_4 + \cdots - 0.109x_{33} \\
 F_2 &= -0.017x_1 + 0.045x_2 - 0.079x_3 + 0.064x_4 + \cdots + 0.093x_{33} \\
 F_3 &= -0.021x_1 + 0.127x_2 - 0.195x_3 + 0.021x_4 + \cdots - 0.068x_{33} \\
 F_4 &= 0.024x_1 + 0.137x_2 + 0.109x_3 + 0.009x_4 + \cdots + 0.094x_{33}
 \end{aligned} \tag{6}$$

利用 SPSS 软件计算得到旋转后的因子贡献及贡献率如表 7 所示：

表 7 旋转后因子贡献及贡献率

因子	贡献	贡献率	累计
F_1	18.238	55.267	55.267
F_2	5.889	17.845	73.112
F_3	4.826	14.623	87.735
F_4	40.470	12.265	99.920

由此计算出的因子得分，可以量化描述城市人才吸引力水平，利用因子得分可以从不同角度对城市人才吸引力水平进行比较分析。为了便于对各城市进行人才吸引力评价，现利用武汉每年的因子得分表计算综合得分，吸引力水平的获取是基于总方差分解表中旋转后各因子的方差贡献率及计算所得的城市各因子得分获取的。其计算公式如下：

$$F = (55.267F_1 + 17.845F_2 + 14.623F_3 + 12.265F_4) / 99.92 \quad (7)$$

经计算，纵向比较武汉每年的综合因子得分后绘制直方图：

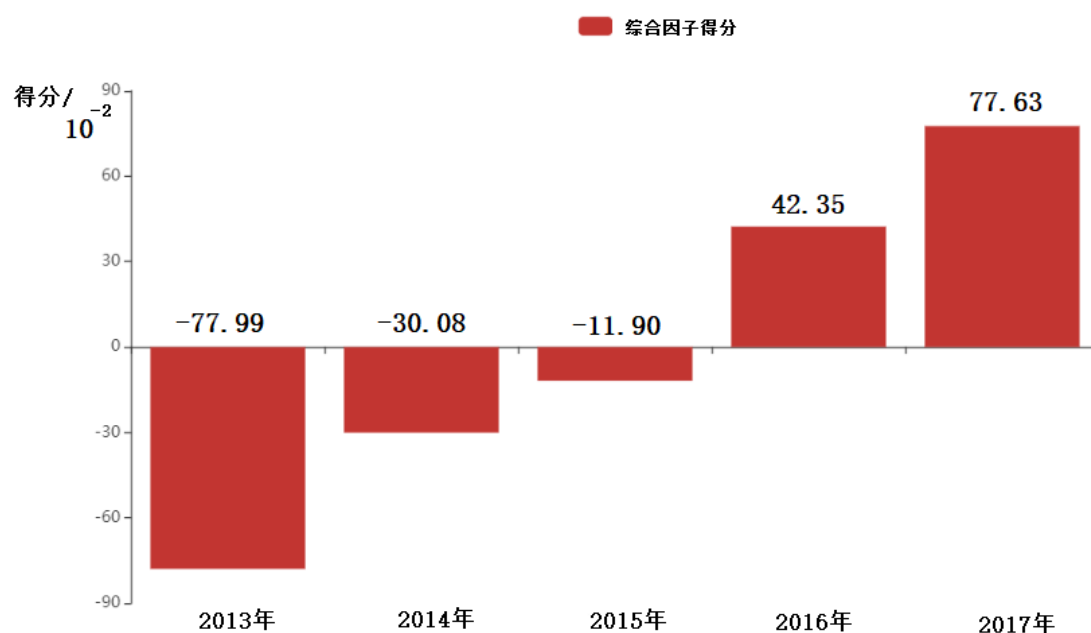


图 3 综合因子得分

4.4 结果分析

(1) 人才吸引力的 33 个指标经因子分析和主成分分析后综合成的 4 个公因子, 由回归法求得的图 6、7 所示, 4 个公因子的累计方差达到 99.92%, 反应的信息量达到比较合适的比例。

(2) 根据武汉每年的因子得分表计算构建出的 2013-2017 年武汉综合因子得分表表明武汉市对于人才吸引力水平整体呈上升趋势, 基本符合近几年武汉人才吸引力情况, 说明武汉市对人才的吸引力随时间呈正向发展。

(3) 由表 4 至表 7 的分析数据显示, 影响人才吸引力最主要因素是工业发展与薪酬因子, 其次是医疗卫生环境因子, 再者是经济贸易因子, 最后是拥挤程度因子。从中可以看出, 吸引人才的首先是一个地区发展前景和薪资水平的高低, 虽然地区交通拥挤程度也很重要, 但对于个人而言, 医疗卫生环境和经济贸易发展更为关键。

(4) 由表 5 可以看出, 武汉市经济贸易因子和工业发展与薪酬因子逐年提高, 这一方面与武汉市提出的《武汉市人力资源和社会保障事业发展“十三五”规划》有关, 另一方面也与武汉市推进人才引进落户政策有关。近几年武汉市医疗卫生环境因子整体呈上升趋势, 这归功于武汉市政府进一步加大对于专业技术人员的教育力度。

五、 问题二模型的建立与求解

5.1 问题的描述与分析

问题二要求针对具体人才类别, 量化分析比较武汉市与其它同类城市的人才吸引力。本文在问题一模型的基础上改进, 将人才分为三类——第一产业人才、第二产业人才和第三产业人才, 并将问题一中的人才吸引力因素通过灰色相关度分成针对各类型人才的主要影响因素, 再对分类后数据进行因子分析得到成都、天津、西安、南京和武汉 2013—2017 年分别对于三类人才吸引力的因子分析模型, 并计算每个城市每年针对于三类人才的人才吸引力综合得分。其分析流程图如图 4 所示:

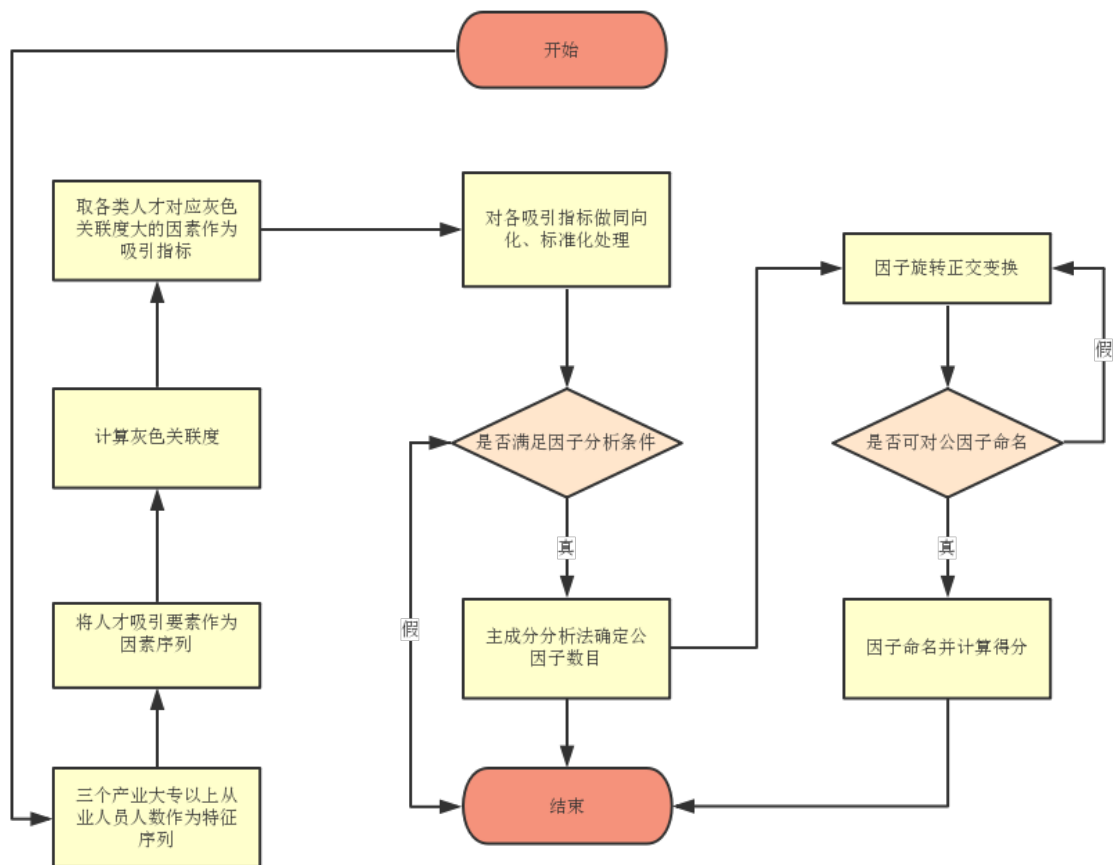


图 4 问题二分析流程图

5.2 模型的建立与求解

5.2.1 人才及其吸引力影响因素的分类

不同产业的人才会有不同的关注焦点，所以城市对不同产业类型的人才吸引力影响因素也存在差异。我们把人才分为三个类型——第一产业人才、第二产业人才和第三产业人才。通过查阅资料 [3] 收集了武汉 2013—2017 年第一、第二和第三产业大专以上学历从业人员人数，所收集数据得到表 8 如下：

表 8 各人才类别从业人数

年份	第一产业人才	第二产业人才	第三产业人才
2013	3800	1022800	958800
2014	3571	1039720	980144
2015	3577	1046760	1022431
2016	3466	1053081	1076044
2017	3229	1037628	1158174

利用灰色关联度理论分别分析三类人才的 33 个影响因素，得到每个影响因子与用电量之间的关联度系数，具体步骤如下：确定因素序列和特征序列。本文分别以武汉 2013—2017 年第一、第二和第三产业大专以上学历从业人员人数作为特征序列，设为 $x_k(p)$ ，采用 $n(n = 15)$ 个数据： $x_k(p) = x_k(1), x_k(2), \dots, x_k(5)(k = 1, 2, 3)$ ；将 33 个人才吸引要素作为因素序列，设为 $x_i(t)$ ，其中有 $m(m = 33)$ 个子序列，每个子序列对应 5 个数据： $x_i(p) = x_i(1), x_i(2), \dots, x_i(5)$ 。

计算两极最小差 k 和最大差 K 。计算出特征序列与因素序列的序列差为 $\Delta_i(t)$ ，找到结果中的最小差和最大差。其中：

$$\begin{aligned}\Delta_1(t) &= |x_0(t) - x_1(t)| \\ k &= \min \min \Delta_i(t) \\ K &= \max \max \Delta_i(t)\end{aligned}\tag{8}$$

灰色关联系数和关联度。因素序列和特征序列在第 T 点的关联系数为：

$$\Phi_{0,(p)} = \frac{k + \varepsilon K}{\Delta_{0i}(t) + \varepsilon K}\tag{9}$$

其中 ε 为分辨系数，取 $\varepsilon = 0.5$ 。

利用 MATLAB 软件分析计算后，取各类人才灰色关联度大小前十的影响因素作为其主要影响因素，如表 9 所示：

表 9 各人才类别从业人数

第一产业	关联度	第二产业	关联度	第三产业	关联度
经济增长率	0.897	农业	0.862	第三产业占比	0.978
旅游业	0.874	第二产业	0.807	邮政业	0.958
第三产业占比	0.840	客运量（万人）	0.787	旅游业	0.933
固定资产总额	0.827	商品房销售价格	0.771	年末电话用户	0.925
房地产和投资业	0.786	工业和建筑业	0.756	剧场、影剧院	0.922
职工平均工资	0.728	执业/助理医师	0.748	地区生产总值	0.921
人均 GDP（元/人）	0.719	固定资产总额	0.742	年末邮政局数	0.915
在岗职工平均工资	0.711	年末邮政局数	0.737	固定资产总额	0.907
地方财政内收入	0.643	地方财政内支出	0.736	货运量（万吨）	0.902
地方财政内支出	0.643	邮政业	0.726	客运量（万人）	0.887
地区生产总值	0.615	消费品零售总额	0.716	地方财政内收入	0.880

5.2.2 分类型人才吸引力分析

将归类后所得的影响第一产业人才吸引力的 10 个影响因素在五个城市中 2013—2017 年的数据，使用 SPSS 进行因子分析，按照累积贡献率的原则提取公因子，以综合性指标来相对全面的反映出全体因子对第一产业人才吸引力的影响情况。按照主因子的提取原则，通过碎石图可看出可以用 3 个主因子来描述此 10 个因子的影响；再通过旋转后因子贡献率计算所得的城市各因子得分得到各城市每年的第一产业人才吸引力综合得分。其计算公式如下：

$$F = (48.328F_1 + 28.667F_2 + 12.283F_3) / 89.277 \quad (10)$$

将得到的五个城市 2013—2017 年的综合得分绘制为折线图的到图 5 如下所示：

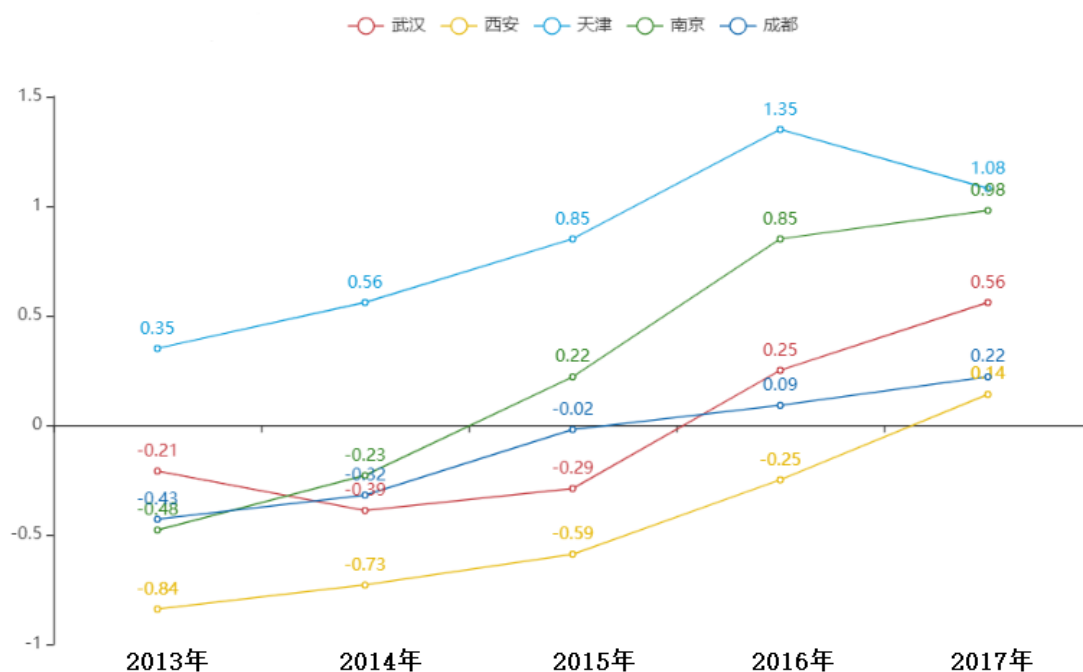


图 5 第一产业人才吸引力综合得分折线图

同理将归类后所得的影响第二产业人才吸引力的 10 个影响因素使用 SPSS 进行因子分析，得到第二产业人才吸引力综合得分计算公式：

$$F = (27.594F_1 + 26.960F_2 + 16.609F_3 + 16.356F_4) / 87.519 \quad (11)$$

第二产业人才吸引力综合得分折线图：

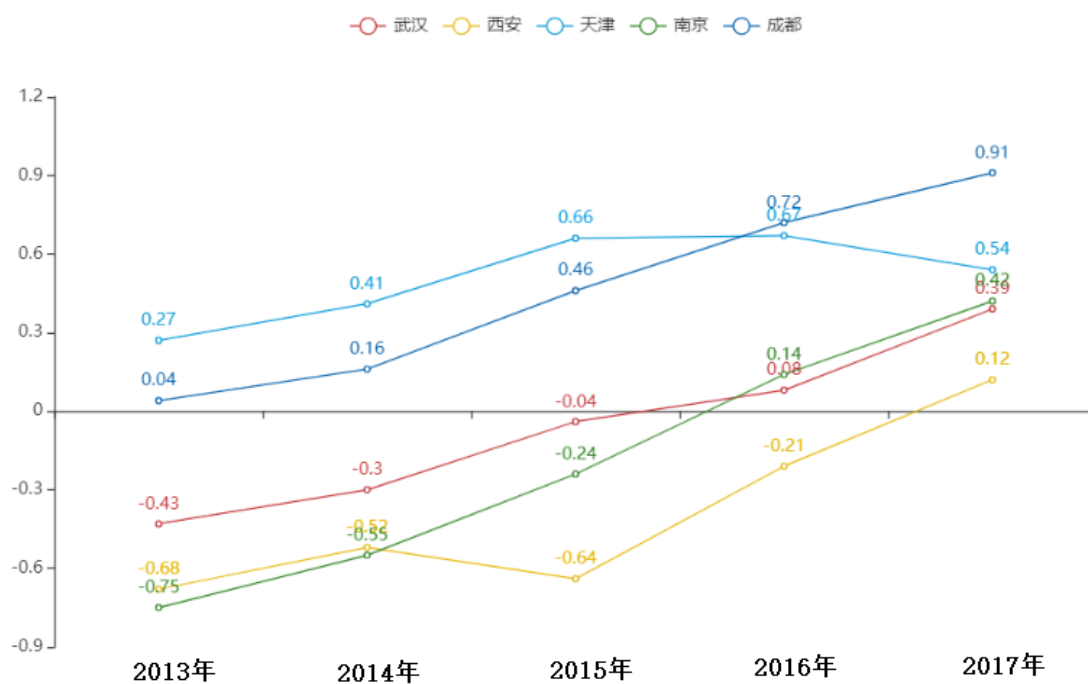


图 6 第二产业人才吸引力综合得分折线图

同理将归类后所得的影响第三产业人才吸引力的 10 个影响因素使用 SPSS 进行因子分析，得到第二产业人才吸引力综合得分计算公式：

$$F = (37.063F_1 + 25.130F_2 + 11.972F_3 + 10.477F_4) / 84.642 \quad (12)$$

第三产业人才吸引力综合得分折线图：

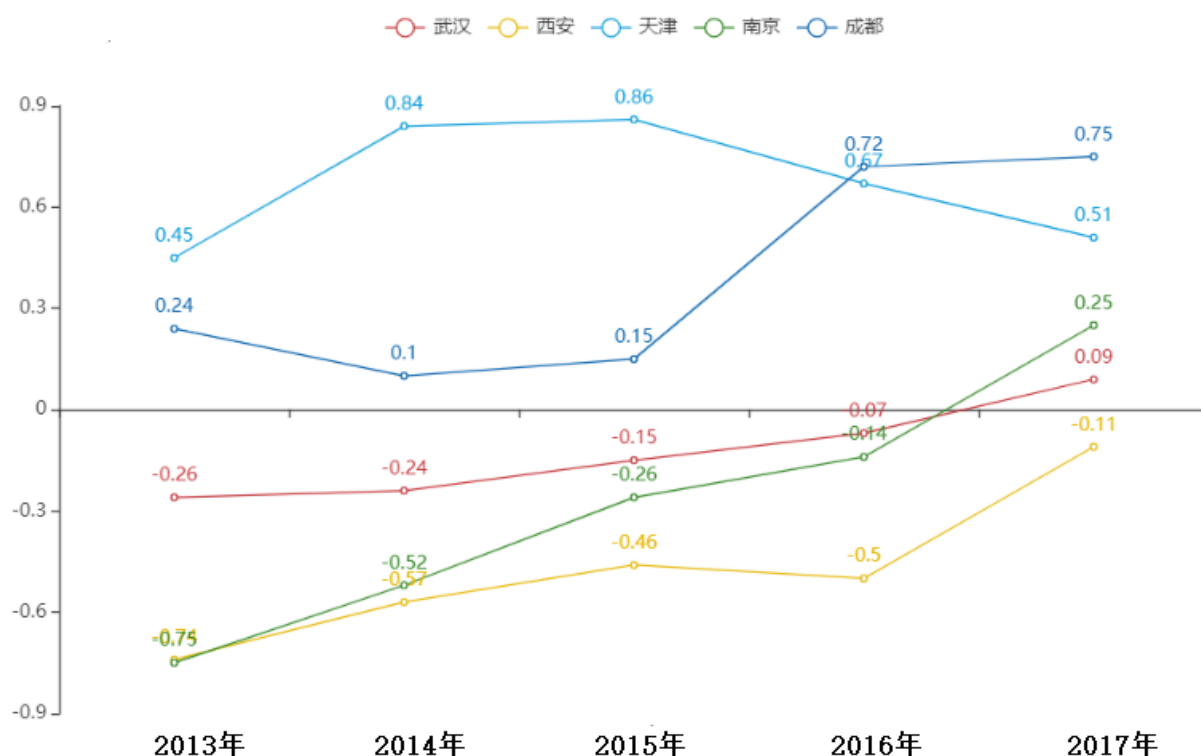


图 7 第三产业人才吸引力综合得分折线图

5.3 结果分析

由图表可以看出，武汉市的人才吸引力水平近几年有较高的水平，尤其是对第二、三产业人才的吸引力表现出逐年增加趋势和较高水平，相对于西安和成都而言，武汉市具有较强的人才吸引力，十分符合地区的经济发展特点。而天津作为老牌北方直辖市，近年发展速度相比略显缓慢，导致其相应人才吸引力也有所下滑。

但是，武汉在发展中也存在一些不足。特别是工业发展与薪酬因子处于人才吸引主导地位，而环境方面的因素尚未完全建立；优质教育与医疗资源分配不均、生活成本高、高校科研机构偏少等因素限制良好人才环境构建等。

提高武汉市的人才吸引力水平，提出以下几点建议：

- (1) 大力发展地方生产总值，为经济社会发展打下坚实的基础。
- (2) 主要行业是加快人才发展的主导力量。建立较为高效的人才市场，通过各行业市场机制，人才较为充分地实现了自身价值，也为武汉的人才持续发展创造了良好条件。

- (3) 注重整合空间结构全面发展，引导各大区块均衡协调稳步发展，建设全面型社会，注重生态建设，贯彻落实以人为本的科学发展理念。
- (4) 加大营商环境改革力度，营造更加开放的贸易投资环境，营造综合成本适宜的产业发展环境，营造更加高效透明的政务环境。
- (5) 政府不断完善就业服务政策，出台大量相关文件，高度重视高校毕业生的就业情况。

六、 写给武汉市人力资源部门的建议报告

七、 模型的评价

7.1 模型的优点

- (1) 从国家统计数据库、武汉市统计年鉴和其他同类城市的统计年鉴中爬取大量数据，并挑选符合实际的 33 个指标与最近五年的数据进行评价。其吸引人才指标从多维度、全方面的考虑，具备科学性、客观性。
- (2) 针对具体人才根据相应产业进行分类。相较于现有城市人才吸引力水平评价模型，我们采用了灰色关联分析法，对三种人才类别分析计算出与指标的相关系数，求出不同人才对 33 种影响因素的的偏好程度，选出前十个合适的指标进行综合打分。

7.2 模型的缺点

未能具体地量化政府政策对人口吸引力的影响，只选取了地方财政收入、支出，固定资产投资总额三个影响因素作为政策影响的指标，与现实情况出现一定偏差，存在一定的局限性。

参考文献

- [1] 张瑞红. 河南省产业集群环境人才吸引力评价研究. 科技管理研究, 32(10):180–184, 2012.
- [2] 张培利宋鸿. 城市人才吸引力的影响因素及提升对策. 湖北社会科学, 2010(2):43–45, 2010.
- [3] 武汉统计局. <http://tjj.wuhan.gov.cn/newslist.aspx?id=2018061817071264>. Website. 武汉市国民经济和社会发展统计公报.
- [4] Meric S Gertler, Kate Geddie, Carolyn Hatch, and Josephine Rekers. Attracting and retaining talent: Evidence from canada’ s city-regions. *Seeking talent for creative cities: The social dynamics of innovation*, pages 3–30, 2014.
- [5] Josh Lepawsky, Chrystal Phan, and Rob Greenwood. Metropolis on the margins: talent attraction and retention to the st. john’s city-region. *The Canadian Geographer/Le Géographe canadien*, 54(3):324–346, 2010.
- [6] EPS 数据平台. <http://olap.epsnet.com.cn/auth/platform.html>. Website. 中国宏观经济数据库.

附录 A 代码

A.1 数据预处理–python 源代码

```
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from factor_analyzer import FactorAnalyzer

#导入数据
df = pd.read_excel("wut.xls")
# print(df.head(5))

#转置
df = pd.DataFrame(df.values.T, index=df.columns, columns=df.index)
# print(df['商品房平均销售价格（元/平方米）'])

数据正向处理
df['商品房平均销售价格（元/平方米）'] =
    -(df['商品房平均销售价格（元/平方米）']-sum(df['商品房平均销售价格（元/平方米）'])/5)
df['工业废水排放量（万吨）'] = -(df['工业废水排放量（万吨）']-sum(df['工业废水排放量（万吨）'])/5)
df['工业二氧化硫排放量（吨）'] =
    -(df['工业二氧化硫排放量（吨）']-sum(df['工业二氧化硫排放量（吨）'])/5)
print(df['商品房平均销售价格（元/平方米）'])
print(df.head(5))
df.to_excel("武汉.xls")

#数据标准化
df = StandardScaler().fit_transform(df)
# print(df)
df=pd.DataFrame(df)
df.to_excel("武汉.xls")
#
#主成分分析
pca = PCA(n_components=4)
newX = pca.fit_transform(df)
print(newX)

#返回所保留的n个成分各自的方差百分比
print("它代表降维后的各主成分的方差值占总方差值的比例，这个比例越大，则越是重要的主成分")
print(pca.explained_variance_ratio_)
print("它代表降维后的各主成分的方差值，方差值越大，则说明越是重要的主成分")
print(pca.explained_variance_)
```