

# Research on Urban Talent Development Based on PCA/SVM Model

Xingzhi Chen, Lizhi Yang, Jia Que, Siying Long

School of Science, Southwest University of Science and Technology, Mianyang Sichuan  
Email: 1451025961@qq.com

Received: Nov. 27<sup>th</sup>, 2019; accepted: Dec. 11<sup>th</sup>, 2019; published: Dec. 18<sup>th</sup>, 2019

---

## Abstract

The cultivation and introduction of human resources is of great significance to the development and prosperity of the regional economy. Based on the research of talent development in urban development, this paper establishes a support vector machine model to predict the potential talent demand in the future. And using the method of objective evaluation-entropy method to give weights to each variable, establish an evaluation model of talent attraction, and quantitatively evaluate the level of talent attraction in a city. The results show that this paper predicts the number of potential talents attracted by urban development each month. The number of people needs to change periodically when the industry is not stable. And the level of talent attraction has been drawn. Finally, this paper provides some suggestions and strategies for talent training for solving the results.

## Keywords

PCA Model, SVM Model, Correlation Index, Talent Attraction Evaluation Model, Entropy Method

---

# 基于PCA模型和SVM模型的城市人才发展研究

陈兴志, 杨栎智, 阙佳, 龙思颖

西南科技大学理学院, 四川 绵阳  
Email: 1451025961@qq.com

收稿日期: 2019年11月27日; 录用日期: 2019年12月11日; 发布日期: 2019年12月18日

---

## 摘要

对人力资源的培养和引入, 对地区经济的发展与繁荣具有十分重要的意义。本文根据城市发展对人才需求的研究, 建立支持向量机模型预测未来潜在的人才需求。并采用客观评价的方法——熵值法赋予各变量权重, 建立了人才吸引力的评价模型, 量化评价一个城市的人才吸引水平。结果显示, 本文预测出了

每个月城市发展对潜在人才吸引的数量, 在开始行业未稳定时人数需求是呈周期性变化。以及得出了城市的人才吸引水平。最后本文针对求解结果, 提供了有关人才培养的一些建议和策略。

## 关键词

PCA模型, SVM模型, 相关指数, 人才吸引力评价模型, 熵值法

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

人才是城市创新和发展的动力, 创新人才能够很快地掌握最新的知识和信息, 特别是掌握知识生产的能力, 以及信息采集、处理、分析、传输与使用能力, 这成为了城市取得发展优势的关键。但是, 目前我国人才资源结构存在专业结构不合理、行业分布不合理等问题。为此, 分析和研究城市就业市场的人才需求对解决人才结构布局不合理、人才的短缺以及城市建设和创新传播等方面具有十分重要的意义。

以往的研究中, 许多学者分别对人才发展进行了不同的研究。在国内, 2003年, 陆晓芳等人[1]建立了评价指标体系, 运用主成分分析法建立了人才要素区域竞争力的评价模型。2006年, 兰艳章等人[2]利用熵值法对人才甄选综合评价中的权重系数问题进行客观赋值, 提出基于熵权值的多层次人才甄选评价法。2008年, 胡瑞卿等人[3]构造了集“模糊综合测评”和“主成分分析测评”优点于一体的“科技人才合理流动程度的模糊层次主成分分析测评模型”。2010年, 黄凤等人[4]构建以品质、学识、能力为核心指标的管理型人力资本价值评估指标体系, 采用层次分析法和熵值法对管理型人力资本价值进行定量测量。2011年, 肖利哲等人[5]构建了由基础层、核心层和结果层组成的多方案人才战略决策模型。李健飞等人[6]运用劳动经济学、多元统计学、社会学等有关理论, 构建了我国不同地区科技人才吸引力评价指标体系及测评模型。2012年, 李朋林等人[7]运用 SVM 方法构建了陕西省人才资源与经济增长的关系模型。2013年, 杨焕海等人[8]提出一种 ARIMA-LSSVM 的人才需求组合预测模型。2016年, 杨璐瑶等人[9]研究我国创新型人才发展当前存在的人才培养、选拔、引进、激励等制度问题。2019年, 高建祥等人[10]研究了城市人才吸引力的评价标准, 深入论证了城市各项指标对人才的吸引力影响。在国外, 1996年, Horowitz 等人[11]推导了一个多方程计量经济模型, 以确定科学人才的空间分布反过来在区域经济活动中的体现程度。2009年, Faggian A 等人[12]使用来自 110 个国家/地区的 1569 个次国家区域的新建数据库, 调查了区域发展的决定因素, 提出了一种新的区域发展模型。2011年, An 等人[13]对区域人才生态环境综合评价体系进行研究。2012年, Xiu-Jun L 等人[14]总结了区域人才开发战略的定义、内涵与结构、模式和研究水平的主要观点和区域人才开发战略研究中存在的问题。2013年, Martindale R J J 等人[15]研究了在通过判别函数分析来调查问卷在现实世界中的适用性。2017年, Cadornin E 等人[16]分析了四个与人才有关的案例, 并且证实了科学园区与学生社区的紧密联系以及与具有公认品牌的国际网络的联系的重要性。2011年, Mellander C 等人[17]围绕着两个核心问题: “有一个如何最好地衡量人力资本的问题”和“关于产生人力资本地域分布的因素”展开了辩论。2015年, HUO Jingbo 等人[18]建立了人才流入区与流出区之间的不对称演化博弈模型。

本文根据城市发展对人才的需求和 A-City 的 Total demand (Pers.)与 Education requirements 之间的关系。建立了支持向量机模型预测未来三年 A-City 潜在的人才需求。并采用客观评价的方法——熵值法赋

予各变量权重,建立了人才吸引力的评价模型,计算 A-City 和上海市的得分来量化评价 A-City 的人才吸引水平。陆晓芳等人[1]、胡瑞卿等人[3]、黄凤等人[4]、李健飞等人[6]、高建祥等人[10]、An 等人[13]通过建立不同的人才发展评价模型,对城市或区域发展与人才之间的引进做出相应的评价和讨论。但是缺乏对未来人才数量的预测和潜在人才需求的研究。兰艳章等人[2]研究了人才的甄选评价,对人才甄选综合评价中的权系数问题进行客观赋值,但却没有解决城市该如何引进人才的问题。肖利哲等人[5]、李朋林等人[7]、杨焕海等人[8]、杨璐瑶等人[9]、Faggian A 等人[12]、Xiu-Jun L 等人[14]、HUO Jingbo 等人[18]通过对政策的研究和区域人才引进战略的讨论,建立了相关人才需求组合预测模型,对地区的发展对人才的需求做出了预测。但却忽略了对人才吸引力水平的评价和对潜在人才需求的研究和讨论。没有结合地区的发展对人才需求进一步深入的探讨。对城市就业市场的人才需求进行研究需要从工作需求,预期职业和所需教育背景等方面考虑,对不同职业,不同学历和教育背景的人才分别进行研究和讨论,并预测未来几年内的潜在人才需求。并根据城市现阶段的人才需求和学生就业情况出发,结合城市的刑侦类别和所在地域以及发展情况、发展前景和产业布局、政策规划,对人才的需求作出进一步的讨论和评价。本文建立了支持向量机模型预测未来三年 A-City 潜在的人才需求。并采用客观评价的方法——熵值法赋予各变量权重,建立了人才吸引力的评价模型,计算 A-City 和上海市的得分来量化评价 A-City 的人才吸引水平。从工作需求、预期职业和所需教育背景等方面考虑,对城市就业市场的人才需求建立支持向量机模型,来预测未来三年 A-City 潜在的人才需求。并结合目前专业人才发展计划中存在的一些问题,提出针对性对策建议,以确保未来市场专业技术人才供需平衡,提升队伍建设水平。

## 2. 问题分析

本文结合 2018 年第八届 APMCM 亚太地区大学生数学建模竞赛 B 题[19]背景和题目所给数据对城市发展及人才需求进行了研究和讨论。并通过建立相关模型对 2018 年第八届 APMCM 亚太地区大学生数学建模竞赛 B 题[19]题目进行求解。

1) 根据附件给出的数据进行分析,从工作需求,预期职业和所需教育背景等方面考虑,对城市就业市场的人才需求进行建模与分析。

a) Education requirements 指标的分析

所给附件中提供了 Total demand (Pers.)的具体人数以及影响 Total demand (Pers.)的指标——Education requirements。本文根据其学历要求,将其重新划分为 4 个 Education requirements 指标,即为 Above Bachelor degree, Bachelor degree, Below Bachelor degree, Unlimited。具体划分结果如表 1 所示:

**Table 1.** Reclassification of education requirements

**表 1.** Education Requirements 的重新分类表

				Education requirements				
				Below Bachelor degree	Bachelor degree	Above Bachelor degree		Unlimited
Junior middle school	Senior middle school	Technical secondary school	Junior college	Bachelor's degree	Master's degree	Doctor's degree	MBA	Unlimited

b) Sector 指标的分析

题目所给的数据中提供了大量的 Sector 的种类,含有 Computer software, Sales management, market/marketing, Sales, Computer hardware 等共 50 种行业。本文根据“ISICRev4.0 2016.9”,将其分为 18 类,具体的关系如图 1 所示:

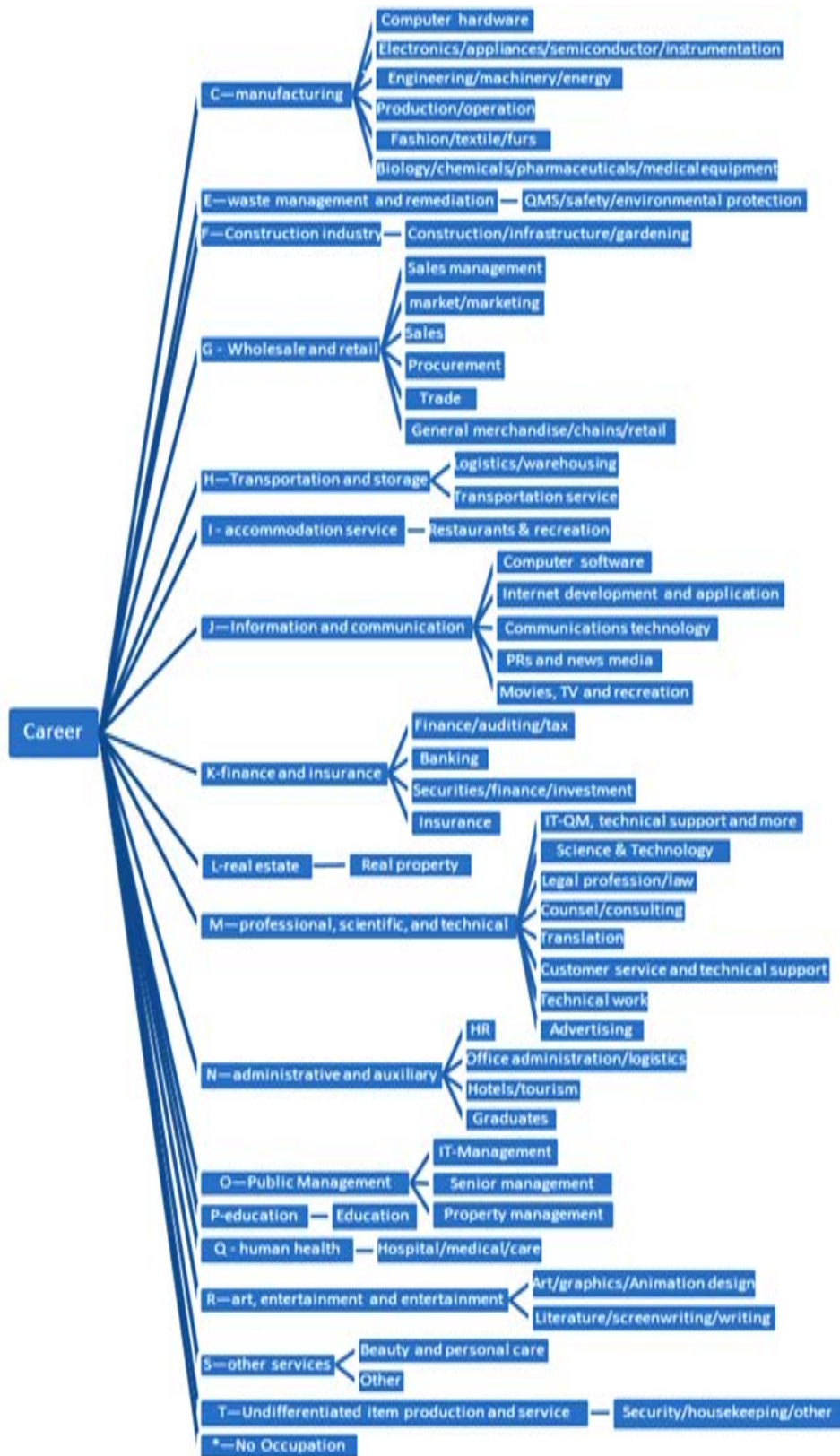


Figure 1. Occupation classification diagram  
图 1. 职业分类关系图



2) 根据附加数据以及 A 城市现阶段的人才需求和学生就业情况, 建立 A 市人才需求模型, 并预测未来三年的人才需求。

由于人才需求量具有周期性、多元化等特点, 为了预测未来三年的潜在人才需求量, 本文利用支持向量机(SVM), 建立了支持向量机模型, 实现了高效的从训练样本到预报样本的推理。

3) 运用问题 2 中的数据和结论来推断 A 城的行政类别, 可能的地理区域, 经济状况和高科技产业发展。

本文结合第一二题中求解的结果。采用最短距离法推测诊断结果, 从而可以得到 A-City 的行政类别, 可能的地理区域, 经济地位和高科技产业发展。

4) 对如今大学毕业生的就业机会多样化现象进行建模和量化, 为 A 城的发展和人才引进提供策略。

本文选取五个指标, 分别是: 发展前景  $A_1$ 、生活水平  $A_2$ 、主要行业增长率  $A_3$ 、城市环境  $A_4$ 、政策影响  $A_5$ , 建立了人才吸引水平模型。采用克隆巴赫系数来衡量。该系数愈高, 各个指标间的信度愈高, 即一致性越好。信度只要达到 0.70 就可接受, 介于 0.70~0.98 均属高信度, 而低于 0.35 则为低信度, 必须予以拒绝。

本文选取评价模型中的客观赋权法——熵值法, 利用熵值法算出的各指标量的权重, 再用各指标量与其权重的乘积和计算得分, 量化吸引力水平。为了更好、更客观的体现打分的可靠性, 选取上海市作为参照。为城市发展和人才引进提供策略。

### 3. 模型的介绍

#### 3.1. 人才需求 PCA 数学模型

1) 对原始数据进行标准化处理

由于各指标代表的性质不同, 因而其计量单位一般也是不同的。为了消除不同单位的影响, 使各个指标具有可比性, 以利于主成分的含义解释, 通常要对搜集的原始数据进行标准化处理。所以对这 Above Bachelor degree, Bachelor degree, Below Bachelor degree, Unlimited 四个指标的观测数据矩阵为:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (1.1.1)$$

其中  $p=4, n=18$ 。那么可以对原始数据进行标准化处理:

$$\begin{cases} x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{Var(x_j)}} \\ \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \\ Var(x_j) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \\ i=1, 2, \dots, n; j=1, 2, \dots, p \end{cases} \quad (1.1.2)$$

2) 计算样本相关系数矩阵

假定原始数据标准化后仍然用  $X$  表示, 则经标准化处理后数据的相关系数为:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{np} \end{bmatrix} \quad (1.1.3)$$

其中

$$r_{ij} = \text{COV}(x_i, x_j) = \frac{\sum_{k=1}^{k=n} (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{n-1}, n > 1 \quad (1.1.4)$$

3) 计算相关系数矩阵  $R$  的特征值和相应的特征向量

特征值:  $\lambda_1, \lambda_2, \dots, \lambda_p$ 。

特征向量:  $a_i = (a_{i1}, a_{i2}, \dots, a_{ip}), i = 1, 2, \dots, p$

4) 选择重要的主成份, 并写出主成份表达式

由主成份分析可以得到  $p$  个主成份, 但是由于各个主成份的方差是递减的, 包含的信息量也是递减的, 所以实际分析时, 根据贡献率选取前  $k$  个主成份。贡献率  $\alpha$  指的是某个主成份的方差占全部方差的比重, 即 Above Bachelor degree, Bachelor degree, Below Bachelor degree, Unlimited 所对应的特征值占全部特征值合计的比重:

$$\alpha = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \quad (1.1.5)$$

贡献率  $\alpha$  越大, 说明该主成份所包含的原始变量的信息越多。

5) 计算主成份得分

根据标准化的原始数据, 按照各个样品, 分别代入主成份表达式, 就可以得到各主成份下的各个样品的新数据, 即主成份得分。具体形式如下:

$$\begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1k} \\ F_{21} & F_{22} & \cdots & F_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ F_{n1} & F_{n2} & \cdots & F_{nk} \end{bmatrix} \quad (1.1.6)$$

其中:  $F_{ij} = a_{j1}x_{i1} + a_{j2}x_{i2} + \cdots + a_{jp}x_{ip}, i = 1, 2, \dots, n; j = 1, 2, \dots, k$ 。

### 3.2. 分析预测潜在人才的支持向量机模型

支持向量机最初用于数据分类问题。在数据分类中, 对于一组数据寻求一个满足分类要求的超平面, 使训练集中的点依照这个平面正确地分为两类, 同时使分开的数据点距离分类超平面最远, 这便是支持向量机的核心思想。

在图 2 中, 圆和三角形分别代表两类数据, 中间虚线  $p_0$  为分类超平面,  $\omega$  为其法向量,  $b$  为偏置。距离分类超平面最近的样本点为支持向量, 过支持向量且平行于分类超平面的平面, 称为分类边界, 它们之间的距离为分类间隔  $s$ 。

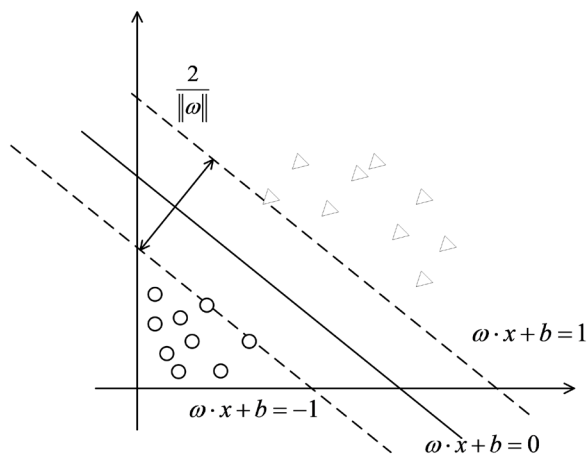


Figure 2. Schematic diagram of the classification boundary of the support vector machine  
图 2. 支持向量机的分类边界示意图

假设给定的数据样本集为  $(x_i, y_i), i = 1, \dots, n, x \in R^n, y \in \{+1, -1\}$  则分类超平面  $p_0$  为:  $\omega \cdot x + b = 0$ , 分类边界  $p_1, p_2$  为:  $\omega \cdot x_1 + b = 1$  或  $\omega \cdot x_2 + b = -1$ 。归一化的样本数据满足:

$$y_i \cdot [(\omega \cdot x_i) + b] \geq 1 \tag{2.1.1}$$

其中分类间隔为:

$$s = \frac{\omega}{\|\omega\|} \cdot (x_1 - x_2) = \frac{2}{\|\omega\|} \tag{2.1.2}$$

即  $p_1, p_2$  之间的间隔为  $\frac{2}{\|\omega\|}$ , 寻求最优超平面意味着最大化  $\frac{2}{\|\omega\|}$ 。于是寻求最优超平面的问题转化为求解二次规划的问题:

$$\begin{aligned} \min(\omega) &= \frac{1}{2} \cdot \|\omega\|^2 \\ \text{s.t.} &\begin{cases} y_i \cdot [(\omega \cdot x_i) + b] \geq 1 \\ i = 1, 2, \dots, n \\ y_i = \pm 1 \end{cases} \end{aligned} \tag{2.1.3}$$

由于实际的数据集中, 数据有的仅近似线性可分甚至非线性可分。对于近似线性可分, 允许极少数点不满足以上分类要求, 对于这种易被误判的点, 加入弱化分类要求的松弛变量  $\xi_i$  和误判时的惩罚因子  $C$ 。则原二次规划变为:

$$\begin{aligned} \min(\omega, \xi_i) &= \frac{1}{2} \cdot \|\omega\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} &\begin{cases} y_i [(\omega \cdot x_i) + b] \geq 1 - \xi_i, \\ \xi_i \geq 0, i = 1, 2, \dots, n \end{cases} \end{aligned} \tag{2.1.4}$$

其中  $C$  是一个大于 0 的常数, 可通过参数调优确定。

而在非线性支持向量回归机中, 可以通过非线性映射到高维空间, 将非线性问题转化为高维空间中的线性或近似线性问题。与分类问题不同, 此处  $y_i$  并不限定取 -1 或 1, 而是可以取任意实数。即寻找  $R^n$  的一个函数  $f(x)$ , 使得:  $y = f(x) = (\omega \cdot x) + b$  以使用  $y = f(x) = (\omega \cdot x) + b$  来推断任一输入的值。其中,  $(\cdot)$  是内积。根据结构风险最小化原则, 函数  $f(\cdot)$  应该使得下式成立:

$$\min \left\{ \frac{1}{2} \cdot \|\omega\|^2 + C \cdot \sum_{i=0}^n L(f(x_i), y_i) \right\} \quad (2.1.5)$$

$L$  为损失函数, 通常采用不敏感区损失函数, 其形式如下:

$$L(f(x_i), y_i) = \begin{cases} |f(x_i) - y_i| - \varepsilon, & \text{其他} \\ 0, & |f(x_i) - y_i| < \varepsilon \end{cases} \quad (2.1.6)$$

然后引入松弛变量  $\xi_i$  和  $\xi_i^*$ , 则函数可以用如下优化来等价:

$$\begin{aligned} & \min \left\{ \frac{1}{2} \cdot \|\omega\|^2 + C \cdot \sum_{i=1}^n (\xi_i + \xi_i^*) \right\} \\ & \text{s.t.} \begin{cases} y_i - f(x_i) \leq \varepsilon + \xi_i \\ f(x_i) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (2.1.7)$$

采用 Langrange 乘数法解决此约束优化问题, 构造 Langrange 方程:

$$\begin{aligned} L(\omega, \xi_i, \xi_i^*) = & \frac{1}{2} \cdot \|\omega\|^2 + C \cdot \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n \alpha_i \cdot (\varepsilon + \xi_i - y_i + \langle \omega, x_i \rangle + b) \\ & - \sum_{i=1}^n \alpha_i^* \cdot (\varepsilon + \xi_i^* + y_i - \langle \omega, x_i \rangle - b) - \sum_{i=1}^n (\beta_i \xi_i + \beta_i^* \xi_i^*) \end{aligned} \quad (2.1.8)$$

式中,  $\alpha_i, \alpha_i^*, \beta_i, \beta_i^* \geq 0$ , 为拉格朗日乘数,  $i=1, 2, \dots, k$ 。

最终求得回归问题变为:

$$f(x) = \sum_{i=1}^{nsv} (\alpha_i - \alpha_i^*) \cdot K(x_i, x_j) + b \quad (2.1.9)$$

式中,  $nsv$  为支持向量的个数,  $(\alpha_i - \alpha_i^*)$  为对应支持向量的系数;  $b$  为阈值。可由任意两个支持向量求解。

### 3.3. 判别模型

在问题一中, 将 Education requirements 指标  $x_i (i=1, 2, 3, 4)$  构成 Education requirements 指标矩阵:

$$X = [x_1, x_2, \dots, x_i] \quad (3.1.1)$$

根据中华人民共和国人力资源和社会保障部中知: 各城市岗位空缺与求职人数的关系均值矩阵为:

$$\bar{X} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_i] \quad (3.1.2)$$

定义 A-City 的 Education requirements 指标矩阵与均值之间的欧氏距离为  $L$ , 则

$$L = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3.1.3)$$

通过计算机循环运算, 找到与样本中欧式距离最短、次短的城市与之匹配, 再结合 A-City 的具体情况, 最后确定其可能的地理区域, 经济地位和高科技产业发展。

### 3.4. 人才吸引水平模型

#### 3.4.1. 一致性检验

在人才吸引力评价体系中, 各个指标必须从不同的方面反映影响人才吸引力的情况, 又不能出现一



种指标的重要程度大于另一种指标的重要程度。所以在建立了人才评价体系之后, 必须进行各个指标间的一致性检验。对每个一级指标分别进行一致性检验如表 2:

**Table 2.** Consistency test results of various first-level indicators

**表 2.** 各类一级指标的一致性检验结果

	克隆巴赫 Alpha	基于标准化项的克隆巴赫 Alpha	项数
发展前景 $A_1$	0.864	0.864	4
生活水平 $A_2$	0.712	0.712	5
主要行业增长率 $A_3$	0.808	0.808	6
城市环境 $A_4$	0.923	0.923	6
政策影响 $A_5$	0.764	0.764	5

根据检验结果来说, 发展前景、生活水平、主要行业增长、城市环境、政策影响这五个一级指标可以对于人才吸引这个评价目标而言, 能够较为准确的衡量。

### 3.4.2. 熵值法模型

人才吸引力评价指标体系中, 各评价指标从不同方面反映了人才吸引力的影响因素, 但它们对人才吸引力的影响力不同。为此, 我们选择使用熵值法来确定各项指标在评价体系中所占的权重, 从而量化各项指标对人才吸引力的影响程度。

熵值法基本原理:

在信息论中, 熵是对不确定性的一种基本度量。信息量越大, 不确定性就越小, 熵也就越小; 信息量越小, 不确定性就越大, 熵也就越大。我们可以用熵值来衡量某个指标的离散程度, 指标离散程度越大, 该指标对综合评价的影响越大。

其基本表达式为:

$$H(x) = \sum_{i=1}^m p(x_i) \ln p(x_i) \quad (4.1.1)$$

其中  $x_i$  为第  $i$  个状态值(共有  $m$  个状态),  $p(x_i)$  为出现第  $i$  个状态值的概率。

由于各个指标的量纲不同, 所以利用以下公式对各指标值进行非负数化、归一化处理。即取  $X$  矩阵中的第  $i$  列各元素与该列值最小的元素作差, 再取该列值最大元素与该列各元素作差, 然后二者作商, 最后加 1。得到新矩阵  $R$  且各元素  $r_{ij}$  的值处于  $[1, 2]$  之间。

$$r_{ij} = \frac{x_{ij} - \min_{1 \leq j \leq m} \{x_{ij}\}}{\max_{1 \leq j \leq m} \{x_{ij}\} - \min_{1 \leq j \leq m} \{x_{ij}\}} + 1 \quad (4.1.2)$$

取  $R$  矩阵第  $i$  列各元素与该列各元素之和作商, 得到  $P$  矩阵,  $p_{ij}$  为其中第  $i$  行, 第  $j$  列元素。

所以熵值为:

$$e_j = -k \sum_{i=1}^n p_{ij} \ln p_{ij} \quad (4.1.3)$$

其中,  $k$  为调节系数,  $k = 1/\ln n$ 。当确定了参与评价的城市数  $n$ ,  $k$  就是常量。  $0 < e_j < 1$ ,  $p_{ij}$  为第  $i$  个城市, 第  $j$  个指标的指标标准化值。

对给定的指标  $j$ ,  $x_{ij}$  的差异性越小,  $e_j$  就越大。当  $x_{ij}$  全部相等时,  $e_j = e_{\max} = 1$ , 此时指标  $j$  没有作

用, 各指标值的差异越大,  $e_j$  越小, 说明该指标起的作用越大。

$d_j$  为第  $j$  个指标评价数据分散程度:

$$d_j = 1 - e_j \quad (4.1.4)$$

将评价指标的熵值转化为权重值  $D$

$$w_j = \frac{1 - e_j}{m - \sum_{j=1}^m e_j}, j = 1, 2, \dots, m \quad (4.1.5)$$

其中  $0 \leq w_j \leq 1$ ,  $\sum_{j=1}^m w_j = 1$ 。

最后, 得出各指标的熵权综合评价。将各指标的权值与其对应的指标值相乘并求和, 其评价模型为:

$$u = \sum_{j=1}^m w_j r_{ij}, i = 1, 2, \dots, n \quad (4.1.6)$$

## 4. 模型的求解及结果分析

### 4.1. 问题 1PCA 数学模型的求解

PCA 算法的算法步骤流程图如下图 3 所示:

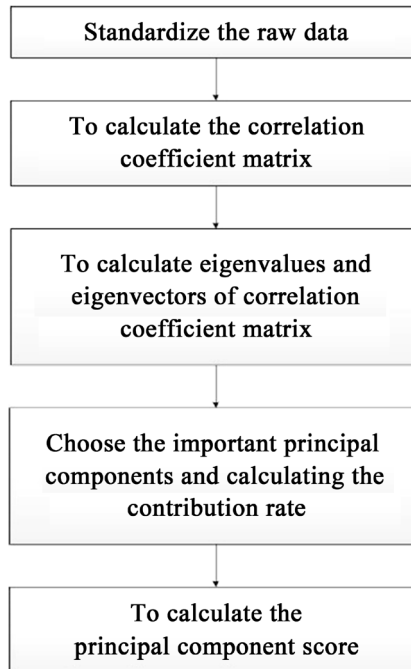


Figure 3. PCA algorithm flow chart

图 3. PCA 算法流程图

### PCA 算法结果

通过 MATLAB 编程, 本文得出 Sector 所对应的 Total demand (Pers.)与 Education requirements 的特征值矩阵如表 3 所示:

**Table 3.** Eigenvalue matrix of class 18 Sector  
**表 3.** 18 类 Sector 的特征值矩阵

Sector	Eigenvalue matrix
C	[2.6697, 0.6752, -0.2795, -0.0089]
E	[-0.6093, -0.0494, -0.0746, -0.0400]
F	[-0.4005, 0.3131, -0.0458, -0.0898]
G	[-0.8219, -0.5312, -0.0086, -0.0356]
H	[-0.7300, -0.1955, 0.1126, 0.1033]
I	[-0.7472, -0.2957, -0.0716, 0.1850]
J	[0.3380, -0.0019, -0.0201, 0.0101]
K	[-0.1147, 0.7670, 0.1108, -0.0038]
L	[-0.6952, -0.2401, -0.1858, 0.0163]
M	[0.0175, -0.0275, -0.1118, 0.0955]
N	[-0.0122, 0.9394, 0.2531, 0.0565]
O	[-0.3682, 0.0700, -0.0232, -0.0030]
P	[3.2284, -0.8308, 0.1939, -0.0061]
Q	[-0.2734, -0.1821, 0.0041, -0.1203]
R	[-0.3653, 0.1222, 0.0754, -0.0470]
S	[0.1269, -0.2891, 0.0549, 0.0424]
T	[-0.5269, -0.0470, -0.0105, 0.0360]
*	[-0.7158, -0.1965, 0.0265, -0.1906]

以上表格是不同的 Sector 所对应的 Education requirements 指标的特征值矩阵  $P(i=1, \dots, 18)$ 。通过上表可以发现：不同的 Education requirements 对应的特征值矩阵差异明显，即本文可以通过计算特征值矩阵的相似度来推测城市期望的职业和所需的教育背景。

#### 4.2. 问题 2 分析预测潜在人才的支持向量机模型的求解

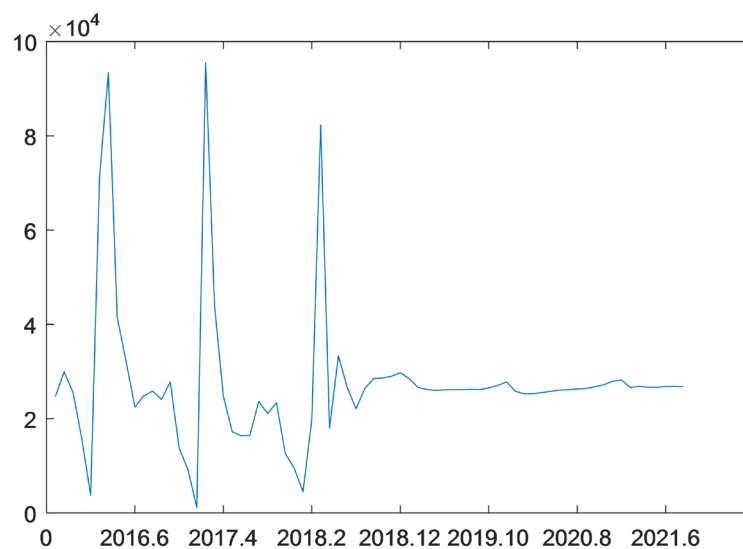
运用 MATLAB 编程, 采用支持向量机算法, 对附件中 2015 年 9 月至 2018 年 8 月中 Total demand (Pers.) 的人数拟合建模。再运用滚动预测的思想, 求出 2018 年 9 月至 2021 年 8 月潜在人才的数量, 其预测值见表 4:

**Table 4.** Predicted value of potential talent from September 2018 to August 2021**表 4.** 2018 年 9 月至 2021 年 8 月潜在人才的预测值

年份	预测值	年份	预测值	年份	预测值
2018.9	28,539	2019.9	26,171	2020.9	263,84
2018.10	28,626	2019.10	26,535	2020.10	267,89
2018.11	29,004	2019.11	27,046	2020.11	27,197
2018.12	29,747	2019.12	27,808	2020.12	27,965
2019.1	28,530	2020.1	25,818	2021.1	28,204
2019.2	26,620	2020.2	25,284	2021.2	26,612
2019.3	26,183	2020.3	25,298	2021.3	26,865
2019.4	26,005	2020.4	25,535	2021.4	26,675
2019.5	26,110	2020.5	25,804	2021.5	26,676
2019.6	26,148	2020.6	26,079	2021.6	26,811
2019.7	26,156	2020.7	26,150	2021.7	26,821
2019.8	26,219	2020.8	26,304	2021.8	26,770

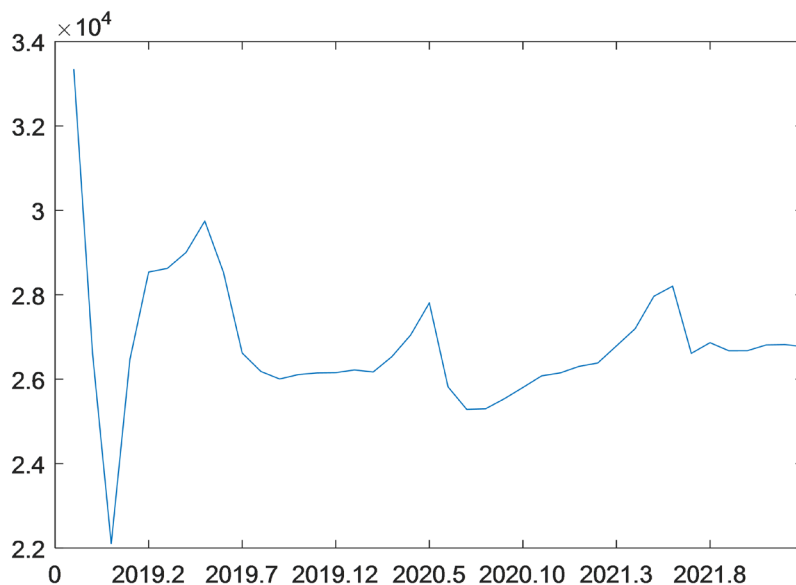
#### 分析预测潜在人才的支持向量机模型结果的分析

将附件中 2015 年 9 月至 2018 年 8 月中 Total demand (Pers.) 的人数拟合建模以及 2018 年 9 月至 2021 年 8 月导入 MATLAB 中, 可做出基于支持向量机的中国 2015 年 9 月~2021 年 8 月潜在人才的预测值与实际值的比较图与 2018 年 9 月至 2021 年 8 月潜在人才的预测值, 如图 4、图 5 所示:



**Figure 4.** Based on support vector machine, the predicted and actual value of potential talents in China from September 2015 to August 2021

**图 4.** 基于支持向量机的中国 2015 年 9 月~2021 年 8 月潜在人才的预测值与实际值



**Figure 5.** Comparison of actual values and predicted values of potential talents from September 2018 to August 2021  
**图 5.** 实际值的比较图与 2018 年 9 月至 2021 年 8 月潜在人才的预测值

从图 4 与图 5 可以看出在开始行业未稳定时 Total demand (Pers.)的人数需求是呈周期性变化,但在后面的预测过程中也有一定程度的周期性变化,但变化没有 2015 年 9 月~2018 年 8 月剧烈。

### 4.3. 问题 3 判别模型的求解

#### 4.3.1. 高科技产业发展判别

首先针对高科技产业发展, 本文认为高科技产业发展的基本全都是高文化背景得人, 因此本文根据高智商人才在 A-City 各行业中的分布情况来判断该市的高科技产业发展情况。

对于 A-City, 总共本科以上学历在各行业中只有 503 人, 占总数的 0.346%, 所以本文可以判断 A-City 的高科技产业发展的速率较低。

#### 4.3.2. 与行政类别判别

对于行政类别和经济地位, 本文认为一个城市的行政类别与其城市的重要贡献专业人才的类型具有相当大的关系, 我们相信一个优秀的城市它的重要贡献人才更偏向于高精尖产业或第一产业。

从附件中给的数据可以体现出 A-City 的第一产业为 G 类(wholesale and retail)产业, 并不属于高精尖技术, 带来的收益也较为低下; 所以 A-City 行政类别和经济地位都不高; 其行政类别可能为副省级市或者地级市这两类。

#### 4.3.3. 地理区域判别

首先在问题一中, 本文得到 18 类 Sector 的特征值矩阵, 所以其指标矩阵  $X$  可表示成:

$$[a_{11} + a_{21} + \dots + a_{n1}, a_{21} + a_{22} + \dots + a_{2n}, \dots, a_{1m} + a_{2m} + \dots + a_{nm}], \text{ 其中 } n = 18, m = 4.$$

将 18 类 Sector 的特征值矩阵带入得:

$$X = [12.7611, 5.7737, 1.6628, 1.0902]$$

根据中华人民共和国人力资源和社会保障部中知: 各地区岗位空缺与求职人数的关系均值矩阵如表 5 所示:

**Table 5.** Mean value matrix of job vacancies and job seekers in different regions**表 5.** 不同地区岗位空缺与求职人数的关系均值矩阵

地区	关系均值矩阵
东部地区	[36.6697,15.6792,11.7355,9.0345]
中部地区	[25.6983,10.0208,10.4776,3.0449]
西部地区	[15.4892,4.8193,2.0450,1.8898]

本文将 8 类 Sector 的特征值矩阵的数据输入程序, 计算出欧式距离。所得结果为西部地区的关系均值矩阵与  $X$  的欧式距离最小, 即 A-City 所处于西部地区。

#### 4.4. 问题 4 人才吸引水平模型的求解

本文根据清华经管学院互联网发展与治理研究中心发布的数据, 建立 4 类第一指标的权重如表 6 所示:

**Table 6.** The weight of each evaluation criterion as a target of evaluation**表 6.** 各评价标准占评价目标的权重

评价目标	评价标准	权重
人才吸引水平	发展前景 $A_1$	0.1657
	生活水平 $A_2$	0.2313
	主要行业增长率 $A_3$	0.2631
	城市环境 $A_4$	0.2728
	政策影响 $A_5$	0.0671

因为上海数字人才净流入量领跑全国, 所以我们本文上海市为基准, 对 A-City 的人才吸引力做出量化评价, 即可得出所以可以得到在没有新职业偏好的 A-City、有新职业偏好的 A-City 与上海的综合得分, 如表 7 所示:

**Table 7.** A-City and Shanghai's comprehensive score sheet**表 7.** A-City 与上海的综合得分表

城市	综合得分
无新职业偏好的 A-City	1.38
有新职业偏好的 A-City	1.46
上海	8.55

从表 7 可以明显的看出, 大学生有了一些新的职业偏好, 对城市发展产生较大的作用。

## 5. 策略和建议

随着社会经济的发展, 人才越来越显示出其不可替代的作用。科教兴国的人才战略, 适应未来经济发展的需要, 在人才方面应该从以下几个方面入手:

第一, 大力提倡多种形式办学, 实现多条腿走路。要实现人才的显著增长, 就必须加强高校的人才培养力度, 同时大力提倡多种办学形式, 加强企业对人才的二次培养以及成人教育, 形成公办教育和民办教育共同发展、社会教育和海外教育共同参与的教育培养新格局。近几年, 我国的实践也同样可以看出这一点。



第二，在人才投资方面，要树立人才是最重要的资本的理念；制定加大人才投资力度的政策，强化人才资源的超前性投资；实行人才投资多元化政策，建立政府、社会、个人相结合的人才投资回报政策；解决人才投资的动力机制和立法问题；在人才数量增加的同时要注重人才质量的不断提升。

第三，注重高质量人才的引进。人才引进成本低、见效快，是一种便捷的人才资源开发方式。政府应该建立规范的人才引进机制，完善人才市场的管理。同时，加强优惠政策，吸引外部人才的流入和防止人才外流。在引才引智时，应防止和避免盲目引进、重复引进和低效引进。

第四，在人才使用方面，政府要积极营造各种以人为本的环境，包括创业环境、市场环境、生活环境和舆论环境；建立良好的用人机制，防止人才外流；要保持人才政策的相对稳定性，注意政策的配套，谨防牵一发而动全身，导致整个人才系统状态的失衡与振荡；充分发挥人才的作用，以便对该城市经济发展产生巨大的拉动作用。

## 基金项目

西南科技大学大学生创新基金项目(项目编号: CX19-061)。

## 参考文献

- [1] 陆晓芳, 王川, 赵树宽. 人才要素区域竞争力评价模型[J]. 吉林大学学报(工学版), 2003, 33(3): 82-85.
- [2] 兰艳章, 柴华奇. 基于熵值法的多层次人才甄选评价模型[J]. 科技进步与对策, 2006, 23(4): 165-167.
- [3] 胡瑞卿, 张岳恒. 科技人才合理流动程度的模糊层次主成分分析测评研究[J]. 科学学与科学技术管理, 2008, 29(8): 189-193.
- [4] 黄凤, 王兆君. 基于层次分析法和熵值法的企业管型人力资本价值的计量[J]. 学术交流, 2010(2): 136-138.
- [5] 肖利哲, 吕璐. 基于熵值法的多方案人才战略决策模型构建[J]. 黑龙江对外经贸, 2011(1): 125-127.
- [6] 李朋林, 赵晶, 马琰玲. 基于 SVM 的人才资源对陕西经济增长影响的预测模型研究[J]. 西安科技大学学报, 2012, 32(5): 637-642.
- [7] 杨焕海. 人才需求组合预测的建模与仿真分析[J]. 计算机仿真, 2013, 30(10): 253-256.
- [8] 杨璐瑶, 张向前. 我国“十三五”适应创新驱动的科技人才发展机制的制度分析[J]. 科技管理研究, 2016, 36(7): 67-72.
- [9] 高建祥, 邱歆芮, 黄昕宇. 人才吸引力评价模型研究——以深圳市为研究对象[J]. 科技经济导刊, 2019, 27(12): 166-169.
- [10] 李健飞. 我国不同地区科技人才吸引力评价研究[D]: [硕士学位论文]. 青岛: 中国海洋大学, 2011.
- [11] Horowitz, I. (1966) Some Aspects of the Effects of the Regional Distribution of Scientific Talent on Regional Economic Activity. *Management Science*, **13**, 217-232. <https://doi.org/10.1287/mnsc.13.3.217>
- [12] Faggian, A. and McCann, P. (2009) Human Capital and Regional Development. *Quarterly Journal of Economics*, **128**, 105-164. <https://doi.org/10.4337/9781848445987.00015>
- [13] An, Q.S. (2011) The Research of Building Regional Talent Ecological Environment Comprehensive Evaluation Index System.
- [14] Lin, X.-J. (2012) Review on Regional Talent Development Strategy. *Journal of Shenyang Institute of Engineering (Social Sciences)*, **8**, 474-477.
- [15] Martindale, R.J.J., Collins, D., Douglas, C. and Whike, A. (2013) Examining the Ecological Validity of the Talent Development Environment Questionnaire. *Journal of Sports Sciences*, **31**, 41-47. <https://doi.org/10.1080/02640414.2012.718443>
- [16] Cadorin, E., Johansson, S.G. and Klofsten, M. (2017) Future Developments for Science Parks. *Industry and Higher Education*, **31**, 156-167. <https://doi.org/10.1177/0950422217700995>
- [17] Mellander, C. and Florida, R. (2011) Creativity, Talent, and Regional Wages in Sweden. *The Annals of Regional Science*, **46**, 637-660. <https://doi.org/10.1007/s00168-009-0354-z>
- [18] Huo, J.B. and You, J.X. (2015) Evolutionary Game Analysis on Regional Mobility of Research and Development Talents. *Journal of Tongji University (Natural Science)*, **43**, 1116-1122.
- [19] 2018 年第八届 APMCM 亚太地区大学生数学建模竞赛[Z]. <http://www.saikr.com/apmcm/2018/>