

基于评价模型对高校综合发展的研究

摘要

高等院校是我国经济社会发展中的重要参与者，本文主要研究各项指标对于高校发展水平和学科发展质量的影响，建立了熵权法模型、TOPSIS 分析模型、秩和比模型、系统聚类模型和灰色关联度模型。

针对问题一，首先考虑高校发展水平的影响主要体现在重要科技奖项和杰出人才产出上，本文选取高质量论文发表数量、国家级科研平台等指标进行评价。通过**熵权法**计算本文选取指标权重，其中高质量论文发表数量和杰出人才产出的权重分别为 0.0948 和 0.0944，表明该指标对高校发展所占权重较大，具有较大的影响力。其次，本文建立**Topsis 综合评价模型**，定量分析各高校指标数据对高校发展水平的影响。最后，得到各高校的综合得分，其中排名前三的高校分别为**西安交通大学、华中科技大学以及哈尔滨工业大学**。

针对问题二，关于对计算机科学与技术学科为对象进行对各高校学科发展水平进行评价，本文主要选取有关计算机学科指标来分析各高校的发展情况，其中本文选择学科评估、生源质量、ACM 排名、学科经费、学科院士人数等作为评价指标。其次，本文通过**秩和比综合评价法**分析了计算机学科各高校的学科能力，采用熵权法对指标进行加权评分，其中学科院士人数权重达到 0.258 占比最高。最后通过**线性回归**结果分析，显示北京大学和清华大学的学科等级值为 3，电子科技大学的学科等级的值是 1，可以得出**北京大学和清华大学**计算机学科发展质量最好，**电子科技大学**计算机学科发展质量相对较差。

针对问题三，为了分析题中所给五所高校的共同点、发展特色与未来前景，本文选取地区 GDP、高校发展特色、政府经费、重点学科数量、重点学科类型共 5 个指标，首先，对收集的数据进行预处理，将高校发展特色、重点学科类型等定性指标转化为定量指标，导入 SPSS26.0 中进行**系统聚类**分析，结果显示深圳大学和广州大学被分为一类，是因为两者均为创新型大学，所在城市 GDP 和其经费较高，重点学科数多。同理，成都大学和江汉大学为同一类，青岛大学单独一类。在研究学校发展与所在城市政策扶持之间的关联度时，选取出台政策作为决定学校发展的主要因素，经费、占地面积、设备总值和产学研情况作为次要因素，建立灰色关联模型，得出 4 个次要因素的**灰色关联度**分析大小分别为 **0.874、0.408、0.863 和 0.931**，因此，对学校发展影响最大的是**产学研情况和经费**。

关键词：熵权法;Topsis;线性回归;秩和比综合评价;灰色关联度分析;

一、问题重述

1.1 问题的背景

高等院校是我国经济社会发展中的重要参与者，一流大学与一流学科的建成是实现社会主义共同富裕的重要基础。随着国家第二轮“双一流”计划评选结果的公布以及第五轮学科评估的有序开展，高校整体发展质量与关键学科建设成效引起教职工、高考生、应届毕业生、准研究生、企事业单位等多方位的热切关注，如何合理地、有效地评判各类高校在过去几年建设发展中的利与弊，是较为困难的，也是有较强现实需求意义的

1.2 解决的问题

问题一：收集题中“西安交通大学、湖中科技大学、哈尔滨工业大学”等 10 所高校的相关数据，从“国家级科研平台、重要科技奖项、重大科研项目”等指标中选出合适的指标对高校发展水平进行评估。

问题二：以“计算机科学与技术”为对象，选择合适的指标，对北京大学、清华大学和浙江大学等高校 2020 年度或 2021 年度的学科发展质量进行评估。

问题三：市属高校着重服务本地经济发展，与当地城市发展息息相关，起步较晚但发展较快是其典型特征。综合评价分析“深圳大学、广州大学、成都大学、江汉大学和青岛大学”之间的共同点、发展特色和未来发展前景，并讨论学校发展与属地城市政策扶持之间的关联度。

二、问题分析

2.1 问题一的分析

本问要求收集“西安交通大学、华中科技大学和哈尔滨工业大学”等 10 所高校的院校相关数据，选取“重要科技奖项数量、杰出人才产出数量和高水平论文数量”等指标对题中 10 个院校进行综合评估。因此，本文根据题中给出的指标收集数据，建立熵权法模型，找出特征指标与院校综合发展水平的权重占比，找出各个指标的权重关系后建立 TOPSIS 综合评价模型，将特征指标作为评估影响因素，计算出各高校的综合评分，并通过得分对高校发展水平进行排序。

2.2 问题二的分析

本问要求收集院校特征数据对“北京大学、清华大学和浙江大学”等 11 所高校的“计算机科学与技术”专业 2020 年度或 2021 年度进行发展质量的评估。本问与问题一不同在于给定了学科定量指标，问题一中熵权法和 TOPSIS 综合评价法是使用多个指标，从多个角度对院校进行综合评价，最终找出院校综合发展水平排名，问题二则是收集计算机科学与技术学科相关指标，对学科进行综合评估。本文通过官方平台收集了与教学质量发展有关的指标，利用熵权法找出指标权重，建立 RSR 法模型，画出线性回归曲线，对院校计算机科学与技术发展质量进行评级。

2.3 问题三的分析

要分析深圳大学、广州大学等五所年轻高校的共同点、发展特色与未来前景和学校发展与属地城市政策扶持之间的关联程度，考虑到这些高校均为地方市属高校，因此，这些高校的发展与所在城市的 GDP 和经费息息相关，同时，其发展定位和学校综合实力决定了自身的发展特色和前景。通过在各高校官网搜集的数据，进行系统聚类分析，可将这五所高校分成相应的类别。在各类别中，可以直观的看出其共同点、发展特色与前景未来。为挖掘出影响学校发展的关键因素，建立灰色关联模型，可根据灰色关联度大小判断与学校发展相关因素的重要程度。

三、模型假设

1. 高校发布信息完全是学校真实数据。
2. 收集的信息适用于所有高校。
3. 编秩时出现的误差不影响后续回归分析。
4. 特征指标与高校评估相关性密切。
5. 假设学科发展不会出现突发外界因素影响。

四、符号说明

象征	意义
x_{ij}	第 i 个样本的第 j 个指数的值
Z_{ij}	归一化矩阵
D_i^+	第 i 个对象与最大值之间的距离
D_i^-	第 i 个对象与最小值之间的距离
C_i	非规范化分数

五、模型的建立与求解

5.1 问题一模型建立与求解

5.1.1 数据分析

本文对“西安交通大学、华中科技大学和哈尔滨工业大学”等高校的发展水平选取指标进行综合评价，通过各大高校官网、地方教育局和中国教育部等官方信息平台，收集了重要科技奖项、重大科研项目和保研率等 10 个特征指标数据，得到 2021 年各高校院士占比数量如图 1 所示，各高校保研率如图 2 所示。有更多完整数据，请参阅附录。

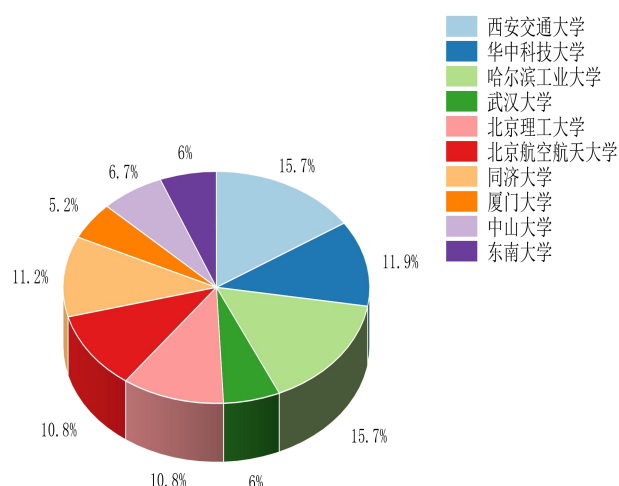


图 1 各高校院士占比

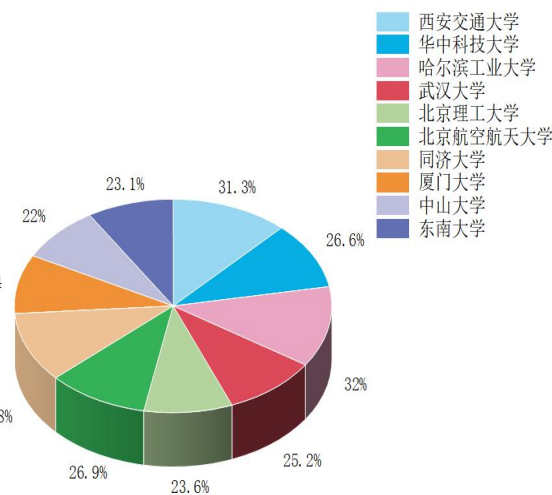


图 2 各高校保研率

由获得的数据和图 1、图 2 所示饼状图可以得出不同的高校的特征指标之间存在一定的差异，哈尔滨工业大学作为传统的工科类院校，经过百年的发展与积淀，哈尔滨工业大学具有较高的 15.9 的院士占比，并有极高的 32% 保研率，但不能只用单个指标作为评估的指标，武汉大学值拥有其他院校较低的 6% 的院士占比，但拥有较高的 25.2% 的保研率。因此院士占比和保研率两个指标不能直接，要评价高校的发展水平还需要其他方面的特征指标进行综合评估。

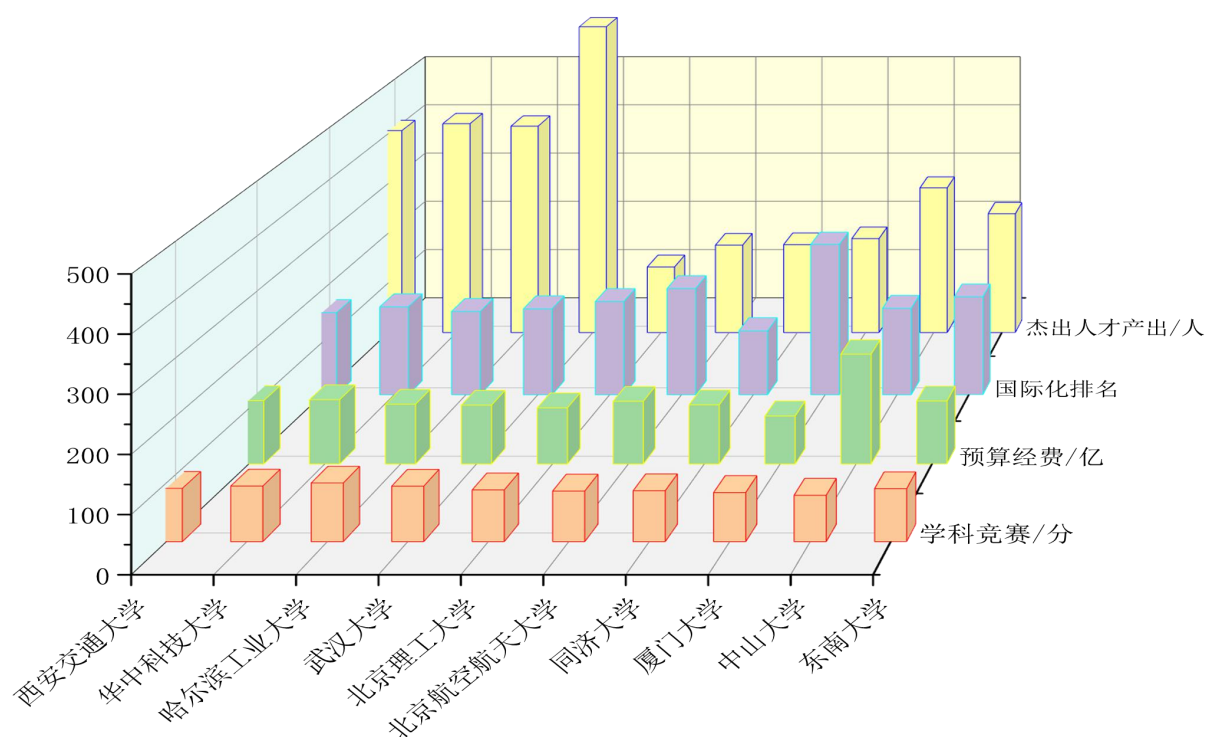


图3 各高校指标数据对比图

收集西安交通大学等10所高校的2021年学科竞赛、预算经费、国际排名和接触人才数量绘制图3，如上图所示，技术类高校如西安交通大学、华中科技大学、哈尔滨工业大学和武汉大学分别拥有406、420、415、615名杰出人才产出。所选的高校中，厦门大学每年86.81亿预算经费，所有院校中最低，中山大学每年拥有198.55亿预算经费，所有院校中最高，但是厦门大学和中山大学的杰出人才产出却只有189和291，因此还需要和其他综合指标进行综合一个学校发展水平。在科研方面，中山大学高质量论文数量以29393篇论文反超技术类高校，拥有最高的高质量论文数量。

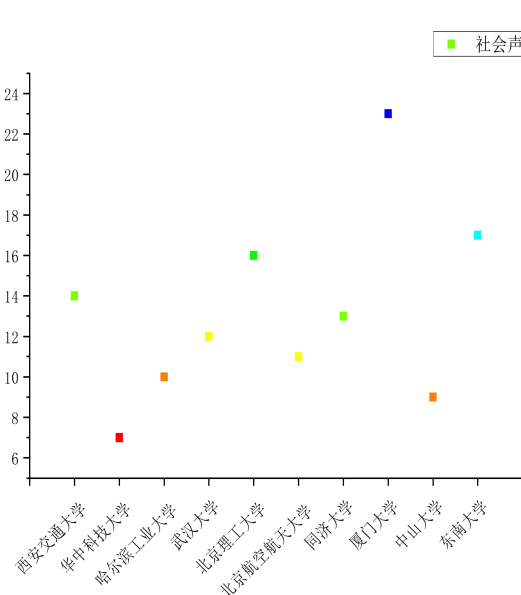


图4 社会声誉散点图

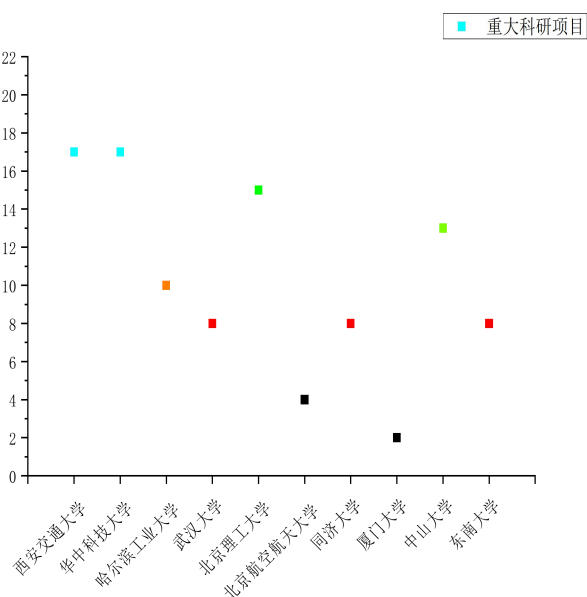


图5 重大科研项目数散点图

通过收集社会声誉排名和重大科研项目数量绘制了散点图如图4、图5所示，根据散点图可以得出，厦门大学社会声誉排名23相对落后，西安交大和华中科大的科研项

目 17 项并列第一，厦门大学 2 项科研项目数量较落后于其他高校，从散点图可以得出初步的比例关系，科研能力越强，社会声誉排名越靠前，但还需考虑其他因素进行高校综合评估，上图显示西安交大和华中科大同样的科研项目数量但社会声誉排名相差较大。

哈尔滨工业大学拥有最高的 15.7% 的院士占比，拥有最高的 32% 的保研率，武汉大学每年拥有每年 106.44 亿想对较低的预算经费，每年培养出最高的 615 名杰出人才，综上所述，高校发展水平分析得从科研水平、师资力量和经费等方面进行综合评估。表 1 列出了高校 21 年的科研数据。

表 1 高校科研能力

	国家级科研平台/分	重要科技奖项/项	重大科研项目/项	社会声誉	高质量论文
西安交通大学	45	91	17	14	26563
华中科技大学	49	88	17	7	14172
哈尔滨工业大学	28	70	10	10	28364
武汉大学	28	51	8	12	24845
北京理工大学	16	18	15	16	16605
北京航空航天大学	32	50	4	11	20767
同济大学	38	36	8	13	27794
厦门大学	33	23	2	23	14127
中山大学	29	46	13	9	29393
东南大学	28	48	8	17	21546

不同高校发展方向和研究对象不同，如表 1 数据表明，哈尔滨工业大学是传统百年工科强校，学校发展科研方向比重大，因此，学校的科技奖项和高质量论文数量都相较于其他院校更多，西安交通大学是综合类强校，2021 年搜集到的数据显示西安交大国家级科研平台综合得分和科技奖项数等特征指标都位列前位。师资水平、培养人才数量和预算经费等单个特征指标不能直接用来评估高校发展水平，本文利用科研能力、学术水平和师资力量等方面指标建立熵权法模型和 TOPSIS 分析模型进行高校发展水平的综合评估。

5.1.2 模型的准备

通过研究国内外学者关于高校发展水平评估影响评价的文献，总结了若干指标。正确合理的指标有助于评价高校发展的真实水平，因此所选择的指标应具有一定程度的典型性和合理性。题中给出的保研率等 10 余个指标能够从多方面对高校进行评估，本文根据指标遵循科学性、重要性、完整性、系统性、实用性的原则，最后将指标归为三类，建模时利用多个指标对高校进行评价排名。

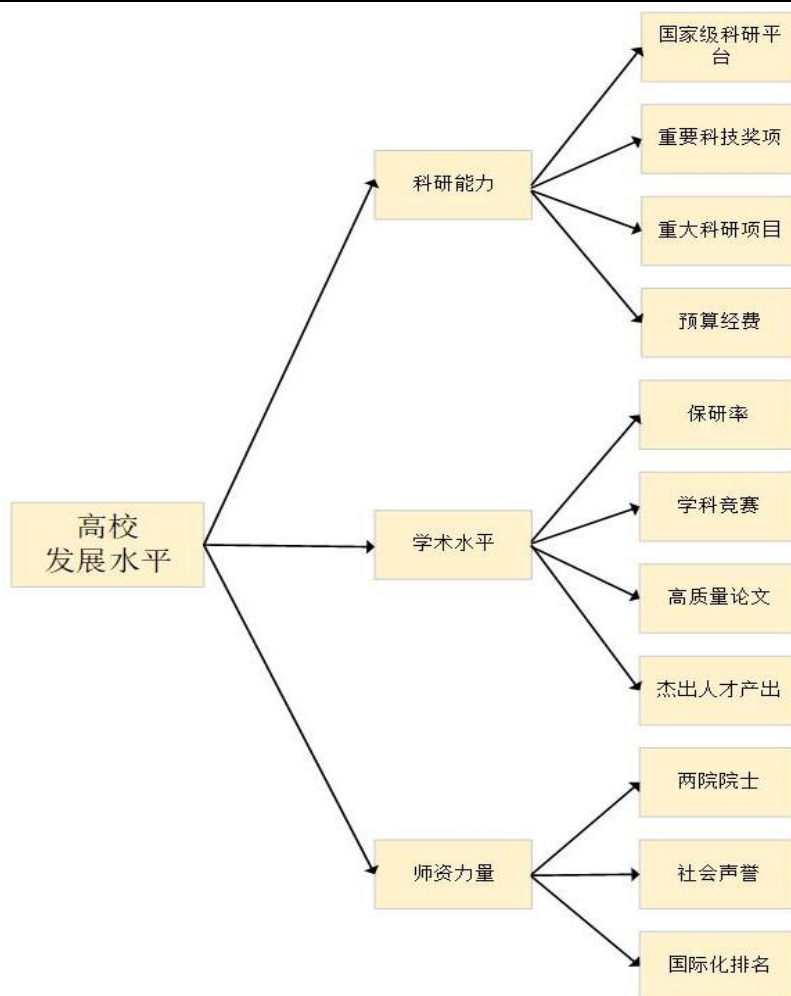


图 6 问题一评价指标

在文献综述的基础上，选取了以上指标对高校发展水平进行评估。

5.1.3 熵权法模型的建立与求解

熵权法是一种综合考虑各种因素的基础上计算关键指标权重的数学方法。也是一种客观的综合权重测定方法，它主要指依据每个指标传递给决策者的信息量来确定指数权重。根据信息论的基本原理，各类信息是能够衡量系统中有序程度的尺度。熵是衡量系统中无序程度的量度。因此，系统熵就用来反映它可以提供给决策者的信息量，而系统熵则通过熵权法获得^[1]。其计算步骤如下：

计算索引中第 j 个指数下第 i 个样本的比例：

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}, i = 1, \dots, n, j = 1, \dots, m \quad (1)$$

计算第 j 个指数的熵值：

$$e_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij}) \quad (2)$$

哪里 $k = \frac{1}{\ln(n)} > 0$. satisfies $e_j \geq 0$;

计算信息熵冗余：

$$d_j = 1 - e_j \quad (3)$$

计算每个指标的权重：

$$w_j = \frac{d_j}{\sum_{j=1}^m d_j} \quad (4)$$

最终的结果于表 2 中

表 2 各指标的权重

指标	权重	指标	权重
国家级科研平台/分	0.0908	杰出人才产出/人	0.0944
重要科技奖项/项	0.0923	国际化排名	0.0930
重大科研项目/项	0.0838	两院院士数/人	0.0906
保研率/%	0.0908	社会声誉	0.0822
学科竞赛/分	0.0932	高质量论文	0.0948
预算经费/亿	0.0941		

根据上表分析可以得到，科研项目数量和社会声誉对高校发展水平影响相对较小，高质量论文和接触人才产出为代表的科研水平和院校培养能力对高校发展水平有极大的影响，综合分析影响指标权重占比可得到，所有指标对高校水平都具有一定的影响，科研水平和经费投入对高校水平影响较大。

5.1.4 TOPSIS 综合评价方法

(1) 数据处理

为了定量评价国家级科研平台等特征指标对高校发展水平的影响，采用 TOPSIS 综合评价方法。TOPSIS 方法的思想流程图如图 7 所示。

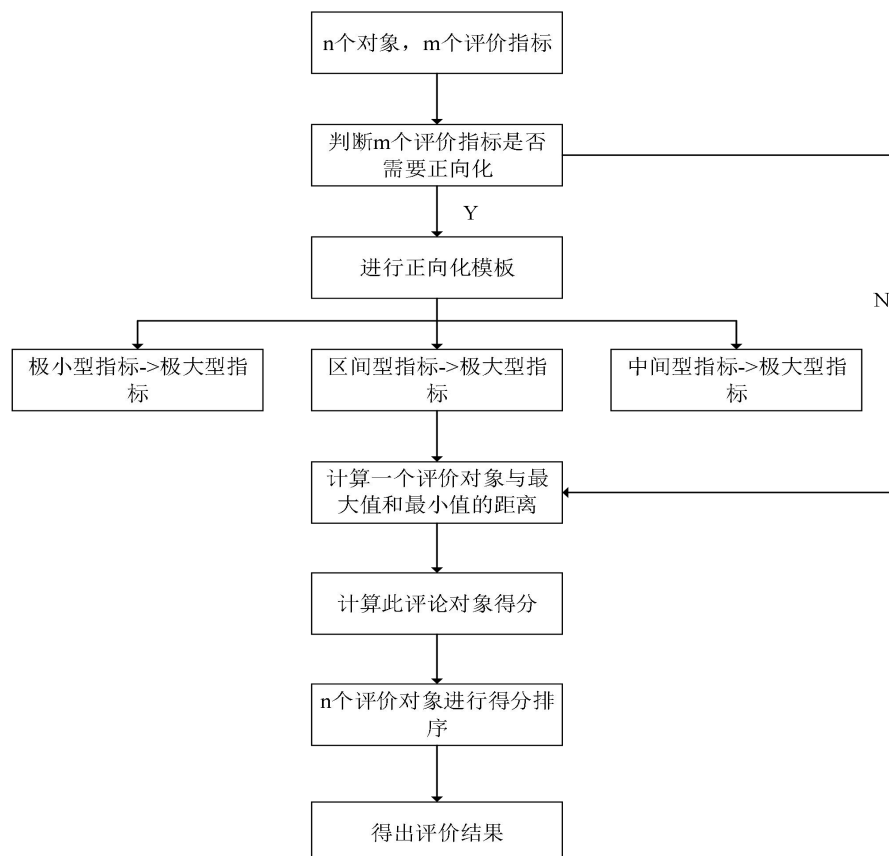


图 7 TOPSIS 运行原理图

TOPSIS 方法要求所有数据都显示上升趋势，即数据值越大，数据越好。使用 TOPSIS 需要数据同化。因此，国际化排名和社会声誉都朝着正方向处理一些数据处理结果如表 3 所示。

表 3 部分归化的一些指标数据

高校	国家级科研平台/分	重大科研项目/项	保研率/%	学科竞赛/分	两院院士数/人
西安交通大学	0.4209	0.4744	0.3750	0.3205	0.4648
华中科技大学	0.4583	0.4744	0.3187	0.3342	0.3541
哈尔滨工业大学	0.2619	0.2791	0.3834	0.3526	0.4648
武汉大学	0.2619	0.2233	0.3020	0.3336	0.1771
北京理工大学	0.1496	0.4186	0.2828	0.3111	0.3209
北京航空航天大学	0.2993	0.1116	0.3223	0.3045	0.3209
同济大学	0.3554	0.2233	0.3331	0.3064	0.3320
厦门大学	0.3086	0.0558	0.2804	0.2955	0.1549
中山大学	0.2712	0.3628	0.2636	0.2791	0.1992
东南大学	0.2619	0.2233	0.2768	0.3184	0.1771

表中数据可以得出，科技类院校的发展水平受基础设施和师资能力影响较大，影响理工类发展水平的主要是师资力量和科研项目数量，综合类影响指标较为宽泛，都有影响，但不受单一特征指标严重影响。

(2) 得到计算指标的理想接近度

索引归一化矩阵是根据上表的归一化数据建立的，该数据定义了第 i 个评估对象与最大值之间的距离：

$$D_i^+ = \sqrt{\sum_{j=1}^m \omega_j (Z_i^+ - z_{ij})^2} \quad (5)$$

定义第 i 个评估对象与最大值的距离：

$$D_i^- = \sqrt{\sum_{j=1}^m \omega_j (Z_i^- - z_{ij})^2} \quad (6)$$

然后，本文计算了未规范化的评估对象 i ($i=1, 2, \dots, n$)：

$$C_i = \frac{D_i^-}{D_i^- + D_i^+} \quad (7)$$

其中，该值在 0 和 1 之间，接近 1 表示整体接收容量和吸引力接近最优方案，反之亦然，它是最差方案。表 4 显示了一些数据计算。

表 4 最大最小值的距离和未归一化的得分

高校	D+	D-	Ci
西安交通大学	0.3775	0.9049	0.7057
华中科技大学	0.3953	0.9343	0.7027
哈尔滨工业大学	0.4474	0.8008	0.6416
武汉大学	0.5815	0.7242	0.5547
北京理工大学	0.8289	0.5428	0.3957
北京航空航天大学	0.7281	0.5303	0.4214
同济大学	0.6675	0.6689	0.5005
厦门大学	1.0787	0.1715	0.1372
中山大学	0.5555	0.7542	0.5759
东南大学	0.7459	0.4494	0.3759

通过熵权法和 TOPSIS 综合评估获得最后的得分，最终的排名如表 5。

表 5 得分及排名

高校	得分	排名
西安交通大学	0.1408	1
华中科技大学	0.1402	2
哈尔滨工业大学	0.1280	3
中山大学	0.1149	4
武汉大学	0.1107	5
同济大学	0.0999	6
北京航空航天大学	0.0841	7
北京理工大学	0.0790	8
东南大学	0.0750	9
厦门大学	0.0274	10

根据以上数据的计算结果，综合国家科研级平台、重要科技奖项和重大科研项目等 10 个指标综合评价分析可以得到选中高校的综合得分和综合排名。由表中数据得出，西安交大得分为 0.1408、华中科技得分为 0.1402，这两所高校分别排名第一和第二。

5.2 问题二模型建立与求解

5.2.1 模型准备

(1) 模型的讲解

秩和比综合评价法的定义：秩和比综合评价法（Rank-sum ratio）是由我国知名学者田凤调教授 1988 年根据统计学提出的一种综合分析模型，曾经广泛运用在医学，经过优化后适合运用于计量资料的综合评价分析。

秩和比法原理：秩和比（RSR）指将效益型指标从小到大排序进行排名、成本型指标从大到小排序进行排名，再计算秩和比，最后统计回归、分档排序。通过秩转换，获得无量纲统计量 RSR，以 RSR 值对评价对象的优劣直接排序或分档排序，从而对评价

对象做出综合评价。



图 8 RSR 建模思路

(2) 数据分析

以计算机科学与技术发展有关水平为研究对象，通过环球信息网、各大高校官网和教育局等机构获取评估“计算机科学与技术”发展质量的因素相关图如下：

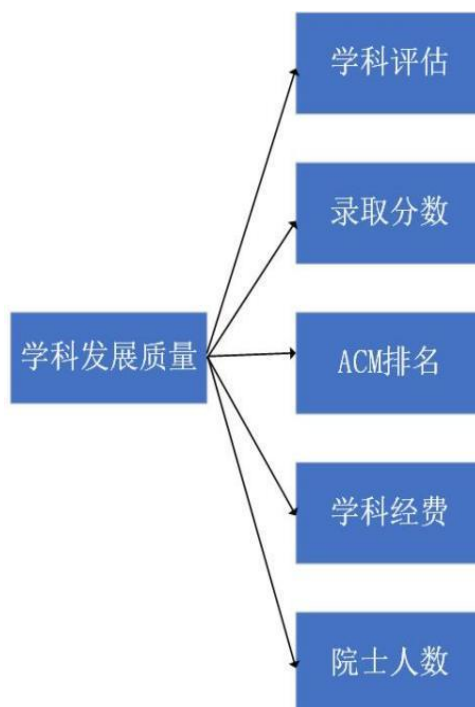


图 9 指标流程图

上图可以得出，计算机科学与技术发展质量与预算科研经费、院士人数和学科排名等因素有关，本文搜集了以上能够进行学科质量评估的五个因数，针对录取分数数据做了柱状图如下：

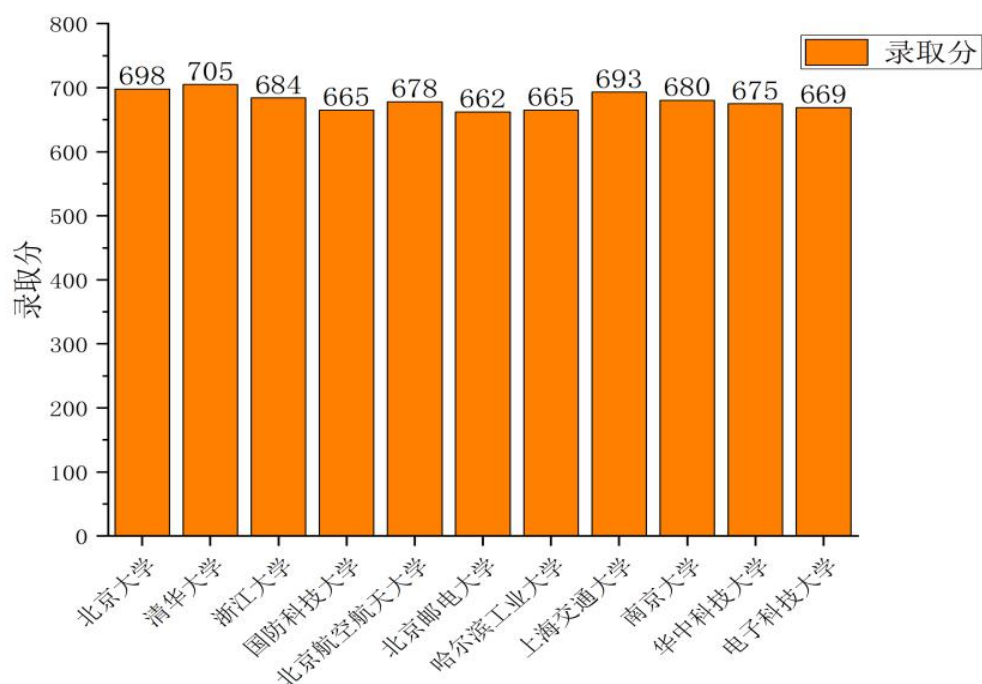


图 10 四川省录取分数线

如上图所示，图中的学校录取分数线针对四川省都具有分数较高的特点，最低都在 660 分以上，清华大学 705 分录取分数线为最高，最低的北京邮电大学也是 662 分，可以得出发展计算机科学与技术学生的自生水平有一定的联系。



图 11 学科特征指标堆叠面积图

院士数量、学科评估排名、预算经费和 ACM 国际排名作为影响计算机科学与技术发展的影响因素，利用院士数量等特征指标建立图 11 和图 12，北京大学学科排名为第一，国防科技大学的 ACM 国际排名相对较低，清华大学 26.44 亿元的预算经费所有高校中最高。本文利用 5 个特征指标对 11 个院校进行学科发展质量评估，建立 RSR 模型

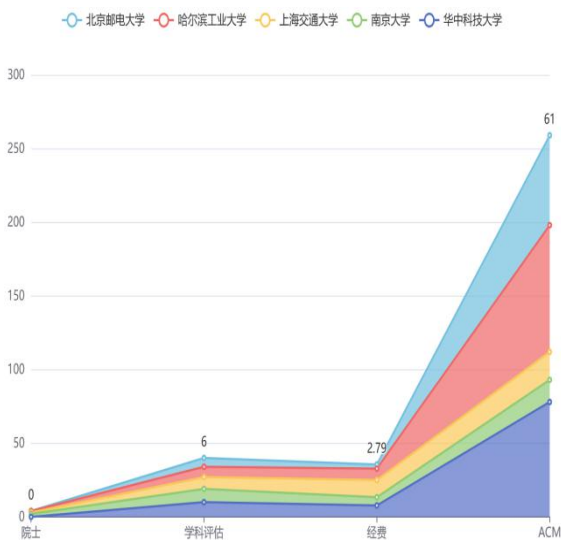


图 12 学科特征指标堆叠面积图

对每个高校进行综合打分，最终得出学科发展高校排名。

5.2.2 模型的建立——RSR 模型

步骤 1：选定北京大学、电子科技大学等 11 个大学作为评价对象，将学科评估排名、预算经费等 5 个数据作为评价指标，构建 11*5 的数据矩阵。

步骤二：

(1) 矩阵进行编制。将 11 个评价对象的 5 个评价指标排列成 11 行 5 列的原始数据表，编出每一个指标和各评价指标对应的秩，得到秩矩阵为 $R = (R_{ij})_{11 \times 5}$ 。

(2) 优化 RSR 模型。为了使得秩和比法模型更加的真实可靠，根据原指标与编制次数之间的线性关系建立线性回归方程，解决综合评价中的问题。指标从小到大编秩的是效益型指标，成本指标根据指标的额大小进行从大到小排序。

对于效益型指标：

$$R_{ij} = 1 + (n - 1) \frac{X_{ij} - \max(X_{1j}, X_{2j}, \dots, X_{nj})}{\max(X_{1j}, X_{nj}, \dots, X_{nj}) - \min(X_{1j}, X_{nj}, \dots, X_{nj})} \quad (8)$$

对于成本型指标：

$$R_{ij} = 1 + (n - 1) \frac{\max(X_{1j}, X_{2j}, \dots, X_{nj}) - X_{ij}}{\max(X_{1j}, X_{nj}, \dots, X_{nj}) - \min(X_{1j}, X_{nj}, \dots, X_{nj})} \quad (9)$$

步骤三：计算评价指标的权重，计算秩和比。通过熵权法确认指标的权重。

计算秩和比：

$$RSR_i = \frac{1}{n \times m} \sum_{j=1}^m R_{ij} \quad (10)$$

当各个指标拥有不同的权重时，计算加权秩和比（WRSR），计算公式：

$$WRSR_i = \frac{1}{n} \sum_{j=1}^m W_j R_{ij} \quad (11)$$

步骤四：计算指标与编秩次数的线性回归方程。将 RSR 编制频率转换为概率单位 Probit_i，将 Probit_i 作为线性回归的自变量，以 RSR_i 值为自变量，计算直线的线性回归方程，即 $RSR = a + b * Probit$ 。

步骤五：综合评估值排序，得出最终评价结果。

按照优化的线性回归方程得出最终的 RSR 评估值，根据过的评价数据对评估对象进行排序。

5.2.3 问题二的求解

本文将 11 个高校的院士数量、学科评估排名作为定量的输入，使用熵权法得出权重如下表所示

表 6 权重计算

指标	信息熵值 e	信息效用值 d	权重
生源质量	0.856	0.144	0.231
经费	0.868	0.132	0.213
院士	0.84	0.16	0.258
学科评估	0.897	0.103	0.166
ACM	0.918	0.082	0.132

上表是熵权法所计算得出的权重结果，根据输入的各个指标对自变量进行权重分析。

熵权法的结果分析得到，生源质量（20 年高考）的权重占比为 23.141%，经费的权重占比 21.297%，学科评估的权重为 16.576%、ACM 排名的权重为 13.198%，其中指标权重最大值为院士（25.789%），最小值为 ACM 排名（13.198%）。

利用熵权法确定权重后，将 RSR 作为因变量数据，根据 $RSR_i = \sum \frac{R}{(MN)}$ 计算得到 n=11 时得线性分析结果如下表：

表 7 线性回归结果分析

	非标准化 系数	标准化 系数	t	p	VIF	R2	调整 R2	F
常数	-0.741	-	-7.755	0.00 0***	-	0.9 52	0.946	F=176.781 P=0.000***
Probit	0.241	0.975	13.296	0.00 0***	1			

注：***、**、*分别代表 1%、5%、10%的显著性水平

上表格展示了本次模型的分析结果，包括模型的标准化系数、t 值、VIF 值、R²、调整 R²等，用于模型的检验，并分析模型的公式。线性回归模型要求总体回归系数不为 0，即变量之间存在回归关系。根据 F 检验结果对模型进行检验，R²代表曲线回归的拟合程度，越接近 1 效果越好，VIF 值代表多重共线性严重程度，用于检验模型是否呈现共线性，即解释变量间存在高度相关的关系（VIF 应小于 10 或者 5，严格为 5）。

从 F 检验的结果分析可以得到，显著性 P 值为 0.000***，水平呈现显著性，拒绝了回归系数为 0 的原假设，同时模型的拟合优度 R²为 0.952，模型表现较为优秀，因此模型基本满足要求。对于变量共线性表现，VIF 全部小于 10，因此模型没有多重共线性问题，模型构建良好。对于变量共线性表现，VIF 全部小于 10，因此模型没有多重共线性问题，模型构建良好。模型的公式如下： $y = -0.741 + 0.241 * Probit$

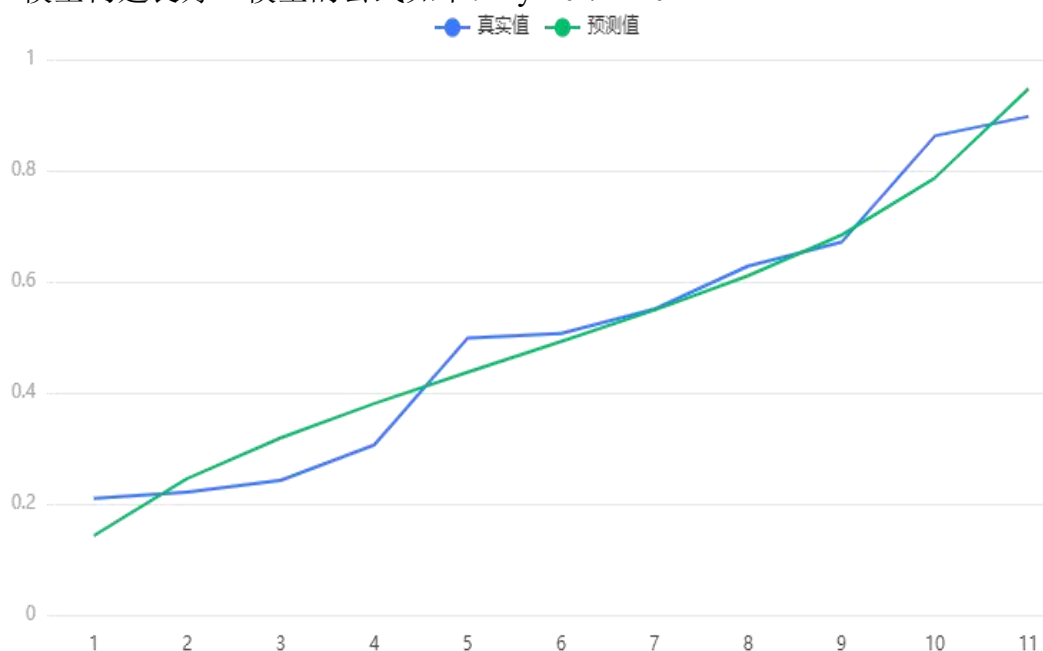


图 13 拟合效果图

上图是 $y = -0.741 + 0.241 * \text{Probit}$ 的数据拟合图，将 Probit 作为因变量时得出的原始数据图、模型拟合值、模型预测值图，图中可以看到真实值和预测值之间趋势相同，误差较小，因此得出秩和比法预估效果较好，能够得出较好的评价结果。按照以上拟合图的回归方程推算出实际的 RSR 临界值并做出临界值表 8 如下：

表 8 分档排序临界值表格

档次	百分位临界值	Probit	RSR 临界值（拟合值）
第 1 档	<15.866	<4	<0.2237
第 2 档	15.866 ~	4 ~	0.2237 ~
第 3 档	84.134 ~	6 ~	0.706 ~

上表中主要数据时得到的分档排序临界值表格，表中 Probit 临界值对应的 RSR 临界值（拟合值），百分位数临界值和 Probit 临界值根据分档水平数量而变化，该两项是固定值且完全一一对应，上表格中 RSR 临界(拟合值)是根据 Probit 临界值代入回归模型计算得到。将 11 所高校通过 RSR 法得出的评价价值按照分档排序临界值进行统一汇总得到最终的“计算机科学与技术”发展质量排名表 9 如下所示：

表 9 RSR 评估结果

高校	RSR_Rank	Probit	RSR Regression	Level
北京大学	2	6.3352	0.7868	3
清华大学	1	7.0004	0.9473	3
浙江大学	3	5.9085	0.6839	2
国防科技大学	7	4.8858	0.4373	2
北京航空航天大学	4	5.6046	0.6107	2
北京邮电大学	10	4.0915	0.2458	2
哈尔滨工业大学	8	4.6512	0.3807	2
上海交通大学	5	5.3488	0.5490	2
南京大学	6	5.1142	0.4924	2
华中科技大学	9	4.3954	0.3190	2
电子科技大学	11	3.6648	0.1428	1

以上的排名是按照回归方程推算所对应的 RSR 估计值对评价对象进行分档排序，通过 RSR 拟合值，以及上一表格中的 RSR 临界(拟合值)进行区间比较，进而得到分档等级水平，分档等级 Level 数字越大表示等级水平越高，即效应越好。表中数据显示北京大学和清华大学的 Level 的值为 3，电子科技大学的 Level 的值是 1，可以得出北京大学和清华大学学科发展质量最好，电子科技大学学科发展质量相对较差。

5.3 问题三模型建立与求解

5.3.1 模型的准备

(1) 数据分析

综合评价市属高校发展的共同点，通过各大高校数据统计网收集评估高校发展的指标，确定城市 GDP、院校发展特色、预估经费、学科评估和重点学科作为分析“深圳大学、广州”等高校的共同点的指标，并绘制了各个高校的指标数据表 10 如下：

表 10 高校数据指标

高校	城市 GDP/亿	发展特色	经费/亿	重点学科数	重点学科
深圳大学	30665	创新型大学	75.33	28	医学
广州大学	28232	创新型大学	40.74	16	通信工程
成都大学	19917	应用型城市大学	24.30	1	医学
江汉大学	17717	综合性城市大学	14.57	1	材料
青岛大学	14136	综合性研究型大学	20.64	25	医学

由表 10 中信息可以得到，五所高校都是建立在经济发展很好的城市中，每一所高校每年都投入较高的资金到学科建设和科研项目中。深圳大学和广州大学每年分别投入 75.33 亿和 40.74 亿元人命币建设高校，学校主要培养创新性人才，并在 2021 年的学科评估中累计评上 28 个和 16 个国家级重点学科。青岛相对其他几个城市 GDP 较低，但每年的预算经费投入一点不低，院校坚持发展综合性研究类院校，在 2021 年的学科评估中取得了 25 个重点学科。

（2）模型的介绍

系统聚类算法的基本思路是，首先将每个样本设定为一个单一的类，并以每两个类之间的距离作为两者之间相似程度的评价指标。距离越大，两个类的差异越大，距离越小，两个类就越相似。将相似度最高的两个类合并成一个新类后，计算新类与其他类的距离，并将相似度最高的类合并到新类。通过这个过程的不迭代，所有的类都被合并为一个类。流程图如下图 14 所示：

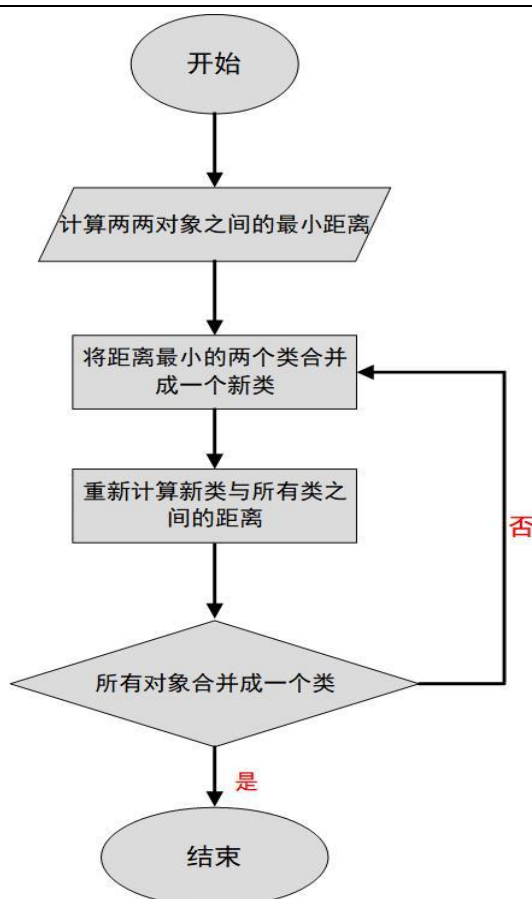


图 14 系统聚类算法流程图

5.3.2 系统聚类模型的建立

(1) 数据预处理

为了便于数据分析和绘图，本文选取了反映高校发展具有代表性的数据进行统计分析。高校共同点主要考虑两个指标：发展特色和重点学科。发展特色是学校主要发展方向，能够直观的反映学校未来的建设目标。重点学科是学校的优势发展科目，与发展特色一样能够直接反映院校的发展方向。由于建立系统聚类模型时需要对数据进行定量分析^[4]，因此将发展特色相同的院校定位一类，重点学科相同的院校定位一类。在软件 SPSS 中选择将定性的数据进行标准化处理，经过标准化转换，数据摆脱了测量单位的限制，能够进行后续的聚类分析。

(2) 确定群组的数量

根据“肘部法”，通过图表可以大致估计出最优的聚类数量。当类别数小于最优聚类数时，随着类别数的增加，聚集系数会出现急剧下降，而当类别数等于最优聚类数时，聚集系数的下降幅度会急剧减小。随着类别数的增加，聚集系数的下降将逐渐变得平缓。因此，与聚集系数线图的弯头位置相对应的类别数是最佳集群数。通过 SPSS 得到的聚集系数与聚类类别数的关系如下图所示。

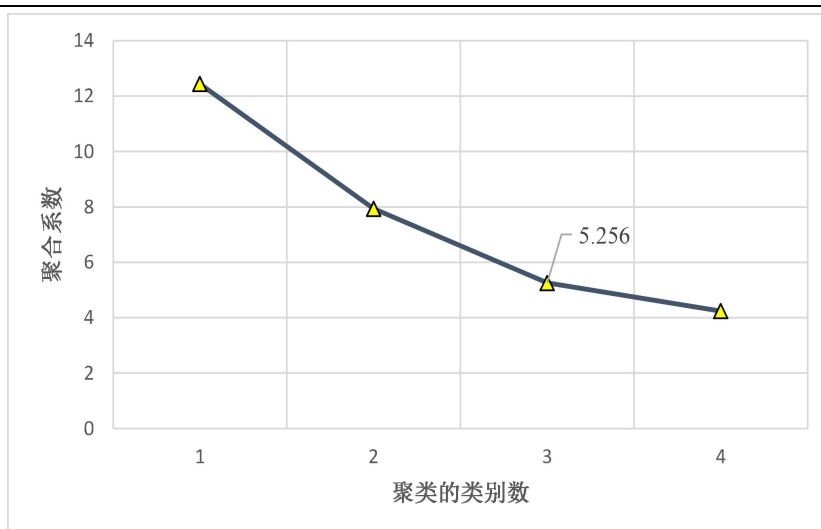


图 15 聚类类别关系图

根据上图，当聚类的数量在 1 到 3 之间时，聚集系数下降较多。当它大于 3 时，聚集系数的变化逐渐趋于平缓。因此，聚类类别的数量 $K=3$ 。

5.3.3 系统聚类模型结果分析

每个类的失真程度等于该类的重心和其内部成员位置之间的距离的平方之和。假设总共有 n 个样本被分为 K 个类 ($K \leq n-1$, 即至少有一个类有两个元素)，而 C_k 代表第 k 个类 ($k = 1, 2, \dots, k$)，而这个类的重心位置被记录为 u_k ，则第 k 类的失真度为： $\sum_{i \in C_k} |x_i - u_k|^2$ ，所有类的失真度为： $\sum_{i \in C_k} |x_i - u_k|^2$ 。在选择样本间距的欧氏距离平方和选择组间联系的聚类方法的基础上，使用 SPSS 绘制了 5 个高校的系统聚类结果的系谱图，如系谱图所示：

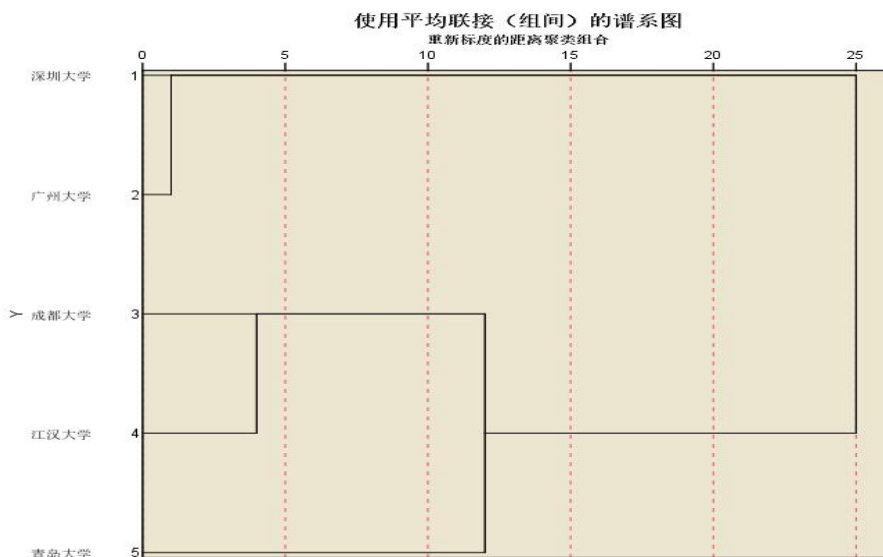


图 16 系统聚类系谱图

根据聚类的群组数量，当 $K=3$ 时，如上图所示，可将五所高校按照共同点分为三类，图中系谱图得出深圳大学和广州大学聚为一类，成都大学和江汉大学聚为一类，青岛大学单独为一类。定量数据得到，深圳大学和广州大学的共同点是所在城市经济发展快、院校经济充足、发展特色相同。成都大学和江汉大学共同点是城市 GDP 相差不大、院校发展特色相同、学科综合评估相同。从发展特色来分析得到，所有学校都是以培养新型技术人才为目的，为国家输送有技术有能力的人才，适应现代化新环境，促

进当地的经济发展，未来的发展符合城市需求，广州和深圳以现代化发展为目标，以发展创新性大学为目标，建设目标符合城市发展方向并能够满足未来需求^[3]，发展前景很好。

5.3.4 灰色关联模型建立

灰色关联是灰色系统理论的重要内容之一，灰色关联能够判别因素之间的关联程度，即因素对因素的影响程度的大小。而灰色关联分析则是对一个系统内变化趋势的定量描述的方法，若两个因素之间的差别较小，则表明两个因素之间的关联度较高，反之，关联度较低^[2]。其计算原理如下：

(1) 首先确定分析数列，由于要挖掘影响高校发展的政策关联度，故本文选择发展政策数量作为特征的母序列 $X_0(k)$

$$X_{0(k)} = \{X_{0(1)}, X_{0(2)}, \dots, X_{0(n)}\} \quad (12)$$

其余 4 个指标作为影响系统行为的因素组成的数据序列即子序列 $X_{1(k)}$ 、 $X_{2(k)} \dots X_{m(k)}$ 。

$$X_{1(k)} = \{x_{1(1)}, x_{1(2)}, \dots, x_{1(n)}\} \quad (13)$$

$$X_{2(k)} = \{x_{2(1)}, x_{2(2)}, \dots, x_{2(n)}\} \quad (14)$$

$$X_{m(k)} = \{x_{m(1)}, x_{m(2)}, \dots, x_{m(n)}\} \quad (15)$$

(2) 为了消除数据之间量纲不同带来的影响，需对原始数据进行预处理，统一数据的量纲，缩小变量的范围从而简化计算，对母序列和子序列中每个指标进行处理，求出每个指标各自的均值，用该指标中的每个元素除以其均值。得到参考数列 $X_0(k)$ ，计算得：

$X_i(k), i=1,2,\dots,m, k=1,2,\dots,n$ 。对各序列消除量纲后得值进行绝对差值计算

$$|X_0 - X_1| \dots |X_0 - X_{27}|。$$

(3) 根据公式计算各个因素得关联系数：

$$e_i(k) = \frac{\Delta(\min) + \rho \Delta(\max)}{\Delta i(k) + \rho \Delta(\max)} = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|} \quad (16)$$

上式中 $e_i(k)$ 为关联系数， $|X_0(k) - X_i(k)|$ 为在 k 时间段，两个序列之间的绝对差值， $\Delta(\max)$ 为某段时间内差的最大值， $\Delta(\min)$ 为某段时间内差的最小值。设分辨系数为 ρ ，一般 $\rho \in (0,1)$ ，通常取 $\rho = 0.5$ 。

(4) 求关联度。关联度为两个序列之间的某个因素进行对比，计算的关联系数的平均值。其计算如下所示：

$$r_i = \frac{1}{m} \sum_{i=1}^m e_i(k) \quad (17)$$

5.3.5 灰色关联模型结果分析

通过将数据导入到 MATLAB 中，通过灰色关联分析模型进行计算，得到各个指标与高校发展关联系数图如下：

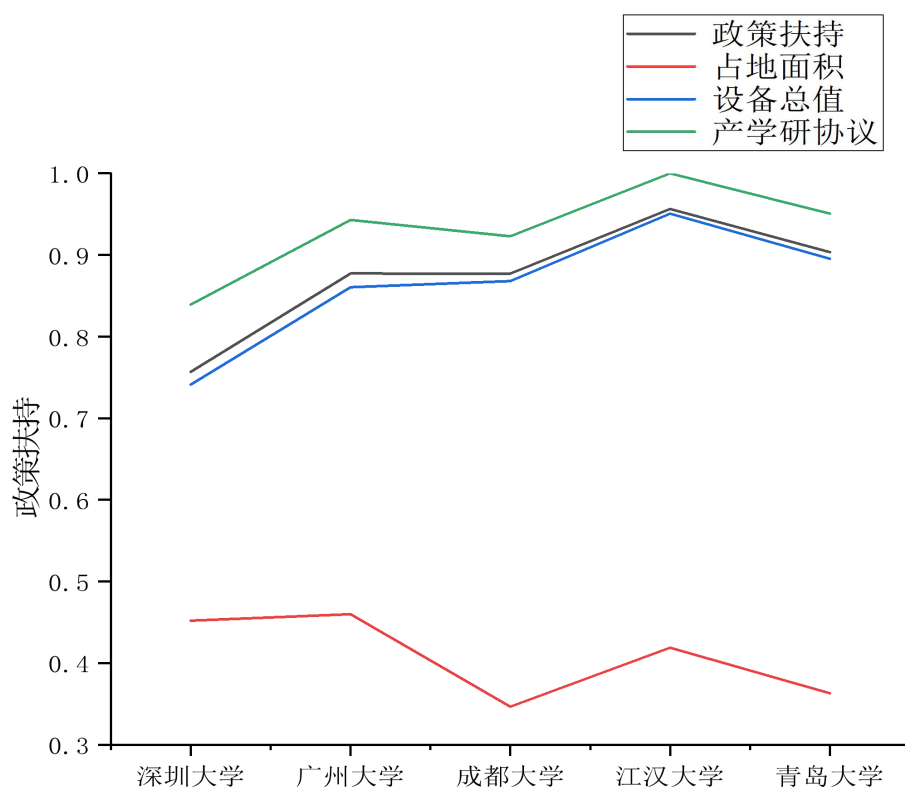


图 17 灰色关联系数图

如上图所示可以得出，市属院校发展与政策扶持、设备总值扶持和产学研协议关系大。以上图江汉大学为例，数据表明产学研协议严重影响未来发展，占地面积扶持对于高校发展影响不大，占地面积扶持对于高校发展影响较小。灰色关联分析影响高校发展关联度排名表汇总如下：

表 11 灰色关联度结果排名

评价项	关联度	排名
产学研协议/项	0.931	1
政策扶持/亿	0.874	2
设备总值/亿	0.863	3
占地面积/亩	0.408	4

从表 11 中可以对于学校发展水平的 4 个评价指标中，产学研协议关联度最高为 0.931，其次是扶持资金关联度为 0.874。说明与政府政策扶持最相关的是产学研协议，其次是政策扶持、设备总值、学校占地面积与政府扶持的相关性较小。

综上所述，若综合分析这五所大学的共同特点，则不难发现，深圳大学和广州大学所在城市均为一线城市，而成都大学、江汉大学和青岛大学所在城市也都是新一线城市，这五个市的 GDP 总量在全国范围内均较为靠前，属于国内的经济发达地区，从而间接地决定了其经费的多少。从学校的发展来看，与传统高校相对比，这些高校均有自己定位和鲜明的特色，有的以创新型为发展目标，有的以应用型城市大学为发展目标，并确立了各自的重点发展学科。在这些因素的共同作用下，使得这五所年轻高校的发展突飞猛进，并在今后成为各自地区具有代表性和带头作用的示范性高校，也将对区域的经济、教育的发展起到重要的推动作用。其中地方高校和地方政府的协同发展，促进双赢是这一政策的根本目的。在政策设计的根本和终止点上，都以这个目的为出发点并作

为评价的标准。学校产学研协议、政策扶持、设备总值发展与当地市政府的扶持政策具有较大的关联度。

六、模型的评价与推广

6.1 模型的评价

本文需要对高校和学科进行综合评估，问题一中熵权法和 TOPSIS 方法对高校进行多方面综合评估，问题二中使用秩和比法对高校进行学科定量评估，问题三中使用系统聚类分析高校发展之间的共同点，本文用到的模型都能够对选中的指标进行分析，找出指标与分析对象之间的关系，最终对分析对象进行评分排名或聚类分类。

6.2 模型的优点

(1) TOPSIS 模型

TOPSIS 模型避免了数据的主观性，不需要目标函数，不必进行测试，也很好的表征了多个指标的综合影响力，对于数据分布和样本量的多少，指标没有严格的限制，不仅适用于小样本数据，也适用于大样本、多模块指标体系，灵活方便。

(2) RSR 模型

RSR 只使用了数据的相对大小关系，而不真正运用数值本身，所以此方法综合性强，可以显示微小变动，对离群值不敏感；能够对各个评价对象进行排序分档，找出优劣，是做比较，找关系的有效手段；能够找出评价指标是否有独立性。

(3) 系统聚类算法

系统聚类算法速度快，对异常值影响很小，不需要预先选择聚类的数量，可根据结果的系数来选择聚类的数量，结果更加可靠。

6.3 模型的缺点

(1) 通过秩替代原始指标值，会损失部分信息；不容易对各个指标进行恰当的编秩。

(2) 整个理论体系目前还不是很完善，其应用受到了某些限制

七、参考文献

- [1]王蔚.基于熵权 TOPSIS 的软件项目风险评估[J].项目管理技术,2022,20(02):102-107.
- [2]赵雨.灰色关联分析法的高校公共英语教学质量评价研究[J].微型电脑应用,2022,38(01):58-61.
- [3]杨纳名,赵彬彬.高等教育评估引导高校发展的现实路径与展望[J].当代教育与文化,2021,13(01):56-60.DOI:10.13749/j.cnki.cn62-1202/g4.2021.01.009.
- [4]梅晓仁,夏青.基于系统聚类分析的高校教师绩效评价模型及实证研究[J].金融理论与教学,2020(04):89-90.DOI:10.13298/j.cnki.ftat.2020.04.016.

八、附录

1.Topsis 代码

```
clear;clc
```

```
load X.mat
```

```
[n,m] = size(X);
```

```
disp(['共有' num2str(n) '个评价对象,' num2str(m) '个评价指标'])
```

```
Judge = input(['这' num2str(m) '个指标是否需要经过正向化处理，需要请输入 1 ，不需要
```

输入 0: ');

if Judge == 1

 position = input('请输入需要正向化处理的指标所在的列，例如第 2、3、6 三列需要处理，那么你需要输入[2,3,6]: ');

 disp('请输入需要处理的这些列的指标类型（1：极小型， 2：中间型， 3：区间型）')

 Type = input('例如：第 2 列是极小型，第 3 列是区间型，第 6 列是中间型，就输入[1,3,2]: ');

 for i = 1 : size(position,2) %这里需要对这些列分别处理，因此我们需要知道一共要处理的次数，即循环的次数

 X(:,position(i)) = positivization(X(:,position(i)),Type(i),position(i));

 end

 disp('正向化后的矩阵 X = ')

 disp(X)

end

Z = X ./ repmat(sum(X.*X) .^ 0.5, n, 1);

disp('标准化矩阵 Z = ')

disp(Z)

D_P = sum([(Z - repmat(max(Z),n,1)) .^ 2],2) .^ 0.5

D_N = sum([(Z - repmat(min(Z),n,1)) .^ 2],2) .^ 0.5

S = D_N ./ (D_P+D_N)

disp('Final Score: ')

stand_S = S / sum(S)

[sorted_S,index] = sort(stand_S,'descend')

function [posit_x] = Mid2max(x,best)

 M = max(abs(x-best));

 posit_x = 1 - abs(x-best) / M;

end

2.DBSCAN code

clc;

clear;

close all;

%% Input Data

X = xlsread('Latitude and longitude.xlsx');

%% Run DBSCAN Clustering Algorithm

```

epsilon=0.5;
MinPts=10;
IDX=DBSCAN(X,epsilon,MinPts);

%% Plot Results

PlotClusterinResult(X, IDX);
title(['DBSCAN Clustering (\epsilon = ' num2str(epsilon) ', MinPts = ' num2str(MinPts)
')]);
%DBSCAN
function [IDX, isnoise]=DBSCAN(X,epsilon,MinPts)

    C=0;

    n=size(X,1);
    IDX=zeros(n,1); % Initialize all to 0, that is, all noise points

    D=pdist2(X,X);

    visited=false(n,1);
    isnoise=false(n,1);

    for i=1:n
        if ~visited(i)
            visited(i)=true;

            Neighbors=RegionQuery(i);
            if numel(Neighbors)<MinPts
                % X(i,:) is NOISE
                isnoise(i)=true;
            else
                C=C+1;
                ExpandCluster(i,Neighbors,C);
            end
        end
    end

end

function ExpandCluster(i,Neighbors,C)
    IDX(i)=C;

    k = 1;
    while true

```

```

        j = Neighbors(k);

        if ~visited(j)
            visited(j)=true;
            Neighbors2=RegionQuery(j);
            if numel(Neighbors2)>=MinPts
                Neighbors=[Neighbors Neighbors2];    %#ok
            end
        end
        if IDX(j)==0
            IDX(j)=C;
        end

        k = k + 1;
        if k > numel(Neighbors)
            break;
        end
    end
end

function Neighbors=RegionQuery(i)
    Neighbors=find(D(i,:)<=epsilon);
end

```

end

2、灰色关联分析代码

```
clear;clc
```

```
load data.mat
```

```
Mean = mean(data);
```

```
data = data ./ repmat(Mean,size(data,1),1);
```

```
disp('预处理后的矩阵为: '); disp(data)
```

```
Y = data(:,1);    % 母序列
```

```
X = data(:,2:end); % 子序列
```

```
absX0_Xi = abs(X - repmat(Y,1,size(X,2)))    % 计算|X0-Xi|
```

```
a = min(min(absX0_Xi))
```

```
b = max(max(absX0_Xi))
```

```
rho = 0.5; % 分辨系数取 0.5
```

```
gamma = (a+rho*b) ./ (absX0_Xi + rho*b)    的关联系数
```

```
disp('子序列中各个指标的灰色关联度分别为: ')
```

```
disp(mean(gamma))
```