

基于灰关联度的面板数据聚类方法 及在空气污染分析中的应用

党耀国, 朱晓月[†], 丁 松, 王俊杰

(南京航空航天大学 经济与管理学院, 南京 211106)

摘 要: 针对面板数据的三维特征、正负相关性以及灰色关联模型缺乏有效检验等问题, 构建面板数据灰色关联模型, 拓展了面板数据聚类方法; 引入灰熵, 提出灰色关联聚类的检验模型. 研究发现, 灰色关联度模型能够反映面板数据的正、负相关关系, 且具有对称性、唯一性和可比性. 通过在苏南 4 市空气质量区域划分中的应用, 表明基于面板数据灰关联模型的聚类方法具有良好的效果, 各类别的灰色关联度差异明显, 层次清晰.

关键词: 灰色关联度; 面板数据; 聚类分析; 斜率

中图分类号: N945.5

文献标志码: A

Panel data clustering method based on grey correlation degree and its application in air pollution analysis

DANG Yao-guo, ZHU Xiao-yue[†], DING Song, WANG Jun-jie

(College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: According to some problems of the grey relational model for panel data, such as the three-dimensional feature, the positive and negative relation of the panel data, and the model being lack of effective testing, the panel data grey relational model is built to expand the panel data clustering method. Grey entropy is introduced into the model and the test model of grey clustering model is proposed. The study shows that the grey relational model can reflect the positive and negative relation of the panel data and possess symmetry property, unique property and comparable property. The application of example in air pollution analysis indicates that the method has good effect, and the grey correlation degree of each category is obvious.

Keywords: grey relational degree; panel data; clustering method; slope

0 引 言

随着我国经济的快速发展和城镇化进程的加快, 汽车保有量急剧上升, 空气污染现象愈演愈烈, 特别是近年来大规模的雾霾现象已严重影响人们的正常生活. 空气污染的形成原因较为复杂, 影响因素众多, 其中污染源排放和当期气象条件对污染物的分布和扩散起主要作用. 空气污染源的排放与地区经济发展速度和水平紧密相关, 而经济发展水平和气象条件都呈现出明显的区域性特征, 因此研究空气污染的区域性划分将有助于提高各地区的空气污染治理效率, 推动空气污染区域协同治理.

在空气污染区域划分研究方面, 段玉森等^[1]运用 EOF 及 Morlet 小波方法分析了全国 47 个环保重点城市 API 指数的时空分布特征; 藺雪芹等^[2]基于空气质量空间效应检验模型对全国受控的 161 个城市的空气质量区域特征进行了研究; 薛安等^[3]以城市作为节点, 以城市间 PM_{2.5} 浓度的相关性与距离的比值作为边的权重, 构建了中国城市 PM_{2.5} 加权网络, 并采用 GN 算法对网络进行划分, 得到不同季节中国 PM_{2.5} 污染的区域分布情况; 张宝春等^[4]利用 GIS 技术, 采用网格化方法研究了珠三角区域近 5 年来的空气质量特征, 解决了空气质量分等定级过程中的区域边界

收稿日期: 2016-11-15; 修回日期: 2017-01-26.

基金项目: 国家自然科学基金项目 (71071077, 71371098); 中央高校基本科研业务费专项资金项目 (NC2012001, NR2013033); 江苏高校哲学社会科学重点研究基地项目 (2012JDXM005); 江苏省普通高校研究生科研创新计划项目 (SJZZ16_0050).

作者简介: 党耀国 (1964—), 男, 教授, 博士生导师, 从事灰色系统理论、数量经济等研究; 朱晓月 (1992—), 女, 硕士生, 从事灰色系统理论及其应用的研究.

[†]通讯作者. E-mail: zxy_luyao@163.com

划分问题. 上述研究为我国空气污染区域划分提供了研究思路, 但目前研究者们主要采用单个空气质量指标来进行区域划分, 综合多维指标信息来进行区域划分的研究尚不多见. 另外, 由于我国对空气污染现象的研究起步较晚, 多维指标数据不完整, 可用数据偏少, 因此本文采用以“小样本、贫信息”的不确定系统为研究对象的面板数据灰色关联分析方法对空气污染区域划分进行研究.

目前, 国内对面板数据灰色关联分析的研究处于起步阶段. 张可等^[5]基于面板数据的曲面簇描述方法, 在三维空间讨论面板数据各指标的几何特征相似性; 郭昆等^[6]将灰关联分析与谱聚类结合, 利用序列的均衡接近度描述数据点的相似程度; 钱吴永等^[7]将表征面板数据时空特征的“水平”距离、“增量”距离和“变异”距离引入灰色关联度计算模型中; 吴利丰等^[8]用海塞矩阵定义凸度, 用数据的凸性表征样本之间的相似程度, 提出了三维灰色凸关联度; 刘震等^[9]利用面板数据的几何特征构建了灰色网格关联度; 崔立志等^[10]用发展指数和增长速度指数度量相关因素矩阵和系统特征行为矩阵的接近程度, 构建了灰色矩阵相似关联度模型; 李雪梅等^[11]构造了所有指标不同对象下时间序列的累加生成序列, 用生成序列的平均生成速率表征原序列的动态变化趋势, 综合偏离、差离和分离的三重差异信息, 构建了指标关联分析模型. 上述研究推动了面板数据灰关联分析的发展, 丰富了面板数据灰色关联分析理论, 但仍存在以下几个问题:

- 1) 用曲面簇描述面板数据, 曲面簇的表示方式不同将影响关联度的大小;
- 2) 通过降维或者提取时空特征来构建灰色关联模型会导致信息缺失;
- 3) 通过平均加权的方式来综合各维度的灰色关联度, 忽视了各维度灰关联度的差异, 导致聚类结果可靠性不高;
- 4) 对模型的性质讨论以及应用有待进一步的研究.

本文基于灰色斜率关联度的基本思想, 对面板数据的三维特征, 将样本的行为矩阵划分为时间维和指标维, 对各维度点的增量进行标准化, 同时综合利用增量的构成比和构成差来定义灰关联系数, 构建面板数据灰关联度模型. 此外, 引入灰熵对模型的聚类结果进行检验, 并将模型应用于苏南4市空气质量区域划分中.

1 面板数据灰色关联模型及检验模型构建

1.1 多指标面板数据的表示

面板数据是多个样本在不同时点的多个指标值构成的数据集, 包含时间维、对象维和指标维3个维度的信息. 每个样本中的各个指标在不同时点的观测值可映射为三维空间中的点.

定义1 设面板数据 X 中第 i 个样本的第 s 个指标在 t 时刻的值记为 $x_i(s, t)$, 其中: $i = 1, 2, \dots, N$, $s = 1, 2, \dots, m$, $t = 1, 2, \dots, n$, 则称

$$X_i(s, t) = \begin{bmatrix} x_i(1, 1) & x_i(1, 2) & \cdots & x_i(1, n) \\ x_i(2, 1) & x_i(2, 2) & \cdots & x_i(2, n) \\ \vdots & \vdots & \ddots & \vdots \\ x_i(m, 1) & x_i(m, 2) & \cdots & x_i(m, n) \end{bmatrix}$$

为样本 i ($i = 1, 2, \dots, N$) 的行为矩阵, 则面板数据可表示为

$$X = \{X_1(s, t), X_2(s, t), \dots, X_N(s, t)\}, \\ s = 1, 2, \dots, m, t = 1, 2, \dots, n.$$

1.2 面板数据灰关联模型构建

本文提出的面板数据灰色关联模型综合了面板数据的三维特征和斜率关联度的建模思想, 将样本的行为矩阵划分为时间维和指标维. 在各维度, 首先将各点的增量进行标准化, 同时综合利用增量的构成比和构成差来定义关联系数, 更充分地利用数据序列所包含的信息, 提升聚类效果. 此外, 引入一个符号函数来反映不同样本在同一维度的正负相关关系, 使得当增量方向一致或者相等时, 关联系数为正; 当增量方向相反时, 关联系数为负.

若样本数据多维指标的量纲不同, 则需要对样本的行为矩阵进行初始化处理.

定义2 设多指标面板数据中第 i 个样本在时刻 t 的第 s 个指标的观测值为 $x_i(s, t)$, 在初始时刻 t_0 的第 s 个指标观测值为 $x_i(s, t_0)$. 若 $x_i(s, t)$ 满足

$$x_i(s, t)D = x_i(s, t)/x_i(s, t_0),$$

则称 D 为初值化算子.

定义3 设样本 i 的行为矩阵为

$$X_i(s, t) = \begin{bmatrix} x_i(1, 1) & x_i(1, 2) & \cdots & x_i(1, n) \\ x_i(2, 1) & x_i(2, 2) & \cdots & x_i(2, n) \\ \vdots & \vdots & \ddots & \vdots \\ x_i(m, 1) & x_i(m, 2) & \cdots & x_i(m, n) \end{bmatrix},$$

其中 $i = 1, 2, \dots, N$. 令

$$H_i(s, t) = x_i(s, t)D - x_i(s-1, t)D, \\ s = 2, 3, \dots, m, t = 1, 2, \dots, n;$$

$$W_i(s, t) = x_i(s, t)D - x_i(s, t - 1)D,$$

$$s = 1, 2, \dots, m, t = 2, 3, \dots, n.$$

则称 $H_i(s, t)$ 为样本 i 在时刻 t 从指标 $s - 1$ 到指标 s 的增量, $W_i(s, t)$ 为样本 i 的 s 指标在时点 $t - 1$ 到时点 t 的增量.

$$C_i(t) = \frac{\sum_{s=2}^m |H_i(s, t)|}{m - 1},$$

$$i = 1, 2, \dots, N, t = 1, 2, \dots, n;$$

$$D_i(s) = \frac{\sum_{t=2}^n |W_i(s, t)|}{n - 1},$$

$$i = 1, 2, \dots, N, s = 1, 2, \dots, m.$$

$C_i(t)$ 表示对象 i 在时刻 t 指标维增量绝对值的平均值, $D_i(s)$ 表示对象 i 的 s 指标在各个时段增量绝对值的平均值.

$$U_i(s, t) = \frac{H_i(s, t)}{C_i(t)},$$

$$i = 1, 2, \dots, N, s = 2, 3, \dots, m, t = 1, 2, \dots, n;$$

$$V_i(s, t) = \frac{W_i(s, t)}{D_i(s)},$$

$$i = 1, 2, \dots, N, s = 1, 2, \dots, m, t = 2, 3, \dots, n.$$

$U_i(s, t)$ 表示对象 i 在时刻 t 从指标 $s - 1$ 到 s 的增量均值化, $V_i(s, t)$ 表示对象 i 的 s 指标在时点 $t - 1$ 到时点 t 的增量均值化.

灰色斜率关联度按照序列曲线的相对变化势态的接近程度来计算关联度, 斜率之间越接近, 关联程度越大. 考虑数据均为单位时距序列, 指标间的距离也视为1, 因此在指标维, 从指标 $s - 1$ 到指标 s 的斜率等于 $H_i(s, t)$; 在时间维, 从时刻 $t - 1$ 到 t 的斜率等于 $W_i(s, t)$. 通过标准化得到 $U_i(s, t)$ 和 $V_i(s, t)$, 则 $|U_i(s, t)|$ 与 $|U_j(s, t)|$ 越接近, 面板数据 X_i 与 X_j 在指标 $s - 1$ 到 s 的关联程度越大; $|V_i(s, t)|$ 与 $|V_j(s, t)|$ 越接近, 面板数据 X_i 与 X_j 在时点 $t - 1$ 到时点 t 的关联程度越大. 因此, 给出如下定义.

定义4 设有面板数据 X_i 和 X_j , 若 $U_i(s, t)$ 和 $U_j(s, t)$ 不同时为0, 令

$$r_{ij}^H(s, t) = \text{sgn}(U_i(s, t) \times U_j(s, t)) / \left[1 + \frac{1}{2} \left(|U_i(s, t)| - |U_j(s, t)| + \frac{1}{2} \left(1 - \frac{\min(|U_i(s, t)|, |U_j(s, t)|)}{\max(|U_i(s, t)|, |U_j(s, t)|)} \right) \right) \right],$$

若 $U_i(s, t)$ 和 $U_j(s, t)$ 同时为0, 令 $r_{ij}^H(s, t) = 1$, 则称 $r_{ij}^H(s, t)$ ($s = 2, 3, \dots, m, t = 1, 2, \dots, n$) 为 X_i 与 X_j 在指标维的灰色关联系数.

同理, 若 $V_i(s, t)$ 和 $V_j(s, t)$ 不同时为0, 令

$$r_{ij}^W(s, t) = \text{sgn}(V_i(s, t) \times V_j(s, t)) / \left[1 + \frac{1}{2} \left(|V_i(s, t)| - |V_j(s, t)| + \frac{1}{2} \left(1 - \frac{\min(|V_i(s, t)|, |V_j(s, t)|)}{\max(|V_i(s, t)|, |V_j(s, t)|)} \right) \right) \right],$$

若 $V_i(s, t)$ 和 $V_j(s, t)$ 同时为0, 令 $r_{ij}^W(s, t) = 1$, 则称 $r_{ij}^W(s, t)$ ($s = 1, 2, \dots, m, t = 2, 3, \dots, n$) 为 X_i 与 X_j 在时间维的灰色关联系数. 其中

$$\text{sgn}(U_i(s, t) \times U_j(s, t)) = \begin{cases} 1, & U_i(s, t) \times U_j(s, t) \geq 0; \\ -1, & U_i(s, t) \times U_j(s, t) < 0; \end{cases}$$

$$\text{sgn}(V_i(s, t) \times V_j(s, t)) = \begin{cases} 1, & V_i(s, t) \times V_j(s, t) \geq 0; \\ -1, & V_i(s, t) \times V_j(s, t) < 0. \end{cases}$$

定义5 设有矩阵数据 X_i 和 X_j , 称

$$\gamma_{ij}^H = \frac{\sum_{s=2}^m \sum_{t=1}^n r_{ij}^H(s, t)}{n(m - 1)}$$

为 X_i 与 X_j 在指标维的灰色关联度, 称

$$\gamma_{ij}^W = \frac{\sum_{s=1}^m \sum_{t=2}^n r_{ij}^W(s, t)}{m(n - 1)}$$

为 X_i 与 X_j 在时间维的灰色关联度.

通过初值化算子将样本行为矩阵无量纲化后, 按文中定义3、定义4和定义5计算得到的增量、灰关联系数和灰关联度均为无量纲数据. 面板数据间的关联度包含多个维度的关联信息, 各维度的关联度 γ_{ij}^H 、 γ_{ij}^W 越大, 则面板数据间的关联程度越大, 由此给出如下定义.

定义6 设有矩阵数据 X_i 和 X_j , 若

$$\gamma_{ij} = \frac{1}{2}(\gamma_{ij}^H + \gamma_{ij}^W),$$

则称 γ_{ij} 为面板数据 X_i 与 X_j 间的灰色综合关联度.

1.3 检验模型构建

从定义6可以看出, 面板数据 X_i 与 X_j 间的灰色综合关联度 γ_{ij} 由时间维、指标维的灰色关联度 γ_{ij}^H 和 γ_{ij}^W 加权平均得到, 当不同维度的灰色关联度 γ_{ij}^H 与 γ_{ij}^W 差异过大时, 灰色综合关联度 γ_{ij} 不能体现其维度的关联程度差异, 导致基于灰色综合关联度 γ_{ij} 的聚类结果可靠性不高. 因此, 引入灰熵来刻画面板数据 X_i 与 X_j 各维度的关联度差异, 进一步验证灰关联聚类结果的合理性.

定义7 设有面板数据 X_i 和 X_j , γ_{ij}^H 和 γ_{ij}^W 分别

为 X_i 与 X_j 在时间维和指标维的灰色关联度, 令 $P_{ij} = |\gamma_{ij}^H| + |\gamma_{ij}^W|$, 则称

$$P_{ij}^H = \frac{|\gamma_{ij}^H|}{P_{ij}},$$

$$P_{ij}^W = \frac{|\gamma_{ij}^W|}{P_{ij}}$$

为面板数据 X_i 与 X_j 灰关联分布映射, 定义 $T_{ij} = -P_{ij}^H \ln P_{ij}^H - P_{ij}^W \ln P_{ij}^W$ 为面板数据 X_i 与 X_j 灰色综合关联度的检验系数.

由定义7可知, 面板数据 X_i 与 X_j 检验系数 T_{ij} 越大, 表明面板数据 X_i 与 X_j 在时间维和指标维的灰色关联度 $|\gamma_{ij}^H|$ 与 $|\gamma_{ij}^W|$ 差异程度越小, 则灰色综合关联度 γ_{ij} 可靠性越高.

1.4 面板数据聚类方法

基于灰色综合关联度和检验模型的面板数据聚类步骤如下:

Step 1: 根据定义1, 将面板数据转化为样本行为矩阵 X_i 与 X_j ;

Step 2: 对原始数据进行初值化处理, 消除量纲影响;

Step 3: 计算样本在时间维和指标维的增量, 并进行标准化;

Step 4: 按照定义4计算样本行为矩阵 X_i 与 X_j 在时间维和指标维的灰色关联矩阵 $\gamma_{ij}^H(s, t)$ 和 $\gamma_{ij}^W(s, t)$.

Step 5: 按照定义5和定义6计算面板数据 X_i 与 X_j 在时间维和指标维的灰色关联度 γ_{ij}^H 、 γ_{ij}^W 及灰色综合关联度 γ_{ij} .

Step 6: 按照定义7计算灰色关联度检验系数 T_{ij} , 判别灰色综合关联度的可靠性.

Step 7: 设定关联度临界值 γ , 根据灰色综合关联度和检验系数对样本进行聚类分析.

2 面板数据灰色综合关联度模型性质

定理1 面板数据灰色综合关联度 γ 具有如下性质: 1) $-1 \leq \gamma \leq 1$; 2) 对称性; 3) 接近性, 即 X_i 与 X_j 越接近, γ 越大; 4) 唯一性; 5) 可比性.

证明 1) 由于 $-1 \leq r_{ij}^H(s, t) \leq 1$, 则

$$-1 \leq \frac{\sum_{s=2}^m \sum_{t=1}^n r_{ij}^H(s, t)}{n(m-1)} \leq 1.$$

同理可证

$$-1 \leq \frac{\sum_{s=1}^m \sum_{t=2}^n r_{ij}^W(s, t)}{m(n-1)} \leq 1.$$

因为

$$1 \leq \frac{1}{2} \left(\frac{\sum_{s=2}^m \sum_{t=1}^n r_{ij}^H(s, t)}{n(m-1)} + \frac{\sum_{s=1}^m \sum_{t=2}^n r_{ij}^W(s, t)}{m(n-1)} \right) \leq 1,$$

所以 $-1 \leq \gamma_{ij} \leq 1$.

2)

$$\gamma_{ij} = \frac{1}{2} \left(\frac{\sum_{s=2}^m \sum_{t=1}^n r_{ij}^H(s, t)}{n(m-1)} + \frac{\sum_{s=1}^m \sum_{t=2}^n r_{ij}^W(s, t)}{m(n-1)} \right),$$

由面板数据灰色综合关联度的定义, 性质2) ~ 性质

5) 显然成立. \square

3 模型应用及分析

3.1 样本数据来源

本文以江苏省为研究区域, 根据2012年中国颁布的《环境空气质量标准》(GB3095-2012), 同时考虑数据的可获取性, 选取市区环境空气质量达二级标准天数(以下简称为DAQ)、二氧化硫年平均浓度/微克每立方米(以下简称为SO₂年平均浓度)、二氧化氮年平均浓度/微克每立方米(以下简称为NO₂年平均浓度)、可吸入颗粒物年平均浓度/微克每立方米(以下简称为PM10年平均浓度)4项指标来综合反映各城市的空气质量状况, 数据来源于2012年~2015年各城市环境状况公报. 由于部分城市统计数据不完整, 选取南京、苏州、无锡、镇江4个城市的空气质量指标建立面板数据, 具体如表1所示. 基于面板数据灰色关联聚类模型和检验模型分析苏南4市空气污染的区域性特征.

表1 南京、苏州、无锡、镇江4个城市空气质量指标

地区	年份	DAQ	SO ₂ 年平均浓度	NO ₂ 年平均浓度	PM10年平均浓度
南京	2015	235	96	50	19
	2014	190	123	54	25
	2013	202	137	55	37
	2012	317	102	51	33
苏州	2015	244	80	54	21
	2014	232	86	53	24
	2013	217	95	53	31
无锡	2012	339	79	49	30
	2015	211	94	41	26
	2014	234	105	45	29
镇江	2013	199	104	47	45
	2012	344	81	44	49
	2015	263	82	42	25
镇江	2014	240	107	46	24
	2013	223	124	42	30
镇江	2012	347	93	34	29

3.2 模型计算及分析

由表1得到样本行为矩阵为

$$X_1 = \begin{bmatrix} 235 & 96 & 50 & 19 \\ 190 & 123 & 54 & 25 \\ 202 & 137 & 55 & 37 \\ 317 & 102 & 51 & 33 \end{bmatrix},$$

$$X_2 = \begin{bmatrix} 244 & 80 & 54 & 21 \\ 232 & 86 & 53 & 24 \\ 217 & 95 & 53 & 31 \\ 339 & 79 & 49 & 30 \end{bmatrix},$$

$$X_3 = \begin{bmatrix} 211 & 94 & 41 & 26 \\ 234 & 105 & 45 & 29 \\ 199 & 104 & 47 & 45 \\ 344 & 81 & 44 & 49 \end{bmatrix},$$

$$X_4 = \begin{bmatrix} 263 & 82 & 42 & 25 \\ 240 & 107 & 46 & 24 \\ 223 & 124 & 42 & 30 \\ 347 & 93 & 34 & 29 \end{bmatrix}.$$

按面板数据聚类Step 1 ~ Step 5计算得到样本行为矩阵在时间维和指标维的灰色关联度矩阵及灰色综合关联度矩阵,结果如下:

$$\gamma_{ij}^H = \begin{bmatrix} 1 & 0.623 & 0.483 & 0.423 \\ & 1 & 0.592 & 0.590 \\ & & 1 & 0.478 \\ & & & 1 \end{bmatrix},$$

$$\gamma_{ij}^W = \begin{bmatrix} 1 & 0.531 & 0.319 & 0.439 \\ & 1 & 0.284 & 0.553 \\ & & 1 & 0.162 \\ & & & 1 \end{bmatrix},$$

$$\gamma_{ij} = \begin{bmatrix} 1 & 0.577 & 0.401 & 0.431 \\ & 1 & 0.438 & 0.571 \\ & & 1 & 0.320 \\ & & & 1 \end{bmatrix}.$$

按照定义7计算灰关联检验系数,得到检验矩阵为

$$T_{ij} = \begin{bmatrix} 1 & 0.690 & 0.672 & 0.693 \\ & 1 & 0.630 & 0.693 \\ & & 1 & 0.566 \\ & & & 1 \end{bmatrix}.$$

按照文献[9]所介绍的基于面板数据的灰色网格关联度模型,计算得到灰关联矩阵

$$\gamma_{ij} = \begin{bmatrix} 1 & 0.595 & 0.434 & 0.497 \\ & 1 & 0.439 & 0.605 \\ & & 1 & 0.336 \\ & & & 1 \end{bmatrix}.$$

取 $\gamma = 0.5$, 将灰色综合关联度 γ_{ij} 大于等于 0.5 的城市划分为相同的空气污染级别, 将灰色综合关联度 γ_{ij} 小于 0.5 的城市划分为不同的污染级别. 由文献[9]灰色网格关联度矩阵得到的聚类结果为南京、苏州、镇江为一类, 无锡为一类. 由灰色综合关联度矩阵得到的聚类结果为南京、苏州、镇江为一类, 无锡为一类, 两个模型结论基本一致. 但当各维度灰关联度的差异过大时, 文献[9]面板数据的灰色网格关联度模型无法有效识别, 本文提出检验模型对灰关联聚类的结果进行评估. 从检验矩阵可以发现, 南京和苏州、苏州和镇江的灰关联度检验系数较高, 因此南京和苏州、苏州和镇江的灰色综合关联度可靠性较高.

结合灰色关联矩阵和检验矩阵给出的信息, 可将南京、苏州和镇江划分为同一类别. 这与现实情况基本相符, 南京与苏州经济发展方式和城市化水平相近, 空气污染状况较为严重; 南京与镇江地理位置相邻, 空气质量相互影响较大; 无锡市空气质量优于其他3个城市. 空气污染的流动性特征决定其治理无法由某一地方政府独立而有效地完成, 基于本文的聚类分析结果, 在空气污染治理方式上, 南京、苏州和镇江的政府企业相互合作、协同治理, 将极大提高空气污染治理效率.

4 结论

本文针对面板数据的三维特征和正负相关性, 基于斜率关联度基本思想, 构建了面板数据灰色关联度模型, 并讨论了模型性质, 拓展了灰色关联度聚类方法; 提出了灰色关联聚类的检验模型, 进一步丰富了面板数据灰色关联分析理论, 并为研究多维数据灰色关联检验模型提供了研究思路; 在此基础上, 应用面板数据灰色关联度模型和检验模型分析苏南4市空气污染的区特征, 划分了空气污染治理区域, 验证了模型的有效性, 为空气污染区域协同治理提供了理论依据.

参考文献(References)

[1] 段玉森, 魏海萍, 伏晴艳, 等. 中国环保重点城市API指数的时空模态区域分异[J]. 环境科学学报, 2008, 28(2): 384-391.

- (Duan Y S, Wei H P, Fu Q Y, et al. Regional spatio-temporal mode differences of air pollution index of key environmental protection cities in China[J]. *Acta Scientiae Circumstantiae*, 2008, 28(2): 384-391.)
- [2] 藺雪芹, 王岱. 中国城市空气质量时空演化特征及社会经济驱动力[J]. *地理学报*, 2016, 71(8): 1357-1371.
(Lin X Q, W D. Spatio-temporal variations and socio-economic driving forces of air quality in Chinese cities[J]. *Acta Geographica Sinica*, 2016, 71(8): 1357-1371.)
- [3] 薛安, 耿恩泽. 基于复杂网络的中国城市MF区域划分[J]. *应用基础与工程科学学报*, 2015, 23(增刊): 68-78.
(Xue A, Geng E Z. Region division study of PM_{2.5} pollution in cities of China based on complex networks[J]. *J of Basic Science and Engineering*, 2015, 23(S): 68-78.)
- [4] 张宝春, 陈彦军, 李伟铿, 等. 基于GIS的珠三角区域空气质量时空特征研究[J]. *生态环境学报*, 2011, 20(4): 600-605.
(Zhang B C, Chen Y J, Li W K, et al. The research of spatial and temporal characteristics on air quality level in Pearl River Delta based on GIS[J]. *Ecology and Environmental Sciences*, 2011, 20(4): 600-605.)
- [5] 张可, 刘思峰. 灰色关联聚类在面板数据中的应用及扩展[J]. *系统工程理论与实践*, 2010, 30(7): 1253-1259.
(Zhang K, Liu S F. Extended clusters of grey incidences for panel data and its application[J]. *Systems Engineering — Theory & Practice*, 2010, 30(7): 1253-1259.)
- [6] 郭昆, 张岐山. 基于灰关联分析的谱聚类[J]. *系统工程理论与实践*, 2010, 30(7): 1260-1265.
(Guo K, Zhang Q S. Spectral clustering based on grey relational analysis[J]. *Systems Engineering — Theory & Practice*, 2010, 30(7): 1260-1265.)
- [7] 钱吴永, 王育红, 党耀国. 基于多指标面板数据的灰色矩阵关联模型及其应用[J]. *系统工程*, 2013, 31(10): 70-74.
(Qian W Y, Wang Y H, Dang Y G. Grey matrix relational modeling and its application based on multivariate panel data[J]. *Systems Engineering*, 2013, 31(10): 70-74.)
- [8] 吴利丰, 刘思峰. 基于灰色凸关联度的面板数据聚类方法及应用[J]. *控制与决策*, 2013, 28(7): 1033-1037.
(Wu L F, Liu S F. Panel data clustering method based on grey convex relation and its application[J]. *Control and Decision*, 2013, 28(7): 1033-1037.)
- [9] 刘震, 党耀国, 钱吴永, 等. 基于面板数据的灰色网络关联度模型[J]. *系统工程理论与实践*, 2014, 34(4): 991-996.
(Liu Z, Dang Y G, Qian W Y, et al. Grey grid incidence model based on panel data[J]. *Systems Engineering — Theory & Practice*, 2014, 34(4): 991-996.)
- [10] 崔立志, 刘思峰. 基于面板数据的灰色矩阵相似关联模型及其应用[J]. *中国管理科学*, 2015, 23(11): 171-176.
(Cui L Z, Liu S F. Grey matrix similarity correlation model based on panel data and its application[J]. *J of Management Science in China*, 2015, 23(11): 171-176.)
- [11] 李雪梅, 党耀国, 王俊杰. 面板数据下的灰色指标关联聚类模型与应用[J]. *控制与决策*, 2015, 30(8): 1447-1452.
(Li X M, Dang Y G, Wang J J. Grey relational clustering model for panel data clustering on indicators and its application[J]. *Control and Decision*, 2015, 30(8): 1447-1452.)
- [12] 孙玉刚, 党耀国. 灰色 T 型关联度的改进[J]. *系统工程理论与实践*, 2008, 28(4): 135-139.
(Sun Y G, Dang Y G. Improvement on grey T 's correlation degree[J]. *Systems Engineering — Theory & Practice*, 2008, 28(4): 135-139.)
- [13] 张岐山, 邓聚龙. 灰关联熵分析方法[J]. *系统工程理论与实践*, 1996, 16(8): 7-11.
(Zhang Q S, Deng J L. The grey relation entropy analysis method[J]. *Systems Engineering — Theory & Practice*, 1996, 16(8): 7-11.)

(责任编辑: 齐 霖)