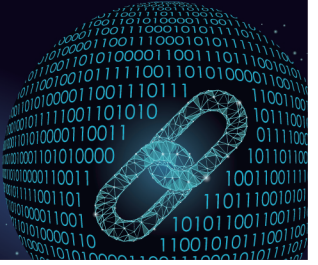




中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



基于面板数据的灰色点集关联分析模型

翟艳丽, 罗格格

引用本文:

翟艳丽, 罗格格. 基于面板数据的灰色点集关联分析模型[J]. 控制与决策, 2024, 39(8): 2801–2809.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.2044>

您可能感兴趣的其他文章

Articles you may be interested in

面向分布式在线学习的共享数据方法

A sharing data approach oriented to distributed online learning

控制与决策. 2021, 36(8): 1871–1880 <https://doi.org/10.13195/j.kzyjc.2019.1811>

基于Fisher Score与最大信息系数的齿轮箱故障特征选择方法

Fault feature selection method of gearbox based on Fisher Score and maximum information coefficient

控制与决策. 2021, 36(9): 2234–2240 <https://doi.org/10.13195/j.kzyjc.2019.1770>

基于WGRA-FCM样本相似性度量的转炉炼钢终点碳温软测量方法

End point carbon temperature measurement method based on WGRA-FCM for sample similarity measurement

控制与决策. 2021, 36(9): 2170–2178 <https://doi.org/10.13195/j.kzyjc.2020.0128>

基于低秩矩阵恢复的视觉显著性目标检测与细化

Saliency object detection and refinement based on low rank matrix recovery

控制与决策. 2021, 36(7): 1707–1713 <https://doi.org/10.13195/j.kzyjc.2019.1795>

大群体应急决策中考虑属性关联的偏好信息融合方法

Preference information fusion method of large groups emergency decision-making based on attributes association

控制与决策. 2021, 36(10): 2537–2546 <https://doi.org/10.13195/j.kzyjc.2020.0117>

基于面板数据的灰色点集关联分析模型

翟艳丽[†], 罗格格

(华北水利水电大学 数学与统计学院, 郑州 450046)

摘要: 针对面板数据样本行为矩阵间的关联性分析, 从接近性视角提出基于面板数据的灰色点集关联分析模型. 首先, 将样本行为矩阵的二元指标子矩阵投射为空间中的点集, 利用对应点的高度差建立对应点集的绝对离差矩阵, 依据此矩阵中数据的平均值和标准差分别建立对应点集的绝对离差平均值关联系数和绝对离差标准差关联系数, 基于此, 构建灰色点集关联分析模型; 然后, 讨论模型的规范性、接近性、正定性、对称性等性质, 对比分析发现, 模型不仅不受指标排序的影响, 且能够感知样本行为矩阵间数据差距的变化并有效反映其接近性; 最后, 利用灰色点集关联分析模型对豫中4个城市的空气质量进行定量评价, 并基于此进行豫中4个城市的关联聚类, 将豫中划分为3类污染区域, 以期对豫中大气污染联防联控的区域划分提供理论支持.

关键词: 面板数据; 接近性关联度; 点集; 绝对离差; 平均值; 标准差

中图分类号: N941.5

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.2044

引用格式: 翟艳丽, 罗格格. 基于面板数据的灰色点集关联分析模型[J]. 控制与决策, 2024, 39(8): 2801-2809.

Grey point set association analysis model based on panel data

ZHAI Yan-li[†], LUO Ge-ge

(School of Mathematics and Statistics, North China University of Water Resources and Electric Power, Zhengzhou 450046, China)

Abstract: Aiming at the correlation analysis between the sample behavior matrices of panel data, a gray point set correlation analysis model based on panel data is proposed from the perspective of proximity. First, the binary index submatrix of the sample behavior matrix is projected into the point set in space, and the absolute deviation matrix of the corresponding point set is established by using the height difference of the corresponding point. Based on the average value and standard deviation of the data in this matrix, the correlation coefficients of the absolute deviation mean value and absolute deviation standard deviation of the corresponding point set are established respectively. Based on this, the grey point set association analysis model is constructed. Then, the properties of the model such as normalization, proximity, positive definiteness, and symmetry are discussed. Comparative analysis shows that the model is not only unaffected by index ranking, but also can perceive the changes of data gaps between sample behavior matrices and effectively reflect their proximity. Finally, the grey point set correlation analysis model is used to quantitatively evaluate the air quality of four cities in central Henan, and based on this, the correlation clustering of four cities in central Henan is carried out, dividing them into three types of pollution areas, in order to provide theoretical support for the regional division of air pollution joint prevention and control in central Henan.

Keywords: panel data; proximity correlation; point set; absolute deviation; average value; standard deviation

0 引言

自邓聚龙教授首创灰色关联模型^[1]以来, 学者们围绕关联模型的构建进行了大量研究, 相关成果在多种社会子系统的因素分析中均有应用^[2-3].

灰关联模型的研究范围从一维序列逐步拓展至面板数据^[4], 目前, 按照研究对象可将基于面板数据的灰关联模型分为两类: 一类将面板数据视为指标

行为矩阵序列, 围绕指标行为矩阵建模, 如灰色矩阵关联模型^[5]、灰色指标关联模型^[6]以及灰色B型关联模型^[7], 这类模型主要利用关联度判断各指标对系统发展的重要程度, 常从相似性视角建模; 另一类将面板数据视为样本行为矩阵序列, 常从接近性视角针对样本行为矩阵建模, 多用于系统优劣评价, 此类模型研究成果有灰色凸关联模型^[8]、新型灰色接近关联模

收稿日期: 2022-11-25; 录用日期: 2023-06-04.

基金项目: 河南省高等学校重点科研项目(18A630032).

责任编委: 刘宝碇.

[†]通讯作者. E-mail: zylncwu@163.com.

型^[9]、灰色曲率关联模型^[10]以及灰色四面体关联模型^[11]等,其中文献[8-10]是在固定指标排序的情况下建模,关联分析结果受指标排序的影响. 文献[9]的模型不能识别样本行为矩阵量级的变化,当两样本行为矩阵扩大相同倍数时,它们的关联度不变,这对接近性关联模型而言不合理. 文献[11]将样本行为矩阵投射为空间四面体,利用四面体有向体积的接近性衡量样本行为矩阵的关联性,当两个样本行为矩阵对应四面体的有向体积相等时,它们的关联度为1,然而,对应四面体有向体积相等并不代表它们的空间位置或几何外形存在特定的关系,模型不满足正定性,不能很好地反映样本行为矩阵的关联性.

基于此,本文构建灰色点集关联分析模型. 首先,将预处理后的面板数据表征为样本行为矩阵序列,考虑到指标的顺序,将样本行为矩阵的二元指标子矩阵投射为空间中的点集. 然后,基于接近性视角,利用对应点的高度差建立对应点集的绝对离差矩阵,依据绝

对离差矩阵中数据的平均值和标准差分别建立对应点集的绝对离差平均值关联系数和绝对离差标准差关联系数,进一步构建对应点集的灰色关联系数和两个样本矩阵的灰色点集关联度. 对比分析发现,模型不受指标排序的影响,能够很好地感知样本行为矩阵间数据差距的变化并有效反映其接近性. 最后,将模型应用于实例,突出模型的实用价值.

1 灰色点集的构建

面板数据涉及样本、指标和时间3个维度. 设样本集 $A = \{a_0, a_1, \dots, a_I\}$ (其中: a_0 为参考样本,其余为比较样本), 指标集 $B = \{b_1, b_2, \dots, b_S\}$, 时间集 $C = \{c_1, c_2, \dots, c_T\}$. 记 u_{it}^s 为第 s 个指标第 i 个样本在时刻 t 的取值, 则面板数据可表示为一个三维表^[4], 如表1所示.

由于不同指标的量纲量级不同, 直接采用表1中的数据建模易发生关联分析结果失真, 本文利用文献[1]中的效果测度函数对表1数据进行预处理.

表1 面板数据三维表

样本	0					...	i					...	I				
指标	时间					...	时间					...	时间				
	1	...	t	...	T		1	...	t	...	T		1	...	t	...	T
1	u_{01}^1	...	u_{0t}^1	...	u_{0T}^1	...	u_{i1}^1	...	u_{it}^1	...	u_{iT}^1	...	u_{I1}^1	...	u_{It}^1	...	u_{IT}^1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
s	u_{01}^s	...	u_{0t}^s	...	u_{0T}^s	...	u_{i1}^s	...	u_{it}^s	...	u_{iT}^s	...	u_{I1}^s	...	u_{It}^s	...	u_{IT}^s
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
S	u_{01}^S	...	u_{0t}^S	...	u_{0T}^S	...	u_{i1}^S	...	u_{it}^S	...	u_{iT}^S	...	u_{I1}^S	...	u_{It}^S	...	u_{IT}^S

定义1^[1] 设面板数据指标行为矩阵序列 $U = \{U^1, U^2, \dots, U^S\}$. 其中: $U^s (s=1, 2, \dots, S)$ 为面板数据在指标 b_s 下的效果样本矩阵, 记

$$U^s = (u_{it}^s) = \begin{bmatrix} u_{01}^s & u_{02}^s & \dots & u_{0T}^s \\ u_{i1}^s & u_{i2}^s & \dots & u_{iT}^s \\ \vdots & \vdots & \ddots & \vdots \\ u_{I1}^s & u_{I2}^s & \dots & u_{IT}^s \end{bmatrix}.$$

1) 若 b_s 为效益型指标, 即指标效果样本值越大越好, b_s 指标下的决策灰靶设为 $u_{it}^s \in [u_{i_0t_0}^s, \max_i \max_t u_{it}^s]$, 其中 $u_{i_0t_0}^s$ 为 b_s 指标效果临界值, 则称

$$x_{it}^s = \frac{u_{it}^s - u_{i_0t_0}^s}{\max_i \max_t u_{it}^s - u_{i_0t_0}^s} \quad (1)$$

为效益型指标效果测度函数.

2) 若 b_s 为成本型指标, 即指标效果样本值越小越好, b_s 指标下的决策灰靶设为 $u_{it}^s \in [\min_i \min_t u_{it}^s, u_{i_0t_0}^s]$, 其中 $u_{i_0t_0}^s$ 为 b_s 指标效果临界值, 则称

$$x_{it}^s = \frac{u_{i_0t_0}^s - u_{it}^s}{u_{i_0t_0}^s - \min_i \min_t u_{it}^s} \quad (2)$$

为成本型指标效果测度函数.

3) 若 b_s 为适中型指标, 指标效果样本值越接近某适中值 w 越好, b_s 指标下的决策灰靶设为 $u_{it}^s \in [w - u_{i_0t_0}^s, w + u_{i_0t_0}^s]$, 其中 $w - u_{i_0t_0}^s, w + u_{i_0t_0}^s$ 分别为 b_s 指标的下限效果临界值和上限效果临界值, 则:

① 当 $u_{it}^s \in [w - u_{i_0t_0}^s, w]$ 时, 称

$$x_{it}^s = \frac{u_{it}^s - w + u_{i_0t_0}^s}{u_{i_0t_0}^s} \quad (3)$$

为适中型指标下限效果测度函数;

② 当 $u_{it}^s \in [w, w + u_{i_0t_0}^s]$ 时, 称

$$x_{it}^s = \frac{w + u_{i_0t_0}^s - u_{it}^s}{u_{i_0t_0}^s} \quad (4)$$

为适中型指标上限效果测度函数.

本文对于效益型指标, 取 $u_{i_0t_0}^s = \min_i \min_t u_{it}^s$ 为指标效果临界值; 对于成本型指标, 取 $u_{i_0t_0}^s = \max_i \max_t u_{it}^s$ 为指标效果临界值; 对于适中型指标, $w - u_{i_0t_0}^s, w + u_{i_0t_0}^s$ 分别为指标的下限效果临界值和上限效果临界值, 取 $u_{i_0t_0}^s = \max_i \max_t |u_{it}^s - w|$. 这样设置与文献[10]中极差变换算子的思想一致, 使得预

处理后的 $x_{it}^s \in [0, 1]$, 且原始指标值 u_{it}^s 越理想, 预处理后 x_{it}^s 越接近 1; 否则, 预处理后 x_{it}^s 越接近 0.

经定义 1 中指标效果测度函数的作用, 表 1 中的 u_{it}^s 变为 x_{it}^s . 由于建模的出发点是样本行为矩阵间的关联性分析, 将预处理后的表 1 按照样本表示为样本行为矩阵序列 $X = \{X_0, X_1, \dots, X_I\}$, 其中

$$X_i = \begin{bmatrix} x_{i1}^1 & x_{i2}^1 & \dots & x_{iT}^1 \\ x_{i1}^2 & x_{i2}^2 & \dots & x_{iT}^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{i1}^S & x_{i2}^S & \dots & x_{iT}^S \end{bmatrix} \quad (5)$$

为第 i ($i = 0, 1, \dots, I$) 个样本的行为矩阵. 记 $x_{it}^s = x_i(s, t)$, 用 $x_i(s, t)$ 表示预处理后第 i 个样本的第 s 个指标在时刻 t 的取值, 可清晰地表示出样本矩阵中元素的含义.

针对指标的排序问题, 本文借鉴文献[11]的方法构造样本行为矩阵的二元指标子矩阵, 并将其中数据投射为空间中的点.

定义 2^[11] 设 X_i 为第 i 个样本的行为矩阵, 称

$$X_{i(j_1, j_2)} = \begin{bmatrix} x_i(j_1, 1) & x_i(j_1, 2) & \dots & x_i(j_1, T) \\ x_i(j_2, 1) & x_i(j_2, 2) & \dots & x_i(j_2, T) \end{bmatrix}$$

为样本行为矩阵 X_i 关于二元指标 b_{j_1} 、 b_{j_2} 的行为子矩阵, 其中 $j_1, j_2 \in \{1, 2, \dots, S\}$ 且 $j_1 \neq j_2$.

由于共有 S 个指标, 其二元组合有 $\frac{S(S-1)}{2}$ 种, 构成集合 $B_S^2 = \{(b_{j_1}, b_{j_2}) | j_1, j_2 \in \{1, 2, \dots, S\}, j_1 \neq j_2\}$.

灰色关联分析的基本思想是从序列的几何特征出发建立关联度模型, 为了体现面板数据的几何特征, 将样本行为矩阵 X_i ($i = 0, 1, \dots, I$) 的二元指标子矩阵 $X_{i(j_1, j_2)}$ 中的数据投射为空间中的点^[11] $(1, t, x_i(j_1, t))$ 、 $(2, t, x_i(j_2, t))$. 其中: $t = 1, 2, \dots, T$, 样本行为矩阵中的数据对应这些点的第 3 个坐标, 反映了这些点在空间中的高度.

定义 3 设 X_i ($i = 0, 1, \dots, I$) 为第 i 个样本行为矩阵, $X_{i(j_1, j_2)}$ 为 X_i 关于二元指标 b_{j_1} 、 b_{j_2} 的行为子矩阵, 则 $X_{i(j_1, j_2)}$ 中两两相邻的 4 个点 $(1, t, x_i(j_1, t))$ 、 $(1, t+1, x_i(j_1, t+1))$ 、 $(2, t, x_i(j_2, t))$ 以及 $(2, t+1, x_i(j_2, t+1))$ 构成一个点集, 称

$$P_i(j_1, j_2, t) = \left[\begin{array}{cc} (1, t, x_i(j_1, t)) & (1, t+1, x_i(j_1, t+1)) \\ (2, t, x_i(j_2, t)) & (2, t+1, x_i(j_2, t+1)) \end{array} \right]$$

为 $X_{i(j_1, j_2)}$ 在时刻 t 的点集.

参考样本行为矩阵 X_0 和比较样本行为矩阵 X_i 关于二元指标 b_{j_1} 、 b_{j_2} 的行为子矩阵分别为 $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$, 称 $X_{0(j_1, j_2)}$ 和 $X_{i(j_1, j_2)}$ 在同一时刻 t 的点集

$P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 为对应点集, 称 $P_0(j_1, j_2, t)$ 和 $P_i(j_1, j_2, t)$ 对应位置的点为对应点.

为了刻画对应点集接近性, 需引入如下概念.

定义 4 设 X_0 为参考样本行为矩阵, X_i 为比较样本行为矩阵, $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 分别为 X_0 、 X_i 关于二元指标 b_{j_1} 、 b_{j_2} 的行为子矩阵, $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 分别为 $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 在时刻 t 的点集, 记 $\Delta_{0i}(j_\omega, v) = |x_0(j_\omega, v) - x_i(j_\omega, v)|$ ($\omega = 1, 2; v = t, t+1$), 称

$$\Delta_{0i}(j_1, j_2, t) = \begin{bmatrix} \Delta_{0i}(j_1, t) & \Delta_{0i}(j_1, t+1) \\ \Delta_{0i}(j_2, t) & \Delta_{0i}(j_2, t+1) \end{bmatrix} \quad (6)$$

为 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的绝对离差矩阵.

对应点集 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 绝对离差矩阵中的元素反映了其对应点在空间中的高度差.

2 灰色点集关联分析模型的构建

借助于样本均值和样本标准差的定义给出绝对离差平均值和绝对离差标准差的概念.

定义 5 设 X_0 为参考样本行为矩阵, X_i 为比较样本行为矩阵, $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 分别为 X_0 、 X_i 关于二元指标 b_{j_1} 、 b_{j_2} 的行为子矩阵, $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 分别为 $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 在时刻 t 的点集, $\Delta_{0i}(j_1, j_2, t)$ 为 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的绝对离差矩阵, 称

$$E_{0i}(j_1, j_2, t) = \frac{1}{4} [\Delta_{0i}(j_1, t) + \Delta_{0i}(j_1, t+1) + \Delta_{0i}(j_2, t) + \Delta_{0i}(j_2, t+1)] \quad (7)$$

为点集 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的绝对离差平均值.

定义 6 设 X_0 为参考样本行为矩阵, X_i 为比较样本行为矩阵, $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 分别为 X_0 、 X_i 关于二元指标 b_{j_1} 、 b_{j_2} 的行为子矩阵, $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 分别为 $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 在时刻 t 的点集, $\Delta_{0i}(j_1, j_2, t)$ 为 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的绝对离差矩阵, $E_{0i}(j_1, j_2, t)$ 为 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的绝对离差平均值, 则称

$$D_{0i}(j_1, j_2, t) = \left\{ \frac{1}{4} [(\Delta_{0i}(j_1, t) - E_{0i}(j_1, j_2, t))^2 + (\Delta_{0i}(j_1, t+1) - E_{0i}(j_1, j_2, t))^2 + (\Delta_{0i}(j_2, t) - E_{0i}(j_1, j_2, t))^2 + (\Delta_{0i}(j_2, t+1) - E_{0i}(j_1, j_2, t))^2] \right\}^{\frac{1}{2}} \quad (8)$$

为点集 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的绝对离差标准差.

当对应点集 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的绝对离差平均值和绝对离差标准差均很小时, 其对应点在空间中的高度差很小, 对应点之间很接近, 此时, 点集 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 应具有高关联性.

最初基于面板数据的灰关联模型用 $1/(1+x)$ 的形式定义关联系数, 但是函数 $1/(1+x)$ 在正半轴的

下降速率较慢,导致模型分辨率不高.与 $1/(1+x)$ 相比, e^{-x} 在正半轴的下降速率较快.本文经数据预处理,样本行为矩阵中的数值 $x_i(s, t) \in [0, 1](\forall i = 0, 1, \dots, I; s = 1, 2, \dots, S; t = 1, 2, \dots, T)$.若用 e^{-x} 构建关联模型,则所得关联度均在 $e^{-1} \approx 0.3679$ 以上.为了提高模型分辨率,本文用 e^{-3x} 构建关联系数,由于 $e^{-3} \approx 0.0498$,所得关联度均在 0.0498 以上.灰色关联公理中的规范性要求灰色关联度介于 $0 \sim 1$ 间且不为 0 ,因此,将关联度限制在 0.0498 以上是可接受的.

定义7 设 X_0 为参考样本行为矩阵, X_i 为比较样本行为矩阵, $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 分别为 X_0 、 X_i 关于二元指标 b_{j_1} 、 b_{j_2} 的行为子矩阵, $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 分别为 $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 在时刻 t 的点集, $E_{0i}(j_1, j_2, t)$ 、 $D_{0i}(j_1, j_2, t)$ 分别为 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的绝对离差平均值和绝对离差标准差,则

$$\alpha_{0i}(j_1, j_2, t) = e^{-3E_{0i}(j_1, j_2, t)}, \quad (9)$$

$$\beta_{0i}(j_1, j_2, t) = e^{-3D_{0i}(j_1, j_2, t)} \quad (10)$$

分别称为点集 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的绝对离差平均值关联系数和绝对离差标准差关联系数.

定义8 设 X_0 为参考样本行为矩阵, X_i 为比较样本行为矩阵, $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 分别为 X_0 、 X_i 关于二元指标 b_{j_1} 、 b_{j_2} 的行为子矩阵, $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 分别为 $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 在时刻 t 的点集, $\alpha_{0i}(j_1, j_2, t)$ 为 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的绝对离差平均值关联系数, $\beta_{0i}(j_1, j_2, t)$ 为 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的绝对离差标准差关联系数,则称

$$\gamma_{0i}(j_1, j_2, t) = \mu_\alpha \alpha_{0i}(j_1, j_2, t) + \mu_\beta \beta_{0i}(j_1, j_2, t) \quad (11)$$

为点集 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的灰色关联系数.其中: μ_α 、 μ_β 分别为 $\alpha_{0i}(j_1, j_2, t)$ 、 $\beta_{0i}(j_1, j_2, t)$ 的权重, $\mu_\alpha, \mu_\beta \geq 0, \mu_\alpha + \mu_\beta = 1$.权重的大小可根据实际情况进行设置,完全不确定信息下取 $\mu_\alpha = \mu_\beta = 0.5$.

定义9 设 X_0 为参考样本行为矩阵, X_i 为比较样本行为矩阵, $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 分别为 X_0 、 X_i 关于二元指标 b_{j_1} 、 b_{j_2} 的行为子矩阵, $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 分别为 $X_{0(j_1, j_2)}$ 、 $X_{i(j_1, j_2)}$ 在时刻 t 的点集, $\gamma_{0i}(j_1, j_2, t)$ 为 $P_0(j_1, j_2, t)$ 、 $P_i(j_1, j_2, t)$ 的灰色关联系数,则

$$\gamma_{0i} = \frac{2}{S(S-1)(T-1)} \sum_{(b_{j_1}, b_{j_2}) \in B_S^2} \sum_{t=1}^{T-1} \gamma_{0i}(j_1, j_2, t) \quad (12)$$

为 X_0 和 X_i 的灰色点集关联度.

3 灰色点集关联分析模型的性质

定理1 灰色点集关联度 γ_{0i} 满足如下规范性^[12]:

$$0 < \gamma_{0i} \leq 1, \gamma_{0i} = 1 \Leftrightarrow X_0 = X_i.$$

证明 显然, $0 < \gamma_{0i} \leq 1$,且当 $X_0 = X_i$ 时, $\gamma_{0i} = 1$ 成立.若 $\gamma_{0i} = 1$,则对于 $\forall t \in \{1, 2, \dots, T-1\}$, $(b_{j_1}, b_{j_2}) \in B_S^2$,均有 $\gamma_{0i}(j_1, j_2, t) = 1$,进一步地,有 $\alpha_{0i}(j_1, j_2, t) = 1, \beta_{0i}(j_1, j_2, t) = 1$,则 $E_{0i}(j_1, j_2, t) = 0, D_{0i}(j_1, j_2, t) = 0$.由式(8)可知, $\Delta_{0i}(j_\omega, v) = 0 (\omega = 1, 2; v = t, t+1)$,因此, $X_0 = X_i$. $\square$

定理2 灰色点集关联度 γ_{0i} 满足如下接近性^[12]:

$$\gamma_{0i} \xrightarrow{i \rightarrow \infty} 1 \Leftrightarrow x_i(s, t) \xrightarrow{i \rightarrow \infty} x_0(s, t), \\ s = 1, 2, \dots, S, t = 1, 2, \dots, T.$$

证明 $x_i(s, t) \xrightarrow{i \rightarrow \infty} x_0(s, t) (s = 1, 2, \dots, S; t = 1, 2, \dots, T)$ 时, $\gamma_{0i} \xrightarrow{i \rightarrow \infty} 1$ 显然成立.若 $\gamma_{0i} \xrightarrow{i \rightarrow \infty} 1$,则 $1 - \gamma_{0i} \xrightarrow{i \rightarrow \infty} 0$,对于 $\forall t \in \{1, 2, \dots, T-1\}$, $(b_{j_1}, b_{j_2}) \in B_S^2$,有 $0 \leq 1 - \gamma_{0i}(j_1, j_2, t) \leq \sum_{(b_{j_1}, b_{j_2}) \in B_S^2} \sum_{t=1}^{T-1} (1 - \gamma_{0i}(j_1, j_2, t)) = \frac{S(S-1)(T-1)}{2} (1 - \gamma_{0i}) \xrightarrow{i \rightarrow \infty} 0$,根据夹逼定理, $1 - \gamma_{0i}(j_1, j_2, t) \xrightarrow{i \rightarrow \infty} 0$,即 $\gamma_{0i}(j_1, j_2, t) \xrightarrow{i \rightarrow \infty} 1$,则 $\alpha_{0i}(j_1, j_2, t) \xrightarrow{i \rightarrow \infty} 1, \beta_{0i}(j_1, j_2, t) \xrightarrow{i \rightarrow \infty} 1$,因此, $E_{0i}(j_1, j_2, t) \xrightarrow{i \rightarrow \infty} 0, D_{0i}(j_1, j_2, t) \xrightarrow{i \rightarrow \infty} 0$,则 $\Delta_{0i}(j_\omega, v) \xrightarrow{i \rightarrow \infty} 0 (\omega = 1, 2; v = t, t+1)$,即对于 $\forall s = 1, 2, \dots, S$ 和 $t = 1, 2, \dots, T$,均有 $|x_0(s, t) - x_i(s, t)| \xrightarrow{i \rightarrow \infty} 0$,即 $x_i(s, t) \xrightarrow{i \rightarrow \infty} x_0(s, t)$. $\square$

定理3 灰色点集关联度 γ_{0i} 满足正定性,即 $\gamma_{0i} = 1$ 当且仅当 X_0 与 X_i 间的关系唯一确定.

证明 由定理1,灰色点集关联度 $\gamma_{0i} = 1$ 当且仅当 $X_0 = X_i$. $\square$

定理4 灰色点集关联度 γ_{0i} 满足对称性,即 $\gamma_{0i} = \gamma_{i0}$.

证明 由于式(6)绝对离差矩阵中的元素满足 $\Delta_{0i}(j_\omega, v) = \Delta_{i0}(j_\omega, v)$,由灰色点集关联模型的建立过程可知 $\gamma_{0i} = \gamma_{i0}$. $\square$

定理5 灰色点集关联度 γ_{0i} 满足唯一性和可比性.

证明 在式(11)中权重系数 μ_α 、 μ_β 取定的情况下,可得到唯一的灰色点集关联度 γ_{0i} ,即 γ_{0i} 满足唯一性.不同样本矩阵对应位置为同一指标在同一时刻的取值,意义、量纲以及量级皆相同,且建模前利用效果测度函数对原始面板数据进行了预处理,消除了不同指标间量纲量级的差异,可有效避免由于二元指标子矩阵中两个指标行的数据量纲量级不同使得绝对离差矩阵中上下两行的数据不是一个量级进而导致关联分析结果失真的现象发生,因此,灰色点集关联度 γ_{0i} 满足可比性. $\square$

4 对比分析

设有如下4个样本行为矩阵:

$$X_0 = \begin{bmatrix} 1.2 & 1.5 & 1.8 \\ 1.4 & 2 & 2.2 \\ 1.8 & 2.1 & 2.5 \end{bmatrix}, X_1 = \begin{bmatrix} 1.4 & 1.9 & 2.1 \\ 1.6 & 0.3 & 2.5 \\ 1.9 & 2.5 & 2.8 \end{bmatrix},$$
$$X_2 = \begin{bmatrix} 1 & 1.8 & 1.6 \\ 1.2 & 2.3 & 3 \\ 1.9 & 2.4 & 4.3 \end{bmatrix}, X_3 = \begin{bmatrix} 1.3 & 1.6 & 1.9 \\ 1.2 & 1.8 & 2 \\ 1.7 & 2 & 2.4 \end{bmatrix}.$$

其中: X_0 为参考样本行为矩阵,其余为比较样本行为矩阵,行为时间维度,列为指标维度,指标均为效益型指标. 下面分别采用所提出模型(式(11)中取 $\mu_\alpha = \mu_\beta = 0.5$)、文献[1]中的三维灰色关联度模型和文献[8-11]中的模型计算 X_0 与 $X_i (i = 1, 2, 3)$ 的关联度,结果如表2所示.

表2 关联度计算结果和排序

关联度模型	关联度			关联序
	γ_{01}	γ_{02}	γ_{03}	
本文	0.5308	0.6354	0.8827	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
文献[1]	0.5949	0.8363	0.7913	$\gamma_{02} \succ \gamma_{03} \succ \gamma_{01}$
文献[8]	0.5756	0.8428	0.8869	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
文献[9]	0.7779	0.8456	0.9319	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
文献[10]	0.8397	0.8566	0.9237	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
文献[11]	0.6793	0.7061	1	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$

由表2可见,所提出模型得到的关联序为 $\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$,与文献[8-11]保持一致,而文献[1]模型所得到的结果为 $\gamma_{02} \succ \gamma_{03} \succ \gamma_{01}$. 所提出模型的建立体现了接近性思想,观察 X_0 和 $X_i (i = 1, 2, 3)$ 可发现, X_0 与 X_3 的对应元素最接近,整体上 X_2 的数据比 X_1 更接近 X_0 . 可见,所提出模型的结果是合理可靠的.

文献[1]的模型公式为

$$\varepsilon_{0i} = \frac{1 + |s_0| + |s_i|}{1 + |s_0| + |s_i| + |s_0 - s_i|},$$

其中 $|s_0|$ 、 $|s_i|$ 、 $|s_0 - s_i|$ 分别为矩阵 X_0 、 X_i 的始边零化曲面与坐标平面所围成曲顶柱体正负体积代数和的绝对值以及两个始边零化曲面所围成曲顶柱体正负体积代数和的绝对值. 此关联度不仅受 $|s_0 - s_i|$ 的影响,还受 $|s_0|$ 、 $|s_i|$ 的影响,当 $|s_0|$ 、 $|s_i|$ 相对于 $|s_0 - s_i|$ 较大时,关联度值接近1,这也是表1中此模型所得到的关联序与其他模型不一致的原因. 相比之下,所提出模型建立在两个矩阵绝对离差矩阵的基础上,关联度仅取决于两个矩阵对应元素的差距,可有效避免这一问题.

为了探究指标排序对以上模型的影响,调换4个样本行为矩阵中指标的顺序,计算结果如表3所示.

由表3可知:不同指标排序下,文献[1]、文献[8]和文献[10]的关联度、关联序均发生了变化;文献[9]

表3 不同指标排序下5种模型的计算结果

模型	指标序	关联度			关联序
		γ_{01}	γ_{02}	γ_{03}	
本文	b_1, b_2, b_3	0.5308	0.6354	0.8827	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_1, b_3, b_2	0.5308	0.6354	0.8827	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_2, b_1, b_3	0.5308	0.6354	0.8827	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_2, b_3, b_1	0.5308	0.6354	0.8827	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_3, b_1, b_2	0.5308	0.6354	0.8827	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_3, b_2, b_1	0.5308	0.6354	0.8827	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
文献[1]	b_1, b_2, b_3	0.5949	0.8363	0.7913	$\gamma_{02} \succ \gamma_{03} \succ \gamma_{01}$
	b_1, b_3, b_2	0.7422	0.8092	0.8359	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_2, b_1, b_3	0.5691	0.9103	0.7083	$\gamma_{02} \succ \gamma_{03} \succ \gamma_{01}$
	b_2, b_3, b_1	0.5773	0.7698	0.7656	$\gamma_{02} \succ \gamma_{03} \succ \gamma_{01}$
	b_3, b_1, b_2	0.8362	0.7917	0.9244	$\gamma_{03} \succ \gamma_{01} \succ \gamma_{02}$
	b_3, b_2, b_1	0.7186	0.7956	1	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
文献[8]	b_1, b_2, b_3	0.5756	0.8428	0.8869	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_1, b_3, b_2	0.6591	0.7731	0.9803	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_2, b_1, b_3	0.7668	0.7096	0.9064	$\gamma_{03} \succ \gamma_{01} \succ \gamma_{02}$
	b_2, b_3, b_1	0.7372	0.7692	0.9830	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_3, b_1, b_2	0.6759	0.8057	0.9167	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_3, b_2, b_1	0.5604	0.7781	0.8933	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
文献[9]	b_1, b_2, b_3	0.7779	0.8456	0.9319	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_1, b_3, b_2	0.8015	0.8430	0.9376	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_2, b_1, b_3	0.7909	0.8511	0.9442	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_2, b_3, b_1	0.8015	0.8430	0.9376	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_3, b_1, b_2	0.7909	0.8511	0.9442	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_3, b_2, b_1	0.7779	0.8456	0.9319	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
文献[10]	b_1, b_2, b_3	0.8397	0.8566	0.9237	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_1, b_3, b_2	0.8087	0.8080	0.9869	$\gamma_{03} \succ \gamma_{01} \succ \gamma_{02}$
	b_2, b_1, b_3	0.7807	0.7799	0.9077	$\gamma_{03} \succ \gamma_{01} \succ \gamma_{02}$
	b_2, b_3, b_1	0.8015	0.8095	0.9557	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_3, b_1, b_2	0.7783	0.7755	0.9322	$\gamma_{03} \succ \gamma_{01} \succ \gamma_{02}$
	b_3, b_2, b_1	0.7992	0.8416	0.9355	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
文献[11]	b_1, b_2, b_3	0.6793	0.7061	1	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_1, b_3, b_2	0.6793	0.7061	1	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_2, b_1, b_3	0.6793	0.7061	1	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_2, b_3, b_1	0.6793	0.7061	1	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_3, b_1, b_2	0.6793	0.7061	1	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$
	b_3, b_2, b_1	0.6793	0.7061	1	$\gamma_{03} \succ \gamma_{02} \succ \gamma_{01}$

的关联序没有变,但是关联度发生了变化;文献[11]和所提出模型得到的关联度、关联序均没有变化. 由此可见,文献[1,8-10]中模型关联分析结果不稳定,所提出模型和文献[11]中模型关联分析结果具有稳定性. 事实上,文献[1,8-10]中模型是在固定指标排序的情况下建立的,因此,关联分析结果受指标顺序的影响.

若令 X_0 与 X_1 中所有元素乘以2,采用文献[9]的模型计算得到 $\gamma(2X_0, 2X_1) = \gamma_{01} \approx 0.7779$,将 X_0 、

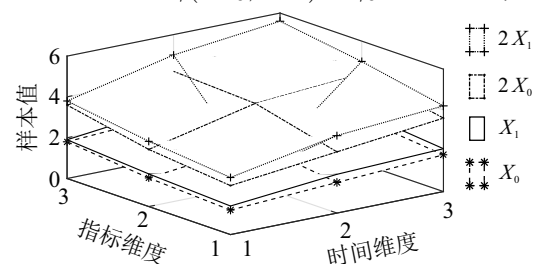


图1 X_0 、 X_1 、 $2X_0$ 以及 $2X_1$ 的三维几何图形

X_1 、 $2X_0$ 以及 $2X_1$ 采用三维图形表示,如图 1 所示.

图 1 中:4 个网格图由下至上依次为 X_0 、 X_1 、 $2X_0$ 、 $2X_1$,显然 $2X_0$ 与 $2X_1$ 的间距大于 X_0 与 X_1 的间距,而采用文献[9]的模型得到 $\gamma(2X_0, 2X_1) = \gamma_{01}$,这个结果对于接近性关联度模型而言并不合理.事实上,文献[9]的模型建立在两个矩阵各维度序列折线相对面积的基础上,相对面积的设置使得任意两个样本矩阵的关联度仅以这两个矩阵本身为量纲,而与其他矩阵无关,消除了整体上样本行为矩阵间量级差异的影响,导致了上述现象的发生.本文数据预处理消除的是同一样本矩阵不同指标行之间量纲量级的差异,两级极差的设置使得所有样本矩阵同一指标行在同一标准下进行变换,不同样本矩阵对应位置数据的量级差异得以保留.采用所提出模型得到 $\gamma(2X_0, 2X_1) = 0.5280$,与 $\gamma_{01} = 0.7200$ 并不相等,且 $\gamma(2X_0, 2X_1) < \gamma_{01}$,符合实际情况.

由表 3 可知,文献[11]的模型在 6 种指标排序下得到的 γ_{03} 均为 1,这是由于文献[11]将样本行为矩阵投射为空间四面体,利用四面体有向体积的差异建立关联度模型,而样本行为矩阵 X_0 与 X_3 在时间维度上相邻两个数据的增量相等,此时 X_0 与 X_3 对应四面体的有向体积相等,所得到的关联度为 1.而两个四面体的有向体积相等并不表示它们的空间位置或几何形状一样, X_0 与 X_3 对应位置的数据不相等,因此关联度为 1 不合理.本文将样本行为矩阵投射为空间中的点集,利用对应点的差距建立灰关联模型,得到的模型满足正定性,关联度为 1 当且仅当两个样本行为矩阵相等,有效避免了此问题的出现.

综上,所提出模型得到的关联度和关联序不受指标排列顺序的影响,能够识别样本行为矩阵量级的变化并有效反映样本行为矩阵间的接近性.

5 实例分析

5.1 豫中 4 个城市空气质量评估

河南省简称“豫”,经济最发达的当属豫中,包括郑州、平顶山、许昌、漯河 4 个城市.2018 年 9 月,河南省出台了《河南省污染防治攻坚战三年行动计划(2018-2020 年)》.为了评估豫中 4 个城市污染防治的效果,现对其 2018 年~2020 年的空气质量状况进行比较.依据《环境空气质量标准》(GB 3095-2012)并结合文献[10],选择 $PM_{2.5}$ 、 PM_{10} 、 SO_2 、 CO 、 NO_2 、 O_3 为评价指标,数据如表 4 所示,来源为中国空气质量在线监测分析平台.记郑州、平顶山、许昌、漯河的样本行为矩阵分别为 X_1^0 、 X_2^0 、 X_3^0 、 X_4^0 ,由于表 4 中数据为成本型,选取 $x_0^0(s, t) = \min_{1 \leq i \leq 4} x_i^0(s, t)$ ($s = 1, 2, \dots, 6; t =$

$1, 2, 3$) 构造参考样本行为矩阵 X_0^0 ,然后运用式(2)消除所有样本行为矩阵构成的面板数据中各指标的量纲,记消除量纲后各样本行为矩阵为 X_i ($i = 0, 1, \dots, 4$).

表 4 2018 年~2020 年豫中 4 个城市空气质量数据

城市	空气指标	2018 年度	2019 年度	2020 年度
郑州	$PM_{2.5}/(\mu g/m^3)$	55	59	51
	$PM_{10}/(\mu g/m^3)$	91	104	89
	$SO_2/(\mu g/m^3)$	14	9	9
	$CO/(mg/m^3)$	0.975	0.896	0.805
	$NO_2/(\mu g/m^3)$	46	45	39
	$O_3/(\mu g/m^3)$	102	111	108
平顶山	$PM_{2.5}/(\mu g/m^3)$	57	60	51
	$PM_{10}/(\mu g/m^3)$	89	95	84
	$SO_2/(\mu g/m^3)$	17	15	12
	$CO/(mg/m^3)$	0.924	0.969	0.875
	$NO_2/(\mu g/m^3)$	35	34	31
	$O_3/(\mu g/m^3)$	103	111	104
许昌	$PM_{2.5}/(\mu g/m^3)$	55	61	53
	$PM_{10}/(\mu g/m^3)$	89	91	79
	$SO_2/(\mu g/m^3)$	14	12	11
	$CO/(mg/m^3)$	1.082	1.008	0.957
	$NO_2/(\mu g/m^3)$	36	34	30
	$O_3/(\mu g/m^3)$	103	108	100
漯河	$PM_{2.5}/(\mu g/m^3)$	54	60	55
	$PM_{10}/(\mu g/m^3)$	91	98	86
	$SO_2/(\mu g/m^3)$	12	10	9
	$CO/(mg/m^3)$	0.786	0.706	0.731
	$NO_2/(\mu g/m^3)$	33	29	26
	$O_3/(\mu g/m^3)$	101	110	101

式(11)中取 $\mu_\alpha = \mu_\beta = 0.5$,由所提出模型得到 X_i ($i = 1, 2, 3, 4$) 与 X_0 的关联度为 $\gamma_{01} \approx 0.4835$ 、 $\gamma_{02} \approx 0.5175$ 、 $\gamma_{03} \approx 0.5830$ 、 $\gamma_{04} \approx 0.7734$,分辨率较高.由各关联度可知,2018 年~2020 年豫中 4 个城市空气质量由好到差为漯河、许昌、平顶山、郑州.

将表 4 中数据按照年度用折线图表示,如图 2 所示.图 2(a)~图 2(c)分别为 2018 年~2020 年的豫中 4 个城市空气指标折线图,数据越大表明空气中对应污染物浓度越高,即折线越靠上的城市空气质量越差.由图 2 可见,郑州的折线整体上在最上方,平顶山的折线也一直很靠上,而漯河和许昌的折线整体上位于下方,这与上面得到的关联序相吻合.

图 2 中 4 个城市的折线比较聚集,为了更清楚地显示出 4 个城市样本矩阵的差异,将表 4 中数据按照指标用折线图表示,如图 3 所示.图 3(a)~图 3(f)分别为豫中 4 个城市 2018 年~2020 年各单一污染物的浓度折线图,折线越靠上表明空气中对应污染物浓度越高,城市空气质量越差.由图 3 可见:漯河的折线仅在图 3(a)和图 3(b)中相对靠上,在其余图中均位于最下方或次下方,表明漯河空气中大部分污染物浓度较低,空气质量较好;许昌的折线在图 3(d)中位于最上

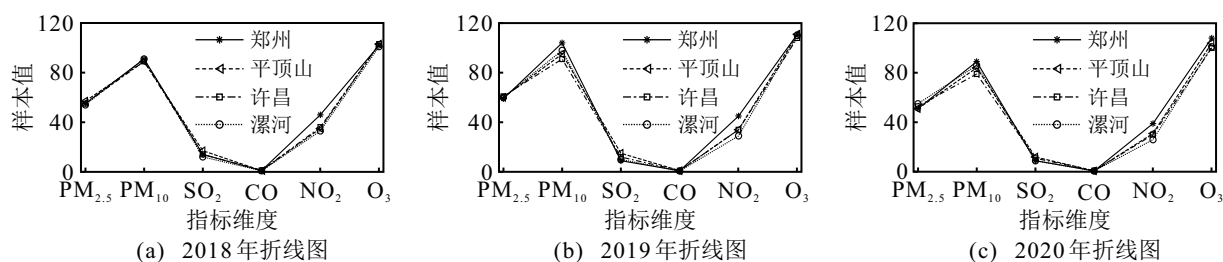


图2 豫中4个城市2018年~2020年年度空气指标折线图

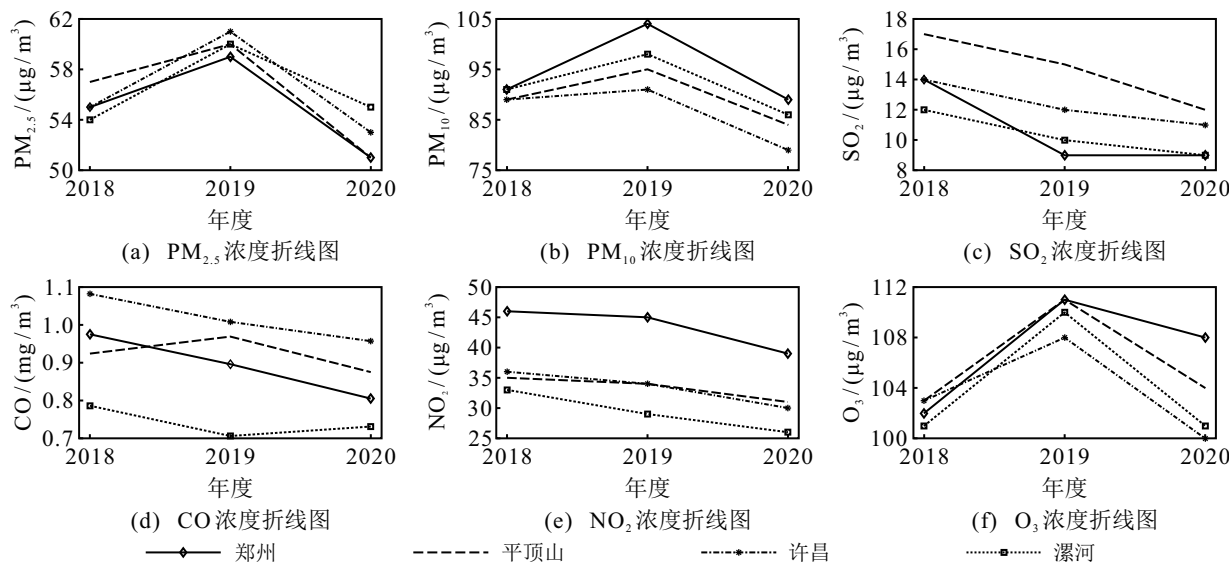


图3 豫中4个城市2018年~2020年空气指标折线图

方,在图3(b)中位于最下方,在其余图中大部分年份位于中间,表明许昌空气中CO的浓度在豫中是最高的,PM₁₀的浓度在豫中是最低的,其余污染物的浓度也相对较低,空气质量相对较好。根据2019年~2021年《河南社会治理发展报告》,2018年~2020年河南省18地市宜居度排名中,漯河和许昌一直排在前5位,平顶山和郑州一直排在后10位,而宜居城市评价的6项一级指标中有一项是环境优美度。此外,漯河2020年11月28日被评为2020年中国宜居宜业城市,由此可见,2018年~2020年漯河的空气质量在豫中排名第1,许昌排名第2名副其实。图3(c)中平顶山的折线始终位于最上方,图3(d)和图3(e)中平顶山的折线先是位于次下方,2019年~2020年位于次上方,图3(f)中平顶山的折线2018年和2019年位于最上方,2020年位于次上方,表明平顶山空气中SO₂的浓度在豫中是最高的,CO、NO₂、O₃的浓度也相对较高。平顶山因煤而兴,属于资源型工业城市,是中部地区重要的能源和重工业基地,重工业的发展对于平顶山的空气质量有很大影响。图3(b)和图3(e)中郑州的折线始终位于最上方,图3(f)中郑州的折线先是位于次上方,2019年和2020年位于最上方,表明郑州空气中PM₁₀、NO₂的浓度在豫中是最高的,O₃的浓度也

相对较高,空气质量较差。郑州是河南省省会,同时也是新一线城市,2018年~2020年共完成生产总值33736.08亿元,每年的生产总值均高于另外3个城市之和,经济发展的优势使得郑州的空气质量受到很大影响。由此可见,2018年~2020年平顶山和郑州的空气质量在豫中分别排名第3、第4符合实际情况。

综上,灰色点集关联分析模型能够对豫中4个城市2018年~2020年的空气质量做出合理的评价,表明所提出模型得到的关联度和关联序具有合理性和有效性,同时也表明所提出模型具有一定的实用价值。为了将所提出模型落到实处,下面基于本节空气质量评估结果,利用所提出灰色点集关联分析模型对豫中4个城市进行聚类,以期对豫中大气污染联防联控的区域划分提供理论支持。

5.2 豫中空气污染区域划分

空气污染受到空气环流的影响,具有跨域性、公共性、外部性等特征,地方政府各自解决各自区域内污染问题事倍功半。北京奥运会、上海世博会、杭州G20峰会等空气质量保障管理的成功经验表明,加强区域联防联控对于改善区域空气质量行之有效。本文以豫中4个城市为研究对象,基于灰色点集关联模型分析豫中空气污染的区域性特征。

根据第5.1节4个城市的样本矩阵 X_1 、 X_2 、 X_3 、 X_4 ,计算它们两两之间的关联度,得到如下关联矩阵^[8]:

$$G = \begin{bmatrix} 1 & 0.5332 & 0.5258 & 0.5119 \\ & 1 & 0.6647 & 0.5292 \\ & & 1 & 0.5815 \\ & & & 1 \end{bmatrix}.$$

其中:第 m 行第 n 列的元素表示样本矩阵 X_m 、 X_n 的关联度 γ_{mn} , $m,n=1,2,3,4$.

由关联矩阵 G 可知,任意两个城市样本矩阵的关联度均在0.5之上,具有较高的关联性和整体性,但是 X_2 、 X_3 的关联度 $\gamma_{23} \approx 0.6647$,其余任意两个不同样本矩阵的关联度均在0.5~0.6之间,平顶山和许昌的关联度明显高于其他城市间的关联度.根据关联矩阵 G 用一次形成法形成谱系图^[13],如图4所示.

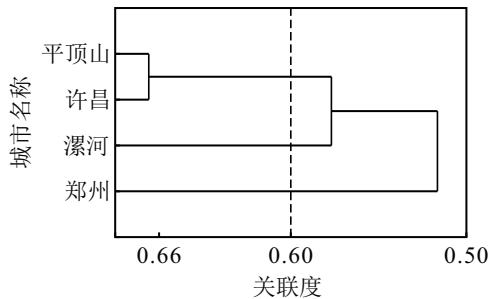


图4 豫中4个城市样本聚类谱系图

根据文献^[1]中关联聚类思想,取临界值 $\gamma=0.6$,将关联度大于等于0.6的城市划分为相同的空气污染级别,将关联度小于0.6的城市划分为不同的污染级别.由图4可见,仅平顶山与许昌是相连的,与漯河、郑州的联系被截断了,因此聚类结果为{平顶山,许昌}、{漯河}、{郑州}.事实上,若取临界值 $\gamma=0.5$,则可将豫中4个城市聚为一类,但是,地理位置、污染程度的不同和经济发展的差异均会影响大气污染联防联控的效果,因此,选取临界值 $\gamma=0.6$,将关联程度绝对高的城市划分在一起,避免聚类结果太泛太粗而忽略了地区差异.

为了对聚类结果形成直观的认识,特绘制豫中行政地图,如图5所示.由图5可见,许昌与平顶山大面积地接壤,郑州和漯河被许昌、平顶山分隔在两侧,聚类结果具有明显的地域性特点,将豫中自北向南分为3个区域.此外,第5.1节空气质量评估结果显示,豫中4个城市空气污染程度由高到低依次为郑州、平顶山、许昌、漯河,即平顶山和许昌的空气质量排名相邻且居中,郑州和漯河位于两端,这表明本节的聚类结果与第5.1节的空气质量评估结果相互一致、相互印证.根据第5.1节空气质量评估结果和本节的聚类结果,

可将{郑州}划分为豫中的第1类污染区域,{平顶山,许昌}划分为豫中的第2类污染区域,{漯河}划分为豫中的第3类污染区域.



图5 豫中4个城市行政区域图

根据河南省统计年鉴,2018年~2020年豫中4个城市的生产总值以及分产业生产总值如表5所示.

表5 2018年~2020年豫中城市
生产总值统计 单位:亿元

城市	生产总值	一产	二产	三产
郑州	33 736.08	444.8	13 827.34	19 463.94
平顶山	6 963.71	539.01	3 217.52	3 207.19
许昌	9 675.53	494.2	5 288.16	3 893.16
漯河	4 388.98	398.75	2 145.22	1 845

由表5可以发现:首先,郑州的经济水平远远高于其他城市,且拉开郑州和其他城市差距的是第2产业和第3产业.作为省会城市,郑州发展较早,由表5可知,郑州第3产业生产总值最高.整体上郑州的经济水平、经济结构与其他城市差距较大,从发展的不一致性考虑,郑州单独作为一类城市治理大气污染效果会比较好.然后,平顶山和许昌的第1产业生产总值在豫中排前两名,第2产业生产总值和第3产业生产总值在豫中排名中间,整体经济水平相差不大.许昌是中原经济区重要的工业城市,装备制造、食品加工、能源电力是许昌的支柱性产业;平顶山因煤而兴,是中原经济区重要的能源和重工业基地,根据《2014年平顶山市支柱产业发展报告》,能源、化工、装备制造、冶金建材、轻纺食品是平顶山的5大支柱产业.总体上平顶山与许昌的经济发展水平很接近,支柱性产业较类似,经济发展的一致性弱化了大气污染区域协同治理的行政壁垒,且平顶山与许昌相邻,针对空气污染的跨域性、流动性和公共性,{平顶山,许昌}作为一类城市,建立大气污染联防联控机制,协同治理大气污染效果会更好.此外,漯河的生产总值与分产业生产总值均排名在豫中末尾,其第1产业生产总值占比9.09%,在豫中排名第1.事实上,漯河是全国消费品工业“三品”战略示范城市和首家中国食品名城,

食品工业是漯河的主导和支柱产业,食品工业属于轻工业,相比于其他城市,漯河产业结构偏轻,这也是漯河的空气质量在豫中排名第1的重要原因。整体上,漯河与其他城市的经济水平、支柱性产业类型差别较大,发展目标、定位与其他城市不同,在大气污染治理上,漯河的需求和侧重点也会与其他城市有较大差别。因此,漯河单独作为一类城市治理大气污染效果会更好。

综上,根据豫中4个城市两两之间的关联度,结合地理位置、污染程度、经济水平等方面的分析,豫中4个城市应划分为3类污染区域:{郑州}为豫中的第1类污染区域,{平顶山,许昌}为豫中的第2类污染区域,{漯河}为豫中的第3类污染区域。平顶山、许昌应建立大气污染联防联控机制,统一规划、统一监测、统一防治,协同进行大气污染治理。

6 结 论

本文从接近性视角出发构建了灰色点集关联分析模型,所提出模型得到的关联度和关联序均不受指标排序的影响,能够很好地识别样本行为矩阵间数据差距的变化并有效地反映其接近性。经验证,所提出模型满足规范性、接近性、正定性、对称性、唯一性和可比性。最后,应用所提出模型进行了豫中4个城市的空气质量评估,基于此,对豫中4个城市进行了关联聚类,将豫中划分为3类污染区域,以期为豫中大气污染联防联控的区域划分提供理论支持。

参考文献(References)

- [1] 刘思峰. 灰色系统理论及其应用[M]. 第8版. 北京: 科学出版社, 2017: 55-159.
(Liu S F. Grey system theory and its application[M]. The 8th edition. Beijing: Science Press, 2017: 55-159.)
- [2] 唐少虎, 刘小明, 朱伟, 等. 基于多学科设计优化的路网交通分布式协同控制[J]. 控制与决策, 2019, 34(9): 1867-1875.
(Tang S H, Liu X M, Zhu W, et al. Distributed cooperative control of road network traffic based on multidisciplinary design optimization[J]. Control and Decision, 2019, 34(9): 1867-1875.)
- [3] 贺利军, 李文锋, 张煜. 基于灰色综合关联分析的多目标优化方法[J]. 控制与决策, 2020, 35(5): 1134-1142.
(He L J, Li W F, Zhang Y. Multi-objective optimization method based on grey synthetic incidence analysis[J]. Control and Decision, 2020, 35(5): 1134-1142.)
- [4] 张可, 刘思峰. 灰色关联聚类在面板数据中的扩展及应用[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.)
- [5] 罗党, 叶莉莉, 韦保磊, 等. 面板数据的灰色矩阵关联

模型及在旱灾脆弱性风险中的应用[J]. 控制与决策, 2018, 33(11): 2051-2056.

- (Luo D, Ye L L, Wei B L, et al. Grey matrix incidence model for panel data and its application in drought vulnerability risk[J]. Control and Decision, 2018, 33(11): 2051-2056.)
- [6] 党耀国, 尚中举, 王俊杰, 等. 基于面板数据的灰色指标关联模型构建及其应用[J]. 控制与决策, 2019, 34(5): 1077-1084.
(Dang Y G, Shang Z J, Wang J J, et al. Grey incidence model for relationship between indicators of panel data and its application[J]. Control and Decision, 2019, 34(5): 1077-1084.)
- [7] 罗党, 张曼曼. 基于面板数据的灰色B型关联模型及其应用[J]. 控制与决策, 2020, 35(6): 1476-1482.
(Luo D, Zhang M M. Grey incidence model of B-mode based on panel data and its application[J]. Control and Decision, 2020, 35(6): 1476-1482.)
- [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): 1033-1036.
(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): 1033-1036.)
- [9] 刘震, 党耀国, 周伟杰, 等. 新型灰色接近关联模型及其拓展[J]. 控制与决策, 2014, 29(6): 1071-1075.
(Liu Z, Dang Y G, Zhou W J, et al. New grey nearness incidence model and its extension[J]. Control and Decision, 2014, 29(6): 1071-1075.)
- [10] 吴鸿华, 屈忠锋. 基于面板数据的灰色曲率关联模型[J]. 控制与决策, 2020, 35(5): 1072-1076.
(Wu H H, Qu Z F. The grey curvature incidence model based on panel data[J]. Control and Decision, 2020, 35(5): 1072-1076.)
- [11] 吴鸿华, 刘思峰, 方志耕. 基于面板数据的灰色四面体网格关联分析模型及应用[J]. 控制与决策, 2022, 37(11): 3033-3041.
(Wu H H, Liu S F, Fang Z G. Grey tetrahedral grid incidence analysis model based on panel data and its application[J]. Control and Decision, 2022, 37(11): 3033-3041.)
- [12] 王慧, 魏勇. 分别基于接近性和相似性的面板数据关联度的公理化[J]. 统计与决策, 2019, 35(3): 13-16.
(Wang H, Wei Y. Axiomatization of correlation degree of panel data respectively based on proximity and similarity[J]. Statistics and Decision, 2019, 35(3): 13-16.)
- [13] 孙效功, 杨作升. 基于灰色关联度的聚类分析方法[J]. 青岛海洋大学学报, 1995(2): 229-232.
(Sun X G, Yang Z S. Cluster analysis based on the gray relational grades[J]. Journal of Qingdao Ocean University, 1995(2): 229-232.)

作者简介

翟艳丽(1982—), 女, 讲师, 博士, 从事灰色系统理论、运筹优化与智能决策等研究, E-mail: zylnwu@163.com;
罗格格(1999—), 女, 硕士生, 从事灰色系统理论与决策分析的研究, E-mail: luogegge2021@163.com.