

基于SBM区间模型的决策单元相似度

黄 衍^{1,2}, 王应明^{1†}, 杨隆浩¹

(1. 福州大学 经济与管理学院, 福州 350108; 2. 福建农林大学 计算机与信息学院, 福州 350002)

摘 要: 在数据包络分析方法的研究中, 当决策单元的输入输出值为区间数时, 区间效率的测算取决于生产参照集的选择. 对此, 针对非径向非角度的松弛测度模型 (SBM), 通过比较计算认为, 以最佳生产状态为统一参照来测算决策单元的区间效率最有利于排序. 同时, 利用区间数相似度理论建立模糊相似矩阵, 进而提出基于参照单元区间效率的相似度排序方法, 对决策单元聚类 and 排序. 最后通过算例表明了所提出方法的可行性和有效性.

关键词: 数据包络分析; 区间数相似度; 聚类分析; 区间数排序

中图分类号: C934; O221 文献标志码: A

Similarity of decision making units based on SBM interval model

HUANG Yan^{1,2}, WANG Ying-ming^{1†}, YANG Long-hao¹

(1. School of Economics & Management, Fuzhou University, Fuzhou 350108, China; 2. College of Information and Computer Science, Fujian Agriculture and Forest University, Fuzhou 350002, China)

Abstract: In the research of the data envelopment analysis method, when the input and output value of the decision making unit is interval number, the calculation of the relative interval efficiency depends on the choice of the front face of the production. Therefore, by the comparison of the interval model of slack based model (SBM) under different production frontier, it is concluded that it's most favorable for the decision making unit ranking by the interval efficiencies which are calculated both by the best production frontier. Then, the interval number similarity theory is introduced, and the fuzzy similarity matrix of interval efficiency is established. The reference interval efficiency unit similarity method is proposed, which is used to cluster and sort the decision making units. Finally, a numerical example is given to illustrate the feasibility and effectiveness of the proposed methods.

Keywords: DEA; interval number similarity; cluster analysis; interval number ranking

0 引 言

随着社会经济的不断发展, 在决策过程中出现了越来越多不确定性的问题. 因此, 人们开始关注并研究基于区间数理论的多属性决策方法^[1]. 数据包络分析方法 (DEA) 是多属性决策理论中一种重要的非参数方法, 其主要用来评估同类型决策单元 (DMU) 在多投入与多产出下的相对有效性. 传统的 DEA 模型 (如 CCR、BCC、SBM) 主要研究关于确定性数据下 DMU 的生产相对效率. 随着不确定性理论逐步完善, 近年来不少学者开始关注不确定性数据的 DEA 理论研究, 其中关于区间数的 DEA 模型拓展是非常重要的一个内容^[2-5]. 目前, Wang 等^[3] 依据不同的生产前沿面将传统的 CCR 模型拓展为 CCR 区间数模型, 并且对不同模型下所得到的 DMU 效率值进行了充分

比较. SBM 模型是基于松弛测度的非径向非角度的 DEA 效率评价方法, 相较于 CCR 模型, 它能更好地反映 DMU 的相对效率值与投入产出松弛量之间的关系. 文献 [4] 和文献 [5] 以不同的参照集构建 SBM 区间模型, 从而在不同的生产前沿面下评价 DMU 的区间相对效率.

对于 DEA 区间数模型而言, DMU 的相对效率值以区间数的形式表示, 因此, 对 DMU 的排序实际上就是对区间数的排序. 区间数的排序不同于实数, 是一个复杂的过程. 目前, 主要的区间数排序方法包含两大类: 一类是确定性排序方法, 另一类是不确定性排序方法. 确定性排序方法舍弃了很多区间数所包含的不确定信息, 使得对区间数不能进行全序排列. 同时, 该类方法在参数选择上还具有很强的主观

收稿日期: 2016-07-26; 修回日期: 2016-10-13.

基金项目: 国家自然科学基金项目 (71371053); 福建省教育厅项目 (JB14001); 教育部人文社科基金项目 (16YJC630064).

作者简介: 黄衍 (1983—), 女, 讲师, 博士生, 从事多属性决策、相似性结构的研究; 王应明 (1964—), 男, 教授, 博士生导师, 从事决策理论与方法、数据包络分析、规则库推理、质量功能展开等研究.

†通讯作者. E-mail: ymwang@fzu.edu.cn

性,如文献[6]所提到的HWC方法. 不确定排序方法是一种基于度的排序方法^[7],其中比较有代表性的是可能度排序方法^[8]、可信度排序方法^[9]、相似度排序方法^[10]. 目前,大部分不确定排序方法都是先对区间数进行两两比较,所得结果通常以矩阵形式表示,从而无法直接对区间数进行排序. 为了解决该问题,需要对排序矩阵进行集结或逐次筛选. 然而,很多集结方法的应用并不合理^[11-16],而采用逐次筛选^[6,9]又会使得计算量过大.

基于以上问题,本文首先将不同前沿面下的SBM区间模型所得到的区间效率进行比较,认为以最优生产状态作为参照构建最优前沿面,并以该前沿面为基准测算DMU的区间效率更为合理. 其次,本文认为用区间数表示DMU的相对效率不仅比实数值更加准确,而且也为决策者提供了更多的信息. 因此,利用这些信息不仅可以对DMU作出排序,还可以利用相似关系对DMU进行聚类分析. 本文引用文献[10]中的区间数相似度定义,计算DMU关于区间效率的相似矩阵,并依据模糊聚类分析方法对其进行聚类. 在聚类结果的基础上,结合DMU区间效率的性质来设定参照DMU,并提出基于参照单元区间效率的相似度排序方法. 该方法通过排序向量的形式给出比较结果,能够直接对所有的决策单元进行排序,且计算量较小.

1 区间数定义及其基本运算

实数域上一个闭区间称为区间数 $[a, \bar{a}]$, 记作 $[a]$. 当 $a = \bar{a}$ 时, 区间数 $[a]$ 退化为实数 a , 用 $I(R)$ 表示实数域上的区间数全体. 对于任意的 $[a], [b] \in I(R)$, 定义区间数之间的计算如下所示^[14]:

$$\begin{aligned}
 [a] + [b] &= [\underline{a} + \underline{b}, \bar{a} + \bar{b}], \\
 [a] - [b] &= [\underline{a} - \underline{b}, \bar{a} - \bar{b}], \\
 [a] \cdot [b] &= [\min\{\underline{a}\underline{b}, \underline{a}\bar{b}, \bar{a}\underline{b}, \bar{a}\bar{b}\}, \max\{\underline{a}\underline{b}, \underline{a}\bar{b}, \bar{a}\underline{b}, \bar{a}\bar{b}\}], \\
 [a]/[b] &= [a] \cdot [1/b], \quad 0 \notin [b].
 \end{aligned}$$

区间数 $[a]$ 的宽度定义为 $m(a) = \bar{a} - \underline{a}$, 区间数 $[a]$ 的中点定义为 $w(a) = (\bar{a} + \underline{a})/2$, 两个区间数相等 $[a] = [b]$ 当且仅当 $\underline{a} = \underline{b}, \bar{a} = \bar{b}$.

2 SBM 区间模型

假设有 n 个 DMU, 每个 DMU 包含 m 个输入和 s 个输出. 令 $X_j = (x_{1j}, \dots, x_{mj}) \geq 0, X_j \neq 0, Y_j = (y_{1j}, \dots, y_{sj}) \geq 0, Y_j \neq 0$ 分别表示 DMU _{j} ($j = 1, 2, \dots, n$) 非负的输入和输出向量, 则其生产可能集定义为

$$\begin{aligned}
 T = \{ (X, Y) \mid & \sum_{j=1}^n \lambda_j X_j \leq X, \sum_{j=1}^n \lambda_j Y_j \geq Y, \\
 & \lambda_j \geq 0, j = 1, 2, \dots, n \}.
 \end{aligned} \tag{1}$$

集合 T 是一个闭凸集, 其边界是生产有效前沿面. SBM 模型是在上述生产可能集下基于 DEA 方法的松弛测度模型^[5]. 该模型不仅更好地说明了效率值随着投入和产出松弛程度变化而严格单调递减, 还解决了投入产出的松弛问题. 其模型如下所示:

$$\begin{aligned}
 \min \rho &= \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{s_i^-}{x_{io}}}{1 + \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_{ro}}}. \\
 \text{s.t. } & \sum_{j=1}^n \lambda_j x_{ij} + s_i^- = x_{io}, \quad i = 1, 2, \dots, m; \\
 & \sum_{j=1}^n \lambda_j y_{rj} - s_r^+ = y_{ro}, \quad r = 1, 2, \dots, s; \\
 & \lambda_j \geq 0, s_i^- \geq 0, s_r^+ \geq 0, \quad i = 1, 2, \dots, m; \\
 & \quad r = 1, 2, \dots, s, j = 1, 2, \dots, n.
 \end{aligned} \tag{2}$$

其中: $x_{io} (i = 1, 2, \dots, m), y_{ro} (r = 1, 2, \dots, s)$ 分别为被评价单元 DMU _{o} 输入输出值. 模型 (2) 可以通过 C-C 变换为如下线性形式 (LSBM)^[4]:

$$\begin{aligned}
 \min \tau &= t - \frac{1}{m} \sum_{i=1}^m \frac{S_i^-}{x_{io}}. \\
 \text{s.t. } 1 &= t + \frac{1}{s} \sum_{r=1}^s \frac{S_r^+}{y_{ro}}; \\
 & \sum_{j=1}^n A_j x_{ij} + S_i^- = t x_{io}, \quad i = 1, 2, \dots, m; \\
 & \sum_{j=1}^n A_j y_{rj} - S_r^+ = t y_{ro}, \quad r = 1, 2, \dots, s; \\
 & A_j \geq 0, S_i^- \geq 0, S_r^+ \geq 0, \quad i = 1, 2, \dots, m; \\
 & \quad r = 1, 2, \dots, s, j = 1, 2, \dots, n.
 \end{aligned} \tag{3}$$

其中

$$\begin{aligned}
 t &= 1 / \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_{ro}} \right), \\
 A_j &= t \lambda_j, \quad S_i^- = t s_i^-, \quad S_r^+ = t s_r^+.
 \end{aligned}$$

为了更好地表示投入与产出之间的关系, 本文将模型 (3) 以对偶形式给出 (DLSBM)^[4]:

$$\begin{aligned}
 \max 1 &+ \sum_{r=1}^s u_r y_{ro} - \sum_{i=1}^m v_i x_{io}. \\
 \text{s.t. } & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \\
 & j = 1, 2, \dots, n;
 \end{aligned}$$

$$\begin{aligned}
 v_i x_{io} &\geq \frac{1}{m}, \quad i = 1, 2, \dots, m; \\
 u_{ro} y_{ro} &\geq \frac{1}{s} \left(1 + \sum_{r=1}^s u_r y_{ro} - \sum_{i=1}^m v_i x_{io} \right), \\
 r &= 1, 2, \dots, s; \\
 u_{ro} &\geq 0, \quad v_{io} \geq 0, \quad r = 1, 2, \dots, s, \\
 i &= 1, 2, \dots, m.
 \end{aligned} \tag{4}$$

当DMU_j的输入输出值为区间数 $[x_{ij}, \bar{x}_{ij}]$, $[y_{rj}, \bar{y}_{rj}]$ 时, 需要构建相应的SBM区间模型. 文献[4]认为基于乐观视角, 以最差的生产状态作为参考集来衡量最好生产状态下的相对效率可以当作区间效率的上限; 而区间效率的下限则可以从悲观视角出发, 以最优的生产状态作为参考来衡量最差生产状态下的相对效率. 其所对应的模型如下(为便于定理1的证明, SBM区间模型不再以LSBM形式给出, 而是统一以DLSBM形式给出):

$$\begin{aligned}
 \bar{H}_o^* &= \max 1 + \sum_{r=1}^s u_r \bar{y}_{ro} - \sum_{i=1}^m v_i \underline{x}_{io}. \\
 \text{s.t.} \\
 \sum_{r=1}^s u_r \underline{y}_{rj} - \sum_{i=1}^m v_i \bar{x}_{ij} &\leq 0, \quad j = 1, 2, \dots, n, \quad j \neq o; \\
 \sum_{r=1}^s u_r \bar{y}_{ro} - \sum_{i=1}^m v_i \underline{x}_{io} &\leq 0; \\
 v_i \underline{x}_{io} &\geq \frac{1}{m}, \quad i = 1, 2, \dots, m; \\
 u_r \bar{y}_{ro} &\geq \frac{1}{s} \left(1 + \sum_{r=1}^s u_r \bar{y}_{ro} - \sum_{i=1}^m v_i \underline{x}_{io} \right), \\
 r &= 1, 2, \dots, s; \\
 u_r &\geq 0, \quad v_i \geq 0, \quad r = 1, 2, \dots, s, \quad i = 1, 2, \dots, m;
 \end{aligned} \tag{5}$$

$$\begin{aligned}
 \underline{H}_o^* &= \max 1 + \sum_{r=1}^s u_r y_{ro} - \sum_{i=1}^m v_i \bar{x}_{io}. \\
 \text{s.t.} \\
 \sum_{r=1}^s u_r \bar{y}_{rj} - \sum_{i=1}^m v_i \underline{x}_{ij} &\leq 0, \quad j = 1, 2, \dots, n; \\
 v_i \bar{x}_{io} &\geq \frac{1}{m}, \quad i = 1, 2, \dots, m; \\
 u_r y_{ro} &\geq \frac{1}{s} \left(1 + \sum_{r=1}^s u_r y_{ro} - \sum_{i=1}^m v_i \bar{x}_{io} \right), \\
 r &= 1, 2, \dots, s; \\
 u_r &\geq 0, \quad v_i \geq 0, \quad r = 1, 2, \dots, s, \quad i = 1, 2, \dots, m.
 \end{aligned} \tag{6}$$

文献[5]认为, 研究区间效率的上下限必须要有

一个统一的参考集. 因此, 本文统一以最优生产状态作为参考集, 分别计算最优生产状态和最差生产状态下的相对效率. 具体模型如下:

$$\begin{aligned}
 \bar{G}_o^* &= \max 1 + \sum_{r=1}^s u_r \bar{y}_{ro} - \sum_{i=1}^m v_i \underline{x}_{io}. \\
 \text{s.t.} \\
 \sum_{r=1}^s u_r \bar{y}_{rj} - \sum_{i=1}^m v_i \underline{x}_{ij} &\leq 0, \quad j = 1, 2, \dots, n; \\
 v_i \underline{x}_{io} &\geq \frac{1}{m}, \quad i = 1, 2, \dots, m; \\
 u_r \bar{y}_{ro} &\geq \frac{1}{s} \left(1 + \sum_{r=1}^s u_r \bar{y}_{ro} - \sum_{i=1}^m v_i \underline{x}_{io} \right), \\
 r &= 1, 2, \dots, s; \\
 u_r &\geq 0, \quad v_i \geq 0, \quad r = 1, 2, \dots, s, \quad i = 1, 2, \dots, m;
 \end{aligned} \tag{7}$$

$$\begin{aligned}
 \underline{G}_o^* &= \max 1 + \sum_{r=1}^s u_r y_{ro} - \sum_{i=1}^m v_i \bar{x}_{io}. \\
 \text{s.t.} \\
 \sum_{r=1}^s u_r \underline{y}_{rj} - \sum_{i=1}^m v_i \bar{x}_{ij} &\leq 0, \quad j = 1, 2, \dots, n; \\
 v_i \bar{x}_{io} &\geq \frac{1}{m}, \quad i = 1, 2, \dots, m; \\
 u_r y_{ro} &\geq \frac{1}{s} \left(1 + \sum_{r=1}^s u_r y_{ro} - \sum_{i=1}^m v_i \bar{x}_{io} \right), \\
 r &= 1, 2, \dots, s; \\
 u_r &\geq 0, \quad v_i \geq 0, \quad r = 1, 2, \dots, s, \quad i = 1, 2, \dots, m.
 \end{aligned} \tag{8}$$

在模型(6)和(8)中, 由于存在

$$\sum_{r=1}^s u_r \bar{y}_{rj} - \sum_{i=1}^m v_i \underline{x}_{ij} \leq 0, \quad j = 1, 2, \dots, n,$$

很容易得到

$$\sum_{r=1}^s u_r \underline{y}_{rj} - \sum_{i=1}^m v_i \bar{x}_{ij} \leq \sum_{r=1}^s u_r \bar{y}_{rj} - \sum_{i=1}^m v_i \underline{x}_{ij} \leq 0, \quad j = 1, 2, \dots, n.$$

因此, 模型(6)和模型(8)中的目标函数在该约束条件下, 其目标值不会超过1.

定理1 设 \bar{G}_o^* , \underline{G}_o^* 分别是模型(7)和模型(8)的最优值, \bar{H}_o^* , \underline{H}_o^* 分别是模型(5)和模型(6)的最优值, 则有 $\bar{G}_o^* \leq \bar{H}_o^*$, $\underline{G}_o^* \leq \underline{H}_o^*$.

证明 因模型(8)与模型(6)相同, 故结论 $\underline{G}_o^* \leq \underline{H}_o^*$ 自然成立, 即 $\underline{G}_o^* = \underline{H}_o^*$. 假设 u_r^* 和 v_i^* 为模型(7)的一组最优权重解, 则有 $\sum_{r=1}^s u_r^* \bar{y}_{rj} - \sum_{i=1}^m v_i^* \underline{x}_{ij} \leq 0, \quad j = 1, 2, \dots, n$. 因此, 对于被评价 DMU_o, $\sum_{r=1}^s u_r^* \bar{y}_{ro} -$

$\sum_{i=1}^m v_i^* \underline{x}_{io} \leq 0$ 成立. 在此基础上, 可知 $\sum_{r=1}^s u_r^* \underline{y}_{rj} - \sum_{i=1}^m v_i^* \underline{x}_{ij} \leq \sum_{r=1}^s u_r^* \overline{y}_{rj} - \sum_{i=1}^m v_i^* \underline{x}_{ij}, j = 1, 2, \dots, n$ 也成立. 除此之外, 模型(5)和模型(7)的另两个约束条件完全相同, 从而可知 u_r^* 和 v_i^* 也是模型(5)的一组可行权重解, 于是有 $\overline{G}_o^* = 1 + \sum_{r=1}^s u_r^* \overline{y}_{ro} - \sum_{i=1}^m v_i^* \underline{x}_{io} = \overline{H}_o \leq \overline{H}_o^*. \square$

在区间SBM模型中, 文献[5]定义区间效率的上限为1时, DMU是有效单元; 否则为无效单元. 生产前沿面由有效DMU组成. 由最优生产状态作为参考集得到的前沿面为最优生产前沿面, 由最差生产状态作为参考集得到的前沿面为最差生产前沿面.

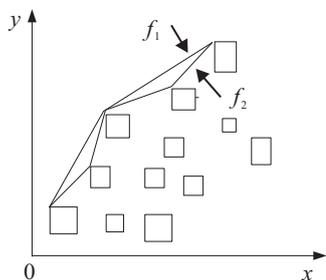


图1 单个输入和单个输出的DMU前沿面

图1中, DMU的输入输出值为区间数, 并只含有单个投入、单个产出. 其中, 每个小方格代表一个DMU, 而 f_1 表示最优生产前沿面, f_2 表示最差生产前沿面. 与定理1的结论一样, 最优生产前沿面所包含的有效单元个数少于最差生产前沿面所包含的单元个数. 在DMU的排序中, 过多的有效DMU会给排序造成极大的困难. 相对于以最差前沿面计算区间效率上限, 选用最优前沿面计算区间效率上限可以适当减少有效DMU的个数. 因此, 模型(7)相较于模型(5)能够更好地区分DMU的优劣.

3 决策单元的聚类 and 排序

近年来, 区间数理论与方法研究得到了学者们的广泛关注, 并取得了很多成果^[7-16]. 当DMU的输入输出信息为区间数时, 其相对效率值只有采用区间数的形式才能保留这些数据所包含的不确定信息. 事实上, 若仅采用确定性排序法对DMU的区间效率进行排序是不完全、不充分的. 而对于当前大多数的不确定性排序方法而言, 主要排序过程是: 首先构建区间数之间关于度的矩阵; 然后通过将矩阵集结成排序向量或对矩阵逐层筛选进行最终排序. 目前, 矩阵的集结方法存在很多不合理的因素^[8], 而当DMU个数较多时, 逐层筛选方法又存在计算量较大的问题. 在DEA研究中, 本文认为具有不同输入输出值的DMU

之间也存在着一定的相似性(或一致性). 这种相似性可以直接地体现在它们的相对效率上. 当DMU的相对效率以区间数形式表示时, 可以利用区间数的不确定性信息获取其相似性信息, 从而进行聚类分析. 通过聚类后再对每一类DMU进行排序, 这样不仅能够减小排序规模, 而且还可以降低不合理性, 进而提高排序的效率.

对DMU进行聚类分析, 首先通过建立相似性关系来描述它们之间的亲疏程度. 区间数相似度是以两个区间数相交部分和相并部分的宽度比值来刻画两者之间的相似程度, 其定义如下^[10]:

定义1 设 $[a_i] = [\underline{a}_i, \overline{a}_i], [a_j] = [\underline{a}_j, \overline{a}_j], i = 1, 2, \dots, n, j = 1, 2, \dots, n;$

$$S_{[a_i],[a_j]} = \frac{m([a_i] \cap [a_j])}{m([a_i] \cup [a_j])}. \quad (9)$$

由定义1可知, 当 $S_{[a_i],[a_j]} = 1$ 时, 区间数 $[a_i]$ 和 $[a_j]$ 完全相似. $S_{[a_i],[a_j]}$ 越大, 二者间的相似程度越大; 反之, 则越小. 因此, 式(9)同样也可以按照区间数在数轴上的位置分布情况来表示, 具体如下:

$$S_{[a_i],[a_j]} = \begin{cases} 0, & \underline{a}_j \leq \overline{a}_i \text{ or } \underline{a}_i \leq \overline{a}_j; \\ \frac{\overline{a}_i - \underline{a}_j}{\overline{a}_j - \underline{a}_i}, & \underline{a}_i < \underline{a}_j \leq \overline{a}_i < \overline{a}_j; \\ \frac{m_{a_j}}{m_{a_i}}, & \underline{a}_i \leq \underline{a}_j < \overline{a}_j \leq \overline{a}_i; \\ \frac{\overline{a}_j - \underline{a}_i}{\overline{a}_i - \underline{a}_j}, & \underline{a}_j < \underline{a}_i \leq \overline{a}_j < \overline{a}_i; \\ \frac{m_{a_i}}{m_{a_j}}, & \underline{a}_j \leq \underline{a}_i < \overline{a}_i \leq \overline{a}_j. \end{cases} \quad (10)$$

该定义满足以下性质^[10]:

- 1) 自反性, $S_{[a_i],[a_j]} = 1 \Leftrightarrow [a_i] = [a_j];$
- 2) 对称性, $S_{[a_i],[a_j]} = S_{[a_j],[a_i]}.$

根据定义1, 可以计算DMU两两间的区间效率相似度, 并建立相似矩阵 $S_{n \times n}$. 由区间数相似度的自反性和对称性可知, $S_{n \times n}$ 也具有自反性和对称性. 区间数相似度反映了两个不确定信息的相似程度, 是一个模糊的概念, 因此, $S_{n \times n}$ 可以看成是模糊集合上的模糊相似矩阵. 本文以模糊集合中的扎德算子 (\vee, \wedge) (其中“ \vee ”表示取加数中的最大者为“和”, “ \wedge ”表示取乘数中的最小者为“积”)作为二元模糊运算的算子, 对 $S_{n \times n}$ 进行模糊复合运算, 从而得到相应的模糊等价矩阵 $R_{n \times n}$. 该矩阵具有自反性、对称性和传递性. 模糊等价矩阵反映了DMU之间的模糊等价关系, 据此可以在一定聚类水平下对其进行分类. 对于模糊等价关系, 给定一个聚类水平 λ . 令

$$r_{ij} = \begin{cases} 0, & r_{ij} < \lambda; \\ 1, & r_{ij} \geq \lambda. \end{cases}$$

则可得到一个对应于模糊等价矩阵的布尔矩阵. 由于模糊等价矩阵具有对称性, 布尔矩阵中各行或列中元素为1的即为一类. 这样, 可将DMU按一定的聚类水平划分成若干类, 直到取得所要求的聚类结果. 为了节省篇幅, 模糊聚类方法的具体步骤见文献[17].

Wang等^[6]将每个DMU的区间效率下限和所有DMU的最大效率作比较, 以此对DMU进行排序, 并称之为最小化最大后悔值法. 然而, 该方法考虑了所有DMU的最大效率以及各个DMU的区间效率的下限, 本文认为可将一组DMU中的上限最大值和下限最大值分别作为参照DMU的效率上下限. 该参照DMU无论是虚拟单元, 还是被排序的DMU, 都被认为是最优的DMU. 同时两个区间效率并集的宽度可用参照DMU的效率上限和被排序DMU的效率下限之差表示, 称为该DMU的最大后悔值^[6], 并作为分母; 两个区间效率交集的宽度可用被排序DMU的效率上限和参照DMU的效率下限之差表示, 称为该DMU的最小获利值, 并作为分子. 特别地, 当DMU的效率上限小于或等于参照DMU的效率下限时, 统一设定其最小获利值为0. 这样该组DMU与参照DMU之间的区间相似度均可表示为最小获利值与最大后悔值的比值. 因此, 最小获利值越大(≥ 0), 最大后悔值越小, 被排序DMU与参照DMU的相似度越高, 排序越靠前. 上述方法称为“基于参照单元区间效率的相似度排序方法”. 下面给出该方法的数学描述:

设 $I(a) = \{[a_i] = [\underline{a}_i, \bar{a}_i] | i = 1, 2, \dots, n, \text{ 且 } \underline{a}_i, \bar{a}_i \in (0, 1)\}$ 为 n 个 DMU 所对应的区间效率集合, 即 DMU _{i} 的区间效率为 $[a_i]$. 定义参照 DMU _{d} 的区间效率为 $[a_d] = [\underline{a}_d, \bar{a}_d]$. 其中: $\underline{a}_d = \max\{\underline{a}_i, i = 1, 2, \dots, n\}$, $\bar{a}_d = \max\{\bar{a}_i, i = 1, 2, \dots, n\}$. 令 $S_{[a_i],[a_d]}$ 表示 DMU _{i} 与 DMU _{d} 之间的相似度, 若 $S_{[a_i],[a_d]} \leq S_{[a_j],[a_d]}$, 则说明 DMU _{j} 相较于 DMU _{i} 更接近目标决策单元, 即 DMU _{i} 不优于 DMU _{j} , 记作 $[a_i] \prec [a_j]$. 特别地, 当 $S_{[a_i],[a_d]} = S_{[a_j],[a_d]}$ 时, 记 $[a_i] \sim [a_j]$.

定理2 根据 $S_{[a_i],[a_d]}$ 所定义的排序关系满足以下性质:

- 1) 自反性. 有 $[a_i] \sim [a_i]$.
- 2) 完全性. 对于 $\forall [a_i], [a_j] \in I(a)$, 有 $[a_i] \prec [a_j]$ 或 $[a_j] \prec [a_i]$.
- 3) 相容性. 若 $[a_i] \prec [a_j]$, 则当 $[a_i], [a_j]$ 退化为实数时, 有 $[a_i] \leq [a_j]$.
- 4) 传递性. 有 $[a_i] \prec [a_j], [a_j] \prec [a_k]$, 则 $[a_i] \prec [a_k]$.
- 5) 分离性. 若 $\bar{a}_i \leq \underline{a}_j$, 则有 $[a_i] \prec [a_j]$.

证明 性质1)~性质3)显然满足.

对于性质4), 由题设可知 $S_{[a_i],[a_d]} > S_{[a_j],[a_d]}$, $S_{[a_j],[a_d]} > S_{[a_k],[a_d]}$, 则有 $S_{[a_i],[a_d]} > S_{[a_k],[a_d]}$.

性质5)分情况进行讨论: 当 $\bar{a}_j \leq \underline{a}_d$ 时, 有 $S_{[a_i],[a_d]} = S_{[a_j],[a_d]} = 0$, 即 $[a_i] \sim [a_j]$; 当 $\underline{a}_j \leq \underline{a}_d < \bar{a}_j$ 时, 有 $S_{[a_i],[a_d]} = 0 < S_{[a_j],[a_d]}$, 即 $[a_j] \succ [a_i]$; 当 $\bar{a}_i > \underline{a}_d$ 时, 由于 $\underline{a}_i < \bar{a}_i < \underline{a}_j < \bar{a}_j$, 有 $S_{[a_i],[a_d]} = \frac{\bar{a}_i - \underline{a}_d}{\underline{a}_d - \underline{a}_i} < S_{[a_j],[a_d]} = \frac{\bar{a}_j - \underline{a}_d}{\underline{a}_d - \underline{a}_j}$, 即 $[a_j] \succ [a_i]$. \square

目前, 关于区间数的不确定性排序方法主要是通过两两比较得到关于度的排序矩阵, 但是排序矩阵并不能直接得到排序结果^[5,7-8,14,16]. 文献[7,14,16]利用模糊判断矩阵的排序方法求解可能度矩阵的排序向量. 然而, 文献[8]却指出, 可能度矩阵不一定是模糊一致性矩阵, 因而以其作为可能度矩阵排序法的理论依据可能导致不合常理的排序结果. 文献[5]通过对可信度矩阵进行逐次筛选的方式获取排序结果, 但是该方法计算量很大, 不利于推广. 针对以上问题, 本文提出一种基于参照单元区间效率的相似度排序方法. 该方法通过计算被排序DMU与参照DMU的相似度, 直接得到被排序DMU的排序向量, 从而避免了集结过程中存在的不合理问题和筛选方法中存在的计算量较大的问题. 相对于两两比较方法而言, 该方法显得更加简单合理.

最后将DMU的聚类 and 排序过程归纳如下:

Step 1: 根据DMU的输入输出值, 利用模型(7)和(8)计算DMU的区间效率集合, 从而可得 $I(a) = \{[a_i] = [\underline{a}_i, \bar{a}_i] | i = 1, 2, \dots, n, \text{ 且 } \underline{a}_i, \bar{a}_i \in (0, 1)\}$;

Step 2: 计算DMU关于区间效率的模糊相似度矩阵 $S_{n \times n}$;

Step 3: 对模糊相似度矩阵进行复合计算, 并得到模糊等价矩阵 $R_{n \times n}$;

Step 4: 按照给定的聚类水平, 对样本进行聚类;

Step 5: 在聚类结果下, 计算每一类的参照DMU的区间效率, 利用基于参照DMU的区间相似度方法对每一类的DMU进行排序;

Step 6: 采用合适的确定性方法对参照DMU进行排序, 确定每一类的先后顺序, 然后再结合Step 5对所有DMU进行全体排序.

4 算例

本文通过对24家台湾银行的相对效率分析来验证上述方法的可行性. 对这24家银行的评价包含3个输入指标和3个输出指标, 它们分别如下^[18]:

x_1 : 总存款(百万元); x_2 : 利息费用(百万元); x_3 : 非利息费用(百万元);

y_1 :总贷款(百万元); y_2 :利息收入(百万元); y_3 :非利息收入(百万元).

输出数据和输入数据以及由式(7)和(8)计算得到的区间效率如表1所示.

表1 24家台湾银行输入输出数据

银行		x_1	x_2	x_3	y_1	y_2	y_3	区间效率
1	L	788 670.6	40 241.93	11 811.94	724 380.14	60 822.39	7 094.72	0.560 4
	U	840 859.3	43 683.96	12 022.59	773 314.72	66 231.62	7 623.20	0.809 6
2	L	926 235.9	42 863.30	15 496.88	786 268.25	66 067.14	12 826.69	0.619 7
	U	1 014 339.3	49 421.62	17 952.67	850 782.56	72 605.03	14 022.96	0.837 3
3	L	895 985.4	40 469.85	13 030.99	77 0236.24	57 395.59	11 691.72	0.649 1
	U	989 805.9	43 463.23	13 986.15	861 676.05	64 209.40	13 079.72	0.909 9
4	L	458 981.8	29 869.43	6 267.73	418 079.49	44 354.53	5 663.31	0.6803
	U	516 654.9	32 997.12	6 924.03	467 712.46	49 620.15	6 335.638	1.000 0
5	L	235 351.1	7 881.37	2 820.19	169 336.03	11 427.47	1 618.14	0.481 7
	U	259 995.1	8 706.64	3 115.50	189 439.03	12 784.10	1 810.24	0.687 2
6	L	256 277.5	8 499.21	1 163.29	200 432.66	11 234.13	2 845.69	0.779 1
	U	283 112.9	9 389.18	1 285.10	224 227.34	12 567.80	3 183.52	1.000 0
7	L	108 792.8	5 421.99	1 405.51	80 058.74	7 848.88	302.15	0.218 9
	U	120 184.7	5 791.03	1 596.83	89 563.04	8 612.78	338.02	0.305 3
8	L	78 795.8	4 052.71	2 488.02	47 904.99	4 975.08	249.43	0.143 4
	U	85 261.1	4 430.68	2 720.07	53 079.75	5 459.29	279.05	0.174 5
9	L	383 560.8	27 531.87	5 352.50	325 799.31	35 344.23	5 393.14	0.179 9
	U	411 675.2	29 694.05	5 744.83	353 146.19	3 831.94	5 963.52	0.224 1
10	L	507 635.2	22 708.68	3 727.91	402 910.43	37 609.65	3 298.46	0.548 0
	U	560 790.8	25 086.55	4 118.27	441 709.21	42 074.53	3 543.68	1.000 0
11	L	166 251.0	8 518.76	3 621.04	147 175.58	11 443.13	1 671.40	0.472 1
	U	182 253.8	8 960.14	3 886.46	167 355.33	12 593.50	1 869.82	0.686 1
12	L	176 709.8	8 324.76	1 554.94	158 536.00	11 591.02	710.44	0.378 1
	U	194 858.3	8 804.19	1 644.49	177 494.95	12 967.06	794.78	1.000 0
13	L	432 487.9	21 002.18	2 693.84	349 537.63	29 012.39	4 799.48	0.708 2
	U	477 774.5	23 201.36	2 975.92	391 033.55	32 456.64	5 369.26	1.000 0
14	L	717 622.8	32 432.93	5 207.24	591 874.45	45 500.26	3 017.95	0.410 8
	U	770 223.5	35 154.16	5 577.30	662 139.76	50 901.89	3 376.23	0.655 7
15	L	101 281.3	5 491.09	4 927.33	78 813.65	7 421.86	578.59	0.202 6
	U	111 886.6	6 066.08	5 443.28	87 607.24	8 302.96	647.27	0.250 3
16	L	126 969.3	7 023.18	3 063.38	122 170.19	9 147.28	1 698.28	0.525 4
	U	141 594.1	7 758.59	3 384.15	136 673.82	10 233.21	1 899.90	1.000 0
17	L	145 850.9	7 933.35	5 981.42	127 122.12	12 139.73	757.21	0.202 1
	U	164 181.9	8 573.00	6 607.75	144 379.11	13 437.21	8 487.11	1.000 0
18	L	143 347.3	8 101.26	2 799.39	126 680.92	11 828.34	2 366.53	0.656 2
	U	165 099.7	8 949.56	2 904.32	144 429.80	13 232.56	2 647.48	0.921 5
19	L	190 173.5	9 307.44	661.98	145 200.48	13 106.07	2 027.19	0.809 2
	U	210 086.9	10 282.04	731.29	162 438.18	14 661.98	2 267.85	1.000 0
20	L	216 899.7	9 514.11	1 910.87	149 165.08	12 403.45	4 760.57	0.800 6
	U	239 611.8	10 510.35	2 110.96	166 873.45	13 875.95	5 016.70	1.000 0
21	L	131 203.4	6 496.85	3 852.75	101 543.04	8 989.75	1 593.30	0.439 5
	U	141 507.4	7 041.09	4 119.01	108 383.65	9 644.39	1 718.14	0.521 1
22	L	214 511.8	10 666.97	4 651.90	171 767.41	19 395.85	3 022.84	0.598 5
	U	239 220.0	11 895.62	5 187.71	194 236.44	21 933.04	3 418.26	1.000 0
23	L	155 200.1	9 242.28	14 248.46	91 728.20	10 657.76	1 038.97	0.189 5
	U	167 934.4	10 000.625	15 577.34	105 261.87	12 003.14	1 190.32	0.230 5
24	L	153 476.5	7 816.278	1 619.78	144 453.15	11 601.73	918.05	0.500 0
	U	166 608.1	8 307.35	1 721.55	157 371.73	12 320.21	969.98	1.000 0

表1中“L”和“U”分别表示区间数的上下限. 根据表1中所算得的关于24家银行的区间效率,可通过

式(9)计算它们之间的两两相似度值,并建立如下模糊相似矩阵:

Table with 24 columns and 24 rows of numerical values representing similarity matrix.

进而通过相容计算可得如下模糊等价矩阵:

Table with 24 columns and 24 rows of numerical values representing fuzzy equivalence matrix.

在模糊等价矩阵的基础上,设定聚类水平 λ = 0.47,对上述结果进行聚类,有

D1 = {7, 8, 9, 15, 23}, D2 = {5, 11, 14, 21}, D3 = {1, 2, 3, 4, 6, 10, 12, 13, 16, 17, 18, 19, 20, 22, 24}.

可以求得相应于每一类的参照DMU,它们依次为

a_d1 = (0.219, 0.305), a_d2 = (0.482, 0.687), a_d3 = (0.809, 1.000).

同时还可以算出每一类DMU与其相应的参照DMU之间的相似度值,结果如表2所示.

由表2可知,三大类的参照DMU的区间效率完

全不相交,根据其在数轴上的位置,可以得到它们的优先顺序为 a_d1 < a_d2 < a_d3. 同时结合各分类内部的排序结果,可以对24家银行进行全排序. 为了能够更好地解释本文排序方法的可行性和合理性,将本文的排序方法与可信度排序方法[5]及可能度排序方法[8]进行比较,其排序结果列于表3.

由于DMU的个数多达24个,采用可信度排序方法[5]需要对排序矩阵进行24次筛选才能得到最终排序结果,计算量较大. 而采用本文的排序方法,只要直接计算24个决策单元和其所对应的参照DMU的区间效率相似度就可以得到排序结果,因此该方法更为简洁明了.

表 2 聚类结果及其基于参照 DMU 的区间效率相似程度

银行号	区间效率	$S_{[a_i],[a_{d_1}]}$	$S_{[a_i],[a_{d_2}]}$	$S_{[a_i],[a_{d_3}]}$
D_1	7	(0.219, 0.305)	1	
	8	(0.143, 0.175)	0	
	9	(0.180, 0.224)	0.041	
	15	(0.203, 0.250)	0.306	
	23	(0.186, 0.231)	0.097	
D_2	5	(0.482, 0.687)		1
	11	(0.472, 0.686)		0.95
	14	(0.411, 0.656)		0.629
	21	(0.440, 0.521)		0.158
D_3	1	(0.560, 0.807)		0
	2	(0.620, 0.837)		0.073 9
	3	(0.649, 0.910)		0.287
	4	(0.680, 1.000)		0.596 8
	6	(0.779, 1.000)		0.863 7
	10	(0.548, 1.000)		0.422 1
	12	(0.378, 1.000)		0.306 8
	13	(0.708, 1.000)		0.653 9
	16	(0.525, 1.000)		0.401 3
	17	(0.202, 1.000)		0.239 1
	18	(0.656, 0.922)		0.326 6
	19	(0.809, 1.000)		1
	20	(0.801, 1.000)		0.956 9
	22	(0.590, 1.000)		0.464 8
24	(0.500, 1.000)		0.381 6	

表 3 3 种方法的排序结果比较

银行号	可信度排序法	本文排序法	可能度排序方法
1	14	15	13
2	12	14	9
3	8	12	7
4	5	5	5
5	16	16	15
6	3	3	3
7	20	20	20
8	24	24	24
9	23	23	23
10	9	7	10
11	17	17	16
12	13	11	14
13	4	4	4
14	18	18	17
15	21	21	21
16	10	8	11
17	15	13	18
18	7	10	6
19	1	1	1
20	2	2	2
21	19	19	19
22	6	6	8
23	22	22	22
24	11	9	12

通过对比相似度排序方法和可信度排序方法所得到的排序结果, D_1 类、 D_2 类、 D_3 类 DMU 的相对位置没有发生改变(D_1 类和 D_2 类的 DMU 相对于 D_3 类

的 DMU 具有较差的区间效率, 在两种排序方法中它们都排在 D_3 类之后), 这表明在聚类分析的基础上进行排序是合理的。

采用可能度排序方法时, 文献 [8] 指出在集结过程中因为可能度矩阵未必是模糊一致性矩阵, 所以依据矩阵行求和法对排序矩阵进行集结, 会令集结的标准和两两可能度比较的标准存在不一致性, 从而影响排序结果的合理性。例如考虑 DMU₂ 和 DMU₁₆, 根据可能度定义^[8]可计算得到 DMU₁₆ 相较于 DMU₂ 的可能度为 0.549, 即两两比较结果为 DMU₁₆ > DMU₂, 而集结后其排序结果为 DMU₁₆ < DMU₂, 二者矛盾。但是采用本文的排序方法, 直接依据被排序 DMU 与参照 DMU 的相似度得到排序向量, 排序标准是唯一的, 因此不存在矛盾。例如 $S_{[a_2],[a_{d_3}]} < S_{[a_{16}],[a_{d_3}]}$, 则相应地有 DMU₁₆ > DMU₂。同时, 可能度排序方法的不合理性会使得排序结果和聚类结果产生不一致性。例如 DMU₁₇ 在聚类结果中属于 D_3 类, 但在可能度排序方法中其排序位置在第 18 位, 低于 D_2 类中的 DMU₁₄ 和 DMU₁₁。相比较而言, 本文的排序方法是在聚类的基础上再进行排序, 不会产生此类问题。

综上所述, 基于参照单元区间效率的相似度排序方法不仅可以通过对 DMU 进行聚类分析, 为决策者提供更为丰富的决策依据, 而且在聚类分析的基础上通过考虑最小获利值与最大后悔值的比值, 对 DMU 进行直接排序, 减少了集结或是筛选的过程, 使得排序标准统一, 排序过程简单, 排序结果更为合理。

5 结 论

本文通过对不同前沿面下对 SBM 区间模型的相对效率分析, 确定了以最优前沿面计算 SBM 区间模型的区间效率更有利于 DMU 的排序, 并且利用区间数相似度理论对 DMU 进行模糊聚类分析, 为决策者提供了关于 DMU 更为丰富的决策信息。同时提出了基于参照单元区间效率的相似度排序方法, 直接、合理、客观地得到排序向量, 快速对 DMU 进行排序。该方法克服了传统的不确定性排序方法中排序标准不一致和计算量较大的问题。最后本文通过实例应用表明了先聚类分析、再对 DMU 进行排序是有效且可行的。

参考文献(References)

[1] 徐泽水. 不确定多属性决策方法及应用[M]. 北京: 清华大学出版社, 2004: 105-155.
(Xu Z S. Uncertain multiple attribute decision making methods and application[M]. Beijing: Tsinghua University Press, 2004: 105-155.)

- [2] Jahanshahloo G R, Lotfi F H, Malkhalifeh M R, et al. A generalized model for data envelopment analysis with interval data[J]. *Applied Mathematical Modelling*, 2009, 33(33): 3237-3244.
- [3] Wang Y M, Greatbanks R, Yang J B. Interval efficiency assessment using data envelopment analysis[J]. *Fuzzy Sets & Systems*, 2005, 153(3): 347-370.
- [4] Hosseinzadeh Lotfi F, Jahanshahloo G R, Esmaeili M. Classification of decision making units with interval data using SBM model[J]. *Applied Mathematical Sciences*, 2007, 1(14): 681-689.
- [5] Hossein Azizia, Sohrab Kordrostami, Alireza Amirteimoori. Slacks-based measures of efficiency in imprecise data envelopment analysis: An approach based on data envelopment analysis with double frontiers[J]. *Computers & Industrial Engineering*, 2014, 79(2): 42-51.
- [6] Wang Y M, Yang J B. Measuring the performances of decision-making units using interval efficiencies[J]. *J of Computational and Applied Mathematics*, 2007, 198(1): 253-267.
- [7] 高峰记. 可能度及区间数综合排序[J]. *系统工程理论与实践*, 2013, 33(8): 2033-2040.
(Gao F J. Possibility degree and comprehensive priority of interval numbers[J]. *Systems Engineering—Theory & Practice*, 2013, 33(8): 2033-2040.)
- [8] 李德清, 韩国柱, 曾文艺, 等. 基于布尔矩阵的区间数排序方法[J]. *控制与决策*, 2016, 31(4): 629-634.
(Li D Q, Han G Z, Zeng W Y, et al. Ranking method of interval numbers based on Boolean matrix[J]. *Control and Decision*, 2016, 31(4): 629-634.)
- [9] Wang Y M, Yang J B, Xu D L. Interval weight generation approaches based on consistency test and interval comparison matrices[J]. *Applied Mathematics & Computation*, 2005, 167(1): 252-273.
- [10] 许瑞丽, 徐泽水. 区间数相似度研究[J]. *数学的实践与认识*, 2007, 24(2): 1-8.
(Xu R L, Xu Z S. On Similarity degrees of interval numbers[J]. *Mathematics in Practice and Theory*, 2007, 24(2): 1-8.)
- [11] 詹棠森, 林卫中. 基于数据最优分区间相似度算法及应用[J]. *数学的实践与认识*, 2009, 39(20): 31-34.
(Zhan T S, Lin W Z. Similarity algorithm and its application on the data optimization partition[J]. *Mathematics in Practice and Theory*, 2009, 39(20): 31-34.)
- [12] 曾文艺, 赵宜宾. 基于区间数度量的区间值模糊集合的归一化距离, 相似度, 模糊度和包含度的关系研究[J]. *模糊系统与数学*, 2012, 26(2): 81-90.
(Zeng W Y, Zhao Y B. Relationship among the normalized distance, the similarity measure, the entropy and the inclusion measure of interval-valued fuzzy sets based on interval-number measurement[J]. *Fuzzy Systems and Mathematics*, 2012, 26(2): 81-90.)
- [13] 肖峻, 张跃, 付川. 基于可能度的区间数排序方法比较[J]. *天津大学学报*, 2011, 44(8): 705-711.
(Xiao J, Zhang Y, Fu C. Comparison between methods of interval number ranking based on possibility[J]. *J of Tianjin University*, 2011, 44(8): 705-711.)
- [14] 徐泽水, 达庆利. 区间数排序的可能度法及其应用[J]. *系统工程学报*, 2003, 18(1): 67-70.
(Xu Z S, Da Q L. Possibility degree method for ranking interval numbers and its application[J]. *J of Systems Engineering*, 2003, 18(1): 67-70.)
- [15] 张吉军. 区间数的排序方法研究[J]. *运筹与管理*, 2003, 12(3): 18-22.
(Zhang J J. Research on method for ranking interval numbers[J]. *Operations Research and Management Science*, 2003, 12(3): 18-22.)
- [16] 李德清, 谷云东. 一种基于可能度的区间数排序方法[J]. *系统工程学报*, 2008, 23(2): 223-226.
(Li D Q, Gu Y D. Method for ranking interval numbers based on possibility degree[J]. *J of Systems Engineering*, 2008, 23(2): 223-226.)
- [17] 高新波. 模糊聚类分析及其应用[M]. 西安: 西安电子科技大学出版, 2004: 42-46.
(Gao X B. Fuzzy cluster analysis and its application[M]. Xi'an: Xi'an Electronic and Science University Publishing, 2004: 42-46.)
- [18] Shiang-Tai, Liu Chiang Kao. Predicting bank performance with financial forecasts: A case of Taiwan commercial banks[J]. *J of Banking & Finance*, 2004, 28(10): 2353-2368.

(责任编辑: 李君玲)