第 30 卷 第 2 期
Vol. 30 No. 2

控 制 与 决 策
Control and Decision

2015 年 2 月
Feb. 2015

文章编号: 1001-0920 (2015) 02-0335-08 DOI: 10.13195/j.kzyjc.2013.1485

决策单元特殊关系的挖掘与建立

木 仁[1], 马占新[2]

(1. 内蒙古工业大学 管理学院，呼和浩特 010051；2. 内蒙古大学 经济管理学院，呼和浩特 010021)

摘 要: 基于偏序集理论的数据包络分析方法, 通过引进适当的偏序关系, 挖掘出决策单元之间的特殊关系. 然而, 随着决策单元所选取的投入产出指标个数的增加, 决策单元之间的偏序关系变得越来越少. 对此, 通过引进决策单元之间的距离和适当的样本决策单元, 建立决策单元之间的特殊关系, 最终生成决策单元之间的格论关系, 并引进相关定理及其算法. 最后通过仿真结果表明了所提出算法的有效性和实用性.

关键词: 数据包络分析方法；样本决策单元；距离；格论；Matlab

中图分类号: N94 **文献标志码:** A

Mine and establish of special relationships between decision making units

MU Ren[1], *MA Zhan-xin*[2]

(1. Management College，Inner Mongolia University of Technology，Hohhot 010051，China；2. School of Economics and Management，Inner Mongolia University，Hohhot 010021，China．Correspondent：MU Ren，E-mail: muren@imut.edu.cn)

Abstract: With the introduction of proper partial order relation in the data envelopment analysis method based on the partial ordered set theory, the special relationships between different decision making units are established. However, with the increase of the number of index of input and output data, the relationships between different decision making units decrease. With the introduction of distances between decision making units and proper sample decision making units, the relationships between different decision making units are established, the relations based on the lattice theory of decision making units are provided, and related theorems and algorithms are introduced. Finally, simulation results show the effectiveness and practicability of the proposed method.

Keywords: data envelopment analysis；sample decision making unit；distance；lattice theory；Matlab

0 引 言

数据包络分析 (DEA)[1]方法是用于评价具有多投入多产出决策问题有效性的一种方法. 以往DEA方法的主要研究工作集中在以下几个方面: 不同类型DEA 模型的引进[2-8], 网络 DEA 模型的引进[9], DEA方法与生产函数理论的结合[10], 交叉效率与广义DEA 方法的引进[11-13], 以及 DEA 方法的各种应用[14]等. 然而, 关于决策单元之间关系的研究却非常少见.

2002 年, 马占新等[15]结合了偏序集理论与数据包络分析方法, 从 Pareto 有效解角度证明了生产可能集中决策单元有效的充分必要条件是该决策单元为相应偏序集中的极大元. 与文献 [15] 不同的是, 文献 [16] 在 CCR 模型中直接引进了决策单元之间的偏序关系, 并分别从定义和实验两个角度建立了决策单元之间的偏序关系. 结果表明, 有效的决策单元必是决策单元构成偏序集中的极大元, 但反之未必成立, 即极大元未必是有效的. 该结论说明了有效决策单元必然能够建立相对其他决策单元的绝对优势, 但决策单元对其他决策单元的相对优势并不一定构成对所有决策单元之间的绝对优势. 该结论同时也表明了决策单元即使对每个决策单元均存在某一方面的优势, 但不一定存在某种优势适合于所有决策单元, 而对有效决策单元而言这种优势却是存在的.

最近, 国外已经有专家开始研究基于半格的数据

收稿日期: 2013-10-26；**修回日期:** 2014-03-18.

基金项目: 内蒙古自然科学基金项目 (2014MS0707)；国家自然科学基金项目 (71401084)；内蒙古自治区高等学校创新团队发展项目 (NMGIRT1404)；内蒙古应用技术研究与开发基金项目 (20130603).

作者简介: 木仁 (1982−), 男, 副教授, 博士, 从事评价与决策理论及数学建模等研究；马占新 (1970−), 男, 教授, 博士生导师, 从事数据包络分析方法等研究.

包络分析方法[17]. 该方法克服了传统数据包络分析方法中凸性的要求, 并提供了基于半格的生产技术理论, 但并未涉及到决策单元交并运算的概念.

虽然研究工作者们对决策单元特殊关系展开了初步的研究, 但对决策单元特殊关系建立方法的研究非常少见. 对此, 本文提出了基于最短距离的决策单元特殊关系建立理论. 首先, 通过引进决策单元之间的不同最短距离定义, 对指定的决策单元引进交并运算; 然后, 引进基于格理论的数据包络分析方法相关性质; 最后, 为了将决策单元的特殊关系利用图形展示出来, 提供了决策单元之间格论关系建立相关算法及其格论关系图绘制程序. 这些方法在为决策单元的投影与改进[18] 提供了最新思路的同时, 也为企业单位的合作与竞争, 以及各种团队的建设提供了相关参考依据.

1 基本概念与定义

本节将简要介绍决策问题及其常用变量表示方式, 并引进偏序集与格的定义等概念.

假设某部门或单位具有 n 个决策单元 (DMU)[14], 每个决策单元有 m 种类型的"投入"(表示对"资源"的耗费) 和 s 种类型的"产出"(消耗"资源"之后, 表明"成效"的一些指标), 第 $i\,(i=1,2,\cdots,n)$ 个决策单元的投入产出向量表示为 $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$. 其中: $\boldsymbol{X}_i = (x_{1i}, x_{2i}, \cdots, x_{mi})^{\mathrm{T}}$, $\boldsymbol{Y}_i = (y_{1i}, y_{2i}, \cdots, y_{si})^{\mathrm{T}}$, 取定产出指标权重 $\boldsymbol{u} = (u_1, u_2, \cdots, u_s)^{\mathrm{T}}$, 投入指标权重 $\boldsymbol{v} = (v_1, v_2, \cdots, v_m)^{\mathrm{T}}$.

定义 1 若集合 P 上的一个二元关系 $\succ$ 满足自反性、反对称性和传递性, 则称二元关系 $\succ$ 为一个偏序关系. 偏序关系的集合 P 称为偏序集[16], 记为 $(P, \succ)$.

定义 2 偏序集 $(P, \succ)$ 称为格[19], 如果对于任意 $a, b \in P$, $\mathrm{Sup}\{a, b\}$ 和 $\mathrm{Inf}\{a, b\}$ 都存在, 则通常将 $\mathrm{Sup}\{a, b\}$ 和 $\mathrm{Inf}\{a, b\}$ 记为 $a \bigvee b$ 和 $a \bigwedge b$, 并读作 a 与 b 的并和交.

由此易知, 所谓格就是任意两个元都存在交并运算的偏序集. 为了加以区别, 将格记为 $L = (L, \bigvee, \bigwedge)$.

2 决策单元之间的偏序关系与交并运算

为了建立决策单元之间的特殊关系, 本节将引进决策单元之间的偏序关系、交并运算以及零元与单位元的概念.

定义 3 (决策单元之间的偏序关系引进) 两个决策单元 $\mathrm{DMU}_i(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ 和 $\mathrm{DMU}_j(\boldsymbol{X}_j, \boldsymbol{Y}_j)$ 存在偏序关系, 如果下述两个条件中有一个成立:

1) 如果 $\boldsymbol{X}_i \geqslant \boldsymbol{X}_j$, $\boldsymbol{Y}_i \leqslant \boldsymbol{Y}_j$, 则将其记为 $\mathrm{DMU}_i \prec \mathrm{DMU}_j$, 并读作 DMU_i 次于 DMU_j; 特别地, 如果 $\boldsymbol{X}_i > \boldsymbol{X}_j$, $\boldsymbol{Y}_i \leqslant \boldsymbol{Y}_j$ 或 $\boldsymbol{X}_i \geqslant \boldsymbol{X}_j$, $\boldsymbol{Y}_i < \boldsymbol{Y}_j$, 则称决策单元 DMU_i 小于决策单元 DMU_j , 并将其记为 $\mathrm{DMU}_i < \mathrm{DMU}_j$.

2) 如果 $\boldsymbol{X}_i \leqslant \boldsymbol{X}_j$, $\boldsymbol{Y}_i \geqslant \boldsymbol{Y}_j$, 则将其记为 $\mathrm{DMU}_i \succ \mathrm{DMU}_j$, 并读作 DMU_i 优于 DMU_j; 特别地, 如果 $\boldsymbol{X}_i < \boldsymbol{X}_j$, $\boldsymbol{Y}_i \geqslant \boldsymbol{Y}_j$ 或 $\boldsymbol{X}_i \leqslant \boldsymbol{X}_j$, $\boldsymbol{Y}_i > \boldsymbol{Y}_j$, 则称决策单元 DMU_i 大于决策单元 DMU_j , 并将其记为 $\mathrm{DMU}_i > \mathrm{DMU}_j$.

定义 4 (决策单元之间的交并运算的引进) 决策单元 $\mathrm{DMU}_i(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ 和 $\mathrm{DMU}_j(\boldsymbol{X}_j, \boldsymbol{Y}_j)$ 的交并运算定义如下:

$$\mathrm{DMU}_i \bigvee \mathrm{DMU}_j = (\boldsymbol{X}_i, \boldsymbol{Y}_i) \bigvee (\boldsymbol{X}_j, \boldsymbol{Y}_j) = (\min\{\boldsymbol{X}_i, \boldsymbol{X}_j\}, \max\{\boldsymbol{Y}_i, \boldsymbol{Y}_j\}),$$

$$\mathrm{DMU}_i \bigwedge \mathrm{DMU}_j = (\boldsymbol{X}_i, \boldsymbol{Y}_i) \bigwedge (\boldsymbol{X}_j, \boldsymbol{Y}_j) = (\max\{\boldsymbol{X}_i, \boldsymbol{X}_j\}, \min\{\boldsymbol{Y}_i, \boldsymbol{Y}_j\}).$$

其中

$$\min\{\boldsymbol{X}_i, \boldsymbol{X}_j\} = (\min\{x_{1i}, x_{1j}\}, \min\{x_{2i}, x_{2j}\}, \cdots, \min\{x_{mi}, x_{mj}\})^{\mathrm{T}},$$

$$\max\{\boldsymbol{Y}_i, \boldsymbol{Y}_j\} = (\max\{y_{1i}, y_{1j}\}, \max\{y_{2i}, y_{2j}\}, \cdots, \max\{y_{si}, y_{sj}\})^{\mathrm{T}},$$

$$\max\{\boldsymbol{X}_i, \boldsymbol{X}_j\} = (\max\{x_{1i}, x_{1j}\}, \max\{x_{2i}, x_{2j}\}, \cdots, \max\{x_{mi}, x_{mj}\})^{\mathrm{T}},$$

$$\min\{\boldsymbol{Y}_i, \boldsymbol{Y}_j\} = (\min\{y_{1i}, y_{1j}\}, \min\{y_{2i}, y_{2j}\}, \cdots, \min\{y_{si}, y_{sj}\})^{\mathrm{T}}.$$

定义 5 (数据包络分析方法中引进交并运算后的零元与单位元) 决策单元 DMU_i 称为零元, 如果对于任意的 DMU_j 均有 $\mathrm{DMU}_i \bigwedge \mathrm{DMU}_j = \mathrm{DMU}_i$; 决策单元 DMU_i 称为单位元, 如果对于任意的 DMU_j 均有 $\mathrm{DMU}_i \bigvee \mathrm{DMU}_j = \mathrm{DMU}_i$.

由定义 5 易知零元与单位元的投入产出数据. 其中: 零元的投入产出数据为

$$\boldsymbol{X} = (\max\{x_{11}, x_{12}, \cdots, x_{1n}\}, \max\{x_{21}, x_{22}, \cdots, x_{2n}\}, \cdots, \max\{x_{m1}, x_{m2}, \cdots, x_{mn}\})^{\mathrm{T}},$$

$$\boldsymbol{Y} = (\min\{y_{11}, y_{12}, \cdots, y_{1n}\}, \min\{y_{21}, y_{22}, \cdots, y_{2n}\}, \cdots, \min\{y_{s1}, y_{s2}, \cdots, y_{sn}\})^{\mathrm{T}};$$

单位元的投入产出数据为

$$\boldsymbol{X} = (\min\{x_{11}, x_{12}, \cdots, x_{1n}\}, \min\{x_{21}, x_{22}, \cdots, x_{2n}\}, \cdots, \min\{x_{m1}, x_{m2}, \cdots, x_{mn}\})^{\mathrm{T}},$$

$$\boldsymbol{Y} = (\max\{y_{11}, y_{12}, \cdots, y_{1n}\}, \max\{y_{21}, y_{22}, \cdots, y_{2n}\}, \cdots, \max\{y_{s1}, y_{s2}, \cdots, y_{sn}\})^{\mathrm{T}}.$$

3　决策单元之间的距离

通过引进决策单元之间的交并运算和相关样本决策单元, 能够建立起各个决策单元之间的特殊关系. 然而, 在实际问题中这种特殊关系的构造需要可靠的依据. 利用数学理论引进决策单元之间的距离概念, 并根据该距离建立决策单元之间的特殊关系, 那么这种特殊关系的建立就有了理论技术支撑.

3.1　决策单元投入产出数据处理

数据包络分析方法在评价决策单元之间的有效性时, 由于所选取的指标权重的任意性, 可不考虑决策单元投入产出数据的量纲; 然而, 在研究决策单元之间的差异、距离等概念时, 不考虑决策单元投入产出数据的量纲显然是不合理的. 在统计学理论中可以通过中心化变换、标准化变换、极差标准化变换及极差正规化变换对数据和变量进行无量纲化处理. 中心化变换和标准化变换虽然对数据进行了无量纲化处理, 但数据的变化区间仍然不是固定区间, 导致在对不同组其他的数据进行比较时并不通用. 极差标准化变换虽然对数据进行了无量纲化处理, 并将数据投影到了 $[-1,1]$ 区间内, 但由于存在负数的情形, 在数据包络分析方法中也不是一种较科学的无量纲化数据处理方法. 极差正规化变换不仅消除了决策单元之间的量纲, 同时也将数据投影到了 $[0,1]$ 区间内, 因此该方法比较适合对数据包络分析方法中的数据进行相关处理. 为了了解该方法, 下面对数据包络分析方法中的投入产出数据进行相关处理.

3.1.1　投入数据的极差正规化变换方法

已知决策单元的投入数据为 $\boldsymbol{X}_i = (x_{1i}, x_{2i}, \cdots, x_{mi})^{\mathrm{T}}$, $i = 1, 2, \cdots, n$, 经极差正规化变换后的第 i 个决策单元的投入数据为

$$\bar{\boldsymbol{X}}_i = (\bar{x}_{1i}, \bar{x}_{2i}, \cdots, \bar{x}_{mi})^{\mathrm{T}},\ i = 1, 2, \cdots, n,$$

其中

$$\bar{x}_{ki} = \frac{x_{ki} - \min\{x_{k1}, x_{k2}, \cdots, x_{kn}\}}{\max\{x_{k1}, x_{k2}, \cdots, x_{kn}\} - \min\{x_{k1}, x_{k2}, \cdots, x_{kn}\}},\quad k = 1, 2, \cdots, m.$$

3.1.2　产出数据的极差正规化变换方法

已知决策单元的产出数据为 $\boldsymbol{Y}_i = (y_{1i}, y_{2i}, \cdots, y_{si})^{\mathrm{T}}$, $i = 1, 2, \cdots, n$, 经极差正规化变换后的第 i 个决策单元的产出数据为

$$\bar{\boldsymbol{Y}}_i = (\bar{y}_{1i}, \bar{y}_{2i}, \cdots, \bar{y}_{si})^{\mathrm{T}},\ i = 1, 2, \cdots, n,$$

其中

$$\bar{y}_{ki} = \frac{y_{ki} - \min\{y_{k1}, y_{k2}, \cdots, y_{kn}\}}{\max\{y_{k1}, y_{k2}, \cdots, y_{kn}\} - \min\{y_{k1}, y_{k2}, \cdots, y_{kn}\}},\quad k = 1, 2, \cdots, s.$$

决策单元的投入产出数据虽然经极差正规化变换后消除了量纲, 并投影到了 $[0,1]$ 区间内, 但当决策单元之间的投入产出数据之间的差距较小时, 即当分母较少时可能将原数据之间的差距扩大化. 在这种情况下, 本文提出对于投入产出数据的相关处理方法. 假设决策单元的投入产出数据均是非负的, 如果存在负数, 则可通过平移等方式将数据均转化为非负数据.

3.1.3　投入数据的放大缩小化处理

已知决策单元的投入数据为 $\boldsymbol{X}_i = (x_{1i}, x_{2i}, \cdots, x_{mi})^{\mathrm{T}}$, $i = 1, 2, \cdots, n$, 该数据的放大缩小化处理后的投入数据变为

$$\bar{\boldsymbol{X}}_i = (\bar{x}_{1i}, \bar{x}_{2i}, \cdots, \bar{x}_{mi})^{\mathrm{T}},\ i = 1, 2, \cdots, n,$$

其中

$$\bar{x}_{ki} = \frac{x_{ki}}{\max\{x_{k1}, x_{k2}, \cdots, x_{kn}\}},\ k = 1, 2, \cdots, m.$$

3.1.4　产出数据的放大缩小化处理

已知决策单元的投入数据为 $\boldsymbol{Y}_i = (y_{1i}, y_{2i}, \cdots, y_{si})^{\mathrm{T}}$, $i = 1, 2, \cdots, n$, 该数据的放大缩小化处理后的产出数据变为

$$\bar{\boldsymbol{Y}}_i = (\bar{y}_{1i}, \bar{y}_{2i}, \cdots, \bar{y}_{si})^{\mathrm{T}},\ i = 1, 2, \cdots, n,$$

其中

$$\bar{y}_{ki} = \frac{y_{ki}}{\max\{y_{k1}, y_{k2}, \cdots, y_{kn}\}},\ k = 1, 2, \cdots, s.$$

决策单元的投入产出数据经放大缩小化处理后不仅投影到了 $[0,1]$ 区间内, 同时消除了决策单元投入产出数据的量纲, 而更为重要的是因为数据是由等比例放大或缩小得到的, 故使用经处理后的数据计算出的决策单元的效率值与原决策单元的效率值是相同的. 从这几点上来讲, 投入产出数据的放大缩小化处理方式更为合理.

3.2　决策单元之间的距离

在本节中假设决策单元的投入产出数据已经根据实际问题的需要进行了必要地处理. 经处理后的决策单元投入产出数据如下: $\bar{\boldsymbol{X}}_i = (\bar{x}_{1i}, \bar{x}_{2i}, \cdots, \bar{x}_{mi})^{\mathrm{T}}$, $\bar{\boldsymbol{Y}}_i = (\bar{y}_{1i}, \bar{y}_{2i}, \cdots, \bar{y}_{si})^{\mathrm{T}}$, $i = 1, 2, \cdots, n$.

1) 绝对值距离.

决策单元 DMU_i 与 DMU_j 之间的绝对值距离定义如下:

$$d_{ij} = \sum_{k=1}^{m} |\bar{x}_{ki} - \bar{x}_{kj}| + \sum_{k=1}^{s} |\bar{y}_{ki} - \bar{y}_{kj}|.$$

2) 欧几里得距离.

决策单元 DMU_i 与 DMU_j 之间的欧几里得距离定义如下:

$$d_{ij}=\sqrt{\sum_{k=1}^{m}(\bar{x}_{ki}-\bar{x}_{kj})^2+\sum_{k=1}^{s}(\bar{y}_{ki}-\bar{y}_{kj})^2}.$$

3) 闵克夫斯基距离.

决策单元 DMU_i 与 DMU_j 之间的闵克夫斯基距离定义如下:

$$d_{ij}=\Big(\sum_{k=1}^{m}(\bar{x}_{ki}-\bar{x}_{kj})^q+\sum_{k=1}^{s}(\bar{y}_{ki}-\bar{y}_{kj})^q\Big)^{\frac{1}{q}}.$$

4) 切比雪夫距离.

决策单元 DMU_i 与 DMU_j 之间的切比雪夫距离定义如下:

$$d_{ij}=\max\{\max_{1\leqslant k\leqslant m}(\bar{x}_{ki}-\bar{x}_{kj}),\max_{1\leqslant k\leqslant s}(\bar{y}_{ki}-\bar{y}_{kj})\}.$$

由此容易看出, 绝对值距离、欧几里得距离和切比雪夫距离均是闵克夫斯基距离的特殊情形, 因此只需给出闵克夫斯基距离的计算方法即可计算出所有的距离.

3.2.1 决策单元之间距离的应用

决策单元在生产前沿面上的投影过程本质上是将一个无效或弱有效的决策单元投影为有效的决策单元的过程, 而决策单元的投影又可分为面向投入的投影、面向产出的投影和面向投入产出的投影. 根据决策单元之间的距离可对决策单元进行投影.

1) 决策单元面向投入产出的最新投影方式.

传统数据包络分析方法中的投影方式所获得的决策单元未必是一个有效的现存决策单元. 如果根据决策单元之间的距离, 将一个无效或弱有效的决策单元投影为与该决策单元最近的有效决策单元, 则这种投影方式所获得的决策单元自然就是一个现存的决策单元了.

2) 决策单元面向投入的改进方式.

决策单元在生产前沿面上的面向投入的投影方式是, 在通过减少投入而不改变产出的前提下, 将一个无效或弱有效的决策单元投影为有效决策单元的过程. 决策单元的面向投入产出的投影方式可分解为面向投入的改进方式和面向产出的改进方式. 通过这两种改进方式最终可将决策单元投影为有效的决策单元. 下面提出决策单元面向产出的改进方式.

在考虑决策单元与决策单元之间的距离时, 由于无需考虑产出数据的具体情况, 本文仅考虑决策单元的投入数据之间的差距, 从而对其进行改进. 此时, 可引进决策单元投入数据之间的距离, 获得决策单元 DMU_i 和 DMU_j 之间的投入数据的闵克夫斯基距离定义如下:

$$d_{ij}=\Big(\sum_{k=1}^{m}(\bar{x}_{ki}-\bar{x}_{kj})^q\Big)^{\frac{1}{q}}.$$

通过计算上述距离便可找到面向投入的决策单元的最易改进方式.

3) 决策单元面向产出的改进方式.

与前面的讨论类似, 可获得决策单元 DMU_i 与 DMU_j 之间产出数据的闵克夫斯基距离定义如下:

$$d_{ij}=\Big(\sum_{k=1}^{s}(\bar{y}_{ki}-\bar{y}_{kj})^q\Big)^{\frac{1}{q}}.$$

通过计算上述距离便可找到面向产出的决策单元的最易改进方式.

3.2.2 总 结

决策单元之间的距离概念为决策单元的投影与改进提供了新的理论依据. 通过决策单元之间距离的大小可对决策单元进行改进. 距离越小表明决策单元之间的差异越小, 在实际中改进起来较容易实现. 不同的距离定义方式提供了决策单元改进的相关信息. 由决策单元之间的绝对值距离可以知道在将一个效率值较低的决策单元改进为效率值较高的决策单元时所需要改进的“总距离”. 决策单元之间的欧几里得距离体现了两个决策单元之间的“实际距离”; 而决策单元之间的切比雪夫距离体现了决策单元改进幅度最大的那一项投入或产出数据; 因此, 可将其定义为决策单元改进的“难度系数”. 在通常情况下, 将一个较次的数据改进为较好的数据实现起来较为困难, 而将一个较好的数据改进为较次的数据相对容易. 例如, 将决策单元的投入数据减少起来相对困难, 但增加投入却比较容易实现. 因此, 在进行决策单元的改进时, 更应考虑将较次的数据改进为较好的数据, 从而在利用决策单元之间的距离对决策单元进行改进时, 更应考虑较次的数据与较好数据之间的差异. 如果决策单元 DMU_i 朝向决策单元 DMU_j 改进, 则定义 DMU_i 与 DMU_j 之间的改进距离为

$$d_{ij}=\Big(\sum_{k=1}^{m}\Big(\frac{1+\mathrm{sig}(\bar{x}_{ki}-\bar{x}_{kj})}{2}\Big)(\bar{x}_{ki}-\bar{x}_{kj})^q+\sum_{k=1}^{s}\Big(\frac{1-\mathrm{sig}(\bar{y}_{ki}-\bar{y}_{kj})}{2}\Big)(\bar{y}_{ki}-\bar{y}_{kj})^q\Big)^{\frac{1}{q}},$$

其中 $\mathrm{sig}(\cdot)$ 为符号函数, 具体定义如下:

$$\mathrm{sig}(x)=\begin{cases}0,\ x=0;\\1,\ x>0;\\-1,\ x<0.\end{cases}$$

在得到决策单元之间的改进距离后, 便可根据实

际问题的需要对决策单元进行交并运算, 争取实现将无效的或弱有效的决策单元投影为有效的决策单元.

4 基于格理论的数据包络分析方法典型性质

为了深入了解在通过引进交并运算建立决策单元格关系后的基于格理论的数据包络分析方法, 下面给出相关定理及其证明.

定理 1 基于格理论的数据包络分析方法中仅存在唯一的单位元与零元, 且有效的决策单元是唯一的.

证明 根据零元与单位元的定义可知二者必是唯一的, 下面证明只有单位元是有效的. 事实上, 假定 DMU_i 是单位元, 因为对于任意的 DMU_j, 均有 $\mathrm{DMU}_i \bigvee \mathrm{DMU}_j = \mathrm{DMU}_i$, 故 $\mathrm{DMU}_i \succ \mathrm{DMU}_j$, 即 $\boldsymbol{X}_i \leqslant \boldsymbol{X}_j$, $\boldsymbol{Y}_i \geqslant \boldsymbol{Y}_j$, 因此只要 $\mathrm{DMU}_j \neq \mathrm{DMU}_i$, 则必存在 x_{ki}, x_{kj} 或 y_{ti}, y_{tj} 使得 $x_{ki} < x_{kj}$ 或 $y_{ti} < y_{tj}$. 由此可知, 如果存在其他有效的决策单元 DMU_j, 则在使该决策单元有效的最优权重下, 因为每个权重都大于 0, 故 DMU_i 的效率值大于 1, 与假设矛盾, 由此定理得证. □

定理 2 如果 $\mathrm{DMU}_i \prec \mathrm{DMU}_j$, $\mathrm{DMU}_i \neq \mathrm{DMU}_j$, 则 DMU_i 至多是弱有效的; 如果 $\mathrm{DMU}_i < \mathrm{DMU}_j$, $\mathrm{DMU}_i \neq \mathrm{DMU}_j$, 则 DMU_i 必是无效的.

证明 如果 $\mathrm{DMU}_i \prec \mathrm{DMU}_j$, $\mathrm{DMU}_i \neq \mathrm{DMU}_j$, 则必存在 x_{ki}, x_{kj} 或 y_{ti}, y_{tj} 使得 $x_{ki} > x_{kj}$ 或 $y_{ti} < y_{tj}$, 从而 DMU_i 的效率值将小于等于 DMU_j 的效率值, 且只有将那些使得 $x_{ki} > x_{kj}$ 和 $y_{ti} < y_{tj}$ 的所有权系数设为 0, 才能保证这两个决策单元的效率值是相等的. 在此情况下, 如果 DMU_j 是有效的, 则 DMU_i 至多是弱有效的; 如果 DMU_j 是弱有效的, 则 DMU_i 至多是弱有效的; 如果 DMU_j 是无效的, 则 DMU_i 必是无效的.

如果 $\mathrm{DMU}_i < \mathrm{DMU}_j$, 则 $\boldsymbol{X}_i > \boldsymbol{X}_j$, $\boldsymbol{Y}_i \geqslant \boldsymbol{Y}_j$, 或 $\boldsymbol{X}_i \leqslant \boldsymbol{X}_j$, $\boldsymbol{Y}_i < \boldsymbol{Y}_j$, 因此在任何一组权重 $\boldsymbol{u}^{\mathrm{T}}$, $\boldsymbol{v}^{\mathrm{T}}$ 下, 均有

$$\frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Y}_i}{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}_i} < \frac{\boldsymbol{u}^{\mathrm{T}}c\boldsymbol{Y}_i}{\boldsymbol{v}^{\mathrm{T}}c\boldsymbol{X}_i} = \frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Y}_j}{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}_j},$$

即决策单元 DMU_i 的效率值小于决策单元 DMU_j 的效率值, 从而 DMU_i 的效率值必小于 1, 这表明 DMU_i 是无效的. □

定理 3 如果 $\mathrm{DMU}_i \bigwedge \mathrm{DMU}_j = \mathrm{DMU}_i$, 则有 $\mathrm{DMU}_i \prec \mathrm{DMU}_j$; 如果 $\mathrm{DMU}_i \bigvee \mathrm{DMU}_j = \mathrm{DMU}_i$, 则有 $\mathrm{DMU}_i \succ \mathrm{DMU}_j$.

证明 如果 $\mathrm{DMU}_i \bigwedge \mathrm{DMU}_j = \mathrm{DMU}_i$, 则有

$$\mathrm{DMU}_i \bigwedge \mathrm{DMU}_j = \mathrm{DMU}_i = (\boldsymbol{X}_i, \boldsymbol{Y}_i) \bigwedge (\boldsymbol{X}_j, \boldsymbol{Y}_j) = (\max(\boldsymbol{X}_i, \boldsymbol{X}_j), (\boldsymbol{Y}_i, \boldsymbol{Y}_j)) = (\boldsymbol{X}_i, \boldsymbol{Y}_i).$$

这表明 $\max(\boldsymbol{X}_i, \boldsymbol{X}_j) = \boldsymbol{X}_i, \min(\boldsymbol{Y}_i, \boldsymbol{Y}_j) = \boldsymbol{Y}_i$, 从而有 $\boldsymbol{X}_i \geqslant \boldsymbol{X}_j$, $\boldsymbol{Y}_i \leqslant \boldsymbol{Y}_j$, 即 $\mathrm{DMU}_i \prec \mathrm{DMU}_j$. 类似可以证明, 如果 $\mathrm{DMU}_i \bigvee \mathrm{DMU}_j = \mathrm{DMU}_i$, 则有 $\mathrm{DMU}_i \succ \mathrm{DMU}_j$. □

定理 4 若 $\mathrm{DMU}_i \bigwedge \mathrm{DMU}_j \neq \mathrm{DMU}_i$ 且 $\mathrm{DMU}_i \bigwedge \mathrm{DMU}_j \neq \mathrm{DMU}_j$, 则 DMU_i 与 DMU_j 互不可比, 且存在某两组权重使其中一个决策单元的效率值大于另一决策单元的效率值.

证明 由于 $\mathrm{DMU}_i \bigwedge \mathrm{DMU}_j \neq \mathrm{DMU}_i$ 且 $\mathrm{DMU}_i \bigwedge \mathrm{DMU}_j \neq \mathrm{DMU}_j$, 由定理 3 可知 DMU_i 与 DMU_j 互不可比, 从而有 $\boldsymbol{X}_i \geqslant \boldsymbol{X}_j$, $\boldsymbol{Y}_i \leqslant \boldsymbol{Y}_j$ 以及 $\boldsymbol{X}_i \leqslant \boldsymbol{X}_j$, $\boldsymbol{Y}_i \geqslant \boldsymbol{Y}_j$ 均不成立. 此时必存在 $x_{ki}, x_{kj}, y_{ti}, y_{tj}$, 使得 $x_{ki} > x_{kj}$, $y_{ti} \leqslant y_{tj}$. 这样可找到一组权重使 DMU_i 的效率值大于 DMU_j 的效率值. 同时也存在 x_{ki}, x_{kj}, y_{ti}, y_{tj} 使得 $x_{ki} \geqslant x_{kj}$, $y_{ti} < y_{tj}$, 这样也能找到一组权重使 DMU_j 的效率值大于 DMU_i 的效率值. □

定理 5 在任意一个具有多投入多产出的决策问题中, 通过构造合理的样本决策单元, 可建立决策单元之间的格论关系.

证明 事实上, 根据决策单元之间交并运算的定义, 只要在原决策问题中再重新增加新引进的经交并运算后获得的决策单元 (称之为样本决策单元[13]), 便可建立起决策单元之间的格论关系. □

5 基于格理论的数据包络分析方法相关算法

根据决策单元偏序关系的定义, 可确定决策单元之间的偏序关系. 然而, 当决策单元个数较多时, 如果采用人工计算的方法不仅速度慢且准确率也不高, 因此需要提供计算决策单元之间偏序关系的相关算法; 其次, 为了计算出各个决策单元之间的交并运算也需要提供相关算法; 最后, 为了将决策单元之间的格论关系图精确地绘制出来, 需要对各个决策单元进行相关数据处理并获得最终效果图. 算法的具体步骤如下.

Step 1: 对决策单元的投入产出数据进行相关处理, 并选好决策单元之间所用的距离;

Step 2: 计算出所有决策单元中距离最小的两个决策单元, 并生成这两个决策单元在交并运算后的决策单元;

Step 3: 判断是否已构造出零元或单位元, 如果没

有则返回 Step 2, 否则转 Step 4;

Step 4: 根据定义确定由 Step 3 所获得的所有决策单元之间的偏序关系;

Step 5: 计算所有决策单元的效率值和弱有效效率值;

Step 6: 计算由 Step 3 获得的所有决策单元的经单位化后的平均投入数据和产出数据;

Step 7: 以各决策单元的经单位化后的平均投入数据和产出数据作为横纵轴, 以决策单元的效率值作为竖轴在空间中绘制各个决策单元, 无效、有效和弱有效的决策单元用不同的符号绘制;

Step 8: 绘制各个决策单元之间的偏序关系图, 原有决策单元的偏序关系和新增决策单元之间的偏序关系将以不同的线段进行连接, 最终将生成决策单元之间的格论关系 Hasse 图.

6 实例演示

试利用基于格理论的数据包络分析方法评价内蒙古自治区自然科学基金分配问题. 表 1 给出了相关数据.

利用 Matlab 软件可计算获得决策单元之间的偏序关系, 为了将结果可视化地显示出来, 图 1 给出了决策单元之间的偏序关系, 表 2 给出了决策单元偏序关系矩阵. 在本文所有图中, 横坐标表示决策单元的投入数据经单位化后求和的平均投入数据, 纵坐标表示产出数据经单位化后求和的平均产出数据, 竖坐标表示决策单元的效率值, 黑点代表无效的决策单元, 星星代表有效的决策单元.

表 1 "十一五"期间自治区批准自然科学基金及获批国家自然科学基金数量、经费分布表

序号	学部	自治区批准数量	自治区批准经费	获批国家数量	获批国家经费
1	数理科学部	94	260	37	1 020
2	化学科学部	93	431	53	1 207
3	生命科学部	688	2 498	262	6 574
4	地球科学部	52	182	38	887
5	工程与材料科学部	165	752	88	2 400
6	信息科学部	88	296	32	753
7	管理科学部	57	187	19	465.8

表 2 CCR 模型中决策单元投入产出数据偏序关系表

学 部	数理科学部	化学科学部	生命科学部	地球科学部	工程与材料科学部	信息科学部	管理科学部
数理科学部	1	0	0	1.279	0	0	0
化学科学部	0	1	0	1.394	0	0	0
生命科学部	1.238	1.349	1	6.819	1.345	0	0
地球科学部	0	0	0	1	0	0	0
工程与材料	0	0	0	1.467	1	0	0
信息科学部	1.071	1.063	0	1.777	1.03	1	0
管理科学部	1.093	1.048	1.002	1.502	1.033	1.013	1

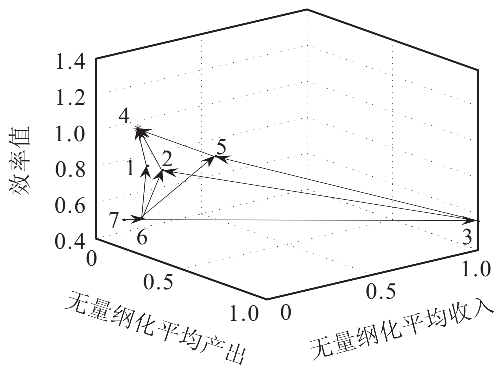

图 1 决策单元偏序关系图

由图 1 可以直观地看出决策单元之间所存在的偏序关系. 由表 2 可观察出决策单元之间存在的差距. 通过偏序关系容易得知决策单元可能的排序结果: 地球科学部确定排在第一; 管理科学部确定排在最后; 生命科学部和信息科学部排在管理科学部的前面, 但前后顺序不确定; 化学科学部、数学科学部和工程与材料科学部将排在生命科学部和信息科学部的前面, 但具体顺序也不确定. 进一步不难得知, 如果利用偏序关系对决策单元进行分类, 则分类结果为: 地球科学部为一类, 管理科学部为一类, 生命科学部和信息科学部为一类, 数理科学部、化学科学部和工程与材料科学部为一类, 每一类的特点是在该类中无法确定前后排序顺序. 最后, 通过偏序关系还可得知决策单元的逐步改进方式和投影方式. 所有学部的有效投影

表 3　增加产出指标后"十一五"期间自治区批准自然科学基金及获批国家自然科学基金数量、经费分布表

序号	学　部	自治区批准数量	自治区批准经费	获批国家数量	获批国家经费	发表论文	经济社会贡献
1	数理科学部	94	260	37	1020	200	2
2	化学科学部	93	431	53	1207	150	2
3	生命科学部	688	2498	262	6574	1800	10
4	地球科学部	52	182	38	887	80	1
5	工程与材料科学部	165	752	88	2400	400	3
6	信息科学部	88	296	32	753	60	5
7	管理科学部	57	187	19	465.8	30	7

模式为地球科学部. 管理科学部可选择不同的逐步改进模式最终投影为地球科学部, 如先投影为信息科学部, 然后再投影为数理科学部, 最终投影为地球科学部. 其他学部也可选择类似的投影模式. 不难发现, 此投影模式的优点在于给决策者提供逐步改进方式及现有的投入产出模式, 从而在实践中更易实现.

容易证明, 在例 1 中决策单元之间构成的偏序关系为格论关系, 没有必要通过引进相关方法来建立格论关系. 然而, 如果再增加两个产出指标, 如发表学术论文和对经济社会的贡献指标, 则如表 3 所示.

图 2 给出了添加发表论文及经济社会贡献指标后决策单元的偏序关系图. 由图 2 可知, 决策单元之间不存在任何偏序关系, 但决策单元 2 和 6 是无效的. 此时决策单元 2 和 6 应如何改进为有效决策单元? 如果利用基于格理论的数据包络分析方法对决策单元 2 和 6 引进交并运算, 则得到如图 3 所示的决策单元特殊关系图. 不难发现, 在理想情况下, 决策单元 2 和 6 合作后改进为有效的决策单元 8, 表明了这种合作方式存在着一定意义; 而在最差情况下可能变坏为决策单元 9, 这表明了合作可能存在的风险. 进一步不难发现决策单元 2, 6, 8 和 9 将构成格.

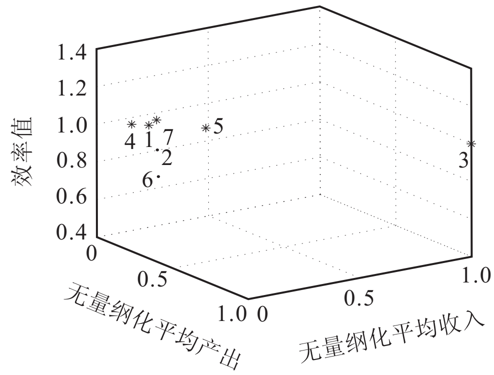

图 2　增加指标后的决策单元偏序关系图

进一步, 如果利用决策单元之间的最短距离概念, 对所有决策单元引进交并运算, 则获得如图 4 所示的决策单元格论关系图.

图 4 中, 新增样本决策单元为原有决策单元经交并运算后生成的决策单元, 其中决策单元 16 是有效的单位元, 决策单元 13 是零元. 通过决策单元最终建立起的格论关系, 可获得基于最短距离决策单元合并方式, 以及经合并后的最优投入产出模式和最差投入产出模式.

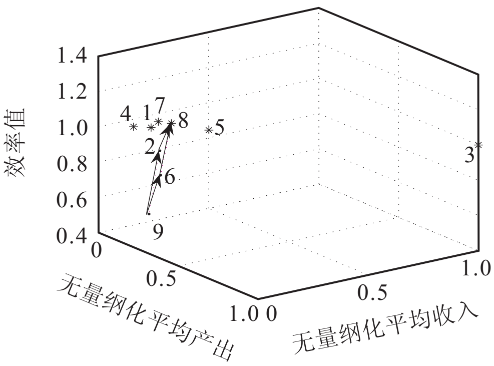

图 3　决策单元格论关系图

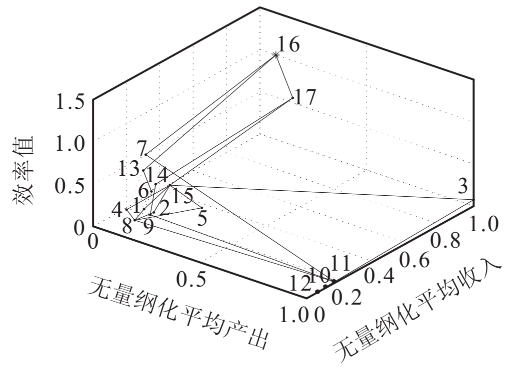

图 4　决策单元总格论关系图

7　结　　论

通过引进决策单元之间的交并运算和距离的概念, 建立了数据包络分析方法与格论之间的特殊关系. 该方法可以为数据包络分析领域提供新的数学理论支撑. 在此过程中, 首先建立的决策单元之间的偏序关系, 不仅为决策者提供了具有较大参考价值的决策信息, 也为传统决策单元的投影方式提供了新思路. 这种投影方法所投影到的是一个现存的决策单元, 而传统投影方式不具备这样性质. 在基于不同距离的决策单元之间引进并运算为企业的合作和各种团队的建设提供了可能理想的投入产出模式, 而交运

算的引进则提供了可能最差的投入产出模式. 与此同时, 基于格理论的数据包络分析方法中的零元和单位元, 给出了今后所有决策单元改进或变坏时的最优及最差情形, 而决策单元之间的偏序关系图和格论关系图的引进, 使得众多决策者可直观地观察出各个决策单元的相关信息, 以便为今后的决策提供更可靠的决策依据.

参考文献(References)

[1] Charnes, Cooper W W, Rhoes E. Measuring the efficiency of decision making units[J]. European J of Operational Research, 1978, 2(6): 429-444.

[2] Banker R D, Charnes A, Cooper W W. Some models for estimating technical and scale inefficiencies in data envelopment analysis[J]. Management Sciences, 1984, 30(9): 1078-1092.

[3] Färe R S, Grosskopf S. A nonparametric cost approach to scale efficiency[J]. Scandinavian J of Economics, 1985, 87 (4): 594-604.

[4] Seiford L M, Thrall R M. Recent development in DEA: The mathematical programming approach to frontier analysis[J]. J of Economics, 1990, 46 (1/2): 7-38.

[5] Sengupta J K. Data envelopment analysis for efficiency measurement in the stochastic case[J]. Computers and Operations Research, 1987, 14(2): 117-129.

[6] Yang Y S. Data envelopment analysis(DEA) model with interval gray numbers[J]. Information and Management Sciences, 1998, 9(4): 11-23.

[7] Wang Y M, Chin K S. Fuzzy data envelopment analysis: A fuzzy expected value approach[J]. Expert Systems with Application, 2011, 38(9): 11678-11685.

[8] Wang Y M, Greatbanks R, Yang J B. Interval efficiency assessment using data envelopment analysis[J]. Fuzzy Sets and Systems, 2005, 153(3): 347-370.

[9] 魏权龄. 评价相对有效性的数据包络分析模型—— DEA 和网络 DEA[M]. 北京: 中国人民大学出版社, 2012: 1-11.
(Wei Q L. Relative efficiency evaluation method of Data envelopment analysis model — DEA and network DEA[M]. Beijing: Renmin University of China, 2012: 1-11.)

[10] 魏权龄. 数据包络分析[M]. 北京: 科学出版社, 2004: 170-183.
(Wei Q L. Data envelopment analysis[M]. Beijing: Science Press, 2006: 170-183.)

[11] Wang Y M, Chin K S. Some alternative models for DEA cross-efficiency evaluation[J]. Int J of Production Economics, 2010, 128(1): 332-338.

[12] Liang L, Wu J, Cook D C, et al. Alternative secondary goals in DEA cross-efficiency evaluation[J]. Int J of Production Economics, 2008, 113(2): 1025-1030.

[13] 马占新. 广义数据包络分析方法[M]. 北京: 科学出版社, 2012：1-12．
(Ma Z X. Generalized data envelopment analysis method[M]. Beijing: Science Press, 2012: 1-12.)

[14] 马占新. 数据包络分析模型与方法[M]. 北京：科学出版社, 2010: 7-10.
(Ma Z X. Data envelopment analysis models and methods[M]. Beijing: Science Press, 2010: 7-10.)

[15] 马占新, 唐焕文, 戴印山. 偏序集理论在数据包络分析中的应用研究[J]. 系统工程, 2002, 17(1): 19-25.
(Ma Z X, Tang H W, Dai Y S. Using theory of partially ordered sets to study data envelopment analysis[J]. J of Systems Engineering, 2002, 17(1): 19-25.)

[16] 木仁, 马占新, 崔巍. 基于偏序集理论的数据包络分析方法[J]. 系统工程与电子技术, 2013, 35(2): 350-356.
(Mu Ren, Ma Z X, Cui W. Fuzzy data envelopment analysis approach based on sample decision making units[J]. J of Systems Engineering and Electronics, 2013, 35(2): 350-356.)

[17] Walter, Liang Q B. On some semilattice structures for production technologies[J]. European J of Operational Research, 2011, 215(3): 740-749.

[18] 王宁沈, 易荣华, 王伟. 决策单元投影研究[J]. 数学的实践与认识, 2009, 39(4): 113-122.
(Wang N S, Yi R H, Wang W. The research on the problem of projections for DMUs[J]. Mathematics Practice and Theory, 2009, 39(4): 113-122.)

[19] Grätzer G. General lattice theory[M]. New York: Academic Press, 1978: 21-30.

(责任编辑: 滕 蓉)