DEAにおける分類データの取扱い方

01001600 成蹊大学 ＊上田 徹 UEDA Tohru

1． まえがき

数値は分からないが、分類（カテゴリ）は分かっている場合に、それが階層的（順序の付いた）カテゴリであれば、最も厳しい環境を意味するカテゴリから順に分析する方法が知られている[1]。また、分類データと量的データが混在する場合の回帰分析法として数量化理論1類がよく知られている。そこで、数量化理論1類、2類での取り扱い方を参考にしてDEAでの分類データの取り込み法を考えてみる。

2． 数量化理論1類からの示唆

数量化理論1類では説明変数に分類データδ_{ij}と量的データx_hが混在すると、回帰式を

$$y^{(k)} = \sum_i \sum_j a_{ij} \delta_{ij}{}^{(k)} + \sum_h b_h x_h{}^{(k)} + c \quad (1)$$

$$\delta_{ij}{}^{(k)} = \begin{cases} 1{:}\text{標本}k\text{がアイテム}i\text{のカテゴリ}j\text{に属するとき} \\ 0{:}\text{その他} \end{cases}$$

のように定式化する。ただし、アイテムiのカテゴリ数をn_iとすると、(n_i-1)個のカテゴリに属さないことが分かれば、その標本は残りのカテゴリに属することが分かることおよび定数項を通じて他の変数の値がゼロの場合の調整ができることからアイテムごとに1カテゴリだけ削除することができる。

DEAでは、分類データδ_{ij}があったときに、1カテゴリだけ削除できるかどうかを考えてみる。ただし、カテゴリ番号が大きいほど大きな数値に対応しているものとする。式(1)の右辺を仮想的入力（または出力）として用いるとき、カテゴリjよりもカテゴリkの方が大きければ

$$a_{ij} \le a_{ik} \quad (2)$$

となる必要がある。たとえば入力10（標本K）と入力100（標本L）の係数がαならば数値データなら仮想的入力において10αと100α（10$\alpha \le$100α）の大きさを占める。入力10をカテゴリ1とし、入力100をカテゴリ2とすると、仮想的入力において標本Kではa_{i1}、標本Lではa_{i2}の大きさを占めるので、数値データと同様の結果を生み出すためには式(1)の条件が必要となる。

CCRモデルでは、式(1)の定数項に相当する項がない。そこですべてのカテゴリを採用すれば、カテゴリ1を持つDMUは

$$a_{i1} \ge 0,\ a_{ij} >> a_{i1} (i \ne j)$$

とすればカテゴリ1でないDMUよりも有利に自分を評価できる。すなわち、カテゴリ1を持つDMUの中で競い合うことになり、それ以外のDMUはより非効率的となる。

また、カテゴリ1を持つDMUの評価では$a_{ij}(j \ne 1)$の値は関係ないため$a_{ij}(j \ne 1)$の値は不定となる。

このとき、$a_{i1}=0$とカテゴリ1を削除することとは同じであり、$a_{i1}>0$なら、a_{i1}という正の定数項を持つBCCモデルをカテゴリ1をもつDMU群に適用したのと同じことになる。

カテゴリk(≥1)を持つDMUは自分を有利にするために$a_{i1}=a_{i2}=\ldots=a_{ik} << a_{ij}\ (j>k)$となる。すなわち、カテゴリ$k$を持つDMUの評価では、$k$以下のカテゴリを持つすべてのDMUが評価対象となる。

２種類以上の分類データがある場合に式(1)を用いると、自分にとって最も有利なカテゴリを持つ分類を重視した評価となる。

BCC モデルでは、もともと定数項があるため、カテゴリ１を削除しても、しなくても同じ結果になる。

このように、元々は数値データであった場合は条件(2)を課せばそれなりの分析ができるが、元々質的データの場合には、カテゴリに関する優劣が定性的に分かっていないと条件(2)を設定できず、カテゴリごとの DMU 群でしか分析できない。たとえば駐車場の有無などは、ある方が集客に有利なので有りを入力のカテゴリ１とすればよい。定性的に判断できない場合には正準相関分析などによりあらかじめカテゴリの順位を決めておかなければならない。

３． __分析例__

図書館データの分析を行う。なお、分析には SAITECH 社の DEA-Solver-Pro 2.0 を用いた。そこでは値をゼロにするとエラーメッセージが出てくるためゼロの代わりに小さな値を与えた。また、式(2)の制約を加えるために AR モデルを用いた。AR-I-C は CCR に式(2)の制約を加えた場合に対応できる。

文献[1]と同じ人口による分類において、カテゴリ数により効率値が異なった例を表１に示す（モデルの最後の数字はカテゴリ数）。これらはすべてカテゴリ２に属している。それらのカテゴリスコア（乗数の値）を表２に示す。AR-I-C-2 は $a_{i1}=0$ と同じであり、この条件を緩められた AR-I-C-3 の方が高い効率値となる。なお、品川区は人口データを削除した BCC-O モデルでは効率値が 0.681 であった。AR-I-C-3 との違いは前者が定数項負であるのに対し、後者が正になっていることである。

職員数も分類データ扱いをし、２種類の分類データありとして分析すると、豊島区は職員数がカテゴリ１、人口がカテゴリ２のため職員数のカテゴリ１のスコアを他より圧倒的に小さくし、人口を無視することで効率値１となる。また、人口、職員数ともにカテゴリ１の台東区は人口のみを分類データとしたときには荒川区を参照 DMU としていたが、荒川区が職員数ではカテゴリ２のため、職員数も分類データ扱いすると効率値１となった。すなわち、分類データのアイテムを増やすごとに競争相手となる DMU は減少するため、アイテム数はむやみに増やしてはいけない。

[1] 刀根：「経営効率性の測定と改善」、節 7.2、日科技連、1993

表１　カテゴリ数により結果が異なった例

	AR-I-C-2	AR-I-C-3	NCN-I-C
新宿	0.653	0.720	0.650
中野	0.838	0.861	0.774
品川	0.839	1.000	0.687
北	0.838	0.864	0.801
葛飾	0.688	0.703	0.766

表２　カテゴリスコア

	AR-I-C-2	AR-I-C-3	
	a_{i2}	a_{i1}	a_{i2}
新宿	0.065	0.572	0.572
中野	0.128	0.920	0.999
品川	0.111	0.605	0.605
北	0.128	0.999	0.999
葛飾	0.065	0.641	0.641