数理解析研究所講究録 1373 巻 2004 年 44-49

ラフ集合によるデータクリーニングを組み込んだFuzzyDEA

奥原 浩之 †, 堂本 絵理 †, 上野 信行 †

† 広島県立大学 経営学部 経営情報学科

Koji OKUHARA†, Eri DOMOTO†, Nobuyuki UENO†

† Department of Management and Information Sciences, Hiroshima Prefectural University

1. はじめに

企業活動において得られる情報から複数の事業体 (Decision Making Unit : DMU) を分析対象として相対的な効率評価を行う手法にデータ包絡分析 (Data Envelopment Analysis : DEA)[1] がある．DEA では支出と収入の比である収支率により経営効率性を算出する有力な手段であり，支出は収入を生み出すための入力，収入はその結果としての出力と見るとき，収入／支出 (収支率の逆数) が大きいほど効率が良いとされる．また，収集された情報が含むさまざまな不正確さや不確実さを考慮するために，入出力データの曖昧さをも考慮できる DEA として，ファジィDEA[2] が考案されている．これらの手法により，DMU 間の相対的な経営効率の評価とともに，非効率な DMU に対する改善策を示すことが可能となる．

近年の情報技術の発展により，大規模かつ分散したデータをまとめて取り扱うことが実現されてきている．それゆえ，収集されたデータに目的とする分析と関連ない情報が混在する傾向も高くなる．このことは不適切なモデル作成によるあやまった分析結果を導く恐れが増すことを意味している．とくに経営情報のような大規模データに対しては人間の経験や知識による識別能力には限界があると考えられるため主観的な分類手法に頼るのは注意が必要である．そのような場合でも客観的に内包されているルールを抽出できる理論としてラフ集合 [3] による分析がある．ラフ集合では情報に内在する曖昧さや識別不能性を数学的に扱うことが可能である．

本研究では，これらを統一的に扱う手法を提案する．そこで，まず絶対値誤差を最小とする回帰分析と DEA を結合した DEARA[4] を拡張したファジィDEA[2] に着目する．DEARA は DMU を平均的な観点からと優れているものからの観点までを同一の枠組みで評価できる．さらに，量的データを質的データへ変換し序数性を考慮するラフ集合による分析法 [5] に着目する．この手法により量的データと質的データが混在する場合でも，ラフ集合によるルール抽出が区間回帰分析を適用することで可能となり，量的データと質的データを同一の枠組みで考慮することができるようになる．つまり，本研究では区間回帰分析を核として，ラフ集合によるルール抽出とファジィDEA の融合を図ることを目的とする．

2. ファジィDEA の概要

N 個の入力変数，M 個の出力変数，L 個の DMU に対する DEARA は，以下の線形計画法により定式化される [4].

$$
\begin{aligned}
\min_{\rho_k,\eta_k,\mu,\nu} \quad & E=\sum_{k=1}^{L}(a_k\rho_k+b_k\eta_k) && (1)\\
\text{s. t.} \quad & \nu^t\mathbf{x}_0=1, \\
& \mu^t\mathbf{y}_k-\nu^t\mathbf{x}_k=\rho_k-\eta_k \quad (k=1,2,3,\cdots,L), \\
& \mu\geq\mathbf{0},\ \nu\geq\mathbf{0}, \\
& \rho_k\geq 0,\ \eta_k\geq 0 \quad (k=1,2,3,\cdots,L).
\end{aligned}
$$

ここで，$\mathbf{x}_k=[x_{k1},x_{k2},\cdots,x_{kN}]^t\in\Re^N\times\Re^1$，$\mathbf{y}_k=[y_{k1},y_{k2},\cdots,y_{kM}]^t\in\Re^M\times\Re^1$ である．ρ_k と η_k はそれぞれ $\mu^t\mathbf{y}_k$ と $\nu^t\mathbf{x}_k$ の間の正の残差，負の残差を表し，a_k と b_k はそれぞれ ρ_k と η_k のウエイト係数を表す．

DEARA を最小絶対値和形の線形回帰分析とする場合は，$a_k = b_k = 1$ $(k = 1, 2, 3, \cdots, L)$ とすることで，以下の定式化と等価なものとみなせる．

$$
\begin{aligned}
\min_{\lambda_k, \mu, \nu} \quad & E = \sum_{k=1}^{L} \lambda_k \\
\text{s.t.} \quad & \nu^t \mathbf{x}_0 = 1, \\
& \|\mu^t \mathbf{y}_k - \nu^t \mathbf{x}_k\| = \lambda_k \quad (k = 1, 2, 3, \cdots, L), \\
& \mu \geq \mathbf{0}, \ \nu \geq \mathbf{0}, \\
& \lambda_k \geq 0 \quad (k = 1, 2, 3, \cdots, L).
\end{aligned} \tag{2}
$$

DEARA を Charnes, Cooper, Rhodes らによって提案された CCR 形 DEA とする場合は，$a_k \to +\infty$ $(k = 1, 2, 3, \cdots, L)$，$b_k \to 0$ $(k \neq 0)$，$b_0 = 1$ とすることで，以下の定式化と等価なものとみなせる．

$$
\begin{aligned}
\max_{\mu, \nu} \quad & E = \mu^t \mathbf{y}_0 \\
\text{s. t.} \quad & \nu^t \mathbf{x}_0 = 1, \\
& \mu^t \mathbf{y}_k \leq \nu^t \mathbf{x}_k \quad (k = 1, 2, 3, \cdots, L), \\
& \mu \geq \mathbf{0}, \ \nu \geq \mathbf{0}.
\end{aligned} \tag{3}
$$

ここで，着目する DMU の出力評価値の総和

$$\mu^t \mathbf{y}_0 = \nu^t \mathbf{x}_0 + \rho_0 - \eta_0 = 1 + \rho_0 - \eta_0 \tag{4}$$

は着目する DMU の一般化効率値といわれる．一般化効率値は入力評価値総和を 1 に正規化したもとでの出力評価値総和の相対的な値を与えるため，最良とは限らない生産関数との相対的な比較の場合には，1 より大きい一般化効率値を持つ DMU も存在しうる．DEARA はパラメータを連続的に変化させることで，入出力間に特定のモデルを想定する必要がないノンパラメトリックな手法で優れ者をベースに評価する DEA と，入出力間にモデルを規定するパラメトリックな手法で平均をベースに評価する回帰分析を包含している効率評価法である．

ところで，企業活動において得られる情報では収集された情報が含むさまざまな不正確さや不確実さを考慮する必要が生じる．入出力データの曖昧さをも考慮した効率評価ができるファジィ DEA は以下のように定式化できる．

$$
\begin{aligned}
\max_{\mu, \nu} \quad & E = \mu^t \mathbf{Y}_0 \\
\text{s. t.} \quad & \nu^t \mathbf{X}_0 \approx \tilde{1}, \\
& \mu^t \mathbf{Y}_k \lesssim \nu^t \mathbf{X}_k \quad (k = 1, 2, 3, \cdots, L), \\
& \mu \geq \mathbf{0}, \ \nu \geq \mathbf{0}.
\end{aligned} \tag{5}
$$

このような場合に対して，DEARA をファジィ入出力データも扱えるように拡張したモデルが田中ら [2] により以下のように提案されている．

$$
\begin{aligned}
\min_{\theta'_k, \theta_k, \psi'_k, \psi_k, \mu, \nu} \quad & E = \sum_{k=1}^{L} (\delta_k \theta_k + \delta'_k \psi_k + \varphi_k \theta'_k + \varphi'_k \psi'_k) \\
\text{s. t.} \quad & \nu^t \mathbf{c}_0 \geq g_0, \\
& \nu^t \mathbf{x}_0 - (1-h) \nu^t \mathbf{c}_0 = 1 - (1-h)e, \\
& \nu^t \mathbf{x}_0 + (1-h) \nu^t \mathbf{c}_0 \leq 1 + (1-h)e, \\
& \mu^t \mathbf{y}_k - (1-h) \mu^t \mathbf{d}_k - \nu^t \mathbf{x}_k + (1-h) \nu^t \mathbf{c}_k = \theta_k - \theta'_k, \\
& \mu^t \mathbf{y}_k + (1-h) \mu^t \mathbf{d}_k - \nu^t \mathbf{x}_k - (1-h) \nu^t \mathbf{c}_k = \psi_k - \psi'_k, \\
& \mu \geq \mathbf{0}, \ \nu \geq \mathbf{0}, \\
& \theta'_k \geq 0, \ \theta_k \geq 0, \ \psi'_k \geq 0, \ \psi'_k \geq 0.
\end{aligned} \tag{6}
$$

ここで，θ_k，θ'_k，ψ_k，ψ'_k $(k=1,2,3,\cdots,L)$ は h レベル集合での $\mu^t\mathbf{Y}_k$ と $\nu^t\mathbf{X}_k$ の間の正と負の左側偏差と右側偏差である．δ_k，δ'_k，φ_k，φ'_k はそれぞれ，k 番目の DMU の θ_k，θ'_k，ψ_k，ψ'_k の非負係数である．

対称な三角型ファジィ入力ベクトル $\mathbf{X}_0=(\mathbf{x}_0,\mathbf{c}_0)$ と出力ベクトル $\mathbf{Y}_0=(\mathbf{y}_0,\mathbf{d}_0)$ での DMU のファジィ効率は非対称な三角型ファジィ数は $\widetilde{\theta}=(w_l,\eta,w_r)$ で表される．

$$\eta=\frac{\mu^{*^t}\mathbf{y}_0}{\nu^{*^t}\mathbf{x}_0},\quad w_l=\eta-\frac{\mu^{*^t}(\mathbf{y}_0-\mathbf{d}_0(1-h))}{\nu^{*^t}(\mathbf{x}_0+\mathbf{c}_0(1-h))},\quad w_r=\frac{\mu^{*^t}(\mathbf{y}_0+\mathbf{d}_0(1-h))}{\nu^{*^t}(\mathbf{x}_0-\mathbf{c}_0(1-h))}-\eta \tag{7}$$

ここで，ν^* と μ^* は係数ベクトル，η，w_l，w_r はファジィ効率 $\widetilde{\theta}$ の中心と左右の幅の値である．

3. ラフ集合によるルール抽出

いま，L 個のサンプルについて N 個の条件属性と M 個の決定属性からなる決定表を考える．ここで，文献 [3] にもとづきラフ集合によるルール抽出について述べる．条件属性のうち序数性が仮定できるものを基準と呼ぶ．このとき，O_q を q 番目の基準に基づいたアウトランキング関係とする．すなわち，xO_qy は q 番目の基準に基づけば，x は少なくとも y と同程度によいことを表す．O_q は比較可能で推移的な関係，すなわち弱順序関係であるとする．さらに m 番目の決定属性において，$s>t$ に対して C^s_m の任意の要素が C^t_m のすべての要素よりも好ましいという性質をもつ U の分割 $\tau=\{C^1_m,C^2_m,\cdots,C^R_m\}$ を考える．ここでは，関係 $(x\in C^s_m,\ y\in O^t_m,\ s>t\Rightarrow xO_my$ かつ $\sim yO_mx)$ が成立している．

いま，m 番目の決定属性を R 個のクラス

$$C^s_m\cap C^t_m=\phi,(s\neq t),\quad C^R_m\succ\cdots\succ C^r_m\succ C^1_m \tag{8}$$

に分類する．このとき，$x\in U$ が与えられると，少なくともクラス C^r_m に属している U の要素の集合である上側累積集合 $C^{\geqq r}_m$ と，たかだかクラス C^r_m に属している U の要素の集合である下側累積集合 $C^{\leqq r}_m$ が，

$$C^{\geqq r}_m=\bigcup_{s\geq r}C^s_m,\quad C^{\leqq r}_m=\bigcup_{s\leq r}C^s_m \tag{9}$$

で定義できる．

すべての基準の集合を W とするとき，$V\subseteq W$ を考える．任意の $v\in V$ について，xO^v_my が成立するとき，x は V において y を支配するといい，xD^V_my で表し

$$xD^V_my\leftrightarrow g(x,n)\geq g(y,n),\quad(\forall v\in V) \tag{10}$$

と定義する．ここで，$g(x,n)$ でサンプル x の属性 n に関する属性値を表す．このとき，m 番目の決定属性において，$x\in U$ が与えられると，V において x を支配する U の要素の集合 $D^{+V}_m(x)$ と，V において x に支配される U の要素の集合 $D^{-V}_m(x)$ が

$$D^{+V}_m(x)=\{y\in U|yD^V_mx\},\quad D^{-V}_m(x)=\{y\in U|xD^V_my\} \tag{11}$$

で定義できる．

支配集合 $D^{+V}_m(x)$ による累積集合 $C^{\geqq r}_m$ の下近似集合 $V_*(C^{\geqq r}_m)$ と上近似集合 $V^*(C^{\geqq r}_m)$ は

$$V_*(C^{\geqq r}_m)=\{x\in U|D^{+V}_m(x)\subseteq C^{\geqq r}_m\},\quad V^*(C^{\geqq r}_m)=\bigcup_{x\in C^{\geqq r}_m}D^{+V}_m(x) \tag{12}$$

で定義できる．この下近似集合から，$V_*(C^{\geqq r}_m)$ に属している x を支配しているデータ x^* は必ずクラス r 以上に属しているというルールが導かれる．すなわち，ある $x^*\in C^{\geqq r}_m$ について，次のような if-then ルールが得られる．

$$\begin{aligned}&\text{IF } g(x^*,n_1)\succeq g(x,n_1)\ \textit{and}\ g(x^*,n_2)\succeq g(x,n_2)\\&\quad\cdots\ \textit{and}\ g(x^*,n_N)\succeq g(x,n_N),\ \text{THEN } x^*\in C^{\geqq r}_m.\end{aligned} \tag{13}$$

同様に支配集合 $D_m^{-V}(x)$ による累積集合 $C_m^{\leq r}$ の下近似集合 $V_*(C_m^{\leq r})$ と上近似集合 $V^*(C_m^{\leq r})$ は

$$V_*(C_m^{\leq r}) = \{x \in U | D_m^{-V}(x) \subseteq C_m^{\leq r}\},\quad V^*(C_m^{\leq r}) = \bigcup_{x \in C_m^{\leq r}} D_m^{-V}(x) \tag{14}$$

で定義できる．やはり，この下近似集合から，$V_*(C_m^{\leq r})$ に属している x を支配しているデータ x^* は必ずクラス r 以下に属しているというルールが導かれる．すなわち，ある $x^* \in C_m^{\leq r}$ について，次のような if-then ルールが得られる．

$$\begin{aligned} &\text{IF } g(x^*, n_1) \preceq g(x, n_1) \textit{ and } g(x^*, n_2) \preceq g(x, n_2) \\ &\quad \cdots \textit{ and } g(x^*, n_N) \preceq g(x, n_N),\ \text{THEN } x^* \in C_m^{\leq r}. \end{aligned} \tag{15}$$

いま，$C_m^{\geq r}$ と $C_m^{\leq r}$ について境界が

$$B_V(C_m^{\geq r}) = V^*(C_m^{\geq r}) - V_*(C_m^{\geq r}),\quad B_V(C_m^{\leq r}) = V^*(C_m^{\leq r}) - V_*(C_m^{\leq r}) \tag{16}$$

で定義できるため，$C_m^{\geq r}$ と $C_m^{\leq r}$ について近似の精度が

$$\alpha_V(C_m^{\geq r}) = \frac{|V^*(C_m^{\geq r}))|}{|V_*(C_m^{\geq r})|},\quad \alpha_V(C_m^{\leq r}) = \frac{|V^*(C_m^{\leq r})|}{|V_*(C_m^{\leq r})|} \tag{17}$$

と定義され，分割 τ が部分基準集合 V によって正しく分類できた対象の割合である近似の質は次のように定められる．

$$\beta_V(\tau) = \frac{|U - (\bigcup_{r=1}^{n} B_V(C_m^{\geq r}) \cup \bigcup_{r=1}^{n} B_V(C_m^{\leq r}))|}{|U|} \tag{18}$$

$\beta_V(\tau) = \beta_W(\tau)$ が成立する極小集合 $V \subseteq W$ を縮約と呼ぶ．縮約は複数が存在し，それらの共通集合を核と呼ぶ．縮約に帰属する属性を用いることにより，近似の質を低下させることなく，決定表を最も簡略化することができる．

条件属性が質的データである場合には，通常の属性値間の序数性を考慮したラフ集合における簡略化手法は適用できないため，ここでは，質的データに対して区間回帰分析を適用することにより，質的データに順序関係を与える方法を説明する [5]．

まず，入力変数 x_i $(i = 1, 2, 3, \cdots, N)$ をそれぞれ s_i 個のカテゴリに分類する．その結果，入力変数 x_i は $\mathbf{x}_i = [x_{i1},\ x_{i2},\ \cdots, x_{is_i}]$ で表される．サンプル s の第 i 入力変数がカテゴリ j $(\leq s_i)$ に属する場合は，x_{ij} のみが 1 となり，その他を 0 とする．このとき，次のような区間回帰モデルを考える．

$$Y_k = \sum_{i=1}^{N} \mathbf{A}_i \mathbf{X}_i \tag{19}$$

ただし，$\mathbf{A}_i$ は区間効用値ベクトル $(\mathbf{a}_i, \mathbf{w}_i)$ を表し，$\mathbf{a}_i = [a_{i1},\ a_{i2},\ \cdots,\ a_{is_i}]^{\mathrm{t}}$ は中心ベクトル，$\mathbf{w}_i = [w_{i1},\ w_{i2},\ \cdots,\ w_{is_i}]^{\mathrm{t}}$ は幅ベクトルである．

区間効用値ベクトル $\mathbf{A}_i$ は次の区間回帰分析により求められる．

$$\begin{aligned} \min_{\mathbf{a},\mathbf{w}} \quad & E = \sum_{k=1}^{L} \mathbf{w} \|\mathbf{x}_k\| \\ \text{s. t.} \quad & y_k \leq= \mathbf{a}\mathbf{x}_k + \mathbf{w}\|\mathbf{x}_k\|, \\ & y_k \geq= \mathbf{a}\mathbf{x}_k - \mathbf{w}\|\mathbf{x}_k\|, \\ & \mathbf{a} \geq \mathbf{0}, \mathbf{w} \geq \mathbf{0}. \end{aligned} \tag{20}$$

ただし，$\mathbf{x}_k = [\mathbf{x}_1,\ \mathbf{x}_2,\ \cdots,\ \mathbf{x}_N]^{\mathrm{t}}$，$\mathbf{a} = [\mathbf{a}_1,\ \mathbf{a}_2,\ \cdots,\ \mathbf{a}_N]^{\mathrm{t}}$，$\mathbf{w} = [\mathbf{w}_1,\ \mathbf{w}_2,\ \cdots,\ \mathbf{w}_N]^{\mathrm{t}}$ である．

区間効用値ベクトル $\mathbf{A}_i$ が得られると，区間値 $S[s_1, s_2]$，$T[t_1, t_2]$ が与えられたとき，属性のカテゴリ間の順序付けは

$$S[s_1, s_2] \succeq T[t_1, t_2] \leftrightarrow s_1 \geq t_1,\ s_2 \geq t_2 \tag{21}$$

で行うことができる．このとき，$g(x,n)$ が区間値であるので，x が V において y を支配することを $xD'^{V}_{m}y$ で表し，以下のように定義する．

$$
\begin{aligned}
xD'^{V}_{m}y &\leftrightarrow g(x,n) \succeq g(y,n), \quad (\forall v \in V) \\
&\leftrightarrow \min_{s_1 \in g(x,n)} s_1 \geq \min_{t_1 \in g(y,n)} t_1 \; and \; \max_{s_2 \in g(x,n)} s_2 \geq \max_{t_2 \in g(y,n)} t_2
\end{aligned} \tag{22}
$$

さらに，$V_*(C^{\geq r}_{m})$ に関する下近似集合から得られたルールと $V_*(C^{\leq r}_{m})$ に関する下近似集合から得られたルールの条件部において，すべての属性においてカテゴリが同じとなる集合から，確実にある $x^* \in U$ がクラス r に属するという次の if-then ルールが得られる．

$$
\begin{aligned}
&\text{IF } g(x^*,n_1) = g(x,n_1) \; and \; g(x^*,n_2) = g(x,n_2) \\
&\quad \cdots \; and \; g(x^*,n_N) = g(x,n_N), \text{ THEN } x^* \in C^{r}_{m}.
\end{aligned}
$$

この抽出されたルールを満たすデータは確実にクラス r に属していることとなる．

4. ラフ集合によるデータの前処理を組み込んだファジィDEA

適切な経営効率の評価のためには，得られたデータすべてをそのまま活用する分析法ではなく，データ自体がもつ説明力を考慮した分析法が望ましい．そのためには分析の目的に応じて，活用する入力変数やサンプルの選択を行なうことが必要である．さらには，量的データと質的データが混在するデータも同時に取り扱いながら分析することも求められる．そこで，区間回帰分析を核として，ラフ集合によるルール抽出とファジィDEA の融合を図る．いま，5つのDMU について，入力として3つの量的データと2つの質的データ，出力として2つの量的データが得られている場合で説明する（表1参照）．

表1　経営情報データの例

	x_1	x_2	x_3	x_4	x_5	y_1	y_2
A	133	24	573	1	0	85	2454
B	232	44	738	0	0	46	3423
C	82	72	472	1	1	63	651
D	210	45	672	0	1	26	1853
E	70	33	426	1	0	13	2564

ここで，例えば x_1，y_2 を最大値と最小値から4つの領域に分割することを考える．x_1 の場合は $x_{11} \in (70,27)$，$x_{12} \in (124,27)$，$x_{13} \in (178,27)$，$x_{14} \in (232,27)$ として0，1変数へ変換される．ただし，表記は(中心，幅)である．同様に x_2，y_1 は3つの領域，x_3 は2つの領域に分割することで0，1変数へ変換する．

このとき，以下のファジィ回帰分析を行なうことを考える．

$$
\min_{\mathbf{w},\mu,\nu} \quad E = \sum_{k=1}^{L} \{\mathbf{w} + \mathbf{w}'\} \|\mathbf{x}_k\| \tag{23}
$$

$$
\begin{aligned}
\text{s. t.} \quad & \nu^t \mathbf{x}_0 = 1, \\
& \|\mu^t \mathbf{y}_k - (1-h)\mu^t \mathbf{d}_k - \nu^t \mathbf{x}_k\| \leq (1-h)\mathbf{w}\|\mathbf{x}_k\|, \\
& \|\mu^t \mathbf{y}_k + (1-h)\mu^t \mathbf{d}_k - \nu^t \mathbf{x}_k\| \leq (1-h)\mathbf{w}'\|\mathbf{x}_k\|, \\
& \mathbf{w} \geq \mathbf{0}, \; \mathbf{w}' \geq \mathbf{0}, \; \mu \geq \mathbf{0}, \; \nu \geq \mathbf{0}.
\end{aligned}
$$

ここで，$\mathbf{x}_k$ は0，1変数に変換された入力データであり，$\mathbf{y}_k$，$\mathbf{d}_k$ はファジィ化された出力データ $\mathbf{y}$ の中心と幅である．パラメータ ν を $\mathbf{1} = [1,1,1,\cdots,1]^t \in \Re^{\{4+3+2+1+1\}} \times \Re^1$ と考えることで，x_{ij} を区間 $(\nu_{ij}, w_{ij} + w'_{ij})$ とみなすことができ，入出力データの中心と幅を得る．このとき，条件属性が質的データである場合に属性値間の序数性を考慮したラフ集合による簡略化手法を適用することでルールを得ることができ，ルールにおいて考慮されていない入力変数とルールで説明できないサンプルを除いた上で経営効率の評価を行うのが適切であることがわかる．ここで，ルールにおいて考慮されていない入力変数とは，$\tilde{x}_{ij} \approx 0$，$(j = 1,2,3,\cdots,s_i)$ となる入力変数 x_i である．ルールで説明できないサンプルとは，最大の近似の質をとる縮約（つまり，全入力変数を用いた場合と同程度の説明力を持つ縮約）で確実ルールが適用できないサンプルのことである．

表2　入力データの0, 1変数と出力データのファジイ化

	x_{11}	x_{12}	x_{13}	x_{14}	x_{21}	x_{22}	x_{23}	x_{31}	x_{32}	x_{41}	x_{51}	y_1	y_2
A	0	1	0	0	1	0	0	1	0	1	0	(85, 18)	(2499, 462)
B	0	0	0	1	0	1	0	0	1	0	0	(49, 18)	(3423, 462)
C	1	0	0	0	0	0	1	1	0	1	1	(49, 18)	(651, 462)
D	0	0	0	1	0	1	0	0	1	0	1	(13, 18)	(1575, 462)
E	1	0	0	0	1	0	0	1	0	1	0	(13, 18)	(2499, 462)

いま，$\xi=\mu^t\mathbf{y}_k-(1-h)\mu^t\mathbf{d}_k-\nu^t\mathbf{x}_k$，$\xi'=\mu^t\mathbf{y}_k+(1-h)\mu^t\mathbf{d}_k-\nu^t\mathbf{x}_k$ として，θ_k，θ'_k，ψ_k，ψ'_k $(k=1,2,3,\cdots,L)$ を

$$\theta_k=\frac{\|\xi\|+\xi}{2},\ \theta'_k=\frac{\|\xi\|-\xi}{2},\ \psi_k=\frac{\|\xi'\|+\xi'}{2},\ \psi'_k=\frac{\|\xi'\|-\xi'}{2}. \tag{24}$$

で与えると

$$\theta_k+\theta'_k=\|\xi\|\le(1-\dot{h})\mathbf{w}\|\mathbf{x}_k\|,\qquad \theta_k-\theta'_k=\xi, \tag{25}$$

$$\psi_k+\psi'_k=\|\xi'\|\le(1-h)\mathbf{w}'\|\mathbf{x}_k\|,\qquad \psi_k-\psi'_k=\xi'. \tag{26}$$

であることから，ルールにおいて考慮されている入力変数 $\mathbf{x}'_k$ とルールで説明できる L' 個のサンプルを用いて，以下のファジィDEA を解くことにより経営効率の評価が行なえる．

$$\min_{\theta'_k,\theta_k,\psi'_k,\psi_k,\mu,\nu}\ E=\sum_{k=1}^{L'}(\delta_k\theta_k+\delta'_k\psi_k+\varphi_k\theta'_k+\varphi'_k\psi'_k) \tag{27}$$

$$\begin{aligned}\text{s. t.}\qquad &\nu^t\mathbf{x}'_0=1,\\ &\mu^t\mathbf{y}_k-(1-h)\mu^t\mathbf{d}_k-\nu^t\mathbf{x}'_k=\theta_k-\theta'_k,\\ &\mu^t\mathbf{y}_k+(1-h)\mu^t\mathbf{d}_k-\nu^t\mathbf{x}'_k=\psi_k-\psi'_k,\\ &\mu\ge\mathbf{0},\ \nu\ge\mathbf{0},\\ &\theta'_k\ge0,\ \theta_k\ge0,\ \psi'_k\ge0,\ \psi'_k\ge0.\end{aligned}$$

ただし，$\mathbf{x}'_k$ は ν で構成される $\mathbf{x}'$ の中心で与えられる．$\delta_k\to+\infty$ $\delta'_k\to+\infty$ $\varphi'_k\to+\infty$ $(k=1,2,3,\cdots,L)$，$\varphi_k\to0$ $(k\neq0)$，$\varphi_0=1$ である．

5. おわりに

本研究では，従来のDMU における経営効率を評価する研究で量的データのみが主に入力として用いられているのに対して，質的データで表される情報も考慮した経営効率について考察した．また，経営効率の評価をより妥当なものとするために，収集されたデータから分析の目的に応じて同一のモデルで扱うべきでない標本の選別といったデータクリーニングや，分析に必要な入出力変数の選択を行なった．

これらを統一的に扱う手法として，絶対値誤差を最小とする回帰分析とDEA を結合したDEARA を拡張したファジィDEA に着目し，量的データを質的データへ変換し序数性を考慮するラフ集合による分析法に着目した．提案する手法により区間回帰分析を核として，ラフ集合によるルール抽出を前処理としてファジィDEA との融合を図ることができる可能性を示した．

参考文献

[1] 刀根薫, “経営効率制の測定と改善 -包絡分析法ＤＥＡによる-” 日科技連出版社, 1993.

[2] P. Guo andH. Tanaka, “Fuzzy DEA: a perceptual evaluation method,” *Fuzzy Sets and Systems*, Vol. 119, pp. 149-160, 2001.

[3] 日本ファジィ学会編, “ファジィとソフトコンピューティングハンドブック,” 共立出版株式会社, 2000.

[4] 篠原正明, “包絡分析と回帰分析を含む性能評価法 DEARA,” オペレーションズ・リサーチ, Vol. 12, pp. 691-695, 1995.

[5] 杉原一臣, 石井博昭, 田中英夫, “ラフ集合による新しいコンジョイント分析の提案,” 日本知能情報ファジィ学会誌, **15**, No. 4, pp. 421-428, 2003.