1995年度日本オペレーションズ・リサーチ学会
春季研究発表会

主成分分析を用いるＤＥＡに関する一考察

01001600 NTT研究所 上田 徹 UEDA Tohru

1．まえがき

主成分分析を用いるＤＥＡについては既に[1],[2]で報告されている。そこでは変数の基準化，分数計画法での分母の正化などが論じられた。それに関連する議論の中で

（1）主成分を求めるための原入力の重みが負のものの扱い

（2）主成分の数を絞ったとき（例えば寄与率80%としたとき）の残りの成分に対する考え方

を明らかにする必要性を感じた。本資料は上記2点に絞って考え方を述べる。

2．負の重みを持つ入力の扱い

第k主成分$\mathbf{Z}_k^t=(Z_{k1},Z_{k2},\ldots,Z_{kn})$は

$$Z_{kj}=\sum_{i=1}^{m}\omega_{ki}x_{ji}\ ;\ \mathbf{w}_k^t=(\omega_{k1},\omega_{k2},\ldots,\omega_{km})$$

で与えられるものとし，重みベクトル$\mathbf{w}_k$の要素の中で非負のものに対応する入力番号の集合をP_k，負の入力番号の集合をN_kとする。

$$P_k=\{h\mid\omega_{kh}\geqq 0\}$$
$$N_k=\{h\mid\omega_{kh}<0\}$$

第k主成分を、P_kに属する入力から作られる$Z_{kj}^{(P)}$と、N_kに属する入力から作られる$Z_{kj}^{(N)}$とに分割し、

$$Z_{kj}^{(P)}=\sum_{h\in P_k}\omega_{kh}x_{jh}$$

$$Z_{kj}^{(N)}=\sum_{h\in N_k}|\omega_{kh}|x_{jh}$$

として，これらを別々の入力と考える方法が刀根先生より示唆された。その方法では重みベクトルの符号の決め方に悩まなくてすむ。

しかし，その方法は次の点で問題がある。

（1）P_kとN_kの要素数には著しくアンバランスな場合がある。

（2）$\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m\}$で表現される空間から$\{\mathbf{Z}_1,\mathbf{Z}_2,\ldots,\mathbf{Z}_p\}$ $(p<<m)$にした効果が薄まる。

（3）$Var(\mathbf{Z}_k)=\lambda_k$であることなど$\mathbf{Z}_k$として扱ったときの主成分分析に関わる議論を再構築しなければならない。

（4）$\mathbf{Z}_k$に比べて$\mathbf{Z}_{k+1}$の分散が極度に小さい場合にも$\mathbf{Z}_{k+1}$を二分割して扱うため結果としてむしろ$\mathbf{Z}_{k+1}$をより重視したことになってしまう。（これについては領域限定法の範囲に分散情報を用いることが考えられる。しかし、DEAの通常入力の扱いについても同様の措置を行うべきかどうか議論の余地がある。）

そこで，主成分の扱いは通常と変えず，重みベクトルの符号の決め方を考える。

主成分のうち最も重視したいのは第１主成分なので、入力（出力）の第２主成分次降の各主成分の符号は出力（入力）の第１主成分との相関が正となるように決める。

3．主成分の数を絞ったときの考え方

例えば累積寄与率rが80%以上となるように主成分の数を$\{\mathbf{Z}_1,\mathbf{Z}_2,\ldots,\mathbf{Z}_p\}$の$p$個に絞ったとき，$(1-r)\times 100\%$の変動は無視されたことになる。そこで$(1-r)\times 100\%$の変動を取り込むために$(p+1)$番目の入力として$\mathbf{Z}_{p+1}$を考える。$\mathbf{Z}_{p+1}$の設定法として平均と分散情報のみを用いる場合と,$m$次元空間内の点から主成分で張られる空間への隔たりとする場合とについて論じる。

3.1 平均と分散情報のみを用いる場合

$\mathbf{Z}_{p+1}$は平均μ_{p+1}，V_{p+1}を持つ確率変数とする。もとの入力変数を$\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m$とするとき，主成分を求めるための変換として

$$\tilde{\mathbf{x}}_i=\frac{\mathbf{x}_i}{S_i},\quad S_i:\text{変数}\mathbf{x}_i\text{の標準偏差}$$

$$\tilde{\mathbf{x}}_i^t=(\tilde{x}_{1i},\tilde{x}_{2i},\ldots,\tilde{x}_{ni})\ ;\ \mathbf{x}_i^t=(x_{1i},x_{2i},\ldots,x_{ni})$$

を使うことを提案した[1],[2]。$\tilde{\mathbf{x}}_i(i=1,2,\ldots,p)$の分散・共分散行列の固有値を$\lambda_j$とすると

$$V_{p+1}=m-\sum_{j=1}^{p}\lambda_j$$

である。ここで第k主成分$\mathbf{Z}_k=(Z_{k1},\ldots,Z_{kn})$は

$$Z_{kj}=\sum_{i=1}^{m}\omega_{ki}\tilde{x}_{ji}$$

で与えられているので，$\mathbf{Z}_k$の平均μ_kは

$$\mu_k = \frac{\sum_{j=1}^{n} Z_{kj}}{n} = \sum_i \omega_{ki} \frac{m_i}{S_i}$$

m_i :変数x_iの平均

で与えられる。$\tilde{x}_i$ の平均の総和$\sum_i \left(\frac{m_i}{S_i} \right)$から第$p$主成分までで表現された量を引いたものを μ_{p+1}と考えると

$$\mu_{p+1} = \sum_{i=1}^{m} (1 - \sum_{k=1}^{p} \omega_{ki}) \frac{m_i}{S_i}$$

となる。このようにして入力，出力ともにDMU jに依存しない確率変数を加えることが考えられる。そのときには，DMU Jにとっては

$$\left(\sum_{r=1}^{q} u_r \eta_{rJ} + u_{q+1} \eta_{q+1} \right) / \left(\sum_{k=1}^{p} v_k Z_{kJ} + v_{p+1} Z_{p+1} \right)$$

ただし，η_{rj} ：出力の第r主成分の第j成分

の最大化を指向すればよいことになる。

しかし、分母、分子の双方に自由に決定できる係数があると、

$u_r, v_k \to 0 \ (r = 1,2,\dots,q \ ; \ k = 1,2,\dots,p)$

$u_{q+1}\eta_{q+1} = v_{p+1} z_{p+1}$

とすれば、どのDMUも近似的に効率値１にできること、入力については分母を正値にするために元々定数を加えているためその中に含まれてしまうことの２点から結局、[2]で提案したモデルを越えるには至らなかった。

3.2 主成分空間との隔たりを用いる場合

p個の主成分で表現されるp次元ベクトル空間V^pへのm次元ベクトル空間の点

$\mathrm{A}: (x_{A1}, x_{A2}, \dots, x_{Am})$

から下ろした垂線の足Hのm次元ベクトル空間での座標を

$\mathbf{x}^{(H)} = (x_{H1}, x_{H2}, \dots, x_{Hm})'$

とする（図１参照）。点Aの主成分はm次元までとると

$(Z_{1A}, Z_{2A}, \dots, Z_{mA})$

であるが、点Hでは(p+1)次元目以降は無視するので

$\mathbf{Z}^{(H)} = (Z_{1H}, Z_{2H}, \dots, Z_{pH}, 0, \dots, 0)'$

である。ここで

$(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_p)' = W(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)'$

$W = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p)'$

なので

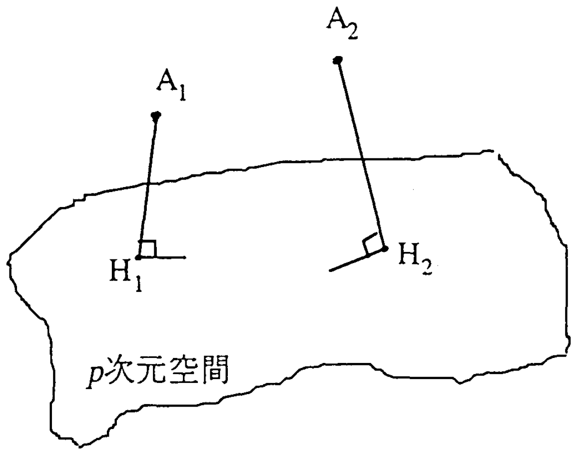

図１　m次元ベクトル空間の点A_iからp次元ベクトル空間への垂線の足H_i

$$\mathbf{x}^{(H)} = W'\mathbf{Z}^{(H)} \tag{1}$$

である。入力の時、

$x_{Ai} > x_{Hi}$であれば$(x_{Ai} - x_{Hi})$を減らしたいし、

$x_{Ai} < x_{Hi}$であれば$(x_{Ai} - x_{Hi})$を増やしたい。

そこで、分数計画における分母

$$\sum_{k=1}^{p} v_k Z_{kj} + c_J$$

の代わりに

$$\sum_{k=1}^{p} v_k Z_{kj} + v_{p+1} \sum_{i=1}^{m} (x_{A_j i} - x_{H_j i}) + c_J \ ; \ v_{p+1} > 0$$

や、$(x_{A_j i} - x_{H_j i})$の符号により別の係数を用いることなどが考えられる。ただし、方向性を考えるためには必ず非負値を取るユークリッドの距離（垂線の脚の長さ）などは使えないと考えている。

この方法を取れば式(1)による目標主成分ベクトルの逆変換が点Hしか表現していないのを変数$(x_{A_j i} - x_{H_j i})$がカバーしてくれることになる。

４．むすび

入力（出力）の主成分の符号は出力（入力）第１主成分との相関係数を見て決定すること、p個の主成分で説明できていない部分を定量化する方法を提案した。実例を用いてさらに詳細に手法の有効性を検討する必要がある。

[1]星合：「DEA法における入出力値に関する一検討」1993年OR学会秋季大会

[2]上田：「多変量解析法を用いたDEA法入出力変換法の検討」1994年OR学会秋季大会