欠測値を含むデータからの局所的な主成分の抽出法*
(Extraction of Local Principal Components from Data with Missing Values)

本多　克宏†・神田　章裕†・市橋　秀友†・山川あす香‡
†大阪府立大学　大学院　工学研究科
Graduate School of Engineering,
Osaka Prefecture University;
1-1 Gakuen-cho, Sakai city, Osaka 599-8531, JAPAN
‡福島学院短期大学　情報ビジネス科
Information and Business Communication,
Fukushima College;
1-1 Chigoike Miyashiro, Fukushima 960-0181, JAPAN

概 要

大規模なデータベースからの知識発見の手法として，ファジィクラスタリングと多変量解析法を組み合わせることにより局所的な特徴量を抽出する研究が近年盛んに行われているが，実データを扱う際にはしばしば欠測値の処理方法が問題となる．本論文では，観測値のみを用いる主成分分析法にFuzzy c-Means 法を融合することにより，欠測値を含むデータから局所的な主成分を抽出する手法を提案する．ファジィ散布行列の固有値問題に帰着される Fuzzy c-Varieties 法が分散共分散行列を用いた主成分分析とクラスタリングの同時適用法ととらえられるのに対して，相関係数行列の固有値問題に帰着される提案法は相関係数行列を用いた主成分分析とクラスタリングの同時適用法であるといえる．

(In many real world applications data sets with missing values are quite common. In this paper, we propose a new approach which extracts local principal components for the feature extraction from a large scale database. The new method is based on a simultaneous approach to principal component analysis and fuzzy clustering with an incomplete data set including missing values. The simultaneous approach extracts local principal components by using the eigenvectors of the correlation coefficient matrix, while Fuzzy c-Varieties(FCV) proposed by Bezdek *et al.* uses the eigenvectors of the fuzzy scatter matrix.)

keywords: fuzzy clustering, principal component analysis, missing value, eigen value problem.

1　はじめに

近年，大規模なデータベースからの知識発見の手法として，ファジィクラスタリングと多変量解析法を組み合わせることにより局所的な特徴量を抽出する研究が盛んに行われている．Bezdek らによる Fuzzy c-Varieties (FCV) 法および Fuzzy c-Elliptotypes (FCE) 法 [1, 2] は，局所的な主成分ベクトルによりいくつかの線形多様体状あるいは楕円体状のクラスタに分類する手法であり，クラスタ分析と主成分分析の同時適用法であるといえる．また，その改良として，クラスタごとの分散共分散行列の固有値をトレードオフパラメータに採用した Adaptive Fuzzy c-Elliptotypes (AFC) 法 [3] や，次元の異なる多様体の発見法 [4, 5] なども提案されている．その他，Bezdek らの Fuzzy c-Means

*システム制御情報学会論文誌, 15, 12, 663-672 (2002)

（FCM）法 [1] の目的関数とその他の多変量解析手法の目的関数を組み合わせることにより，様々な局所的特徴量を抽出する試みもなされている [6, 7].

しかし，実データを分析する際には，データの中に観測されなかった部分，すなわち欠測値を含み，直接，これらの手法を適用できないことがある．多変量解析法の適用において，このような不完全データを処理する方法がいくつか考えられる [8]．最も簡単な手法として，欠測値を含むデータを削除した後に分析するということが考えられるが，削除することで多くの情報を切り捨ててしまうという問題点がある．また，欠測値そのものを何らかの方法によって推定し，データの欠落部分を補完してから分析する手法なども考えられるが，補完の際に用いられるモデルの妥当性などの問題がある．そこで，モデルの仮定を設けず，欠測部分の補完という手続きを避けながら，なおかつデータ行列のすべての観測値を直接利用して，多変量データの分析を行うことが望まれる．

欠測値を含む多変量データに対する主成分分析法として，Ruhe [9] や Wiberg [10]，柴山 [11] は，データ行列の最小 2 乗近似において観測値に対応する要素については誤差が小さくなるようにするものの，欠測値に対応する要素については変化するに任せることにより，多変量データの主たる変動成分を抽出する方法を提案している．また，高根 [12] は柴山ら [13] の欠測値がある場合の線形等化法を多次元に拡張することにより，標準化したデータの主成分を抽出する手法を提案している．

一方，ファジィクラスタリングの適用においても，同様に欠測値の処理法が問題となる．宮本ら [14] や Timm ら [15] は，重み付き平均を欠測値の推定値として用いたり，欠測値を無視したりすることにより，FCM 法において欠測値を処理する手法を提案している．また，本多ら [16] は柴山 [11] の最小 2 乗近似に基づく主成分分析法に宮本らの手法を融合することにより FCV 法において欠測値を処理する手法を提案し，協調フィルタリングへの応用を行っている [17] が，データ行列と同じ要素数の近似行列を繰り返しごとに求める必要があり，データ数が増えるにつれてメモリの所要量や計算量が増加するという欠点がある．

そこで本論文では，高根の欠測値を含むデータの主成分分析法にファジィクラスタリングで用いるメンバシップを導入することにより，欠測値を含むデータに対応できるクラスタリング法を提案する．提案法はラグランジュ乗数法を FCM 法と高根による主成分分析法の同時分析に応用したもので，K-L 情報量正則化による FCM クラスタリング [18] により複数のクラスタに分類して分析する．FCV 法が分散共分散行列を用いた主成分分析とファジィクラスタリングの同時適用法とみなされるのに対して，提案法は標準化されたデータによる相関係数行列を用いた主成分分析にクラスタリングを融合した手法である．

数値例では，人工データやスーパーマーケットで収集された POS データを用いて，提案法の特徴を検証する．

2　欠測値を含むデータの主成分分析とクラスタリングの同時適用法

n 次元の N 個の標本データからなる（$N \times n$）データ行列 $X = (x_{ij})$ が与えられたときに，N 個の標本データを C 個のクラスタに分割しながら，クラスタごとに局所的な主成分を抽出する問題を考える．

Bezdek らの Fuzzy c-Elliptotypes（FCE）法 [1, 2] は，線形のクラスタを抽出するためにクラスタのプロトタイプとして線形多様体を用いる Fuzzy c-Varieties（FCV）法の目的関数と，Fuzzy c-Means（FCM）法の目的関数を重み付きで足し合わせることにより楕円体状のクラスタを得る手法であり，以下の目的関数の最小化問題に定式化される．

$$
\begin{aligned}
L \;=\; & \sum_{i=1}^{N}\sum_{c=1}^{C} {u_{ci}}^{m}\Big((\boldsymbol{x}_i-\boldsymbol{b}_c)^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{b}_c) \\
& -\alpha\sum_{k=1}^{t}\boldsymbol{v}_{ck}^{\mathrm{T}}B_{ci}\boldsymbol{v}_{ck}\Big) \qquad (1)
\end{aligned}
$$

ただし，$\boldsymbol{x}_i = (x_{i1}, \cdots, x_{in})^{\mathrm{T}}$ は n 次元の第 i 標本データを表す．u_{ci} は第 i データがクラスタ c に属する度合いを表すメンバシップで，

$$
\boldsymbol{u}_c = \Big\{(u_{ci}) | \sum_{c=1}^{C} u_{ci} = 1,\ u_{ci} \in [0,1]\Big\}
$$

を満たすものとする．$\boldsymbol{b}_c = (b_{c1}, \cdots, b_{cn})^{\mathrm{T}}$ はクラスタ c の中心である．また，$\boldsymbol{v}_{ck}$ はクラスタ c のプロトタイプを表す線形多様体を張るベクトルであり，t は線形多様体の次元を表す．

ここで，

$$B_{ci} = (\boldsymbol{x}_i - \boldsymbol{b}_c)(\boldsymbol{x}_i - \boldsymbol{b}_c)^{\mathrm{T}} \tag{2}$$

であり，$\boldsymbol{v}_{ck}^{\mathrm{T}} B_{ci} \boldsymbol{v}_{ck}$ は，

$$\begin{aligned} \boldsymbol{v}_{ck}^{\mathrm{T}} B_{ci} \boldsymbol{v}_{ck} &= \boldsymbol{v}_{ck}^{\mathrm{T}} (\boldsymbol{x}_i - \boldsymbol{b}_c)(\boldsymbol{x}_i - \boldsymbol{b}_c)^{\mathrm{T}} \boldsymbol{v}_{ck} \\ &= |(\boldsymbol{x}_i - \boldsymbol{b}_c)^{\mathrm{T}} \boldsymbol{v}_{ck}|^2 \end{aligned} \tag{3}$$

のように表すことができ，$\boldsymbol{v}_{ck}$ により定まる射影軸に射影されたデータとクラスタ中心（平均）との 2 乗距離となる．$\boldsymbol{v}_{ck}$ を主成分ベクトルと考えると $\boldsymbol{v}_{ck}^{\mathrm{T}} B_{ci} \boldsymbol{v}_{ck}$ はクラスタごとにメンバシップを考慮した主成分分析のための目的関数であるととらえられるので，最大化することにより局所的な主成分を抽出できる．α は主成分分析に対する重み係数で，L は $\alpha = 0$ のときは FCM 法の目的関数に，$\alpha = 1$ のときは線形多様体とデータ点との距離の総和を表す FCV 法の目的関数に等しくなる．このように，FCV 法および FCE 法の目的関数は，クラスタごとのメンバシップを考慮した主成分分析の目的関数に FCM 法の目的関数を重み付きで足し合わせることにより，データの局所的な構造を考慮しながら主成分を抽出する手法であると考えられることから，多変量解析手法と FCM 法の目的関数を重み付きで足し合わせることにより，大規模なデータベースから部分構造を考慮した局所的な特徴量を抽出する研究が行われている [6, 7].

しかし，データ行列に欠測値が含まれる場合には，FCE 法の目的関数は定義できない．そこで本章では，高根 [12] による欠測値を含む場合の主成分分析法の目的関数に，FCM 法の目的関数を融合することにより，欠測値を含むデータから局所的な主成分を抽出する手法を提案する．

2.1 欠測値を含むデータからの局所的な主成分の抽出法

高根は柴山ら [13] の欠測値が含まれる場合の線形等化法を多次元に拡張することにより，欠測値を含むデータから主成分を抽出する手法を提案している．この手法は標準化されたデータの主成分分析にしか適用できないが，解が解析的に求まるという特長を持っている．本節では，高根の手法をメンバシップが与えられた場合に拡張し，クラスタごとにメンバシップを考慮しながら局所的な主成分を抽出する方法を提案する．

欠測値を含む（$N \times n$）データ行列 $X = (x_{ij})$ の i 行目の要素を対角要素とする対角行列を D_{xi} で表し，クラスタ $(c = 1, \cdots, C)$ ごとに次のモデルを定義する．

$$Y_{ci} = D_{xi} V_c + V_{0c} \quad (i = 1, \cdots, N) \tag{4}$$

ただし，求める主成分ベクトルの数を t とし，Y_{ci} は（$n \times t$）行列，V_c ， V_{0c} は（$n \times t$）の重み行列であり，V_c は次の制約条件を満たすものとする．

$$V_c^{\mathrm{T}} S_c V_c = I \tag{5}$$

ここで，S_c は観測された値のみから計算されたクラスタ c における各変量の変動を対角要素とする対角行列で，x_{ij} が観測されていれば 1，欠測値であれば 0 をとる 2 値変量 d_{ij} を用いると， j 番目の対角要素 s_{cj} は，

$$s_{cj} = \sum_{i=1}^{N} d_{ij} u_{ci} (x_{ij} - b_{cj})^2$$

と表される．クラスタごとの局所的な主成分分析のための目的関数を次のように定義し，その最小化問題を考える．

$$\begin{aligned} J_c &= \sum_{i=1}^{N} u_{ci} \mathrm{tr}\Big((Y_{ci} - \mathbf{1}_n \boldsymbol{w}_{ci}^{\mathrm{T}})^{\mathrm{T}} D_{wi} \\ &\quad \times (Y_{ci} - \mathbf{1}_n \boldsymbol{w}_{ci}^{\mathrm{T}}) \Big) \end{aligned} \tag{6}$$

ただし，tr は行列の対角要素の和（トレース）を表す．$\mathbf{1}_n$ は要素がすべて 1 の n 次元ベクトル，D_{wi} は第 j 対角要素が d_{ij} である対角行列とし，$\boldsymbol{w}_{ci}$ はデータごとに与えられる t 次元ベクトルとする．$t = 1$ のときは，(6) 式の最小化は (4) 式で表されるベクトル Y_{ci} の要素をすべて w_{ci} に等しくすることを表すので，クラスタごとにデータを線形等化することに相当する．

まず，V_c，V_{0c} を暫時既知として (6) 式を $\boldsymbol{w}_{ci}$ についてのみ最小化すると，求める $\hat{\boldsymbol{w}}_{ci}^{\mathrm{T}}$ は，

$$\hat{\boldsymbol{w}}_{ci}^{\mathrm{T}} = \mathbf{1}_n^{\mathrm{T}} D_{wi} Y_{ci} / n_i \tag{7}$$

$$n_i = \mathbf{1}_n^{\mathrm{T}} D_{wi} \mathbf{1}_n \tag{8}$$

となる．したがって $\boldsymbol{w}_{ci}$ について最小化した (6) 式は，

$$J_c^*(V_c, V_{0c}) = \sum_{i=1}^{N} u_{ci} \mathrm{tr}(Y_{ci}^{\mathrm{T}} C_i Y_{ci}) \tag{9}$$

と表せる．ただし

$$C_i = (Q_{n/Dwi})^{\mathrm{T}} D_{wi} Q_{n/Dwi} \tag{10}$$

$$Q_{n/Dwi} = I - \mathbf{1}_n \mathbf{1}_n^{\mathrm{T}} D_{wi} / n_i \tag{11}$$

とおいた．ここで，(4) 式を (9) 式に代入すると，

$$\begin{aligned} J_c^*(V_c, V_{0c}) &= \mathrm{tr}(V_c^{\mathrm{T}} A_{c1} V_c) + 2\mathrm{tr}(V_c^{\mathrm{T}} A_{c2} V_{0c}) \\ &\quad + \mathrm{tr}(V_{0c}^{\mathrm{T}} A_{c3} V_{0c}) \end{aligned} \tag{12}$$

と書き換えることができる．ただし，

$$A_{c1} = \sum_{i=1}^{N} u_{ci} D_{xi} C_i D_{xi} \tag{13}$$

$$A_{c2} = \sum_{i=1}^{N} u_{ci} D_{xi} C_i \tag{14}$$

$$A_{c3} = \sum_{i=1}^{N} u_{ci} C_i \tag{15}$$

とする．さらに，$J_c^*(V_c, V_{0c})$ を V_{0c} に関して最小化する．$\partial J_c^* / \partial V_{0c} = O$ から，次の関係が得られる．

$$A_{c3} V_{0c} = -A_{c2}^{\mathrm{T}} V_c \tag{16}$$

これを解いて最適解 $\hat{V}_{0c}$ を求めればよい．しかし，データ行列に欠測値が含まれる場合には，A_{c3} がランク落ちした行列となるため，通常の逆行列は存在しない．そのため， A_{c3} のムーア・ペンローズ逆行列 A_{c3}^{+} を使って (16) 式から $\hat{V}_{c0}$ を求めると，

$$\hat{V}_{0c} = -A_{c3}^{+} A_{c2}^{\mathrm{T}} V_c \tag{17}$$

となる．よって，V_{0c} について最小化した (12) 式は次のようになる．

$$\begin{aligned} J_c^{**}(V_c) &= \mathrm{tr}\Big(V_c^{\mathrm{T}} (A_{c1} - A_{c2} A_{c3}^{+} A_{c2}^{\mathrm{T}}) V_c\Big) \\ &= \mathrm{tr}(V_c^{\mathrm{T}} A_c V_c) \end{aligned} \tag{18}$$

ただし $A_c = A_{c1} - A_{c2} A_{c3}^{+} A_{c2}^{\mathrm{T}}$ とおいた．ここで，

$$V_c^{\mathrm{T}} S_c V_c = V_c^{\mathrm{T}} (S_c^{1/2})^{\mathrm{T}} (S_c^{1/2}) V_c = I \tag{19}$$

と書き換えられるので $\tilde{V}_c = S_c^{1/2} V_c$ とおくと制約条件は，

$$\tilde{V}_c^{\mathrm{T}} \tilde{V}_c = I \tag{20}$$

となり， (18) 式は，

$$J_c^{**}(\tilde{V}_c) = \mathrm{tr}\Big(\tilde{V}_c^{\mathrm{T}} (S_c^{-1/2})^{\mathrm{T}} A_c (S_c^{-1/2}) \tilde{V}_c\Big) \tag{21}$$

と表すことができ，(20) 式の制約条件のもとで J_c^{**} を $\tilde{V}_c$ について最小化すると，次の一般化固有方程式が導かれる．

$$S_c^{-1/2} A_c S_c^{-1/2} \tilde{V}_c = \tilde{V}_c \Delta_c \tag{22}$$

ただし，Δ_c は固有値を対角要素に持つ対角行列である．(22) 式の固有値問題を解くことにより，J_c を最小とする V_c を求めることができる．

ここで，(22) 式の固有値問題と主成分分析との関係を調べるために，データ行列に欠測値が含まれない場合について考察する．いま，クラスタ c において 1 本の主成分ベクトル $\boldsymbol{v}_c = (v_{c1}, \cdots, v_{cn})$ を推定するとし，

$$y_{cij} = v_{cj}x_{ij} + v_{0cj} \ \ ; i = 1, \cdots, N, j = 1, \cdots, n \tag{23}$$

なるモデルを仮定すると，目的関数 J_c は，

$$J_c = \sum_{i=1}^{N}\sum_{j=1}^{n} u_{ci}(y_{cij} - w_{ci})^2 \tag{24}$$

と定義され，最適性の必要条件から，

$$\hat{w}_{ci} = \frac{1}{n}\sum_{j=1}^{n} y_{cij} = \bar{y}_{ci} \tag{25}$$

$$\hat{\boldsymbol{v}}_{0c} = -\frac{\displaystyle\sum_{i=1}^{N} u_{ci}D_{xi}}{\displaystyle\sum_{i=1}^{N} u_{ci}}\boldsymbol{v}_c \tag{26}$$

となり，

$$\begin{aligned} J_c^{**}(\boldsymbol{v}_c) \quad = \quad & \boldsymbol{v}_c^{\mathrm{T}}\Bigg(\sum_{i=1}^{N} u_{ci}D_{xi}C_iD_{xi} \\ & -\frac{\Big(\displaystyle\sum_{i=1}^{N} u_{ci}D_{xi}\Big)C_i\Big(\displaystyle\sum_{i=1}^{N} u_{ci}D_{xi}\Big)}{\displaystyle\sum_{i=1}^{N} u_{ci}}\Bigg)\boldsymbol{v}_c \end{aligned} \tag{27}$$

と変形される．ここで，

$$C_i = \Big(I - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\Big)^{\mathrm{T}}\Big(I - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\Big) \tag{28}$$

であり，$\boldsymbol{b}_c$ の要素を対角要素にもつ対角行列を D_{bc}，

$$\boldsymbol{b}_c = \frac{\displaystyle\sum_{i=1}^{N} u_{ci}\boldsymbol{x}_i}{\displaystyle\sum_{i=1}^{N} u_{ci}} \tag{29}$$

とすると，

$$D_{xi}C_iD_{xi} = D_{xi}^2 - \frac{1}{n}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}} \tag{30}$$

$$\begin{aligned} & \frac{\Big(\displaystyle\sum_{i=1}^{N} u_{ci}D_{xi}\Big)C_i\Big(\displaystyle\sum_{i=1}^{N} u_{ci}D_{xi}\Big)}{\displaystyle\sum_{i=1}^{N} u_{ci}} \\ & = \sum_{i=1}^{N} u_{ci}D_{bc}^2 - \frac{1}{n}\sum_{i=1}^{N} u_{ci}\boldsymbol{b}_c\boldsymbol{b}_c^{\mathrm{T}} \end{aligned} \tag{31}$$

となるので，$J_c^{**} = \boldsymbol{v}_c^{\mathrm{T}} A_c \boldsymbol{v}_c$ とおくと，

$$\begin{aligned} A_c &= \sum_{i=1}^{N} u_{ci}(D_{xi}^2 - D_{bc}^2) \\ &\quad -\frac{1}{n}\sum_{i=1}^{N} u_{ci}(\boldsymbol{x}_i - \boldsymbol{b}_c)(\boldsymbol{x}_i - \boldsymbol{b}_c)^{\mathrm{T}} \\ &= \mathrm{diag}(S_{fc}) - \frac{1}{n}S_{fc} \end{aligned} \tag{32}$$

となる．ここで，S_{fc} はクラスタ c におけるファジィ散布行列であり，diag は対角要素のみからなる対角行列を表す．したがって，$\mathrm{diag}(S_{fc}) = S_c$ から，$S_c^{-1/2} A_c S_c^{-1/2}$ は，

$$S_c^{-1/2} A_c S_c^{-1/2} = I - S_c^{-1/2} S_{fc} S_c^{-1/2}/n \tag{33}$$

$$= I - R_c/n \tag{34}$$

とできる．ここで，R_c はファジィクラスタ c における相関係数行列であり，その大きい固有値に対応する固有ベクトルはメンバシップを考慮しながら標準化したデータの主成分ベクトルに相当する [19]．(22) 式の固有値問題は，$t = 1$ のとき，固有値を λ とおくと，

$$S_c^{-1/2} A_c S_c^{-1/2} \tilde{\boldsymbol{v}}_c = (I - R_c/n)\tilde{\boldsymbol{v}}_c = \lambda \tilde{\boldsymbol{v}}_c \tag{35}$$

$$R_c \tilde{\boldsymbol{v}}_c = (1-\lambda) n \tilde{\boldsymbol{v}}_c \tag{36}$$

となることから，相関係数行列の最大固有値に対応する固有ベクトルが，$S_c^{-1/2} A_c S_c^{-1/2}$ の最小固有値に対応する固有ベクトルに等しいことがわかる．同様に，2 本以上の主成分ベクトルを求める際にも，$S_c^{-1/2} A_c S_c^{-1/2}$ の小さい固有値に対応する固有ベクトルを求めることにより，相関係数行列の大きい固有値に対応する固有ベクトルを求めることができる．

2.2 欠測値を含むデータの主成分分析とファジィクラスタリングの同時適用法

本節では，前節で提案した局所的な主成分分析のための目的関数に欠測値を含む場合に拡張した FCM 法の目的関数を融合することにより，欠測値を含むデータの主成分分析とファジィクラスタリングの同時適用法を提案する．

FCM 法における欠測値の処理法としては，宮本ら [14] がメンバシップのべき乗を用いる標準的な方法とエントロピー正則化 [20] を用いる方法の二つのバリエーションごとに，欠測値を無視して残りの座標で距離を定義する手法や欠測値に重み付きの平均値を代入する手法を提案している．また，Timm ら [15] も標準的な方法について同様の考察を行い，欠測値を無視する手法の方が，代入する場合よりもよりあいまいなメンバシップの割り当てを行う傾向があると報告している．

欠測値を無視することにより，標準的な FCM 法の目的関数は，次のように書き換えられる．

$$\psi = \sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci}^{m} \sum_{j=1}^{n} d_{ij}(x_{ij} - b_{cj})^2 \tag{37}$$

前節の局所的な主成分分析の目的関数と欠測値を含むデータの FCM 法の目的関数を融合することにより，欠測値を含むデータのファジィクラスタリングと主成分分析の同時適用法の定式化を考える．ただし，前節の局所的な主成分分析ではデータをクラスタごとに標準化したデータの主成分の抽出を考えていたのに対し，前述の標準的な FCM 法もしくはエントロピー正則化を用いる FCM 法では，クラスタごとのデータの標準化を考慮していない．そこで，クラスタごとの容量の最適化も考慮に入れながらファジィクラスタを得ることができる K-L 情報量正則化 [18] を導入することにより，ラグランジュ乗数法による目的関数を以下のように定義する．

$$\begin{aligned} L &= \sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci} \sum_{j=1}^{n} d_{ij}\Big(\alpha \sum_{k=1}^{t}(y_{cijk} - w_{cik})^2 \\ &\quad +(1-\alpha)(x_{ij} - b_{cj})^2\Big) \end{aligned}$$

$$+\beta\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ci}\log\frac{u_{ci}}{\pi_c}$$
$$-\sum_{i=1}^{N}\gamma_i\Big(\sum_{c=1}^{C}u_{ci}-1\Big)-\tau\Big(\sum_{c=1}^{C}\pi_c-1\Big) \tag{38}$$

ここで，π_c はクラスタ c の容量を表す変量で，γ_i および τ はラグランジュ乗数である．第 1 項目の $(y_{cijk}-w_{cik})^2$ は前節で述べた主成分分析のための項であり，これを最小化することによって局所的な主成分の抽出を行う．$(x_{ij}-b_{cj})^2$ は FCM 法の目的関数を表す．α は主成分分析とクラスタリングのトレードオフを表すパラメータであり，$\alpha=0$ のときには FCM 法と同様の球状のクラスタが得られ，α を 1 に近づけるにつれて，主成分の抽出を重視した線形多様体状のクラスタが得られるようになる．第 2 項目はファジィクラスタを得るための K-L 情報量を表す項であり，メンバシップ u_{ci} とクラスタ容量 π_c を近づけることにより，クラスタ容量を考慮しながらクラスタリングが行われる．重み β はファジィ度を制御する係数で，標準的な FCM 法における m と同様に，大きくするにしたがってよりあいまいなデータ分割が得られるようになる．第 3，4 項目はそれぞれ，メンバシップの和が 1 という制約条件，π_c の和が 1 という制約条件を表す項である．文献 [18] ではマハラノビス距離を用い，変量間の分散共分散行列の要素も決定変数としているが，ここでは変量間の相関関係は主成分によって表されるという観点から，クラスタ容量のみを決定変数として用いている．

L の最適性の必要条件 $\partial L/\partial u_{ci}=0$ より，

$$u_{ci}=\frac{\pi_c\exp(E_{ci})}{\sum_{l=1}^{C}\pi_l\exp(E_{li})} \tag{39}$$

が得られる．ただし，

$$E_{li} = -\frac{1}{\beta}\sum_{j=1}^{n}d_{ij}\Big(\alpha\sum_{k=1}^{t}(y_{lijk}-w_{lik})^2 +(1-\alpha)(x_{ij}-b_{lj})^2\Big) \tag{40}$$

である．また，$\partial L/\partial\pi_c=0$，$\partial L/\partial b_{cj}=0$ から，それぞれ，

$$\pi_c = \frac{1}{N}\sum_{i=1}^{N}u_{ci} \tag{41}$$

$$b_{cj} = \frac{\sum_{i=1}^{N}u_{ci}d_{ij}x_{ij}}{\sum_{i=1}^{N}u_{ci}d_{ij}} \tag{42}$$

である．

クラスタリングアルゴリズムを以下に示す．

step:1 メンバシップ u_{ci}, $i=1,\cdots,N$, $c=1,\cdots,C$ を乱数により $\sum_{c=1}^{C}u_{ci}=1$ となるように定める．

step:2 (22) 式の固有値問題を解き，局所的な主成分ベクトル， $\boldsymbol{y}_{cij}$，$\boldsymbol{w}_{ci}$ を求める．

step:3 (41) 式より，π_c を更新する．

step:4 (42) 式より，b_{cj} を更新する．

step:5 (39) 式より，u_{ci} を更新する．

step:6 小さな正数 ε に対して収束判定条件
$$\max_{c,i}|u_{ci}^{NEW}-u_{ci}^{OLD}|<\varepsilon$$
を満たせば終了．そうでなければ，ステップ 2 へ．

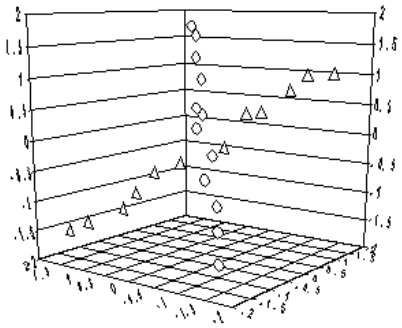

図 1: Clustering result with no missing value

2.3 因子負荷量の求め方

クラスタごとに抽出された主成分と各変量の関係を定量的に示すためには，第 j 変量 $x_{\cdot j}$ のクラスタ c における第 k 主成分 $z_{c\cdot k}$ に対するファジィ因子負荷量 η_{ckj} [19] を用いることができる．

$$\eta_{ckj} = \frac{Cov_c\{z_{c\cdot k}, x_{\cdot j}\}}{\sqrt{V_c\{z_{c\cdot k}\}V_c\{x_{\cdot j}\}}} \tag{43}$$

ただし，$Cov_c\{z_{c\cdot k}, x_{\cdot j}\}$，$V_c\{z_{c\cdot k}\}$，$V_c\{x_{\cdot j}\}$ はそれぞれ，クラスタ c における主成分 $z_{c\cdot k}$ と変量 $x_{\cdot j}$ の共分散，主成分の分散，および変量の分散である．

第 i データ点の主成分得点を要素とするベクトル $\boldsymbol{z}_{ci}$ は，欠測値を含まない場合については，主成分行列 V_c とクラスタ中心 $\boldsymbol{b}_c$ を用いて，

$$\boldsymbol{z}_{ci} = V_c^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{b}_c) \tag{44}$$

により求められるが，欠測値を含むデータについては適用することができない．そこで，欠測値を含むデータについては，データの低階数近似の立場 [11] から，

$$\boldsymbol{x}_i = V_c\boldsymbol{z}_{ci} + \boldsymbol{b}_c + \boldsymbol{e}_i \tag{45}$$

とおいたときに，観測値の 2 乗誤差の和 ${\boldsymbol{e}_i}^{\mathrm{T}} D_{wi}\boldsymbol{e}_i$ が最小となるように，

$$\boldsymbol{z}_{ci} = ({V_c}^{\mathrm{T}} D_{wi} V_c)^{-1} {V_c}^{\mathrm{T}} D_{wi}(\boldsymbol{x}_i - \boldsymbol{b}_c) \tag{46}$$

と求める．これは，欠測値を含むデータが主成分ベクトルにより張られる線形多様体上，もしくはその線形多様体からの距離が最も近くなる点に存在すると仮定して欠測値を補完し，(44) 式の値を求めることに等しい．

3 数値例

本章では，提案手法の有効性を検証するために，以下の二つの数値例を示す．

3.1 人工データを用いた数値実験

まず，表 1 に示す 3 次元 24 個の人工データを用いて，実験を行った．データ集合は 3 次元空間内で二つの直線状に布置されており，欠測値がない場合には図 1 に示されるように提案法を用いて○と△で表される二つの線形クラスタに分割され，表 2 の因子負荷量を持つ主成分とクラスタ中心が得られた．ただし，パラメータは $\alpha = 0.99$，$\beta = 0.1$ とした．この人工データではクラスタごとに各変量がほぼ等しい分散を有しているために，得られるクラスタリング結果はデータの標準化を考慮しない FCV 法と同様なものとなっている．

つぎに，このデータ集合に対して，24 個すべてのサンプルからランダムに一つずつ，表 1 で太字で示された値を欠落させることにより，データ行列の要素の 1/3 が欠測値となるテストデータ集合を作成した．すべてのサンプルデータに欠測値が含まれるために，欠測値を含むサンプルを削除すると全く分析することができず，回帰分析などによりデータを補完して分析することも困難な例となっている．作成した不完全データを欠測値が無い場合と同じパラメータを用いて提案法により二つのクラ

表 1: Artificial data set

—	x_1	x_2	x_3
1	-1.43	-1.28	$\mathbf{-1.56}$
2	$\mathbf{-0.96}$	-0.93	-1.21
3	$\mathbf{-0.80}$	-0.77	-0.87
4	$\mathbf{-0.63}$	-0.41	-0.52
5	-0.36	-0.35	$\mathbf{-0.17}$
6	0.18	0.26	**0.38**
7	**0.40**	0.55	0.43
8	0.82	0.77	**0.88**
9	0.99	**1.02**	1.23
10	**1.33**	1.28	1.58
11	1.53	**1.54**	1.73
12	$\mathbf{-0.06}$	0.20	0.18
13	$\mathbf{-1.71}$	1.54	-1.46
14	-1.54	**1.28**	-1.33
15	-1.17	**0.83**	-1.10
16	$\mathbf{-0.90}$	0.77	-0.87
17	-0.73	**0.51**	-0.53
18	-0.36	**0.25**	-0.40
19	0.18	-0.26	$\mathbf{-0.16}$
20	0.45	-0.51	**0.39**
21	0.52	-0.77	**0.43**
22	0.99	$\mathbf{-1.03}$	0.76
23	1.26	$\mathbf{-1.18}$	0.99
24	1.53	-1.54	**1.03**

表 2: Result of analysis with no missing value

cluster center

—	x_1	x_2	x_3
$c = 1$	-0.13	-0.01	-0.17
$c = 2$	0.11	0.16	0.16

factor loading

—	x_1	x_2	x_3
$c = 1$	0.94	-0.93	0.94
$c = 2$	0.94	0.94	0.95

スタに分割した結果を表 3 に示す．欠測値の有無に関わらず，ほぼ等しいクラスタ中心および主成分が得られており，提案法が欠測値を含む不完全データから局所的な構造を考慮しながら主成分を抽出するのに有効であることが分かる．

表 3: Result of analysis with missing values

cluster center

—	x_1	x_2	x_3
$c=1$	0.38	-0.07	-0.22
$c=2$	-0.02	-0.11	0.01

factor loading

—	x_1	x_2	x_3
$c=1$	0.85	-0.57	0.84
$c=2$	0.82	0.65	0.84

表 4: Result of FCV

———	cluster center		factor loading($c=1$)		factor loading($c=2$)	
variable	$c=1$	c=2	1st PC	2nd PC	1st PC	2nd PC
1.	-0.197	0.198	0.000	-0.003	-0.080	0.117
2.	-0.022	-0.022	0.223	**-0.383**	-0.227	**-0.449**
3.	0.000	0.000	-0.043	**0.456**	0.087	0.338
4.	0.021	0.021	-0.273	0.289	0.332	**0.428**
5.	**-0.786**	**0.790**	0.252	**0.471**	0.257	**-0.404**
6.	**-0.775**	**0.779**	0.291	**0.481**	0.255	**-0.378**
7.	**-0.765**	**0.769**	0.205	**0.468**	**0.397**	**-0.443**
8.	**-0.766**	**0.770**	0.112	**0.413**	**0.381**	**-0.472**
9.	**-0.768**	**0.771**	0.164	**0.412**	0.326	**-0.444**
10.	-0.344	0.345	**0.568**	0.262	**-0.647**	0.335
11.	-0.177	0.178	**0.687**	0.246	**-0.750**	**0.453**
12.	-0.105	0.105	**0.391**	0.034	**-0.422**	**0.421**
13.	-0.140	0.140	**0.830**	**0.449**	**-0.669**	**0.393**
14.	0.058	-0.059	**0.426**	0.040	**-0.693**	0.260
15.	0.033	-0.034	**0.698**	**0.435**	**-0.522**	0.182
16.	0.033	-0.034	**0.641**	**0.387**	-0.306	0.258
17.	-0.140	0.140	**-0.601**	**0.552**	**0.608**	**0.490**
18.	-0.162	0.163	**-0.577**	**0.687**	**0.630**	**0.608**
19.	-0.197	0.198	**-0.628**	**0.709**	**0.645**	**0.658**
20.	-0.089	0.089	**-0.554**	**0.712**	**0.562**	**0.740**

3.2 POS データの分析

つぎに，文献 [7] で用いられた POS(Point of Sales) データによる数値実験を通して，提案法の特徴を示す．データは 2 店舗のスーパーマーケットで 1 年間にわたって収集・蓄積されたもので，12 月 31 日などの特異な傾向を示す日を除いた 333 日分について，各変量ごとに平均 0，分散 1 に基準化したものを用いた．用いた変量は，2 店舗の来客数，日配品点数に暦，気象情報を加えた以下の 20 個の変量である．

1. 祝日，2. 金曜，3. 土曜，4. 日曜，5. 平均気温，
6-9. 気温 (9 時,12 時,15 時,18 時)，10. 相対湿度，
11,12. 天気概況 (昼, 夜)，13. 日降水量，
14-16. 降水量 (9-12 時,12-15 時,15-18 時),

表 5: Result of proposed method with no missing value

———	cluster center		factor loading($c=1$)		factor loading($c=2$)	
variable	$c=1$	c=2	1st PC	2nd PC	1st PC	2nd PC
1.	−0.021	0.020	0.056	0.005	0.036	0.014
2.	0.033	−0.031	0.192	−0.163	−0.120	−0.203
3.	−0.070	0.065	−0.072	0.128	0.106	0.081
4.	−0.038	0.035	−0.120	0.144	0.077	**0.352**
5.	**−0.829**	**0.777**	**0.669**	**0.660**	**−0.841**	−0.022
6.	**−0.823**	**0.771**	**0.666**	**0.646**	**−0.822**	−0.033
7.	**−0.820**	**0.768**	**0.641**	**0.674**	**−0.854**	0.047
8.	**−0.812**	**0.761**	**0.624**	**0.678**	**−0.866**	0.065
9.	**−0.814**	**0.762**	**0.634**	**0.661**	**−0.857**	0.028
10.	−0.279	0.261	**0.488**	−0.034	0.035	**−0.497**
11.	−0.119	0.112	0.339	−0.200	0.231	**−0.544**
12.	−0.107	0.100	0.330	−0.103	0.180	−0.246
13.	−0.155	0.145	**0.488**	−0.105	0.165	**−0.525**
14.	−0.093	0.087	**0.387**	−0.094	0.113	**−0.411**
15.	−0.089	0.083	**0.430**	−0.083	0.109	**−0.408**
16.	−0.088	0.082	0.296	−0.181	0.097	−0.290
17.	−0.194	0.182	−0.195	**0.442**	−0.008	**0.588**
18.	−0.227	0.213	−0.188	**0.457**	0.025	**0.607**
19.	−0.257	0.241	−0.168	**0.536**	0.024	**0.639**
20.	−0.138	0.130	−0.252	**0.383**	0.143	**0.588**

17. 来客数 (A 店)，18. 日配品点数 (A 店)，
19. 来客数 (B 店)，20. 日配品点数 (B 店)

ただし，日配品とは工場で生産される食品のうち，数日中に消費されるものを指し，その総売上点数が日配品点数である．また，暦に関する変数はおのおのあてはまるか否かを表すダミー変数であり，天気概況は降水の状況により 0（晴れまたは曇り）から 3（大雨）の値を与えた．

まず，相関係数行列を用いる提案法と分散共分散行列を用いる FCV 法の違いを調べるために，333 日分の POS データのすべてを用いて提案法で分析した結果と文献 [7] で得られた FCV 法による分析結果の比較を行った．おのおのの手法で得られたクラスタ中心と二つの主成分のファジィ因子負荷量を表 4，表 5 に示す．表中，絶対値が 0.35 以上のものは太字で記載している．ただし，提案法におけるパラメータは $\alpha = 0.9$，$\beta = 1.0$ とした．

表 4，表 5 の 2〜3 列目から，いずれの手法においても太字で示された気温に関する変量でクラスタ中心が分かれており，温暖期と寒冷期の二つのクラスタに分類されている．一方，表中の 4〜7 列目に示されたファジィ因子負荷量を比較すると，FCV 法ではいずれのクラスタにおいても第 1 主成分が降水量や来店客数・日配品点数といった変量と相関が大きいのに対して，提案法では第 1 主成分が気温の影響を強く反映している．これは，主に気温の高低を加味したクラスタリングの結果，おのおののクラスタに属するデータの気温の分散が小さくなっているために，FCV 法ではその影響が弱まっているのに対して，提案法ではメンバシップを考慮した標準化を行っているので FCV 法では抽出できない特徴量が抽出できていると考えられる．このように，相関係数行列を用いる提案法は，クラスタリングによる層別の影響で FCV 法が無視してしまう変量についても，標準化して考慮することができるという利点があることが分かる．また，この特徴は主成分分析において解が変量の測定単位により影響を受けるという性質，すなわち尺度不変（scale invariant）でないという性質に対応するものである．

つぎに，観測データに欠測値が含まれる場合を考える．333 日分のサンプルデータのすべてについて，気象に関する変量（5〜16）からランダムに二つずつの値を欠落させることにより，10%の欠測値が含まれるデータ行列を作成した．前節と同様に，すべてのサンプルが欠測値を有するために，

表 6: Result of FCV with missing values

———	cluster center		factor loading($c=1$)		factor loading($c=2$)	
variable	$c=1$	c=2	1st PC	2nd PC	1st PC	2nd PC
1.	−0.018	0.021	0.042	0.032	0.001	0.020
2.	0.016	−0.026	0.164	**−0.395**	−0.327	0.114
3.	−0.040	0.052	0.002	**0.372**	0.169	0.333
4.	−0.054	0.056	0.233	0.333	**0.379**	0.270
5.	**−0.791**	**0.748**	0.300	**0.366**	−0.010	−0.176
6.	**−0.761**	**0.720**	0.302	**0.375**	−0.067	−0.166
7.	**−0.771**	**0.760**	0.190	**0.391**	0.051	−0.219
8.	**−0.775**	**0.707**	0.224	**0.369**	−0.007	−0.239
9.	**−0.797**	**0.738**	0.279	**0.359**	0.012	−0.210
10.	−0.274	0.208	**0.580**	0.195	**−0.418**	**0.452**
11.	−0.054	0.112	**0.611**	0.204	**−0.445**	**0.541**
12.	−0.107	0.099	**0.391**	0.068	−0.305	**0.397**
13.	−0.036	0.054	**0.659**	0.195	**−0.466**	**0.564**
14.	−0.067	0.104	**0.472**	0.028	**−0.391**	**0.398**
15.	0.093	−0.066	**0.505**	0.200	−0.269	0.280
16.	0.006	0.016	**0.493**	0.151	−0.269	0.280
17.	−0.206	0.229	**−0.454**	**0.611**	**0.677**	0.249
18.	−0.263	0.261	**−0.411**	**0.718**	**0.749**	**0.379**
19.	−0.263	0.261	**−0.445**	**0.739**	**0.749**	**0.366**
20.	−0.141	0.148	**−0.410**	**0.725**	**0.705**	**0.491**

欠測値を含むサンプルを削除すると全く分析することができない例となっている．文献 [16] の FCV 法を不完全データに拡張した手法および提案法を用いて分析した結果を表 6，表 7 に示す．ただし，パラメータは欠測値が無い場合と同じものを用い，表中，絶対値が 0.35 以上のものを太字で示した．

表 6，表 7 の 2〜3 列目から，欠測値の有無に関わらず，温暖期と寒冷期に分割されていることが分かる．また，表中の 4〜7 列目に示されたファジィ因子負荷量には，以下のような特徴が見られる．欠測値がない場合には目的関数が FCV 法と等しくなる文献 [16] の手法では，気温に関する変量の因子負荷量がさらに小さくなったものの，欠測値がない場合の FCV 法と類似した結果が得られている．一方，提案法では，第 1 主成分において欠測値を含む気象関連の変量の因子負荷量が小さくなり，FCV 法の結果に近づく傾向が見られたものの，データの層別に用いられた気温に関する変量の影響も無視されておらず，欠測値が無い場合と類似した相関関係がとらえられている．これらの違いは，文献 [16] の手法がファジィ散布行列を用いる FCV 法を欠測値が含まれるデータに拡張したものであるのに対して，提案手法が相関係数行列を用いることによりクラスタごとにデータの基準化を考慮していることを示している．以上のように，提案法はファジィ分散共分散行列の固有値問題に帰着される FCV 法とは若干異なる主成分を抽出するものの，欠測値を含む場合にも妥当な結果を得られる手法であることが分かる．

4 おわりに

欠測値の補完などの手続きをせずに，欠測値を含むデータから局所的な主成分を抽出する手法を提案した．Bezdek らの FCV 法がデータの分散共分散行列を用いた主成分分析とファジィクラスタリングの同時適用法ととらえられるのに対して，提案法は相関係数行列を用いた主成分分析とファジィクラスタリングの同時適用法である．FCV 法を不完全データに拡張した本多ら [16] の手法ではデータ行列の低階数近似行列を保持するために，サンプルデータ数が多くなるにつれてメモリの消費量が増すのに対して，提案法では欠測値があるにもかかわらず固有値問題を解くことにより解析的に主成

表 7: Result of proposed method with missing values

———	cluster center		factor loading($c=1$)		factor loading($c=2$)	
variable	$c=1$	c=2	1st PC	2nd PC	1st PC	2nd PC
1.	−0.016	0.016	0.046	0.040	0.038	−0.022
2.	0.049	−0.047	0.280	−0.073	−0.247	−0.114
3.	−0.083	0.080	−0.166	0.100	0.159	−0.001
4.	−0.062	0.059	−0.211	0.082	0.266	0.268
5.	**−0.793**	**0.742**	0.286	**0.783**	**−0.680**	**0.354**
6.	**−0.764**	**0.718**	0.295	**0.751**	**−0.664**	0.304
7.	**−0.759**	**0.755**	0.227	**0.785**	**−0.630**	**0.454**
8.	**−0.762**	**0.698**	0.269	**0.755**	**−0.647**	**0.411**
9.	**−0.791**	**0.736**	0.271	**0.769**	**−0.674**	**0.391**
10.	−0.314	0.229	**0.377**	0.163	−0.174	**−0.444**
11.	−0.106	0.138	**0.354**	0.038	−0.022	**−0.538**
12.	−0.109	0.091	0.336	0.056	0.017	−0.277
13.	−0.132	0.114	**0.421**	0.097	−0.068	**−0.504**
14.	−0.075	0.104	**0.352**	0.044	−0.073	**−0.363**
15.	−0.098	0.077	0.323	0.111	−0.059	−0.336
16.	−0.084	0.079	0.311	0.058	−0.055	−0.288
17.	−0.228	0.219	**−0.433**	0.335	0.332	**0.512**
18.	−0.263	0.252	**−0.448**	**0.359**	**0.380**	**0.505**
19.	−0.294	0.283	**−0.463**	**0.431**	**0.392**	**0.525**
20.	−0.185	0.177	**−0.479**	0.261	**0.476**	**0.421**

分ベクトルが求まるため，データ数が多い場合でも効率的に分析できるという特長がある．ただし，主成分分析の重みやファジィ度を制御するパラメータなどを試行錯誤により定めなければならない点は従来法と同様に問題点として残っており，適応的にクラスタ形状を決定する手法の開発が課題である．数値例では，FCV 法と提案法での欠測値がない場合の結果の比較をとおして，クラスタごとのデータの標準化を考慮するか否かで結果に違いを生じることを示した．また，欠測値がある場合についても妥当な主成分が抽出されることを示した．提案法は欠測値がランダムに発生しているという仮定に基づいて定式化がなされているが，実データを扱う際には欠測値の発生に偏りがある場合も少なくない．欠測値の発生機構を考慮に入れたモデルの作成が今後の課題である．

参考文献

[1] J. C. Bezdek: *Pattern recognition with fuzzy objective function algorithms,* Plenum press (1981)

[2] J. C. Bezdek, C. Coray, R. Gunderson and J. Watson: Detection and characterization of cluster substructure II. Fuzzy c-varieties and convex combinations thereof; *SIAM Jour. of Appl. Math.*, Vol.40, No.2, pp.358–372 (1981)

[3] R. N. Dave: An adaptive fuzzy c-elliptotype clustering algorithm; *Proc. of the North American Fuzzy Information Processing Society :Quater Century of Fuzziness*, Vol.1, pp.9–12 (1990)

[4] 馬屋原, 中森: 線形多様体クラスタリングと楕円形ファジィモデル; 日本ファジィ学会誌, Vol.10, No.1, pp.142-149 (1998)

[5] 馬屋原, 宮本: 次元係数の正則化による線形ファジィクラスタリング; 日本ファジィ学会誌, Vol.12, No.4, pp.552-561 (2000)

[6] 本多, 山川, 市橋, 三好, 奥山: ファジィクラスタリングと回帰と主成分の同時分析法; システム制御情報学会論文誌, Vol. 13, No. 5, pp. 236–243 (2000)

[7] 呉, 本多, 新山, 市橋: ファジィクラスタリングを用いる外的基準に独立な局所的主成分の抽出法; 日本ファジィ学会誌, Vol. 12, No. 6, pp. 826–834 (2000)

[8] 竹内: 統計学辞典, 東洋経済新報社 (1989)

[9] A. Ruhe: Numerical computation of principal components when several observations are missing; *Tech Rep. UMINF-48-74, Dept. Information Processing*, Umea Univ. (1974)

[10] T. Wiberg: Computation of principal components when data are missing; *Proc. of Second Symp. Computational Statistics*, pp. 229–236 (1976)

[11] 柴山: 欠測値を含む多変量データのための主成分分析的方法; 教育心理学研究, Vol. 40, No. 3, pp. 257–265 (1992)

[12] 高根: 制約つき主成分分析法, 朝倉書店 (1995)

[13] 柴山, 芝: 欠測値がある場合の線形等化法; 教育心理学研究, Vol.35, No.1, pp.86–89 (1987)

[14] S. Miyamoto, O. Takata and K. Umayahara: Handling missing values in fuzzy c-means; *Proc. of Third Asian Fuzzy Systems Symp.*, pp. 139–142 (1998)

[15] H. Timm and R. Kruse: Fuzzy cluster analysis with missing values; *Proc. of 17th International Conf. of the North American Fuzzy Information Processing Society*, pp. 242–246 (1998)

[16] 本多, 杉浦, 市橋, 荒木, 久津見: 最小 2 乗基準を用いた Fuzzy c-Varieties 法における欠測値の処理法; 日本ファジィ学会誌, Vol. 13, No. 6, pp. 680–688 (2001)

[17] K. Honda, N. Sugiura, H. Ichihashi and S. Araki: Collaborative filtering using principal component analysis and fuzzy clustering; *Web Intelligence: Research and Development*, Lecture Notes in Artificial Intelligence 2198, Springer, pp. 394-402 (2001)

[18] 宮岸, 市橋, 本多: K-L 情報量正則化 FCM クラスタリング法; 日本ファジィ学会誌, Vol. 13, No. 4, pp. 406–417 (2001)

[19] Y. Yabuuchi and J. Watada: Fuzzy principal component analysis and its application; *Biomedical Fuzzy and Human Sciences*, Vol.3, No.1, pp.83–92 (1997)

[20] 宮本: クラスター分析入門, 森北出版 (1999)