ロバストな線形ファジィクラスタリング法の提案と協調フィルタリングシステムへの応用*

(Extraction of Local Principal Components from Data with Missing Values)

本多　克宏†・杉浦　伸和†・市橋　秀友†
†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

概 要

線形多様体をプロトタイプとするFuzzy c-Varieties（FCV）法は，局所的な主成分分析法とみなされるクラスタリング法であるが，最小2乗法の原理に基づいており，ノイズの影響を受けやすいという欠点がある．本研究では，サンプル内のノイズを処理することによりロバストな線形クラスタリングを行う手法を提案する．おのおのの要素が外れ値であるか否かを定める重みを付加することによってロバストなモデルの推定を行う提案法は，欠測値に対応する重みを0に固定することにより，欠測値のロバストな推定が可能である．協調フィルタリングへの応用実験では，メモリの所要量が少ないにもかかわらず従来法以上の推薦性能があることを示す．

(Non-linear extensions of Principal Component Analysis (PCA) have been developed for detecting the lower-dimensional representations of real world data sets. Fuzzy c-Varieties (FCV) is the linear fuzzy clustering algorithm that can be regarded as a Local PCA technique. However least squares techniques often fail to account for "outliers". This paper proposes a technique for making the FCV algorithm robust to intra-sample outliers. The objective function based on the lower rank approximation of the data matrix is minimized by a robust M-estimation algorithm that is similar to FCM-type iterative procedures. The new method is also useful for estimating missing values and a numerical experiment of Collaborative Filtering reveals an improvement in recommendation performance.)
keywords: robust clustering, principal component analysis, missing value, collaborative filtering.

1　はじめに

多変量データの情報圧縮の際に線形モデルを用いる主成分分析（Principal Component Analysis: PCA）は，データベースからの知識発見や画像処理など，広範な分野で応用されている基礎的な手法であるが，そのモデルは最小2乗法の原理に基づくものであり，外れ値やノイズの影響を受けやすいという欠点がある．実データを分析する際に問題となるノイズには，2種類ある．一つは標本データの集合にいくつかのノイズサンプルが紛れ込んでいる場合であり，この場合にはノイズサンプルに関するすべての要素をデータ行列から取り除いて分析する必要がある．一方，標本データそのものは不良なサンプルではないものの，その要素のいくつかのみが測定ミスなどのために外れ値となっている"サンプル内のノイズ"も存在する．これらのサンプル内のノイズを扱う際には，外れ値を含むサンプルをすべて削除してしまうのではなく，外れ値のみを無視して分析することが望ましい．そこで，

*システム制御情報学会論文誌, 16, 11, 597-605 (2003)

De la Torre ら [1, 2] は画像の分析においてピクセルレベルで生じるノイズを処理しながらロバストな主成分を得るための方法として，要素ごとのロバスト M 推定に基づく部分空間の抽出法を提案している．

しかしながら，実世界で収集される大規模なデータ集合からの特徴抽出においては，単一の線形モデルでは高次元データの詳細な特徴をとらえられない場合も多い．そこで，非線形な分布形状を有する多変量データに対して局所的に線形モデルを当てはめることにより，部分構造を考慮した分析を行う研究がいくつかなされている [3, 4, 5]．また，局所的な線形構造をとらえるためのファジィクラスタリング法である Bezdek らの Fuzzy c-Varieties（FCV）法 [6, 7] も，部分空間への帰属度（メンバシップ）を考慮した散布行列の固有値問題から線形多様体を張る直交基底系を求める手法であり，局所的な主成分分析法ととらえられるものである．これらは非線形なデータ分析法の一種とみなされるが，計算の簡略さやモデルの理解の容易さの点で，Principal Curves [8] やニューラルネットワークを用いる方法 [9, 10] のような非線形手法よりも利用しやすい．

ファジィクラスタリングもまた最小 2 乗法に基づく定式化が一般的であり，ノイズの影響を避けながらメンバシップを推定するための手法がいくつか考えられている．Dave らのノイズクラスタリング [11] では C 個のクラスタの他に，すべてのノイズサンプルを吸収する $C+1$ 個目の “ノイズクラスタ” を用意し，ノイズの影響を受けない C 個のクラスタにデータを分割している．そのため，おのおのの標本データに関する C 個のクラスタへのメンバシップの和は 1 以下となる制約が付加されたモデルとなっている．また，可能性的な制約を用いる Krishnapuram らの可能性的クラスタリング [12] では，メンバシップの和に関する制約が取り除かれ，おのおののメンバシップが $[0, 1]$ の値を持つことだけが定められている．そのため，メンバシップはおのおのの標本データがクラスタに所属する可能性を表しているとみなされる．さらに，線形クラスタリングにおいてロバストな局所的モデルを推定する手法としては，プロトタイプが超平面で表されるモデルに絶対値距離によるクラスタリング基準を導入した研究もなされている [13]．しかし，これらの手法ではノイズとみなされたサンプルの要素すべてを無視して分析を行うため，多くの変量のいくつかのみが外れ値である場合には情報の損失が大きい．

本論文では，サンプル内のノイズを処理することによりロバストな線形クラスタリングを行う手法を提案する．データ点と線形多様体との距離をクラスタリング基準とする FCV 法の目的関数は，主成分分析で用いられるデータ行列の低階数近似 [14] に基づく手法と類似点がある [15, 16]．そこで本論文では，要素ごとの低階数近似に M 推定で用いられる手法を導入することにより，サンプル内の外れ値を無視しながら線形クラスタリングを施す手法を提案する．提案法のアルゴリズムは反復重み付け最小 2 乗（Iteratively Reweighted Least Squares: IRLS）法 [17] の原理に基づいており，おのおのの要素が外れ値であるか否かを定める重みを付加することで，非線形方程式を解くことなしにロバストなモデルの推定を行う．

また，ノイズと同様に実データの分析の障害となる欠測値を処理する方法として，文献 [15] では要素ごとにデータ行列を近似する最小 2 乗基準の利点を活かした欠測値処理法が提案されており，近似行列の要素を欠測値の推定値とすることにより協調フィルタリングへの応用が試みられている [16]．提案法ではおのおのの要素に対する重みも決定変数であるが，事前知識としてデータの観測の有無が与えられる場合には，欠測値に対応する重みを 0 に固定することにより，文献 [15] と同様に欠測値を処理することができ，ロバストな局所的モデルを用いた欠測値の推定が可能である．数値実験では，人工データを用いて提案法の特長を検証した後に，情報フィルタリング技術の一種である協調フィルタリングシステムでの比較により推薦性能が向上することを示す．

2　ロバストな主成分分析とクラスタリングの同時適用法

2.1　最小 2 乗基準を用いた線形ファジィクラスタリング

m 次元の n 個の標本データからなる（$n \times m$）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する問題を考える．データ行列は適宜，変量 j に係る n 個の要素を並べた縦ベクトル $\boldsymbol{x}_j$ を用いて $X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_j, \cdots, \boldsymbol{x}_m)$，または標本 i に係る m 個の要素を並べた縦ベクトル $\tilde{\boldsymbol{x}}_i$ を用いて $X = (\tilde{\boldsymbol{x}}_1, \cdots, \tilde{\boldsymbol{x}}_i, \cdots, \tilde{\boldsymbol{x}}_n)^\top$ と表す．ただし，$\top$ は転置を表す．(以降，太字はすべて縦ベクトルを表し，行列の行方向の要素を並べた縦ベクトルは˜を付して表記することとする.)

主成分分析とファジィクラスタリングを同時に適用する場合には，データ集合の部分構造をよく表現する局所的な主成分ベクトルを用いてデータを分割することが目的となる．FCV 法では互いに線形独立な単位ベクトル $\boldsymbol{a}_{cj}$ により張られる p 次元の線形多様体をクラスタのプロトタイプとし，標本

データと線形多様体との距離を分類尺度とすることにより，以下の目的関数の最小化を考える [6, 7].

$$L_{fcv} = \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}^{\theta}\Big\{(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)^{\top}(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c) - \sum_{j=1}^{p} \boldsymbol{a}_{cj}^{\top} R_{ci} \boldsymbol{a}_{cj}\Big\} \tag{1}$$

$$R_{ci} = (\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c)^{\top} \tag{2}$$

ただし，u_{ci} は標本データ $\tilde{\boldsymbol{x}}_i$ の第 c クラスタへの帰属度を表すメンバシップであり，c に関する和が 1 となる確率的制約を満たすものとする．$\boldsymbol{b}_c$ は第 c クラスタの中心である．メンバシップのべき乗はファジィ分割を得るために用いられたもので，θ が大きいほどおのおのの標本データの所属が明確でなくなり，あいまいなデータ分割が得られるようになる．

最適性の必要条件 $\partial L_{fcv}/\partial \boldsymbol{a}_{cj} = \boldsymbol{0}$ から，最適なベクトル $\boldsymbol{a}_{cj}$ は以下の固有値問題を解くことで得られる．

$$S_{fc}\boldsymbol{a}_{cj} = \mu_{cj}\boldsymbol{a}_{cj} \tag{3}$$

ここで，S_{fc} は以下の一般化されたファジィ散布行列である．

$$S_{fc} = \sum_{i=1}^{n} u_{ci}^{\theta} R_{ci} \tag{4}$$

最適な $\boldsymbol{a}_{cj}$ は大きな固有値に対応する固有ベクトルとして算出されることから，クラスタごとに帰属度を考慮しながら求められるファジィ主成分ベクトル [18] であるとみなされる．

同様に，クラスタ中心およびメンバシップもおのおの最適性の必要条件 $\partial L_{fcv}/\partial \boldsymbol{b}_c = \boldsymbol{0}$ および $\partial L_{fcv}/\partial u_{ci} = 0$ から求められ，繰り返しアルゴリズムによりクラスタリング結果が得られる．

FCV 法の目的関数はデータ点と線形多様体との距離により定義されるが，本多ら [15] は主成分分析で用いられる最小 2 乗基準にメンバシップを導入することにより，データ行列の低階数近似の立場から線形ファジィクラスタリングを論じている．最小 2 乗基準を用いた局所的な主成分分析の目的関数は，以下のように定義される．

$$L_{lsc} = \sum_{c=1}^{C} \mathrm{tr}\Big\{(X - Y_c)^{\top} U_c^{\theta} (X - Y_c)\Big\} \tag{5}$$

ただし，$U_c = \mathrm{diag}(u_{c1}, \cdots, u_{cn})$ であり，tr は対角要素の和（トレース）を表す．$Y_c = (y_{cij})$ は第 c クラスタにおけるデータ行列 X の低階数近似行列を表し，

$$Y_c = F_c A_c^{\top} + \boldsymbol{1}_n \boldsymbol{b}_c^{\top} \tag{6}$$

である．ここで，$F_c = (\tilde{\boldsymbol{f}}_{c1}, \cdots, \tilde{\boldsymbol{f}}_{cn})^{\top}$ はおのおのの標本データの主成分得点からなる $(n \times p)$ の成分得点行列，$A_c = (\boldsymbol{a}_{c1}, \cdots, \boldsymbol{a}_{cp})$ は主成分ベクトルを並べた $(m \times p)$ の主成分行列である．また，$\boldsymbol{1}_n$ はすべての要素が 1 である n 次元ベクトルである．

メンバシップを固定して考えた場合，クラスタごとの局所的な主成分の抽出は (5) 式を最小とする F_c，A_c および $\boldsymbol{b}_c$ を求めることとなる．目的関数の最適性の必要条件 $\partial L_{lsc}/\partial \boldsymbol{b}_c = \boldsymbol{0}$ から，最適なクラスタ中心は

$$\boldsymbol{b}_c = (\boldsymbol{1}_n^{\top} U_c^{\theta} \boldsymbol{1}_n)^{-1} (X^{\top} - A_c F_c^{\top}) U_c^{\theta} \boldsymbol{1}_n \tag{7}$$

となり，制約条件として $F_c^{\top} U_c^{\theta} \boldsymbol{1}_n = \boldsymbol{0}$ を考えると，

$$\boldsymbol{b}_c = (\boldsymbol{1}_n^{\top} U_c^{\theta} \boldsymbol{1}_n)^{-1} X^{\top} U_c^{\theta} \boldsymbol{1}_n \tag{8}$$

のように求まる．(8) 式は FCV 法のアルゴリズムにおけるクラスタ中心 $\boldsymbol{b}_c$ の更新則と等しくなる．(6) 式を代入することにより，(5) 式は

$$L_{lsc} = \sum_{c=1}^{C}\Big\{\mathrm{tr}(X_c^{\top} U_c^{\theta} X_c) - 2\mathrm{tr}(X_c^{\top} U_c^{\theta} F_c A_c^{\top}) + \mathrm{tr}(A_c F_c^{\top} U_c^{\theta} F_c A_c^{\top})\Big\} \tag{9}$$

のようになる．ただし，$X_c = X - \boldsymbol{1}_n \boldsymbol{b}_c^{\top}$ とおいた．

さらに，$\partial L_{lsc}/\partial F_c = O$ から，以下の関係が導かれる．

$$F_c A_c^\top A_c = X_c A_c \tag{10}$$

ここで，FCV 法と同様に局所的な主成分ベクトルに $A_c^\top A_c = I$ なる制約を設けるとすると，成分得点行列は $F_c = X_c A_c$ と求まり，目的関数は以下のように書き換えられる．

$$\begin{aligned} L_{lsc} &= \sum_{c=1}^{C} \Big\{ \mathrm{tr}(X_c^\top U_c^\theta X_c) - \mathrm{tr}(A_c^\top X_c^\top U_c^\theta X_c A_c) \Big\} \\ &= L_{fcv} \end{aligned} \tag{11}$$

このように，(5) 式は FCV 法と同じ最適解を与える目的関数であるといえる．

また，(5) 式のようなデータ行列の低階数近似に基づく目的関数は，要素ごとに書き表すことにより，

$$L_{lsc} = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci}^\theta \sum_{j=1}^{m} (x_{ij} - \sum_{k=1}^{p} f_{cik} a_{cjk} - b_{cj})^2 \tag{12}$$

のように変形できることから，要素ごとの最小 2 乗近似に基づく定式化であるともみなされる．ただし，クラスタ数が 1 の場合はすべてのメンバシップが 1 となり，最小 2 乗基準に基づく主成分分析の目的関数に一致する．

2.2 ロバストな局所的主成分分析法

ファジィクラスタリングにおいてノイズの影響を無視したデータ分割を得るための方法として，Dave[11] はすべてのノイズサンプルを吸収する $C+1$ 個目の “ノイズクラスタ” を用意し，以下の目的関数を最小化するノイズクラスタリングを提案している．

$$L_{nc} = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci}^\theta d_{ci}^2 + \sum_{i=1}^{n} \delta^2 \left(1 - \sum_{c=1}^{C} u_{ci} \right)^\theta \tag{13}$$

ただし，d_{ci} はクラスタリング基準を表し，FCM 法ではクラスタ中心と標本データ点との距離，FCV 法ではプロトタイプとなる線形多様体との距離である．$u_{*i} = 1 - \sum_{c=1}^{C} u_{ci}$ はノイズクラスタへの帰属度を表し，すべての標本データとノイズクラスタとの距離を固定値 δ とすることにより，距離が δ よりも小さいクラスタが存在しない標本データをノイズクラスタに吸収している．また，Krishnapuram らの可能性的クラスタリング [12] では，

$$L_{pos} = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci}^\theta d_{ci}^2 + \sum_{c=1}^{C} \eta_c \sum_{i=1}^{n} (1 - u_{ci})^\theta \tag{14}$$

なる目的関数を，おのおののメンバシップが $[0, 1]$ の値を持つことだけを制約条件として最小化することを提案している．ただし，η_c がノイズの範囲を定める定数であり，ノイズクラスタリングにおける δ と同様の働きを担う．しかし，これらの手法ではノイズとみなされたサンプルの要素すべてを無視して分析がなされるため，多くの変量のいくつかのみが外れ値となっている場合には情報の損失が大きい．

本節では，ロバスト推定法における M 推定の手法を導入することにより，サンプル内ノイズを含む不完全データのための局所的な主成分分析法を提案する．

最小 2 乗基準を用いる主成分分析の目的関数は，(12) 式と同様に，要素ごとに最小 2 乗近似を行っているとみなすことができるため，サンプル内ノイズに対してロバストではない．そこで，サンプル内ノイズを含むデータ集合からの部分空間の抽出法として，De la Torre ら [1, 2] は，要素ごとの近似を考える際に，最小 2 乗基準の代わりにロバストな基準を導入することにより，以下の目的関数の最小化問題を提案している．

$$L_{rpca} = \sum_{i=1}^{n} \sum_{j=1}^{m} \rho(x_{ij} - \sum_{k=1}^{p} f_{ik} a_{jk} - b_j) \tag{15}$$

ここで，$\rho(\cdot)$ はロバスト ρ 関数 [17] を表しており，文献 [1, 2] では Geman-McClure 関数 [19]

$$\rho(x) = \frac{x^2}{x^2 + \sigma_j^2} \tag{16}$$

が用いられている．通常，最適化の過程において，ロバスト関数の形状を決定する σ_j をアニーリング手法により順次減少させることにより，初期値に依存しない結果を得ることが試みられる．非線形関数を含む (15) 式を最小化するために，De la Torre らは反復重み付け最小 2 乗法（Iteratively Reweighted Least Squares: IRLS） [17] に基づく方法の他，擬似的な 2 次導関数を用いる勾配法を提案している．

以下では，最小 2 乗基準を用いる線形ファジィクラスタリングの目的関数に ρ 関数を導入することにより，ロバストな局所的線形モデルを抽出する．標準的な手法におけるメンバシップのべき乗の代わりに，エントロピー正則化 [20] を考慮することにより，ロバストな FCV 法の目的関数は，以下のように定義される．

$$\begin{aligned} L_{rfcv} &= \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci} \sum_{j=1}^{m} \rho(x_{ij} - \sum_{k=1}^{p} f_{cik} a_{cjk} - b_{cj}) \\ &\quad + \lambda \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci} \log u_{ci} \end{aligned} \tag{17}$$

ここで，エントロピー項が標準的な手法におけるメンバシップのべき乗と同様の働きを担い，λ を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．エントロピー正則化には，標準的な方法で必要となる例外処理が不要であるなどの特長があることから，提案法ではエントロピー正則化を採用して定式化を行う．ただし，エントロピー正則化を用いる手法とメンバシップのべき乗を用いる手法とでは，メンバシップ関数を分類関数として用いる場合に性質の相違がある [20]．

(17) 式の目的関数を最小とする解を唯一に求めるために，

$$F_c^\top U_c F_c = I \quad ; \quad c = 1, \cdots, C \tag{18}$$

$$F_c^\top U_c \mathbf{1}_n = \mathbf{0} \quad ; \quad c = 1, \cdots, C \tag{19}$$

$$\sum_{c=1}^{C} u_{ci} = 1 \quad ; \quad i = 1, \cdots, n \tag{20}$$

および $A_c^\top A_c$ が対角行列であるという条件を付加する．（文献 [15] では (19) 式の代わりに $F_c^\top \mathbf{1}_n = \mathbf{0}$ を用いているが，(7) 式が FCV 法での更新式に一致するための制約を考慮して，本論文では (19) 式を用いることとする．）

(17) 式の目的関数では非線形の ρ 関数によりクラスタリング基準が変換されているために，固有値問題に帰着させることができない．そこで，クラスタごとに $(n \times m)$ の重み行列 $W_c = (w_{cij})$ を考え，IRLS 法の原理に従って最適解を求める．W_c の要素 w_{cij} は近似誤差

$$e_{cij} = x_{ij} - \sum_{k=1}^{p} f_{cik} a_{cjk} - b_{cj} \tag{21}$$

に対する重みを表し，Geman-McClure の ρ 関数を用いる場合，

$$\begin{aligned} w_{cij} &= \frac{1}{e_{cij}} \cdot \frac{\partial \rho(e_{cij})}{\partial e_{cij}} \\ &= \frac{2\sigma_j^2}{(e_{cij}^2 + \sigma_j^2)^2} \end{aligned} \tag{22}$$

と定められる．このとき，(17) 式の目的関数の代わりに，

$$\begin{aligned} L_{rfcv'} &= \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci} \sum_{j=1}^{m} w_{cij} e_{cij}^2 \\ &\quad + \lambda \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci} \log u_{ci} \end{aligned} \tag{23}$$

を用いたとしても，

$$\begin{aligned} \frac{1}{2}\frac{\partial L_{rfcv'}}{\partial e_{cij}} &= u_{ci} w_{cij} e_{cij} \\ &\cong \frac{2u_{ci}\sigma_j^2 e_{cij}}{(e_{cij}^2 + \sigma_j^2)^2} = \frac{\partial L_{rfcv}}{\partial e_{cij}} \end{aligned} \tag{24}$$

となるために，w_{cij} を暫時定数として扱っても，不動点繰り返しアルゴリズムによって (17) 式を擬似的に最適化することができる．σ_j として大きな値を与えた場合にはすべての重み w_{cij} がほぼ等しい値を持ち，(23) 式は FCV 法の目的関数と同様になる．一方，σ_j が小さい場合は，大きな近似誤差 e_{cij} には小さな重みが割り当てられ，分析において無視されるようになる．

最適な A_c および $\boldsymbol{b}_c$ を求めるために，(23) 式を以下のように書き換える．

$$\begin{aligned} L_{rfcv'} &= \sum_{c=1}^{C}\sum_{j=1}^{m}(\boldsymbol{x}_j - F_c\tilde{\boldsymbol{a}}_{cj} - \mathbf{1}_n b_{cj})^\top U_c W_{cj} \\ &\quad \times(\boldsymbol{x}_j - F_c\tilde{\boldsymbol{a}}_{cj} - \mathbf{1}_n b_{cj}) \\ &\quad +\lambda\sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}\log u_{ci} \end{aligned} \tag{25}$$

ただし，

$$A_c = (\tilde{\boldsymbol{a}}_{c1},\cdots,\tilde{\boldsymbol{a}}_{cm})^\top$$

$$W_{cj} = \mathrm{diag}(w_{c1j},\cdots,w_{cnj})$$

である．最適性の必要条件 $\partial L_{rfcv'}/\partial\tilde{\boldsymbol{a}}_{cj} = \mathbf{0}$ および $\partial L_{rfcv'}/\partial b_{cj} = 0$ から，

$$\tilde{\boldsymbol{a}}_{cj} = (F_c^\top U_c W_{cj} F_c)^{-1} F_c^\top U_c W_{cj}(\boldsymbol{x}_j - \mathbf{1}_n b_{cj}) \tag{26}$$

$$b_{cj} = (\mathbf{1}_n^\top U_c W_{cj}\mathbf{1}_n)^{-1}\mathbf{1}_n^\top U_c W_{cj}(\boldsymbol{x}_j - F_c\tilde{\boldsymbol{a}}_{cj}) \tag{27}$$

が求まる．

同様に，最適な F_c および u_{ci} を求める際には，(23) 式を

$$\begin{aligned} L_{rfcv'} &= \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}(\tilde{\boldsymbol{x}}_i - A_c\tilde{\boldsymbol{f}}_{ci} - \boldsymbol{b}_c)^\top \tilde{W}_{ci} \\ &\quad \times(\tilde{\boldsymbol{x}}_i - A_c\tilde{\boldsymbol{f}}_{ci} - \boldsymbol{b}_c) \\ &\quad +\lambda\sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}\log u_{ci} \end{aligned} \tag{28}$$

のように書き換えることにより，$\partial L_{rfcv'}/\partial\tilde{\boldsymbol{f}}_{ci} = \mathbf{0}$ および $\partial L_{rfcv'}/\partial u_{ci} = 0$ を満たすパラメータは

$$\tilde{\boldsymbol{f}}_{ci} = (A_c^\top\tilde{W}_{ci}A_c)^{-1}A_c^\top\tilde{W}_{ci}(\tilde{\boldsymbol{x}}_i - \boldsymbol{b}_c) \tag{29}$$

$$\begin{aligned} u_{ci} &= \exp\Big\{-(\tilde{\boldsymbol{x}}_i - A_c\tilde{\boldsymbol{f}}_{ci} - \boldsymbol{b}_c)^\top\tilde{W}_{ci} \\ &\quad \times(\tilde{\boldsymbol{x}}_i - A_c\tilde{\boldsymbol{f}}_{ci} - \boldsymbol{b}_c)/\lambda - 1\Big\} \end{aligned} \tag{30}$$

と求められる．ただし，

$$\tilde{W}_{ci} = \mathrm{diag}(w_{ci1},\cdots,w_{cim}) \tag{31}$$

である．

以上の更新則を用いた繰り返しアルゴリズムは，次のとおりである．

Robust Fuzzy c-Varieties (Robust FCV) Algorithm

Step 1　乱数を用いて $U_c, A_c, \boldsymbol{b}_c, F_c$ を初期化し，(18)〜(20) 式および $A_c^\top A_c$ が対角行列となる制約条件を満たすように基準化する．

Step 2　W_c の初期値を求める．

Step 3　(26) 式により A_c を更新し，$A_c^\top A_c$ が対角行列となるように変換する．

Step 4　(29) 式により F_c を更新し，(18) 式および (19) 式の制約条件を満たすように基準化する．

Step 5　(27) 式により $\boldsymbol{b}_c$ を更新する．

Step 6　(30) 式により U_c を更新し，(20) 式を満たすように基準化する．

Step 7　メンバシップの収束判定条件

$$\max_{c,i} \mid u_{ci}^{NEW} - u_{ci}^{OLD} \mid < \epsilon_1,$$

を満たせば Step 8 へ．それ以外は Step 3 へ戻る．

Step 8　(22) 式により重み W_c を更新する．収束判定条件

$$\max_{c,i,j} \mid w_{cij}^{NEW} - w_{cij}^{OLD} \mid < \epsilon_2,$$

を満たせば終了．それ以外は Step 3 へ戻る．

提案法はすべての w_{cij} が等しいときは FCV 法と同様のデータ分割を与えるため，アニーリングの初期段階は FCV 法で初期値を定めているとみなすことができる．

3　局所的な線形モデルを用いた欠測値の推定と協調フィルタリングへの応用

膨大な量の情報の中からユーザの目的にあった情報を抽出する技術として情報フィルタリング [21] が注目されている．情報の選別法としては，情報そのものの内容とユーザのニーズとを比較して関連するものを推薦する方法などが考案されているが，自然言語の処理を通して情報の持つ内容を自動的に理解するシステムを構築することは容易ではない．そこで，ユーザ同士が協力し合いながら自分の選択した情報の印象を記録していくことにより，他のユーザが情報を選別するのを助ける協調フィルタリングシステムが提案されており，インターネット上の書店や CD ショップなどでの実用化が進められている．そこで用いられるデータは多くのユーザが様々なコンテンツを評価した値の行列であり，ユーザ間の評価値の比較により，未評価の情報に対する嗜好の度合いを予測する．したがって，他者との関係に基づいて情報の選別を行う協調フィルタリングは，データ行列中の欠測値を推定する問題とみなすことができる [22]．システムは未評価の選択肢の中から大きな予測値をもつものを対応するユーザに推薦することにより，情報の選別を行う．

代表的な協調フィルタリング手法である GroupLens [23] では，Peason 相関係数を用いたユーザ間の類似度による重み付け平均値

$$p_{ij} = \bar{x}_i + \frac{\sum_{u=1}^{n}(x_{uj} - \bar{x}_u) \times \omega_{iu}}{\sum_{u=1}^{n} \omega_{iu}} \tag{32}$$

により欠測値 x_{ij} の推定値 p_{ij} を求める方法を提案している．ただし，$\bar{x}_i$ は第 i ユーザの評価値の平均値であり，ω_{iu} は第 i ユーザと第 u ユーザの相関係数である．このように相関係数を用いて嗜好の類似した（もしくは逆の傾向を示す）ユーザからなる近傍を選択し，近傍に含まれるユーザの平均値として推定値を算出していることから，GroupLens は近傍に基づく手法とよばれている．近傍に基づく手法は，各ユーザが共通に評価したコンテンツが少なく信頼性の高い相関係数を求められない場合は利用できないが，十分なデータが収集されているならば全体の平均値を用いる場合（(32) 式ですべての重み ω_{iu} を 1 とおいた場合）よりも高い推薦能力を持つことが知られている [22]．しかし，予測値の算出のためには相関係数行列のすべての要素を保持する必要があり，メモリ容量が少ない場合には実装が困難となる．

そこで，本多ら [16] はデータの変動を良く表す局所的な線形モデルを用いて予測を行うことにより，少ないメモリ容量でも従来法と同程度の推薦能力を有するシステムが構築できることを示している．不完全なデータ行列から，欠測値のみを無視しながら局所的な線形モデルを推定する際には，最小 2 乗基準に基づく FCV 法の目的関数に，データ x_{ij} が観測されていれば 1，欠測値ならば 0 を持つ 2 値変数 d_{ij} を導入し，エントロピー正則化を考慮することで，

$$\begin{aligned} L_{fcvm} \quad = \quad & \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci} \sum_{j=1}^{m} d_{ij}(x_{ij} - \sum_{k=1}^{p} f_{cik} a_{cjk} - b_{cj})^2 \\ & + \lambda \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci} \log u_{ci} \end{aligned} \tag{33}$$

なる目的関数の最小化が図られる [15]．

本節では，観測の有無に関する事前知識が得られた場合に，重みを適切に定めて欠測値を処理することにより，不完全データからノイズの影響を無視した局所的な線形モデルを抽出し，欠測値を推定することを考える．観測の有無に応じて重み w_{cij} を，

$$w_{cij} = \begin{cases} \frac{2\sigma_j^2}{(e_{cij}^2+\sigma_j^2)^2} & ; x_{ij} \text{ is observed.} \\ 0 & ; x_{ij} \text{ is missing.} \end{cases} \tag{34}$$

のように定めることにより，データ行列を要素ごとに低階数近似する際に，観測値に対応する要素についてはなるべく誤差を小さくするものの，欠測値に対応する要素については流れに任せたパラメータ推定が行われる．ここで，観測値に対応するすべての重み w_{cij} を 1 に固定した場合は，提案法は文献 [15] の手法と等しくなる．

いったん，局所的な線形モデルが推定されると，クラスタごとの近似行列 Y_c には欠落がないことから，X の欠測値に対応する Y_c の要素を用いて予測が可能となる．これは，欠測値を含むデータ点がプロトタイプとなる線形多様体上か，もしくはプロトタイプに最も近い点に存在すると仮定して欠測値を推定することに相当し，観測値の持つ情報を活かした推定法となっている．予測の際には，まず，メンバシップが最大となるクラスタを当該ユーザの所属クラスタとし，そのクラスタの近似行列中の値を予測値として用いる．

また，モデルの推定に用いられなかった新たな第 τ ユーザが与えられた場合には，(30) 式をファジィ分類関数 [20] として用いることにより所属クラスタを決定し，(22) 式により $w_{c\tau j}$ を，(29) 式により $\tilde{\boldsymbol{f}}_{c\tau}$ を計算することにより，予測値 $y_{c\tau j}$ を，

$$y_{c\tau j} = \sum_{k=1}^{p} f_{c\tau k} a_{cjk} + b_{cj} \tag{35}$$

のように推定することができる．

このように，局所的な線形モデルを用いる協調フィルタリングシステムは，線形モデルがいったん求まればクラスタごとにファジィ分類関数およびパラメータの更新式を保持するだけで新たなユーザに対しても予測値が計算できるので，メモリの所要量の少ない効率的な手法であるといえる．また，所属クラスタの決定を近傍の選択，線形モデルによる予測を近傍に含まれるユーザの評価値による予測ととらえると，提案手法を用いた協調フィルタリングシステムは近傍に基づく手法に含まれる．

4 数値実験

本章では，提案手法の特長を確かめるために人工データを用いた数値例を示した後に，協調フィルタリングシステムとしての性能を評価する．

4.1 人工データを用いた数値実験

3 次元空間で 2 本の直線状に分布するデータ集合を用いて提案法の特長を確かめる数値実験を行った．ノイズを含まないデータ集合に $\theta = 2$ として FCV 法を適用した結果，24 個の標本データは図 1 中で●と○で示されるクラスタに分割され，それぞれ破線で示すプロトタイプが得られた．以下では，データ集合にノイズや欠測値が含まれる場合でも同じプロトタイプが得られるかどうかを検証する．

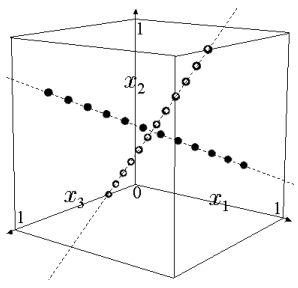


図 1: 3-D plots of original data set

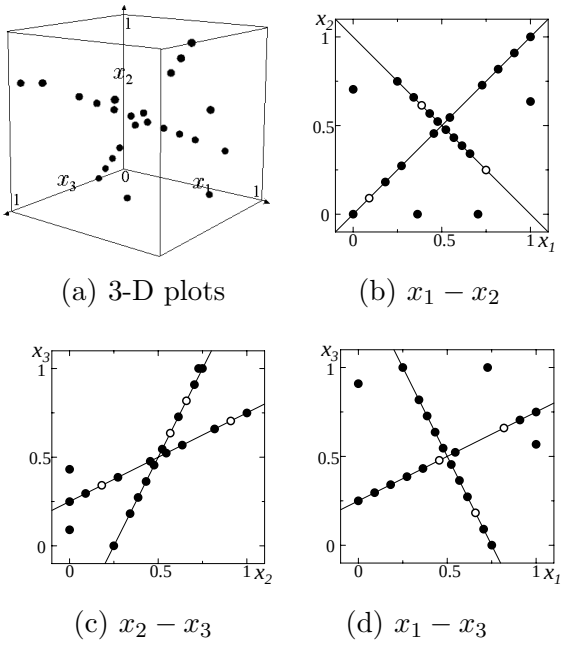


図 2: 3-D plots and 2-D projections of noisy data set (21%)

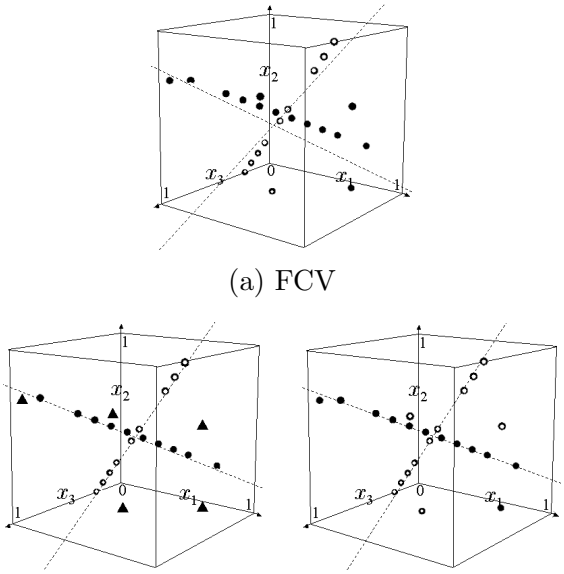
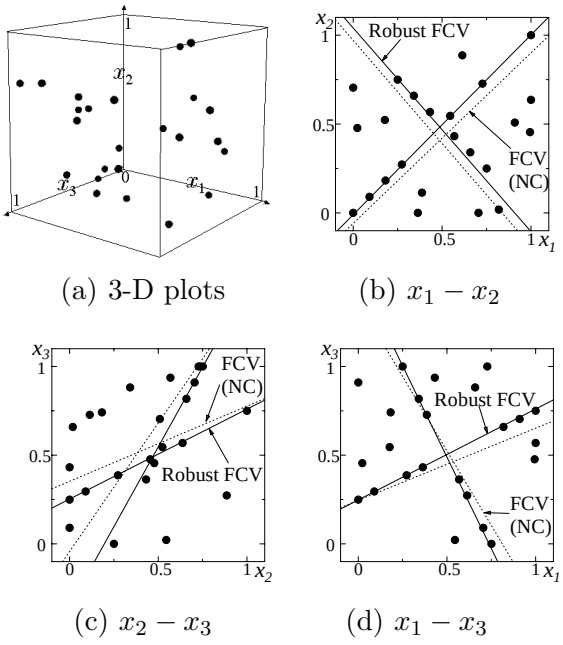


図 3: Clustering result with noisy data set (21%)

まず，ノイズに対するロバストさを検証するために，24 個のうちから 5 個のデータ点を選び，それぞれ一つの要素をノイズに置き換えることにより，21%のデータ点がノイズを含むデータ集合を作成した．ノイズを含むデータ集合の散布図および 2 次元平面への射影を図 2 に示す．ここで，2 次元平面への射影 (b)〜(d) の直線は元のデータが乗る直線を表しており，ノイズを含むサンプルも (b)〜(d) のいずれかの図中では直線に乗っている．比較のために，提案法のほかに，FCV 法および FCV 法にノイズクラスタリングを導入したアルゴリズムを適用した．三つの手法で得られた分割結果およびプロトタイプを図 3 に示す．ただし，提案法では $\lambda = 0.05$ とした．ノイズの範囲を決めるパラメータ σ_j はすべての変量について共通とし，繰り返し回数 t を用いて，

$$\sigma_j^2 = \frac{0.5}{\log(t+2)} \tag{36}$$

のようにアニーリングを行った．ノイズの影響のために FCV 法では 2 本の直線が正しくとらえられていないのに対して，ノイズクラスタリングと提案法ではノイズが含まれない場合とほぼ等しいプロ



(a) 3-D plots (b) $x_1 - x_2$

(c) $x_2 - x_3$ (d) $x_1 - x_3$

図 4: 3-D plots and 2-D projections of noisy data set (67%)

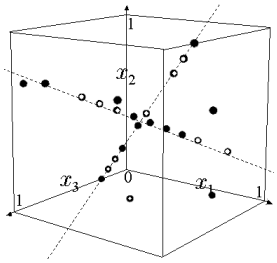


図 5: Clustering result with incomplete data set

トタイプが得られていることが分かる．ここで，ノイズクラスタリングでは▲で示されたノイズを含むサンプルの要素すべてを無視してクラスタリングが行われているのに対して，提案法ではノイズ要素のみを無視して，ノイズを含むサンプルもおのおののクラスタに分類されている．このように，データ行列に少数のノイズ要素しか含まれない場合には，提案法とノイズクラスタリングは同様のプロトタイプを与えるといえる．

つぎに，ノイズの割合を増やして同様の実験を行った．用いたデータは，67%のサンプルデータが一つずつのノイズを含む図 4 の (a) のデータである．ノイズサンプルの割合が大きいために，散布図からは直線構造がとらえられないが，前記の例と同様に，2 次元平面への射影を考えると，ノイズサンプルもいずれかの平面で直線に乗っている．得られたプロトタイプを，図 4 の (b)〜(d) の 2 次元平面への射影に重ねて示す．ただし，実線が Robust FCV 法で得られたプロトタイプを，破線がノイズクラスタリングを導入した FCV 法で得られたプロトタイプを表す．ノイズサンプルの割合が大きく，それらを無視した分析が困難となっているために，ノイズクラスタリングでも直線がとらえられなかったが，提案法ではノイズ要素のみを無視しているために，ノイズの影響を無視した分析ができている．

さらに，ノイズを含むデータ行列からいくらかの情報を欠落させた場合の実験を行った．データは図 2 の○で表された 10 個のサンプルデータからおのおの一つの要素を欠落させて作成した．たとえば，$x_1 - x_2$ 平面への射影における○は，そのサンプルデータの x_3 要素が欠測値となっていることを表している．図 5 に得られたプロトタイプを示す．ただし，欠測値を含むサンプルデータを○で表しており，従来法を適用する場合ではそれらをすべて取り除いて分析しなければならないため，望ましい結果が得られない例となっている．欠測値のみを無視し，観測された要素のすべてを利用する提案法では，欠測値やノイズの影響を受けずに，元のデータを用いた場合とほぼ等しい直線構造をとらえられている．以上のように，提案法がサンプル内のノイズや欠測値の影響を無視しながら局所的な線形構造を抽出するのに有効な手法であることが分かる．

4.2 協調フィルタリングへの応用

つづいて，MovieLens ウェブサイト [24] で収集された映画評価データを用いて，提案手法で得られる局所的な線形モデルによる協調フィルタリングシステムの性能を比較する実験を行った．用いたデータは，943 人のユーザが 1682 種類の映画の中から各自が見た映画に対して 5 段階（1〜5 点）評価したデータで，100,000 件の評価値からなる．ここで，各ユーザは少なくとも 20 種類以上の映画に対して評価を行っている．この 100,000 件のデータを 80,000 件のトレーニングデータと 20,000 件のテストデータに分け，トレーニングデータを用いて作成した予測モデルを用いてテストデータを予測する実験を行った．ただし，提案法の予測精度を正しく評価するために，3 人以下のユーザしか評価していない映画を除いて 1240 種類の映画について評価値の予測実験を行った．他者との相関を考慮せずに全体の平均値を用いる方法，(32) 式により予測を行う GroupLens および文献 [16] の不完全データのための FCV 法を用いる場合との比較の結果を表 1 に示す．クラスタリングでは 1 本の主成分ベクトルを用いてユーザを二つのクラスタに分割した．また，ファジィ度は $\lambda = 6.0$ とし，σ_j は以下により与えた．

$$\sigma_j^2 = \frac{5.0}{\log(t+2)} \tag{37}$$

表 1: Comparison of prediction algorithms

Algorithm	**MAE**	**ROC**
Non-personalized Method	0.821	0.714
GroupLens	0.762	0.762
FCV with Missing Values	0.754	0.777
Robust FCV	0.751	0.789

性能の比較には，予測値 y_{cij} と評価値 x_{ij} との絶対誤差 $|y_{cij} - x_{ij}|$ の平均値（平均絶対誤差：MAE）の他に，推薦システムとしての性能を評価するための ROC(Receiver Operating Characteristic) 感度 [25] を用いた．ROC 感度は推薦されるべきアイテムが推薦システムから正しく推薦される割合を表しており，ここではテストデータのうち 4 以上の評価値を持つアイテムにシステムが 3.5 以上の予測値を与えた割合とした．比較の結果，提案法が最小の MAE を持つとともに ROC 感度も最大となっており，予測の際のメモリ所要量が従来法よりも少ないながら，従来法以上の推薦能力をもつシステムを構築できることが分かった．

なお，本実験においては，クラスタの数および主成分ベクトルの数を増やして分析を行っても，MAE および ROC ともに改善は見られなかった．これは，ここで用いたデータでは映画に対する嗜好に極端な偏りがなく，少数の主成分で全体の傾向をとらえることができたためと考えられる．ただし，クラスタや主成分ベクトルの最適な数はデータに依存するものであり，実装に際しては，cross-validation 法などにより決定する必要がある．

5 おわりに

本論文では，サンプル内のノイズを含むデータからロバストな局所的線形モデルを抽出する手法を提案した．提案法では，最小 2 乗基準を用いた線形ファジィクラスタリングの目的関数において，要素ごとにノイズであるか否かを表す重みを導入することにより，FCM 法と同様の繰り返しアルゴリズムを用いた最適化が行われる．要素ごとにノイズか否かを判定する提案法は，ノイズを含むサンプルすべてを分析対象から取り除いてしまうノイズクラスタリングや可能性的クラスタリングに比較して，情報の損失が少ない手法であり，要素ごとのノイズが含まれやすい多変量データの分析において有効である．また，観測の有無について事前知識が得られる場合には，欠測値に対応する重みを 0 に固定することにより，不完全データに対して欠測値の影響を無視した分析を施すことができる．得られた局所的な線形モデルはデータの変動を良く表すモデルとなっていることから，欠測値の推定においても有効であり，協調フィルタリングへの応用の実験においては，より簡潔な予測モデルを用いても従来法以上のコンテンツ推薦能力を持つフィルタリングシステムが開発できることを示した．ただし，分析の初期段階で用いる FCV 法では初期分割に依存して異なる結果が得られることがあるため，いくつかの初期分割で分析を繰り返すことにより最適な結果を探索することが望ましい．また，

従来法と同様にファジィ度を分析者が決定しなければならないのに加えて，提案法ではロバスト関数の形状を決定する σ を適正に与えなければノイズでない要素をもノイズとみなして無視することにより，本来とは異なるクラスタに分類されるサンプルが生じることがある．ノイズクラスタリングにおけるノイズの範囲を決定するパラメータと同様に，これらのパラメータを定める基準を考案することは今後の課題である．

参考文献

[1] F. de la Torre and M. J. Black: Robust principal component analysis for computer vision; *Proc. of International Conference on Computer Vision*, pp. 362-369 (2001)

[2] F. de la Torre and M. J. Black: A framework for robust subspace learning; *International Journal of Computer Vision*, Vol. 54, pp. 117-142 (2003)

[3] G. E. Hinton, P. Dayan and M. Revow: Modeling the manifolds of images of handwritten digits; *IEEE Trans. on Neural Networks*, Vol. 8, No. 1, pp. 65-74 (1997)

[4] N. Kambhatla and T. K. Leen: Dimension reduction by local principal component analysis; *Neural Computation*, Vol. 9, No. 7, pp. 1493-1516 (1997)

[5] M. E. Tipping and C. M. Bishop: Mixtures of probabilistic principal component analysers; *Neural Computation*, Vol. 11, No. 2, pp. 443-482 (1999)

[6] J. C. Bezdek: *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press (1981)

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

[8] T. Hastie and W. Stuetzle: Principal curves; *J. of American Statistical Association*, Vol. 84, pp. 502-516 (1989)

[9] M. A. Kramer: Nonlinear principal component analysis using autoassociative neural networks; *AIChE Journal*, Vol. 37, No. 2, pp. 233–243 (1991)

[10] E. Oja: Data compression, feature extraction, and autoassociation in feedforward neural networks; *Artificial Neural Networks*, Elsevier Science Publishers, pp.737-745 (1991)

[11] R. N. Dave: Characterization and detection of noise in clustering; *Pattern Recognition Letters*, Vol.12, No.11, pp. 657-664 (1991)

[12] R. Krishnapuram and J. M. Keller: A possibilistic approach to clustering; *IEEE Trans. on Fuzzy Systems*, Vol.1, pp. 98-110 (1993)

[13] 本多, 東江, 市橋: 最小絶対誤差に基づく線形ファジークラスタリング; 電子情報通信学会論文誌 (D-II), Vol. J86-D-II, No. 1, pp. 12-21 (2003)

[14] P. Whittle: On principal components and least square methods of factor analysis; *Skand. Akt.*, Vol.35, pp. 223-239 (1952)

[15] 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)

[16] 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)

[17] P. W. Holland and R. E. Welsch: Robust regression using iteratively reweighted least-squares; *Communications in Statistics*, Vol. A6, No. 9, pp. 813-827 (1977)

[18] 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)

[19] S. Geman and D. E. McClure: Statistical methods for tomographic image reconstruction; *Bulletin of International Statistical Institute*, Vol.LII-4, pp.5-21 (1987)

[20] 宮本, 馬屋原, 向殿: ファジィ c-平均法とエントロピー正則化法におけるファジィ分類関数; 日本ファジィ学会誌, Vol.10, No.3, pp.156-164 (1998)

[21] 森田, 速水: 情報フィルタリングシステム -情報洪水への処方箋-; 情報処理, Vol.37, No.8, pp.751-757 (1996)

[22] J. L. Herlocker, J. A. Konstan, A. Borchers and J. Riedl: An algorithmic framework for performing collaborative filtering; *Proc. of Conference on Research and Development in Information Retrieval* (1999)

[23] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gardon and J. Riedl: Grouplens: Applying collaborative filtering to usenet news; *Communications of the ACM*, Vol.40, No.3, pp.77–87 (1997)

[24] MovieLens Web Page; http://www.movielens.org/

[25] J. A. Swets: Measuring the accuracy of diagnostic systems; *Science*, Vol. 240, No. 4857, pp. 1285–1289 (1988)