逐次学習を伴う線形ファジィクラスタリングによる 適応的な協調フィルタリング*

本多 克宏[†]・市橋 秀友[†]・野津 亮[†]

Adaptive Collaborative Filtering Using Linear Fuzzy Clustering with Sequential Learning*

Katsuhiko HONDA[†], Hidetomo ICHIHASHI[†] and Akira NOTSU[†]

Collaborative filtering is a technique for reducing information overload, and personalized recommendation is performed by predicting missing values in a data matrix. Linear fuzzy clustering is a technique for local principal component analysis and can be used for constructing local prediction models considering data substructures. This paper proposes a new algorithm for constructing local linear models that performs a simultaneous application of fuzzy clustering and principal component analysis based on sequential subspace learning. In numerical experiments, the diagnostic power of the filtering system is shown to be improved by predicting missing values using the proposed local linear models.

1. はじめに

ユーザやコンテンツ間の関連をもとに個人に特化した推奨コンテンツを探索する協調フィルタリング [1,2] は、情報過多に対する有効な対処法として広く用いられている。協調フィルタリングでは、ユーザが多くのコンテンツを評価した評価値行列が与えられた際に、未評価のコンテンツのうち高い評価値が得られると期待されるものを探索して推薦することが目的となる。したがって、評価値行列中の未評価値（欠測値）を予測する問題ともとらえられる [3]。未評価値の予測法には、大きく分けて Memory-based アルゴリズムと Model-based アルゴリズムの 2 種類がある [4]。商用 web サイトでの応用も多い Memory-based アルゴリズムは近傍に基づく手法の一種で、蓄積された評価値データから推薦対象と類似したユーザ（またはコンテンツ）を探索し、それらの重み付き平均として未評価値を推定する。事前のモデル学習を必要としない Memory-based アルゴリズムはシンプルでかつ学習用データが少ない場合に有用な手法であるが、蓄積されたすべてのデータを保持する必要があるため、推薦システムのメモリの所要量が多いという欠点もある。一

方、Model-based アルゴリズムでは、ユーザやコンテンツ間の関連を少数の予測モデルで代表させることにより、簡便な推薦システムの構築を可能とする。

線形ファジィクラスタリングで得られる局所的な線形モデルを協調フィルタリングにおける未評価値の予測問題に適用する研究がなされている [5,6]。線形ファジィクラスタリングはデータをいくつかの線形多様体状のクラスタに分割する手法であるが、各クラスタの代表であるプロトタイプとして線形多様体を用い、プロトタイプからの距離の総和が最小となるようにデータ分割を行う一方で、各クラスタのプロトタイプを張るベクトルが局所領域における主成分ベクトルとみなされることから、局所的な主成分分析法の一種であるといえる。本多らは、線形多様体とデータ点との距離からなる Fuzzy *c*-Varieties (FCV) 法 [7] のクラスタリング基準を要素ごと近似に基づく最小 2 乗基準で定義することにより、欠測値の取り扱いを可能とし [8]、近似行列の要素を用いた未評価値の推定法を提案している [5,6]。多次元データの構造を単一の線形モデルで近似することが困難な場合でも、クラスタリングによりデータ領域を分割し複数の予測モデルを併用することで Memory-based アルゴリズムと同等の予測精度を得ている。また、コンテンツの数（データ次元）に比してクラスタ数は少数に設定されることから、推薦システムのメモリの所要量が少ないという利点がある。

近年、ユーザの嗜好の動的な変化を考慮して推奨コンテンツを探索する研究がいくつかなされている。従

* 原稿受付 2006 年 8 月 29 日

[†] 大阪府立大学 大学院 工学研究科 Graduate School of Engineering, Osaka Prefecture University; 1-1 Gakuencho, Nakaku, Sakai, Osaka 599-8531, JAPAN

Key Words: collaborative filtering, fuzzy clustering, sequential learning.

来の協調フィルタリングシステムでは蓄積されたすべてのデータを同等に扱っていたが、実世界の問題ではユーザの嗜好が時間とともに変化することも多い。Dingら[9,10]は、Memory-based アルゴリズムにおける未評価値の推定に際して評価してから時間の重みを考慮することで、未来の事象の予測に適した予測モデルを提案している。また、小澤ら[11]は日時情報から未評価値の推定に用いる期間を制限する方法を提案し、携帯電話向け着信メロディのダウンロード履歴に適用している。そこで本研究では、線形ファジィクラスタリングによる協調フィルタリングシステムにおいて、局所的な線形モデルをユーザの嗜好の変化に動的に適応させる手法を提案する。「評価データが1件ずつ逐次的に得られる」という問題の特性を考慮して逐次学習アルゴリズムを開発し、ユーザの嗜好の動的な変化に対応したコンテンツ推薦システムを提案する。

以下では、2章で既存のクラスタリング手法とその協調フィルタリングへの応用を概説した後に、3章で逐次学習に基づく適応的な協調フィルタリング手法を提案する。4章では数値実験を通して提案手法の特徴を示し、5章にまとめを述べる。

2. 線形ファジィクラスタリングを用いた協調フィルタリング

n 人のユーザが m 個のコンテンツを評価した ($n \times m$) データ行列 $X = \{x_{ij}\}$ が与えられたときに、 n 人のユーザを C 個のクラスタに分割しながら、クラスタごとにデータの特徴をよく表す線形モデルを推定し、得られた線形モデルを用いて未評価値を推定する問題を考える。

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

Bezdek らの Fuzzy c -Varieties (FCV) 法[7]は互いに独立な p 本の単位ベクトル $\mathbf{a}_{c1}, \dots, \mathbf{a}_{cp}$ により張られる線形多様体を第 c クラスタのプロトタイプとし、プロトタイプからの距離の総和が最小となるクラスタ分割を得る手法であり、以下の目的関数の最小化問題として定式化される。

$$L_{fcv} = \sum_{c=1}^C \sum_{i=1}^n u_{ci}^\theta \left\{ (\tilde{\mathbf{x}}_i - \mathbf{b}_c)^\top (\tilde{\mathbf{x}}_i - \mathbf{b}_c) - \sum_{k=1}^p \mathbf{a}_{ck}^\top R_{ci} \mathbf{a}_{ck} \right\} \quad (1)$$

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

ここで、 $\tilde{\mathbf{x}}_i$ は第 i ユーザの評価値を並べたベクトルであり、 $\mathbf{b}_c$ は第 c クラスタの中心を表す。(1) 式の $\{\}$ 内が第 i ユーザと第 c クラスタのプロトタイプとの距離を表すクラスタリング基準である。 u_{ci} は第 i ユーザの第 c クラスタへの所属度合いを表すメンバシップで、

$$\sum_{c=1}^C u_{ci} = 1 \quad ; i = 1, \dots, n \quad (3)$$

を満たすものとする。(3) 式によりメンバシップを混合分布モデルにおける事後確率と同様に解釈できることから、確率的制約に基づくデータ分割とよばれる[12]。 θ はメンバシップのファジィ度を定める定数である。 $\theta = 1$ のときはクリスプなデータ分割を与え、 θ が大きくなるにつれてデータの所属があいまいとなるファジィな分割になる。メンバシップとプロトタイプの交互最適化により、(局所的な)最適解が探索される[13]。メンバシップとクラスタ中心を固定して考えた場合、(1) 式の最小化問題は、一般化されたメンバシップの重み付きファジィ散布行列

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

の大きい p 個の固有値と対応する固有ベクトルを求める問題に帰着されることから、 $\mathbf{a}_{ck}$ は局所的な主成分ベクトルとみなされる[14]。

また、FCV 法を潜在変量の抽出に基づく局所的な主成分分析の一種とみなす研究もなされている。クラスタごとに p 個の潜在変量を用いてデータ行列を低階数近似する問題ととらえることにより、(1) 式は、

$$L_{fcv} = \sum_{c=1}^C \text{tr} \left\{ (X - Y_c)^\top U_c^\theta (X - Y_c) \right\} \quad (5)$$

のように書き換えられる[5,8]。ただし、 U_c は $(n \times n)$ 対角行列 $U_c = \text{diag}(u_{c1}, \dots, u_{cn})$ であり、 tr は行列の対角要素の和 (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 の低階数近似行列であり、主成分得点からなる $(n \times p)$ 成分得点行列 $F_c = \{f_{cij}\}$ と p 本の局所的な主成分ベクトルを並べた $(m \times p)$ 主成分行列 $A_c = (\mathbf{a}_{c1}, \dots, \mathbf{a}_{cp})$ を用いて、 $Y_c = F_c A_c^\top + \mathbf{1}_n \mathbf{b}_c^\top$ のように表されるものとする。ただし、 $\mathbf{1}_n$ はすべての要素が1の n 次元ベクトルである。FCV 法と同様にファジィ散布行列の固有値問題を解くことで主成分ベクトルを求める方法のほか、交互最小2乗法に基づく主成分分析[15]と同様に交互最適化により局所的な主成分を求められる。ただし、交互最適化による場合は、回転の任意性が残る。

要素ごとの近似としての定式化を利用して、本多ら[5,6,8]はデータ行列中の要素ごとの欠測値やサンプル内ノイズを処理する手法を提案している。有意な観測値に対応する要素については2乗誤差を最小にするようにし、欠測値やノイズに対応する要素については他の観測値の影響に任せることにより、欠測値やノイズの影響を無視しながら線形クラスタを抽出するための目的関数を以下のように定式化している。

$$L_{rfce} = \sum_{c=1}^C \sum_{i=1}^n u_{ci}^\theta \sum_{j=1}^m w_{cij} \left\{ \alpha \left(x_{ij} - \sum_{k=1}^p f_{cik} a_{ckj} - b_{cj} \right)^2 + (1-\alpha)(x_{ij} - b_{cj})^2 \right\} \quad (6)$$

ただし、 w_{cij} はデータ行列 X の i 行 j 列の要素 x_{ij} に対する重みを表す。欠測値の処理が目的の場合は、 x_{ij} が観測されていれば $w_{cij} = 1$ 、欠測値の場合は $w_{cij} = 0$ として、対応する 2 乗誤差が無視される [5,8]。また、第 c クラスタにおいて x_{ij} がサンプル内のノイズとみなされる場合は、 w_{cij} として小さな重みを与えることで局所的な主成分の抽出への影響を抑える [6]。(6) 式の目的関数では、線形多様体からの距離だけでなくクラスタ中心からの距離も考慮するために、クラスタ中心（点）をプロトタイプとする Fuzzy c -Means (FCM) 法 [7] の目的関数との線形結合が用いられており、Fuzzy c -Elliptotypes (FCE) 法 [7] の発展となっている。 α が主成分分析と FCM クラスタリングのトレードオフ係数である。 α が 0 のときは FCM 法と同様の点を中心とするクラスタが得られ、 α が 1 に近づくにつれて局所的な主成分の抽出を重視して、主成分軸の方向に広がりをもつ形状のクラスタとなる。また、 $\alpha = 1$ のときは、(5) 式を要素ごとに書き下して重み w_{cij} を導入したものとなる。ただし、唯一の解を得るために、

$$F_c^\top U_c^\theta F_c = I \quad ; c = 1, \dots, C \quad (7)$$

$$F_c^\top U_c^\theta \mathbf{1}_n = \mathbf{0} \quad ; c = 1, \dots, C \quad (8)$$

$$\sum_{c=1}^C u_{ci} = 1 \quad ; i = 1, \dots, n \quad (9)$$

および、 $A_c^\top A_c$ が対角行列となる制約条件を付加する。

(6) 式の目的関数を最小とするパラメータを同定する際には、標本データの中に要素ごとの欠測値が含まれているために、FCV 法のようにファジィ散布行列の固有値問題に帰着させることができないため、交互最小 2 乗法に基づいて最適化が行われる。最適性の必要条件から、おのおののパラメータの更新則は、以下のように求められる。

$$\tilde{a}_{cj} = (F_c^\top U_c^\theta W_{cj} F_c)^{-1} F_c^\top U_c^\theta W_{cj} (\mathbf{x}_j - \mathbf{1}_n b_{cj}) \quad (10)$$

$$b_{cj} = (\mathbf{1}_n^\top U_c^\theta W_{cj} \mathbf{1}_n)^{-1} \mathbf{1}_n^\top U_c^\theta W_{cj} (\mathbf{x}_j - \alpha F_c \tilde{a}_{cj}) \quad (11)$$

$$\tilde{\mathbf{f}}_{ci} = (A_c^\top \tilde{W}_{ci} A_c)^{-1} A_c^\top \tilde{W}_{ci} (\tilde{\mathbf{x}}_i - \mathbf{b}_c) \quad (12)$$

$$u_{ci} = \left[\sum_{l=1}^C \left(\frac{d_{ci}^2}{d_{li}^2} \right)^{1/(\theta-1)} \right]^{-1} \quad (13)$$

$$d_{ci}^2 = \alpha (\tilde{\mathbf{x}}_i - A_c \tilde{\mathbf{f}}_{ci} - \mathbf{b}_c)^\top \tilde{W}_{ci} (\tilde{\mathbf{x}}_i - A_c \tilde{\mathbf{f}}_{ci} - \mathbf{b}_c) + (1-\alpha) (\tilde{\mathbf{x}}_i - \mathbf{b}_c)^\top \tilde{W}_{ci} (\tilde{\mathbf{x}}_i - \mathbf{b}_c) \quad (14)$$

ただし、 $\mathbf{x}_j = (x_{1j}, \dots, x_{nj})^\top$, $A_c = (\tilde{a}_{c1}, \dots, \tilde{a}_{cm})^\top$, $F_c = (\tilde{\mathbf{f}}_{c1}, \dots, \tilde{\mathbf{f}}_{cn})^\top$, $W_{cj} = \text{diag}(w_{c1j}, \dots, w_{cnj})$ および $\tilde{W}_{ci} = \text{diag}(w_{ci1}, \dots, w_{cim})$ である。

w_{cij} を固定した場合、繰り返しアルゴリズムは以下のように与えられる。

Algorithm: Robust Fuzzy c -Elliptotypes (Robust FCE)

Step 1 クラスタごとに $U_c, A_c, \mathbf{b}_c, F_c$ を乱数により初期化し、(7)-(9) 式の制約条件と $A_c^\top A_c$ が対角行列となる制約を満たすように基準化する。収束判定条件の閾値 ϵ を小さな正数により定める。

Step 2 (10) 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形する。

Step 3 (12) 式により F_c を更新し、(7) 式および (8) 式を満たすように基準化する。

Step 4 (11) 式により $\mathbf{b}_c$ を更新する。

Step 5 (13) 式により U_c を更新する。

Step 6 収束判定条件

$$\max_{c,i} |u_{ci}^{NEW} - u_{ci}^{OLD}| < \epsilon,$$

を満たせば終了。それ以外は Step 2 へ戻る。

2.2 局所的な主成分を用いた未評価値の推定

いったん、局所的な線形モデルが推定されると、それらを用いて未評価値を予測することができる。まず、線形モデルの推定に用いられたユーザについては、メンバーシップが最大となるクラスタにおいて、データの欠測値に対応する低階数近似行列中の要素の値を未評価値の推定値とする。

また、モデルの推定に用いられなかった新たなユーザ ω の未評価値 $x_{\omega\tau}$ についても、(12) 式から観測値のみを考慮して主成分得点 $f_{c\omega k}$ を求め、(13) 式のメンバーシップの更新則をファジィ分類関数として用いて所属するクラスタ c を決定し、

$$y_{c\omega\tau} = \sum_{k=1}^p f_{c\omega k} a_{c\tau k} + b_{c\tau} \quad (15)$$

のように推定できる。予測値の大きいものを推奨コンテンツとして提示することで、情報のフィルタリングが行われる。

3 ユーザの嗜好の変化に適応した予測モデルの同定

3.1 評価時刻を考慮した協調フィルタリング

代表的な協調フィルタリング手法である GroupLens[2] では、Pearson 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

$$p_{ij} = \bar{x}_i + \frac{\sum_{u=1}^n (x_{uj} - \bar{x}_u) \times \omega_{iu}}{\sum_{u=1}^n \omega_{iu}} \quad (16)$$

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。文献 [5,6] の線形ファジィクラスタリングを用いた協調フィルタリング手法は、ユーザの分割を近傍の探索、線形モデルの推定を近傍における予測式の推定とみなすと、GroupLens に類した近傍に基づく手法の一種であるといえる。また、ユーザではなくコンテンツ間の類似度をもとに近傍を探索し、近傍に含まれるコンテンツの平均値として推定値を算出する方法も提案されている [16]。

$$p_{ij} = \frac{\sum_{v=1}^m x_{iv} \times r_{jv}}{\sum_{v=1}^m r_{jv}} \quad (17)$$

ただし、 r_{jv} は第 j コンテンツと第 v コンテンツの相関係数である。ユーザからなる近傍を用いるよりも精度が改善されるという報告もなされている。

しかし、ユーザの嗜好に変化が生じた際には、すべての評価値を同等に扱うのではなく、評価してから時間が短いものを重視する方が将来の事象を予測するという目的にふさわしいといえる。Ding と Li [9] は類似度による重み付け平均値を計算する際に、さらに評価してから時間に基づく重みを加味した予測モデルを提案している。文献 [9] ではコンテンツ間の類似度に基づくモデルが用いられており、以下の予測式となる。

$$p_{ij} = \frac{\sum_{v=1}^m x_{iv} \times r_{jv} \times \psi(t_{iv})}{\sum_{v=1}^m r_{jv} \times \psi(t_{iv})} \quad (18)$$

ただし、 t_{ij} がユーザ i のコンテンツ j を評価した時刻であり、 $\psi(t_{ij})$ が評価してから時間を考慮した重みである。評価してから時間が長いほど重み $\psi(t_{ij})$ を小さくすることで、ユーザの嗜好の変化に適応した予測が可能となる。つまり、特定のユーザの嗜好がある時刻から変化した状況を想定し、変化後にそのユーザが評価した値を重視した予測を行っている。また、ユーザ間の類似度に基づくモデルへの適用も考えることができ、(16) 式は、

$$p_{ij} = \bar{x}_i + \frac{\sum_{u=1}^n (x_{uj} - \bar{x}_u) \times \omega_{iu} \times \psi(t_{uj})}{\sum_{u=1}^n \omega_{iu} \times \psi(t_{uj})} \quad (19)$$

となる。このモデルでは、特定のコンテンツに対する評価値の特性がある時刻から変化した状況を想定し、変化後にそのコンテンツが評価された値を重視した予測を行っている。また、嗜好の変化が激しい場合には信頼性が低いと判断して小さい重みを与えることで将来の事象の予測精度を高める試み [10] や、日時情報から未評価値の推定に用いる期間を制限する方法 [11] も提案されている。ただし、評価時刻に基づく重みを与えるためには、評価値行列と合わせて評価時刻に関する情報も保持する必要があり、計算量とともにメモリの所要量が増大する

という欠点がある。

以下では、線形ファジィクラスタリングに基づく協調フィルタリングシステムにおいてユーザの嗜好の変化に適応した予測モデルを同定することを考える。

3.2 逐次学習に基づくアルゴリズムを用いた予測モデルの更新

協調フィルタリングの実装においては、入力データは必ずしも一括で与えられるとは限らず、むしろ、あるユーザがあるコンテンツを選択した際にその評価値を一つずつ逐次的に入力すると考える方が自然である。さらに、ユーザの嗜好は時間とともに変化する傾向を持つことから、同じコンテンツに関する評価値も評価してから時間が短いものほど重視されるべきであるといえる。そこで、本節では評価値が1件ずつ与えられる場合に、局所的な線形モデルを逐次的に更新するアルゴリズムを提案する。逐次学習に基づくクラスタリング法としては、学習ベクトル量子化 (LVQ) を発展させた LVQ クラスタリング [17] が提案されている。LVQ クラスタリングは点をプロトタイプとするクラスタリングに逐次学習を応用したもので、入力データとして標本が1件ずつ与えられる場合を想定し、最近隣クラスタ中心を逐次的に更新することで入力空間の分割を行っている。ここでは、この考えをさらに発展させ、ユーザの評価値が1件ずつ与えられる場合の更新則を考える。ファジィ分割を行うため、最近隣プロトタイプの探索の代わりにすべてのクラスタへの帰属度を求める。また、評価値1件のみを用いて線形モデルのパラメータ（クラスタ中心、主成分ベクトル、主成分得点）を更新する。

時刻 t ($t=1, 2, \dots$) に第 i ユーザの第 j アイテムに関する評価値 x_{ij} が1件だけ与えられるものとし、逐次的に局所的な線形モデルを更新する。以下では、クラスタのプロトタイプが直線 ($p=1$) の場合、すなわち、Fuzzy c -Lines (FCL) 法 [7] の発展を考える。(6) 式の目的関数の中で、おのおのの時刻において考慮すべき（部分的な）目的関数は以下のように与えられる。

$$L_{f_{cem}}^{(ij)} = \sum_{c=1}^C u_{ci}^0 \left\{ \alpha (x_{ij} - f_{ci} a_{cj} - b_{cj})^2 + (1 - \alpha) (x_{ij} - b_{cj})^2 \right\} \quad (20)$$

ここで、 f_{ci}, a_{cj}, b_{cj} は線形モデルのパラメータであり、 u_{ci} はメンバシップである。これらのパラメータを LVQ [18] の枠組みで更新するために、以下の更新モデルを用いる。いま、ある未知パラメータ $\phi \in \{f_{ci}, a_{cj}, b_{cj}\}$ を時刻 t に

$$\phi(t+1) = \phi(t) + \gamma (\hat{\phi}(t) - \phi(t)) \quad (21)$$

のように更新するものとする。ただし、 $\hat{\phi}(t)$ が (20) 式の部分的な目的関数を最小化する値とし、学習率 γ により更新の幅を定める。

おのおののパラメータについて、部分的な目的関数を

最小化する値は、

$$\frac{\partial L_{f_{cem}}^{(ij)}}{\partial b_{cj}} = -2u_{ci}^\theta (x_{ij} - \alpha f_{ci} a_{cj} - b_{cj}) = 0 \quad (22)$$

$$\frac{\partial L_{f_{cem}}^{(ij)}}{\partial a_{cj}} = -2\alpha u_{ci}^\theta f_{ci} (x_{ij} - f_{ci} a_{cj} - b_{cj}) = 0 \quad (23)$$

$$\frac{\partial L_{f_{cem}}^{(ij)}}{\partial f_{ci}} = -2\alpha u_{ci}^\theta a_{cj} (x_{ij} - f_{ci} a_{cj} - b_{cj}) = 0 \quad (24)$$

なる条件から求められ、以下の更新則にまとめられる。

$$b_{cj} = b_{cj} + u_{ci}^\theta \gamma (x_{ij} - \alpha f_{ci} a_{cj} - b_{cj}) \quad (25)$$

$$a_{cj} = a_{cj} + u_{ci}^\theta \gamma \left(\frac{1}{f_{ci}} (x_{ij} - b_{cj}) - a_{cj} \right) \quad (26)$$

$$f_{ci} = f_{ci} + \gamma \left(\frac{1}{a_{cj}} (x_{ij} - b_{cj}) - f_{ci} \right) \quad (27)$$

ここで、 b_{cj} および a_{cj} について、学習率に u_{ci}^θ をかけたのは、文献 [5,8] ではファジィ分割を考慮しながらパラメータが更新されていたためである。 f_{ci} については、最小2乗近似の観点で近似点を求めるパラメータであり、メンバシップとは無関係であるため、学習率は γ のみとしている。すなわち、(10)式と(11)式には U_c^θ が含まれるのに対して、(12)式は U_c^θ の影響を受けないことに対応している。

一方、メンバシップ u_{ci} については、確率的な制約から相対的な比率を表すために、 $u_{ci}, c=1, \dots, C$ が互いに独立しておらず、逐次学習により個々に更新することができない。そこで、 u_{ci} を直接保持するのではなく、 u_{ci} の更新に用いられる距離尺度 d_{ci}^2 を逐次更新しながら保持することを考える。いま、 x_{ij} が一つ与えられた場合の部分的な目的関数における d_{ci}^2 は

$$d_{ci}^2 = \alpha (x_{ij} - f_{ci} a_{cj} - b_{cj})^2 + (1 - \alpha) (x_{ij} - b_{cj})^2 \quad (28)$$

と与えられるので、更新則は以下ようになる。

$$d_{ci}^2 = d_{ci}^2 + \gamma \left(\alpha (x_{ij} - f_{ci} a_{cj} - b_{cj})^2 + (1 - \alpha) (x_{ij} - b_{cj})^2 - \frac{1}{m} d_{ci}^2 \right) \quad (29)$$

以上の更新則を用いたアルゴリズムの一例を以下に示す。

Algorithm: Component-wise Sequential Fuzzy c-Elliptotypes (CS-FCE)

Step 1 時刻 $t=0$ に利用可能な初期データ集合に従来法を適用し、 $f_{ci}, a_{cj}, b_{cj}, d_{ci}^2, c=1, \dots, C, i=1, \dots, n, j=1, \dots, m$ の初期値を求める。

Step 2 $t=1, \dots$ について、以下の手順を行う。

Step 2-1 d_{ci}^2 から、(13)式により u_{ci} を求める。

Step 2-2 (26)式により、 a_{cj} を更新する。

Step 2-3 (27)式により f_{ci} を更新し、(7)-(8)式の制約を満たすように基準化する。

Step 2-4 (25)式により、 b_{cj} を更新する。

Step 2-5 (29)式により d_{ci}^2 を更新する。

また、充分な量の初期データ集合が与えられている場合には、ランダムに作成した初期値から逐次学習により Step 1 の初期モデルを推定することも可能である。

ここで、(26)式および(27)式の更新則は、 f_{ci} や a_{cj} が小さいときには更新幅が過剰に大きすぎることがある。そこで、数値実験では、 $|\hat{\phi}(t) - \phi(t)|$ の最大値を 10 とする制限を加えて実行した。

この局所的な主成分を用いた未評価値の予測システムでは、最大で mn 個の要素からなるデータ行列のもつ情報を、 C 個の主部分空間を表す $2Cm$ 個の要素 (a_{cj}, b_{cj}) と C 個の主部分空間上でのユーザの位置を表す Cn 個の要素 (f_{ci}) および C 個のクラスタまでの距離を表す Cn 個の要素 (d_{ci}^2) からなる $2C(m+n)$ 個の要素のみを用いて表現していることになる。

4. 数値実験

4.1 人工データを用いた実験

1,000 個の標本からなる 3 次元 ($m=3$) データを用いて実験を行った。データ集合は $t=0$ で利用可能な初期集合と $t=1, 2, \dots$ に逐次的に与えられる集合からなり、それぞれ異なる生成モデルから作成した。また、各生成モデルには二つの線形構造が含まれるものとした。Fig. 1 に 3 次元散布図を示す。小さい点が初期集合を、大きい点が $t=1$ 以降に逐次的に与えられるデータの集合を示す。いずれのデータ集合も 3 次元空間内で 2 本の直線状に散布しているが、クラスタごとの直線状の構造が $t=1$ の時点で変化したものとする。

まず、初期集合に FCE 法を適用して初期分割モデルを作成した。その後、 $t=1, 2, \dots, 3000$ としてランダムに選んだデータ点の一つの観測値のみを用いて逐次学習を行い、局所的な線形モデルを更新した。ただし、クラスタリングにおけるパラメータは $(C, p, \alpha, \theta, \gamma) = (2, 1, 1.0, 2.0, 0.01)$ とした。Fig. 2 に初期モデル ($t=0$)、 $t=500$ および $t=1000$ のモデル、更新後のモデル ($t=3000$) で得られたプロトタイプをそれぞれ示す。図中、2 本の直線がおのおののクラスタのプロトタイプである。図から、クラスタごとのプロトタイプが時間とともに適応し、逐次更新アルゴリズムにより線形モデルが適切に更新されていることがわかる。

つぎに、学習率 γ の変化と性能の関係について考察する。学習率を変化させた場合の逐次学習用データ 3000 件に対する (20) 式の部分的な目的関数の平均値 (平均 2 乗誤差) の推移を Fig. 3 に示す。図から、学習率が小さい (0.001 のとき) と収束が遅く所与のデータのみでは構造の変化をとらえられず、大きいほど収束が早まるといえる。一方、大きすぎる場合 (0.5 のとき) は構造が正しくとらえられていないことがわかる。したがって、収束に適した学習率の設定が必要であるといえる。

Fig. 1 3-D artificial data set

Fig. 2 Prototypes derived by each model

Fig. 3 Comparison of different learning parameters

4.2 協調フィルタリングへの応用

協調フィルタリングシステムとしての推薦能力の検証のために、MovieLens 映画推薦システムの web サイト [19] で公開されているデータを用いて実験を行った。データは 943 人のユーザが種々の映画を SD 法により 5

段階評価した 100,000 件の評価値からなるが、本実験ではデータに内在する共分散構造を考慮するために、各映画が少なくとも 50 人以上から評価され、かつ、各ユーザが少なくとも 20 種類の映画を評価しているという条件を満たすように、874 人のユーザと 598 種類の映画を選び、計 82,275 件の評価値のみを用いた。また、ランダムに選んだ 40,000 件を初期データとし、残りを逐次学習用のデータ集合とした。さらに、逐次学習用データには一部の映画でユーザの嗜好の変化があったものとし、150 種類の映画に $x_{ij} = \min\{x_{ij} + 1, 5\}$ なる変換を、別の 150 種類の映画に $x_{ij} = \max\{x_{ij} - 1, 1\}$ なる変換を行い、残りの 298 種類の映画については変換を行わなかった（たとえば、著名な映画賞の受賞などによりユーザの評価が変化した状況を想定している）。これらの評価値データを用いて、提案法による適応的な協調フィルタリングの効果を検証した。比較の際の基準としては、予測モデルとしての性能を比較するための MAE（平均絶対誤差：実際の評価値と予測値（1 以上～5 以下）との誤差の絶対値の平均値）と VAR（絶対誤差の分散）、推薦器としての性能を比較するための ROC（Receiver Operating Characteristic）[20] を用いた。ここでは、実際の評価値が 4 以上であったものを推薦すべき映画とみなし、4 以上の評価値に 3.5 以上の予測値を与えた割合である ROC1 と、3 以下の評価値に 3.5 未満の予測値を与えた割合である ROC2 を定義した。ROC1 および ROC2 がともに大きいほど良い推薦が行われていることになり、推薦器としての性能が良いといえる。また、上記のデータ変換の結果、逐次学習用データ 42,275 件について、評価値の平均が 3.63 から 3.58 へと減少し、4 以上の評価値が 24,990 件から 23,788 件へと減少していることから、ROC2 の算出の対象となるデータが増えている。

4.2.1 提案法における傾向の変化への適応

まず、文献 [5] の欠測値の処理を考慮した FCE 法を用いて初期データ ($t=0$) から選んだ 36,000 件のトレーニングデータにより予測モデルを作成した。ただし、クラスタリングのパラメータは、 $C \in \{1, 2, 3\}$, $p=1$, $\alpha \in \{1.0, 0.9, 0.8, 0.7, 0.6\}$, $\theta \in \{2.0, 1.5\}$ を用い、残りの 4,000 件のテストデータについての ROC1 と ROC2 の和が最大となる組合せとして、 $(C, p, \alpha, \theta) = (2, 1, 0.9, 1.5)$ を採用した。得られた性能を Table 1 に示す。ここでは、これらの値を Model-based アルゴリズムの精度の指標とみなし、傾向が変化した後のデータに対しても逐次学習によるモデル更新で同程度の精度を得ることを目的とする。学習率 γ としていくつかの値を用いて逐次学習用データ（変換後のデータ）から選んだ 38,275 件のトレーニングデータについて $t=1, 2, \dots, 38275$ の逐次学習を行い、残りの 4,000 件のテストデータについての予測性能を調べた。モデル更新を行わなかった場合との比較を Table 2 に示す。Table 1 および Table 2 より、初期モデルを更新しない場合は Table 1 に比べて精度が劣

Table 1 Performance of conventional FCE model (before transformation)

	MAE	VAR	ROC1	ROC2
Base data	0.751	0.358	0.780	0.586

Table 2 Comparison of performance with various learning parameters (after transformation)

γ	MAE	VAR	ROC1	ROC2
0	0.868	0.382	0.767	0.547
0.001	0.857	0.375	0.781	0.528
0.005	0.827	0.361	0.778	0.563
0.01	0.814	0.349	0.777	0.584
0.05	0.750	0.318	0.740	0.678
0.1	0.744	0.325	0.735	0.699
0.5	0.770	0.360	0.735	0.669

化しており、モデルの更新が必要であることがわかる。そして、学習率を増加させるにしたがって MAE が減少し、ROC2 が上昇していることから、逐次学習によるモデルの更新が嗜好の変化をとらえた精度向上に寄与していることがわかる。ROC1 が上昇していないのは、データ変換により 4 以上の評価値の数が減少したために予測値が全体的に減少したことに起因するが、ROC1 の変動に比べて ROC2 の上昇が大きいので、推薦性能は向上したといえる。ただし、学習率が大きすぎる場合は精度が劣化していることから、本推薦システムでは学習率を MAE に有意な改善が見られた 0.05～0.1 程度にするべきである。また、比較として、利用可能なすべてのトレーニングデータ (74,275 件) および 38,275 件のトレーニングデータに文献 [5] の手法を適用した場合の精度を Table 3 に示す。ここで、38,275 件のみを用いる手法は、Model-based アルゴリズムに小澤ら [11] の期間を制限する考えを取り入れたモデルの一種といえる。逐次学習したモデルに比べておおむね良い結果となっており、実運用では定期的に一括学習したモデルと精度の比較を行うことが望ましいといえる。また、一括学習においても、用いる評価値の期間を制限することで傾向の変化をとらえられると考えられる。ただし、実システムでは傾向の変化時点を特定することは容易ではないため、どのような指標で期間を設定すればよいかが問題となる。

4.2.2 時間重みを用いる GroupLens との比較

つづいて、Memory-based アルゴリズムの GroupLens に評価してから時間に基づく重みを用いた手法との比較を行った。未評価値の推定には、初期データ 36,000 件と逐次学習に用いた 38,275 件のトレーニングデータのすべてを用い、精度の比較を行う時点 ($t=38275$) と評価時刻の差 τ に基づいて以下の重み [9] を与えた。

$$\psi(\tau) = \exp\left(-\frac{\tau}{T_0}\right) \quad (30)$$

Table 3 Performance of conventional FCE model (after transformation)

	MAE	VAR	ROC1	ROC2
All data	0.738	0.294	0.809	0.643
38,275	0.727	0.292	0.734	0.720

Table 4 Result of GroupLens with user-based algorithm (after transformation)

T_0	MAE	VAR	ROC1	ROC2
original	0.810	0.454	0.751	0.622
35000	0.775	0.435	0.754	0.664
30000	0.772	0.436	0.752	0.668
25000	0.765	0.433	0.755	0.674
20000	0.757	0.425	0.753	0.682
15000	0.751	0.426	0.753	0.691
10000	0.753	0.439	0.748	0.702
5000	0.764	0.454	0.736	0.700

パラメータ T_0 が小さいほど直近のデータを重視する傾向が強く、大きいほど重みを用いないモデル (original) に近づく。なお、初期データについてもランダムに並べ替えて評価時刻 ($t=0, -1, \dots, -35999$) を定めた。まず、ユーザ間の類似度に基づいて映画ごとの評価値の重みつき平均を推定する user-based アルゴリズムにおいて、 T_0 を種々に変化させたときの精度の比較を Table 4 に示す。表から、評価してから時間が短いデータを重視することにより、傾向の変化がとらえられて精度が向上していることがわかるが、 T_0 が 10000～5000 程度になると予測に用いられる評価値の数が極端に少なくなり、精度が劣化している。また、 T_0 が 20000～15000 程度のときに精度が最も良く、重みを用いない場合と比較して MAE にも有意な改善が見られたが、提案手法と同程度の精度といえる。

つぎに、映画間の類似度に基づいてユーザごとの評価値の重みつき平均を推定する item-based アルゴリズムの精度の比較を Table 5 に示す。表から、item-based アルゴリズムでは重みによる精度の向上が見られないことがわかる。ここでは、個々のユーザの嗜好の変化ではなく映画ごとの評価が変化する状況であるために、当該ユーザの評価時刻に基づく重みでは傾向の変化をとらえられないことを意味している。ただし、重みを考慮しない場合でも提案法以上の精度を持っていることから、提案法においてもアイテム間の類似度によるデータ分割を考慮したモデル改良が今後の課題として残っている。

4.2.3 メモリ所要量および計算量

最後に、メモリ所要量および計算量について比較する。GroupLens では個々の予測値の算出に 74,275 件の評価値すべてを用いているのに対して、提案の予測モデルでは $2C(m+n) = 2 \times 2 \times (598 + 874) = 5,888$ 個の

Table 5 Result of GroupLens with item-based algorithm (after transformation)

T_0	MAE	VAR	ROC1	ROC2
original	0.721	0.327	0.806	0.671
35000	0.725	0.326	0.798	0.671
30000	0.728	0.332	0.795	0.671
25000	0.730	0.333	0.794	0.668
20000	0.731	0.332	0.791	0.672
15000	0.747	0.357	0.785	0.667
10000	0.769	0.392	0.770	0.662
5000	0.807	0.461	0.743	0.646

値を保持するのみであり、Model-based アルゴリズムがメモリ所要量の少ない手法であることがわかる。また、GroupLens ではある一人のユーザの特定の1個の未評価値の予測に、他の873ユーザ（または597アイテム）との相関係数の算出 ($O(mn)$) とそれらの重みを用いた平均値の算出 ($O(m)$ または $O(n)$) が必要であるが、提案法では C 個のメンバシップの算出 ($O(C^2)$) による所属クラスタの推定と当該クラスタにおける予測値の算出 ($y_{cij} = f_{ci}a_{cj} + b_{cj}$) のみ必要で、通常 $C \ll m, n$ であることから、予測における計算量が大幅に軽減できる。さらに、逐次学習においても、1件の評価値の入力に対して (25)~(27) 式および (29) 式をクラスタ数 C 回求めるだけでよく、計算量は $O(CT)$ である。ただし、 T は入力件数であり、通常 $T \ll mn$ である。一方、初期モデルの算出に用いた FCE 法では、(10)~(13) 式の更新 (おのおの $O(Cm)$ または $O(Cn)$) をクラスタごとにすべてのユーザとアイテムについて行う必要があり、さらにアルゴリズムが収束するまで繰り返す。本実験では、 $\epsilon = 0.000001$ として収束するまでに169回の繰り返しの要した。簡略化のために初期モデルを逐次学習で求めることも可能である。本実験への適用では、たとえば FCE 法で1回のみ更新した後に逐次更新で初期モデルを作成した場合、 $\gamma = 0.05$ としたとき、Table 1 に対応する初期モデルの初期テストデータに対する精度が (MAE, VAR, ROC1, ROC2) = (0.838, 0.410, 0.740, 0.502)、Table 2 に対応する更新後のモデルの精度が (MAE, VAR, ROC1, ROC2) = (0.788, 0.351, 0.725, 0.674) となっており、簡略なアプローチでもある程度の精度が期待できるといえる。

5. おわりに

本論文では、線形ファジィクラスタリングを用いた協調フィルタリング法の発展として、評価値の逐次学習による動的な適応を可能とする予測モデルの提案を行った。人工データを用いた実験により、逐次学習モデルによっても線形クラスタへのデータの分類が可能であることを示した。また、映画評価データを用いた実験により、逐

次学習を用いる提案法がユーザの嗜好の変化をとらえた推薦を行えることを示した。Memory-based アルゴリズムの GroupLens との精度の比較では、ユーザ間類似度に基づくユーザの分割を伴う提案法と user-based アルゴリズムで同程度の精度が得られており、メモリ所要量や計算量が少ない提案手法でも十分な精度が期待できる。ただし、item-based アルゴリズムには劣っていたことから、コンテンツ分割を伴う局所的な線形モデルによる改良の効果検証が検討課題として残っている。

メモリ所要量や計算量が少ない提案手法は、ユーザとのデータ入出力に関わるリコメンダ端末と基幹サーバを分離した実装において効果が期待される。従来の協調フィルタリングシステムでは基幹サーバにおける運用が前提であったが、基幹サーバの機能をデータの蓄積と初期モデルの作成 (および定期的な一括更新) に限定し、ユーザとの逐次的な入出力が必要な推薦システムを簡易なりコメンダ端末上で運用することが可能となり、低コストで基幹サーバの負荷を軽減できる。

実システムへの応用により、実問題におけるユーザの嗜好の変化への動的適応の効果を検証することが今後の課題である。

参考文献

- [1] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl: Grouplens: an open architecture for collaborative filtering of netnews; *Proc. of ACM Conference on Computer-Supported Cooperative Work*, pp. 175–186 (1994)
- [2] 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)
- [3] 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)
- [4] J. Breese, D. Heckerman and C. Kadie: Empirical analysis of predictive algorithms for collaborative filtering; *Proc. of 14th Conference on Uncertainty in Artificial Intelligence*, pp. 43–52 (1998)
- [5] 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)
- [6] K. Honda and H. Ichihashi: Component-wise robust linear fuzzy clustering for collaborative filtering; *International Journal of Approximate Reasoning*, Vol. 37, No. 2, pp. 127–144 (2004)
- [7] J. C. Bezdek: *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press (1981)
- [8] K. Honda and H. Ichihashi: Linear fuzzy clustering

- techniques with missing values and their application to local principal component analysis; *IEEE Trans. on Fuzzy Systems*, Vol. 12, No. 2, pp. 183–193 (2004)
- [9] Y. Ding and X. Li: Time weight collaborative filtering; *Proc. 14th ACM international conference on Information and knowledge management*, pp. 485–492 (2005)
- [10] Y. Ding, X. Li and M. E. Orlowska: Recency-based collaborative filtering; *Proc. 17th Australian Database Conference (ADC2006)*, CRPIT, Vol. 49, pp. 79–88 (2006)
- [11] 小澤, 内藤, コストブ: アイテムが選択された日時を用いた協調フィルタリングと着信メロディのダウンロード履歴を用いた評価; 日本知能情報ファジィ学会誌, Vol. 18, No. 5, pp. 799–808 (2006)
- [12] F. Höppner, F. Klawonn, R. Kruse and T. Runkler: *Fuzzy Cluster Analysis*, Jhon Wiley & Sons (1999)
- [13] 宮本: クラスタ分析入門, 森北出版 (1999)
- [14] K. Honda and H. Ichihashi: Regularized linear fuzzy clustering and probabilistic PCA mixture models; *IEEE Trans. on Fuzzy Systems*, Vol. 13, No. 4, pp. 508–516 (2005)
- [15] 竹内: 統計学辞典, 東洋経済新報社, pp. 356–357 (1989)
- [16] B. M. Sarwar, G. Karypis, J. A. Konstan and J. Riedl: Item-based collaborative filtering recommendation algorithms; *Proc. of 10th International World Wide Web Conference (WWW10)*, pp. 285–295 (2001)
- [17] 井口, 宮本: カーネル関数を利用した LVQ クラスタリングと SOM; 日本知能情報ファジィ学会誌, Vol. 17, No. 1, pp. 88–94 (2005)
- [18] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag (1995)
- [19] MovieLens Web Page; <http://www.movielens.org/>
- [20] J. A. Swets: Measuring the accuracy of diagnostic systems; *Science*, Vol. 240, No. 4857, pp. 1285–1289 (1988)
-