

購買順序を効率的に用いた協調フィルタリング

岩田 具治^{†1} 山田 武士^{†1} 上田 修功^{†1}

高速にパラメータ推定可能な購買順序を考慮した協調フィルタリング手法を提案する．これまでに購買履歴を入力とし次の購入商品を予測する協調フィルタリング問題において，購買順序を考慮したマルコフモデルや最大エントロピーモデルが適用されている．マルコフモデルはパラメータの推定，更新を高速にできるが，予測精度が低いという問題点がある．一方，最大エントロピーモデルはパラメータ推定，更新に多くの計算量を必要とするが，予測精度は高い．提案法は，複数のマルコフモデルを最大エントロピー原理を用いて統合することにより，高速性と高予測精度を同時に実現する．音楽，動画，漫画配信サービスの実ログデータを用い，提案法の有効性を示す．

Collaborative Filtering Efficiently Using Purchase Orders

TOMOHARU IWATA,^{†1} TAKESHI YAMADA^{†1} and NAONORI UEDA^{†1}

We propose a new collaborative filtering method that can predict the next purchase item by efficiently using the sequential information in purchase histories for recommendations. Markov models and maximum entropy models are both widely used techniques for such recommendations. In Markov models, parameters can be estimated and updated fast and efficiently, but predictions may not be accurate. On the other hand, the accuracy of maximum entropy models is generally high, however parameter estimation incurs a high computational cost. We achieve both fast parameter estimation and high predictive accuracy by combining multiple simple Markov models based on the maximum entropy principle. Experiments using real log data sets of music, movie and cartoon distribution services show that the proposed method outperforms other conventional methods found in the literature.

1. はじめに

リコメンデーションシステムはオンラインストアにおいて広く用いられている¹⁸⁾．ユーザが興味のあると思われる商品を推薦することにより，ユーザは欲しい情報により早くたどり着くことができ，また，オンラインストアにとっては収益の増加が見込まれるという利点がある．リコメンデーションを実現するための手法として，協調フィルタリング^{13),17)}，コンテンツフィルタリング¹¹⁾，および，そのハイブリッド方式⁹⁾など数多く提案されている．協調フィルタリングでは，似た興味を持つ他のユーザの評点履歴や購買履歴を用いて興味を推定する．また，コンテンツフィルタリングでは，興味を推定するために商品情報を利用する．商品情報として，たとえば本の著者や映画のジャンルなどが用いられる．

本論文では，購買履歴のみを入力とし，次に購入す

る商品を予測する協調フィルタリング問題を扱う．次に購入する商品はユーザのその時点の興味を最も表していると考えられ，購入商品予測の精度が高い手法が，興味の予測精度が高い手法であると考ええる．購買履歴は実際にサービスを展開しているオンラインストアであれば蓄積されるものであり，基本的なデータであるといえる．一方，評点情報や商品情報は予測のための有用な情報となりうるが，前者はユーザに評価の負担を課し，後者は新商品の発売ごとに人手でデータ登録しなければならないという，実用上の問題点がある．そこで本論文では購買履歴のみに着目する．

これまでに提案されている協調フィルタリング手法の大半は，商品の購買順序を考慮していない¹⁷⁾．しかし，昔購入した商品よりも最近購入した商品の方が次に購入する商品を予測するための情報がより多く含まれると考えられる．簡単な例をあげると，ドラマの第1話のDVDを借りたユーザは，次にそのドラマの第2話を借りる確率が高いだろうし，興味は変化することもあるため，昔の購買履歴があまり参考にならない場合もあるだろう．また，直前に購入された商品のみを考慮する手法^{6),9)}もしばしば使われるが，過去に購

^{†1} 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, NTT Corporation

入した多くの商品の情報を無視してしまっている．

本論文では、購買順序に着目した新たな協調フィルタリング手法を提案する．これまでに購買順序を考慮する手法として、マルコフモデルや最大エントロピーモデルが協調フィルタリングに適用されている⁸⁾．マルコフモデルはパラメータの推定、更新を高速にできるが、予測精度が低いという問題点がある．一方、最大エントロピーモデルはパラメータ推定、更新に多くの計算量を必要とするが、予測精度は高い．提案法では、順序情報を扱うための複数の単純マルコフモデルを最大エントロピー原理を用いて統合させることにより、高速なパラメータ推定かつ高い予測精度を実現する．提案法は新規データ追加によるパラメータ更新も高速にできる．オンラインストアでは時々刻々と購買履歴が蓄積されるため、高速なパラメータ更新は重要である．漫画、音楽、動画の配信サービスにおける実購買履歴データを用い、提案法の有効性を評価する．

以下の本文では、まず 2 章で提案法の基礎となるマルコフモデルと最大エントロピーモデルについて説明し、3 章で本論文で提案する効率的に順序情報を利用する協調フィルタリング法について詳述する．4 章では関連研究について述べる．そして 5 章でオンラインストアの実購買履歴データを用いて提案法の評価を行い、6 章で結論と今後の課題を示す．

2. 従来法

ユーザ集合を U 、商品集合を S 、ユーザ n が k 番目に購入した商品を $x_{n,k} \in S$ 、そのときの購買履歴を $\mathbf{u}_{n,k} = (x_{n,k-1}, \dots, x_{n,1})$ とする．また N を全ユーザ数、 V を全商品数、 K_n をユーザ n が購入した商品数とする．

2.1 マルコフモデル

マルコフモデルは順序情報を扱う確率モデルとして広く用いられている． L 次マルコフモデルでは、次に購入する商品は L 個前までの購買の履歴に依存し、

$$P(x_{n,k} | \mathbf{u}_{n,k}) = P(x_{n,k} | x_{n,k-1}, \dots, x_{n,k-L}), \quad (1)$$

と表される． L 個前までの購買の履歴を $\mathbf{u} = (x_1, \dots, x_L)$ とする．ここで x_l は l 個前の購買商品を表す． L 次マルコフモデルにおける、購買履歴 \mathbf{u} が与えられたときの商品 x を購入する確率は、MAP 推定により

$$\hat{P}(x | \mathbf{u}) = \frac{M(x, \mathbf{u}) + \delta}{M(\mathbf{u}) + \delta V}, \quad (2)$$

となる．ここで δ はハイパーパラメータであり、leave-one-out 交差検定法により推定することができる．ま

た $M(x, \mathbf{u})$ は全ユーザの購買履歴中に含まれる系列 $(x, \mathbf{u}) = (x, x_1, \dots, x_L)$ の数

$$M(x, \mathbf{u}) = \sum_{n \in U} \sum_{k=L+1}^{K_n} I((x_{n,k} = x) \wedge (x_{n,k-1} = x_1) \wedge \dots \wedge (x_{n,k-L} = x_L)), \quad (3)$$

を表し、 $M(\mathbf{u})$ は全ユーザの購買履歴中に含まれる系列 $\mathbf{u} = (x_1, \dots, x_L)$ の数

$$M(\mathbf{u}) = \sum_{n \in U} \sum_{k=L+1}^{K_n} I((x_{n,k-1} = x_1) \wedge \dots \wedge (x_{n,k-L} = x_L)), \quad (4)$$

を表す． $I(y)$ は y が真ならば 1、そうでなければ 0 をとる特性関数を表す．

マルコフモデルにおけるパラメータは、上式のように閉形式で書くことができ、かつ、単純な和のみで計算できるため、推定に必要な計算量は少ない．また、新たなデータが得られた際のパラメータ更新が以下のように容易にできる．まず、すでにある購買データに含まれる式 (2) の分子、分母に相当する系列の数をそれぞれ記憶する．そして、新たな購買データが追加されたとき、追加購買データに含まれるそれぞれの系列数を数え、記憶しておいた系列数に足すことで、パラメータを更新できる．このとき既存の購買データを改めて読み込む必要はない．しかし、 L 次マルコフモデルのパラメータ数は $O(V^{L+1})$ であり、高次マルコフモデルではパラメータ数が膨大になるため、頑健なパラメータ推定は困難であるという問題点がある．

商品 x を購入したという条件のもと l 個前に購入した商品と l' 個前に購入した商品は独立であると仮定すると、下式

$$P(x_1, \dots, x_L | x) = \prod_{l=1}^L P_l(x_l | x), \quad (5)$$

のように分解でき、 L 個前までの購買順序情報を用い、かつ、パラメータ数を少なくすることができる．ここで $P_l(x' | x)$ は商品 x を購入した l 個前の商品が x' である確率を表し、 l ギャップマルコフモデル¹⁴⁾ と呼ぶ．このとき購買履歴 \mathbf{u} が与えられたもとで商品 x を購入する確率はベイズの定理より、

$$P(x | \mathbf{u}) = \frac{P(x)P(\mathbf{u} | x)}{\sum_{x'} P(x')P(\mathbf{u} | x')} = \frac{P(x) \prod_{l=1}^L P_l(x_l | x)}{\sum_{x'} P(x') \prod_{l=1}^L P_l(x_l | x')}, \quad (6)$$

となる．ここで $P(x)$ は商品 x を購入する事前確率

を表す．このときのパラメータ数は $O(LV^2)$ となる．事前確率の MAP 推定値は

$$\hat{P}(x) = \frac{\sum_n \sum_k I(x_{n,k} = x) + \delta}{\sum_n K_n + \delta V}, \quad (7)$$

となる．また同様に， l ギャップマルコフモデルの MAP 推定値は，

$$\begin{aligned} & \hat{P}_l(x'|x) \\ &= \frac{\sum_n \sum_k I((x_{n,k} = x) \wedge (x_{n,k-l} = x')) + \delta}{\sum_n \sum_k I(x_{n,k} = x) + \delta V}, \end{aligned} \quad (8)$$

となる．ここでハイパーパラメータの項を除いて考えると，分母は全ユーザの購買履歴中の商品 x の購買数，分子は商品 x の l 個前の購買が商品 x_l である数を表す．

2.2 最大エントロピーモデル

最大エントロピーモデルは，与えられたデータに関する制約を満たす確率分布のなかでエントロピーが最大となる分布を求める手法であり，協調フィルタリング^{9),13),19)} だけでなく，自然言語処理¹²⁾ など幅広い分野で応用されている．最大エントロピーモデルでは，購買履歴 \mathbf{u} が与えられたとき商品 x を購入する確率は

$$P(x|\mathbf{u}) = \frac{1}{Z(\mathbf{u})} \exp\left(\sum_{f \in F} \alpha_f y_f(x, \mathbf{u})\right), \quad (9)$$

となる．ここで， $Z(\mathbf{u}) = \sum_x \exp\left(\sum_f \alpha_f y_f(x, \mathbf{u})\right)$ は正規化項， y_f は特徴量， F は特徴量集合を表す．購買順序を考慮するため，特徴量として l 個前に購入した商品情報

$$y_{l,i,j}(x, \mathbf{u}) = \begin{cases} 1 & \text{if } i = x_l \text{ and } j = x, \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

を用いる．このときのパラメータ数は LV^2 となる．

未知パラメータ集合 $\{\alpha_f\}_{f \in F}$ は，対数尤度

$$J = \sum_n \sum_k \log P(x_{n,k} | \mathbf{u}_{n,k}), \quad (11)$$

を準ニュートン法¹⁰⁾ などの最適化手法を用いて最大化することにより推定できる．なお，目的関数 J の凸性より，解の大域的最適性が保証される．過学習を抑えるため，実験では未知パラメータの事前分布として平均 0 の正規分布を用いた¹⁾．

一般に，識別モデルである最大エントロピーモデルは，生成モデルであるマルコフモデルに比べ性能が高いといわれており¹⁵⁾，協調フィルタリング問題におい

ても最大エントロピーモデルの優位性が確認されている⁸⁾．しかし，最大エントロピーモデルのパラメータの推定値は，マルコフモデルとは異なり閉形式で求まらず，計算量の大きい準ニュートン法などの繰返し最適化法を用いる必要がある．また，新たなデータが追加されたとき，既存のデータも読み直す必要があり，パラメータ更新も容易ではない．

3. 提案法

提案法では，複数のギャップマルコフモデルを最大エントロピー原理により統合することにより，マルコフモデルと同程度のパラメータ更新の容易性と，最大エントロピーモデルと同程度の高い予測精度を同時に実現する．

まず，事前確率 $P(x)$ を式 (7) により， l ギャップマルコフモデル $P_l(x'|x)$ ， $l = 1, \dots, L$ を式 (8) により推定する．そして，購買履歴 \mathbf{u} のとき商品 x を購入する確率 $P(x|\mathbf{u})$ を，購買順序が反映されたものにするため， $P(x|\mathbf{u})$ によるギャップマルコフモデルの対数尤度の期待値は，与えられたデータにおける対数尤度の経験的期待値と一致するという制約

$$\begin{aligned} & \sum_n \sum_k \sum_x P(x|\mathbf{u}_{n,k}) \log \hat{P}_l(x_{n,k-l}|x) \\ &= \sum_n \sum_k \log \hat{P}_l(x_{n,k-l}|x_{n,k}), \quad l = 1, \dots, L, \end{aligned} \quad (12)$$

を設定する．また，同様に全体的な人気を表す事前確率の対数尤度の期待値も一致するという制約

$$\begin{aligned} & \sum_n \sum_k \sum_x P(x|\mathbf{u}_{n,k}) \log \hat{P}(x) \\ &= \sum_n \sum_k \log \hat{P}(x_{n,k}), \end{aligned} \quad (13)$$

を設定する．最大エントロピー原理に従い，式 (12)，(13) を制約条件として， $P(x|\mathbf{u})$ のエントロピーを最大にする分布を求めると

$$\begin{aligned} & P(x|\mathbf{u}) \\ &= \frac{1}{Z(\mathbf{u})} \exp\left(\alpha_0 \log \hat{P}(x) + \sum_{l=1}^L \alpha_l \log \hat{P}_l(x_l|x)\right) \\ &= \frac{1}{Z(\mathbf{u})} \hat{P}(x)^{\alpha_0} \prod_{l=1}^L \hat{P}_l(x_l|x)^{\alpha_l}, \end{aligned} \quad (14)$$

を得る．ここで $Z(\mathbf{u}) = \sum_x \exp\left(\alpha_0 \log \hat{P}(x) + \sum_l \alpha_l \log \hat{P}_l(x_l|x)\right)$ は正規化項， $\alpha = \{\alpha_l\}_{l=0}^L$ は未知パラメータ集合であり， α_0 は事前確率の重み， α_l は l ギャップマルコフモデルの重みを表す．提案モデ

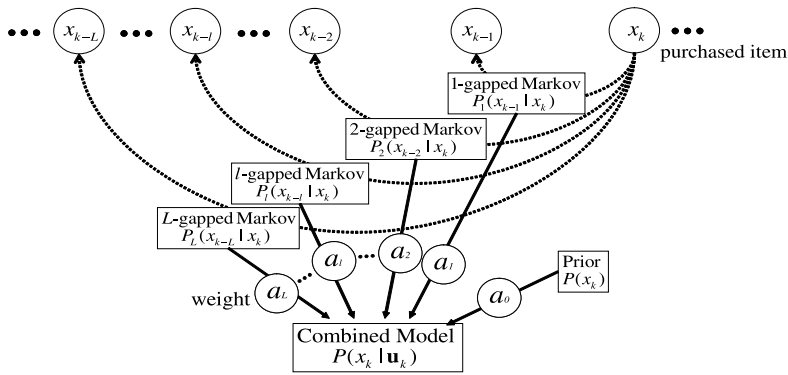


図 1 提案モデル：複数のギャップマルコフモデルの重み付き統合
 Fig. 1 The proposed model: multiple gapped Markov models are combined with weights.

ルは L 個のギャップマルコフモデルを重みを付けて統合した形となっており (図 1), 重みがすべて 1 のとき ($\alpha_l = 1, l = 0, \dots, L$), 各ギャップマルコフモデルが独立と仮定したモデルである式 (6) と一致する. 換言すれば, 提案モデルは独立ギャップマルコフモデルを特殊なケースとして包含する.

最大エントロピーモデルと同様に, 式 (11) の対数尤度を最大化することにより未知パラメータ集合 α を推定できる. ここで, 事前確率およびギャップマルコフモデルの推定に用いた学習データを α の推定に用いると過学習を起こす可能性がある. そのため, M 重交差検定法により α を推定する. まず, ユーザ集合 U を M 個の部分集合 $\{U_m\}_{m=1}^M$ に分割し, $U_{-m} = \{U_j\}_{j \neq m}$ を用いて事前確率 $\hat{P}(x; U_{-m})$ およびギャップマルコフモデル $\hat{P}_l(x'|x; U_{-m})$ を式 (15), (16) で各々推定する.

$$\hat{P}(x; U_{-m}) = \frac{\sum_{n \in U_{-m}} \sum_k I(x_{n,k} = x) + \delta}{\sum_{n \in U_{-m}} \sum_k 1 + \delta V}, \quad (15)$$

$$\begin{aligned} \hat{P}_l(x'|x; U_{-m}) &= \frac{\sum_{n \in U_{-m}} \sum_k I(x_{n,k} = x \wedge x_{n,k-l} = x') + \delta}{\sum_{n \in U_{-m}} \sum_k I(x_{n,k} = x) + \delta V}. \end{aligned} \quad (16)$$

そして, 学習に用いなかった U_m の対数尤度の和

$$L = \sum_m \sum_{n \in U_m} \sum_k \log P(x_{n,k} | \mathbf{u}_{n,k}; U_{-m}), \quad (17)$$

を最大化することにより α を推定する. ここで

$$\begin{aligned} P(x | \mathbf{u}; U_{-m}) &= \frac{1}{Z(\mathbf{u})} \hat{P}(x; U_{-m})^{\alpha_0} \prod_{l=1}^L \hat{P}_l(x_l | x; U_{-m})^{\alpha_l}, \end{aligned} \quad (18)$$

である. 実験では α_l の事前分布として平均 0 の正規分布を用いた.

提案モデルのパラメータ数は $LV(V-1) + V + L$ である. このうち L 個のギャップマルコフモデルと事前確率に対応する $LV(V-1) + V - 1$ のパラメータは, 上述のように高速に求めることができ, また, データ追加にともなうパラメータ更新も容易である. 多くの計算時間がかかる繰返し最適化が必要なのは $L + 1$ 個の重み α のみである. そのため, $O(LV^2)$ のパラメータの繰返し最適化が必要な最大エントロピーモデルと比べ, 少ない計算量でパラメータを求めることができる.

また, 重み α は以前に学習したものをを用い, 事前確率とギャップマルコフモデルのみを更新する場合, さらに計算量を削減することができる. 事前確率やギャップマルコフモデルは新商品の発売開始の際など, 大きくそのパラメータが変化するが, l 個前の購買商品が次の購買に与える影響を表すギャップマルコフモデルの重みは新たにデータが追加されても急激には変化しないと考えられる. なお, α の再学習のために全データを保持している必要がある.

4. 関連研究

複数の生成モデルの最大エントロピー原理による統合は, タイトルや参考文献など複数の付加情報を含むテキストデータの分類器設計^{3),15)} で提案されており, 単体の生成モデル, 最大エントロピーモデルと比べ高い分類性能が確認されている. 提案法では, 付加情報ではなく, 順序情報を考慮するためにモデルを統合している. また, 協調フィルタリング問題においてはデータが時々刻々と蓄積されるため, パラメータ更新が高速に行えることが重要であるため, 提案法では性能を下げることなくパラメータ更新を容易にすること

表 1 各データの開始日, 終了日, ユーザ数, 商品数, 購買数

Table 1 The start date, end date, the number of users, items and transactions.

	開始日	終了日	ユーザ数	商品数	購買数
Music4	2005/4/1	2005/4/30	247	132	1,508
Music5	2005/4/1	2005/5/31	1,120	348	7,588
Music6	2005/4/1	2005/6/30	2,104	561	15,216
Movie	2007/1/1	2007/1/1	3,085	1,569	25,363
Cartoon1	2005/4/1	2006/1/31	42,184	153	453,386
Cartoon2	2005/4/1	2006/2/28	53,830	161	599,196
Cartoon3	2005/4/1	2006/3/31	69,217	175	808,182

を目的に, 生成モデルの統合を利用している.

遠い過去の購買情報を扱うため, 共起関係が強い商品ペアのみをトリガとして素性に組み込んだ最大エントロピーモデルが提案されている¹³⁾. しかし, トリガモデルでは過去の情報を扱うことはできるが, 順序情報が無視されてしまう. 購買時間を考慮した協調フィルタリング手法も提案されている²⁾. 彼らは経過時間に応じて, パラメータの重みを指数減衰させることで, 最近のデータに重みを置いている. それゆえ, 過去の情報が有用な場合や, 減衰が指数関数で近似できない場合には適切でない.

また言語処理の分野で, 高次のマルコフモデルの頑健な推定のため, 低次のマルコフモデルの線形形で補間する, 削除補間法が提案されている⁷⁾. 削除補間法では, 高次のマルコフモデルが必要であるため, 膨大な数のパラメータを記憶する容量が必要になるという問題点がある. 言語では文法などで単語の並びに規則性が高いため高次のマルコフモデルが必要であると考えられる. 一方, 購買履歴では言語ほど規則性が高くないため, l 個前に購入した商品の情報だけを持つギャップマルコフモデルでも実用上十分と考えられる. 提案法はギャップマルコフモデルを最大エントロピー原理に基づき統合したが, ギャップマルコフモデルを線形結合する手法が提案されている¹⁴⁾. 線形結合の場合, 混合係数およびギャップマルコフモデルを EM アルゴリズムを用いて, 収束するまで交互に推定する必要がある.

5. 実験

提案法を評価するため, 音楽, 動画, 漫画の 3 種類の実購買履歴データを用い実験を行った.

5.1 データセット

音楽データは, 音楽配信サービスにおける購買履歴であり, 2005 年 4 月 1 日から 2005 年 4 月 30 日まで, 2005 年 5 月 31 日まで, 2005 年 6 月 30 日までの 3 種類のデータセットを作成した. それぞれ Music4, Music5, Music6 と呼ぶ. 動画データは, 2007

年 1 月 1 日の動画配信サービスにおける購買履歴であり, Movie と呼ぶ. 漫画データは, 携帯電話用漫画配信サービスにおける購買履歴であり, 2005 年 4 月 1 日から 2006 年 1 月 31 日まで, 2006 年 2 月 28 日まで, 2006 年 3 月 31 日までの 3 種類のデータセットを作成した. それぞれ Cartoon1, Cartoon2, Cartoon3 と呼ぶ. ここで 1 つの漫画が複数巻あるものは同一の商品として扱った. 各データともに, 売上数が 10 未満の商品, および, 購買数が 5 未満のユーザは省いた. また, ある商品が 2 回以上同一ユーザに購入された場合, 2 回目以降の購買は履歴から省いた. このときの, 各データのユーザ数, 商品数, 購買数を表 1 に示す. 各ユーザの最後に購入した商品をテストデータ, それ以前の購買履歴を学習データとして用いる. ここで, 最後に購入した商品が学習データに含まれないものはテストデータから省いた.

5.2 比較手法

提案法を評価するため, 以下の 9 つのモデルを用いて実験を行った.

OurMethod: 提案法

Markov1: 1 次マルコフモデル. 式 (2) $L = 1$.

Markov2: 2 次マルコフモデル. 式 (2) $L = 2$.

Markov3: 3 次マルコフモデル. 式 (2) $L = 3$.

GapMarkov: 各ギャップマルコフモデルがそれぞれ独立と仮定したモデル. 式 (6).

MaxEntSeq: 特徴量として式 (10) を用いた, 購買順序を考慮した最大エントロピーモデル.

MaxEnt: 特徴量として下式を用いた, 購買順序を考慮しない最大エントロピーモデル.

$$y_{i,j}(x, \mathbf{u}) = \begin{cases} 1 & \text{if } i \in \mathbf{u} \text{ and } j = x, \\ 0 & \text{otherwise.} \end{cases} \quad (19)$$

Cosine: コサイン類似度に基づくアイテムベース協調フィルタリング¹⁷⁾. 商品 x と x' のコサイン類似度は

$$\text{sim}(x, x') = \cos(\mathbf{r}, \mathbf{r}') = \frac{\mathbf{r}^T \mathbf{r}'}{\|\mathbf{r}\| \|\mathbf{r}'\|}, \quad (20)$$

で表される．ここで $r = (r_1, \dots, r_N)^T$ はユーザ n が商品 x を購入したことがあれば $r_n = 1$ ，そうでなければ $r_n = 0$ ，を要素に持つ列ベクトル， r^T はベクトル r の転置， $\|\cdot\|$ はユークリッドノルムを表す．商品 x を購入する確率は購買履歴に含まれる商品との類似度の和

$$P(x|\mathbf{u}) \propto \sum_{x' \in \mathbf{u}} \text{sim}(x, x'), \quad (21)$$

によって推定される．つまり，これまでに購入した商品と似た商品は購入する確率が高くなる．この手法は主に評点を予測する場合に利用される．パラメータ数は $O(V^2)$ である．商品の類似度ではなくユーザの類似度を基にした協調フィルタリング手法も提案されている¹⁶⁾ が，一般にユーザ数は商品数に比べて多いため，より多くの計算量が必要となる．

PLSA：Probabilistic latent semantic analysis⁴⁾．ユーザ n が商品 x を購入する確率は

$$P(x|n) = \sum_z P(x|z)P(z|n), \quad (22)$$

で与えられる．ここで z は隠れクラスであり， $P(x|z)$ はクラス z に属するユーザが商品 x を購入する確率， $P(z|n)$ はユーザ n がクラス z に属する確率を表す．確率 $P(x|z)$ および $P(z|n)$ は対数尤度

$$J = \sum_n \sum_k \log P(x_{n,k}, n), \quad (23)$$

を EM アルゴリズムを用いて最大化することにより推定できる．PLSA は協調フィルタリングだけでなく，情報検索や自然言語処理にも応用されている⁵⁾．隠れクラスの数 Z としたとき，パラメータ数は $O(NZ + VZ)$ である．

Mixture：混合ギャップマルコフモデル¹⁴⁾．商品 x を購入する確率はギャップマルコフモデルの線形結合

$$P(x|\mathbf{u}) = \sum_{l=1}^L P(l)P_l(x|\mathbf{x}_l), \quad (24)$$

で表される．ここで $P(l) \geq 0$ は混合係数で $\sum_{l=1}^L P(l) = 1$ ， $P_l(x|\mathbf{x}_l)$ は l 個前の購買商品が \mathbf{x}_l のとき商品 x を購入する確率を表す．混合係数 $P(l)$ および確率 $P_l(x|\mathbf{x}_l)$ は対数尤度

$$J = \sum_n \sum_k \log P(x_{n,k}|\mathbf{u}_{n,k}), \quad (25)$$

を EM アルゴリズムによって最大化することにより推定できる．パラメータ数は $O(LV^2)$ であり，提案法と同程度である．

Markov, GapMarkov, MaxEntSeq, Mixture, 提案法は購買順序を考慮する手法であり，MaxEnt, Co-

sine, PLSA は購買順序を考慮しない手法である．MaxEnt は何個前にどの商品を購入したかという順序は考慮しないが，購買履歴 \mathbf{u} が与えられたときの条件付き確率 $P(x|\mathbf{u})$ をモデル化するため，前後関係は考慮する．つまり，商品 x の後に商品 x' に購入することはあるが，商品 x' の後に商品 x に購入する可能性は低いなどをモデル化できる．一方，コサイン類似度は対称であり $\text{sim}(x, x') = \text{sim}(x', x)$ ，また PLSA は商品 x とユーザ n の同時分布 $P(x, n)$ をモデル化しているため，Cosine と PLSA は前後関係も考慮しない．

提案法における重み α は 10 重交差検定により推定した．マルコフモデル，事前確率およびギャップマルコフモデルにおけるハイパーパラメータ δ は leave-one-out 交差検定法により求めた．また，最大エントロピーモデルにおいて，上記の特徴量に加え履歴に無関係な全体的な商品の人気を表すための特徴量

$$y_i(x, \mathbf{u}) = \begin{cases} 1 & \text{if } i = x, \\ 0 & \text{otherwise,} \end{cases} \quad (26)$$

もあわせて用い，パラメータの事前分布は分散 1 の正規分布とした．

5.3 結果

漫画，音楽，動画データにおける各手法の正答率を表 2 に，トップ 3 正答率を表 3 に示す．正答率は下式により計算した．

$$A = \frac{1}{N} \sum_n I(x_{n,K_n} = \hat{x}(u_{n,K_n})), \quad (27)$$

ここで x_{n,K_n} はユーザ n の最後の購買商品を表す．また， $\hat{x}(u)$ は購買履歴 u が与えられたときの予測商品を表し，購買確率が最大の商品とする．

$$\hat{x}(u) = \arg \max_{x \in S_{-u}} P(x|u), \quad (28)$$

ここで S_{-u} は購買履歴 u に含まれない商品の集合を表す．また，トップ 3 正答率とは購買確率が高い 3 商品中に実際の購買商品が含まれる割合を表す．GapMarkov, MaxEntSeq, Mixture, 提案法における最大ギャップ L ，および，PLSA における隠れクラス数 Z は本来交差検定などによりデータから学習すべきであるが，本実験ではそれぞれ $L \in \{1, \dots, 10\}$ ， $Z \in \{5, 10, \dots, 50\}$ として実験を行い，最も良い正答率を表示している．括弧内はそのときの L もしくは Z の値である．なお，後述のように，提案法では最大ギャップ L を大きくしても予測性能はあまり下がらないため，計算資源が許す限り L を大きくするという設定法も可能である．

表 2 正答率 (%)
Table 2 Accuracies (%).

	Markov1	Markov2	Markov3	GapMarkov	MaxEntSeq	MaxEnt	Cosine	PLSA	Mixture	OurMethod
Music4	15.7	10.6	2.1	19.5 (2)	19.5 (1)	15.7	8.9	12.3(5)	16.1(3)	19.5 (2)
Music5	12.8	6.4	1.4	12.9(1)	14.7 (8)	10.4	5.7	8.3(5)	12.8(1)	14.0(6)
Music6	11.4	7.1	2.0	10.9(2)	12.5(6)	9.1	4.6	6.9(15)	11.5(2)	12.9 (8)
Movie	39.1	31.1	30.1	39.4(1)	38.7(3)	27.7	5.4	6.1(50)	39.5(7)	39.6 (3)
Cartoon1	15.8	16.2	10.4	17.7(3)	19.1(5)	17.3	8.9	11.3(5)	16.4(2)	19.5 (10)
Cartoon2	16.2	18.3	14.4	20.7(4)	21.8 (6)	19.1	8.4	8.8(5)	16.7(2)	21.7(10)
Cartoon3	16.1	17.1	14.4	18.3(5)	20.9 (7)	18.8	8.4	9.6(30)	16.1(1)	19.6(10)

表 3 トップ 3 正答率 (%)
Table 3 Top-3 accuracies (%).

	Markov1	Markov2	Markov3	GapMarkov	MaxEntSeq	MaxEnt	Cosine	PLSA	Mixture	OurMethod
Music4	30.5	15.3	2.1	32.2(2)	36.4 (3)	25.4	17.4	26.7(10)	31.4(3)	34.7(6)
Music5	24.7	12.0	3.2	24.3(2)	28.0 (3)	22.6	14.4	19.6(25)	24.7(1)	27.6(7)
Music6	21.9	11.4	3.3	22.2(1)	25.0(3)	18.4	12.1	14.7(15)	23.2(6)	25.6 (10)
Movie	44.4	32.7	31.0	44.3(1)	43.9(3)	35.3	10.3	13.5(50)	44.7(10)	45.0 (6)
Cartoon1	29.1	28.3	17.8	31.7(4)	34.5 (7)	34.1	17.9	22.9(40)	30.1(2)	34.4(10)
Cartoon2	29.6	31.0	22.6	34.4(4)	36.7 (7)	35.0	17.3	21.5(5)	31.6(2)	42.7 (4)
Cartoon3	28.8	30.2	23.5	32.8(6)	35.6 (8)	33.4	18.3	23.8(30)	30.1(2)	35.1(10)

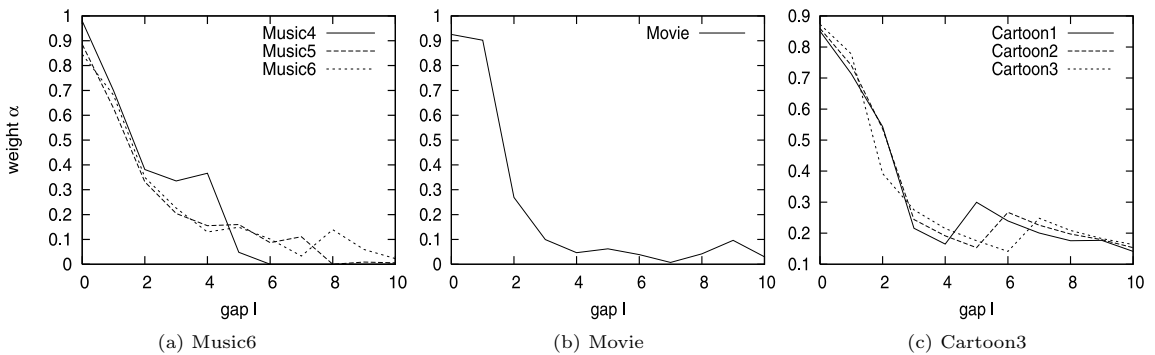


図 2 提案法における重み α_l
Fig. 2 Weights in our method α_l .

提案法と MaxEntSeq は他と比べ同程度に正答率が高い。GapMarkov は提案法に比べ正答率が低くなっている。これは各ギャップマルコフモデルが独立であるという仮定が適切でないためと考えられる。すべてのデータセットで Markov3 は Markov1, Markov2 に比べ正答率が低くなっている。これは高次になるとパラメータ数がデータ数に比べ多くなり、頑健な推定ができないためと考えられる。また、MaxEntSeq と MaxEnt を比較すると、全データセットで MaxEntSeq の方が正答率が高く、購買順序は正答率を上げるために重要であることが示唆される。Cosine と PLSA の正答率は低い。これは購買順序を考慮していないためであると考えられる。Cosine や PLSA は評点予測のためによく用いられるが、今回の実験で用いた購買商品予測問題の場合、順序は大きな情報となる

ことが示唆される。Mixture の正答率は Markov1 と同程度であった。Mixture において推定された混合係数は $P(l = 1) \simeq 1, P(l \neq 1) \simeq 0$ であることが多かった。つまり、ギャップマルコフモデルを線形和で結合した場合、1 ギャップマルコフモデルのみが優勢になり、2 以上個前の購買商品情報を有効に利用できていない。

図 2 に提案法 ($L = 10$) における事前確率およびギャップマルコフモデルの重み α を示す。全データセットでギャップが小さいマルコフモデルの重みが大きい傾向があり、最近の履歴が購買予測に関する大きな情報を与えるという直感と一致する。2 ギャップ以上のマルコフモデルの重みもある程度あり、過去の購買商品情報も有効に利用していることが分かる。また、終了日が異なるデータでも、それほど重みは異なっ

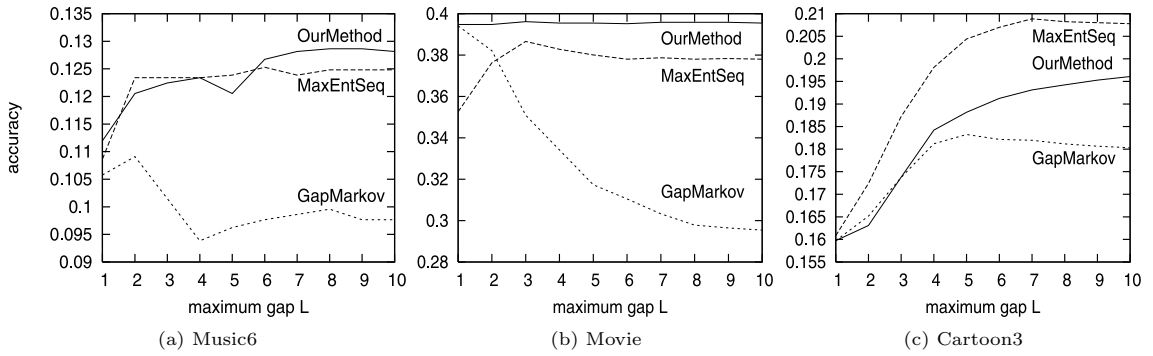


図 3 提案法, MaxEntSeq, GapMarkov の最大ギャップ L に対する正答率
 Fig. 3 Accuracies of our method, MaxEntSeq and GapMarkov with maximum gap L .

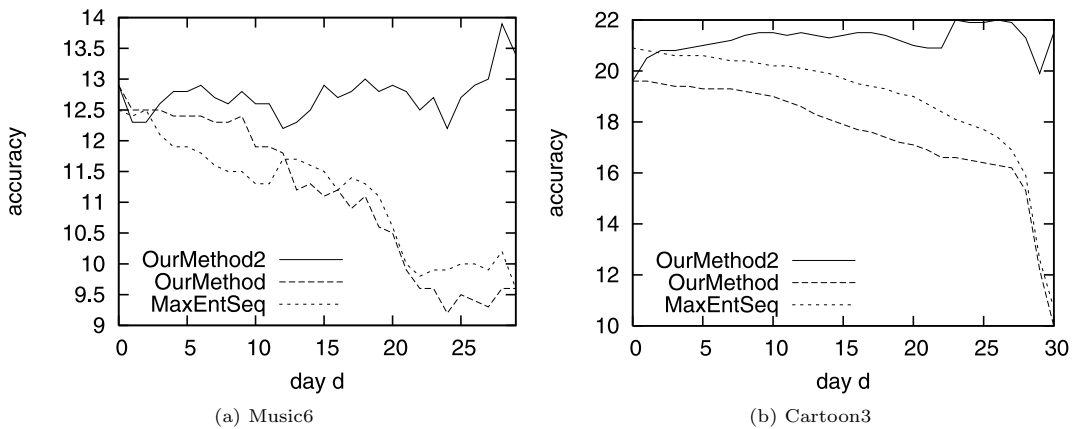


図 4 d 日前までのデータで推定した重み α を用いたときの提案法の正答率 (OurMethod2). OurMethod および MaxEntSeq は d 日前のデータを用いて全パラメータを推定したときの正答率を表す
 Fig. 4 Accuracies of our method, in which weights α are estimated by using data until d days before the end date (OurMethod2). OurMethod and MaxEntSeq represent accuracies when all parameters are estimated using data until d days before the end date.

いない。

図 3 に提案法, MaxEntSeq, GapMarkov の最大ギャップ L に対する正答率を示す。提案法は L を増加させた場合に正答率がそれほど下がることはないが, MaxEntSeq, GapMarkov は正答率の低下が見られる。MaxEntSeq では過学習が起こっているため, また, GapMarkov では L を大きくしたとき各ギャップマルコフモデルが独立であるという仮定が不適切になりやすいためと考えられる。

d 日前までのデータで推定した重み α を用い, 事前確率, ギャップマルコフモデルのみ全データで推定したときの提案法 (OurMethod2) の正答率を図 4 に示す。ここで OurMethod および MaxEntSeq は d 日前のデータを用いて全パラメータを推定したときの正答

率を表す。OurMethod2 において, 事前確率, ギャップマルコフモデルのハイパーパラメータ δ は d 日前までのデータを用いて推定したものをを用いた。漫画, 音楽データとともに, 過去の重みを用いても正答率はそれほど変わらず, 全データを用いて推定した提案法, MaxEntSeq と同程度の性能であった。この結果は, 新規データ追加によるパラメータの更新に多くの計算量が必要な重み α の更新頻度は少なくし (たとえば 1 月に 1 回程度), 更新が容易な事前確率, ギャップマルコフモデルのみ短い間隔で更新するだけでも, 高い予測精度を実現できることを示している。

各手法のパラメータ学習に必要な計算時間を比較した。Xeon3.6 GHz, メモリ 2 GB の計算機での計算時間を表 4 に示す。OurMethod2 は, 重み α は過去のデー

表 4 計算時間 (秒)

Table 4 Computational time (second).

	Markov1	Markov2	Markov3	GapMarkov	MaxEntSeq	MaxEnt	Cosine	PLSA	Mixture	OurMethod	OurMethod2
Music4	0.03	0.06	0.05	0.04	12.77	6.28	0.01	1.07	0.17	3.66	0.08
Music5	0.12	0.69	0.76	0.23	456.56	217.54	0.06	10.31	1.52	87.21	0.45
Music6	0.30	2.15	2.65	0.67	1969.55	916.47	0.18	29.05	3.87	345.90	1.12
Movie	2.00	11.04	9.35	5.16	9316.09	8267.11	0.97	117.69	14.53	2968.72	7.49
Cartoon1	1.61	4.50	20.45	2.82	132001.62	16046.90	7.97	239.01	753.77	5267.85	4.39
Cartoon2	2.13	5.66	27.42	3.83	216770.25	21597.25	11.36	356.28	1242.36	7931.55	5.59
Cartoon3	2.96	7.58	40.05	5.23	347511.57	42994.86	16.08	478.59	1553.69	10444.11	7.50

タでの計算済のものを用い、事前確率とギャップマルコフモデルのみ学習する手法である。GapMarkov, MaxEntSeq, Mixture, OurMethod, OurMethod2 での最大ギャップは $L = 10$ とし、PLSA での隠れクラス数は $Z = 10$ とした。Markov, GapMarkov, Cosine, OurMethod2 は計算時間が短く、最も規模が大きいデータである Cartoon3 でも数十秒以内で終了している。重み計算まで含めた提案法 (OurMethod) は、上記の 4 手法と比べると多くの計算量を必要とするが、正答率が同程度に高い MaxEntSeq と比べると、漫画データの場合約 30 倍高速である。

6. む す び

本論文では、購買順序を考慮した複数のマルコフモデルを最大エントロピー原理で統合した協調フィルタリング手法を提案した。提案法は、パラメータ推定および新規データ追加によるパラメータ更新を高速にできるという利点を持つ。実データを用いた実験で、提案法は多くの計算量が必要な購買順序を考慮した従来法と同程度の性能であること、また、計算量が少ない手法に比べ高い性能であることを確認した。今回は購買履歴のみを入力としたが、最大エントロピー原理の枠組みで、コンテンツ情報やユーザ情報も組み込むことも可能であるため、今後はそれらの情報も考慮しモデルの性能向上を進めていきたい。また、今回は重みをそれぞれ未知パラメータとして個別に推定したが、重みを指数分布 $\alpha_l = \lambda \exp(-\lambda l)$ などによりモデル化することも考えられる。個別に推定することにより、事前知識を必要とせず各重みをデータから適応的に推定することができ、また、通常の最大エントロピーモデルと同様に大域的最適解が得られるという利点がある。一方、モデル化する場合にもパラメータの削減ができるという利点があるため、今後、重みのモデル化についても検討していきたい。

参 考 文 献

- 1) Chen, S.F. and Rosenfeld, R.: A Gaussian prior for smoothing maximum entropy models, Technical report, CMUCS-99-108 (1999).
- 2) Ding, Y. and Li, X.: Time weight collaborative filtering, *CIKM '05: Proc. 14th ACM international conference on Information and knowledge management*, New York, NY, USA, pp.485–492, ACM Press (2005).
- 3) Fujino, A., Ueda, N. and Saito, K.: A hybrid generative/discriminative approach to text classification with additional information, *Information Processing and Management*, Vol.43, pp.379–392 (2007).
- 4) Hofmann, T.: Probabilistic Latent Semantic Analysis, *Proc. 15th Conference on Uncertainty in Artificial Intelligence, UAI-99*, San Francisco, CA, pp.289–296, Morgan Kaufmann Publishers (1999).
- 5) Hofmann, T.: Probabilistic Latent Semantic Indexing, *Proc. Annual International SIGIR Conference on Research and Development in Information Retrieval, SIGIR-99*, pp.50–57 (1999).
- 6) Iwata, T., Saito, K. and Yamada, T.: Recommendation methods for extending subscription periods, *Proc. Conference on Knowledge Discovery and Data Mining*, pp.574–579 (2006).
- 7) Jelinek, F. and Mercer, R.L.: Interpolated estimation of Markov source parameters from sparse data, *Proc. Workshop on Pattern Recognition in Practice* (1980).
- 8) Jin, X., Mobasher, B. and Zhou, Y.: A Web Recommendation System Based on Maximum Entropy, *Proc. International Conference on Information Technology: Coding and Computing (ITCC'05), Vol.1*, Washington, DC, USA, pp.213–218, IEEE Computer Society (2005).
- 9) Jin, X., Zhou, Y. and Mobasher, B.: A maximum entropy web recommendation system: combining collaborative and content features, *Proc. ACM SIGKDD Conference on Knowl-*

edge Discovery and Data Mining (2005).

- 10) Liu, D.C. and Nocedal, J.: On the limited memory BFGS method for large scale optimization, *Math. Programming*, Vol.45, No.3, pp.503–528 (1989).
- 11) Mooney, R.J. and Roy, L.: Content-based book recommending using learning for text categorization, *Proc. 5th ACM Conference on Digital Libraries*, pp.195–204 (2000).
- 12) Nigam, K., Lafferty, J. and McCallum, A.: Using maximum entropy for text classification, *Proc. IJCAI-99 Workshop on Machine Learning for Information Filtering*, pp.61–67 (1999).
- 13) Pavlov, D.Y. and Pennock, D.M.: A maximum entropy approach to collaborative filtering in dynamic, sparse, high-dimensional domains, *Proc. Neural Information Processing Systems*, pp.1441–1448 (2002).
- 14) Raftery, A.E.: A model for high-order Markov chains, *Journal of the Royal Statistical Society B*, Vol.47, No.3, pp.528–539 (1985).
- 15) Raina, R., Shen, Y., Ng, A.Y. and McCallum, A.: Classification with Hybrid Generative / Discriminative Models, *Proc. Neural Information Processing Systems* (2004).
- 16) Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: Grouplens: an open architecture for collaborative filtering of netnews, *Proc. ACM Conference on Computer Supported Cooperative Work*, pp.175–186 (1994).
- 17) Sarwar, B., Karypis, G., Konstan, J. and Riedl, J.: Item-based collaborative filtering recommendation algorithms, *Proc. 10th International Conference on World Wide Web*, New York, NY, USA, pp.285–295, ACM Press (2001).
- 18) Schafer, J.B., Konstan, J.A. and Riedl, J.: E-commerce recommendation applications, *Data Mining and Knowledge Discovery*, Vol.5, pp.115–153 (2001).
- 19) Zitnick, L. and Kanade, T.: Maximum entropy for collaborative filtering, *Proc. 20th Conference on Uncertainty in Artificial Intelligence*, pp.636–643 (2004).

(平成 19 年 11 月 22 日受付)

(平成 19 年 12 月 20 日採録)



岩田 具治 (正会員)

昭和 54 年生。平成 13 年慶應義塾大学環境情報学部環境情報学科卒業。平成 15 年東京大学大学院総合文化研究科広域科学専攻修了。同年 NTT 入社。機械学習，可視化の研究に従事。現在，NTT コミュニケーション科学基礎研究所に所属。FIT 船井ベストペーパー賞（平成 16 年），FIT ヤングリサーチャー賞（平成 19 年）受賞。電子情報通信学会会員。



山田 武士 (正会員)

昭和 39 年生。昭和 63 年 3 月東京大学理学部数学科卒業。同年 NTT 入社。平成 8 年より 1 年間英国コベントリー大学客員研究員。現在，NTT コミュニケーション科学基礎研究所創発環境研究グループリーダー。主として機械学習，データマイニング，メタヒューリスティクスによる組合せ最適化等の研究に従事。博士（情報学）。電子情報通信学会，ACM，IEEE 各会員。



上田 修功 (正会員)

昭和 33 年生。昭和 57 年大阪大学工学部通信工学科卒業。昭和 59 年同大学大学院修士課程修了。工学博士。同年 NTT 入社。平成 5 年より 1 年間 Purdue 大学客員研究員。画像処理，パターン認識・学習，ニューラルネットワーク，統計的学習，Web 統計解析の研究に従事。現在，NTT コミュニケーション科学基礎研究所協創情報研究部長，奈良先端科学技術大学院大学客員教授。電気通信普及財団賞受賞，電子情報通信学会論文賞，FIT 船井ベストペーパー賞等受賞。電子情報通信学会，IEEE 各会員。