

# 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, 山田 武士, 上田 修功

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

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。

## An efficient collaborative filtering using purchase orders

Tomoharu Iwata, Takeshi Yamada, Naonori Ueda

NTT Communication Science Laboratories, NTT Cooperation

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。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。本稿では、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られる。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言える。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得るが、前者はユーザに評価の負担を課し、後者は新商品の発売毎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。そこで本稿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[7]。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。簡単な例を挙げると、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[2, 4] しばしば使われるが、過去に購入した多くの商品の情報を無視してしまっている。

本稿では、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[3]。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。

## 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する。  $L$  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

$$\hat{P}(x|\mathbf{u}) = \frac{M(x, \mathbf{u}) + \delta}{M(\mathbf{u}) + \delta V}, \quad (1)$$

となる。ここで  $\delta$  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く、また、新たなデータが得られた際のパラメータ更新が容易にできる。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、 $P(x_1, \dots, x_L|x) = \prod_{l=1}^L P_l(x_l|x)$  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 [5] と呼ぶ。  $l$  ギャップマルコフモデルの MAP 推定値は、

$$\hat{P}_l(x'|x) = \frac{\sum_n \sum_k I((x_{n,k} = x) \wedge (x_{n,k-l} = x')) + \delta}{\sum_n \sum_k I(x_{n,k} = x) + \delta V}, \quad (2)$$

となる。  $I(y)$  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。

### 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や自然言語処理など幅広い分野で応用されている [3]。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

$$P(x|\mathbf{u}) = \frac{1}{Z(\mathbf{u})} \exp\left(\sum_{f \in F} \alpha_f y_f(x, \mathbf{u})\right), \quad (3)$$

となる。ここで、 $Z(\mathbf{u})$  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。

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言われており [6]、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 [3]。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り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時、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  $P(x)$  および  $l$  ギャップマルコフモデル  $P_l(x'|x)$  を推定する。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を設定する。最大エントロピー原理に従い、

上記の制約を条件として、 $P(x|\mathbf{u})$  のエントロピーを最大にする分布を求めると

$$\begin{aligned} P(x|\mathbf{u}) &= \frac{1}{Z(\mathbf{u})} \exp\left(\alpha_0 \log \hat{P}(x) + \sum_{l=1}^L \alpha_l \log \hat{P}_l(x_l|x)\right) \\ &= \frac{1}{Z(\mathbf{u})} \hat{P}(x)^{\alpha_0} \prod_{l=1}^L \hat{P}_l(x_l|x)^{\alpha_l}, \end{aligned} \quad (4)$$

を得る。ここで  $Z(\mathbf{u})$  は正規化項、 $\alpha = \{\alpha_l\}_{l=0}^L$  は未知パラメータ集合であり、 $\alpha_0$  は事前確率の重み、 $\alpha_l$  は  $l$  ギャップマルコフモデルの重みを表す。提案モデルは  $L$  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、重みが全て1のとき、各ギャップマルコフモデルが独立と仮定したモデルと一致する。

最大エントロピーモデルと同様に、対数尤度を最大化することにより未知パラメータ集合  $\alpha$  を推定できる。なお、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  $\alpha$  の推定に用いると過学習を起こす可能性があるため、交差検定法により  $\alpha$  を推定する必要がある。

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り返し最適化が必要なのは  $L+1$  個の重み  $\alpha$  のみである。そのため、 $O(LV^2)$  のパラメータの繰り返し最適化が必要な最大エントロピーモデルと比較し、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  $\alpha$  は以前に学習したものを使い、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。実データを用いた実験により、1ヵ月前までのデータで学習した重みを用いた場合でも、予測精度はそれほど低下しないことを確認した。

## 4 実験

提案法を評価するため、音楽、動画、漫画配信の3種類の実購買履歴データを用いた実験を行った。音楽、漫画データでは、データの最終日が異なる3つのデータセットを作成した。各データの開始日、終了日、ユーザ数、商品数、購買数を表1に示す。こ

表 1: 各データの開始日、最終日、ユーザ数、商品数、購買数。

	開始日	最終日	ユーザ数	商品数	購買数
Music4	05/4/1	05/4/30	247	132	1,508
Music5	05/4/1	05/5/31	1,120	348	7,588
Music6	05/4/1	05/6/30	2,104	561	15,216
Movie	07/1/1	07/1/1	3,085	1,569	25,363
Cartoon1	05/4/1	06/1/31	42,184	153	453,386
Cartoon2	05/4/1	06/2/28	53,830	161	599,196
Cartoon3	05/4/1	06/3/31	69,217	175	808,182

こで、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。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。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。

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。

**Markov2:** 2次マルコフモデル。

**Markov3:** 3次マルコフモデル。

**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。

**MaxEntSeq:** 購買順序を考慮した最大エントロピーモデル。

**MaxEnt:** 特徴量として購入済商品集合を用いた、購買順序を考慮しない最大エントロピーモデル。

**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 [7]。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。

**PLSA:** Probabilistic latent semantic analysis [1]。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 [5]。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  $P(x|\mathbf{u}) = \sum_{l=1}^L P(l)P_l(x|x_l)$  で表される。ここで  $P(l) \geq 0$  は混合係数で  $\sum_{l=1}^L P(l) = 1$ 。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, および, OurMethod は購買順序を考慮する手法であり, MaxEnt, Cosine, PLSA は購買順序を考慮しない手法である。

購買履歴  $\mathbf{u}$  が与えられたときの予測商品は、購買確率が最大の商品とする。評価尺度として正答率を用いた。表2に漫画、音楽、動画データにおける各手法の正答率を示す。GapMarkov, MaxEntSeq,

表 2: 正答率 (%)

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

Mixture, 提案法においては, 最大ギャップ  $L$  を 1 から 10 までで実験し, 最もよい  $L$  の正答率を表示している. 括弧内の値はそのときの  $L$  である. PLSA においては, 隠れクラス数  $Z$  を 5, 10,  $\dots$ , 50 とし て実験し, 最もよい  $Z$  の正答率を表示している. 括弧内の値はそのときの  $Z$  である.

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. 全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. つまり, ギャップマルコフモデルを線形形で結合した場合, 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.

Xeon3.6GHz, メモリ 2GB の計算機でパラメータ学習に必要な計算時間を比較した結果, 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 でも数十秒で終了した. 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が必要とするが, 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 倍高速であった.

## 5 むすび

本稿では, 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が高速にできるという利点を持つ. 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  $\alpha_i = \lambda \exp(-\lambda i)$  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] Hofmann, T.: Probabilistic Latent Semantic Analysis, *Proc. UAI*, pp. 289–296 (1999).
- [2] Iwata, T., Saito, K. and Yamada, T.: Recommendation methods for extending subscription periods, *Proc. KDD*, pp. 574–579 (2006).
- [3] Jin, X., Mobasher, B. and Zhou, Y.: A Web Recommendation System Based on Maximum Entropy, *Proc. ITCC*, pp. 213–218 (2005).
- [4] Jin, X., Zhou, Y. and Mobasher, B.: A maximum entropy web recommendation system: combining collaborative and content features, *Proc. KDD* (2005).
- [5] 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).
- [6] Raina, R., Shen, Y., Ng, A. Y. and McCallum, A.: Classification with Hybrid Generative / Discriminative Models, *Proc. NIPS* (2004).
- [7] Sarwar, B., Karypis, G., Konstan, J. and Reidl, J.: Item-based collaborative filtering recommendation algorithms, *Proc. WWW*, pp. 285–295 (2001).