

# 服飾品 EC サイトのデータに対する売上高最大化を目的とした推薦手法の提案

1X10C117-4 山上 敢  
指導教員 後藤 正幸

## 1 研究背景・目的

近年、情報技術の発展によりインターネット上の EC サイトで扱われるアイテム数が増えており、ユーザの嗜好に合ったアイテムを自動的に提示する推薦システムの重要性が増している。一般的な推薦システムでは、ユーザの嗜好に合ったアイテムを推薦することで購入されるアイテム点数の増加を目的としている。しかし、経営の視点に立つと購入されるアイテム点数を増加させることよりも、売上高を向上させることが重要であると考えられる。本研究では、売上高を最大化する推薦手法の提案を目的とする。

売上高を最大化するために、本研究では、各ユーザにアイテムを推薦するときに期待できる売上高を最大化する方法を考える。これはユーザが推薦されるアイテムを購入する確率を推定し、アイテムの価格との積和をとることで計算される。そのため、(購入確率) × (価格) の値が大きいアイテムから優先的に推薦することで、売上高最大化が達成できる。この値の推定精度は、アイテムの価格が不変であることから、購入確率の推定精度に依存する。そこで、一般的に高精度とされる購入確率の推定モデルである Aspect Model(以下 AM) [1] を導入すると共に、購入確率の推定精度向上により売上高最大化を目指す。

本研究では、アイテムの価格情報を持つ服飾品 EC サイトのデータを対象とし、対象データの特徴を考慮することにより、購入確率の推定精度を高め、売上高の向上を目指す。対象データの分析より、アイテムには流行があることと、ユーザは過去に閲覧したアイテムを購入する傾向があることが示されたことから、このアイテムの流行と閲覧履歴を考慮した購入確率の推定法を提案する。提案手法をこの服飾品 EC サイトのデータに適用し、その有効性を示す。

## 2 準備

### 2.1 Aspect Model を用いた推薦手法

AM はユーザとアイテムの間に潜在クラスを設ける確率的潜在クラスモデルである。このモデルでは、潜在クラスは典型的な嗜好のパターンを表し、その嗜好パターンとユーザ、アイテムとの関係を条件付き確率で表現する。

いま、 $I$  人のユーザの集合を  $\mathcal{X} = \{x_i : 1 \leq i \leq I\}$ 、 $J$  個のアイテムの集合を  $\mathcal{Y} = \{y_j : 1 \leq j \leq J\}$ 、 $K$  個の潜在クラスの集合を  $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$  とする。AM の確率モデルは式 (1) で示される。

$$P(x_i, y_j) = \sum_k P(z_k) P(x_i | z_k) P(y_j | z_k) \quad (1)$$

パラメータ  $P(z_k)$ ,  $P(x_i | z_k)$ ,  $P(y_j | z_k)$  は EM アルゴリズム [2] により推定され、ユーザ  $x_i$  に対するアイテム  $y_j$  の購入確率はこれらの推定値を用いて式 (2) で求められる。

$$\hat{P}(y_j | x_i) = \frac{\hat{P}(x_i, y_j)}{\sum_j \hat{P}(x_i, y_j)} \quad (2)$$

ただし、 $\hat{P}(x_i, y_j)$  は  $P(x_i, y_j)$  の推定値を表す。この購入確率の高いアイテム  $y_j$  を被推薦ユーザ  $x_i$  に推薦する。

### 2.2 データの特徴

本研究が対象とする服飾品 EC サイトには、購入履歴と閲覧履歴に加え、受注日やアイテムの価格といった様々な情報が蓄積されている。これらのデータの情報から、データに

適した購入確率の推定法を構築するために、対象データの詳細な分析を行った。

その結果から、二つの特徴が明確となった。一つ目は、服飾品は多く売れる時期と全く売れない時期があるアイテムが多いという点で、これを流行と呼ぶことにする。二つ目は、ユーザはアイテムを購入する前に何度かそのアイテムを閲覧する傾向があるという点である。これら二つの特徴を考慮することで、より精度の高い購入確率の推定を行うことができると考えられる。

## 3 提案手法

### 3.1 概要

売上高最大化の推薦を行うためには、各ユーザにアイテムを  $N$  個ずつ推薦するときの期待売上高が最大となる推薦を行う必要がある。いま、ユーザ  $x_i$  に対して推薦する  $N$  個のアイテム集合を  $\mathcal{S}_i$ 、アイテム  $y_j$  の価格を  $M(y_j)$  とすると、ユーザ  $x_i$  に  $N$  個のアイテムを推薦するときの期待売上高  $E(x_i)$  は式 (3) で表される。

$$E(x_i) = \sum_{y_j \in \mathcal{S}_i} P(y_j | x_i) M(y_j) \quad (3)$$

期待売上高  $E(x_i)$  はユーザ  $x_i$  に対して推薦するアイテム集合  $\mathcal{S}_i$  によって変化するが、 $P(y_j | x_i) M(y_j)$  を推薦基準とし、この値が大きいアイテムを順に  $N$  個推薦したとき、 $E(x_i)$  は最大となる。各ユーザの  $E(x_i)$  が最大となる推薦を行うことで、全ユーザについての売上高も最大となる。この推薦基準を求める際に、購入確率  $P(y_j | x_i)$  を精度よく推定する必要があるため、上述のデータの二つの特徴を考慮した購入確率の推定を行う。

一つ目に対しては、AM にアイテムの流行を考慮する方法を考える。そのために、AM のパラメータ学習に用いるデータの期間を分割して考える手法を提案する。「ユーザ間の嗜好の類似性は年間を通して変わらないがアイテムには流行がある」という仮定をおき、潜在クラスとユーザとの関係を表すパラメータは全期間の学習データを用いて推定するのに対し、潜在クラスとアイテムとの関係を表すパラメータは予測を行う期間の直近のデータのみを用いて再推定する。

二つ目に対しては、ユーザが閲覧したアイテムの購入確率が高くなるようなモデルを考える。閲覧されたアイテムに対しては、上述の流行を考慮した AM による購入確率に重みを加算することで、購入確率が高まるようにする。

### 3.2 流行を考慮した購入確率の推定方法

AM のパラメータ  $P(z_k)$ ,  $P(x_i | z_k)$  は、潜在クラスの生起確率とユーザの購入確率であり、なるべく多くのデータから推定した方が精度が高まる一方、アイテムの購入確率  $P(y_j | z_k)$  には流行の存在を考慮する必要がある。そこで、 $P(z_k)$ ,  $P(x_i | z_k)$  は全期間の購入履歴データを用いて推定し、一方、 $P(y_j | z_k)$  は予測期間の直近  $B$  ヶ月の購入履歴データのみを用いて推定し直す。これにより、潜在クラスとアイテムとの関係を表すパラメータは直近の流行を考慮することが可能となり、購入確率の推定精度が向上すると考えられる。

全期間の購入履歴データを用いて AM のパラメータ  $P(z_k)$ ,  $P(x_i | z_k)$ ,  $P(y_j | z_k)$  を推定した後、推定した  $\hat{P}(z_k)$ ,  $\hat{P}(x_i | z_k)$

と直近  $B$ ヶ月の購入履歴データを用いて  $\hat{P}(y_j|z_k)$  のみを式 (4) で推定し直す．

$$\hat{P}(y_j|z_k) = \frac{\sum_i n_B(x_i, y_j) \hat{P}(z_k|x_i)}{\sum_j \sum_i n_B(x_i, y_j) \hat{P}(z_k|x_i)} \quad (4)$$

$n_B(x_i, y_j)$  は直近  $B$ ヶ月の購入履歴データ内でユーザ  $x_i$  がアイテム  $y_j$  を購入している場合は 1, 購入していない場合は 0 の値を示すインジケータ関数であり,  $\hat{P}(z_k|x_i)$  はユーザ  $x_i$  の潜在クラス  $z_k$  への所属確率を表す．

全期間の購入履歴から推定した  $\hat{P}(z_k)$ ,  $\hat{P}(x_i|z_k)$ , 直近の購入履歴データを用いて式 (4) で推定した  $\hat{P}(y_j|z_k)$  を用いて式 (1), (2) によりユーザ  $x_i$  に対するアイテム  $y_j$  の購入確率を算出する．

### 3.3 閲覧履歴の考慮

ユーザの閲覧アイテムは購入されやすいという特徴から, 閲覧アイテムの購入確率を高めるモデルを考える．そこで, 3.2 節の方法で推定された購入確率  $\hat{P}(y_j|x_i)$  に重み  $\alpha$  を加算し, 基準化を行うことで新たな購入確率を算出する． $\alpha$  はユーザの閲覧アイテムをどれだけ優先するかを決める値であり,  $\alpha \geq 0$  とする． $\alpha = 0$  のとき, 従来法と同等になり,  $\alpha$  が大きいほど閲覧アイテムを優先的に推薦することになる．

いま, ユーザ  $x_i$  とアイテム  $y_j$  のペアを  $(x_i, y_j)$ , 予測期間の直近  $V$ カ月の閲覧履歴データ内に存在するユーザとアイテムのペアの集合を  $\mathcal{L}_V$ , ユーザ  $x_i$  の閲覧アイテム数を  $H_V(x_i)$  とすると, 閲覧を考慮した新たな購入確率  $\tilde{P}(y_j|x_i)$  を式 (5) で算出する．

$$\tilde{P}(y_j|x_i) = \begin{cases} \frac{\hat{P}(y_j|x_i) + \alpha}{1 + H_V(x_i)\alpha}, & (x_i, y_j) \in \mathcal{L}_V \\ \frac{\hat{P}(y_j|x_i)}{1 + H_V(x_i)\alpha}, & (x_i, y_j) \notin \mathcal{L}_V \end{cases} \quad (5)$$

### 3.4 売上高最大化のための推薦アルゴリズム

具体的な手順を以下に示す．

Step1) 全期間の学習用購入履歴データを用いて AM のパラメータ  $P(z_k)$ ,  $P(x_i|z_k)$ ,  $P(y_j|z_k)$  を推定する．

Step2) 直近  $B$ カ月の購入履歴データを用いて式 (4) によりパラメータ  $P(y_j|z_k)$  を推定し直し, 式 (2) により, 購入確率を算出する．

Step3) 閲覧の重みを考慮した新たな購入確率  $\tilde{P}(y_j|x_i)$  を式 (5) により算出する．

Step4) 購入確率  $\tilde{P}(y_j|x_i)$  と価格  $M(y_j)$  の積が大きいアイテム  $y_j$  をユーザ  $x_i$  に推薦する． □

## 4 実験

提案手法の有効性を評価するため, 経営科学系研究部会連合協議会主催, 平成 25 年度データ解析コンペティションで提供されたデータの大手服飾品 EC サイトの購入履歴データと閲覧履歴データを用いて実験を行う．

### 4.1 実験条件

学習には 2011 年 9 月 1 日から 2013 年 2 月 28 日までの購入履歴データを使用し, ユーザ数  $I = 93,865$ , アイテム数  $J = 292,635$ , 購入数は 715,704 件である．予測には 2013 年 3 月 1 日から 3 月 31 日までの購入履歴データを使用し, 購入数は 10,635 件である．流行を考慮するためのデータ期間, ユーザの過去の閲覧アイテムのデータ期間は, それぞれ  $B=1, V=1$  とし, 2013 年 2 月 1 日から 2 月 28 日までの購入履歴データ, 閲覧履歴データを使用する．予備実験により, 潜在クラス数  $K = 10$ , 閲覧の重み  $\alpha = 0.01$  とした．

通常の AM による推薦を従来手法, アイテムの流行と閲覧の重みを考慮した購入確率を推薦基準とする推薦を提案 1,

さらにその購入確率と価格の積を推薦基準とする推薦を提案 2 として比較を行った．

### 4.2 評価方法

本研究の目的は売上高最大化であるので, 評価指標は推薦されたアイテムの中でテストデータ中でも実際に購入されているアイテムの価格の総和を用いる．売上高  $R$  は, 式 (6) で算出する．

$$R = \sum_i \sum_j \delta(x_i, y_j) M(y_j) \quad (6)$$

ここで,  $\delta(x_i, y_j)$  はユーザ  $x_i$  に推薦したアイテム  $y_j$  の中でテストデータ中で実際に購入されている場合は 1, 購入されていない場合は 0 の値を示すインジケータ関数である．

また, 購入確率の推定精度を測るために TopN 精度を用いる．TopN 精度は, 推薦された全  $NI$  個のアイテムの中でテストデータ中でも実際に購入されているアイテムの割合である．

### 4.3 実験結果と考察

各手法での Top10 精度を図 1, 売上高を図 2 に示す．

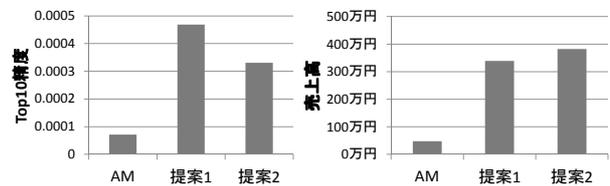


図 1. Top10 精度の比較

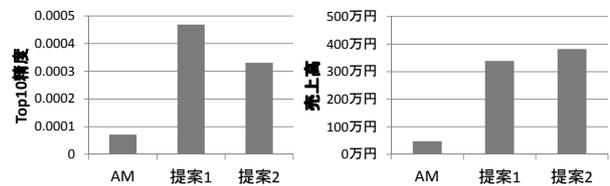


図 2. 売上高の比較

図 1 より, 従来手法の AM よりも両提案の精度が高いことがわかる．これは, 服飾品データの特徴を考慮することで購入確率の推定精度が向上し, 購入された個数を増加させることができたためであると考えられる．また, 提案 1 と提案 2 では, 提案 1 の方が推薦精度は高い．これは, 単純にアイテムの予測の精度のみを考えた場合には, 購入確率  $\tilde{P}(y_j|x_i)$  の高いアイテム  $y_j$  を推薦することが最良であるためである．

図 2 より, 従来手法よりも両提案の売上高が高いことがわかる．これは, 購入された個数が増加しているためであると考えられる．また, 提案 2 は提案 1 に比べ, 購入された個数が少なくなった反面, 全体の売上高は高くなっている．売上高を購入された個数で割ったアイテムの平均単価を調べると, 提案 1 では 7,708 円, 提案 2 では 12,308 円と提案 2 の方が 1.6 倍程度高かった．推薦基準の大小関係が価格の乗算により入れ替わり, 価格の高いアイテムが多く推薦されていると考えられる．提案 2 によりアイテムの価格を考慮することが売上高の向上に有効であるといえる．

### 5 まとめと今後の課題

本研究では服飾品 EC サイトの推薦システムにおいて売上高の最大化を目的とし, アイテムの流行と閲覧履歴を考慮した購入確率の推定方法, 並びにアイテムの価格情報を用いた推薦基準を提案し, 実データを用いた実験により有効性を示した．今後の課題は, 購入にどれだけ閲覧情報が寄与しているかを表す閲覧の重みの決定法の検討などが挙げられる．

#### 参考文献

- [1] T. Hofmann, and J. Puzicha, “Latent Class Models for Collaborative Filtering,” *Proc. 16th International Joint Conference on Artificial Intelligence*, pp.688–693, 1999.
- [2] 宮川雅巳, “EM アルゴリズムとその周辺,” *応用統計学*, Vol.16, No.1, pp.1–21, 1987.