

クーポン付購買履歴データを用いた顧客購買行動分析

1X11C101-5 松崎 祐樹
指導教員 後藤 正幸

1 研究背景・目的

近年、実際の小売店舗においても、ICカードによるポイント付与システムなどを活用し、顧客の購買履歴データが取得可能になっている。その活用の1つとして、ユーザ別にカスタマイズされたアイテムの割引クーポンを発券し、次の来店を促す購買促進施策がある。このようなクーポン付アイテムは、発券数が限定的である上に値引販売されるため、それ自体は売上高に占める割合が高くない。そのため、クーポン付アイテムの購入が他の通常アイテム購入に結び付くか否かが全体の売上高向上に重要となる。従って、このようなクーポン付アイテムと同時購入される通常アイテムに関するユーザの購買行動をモデル化し、通常アイテムを含めたレシート1枚あたりの購買金額の期待値を高くするようにクーポンを発券することができれば、売上高の向上が期待できる。

他方、ユーザやアイテムの異質性を表現できる手法として、ユーザの購買行動が潜在的な嗜好に従うと仮定し、潜在クラスを用いてモデル化した手法 [1] が存在する。しかしながら、これはクーポン付アイテムと通常アイテムの同時購買行動を表現したモデルにはなっておらず、本研究で対象とするユーザの購買行動を正しく表現することができない。

そこで本研究では、クーポン付アイテムと通常アイテムの同時購買を含むユーザの購買行動は潜在的な嗜好によるものであると仮定し、潜在クラスモデルによって両者の同時購買を表現可能なモデルを構築する。このモデルにより、クーポン付アイテムと通常アイテムの同時購買確率が算出可能となる。その上で、クーポンの発券によるアイテムの値引き購買をクーポン付アイテム推薦による購買と捉え、クーポン付アイテムによる通常アイテムの同時購買金額を最大化する推薦手法を提案する。また、大手スーパーの購買履歴データに対し提案手法によるモデル化と推薦を行い、現状の推薦アイテムよりもよいクーポン付アイテムの存在を示すことにより、その有効性を示す。

2 準備

2.1 基本データ分析

本研究が対象とする大手スーパーでは、蓄積された購買履歴を用いて毎月ユーザごとに異なるアイテムの割引クーポンを発券している。このクーポンの利用有無に着目し、クーポン付アイテムがユーザの通常アイテムの同時購買に与える影響について分析を行った。その結果から、クーポン付アイテムを含んだ購買の平均単価が高いこと、クーポン付アイテムごとに同時に購入される合計金額の平均値に差があることが明らかとなった。これらのことから、クーポン付アイテムが同時購買金額増加に対して効果的であること、同時購買金額を増加させるために適したクーポン付アイテムが存在することが示唆されている。

2.2 潜在クラスを用いた購買行動のモデル化

ユーザの嗜好やアイテムの特徴の類似性を潜在クラスを用いて定式化し、ユーザの購買行動をモデル化した手法として Goto らによる手法 [1] が存在する。いま、 I 人からなるユーザ集合を $U = \{u_i : 1 \leq i \leq I\}$ 、 N 個のアイテム集合を $C = \{c_n : 1 \leq n \leq N\}$ 、 L 個の潜在クラス集合を $Z = \{z_l : 1 \leq l \leq L\}$ 、購買履歴を $W = \{w_i : 1 \leq i \leq I\}$ とする。ただし、 $w_i = (w_1^i, w_2^i, \dots, w_N^i)$ はユーザ u_i の購買

有無ベクトルとし、 w_n^i は、ユーザ u_i がアイテム c_n を購買していれば 1、それ以外は 0 をとる二値変数である。また、ユーザ u_i と購買履歴 w_i に対応する潜在クラスを $v_i \in Z$ とすれば、 i 番目の完全データ (u_i, w_i, v_i) の確率モデルは式 (1) で示される。

$$\begin{aligned} P(u_i, w_i, v_i) &= P(v_i)P(u_i|v_i)P(w_i|v_i) \\ &= P(v_i)P(u_i|v_i) \prod_{n=1}^N P(c_n|v_i)w_n^i P(\bar{c}_n|v_i)^{1-w_n^i} \quad (1) \end{aligned}$$

ただし、 c_n は「アイテム c_n を購買する」、 \bar{c}_n は「アイテム c_n を購買しない」という事象であり $P(c_n|v_i)$ には二項分布を仮定している。ここでは、 $v_i \in Z$ であるため、式 (1) のパラメータは $P(z_l)$ 、 $P(u_i|z_l)$ 、 $P(c_n|z_l)$ となり、これらのパラメータを EM アルゴリズム [2] により推定する。

3 提案手法

3.1 概要

本研究で対象とするデータは、各購買行動ごとに、クーポン付アイテムの購買と通常アイテムの購買があり、ユーザの購買行動ごとに購買履歴（以下レシート）が存在する。これらの特徴を適切に表現するため、クーポン付アイテムと通常アイテムの同時購買と、レシートごとの購買行動が表現可能なモデルを構築する。このため、従来手法 [1] を以下の2点で拡張する。1点目として、従来手法で定義されていたアイテム集合 C を、クーポン付アイテム集合と通常アイテム集合の、互いに排反な2つの集合として定義する。2点目として、各ユーザに対し全購買履歴を1つの事象と定義する従来手法に対して、複数回の購買履歴を表現できるモデルとする。これにより、売上高最大化のためにクーポン付アイテム推薦時の期待同時購買金額が最大となるような推薦が可能となる。

3.2 提案モデルへの拡張

従来手法では、アイテム集合 C は全てのアイテムを要素とする集合であるため、クーポンの有無を考慮できない。そこで、アイテム集合 C を、 J 個の通常アイテム集合 $A = \{a_j : 1 \leq j \leq J\}$ と K 個のクーポン付アイテム集合 $B = \{b_k : 1 \leq k \leq K\}$ として定義する。これによってクーポン付アイテムと通常アイテムを区別したモデル化が可能となる。加えて、全ユーザの全レシートを要素とする購買履歴集合を $X = \{x_r : 1 \leq r \leq R\}$ と定義する。ただし、 x_r は r 番目のレシートを表すものとし、通常購買ベクトル x_{rA} 、クーポン付購買ベクトル x_{rB} により $x_r = (x_{rA}, x_{rB})$ と定義する。また、ユーザは少なくとも1つのレシートを持つため $R \geq I$ である。ここで、 r 番目のレシートに着目し、購買したユーザを $y_r \in U$ 、所属する潜在クラスを $v_r \in Z$ とすると、 r 番目の完全データは、 $(y_r, x_{rA}, x_{rB}, v_r)$ と表され、その確率モデルは式 (2) のようになる。式 (2) における、 $x_{a_j}^r, x_{b_k}^r$ は、 r 番目のレシートでアイテム a_j, b_k を購買した場合 1、それ以外は 0 をとる二値変数である。

$$\begin{aligned} P(y_r, x_{rA}, x_{rB}, v_r) &= P(v_r)P(y_r|v_r)P(x_{rA}|v_r)P(x_{rB}|v_r) \\ &= P(v_r)P(y_r|v_r) \prod_{j=1}^J P(a_j|v_r)x_{a_j}^r P(\bar{a}_j|v_r)^{1-x_{a_j}^r} \\ &\quad \times \prod_{k=1}^K P(b_k|v_r)x_{b_k}^r P(\bar{b}_k|v_r)^{1-x_{b_k}^r} \quad (2) \end{aligned}$$

さらに、合計 R 枚ある全レシートを表現した確率モデルは、式 (2) の積により表され、式 (3) により表現できる。

$$P(\mathbf{Y}, \mathbf{X}_A, \mathbf{X}_B, \mathbf{V}) = \prod_{r=1}^R P(y_r, \mathbf{x}_{rA}, \mathbf{x}_{rB}, v_r) \\ = \prod_{r=1}^R P(v_r)P(y_r|v_r)P(\mathbf{x}_{rA}|v_r)P(\mathbf{x}_{rB}|v_r) \quad (3)$$

ただし、 $\mathbf{Y} = (y_1, y_2, \dots, y_R)$, $\mathbf{X}_A = (\mathbf{x}_{1A}, \mathbf{x}_{2A}, \dots, \mathbf{x}_{RA})$, $\mathbf{X}_B = (\mathbf{x}_{1B}, \mathbf{x}_{2B}, \dots, \mathbf{x}_{RB})$, $\mathbf{V} = (v_1, v_2, \dots, v_R)$ とする。 $\mathbf{Y} = (y_1, y_2, \dots, y_R)$ の中でユーザ u_i は重複して出現し得るため、ユーザの購買行動をレシートごとに表現できる。

3.3 EM アルゴリズムによるパラメータの推定

提案モデルにおけるパラメータ $P(z_i)$, $P(u_i|z_i)$, $P(a_j|z_i)$, $P(b_k|z_i)$ の推定は EM アルゴリズム [2] により行う。E-step, M-step の更新はそれぞれ、式 (4), 式 (5)–(8) で表現できる。

【E-step】

$$P(z_i|y_r, \mathbf{x}_{rA}, \mathbf{x}_{rB}) = \frac{P(y_r, \mathbf{x}_{rA}, \mathbf{x}_{rB}, z_i)}{\sum_{z_i \in Z} P(y_r, \mathbf{x}_{rA}, \mathbf{x}_{rB}, z_i)} \quad (4)$$

【M-step】

$$P(z_i) = \frac{1}{R} \sum_{r=1}^R P(z_i|y_r, \mathbf{x}_{rA}, \mathbf{x}_{rB}) \quad (5)$$

$$P(u_i|z_i) = \frac{1}{RP(z_i)} \sum_{r=1}^R \delta(y_r = u_i) P(z_i|y_r, \mathbf{x}_{rA}, \mathbf{x}_{rB}) \quad (6)$$

$$P(a_j|z_i) = \frac{1}{RP(z_i)} \sum_{r=1}^R P(z_i|y_r, \mathbf{x}_{rA}, \mathbf{x}_{rB}) x_{rA_j}^r \quad (7)$$

$$P(b_k|z_i) = \frac{1}{RP(z_i)} \sum_{r=1}^R P(z_i|y_r, \mathbf{x}_{rA}, \mathbf{x}_{rB}) x_{rB_k}^r \quad (8)$$

ただし、 $\delta(y_r = u_i)$ は $y_r = u_i$ ならば 1, それ以外は 0 をとるインジケータ関数である。E-step, M-step を完全データの対数尤度が収束するまで繰り返し、パラメータを推定する。

3.4 期待同時購買金額を最大化する推薦

期待同時購買金額最大化のために、提案モデルを用いて、クーポン付アイテムと同時購買されるアイテムの合計金額の期待値を算出する。いま、 $M(a_j)$ をアイテム a_j の価格とし、ユーザ u_i にクーポン付アイテム b_k を推薦した際に購買される通常アイテムの期待同時購買金額を式 (9) で定義する。

$$E(u_i, b_k) = \sum_j M(a_j) \times P(a_j|u_i, b_k) \quad (9)$$

ただし、 $P(a_j|u_i, b_k)$ は、ユーザ u_i がクーポン付アイテム b_k を購買しているもとで通常アイテム a_j を同時購買する確率であり、式 (10) で求められる。

$$P(a_j|u_i, b_k) = \frac{\sum_z P(z)P(a_j|z)P(u_i|z)P(b_k|z)}{\sum_z P(z)P(u_i|z)P(b_k|z)} \quad (10)$$

推薦アイテムを選出する際には、式 (9) に加えて、ユーザ u_i がクーポン付アイテム b_k を購買する確率 $P(b_k|u_i)$ を考慮し、式 (11) の値が高いアイテムから順に推薦する。

$$E^l(u_i, b_k) = E(u_i, b_k) \times P(b_k|u_i) \quad (11)$$

4 実験・考察

提案モデルの有効性評価のため、経営科学系研究部会連合協議会主催、平成 26 年度データ解析コンペティションで提供された大手スーパーのある 1 店舗における 2013 年 7 月から 2014 年 6 月のある店舗の購買履歴データを用いて実験を行った。

提案モデルの適用により、実際に推薦されたクーポン付アイテムよりも、期待同時購買金額を増加させるという点で、適切なクーポン付アイテムが存在することを示す。テスト期間において、各ユーザに実際に推薦された 20 個のクーポン付アイテムと、式 (11) の値が高い順に選出した 20 アイテムについて、式 (9) を用い、ユーザごとに期待同時購買金額の平均値を算出し、その平均値どうしを比較した。

実験には、2013 年 7 月 1 日から 2014 年 5 月 31 日までの 11ヶ月間を学習期間、2014 年 6 月 1 日から 30 日までをテスト期間とする。学習のために 1,000 人のユーザをランダムサンプリングし、上述の期間によって学習データとテストデータに分割した。なお、学習データにおけるユーザ数は $I=942$, 通常アイテム数は $J=10,116$, クーポン付アイテムは $K=390$, レシート数は $R=18,477$ である。

実験結果として、潜在クラス数を 2 から 10, 15, 20, 25, 30 とした際の、実際に推薦されたアイテム (従来) と期待同時金額が高い順に選出したアイテム (提案) のそれぞれについて算出した期待同時購買金額の平均値を図 1 に示す。

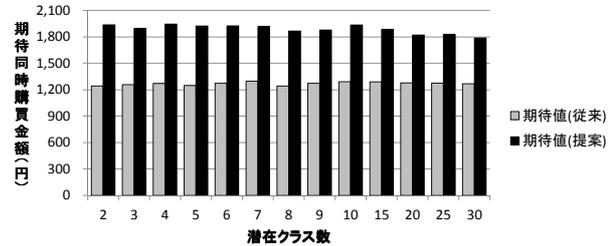


図 1. 実験結果

図 1 から、提案手法によるクーポン付アイテム推薦により、アイテム推薦時の 1 レシートにおける期待同時購買金額が実際に推薦されたクーポン付アイテムの期待同時購買金額に比べ 540 円から 690 円程度向上することが分かる。これにより、実際にユーザに推薦されたクーポン付アイテムよりも、同時購買金額を増加させるようなクーポン付アイテムが存在することが明らかとなった。これにより、提案手法によるクーポン付アイテムの推薦を行うことで、同時購買金額の向上、延いては、売上高の向上が期待できると言える。また、潜在クラス数 4 の場合において期待値 (提案) が最も高くなるが、これは潜在クラス数を適切な値とすることにより、潜在的なユーザの嗜好の異質性やアイテムの類似性をよりよく表現できたためであると考えられる。

5 まとめと今後の課題

本研究では、大手スーパーの購買履歴データを用いて通常アイテムとクーポン付アイテムの同時購買を考慮したモデルを提案した。また、実際の購買履歴データを用いて、同時購買金額を増加させるという点でよりよいクーポン付アイテムの存在を明らかにした。今後の課題として、実際にクーポンを発券した際の顧客の反応や、より正確な同時購買金額の算出などが挙げられる。

参考文献

- [1] M. Goto, K. Minetoma, K. Mikawa, M. Kobayashi, S. Hirasawa, "A Modified Aspect Model for Simulation Analysis," *Proc. IEEE International Conference on Systems Man and Cybernetics*, pp.1306-1311, 2014
- [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.