

レコメンデーションに誘導されやすい顧客の抽出方式と評価

恵津森 真仁[†] 高山 毅[†] 村田 嘉利[†] 佐藤 永欣[†]
岩手県立大学ソフトウェア情報学部[†]

1. はじめに

近年、売り上げの向上を目的とする「レコメンデーション」への注目が高まっている。これは商品や顧客ごとの特性に注目して、購入される可能性が相対的に高い商品を、店側からのプッシュ型サービスとしてお勧めするものである。以降では、レコメンデーションを行なうことを単に、「レコメンドする」と書く。

著者らはこれまで、商品購入の前後関係を考慮してレコメンドする方式を提案している。そして、商品の購入履歴のアンケートを用いたシミュレーションで、有効との評価を得ている¹⁾。それを受けて著者らは、デパートの購入履歴の実データからレコメンドし、有効な結果を得ることを目指している²⁾。デパートの顧客へ DM を送付してレコメンドする場合、時系列パターンマイニング³⁾の知見を利用することが考えられる。すなわち、 $A \rightarrow B$ という時系列パターン(以降単にパターンと呼ぶ)が頻出していれば、 A の購入者集合 $Cust(A)$ に B をレコメンドする。しかしながら、 $Cust(A)$ のうちどの顧客へレコメンドするのが有効かは、これまで明らかにされていない。

$Cust(A)$ の中には、 $A \rightarrow B$ のレコメンデーションに誘導されやすい顧客 $Cust_Y(A \rightarrow B)$ と、そうでない顧客 $Cust_N(A \rightarrow B)$ が混在している。実際にはこれら二者は相対的であるが、境界があると仮定して、 $Cust(A)$ を $Cust_Y(A \rightarrow B)$ と $Cust_N(A \rightarrow B)$ に二分して考えることにする。レコメンデーションの実施前に $Cust_Y(A \rightarrow B)$ を抽出できれば、レコメンド精度の向上に有用である。本稿では、パターンごとに誘導されやすい顧客 $Cust_Y(A \rightarrow B)$ を抽出する方式を提案、評価する。

2. 誘導されやすい顧客の抽出方式

2.1 分析単位

本稿では、データの単位として「品番」という概念を採用する。一部例外はあるが、品番とはデパート内の個々のお店、売場と考えて良い。本稿では、共同研究中のデパート X 社との協議より、個々の商品単位ではなく品番単位で議論を進める。

2.2 参考期と誘導期

デパートの場合、商品に季節性が含まれるため、頻出するパターン $A \rightarrow B$ は、時期によって変動する。そこで、パターンの前側として考える期間を「参考

期」、後ろ側として考える期間を「誘導期」と呼ぶことにする。たとえば、参考期を 3/1~4/30、誘導期を 7/1~8/31 としたときに、過去の年に頻出するパターン $A \rightarrow B$ を使うと、以下のことが考えられる。すなわち、夏期に直近の春期の購入履歴をもとにしてレコメンデーションを実施することである。

2.3 吉兆度の導入

一般に、頻出パターン $A \rightarrow B$ の導出は過去のデータを用いて行なわれる。各顧客は、参考期に A に加えて A, B 以外の品番 $H_i (i=1, \dots, n-2; n$ は品番の総数) のいずれかでも購入していることが少なくない。そこで本稿では、参考期にどの品番 H_i でも購入していると、誘導期に B で購入しやすいかを考える。まず、変数 $Y_num(A \rightarrow B, H_i)$, $N_num(A \rightarrow B, H_i)$ を以下で定義する。

- $Y_num(A \rightarrow B, H_i)$: 「ある参考期に品番 A かつ H_i で購入した顧客が、対応する誘導期に品番 B で購入した」という事実が発生した回数
- $N_num(A \rightarrow B, H_i)$: 「ある参考期に品番 A かつ H_i で購入した顧客が、対応する誘導期に品番 B で購入しなかった」という事実が発生した回数

また、パターン $A \rightarrow B$ に対する品番ごとの吉兆度 $K(A \rightarrow B, H_i)$ を、以下により定義する。

$$K(A \rightarrow B, H_i) = Y_num(A \rightarrow B, H_i) / N_num(A \rightarrow B, H_i)$$

ただし、 $Y_num(A \rightarrow B, H_i) = N_num(A \rightarrow B, H_i) = 0$ の場合には、 $K(A \rightarrow B, H_i) = 0$
 $Y_num(A \rightarrow B, H_i) \neq 0$ かつ $N_num(A \rightarrow B, H_i) = 0$ の場合には、 $K(A \rightarrow B, H_i) = \infty$

簡素化した例で、具体例を示す。

- A, B 以外の品番としては、全体で $H_1 \sim H_4$ の四つのみ
- $Cust(A)$ の要素は、 $C_1 \sim C_3$ の 3 名のみ
- $C_1 \sim C_3$ の各購入履歴を表 1 とする。その場合、品番ごとの $Y_num(A \rightarrow B, H_i)$, $N_num(A \rightarrow B, H_i)$, 吉兆度 $K(A \rightarrow B, H_i)$ の各値は、表 2 の通りである。

表 1 $Cust(A)$ に含まれる全顧客 $C_1 \sim C_3$ の購入履歴

顧客	参考期	誘導期
C_1	A, H_1	B, H_2
C_2	H_1, A, H_2, H_1	B, B, H_3
C_3	A, H_3, H_2	H_2, H_4

2.4 顧客抽出とレコメンデーション

表 2 の $K(A \rightarrow B, H_i)$ より、たとえば参考期に A の他に H_3 でも購入した顧客よりは、 H_1 でも購入した顧客の方が、誘導期の B での購入に期待が持てる。そ

Proposition and Evaluation of Expectable Customers Picking up Method in a Recommendation.

[†] M. Etsumori, T. Takayama, Y. Murata, and N. Sato (Faculty of Software and Information Science, Iwate Prefectural University)

ここで吉兆度が相対的に上位にあり、レコメンデーション時に吉兆と見なして利用する品番のことを、「吉兆品番」と呼ぶことにする。上位のどこまでを吉兆品番と見なすかは、抽出すべき人数によって変化させる。本稿では、誘導期のBでの購入のレコメンデーションを、参考期にAに加え吉兆品番で購入している顧客へ行なうことを提案する。

表2 パターンA→Bに対する品番ごとの Y_num, N_num, K の値

品番	Y_num	N_num	K
H ₁	3	0	∞
H ₂	2	1	2
H ₃	0	1	0
H ₄	0	0	0

3. 評価実験

3.1 実験方法

表3の条件で抽出した表4の頻出4パターンでレコメンドする。

表3 評価実験の各種条件

分析対象顧客	X社のメンバーズカード保持者で、Z店での購入履歴のある顧客
参考期	3/1~8/31
誘導期	11/1~11/30
頻出パターン、吉兆度の導出	2005、2006年のデータから
Cust(A)の作成	2007年のデータから

表4 利用する4パターンと吉兆品番

	参考期側品番	誘導期側品番	吉兆品番
(1)	1(婦人・子供衣料系)	2(ファミリー衣料系)	6(婦人キャリア系)
(2)	1(婦人・子供衣料系)	3(婦人肌着・靴下系)	7(婦人スポーツ系) 8(婦人カジュアル系)
(3)	3(婦人肌着・靴下系)	4(婦人ミセス系)	9(服飾雑貨系)
(4)	5(婦人下着系)	4(婦人ミセス系)	9(服飾雑貨系) 10(身の回り品系)

各パターンごとに、参考期側品番で購入しかつ、

- 吉兆品番のいずれかで購入している顧客 100名
- 吉兆品番のいずれでも購入していない顧客 100名

を用意し、両者に誘導期側品番をDM送付によりレコメンドする。そして誘導期での誘導期側品番の購入者数mを実測し、

$$\text{レコメンド精度(\%)} = (m/100) \times 100$$

を比較する。

3.2 実験結果

実験結果をパターンごとに表5に示す。4パター

ンとも、吉兆品番での購入の方がレコメンド精度が高いという結果を得た。パターン(3)、(4)のレコメンド精度が(1)、(2)より相対的に低い原因としては、以下のことが考えられる。すなわち、過去の年においてもこれらのパターンのサポート値³⁾は相対的に低い。

表5 実験結果

	参考期の吉兆品番での購入	レコメンド精度(%)
(1)	あり	15
	なし	7
(2)	あり	25
	なし	5
(3)	あり	1
	なし	0
(4)	あり	2
	なし	0

4. まとめと今後の展望

本稿では、デパートの顧客へDMを送付してレコメンデーションを行なう場合に、時系列パターンごとに誘導されやすい顧客を抽出する方式を提案した。具体的には、その抽出に吉兆度という概念を用いることを提案した。実験結果によれば、提案方式は有効である。

関連研究として、繰返し購入の促進に有効な文献⁴⁾が挙げられる。一方、本稿の提案方式は、各顧客が購入する品番の範囲を拡充するために有効である。

今後の展望として、(i)参考期における顧客ごとの全購入履歴からの各顧客の序列化、(ii)(i)に基づき、ヒット率を最適化しての、DM送付の効率化、(iii)データ量の拡充、等が考えられる。

参考文献

- 1) 高山毅, 小熊浩史, 池田哲夫, 三浦良介: 商品購入の前後関係を考慮したレコメンド方式と評価, 電子情報通信学会「データ工学」研究会技術報告, DE2004-5, pp. 25-30, 2004.
- 2) 川口健悦, 徳野成之, 高山毅, 村田嘉利, 池田哲夫: 商品購入の前後関係を考慮したレコメンド方式のデパートへの応用と評価実験, 情報処理学会第69回全国大会 2S-8, 2007.
- 3) Agrawal, R. and Srikant, R.: "Mining Sequential Patterns," Proceedings of the 11th International Conference on Data Engineering, pp. 3-14, 1995.
- 4) 裴明花, 谷口伸一, 原隆浩, 西尾章治郎: 重要な顧客層および相関ルール発見のための繰返し購買パターンを考慮した相関ルールマイニング, 情報処理学会論文誌, Vol. 47, No. 12, pp. 3352-3364, 2006.