

レコメンデーションに誘導されやすい顧客を抽出する「吉兆度方式」の効率化

高島 隼也† 高山 毅† 村田 嘉利† 佐藤 永欣† 加藤 大樹†
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。レコメンデーションとは、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、店側からプッシュ型サービスとして顧客側へお勧めするものである。著者らの研究グループでは、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して、吉兆度方式を提案している[1]。本稿では、この吉兆度方式の効率化を図る。具体的には、RFM 分析[2]を吉兆度方式へ融合して、レコメンデーションに誘導されやすい顧客の抽出精度の向上を目指す。

ここで本研究では、データの単位として「品番」という概念を採用している。一部例外はあるが、品番とはデパート内の個々のお店、売り場と考えるとよい。本研究では、共同研究中のデパート X 社との協議により、個々の商品単位ではなく品番単位で議論を進める。

2. 先行研究

2.1 吉兆度方式[1]

2.1.1 参考期と誘導期

顧客の購入履歴で頻出する前後関係パターン(以降単に「パターン」と呼ぶ)A→B は、時期によって変動し得る。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」と呼んでいる。例えば、参考期を 3/1~5/31、誘導期を 7/1~8/31 とし、過年に頻出するパターン A→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基にしたレコメンデーションの実施である。

2.1.2 両側圧縮法と参考期圧縮法

単一の顧客の購入履歴で、参考期に A が count(A)回、誘導期に B が count(B)回登場する場合、パターン A→B の発生回数のカウント法として四通りを提案している。本稿ではこのうち、以降に示す両側圧縮法と参考期圧縮法を用いる。両側圧縮法とは、「count(A)≥1 かつ count(B)≥1」を満たす顧客人数を、パターン A→B の発生回数とする。また参考期圧縮法とは、「count(A)≥1」である顧客の count(B)を全顧客について足した値を、パターン A→B の発生回数とする。

2.1.3 吉兆度

参考期に A に加えてどの品番 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示す尺度として、以下に示す「吉兆度」を提案している。すなわち、ある参考期に品番 A かつ H_i で購入した顧客 C_{AHi} が、対応する誘導期に品番 B で

- 購入という事実が発生した回数: $Y_num(A \rightarrow B, H_i)$
 - 非購入との事実が発生した回数: $N_num(A \rightarrow B, H_i)$
- とした上で、パターン A→B に対する品番 H_i の吉兆度 $K(A \rightarrow B, H_i)$ は、 $C_j \in C_{AHi}$ なるすべての顧客 u 名についての下式である。

$$K(A \rightarrow B, H_i) = \sum_{j=1}^u \frac{Y_num_j(A \rightarrow B, H_i)}{Y_num_j(A \rightarrow B, H_i) + N_num_j(A \rightarrow B, H_i)} \quad \dots (1)$$

そして H_i は、より粗粒度化して品番群としても良いこととしている。品番群とは、デパート側の組織構成に基づき、同一カ

テゴリに属する品番を 1 グループにまとめたものである。

2.1.4 パターン A→B での顧客 C_k の有望度合

直近の参考期に品番 A で顧客 C_k が購入した場合、直後の誘導期での C_k の品番 B での購入の有望度合 E_k は、直近の参考期に C_k が購入した A 以外の品番を H_{ik} とするとき、

$$E_k = \sum_{ik} K(A \rightarrow B, H_{ik}) \quad \dots (2)$$

とする。

2.2 RFM 分析[2]

RFM 分析では、

- R(recency): 一番最後に買ったのはいつか
- F(frequency): どのくらいの頻度で買っているか
- M(monetary): いくら使っているか

の三つの観点から顧客を指標化し、優良顧客を選出する。三尺度のいずれに注目するかや、それぞれをどう重み付けするかは、通常、店側が判断し定める。

3. 「RFM 分析」の「吉兆度方式」への融合

ある品番のレコメンデーションを行なう場合、RFM 分析を用いると有望顧客 N 名を抽出することができる。このとき、顧客 k の有望度合を、0.0~1.0 を定義域とする関数 $r(k)$ で表現することができる。一方、吉兆度方式を用いても、有望顧客 N 名を抽出することができる。この場合、顧客 k の有望度合 $E(k)$ も、0.0~1.0 で表現することができる。

そこで、吉兆度方式へ RFM 分析を融合した場合の、顧客 k の有望度合 $rE(k)$ を、

$$rE(k) = r(k) \times E(k) \quad \dots (3)$$

によって定める。

なお、各手法で抽出される有望顧客 N 名は、一般に完全には一致しない。そこで、融合した場合の有望顧客 $M(\leq N)$ 名を、 $rE(k)$ の値に基づき定める。

4. 評価実験

4.1 方法

顧客へ「その顧客が過去に未購入の品番」をレコメンドし、そこでの購入の実現、すなわちその顧客の購入品番の範囲を拡張するタスクで実験する。なお一般には、このタスクの成功は容易とは言えない。X社側との協議に基づき、表 1 の条件で実験を進める。

表 1 実験条件

対象顧客の母集団	X社のカード会員かつ友の会会員で、Z店での購入履歴のある顧客 40,597 名
対象品番の母集団	294 品番
参考期	06/01~08/31
誘導期	11/1~11/30
レコメンデーション前の分析	2008, 9年の各参考期, 誘導期, および 2010年の参考期から

表 2 の誘導期側品番の列は、レコメンデーションを行なう 16 の品番を示している。以下、

- ①RFM 分析の結果に基づく有望顧客
- ②吉兆度方式での有望顧客
- ③RFM 分析を吉兆度方式へ融合した場合の有望顧客

の選定方法を述べる。なお、三手法とも始めに、過去半年以内に誘導期側の品番 B_j で購入している顧客を除外し、残った顧客を分析対

Efficiency Improvement of Auspicious Level Method Which Picks up Expectable Customers in a Recommendation
†J.Takahima, T. Takayama, Y. Murata, N. Sato, and H.Kato
(Faculty of Software and Information Science, Iwate Prefectural University)

象顧客集合 $\{C(B_v)\}$ とする。

①RFM 分析の結果に基づく有望顧客

X 社では、RFM を各 5 段階で評価している。X 社との協議に基づき、M のみに 2 倍の重みを持たせる。その結果、R, F, M の三つの積は、2~250 点となる。これを正規化し、

$$r(k) = \frac{(R \times F \times 2M) - 2}{248} \quad \dots(4)$$

により、各顧客の有望度を定める。

②吉兆度方式に基づく有望顧客

(2)式の値を正規化し、品番 B_v の誘導における顧客 k の有望度合

$$E(B_v, k) = \frac{E_k}{\max(E_k, k \in \{C(B_v)\})} \quad \dots(5)$$

の値によって、品番 B_v への誘導対象顧客を選定する。

③RFM 分析と吉兆度方式の融合に基づく有望顧客

選定方法については 3.1 節に記載した通りである。なお表 2 中では、秘密保持のため品番を系統で記載している。

4.2 結果

表 2 が実験結果である。ヒット人数が小数になっているのは、有望顧客の同順位に複数人が居る場合の補正をしてい

るためである。吉兆度方式に RFM 分析を融合(③)すること

- 購入確率で見ると、15/16(通り)の 93.75%
 - ヒット者の平均順位で見ると、11/16(通り)の 68.75%
- で、吉兆度方式単独(②)と同等または、それより良い結果が得られた。

5. 結論と今後の展望

本稿では、レコメンデーションに誘導されやすい顧客を抽出する吉兆度方式の効率化として、RFM 分析の融合を提案した。評価実験の結果によれば、提案手法は有用である。

今後の展望として、以下の三点が考えられる:(i)R, F, M の重みを変えた場合の分析、(ii)融合方法を $rE(k)=r(k)+E(k)$ とした場合の分析、(iii)本稿での提案方式のオンラインショッピングでの適用。

参考文献

- [1] 工藤亮祐, 高山毅, 村田嘉利, 佐藤永欣, 高島隼也:レコメンデーションに誘導されやすい顧客を抽出する「吉兆度方式」と「協調フィルタリング方式」の相対評価, 第 72 回情処全大, 3S-1, 2010.
- [2] 裴明花, 谷口伸一, 原隆浩, 西尾章治郎:重要な顧客層および関連ルール発見のための繰り返し購買パターンを考慮した関連ルールマイニング, 情処論誌, Vol.47, No.12, pp.3352-3364, 2006.

表 2 実験を行なったレコメンデーションの 16 パターンと、その結果

圧縮法	パターン番号	参考期側品番	誘導期側品番	①RFM分析単独				②吉兆度方式単独				③吉兆度方式へRFM分析を融合			
				有望顧客(人)	ヒット人数(人)	購入確率(%)	ヒット者の平均順位(位)	有望顧客(人)	ヒット人数(人)	購入確率(%)	ヒット者の平均順位(位)	有望顧客(人)	ヒット人数(人)	購入確率(%)	ヒット者の平均順位(位)
両側圧縮法	(1)	1(婦人服系(a))	2(婦人服系(b))	84	3.00	3.57	26.00	84	2.00	2.38	31.50	84	2.00	2.38	29.50
	(2)	3(婦人服洋品系(a))	4(婦人服系(c))	97	1.00	1.03	27.00	97	1.00	1.03	40.00	97	1.00	1.03	27.00
	(3)	5(紳士服系(a))	6(紳士服系(b))	110	0.00	0.00	0.00	110	0.00	0.00	0.00	110	0.00	0.00	0.00
	(4)	7(特選系(a))	8(紳士雑貨系(a))	111	3.00	2.70	27.67	111	2.06	1.86	32.33	111	2.00	1.80	48.00
	(5)	9(服飾雑貨系(a))	10(家庭用品系(a))	63	1.00	1.59	21.00	63	2.00	3.17	28.00	63	2.00	3.17	26.50
	(6)	2(婦人服系(b))	11(家庭用品系(b))	119	4.03	3.39	52.50	119	4.00	3.36	79.00	119	4.00	3.36	51.25
	(7)	12(家庭用品系(c))	13(家庭用品系(d))	102	4.00	3.92	20.00	102	4.00	3.92	34.75	102	4.00	3.92	33.50
	(8)	14(家庭用品系(e))	15(雑貨系(a))	109	3.00	2.75	39.67	109	3.00	2.75	57.67	109	3.00	2.75	45.00
参考期圧縮法	(9)	16(服飾雑貨系(b))	17(婦人服系(d))	57	0.00	0.00	0.00	57	0.00	0.00	0.00	57	0.00	0.00	0.00
	(10)	18(婦人靴系(a))	2(婦人服系(b))	136	1.00	0.74	54.00	136	1.01	0.74	46.50	136	1.05	0.77	77.00
	(11)	19(服飾雑貨系(c))	20(スポーツ系(a))	134	1.00	0.75	82.00	134	0.00	0.00	0.00	134	0.00	0.00	0.00
	(12)	18(婦人靴系(a))	8(紳士雑貨系(a))	109	1.06	0.97	89.00	109	3.00	2.75	31.00	109	3.00	2.75	72.00
	(13)	21(紳士服系(c))	10(家庭用品系(a))	64	1.00	1.56	27.00	64	2.00	3.13	17.50	64	2.00	3.13	23.00
	(14)	2(婦人服系(b))	11(家庭用品系(b))	114	4.03	3.54	52.50	114	4.00	3.51	74.00	114	4.00	3.51	47.50
	(15)	16(服飾雑貨系(b))	13(家庭用品系(d))	22	0.00	0.00	0.00	22	0.00	0.00	0.00	22	0.00	0.00	0.00
	(16)	14(家庭用品系(e))	15(雑貨系(a))	97	3.00	3.09	39.67	97	2.00	2.06	31.00	97	3.00	3.09	37.33
合計				1528	30.12	1.97	42.92	1528	30.07	1.97	41.94	1528	31.05	2.03	43.13