

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

工藤 亮祐[†] 高山 毅[†] 村田 嘉利[†] 佐藤 永欣[†] 高島 隼也[†]
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][2]。一方、これまでレコメンド手法では協調フィルタリング方式が代表的である。本稿では吉兆度方式と協調フィルタリング方式を定量的に比較し、その優劣を評価する。

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。本研究では、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. 吉兆度方式[1][2]

以下、吉兆度方式を文献[1], [2]に基づき概括する。

2.1 参考期と誘導期

頻出する時系列パターン(以降単に「パターン」と呼ぶ) $A \rightarrow B$ 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 としたときに、過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基にしてレコメンデーションを実施することである。

2.2 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 $\text{count}(A)$ 回、誘導期に B が $\text{count}(B)$ 回登場する場合、パターン $A \rightarrow B$ 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、参考期の品番 A 、誘導期の品番 B の購入の有無のみを考慮し、「 $\text{count}(A) \geq 1$ かつ $\text{count}(B) \geq 1$ 」を満たす顧客人数を、パターン $A \rightarrow B$ の発生回数とする。また参考期圧縮法とは、参考期側の品番 A での購入の有無のみ考慮し、「 $\text{count}(A) \geq 1$ 」である顧客の $\text{count}(B)$ を全顧客について足した値を、パターン $A \rightarrow B$ の発生回数とする。

2.3 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: $Y_num(A \rightarrow B, H_i)$
- 非購入という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 $A \rightarrow B$ 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。

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

ここで $Y_num(A \rightarrow B, H_i)$, $N_num(A \rightarrow B, H_i)$ は、前項で述べた「パターン発生数のカウント法」により変化させている。紙幅の都合により導出過程の詳細は省くが、表1の通りとしている。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。

表1 単一の顧客ごとのパターン $A \rightarrow B$ に対する品番群 H_i の Y_num と N_num

	$Y_num(A \rightarrow B, H_i)$	$N_num(A \rightarrow B, H_i)$
両側圧縮法	1	1
参考期圧縮法	$\text{count}(B)$	1

2.4 顧客抽出とレコメンデーション

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群のことを、「吉兆品番群」と呼んでいる。そして、頻出パターン $A \rightarrow B$ が存在した時に、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 に加え吉兆品番群で購入している顧客へ行なうことを提案している。

3. 協調フィルタリング方式

以下、協調フィルタリング方式を文献[3]に基づき概括する。一般に、本方式では顧客の購入履歴から類似嗜好の顧客をグルーピングする。その上で、対象とする顧客 X の未購入商品のうち、同一グループの顧客が既購入の商品をレコメンドする。

図1の購入履歴行列は、行を顧客、列を商品とする 2 値行列 N であり、 i 行の顧客が j 列の商品を購入した場合、要素 n_{ij} を 1 とする。値を 0/1 ではなく評価値とす

Comparison Evaluation of Auspicious Level Method from Collaborative Filtering one Which Picking up Expectable Customers in a Recommendation
†R. Kudo, T. Takayama, Y. Murata, N. Sato, and J. Takashima
(Faculty of Software and Information Science, Iwate Prefectural University)

る方法もあるが、ここではデータ取得がより容易な 0/1 とする。図 1 では、顧客 X の類似顧客は顧客 B、D と見なせる。顧客 X の未購入商品中でのレコメンドすべき順位は、0.0~1.0 のいずれかの値を取る購入の期待度の高さから、商品 a, b, e の順と考える。

	商品				
	a	b	c	d	e
顧客 A	0	0	0	0	1
B	1	0	1	1	0
C	0	1	1	0	0
D	1	1	1	1	0
X	0	0	1	1	0

図 1 購入履歴行列.

4. 吉兆度方式と協調フィルタリング方式の比較

4.1 実験方法

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の拡張を目指すタスクで実験する。なお一般には、このタスクの成功は容易とは言えない。吉兆度方式と協調フィルタリング方式の各方式で、購入が有望と推定されるそれぞれ k 個の品番とその各有望顧客 l 名にレコメンドを実施する。そして実際に購入、すなわちヒットに結びつけられるか否かを分析する。

表 2 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 40,650名
抽出する対象顧客数	1600名 (2×k×l)
対象品番の母集団	298品番
参考期	06/01~08/31
誘導期	11/14~11/30
レコメンド前の分析	2007, 8年の各参考期, 誘導期, および 2009年の参考期から

表 3 吉兆度方式での分析結果で有望と推定して利用する頻出 5 パターンと吉兆品番群

カウント法	参考期側品番	誘導期側品番	吉兆品番群
(6) 両側圧縮法	6 (書籍系)	7 (音楽映像ソフト系)	13(輸入雑貨系) 14(飲食系) 15(化粧品系)
	8(婦人肌着系)	9 (文具系)	16(家庭用品-3 群系) 17(玩具系) 18(家庭用品-2 群系) 19(家庭用品-4 群系)
(8) 参考期圧縮法	8(婦人肌着系)	10(家庭用品系)	20(L サイズ婦人服系) 21(靴バッグ系)
(9) 参考期圧縮法	6 (書籍系)	11(化粧品系)	22(ヤング婦人服系) 21(靴バッグ系)
(10) 参考期圧縮法	8(婦人肌着系)	12(家庭用品系)	23(家庭用品-3 群系) 24(家具インテリア系) 19(家庭用品-4 群系)

表 4 協調フィルタリング方式での分析結果で有望と推定してレコメンドする 5 品番

(1)	(2)	(3)	(4)	(5)
1(化粧品系)	2(特選婦人服系)	3(婦人靴系)	4(陶磁器系)	5(雑貨系)

レコメンドは、2k 個の品番を記載した広告を 2k \times l 名の顧客へ y 年の誘導期直前に郵送し実現する。

今回は、k=5, l=160, y=2009 とする。また、その他の実験条件は表 2~4 の通りである。なお表 3, 4 中では、秘密保持のため品番をコード化し、系統のみ記載する。

4.2 実験結果 (表 5)

表 5 有望品番(表中上段)ごとの各有望顧客 160 名中でのヒット人数(表中下段, 単位=人)

協調フィルタリング方式	品番	1	2	3	4	5	5品番計
	ヒット人数	0	0	0	2	0	2
吉兆度方式	品番	7	9	10	11	12	5品番計
	ヒット人数	4	6	2	1	2	15

まず、最右欄の 5 品番計でのヒット人数で比較すると、協調フィルタリング方式でのレコメンド成功が 2 名に対して、吉兆度方式側が 15 名と、優位な結果が得られている。前者の 2 名について更に分析したところ、うち 1 名は吉兆度方式でも有望顧客と推定可能なことがわかった。一方、後者の 15 名については、協調フィルタリング方式でも有望と推定可能なのはわずか 1 名とわかった。すなわちヒット者 17 名中で、吉兆度方式で推定可能なのが 16 名に対し、協調フィルタリング方式では 3 名に留まっている。以上より、吉兆度方式が相対的に優位と評価できる。

吉兆度方式の中では品番 7, 9 の、両側圧縮法での分析で有望と推定されたパターンの方が、参考期圧縮法のそれより良い結果を得ている。

5. まとめと今後の展望

本稿では、レコメンドに誘導されやすい顧客を抽出する二つの方式を相対評価した。そして、吉兆度方式の方が協調フィルタリング方式より優れているという結果を得た。今後の展望として、以下のことが考えられる: 吉兆度方式で、(i)分析を年代別に行なうことによるレコメンド精度の向上、(ii)レコメンドする品番の先決めによる、ビジネスとしての有用性の向上。

参考文献

- [1] 太田光雄, 高山毅, 村田嘉利, 佐藤永欣ほか: レコメンドに誘導されやすい顧客の抽出方式の拡張と評価, 第 71 回情処全大 6N-4, 2009.
- [2] 恵津森真仁, 高山毅, 村田嘉利, 佐藤永欣: レコメンドに誘導されやすい顧客の抽出方式と評価, 第 70 回情処全大 2T-7, 2008.
- [3] 神島敏弘: 推薦システムのアルゴリズム(2), 人工知能学会誌, Vol.23, No.1, pp.89-95, 2007.