

購買推奨確率モデルによるビッグデータの解析

石田 実, 西尾 チヅル, 佐藤 忠彦

1. はじめに

オンラインショップでは、「類似した消費者が購入した製品を推奨する」あるいは「併買されやすい製品を推奨する」という手法をシステム化した推奨（レコメンド）システムが、重要な販促ツールとなっている。分類や判別の統計モデルを事例データに当てはめて推奨の予測精度を比較するさまざまな実証研究が行われているが [1], 消費者の購買推奨に適した統計モデルはまだ明らかになっていない。その理由として、推奨システムで表現すべき消費者の購買行動の特徴が示されておらず、データの特徴をモデル化するという観点で議論できていないという問題がある。購買履歴は実務マーケティングを高度化するために、その高度活用が期待されている。ビッグデータ解析は、大規模で鮮度の高い多様なデータを扱う点で従来のデータ分析とは異なる [2]。従来のデータ分析では、分析仮説に沿ったデータ構造を検討し、計画的にデータを収集し、そのデータに基づき解析を進める。一方で、購買履歴のようにログとして得たデータを分析する場合、データの特徴を捉えて適した統計モデルを利用することが特に重要である。そこで本研究では、推奨システムで表現すべき購買行動の特徴を議論するため、個人ごとの製品別購買予測確率を陽に数式として導出する新たなアプローチを提案する。購買予測確率の導出に用いる購買行動の仮定を可能な限り厳選し、推奨システムと購買行動との関係を示す。数式を導く手順は単純で、消費者の購買経験が次の購買に与える影響を条件付き確率で表し、複数の購買履歴を条件とする購買予測確率を導く。購買履歴の製品を 1 個から複数製品に拡張す

る際に、弱学習を適用する方法と条件付き独立を仮定する 2 通りの数式を提案して比較する。実証実験として、音楽 CD の販売 POS データを用いて購買予測力を比較したところ、予測対象を購買数の多い消費者と製品に限定した場合を除いて、提案する 2 つのモデルとも比較モデルとしたカーネル法サポートベクターマシンより高い予測力を示した。本研究で提案する数式は購買確率を表すので、順序統計量としての推奨指標に用いるだけでなく、確率の合計を売上への予測に利用できる利点もある。売上を予測するマーケティング活動への応用事例として、音楽業界の商習慣であるレコードレーベルを提案モデルに基づいて考察し、企業の立場から製品の類似をマネジメントする効果について議論する。

2. 先行研究

製品と消費者の類似関係と購買との関連の研究には、認知論あるいは購買行動の立場の研究と、計算機科学における研究がある。

認知心理学の研究として、Tversky [3] は類似認識を用いたカテゴリーベース処理のモデル（対比モデル：Contrast Model）を提案した。カテゴリーベース処理は、人がさまざまな情報や製品をカテゴリー化することでその情報や製品を理解し、評価する情報処理手法である。消費者のライフスタイルが多様化し、また膨大な情報に接する社会では、消費者がさまざまな情報を判断することが多くなるため、認知努力を節約する仕組みとして類似認識に基づくカテゴリーベース処理に頼る状況が多くなる [4]。澁谷 [5] は、消費者が自身に類似する他人の体験を参考に行動し、類似認識が製品の評価や購買に影響することを実験で検証した。製品を陳列する物理的制約がないオンラインショップでは、消費者はさまざまな製品を選択できる利便がある一方で、膨大な情報に接して情報探索の負荷を感じたり、購入製品を合理的に判断できない不安（知覚リスク）が大きくなる。Häubl & Trifts [6] は推奨システ

いしだ みのる
株式会社アークエンジン
〒798-0013 愛媛県宇和島市御幸町 2-1-6
にしお ちづる, さとう ただひこ
筑波大学
〒112-0012 東京都文京区大塚 3-29-1
受付 13.12.11 採択 14.4.4

表 1 推奨システムの手法の分類

	手法
コンテンツベースフィルタリング	頻度に基づく評価, クラスタリング, 潜在クラスモデル, 決定木, ニューラルネットワーク
協調フィルタリング	近傍法 (類似した利用者・アイテムの評価を利用), グラフ理論, ペイジアンネットワーク, クラスタリング, ニューラルネットワーク, 線形回帰, カーネル法 SVM, 確率モデル
ハイブリッドフィルタリング	上記手法を併用あるいは結合

Adomavicius and Tuzhilin [9] の表 2 を参考に作成.

ムの機能として, 消費者ごとに購入候補の考慮集合を提案することと, 購入候補の比較検討を支援することの 2 点を挙げて, 消費者の満足度を高める効果があると指摘している.

計算機科学の分野の研究として, 米国のゼロックス研究所の Tapestry [7] と GroupLens [8] を端緒とする, 情報の選択や評価を推測する情報フィルタリングに関する研究がある. レコメンドシステムのサーベイ研究である Adomavicius & Tuzhilin [9], 神島 [10, 11] は, 利用するデータの違いにより推奨システムをコンテンツベースフィルタリング, 協調フィルタリング, ハイブリッドフィルタリングに分類した. 表 1 に, それらを総括的に整理して示した. コンテンツベースフィルタリングは, 利用者 (顧客) とアイテム (評価対象や商品) の属性を対応づける手法である. 協調フィルタリングは, 利用者のアイテムに対する評価履歴や購買履歴の類似関係を手掛かりとする手法である. ハイブリッドフィルタリングは両者を併用する手法である. 表 1 のほかの分類として, 評価値が 5 得点評価のような順序データか, あるいは購買の有無のような 2 値データかという分類も重要になる. 順序データの場合は, 正規分布とみなして相関係数で類似評価できる [12]. 2 値データの場合は, 相関係数に代わる多数の類似係数が提案されているが [13, 14, 15], どの類似係数が優れているか定説はない. 石田・西尾・椿 [16] は, 購買履歴を用いて消費者と製品を教師なし分類する際に, 2 項分布の相関係数にあたる交互作用統計量がほかの類似係数に対して優位となることを実証し, その理由を考察している. 推奨システムにおける類似係数の利用について, Iacobucci et al. [17] は消費者と製品の分類をそれぞれ作成して推奨システムに応用する概念を示し

た. 本多 [18] は, 消費者と製品の分類を同時に行う共クラスタリングを協調フィルタリングに適用する手法を示している. Breese et al. [1] は, 利用者 i のアイテム a_j への評価得点 $v_{i,a_j} (\in 0, \dots, m)$ の期待値を,

$$E(v_{i,a_j}) = \sum_{l=0}^m Pr(v_{i,a_j} = l | v_{i,a_k}, a_k \in I_i) \cdot l$$

ただし I_i は利用者 i の評価済みアイテムの集合として記述した. Breese et al. [1] は, $Pr(v_{i,a_j} | v_{i,a_k})$ の計算を陽に評価しているわけではなく, 潜在クラスモデルを用いて推定する手法と, ペイジアンネットワークを用いて推定する手法を示して比較検証している. 本研究では, 陽に条件付き確率の算式を示す.

推奨システムの評価指標は多数あり, 広く優位と認められる指標はない [12]. 消費者が推奨後に購入した製品を正解とする一致率の概念には, 推奨リストにおける正解の比率である精度 (または適合率: precision) と, 正解リストにおける推奨した比率である再現率 (recall) がある. 推奨リストを短くすると精度が高くなる一方で, 再現率が低くなるトレードオフの関係がある. Breese et al. [1] は, 評価指標に関する議論を踏まえて推奨リストの長さ指数関数でペナルティを与える Rank Scoring 指標を提案した.

本研究の実証対象とする音楽市場について, Kubacki [19] は聞き取り調査を行い, 音楽家には芸術家タイプと, 音楽を商品として扱うプロモータータイプの 2 通りあることを示した. 音楽 CD をプロモーションするレコード会社 (レーベル) では, A&R (Artist & Repertoire) 部門において音楽家を発掘し, 育成してレコードを制作し, 楽曲の宣伝を行うのが主要な業務となる [20]. 勝又・阿部 [21] は音楽 CD の ID 付き POS データを用いた実証研究において, 音楽 CD は嗜好の異質性が大きく, 顧客セグメントや個人単位の嗜好の推定が重要であると指摘している. なお, オペレーションズ・リサーチ, 第 52 巻 (2007) は CD 販売データの分析の特集号を組んでいる.

3. 提案モデル

3.1 モデルの意図

協調フィルタリングのアプローチを踏まえて, 購買行動のみを観測して評価するモデルを考える. 消費者はある製品を買うという経験を通して異質性を獲得し, 同時に消費者間の類似が生まれ, 次に買う製品の購買確率が购买前に比べて変化するさまを条件付き確率を用いてモデル化する. モデルの設定として, 消費者がど

の製品を購入したかが次の購買確率の条件となり、次の購買に影響を与えると考え、このため、購買履歴のない初期状態ですべての消費者が完全に同質であると設定しても、確率的に起きる購買経験を通して消費者に異質性が生まれる。モデルの構築では、次の3つの仮定を設ける。3番目の仮定は2通り設けるため、提案モデルも2つある。(1) 購買確率に2項分布を仮定、(2) 購買確率の相関係数が消費者に共通と仮定、(3) 購買確率の条件としての購買履歴を1製品から複数製品に増やす際に、(3-A) 弱学習の手法が有効と仮定する、あるいは、(3-B) 条件付き独立を仮定する。それぞれの仮定の根拠は、モデル構築の際に記述する。モデル化する変数として、個々の消費者と製品は区別するが、購買数量については考慮せず、一定期間の過去の購買の有無のみを0か1とする。時間については、過去と将来の2時点のみ考える。このためモデルを適用する製品として、耐久財や本研究で扱う音楽CDのように、通常1回に1個ずつ購入され、繰り返し購買が起きない製品を想定する。

3.2 理論モデル

モデルの記述のために消費者と製品の集合、および所与の期間の購買履歴と購買確率を次のとおり記す。

$I = \{1, \dots, i, \dots, n\}$ n 人の消費者の集合

$A = \{a_1, \dots, a_j, \dots, a_m\}$ m 個の製品の集合

$U = \{u_{i,j} \mid \text{消費者 } i(i \in I) \text{ が製品 } a_j(a_j \in A) \text{ を過去に1個以上購入した場合に限り } u_{i,j} = 1, \text{ さもなくば } u_{i,j} = 0\}$

消費者 $i(i \in I)$ の製品 $a, b(a \in A)$ の購買確率について、

$P_i(a)$ 消費者 i が製品 a を購買する確率

$P_i(b|a)$ 製品 a の購買の有無の条件下で、消費者 i が製品 b を購買する条件付き確率

$P_i(a \cap b)$ 消費者 i が製品 a と b を共に購買する同時確率

$P_i(a_j|A)$ 過去の購買経験を与えた場合の、消費者 i が製品 a_j を購買する条件付き確率

本研究では、 $P_i(a_j|A)$ をモデル化する。表2に、消費者 i の製品 a と b の購買に関する4通りの事象の発生確率をまとめる。

仮定1：購買発生の有無 $P_i(a)$ は2項分布に従う。

以下の数式の展開において、購買発生の有無に2項分布を仮定する。2項分布を仮定するのは、第1に2値データの分布として最も単純な分布であり、第2に先行研究 [16] で購買発生の有無に2項分布を仮定すると妥当な分類が得られると示されているからである。

表2 製品 a と b の購買に関する発生確率

		製品 b		
		購買する	購買しない	計
製品 a	購買する	$P_i(a \cap b)$	$\begin{matrix} P_i(a) \\ -P_i(a \cap b) \end{matrix}$	$P_i(a)$
	購買しない	$\begin{matrix} P_i(b) \\ -P_i(a \cap b) \end{matrix}$	$\begin{matrix} 1 + P_i(a \cap b) \\ -P_i(a) \\ -P_i(b) \end{matrix}$	$1 - P_i(a)$
	計	$P_i(b)$	$1 - P_i(b)$	1

この仮定により、消費者 i の2製品 a, b の購買確率のピアソンの相関係数は、

$$s_i(a, b) = \frac{P_i(a \cap b) - P_i(a)P_i(b)}{\sqrt{P_i(a)(1 - P_i(a))P_i(b)(1 - P_i(b))}} \quad (1)$$

(1) 式の $s_i(a, b)$ は、独立性の検定指標であるカイ二乗値を試行回数で基準化して平方根をとった値に一致し、交互作用統計量とも呼ばれる。 $s_i(a, b)$ を用いて、消費者 i が製品 a を購買した場合に製品 b を購買する条件付き確率 $P_i(b|a = 1)$ は、(1) 式を変形して $s_i(a, b)$ を用いて表すことができる。

$$\begin{aligned} P_i(b|a = 1) &= \frac{P_i(a \cap b)}{P_i(a)} \\ &= P_i(b) \\ &\quad + s_i(a, b) \sqrt{\frac{1 - P_i(a)}{P_i(a)} P_i(b)(1 - P_i(b))} \end{aligned} \quad (2)$$

同様に製品 a を購買しなかった場合に製品 b を購買する条件付き確率は

$$\begin{aligned} P_i(b|a = 0) &= \frac{P_i(b) - P_i(a \cap b)}{1 - P_i(a)} \\ &= P_i(b) - s_i(a, b) \sqrt{\frac{P_i(a)}{1 - P_i(a)} P_i(b)(1 - P_i(b))} \end{aligned} \quad (3)$$

(2) 式と (3) 式をまとめて、

$$\begin{aligned} P_i(b|a) &= P_i(b) \left\{ 1 + t_{i,a} s_i(a, b) \right. \\ &\quad \left. \times \sqrt{\left(\frac{1 - P_i(a)}{P_i(a)} \right)^{t_{i,a}} \frac{1 - P_i(b)}{P_i(b)}} \right\} \end{aligned} \quad (4)$$

ただし、

$$t_{i,a} = \begin{cases} 1 & (a \text{ を購買した場合}) \\ -1 & (a \text{ を購買しなかった場合}) \end{cases}$$

(4) 式を変形すると

$$\frac{P_i(b|a)}{P_i(b)} - 1 = t_{i,a} s_i(a, b) \sqrt{\left(\frac{1 - P_i(a)}{P_i(a)}\right)^{t_{i,a}} \frac{1 - P_i(b)}{P_i(b)}} \quad (5)$$

となるので、製品 a の購買の有無により製品 b を購買する条件付き確率が増減する比率は、 $\sqrt{(1 - P_i(a))/P_i(a)}^{t_{i,a}}$ に比例することがわかる。製品 a を購買した場合は、購買確率 $P_i(a)$ が小さく買う人の少ない製品 a を購入した場合に影響が大きく、逆に、多くの人が買う製品を購入した場合の影響は小さくなる。また、製品 a を購買しなかった場合は逆数の $\sqrt{P_i(a)/(1 - P_i(a))}$ に比例し、多くの人が買う製品を買っていないという情報の影響が大きくなるので、稀な事象ほど情報量が大きいと解釈できる。次に、プロダクトマップとして市場の有り様が消費者に同じに見えることを表す仮定を置く。

仮定 2：2 製品 a, b の購買確率の相関係数 $s_i(a, b)$ は消費者 i に依存しないと仮定。

これは非常に強い仮定であるが、マーケティングの分析で暗黙に用いる自然な仮定である。すなわち、製品の類似評価から構成するプロダクトマップを消費者ごとに複数作成しないのと同じで、消費者の認知する製品間の関係に異質性がないと仮定する。

最後の仮定として、1 個の製品の購買の有無を条件とする確率から、複数製品の購買履歴を条件とする確率を計算するための仮定を 2 通り設ける。

仮定 3-A (弱学習器)：消費者 i が製品 a_j の過去の購買の有無のみ知り得た場合に製品 b の将来の購買確率を推定するモデルを $M_{i,b}(a_j)$ として、 m 個すべての製品の購買履歴を知り得た場合に同様に製品 b の購買確率を推定するモデルを $M_{i,b}(\mathbf{A})$ とする。このとき、 $M_{i,b}(\mathbf{A})$ は m 個のモデル $M_{i,b}(a_j), j = 1, \dots, m$ の弱学習により、 $M_{i,b}(a_j)$ の単純平均で推計できる。

仮定 3-B (条件付き独立)：任意の 3 個の製品 a, b, c の購買について条件付き独立が成り立つ。すなわち、

$$P_i(a, b|c) = P_i(a|c)P_i(b|c).$$

まず仮定 3-A について、モデル $M_{i,b}(\mathbf{A})$ を全サンプル \mathbf{A} が与えられるモデルと考え、モデル $M_{i,b}(a_j)$ をサンプルが 1 個しか与えられない精度の低い弱い学習器と考える。弱い学習器を多数結合する手法はコミッティ(committee)と呼ばれ [22]、結合手法にはブースティングなどのよく知られた重み付けの最適化手法もあるが、提案モデルでは確率が重みの役割を果たす

ため、重み付けの重複を避けて簡易に単純平均を用いる。モデル $M_{i,b}(a_j)$ には (4) 式の $P_i(b|a_j)$ の算式を用い、モデル $M_{i,b}(\mathbf{A})$ は $P_i(b|\mathbf{A})$ を推定すると考える。

また、仮定 3-B の条件付き独立は、条件が同じ 2 つの事象の発生確率が独立であり、それぞれの発生確率の積が同時確率と等しくなることを表している。また、分類モデルの単純ベイズ分類器(Naive Bayes classifier)で用いる仮定でもある。本研究では、この仮定により導かれる算式の妥当性を実証検証し、間接的に提示した仮定の妥当性を評価する。

3.3 仮定 3-A：弱学習器モデル

仮定 2 と 3-A の弱学習器に関する仮定を用いて、数式の展開を進める。仮定 2 により、以降、消費者 i に依存しない相関係数 $s_i(a, b)$ に代えて $s(a, b)$ と表記する。仮定 3-A により、モデル $M_{i,b}(a_j)$ を $a_j \in \mathbf{A}$ について単純平均で結合してモデル $M_{i,b}(\mathbf{A})$ を求める。モデル $M_{i,b}(a_j)$ に (4) 式を当てはめ、モデル $M_{i,b}(\mathbf{A})$ の推定値となる次式 (6) を得る。

$$\begin{aligned} P_i(b|\mathbf{A}) &= \frac{1}{m} \sum_{j=1}^m \left[P_i(b) \left\{ 1 + t_{i,a_j} s(a_j, b) \sqrt{\left(\frac{1 - P_i(a_j)}{P_i(a_j)}\right)^{t_{i,a_j}} \frac{1 - P_i(b)}{P_i(b)}} \right\} \right] \\ &= P_i(b) + \frac{\sqrt{P_i(b)(1 - P_i(b))}}{m} \\ &\quad \times \sum_{j=1}^m \left\{ t_{i,a_j} s(a_j, b) \sqrt{\left(\frac{1 - P_i(a_j)}{P_i(a_j)}\right)^{t_{i,a_j}}} \right\} \quad (6) \end{aligned}$$

(6) 式の右辺を計算するには、相関係数 $s(a, b)$ と確率 $P_i(a)$ の値を知る必要がある。相関係数 $s(a, b)$ については、仮定 2 により消費者を区別しないで表 2 の各事象の度数をクロス集計して推計値とする。実データが条件付き確率に従う点については、集計により周辺確率の相関係数を推計できると考える。一方、確率 $P_i(a)$ の推計は困難である。消費者の購買は、購買履歴を条件とする条件付き確率に従うと考えるため、購買履歴を観測して直接的に $P_i(a)$ を推計することができない。そこで筆者らは、 $P_i(a)$ の簡易な推定法を模索したが、その効果は $P_i(a)$ を定数を置いた場合に比べて限定的であり、効果的な $P_i(a)$ の推定法を議論して本研究を煩雑にする利点は小さいと考えるに至った。このため、消費者 i と製品 a に依らず、購買確率 $P_i(a) = \text{定数}$ とする。 $P_i(a)$ を個別に推定するか否かはモデルの議論の本質ではないため、この設定は仮定として扱わない。

$$P_i(a) = c, \forall a \in \mathbf{A}, c \text{ は定数} \quad (7)$$

これを (6) 式に代入すると、

$$\begin{aligned} & P_i(b|\mathbf{A}) \\ &= c + \frac{\sqrt{c(1-c)}}{m} \left\{ \sqrt{\frac{1-c}{c}} \sum_{a_j \in A_i} s_i(a, b) \right. \\ & \quad \left. - \sqrt{\frac{c}{1-c}} \sum_{a_j \in \bar{A}_i} s_i(a, b) \right\} \end{aligned}$$

ただし、 A_i は消費者 i が購買した製品の添え字で $A_i = \{k | u_{i,k} = 1\}$ 、購買しなかった製品の添え字を $\bar{A}_i = A - A_i$ とし、

$$\begin{aligned} &= c + \frac{1-c}{m} \sum_{a_j \in A_i} s(a_j, b) - \frac{c}{m} \sum_{a_j \in \bar{A}_i} s(a_j, b) \\ & c \text{ が } 1 \text{ に対して十分に小さいとして} \\ & \approx \frac{1}{m} \sum_{a_j \in A_i} s(a_j, b) \quad (8) \end{aligned}$$

(8) 式の $s(a_j, b)$ を相関係数行列 \mathbf{S} で置き換えて行列表記すると、提案する弱学習器モデルは以下の行列計算式となる。

$$\mathbf{P} = \frac{1}{m} \mathbf{U} \mathbf{S} \quad (9)$$

購買履歴の行列 \mathbf{U} に相関係数行列 \mathbf{S} を掛けて購買指標とする (9) 式は、筆者が経験的に知る最も簡易な協調フィルタリングの算式に一致する。

3.4 仮定 3-B : 条件付き独立モデル

今度は、仮定 2 と 3-B の条件付き独立の仮定を用いて、数式の展開を進める。仮定 3-B の条件付き独立を繰り返し適用して次式を得る。

$$P_i(\mathbf{A}|b) = \prod_k P_i(a_k|b) \quad (10)$$

ベイズの定理を用いて、複数製品の購買履歴条件付き確率 $P_i(b|\mathbf{A})$ を次のとおり計算する。次式に限り、計算の意図を明確にするため $P_i(b)$ を $P_i(b=1)$ 、 $1-P_i(b)$ を $P_i(b=0)$ と記述する。

$$\begin{aligned} & P_i(b=1|\mathbf{A}) \\ &= \frac{P_i(\mathbf{A}|b=1)P_i(b=1)}{P_i(\mathbf{A})} \\ &= \frac{P_i(\mathbf{A}|b=1)P_i(b=1)}{P_i(\mathbf{A}|b=0)P_i(b=0) + P_i(\mathbf{A}|b=1)P_i(b=1)} \end{aligned}$$

(10) 式を用いて、

$$= \frac{P_i(b=1) \prod_k P_i(a_k|b=1)}{P_i(b=0) \prod_k P_i(a_k|b=0) + P_i(b=1) \prod_k P_i(a_k|b=1)} \quad (11)$$

また、仮定 2 により、消費者 i に依存しない相関係数 $s_i(a, b)$ に代えて $s(a, b)$ と表記し、(4) 式の a と b を逆転して次式を得る。

$$\begin{aligned} P_i(a|b) &= P_i(a) \left\{ 1 + t_{i,b} s(a, b) \right. \\ & \quad \left. \times \sqrt{\left(\frac{1-P_i(b)}{P_i(b)} \right)^{t_{i,b}} \frac{1-P_i(a)}{P_i(a)}} \right\} \quad (12) \end{aligned}$$

数値計算において、弱学習器モデルの (7) 式と同様に $P_i(a) = c$ を (12) 式に代入すると、

$$P_i(a|b=1) = c + (1-c)s(a, b) \quad (13)$$

$$P_i(a|b=0) = c(1-s(a, b)) \quad (14)$$

(13) 式と (14) 式を用いて (11) 式の各項を表すと、

$$\begin{aligned} & P_i(b) \prod_k P_i(a_k|b=1) \\ &= c \prod_{k \in A_i} \{c + (1-c)s(a_k, b)\} \\ & \quad \times \prod_{k \in \bar{A}_i} \{(1-c)(1-s(a_k, b))\} \quad (15) \\ & (1-P_i(b)) \prod_k P_i(a_k|b=0) \\ &= (1-c) \prod_{k \in A_i} \{c(1-s(a_k, b))\} \\ & \quad \prod_{k \in \bar{A}_i} \{1-c(1-s(a_k, b))\} \quad (16) \end{aligned}$$

(15) 式と (16) 式を (11) 式に代入して、条件付き独立モデルとなる $P_i(b=1|\mathbf{A})$ の数式を得る。この数式を書き下すのは冗長なので割愛する。

4. 実証分析

4.1 目的と手順

4.1.1 購買予測力の検証

第 1 の実証分析として、前節で提案したモデルの購買予測力を 2 つの指標で評価して比較する。検証データには音楽 CD 販売の ID 付き POS データを用い、結果の再現性を確認するため、データセットの作成期間を 8 通りと抽出条件を 5 通りの 8×5 パターンの組合せを設ける。学習データの期間は 3 カ月間、続く 3 カ月間を検証期間とする。第 1 期の学習データを 2002 年 11 月から 2003 年 1 月までとし、検証データを 2003 年 2 月から 4 月までとして購買を予測する。第 2 期以降は期間を 1 カ月ずらして繰り返し計 8 期の検証を行う。抽出条件は、相関係数や確率の推計が満足にできない評価対象を除く目的で設け、購入枚数が $K (= 3, \dots, 7)$

枚以上の消費者と、販売枚数が $2K$ 枚以上の製品の組合せの購買履歴に限定する。 K の値に幅を設けたのは、消費者 \times 製品のデータ行列でゼロ以外の要素の多少のスパース性がモデルの説明力に与える影響を評価するためである。 検証の手順は次のとおりである。 各データセットについて、学習用データを用いて 2 製品 a, b の購買確率の相関係数 $s(a, b)$ を計算する。 また、 $P_i(a) = c$ (一定値) となる c を、 $c = \sum_{i,j} u_{i,j} / mn \times$ 検証期間月数/学習期間月数と計算する。 次に、これらの値を提案モデルに代入する。 弱学習器モデルの場合は、(9) 式を用いて購買履歴の行列 \mathbf{U} に相関係数の行列 $\mathbf{S} = \{s(a, b)\}$ を掛け、製品数 m で割って購買予測確率 $P_i(b|\mathbf{A})$ を求める。 条件付き独立モデルの場合は、 c と $s(a, b)$ の値を用いて (15) 式と (16) 式の値を求め、これを (11) 式に代入して目的の購買確率の推定値を求める。 最後に予測確率 $P_i(b|\mathbf{A})$ を、学習期間で購買のあった消費者と製品の組合せを除いて、検証期間の購買結果に照らして評価する。

提案する 2 モデルとの比較のため、カーネル法 SVM (サポートベクターマシン) を用いた購買確率の推定も行う。 カーネル法 SVM では、共分散行列と同じく半正定値行列の性質を持つカーネル \mathbf{S}' を使い、消費者 i ごとに次式で定める座標の判別の境界からのマージンと呼ばれる距離を最大化する。

$$\mathbf{w}_i \cdot \mathbf{S}' + b_i \cdot e \quad (17)$$

ただし、 \mathbf{w}_i はサポートベクターマシンで推定する重み (方向) を表す行ベクトル、 b_i を切片のスカラー値、 e をすべての要素が 1 の行ベクトルとする。 消費者 i ごとに数件の購買履歴から \mathbf{w}_i を推定すると、データの情報量の観点からオーバーフィットのリスクがある。 このとき、重み \mathbf{w}_i を推定する代わりに、簡易に消費者 i の製品 a_j の購買の有無 $u_{i,j}/m$ を用いると、切片を除いて (17) 式は (9) 式 $\mathbf{P} = 1/m \cdot \mathbf{U}\mathbf{S}$ の第 i 行に一致する。 また、(9) 式の相関係数行列 \mathbf{S} はカーネルの性質を備える。 本研究の議論が妥当であれば、パラメータ推定の必要はなく、提案モデルは良い結果を示すであろう。 カーネル法 SVM の実行には、統計ソフト R の kernlab パッケージ [23]、Chang & Lin [24] を使い、モデルの適用は各消費者ごとに行う。 カーネルには半正定値となる代表的な類似係数である Jaccard 係数 [15] を用いる。

モデルの評価指標として、Breese et al. [1] の Rank Scoring 指標とコルモゴロフスミルノフ検定の 2 つを併用する。 Rank Scoring 指標は、推奨リストの長さ

に指数関数でペナルティを与えて精度-再現率のトレードオフを評価する指標である。 まず、消費者 i ごとに全製品の購買予測確率を降順に並べ、検証期間で購買された製品が出現した順番の配列 \mathbf{R}_i を作成する。 仮に 3 番目と 5 番目、... に購買予測確率が高いと評価した製品が購買された場合は $\mathbf{R}_i = \{3, 5, \dots\}$ となり、配列 \mathbf{R}_i の要素の個数は消費者 i が検証期間に購買した製品の個数 v_i に一致する。 最も良い予測ができた場合は $\mathbf{R}_i^{max} = \{1, 2, \dots, v_i\}$ となる。 まず、次式 (18) で消費者 i の予測の一致率を評価する。

$$\text{Rank}(\mathbf{R}_i) = \sum_{k \in \mathbf{R}_i} \frac{1}{2^{\frac{k-1}{T-1}}} \quad (18)$$

半減期 T の値は評価者が設定する指数関数を定めるパラメータである。 Rank Scoring 指標は、(19) 式のとおり $\text{Rank}(\mathbf{R}_i)$ を消費者について合算して最大 100 に基準化した値とする。

$$\text{Rank Scoring} = 100 \frac{\sum_{i \in \mathbf{I}} \text{Rank}(\mathbf{R}_i)}{\sum_{i \in \mathbf{I}} \text{Rank}(\mathbf{R}_i^{max})} \quad (19)$$

Breese et al. [1] は半減期 T の値は比較結果にあまり影響しないと指摘している。 本実験でも、半減期 T を 2 倍にすると評価値が約 2 倍になり比較順序には影響がないことを確認し、半減期 $T = 10$ とした。 Rank Scoring 指標は推奨リストの長さにはペナルティを与える利点があるが、購買予測確率を消費者ごとの順序尺度としてのみ評価しているという課題がある。 特定の消費者に推奨する製品の順序付けは評価しているが、特定の製品を推奨する消費者の順序付けは評価していない。 そこで、消費者 \times 製品の行列の形で得られる購買予測確率を消費者と製品の区別なく 1 列に並べた後、検証データにおける購買の有無で 2 グループに分けて、2 グループの確率の分布の水準の差異をコルモゴロフスミルノフ検定で片側検定することも併せて行う。

4.1.2 相関係数の同質性の検証

本項では、「2 製品の購買確率の相関係数は消費者に依存しない」とする仮定 2 の妥当性を検証する。 購買が類似する消費者の情報を用いて製品を推奨する最近傍法 (あるいは近傍選択 [12]) は、購買行動に基づく消費者セグメンテーションの有用性を示唆している。 本項の検証では、購買の類似性により 2 つの消費者セグメントに分類し、それぞれの購買履歴を用いて 2 通りの製品間の相関係数行列を計算し、先の購買予測力の検証と同様の手順で各消費者セグメントごとに弱学習器モデルの購買予測力を評価する。 次に、異なる消

費者セグメントの購買履歴から計算した相関係数行列で差し替えて、同様にモデルの評価を行う。自身の購買履歴から計算した相関係数行列を用いた場合と、他セグメントの購買履歴から計算した相関係数行列を用いた場合でモデルの説明力に有意な差がなければ、消費者セグメントを設ける有用性がなく、仮定2が妥当な仮定だと判断できることになる。データとして、購買予測力の検証の $K = 3$ の条件で抽出した8期間のデータセットを用いる。まず、消費者間の相関係数行列を計算し、これを類似指標としてPAM (Partitioning Around Medoids) 法により消費者を2つのセグメントに分類する。PAM法はKaufman & Rousseeuw [25] によって提案された非階層的クラスタリング手法で、非階層的であるため分類数があらかじめ決まった大規模要素の分類に適している。評価対象製品は、相関係数行列を差し替えるために $K = 3$ の条件で2つの消費者セグメントに共通して購買のある製品とする。

表3 購買予測力の検証のデータセットの特徴
第1期が抽出条件 $K = 3$ のサンプルのみ掲載

期間	K	消費者数	製品数	学習期間 購買数/人	検証期間 購買数/人
1	3	8,610	899	4.60	1.43
1	4	4,333	560	5.78	1.64
1	5	2,025	317	6.97	1.63
1	6	843	156	8.10	1.49
1	7	165	46	9.47	1.04
2	3	8,281	849	5.78	1.64
3	3	8,106	856	6.97	1.63
4	3	7,173	803	8.10	1.49
5	3	6,844	831	9.47	1.04
6	3	5,866	800	4.48	1.41
7	3	6,331	841	5.63	1.66
8	3	6,216	824	6.76	1.64

表4 相関係数の同質性の検証のデータセットの特徴

期間	消費者数 Gr. 1/Gr. 2	製品数	学習期間 購買数/人	検証期間 購買数/人
1	5,256/3,354	295	3.69	0.99
2	5,216/3,065	276	3.57	0.97
3	5,714/2,392	247	3.41	0.75
4	4,954/2,216	249	3.45	0.88
5	4,724/2,120	260	3.45	0.90
6	1,156/4,710	212	3.14	0.87
7	4,723/1,608	249	3.30	0.85
8	5,373/843	158	2.64	0.72

4.2 データ

大手セル・レンタル CD チェーン店から個人が特定できないように加工したうえで提供いただいた、音楽 CD 販売の購買履歴 417,380 件の ID 付き購買 POS データを検証に用いる。音楽 CD は衝動買いが少ない製品で、事前に楽曲を聴いて購入製品を決めてから店舗に行く買い方が全体の 82.5% を占めるという調査結果 [26] が示すように、探索コストが大きく繰り返し購買がない経験財という特徴を持っている。購買の期間は 2002 年 11 月 1 日から 2003 年 12 月 21 日までの 1 年あまりで、東京と大阪エリアの計 5 店の購買履歴を評価対象とする。表 3 には、実証実験 1 の第 1 期のデータセットか、あるいは $K = 3$ の条件で「学習期間に 3 枚 ($= K$) 以上の製品を購入した消費者」と「6 枚 ($= 2K$) 以上販売した製品」の組合せの 12 通りのデータセットの特徴を示した。 K の値が大きくなると、条件に該当する消費者数と製品数は減少する。抽出条件 K の値が同じ場合、各期のデータの特徴は同様であった。第 2 期以降では、 K が 6 または 7 で条件に該当するデータが残らない場合が 7 回あり、33 通りの有効なデータセットを構成し、分析対象とした。よって表 3 から割愛した実証実験 1 のデータセットが 21 ($= 33 - 12$) 通りある。1 人当たりの購買数平均が学習期間より検証期間で小さいのは、学習期間のデータ抽出に購買数の下限を設定したからである。表 4 には、実証実験 2 で用いるデータの各期の特徴をまとめた。2 つの消費者セグメントに共通して購買のある製品を対象とするため、実証実験 1 の第 1 期の場合に比べて製品数が少ない。

4.3 結果と考察

4.3.1 購買予測力の検証

表 5 に、抽出条件の最も緩い $K = 3$ のケースの Rank Scoring 指標とコルモゴロフ-スミルノフ検定指標の評価結果をまとめた。どちらの評価指標ともに値が大きいほうが予測精度が良い。期間による評価値の差異はモデルによる差異に比べて十分に小さく、2 つの評価法ともにすべての期間で評価の良い順に弱学習器モデル > 条件付き独立モデル > カーネル法 SVM となった。また、コルモゴロフ-スミルノフ検定で 3 つのモデルともに購買の有無の 2 グループの購買予測確率の分布に 0.1% 水準で有意な差があった。 $K = 4$ または 5 の抽出条件による各期のモデルの優位比較でも同様の結果を得た。表 6 には、第 1 期の抽出条件 K の異なるデータセットの評価結果を示す。抽出条件の K の値を大きくすると、学習期間で購買数の多い消費者

表 5 購買予測力の検証結果 (抽出条件 $K = 3$ の場合)
 表中の値: Rank Scoring 指標/コルモゴロフスミル
 ノフ検定指標

期間	K	カーネル法 SVM	条件付き 独立モデル	弱学習器 モデル
1	3	2.12/0.10	3.93/0.16	5.43/0.40
2	3	2.37/0.10	3.58/0.15	5.21/0.38
3	3	1.80/0.08	3.50/0.16	4.46/0.35
4	3	2.19/0.09	3.64/0.15	4.71/0.37
5	3	2.10/0.08	3.47/0.15	4.66/0.37
6	3	2.56/0.09	4.09/0.17	5.40/0.39
7	3	2.40/0.09	3.71/0.15	5.34/0.39
8	3	3.03/0.10	4.11/0.16	6.08/0.39
平均		2.32/0.09	3.75/0.16	5.16/0.38
標準偏差		0.37/0.01	0.26/0.01	0.53/0.02

表 6 購買予測力の検証結果 (第 1 期)
 表中の値: Rank Scoring 指標/コルモゴロフスミル
 ノフ検定指標

期間	K	カーネル法 SVM	条件付き 独立モデル	弱学習器 モデル
1	3	2.12/0.10	3.93/0.16	5.43/0.40
1	4	2.82/0.09	4.84/0.15	6.75/0.39
1	5	3.82/0.08	5.66/0.13	8.00/0.39
1	6	6.21/0.09	5.85/0.08	8.94/0.34
1	7	11.49/n.a.	6.61/n.a.	12.78/0.24

表中 n.a. は p 値が 5% 以上, 他は 0.1% 水準で有意.

と製品が評価対象となり, Rank Scoring 指標による一致率評価は良くなるが, 一方で, コルモゴロフスミルノフ検定指標による一致率評価は低下し, $K = 7$ でカーネル法 SVM と条件付き独立モデルの判別は有意でなかった. 2 つの指標の傾向が逆転したのは, コルモゴロフスミルノフ検定指標のみが評価する特定の製品を購入する消費者の順序の推計力が低下したのが要因である. 表 3 を確認してもらえれば, 学習期間に購買数の多い消費者の検証期間の購買数が多いとは限らないとわかる. 提案モデルは, 購買履歴数の多い消費者の予測購買確率の合計を高く推計するが, この水準が検証期間に持続しないことが消費者順序の推計を低下させた要因である. 消費者ごとにパラメータ推定するカーネル法 SVM でも同様の影響があり, 製品数が少ない場合に相対的に良い評価となったのは, 購買履歴数に対して推定するパラメータ数が減りオーバーフィットのリスクが減ったためと考える.

4.3.2 相関係数の同質性の検証

表 7 には, 購買の類似する 2 つの消費者セグメント

表 7 相関係数の同質性の検証結果: 弱学習器モデル
 表中の値: Rank Scoring 指標/コルモゴロフスミル
 ノフ検定指標

期間	自身の相関係数	交換した相関係数
1	2.87/0.02	2.89/0.07
2	2.81/0.02	2.89/0.07
3	3.07/0.03	3.07/0.05
4	2.88/n.a.	2.92/0.04
5	2.58/n.a.	2.73/0.05
6	3.69/0.07	3.60/0.06
7	2.77/0.02	2.83/0.05
8	4.40/0.05	4.09/0.05
平均	3.13/0.04	3.13/0.05
標準偏差	0.61/0.02	0.47/0.01

表中 n.a. は p 値が 5% 以上, 他は 1% 水準で有意.

について, それぞれの購買履歴から計算した相関係数を用いた場合と, 相関係数を交換して用いた場合の弱学習器モデルの評価結果を示す. 交換しても予測一致率に差異は見られない. また, 同じ抽出条件 $K = 3$ の実証実験 1 の結果に比べて評価値が悪化している. 相関係数を交換する必要から, 2 つの消費者セグメントで共通に購買履歴のある製品を評価対象としたため, 分析対象製品が広く買われる製品に限定されて購買予測が難しくなったと考える. 条件付き独立モデルについて同じ分析をしたところ, 同様の結果であった.

5. 製品の類似を利用したプロモーションの考察

5.1 製品の類似と売上の交差弾力性

音楽 CD の市場におけるプロモーションの特徴として, レーベルという商習慣がある. 昨今は, 出版社も叢書 (book series) をレーベルと称するのが一般化している. レコード会社や出版社は, 音楽性や作風の類似したアーティストや作者を同じレーベルとしてグループ化し, 会社の部門名に相当する組織でプロモーションを行う [20]. この仕組みは, 同じ企業ブランドの傘下にあるブランド・ポートフォリオ [27] に類似しているが, レコード会社は, 企業名やレコードレーベル名をブランドとして消費者に認知させる意図に乏しいという点でブランド戦略とは異なる. 実際, 音楽 CD をレーベルで判断する消費者は稀であろう. 類似認識について, アーティスト達をプロモーションする組織が同じなら, 同じメディアやイベントに所属アーティストが露出する機会が増えて, 消費者は無意識に類似関係を認識しやすくなる. 例えばレコード会社アップ

ロントワークスに所属するアーティストが、テレビ番組のハロー!モーニング（テレビ東京で2000年4月～2007年4月放送）に出演しているのを観た視聴者は、特に説明がなくても「モーニング娘。」に代表されるハロー!プロジェクトのメンバーと認識したであろう。音楽性の似通ったアーティストは競合する可能性も高いが、同じレコードレーベルに所属していれば、新譜の発売時期をずらすなど、競合要因を打ち消しやすい。

ある2製品の類似係数の指標が大きいということは、それらの製品は共に購入されやすい（並買されやすい）ことを表しており、補完財の関係にあるといえる。よって、2製品が類似していると消費者が認識することで、補完の性質が強まり、1つの製品の売上増加が、ほかの製品の売上を波及的に増やすと考えられる。レーベルの類似を認識させる機能に着目し、その効果を提案モデルを用いて考察する。

5.2 評価手法

本節では、提案モデルを用いて類似によるプロモーション効果の高いアーティストの組と、所属レコードレーベルとの関係について分析する。用いるデータとモデルは次のとおり。

データ 前節と同じ音楽CD販売データの2002年11月～2003年4月の期間の、月間平均1枚以上購入した消費者と2枚以上販売した製品の購買履歴

モデル 弱学習器モデル

アーティストの関連の強さとして、アーティストAの売上が増えることによる、アーティストBの売上の増分の影響を、需要の交差価格弾力性の算式に倣い、次のとおり売上の交差弾力性 $\delta_{A,B}$ を定義して評価する。

$$\begin{aligned} & \text{売上の交差弾力性: } \delta_{A,B} \\ & = \frac{\text{アーティストBの売上数量の変化率}}{\text{アーティストAの売上数量の変化率}} \quad (20) \end{aligned}$$

まず、弱学習器モデルにデータを当てはめて各消費者の購買確率を評価し、まだアーティストAを購入していないが購入すると予測される購買確率が高い消費者から順に実際の売上の10%に相当する消費者数を抽出し、その購買が新たに発生すると見込む。この見込み購買を購買履歴に追加し、購買確率を再度評価するとアーティストBの購買確率が増減する。この増分を、補完関係による購買の効果として先に定義した売上の交差弾力性として評価する。

仮定として追加する購買履歴数を10%という比率にしたのは、プロモーションの目標として売上の10%増

は現実的な設定と考えたためである。このように、類似関係による売上の増分を評価できるのは、順序統計量を推奨の指標とするモデルと異なり、提案モデルが確率モデルであることの利点である。ただし、ブランド選択確率をロジットモデルで推計する場合のように、購買確率の合計が売上に一致するような水準調整のロジックを提案モデルに組み込んでいない。本実験では影響の順序に関心があるが、絶対的な水準に着目する場合は水準調整のロジックが必要となろう。

5.3 結果

売上の交差弾力性の影響を、まずアーティストごとに集計し、次に市場全体で集計する。表8は、売上の交差弾力性の影響による売上の増加が高いアーティスト上位10組の表である。表の売上順位は、売上増を見込む前の実際の売上順位を示している。第8位の「浜崎あゆみ」から「B'z」への影響の組合せ以外の9組は、同じレコード会社・レーベルに所属しており、同じレコードレーベル内で、売上増加のシナジー効果が高いと評価できた。上位10組の特徴として、第6位（1位：CHEMISTRY-9位：平井堅）のようにベースとなる元の売上数量が多い組と、第1位（25位：松浦亜弥-79位：後藤真希）のように元の順位以上に補完関係が強い組が見られる。

次に、各アーティストの売上の交差弾力性の市場全

表8 売上の交差弾力性の影響の大きいアーティスト上位10組

No.	売上順位	要因 [所属]	売上順位	結果 [所属]
1	25	松浦亜弥 [UFW ^a]	79	後藤真希 [UFW ^a]
2	2	浜崎あゆみ [avex]	6	BoA [avex]
3	23	KinKi Kids [ジャニーズエンターテイメント]	90	J-FRIENDS [(ジャニーズ)]
4	2	浜崎あゆみ [avex]	13	Every Little Thing [avex]
5	25	松浦亜弥 [UFW ^a]	147	ごまっとう (後藤真希, 松浦亜弥, 藤本美貴) [UFW ^a]
6	1	CHEMISTRY [SME(J) ^b]	9	平井堅 [SME(J) ^b]
7	25	松浦亜弥 [UFW ^a]	74	藤本美貴 [UFW ^a]
8	2	浜崎あゆみ [avex]	3	B'z [ビーイング]
9	79	後藤真希 [UFW ^a]	147	ごまっとう (後藤真希, 松浦亜弥, 藤本美貴) [UFW ^a]
10	38	モーニング娘。 [UFW ^a]	79	後藤真希 [UFW ^a]

UFW^a : アップフロントワークス, SME(J)^b : ソニー・ミュージックエンタテインメント (日本).

表 9 売上の交差弾力性の市場への影響が大きいアーティスト上位 10 位

No.	売上順位	要因アーティスト名	所属レコード会社名
1	2	浜崎あゆみ	avex
2	6	BoA	avex
3	16	TLC	Sony Music Entertainment
4	13	Every Little Thing	avex
5	1	CHEMISTRY	SME(J) ^a
6	11	宇多田ヒカル	EMI ミュージック・ジャパン
7	25	松浦亜弥	アップフロントワークス
8	22	ジュニアファー・ロペス	Sony Music Entertainment
9	26	中島美嘉	SME(J) ^a
10	44	175R	EMI ミュージック・ジャパン

SME(J)^a: ソニー・ミュージックエンタテインメント(日本).

体に対する影響を集計し、表 9 に上位 10 位示す。表 8 と比較すると、限られた消費者に訴求する女性アイドルがランキング外になり、売上の交差弾力性の大きさと潜在顧客の多さの広がりバランスがとれたアーティストが上位になった。アイドルは特定のファンを持つアーティストとして、レコードレーベルの仕組みが特に有効に機能していると考ええる。

6. 議論と今後の課題

6.1 議論

過去の購買履歴から将来の消費者の購買確率を予測する算式を、明示的な仮定に基づいて導き、従来のアプローチで課題とされた類似評価の基準を購買履歴データの分布の問題に帰着させ、2 項分布の相関係数で類似評価する理論的な根拠を示した。また、分布に基づいて議論したことで、推奨の評価を順序統計量ではなく確率として示せた。購買予測力の検証の結果、提案モデルは予測対象を購買数の多い消費者と製品に限定した場合を除いて、カーネル法 SVM より高い購買予測力を示した。購買頻度の高いデータに対して説明力が劣るのは、類似情報の利用のみモデル化し、消費者ごとの購買数を予測するロジックを組み入れていないためと考える。音楽 CD 市場のように繰り返し購買がない購買履歴データに対しては、購買頻度の情報に比べて類似情報による説明力の役割が大きく、平均より相関係数に着目する提案モデルが有用である。一般的

な判別の統計モデルが共分散行列を利用するのに対し、基準化した共分散の相関係数のみならず他消費者のデータから推計できるとの仮定に基づいて、相関係数のみ利用する点が提案モデルの特徴であり、協調フィルタリングのアプローチである。一方、繰り返し購買があり購買頻度の高い市場では、個人属性や製品属性を用いて購買頻度の平均や分散を説明する統計モデルやコンテンツベースフィルタリングの有用性が増し、2 つのアプローチを統合するハイブリッドフィルタリングが有用となろう。相関係数の同質性の検証の結果から、製品間の相関係数の推計に消費者の異質性を考慮しない仮定 2 が妥当とわかった。製品間の類似を計算したデータ行列を転置して同じ計算をすれば消費者間の類似が求まるように、製品と消費者の類似は表裏一体の関係にあり、どちらかの情報を用いれば重複して用いる効果は少ないと解釈する。

実務への応用として、類似関係が購買に与える影響の高いアーティストの組を評価したところ、上位 10 組の中で 9 組が同じレコードレーベルに所属していることがわかった。このため同じレコードレーベルに属するアーティストのプロモーションには、補完関係による売上増加効果があると解釈できる。特に、アイドルの音楽ジャンルで有効であったのは、このジャンルは Kubacki [19] の指摘するプロモータータイプの特徴が強く、宣伝の効果が現れやすかったと解釈する。昨今、文学的であるより広く読まれることを第一に考える書籍のジャンルのライトノベルでレーベル(叢書)の用語が使われるようになった。レーベルは、財やサービスの提供者とは別に作者が存在する音楽や書籍などの製品をグループ化し、消費者に暗黙的に類似を認識させ、波及的に売上を伸ばすマーケティング手法として有効に機能していると考ええる。

6.2 今後の課題

提案モデルの課題として、協調フィルタリングのアプローチに沿って相関係数に着目する一方で、平均値となる頻度の予測を組み入れていない点がある。繰り返し購買のない市場では相関係数のみに着目しても有効であったが、繰り返し購買のある市場では頻度の推定の重要度が増す。2 項分布以外の分布を仮定して提案モデルを構築し直し、消費者や製品の属性情報を用いるコンテンツベースフィルタリングのアプローチを取り入れたり、時系列情報などの変数を組み入れて、頻度を予測するロジックを組み入れることが今後の課題となる。レコード会社のレコードレーベルを用いたプロモーションについては、アーティストの類似性がレ

コード会社のプロモーションによって高められたのか、あるいはレコード会社が類似性の高いアーティストを集めた結果なのかについて、議論できていない点が課題として残る。消費者が製品の類似を認識する要因を整理し、類似を利用したプロモーションの実際を評価する事が今後の課題である。

参考文献

- [1] J. S. Breese, D. Heckerman and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proceedings of the fourteenth Annual Conference on Uncertainty in Artificial Intelligence*, 43–52, 1998.
- [2] E. Brynjolfsson and A. McAfee, "Big data: The management revolution," *Harvard Business Review*, **60**, 60–68, 2012.
- [3] A. Tversky, "Features of similarity," *Psychological Review*, **84**, 327–352, 1977.
- [4] 清水聰, 『新しい消費者行動』, 千倉書房, 1999.
- [5] 澁谷覚, 『類似性の構造と判断—他者との比較が消費者行動を変える—』, 有斐閣, 2013.
- [6] G. Häubl and V. Trifts, "Consumer decision making in online shopping environments: The effects of interactive decision aids," *Marketing Science*, **19**, 4–21, 2000.
- [7] D. Goldberg, D. Nichols, B. M. Oki and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, **35**, 61–70, 1992.
- [8] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, "GroupLens: An open architecture for collaborative filtering of netnews," in *Proceedings of the 1994 International ACM Conference on Computer Supported Collaborative Work Conference*, 175–186, 1994.
- [9] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, **17**, 734–749, 2006.
- [10] 神鷹敏弘, "推薦システムのアルゴリズム (1)," 人工知能学会誌, **22**, 826–837, 2007.
- [11] 神鷹敏弘, "推薦システムのアルゴリズム (2)(3)," 人工知能学会誌, **23**, 89–103, 248–263, 2008.
- [12] D. Jannach and M. Zanker, A. Felfernig and G. Friedrich, *Recommender systems: An introduction*, Cambridge University Press, 2010 (田中克己・角谷和俊 監訳), 『情報推薦システム入門—理論と実践—』, 共立出版, 2012.
- [13] M. Wedel and K. A. Wagner, *Market Segmentation: Conceptual and Methodological Foundations 2nd ed.*, Kluwer Academic Publisher, 2000.
- [14] A. N. Albatineh, M. Niewiadomska-Bugaj and D. Mihalko, "On similarity indices and correction for chance agreement," *Journal of Classification*, **23**, 301–313, 2006.
- [15] J. C. Gower and P. Legendre, "Metric and Euclidean properties of dissimilarity coefficients," *Journal of Classification*, **3**, 5–48, 1986.
- [16] 石田実, 西尾チヅル, 椿広計, "2 値変量に基づく教師無し分類における類似係数の選択," 行動計量学, **38**, 65–81, 2011.
- [17] D. Iacobucci, P. Arabie and A. Bodapati, "Recommendation agents on the internet," *Journal of Interactive Marketing*, **14**, 2–11, 2000.
- [18] 本多克宏, "クラスタリングの概念と意思決定支援への応用", オペレーションズ・リサーチ, **57**, 250–254, 2012.
- [19] K. Kubacki and R. Croft, "Mass marketing, music, and morality," *Journal of Marketing Management*, **20**, 577–590, 2004.
- [20] H. L. Vogel, *Entertainment Industry Economics: A Guide for Financial Analysis, 8th ed.*, Cambridge University Press, 2011 (助川たかね 訳), ハロルド・ヴォーゲルのエンタテインメント・ビジネス—その産業構造と経済・金融・マーケティング—, 慶応義塾大学出版会, 2013.
- [21] 勝又壮太郎, 阿部誠, "顧客嗜好の時間的変化を組み込んだ音楽 CD 選考モデルの構築と CRM への応用," マーケティング・サイエンス, **16**, 25–47, 2007.
- [22] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006 (元田浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田昇 訳), 『パターン認識と機械学習 上, 下—ベイジ理論による統計的予測—』, シュプリンガー・ジャパン, 2007, 2008.
- [23] R Core Team, R: A language and environment for statistical computing, *R Foundation for Statistical Computing, Vienna, Austria*, <http://www.R-project.org/> (Accessed 12 June 2013).
- [24] C.-C. Chang and C.-J. Lin, LIBSVM: A library for support vector machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [25] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data*, John Wiley & Sons, Inc., 1990.
- [26] 一般社団法人 日本レコード協会, 「2009 年度音楽メディアユーザー実態調査」, <http://www.riaj.or.jp/report/mediauser/pdf/softuser2009.pdf>
- [27] D. A. Aaker, *Brand Portfolio Strategy*, Free Press, 2004 (阿久津聡 訳), 『ブランド・ポートフォリオ戦略』, ダイヤモンド社, 2005.