

未購買アイテムを考慮した 顧客潜在価値による販売促進への活用

中原 孝信

1. はじめに

小売業にとって、売上に影響を与える顧客の購買行動を理解することは、非常に関心の高いテーマである。これまでに、購買行動モデルを用いたブランド選択に関する研究[1]や、販売促進の効果に関する研究[2]、そしてデータマイニングを用いた購買行動に関する研究[3][5]など、数多くの研究が行われてきた。また、実務では、RFM分析やデシル分析など、顧客のランク付けを行うために、過去に購買した顧客の購買金額や来店頻度を利用した分析が用いられてきた。

これまでに行われてきた顧客購買履歴データ（ID付きPOSデータ）を活用した研究は、非常に有用なものであるが、多くの研究は、購買したという過去の結果にのみ焦点をあてたものである。しかし、現実には、迷ったが購買には至らなかった、自分の欲しい商品が品切れで購買には至らなかったなど、最終的には購買に結び付かなかったが、購買意向を持つ顧客は多く存在している。そのような顧客を発見することは、効率的な販売促進の実施や、売上増加のための対象顧客として選択することがきるため、小売店にとって非常に重要なテーマである。

本研究では、商品を購買する可能性の高い顧客を発見するために、推薦システム[4][6]の枠組みを応用して、これまでに顧客が購買した商品の情報から、他の商品を購買する可能性を考慮した顧客潜在価値（Customer Potential Value）という指標を提案する。そして最終的に顧客潜在価値を用いた販売促進に関する提案を行う。

本稿の構成は、2節で今回の事例研究で利用したデ

ータについての簡単な基礎分析の結果を示す。3節では、顧客の潜在価値を求める方法について説明する。そして4節にて、提案する手法を実際のデータに適用して得られた結果を示し、その考察を行う。最後に5節にて、まとめと課題について述べる。

2. 分析対象データ

分析対象のデータは、マルイ3店舗（有楽町店、渋谷店、池袋店）の1年間（2008年4月～2009年3月）の顧客購買履歴データ¹である。レコード数は約312万件であった。マルイでは、会員カードを配布しており、今回利用した購買データに含まれる顧客数は約55万人で、そのうち女性会員が約75%を占めている。利用できる項目は、ロックコードと呼ばれるブランドを識別できるコード、そして商品を識別できるアイテムコードであり、その売上数量と金額などが利用できる。ただしブランド名や商品名は識別できないデータである。

基礎的な分析の結果から、売上金額は、3店舗の中では有楽町店が最も高く、残りの2店舗は同程度であることが確認できた。また、7月、9月そして1月の売上金額が各店舗で高く、セールや初売りの効果が顕著に表れており、月によって売り上げに変動がある。次に曜日別では、平日と休日の売り上げに違いがあり、休日は平日に比べて倍程度の売上金額となっている。年代別では、20代で全顧客の約60%、30代を含めると全顧客の約85%を占めており、その大部分が女性顧客である。渋谷店は他の2店舗に比べて来店あたりの平均購買額が高くなっている。また、どの店舗でも30代以上の顧客は来店あたりの平均購買額が高い。

これらのことから、20代は客数が多く最も売り上げに貢献している顧客であるが、30代以上の顧客は、

¹ 平成21年度データ解析コンペティションで提供していただいたデータを用いている。

20代よりも平均購買額が高く購買力を持っているため、今後、店舗の売り上げに貢献する可能性を持っていると見なすことができる。このように何らかの方法で売り上げに貢献する可能性を持った顧客を発見し、アプローチすることで売り上げの増加につなげることができるために、潜在的な購買力を持つ顧客を見つけることは重要である。

次節では、潜在的な購買力を持つ顧客を発見するために、本研究で提案する顧客潜在価値モデルについて説明する。

3. 顧客潜在価値モデル

これまで行われてきた多くの研究は、購買した事実にのみ着目していたため、その時点では購買には至らなかつたが、今後商品を購買する可能性のある顧客を見過ごしていた。顧客が本来持っている価値を測定するためには、購買した事実と購買する可能性を考慮した潜在的な価値を明らかにすることが重要である。本研究の目的は、購買したアイテム²の情報に加えて、未購買のアイテムを購買する可能性を考慮し、顧客の持つ潜在的な購買力を表す顧客潜在価値を明らかにすることである。提案するモデルは、対象顧客が購買したアイテムとその顧客の未購買のアイテムに対する購買の可能性を、全顧客の購買履歴から決定する。そのために、任意の期間内で購買されたアイテム間の関係と、その併買確率を表したグラフを全顧客の購買履歴データから作成する。これを潜在価値グラフと呼ぶことにする。次に、潜在価値グラフを利用し、対象顧客が購買したアイテムと関連する未購買のアイテムの購買確率を算出する。これをアイテム潜在価値と呼ぶ。そして顧客潜在価値は、各顧客が持つアイテム潜在価値を利用して算出する。

3.1 潜在価値グラフの生成

潜在価値グラフを作成するに際して、任意の期間 t で顧客集合 $C = \{c_1, c_2, \dots, c_n\}$ により購買されたアイテム集合 I を頂点集合 V で表す。そして、ある顧客に併買された 2 つのアイテム $x, y \in I$ を頂点 $v_x, v_y \in V$ で表し、その接続を無向枝 $e(v_x, v_y)$ で表す。このとき、無向枝 $e(v_x, v_y)$ に与える重みを $w(v_x, v_y)$ で表し、全顧客によって購買されたアイテム同士の併買関係を重み付きの無向グラフで表現する。

ここで、無向枝に与える重みを分析対象データから

² ここでアイテムとは、商品、ブランド、そして分類など分析を実施する上で用いる対象を意味する。

検討する。枝に与える重みは、各アイテム間の関係の強さを表す値が望ましい。アイテム間の関連の強さを表す代表的な値としては、確信度 (*confidence*)、支持度 (*support*)、そしてリフト値 (*lift*) などの指標が提案されている[7]。2つのアイテム x, y の確信度は式(1)で与えられる。

$$confidence(x, y) = \frac{x, y \text{ の併買者数}}{x \text{ の購買者数}} \quad (1)$$

これは、アイテム x の購買を条件としたときの、アイテム y が購買される条件付き確率を意味する。したがって、確信度を計算する際には、アイテム間の方向性を考慮する必要があり、無向枝に対する重みには適切ではない。次に、支持度は式(2)で与えられる。

$$support(x, y) = \frac{x, y \text{ の併買者数}}{\text{対象全顧客数}} \quad (2)$$

支持度は分子の数にその値が依存するため、あるアイテム間だけの関係に着目した場合に、高い確率で発生する併買関係を過小評価してしまう傾向がある。

以上の考察の結果、本データに対してはマーケットバスケット分析で利用されることの多い、リフト値を枝に与える重みとして利用する。ここでは、リフト値を用いた無向枝の重みを以下の式で定義する。

$$w(v_x, v_y) = \frac{P(v_x \cap v_y)}{P(v_x) \cdot P(v_y)} \quad (3)$$

ここで、 $P(v_x) = x$ の購買者数/対象全顧客数、 $P(v_y) = y$ の購買者数/対象全顧客数とする。そして、 $P(v_x \cap v_y) = x, y$ の併買者数/対象全顧客数である。したがって、リフト値を重みに持つ $w(v_x, v_y)$ は、併買確率の実測値と期待値を比較した値で $w(v_x, v_y) > 1$ ならば 2 つのアイテム間の関連性が高く（正の関連）、1 以下ならば関連性が低い（負の関連）ことを意味する。

リフト値は、アイテム x と y の購買者が少ない場合に分母の値は小さくなり、分母が非常に小さい数の場合は、リフト値が極端に大きくなってしまう。この問題を回避するために、 $P(v_x) \cdot P(v_y)$ がある閾値 α より小さいときは、リフト値を 1 とすることで、次に説明する顧客潜在価値を算出する際に一切影響を与えないこととする。この点を反映させた無向枝の重みを以下の式(4)で定義します。

$$w(v_x, v_y) = \begin{cases} \frac{P(v_x \cap v_y)}{P(v_x) \cdot P(v_y)} & \text{if } P(v_x) \cdot P(v_y) \geq \alpha \\ 1 & \text{if } P(v_x) \cdot P(v_y) < \alpha \end{cases} \quad (4)$$

この定義から、アイテムの購買確率の積

$P(v_x) \cdot P(v_y)$ が低いアイテム間の関連性を考慮しないことになる。閾値 α の設定法は、後述する適用例で示す。

3.2 アイテム潜在価値と顧客潜在価値

次に、アイテム潜在価値と顧客潜在価値を算出する。ここでは、潜在価値グラフを利用し、以下の方法で顧客 c_j の未購買のアイテムに対するアイテム潜在価値と、顧客 c_j の顧客潜在価値を計算する。

- 1) 潜在価値グラフから、重み $w(v_x, v_y) \leq 1$ の枝を除去
- 2) 潜在価値グラフから、顧客 c_j の購買済みアイテム同士が接続された枝を除去
- 3) 潜在価値グラフから、顧客 c_j の購買済みアイテムと接続された枝以外を除去
- 4) 顧客 c_j の未購買アイテムと接続された枝の重みの合計値を、そのアイテムのアイテム潜在価値として計算
- 5) 顧客 c_j のアイテム潜在価値を合計したものを顧客 c_j の顧客潜在価値とする

図1は、潜在価値グラフのイメージを示しており、点の番号はアイテム、そして枝の横の数値は枝の重みを表している。上記の手順によって、このグラフからアイテム潜在価値が算出される。太い丸で囲まれた点は、顧客 c_j によって購買された各アイテムを示している。ここで1)に従い、枝の重みが1以下である点線で示した $e(2, 4), e(6, 7)$ の枝は除去される。次に2)より、購買済みアイテム同士を結ぶ枝 $e(3, 6)$ は除去され、また3)より枝 $e(8, 9)$ が除去される。そして4)より、購買済みアイテムと枝で結ばれた未購買アイテム集合 $\{2, 5, 8\}$ に対して、各アイテムのアイテム潜在価値が計算される。例えば、アイテム5の場合は、点1, 4, 6と接続があるので、各枝の重みの合計値 ($1.1 + 1.5 + 1.8 = 4.4$) がアイテム5のアイテム潜在価値となる。このようにして、すべての購買済みアイテムと接続のある未購買アイテムに対してアイテ

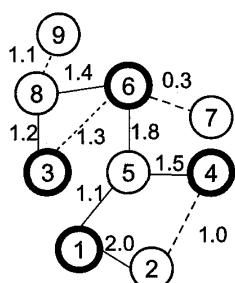


図1 潜在価値グラフを用いたアイテム潜在価値の算出

ム潜在価値が計算される。最後に顧客 c_j の顧客潜在価値は、顧客 c_j が持つ各アイテムのアイテム潜在価値を合計した値であり、この場合は、アイテム2の2.0、アイテム5の4.4、そしてアイテム8の2.6が合計された値で9となる。

これらの手順によって算出されたアイテム潜在価値は、他人が併買しやすいアイテム群の1つを対象顧客が購入していると上昇する傾向にある。また、アイテム潜在価値の高いアイテムは、対象顧客がこれまでに購入した多くのアイテムと強い関連を持っているため、薦めると購買につながる可能性の高いアイテムと見なすことができる。また、顧客潜在価値は、未購買アイテム群を購買する可能性を示した値であり、顧客潜在価値の高い顧客は、潜在的な購買力を持っていることを意味する。

4. 百貨店データからの顧客潜在価値の算出

提案した顧客潜在価値モデルをもともと売り上げの高い有楽町店のデータに適用した。その際、3分の2の顧客は来店回数が1回であったため、前期半年間の購買履歴データから、2回以上来店した顧客約6万人を分析の対象とした。また、後期半年間のデータは、販売促進を想定した顧客潜在価値によるアイテム推薦の検証用データとして利用する。ここでアイテムは、ブランドとして識別できるブロックコードを利用し、合計で252種類のブロックコードを扱う。潜在価値グラフは、半年間の各月で $t=1, 2, \dots, 6$ として生成する。これは基礎分析の結果から、各月ごとに売上金額は異なっており、百貨店という業態から同一の商品を長期間販売することはないため、月ごとの購買行動を併買の基準としてモデルを適用することにした。したがって、顧客 c_j が半年間の中で購買した t ごとにアイテム潜在価値と顧客潜在価値が計算され、最終的に顧客 c_j の顧客潜在価値は、各 t で計算した顧客潜在価値の合計値になる。

ここで潜在価値グラフを生成する際の枝の重みに関して、3.1節で示したように閾値 α を定める必要がある。図2は、全アイテムのペアについてのリフト値と $P(v_x) \cdot P(v_y)$ の関係を散布図で示している。図中の各点は2アイテムのペアに対応している。 $P(v_x) \cdot P(v_y)$ の値が0.001を下回る（図中の横線より下側）と、リフト値の値が極端に大きくなるケースが多くなっていることが確認できる。したがって、ここでは閾値 α

を 0.001 として枝の重みを計算する。 $P(v_x) \cdot P(v_y)$ が α より小さいアイテムペアの重みは 1 となり、これらの関連性はアイテム潜在価値を算出する際には、無視されることになる。

図 3 は、得られた潜在価値グラフを図示しており、6月の購買履歴データを用いて作成したものである。点はアイテム、各枝についている数値は枝の重み、そして枝の太さは枝の重みを相対的に表している。ここでは、図が複雑になるため、枝の重みが 1 よりも大きいアイテム間の関係だけを示している。婦人雑貨 1525 は、多くのアイテムと接続があり、併買されやすいアイテムである。また、婦人アパレル 0459 と婦人アパレル 0117 は太い枝で接続されており、この 2 つは、併買されやすいアイテムである。このような潜在価値グラフからアイテム潜在価値と顧客潜在価値が算出される。

図 4 は、得られた顧客潜在価値と前期半年間の購買

数量との関係を示している。散布図の各点は一人の顧客を示している。顧客潜在価値と購買数量の相関係数は 0.36 で、購買金額との相関係数は 0.025 であった。これは購買済みアイテムに基づき、顧客潜在価値を計算しているため、金額よりも購買数量との関係が強くなっていると考えられる。顧客潜在価値は、ある顧客が多くのアイテムを購入した場合に、潜在価値グラフで残される枝が増加することから、その顧客の顧客潜在価値は増加する傾向にある。しかし、図中の丸で囲んだ部分は、購買数量が多いが顧客潜在価値の低い顧客が集まっている。このような顧客は、併買されやすいアイテムペアの両方をたくさん購入しており、未購買のアイテムとの接続が少なくなっているため、顧客潜在価値が低くなっていると考えられる。一方で、購買数量はさほど多くないが、顧客潜在価値の高い顧客は、多くのアイテムと併買されやすいアイテムを購入しており、未購買のアイテムとの接続が多くなるため、

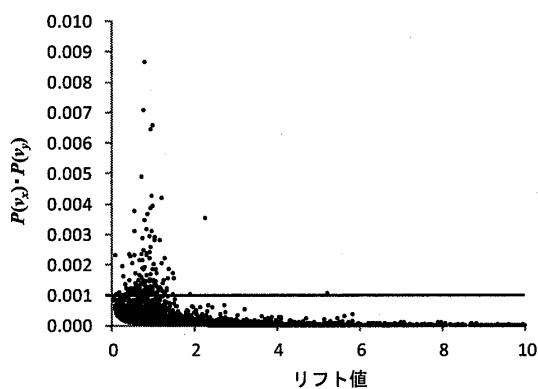


図 2 リフト値と $P(v_x) \cdot P(v_y)$ の関係

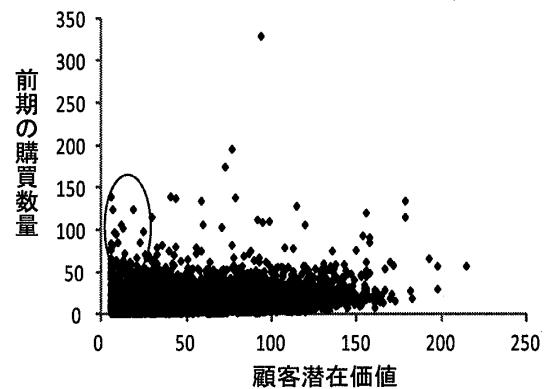


図 4 顧客潜在価値と購買数量の関係

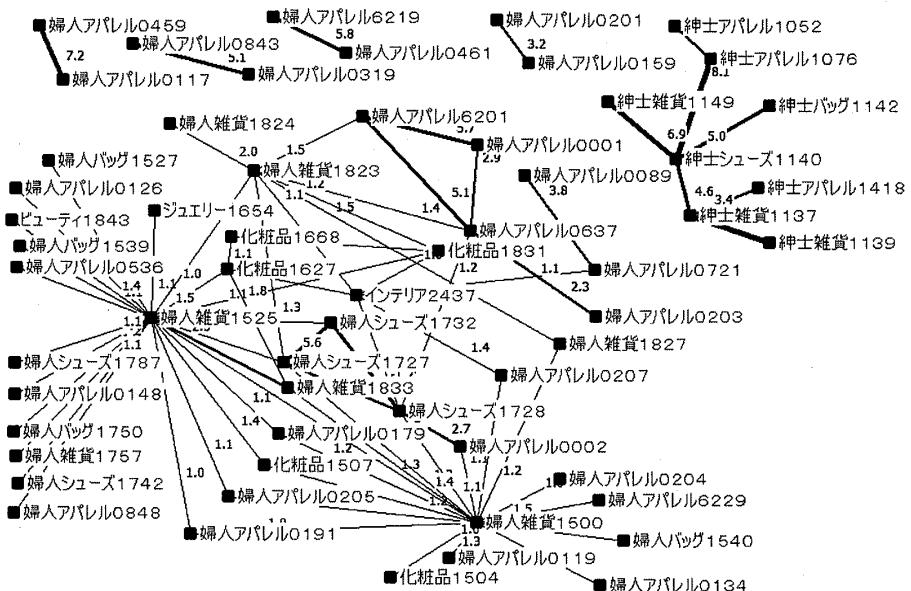


図 3 潜在価値グラフ

顧客潜在価値が高くなると考えられる。

算出した顧客潜在価値は、各顧客の購買済みアイテムと関連のある未購買のアイテム集合を購買する可能性を示した値であり、この値が高い顧客は、潜在的な購買力を持っている。したがって、この値を利用することで、効率的なプロモーションを行うことが可能になると考えられる。

4.1 顧客潜在価値による顧客のランク付け

CRM (Customer relationship management) では、売り上げに貢献する優良顧客や、今後、優良顧客になることが期待される顧客を特定し、それらの顧客と長期的に良好な関係を維持することで、継続的に利益を獲得しようという試みが行われている。実務では、そのような顧客を特定するために、RFM分析を用いた方法が実施されている。RFM分析は、Recency, Frequency, Monetary の3つの軸で顧客のランク付けを行う方法である。Recency (以下 R と呼ぶ) は、もっとも近い購買日からの経過日数、Frequency (以下 F と呼ぶ) は、期間内の購買頻度、そして Monetary (以下 M と呼ぶ) は、期間内の購買総額を表しており、それら3つの値をそれぞれ5等分に分類して、R, F, M のランクによって顧客がランク付けされる。

RFM分析で利用される3つの評価軸は、いずれも購買が行われた事実のみを扱ったものであり、顧客の潜在的な価値を評価することはできない。そこで、顧客の潜在的な価値を考慮した顧客のランク付けを実施するために、RFM それぞれの値と顧客潜在価値 (CPV) との相関係数を計算した (表1)。表1から、R, F, M と顧客潜在価値の相関係数でもっとも高い値は、F との相関で 0.319 である。したがって、ここでは F の代わりに CPV を利用することで頻度の代わりに顧客潜在価値を考慮した顧客のランク付けを行う。そして、ターゲット顧客を選択し、その顧客集合に対して販売促進を想定したアイテムの推薦を実施する。

4.2 顧客潜在価値を用いた販売促進方法の提案

図5は、R, CPV, M による顧客のセグメントを示

表1 RFM と顧客潜在価値の相関係数

指標	指標	相関係数
R	F	-0.282
R	M	-0.154
R	CPV	-0.113
F	M	0.464
F	CPV	0.319
M	CPV	0.022

しており、それぞれの値を人数が均等になるように高中低の3等分に分けて立方体で表現している。手前の色のついているマスは、R, CPV, M がいずれももっとも高いセグメントであり、前期半年間のデータでは、4,177人が属していた。このセグメントに属する顧客は、直近の来店があり、期間内の購買額が高く、さらに潜在的な購買力を持った顧客を意味している。一方で、奥の色のついているマスは、R と CPV が高く M の低いセグメントであり、前期半年間のデータでは、1,418人が属していた。このセグメントに属する顧客は、直近の来店があり、潜在的な購買力は高いが、実際の購買額は低い顧客である。したがって、このような顧客は、適切に販売促進を実施することで、売り上げの増加に結びつく可能性が高い顧客であり、魅力的なセグメントであると見なすことができる。このように、CPV を利用したセグメントを行うことで、購買した過去の事実に基づくだけではなく、顧客の潜在的な購買力を考慮したターゲティングが可能となる。

次に、この R と CPV が高く M の低いセグメントに属する 1,418人の顧客集合の中で、後期半年間に5個以上のブロックを購買した 192人に対して、アイテム潜在価値を利用してアイテム (ブロックコード) の推薦対象を選択する。アイテム潜在価値は、ある顧客が購買したアイテムと未購買のアイテムとの関連性の強さによって値が高くなるため、アイテム潜在価値の高いアイテムを推薦することで、未購買のアイテムの購買が期待される。ここでは、顧客ごとにアイテム潜在価値の高いブロックを5つ推薦した際に、後期半年間のデータでその顧客に推薦したアイテムが購買されていれば正解とする。アイテム潜在価値は、半年の各月で生成した潜在価値グラフを利用して算出されるため、同一のアイテムが複数の潜在価値グラフで出現している場合は、それらの合計値がそのアイテムのアイ

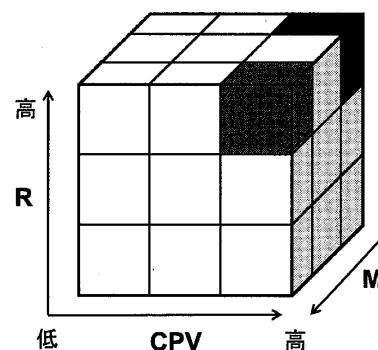


図5 顧客潜在価値によるセグメンテーション

表2 推薦結果

	正解数	不正解数	正解割合
アイテム PV	119	841	12.4%
ランダム	24	936	2.5%

テム潜在価値になる。

アイテム潜在価値を用いた推薦は、前期に併買の少ないアイテムだけを購買するなど、アイテム間の接続関係によっては、必ずしも5つの推薦アイテムを選択することができない場合がある。本データでは、対象顧客192人の中で、推薦アイテム数が5個よりも少なくなる顧客が3人いることを確認している。本実験では、それらの顧客に対しては、5個になるまで残りのアイテムをランダムに選択することにした。実際の適用で推薦アイテム数が少ない顧客が多く存在するならば、併買として扱う期間を長くするなどの工夫により、推薦アイテム数を増やすことなどが必要となる。

最後に、アイテム潜在価値を用いた推薦の効果を検証するために、ランダム推薦との正解数を比較した。ランダム推薦は、前期に購買された252種類のブロックを対象に、各顧客が前期に未購買であったブロックから5つのブロックをランダムに選択して推薦する。したがって、いずれの推薦方法も合計960個のブロックが推薦される。

表2は、アイテム潜在価値（アイテムPVと表記する）を用いた推薦の結果とランダム推薦による結果をそれぞれ示している。アイテム潜在価値を利用した推薦は、ランダムによる推薦と比較して約5倍の正解数であり、正解割合から、推薦したアイテムの約1割は実際に購買が行われていることが確認できた。これは、ある程度の精度を示しており、特に正解が多かったブロックは婦人雑貨や婦人アパレルなどの婦人関係のブロックであった。検証用に用いた後期半年間のデータは、推薦を行わない状況で得られたデータであり、実際に推薦を実施することでこれらの値よりも、さらに多くの購買が行われると考えられる。

5. おわりに

本研究では、これまでに購買されたアイテム間の関係を表現した潜在価値グラフから、各顧客の購買した商品と未購買の商品との購買確率を計算し、顧客潜在価値を算出する方法を提案した。顧客潜在価値は、未購買の商品に対して顧客が持っている潜在的な購買力を示した値であり、顧客潜在価値の応用例として、セ

グメントとターゲティングに顧客潜在価値を利用して、より効率のよい販売促進を実施するための提案を行った。アイテムの推薦では、ランダム推薦に比べて、アイテム潜在価値を利用した推薦は約5倍の正解率であり、その有効性をある程度示すことができた。今後は、今回の研究をベースに、同一商品の繰り返し購買が存在する食品などに対して顧客潜在価値モデルを適用し、より実用性の高い方法を構築していきたいと考えている。

謝辞 本研究の一部は、「文部科学省私立大学戦略的研究基盤形成支援事業（平成21年度～平成25年度）」、文部科学省の科研費22243033の助成、そして、平成22年度関西大学学術研究助成基金（奨励研究）において、研究費を受けたものの成果として公表するものである。

参考文献

- [1] Guadagni, P.M. and Little, J.D.C.: "A logit model of brand choice, calibrated on scanner data," *Marketing Science*, Vol. 2, pp. 203-238 (1983).
- [2] Gupta, S.: "Impact of sales promotions on when, what, and how much to buy," *Journal of Marketing Research*, Vol. 25, pp. 324-355 (1988).
- [3] Hamuro, Y., Katoh, N., Matsuda, Y. and Yada, K.: "Mining pharmacy data helps to make profits," *Data Mining and Knowledge Discovery*, Vol. 2, pp. 391-398 (1998).
- [4] Nakahara, T. and Morita, H.: "Recommender System for Music CDs Using a Graph Partitioning Method," *Knowledge-Based and Intelligent Information and Engineering Systems*, Vol. 5712, pp. 259-269 (2009).
- [5] 中原孝信、森田裕之：「百貨店のクレジット購買データを用いた関連購買による顧客特徴分析」、オペレーションズ・リサーチ、Vol. 50, No. 7, pp. 488-494 (2005).
- [6] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: "GroupLens: An open architecture for collaborative filtering of netnews," *In proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, pp. 175-186 (1994).
- [7] Tan, P.-N., Kumar, V. and Srivastava, J.: "Selecting the right Interestingness Measure for Association Patterns," *Proc. of the Eight International Conference on Knowledge Discovery and Data Mining*, pp. 32-41 (2002).