

研究種目：基盤研究 (C)
研究期間：2005 ～ 2008
課題番号：17510114
研究課題名 (和文) CRM とインターネット：優良顧客識別アルゴリズムの評価とブラウジング行動解析
研究課題名 (英文) CRM and the Internet: Evaluation of Customer Segmentation Algorithms and Analysis of Browsing Behavior
研究代表者 住田 潮 (SUMITA USHIO)
筑波大学・大学院システム情報工学研究科・ 教授
研究者番号：10236044

研究成果の概要：顧客層をより精密に類型化し、購買前段階行動を反映する新たなマルコフ・モデルを確立した。ほぼ完璧な形で現実データに対する適応性を実現すると同時に、パラメーターの推定値そのものによって、顧客行動やブランド力を構造的に把握することを可能にした。本研究は、2005年度データ解析コンペティション社会人の部、最優秀賞（第1位）を受賞。さらに、2006年度日本オペレーションズ・リサーチ学会事例研究賞も受賞した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2005年度	2,100,000	0	2,100,000
2006年度	600,000	0	600,000
2007年度	600,000	180,000	780,000
2008年度	300,000	90,000	390,000
年度			
総計	3,600,000	270,000	3,870,000

研究分野：マーケティング応用確率論

科研費の分科・細目：社会・安全システム科学・社会システム工学・安全システム

キーワード：マーケティング、CRM、識別手法、市場ライフサイクル、顧客ロイヤルティ、確率的数値解法

1. 研究開始当初の背景

インターネットの本格的普及に伴い、消費者を取り巻く環境は大きく変化し、マーケティングの分野では、その中心がマス・マーケティングから1対1マーケティングへと移行しつつある。この基本的動向の中で、CRM (Customer Relationship Management) がますます重要視されるようになった。CRM とは、個々の顧客と企業とが双方向で情報の遣り取りを継続する密接な関係を維持し、顧客の

生涯価値を拡大するための経営手法である。CRM の主要な機能としては、ポイント・インセンティブなどによる顧客属性や購買データの収集とデータベースの構築、収集されたデータの分析に基づくアクションプランの作成と実施を実現する仕組み等が挙げられるが、とりわけ重要となるのが、優良顧客の識別分析である。顧客の順位付けに関する分析 (例えば RMF 分析やクロス分析) を通して利益貢献が高いとされる顧客層を探り当て、

将来の購買行動を予測することによって優良顧客を識別する。さらに、識別された優良顧客に対してプロモーションを実施する等のアクションを起こし、優良顧客の維持や収益の増加を目指すことになる。顧客の識別手法としては、データマイニングの分野でよく用いられる決定木やニューラルネット、統計的手法であるロジスティック回帰分析、あるいは、最近、脚光を浴びつつある SVM (Support Vector Machine) など様々なものが存在する。それらをどのように評価し選択するかという問題や、さらには、個々の手法においても多くのパラメータ設定が必要であり、その評価・選択をどのように行うかというモデル選択の問題が生じる。

2. 研究の目的

本研究の目的は、リコールと精度を基礎として、識別された優良顧客に対してプロモーションを行う際の利益効果をモデル化し、新たな識別手法の比較評価を行う方法論を確立することにある。具体的には、優良でない顧客を優良顧客と識別してしまう過ちを第1種の間違い、優良顧客を優良でない顧客と識別してしまう過ちを第2種の過ちと定義する。識別された優良顧客に対してプロモーション・キャンペーンを行うことを考えると、第1種の過ちは非効率的なキャンペーン費用の増大をもたらす。第2種の過ちは機会損失の増大をもたらす。優良顧客の識別基準を緩めると、第1種の過ちが大きくなる一方、第2種の過ちが小さくなる。逆に、優良顧客の識別基準を厳しくすると、第1種の過ちが小さくなり第2種の過ちが大きくなる。このトレード・オフ効果を評価するため、プロモーション・キャンペーンの結果に対する利益関数を導入し、利益関数を最大にする判別基準をラグランジュ乗数法によって求める。さらに、具体的な識別アルゴリズムに対して、リコールや精度を推定し、その結果に基づく最適利益関数の値によって比較評価する。また、顧客のブラウジング行動の解析を通して e-マーケティングの新たな方法論の確立に寄与する。

3. 研究の方法

X 人の顧客集合を考え、識別の閾値 $z(0 \leq z \leq 1)$ が与えられたとき、顧客集合を優良顧客 (+1) と非優良顧客 (-1) とに識別する任意のアルゴリズムを考える。例えば、各顧客について過去の月別購買額デシルが与えられたとき、翌月に購買額でトップ 10%に入る顧客を優良顧客として識別するといった問題である。識別アルゴリズムには様々なものがあるが、どのようなアルゴリズムを用いたとしても、結果は下記の表のよう

にまとめられる。

		識別		合計
		-1	+1	
実際	-1	$x_{11}(z)$	$x_{12}(z)$	X^-
	+1	$x_{21}(z)$	$x_{22}(z)$	X^+
合計		$\hat{X}^-(z)$	$\hat{X}^+(z)$	X

ここで、 $x_{12}(z)$ が第1種の過ちを表し、 $x_{21}(z)$ が第2種の過ちを表すことに注意しておく。識別評価を行うための指標として、リコール、精度、正答率を定義する。記法の便宜上、 $x_{ij}(z)$ をセル関数と呼ぶことにする。リコール $R(z)$ とは、実際の優良顧客のうち正しく優良顧客と識別された顧客の割合であり、

$$(1) \quad R(z) = \frac{x_{22}(z)}{x_{21}(z) + x_{22}(z)} = \frac{x_{22}(z)}{X^+}$$

で与えられる。精度 $P(z)$ とは、優良顧客と識別された顧客のうち実際に優良顧客となった顧客の割合であり、

$$(2) \quad P(z) = \frac{x_{12}(z)}{x_{11}(z) + x_{12}(z)} = \frac{x_{12}(z)}{\hat{X}^-(z)}$$

となる。正答率 $A(z)$ は、全顧客のうち正しく識別された顧客の割合であり、

$$(3) \quad A(z) = \frac{x_{11}(z) + x_{22}(z)}{x_{11}(z) + x_{12}(z) + x_{21}(z) + x_{22}(z)} = \frac{x_{11}(z) + x_{22}(z)}{X}$$

で与えられる。

また、 $A(z)$ を $P(z)$ と $R(z)$ によって表すと、

$$(4) \quad A(z) = \frac{X^-}{X} + \frac{X^+}{X} R(z) \left(2 - \frac{1}{P(z)} \right)$$

となることに注意しておく。

翌月の優良顧客 (非優良顧客) の期待購買額を $\theta^+(\theta^-)$ 、プロモーションへの反応を $\eta^+(\eta^-)$ 、一人当たりのプロモーション費用を v とすると、優良顧客の識別が与えられた際のプロモーション・キャンペーンによる利益の増分 $\Delta V(z)$ は、次式で与えられる。

$$(5) \quad \Delta V(z) = \left(\frac{\theta^- \eta^-}{100} - v \right) x_{12}(z) + \left(\frac{\theta^+ \eta^+}{100} - v \right) x_{22}(z) \\ = X^+ R(z) \left\{ \frac{\gamma^-(1-P(z))}{P(z)} + \gamma^+ \right\}$$

ただし、

$$(6) \quad \gamma^- = \frac{\theta^- \eta^-}{100} - v, \quad \gamma^+ = \frac{\theta^+ \eta^+}{100} - v$$

である。従って、任意の識別アルゴリズムが与えられたとき、(5)式を最大化する z^* を求めることが問題となる。

一般的にこの問題を解くことは困難であり、先ず、 $x_{ij}(z)$ を閾値 z の関数として見たとき、以下の自然な仮定を置く。

<仮定1> $x_{11}(z)$ と $x_{21}(z)$ は z の増加関数で $\lim_{z \rightarrow 0} x_{11}(z) = \lim_{z \rightarrow 0} x_{21}(z) = 0$ かつ、 $\lim_{z \rightarrow 1} x_{11}(z) = X^-$ $\lim_{z \rightarrow 1} x_{21}(z) = X^+$ が成立する。

<仮定2>

$x_{12}(z)$ と $x_{22}(z)$ は z の減少関数で、 $\lim_{z \rightarrow 0} x_{12}(z) = X^-$ $\lim_{z \rightarrow 0} x_{22}(z) = X^+$ かつ $\lim_{z \rightarrow 1} x_{12}(z) = \lim_{z \rightarrow 1} x_{22}(z) = 0$ が成立する。

この仮定を満たす関数で、例えば、以下のよう $x_{ij}(z)$ を考える。

(7)

$$x_{11}(z) = X^- \left\{ 1 - \exp \left(-\hat{\alpha}_{11} \frac{z}{1-z} \right) \right\} \\ x_{12}(z) = X^- \exp \left(-\hat{\alpha}_{11} \frac{z}{1-z} \right) \\ x_{21}(z) = X^+ \left\{ 1 - \exp \left(-\hat{\alpha}_{21} \frac{z}{1-z} \right) \right\} \\ x_{22}(z) = X^+ \exp \left(-\hat{\alpha}_{21} \frac{z}{1-z} \right)$$

このとき(1)と(7)より、リコールは

$$(8) \quad R(z) = \exp \left(-\hat{\alpha}_{21} \frac{z}{1-z} \right)$$

と与えられ、これから

$$(9) \quad \frac{z}{1-z} = -\frac{1}{\hat{\alpha}_{21}} \log R(z)$$

が成立する。さらに(2)と(9)より精度は

$$(10) \quad P(z) = \left\{ 1 + \left(\frac{1}{\lambda} - 1 \right) R(z)^{\frac{\hat{\alpha}_{11}-1}{\hat{\alpha}_{21}}} \right\}^{-1}$$

となる。ここで、 $\lambda = X^+ / X^-$ である。(5)と(10)より、最適化問題は、先ずリコールと精度のみを変数として含む最適化問題を解き、その最適解から(9)を用いて(5)式を最大化する z^* を求めるという2段階に分解できることが分かる。すなわち、

<最適化問題>

$$\max f(P, R) = \frac{\Delta V(z)}{X^+} = R \left\{ \frac{\gamma^-(1-P)}{P} + \gamma^+ \right\}$$

subject to

$$G(P, R) = P \left\{ 1 + \left(\frac{1}{\lambda} - 1 \right) R^{\frac{\hat{\alpha}_{11}-1}{\hat{\alpha}_{21}}} \right\} - 1$$

$$0 \leq P, R \leq 1$$

4. 研究成果

音楽CD市場は、作品のリリース直後に売上が急上昇し、20週間前後ではほぼ0に収束するという特異な商品ライフサイクルを持つという点に特徴がある。この特異性を捉えるため、既存研究に対してマーケットセグメンテーションをより精緻化し、顧客の前段階行動を詳細に反映する新たなマルコフ・モデルの確立に成功した。マルコフ・モデルはPOSデータから推定される6つのパラメータを持ち、ほぼ完璧な形で現実データに対する適応性を実現すると同時に、パラメータの推定値そのものによって、音楽CD市場における顧客行動やブランド力を構造的に把握することを可能にした。本研究は、2005年度データ解析コンペティション社会人の部、最優秀賞(第1位)を受賞。さらに、2006年度日本オペレーションズ・リサーチ学会事例研究賞も受賞した。顧客のブラウジング行動の解析を通してeマーケティングの新たな方法論の確立に寄与する目的に対しては、以下の成果を挙げた。Webサイトのページアクセスに対しては、サ

イト訪問の日時、どの検索エンジンやキーワードを用いて訪れてきたか、閲覧ページの時系列データ等の情報が Web サーバにログデータとして蓄積される。このログデータを活用することにより、閲覧顧客のページアクセスの時間的変化をリアルタイムで捉えることが可能となり、サイト訪問の進行中に顧客の行動特性を解析し、その結果に基づいてマーケティングの観点から個別的な対応を実施する道が拓かれた。e-ビジネスにおいては、このようなリアルタイム・リコメンデーションエンジンを開発し実装することが極めて重要になってきている。EC(e-Commerce)サイトにおいて商品・サービスが繰り返し購買される場合には、閲覧顧客の購買履歴や属性データに基づいて顧客の行動特性を把握することが可能であり、リアルタイム・リコメンデーションエンジン開発の基礎となる情報源も豊富となる。それに対し、商品・サービスが一度しか購買されない場合には、蓄積された豊富なデータから閲覧顧客の行動特性を把握することが不可能となり、現在継続中のセッション履歴からのみ重要顧客であるか否かの判別を行う必要がある。このような環境におけるリアルタイム・リコメンデーションエンジンの開発は極めて困難であるが、本プロジェクトでは、この問題が克服された。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 8 件)

- ① Mizuno. M, Saji. A, Sumita. U and Suzuki. H, “ Optimal Threshold Analysis of Segmentation Methods for Identifying Target Customers,” *European Journal of Operational Research*, Vol.186, Issue 1, pp. 358-379 (2008) 査読有
- ② 住田 潮, 齊藤 晃一, 高橋 一樹, 小池雄平, 菅谷 健人, 『音楽CD市場におけるブランド力の構造分析』, *オペレーションズ・リサーチ*, 52 巻, 2 号, pp. 69-78 (2007) 査読有
- ③ Sumita. U, Ise. T and Yonezawa. K, “ Stochastic Analysis of Number of Corporations in a Market Derived from Strategic Policies of Individual Corporations for Market Entry and Retreat, ” *Journal of Operations Research Society of Japan*, Vol. 46, No. 1, pp. 1-18 (2006) 査読有

[学会発表] (計 8 件)

- ① Ushio Sumita and Rina Isogai, Development of e-Marketing Contract Structure Based on Consumer-Generated Contents and Its Optimal Strategy, The 2009 Winter Conference on Business Intelligence, U.S.A. , February 27, (2009)
- ② Ushio Sumita and Jinshui Zuo, Analysis of a Correlated Multivariate Shock Model Generated from a Renewal Sequence, International Workshop on Applied Probability, France, July 8, (2008)

6. 研究組織

(1) 研究代表者

住田 潮 (SUMITA USHIO)

筑波大学・大学院システム情報工学研究科・教授

研究者番号 : 10236044