

消費者向け電子商取引サイトにおける顧客の来店有無を考慮した購買モデル

佐藤 翔太, 朝日 弓未

近年、消費者向け電子商取引サイト運営者にとって顧客の購買予測の必要性が高まっている。本研究では顧客の来店を考慮して、顧客の購買予測モデルを構築した。分析の結果、来店を考慮したモデルのほうが来店を考慮しないモデルより予測精度が高かった。また、直近1週間で来店回数が多いほど来店しやすくなるなどの顧客の傾向も明らかになった。

キーワード：階層ベイズ、ロジットモデル、サイト来店、アクセスログデータ

1. はじめに

近年、消費者向け電子商取引（EC: Electronic Commerce）サイトの市場規模が増加している [5]。ここで、電子商取引とは「インターネット上で行われる財またはサービスの販売または購入 [6]」を指す。EC サイトの市場規模の増加に伴い、EC サイトを運営する企業の増加が予想される。その結果、EC サイト間での競争が激化していくと考えられ、多くの EC サイトでの顧客の奪い合いが起きていると考えられる。そのため、EC サイトの運営者は売上を上げるために、どの顧客が近い将来に利益をもたらす顧客となりうるかを予測し、顧客にアプローチを行う必要がある。つまり、顧客ごとの EC サイト上での行動を予測する必要性が高まっていると言える。

EC サイトに利益をもたらす顧客の指標として、顧客のサイト滞在時間や購買回数などさまざまな指標が考えられる。サイト滞在時間に着目した研究として、例えば勝又 [3] では、訪問者のコンテンツ訪問と滞在時間を推定するモデルを構築している。顧客のサイト滞在時間も重要な指標であると考えられるが、顧客の購買は EC サイトの利益に直結すると考え、本研究では顧客の購買に着目する。

EC サイト上の顧客の購買行動を分析した研究として、例えば Van den Poel and Buckinx [11] がある。この研究ではサイト訪問を購買機会ととらえ、サイト

への来店が行われたもとの購買行動を分析し、次回来店時の購買有無を予測している。また、Sismeiro and Bucklin [10] ではサイト上において購買に至るまでの過程を「個人情報の入力」などいくつかの段階に分けて顧客の購買モデルを構築している。また、Moe and Fader [8] では顧客の過去の来店と購買の傾向から顧客の購買を予測するモデルを構築している。これらの研究は来店が行われたもとの顧客の購買確率を算出しているため、来店時における顧客の購買確率を算出することになる。しかし、来店時の購買確率が算出できても、来店がいつ行われるかは不明なため、顧客がいつ購買するかを把握することは困難であると考えられる。これらの研究は顧客のサイトへの来店と購買の両方の行動を同時にとらえていないという点で課題を残している [8]。

顧客の来店と購買を同時に扱った研究として佐藤・樋口 [9] がある。佐藤・樋口 [9] では一般状態空間モデルを用いて、スーパーマーケットにおける来店と牛乳カテゴリーの購買を分析している。また、勝又・糸久 [4] では、顧客の購買の意思決定を店舗での購買と来店に分解して表現することで、詳細に顧客の行動を理解でき、より深い知見を得られると指摘している。EC サイトのアクセスログデータに文献 [9] と類似のモデルを適用することにより、顧客のサイトでの購買確率をサイトへの来店確率と購買確率に分解して推定することができる。EC サイトでは、商品を購入するために来店するスーパーマーケットなどとは異なり、店舗に来店はしたが、購買しようとは思わず何も購買しなかったという行動が起こりうる。顧客の購買と来店につながる要因を分解して明らかにすることで、顧客の購買意向、来店意向に沿ったアプローチを行うことが

さとう しょうた

東京理科大学大学院工学研究科経営工学専攻

〒162-8601 東京都新宿区神楽坂1-3

あさひ ゆみ

静岡大学工学研究科事業開発マネジメント専攻

〒432-8561 静岡県浜松市中区城北3-5-1

できる。例えば、来店意向のみある顧客ならば、来店してはくれるため衝動買いを促すアプローチなどが考えられる。そのため、顧客の購買と来店を分解してとらえるモデルは、特に EC サイトにおいては、顧客へのプロモーション施策における有用な示唆を得ることができると考えられる。

また、顧客の購買に至る行動は人それぞれ異なることと考えることで顧客ごとにアプローチ戦略を考えることができる。そのため、本研究では顧客の異質性と来店有無を考慮して、顧客の購買モデルを構築する。

本研究の目的は、顧客の異質性と来店有無を考慮した顧客の購買モデルを構築し、顧客の購買と来店の傾向を明らかにし、顧客ごとのアプローチに関する示唆を得ることである。

2. 分析モデル

2.1 購買と来店の同時確率モデル

本研究では、データ観測開始日から数えて第 t 日 ($t = 1, \dots, T$) における顧客 i ($i = 1, \dots, N$) のサイトへの来店有無と購買有無を考える。また、本研究ではサイトへの来店が行われない限り、このサイト上での購買は行われたいとする。

y_{it} を第 t 日において顧客 i が購買していたら 1、そうでなければ 0 を取る二値変数とする。 v_{it} を第 t 日において顧客 i が来店していたら 1、そうでなければ 0 を取る二値変数とする。本研究ではサイトへの来店と購買という 2 つの二項選択を考えてモデル化を行う。第 t 日において顧客 i の来店と購買の同時確率を $P(y_{it}, v_{it})$ とする。購買と来店の同時確率は、来店確率と来店が与えられたもとの購買の条件付き確率に分解できるため、(1) 式のように定式化する [9]。

$$\begin{aligned} P(y_{it}, v_{it}) &= P(y_{it}|v_{it})P(v_{it}) \\ &= (pp_{it})^{y_{it}}(1 - pp_{it})^{(1-y_{it})}(vp_{it})^{v_{it}} \\ &\quad \times (1 - vp_{it})^{(1-v_{it})} \end{aligned} \quad (1)$$

ここで pp_{it} は第 t 日に顧客 i がサイトに来店したという条件のもとで購買する確率であり、 vp_{it} は第 t 日において顧客 i がサイトに来店する確率である。なお、第 t 日に顧客 i がサイトに来店しなかった場合、このサイトでの購買は起こらないとしているため、 $v_{it} = 0$ のとき、購買が起こる確率 pp_{it} は 0 とする。尤度関数は (2) 式で表せる。

$$L = \prod_{i=1}^N \prod_{t=1}^T P(y_{it}, v_{it}) \quad (2)$$

(2) 式の尤度関数を利用してパラメータを推定する。

2.2 購買有無のモデル化

第 t 日における顧客 i の購買に対する効用を u_{it} とし、(1) 式中の購買確率 pp_{it} を (3)、(4) 式のようにロジットモデルで定式化する [9]。

$$pp_{it} = \frac{v_{it} \exp(u_{it})}{1 + \exp(u_{it})} \quad (3)$$

$$u_{it} = x'_{it}\beta_i \quad (4)$$

ここで x_{it} は顧客 i の第 t 日における購買有無に関する切片を含む説明変数であり、 β_i はパラメータベクトルである。また、 v_{it} は第 t 日において顧客 i が来店していたら 1、そうでなければ 0 を取る二値変数である。本研究では顧客の異質性を考慮するため、パラメータは顧客ごとに異なると考えている。また、本研究ではサイトへの来店が行われない限り購買は起こらないと考えている。

2.3 サイトへの来店有無のモデル化

第 t 日における顧客 i のサイトへの来店に対する効用を z_{it} とし、(1) 式中の来店確率 vp_{it} を (5)、(6) 式のようにロジットモデルで定式化する [9]。

$$vp_{it} = \frac{\exp(z_{it})}{1 + \exp(z_{it})} \quad (5)$$

$$z_{it} = w'_{it}\gamma_i \quad (6)$$

ここで w_{it} は顧客 i の第 t 日における来店有無に関する切片を含む説明変数であり、 γ_i はパラメータベクトルである。来店行動についても顧客の異質性を考慮するため、パラメータは顧客ごとに異なるとする。

2.4 パラメータの階層化

顧客の来店と購買をモデル化した佐藤・樋口 [9] では、比較的頻繁に買われる牛乳カテゴリーを対象とし、3 年以上もの長いデータ期間があるために個人ごとに十分な量のデータがある。そのため、個人ごとにモデルの推定が行われている。また、個人ごとに十分な量のデータがない場合、階層化して異質性を考慮するモデルが有用である [9] とされている。本研究では佐藤・樋口 [9] とは異なり、顧客の異質性を、パラメータを階層化することによって考慮する。パラメータベクトル β_i 、 γ_i を顧客ごとにひとまとまりのベクトルとし、 θ_i とする。つまり

表 1 説明変数の候補

変数名	説明
1 週間累積商品ページ	($t-7$) 日から ($t-1$) 日までの商品ページ閲覧回数の合計
1 週間累積来店	($t-7$) 日から ($t-1$) 日までの来店回数の合計
前日来店	($t-1$) 日に来店していれば 1, そうでなければ 0
前日商品ページ	($t-1$) 日の商品ページ閲覧回数の合計
1 週間累積金額	($t-7$) 日から ($t-1$) 日までの購買金額の合計 (万円)
1 カ月累積金額	($t-30$) 日から ($t-1$) 日までの購買金額の合計 (万円)
土日祝ダミー	t 日が土日祝日であったら 1, そうでなければ 0

$$\begin{aligned} \beta_i &= (\beta_{i1}, \dots, \beta_{ik_1})' \\ \gamma_i &= (\gamma_{i1}, \dots, \gamma_{ik_2})' \\ \theta_i &= (\beta_{i1}, \dots, \beta_{ik_1}, \gamma_{i1}, \dots, \gamma_{ik_2})' \end{aligned} \quad (7)$$

となる。ここで k_1, k_2 はそれぞれ購買と来店に関する切片を含む説明変数の数である。顧客ごとのパラメータベクトル θ_i を (8) 式のように定式化する。

$$\theta_i = \Delta' d_i + \eta_i, \eta_i \sim \text{MVN}(0, V_\theta) \quad (8)$$

ここで $\text{MVN}(0, V_\theta)$ は平均が零ベクトル、分散共分散行列 V_θ の多変量正規分布を表す。 Δ の事前分布としては多変量正規分布を、 V_θ の事前分布としては逆ウィシャート分布を設定した。 Δ, V_θ の事前分布のパラメータは十分に拡散した事前分布となるような値を設定した。また d_i は性別などの顧客 i に関する情報を表す変数である。

3. 使用データ

3.1 データ概要

2 章のモデルを用いて分析を行う。本研究で用いるデータは経営科学系研究部会連合協議会が主催する平成 23 年度データ解析コンペティションにおいて提供いただいたデータである。このデータはゴルフ用品を扱う EC サイトのデータである。期間は 2010 年 7 月 1 日から 2011 年 6 月末日までである。東日本大震災の影響を考慮し、また、直近 1 カ月の行動を説明変数として利用するため、本研究ではデータ期間のうち、2010 年 8 月 1 日から 2011 年 3 月 10 日までのデータを用いる。本研究では学習データと検証データを勝又・糸久 [4] のように時期で分ける。データ期間のうち、2010 年 8 月 1 日から 2011 年 1 月末日までを学習用データとし、2011 年 2 月 1 日から 2011 年 3 月 10 日のデータを検証用データとして用いる。分析対象は学習用データで購買を行った日が 5 日以上ある顧客の 96

表 2 購買に関わる変数と来店に関わる変数

説明変数ベクトル	変数名
購買に関わる変数 x_{it}	前日商品ページ (対数)
	土日祝ダミー
来店に関わる変数 w_{it}	1 週間累積来店 (対数)
	前日来店
	前日商品ページ (対数)
	1 週間累積金額 (土日祝ダミー)

人とする。本研究では購買機会をカレンダー上の日にちととらえているため、データを日ごとに集計して分析に用いる。

3.2 説明変数

Moe and Fader [8] では購買の履歴と来店の履歴により購買を予測している。そこで本研究では購買の履歴と来店の履歴に当たる変数として表 1 の 7 種類の変数を用意した。また、先行研究 [2] において来店回数とページ閲覧は対数をとってモデルに用いられているため、本研究でも表 1 中の「1 週間累積商品ページ」、「1 週間累積来店」、「前日商品ページ」は 1 を足して対数をとって分析に利用する。

まず購買に関する変数を考える。モデルによる分析に先立ち、表 1 の変数を説明変数、来店ごとの購買有無を目的変数としたロジスティック回帰分析を行い、変数選択を行った結果、「1 週間累積来店」、「土日祝ダミー」、「前日来店」、「前日商品ページ」の変数が選択された。購買については価格帯の高いゴルフ用品ということもあり、どの顧客もデータ期間中の購買回数は数回程度である。購買回数が少ないため、説明変数が多いと過剰適合してしまう恐れがある。Moe [7] では商品ページの閲覧と購買との関係性が指摘されているため、本研究では商品ページの閲覧に着目し、選択された 4 変数のうち「前日商品ページ」、「土日祝ダミー」の 2 つをモデルに用いる。

表 5 購買モデルにおけるパラメータ Δ の推定結果

	切片	前日商品ページ	土日祝ダミー
切片	-2.82 (-3.94, -1.69)	0.08 (-0.71, 0.82)	0.74 (-0.67, 1.96)
女性ダミー	0.07 (-1.39, 1.17)	0.19 (-0.52, 0.86)	0.41 (-0.82, 1.94)
年齢	0.01 (-0.01, 0.03)	0.00 (-0.01, 0.02)	-0.01 (-0.04, 0.01)

表 6 来店モデルにおけるパラメータ Δ の推定結果

	切片	1週間累積来店	前日商品ページ	前日来店	1週間累積金額	土日祝ダミー
切片	-2.81 (-3.85, -1.78)	<u>1.09</u> (0.44, 1.76)	0.21 (-0.34, 0.79)	-0.04 (-0.93, 0.96)	-0.34 (-0.94, 0.25)	0.81 (-0.54, 2.04)
女性ダミー	-1.03 (-2.09, -0.08)	0.09 (-0.61, 0.79)	0.10 (-0.47, 0.67)	-0.41 (-1.40, 0.733)	-0.39 (-1.03, 0.23)	<u>1.41</u> (0.09, 2.57)
年齢	0.03 (0.01, 0.05)	-0.01 (-0.02, 0.00)	0.00 (-0.01, 0.01)	0.00 (-0.02, 0.02)	0.01 (0.00, 0.02)	-0.02 (-0.04, 0.01)

本研究の購買モデルと同じ変数を利用し、顧客の異質性を考慮せず来店かつ購買を目的変数とした通常のロジットモデルを利用する。各モデルを整理したものとROC曲線下の面積を表4に示す。

表4の本研究モデル1は「土日祝ダミー」の説明変数を来店モデルに含めたモデル、本研究モデル2は「土日祝ダミー」の説明変数を来店モデルに含めないモデルである。表4より、いずれのモデルもROC曲線下の面積が0.5を超えているため、いずれのモデルもランダムな予測よりは良いと言える。表4を見ると、「土日祝ダミー」の説明変数を来店モデルに含めた本研究モデル1は本研究モデル2に比べROC曲線下の面積が大きいことがわかる。そのため、本研究モデル1のほうが本研究モデル2に比べモデルの性能が良いと言える。

続いて、本研究モデル1と比較モデル1を比較する。本研究モデル1と比較モデル1はどちらも顧客の異質性を考慮しており、また購買に対する説明変数も同一である。表4を見ると来店をモデル化している本研究モデル1は、来店をモデル化しない比較モデル1と比べROC曲線下の面積が大きいいため、モデルの性能が良いことがわかる。よって以降では、来店をモデル化し、「土日祝ダミー」の説明変数を来店モデルに含めた本研究モデル1について結果を考察する。

4.2.2 購買と来店に関する傾向

次に、購買と来店に関する傾向を観察する。階層モデルを用いることで、パラメータと顧客属性の関係性を明らかにすることができる。パラメータ Δ は各説明

変数の購買・来店への影響度合いへの性別と年齢の影響を示している。パラメータ Δ の推定値を表5、表6に示す。推定値の下の()内の数値は95%信用区間の下限値と上限値である。

表5の推定値、信用区間を見ると、購買にかかるパラメータと顧客属性の間に強い関連は見られない。

表6の「1週間累積来店」の推定値や信用区間を見ると、切片が正である。そのため、直近1週間の来店回数の合計が多いほど、顧客は来店する傾向にあると言える。直近1週間であれ来店が多いとき、顧客はそのサイトの扱っている商材に興味を持っている時期と考えられる。扱っている商材に興味を持っているために、また来店するのではないかと考えられる。

また、表6の「土日祝ダミー」を見ると、女性ダミーの推定値が正である。そのため、女性は土日祝日に来店する傾向にあると言える。

続けて顧客別パラメータについて考察を行う。分析結果としてサイトへの来店と購買に関するパラメータが得られる。まず「前日商品ページ」のパラメータに関して考察を行う。「前日商品ページ」は購買、来店に対する第 t 日の前日の商品ページの閲覧数合計の影響を示している。図1に「1週間累積商品ページ(購買)」と「1週間累積商品ページ(来店)」の顧客ごとのパラメータ推定値を示す。

「前日商品ページ(購買)」はパラメータ β_i のうち「前日商品ページ」にかかるパラメータ、「前日商品ページ(来店)」はパラメータ γ_i のうち「前日商品ページ」にかかるパラメータを示している。図1中の点は顧客

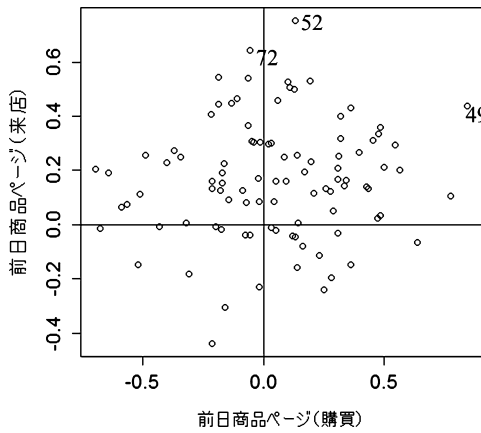


図1 「前日商品ページ」のパラメータ

を表しており、点の右下に添えてある番号は、それぞれの顧客に振った連番を示している。図1中の横軸は「前日商品ページ（購買）」のパラメータの値を示し、右側にプロットされている顧客ほど、前日に商品ページの閲覧が多い日ほどその次の日に購買する傾向のある顧客と解釈できる。図1中の縦軸は「前日商品ページ（来店）」のパラメータの値を示し、上側にプロットされている顧客ほど、前日に商品ページの閲覧が多い日ほどその次の日に来店する傾向のある顧客と解釈できる。例えば図1の右上にプロットされているNo. 49の顧客は、前日に商品ページの閲覧が多いほど来店し、購買する傾向のある顧客であると言える。図1の右上にプロットされている顧客は、商品ページの閲覧数が多い次の日に来店、購買する傾向があるため、商品ページの閲覧数が多い次の日に、購買意向を持って来店する可能性がある。

また、No. 52やNo. 72のように図1の中央上にプロットされている顧客は、商品ページの閲覧数が多い次の日に来店する傾向にあるが、購買する傾向はない顧客である。そのため、図1の中央上付近にプロットされている顧客は、商品ページの閲覧数が多い次の日は来店をしてくれる傾向はあるため、商品ページの閲覧数が多い次の日に衝動買いを促すアプローチが有効ではないかと考えられる。

次に、「土日祝ダミー」のパラメータに関して考察を行う。図2に「土日祝ダミー（購買）」と「土日祝ダミー（来店）」の顧客ごとのパラメータ推定値を示す。

「土日祝ダミー（購買）」はパラメータ β_i のうち「土日祝ダミー」にかかるパラメータ、「土日祝ダミー（来店）」はパラメータ γ_i のうち「土日祝ダミー」にかかるパラメータを示している。図2中の横軸は「土日祝

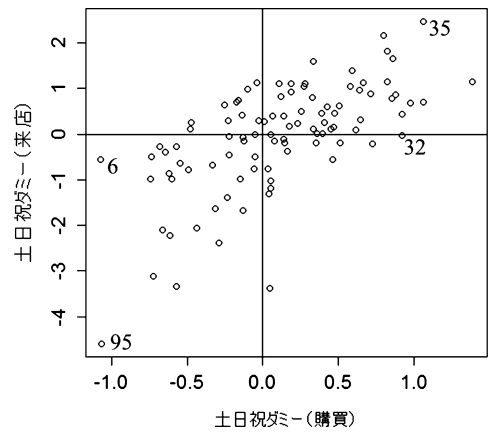


図2 「土日祝ダミー」のパラメータ

ダミー（購買）」はパラメータの値を示し、図2中の縦軸は「土日祝ダミー（来店）」のパラメータの値を示している。図2を見ると、「土日祝ダミー（購買）」のパラメータの値と「土日祝ダミー（来店）」のパラメータの値にはおおむね正の相関があることがわかる。No. 35のように図2の右上にプロットされている顧客は土日祝日に来店し購買する傾向があり、No. 95のように図2の左下にプロットされている顧客は平日に来店し購買する傾向があると解釈できる。No. 32の顧客のように、図2の右側中央付近にプロットされている顧客は、来店には土日祝や平日は関係ないが、購買は土日祝にする傾向のある顧客と解釈できる。また、No. 6の顧客のように、図2の左側中央付近にプロットされている顧客は、来店には土日祝や平日は関係ないが、購買は平日にする傾向のある顧客と解釈できる。

5. まとめ

本研究では、来店有無を考慮した顧客の購買モデルを構築し、モデルを用いた分析を行った。

まず、顧客の来店と購買を二段階に分けてモデル化した本研究のモデル、来店かつ購買を予測する一段階のモデル、通常のロジットモデルの予測精度を比較した。比較した結果、来店を考慮した本研究のモデルの方が購買を直接推定するモデルに比べ、モデルの性能が良かった。

次に、各説明変数の購買、来店への影響度合いと顧客属性との関連を分析した。分析の結果、顧客は直近1週間での来店回数が多いと来店する傾向にあることや、女性は土日祝日に来店する傾向があることがわかった。

最後にモデルを用いて顧客ごとのパラメータを推定し、顧客ごとの購買と来店につながる傾向を分析した。

分析の結果、「商品ページを多く閲覧した次の日に来店し購買する顧客」や、「商品ページを多く閲覧した次日は来店はするが、必ずしも購買はしない顧客」、「来店には土日祝や平日は関係ないが、購買は平日にする傾向のある顧客」などといった顧客の存在を確認することができた。

今後の課題としては、購買の予測精度の向上が挙げられる。本研究では得ることのできなかったゴルフのプレー頻度やプレー時期に関するデータを説明変数として利用することで、予測精度を高めることが期待できる。また、本研究ではゴルフ用品のデータを用いている。そのため、例えばファッション用品などゴルフ用品以外の EC サイトのデータを用いて、モデルの実証分析を行い、本研究で得られた結果と比較し考察を行うことも今後の課題である。

謝辞 本研究をまとめるにあたり、山口俊和先生（東京理科大学）には大変お世話になりました。この場をお借りして御礼申し上げます。

参考文献

- [1] R. C. Blattberg, B. Kim and S. A. Neslin, “Database Marketing: Analyzing and Managing Customers,” *Springer*, 2008.
- [2] R. E. Bucklin and C. Sismeiro, “A Model of Web Site Browsing Behavior Estimated on Clickstream Data,” *Journal of Marketing Research*, **40**(3), 249–267, 2003.
- [3] 勝又壮太郎, 「ウェブコンテンツ評価モデルの比較と活用—滞在時間とページビュー双方の観点から—」, マーケティングサイエンス, **18**(1), 1–27, 2010.
- [4] 勝又壮太郎, 糸久正人, 「需要要因分解モデルによる予測と多面的な顧客理解への活用」, オペレーションズ・リサーチ, **55**(2), 106–112, 2010.
- [5] 経済産業省, 「平成 21 年度我が国情報経済社会における基盤整備（電子商取引に関する市場調査）報告書」, 経済産業省, 2010.
- [6] 経済産業省, (<http://www.meti.go.jp/policy/it-policy/statistics/outlook/h22houkoku.pdf>), 最終閲覧日：2011/9/27.
- [7] W. W. Moe, “Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream,” *Journal of Consumer Psychology*, **13**(1&2), 29–39, 2003.
- [8] W. W. Moe and P. S. Fader, “Dynamic Conversion Behavior at E-Commerce Sites,” *Management Science*, **50**(3), 326–335, 2004.
- [9] 佐藤忠彦, 樋口知之, 「動的個人モデルによる購買生起行動の解析」, マーケティング・サイエンス, **16**(1・2), 49–73, 2009.
- [10] C. Sismeiro and R. E. Bucklin, “Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach,” *Journal of Marketing Research*, **41**(3), 306–323, 2004.
- [11] D. Van den Poel and W. Buckinx, “Predicting online-purchasing behaviour,” *European Journal of Operational Research*, **166**(2), 557–575, 2005.