

消費者の閲覧・購買行動の予測における多様性変数の提案 — 行動の多様性の理論的背景と他指標との検討 —

新 美 潤一郎

Both in academia and practice various methods of Customer Relationship Management (CRM) have been in use especially RFMC (Zhang et al., 2014) which is recently proposed as a novel method due to its high performance of predicting competitive purchase. However, it still has difficulty identifying consumer heterogeneity compared to variety variables proposed by NIIMI and HOSHINO (2015, 2017a). The aim of this study is to examine the theoretical background of the variety of consumers online and offline behaviors according to the existing literature. We expand RFMC with the advantage of variety variables and compare the performances between three models based on RFMC and variety variables.

Keywords: marketing, applied statistics, share-of-wallet, clickstream data, E-Commerce

I. はじめに

1. 本研究の背景

現在では消費者の趣味嗜好が多様化し、なおかつIT技術の発達によりインターネット上のECサイトやスマートフォン上のモバイルアプリを利用したオンラインショッピングが活発化している。企業にはOne-to-Oneマーケティングをはじめとして各顧客の嗜好)や状況に最適化したマーケティング戦略が求められている。

こういった現状で、企業が自社で保有する顧客を適切に管理する顧客関係管理 (CRM) において、マーケティングを顧客のロイヤルティ (自社への愛着心) の高さに従って実施するロイヤルティマーケティングが一般的に広く実施されている。Jacoby and Chestnut (1978) ではロイヤルティを行動的ロイヤルティ (behavioral loyalty) と態度的ロイヤルティ (attitudinal loyalty) に分類しており、前者は再訪問や再購買といった自社の利益に直接的に貢献する実行動に基づくもの、後者は自社ブランドへの好印象など他の消費者へのレコメンデーションにも繋がりを認知・態度としている。

このように実務・学問分野ともに消費者の企業に対するロイヤルティの高さの把握がCRMの観点から重要視されている中で、新美・星野 (2015) では企業が自社内で得られるWeb情報のみから、新美・

星野 (2017a) では自社で保持する複数の大規模データから顧客の社内外での行動を予測することで行動的ロイヤルティとしての将来的な購買回数や競合他社での行動の推定を実施している。さらに予測の際には、消費者の行動がどれほど多様であるかを多様性変数として考慮することで予測精度が向上することを示した。

2. 本研究の目的

本研究ではこのような背景において、まずマーケティング系の先行研究について網羅的なレビューを行い、先行研究に対する多様性変数の位置付けについて確認する。また、このような消費者の行動の多様性を考慮することで競合他社の購買を高い精度で予測できる点に関する理論的背景の考察を行う。また近年着目されているRFMC分析 (II節にて詳述) と比較した行動の多様性変数の有用性について議論する。そしてRFMC分析の持つ課題について議論を行った上で、多様性変数が有効となる背景に基づいてRFMC分析を拡張したモデルについても検証する。

したがって本論文の構成は、まずII節でWebデータを用いて顧客行動を予測しているマーケティング系の先行研究について俯瞰的なレビューを行う。次にIII節で多様性変数が顧客行動の説明に有用である点の理論的背景や、関連する指標との比較を行う。

そして第IV節では多様性変数や代表的なCRMの手法を用いて顧客の行動予測のシミュレーションと実データ解析を実施する。最後にV節にてまとめと今後の課題について述べる。

II. 関連する先行研究のレビュー

I. CRMの手法

本節では関連する先行研究について、一般的な顧客関係管理 (CRM) の手法、オンラインの消費者の行動の解析に用いられる Web データ、そして近年活用されている機械学習を用いた事例についてのレビューを行う。

まず企業が自社の顧客のロイヤルティを測るための代表的な指標には (1) 顧客生涯価値, (2) RFM 分析, (3) シェア・オブ・ウォレットの 3 つが存在する。

(1) 顧客生涯価値

企業における顧客のロイヤルティの把握には様々な手法が開発されているが、特に行動的ロイヤルティの把握のために実務においても積極的に用いられている代表的な手法として、顧客生涯価値 (Customer Lifetime Value, CLV, ただし LTV と表記される場合もある) がある (Blattberg et al., 2008)。CLV では顧客が自社に生涯的にもたらす利益の総和を算出することによる行動的ロイヤルティの推定に加えて、将来的な離反時期の予測 (Berger and Nasr, 1998) 等が行われる。

(2) RFM 分析

CRM の手法としては、CLV 以外にも RFM 分析 (Gupta and Lehmann, 2006) が存在する。RFM 分析では、各顧客の過去一定期間における購買行動から集計した最終購買日 (recency) ・購買頻度 (frequency) ・支出額 (monetary) を用いた直近期間の行動的ロイヤルティの高さの算出 (Gupta and Lehmann, 2006) 等が行われており、実務でも積極的に活用されている。

しかしながらこの RFM 情報を用いた顧客管理には、次に示すような問題が存在する。例として図 1 のような購買パターン I ・ II が存在した場合に、その利用パターンとしての購買発生のタイミングのばらつきが大ききから、パターン II は明らかな定期利用であるがパターン I はそうではないなど自社利用の動機が異なっている可能性がある。それでも RFM 分析の指標の上では期間中の購買回数や最終

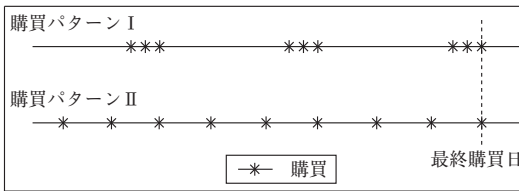
購買日が同一であることから、支出額が同一であった場合には RFM の数値上では完全に同一の行動的ロイヤルティとみなされてしまう。さらに購買パターン I では自社を利用していない空白期間が複数存在しているが、この空白期間に顧客は競合他社を利用しており、データが取れていないだけで顧客は実際には継続的に利用を行なっている可能性がある。

そこで RFM 分析において区別できないこうした購買の等間隔性を指標化した概念として、近年では RFM の各指標に新たに C 指標を加えた RFMC 分析 (Zhang et al., 2014; Platzer and Reutterer, 2016) が提案されている。C 指標は Clumpiness (購買や訪問といったイベントが期間全体のうちのある短い期間に集中して発生している事象) の有無を表す指標である。具体的には購買等のイベントを時系列に集計し、イベント発生がランダムに行われているかを検定することでその不均一性が特定の有意水準を超えた場合に Clumpiness の発生と判断する。C 指標は従来の RFM に比べて消費者の購買の異質性の考慮によって購買の予測精度を向上させるが、さらに Clumpiness が発生している場合には必然的にその前後には自社を利用していない空白期間が存在していることになることから、指標化した Clumpiness の有無を競合購買の予測に考慮することで消費者が競合他社にブランドスイッチしている可能性を適切に発見できる可能性がある。先述の Zhang (2013) では Hulu や Netflix をはじめとして近年台頭する複数の有料動画ストリーミングサービスの利用履歴の時系列データから、Clumpiness を用いて競合の利用を推定した場合に C 指標が大きな説明力を持ったことが示されている。

(3) シェア・オブ・ウォレット

一方近年では競合他社での購買を考慮した指標としてのシェア・オブ・ウォレット (SOW) の活用も盛んである。SOW は消費者ごとの特定の市場や商品カテゴリへの支出額全体に占める自社への支出額の割合として計算される。その定義から、算出することで各顧客について自社への行動的ロイヤルティが計算できる (Jones et al., 1995) ことに加えて、競合他社に支出している金額から潜在的な収益性 (potential profitability) も同時に把握できる (Gladly and Croux, 2009) という点で有用とされる。Chen and Steckel (2012) では SOW を得ることにより顧客ごとの追加的な購買余地を把握する

図 1：RFMでは区別できない購買パターンの例



ことで、非ロイヤル顧客のみに対してクロスセリングやアップセリングを用いた効率的なプロモーションを実施できるとしている。

新美・星野（2015）では実データを用いてECサイト・ニュースサイトにおけるSOWの推定を行っているが、特定の市場や商品カテゴリでの自社SOWの推定は様々な市場を対象に実施されており、自社で得られない競合他社での購買情報は欠損データ予測の文脈として統計的データ融合¹⁾の枠組みで議論されることが多い。代表的なものとして因子モデルを欠損変数の予測に応用することでデータ融合を実施したKamakura and Wedel（1997）があるが、具体的に特定業種でのデータを用いてSOWを予測したものとしては、銀行への支出に関する調査データを用いて複数銀行間での金融商品のSOWの推定を行ったDu et al.（2007）、個人ごとの支出パターンの異質性や企業間・カテゴリ間での支出の関係の同時性を考慮しながら銀行内でのカテゴリごとの支出と銀行ごとの支出の同時モデリングを行うことでSOWの推定を行ったJang et al.（2016）、調査データ等の追加的な情報を用いずに自社への支出情報としてのRFMとデモグラフィック情報のみで銀行の利用額としての預金残高のSOWを算出したGladly and Croux（2009）、複数のクレジットカード間での自社カードへの支出額に関するSOWを算出したChen and Steckel（2012）、小売企業を対象としたものとしてMagi（2003）などが存在する。また機械学習や深層学習を用いた先行研究として、欠損値の予測に用いることが可能なベイジアンネットワークで購買履歴のID-POSデータと調査パネルのデータ融合を実施した石垣他（2011）や、深層学習の代表的な手法の一つであるDeep Boltzmann Machine（Salakhutdinov and Hinton, 2009）を因子モデルの拡張として非線形の深層モデルに発展させてデータ融合に応用した新美・星野（2017b）などが存在する。

これらデータ融合系の先行研究を俯瞰した示唆は次のような 1) 消費者について得られている自社で

の支出額は競合での支出額とほとんど相関を持たないこと、2) ごく一部の消費者が競合利用の大部分を占めていること、の2点にまとめることができる。すなわち自社での購買額が小さいからといって競合を積極的に利用しているわけではなく、そもそもカテゴリへの支出額自体が小さい消費者である場合も多いことから、SOWの低い顧客を適切に発見することが自社のシェアを改善するための有効な手立てとなる。

2. インターネット閲覧履歴データの活用

ウェブ系の論文について網羅的にレビューしたBucklin et al.（2002）によれば、本研究のようにオンラインストアでの購買をはじめとしたインターネット上の消費者の行動予測を扱ったマーケティング研究の多くで、ウェブ閲覧履歴データであるClickstream Dataが活用されている。Clickstream Dataはユーザの閲覧したウェブページについてそのURL・アクセス日時・滞在秒数・リファラ情報²⁾等の情報を逐次サーバで収集したものであり、各ユーザについてウェブサイト横断的に閲覧情報を収集したUser-Centric Clickstream Dataと、特定のウェブサイトについてそこにアクセスした全ユーザのサイト内での閲覧を収集したSite-Centric Clickstream Dataに分類できる。企業が保有するECサイト等のサーバから自社データとして得られるのはSite-Centricなデータのみであり、自社サイトにアクセスした顧客について、自社サイト外でのウェブ閲覧に関する情報を得ることはできない。そこでインターネットの利用に関する人口統計に基づいて収集されたUser-Centricなデータを特定のウェブサイトへのアクセスに限定することにより、企業が得られうる擬似的なSite-Centric Clickstream Dataとすることで様々な研究に応用されている。

またSrivastava et al.（2000）ではウェブ上で収集できるデータを i) サーバで収集できるClickstream Dataなどのweb usage data, ii) 年齢性別といったデモグラフィック情報や嗜好などのuser profile data, iii) ECサイトにおける商品情報など実際のテキストや画像データなどのcontents data, iv) HTML等を用いて表されるウェブサイトの構造を扱ったstructure dataに分類しており、本研究では特にweb usage dataから得られる消費者のウェブ閲覧の特徴をアクセス・パターン情報³⁾と呼ぶ。web usage dataからのアクセス・

パターン情報の抽出と顧客に関する知見の創出は Buchner and Mulvenna (1998) をはじめとして古くから行われており、従来の実店舗では得ることの難しかった消費者の店舗内での動線や、その際にどういった商品が閲覧・検討されたのかなど、購買までの過程の行動を詳細に把握することが可能な点で実務・アカデミック共に大変有用な情報であるといえる。しかしながら先行研究におけるアクセス・パターン情報の活用には次に指摘する 2 点の特徴として、1) アクセス・パターン情報の抽出にあたっては多くの場合に量的な大きさ・平均値・最頻値が主に利用される点、2) 企業にとって貴重な自社外での消費者の行動が得られるリファラ情報の活用が限定的である点、が挙げられる。まず 1) については先の Srivastava でも言及されているように、web usage data からマーケティングに有用なパターンを発見するには、アクセスされたページ数や滞在時間等の集計に対しての記述統計量の算出、最もよく閲覧されたページなどのカテゴリカルな情報としての最頻値変数⁴⁾の計測などが行われる。しかしこの場合には新美・星野 (2015) で指摘したように、同じ頻度で自社に流入する複数の顧客について各リファラへの依存度の区別ができないなどの課題が存在する。

次に 2) については、企業のマーケティング戦略策定にあたっては、アンケートや競合他社の利用状況などを得るには追加的な情報収集コストがかかることが一般的であり、さらにそういったデータを外部から入手した場合でも、企業が自社で保有する顧客データとは収録されている対象者が異なることや仮に一部が重複していてもユーザ ID 等が異なることから紐づけて解析を行うことは困難という問題が存在する。したがってマーケティングのための顧客情報の推定にあたっては、自社で得られる情報のみから顧客行動の予測精度を向上させることに実務的にも高い重要性が存在する。

本研究と同様に Clickstream Data をマーケティング系に応用する研究は様々な目的で実施されており、主に 1) 消費者の購買行動などウェブ上でのコンバージョン、2) 特定ウェブサイト内での回遊行動、3) ウェブサイト内もしくは検索ポータルでの検索行動、4) オンライン広告のコンバージョンや広告効果の推定、の 4 分野が扱われる。まず 1) のコンバージョン予測の代表的なものとして、過去の購買行動からウェブサイトへの訪問パターンや購買

パターンを発見することで将来的な購買を予測した Park and Park (2016) や、自社 EC への次期アクセスの際の購買有無の予測にあたって Clickstream Data や購買履歴データなどを組み合わせてロジスティック回帰分析で予測した Van den Poel and Buckinx (2005) など、2) については自社ウェブサイト内でのページ遷移の経路の考慮が購買有無の予測においても有用であることを示した Montgomery et al. (2004)、消費者の EC サイトへの購買にあたって各ユーザのアクセスの目的を購買や情報収集など 4 つに分類して購買行動の予測に考慮した Moe and Fader (2004) などが挙げられる。3) については Clickstream Data を消費者の情報探索としての EC サイト内での検索・閲覧行動の把握に応用した Moe (2003) や米 Amazon でのカメラの検索行動から購買をモデル化した Kim et al. (2010) のように単一 EC サイト内での検索行動を扱ったものが存在する一方で、消費者の検索ポータルでの検索行動を depth (検索した結果として訪問するサイト数)・dynamics (depth の時系列での変化)・activity (結果として購買する商品カテゴリ数) として指標化し、検索後にアクセスする EC サイトでの購買行動との関係を明らかにした Johnson et al. (2004) のように、検索ポータルを対象として複数の EC サイトへの流入や購買を扱っているものも存在する。4) のオンライン広告を扱ったものとして、ウェブ広告への接触から購買に至るまでの期間の推定を行なった Manchanda et al. (2006) や、複数のオンライン広告出稿先での消費者の異なる広告接触傾向と広告の形式によるクリック率の変化の関係を明らかにした Nottorf (2014) 等が存在する。

3. 機械学習・深層学習のマーケティングへの応用

こういったマーケティング系の行動予測には、統計モデルを用いたものに加え近年では機械学習を用いたものも増加している。中でも深層学習は潜在変数と非線形の活性化関数を複数重ねた深層構造が特徴的であり、分布を仮定しないノンパラメトリックな推定器である (岡谷, 2015)。

機械学習を用いたマーケティングの先行研究として、Vieira (2015) は web usage data としての Clickstream Data から消費者の EC サイトでの行動データを取り出し、あるセッションでの購買有無の予測とその際に購買される商品の予測をロジスティック回帰分析、機械学習の一手法である Random

Forest (Breiman, 2001), ロジスティック回帰を深層学習的に発展させた Deep Belief Network (Bengio et al., 2007) を用いて予測しその精度を比較している。その際には商品情報を加味するため、Word2Vec³⁾を用いて web contents data としての商品情報データをベクトルに変換し特徴量として投入して解析を行っている。

マーケティング分野においては顧客データが完全に得られることはほとんどないという点で、完全データが得られることが一般的である画像認識等とは大きく異なる点が多い。そのため完全データの使用が前提とされる深層学習が活用されることは少なかったが、近年では Lopes and Ribeiro (2012) で一般的なニューラルネットワークを欠損値を扱えるよう拡張した Neural Selective Input Model (NSIM) が提案されるなど徐々にマーケティングモデルに適した手法も登場している。先述の新美・星野 (2017b) は NSIM を参考に DBM で欠損データを扱えるよう改良したモデルである。

こういった深層学習では推定器の柔軟さから比較的予測力が高まる傾向にある。しかし一方で非線形の活性化関数と何層にもわたる潜在変数を持った複雑な構造であることから、特に本研究のような社会科学データへの応用においては実際の変数間の関係や構造を捉えているとは考えにくい。そのため一般の統計モデルが説明変数と目的変数の構造的な関係を把握することを目的に使用される一方で、深層学習は予測精度が高まることを目的に活用されることが一般的である。

また深層学習においては特徴選択もひとつの課題として考えられている。特徴選択とは与えられた説明変数を様々組み合わせることにより目的変数を最もよく説明できる潜在変数を自動的に作成する手法であり、多くの深層学習モデルに組み込まれている (岡谷, 2015)。深層学習ではこの特徴選択が利用できることから変数選択が不要とされるが、実際には新美・星野 (2017a) において深層学習器に本研究でも扱う多様性変数を投入することで行動予測の精度が向上することが示されており、特徴選択でも未だあらゆる変数を加味できるわけではないといえる。

4. 先行研究と行動の多様性

このようにマーケティング系の行動予測を扱った先行研究には様々なものが存在しているが、これらを網羅的に調査した結果として、新美・星野

(2015, 2017a) で扱っているような消費者の行動の多様性と行動予測の関係について着目しているような事例は存在していなかった。こういった点からも消費者の行動の幅広さに着目してその多様性を指標化すること自体の新規性に加え、それを用いることで行動予測の精度が向上する点についても十分な有用性があると考えられる。

III. 消費者の行動の多様性

1. Web 上での行動の多様性が有用な背景

消費者の行動の多様性に関しては、web 上の行動の多様性と web に限らない実生活の多様性までを考慮している。II 節でも述べたように多様性に類似した概念について着目している先行研究が非常に限定的であることから、多様性変数が消費者の行動に関する説明力を持っている一方で、その有用性に関する構造は明らかになっていない。そこで本節では、一般消費者の Web 利用について調査している文献から多様性変数が有用となる理論的背景について考察する。

まず考えられる理由として、多様性変数を用いることで 1) ウェブ利用における消費者の異質性を部分的に把握できるため、2) ウェブ利用の経験や習熟度を反映しているため、の 2 点が挙げられる。1) に関しては、Dembczynski et al. (2008) においてそもそも Web 上での消費者の閲覧行動には大きな個人差が存在する可能性が指摘されている。星野 (2009) では、マーケティングの実証研究で用いられている web usage data や購買履歴データはいずれも結果のみを記録した行動データであり、その行動がどういった嗜好や属性に起因して発生しているかは不明であることから、Clickstream Data に関しても個人の異質性が発生しているとしている。新美・星野 (2015) では多様性変数を用いない従来の変数では Clickstream Data 上に捉えられない Web 閲覧パターンの差異が存在することについて指摘しているが、アクセス・パターン情報の多様性を用いることでこういった個人の Web 利用の異質性を部分的に捉えていることが予測精度の改善に繋がっていると考えられる。

2) については Johnson et al. (2004) において、書籍や CD といったコモディティ類似型商品 (commodity-like products) では Web 利用の習熟度が高く EC サイトの利用経験が多い消費者ほど購買に

利用するウェブサイト数が減少し、特定サイトに集中することが示されている。Web 利用の習熟度が高まることで習慣的に特定のサイトで購買を済ませやすいという傾向から、よく訪れるウェブサイトはブラウザのお気に入り登録して簡単にアクセスできるようにするといった行動に起因して Web 行動の多様性に変化（特にリファラの多様性の減少や特定時間帯への集中等）が発生していると考えられる。Web の習熟度が効果を持つ理由として、商品購買の意思決定のために必要な情報の探索コストが異なってくる点がある。というのも実店舗で競合する複数の店舗に赴いて商品を比較検討する必要があるのと同様に、オンラインショッピングにおいても複数店舗を Web 上で閲覧し商品の品質や価格について限られた情報と選択肢の中で最適な意思決定を行わなければならない。そこで Web への習熟度の低い消費者では意思決定のための情報探索コストが高まることから様々なウェブサイトを経由して自社サイトに訪問することが考えられる。これは経済学におけるサーチ理論（今井他, 2007）等から考えても自然な構造である。

2. 実行動の多様性が有用な背景

次に新美・星野 (2017a) で扱っている、実行動における多様性の有効性についてその理論的背景を考察する。理由としては 1) 消費者の異質性の一部を捉えられるため、2) デモグラフィック情報の一部を捉えられるための2点が考えられる。1) については、特に実店舗での購買においては行動の結果としての購買履歴データしか得られないことが一般的であり、その購買プロセスとしての比較検討の過程を得ることは困難であった。しかしながら新美・星野 (2017a) で扱った小売ブランドでは、仮に実店舗で購買を行わずとも訪問した際にモバイルアプリから「チェックイン」を行うことでポイントが得られるインセンティブを与えたり、店内に置かれた商品についてモバイルアプリから詳細な情報を得られたりといった形で、実購買を伴わない訪問についても情報を得られる仕組みが整備されている。こういった購買プロセス中の消費者ごとに異なる行動パターンを多様性で捉えることで、消費者の異質性を部分的に捉えている可能性が考えられる。2) については、消費者の職業などの属性や店舗の扱う商品カテゴリによっては、平日の日中に実店舗へ訪問し購買することは難しいなど、年齢層や職業といったデモ

グラフィック情報とアクセスできる曜日や時間帯の多様性が関係している可能性がある。特に新美・星野 (2017a) においては深層学習を用いているため、実際のデモグラフィック情報は投入していないものの隠れ層においては潜在変数としての特徴量で擬似的なデモグラフィック情報が得られている可能性がある。

3. 多様性変数と RFMC の関係

競合他社での行動を考慮した解析を行うにあたって、Park and Fader (2004) では競合するウェブサイトでは訪問期間が互いに影響し合う点について指摘しており、これは複数サイトにおける価格や商品の比較検討が目的とされている。II 節でも述べたように、ある企業での利用が空白になっている期間には競合する他企業の利用を行なっている可能性があることから、自社の利用間隔の不均一性を示すことのできる RFMC では競合他社での利用についても予測が可能であると考えられる。ここでは多様性変数が有効となる主要因が消費者の異質性の判別能力の高さであることを確認するために、近年着目されている RFMC における Clumpiness の有無を表す C 指標の課題について議論した上で、C 指標における消費者の異質性の判別能力を高めるよう拡張することで予測精度が向上する可能性について検討する。

そもそも RFMC の C 指標は、ある調査期間における潜在的な購買機会（日々や週次で基準化された調査期間）において、ある顧客 i の実購買の時期から算出される購買間隔のばらつきの指標 H_i と、ランダムに同様の回数の購買が行われた場合のシミュレーションから計算できる一般的な購買間隔のばらつきの上側 $\alpha\%$ の臨界値 H_0 を比較した場合に、当該顧客の C 指標は

$$C_i = \begin{cases} 1(H_i \geq H_0) \\ 0(\text{otherwise}) \end{cases}$$

と定義される。これは消費者 i の購買行動が $\alpha\%$ の統計的有意差をもって非ランダムであることを示していることから、短期間のまとまったイベント発生 (Clumpiness) と結論づけられる。

II 節で述べたように、先行研究においては確かにこの手法で競合の利用を高い精度で見出せるという示唆が得られている。しかしながらこの C 指標を用いてもなお購買パターンの異質性は考慮しきれていない。C 指標の定義について中山 (2016) では

「(イベントが) 均等な間隔に従わない度合い」とされているものの、式(1)からもわかるように実際にはC指標は離散値であり、仮に期間中に複数のClumpinessが存在したり極端に偏った時期に購買が行われたりする場合(いずれも H_i の値が大きく上昇)にも通常のClumpinessの発生と同様に $C_i = 1$ と判定されてしまう。購買の間隔の不均一性が強まるほど自社での購買回数に対しての自社を利用していない継続的な空白期間が長くなることから、購買の不均一性についても他のRFMの3指標と同様に連続値で比較できることが望ましい。しかし中山(2016)でも指摘されている通り、C指標の算出のための H_i を複数のサンプル同士で比較する場合は、サンプル同士で潜在的な購買機会と実際の購買回数が同一である必要がある。特定の期間における複数の顧客に対して解析を行う場合には潜在的な購買機会が異なることは考えにくい、実際の購買回数についてはサンプル間で同一になることは少なく、同一条件のみでの比較では解析に用いるにあたって不便である。そこで検定時に算出した臨界値 H_0 に対する H_i の比率として新たな指標であるC比率を

$$C_i^* = \frac{H_i}{H_0} \quad (1)$$

と定義する。C比率では同一条件下の「不均一さの基準」としての臨界値と実際のばらつきの大さの比をとっているため、平均的な基準から比較してどれほど購買が不均一かについてその度合いを連続値で定義することが可能である。この場合にはC指標と比較して明らかに購買パターンの異質性が詳細に把握できることから、多様性変数を用いる場合と同様に通常のRFMCに比べて購買予測の精度が向上すると考えられる。そこで本研究では通常のRFMCや提案する多様性変数に加えて、このC比率を用いたモデルについても比較を行う。

IV. シミュレーションと実データ解析

1. シミュレーションによるC指標とC比率の比較

ここで擬似的に作成した購買データを用いたシミュレーションと、実データを用いた解析を行う。本研究ではアクセス履歴データを用いることを前提としているため、RFM分析における支出額としてのM指標は考慮しない。

多様性変数の有用性について実データで確認する

前に、ここでC指標とC比率の有効性について簡易的に作成した消費者の疑似購買データを用いて、シミュレーションを行う。シミュレーションで多様性変数を用いないのは、購買行動と紐づいた疑似的なweb usage dataの作成が困難であるためである。

ここでは疑似データを作成し、それをRFMC情報のみで予測することを考える。その際に説明変数を変更した2つのモデルを作成しており、モデル1(通常のRFMC情報を用いた場合)とモデル2(RFMCでC指標の代わりにC比率を用いた場合)それぞれでの情報量基準を比較することでその妥当性について検証する。

まず市場としてECサイトA社と競合B社を仮定し、潜在的な購買機会 N におけるある消費者 i の市場全体での購買回数(トータルウォレット)を非負の整数 Y_i で定義した。 Y_i はポアソン回帰関数を用いた3因子 $f_{ij}(j=1, 2, 3, i=1, \dots, 3000)$ に従う線形モデルとして、

$$P(Y_i = k | \beta) = \lambda^k \frac{e^{-\lambda}}{k!} \quad (2)$$

$$\lambda = \beta_0 + \sum_{j=1}^3 \beta_j f_{ij} + \epsilon_i \quad (3)$$

で作成した。消費者 i の t 回目の購買(ただし $t=1, \dots, Y_i$)の発生時期 T_{it} (ただし $T_{it} \leq N$)を期間 N に対する一様分布で設定することで、期間中のどのタイミングでも同様の確率で購買が発生することを仮定している。また購買 t での利用企業 q_{it} については、購買にA社を利用した場合に1、競合の場合に0をとる離散値として定義し、その確率にはベルヌーイ分布を用いて、

$$P(q_{it} = 1) = \begin{cases} 0.5(t=1) \\ 0.4+0.2q_{it}(\text{otherwise}) \end{cases} \quad (4)$$

とした。これは初回購買($t=1$)は完全にランダムに決定されるものの、2回目以降の購買では前回利用した企業を再度利用する確率が微増するモデルであり、ブランドスイッチにより購買が連続しやすいことからClumpinessが発生しやすくなることを目的としている。

作成したシミュレーションデータから各消費者についてのA社でのR指標、F指標、C指標、C比率を算出し、競合B社での購買回数についてポアソン重回帰分析を行なった結果を表1に示す。ここでは作成した2つのモデルの適合度について情報量規準AIC・BICで比較した結果を用いており、競合での

表 1：シミュレーション結果

	モデル 1	モデル 2
AIC	9298.15	9297.89
BIC	9327.55	9327.29

購買回数の予測ではC比率を用いたモデル2の場合でモデルの適合度が向上したことがわかる。

2. 実データ解析によるC指標、C比率、多様性変数の比較

次に実データを用いた解析を行なう。使用したデータは株式会社ビデオリサーチインタラクティブより提供いただいた web usage data としての Web Report で、これはインターネット利用の人口統計に基づいて収集された大規模なウェブ閲覧履歴のパネルデータである。データ内から2012年1月の1ヶ月間の通信販売の大手2社（A社・R社）へのアクセスデータのみを抽出したところ、期間中1度でもA社もしくはR社にアクセスしたユーザは5677人であった。企業が自社データを用いて競合他社での購買状況を予測することを想定し、自社企業をA社、競合をR社として、A社で1度でも購買している顧客に対してR社での購買を予測するモデルを作成したところ、対象者は450人となった。

解析にあたっては該当期間のA社での閲覧履歴からRFMC情報、C比率、多様性変数を計測し、シミュレーションと同様に説明変数のみを差し替えた複数のモデルを作成し、各モデルを用いて競合R社での同じ期間の購買回数についてポアソン重回

帰分析を用いて予測する。その結果から情報量基準を用いて変数の妥当性について議論する。A社・R社での購買情報やその他の準備した変数について、モデル1（一般的なR、F、C指標を使用）、モデル2（一般的なR、F指標とC比率を使用）、モデル3（多様性変数を使用）の3モデルを作成した結果の記述統計量について表2に示す。（ただしA社顧客にはR社では購買していないサンプルも存在しているため、R社に関する変数の一部の統計量は記載していない。）

C指標が{0, 1}の離散値である一方で、C比率では最大値が2.97となっており、基準値と比較しておよそ3倍程度の不均一性が存在していることがわかる。C比率が1以上のサンプルについてはC指標では全て1と判定されてしまうことから、C指標を用いることで従来のRFMCと比較して高い精度で競合購買を予測できると考えられる。さらに多様性変数については、RFMCのような購買に関わる情報を一切用いていないものの、アクセス・パターン情報として消費者のウェブ閲覧パターンの異質性について詳細に把握できることからさらに高い精度で顧客行動の予測が可能であることが予想される。

実データ解析の結果について表3に示す。まずはC比率を用いて拡張したRFMCを用いた場合には、シミュレーションと同様に通常のRFMCに比べて情報量基準を改善させることができた。また特筆すべき点として、R、F、C指標という自社での購買に関する情報を一切説明変数に用いずにWebの閲覧行動としての流入経路や閲覧の時間帯・曜日の多

表 2：実データ解析の記述統計量と使用した変数一覧

変数名	投入モデル	平均値	標準偏差	最小値	最大値
A社閲覧ページ数	—	122.44	162.62	4.00	1412.00
R社閲覧ページ数	—	284.46	775.40	0.00	8013.00
A社購買回数（F指標）	1, 2	1.44	0.90	1.00	8.00
R社購買回数	目的変数	0.20	0.56	0.00	4.00
A社最終購買（R指標）	1, 2	14.50	8.70	1.00	30.00
R社最終購買	—	11.11	8.40	—	29.00
C指標	1	0.56	0.50	0.00	1.00
C比率	2	1.23	0.69	0.40	2.97
リファラHHI	3	0.66	0.25	0.10	1.00
時間帯HHI	3	0.27	0.22	0.05	1.00
曜日HHI	3	0.36	0.23	0.15	1.00

様さについてのみ考慮したモデル3において最良の結果が得られている。

3. 結果の考察

まず近年着目されているRFMCについては、消費者の行動の多様性の概念を応用して連続値に拡張することでシミュレーション、実データ解析ともにその予測精度を高めることができた。これは多様性変数の有用性でも触れた消費者の行動の異質性を詳細に考慮できたことが要因であると考えられる。しかしながら、この拡張したRFMCと比較しても、なお多様性変数のみを用いた場合の方が競合他社での購買の予測精度を高めることができた。RFMCと比較した多様性変数の有用性に関しては、RFMCに比べても詳細に消費者の異質性を考慮できる点に加えて、RFMC情報があくまで購買履歴という売上の結果データを集計したものに過ぎない一方で、多様性変数はその多くがアクセス履歴データに由来するものであり、購買に至る前に商品を検討する過程を考慮できる点に起因すると考えられる。

表3：実データ解析の結果

	モデル1	モデル2	モデル3
AIC	466.68	465.10	458.78
BIC	486.86	485.29	478.97

V. まとめ

本論文では新美・星野（2015, 2017a）にて提案してきた企業の持つ顧客の行動の多様性について、先行研究と比較した際のその理論的背景について考察した。そこでは多様性変数の有用性の構造としての消費者の異質性について詳細に考慮できる点や、Web利用の習熟度との関係、デモグラフィック情報と相関している可能性等について検討した。

また近年積極的に用いられているRFMCのC指標に着目しその課題である離散変数としての定義について、多様性変数に倣って消費者の異質性を判別できるよう連続変数として拡張することで、シミュレーション・実データ解析ともに通常のRFMCと比較してモデルの適合度が改善されたことから、消費者の異質性の考慮により競合他社の購買を高い精度で予測できる点についても示すことができた。それに加えて実データ解析ではRFMCと比較して多様性変数が高い説明力を持ったことから、学問研究

に限らず実務的にも幅広く活用できる指標であるといえる。

今後の課題として、まず本研究ではデータの都合上アクセス履歴データのみを用いての解析を想定しているため、RFMC分析において支出額を加味することができなかった。これに関しては、Clickstream DataとID-POSが共通のIDで紐付けされているようなデータセットを用いることで考慮できると考えられる。またRFMCにおけるC指標の問題は今回扱ったもの以外にもまだ存在している。Clumpinessの発生有無の検定においては、潜在的な購買機会と実際の購買回数が一致した場合など、極めて高頻度に購買が行われている場合には、購買の回数が均一にも関わらずClumpinessの発生確率が大きく上昇してしまう。これは基準化されている潜在的な購買機会の設定について、その基準化の方法の恣意性に加えてClumpinessの算出方法の問題でもあると考えられる。これに関しては調査期間内での購買タイミングの偏りについて、多様性変数の算出のために用いているHHIなどの指標を応用するなどすることで柔軟な連続値として定義できる可能性がある。

注

- 1) statistical data fusion. 経済学分野においてはdata combinationと呼ばれる場合もある(Ridder and Moffitt, 2007)。
- 2) あるウェブページを閲覧した際にその直前に閲覧していたWebページのURL情報であり、ウェブサイトへの流入経路等の判別に活用されることが多い。
- 3) Srivastava et al. (2000); Cooley et al. (1997)等におけるaccess patternsの著者訳。
- 4) 本研究では消費者のweb usage dataから計測されたアクセス・パターン情報のうちカテゴリカルな変数としての「最もよく購買された商品カテゴリ」「最もよく用いられた流入経路」等を最頻値変数と呼ぶ。
- 5) 自然言語の文章から単語同士の関係をベクトル表現に変換することで、深層学習等の識別器に特徴量として投入するためのアルゴリズム (Mikolov et al., 2013)。

引用文献

- 石垣司・竹中毅・本村陽一（2011）「日常購買行動に関する大規模データの融合による顧客行動予測システム実サービス支援のためのカテゴリマイニング技術」『人工知能学会論文誌』, 26(6), 670-681頁。
- 今井亮一・工藤教孝・佐々木勝・清水崇（2007）『サーチ理論—分権的取引の経済学』, 東京大学出版会。
- 岡谷貴之（2015）『深層学習』, 講談社。

- 中山雄司 (2016) 「顧客関係管理研究の新動向: 来店/購買間隔の不均一性を測るクラシビネス指標」『甲南経営研究』, 57(2), 161-181頁。
- 新美潤一郎・星野崇宏 (2015) 「ユーザ別アクセス・パターン情報の多様性を用いた顧客行動の予測とモデリング」『応用統計学』, 44(3), 121-143頁。
- 新美潤一郎・星野崇宏 (2017a) 「顧客行動の多様性変数を利用した購買行動の予測」『人工知能学会論文誌』, 32(2), B-G63, 1-9頁。
- 新美潤一郎・星野崇宏 (2017b) 「Deep Boltzmann Machine を用いたデータ融合手法の提案」, 2017年度人工知能学会全国大会 (於ウインクあいち)。
- 星野崇宏 (2009) 『調査観察データの統計科学: 因果推論・選択バイアス・データ融合』, 岩波書店。
- Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. (2007), "Greedy layerwise training of deep networks," *Advances in neural information processing systems*, Vol.19, pp.153-160.
- Berger, P. D., and Nasr, N. I. (1998), "Customer lifetime value: Marketing models and applications," *Journal of interactive marketing*, Vol.12, No.1, pp.17-30.
- Blattberg, R. C., Kim, B. D., and Neslin, S. A. (2008), *Database Marketing: Analyzing and Managing Customers*, Springer.
- Breiman, L. (2001), "Random forests," *Machine learning*, Vol.45, No.1, pp.5-32.
- Buchner, A. G., and Mulvenna, M. D. (1998), "Discovering internet marketing intelligence through online analytical web usage mining," *ACM Sigmod Record*, Vol.27, No.4, pp.54-61.
- Bucklin, R.E. Lattin, J. M., Ansari, A., Gupta, S., Bell, D., Coupey, E., Little, J. D. C., Mela, C., Montgomery, A., and Steckel, J. (2002), "Choice and the Internet: From clickstream to research stream," *Marketing Letters*, Vol.13, No.3, pp.245-258.
- Chen, Y., and Steckel, J. H. (2012), "Modeling credit card share of wallet: Solving the incomplete information problem," *Journal of Marketing Research*, Vol.49, No.5, pp.655-669.
- Cooley, R., Mobasher, B., and Srivastava, J. (1997), "Web mining: Information and pattern discovery on the world wide web," in *Tools with Artificial Intelligence, 1997. Proceedings., Ninth IEEE International Conference on*, pp.558-567, IEEE.
- Dembczynski, K., Kotlowski, W., and Weiss, D. (2008), "Predicting ads clickthrough rate with decision rules," in *Workshop on targeting and ranking in online advertising*, Vol.2008.
- Du, R. Y., Kamakura, W. A., and Mela, C. F. (2007), "Size and share of customer wallet," *Journal of Marketing*, Vol.71, No.2, pp.94-113.
- Gladly, N., and Croux, C. (2009), "Predicting customer wallet without survey data," *Journal of Service Research*, Vol.11, No.3, pp.219-231.
- Gupta, S., & Lehmann, D. R. (2006), "Customer lifetime value and firm valuation," *Journal of Relationship Marketing*, Vol.5, No.2-3, pp.87-110.
- Berkowitz, E. N. (1978), *Brand loyalty: Measurement and management*, John Wiley & Sons Incorporated.
- Jang, S., Prasad, A., and Ratchford, B. T. (2016), "Consumer spending patterns across firms and categories: Application to the size-and share-of-wallet," *International Journal of Research in Marketing*, Vol.33, No.1, pp.123-139.
- Johnson, E. J., Moe, W. W., Fader, P. S., Bellman, S., and Lohse, G. L. (2004), "On the depth and dynamics of online search behavior," *Management science*, Vol.50, No.3, pp.299-308.
- Jones, T. O., and Sasser, W. E. (1995), "Why satisfied customers defect," *Harvard business review*, Vol.73, No.6, p.88.
- Kamakura, W. A., and Wedel, M. (1997), "Statistical data fusion for cross-tabulation," *Journal of Marketing Research*, pp.485-498.
- Kim, J. B., Albuquerque, P., and Bronnenberg, B. J. (2010), "Online demand under limited consumer search," *Marketing science*, Vol.29, No.6, pp.1001-1023.
- Lopes, N., & Ribeiro, B. (2012), "Handling missing values via a neural selective input model," *Neural Network World*, Vol.22, No.4, p.357.
- Magi, A. W. (2003), "Share of wallet in retailing: the effects of customer satisfaction, loyalty cards and shopper characteristics," *Journal of Retailing*, Vol.79, No.2, pp.97-106.
- Manchanda, P., Dube, J. P., Goh, K. Y., and Chintagunta, P. K. (2006), "The effect of banner advertising on internet purchasing," *Journal of Marketing Research*, Vol.43, No.1, pp.98-108.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013), "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, pp.3111-3119.
- Moe, W. W. (2003), "Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream," *Journal of consumer psychology*, Vol.13, No.1, pp.29-39.
- Moe, W. W., and Fader, P. S. (2004), "Dynamic conversion behavior at e-commerce sites," *Management Science*, Vol.50, No.3, pp.326-335.
- Montgomery, A. L., Li, S., Srinivasan, K., and Liechty, J. C. (2004), "Modeling online browsing and path analysis using clickstream data,"

- Marketing science*, Vol.23, No.4, pp.579-595.
- Nottorf, F. (2014), "Modeling the clickstream across multiple online advertising channels using a binary logit with Bayesian mixture of normals," *Electronic Commerce Research and Applications*, Vol.13, No.1, pp.45-55.
- Park, C. H., and Park, Y. H. (2016), "Investigating purchase conversion by uncovering online visit patterns," *Marketing Science*, Vol.35, No.6, pp. 894-914.
- Park, S., and Lee, D. (2017), "An Empirical Study on Consumer Online Shopping Channel Choice Behavior in Omni-channel Environment," *Telematics and Informatics*.
- Park, Y. H., and Fader, P. S. (2004), "Modeling browsing behavior at multiple websites," *Marketing Science*, Vol.23, No.3, pp.280-303.
- Platzer, M., and Reutterer, T. (2016), "Ticking away the moments: Timing regularity helps to better predict customer activity," *Marketing Science*, Vol.35, No.5, pp.779-799.
- Van den Poel, D., and Buckinx, W. (2005), "Predicting online-purchasing behaviour," *European Journal of Operational Research*, Vol.166, No.2, pp.557-575.
- Salakhutdinov, R., and Hinton, G. (2009), "Deep boltzmann machines," in *Artificial Intelligence and Statistics*, pp.448-455.
- Srivastava, J., Cooley, R., Deshpande, M., and Tan, P. N. (2000), "Web usage mining: Discovery and applications of usage patterns from web data," *Acm Sigkdd Explorations Newsletter*, Vol.1, No.2, pp.12-23.
- Verhoef, P. C., Kannan, P. K., and Inman, J. J. (2015), "From multi-channel retailing to omni-channel retailing: introduction to the special issue on multi-channel retailing," *Journal of retailing*, Vol.91, No.2, pp.174-181.
- Vieira, A. (2015), "Predicting online user behaviour using deep learning algorithms," *CoRR*, Vol. abs/1511.06247.
- Zhang, Y. (2013), "New clumpiness measures and their application in customer evaluation," *Journal of Applied Statistics*, Vol.40, No.11, pp.2533-2548.
- Zhang, Y., Bradlow, E. T., & Small, D. S. (2014), "Predicting customer value using clumpiness: From RFM to RFMC," *Marketing Science*, Vol. 34, No.2, pp.195-208.

(名古屋大学大学院経済学研究科博士後期課程)