

別紙 4

報告番号	※	第	号
------	---	---	---

主 論 文 の 要 旨

論文題目 顧客の行動の多様性を用いた行動予測手法の提案 -多様性変数を活用した消費者の閲覧・購買の予測と精度の検証-

氏 名 新美 潤一郎

論 文 内 容 の 要 旨

(1) 本論文の目的

本論文の目的は、企業がマーケティング等の戦略策定を行うにあたって、企業が保有している顧客の将来の行動や競合他社の利用状況など、通常の企業活動の中では得ることの難しい顧客情報の予測にあたりその精度を改善することにある。インターネット等ICTの発展やスマートフォン等デバイスの普及に起因するライフスタイルの変化により、eコマース（電子商取引）の市場は年々拡大している。また消費者の嗜好の多様化に伴い、より個々人の嗜好に最適化したOne-to-Oneマーケティングなどの戦略が必要とされてきている。こういった背景から、顧客情報の予測はますます重要度を高めている。

一方で先行研究や実務で用いられるマーケティング分析では捉えられない行動の異質性が存在している。ビッグデータ活用の潮流に伴い様々な企業内の大規模データの活用の流れになっているが、安易な大量データの活用は予測モデルの大規模化による解析の所要時間の増大に繋がり、迅速な意思決定の障壁となりかねない。先行研究で捉えられていない異質性を捉えつつ、少ない変数で効率的に予測を行うことが必要である。さらに近年では機械学習、とりわけ深層学習が実務でも浸透し始めたが、解析はブラックボックス化しやすく、最大の特徴である特徴選択についても、社会科学データにおいてどれほど複雑な変数が自動で生成できるのかは議論されていない。

そこで本論文では企業が保有する顧客データ（顧客の購買履歴データ・インターネットのアクセス履歴データ・調査データ等）から行動や嗜好の「多様性」を指標化し、分析に加味することで顧客行動の予測精度を向上させることができる可能性について検証している。さらにその予測に用いるモデルについても、一般的な統計解析で用いられる一般化線形モデルだけでなく、近年企業で積極的に活用されている深層学習の代表的な手法を用いた解析例についても示し、マーケティングデータなどの社会科学

データにおける特徴選択の精度を含め、多様性変数の有用性について企業から提供された様々な実データを用いて網羅的に検証する。

(2) 本論文の構成と内容

本論文は5つの章で構成されている。まず本研究では主にオンラインショッピングのウェブサイトを主な対象として消費者の行動の解析を行うが、第1章「はじめに」では、研究の背景としてマーケティング・サイエンス分野において統計的に消費者の行動の予測を行うことの重要性について述べている。具体的には、企業は消費者について自社内の過去の行動（閲覧・購買等）は各種データを収集することで容易に把握でき、すなわち自社に関しては将来の行動についても過去のパターンからある程度の精度では予測することが可能である。その一方で、自社外のデータ（競合他社の利用頻度や購買額など）については把握することが難しい。自社外での行動はデータが得られにくく予測も困難であるが、これらの情報を把握することは顧客ごとの自社での支出額の割合（シェア・オブ・ウォレット）を把握しマーケティング戦略を最適化するにあたって非常に重要な情報である。

また深層学習活用にあたってのマーケティング実データを用いた包括的な検証が必要となる背景について詳述している。これは社会的にデータ解析の需要が高まりビッグデータと深層学習の活用の流れが加速している一方で、解析はブラックボックス化しがちであり、社会科学データにおける統計モデリングと比較した深層学習の有用性が現状では不明瞭なためである。

次に第2章「先行研究と多様性の理論的背景の考察」では、まず関連するマーケティング・サイエンス系の先行研究について、一般的に用いられる顧客関係管理（CRM）の手法、ウェブサイトを対象とした消費者の行動予測、機械学習手法のマーケティング分野への応用の3点に分けた網羅的なレビューを実施することで本研究の位置付けについて確認している。

次に本研究で提案する指標である多様性変数を分析に用いることで想定される優位性について述べている。既存研究や実際の企業での解析において、従来の予測モデルでは「どれほど多くのカテゴリーの商品を購入しているか」「どれほど多くのウェブサイトから自社にアクセスしているか」等の量的変数をそのまま投入したり、「特定のウェブサイトAから企業サイトに流入した際に購買確率が上昇する」等データをカテゴリカルに集計した質的変数をそのまま投入したりといった方法が一般に用いられている。しかしながら、多様化する消費者の嗜好を適切に把握し解析するためには、これら情報を組み合わせ「どれほど多くのカテゴリーの商品を購入しながら、どれほど特定のカテゴリに依存しているか」を指標化することが重要となる。したがって市場の集中度を示すハフィンダール・ハーシュマン指数をはじめとした既存の経済指標から要素ごとのシェアと依存度を指標化できるようなものを複数検討し、これらを応用して多様性変数を定義した。

この多様性変数が有用となりうる構造には消費者の行動の異質性を詳細に捉えられる点がある。近年ではCRMの新たな手法として顧客の購買履歴情報を用いたRFMC分析が提案されているが、特にRFMCのC指標は離散値として定義されているという点で識別性が低いという問題がある。そこでRFMC分析を識別性を高める形で拡張させ、さらに多様性変数と予測精度を比較することにより、多様性変数が購買情報を用いていないにも関わらず高い精度で顧客行動を予測できることを示す。

第3章「Web 行動の多様性の検討」では、サイト横断的に取得された大規模なアクセス履歴データを用いて、日本の代表的なニュースサイトとECサイトを対象として、擬似的に設定した自社企業とその競合他社における消費者の訪問期間・購買回数の予測を一般化線形モデル（GLM）で予測している。その際には消費者のウェブ上での行動を記録したアクセス・パターン情報についての多様性を算出しモデルに投入しているが、その中でも「様々なウェブサイトから自社サイトへアクセスしているような消費者は、同様の経路で競合他社にも積極的にアクセスし利用している」という仮説の元に特にリファラ情報に着目している。リファラ情報とは、あるウェブページを閲覧した際に直前に閲覧していたページのURLの記録であり、企業が自社顧客についてアクセス履歴データから得られる数少ない自社外の情報でもある。解析の結果から、一般化線形モデルでの顧客の将来・累積の訪問・購買の頻度について、多様性変数を投入したモデルと従来一般的に用いられている変数のみのモデルにて予測を行った結果、全ての場合において多様性変数を投入した交差検証モデルで最も予測精度が向上する結果となった。

しかしながらこの解析には次に挙げる2つの問題点が考えられる。i) 個人の異質性の考慮ができていない点、ii) 閲覧と購買を同時に考慮できていない点である。i) については、多様性変数が消費者の異質性を捉えるとはいえ、実際にはモデルに変量効果として個人の異質性を投入することで異質性は考慮が可能であり、多様性変数の有用性がなくなってしまう可能性がある。ii) に関しては、実務での有用性を加味すると、複数回発生する訪問と購買について同時に考慮できるモデルの方が貢献は高いといえる。そこで個人の異質性を変量効果として投入し、なおかつ各消費者について複数回ある訪問期間の予測とその際の購買の有無に関する予測としてWeibull分布に基づく生存時間分析とロジスティック回帰分析での同時モデリングを行なった。通常このような解析にはマルコフ連鎖モンテカルロ法（MCMC）等のシミュレーションモデルを用いることが一般的だが、近年のビッグデータの活用の潮流に起因する顧客データの蓄積により、MCMCを用いたこのような解析には非常に時間がかかるという問題がある。そこでこのような大規模データに適用しても短時間で効率的にモデリングが行える手法として大きく注目されているモデルである積分段階的ラプラス近似法（INLA）を用いて解析を行った。INLAは近年混合線形モデルへの一般化が行われたことで様々な応用研究に用いられている。

結果として、個人の異質性を変量効果として投入しなおかつ説明変数としてモデルに多様性変数を投入した場合に、モデルの情報量規準が最も向上した。またパラメー

タ推定値においても本研究で提案する多様性変数の多くが有意に説明力を持った一方で、既存の質的情報の説明力が低下した。この結果から多様性変数の有用性は単に消費者の異質性を捉えるだけではなく、第2章で考察したように Web 利用の習熟度やデモグラフィック変数の代替などの効果が発生していることが予想された。

第3章までの解析においては、本研究で着目している多様性変数について主にウェブ上で得られるデータのみを対象としてきた。しかしながらこれらの多様性の情報はウェブ上に限らず消費者の日常生活の嗜好の把握とそれに基づくより幅広い購買予測にも有用であることが考えられる。さらに深層学習での多様性変数の利用による予測精度の変化についても検証が必要である。そこで第4章「深層学習を用いた Web と実行動の多様性の検討」では、近年積極的に行われているオムニチャネル（実店舗・オンライン上の EC サイト・モバイルアプリなど複数の販売チャネルを組み合わせることで、より各消費者の生活に最適化したマーケティングを行う手法）での行動予測にあたり、実際の小売ブランドの購買履歴データなど複数の大規模データを組み合わせ、第3章と同様に変数群を差し替えての予測精度の変化に加え、深層学習と GLM で同一の解析を行なった場合の比較を行なった。その結果として、いずれも深層学習モデルに多様性変数を投入した場合において最も予測精度が改善する結果となり、深層学習を用いた柔軟な予測モデルの有用性や、オンライン上の行動に限らない幅広い行動についての多様性の考慮が予測精度に貢献することを示した。

第5章「まとめ」では各章で実施した解析の目的・内容・結果を俯瞰して比較するとともに、既存指標に対する多様性変数の有用性やマーケティングにおける深層学習の活用についても結果をもとに改めてまとめている。最後に今後の課題として得られるデータの制約に依存する問題など、複数の課題について議論を行っている。

