

(資料 8-1)

2023年 1月 25日

## 博士學位論文審査結果報告書

論文題目：機械学習に基づくイーコマースにおける顧客流出予測に関する研究  
Research on Consumer Churn Prediction in E-commerce Based on Machine Learning

申請者：夏侯 賢城 (カコウ ケンジョウ)

学歴：令和 2年 4月 大阪産業大学大学院研究科  
経営・流通学専攻修士課程入学

令和 2年 3月 同上修了

令和 2年 4月 同上経営・流通学専攻博士後期課程進学

令和 5年 3月 同上修了見込み

学力確認：令和 5年 1月 25日 (博士論文のみ)

研究業績 国際ジャーナル 2件、アメリカンジャーナル 1件  
(計ジャーナル 3件採録)、人工知能学会全国大会 1件

発表論文 論文目録を参照のこと

審査委員：大阪産業大学大学院経営・流通学研究科

教授：原田 良雄



審査委員：大阪産業大学大学院経営・流通学研究科

教授：中村 徹



審査委員：大阪産業大学大学院経営・流通学研究科

教授：藤岡 芳郎



申請者氏名	夏侯 賢城 (カコウケンジョウ) (20DC02)
<p style="text-align: center;">(論文内容の要旨)</p> <p>インターネット上で商品を売買する EC サイト市場の拡大を背景に、市場内での顧客獲得競争が起きている。顧客は企業の最も重要な資産の一つとして、企業の市場競争力と業績の向上に非常に重要な役割を果たしている。しかし、激しい市場競争の中で顧客は多くの製品やサービスプロバイダから容易に選択することができ、企業では顧客の流出が発生する可能性がある。一方、パレート法則によると、企業の 80%の利益は 20%の顧客から生み出す。企業が市場の優位性を維持するために、既存の顧客資源をどのように流出を避けるか、そして、既存の顧客をどのように利用すればよいかは企業が直面している重要な問題となっている。</p> <p>顧客関係管理 (Customer Relationship Management : CRM) とは情報技術の援用、すなわちデータを用いて顧客を識別し、顧客に対して顧客ごとに合ったダイレクト・メッセージを配信するなどで満足させ、顧客維持につなげていくものとされて、顧客と企業とのあらゆる接点や接触履歴を全て管理し、それをもとに顧客との関係性を深めて維持することで顧客満足度や顧客のロイヤルティを向上させ、収益の拡大を目指すマネジメント手法である。顧客ロイヤルティを高めるためには顧客データの分析が不可欠である。顧客関係管理に関する先行研究によると、顧客流出予測技術を用いて流出する可能性がある顧客を識別して、予測結果に基づいてマーケティング戦略を改善し、既存の顧客を維持することで業績損失を効果的に防止することができる。そのため、実務と理論の両側面から見ると、企業と顧客の関係性を構築する CRM システムを重視しなければならない。</p> <p>ほとんどのショッピングサイトでは、Webサーバとデータベースサーバが連携して動作している。データベースサーバには、顧客情報、商品情報、在庫情報、販売情報などが保管され、Webサイトの訪問者が入力した情報が、リアルタイムにデータベースに書き込まれて更新される。一方、ビッグデータとクラウドコンピューティングなど情報技術の発達によって、消費者の購買データなどの消費者行動データは容易に収集、保存できるようになった。EC企業にとっても CRM における顧客の購買行動データの分析と顧客流出予測に関する研究はあまり進んでいない。多様なデータに顧客流出予測手法が完全に対応されないため、各業界が多くのデータを保有しているにも関わらず適切な活用方法が</p>	

まだ確立されていない。そのため、本稿ではCRMに関する先行研究文献の調査をもとに、機械学習に基づくイーコマースにおける新たな顧客流出予測モデルを提案した。

本論文は、イーコマースの発展と EC 顧客購買行動の多様性及びビッグデータ情報技術の背景に基づいて、オンライン購買行動と購買時間の特徴を分析し、B2C イーコマースのマーケティングの顧客セグメンテーションと顧客流出予測の学術論文におけるセグメンテーションと予測実施の手法を総括した上で、RF-PACV モデルと「セグメンテーション・ファースト」モデルを提案した。

多種類の機械学習アルゴリズムを用いた顧客流出予測の精度を考察して、その後、各種の機械学習アルゴリズムの予測性能を比較し、B2C イーコマースデータ分析における機械学習の有効性・実用性を議論する。

第 I 章「顧客セグメンテーションと CRM に関する先行研究」では、従来の研究はRFMモデル分析に基づいて顧客セグメンテーションを行うことが多かったが、顧客の購買行動、購買意識と意思決定などの情報を把握することが難しい。このため、本章はKotler と Keller (2008) が提示した顧客セグメンテーションの基準に基づき、各種顧客セグメンテーション方法の特徴を詳しく分析し、B2C イーコマース顧客セグメンテーション方法を提示し顧客セグメンテーションの研究構想を明示する。

第 II 章「顧客流出予測に関する先行研究」では、まず機械学習の基本概念、機械学習におけるデータ活用の流れ、および基本手法とその応用を述べる。

第 III 章「顧客流出予測モデリングの基礎理論」では、予測モデルを構築するためには、多種類の機械学習アルゴリズムが必要になる。そこで、本章は既存の顧客流出予測アルゴリズムの理論その方法を述べる。各種のアルゴリズムの手順をそれぞれ紹介する。本章で用いたアルゴリズムは、k 平均法 (k-means) 、ランダムフォレスト (Random Forest : RF) 、ロジスティック回帰 (Logistic Regression:LR) 、サポートベクターマシン (Support Vector Machine: SVM) 、誤差逆伝播法によるニューラルネットワーク (Back Propagation Neural Network: BPNN) 、アダブースト (AdaBoost) である。

第 IV 章「顧客流出予測の実証研究」では、Alibaba Cloud Tianchi プラットフォームのデータベースを原始データセットとして用いて、4 種類の予測アルゴリズムを採用し、それぞれロジスティック回帰 (Logistic Regression : LR) 、サポートベクターマシン (Support Vector Machine : SVM) 、誤差逆伝播法によるニューラルネットワーク (Back Propagation Neural Network:BPNN) と AdaBoost アルゴリズムの 4 種類の予測アルゴリズムを用いた。そして、この 4 つのアルゴリズムの予測性能を比較する。

第 V 章考察では、顧客セグメンテーションの結果に基づいて、顧客セグメン

テーション後の顧客群の消費行動の特徴を検討する。また、4種類のアルゴリズムの予測モデルの結果を基に、予測モデルの有効性・有用性と理論価値を検討する。主に以下の3つの観点から考察を行っている。

### 1. 顧客流出予測の有用性

本研究は顧客流出率を予測する流れの中に縦断的な時間変数とその他の変数（クリック、追加購入、コレクションなど）を採用して、顧客セグメンテーションを行って、各顧客クラスターの流出の割合を計算した。顧客そして、流出顧客のデータに基づいて、流出に対する変数の影響重みを計算し、顧客が流出した具体的な原因を明らかにする。これにより、企業が顧客関係管理レベルを高めると同時に、顧客流出管理をより細かく行うことができ、顧客関係管理効率を高め、管理財務コストを減らすと考えられる。また、マーケティングの面でも積極的な影響を与えることも期待される。

### 2. 変数の選択と顧客セグメンテーション

B2Cイーコマースの顧客行動データの縦断的な時間性と多様な行動情報（クリック数、購買回数、カートに入れ数、お気に入り数）の特徴を重要な変数として、消費者行動を4つの時間帯（未明、午前、午後、夜）に基づいて細分化することが本研究の新しい見解のひとつである。実証実験では、時間（期間）変数の観点から、時間に関する変数を「時間」と定義される。予測データセットでは商品種類変数に加えて、データ変数は時間変数（未明、午前、午後、夜）と4つの動的な変数（クリック数、購買回数、カートに入れ数、お気に入り数）と組合せて、RF-PACVモデル（人工知能学会全国大会にて発表）を新たに定義した。

そして、このモデルについてセグメンテーションを行って、3つの顧客クラスターに分けている。第IV章の予測実験結果により、夜変数と午後変数は流出予測に大きな影響を及ぼし、「夜Buy」と「午後Buy」の変数重要性が上位2位にランクされていることが分かった。研究結果は、ショッピング時間変数が流出予測の重要な変数であることを示すことができた。有効であると考えられる。

### 3. 予測モデル

本研究で構築している「セグメンテーション・ファースト」予測モデルは、二つの機械学習アルゴリズムを利用して、顧客流出について流出予測を行った。その結果、各アルゴリズムのセグメンテーション後の予測性能はセグメンテーション前より高いことを示し、「セグメンテーション・ファースト」予測モデルの有効性をデータを用いて示すことが出来た。

結論では、顧客セグメンテーションと各モデルの予測性能の結果について、顧客セグメンテーション方法と予測モデルがイーコマース企業のマーケティング戦略の制定に与える影響を議論する。実証研究の結果についての考察に基

づき、本研究の学術的な貢献と B2C イーコマース企業の顧客関係管理の現実的意義についてまとめ、今後の課題を述べる。

結果から見ると、AdaBoost のセグメンテーション後の予測の確率はセグメンテーション前の確率より高いだけでなく、セグメンテーション後の予測の確率は他のアルゴリズムよりも高い。その結果、AdaBoost を用いて構築している顧客流出の予測モデルが優れていることを示した。つまり、B2C イーコマースにおいて、AdaBoost は適用でき有用なことを示した。。また、本研究で構築している「セグメンテーション・ファースト」予測モデルは、二つの機械学習アルゴリズムを利用して、顧客流出について流出予測を行った。その結果、各アルゴリズムのセグメンテーション後の予測性能はセグメンテーション前より高い。これで、本研究で構築している「セグメンテーション・ファースト」モデルは有効であることをしめすことができた。（採録された論文より）

申請者氏名	夏侯 賢城 (カコウケンジョウ) (20DC02)
-------	------------------------------

(論文審査結果の要旨)

本研究は、イーコマースの発展と EC 顧客購買行動の多様性及びビッグデータ情報技術の背景に基づいて、オンライン購買行動と購買時間の特徴を分析し、B2C イーコマースのマーケティングの顧客セグメンテーションと顧客流出予測の学術論文におけるセグメンテーションと予測実施の手法を総括した上で、RF-PACV モデルと「セグメンテーション・ファースト」モデルを提案した。多種類の機械学習アルゴリズムを用いた顧客流出予測の精度を考察して、その後、各種の機械学習アルゴリズムの予測性能を比較し、B2C イーコマースデータ分析における機械学習の有効性・実用性を議論する。

B2Cイーコマースに対する顧客セグメンテーション研究は比較的少なく、本研究では従来の研究者の顧客セグメンテーションに関する方法を参考にして、主にイーコマースショッピングサイトで商品購買行為を展開する顧客に対してセグメンテーション研究を行い、その後、セグメンテーション結果に基づいて異なる顧客群に対して流出予測の研究を行う。セグメンテーション生成においては、どの手法が EC に適合的であるかを実証的に比較評価している。実証評価に利用するデータは、元のデータセットはAlibaba Cloud「TIANCHI天池」プラットフォーム(2014)が公開した科学研究とビッグデータコンテスト用のデータセットである。

流出予測には4種類の予測アルゴリズムを採用し、それぞれロジスティック回帰(Logistic Regression: LR)、サポートベクターマシン(Support Vector Machine: SVM)、誤差逆伝播法によるニューラルネットワーク(Back Propagation Neural Network: BPNN)とAdaBoostアルゴリズムを用いている。そして、この4つのアルゴリズムの予測性能を比較し結果を評価した。

本稿は、上記を研究した結果についてまとめたものであり、得られた結果は次のとおりである。

1. 顧客流出予測の予測アルゴリズムについて

流出予測において、4種類の予測アルゴリズムを比較検討した結果、AdaBoostアルゴリズムを用いた実験結果が一番優れていた。

2. 変数の選択と顧客セグメンテーションについて

B2Cイーコマースの顧客行動データの縦断的な時間性と多様な行動情報(クリック数、購買回数、カートに入れ数、お気に入り数)の特徴を重要な変数として、消費者

行動を4つの時間帯(未明、午前、午後、夜)に基づいて細分化することが本研究の新しい見解のひとつである。実証実験では、時間(期間)変数の観点から、時間に関する変数を「時間」と定義される。予測データセットでは商品種類変数に加えて、データ変数は時間変数(未明、午前、午後、夜)と4つの動的な変数(クリック数、購買回数、カートに入れ数、お気に入り数)と組合せて、RF-PACVモデル(人工知能学会全国大会にて発表)を新たに提案した。

### 3. セグメンテーション・ファーストモデルの有効性

顧客流出予測モデルを構築する時、二つの機械学習のアルゴリズムを採用するが、ひとつの機械学習アルゴリズムで顧客セグメンテーションを先にして、その結果をもう一つの機械学習アルゴリズムに代入して顧客流出予測を行うこと。つまり、「セグメンテーション・ファースト」モデルである。これで、セグメンテーション前とセグメンテーション後のモデル予測結果の比較により、顧客セグメンテーションの効果を測定する。予測結果によると、k-meansクラスタリングを用いて顧客のセグメンテーション後の各予測指標が明らかに向上し、セグメンテーションが必要であることを証明した。

### 4. 研究の意義

本研究はB2Cイーコマースの顧客流出予測モデルの開発と企業の顧客関係管理に啓示がある。例えば、イーコマース企業が顧客関係管理を展開する際に、顧客をどのように細分化し、予測アルゴリズムをどのように選択するかは、本研究の成果を参考にすることができる。

B2Cイーコマース業界に関する顧客のセグメンテーションと流出予測モデルに関する研究の発展に積極的な貢献をした。

B2Cイーコマース環境の下でどのように変数を選択するか、どのように顧客のセグメンテーションを行うか、どのように変数の数を減して次元を下げるか、新しい見解を提示した。

「セグメンテーション・ファースト」という予測モデルの構築によって、B2Cイーコマースにおいて顧客流出に対する新しい予測方法を提案した。これも本研究のもうひとつの貢献である。

#### (審査委員会の所見)

本論文は、International Journal (2件)、American Journalに掲載され、また人工知能学会に発表した複数の論文をベースに構成されており、個々の提案や有用性については評価を受けており、学術上、實際上寄与するところが少なくない。よって、本論文は博士(経営学)の学位論文として価値あるものと認める。