

位置情報データによる競合店舗の利用状況の多様性を 用いた購買予測手法の提案

A Novel Approach of Purchase Prediction Considering the Purchase on Competing
Company by Using ID-POS Combined with Geofencing Location Data

新美潤一郎*¹
Junichiro NIIMI

星野崇宏*²
Takahiro HOSHINO

*¹理化学研究所革新知能統合研究センター
RIKEN Center for Advanced Intelligence Project

*²慶應義塾大学経済学部, 理化学研究所革新知能統合研究センター
Faculty of Economics, Keio University, RIKEN Center for Advanced Intelligence Project

Nowadays various kinds of data about customers behavior on online websites or applications can be collected for the marketing analysis. However, especially for the physical retailing stores, it is still difficult to acquire their behaviors on competing firms. In this research, we develop a novel approach of considering the usage of competitors. We present the way to predict customers future purchase on retailing stores with ID-POS data combined with the third-party large-scale GPS locational data collected from the smartphone apps.

1. はじめに

近年ではEC（電子商取引）サイトの発達やネットスーパーの普及などの要因により実店舗での小売販売も競争が激化している。マーケティングの戦略策定にあたり自社で保有する顧客について正しく情報を把握することは、正確なターゲティングやプロモーション展開にあたって必須である。中でも自社ロイヤルティの推定にあたっては消費者のシェア・オブ・ウォレット（SOW）*¹を考慮することが重要となる。

このようなSOWの予測にあたっては、EC市場においてはClickstream Dataをはじめとしたアクセス履歴データの収集によるウェブサイト横断的なインターネットの閲覧履歴データを用いることで、消費者の競合企業での行動を考慮した解析が行われている。しかし実店舗企業でのSOWは、企業横断的に利用履歴を得ることが困難であることから未だ事例が少ないというのが現状である。実店舗での購買を把握する代表的な手法として、これまでもアンケートを用いた市場調査や、家計簿サービスにおける購買後のレシートのスキャンなどの方法で消費者の支出を把握することが試みられてきたが、これらの手法では認知バイアスによる回答の歪みやレシートスキャンにあたって調査対象者への負担が大きいといった問題があり、正確なデータを大規模に収集することは困難であった。

近年では広く普及したスマートフォン端末で、複数のアプリケーションを用いて特定の消費者の位置情報を継続的に収集する試みが大規模に行われている。しかし収集された位置情報は企業の小売データ等と共通のIDで連携しているわけではなく、マーケティ

ング分析への活用にはまだ課題が残っている。そこで本研究では、大規模に収集された位置情報データと特定の小売企業で収集されたID-POSデータを紐づけることで、実店舗での競合他社を考慮したマーケティング解析の手法を提示する。

2. 先行研究

2.1 機械学習を用いた購買予測

機械学習、とりわけ深層学習はその急速な発展とともに様々な分野に応用されているが、マーケティング分野としての消費者の購買予測に応用した事例は画像認識等に比べると少数である。代表的なものとしてオンラインストアでの購買を扱ったものには、消費者のECサイト上での行動についてのClickstream Dataから、あるECサイトへの訪問ごとの購買有無の予測とその際に購買された商品を予測し、その精度をロジスティック回帰分析、ランダムフォレスト、Deep Belief Network (DBN)で比較した[Vieira 15]がある。一方で実店舗での購買データを外部データと紐付けて解析したものとしては、購買履歴データ、アクセス履歴データ、アプリ操作履歴データから実店舗・オンラインストアでの購買を横断的に予測し、Feed-Forward Neural Network (FFNN) [Svozil 97]と一般化線形モデル (GLM) で結果を比較した[新美 17]などがある。

2.2 行動の多様性の考慮

本研究では行動の多様性にも焦点を当てている。行動の多様性とは、消費者の行動の幅広さについて特定要素への依存度を測ることで考慮したものである。消費者の行動の多様性を購買予測に考慮することについては、[新美 15]においてオンラインストアで購買された商品カテゴリや自社店舗への流入経路（リファラ情報）の幅広さを多様性指標として投入す

連絡先: 慶應義塾大学経済学部星野崇宏 (e-mail: hoshino@econ.keio.ac.jp)

*¹ 顧客のある市場全体への支出に対する自社への支出額の割合

表 1: ある消費者 A,B が利用した店舗数の例

	消費者 A	消費者 B
利用総数	50 回	50 回
T 社	10 回	10 回
競合 A 社	10 回	37 回
競合 B 社	10 回	1 回
競合 C 社	10 回	1 回
競合 D 社	10 回	1 回
T 社利用の割合	20%	20%
特定ブランドへの依存度	低	高
HHI	2000	5888

ることで購買予測の精度が改善することが示されている。多様性指標の定式化には、経済指標としてある市場の寡占の度合いを表すハフィンダール・ハーシマン指数 (HHI) [Davis 09] を用いている。HHI は市場に参入する企業 i の市場シェア s_i の二乗和として次の式 1 のように表される。

$$HHI = \sum_{i=1}^k s_i^2 \quad (1)$$

HHI は一要素が完全に独占している場合に最大値 $100^2 = 10000$ を取る非負の値であるが、本研究では解析で扱いやすいよう 10000 で除し 0 から 1 の値をとる連続変数として投入している。

位置情報から得られる行動の多様性の例として、次の表 1 にある 2 人の消費者 A, B が期間中に訪問した店舗についての集計を示すと、消費者 A, B はともに自社店舗の利用の割合は同一であるものの、全体として特定ブランドへの依存度が大きく異なっている。市場全体の利用回数、自社利用回数、自社 SOW が同一な場合でも HHI を用いることで購買行動の異質性を把握することができる。消費者 A が特定のブランドにロイヤルティを持たずに様々なスーパーマーケットを利用している一方で、消費者 B の場合には競合 A 社を中心に利用しつつ T 社を補助的に利用しているような状況が推察できることから、多様性変数を用いることで顧客のターゲティングをより最適化することが可能である。

また本研究と同様にマーケティングデータに対して深層学習手法を適用する場合には、企業規模が増大するにつれてカテゴリカル変数 (例として「最もよく利用した時間帯」「最もよく購買された商品カテゴリ」等) の増加により予測モデルの大規模化に繋がらう。これは結果としてデータ解析の構築コストの増大や意思決定の遅延をもたらす可能性もある。[新美 17] ではこのような大規模なカテゴリカル変数の代替として多様性変数を投入することで、予測器の次元数を大きく削減した上で高い予測精度を維持した効率的な解析が実現できることが示されている。

3. 解析手法

3.1 使用データと分析対象

本研究では、実店舗における購買行動の予測を目的として九州地方を中心に展開する大手スーパーマーケットチェーン T 社の消費者を分析対象としている。本研究では 2 つのデータを組み合わせる解析に用いているが、まずは T 社から提供いただいた 2017 年 8,9,10 月の ID-POS データを用いて、8,9 月の購買から 10 月の購買を予測する解析を行う。したがって 2017 年 8,9 月に店舗で購買を行った消費者を T 社における自社顧客と定義している。ただし位置情報とのマッチングのため、8 から 10 月の期間全体にわたって最低でも 3 回の購買が行われていないものについては対象から除外している。

2 つ目のデータとして本研究では位置情報データを活用している。この位置情報データは一般に消費者がスマートフォンで特定の複数のアプリケーションを利用した際にバックグラウンドで取得された位置情報がサーバ上で集計されるものである。ただし GPS に基づく正確な位置情報ではなく、地域を特定の距離で区切った仮想の区画であるジオフェンス内への進入・退出を記録したものである。これを用いて T 社の各店舗および競合すると考えられる近隣の複数の店舗のジオフェンスを特定し、収集した位置情報から消費者の店舗訪問の履歴を得られることを確認した。最後に T 社店舗での購買履歴とその時間帯に該当店舗に訪問していることを条件としたマッチングを実施し、ID-POS と位置情報の紐付けを行った。マッチングの方法については 3.2 節で詳述。

3.2 購買履歴と位置情報のマッチング

本研究で用いている購買履歴データと位置情報データは異なる消費者を対象として異なる手法で収集されたものであり、必然的に異なる ID を用いて集計されている。したがって解析に先立ちこれらデータの紐付けが必要となるため、ここで本研究でのマッチング方法について述べる。位置情報データでは特定ジオフェンスへの進入・退出の時間が、購買履歴データでは ID で継続的に特定の消費者の購買日時が追跡されていることから、決済を行った日時の前後 30 分以内に滞在情報が得られていることをもって同一購買とみなしている。次に位置情報と購買履歴の各 ID について、期間中に少なくとも 3 回以上購買を行っている顧客に対して、期間中全体の購買回数に対するこの一致回数の割合の多いものから順にマッチングさせた。結果としてマッチングデータは 1,000 程度が得られ、そこからさらに購買条件に従い顧客を抽出した。

3.3 予測手法

予測には研究課題として近い手法である [新美 17] と同様に FFNN による 3 層の一般的な深層学習を用いた。複数の変数群を用意し、使用する特徴量を差し替えることによる予測精度の変化について確認し、位置情報の多様性を投入して予測精度が改善することをもって位置情報の多様性の有効性を示す。

解析結果については当日示す。

参考文献

- [Davis 09] Davis, P. and Garcés, E.: *Quantitative techniques for competition and antitrust analysis*, Princeton University Press (2009)
- [新美 15] 新美 潤一郎, 星野 崇宏: 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*, Vol. 44, No. 3 (2015)
- [Svozil 97] Svozil, D., Kvasnicka, V., and Pospichal, J.: Introduction to multi-layer feed-forward neural networks, *Chemometrics and intelligent laboratory systems*, Vol. 39, No. 1, pp. 43–62 (1997)
- [Vieira 15] Vieira, A.: Predicting online user behaviour using deep learning algorithms, *CoRR*, Vol. abs/1511.06247, (2015)
- [新美 17] 新美 潤一郎, 星野 崇宏: 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*, Vol. 32, No. 2, pp. B–G63_1–9 (2017)