

テナント型 EC モールを対象とした 商品の安価タイミング予測モデルの提案 ータイムセール周辺期間を対象としてー

石井 楽志[†] 松本 恒雄^{††} 福島 直央^{†††} 木戸 冬子^{††††} 山名 早人^{††††}

[†] 早稲田大学基幹理工学研究科 〒169-8555 東京都新宿区大久保 3-4-1

^{†††††} 早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1

^{†††††} 早稲田大学理工学術院総合研究所 〒169-8555 東京都新宿区大久保 3-4-11

^{††} 国民生活センター 〒252-0229 神奈川県相模原市中央区弥栄 3-1-1

^{†††} LINE 株式会社 公共戦略室 〒160-0004 東京都新宿区四谷 1-6-1

E-mail: [†], ^{†††††} {g-ishii, yamana}@yama.info.waseda.ac.jp, ^{†††} nao.fukushima@linecorp.com, ^{†††††} fkido@aoni.waseda.jp

あらまし 近年, EC 市場は急速的な成長を遂げている. EC 市場の商品は実市場と比べ価格変動が激しく, ユーザは商品を購入するベストなタイミングを予測しづらい. そこで本研究では, ユーザが商品の購入タイミングを決定する一助となるべく, EC 商品の安価タイミング予測モデルを提案する. 運営側が行うセールに対して価格・ポイントに特徴的な変動があるという仮説のもと, 過去のセール周辺期間における価格・ポイント倍率の変動を学習し, 安価に商品を購入できる期間を予測するモデルを構築した. 評価実験では, 安価購入タイミングを 4 クラス分類のタスクとして定義 (セール期間中, セール終了後 3 日間, セール終了後 4 日間~セール終了後 1 週間, 何れも同一) し, 延べ 7,200 商品の価格・ポイント変動に対して, 正解率 0.601 を達成した. また, ポイント倍率を特徴量に追加することで正解率を向上させることができることを示した.

キーワード 電子商取引, 消費者問題, 時系列データ処理, 深層学習

1. はじめに

インターネットの普及とともに EC 市場は急速に成長を遂げている. EC 市場の成長加速の背景には「プラットフォームビジネス」があり, 売り手と買い手をつなぐ場を提供し, 多くのネット店舗が出店している. プラットフォームビジネスを展開している代表的な企業に楽天, Amazon, Yahoo! JAPAN がある. ショッピングモール型の EC プラットフォームを EC モールと呼び, ショップ単位で出店する EC モールを「テナント型 EC モール」と呼ぶ. 楽天市場や Yahoo!ショッピングがこれに該当する.

ネット上で消費者が商品を探す際に, 商品価格情報は重要な検索条件の一つである[1]. EC 市場は実市場と比較して, 価格調整コストの低下とリードタイムの短縮により, 商品価格の調整に柔軟性がある. したがって, EC 市場の商品価格は実市場に比較して高頻度で変化し, その分散も大きい[2]. つまり, 消費者視点で考えた場合, 価格変動の激しいオンライン商品を安価に購入するタイミングを見極めることは困難である. 加えて, 近年の代表的な EC モールでは, 各プラットフォーム上で利用できるポイントを運用しており, ポイントを加味した上で安価となるタイミングを判断することはさらに難しい. Yahoo!ショッピングでは T ポイント, 楽天市場の利用では楽天ポイントが購入金額

に応じて加算され, 消費者のロイヤリティ形成に有効であることが報告されている[3]. 多くの消費者にとってポイントは, 企業通貨・疑似通貨として重要な価値があると認識されており, 購買行動に影響を与えている[4]. 売り手は効果的なポイント戦略を打ち立てるため, 価格と同様, 付与ポイントの倍率も激しい変動があると予想される.

こうした背景のもと, 本稿は, 消費者が商品購入タイミングを決定する一助となるべく, 商品価格・ポイント倍率の変動パターンを分析し, 予測モデルの構築を検討する. 特に, 価格や変ポイント変動が大きい「EC モールのセール周辺期間」を対象とした予測モデル構築に取り組む.

EC 商品の価格予測に関する研究としては, 2018 年に Tseng ら[5]が, ニュースの感情分析を利用し, ARIMA (自己回帰和分移動平均) モデルを用いた SSA-ARMA (signal, sentiment, autoregressive and moving average) モデルを提案し, 携帯電話の価格上昇・下落・変動無の 3 クラス分類を行った. 2019 年には, Carta ら[6]が, ARIMA モデルの外生変数として Google トレンドを用いて Amazon の商品を対象に価格予測を行い, 10 日後までの予測で平均絶対誤差 1.98%, 30 日後までの予測で 6.64%を達成した. 2020 年に Fathalla ら[7]は, 中古品を対象に, 画像と説明を用い, LSTM と CNN を

含むニューラルネットワークを利用した中古品価格予測を行い、平均絶対誤差 2%を達成している。しかし、何れもセール期間やポイントを加味した予測は実現されておらず、セール期間やポイントを加味した価格予測としては、本稿が最初の試みとなる。特にセール期間における価格やポイントの変動は大きいことが予想され、こうした変動が大きい期間に対する予測を実現することに本研究の意義がある。

本稿では、「EC モールのセール周辺期間（以降セール周辺期間とは、セール 1 週間前からセール 1 週間後の期間を指す）において特徴的な価格・ポイント倍率変化がある」という仮説のもと、セール周辺期間の安価購入タイミング予測に取り組む。予測では、ポイント時系列データを含めた価格予測をもとに、安価タイミングを 4 クラス分類問題（セール期間、セール終了後 3 日間、セール終了後 4 日間～セール終了後 1 週間、何れの期間も同一）として取り扱う。なお、従来の研究[5][6][7]では、予測対象商品毎の長期に渡る価格変動を用いて予測モデル構築を行っているが、本稿では、商品毎に学習器を構築せず、類似する価格変動を伴う商品群毎に分類器を構築する。これは、テナント型 EC モールを対象とした場合、膨大な商品数、テナント数を扱うことと、商品の入れ替わりが激しいため、各々に分類器を用意することが困難であるからである。

本稿の貢献は以下の 3 点である。

- セール周辺期間における商品の安価タイミング予測では、ポイント倍率のデータが特徴量として有効であることを明らかにした点。
- 上記をもとに、EC モールのセール周辺期間での購入タイミング予測モデルを 4 クラス分類により構築した点。
- のべ 7,200 商品（楽天市場）を対象とした実証実験により、4 クラス分類の正解率 0.61 を達成した点（学習用に 4,800 商品、テスト用に 2,400 商品を使用）。

以下、2 節で関連研究を紹介する。3 節ではリサーチクエスチョンを定義する。4 節では対象となる EC モールを紹介する。5 節では事前実験の手法と結果について述べ、6 節において予測モデルを提案する。7 節において実証実験結果を示し、8 節で結論を示す。

2. 関連研究

EC 商品の価格予測に関する研究としては、2018 年に Tseng ら[5]が、ニュースの感情分析を利用し、ARIMA（自己回帰和分移動平均）モデルを用いた SSA-ARMA (signal, sentiment, autoregressive and moving average) モデルを提案した。Tseng らは、提案モデルを用いて、携帯電話の価格上昇・下落・変動無を予測し

ている。しかし、その精度は示されていない。

2019 年には、Carta ら[6]が、ARIMA モデルの外生変数として Google トレンドを用いて Amazon の商品を対象に、商品情報と商品メーカーの情報を用いた価格予測を行った。10 日後までの予測で平均絶対誤差 1.98%、30 日後までの予測で 6.64%を達成した。Carta らは、約 900 万もの商品について 2016 年から 2017 年の間の 1 日毎ごとの価格を元に価格予測を行っており、膨大な過去の商品毎の価格変動データを必要とする。

2020 年に Fathalla ら[7]は、中古品を対象に画像と説明を用い、LSTM と CNN を含むニューラルネットワークを利用した中古品価格予測を行い、平均絶対誤差 2%を達成している。しかし、単一の EC サイトを対象としており、かつ、価格変動の大きいセール周辺期間は対象となっていない。

2018 年の Bakir ら[8]の研究は、上記の研究と異なり、商品の直前の 7 日間の価格変動を入力として翌日の価格を予測するモデルを提案している。amazon.fr におけるスマートフォンの販売価格を予測対象としており、LSTM ニューラルネットワークが、SVR モデルに比較して RMSE を 36.9 ユーロから 23.64 ユーロに小さくできることを確認している。

上記で紹介した従来の研究では、商品毎の長期に渡る膨大な価格変動データを必要とするだけでなく、本稿で対象とする価格変動が激しいセール期間を対象としていない[5][6][7]。一方で、Bakir らの研究[8]では、商品に対する直近 1 週間のデータから翌日の価格を予測しているが、2 日目以降の予測についての正確性は議論されていない。

3. 本論文の貢献とリサーチクエスチョン

本稿は、消費者が商品購入タイミングを決定する一助となるべく、商品価格・ポイント倍率の変動パターンを分析し、特に、価格や変ポイント変動が大きい「EC モールのセール周辺期間」を対象とした予測モデルの構築を目的としている。

EC モール（テナント型 EC モール）を対象とした場合、従来研究[5][6][7]のように長期に渡って商品毎の価格変動データを取得し予測することは困難である。そこで、本稿では、Bakir らの研究[8]と同様、予測時点の直近の価格変動データから当該商品の価格変動を予測することに取り組む。Bakir らの手法では翌日の価格予測を行っているのに対し、本稿では、長期に渡る価格変動を 4 クラス分類（セール期間、セール終了後 3 日間、セール終了後 4 日間～セール終了後 1 週間、何れの期間も同一）として実現する。

また、従来の研究[5][6][7][8]は、何れも日常的な価格変動予測を対象としているのに対し、本稿では、EC

モール特有のセール周辺期間での大きな価格変動が発生する期間での商品購入タイミング予測を目指しており、従来の研究が対象としなかった領域である。さらに、ポイントを含め購入タイミングを予測する点にも新規性がある。

上記を実現するため、本稿では、商品毎に学習器を構築せず、類似する価格変動を伴う商品群毎に分類器を構築する。これは、テナント型 EC モールを対象とした場合、膨大な商品数、テナント数を扱うことと、商品の入れ替わりが激しいため、各々に分類器を用意することが困難であるからである。本手法により、過去の価格変動データを学習データとして持たない商品に対する予測を可能とする。

以上を踏まえ、EC 市場でお得に商品を購入したいと考える消費者にとって有益となるような分析、予測の実現のため、以下のリサーチクエスチョンを設定する。

RQ1. EC モールで活用されているポイントと価格の変化に関連性はあるか？（5 節）

RQ2. セール周辺期間の短期間時系列を用いて、どの程度の精度で商品の安価タイミング予測が可能か？（7 節）

RQ3. 学習に使用する特徴量によって予測精度に差異はあるか？（7 節）

近年の EC モールでは商品購入時にポイントが付与されることが多く、実際に消費者がお得に商品を購入しようと考えた時、「価格の変化に注目し、付与されるポイントを考えていないことが多い」と予想され、RQ1 ではポイントと価格との関係を明らかにする。販売事業者の視点から、「商品の売れ行きの変動を捉えた値引き戦略」[9]の研究は実施されているが、消費者の視点から、価格の値引きや付与ポイントの増加が頻繁に発生する「タイムセール周辺期間の価格・ポイント変動」を調査した報告は、我々の知る限り存在しない。

また、RQ2, RQ3 を設定し、事前実験の結果を踏まえた商品の安価タイミング予測手法を提案する。

4. 対象とする EC モール

対象とする EC モールとして、以下の 3 条件を満たす EC モールを対象とした。

- 1) テナント型 EC モールである。
- 2) EC モール内でセールが実施される。
- 3) 商品購入時に EC モール上で利用可能なポイントが付与される。

1)の条件については、EC モールの運営側が実施するセールに対して「出店事業者が特徴的な価格設定を行うであろう」という仮説を立てたため、ショップ単位で出店しているテナント型 EC モールを対象とした。2)及び3)については、3 節で示した RQ を満たすために、

「EC モールがセールを開催する」、「ポイントを運用している」という条件が必要となるためである。

上記 3 点を踏まえ、「API を用いデータを取得できること」「大規模な EC モールであること」を前提として楽天市場を対象とした。以下に、楽天市場で実施されている代表的な大規模セールを示す。

1. 楽天スーパーセール
2. 楽天お買い物マラソン
3. 楽天大感謝祭
4. 楽天超ポイントバック祭
5. 楽天ブラックフライデー
6. 楽天イーグルス感謝祭

これらのセールは、ポイントの還元方法が「ショップ数カウント方式」と「購入金額カウント方式」の 2 種類に分けられる。1,2,3,5 のセールは商品を購入したショップの数によってポイント倍率が決まる「ショップ数カウント方式」となっており、4,6 のセールは、セール期間中に購入した商品の合計金額によってポイント倍率が決まる「購入金額カウント方式」である。

5. セール周辺期間での価格変動調査（事前実験）

事前実験では、「セール周辺期間では、どのような価格変動パターンが多いか」、「ポイント倍率が価格予測に有効な特徴量となるか」の二点を調査する。具体的には、セール周辺期間で価格変動が類似する時系列を分類し、変動パターンの特徴を調査する。特徴的な価格変動が見られる商品を対象に、ポイント倍率の変動の調査を含めて実施する。本事前実験の結果を踏まえ、安価タイミング予測モデルを次節で提案する。

5.1 概要と方法

「セール期間の前」「セール期間中」「セール期間後」といった、セールの実施に伴ったポイントと価格の変動に着目する。具体的には、以下の 5 つの期間での価格変動を分類し、セール周辺期間における価格変動パターンを調査する。これは、セール期間及びその前後の期間において価格及びポイント変動が発生するという仮定に基づく。なお、変動を把握するために、各期間は重なりを持つ分割となっている。

ステップ 1: セール期間前後でのデータの 5 分割

- ① セール開始 1 週間前からセール開始前日まで
- ② セール開始前日とセール開始日の 2 日間
- ③ セール期間
- ④ セール最終日とセール終了翌日の 2 日間
- ⑤ セール終了翌日からセール終了日 1 週間後まで

ステップ 2: 上記各期間内での回帰係数 β を求め、 β に応じて以下の 3 値分類を行う。閾値 0.01 は、事前の調査により、各パターンに分類されるデータ数が収束した（変化しなくなった）時の値を設定した。

- up ($\beta > 0.01$)
- stay ($0.01 \geq \beta \geq -0.01$)
- down ($\beta < -0.01$)

以上の2ステップを価格データに適用し、回帰係数の3パターン、期間の5パターンの組み合わせ、243 (= 3⁵) のパターンに商品の価格変動を分割する。

次に、頻繁に発生する価格の時系列変動パターンを対象に解析を行う。ここでは、全商品時系列データの70%に含まれるパターン群 (図1の赤枠部分) をデータ数が多い順に抽出する。これは、「セール周辺期間では、どのような価格変動パターンが多いか」「ポイント倍率が価格予測に有効な特徴量となるか」を確認するためであり、上位パターンのみを対象とする。

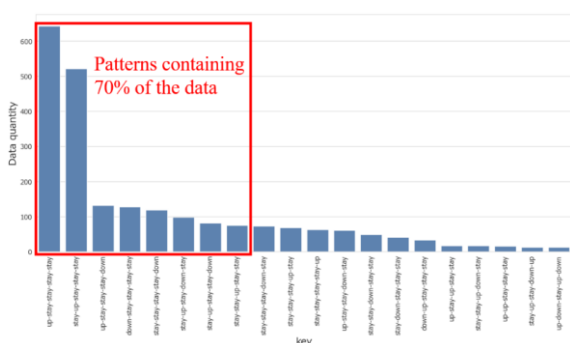


図1. 楽天超ポイントバック祭 (2021年12月13日開催) でデータ量の多い上位20個の価格変動分割パターンのデータ数 (5つの期間とも stay のパターンは除く)。

上記のパターン抽出をセール毎に行い、上位70%内の変動パターンに含まれる価格変動データを対象に詳細な解析を5.3以降で行う。

5.2 ECモールからのデータ収集

楽天市場の検索キーワードランキングトップ100 (2021年10月時点) を抽出し、同100キーワード各々に対して検索結果500件の商品情報 (商品名) をスクレイピングし取得した¹。合計5万件 (=100×500) のデータの内、商品名が重複しているものは削除する。次に、楽天商品検索API²を利用し、5万件の商品に対して一日1回、価格・ポイント倍率の時系列データを収集した。ポイント時系列データは、全てポイント倍率に置換し、データを統一する。例えば、1,000円の商品に対して100ポイントであれば、10倍に置換する。

以下では、2021/11/26~2022/6/2の19,386商品×189日の時系列データセットを扱う。

セール情報については、Rakutenカレンダー³から全ユーザ対象のセール情報を抽出した。なお、対象セー

ルは、4節で示した大型セールから、性質の異なるセールの2種類ずつ (「お買い物マラソン」, 「スーパーSALE」, 「超ポイントバック祭」, 「大感謝祭」) 選択した。なお、2021/11/26~2022/6/2内に実施された合計12セールの対象とする。

5.3 価格・ポイントデータに対する前処理

価格及びポイントの時系列データ (前項で作成) に対して、以下の前処理を施す。

(1) 欠損値の補完

価格、ポイントデータは毎日取得してるが、欠損が生じる。欠損データは、以下の手法により補完する。

- 欠損値は前日の価格、または前日のポイント倍率で補完する。
- 2日以上連続して欠損している値については補完せず、最初の日欠損値だけ補完し、それ以降は欠損値のままとする。

(2) 有効データを持つ商品の抽出

対象12セールにおけるセール周辺期間すべてにおいて欠損値がない商品を抽出する。

(3) 正規化

商品に依存することなく、価格変動を扱うため、商品毎に、取得データ期間 (2021/11/26~2022/6/2) 内の最高価格と最低価格を用いて範囲[-1,1]のMin-max normalizationを行う。ポイント時系列についてはポイント倍率をそのまま扱う。

上記の前処理の結果として、データ収集時の19,386商品から2,004商品を抽出した。

5.4 K-means 適用方法

5.2で抽出したセール毎の頻出価格変動パターン上位70%について、クラスタリングを用いた詳細な解析を行う。具体的には、商品別の価格変動をk-means⁴により、時系列クラスタリングした。k-means適用時の距離としては、ユークリッド距離を採用し、k-meansアルゴリズムの内部の最大イテレーション回数は100回、セントロイドの初期化に用いる乱数発生器の状態は27とした。なお、クラスタ数kは、エルボー法により算出した。

5.5 事前実験結果

セール周辺期間内での短期間時系列の変動パターンは、5.1で示した5つの期間 (①~⑤) おける3値分類の結果を、次のように表現する。

(期間①の分類) - (期間②の分類) - (期間③の分類) - (期間④の分類) - (期間⑤の分類)

各セール周辺期間において、データ数が多い順番に

¹ スクレイピングには Beautiful Soup 4.10.0 を利用。

² Rakuten Developers, “楽天商品検索API (version:2017-07-06)”, API詳細。
<https://webservice.rakuten.co.jp/documentation/jchiba-item-search>

³ Rakuten, “楽天ポイント セール情報まとめ”, Rakuten カレンダー

(参照 2022-11-23)。

⁴ <https://calendar.rakuten.co.jp/cal/8196>

tslearn 0.5.3.2 を利用。
<https://pypi.org/project/tslearn/>

抽出した全商品時系列データの 70%に含まれる商品の時系列データ群を抽出し、クラスタリングを行った結果をクラスタ内時系列の平均時系列として、図 2 に示す。なお、どのセールについても「stay-stay-stay-stay」変動パターンが一番多く、全体時系列の約 8 割を占めており、分析から除外している。「stay-stay-stay-stay」の次に多い変動パターンは、「up-stay-stay-stay」変動パターンである。

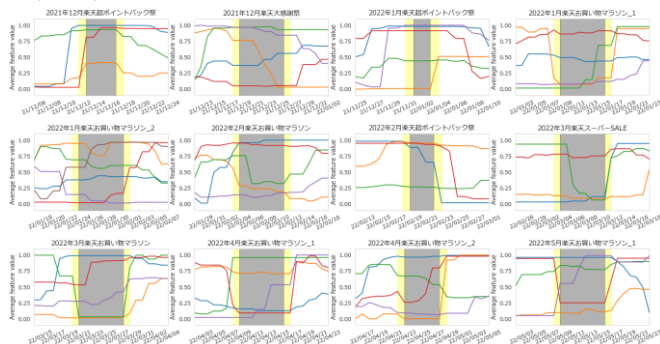


図 2. 各セール周辺期間のデータ数上位の価格変動分割パターンの各クラスタ内時系列データの平均時系列。背景色が白の部分は 5. 1 に示した期間①と期間⑤、黄色の部分は期間②と期間④、灰色の部分は期間③を示す。

クラスタリング結果より以下のことがわかる。

- セール周辺期間で価格変動がある商品については、ほとんどのセールにおいて、セール前 1 週間に価格が上がるクラスタが存在する。
- 図 2 には表現されていないが、各セール周辺期間での価格変動において、セール前 1 週間に価格が上昇するクラスタ内の商品数が最も多い。その他の期間において特異なパターンは発見できない。

上記の結果から、データ数の多い変動クラスタに対して、ポイント変動を調査するために、各セールにおいて「up-stay-stay-stay」に分類された商品のポイント倍率時系列に対してクラスタリングを行った結果を図 3 に示す。ここで、クラスタ内のデータが少ないクラスタ（データ数 4 未満のクラスタ）は外れ値として扱い、グラフから除外している。

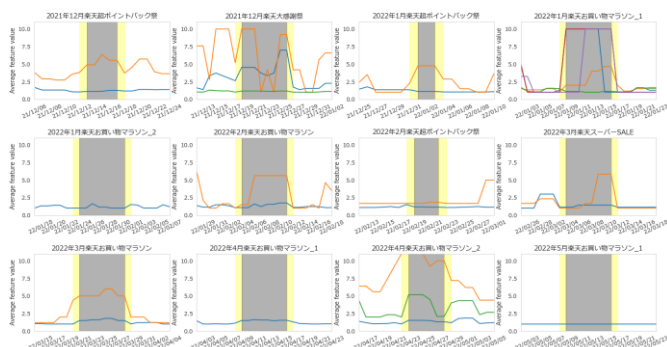


図 3. 各セール周辺期間の価格変動パターンにおいて「up-stay-stay-stay」に分類された商品のポイント倍率時系列クラスタリング結果。各クラスタ内時系列データの平均時系列を示す。

図 3 のクラスタリング結果より、多くのセールで、セール期間にポイント倍率が高くなっているクラスタが存在することがわかる。これは、期間①で価格が上昇した商品群に対し、セール期間③でポイント倍率が上がった商品が含まれていると予想される。つまり、ポイント倍率の変化と価格の変化に何らかの関連性があると考えられる。

ここで、事前実験の結果を基に、RQ1「EC モードで活用されているポイントと価格の変化に関連性はあるか？」に回答する。

特徴的な価格変動パターンを持つ商品群のポイント倍率が、セール期間に高くなる傾向が見られたため、価格とポイント倍率には関連性があると判断できる。セール周辺期間では、ポイント倍率の変動に伴って、出店事業者が特徴的な価格設定を行う可能性がある。

6. 商品購入タイミング予測手法の提案

事前実験の結果を踏まえ、対象商品の安価タイミング予測モデルを提案する。価格とポイント倍率を使用し、商品ごとの付与ポイントを求め、価格からポイントを引いた（価格-ポイント）の時系列データを生成する。本データをもとに、のべ商品数 7,200 件の短期間時系列データを用いて学習モデルを構築する。

6.1 問題設定

本稿は、消費者がセール周辺期間で商品の購入タイミングを判断する一助となるために、お得に商品を購入できるタイミングを予測することを目指している。そのため、（価格-ポイント）が安価になる期間を予測することが本手法の目的となる。消費者が実際に利用することを想定した場合、「予測に用いる時間依存データを長期的に集めることは、消費者への負荷となる」ため、短期間での時系列データを用いた予測を行う。

特徴量として用いる時系列データの期間はセール開始日 1 週間前からセール開始日の合計 8 日間とし、以下の 3 期間のうち、いつ購入するとお得に商品を購入できる可能性が高いか、それ以外を判定する単純な 4 クラス分類問題として扱う。

予測対象とする 3 期間

期間 A：現在のセール期間

期間 B：セール終了後 3 日間

期間 C：セール終了後 4 日間～セール終了後 1 週間

クラスの定義

クラス 1. 期間 A でお得に商品を購入できる可能性が高い。

クラス 2. 期間 B でお得に商品を購入できる可能性が高い。

クラス 3. 期間 C でお得に商品を購入できる可能性が高い。

クラス4. お得に商品を購入できる期間を予測できない。(変動がない商品であり何れの期間も同一)

分類タスクに用いる特徴量

(価格-ポイント) 及び価格の時系列データについて、商品間の価格差を統一するためにセール開始日の値を基準値とし、基準値を 100 とした値に正規化する。提案手法では、事前実験の結果を踏まえて(価格-ポイント)の安価購入期間の分類予測に、以下の3つの多変量時系列を用いる。

- (価格-ポイント)の時系列データ
- 価格の時系列データ
- ポイント倍率の時系列データ

評価実験では、(価格-ポイント)の時系列データのみを特徴量として用いた場合、(価格-ポイント)の時系列データと価格の時系列データを特徴量として用いた場合、(価格-ポイント)の時系列データとポイント倍率の時系列データを特徴量として用いた場合の3種類の特徴量を対象として比較する。

6.2 ラベリング手法

4クラス分類を行うために、時系列データのラベリングを行う。6.1で述べた3期間で全く変動のないデータをクラス4へラベリングし、それ以外のデータについて、以下の4つのフェーズで正解データ(予測値)の生成を行う。正解データは one-hot ベクトルで表現する。

フェーズ 1. 正解データとして、要素が全て 0 で初期化された(データ数×クラス数)の2次元配列を生成する。

フェーズ 2. データごとに3期間で最も安価である日を含む期間のクラス(複数存在する場合もある)に1を割り当てる(1が正解となる期間を表す)。

フェーズ 3. 正解データ内で1が重複している列(データ)に対して、正解となる1を一つに絞る。具体的には、当該期間内のデータ変動の最大値が最も小さい期間のみを1のままとし、他の期間を0とする(等しい場合は1のまま)。

フェーズ 4. フェーズ3後も正解データ内で1が重複している列(データ)がある場合、すでに1が割り当てられている期間のうち、ランダムで1つの期間のみを1とし他を0とする。

例として、セール周辺期間における、ある3商品(item_a,b,c)の(価格-ポイント)変動値を用いて説明する(図4)。背景色が黄色の部分(期間A)、灰色の部分(期間B)、緑色の部分(期間C)を示している。黒縦線はセール開始日とセール終了日を示す。また、各商品のA~C各期間内の(価格-ポイント)変動値の最小値と最大値を以下の表1に示す。

はじめに、商品ごとに全期間の最小値を持つ期間を正解とする(フェーズ2)。表1より全期間の最小値を

持つ期間を1つに絞れる商品はitem_c(期間Aが最小)のみであることがわかる。他2商品は、全期間の最小値を持つ期間が複数個存在するため、これらの期間で優劣をつける必要がある。

そこで、フェーズ3では、フェーズ2で抽出した「全期間の最小値を持つ期間」を比較対象として、各期間内の(価格-ポイント)変動値の最大値が最も小さい期間を正解とする。item_aは3期間すべての比較、item_bは、期間Bと期間Cの比較を行う。表1よりitem_bは期間Cの最大値の方が期間Bの最大値より小さいため、クラス3へ分類される。一方、item_aは、期間Aと期間Cの最大値が最も小さく、等しい。このため、フェーズ4に進み、ランダムにクラスを割り当てる。ただし、フェーズ2及び3の過程で正解とされなかった期間は除外する。このランダムの割り当て処理は、フェーズ4まで残った期間はどれも購入タイミングに適していると判断したためである。

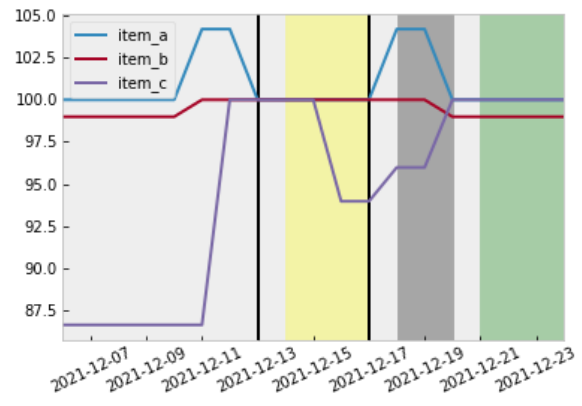


図4. セール周辺期間における、ある3商品の(価格-ポイント)変動値

表1 図4の商品例における期間A~Cの(価格-ポイント)変動値の最小値と最大値

	期間 A		期間 B		期間 C	
	min	max	min	max	min	max
item_a	100	100	100	105	100	100
item_b	100	100	99.0	100	99.0	99.0
item_c	94.0	100	96.0	100	100	100

6.3 データの前処理(正規化)

6.1では、各特徴量における商品間の正規化を行った。本項では、用いる特徴量間での正規化を行う。これは、多変量時系列を特徴量とする場合、特徴量間での正規化が必要となるためである。のべ商品数7,200件の短期間時系列ごとに、セール周辺期間中における(価格-ポイント)[%]、価格[%]、ポイント倍率の最大値と最小値を図5に示す。図5のX軸原点は、セール期間1週間前を示し、価格のY軸はセール開始日を100と正規化している。

- (価格-ポイント)の正規化

短期間時系列内での最大値が150%を超える商品は全体の約0.53%、最小値が150%を超える商品は全体の

約 0.22%であった。外れ値を補完するため、値が 150%以上の商品は 150 に、50%以下の商品は 50 に置換する。その後、全短期間時系列を対象に 0~1 の Min-max Normalization を行う。

- 価格の正規化

図 5 に示される通り、価格の最大値、最小値の関係も(価格-ポイント)とほぼ一定であるため、値が 150%以上の商品は 150 に、50%以下の商品は 50 に置換する。その後、全短期間時系列を対象に 0~1 の Min-max Normalization を行う。

- ポイント倍率の正規化

全短期間時系列の最小値が 1, 最大値が 20 であったため、そのまま全短期間時系列を対象に 0~1 の Min-max Normalization を行う。

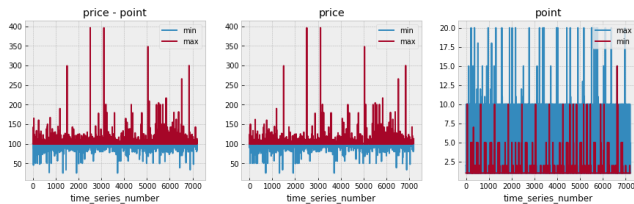


図 5. 短期間時系列ごとの(価格-ポイント) [%]、価格 [%]、ポイント倍率の最大値と最小値

6.4 時系列分類モデル

提案手法の概要を図 6 に示す。図 6 に示す通り、セール 1 週間前~セール開始日の期間における多変量時系列データは、6.3 で述べた正規化を適用し、セール開始日の次の日~セール後 1 週間の期間における(価格-ポイント)の時系列データを 6.2 で述べたラベリング手法を用い、one-hot ベクトルに変換する。学習用データセットは全 7,200 個のデータセットを 2/3 使用し、残りの 1/3 のデータセットをテストとする。

ラベリングの結果、クラスごとのデータ数は表 2 に示す分布となった。

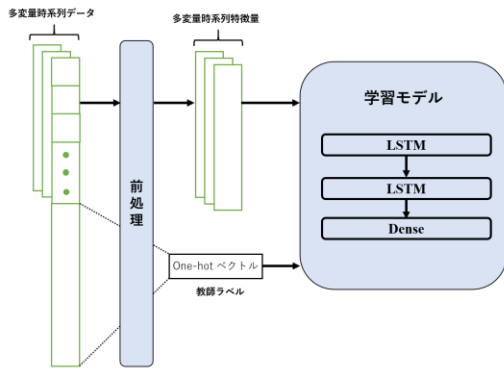


図 6. 提案手法の概要図

表 2. 学習データとテストデータのクラス別データ数.

	class1	class2	class3	class4
学習データ	1,092	878	613	2,217
テストデータ	565	462	340	1,033

学習モデルについて、1 層目の LSTM 層の入力サイズは (8, 3) となっており、隠れ状態の数は 128 であり、出力をシーケンスで返す。2 層目の LSTM 層の入力サイズは (8, 128) となっており、隠れ状態の数は 32 であり、出力をシーケンスで返さない。活性化関数はどちらも ReLU を適用している。全結合層ではソフトマックス関数を適用して、4 クラスの確率を算出する。オプティマイザには Adam を使用する。学習のバッチサイズは 128 に、エポック数は 200 に設定し、学習用データセットのうち 10%を検証用データに用いる。

モデル構築には、Python ライブラリの Keras⁵と呼ばれるニューラルネットワークを用いた。

7. 評価実験

7.1 評価方法

(価格-ポイント)、価格、ポイント倍率の 3 つの時系列データの特徴量として用いた学習モデルを $M_{pr,po}$ 、(価格-ポイント) と価格の 2 つの時系列データの特徴量として用いた学習モデルを M_{pr} 、(価格-ポイント) とポイント倍率の 2 つの時系列データの特徴量として用いた学習モデルを M_{po} とした。

作成したモデルについては、混同行列から一般的にモデル評価に用いられる Recall (再現率), Precision (適合率), F1-score (調和平均), Accuracy (正解率) を算出して評価を行う。

7.2 実験結果

$M_{pr,po}$, M_{po} , M_{pr} の実験結果について表 3~8 に示す。

表 3. $M_{pr,po}$ モデルの性能評価結果 (混合行列)

予測値	class1	class2	class 3	class 4
class1	357	40	22	146
class2	151	85	18	208
class3	102	38	31	169
class4	20	27	18	968

表 4. $M_{pr,po}$ モデルの性能評価結果

	class1	class2	class 3	class 4
recall	0.632	0.184	0.091	0.937
precision	0.567	0.447	0.348	0.649
F1-score	0.597	0.261	0.145	0.767
accuracy	0.601			

表 5. M_{pr} モデルの性能評価結果 (混合行列)

予測値	class1	class2	class 3	class 4
class1	44	0	10	511
class2	14	2	16	430
class3	11	6	17	306
class4	1	3	13	1,016

⁵ keras 2.11.0

<https://pypi.org/project/keras/>

表 6. M_{pr} モデルの性能評価結果

	class1	class2	class 3	class 4
recall	0.078	0.004	0.050	0.984
precision	0.629	0.182	0.304	0.449
F1-score	0.139	0.009	0.086	0.617
accuracy	0.450			

表 7. M_{po} モデルの性能評価結果 (混合行列)

	予測値	class1	class2	class 3	class 4
真 値	class1	361	23	30	151
	class2	147	74	27	214
	class3	102	35	37	166
	class4	22	17	33	961

表 8. M_{po} モデルの性能評価結果

	class1	class2	class 3	class 4
recall	0.639	0.160	0.109	0.930
precision	0.571	0.497	0.291	0.644
F-score	0.603	0.242	0.158	0.761
accuracy	0.597			

上記の結果から、(価格-ポイント)、価格、ポイント倍率の時系列データを3つの特徴量を用いた場合、最も高い Accuracy (0.601) となることが分かった。

7.3 考察

実験結果をもとに、**RQ2**「セール周辺期間の短期間時系列を用いて、どの程度の精度で商品の安価タイミング予測が可能か?」、**RQ3**「学習に使用する特徴量によって予測精度に差異はあるか?」に回答する。

まず、**RQ2**については、0.601の正解率で安価タイミングを推定できる(4クラス分類タスク)ことを確認した。一方で、クラス1「現在のセール期間での購入が安価である」とクラス4「何れの期間でも価格は変動しない」のF1スコアはそれぞれ、0.597、0.767であるのに対し、クラス2「セール終了後3日間」及びクラス3「セール終了後4日間~セール終了後1週間」の判定では、F1スコアが0.261、0.145と低く正しく予測できているという結論には至らなかった。

そこで、クラス2とクラス3を「セール終了後1週間」という一つのクラスにまとめた3クラス分類モデルの性能評価を行った(クラス2-3と表す)。

表 9. $M_{pr po}$ モデルの性能評価結果 (3クラス分類モデル)

	class1	class2-3	class 4
recall	0.6	0.305	0.903
precision	0.577	0.592	0.667
F1-score	0.588	0.403	0.768
accuracy	0.632		

上記の結果から、「セール終了後1週間」にまとめた3クラス分類モデルも精度は大きく向上しなかった。

次に、**RQ3**については、ポイント倍率時系列データの特徴量に追加することで、絶対値で0.151の正解率向上を確認できた($M_{pr po}$ と M_{po} との比較)。このことから、過去のポイント倍率は、安価タイミング予測に有効な特徴量であることが考えられる。

8. おわりに

本研究では、消費者が実際にモデルを扱うことを想定して、短期間時系列データ(セール直近7日間とセール初日の価格及びポイント変動データ)を用いて、セール商品の安価タイミング予測を行った。予測期間を分割して4クラス分類の問題としてモデルの提案を行ったところ、0.601の正解率を達成した。また、ポイント倍率データがセール周辺の安価タイミング予測の特徴量として有効であることを示した。これは、セール周辺期間において価格の変動とポイント倍率との間に関係性があることを示している。また、ポイント変動は、価格変動に比較しセール期間内での変動が大きいことも分かった。

今後は、商品カテゴリなどの特徴量の追加やラベリング手法の再検討、季節性を考慮した予測などにより、正解率のさらなる向上を目指したい。なお、商品カテゴリの利用については、国立情報学研究所を通じて配布される「楽天データセット」の内、楽天市場データに含まれる商品カテゴリ分類等を参考としたい。

参 考 文 献

- [1] Dinerstein, M., Einav, L., Levin, J., and Sundaresan, N., "Consumer price search and platform design in Internet commerce," *American Economic Review*, vol.108, no.7, pp.1820-1859, 2018.
- [2] Zheng, X., Men, J., Yang, F., and Gong, X., "Understanding impulse buying in mobile commerce: An investigation into hedonic and utilitarian browsing," *Int. J. of Information Management*, vol.48, pp.151-160, 2019.
- [3] 田部溪哉, ECサイト利用者の知覚価値によるロイヤルティ形成-ウェブサイトへの満足の媒介効果に焦点を合わせ-, 城西大学経営紀要, 第14号, pp. 63-79, 2018.
- [4] 寺地一浩, ポイントベネフィットに対する消費者行動研究. 大阪市立大学大学院創造都市研究科, 博士学位論文, 2016.
- [5] Tseng, KK., Lin, RY., Zhou, H., Kurniajaya, KJ., and Li, Q., "Price prediction of e-commerce products through Internet sentiment analysis," *Electronic Commerce Research*, vol.18, pp. 65-88. 2017.
- [6] Carta, S., Medda, A., Pili, A., Reforgiato Recupero, D., and Saia, R., "Forecasting e-commerce products prices by combining an autoregressive integrated moving average (ARIMA) model and google trends data," *Future Internet*, vol.11, no.1, 5, pp. 1-19, 2019.
- [7] Fathalla, A., Salah, A., Li, K., Li, K., and Francesco, P., "Deep end-to-end learning for price prediction of second-hand items," *Knowl. Inf. Syst.*, vol.62, pp. 4541-4568, 2020.
- [8] Bakir, H., Chniti, G., and Zaher, H., "E-Commerce Price Forecasting Using LSTM Neural Networks," *Int. J. Mach. Learn. Comput*, vol.8, pp.169-174, 2018.
- [9] Kumar, A., "A conceptual model for automation of product dynamic pricing and sales promotion for a retail organization," *Kybernetes*, vol.45, no.3, pp. 490-507. 2016.