

# 販売履歴データに基づく中古ファッションアイテムの 出品価格推定モデルの提案

桑田 和<sup>1,a)</sup> 杉崎 智哉<sup>1,b)</sup> 三川 健太<sup>2,c)</sup> 後藤 正幸<sup>1,d)</sup>

受付日 2020年5月20日, 採録日 2020年10月6日

**概要:** 本研究では中古ファッション EC サイトにおける出品価格の決定問題を対象とする。このサイトではユーザから買い取ったアイテムを中古品として再販売しているが、その買取価格はアイテムの出品価格と連動して決定されるため、適切な出品価格の決定が重要な課題となっている。さらに、このサイトでは一定期間売れなかったアイテムの出品価格を値下げする仕組みにより、出品価格のまま購入される場合と値下げされて購入される場合の2パターンが存在する。値下げにより、在庫コストを低減できる反面、収益性の観点からは出品価格のまま購入されることが望ましい。しかし、出品価格のまま購入されたアイテムはさらに高い出品価格でも購入された可能性があり、値下げされたアイテムは仮に出品価格を低くしていれば出品価格のまま購入された可能性がある。そのため、販売履歴における購入時の価格が適切とは限らず、適切な出品価格を推定する際に考慮が必要である。そこで、本研究ではアイテムが値下がりせずに購入される最大出品価格の推定方法を提案する。提案手法では、出品価格のまま購入されるか否かの二値分類器を構築する。そして、出品価格のまま購入されると予測されたアイテムは出品価格を段階的に上げながら、値下げされて購入されると予測されたアイテムは出品価格を段階的に下げながら適切な出品価格を推定する。最後に、このサイトの販売履歴データを用いて提案手法の有効性を示す。

キーワード：中古ファッションアイテム, EC サイト, 出品価格推定, 機械学習, 販売履歴データ

## An Estimation Model of Exhibits Price for Second-hand Fashion Items Based on Sales History Data

IZUMI KUWATA<sup>1,a)</sup> TOMOYA SUGISAKI<sup>1,b)</sup> KENTA MIKAWA<sup>2,c)</sup> MASAYUKI GOTO<sup>1,d)</sup>

Received: May 20, 2020, Accepted: October 6, 2020

**Abstract:** In this study, we focus on a problem of pricing used items on a second-hand fashion EC site. The buying price of a used fashion item is determined according to its open price for selling on the site. Therefore, the decision of an appropriate open price is so important. In addition, this site automatically marks down open prices of items that have not been purchased for a certain period of time step by step. From the viewpoint of business model, high open price is desirable to earn more profits but the marking down should be avoided. For that reason, it is difficult to determine the proper open price. Therefore, we proposed a model to estimate the maximum open price under avoiding a markdown for each item. In our proposed model, we construct a binary classifier predicting if an item is purchased for open price or with marking down. The proposed model can estimate the proper price with raising the open price of non-marking down items and lowering the open price of marking down items little by little. In order to evaluate our proposed model, we demonstrate analytical experiments by using actual purchasing history data on a fashion EC site.

**Keywords:** second-hand fashion item, EC site, estimation of exhibits price, machine learning, sales history data

<sup>1</sup> 早稲田大学  
Waseda University, Shinjuku, Tokyo 169-8555, Japan

<sup>2</sup> 湘南工科大学  
Shonan Institute of Technology, Fujisawa, Kanagawa 251-8511, Japan

a) ikdsk14243150@toki.waseda.jp

b) sugi5524@ruri.waseda.jp

c) mikawa@info.shonan-it.ac.jp

d) masagoto@waseda.jp

## 1. はじめに

近年、リユース商品（以下、アイテム）を購入するユーザが増加し、リユース市場は拡大の一途をたどっている。なかでも、ECサイトにおけるリユース市場は大きな割合を占めており、膨大な量のアイテムの販売履歴データが蓄積されている。一方、リユース事業を展開する企業の多くは、ユーザから提供される中古アイテムに対し、出品者が買取価格と出品価格を決定し、再販売を行う。その際、買取価格は、収益性や売れ残りリスクなどの観点から出品価格と連動して決められることが多い。そのため、多様なリユース商品に対して適正な出品価格を決定することはビジネス上大変重要な意思決定であるといえる。以下、本研究では出品価格を、アイテムをECサイト上で出品する際の初期価格と定義する。この出品価格は一般的に出品者の経験的な意思決定に基づいて行われており、必ずしも最適な価格決定がされているとは限らないため、最適な出品価格の決定は重要な課題の1つである。

一方、リユース市場では、商品の売れ残りを防ぐために値下げの仕組みを導入している企業も多い。値下げを導入する場合、収益性の観点からは、出品されてから誰かが購入するまでの期間（以下、掲載期間）の長期化にともなう在庫費の増加や、アイテムの値下げにともなう売上額の減少を避けるため、可能な限り値下げをせずに早い段階で購入されることが望ましい。しかしながら、アイテム購入の意思決定はユーザ側にあるため、個々のアイテムによって特徴が異なるような多様性の高いアイテムでは、実際に出品して購入されるまで、値下げが必要になってしまうか否かという結果を予想できない場合も多い。

ECサイトを運営する企業では過去に出品し、購入されたアイテムの出品価格や購入された際の価格といった販売履歴データが膨大に蓄積されている。そこで、それらのデータを分析することが各アイテムに対する適切な出品価格の決定につながると考える。しかしながら、蓄積されている販売履歴データには、「出品後間もなく出品価格で売れたアイテム」と、「掲載されている期間が長くなったために、出品価格から値下げして売れたアイテム」の2パターンの販売結果が混在している。前者のアイテムは、出品価格をさらに高く設定していたとしても、掲載期間がそれほど伸びずに値下げをとまわず購入され、より多くの売上げが期待できた可能性がある。一方、後者のアイテムは、出品価格を低く設定することにより掲載期間が短縮され、出品価格で売れることで在庫費の削減が期待できた可能性がある。すなわち、アイテムが購入された時点での販売価格が値下げが起こらない最適な出品価格とは限らず、アイテムが購入される際の販売価格を予測するだけでは、最適な出品価格は決定できない。

そこで本研究では、アイテムが値下がりせずに購入され

る出品価格の最大値、すなわち、「それより高い出品価格であると売れずに値下がりしてしまい、それ以下であれば出品価格のまま売れるような価格」が個々のアイテムに対して存在すると仮定する。そして、この閾値価格を推定するために、機械学習モデルを用いてアイテムの出品価格のみを変化させた場合の販売結果を予測することで、そのアイテムの値下がりが起こらない出品価格の最大値を推定する手法を提案する。この価格を推定することができれば、最適な出品価格の決定の一助となる可能性がある。提案手法でははじめに、過去の販売履歴データによる機械学習モデルを用いて、出品価格で売れるか、値下げしないと売れないかの2パターンの販売結果を判別する分類器を構築する。ここでは、高精度な分類器の1つである Random Forests (RF) [1] を用いて2値分類器を構築し、他のいくつかの機械学習モデルと比較してその優位性を示す。次に、構築した分類器を用いて、アイテムの出品価格のみを変化させた際の販売結果の予測値を分析し、元の販売結果とは異なる結果が得られる閾値の出品価格を求める。この方法により、値下げが起こらない範囲での出品価格の最大値を推定することができ、実際のアイテムの出品価格決定の基準値となることが期待できる。本研究では、某ファッションECサイト（以下、サイトA）の販売履歴データに提案手法を適用し、分析結果に対する考察を与える。

## 2. 関連研究

### 2.1 アパレル業界を対象とした研究

アパレル業界におけるデータ活用に関する研究は、売上予測や需要予測を目的としたものが多く、様々な手法が用いられている [2]。Thomasseyら [3]、Niら [4] は、ファッションアイテムの特徴である多様性やトレンド傾向、季節性といった要因を考慮した売上予測について研究している。前者はクラスタリング手法と決定木を用いて販売履歴データに基づく中期的な売上予測を、後者は購買履歴データと外部要因に基づく長期売上予測と短期売上予測を組み合わせた2段階予測を行っている。一方、ファッションアイテムの売上予測を実用化するために、モデルの計算コストの観点から研究も報告されている。たとえば、Auら [5] は遺伝的アルゴリズムを用いたニューラルネットワークによる短期的な売上予測を行っており、Yuら [6] は extreme learning machine と統計的手法を用いた、実応用を目的とした売上予測の高速化を行っている。さらに、Aksoyら [7] の適応型ネットワークに基づくファジィ推論システムを用いた需要予測や、Choiら [8] の SARIMA とウェーブレット変換を組み合わせた手法による販売時系列データに基づく売上予測など、多様な手法を用いた研究が行われている。

これらの手法は売上予測に終始しており、機械学習を用いた価格決定に焦点を当てた研究は少なく、議論の余地がある。そのため、アイテムの出品価格の決定についての議

論は応用上、大きな意義を持つと考えられる。

## 2.2 価格設定に関する研究

マーケティング分野において、製品やサービスの価格決定は企業の利益獲得のための重要な手段であり、これに関する研究は多くなされている [9]。これらの研究から、価格決定には様々な要因が影響を与えることが分かっており、その 1 つが消費者の価格意識の複雑さである。Erickson ら [10] は、製品に対する消費者の評価は、負の直接効果や正の間接効果に影響を受けることを主張している。負の直接効果とは、金銭的な負担であり、価格が高くなるほど購入されにくくなることを意味する。一方、正の間接効果とは、品質に関するポジティブな評価であり、価格が高くなるほど消費者は高品質であると評価することを意味する。また、Tellis ら [11] は、企業の目的と消費者の特性という 2 つの観点から、価格戦略を 9 つに分類している。企業の目的という観点では、値引きなどを用いて消費者のセグメントごとに異なった価格を提示する価格決定、競争的地位を獲得するための価格設定、製品ラインを持つ企業による価格決定の 3 つが存在する。一方、消費者の特性は、より安価な製品や高品質な製品の探索に高いコストをかける消費者、製品に対する上限価格が低い消費者、移動コストや投資リスクといった取引コストを持つ消費者の 3 つが存在すると主張している。

このようなマーケティング分野における価格決定問題に加えて、製品の価格決定後、価格変更を行う際の戦略に関する研究も行われている [12]。アイテムの価格を自動的に値下げする仕組みを採用している場合、価格変更の戦略もふまえて出品価格の決定を行うことが望ましいと考えられる。しかしながら、その場合、問題設定が複雑化してしまうという問題点が生じる。そこで本研究では、値下げが起こる前にアイテムが購入されるための価格決定を目的とする。さらに、機械学習を用いて動的に変動する需要をとらえた価格決定を行う、ダイナミックプライシングに関する研究も多くなされている。Cheung ら [13] は価格変更回数を考慮して、過去の需要に基づいた商品の価格決定アルゴリズムを提案している。このようにダイナミックプライシングは一般的に過去の需要量に基づいて時間変化にともなって価格を決定する。しかしながら、本研究で対象とする中古ファッションアイテムという問題設定では、商品の多様性が非常に高く、その需要量は時間変化のほかにもブランドやカテゴリといった多様な特徴量によっても変化してしまう。このようなほぼ「1 点もの」のアイテム群に対して、過去のデータから個々のアイテムの需要を推定して動的な価格決定に結び付けようとするダイナミックプライシングの仕組みをそのまま適用することは難しい。また、一般的なダイナミックプライシングでは同業他社が設定している価格や流行なども考慮に入れることもある

ため、その膨大なデータを扱う分析システムの導入には大きなコストが生じることが想定される。これに対し本研究では、多様性がきわめて高いファッションアイテムを対象として、自社保有のデータのみを利用した価格決定方法について検討を行うものとする。すなわち、一般的なダイナミックプライシングでとられるアプローチではなく、機械学習を用いて販売結果を予測しながら適切な出品価格を決定する手法を提案する。

一方、過去の販売履歴データを活用した価格決定問題に対する解決策の提案は 1 つの有効なアプローチと考えられる。仁ノ平ら [14] は、本研究と同様に販売履歴データを活用して、設定した出品価格のもとでアイテムを販売した場合の最終的な販売価格を予測するモデルを構築し、実販売データに対する予測精度の面からその有効性を示している。しかしながら、サイト側で設定された出品価格に対する販売価格の予測モデルであり、アイテムの販売結果は出品価格の設定によって異なる可能性がある。そのため、購買履歴データにおける最終的な販売価格が適切な出品価格とは限らない。そこで、本研究では各アイテムの出品価格を変更しながら販売結果を予測することで適切な出品価格を決定する方法を提案する。

## 3. 対象問題

### 3.1 問題設定

本研究ではユーザからアイテムを買い取り、中古品として再販する事業を展開するサイト A における出品価格の決定問題を対象とする。サイト A では、アイテムの出品価格を高価格にすると、それに比例して購入時の価格も高価格になり、売上げは増加することが多い [14]。しかし、ユーザに購入されにくくなるため、出品から販売までの期間が長くなり、在庫費の増加や売れ残り、サイト上に同一アイテムを長期間掲載することによるサイトの魅力低下といった問題が発生する。一方、アイテムを低価格で出品した場合、多くのアイテムはすぐにユーザに購入される反面、より高価格で購入される機会を失う。そのため、掲載期間と売上げの双方を考慮した出品価格の推定が望まれる。

過去に出品され、購入されたアイテムには、前述のように「出品後間もなく出品価格で売れたケース」と「掲載されている期間が長くなったために、出品価格から値下げして売れたケース」の 2 パターンの販売結果が存在する。出品価格で売れたアイテムは、より高い出品価格であってもユーザの購買意欲やアイテムの価値により、その価格で購入され、掲載期間に大きな変化が生じなかった可能性がある。一方、値下げして売れたアイテムは、より低い出品価格であればアイテムに対してユーザが感じる価値より価格が下回り、より早く購入され、掲載期間が短縮された可能性がある。このように、サイト A で扱われるアイテムは出品価格によって販売結果が異なるため、単に過去の履歴

表 1 分析データの詳細

Table 1 Details of the data to be analyzed.

出品期間	2017年1月1日～2018年12月31日
販売期間	2018年1月1日～2018年12月31日
データ数	3,080,621件

表 2 オフ率の内訳

Table 2 Rate of price off rate.

オフ率	データ数 (個)	割合 (%)
0	953,104	30.94
10	158,897	5.16
20	130,037	4.22
30	217,177	7.05
40	144,037	4.68
50	290,379	9.43
60	232,441	7.55
70	711,495	23.10
80	123,935	4.02
90	119,119	3.87
計	3,080,621	100.00

データから販売結果を予測するモデルを構築するだけでは適切な出品価格を推定することはできない。したがって、アイテムの販売結果の差異を考慮に入れて、適切な出品価格を推定する方法を考える必要がある。そこで本研究では、出品価格からの値下がり率（以下、オフ率）と掲載期間を用いて販売結果の判別モデルを構築する。次に、予測される販売結果を基に、値下がりしない出品価格の最大値を推定する。これにより、個々のアイテムに対する値下げを可能な限り回避しつつ、収益性の高い出品価格を見積もることが可能となる。

### 3.2 事前分析

本研究で対象とするデータには、あるアイテムが買い取られてからユーザに購入されるまでの情報が蓄積されており、出品したアイテムの種類やブランド、販売時期や価格およびそのときのオフ率などといった詳細な情報まで得ることができる。本節では、対象データの特徴を把握するため、オフ率に関する分析、掲載期間に関する分析および販売結果に関する分析を行う。分析に用いた、過去の販売履歴データの詳細を表 1 に示す。

はじめに、過去にサイト A で出品されたアイテムがどの程度値下げされて購入されたのかを調べるため、アイテムのオフ率に関する分析を行う。サイト A では、10%ずつオフ率を変化させている。全アイテムにおける各オフ率の割合を表 2 に示す。

表 2 より、オフ率 0% で購入されるアイテムが最も多く、全体の約 3 割を占めていることが分かる。次いで、オフ率 70% で購入されるアイテムが多く、全体の約 2 割を占めていることが分かる。0%、70% 以外の各オフ率では、所属

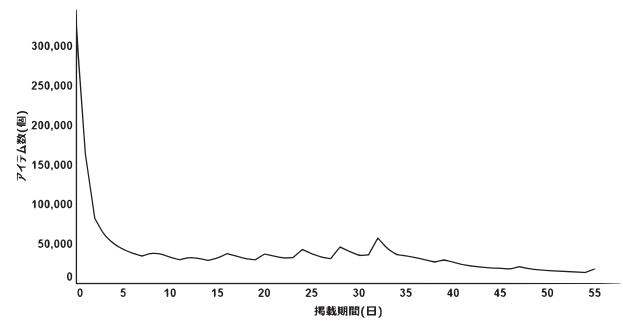


図 1 掲載期間

Fig. 1 The sales periods for all items.

するアイテムの割合が 3%~9% 程度で大きく変化していない。このことからサイト A で扱われるアイテムは、オフ率が 0% の販売結果となるアイテムと、値下げされてオフ率 0% 以外の販売結果となるアイテムの 2 種類のアイテムに分類できると考えられる。

次に、過去にサイト A で出品されたアイテムが出品されてから購入されるまでに要した期間を調べるため、アイテムの掲載期間に関する分析を行う。ここでは、全アイテムにおける各掲載期間がどのような分布になっているかを分析した結果を図 1 に示す。ただし、アイテムが多く販売されている 0 日~55 日を対象とした。図 1 より、購入アイテム数が最も多いのは掲載期間 0 日であることが分かった。また、アイテム数は 0 日から減少し、7 日と 8 日を境に増加傾向に転じ、傾向が変化している。すなわち、サイト A で扱われるアイテムは、掲載期間 8 日未満の早い段階で購入されてしまうアイテムと、8 日以降と出品して一定期間経過した後に値下げされてから購入されるアイテムの 2 種類のアイテム群が混ざっていることがうかがえる。

最後に、前述の 2 パターンの販売結果の境界を考えるため、出品価格の決定に大きく依存している要因の 1 つであるブランドに着目して、出品価格で売れるアイテムと値下げしないと売れないアイテムの特徴について分析した結果を示す。サイト A では多様なブランドを扱っているため、表 3 のように 8 つのブランドクラスに分類し、管理している。ブランドの管理において、その類似性に最も強く寄与している要因は価格帯であるため、ブランドクラスは価格帯を中心として分類されている\*1。ただし、クラス 8 は在庫期間を短くしたいアイテムとして、戦略的な観点から人為的に選別したブランドクラスである。なお、これらのブランドクラスは、前年の各ブランドと使用状態に基づいたアイテムの平均販売単価により、クラスを作成している。したがって、各ブランドの全アイテムが複数のクラスに所

\*1 平均販売単価によってカテゴライズしたブランドクラスは、サイト A で実際にアイテムの管理のために利用されているため、本研究でもこのブランドクラスを分析に用いる。ただし、平均販売単価によってカテゴライズされているため、以後の出品価格や販売価格の分析の際には、価格帯による選択バイアスが生じることを考慮に入れた検討を行う。

表 3 価格帯によるブランドクラス分類条件

Table 3 Conditions for classifying items into brand clusters.

クラス	条件
1	アイテムの価値が落ちない (人為的に選別)
2	平均販売単価 10,001 円～
3	平均販売単価 6,001 円～10,000 円
4	平均販売単価 4,001 円～6,000 円
5	平均販売単価 2,001 円～4,000 円
6	平均販売単価 1,001 円～2,000 円
7	平均販売単価～1,000 円
8	平均販売単価 4,000 円～かつ 平均掲載期間～15 日

表 4 価格帯ブランドクラスに対する販売結果別の平均出品価格 (円)

Table 4 Average open price in brand cluster for each sales result (JPY).

クラス	出品価格で売れるアイテム	値下げしないと売れないアイテム
1	6,198	<b>7,030</b>
2	16,542	<b>21,945</b>
3	10,590	<b>13,535</b>
4	7,311	<b>9,207</b>
5	4,500	<b>5,641</b>
6	3,037	<b>3,709</b>
7	2,580	<b>3,017</b>
8	3,244	<b>4,262</b>

太字は 2 つのアイテム群を比較して大きい値

属する場合がある。

上記のオフ率および、掲載期間に関する分析結果より、「出品価格で売れるアイテム」をオフ率 0%かつ掲載期間 8 日未満で購入されたもの、「値下げしないと売れないアイテム」をオフ率 0%以外かつ掲載期間 8 日以降で購入されたものとそれぞれ定義する。各販売結果の、各価格帯ブランドクラスにおける出品価格の平均値を表 4 に示す。

表 4 より、すべてのクラスにおいて、その平均出品価格は、値下げしないと売れないアイテムの方が高い。このことから、同じような価格帯のアイテムでも、出品価格で売れるアイテムはその出品価格をより高く設定することで、値下げしないと売れないアイテムとして販売された可能性があり、値下げしないと売れないアイテムはその出品価格をより低く設定することで、出品価格で売れるアイテムとして販売された可能性があるといえる。したがって、値下がりしない最大出品価格は、前者の出品価格をより高く、後者の出品価格をより低くすることで得られると考えられる。

## 4. 提案手法

### 4.1 着想

サイト A で出品されるアイテムは、ここまで述べてきたとおり、出品価格によって 2 パターンの販売結果になると考えられる。出品価格で売れるアイテムは、ユーザに「そ

の出品価格は妥当、または安価である」と判断されたアイテムである。それらをより高い価格で出品することを想定し出品価格を徐々に上げていくと、ある価格を境に出品価格のままでは売れなくなり、値下げ後に買われるようになると考えられる。一方、値下げしないと売れないアイテムは、ユーザに「出品価格が高く、妥当でない」と判断されたアイテムである。それらをより低い価格で出品することを想定し出品価格を徐々に下げていくと、ある価格を境に値下げの必要がなくなり、出品価格で買われることが予想される。すなわち、アイテムが値下がりせずに購入される出品価格の最大値は、2 パターンの販売結果の境界に存在すると考えられる。

そこで本研究では、それより高い出品価格であると値下がりにつながり、それ以下であると出品価格で売れるような価格が個々のアイテムに対して存在すると仮定し、この閾値価格を推定する方法を提案する。ここで、この閾値価格は「アイテムが値下がりせずに購入されるような出品価格の最大値」と定義する。具体的には、出品価格で売れると予測されるアイテムに対しては、出品価格を段階的に上げながら販売結果の予測を繰り返し、「値下げしないと売れなくなると予測される価格の直前の価格」を「値下がりしない出品価格の最大値」として推定する。一方、現状の出品価格では値下げしないと売れないと予測されるアイテムに対しては、出品価格を段階的に下げながら販売結果の予測を繰り返し、「出品価格で売れるという予測に変化する際の価格」を「値下がりしない出品価格の最大値」として推定する。これにより、個々のアイテムに対する値下げを可能な限り回避しつつ、収益性の高い出品価格を見積もることが可能となる。

ここで、個々のアイテムに対し、出品価格から販売価格を予測する問題を 1 つの回帰問題としてとらえ、推定された回帰モデルを活用して逆算的に出品価格を決定する方法も考えられる [14]。しかしながら、個々のアイテムの販売価格を予測する回帰モデルを構築することは可能であるが、その予測精度をある程度以上に高めることは非常に困難であるという課題がある。その要因の 1 つに、個々のアイテムの販売価格が離散的に変化するため、誤差として正規ノイズを仮定することが難しいという点があげられる。本研究で対象とする個々のアイテムの販売価格は、EC サイト上に出品された際の出品価格から段階的に値下げされ、購入者が現れた時点での価格が販売価格となる。その際の値下げは連続的に行われるわけではなく、10%OFF、20%OFF のような形で離散的に値下げが行われる。すなわち、たとえば各アイテムの値下げ幅が「5,000 円から 4,000 円」のように 1,000 円単位で行われれば、その間の販売価格はとれないことになるため、必然的に、回帰モデルの予測誤差の標準偏差を 1,000 円以下にすることは難しくなってしまう。この点において、「観測された販売価格は、その期待

値に連続値の誤差が加わったものである」という仮定は厳密には成り立たないため、回帰モデルを適用する場合はこの点を考慮したうえで扱う必要がある。

そこで、本研究では、連続的に変更可能な出品価格を説明変数として、販売結果であるオフ率と掲載期間が離散的に変化することを想定した予測モデルを考える。具体的には、「値下げされるか否か」、「出品してすぐに売れるか否か」といった離散事象を目的変数として、説明変数の出品価格を連続的に変化させることで適正な出品価格を推定するモデルを構築する。

## 4.2 概要

提案手法ではまず、過去に販売されたアイテムの販売履歴データを活用し、機械学習モデルによって「値下げせずに販売できるか否か」を判別する分類器を構築することを考える。しかしその際、アイテムの一部には、戦略的な観点から、すぐに売りたいアイテムや長くサイトに掲載しておきたいアイテムが存在するため、このような戦略的アイテムがデータに混在することを考慮する必要がある。すなわち、一般的なアイテムでは、「小さいオフ率のうち、短い掲載期間で販売される」か、「長い掲載期間となり、大きいオフ率で販売される」かのどちらかである。しかし、「掲載期間は長いにもかかわらずオフ率がきわめて小さいアイテム」、あるいは「短い掲載期間に大きなオフ率で購入されているアイテム」というきわめて戦略的に扱われているアイテムが存在しており、これらを他の通常アイテムと合わせて、機械学習モデルに学習させると外れ値となってしまう、良い影響を及ぼさない。そこで本研究では、そのような戦略的なアイテム以外のアイテムの出品価格を推定するため、販売結果の予測においては、オフ率による判別規則と掲載期間による判別規則の2つの分類器を導入することを考える。すなわち、過去の販売実績データから、オフ率と掲載期間の2つの判別基準に対して分類器を構成し、2つの分類器の両方の判別結果の一致をもって、「出品後間もなく出品価格で売れたケース」と「掲載されている期間が長くなったために、出品価格から値下げして売れたケース」の2パターンの販売結果を判別する判別規則を構築する。これにより、出品価格を設定した新規アイテムに、2つの判別規則を適用することで、一般的な商品のみに対して販売結果を予測することが可能となる。

次に、学習によって構築された分類器を活用し、アイテムが値下がりしないと予測される出品価格の最大値を推定する。そのため、出品価格で売れると予測されるアイテムに対しては、出品価格を段階的に上げながらオフ率の判別規則を適用し、値下げしないと売れないと予測される出品価格を値下がりしない出品価格の最大値と推定する。一方、値下げしないと売れないと予測されるアイテムに対しては、出品価格を段階的に下げながらオフ率の判別規則を

### Step1. 2通りの販売結果を判別する分類規則の構築



### Step2. 新規アイテムの販売結果の予測



### Step3. 新規アイテムの値下がりしない出品価格の最大値の推定

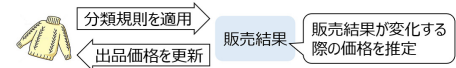


図 2 提案手法の概要

Fig. 2 Overview of our proposed method.

適用し、出品価格で売れると予測される出品価格を値下がりしない出品価格の最大値と推定する。提案手法のイメージを図 2 に示す。

ここで、出品価格の分布はアイテムのカテゴリにより異なることが考えられるが、特定のカテゴリに絞ることで季節性やデータ数にも偏りが生じてしまう。そのため、カテゴリ別に分類モデルを構築しようとする、適切に学習できないモデルが生じてしまう可能性がある。また、分類対象としているオフ率と掲載期間は、カテゴリの出品価格分布によって大きく変化しないことも確認しており、本研究では全カテゴリに対して共通のモデルに適用することとした。

## 4.3 アルゴリズム

いま、アイテムの定義域を  $\mathcal{X}$  と定義し、これはアイテムの特徴ベクトル  $\mathbf{x}$  がとりうる値の集合とする。すなわち、 $\mathbf{x} \in \mathcal{X}$  である。特徴ベクトル  $\mathbf{x}$  の要素である特徴量としては、連続値変数だけでなく、質的変数のダミー変数であってもよいが、出品価格以外の特徴量で構成されているものとする。各アイテムを EC サイト上で販売した際の販売結果は、アイテムの特徴ベクトル  $\mathbf{x}$  と出品価格  $p$  によって決まると考えられる。すなわち、あるアイテムの出品価格  $p$  と特徴量  $\mathbf{x}$  に対し、オフ率が 0%か 0%より高いかを二値で判別する関数を  $f(p, \mathbf{x})$ 、掲載期間が  $k$  日未満か  $k$  日以上となるかを判別する関数を  $g(p, \mathbf{x})$  とすれば、これらの分類器を過去の販売実績データから学習することができる。

いま、 $N$  個の学習データ集合を  $\mathcal{I} = \{(p_n, \mathbf{x}_n, o_n, d_n) | p_n \in \mathbb{R}^+, \mathbf{x}_n \in \mathcal{X}, o_n \in \{0, 1\}, d_n \in \{0, 1\}\}_{n=1}^N$  とし、 $n$  番目のアイテムの出品価格を  $p_n$ 、出品価格以外の特徴量を  $\mathbf{x}_n$  とする。ただし、 $o_n$  は  $n$  番目のアイテムのオフ率が 0%ならば 0、0%より高ければ 1 をとる 2 値変数であり、 $d_n$  は  $n$  番目のアイテムの掲載期間が  $k$  日未満ならば 0、 $k$  日以上ならば 1 をとる 2 値変数とする。この  $N$  個の学習データを用いて、 $f(p, \mathbf{x})$  と  $g(p, \mathbf{x})$  の 2 つの判別モデルを学習することが可能となる。

次に、学習が終了した 2 つの分類器を用いて、新たに出品するアイテム  $\tilde{\mathbf{x}}$  を出品したときの販売結果を予測する

が、この予測には出品価格の入力が必要となる。本研究では、サイト A が現在運用しているプライシングの仕組みに従って算出された（適切と推測される）出品価格  $\tilde{p}$  を仮の出品価格とし、その際の販売結果を予測しながら、出品価格を調整する方法を考える。いま、出品価格を決定したい新たな新規アイテムを  $\tilde{x}$  とし、付与された暫定的価格  $\tilde{p}$  との組合せを  $(\tilde{p}, \tilde{x})$  と記述する。ただし、 $\tilde{p}$  はサイト A の値付けシステムによって、仮の価格が与えられるものとする。このように、暫定的価格を付与する形で与えられた仮出品アイテム集合を  $\mathcal{J} = \{(\tilde{p}, \tilde{x}) | \tilde{p} \in \mathbb{R}^+, \tilde{x} \in \mathcal{X}\}$  とする。このとき、分類器  $f(p, \mathbf{x})$  と  $g(p, \mathbf{x})$  の学習アルゴリズムと、新たに出品する仮出品アイテム  $(\tilde{p}, \tilde{x}) \in \mathcal{J}$  の販売結果を予測し、値下がりしない出品価格の最大値を推定する方法を以下に示す。ここでは、仮出品アイテムのうち、オフ率が 0% ( $f(\tilde{p}, \tilde{x}) = 0$ ) かつ掲載期間が  $k$  日未満 ( $g(\tilde{p}, \tilde{x}) = 0$ ) と予測されるものを「出品価格で売れるアイテム」とし、そのアイテム集合を  $\mathcal{J}_0 \subseteq \mathcal{J}$  と定義する。一方、仮出品アイテムのうち、オフ率が 0% ( $f(\tilde{p}, \tilde{x}) = 1$ ) より高く、かつ掲載期間が  $k$  日以降 ( $g(\tilde{p}, \tilde{x}) = 1$ ) と予測されるものを「値下げしないと売れないアイテム」とし、そのアイテム集合を  $\mathcal{J}_1 \subseteq \mathcal{J}$  と定義する。ただし、どちらにも含まれないアイテムについては、例外アイテム（人手で値付けされる戦略的なアイテム）であると判定し「値下がり起きない出品価格の最大値」は求めない。また、出品価格の値上げ幅  $s$ 、値下げ幅  $t$ 、さらに、出品価格を無限に変化させないための上限値と下限値をそれぞれあらかじめ設定しておく。

以下に、提案手法のアルゴリズムを示す。初期条件として、 $\mathcal{J}_0 = \mathcal{J}_1 = \phi$  とし、仮の暫定的価格をともなう新規アイテム集合  $\mathcal{J}$  は与えられているものとする。

### Step1) 分類器の学習

**Step1-1)** 分類器  $f(p, \mathbf{x})$  は、入力アイテムのオフ率が 0% ならば 0、0% より高ければ 1 となるように  $\{(p_n, \mathbf{x}_n, o_n)\}_{n=1}^N$  を学習データとして学習する。

**Step1-2)** 分類器  $g(p, \mathbf{x})$  は、入力アイテムの掲載期間が  $k$  日未満ならば 0、 $k$  日以上ならば 1 となるように  $\{(p_n, \mathbf{x}_n, d_n)\}_{n=1}^N$  を学習データとして学習する。

### Step2) 販売結果の予測

**Step2-1)** 新規アイテム  $(\tilde{p}, \tilde{x}) \in \mathcal{J}$  に対し、 $f(\tilde{p}, \tilde{x}) = 0$  かつ  $g(\tilde{p}, \tilde{x}) = 0$  であれば、 $(\tilde{p}, \tilde{x})$  を  $\mathcal{J}_0$  に追加する。それ以外の場合は Step2-2 へ。すべての新規アイテム  $(\tilde{p}, \tilde{x})$  の予測が終わってれば、Step3 へ。さもなくば、Step2-4 へ。

**Step2-2)** 新規アイテム  $(\tilde{p}, \tilde{x}) \in \mathcal{J}$  に対し、 $f(\tilde{p}, \tilde{x}) = 1$  かつ  $g(\tilde{p}, \tilde{x}) = 1$  であれば、 $(\tilde{p}, \tilde{x})$  を  $\mathcal{J}_1$  に追加する。それ以外の場合は Step2-3 へ。すべての新規アイテム  $(\tilde{p}, \tilde{x})$  の予測が終わってれば、Step3 へ。さもなくば、Step2-4 へ。

**Step2-3)** 新規アイテム  $(\tilde{p}, \tilde{x}) \in \mathcal{J}$  は、例外アイテム（人

手で値付けされる戦略的アイテム）とする。

**Step2-4)**  $\mathcal{J}$  内の新規アイテムが残っていれば、次の新規アイテムを対象として Step2-1 に戻り、新規アイテムの判定を繰り返す。

**Step3)** 出品価格で売れるアイテム ( $(\tilde{p}, \tilde{x}) \in \mathcal{J}_0$ ) の値下がりしない出品価格の最大値の推定

**Step3-1)**  $j = 0, \tilde{p}^{(0)} = \tilde{p}$  とする。

**Step3-2)**  $\tilde{p}^{(j+1)} = \tilde{p}^{(j)} + \tilde{p}^{(j)} \times s$  とする。 $\tilde{p}^{(j+1)}$  が上限値以上の場合、上限値を  $\tilde{x}$  の最大出品価格とする。

**Step3-3)**  $f(\tilde{p}^{(j+1)}, \tilde{x}) = 0$  ならば、 $j = j + 1$  とし、Step3-2 に戻る。

**Step3-4)**  $f(\tilde{p}^{(j+1)}, \tilde{x}) = 1$  ならば、 $\tilde{p}^{(j)}$  を  $\tilde{x}$  の最大出品価格とする。 $\mathcal{J}_0$  内の新規アイテムが残っていれば、Step3-1 に戻って、次の新規アイテムに対して最大出品価格の推定を繰り返す。

**Step4)** 値下げしないと売れないアイテム ( $(\tilde{p}, \tilde{x}) \in \mathcal{J}_1$ ) の値下がりしない出品価格の最大値の推定

**Step4-1)**  $j = 0, \tilde{p}^{(0)} = \tilde{p}$  とする。

**Step4-2)**  $\tilde{p}^{(j+1)} = \tilde{p}^{(j)} - \tilde{p}^{(j)} \times t$  とする。 $\tilde{p}^{(j+1)}$  が下限値以下の場合、下限値を最大出品価格として終了。

**Step4-3)**  $f(\tilde{p}^{(j+1)}, \tilde{x}) = 1$  ならば、 $j = j + 1$  とし、Step4-2 に戻る。

**Step4-3)**  $f(\tilde{p}^{(j+1)}, \tilde{x}) = 0$  ならば、 $\tilde{p}^{(j+1)}$  を  $\tilde{x}$  の最大出品価格とする。 $\mathcal{J}_1$  内の新規アイテムが残っていれば、Step4-1 に戻って、次の新規アイテムに対して最大出品価格の推定を繰り返す。

本研究では、オフ率の分類器  $f(p, \mathbf{x})$  と掲載期間の分類器  $g(p, \mathbf{x})$  に、高精度な分類器の 1 つである Random Forests (RF) を用いる。RF は、Breiman によって提唱されたバギング [15] という手法を拡張させた手法である。複数の決定木 [16] のアンサンブルをベースとした手法であり、複数の決定木でそれぞれの結果を補うことにより高い予測性能を可能にする [17]。具体的には、学習データから得られるブートストラップサンプルを用いて弱学習器となる決定木を複数構築する。そして、新規入力データのクラスを複数の決定木による結果の多数決により予測する。RF には、相関の低い決定木を構築することができ、各決定木において多様性が生じるという特徴がある。このため、RF の汎化性能は高いことが知られている。

## 5. 実データ分析

サイト A の販売履歴データに提案手法を適用し、2 パターンの販売結果に判別するための判別規則の学習を行う。そして、判別規則を用いて、新規アイテムの販売結果を予測し、値下がりせずに購入される出品価格の最大値を推定する。得られた結果から、各アイテムと推定出品価格との関係分析を行う。

5.1 分析条件

サイト A において 2017 年 1 月 1 日から 2018 年 12 月 31 日までに出品されたアイテムのうち、2018 年 1 月 1 日から 12 月 31 日までに販売されたアイテムの販売履歴データに提案手法を適用する。データ件数は 150,000 件（学習データ：100,000 件，テストデータ：50,000 件）であり，アイテムの特徴量として，出品価格，定価，買取価格，在庫日数，6 種類のアイテムカテゴリ，価格帯ブランドクラス，使用状態，買取月，出品月を用いる ( $D = 42$ )。本分析における  $f(p, \mathbf{x})$ ,  $g(p, \mathbf{x})$  の学習フェーズでは， $o_n$  および  $d_n$  の 0 クラスと 1 クラスの学習サンプルサイズを一致させるため，SMOTE [18] を用いたオーバサンプリングを適用した。また，分類の評価指標として式 (1) から式 (4) に示す正解率，適合率，再現率，F 値を用いる。

$$\text{正解率} = \frac{\text{正しく判別されたデータ数}}{\text{全データ数}} \quad (1)$$

$$\text{適合率} = \frac{\text{正しく判別された正例のデータ数}}{\text{正例と予測されたデータ数}} \quad (2)$$

$$\text{再現率} = \frac{\text{正しく判別された正例のデータ数}}{\text{正例のデータ数}} \quad (3)$$

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (4)$$

本研究では，事前分析を通じ，掲載期間の判別基準を  $k = 8$  とし，出品価格の値上げ幅  $s = 0.01$ ，値下げ幅  $t = 0.01$  と設定した。また，オフ率の分類器  $f(p, \mathbf{x})$  と掲載期間の分類器  $g(p, \mathbf{x})$  には RF を用いる。ここで，各分類器における木の数，木の深さの最大値は，それぞれグリッドサーチを行うことにより決定した。分類器  $f(p, \mathbf{x})$ ,  $g(p, \mathbf{x})$  における木の数については，100 から 900 の範囲で 100 ずつ値を増やししながら，それぞれ正解率が最も高くなった 900, 800 とした。木の深さの最大値については，10 から 90 の範囲で 10 ずつ値を増やししながら，それぞれ正解率が最も高くなった 60, 30 とした。さらに，各ノードに付与する分岐条件の決定においては，使用した特徴量すべてからランダムに特徴量選択を行い，各ノードの分割基準にはジニ係数を用いるものとした。また，出品価格を変化させる際の制約条件として，出品価格の上限値を各アイテムの定価，出品価格の下限値を各アイテムの定価の 10% と設定した。

分類器に用いる RF の比較手法として，高い分類精度で知られている XGBoost [19], LightGBM [20] を用いて同様の評価を行う。比較手法のパラメータについてもグリッドサーチを行い，XGBoost については両分類器ともに，木の深さの最大値を 10，学習率を 0.2 とした。LightGBM については，各分類器  $f(p, \mathbf{x})$ ,  $g(p, \mathbf{x})$  の木の深さの最大値を 20, 10，分岐の個数を 90, 50，学習率を 0.1, 0.3 とした。

5.2 2 パターンの販売結果への判別

2 パターンの販売結果への判別において，上記のモデル

表 5 オフ率の分類器  $f(p, \mathbf{x})$  による各評価指標

Table 5 Evaluation indexes by  $f(p, \mathbf{x})$ .

評価指標	RF	XGBoost	LightGBM
正解率	0.68	0.42	0.45
適合率	0.49	0.34	0.35
再現率	0.53	0.93	0.91
F 値	0.51	0.50	0.51

表 6 掲載期間の分類器  $g(p, \mathbf{x})$  による各評価指標

Table 6 Evaluation indexes by  $g(p, \mathbf{x})$ .

評価指標	RF	XGBoost	LightGBM
正解率	0.69	0.38	0.38
適合率	0.43	0.29	0.28
再現率	0.52	0.95	0.94
F 値	0.47	0.44	0.44

表 7 テストデータの各販売結果における正解率

Table 7 Accuracy by each sales result.

	出品価格で 売れるアイテム	値下げしないと 売れないアイテム
正解率	0.45	0.79
データ数 (件)	12,950	34,594

を用いて構築したオフ率の分類器にテストデータを適用した結果を表 5，掲載期間の分類器にテストデータを適用した結果を表 6 に示す。

表 5 と表 6 より，オフ率と掲載期間の分類器ともに，正解率，適合率，F 値は RF が最も優れていることが分かる。再現率については両分類器とも XGBoost が最も優れているが，これは学習済みの弱学習器による誤差にあてはまるように学習を行うことにより過学習が起きていることが考えられる。したがって，オフ率と掲載期間の分類器には RF が適切であると考えられる。

次に，式 (5) を用いて算出した，テストデータの各販売結果における正解率を表 7 に示す。データ数とは，テストデータにおける正解ラベルを用いて算出した正解データ数である。ただし，ここでは正解ラベルとして，出品価格で売れるアイテムまたは値下げしないと売れないアイテムと定義されなかったアイテムは分析対象外とした。

$$\text{正解率} = \frac{\text{正しく判別されたデータ数}}{\text{各販売結果に予測されたアイテム数}} \quad (5)$$

表 7 より，出品価格で売れるアイテムへの判別は 45%，値下げしないと売れないアイテムへの判別は 79% の精度が得られていることが分かる。実際に出品価格で売れたアイテムは全体の約 26% 程度であり，値下げして売れたアイテムに比べてアイテム数が少ないため，予測が困難になることが考えられる。このことから，オフ率と掲載期間を用いて新規アイテムの販売結果をある程度予測できることが示された。



表 8 各グループの内訳

Table 8 Breakdown of each group.

	出品価格で 売れるアイテム		値下げしないと 売れないアイテム	
	上限値未達	上限値到達	下限値未達	下限値到達
データ数 (件)	11,494	1,740	26,037	5,004
割合 (%)	25.96	3.93	58.81	11.30

表 9 カテゴリ別の各グループの割合 [%]

Table 9 Percentages of each group in each category [%].

カテゴリ (割合)	出品価格で 売れるアイテム		値下げしないと 売れないアイテム	
	上限値未達	上限値到達	下限値未達	下限値到達
全体 (100.00)	25.96	3.93	58.81	11.30
アウト (11.37)	16.78	<b>7.19</b>	<b>61.42</b>	<b>14.61</b>
グッズ (15.68)	21.89	3.37	<b>64.07</b>	10.67
トップス (38.72)	<b>29.00</b>	<b>4.04</b>	57.20	9.76
ボトムス (24.04)	<b>27.90</b>	2.67	57.72	<b>11.70</b>
ワンピース (10.23)	<b>27.34</b>	3.93	55.17	<b>13.57</b>

※太字の割合は各グループが全体に占める割合より大きいカテゴリ

### 5.3 値下がりせずに購入される出品価格の最大値について

本節では、出品価格で売れると予測されたアイテムと値下げしないと売れないと予測されたアイテムについて、値下がりせずに購入される出品価格の最大値を推定した結果をアイテムのカテゴリとブランドに基づいて分析を行う。出品価格で売れると予測されたアイテムについては、推定された出品価格が、上限値に到達しなかったアイテムと上限値に到達したアイテムが存在する。一方、値下げしないと売れないと予測されたアイテムについては、推定された出品価格が、下限値に到達しなかったアイテムと下限値に到達したアイテムが存在する。すなわち、上限値に到達しなかったアイテムと下限値に到達しなかったアイテムについては、適切な出品価格が推定されたアイテム、上限値に到達したアイテムと下限値に到達したアイテムについては、設定したパラメータの範囲内では適切な出品価格が推定できないと判断されたアイテムと考えることができる。これら4つのグループの割合を表8に示す。

表8より、下限値に到達しなかったアイテムが最も多いことが分かる。このことから、出品価格を多少低く設定することにより、オフ率0%で販売可能なアイテムが多いことが理解できる。また、出品価格で売れると予測されたアイテムについては、推定された出品価格が上限値に到達しなかったアイテムが多いことが読み取れる。このことから、出品価格をある程度高く設定しても販売結果が変わらないアイテムが多いことが考えられる。

#### 5.3.1 カテゴリに着目した分析

上述した4つのグループに属するアイテムについて、カテゴリに着目して分析を行う。分析結果を表9に示す。ただし、全データを用いた事前分析において、6種類のカテ

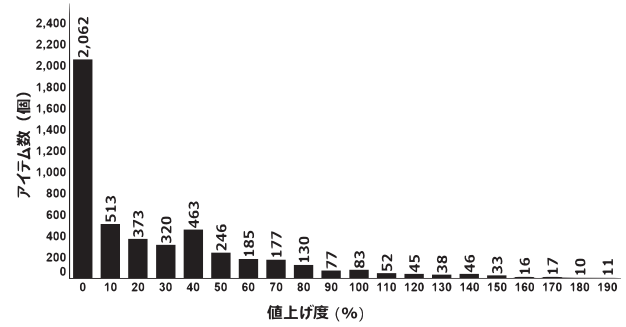


図 3 値上げ度の分布 (トップス)

Fig. 3 The degree of price increase (Tops).

ゴリのうち、アイテムの所属割合が1%未満であったカテゴリはデータ数が非常に少なく、分析する上で偏った特徴が見られると考え、今回は分析対象外とした。

表9より、各カテゴリにおける各グループへの所属割合を比較すると、上限値に到達しなかったアイテムには、トップス、ボトムス、ワンピースが多く含まれていることが読み取れる。それらのアイテムは出品価格を上げてオフ率0%を達成できるアイテムと考えられる。また、上限値に到達したアイテムにはアウト、トップスが多く含まれ、それらは定価のままでもオフ率0%を達成できるアイテムであり、他のアイテムに比べて売上増加が期待できる。

一方、下限値に到達しなかったアイテムには、アウト、グッズが多く含まれる。これらは出品価格を下げることでオフ率0%を達成できると考えられ、掲載期間の短縮が見込まれる。また、下限値に到達したアイテムにはアウト、ボトムス、ワンピースが多く含まれ、これらは出品価格を大きく下げてもオフ率0%の達成が難しいアイテムが多く、アイテムの買取自体を検討する必要があると考えられる。

次に、上限に到達しなかったアイテムのカテゴリにおいて、最も割合が大きくなったトップスに対し、提案手法で推定した出品価格とその値上がりした割合（以下、値上げ度）を式(6)のように定義し、得られた度数分布を図3に示す。ただし、見やすいよう、度数が2桁となった値上げ度を表示する。また、下限に到達しなかったアイテムのカテゴリにおいて、最も割合が大きくなったグッズに対し、推定した出品価格とその値下がりした割合（以下、値下げ度）を式(7)のように定義し、得られた度数分布を図4に示す。ただし、図中の%表示は、q%以上q+10%未満の値上げ度・値下げ度となったアイテムをq%と表記した。

$$\text{値上げ度} = \left( \frac{\text{最終出品価格} - \text{元の出品価格}}{\text{元の出品価格}} \right) \times 100 \quad (6)$$

$$\text{値下げ度} = \left( \frac{\text{元の出品価格} - \text{最終出品価格}}{\text{元の出品価格}} \right) \times 100 \quad (7)$$

図3より、値上げ度が0%から9%のアイテムが多く存

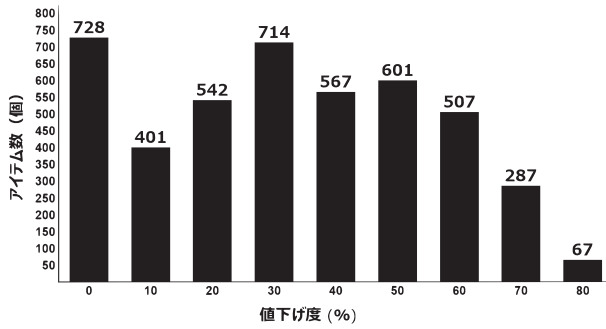


図 4 値下げ度の分布 (グッズ)

Fig. 4 The degree of price reduction (Goods).

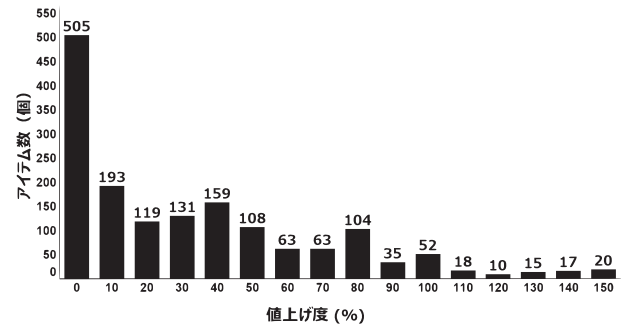


図 5 値上げ度の分布 (価格帯ブランドクラス 8)

Fig. 5 The degree of price increase (price range class 8).

表 10 価格帯ブランドクラス別の各グループの割合 [%]

Table 10 Percentages of each group in each brand cluster [%].

クラス (割合)	出品価格で売れるアイテム		値下げしないと売れないアイテム	
	上限値未達	上限値到達	下限値未達	下限値到達
全体 (100.00)	25.96	3.93	58.81	11.30
1 (1.59)	<b>32.28</b>	<b>12.34</b>	53.23	2.15
3 (2.62)	14.07	<b>8.86</b>	<b>66.72</b>	10.34
4 (7.65)	19.86	<b>6.07</b>	<b>66.07</b>	8.00
5 (29.95)	23.35	3.61	<b>64.54</b>	8.50
6 (48.16)	<b>26.63</b>	2.73	<b>59.22</b>	<b>11.42</b>
7 (3.22)	5.23	0.07	29.77	<b>64.92</b>
8 (6.81)	<b>54.13</b>	<b>7.59</b>	35.37	2.91

※太字の割合は各グループへの所属割合が最も大きいクラス

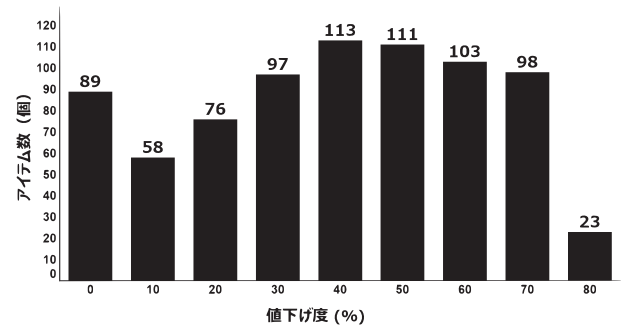


図 6 値下げ度の分布 (価格帯ブランドクラス 3)

Fig. 6 The degree of price reduction (price range class 3).

在していることから、上限値に到達しなかったトップスの多くは出品価格を0%から9%値上げしても出品価格で売れると考えられ、売上増加が期待できる。また図4より、値下げ度が0%から9%と30%から39%のアイテムが多く存在していることから、下限値に到達しなかったデニムの多くは出品価格を0%から9%または30%から39%値下げすることで出品価格で売れると考えられ、掲載期間短縮が見込まれる。

### 5.3.2 ブランドに着目した分析

前項と同様に、4つのグループに属するアイテムについて、アイテムのブランドに着目して分析を行う。表3に示した価格帯ブランドクラスを用いて、各価格帯ブランドクラスにおけるグループの割合を分析した結果を表10に示す。ただし、前節と同様に全データを用いた事前分析において、8つの価格帯ブランドクラスのうち、アイテムの所属割合が1%未満であったクラスは分析対象外とした。

表10より、クラス8は他のクラスに比べて上限値未達の割合が大きい。このことから、クラス8に所属するアイテムの多くは出品価格を上げてオフ率0%を達成できるアイテムと考えられる。また、クラス1は他のクラスに比べて上限値到達の割合が大きい。このことから、クラス1に所属するアイテムの多くは定価のままでもオフ率0%を達成できるアイテムであり、他のアイテムに比べて売上増

加が期待できる。

一方、クラス3は他のクラスに比べて下限値未達の割合が大きい。このことから、クラス3に所属するアイテムの多くは出品価格を下げることでオフ率0%を達成できると考えられ、掲載期間の短縮が見込まれる。また、クラス7は他のクラスに比べて下限値到達の割合が大きい。このことから、クラス7に所属するアイテムの多くは出品価格を大きく下げてもオフ率0%を達成できないアイテムが多く、アイテムの買取自体を検討する必要があると考えられる。

次に、図5に上限値未達の割合が最も大きかったクラス8に所属するアイテムに対し、式(6)により値上げ度を求めた度数分布を示す。ただし、見やすいよう、度数が2桁となった値上げ度を表示する。また、図6に下限値未達の割合が最も大きかったクラス3に所属するアイテムに対し、式(7)により値下げ度を求めた度数分布を示す。

図5より、値上げ度が0%から9%のアイテムが多く存在していることから、上限値に到達しなかったクラス8の多くは出品価格を0%から9%値上げしても出品価格で売れると考えられ、売上増加が期待できる。また図6より、値下げ度が40%から59%のアイテムが多く存在していることから、下限値に到達しなかったクラス3の多くは出品価格を40%から59%値下げすることで出品価格で売れると考えられ、掲載期間短縮が見込まれる。

## 5.4 収益性における効果

本節では、提案手法を適用することによる効果について、推定された出品価格と実際の販売価格を比較することで考察を行う。

5.3節で述べたように、適切な出品価格が推定されたアイテムには、出品価格で売れるアイテムと値下げしないと売れないアイテムが存在する。その中でも上限値・下限値に到達したか否かによって「上限値未達・上限値到達」、「下限値未達・下限値到達」の4つのグループに分けられる。「上限値未達・上限値到達・下限値未達」のアイテムは、提案手法で設定した出品価格で売れる価格が推定されたが、下限値到達のアイテムは、下限値まで大幅に出品価格を下げて、その価格では売れないアイテムであると解釈できる。しかし、それらの下限値到達のアイテムは、出品後に値下げを行うことで出品価格からのオフ率にユーザが魅力を感じて購入に至る場合があると考えられる。すなわち、出品価格では購入に至らないが、出品後に値下げを行うことで購入される可能性があるといえるため、購入されないアイテムとはいえ、損失として換算することはできない。また、このアイテム群は出品以前に買取段階での意思決定によって損失を防ぐことをあわせて考慮すべきであろう。

そこで、上記の分析に用いた5万件のテストデータのうち、本研究の提案手法で販売結果が予測可能な「上限値未達・上限値到達・下限値未達」のアイテムについて、推定された出品価格を用いた場合の総売上額へのインパクトについて評価を行う。これらのアイテムの推定出品価格は出品後すぐに売れる価格であり、本研究では「オフ率0%かつ掲載期間8日未満」のアイテムと定義している。そのため、過去の販売履歴データにおける「オフ率0%かつ掲載期間8日未満」で売れたアイテムの販売価格との比較を行う。ただし、上限値未達・上限値到達のアイテムに対しては、出品価格で売れると予測されたアイテムの正解率である0.45、下限値未達のアイテムに対しては値下げしないと売れないと予測されたアイテムの正解率である0.79をかけることで推定した出品価格の信頼度を考慮している。

実際に、上限値未達・上限値到達・下限値未達の全アイテムの推定出品価格に正解率をかけた値（提案手法で想定される総売上額）と、過去の販売履歴データにおける「オフ率0%かつ掲載期間8日未満」で売れたアイテムの総売上額の実績との差は51,471,326（円）であった。つまり、提案手法を用いることで約5千万円の売上げ増加が期待できるといえる。加えて、本分析では5万件のテストデータに対して適切な出品価格を推定したが、実運用を行う際にはさらに多くのアイテムに対して適用されることが想定される。そのため、より大きな売上げの増加量が期待できる。

## 6. 考察

ここでは、特徴量選択、ハイパーパラメータの決定、閾

値の決定の観点から、本研究の提案モデルについて考察を行う。

### 6.1 特徴量選択

本研究では、オフ率の分類器、掲載期間の分類器のどちらも、出品価格、定価、6種類のアイテムカテゴリといった42次元の変数を用いて分類器の構築を行った。分類精度を向上させるという観点においては、既存の特徴量から有用な特徴量選択や新しい特徴量を生成する特徴量エンジニアリング [21] や、特徴量間の交互作用を考慮した回帰モデルである Factorization Machines [22] といった手法を用いてより適切な特徴量の組合せの検討することで、各分類器の精度向上が期待できる。しかしながら、分類精度がある程度得られていることから、本研究で行った分類器の構築は適切であったと考えられる。

### 6.2 ハイパーパラメータの決定

本研究の提案手法の実装にあたり、値下がりせずに購入される出品価格の最大値を推定する際の出品価格の上限値と下限値は売上げやブランドイメージなどを考慮して検討すべきである。また、出品価格の値上げ幅、値下げ幅についても、より小さな値に設定することで、正確に適切な出品価格の推定が期待できる。しかし、この2種類のパラメータの設定は、正確な出品価格の推定だけではなく、処理スピードにも大きな影響をもたらすことが考えられる。本研究で行った分析では、出品価格の上限値をアイテムの定価、下限値をアイテムの定価の10%、出品価格の値上げ幅と値下げ幅をそれぞれ1%と設定した。分析の結果、上述したように売上げの増加を期待できる結果が得られた。このことから、本研究で設定したハイパーパラメータは適切であったと考えられる。

### 6.3 閾値の決定

本研究では、「出品価格で売れるアイテム」をオフ率0%かつ掲載期間8日未満で購入されたもの、「値下げしないと売れないアイテム」をオフ率0%以外かつ掲載期間8日以降で購入されたものとそれぞれ定義した。この「0%」と「8日」という閾値は事前分析により、性質が異なると考えられるアイテムを分けられる適切な値と判断した値であり、性質の異なるデータを分けるうえで重要な決定であると考えられる。加えて、5章の実データ分析においてはオーバースAMPLINGを導入しているため、データインバランスは発生していない。

提案手法の他の事例への適用を考える場合、事前分析を行わず、意思決定者がランダムな閾値決定を行うと、データインバランスの問題が発生する可能性が考えられる。一方で、分類器構築の際にオーバースAMPLINGを導入することでデータインバランスの問題は軽減すると考えられるが、

この閾値決定は提案手法を実装するうえで重要である。

## 7. まとめと今後の課題

本研究ではサイト A におけるアイテムの出品価格決定問題を対象とし、販売履歴データを活用した機械学習モデルを駆使した方法について検討を行った。新たに出品するアイテムの値下がり不起こらない出品価格の最大値を推定することを目的とし、アイテムの販売結果を予測することで、値下がりしない出品価格の最大値を推定する手法を提案した。提案手法を実際にサイト A の販売履歴データに適用し出品価格を推定することで、提案手法の有用性を示した。これにより、サイト A において、各アイテムが値下がりせず購入される出品価格の最大値の推定を可能とした。この推定出品価格はサイト A における出品価格決定の際の新たな基準値となることが期待できる。

今後の課題としては、値下がりしない出品価格の最大値を求めなかったアイテムの解釈があげられる。本研究では、新規アイテムの出品価格を推定する際、オフ率が 0% 以下かつ掲載期間が  $k$  日以降と予測されたアイテムと、オフ率が 0% より高くかつ掲載期間が  $k$  日未満と予測されたアイテムは、推定対象外のアイテムとした。そのため、実応用の観点では、それらのアイテムを考慮した出品価格の推定モデルも必要である。

謝辞 多くの有益なコメントを多数いただき、また本論文の事後分析の方向性についてもご示唆をいただいた査読者に深く感謝いたします。

## 参考文献

[1] Breiman, L.: Random Forests, *Machine Learning*, Vol.45, No.1, pp.5–32 (2001).

[2] Liu, N., Ren, S., Choi, T.M., Hui, C.L. and Ng, S.F.: Sales Forecasting for Fashion Retailing Service Industry: A Review, *Mathematical Problems in Engineering*, Vol.2013, No.738675 (2013).

[3] Thomassey, S. and Fiordaliso, A.: A hybrid sales forecasting system based on clustering and decision trees, *Decision Support Systems*, Vol.42, No.1, pp.408–421 (2006).

[4] Ni, Y. and Fan, F.: A two-stage dynamic sales forecasting model for the fashion retail, *Expert Systems with Applications*, Vol.38, No.3, pp.1529–1536 (2011).

[5] Au, K.-F., Choi, T.-M. and Yu, Y.: Fashion retail forecasting by evolutionary neural networks, *International Journal of Production Economics*, Vol.114, No.2, pp.615–630 (2008).

[6] Yu, Y., Choi, T.-M. and Hui, C.-L.: An intelligent fast sales forecasting model for fashion products, *Expert Systems with Applications*, Vol.38, No.6, pp.7373–7379 (2011).

[7] Aksoy, A., Ozturk, N. and Sucky, E.: A decision support system for demand forecasting in the clothing industry, *International Journal of Clothing Science and Technology*, Vol.24, No.4, pp.221–236 (2012).

[8] Choi, T.-M., Yu, Y. and Au, K.-F.: A hybrid SARIMA wavelet transform method for sales forecasting, *Decision*

*Support Systems*, Vol.51, No.1, pp.130–140 (2011).

[9] 上田隆穂：価格決定におけるマーケティング戦略，学習院大学経済論集，Vol.31, No.4, pp.185–208 (1995).

[10] Erickson, G.M. and Johansson, J.K.: The Role of Price in Multi-attribute Product Evaluation, *Journal of Consumer Research*, Vol.12, No.2, pp.195–199 (1985).

[11] Tellis, G.J.: Beyond The Many Faces of Price: An Integration of Pricing Strategies, *Journal of Marketing*, Vol.50, No.4, pp.146–160 (1986).

[12] 岩本明憲：価格変更戦略の理論的フレームワークの構築，日本マーケティング学会マーケティングジャーナル，Vol.39, No.3, pp.89–103 (2020).

[13] Cheung, W.C., Simchi-Levi, D. and Wang, H.: Dynamic Pricing and Demand Learning with Limited Price Experimentation, *Operations Research*, Vol.65, No.6, pp.1722–1731 (2017).

[14] 仁ノ平将人, 三川健太, 後藤正幸：販売履歴データに基づく中古ファッションアイテムの販売価格予測モデルに関する一考察，情報処理学会論文誌，Vol.60, No.4, pp.1151–1161 (2019).

[15] Breiman, L.: Bagging predictors, *Machine Learning*, Vol.24, No.2, pp.123–140 (1996).

[16] Breiman, L., Friedman, J., Olshen, R.A. and Stone, C.J.: *Classification and Regression Trees*, Wadsworth, Belmont, CA (1984).

[17] 波部 斉：ランダムフォレスト，情報処理学会研究報告，Vol.2012-CVIM-182, No.31, pp.1–8 (2012).

[18] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P.: SMOTE: Synthetic Minority Over-sampling Technique, *Journal of Artificial Intelligence Research*, Vol.16, pp.321–357 (2002).

[19] Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.785–794 (2016).

[20] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y.: LightGBM: A Highly Efficient Gradient Boosting Decision Tree, *Advances in Neural Information Processing Systems* (2017).

[21] Nargesian, F., Samulowitz, H., Khurana, U., Khalil, E.B. and Turaga, D.: Learning Feature Engineering for Classification, *Proc. 26th International Joint Conference on Artificial Intelligence*, pp.2529–2535 (2017).

[22] Rendle, S.: Factorization Machines, *Proc. 2010 IEEE International Conference on Data Mining*, pp.995–1000 (2010).



桑田 和

1996年生。2019年早稲田大学創造理工学部経営システム工学科卒業。現在、同大学大学院創造理工学研究科経営システム工学専攻在学。機械学習を用いたデータ分析に関する研究に興味を持つ。



杉崎 智哉

1994年生。2018年早稲田大学創造理工学部経営システム工学科卒業。2020年同大学大学院修士課程修了。在学時 Factorization Machines のモデル構造に関する研究に従事。



三川 健太 (正会員)

1981年生。2005年武蔵工業大学環境情報学部環境情報学科卒業。2007年同大学大学院修士課程修了。2016年早稲田大学大学院博士後期課程修了。博士(工学)。2013年早稲田大学助手。2016年湘南工科大学工学部情報工学科講師。2019年同大学准教授。機械学習とその応用に関する研究に従事。IEEE, 電子情報通信学会, 日本経営工学会等各会員。



後藤 正幸 (正会員)

1969年生。1994年武蔵工業大学大学院修士課程修了。2000年早稲田大学大学院理工学研究科博士課程修了。博士(工学)。1997年同大学理工学部助手。2000年東京大学大学院工学研究科助手。2002年武蔵工業大学環境情報学部助教授。2008年早稲田大学創造理工学部経営システム工学科准教授。2011年同大学教授。情報数理応用とデータサイエンス, ならびにビジネスアナリティクスの研究に従事。著書に、『入門パターン認識と機械学習』コロナ社(2014), 『ビジネス統計—統計基礎とエクセル分析』オデッセイコミュニケーションズ(2015)等。IEEE, INFORMS, 電子情報通信学会, 人工知能学会, 日本経営工学会, 経営情報学会等各会員。