

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決定モデルに関する研究

1X15C036-1 桑田 和
指導教員 後藤 正幸

1 研究背景・目的

本研究では、ユーザからアイテムを買い取り、中古品として再販売する事業を展開する某ファッション EC サイト (以下、サイト A) における出品価格の決定問題を対象とする。出品価格とはアイテムを EC サイト上で出品する際の価格であり、これと連動して買取価格も決まるため、その最適化はビジネス上大変重要な課題の 1 つとなっている。

サイト A では、アイテムの売れ残りを防ぐため、出品から一定期間経過した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。収益性の観点からは、可能な限り値下げをせずに販売することが望ましいが、購入の意思決定はユーザ側にあるため、ブランドや季節性といった多様性の高いファッションアイテムは実際に出品するまで購入結果が予想できない場合も多い。一方、サイト A では過去に出品したアイテムの出品価格や購入価格といった販売履歴データが膨大に蓄積されており、それらのデータの分析が各アイテムに対する最適な出品価格の決定につながると考えられる。

しかし、サイト A に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のまま購入されたアイテムと、出品から購入されるまでの期間 (以下、販売期間) が長くなったために、出品価格から値下げされて購入されたアイテムの 2 通りの販売結果が混在している。前者は出品価格を高く設定しても出品価格のまま購入された可能性、後者は出品価格を低く設定することで出品価格のまま購入された可能性がある。そのため、購入された時点での販売価格がアイテムに対する最適な価格とは限らない。

そこで本研究では、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し、その結果から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モデルを提案する。提案モデルではまず、過去の販売履歴データを用いて、アイテムの出品価格を基に、出品価格で購入されるか、値下げされて購入されるかの 2 通りの販売結果を判別する分類器を構築する。次に、推定された分類器を用い、アイテムの出品価格のみを変化させて結果の差異を分析する。そして、元の販売結果とは異なる結果となった際の価格を求め、これを出品価格とする。この方法により、値下がり不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、実際の価格決定の基準値となることが期待できる。本研究では、サイト A の販売履歴データに提案モデルを適用し、結果に対する考察を行う。

2 問題設定

サイト A では、出品価格が高価格であると販売期間が長くなり、在庫費の増加や売れ残り、同一アイテムの長期間掲載によるサイトの魅力低下などの問題が発生する。一方、アイテムを低価格で出品した場合、すぐに顧客に購入される反面、より高価格で購入される機会を失う。そのため、販売期間と売上の双方を考慮した出品価格の推定が望まれる。

過去に出品されたアイテムのうち、出品価格のまま購入されたアイテムは、より高い出品価格であってもその価格で購入され、販売期間に大きな変化が生じなかった可能性がある。一方、値下げされたアイテムは、より低い出品価格であればより早く購入され、販売期間が短縮された可能性がある。そのため、これらの結果の差異を考慮に入れて、より適切な出品価格を推定するモデルを構築する必要がある。

そこで本研究では、まず、出品価格からの値下がり率 (以

下、オフ率) と販売期間を用い、出品価格を設定したアイテムが出品価格で購入されるか、値下げされるかを判別する分類モデルを構築する。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方法を示す。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 提案モデル

3.1 概要

サイト A で扱われるアイテムは、出品価格によって 2 通りの販売結果が得られる。出品価格で購入されたアイテムは、ユーザに出品価格が妥当または安価であると判断された場合であり、それらは出品価格を上げると、ある価格を境に出品価格のまま購入されなくなり、値下げに結び付くと考えられる。一方、値下げされたアイテムは、ユーザに出品価格が高く、妥当でないと判断された場合であり、それらは出品価格を下げると、ある価格を境に出品価格で購入されるようになると考えられる。

そこで本研究では、出品価格で購入さ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、値下げされると予測される直前の価格を、値下がりしない出品価格の最大値と推定する。一方、値下げさ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下げ、出品価格で購入されると予測される際の価格を、値下がりしない出品価格の最大値と推定する。

提案モデルではまず、過去に購入されたアイテムの販売履歴データにおける、オフ率と販売期間を分類基準として学習することで、出品価格で購入される場合と値下げされる場合の 2 通りの販売結果を判別する分類規則を構築する。そして出品価格を設定した新規アイテムに分類規則を適用し、販売結果を予測する。次に、アイテムの値下がりしない出品価格の最大値を推定する。出品価格で購入されると予測されるアイテムに対しては、出品価格を上げながらオフ率の分類規則を適用し、値下げされると予測されるアイテムに対しては、出品価格を下げながらオフ率の分類規則を適用し、出品価格で購入されるアイテムに変化する出品価格を見つける。

3.2 提案モデルのアルゴリズム

N 個の学習データのうち、 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 $\mathbf{x}_n \in \mathbb{R}^D$ 、オフ率に関する分類ラベルを表す離散変数を $o_n \in \{0, 1\}$ 、販売期間に関する分類ラベルを表す離散変数を $d_n \in \{0, 1\}$ と定義する。ただし、 o_n は n 番目のアイテムのオフ率が $h\%$ 以下ならば 0、 $h\%$ より高ければ 1 をとる変数であり、 d_n は n 番目のアイテムの販売期間が k 日未満ならば 0、 k 日以降ならば 1 をとる変数とする。また、オフ率が $h\%$ 以下と $h\%$ より高いクラスを識別する関数を $f(\cdot)$ 、販売期間が k 日未満と k 日以降を識別する関数を $g(\cdot)$ とする。

このとき、 $f(\cdot)$ 、 $g(\cdot)$ の学習アルゴリズムと、新たに出品する新規アイテム $\tilde{\mathbf{x}}$ 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アルゴリズムを以下に示す。ここでは、新規アイテムのうち、オフ率が $h\%$ 以下かつ販売期間が k 日未満と予測されるアイテムを出品価格で購入されるアイテム ($I_0 \in \mathcal{I}$)、オフ率が $h\%$ より高かつ販売期間が k 日以降と予測されるアイテムを値下げされるアイテム ($I_1 \in \mathcal{I}$) 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

ては値下がりしない出品価格の最大値は求めない。また、新規アイテム \tilde{x} の暫定出品価格を \tilde{p} 、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予め設定しておく。

Step1) 販売結果の予測

Step1-1) $f(\cdot)$ は、入力アイテムのオフ率が $h\%$ 以下ならば 0 、 $h\%$ より高ければ 1 となるように

$\{p_n, x_n, o_n\}_{n=1}^N$ を学習データとして学習する。

Step1-2) $g(\cdot)$ は、入力アイテムの販売期間が k 日未満ならば 0 、 k 日以降ならば 1 となるように

$\{p_n, x_n, d_n\}_{n=1}^N$ を学習データとして学習する。

Step1-3) 新規アイテム \tilde{x} に対し、 $f(\tilde{p}, \tilde{x}) = 0$ かつ $g(\tilde{p}, \tilde{x}) = 0$ のとき I_0 へ分類し、Step2 へ。

$f(\tilde{p}, \tilde{x}) = 1$ かつ $g(\tilde{p}, \tilde{x}) = 1$ のとき、 I_1 へ分類し、Step3 へ。

Step2) 出品価格で購入されるアイテム ($\{\tilde{p}, \tilde{x}\} \in I_0$) の値下がりしない出品価格の最大値の推定

Step2-1) $j = 0$, $\tilde{p}^{(0)} = \tilde{p}$ とする。

Step2-2) $\tilde{p}^{(j+1)} = \tilde{p}^{(j)} + \tilde{p}^{(j)} \times s$ とする。 $\tilde{p}^{(j+1)}$ が上限値以上の場合、上限値を最大出品価格として終了。

Step2-3) $f(\tilde{p}^{(j+1)}, \tilde{x}) = 1$ ならば、 $\tilde{p}^{(j)}$ を最大出品価格として終了。 $f(\tilde{p}^{(j+1)}, \tilde{x}) = 0$ ならば、 $j = j + 1$ とし、Step2-2 に戻る。

Step3) 値下げされるアイテム ($\{\tilde{p}, \tilde{x}\} \in I_1$) 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) $\tilde{p}^{(j+1)} = \tilde{p}^{(j)} - \tilde{p}^{(j)} \times t$ 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。

Step3-3) $f(\tilde{p}^{(j+1)}, \tilde{x}) = 0$ ならば、 $\tilde{p}^{(j+1)}$ を最大出品価格として終了。 $f(\tilde{p}^{(j+1)}, \tilde{x}) = 1$ ならば、 $j = j + 1$ とし、Step3-2 に戻る。

4 実験

サイト A の販売履歴データに提案モデルを適用し、2通りの販売結果に判別するための分類規則の学習を行う。そして、分類規則を用いて、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。

4.1 実験条件

提案モデルを、サイト A における 2017 年 1 月 1 日から 12 月 31 日の販売履歴データに適用する。データ件数は 284,076 件 (分類器を構築する際の学習データ: 255,668 件, 新規アイテムとして使うテストデータ: 28,408 件) であり、アイテムの特徴量として、出品価格に加えて、定価、20 種類のアイテムカテゴリ、51 種類の補助変数を用いる ($D = 72$)。また、分類の評価指標として正解率を用いる。

本研究では、事前分析を通じ、オフ率の分類基準を $h = 0$ 、販売期間の分類基準を $k = 8$ 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 $f(\cdot)$ と販売期間の分類器 $g(\cdot)$ には Random Forests[1] を用いる。ここで、分類器 $f(\cdot)$ 、 $g(\cdot)$ における木の数はそれぞれ 200, 700 とし、木の深さの最大値はどちらも 100 とした。さらに、各ノードの分割基準にはジニ係数を用いるものとした。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。

4.2 実験結果

(1) 2通りの販売結果への分類結果

2通りの販売結果への分類において、オフ率と販売期間の分類器の正解率を表 1 に、テストデータの各販売結果にお

ける正解率を表 2 に示す。

表 1. 各分類器による正解率

	$f(\cdot)$ (オフ率)	$g(\cdot)$ (販売期間)
学習データ	0.999	0.999
テストデータ	0.690	0.702

表 2. テストデータの各販売結果における正解率

	出品価格で購入されるアイテム	値下げされるアイテム
正解率	0.479	0.753

表 1 より、オフ率と販売期間の分類器はテストデータに対して約 7 割の正解率が得られていることが分かる。また表 2 より、出品価格で購入されるアイテムへの分類は約 48%、値下げされるアイテムへの分類は約 75%の精度が得られていることが分かる。以上より、オフ率と販売期間を用いることで、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。

(2) 値下がりしない出品価格の最大値についての分析

出品価格で購入されると予測されたアイテムのうち、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、値下げされると予測されたアイテムのうち、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についてそれぞれ分析する。各アイテムのカテゴリの割合を表 3 に示す。

表 3. カテゴリ別の各グループの割合 [%]

カテゴリ (割合)	出品価格で購入されるアイテム		値下げされるアイテム	
	上限値未達	上限値到達	下限値未達	下限値到達
すべて (100.00)	6.16	15.80	42.64	35.41
カットソー (15.84)	7.20	12.51	44.16	36.13
コート (3.64)	3.52	41.71	25.13	29.65
シャツ (10.85)	8.40	17.18	44.15	30.27
シューズ (10.02)	3.24	5.76	38.97	52.03
スカート (7.26)	5.42	14.44	48.93	31.21
デニム (4.30)	9.48	13.95	48.03	28.54
ニット (13.51)	4.10	14.41	34.54	46.95
バッグ・ケース類 (9.86)	6.27	22.06	41.57	30.10
パンツ (13.36)	7.37	11.00	48.78	32.85
ワンピース (11.36)	6.61	15.79	49.44	28.16

(※太字は割合が大きい上位 3 つのカテゴリ)

表 3 より、全アイテムの所属割合と比較すると、上限値に到達しなかったアイテムには、シャツ、デニム、パンツが多く含まれ、それら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。上限値に到達したアイテムにはコート、シャツ、バッグ・ケース類が多く含まれ、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を見込めると考えられる。

一方、下限値に到達しなかったアイテムには、スカート、パンツ、ワンピースが多く含まれ、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、販売期間の短縮が見込める。下限値に到達したアイテムにはカットソー、シューズ、ニットが多く含まれ、これらは出品価格を大きく下げてもオフ率 0% 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。

5 まとめと今後の課題

本研究では値下がりしない出品価格の最大値を推定するモデルを提案し、サイト A における実際の販売履歴データを用いて出品価格を推定し、分析することでその有効性を示した。今後の課題として、実験で用いたテストデータのうち、二値分類後に値下がりしない出品価格の最大値を求めなかった、オフ率「0%」かつ販売期間「8 日以降」とオフ率「0% 以外」かつ販売期間「8 日未満」と予測されたアイテムを検討することが挙げられる。

参考文献

- [1] L. Breiman, "Random Forests," *Machine Learning*, Vol. 45, pp. 5–32, 2001.