

# 中古ファッション EC サイトにおける出品価格と販売価格の 関係分析モデルに関する一考察

情報数理応用研究

5218C007-9 金澤真平

指導教員 後藤正幸

## An Analytical Model of Exhibition Price Change Effects on Second-hand Fashion EC Site

KANAZAWA Shimpei

### 1. 研究背景と目的

スマートフォンの爆発的な普及により、インターネットを介して商品（以下、アイテム）の売買を行う EC サイトが広く利用されるようになった。特にここ数年では、ファッションアイテムを商材とする EC サイトの市場規模が拡大し、EC サイトの運営企業は膨大な販売履歴データを蓄積することが可能となっている。このような背景のもと、様々な機械学習手法に基づくデータ分析の試みがなされており、顧客分析や商品推薦などで成果を挙げつつある [1]。

本研究では、ユーザからファッションアイテムの買取、値付け、および再販売のビジネスを展開する中古ファッション EC サイト（以下、EC サイト A）を対象とする。EC サイト A では出品時の価格を出品価格としており、最終的にユーザが購入した時点の価格を販売価格としている。また、アイテムの売れ残りを防ぐため、出品してから一定期間売れなかったアイテムに対して自動的に値下げを行うシステムを採用している。ビジネス的観点からは、各アイテムが値下げしない状態で売ることが望ましく、適切な出品価格の設定が重要な課題のひとつとなっている。そこで、各アイテムに関して、出品価格をいくらに設定すると、最終的にいくらで売れるのかを予測するモデルが出品価格設定の一助になると考えられる。すなわち、膨大な販売履歴データに基づいた販売価格予測モデルの構築が有効な手段である。

EC サイト A の各アイテムにはブランドやカテゴリ、価格情報など様々な情報があり、各アイテムの販売結果を正確に予測することは非常に難しく、EC サイト A で採用している値付けシステム（以下、現行システム）によって適切な価格設定となっているか否かは定かでない。実際、売れ残ったことで大きく値下げされて販売されるケースも多々生じている。そのため、現行システムによって設定された価格から出品価格を変更した場合に販売価格が上がるアイテムや、逆に出品価格を変更することで販売価格が下がるアイテムも存在すると考えられる。そのため、出品価格を変更した場合の販売価格を高い精度で予測可能なモデルの構築が望まれる。すなわち、出品価格と販売価格の関係性を分析可能なモデルは有用であると考えられる。

この問題に対し、仁ノ平ら [2] はアイテムの販売傾向をもとにクラスタリングを行い、混合回帰を導入した予測モデルを提案しており、ある程度の予測精度が得られることを示した。また、アイテムの特徴によっては高い精度で予測することが難しいことも示している。しかし、現行システムで設定

された出品価格で販売した販売履歴データのみを学習に用いており、出品価格を変更した際の販売価格を取り込んでいないため、出品価格を変更した場合の販売価格を予測するモデルとして利用可能かどうかは定かではない。

一方、出品価格を変更した際の販売価格を予測するモデルを構築するためには、実際に出品価格を変更して販売価格を観察する必要がある。しかし、事業上の制約のため、全アイテムを対象として価格変更テストを実施するのは困難であり、価格変更テストを通じて得られる結果データは少数に限られる。一般に、比較的少数の学習データから構築された機械学習モデルは、精度面の問題が起る可能性があり、加えて、大量に存在している過去の販売履歴データを全く活用しないことも不合理であると考えられる。そのため、過去の膨大な販売履歴データを最大限に活用しつつ、限られた価格変更テストの結果データを統合的に活用し、精度の高い予測モデルを構築するとともに、出品価格と販売価格の関係性についての分析を可能とすることが望ましい。

そこで本研究では、過去の販売履歴データと価格変更テストの結果データを統合的に活用し、出品価格と販売価格の関係性を分析するモデルを提案する。本研究ではまず、現行システムで設定された出品価格を変更した場合の販売価格を観察するため、価格変更テストを設計する。しかし、データ数に限りがあり、予測モデルの学習には不十分である。そのため、十分なデータ数のある過去の販売履歴データを最大限に活用し、予測モデルを構築する。一方、全てのアイテムを高い精度で予測することは難しく、アイテムの特徴によって予測の難易度が異なることが分かっている。そこで本研究では、クラスタリングを導入し、価格変更テストの結果データをモデルの検証データとして活用することで、高い精度で予測可能なアイテム群を発見する。そして、これらのアイテムを提案モデルを適用可能なアイテム群とし、出品価格を変更した場合の販売価格を予測する。また、出品価格と販売価格の関係性を捉え、適切な出品価格設定に関する分析を行う。提案分析モデルを実際のデータに適用することで、出品価格と販売価格の関係性の分析モデルとして有用であることを示す。

### 2. 対象事例 (EC サイト A) の概要

#### 2.1. 現行の出品価格設定と販売価格決定メカニズム

EC サイト A では各アイテムの販売実績に基づいて出品価格を設定するシステムを採用しており、各アイテムはこの現行システムに基づいて出品されている。その中には出品後、すぐに売れるアイテムもあれば、売れ残ってしまうアイテム

もある。EC サイト A では一定期間売れ残ったアイテムに対して自動的に値下げをするシステムを採用しているため、すぐに売れるアイテムの販売価格はもとの出品価格と同じ値となるのに対し、売れ残ったアイテムの販売価格は出品価格と大きく異なる傾向がある。前者は販売戦略的には望ましい結果とも言えるが、より高い出品価格でも即売された可能性もあり、出品価格の設定を見直すことで売上が向上する余地が考えられる。一方、後者の売れ残りアイテムは価格設定が適切でない可能性があり、価格設定を見直すことで売上の向上につながると考えられる。そのため、出品価格と販売価格の関係性をモデル化することで、適切な出品価格の設定に役立つと考えられる。

## 2.2. 蓄積している販売履歴データ

EC サイト A の販売履歴データにおける各アイテムにはブランド区分やアイテム区分、出品価格、販売価格やその際の値下げ率（オフ率）など様々な情報が付与されている。EC サイト A におけるデータの概要を表 1 に示す。

表 1: EC サイト A のデータ概要

変数名 (連続変数)	説明
定価	アイテムの定価
買取価格 *	商品の買取価格
出品価格 *	出品時の価格
販売価格	最終的に売れた価格
オフ率	値下げ率 (%)
変数名 (カテゴリ変数)	説明
ブランド区分 *	ブランドを階級化
カテゴリ区分	商品のカテゴリ情報
コンディション *	商品状態 (数段階)
性別	アイテムの対象性別

表 1 の \* 付きの変数は EC サイト A が独自の基準で定めている変数である。特に、定価とは各アイテムの元々の価格を表しており、独自に設定される出品価格とは異なるものである。ブランド区分は多数あるファッションブランドをそれぞれ数種類 (A,B,C,...) にグルーピングして区分化している。また、コンディションは商品状態 (ほぼ新品, 多少の使用感等) を数段階 (A,B,C,...) で表しており、カテゴリ区分はアイテムのカテゴリ情報 (アウター, トップス, ボトムスなど) を表している。

## 2.3. 販売価格予測モデルの難点と本研究の発想

販売履歴データ中の各アイテムは現行システムで値付けされて販売されたものであり、そのデータを用いて学習した販売価格予測モデルをもとに出品価格を変えた時の販売価格を予測した場合、その予測値が適切であるかどうかは定かではない。そのため、現行システムから価格を変更し、実際に出品・販売した結果のデータをモデルに導入することが有効であると考えられる。

そこで本研究では、価格変更テストを設計し、現行システムとは異なる価格設定による販売結果を獲得する。そして、過去の販売履歴データと価格変更テストの結果データを統合

的に活用するとともに、クラスタリングによって高精度で予測可能なアイテム群を発見する手法を提案する。この方法により、分析対象アイテム群を特定し、適切な出品価格設定に関する分析を行う。

## 3. 提案分析モデル

### 3.1. 提案分析モデルの概略

提案分析モデルでは、出品価格と販売価格の関係性を捉えることでアイテムごとの出品価格設定に関する分析を行うことを目的としており、本研究では提案分析モデル構築の一環として価格変更テストを実施した。そして、販売履歴データをもとに学習した予測モデルに結果データを組み合わせ、さらに、提案分析モデルの適用が可能なアイテムを発見するために潜在クラスモデル [3] を適用したクラスタリングを行う。そして最後に、分析対象アイテムの出品価格を変更した時の販売価格に関して分析する。

提案分析モデルは価格変更テスト、販売履歴データに基づく販売価格予測モデルの構築、価格変更テストの対象アイテムに対する予測誤差の算出とクラスタリングによる分析対象アイテムの発見、そして分析対象アイテムの価格設定に関する分析という 4 つのステップで構成される。

### 3.2. 提案分析モデルの詳細

#### 3.2.1. 価格変更テストの概要とその結果

EC サイト A の現行システムで設定された各アイテムの出品価格を変更し、実際に販売した結果を観察するため、本研究では 2018 年 12 月から 2019 年 1 月に出品予定のアイテムの一部を対象として価格変更テストを実施する。本来、全アイテムの価格を変更した販売結果をもとにモデル化することが望ましいが、事業面での制約上、ランダムに選出した 4,500 件のアイテムを対象とした。そして、1,500 件ずつに対し「0%(変更なし)・10%UP・20%UP」という 3 段階の価格変更を行ない、実際に出品した。価格変更テストの結果を表 2 に示す。

表 2: 価格変更テストの販売結果 (件)

変更法	0%	10%UP	20%UP	全体
対象数	1,500	1,500	1,500	4,500
販売済数	1,123	1,071	1,063	3,257
消化率	74.87%	71.4%	70.87%	72.38%

表 2 のように、「10%UP・20%UP」を適用したアイテムの販売消化率は「0%」より小さくなった。本研究では実際に販売された 3,257 件をモデルに活用する。

#### 3.2.2. 販売価格予測モデル

あるひとつの値を予測する回帰モデルとして重回帰分析やランダムフォレストなどが知られているが、本研究では高速かつ高精度なモデルとして Lightgbm [4] を用いる。Lightgbm は、複数の決定木を一つにまとめるアンサンブル学習の勾配ブースティングを用いており、高い精度を発揮することが知られている。また、学習データの特徴量を階級に分けてヒストグラム化することで、計算コストを軽減している。そのため、Lightgbm は予測精度や計算効率に優れており、大規模なデータセットにも適用可能なモデルとなっている。

### 3.2.3. 誤差算出とクラスタリングによる分析対象発見

続いて分析対象となるアイテムを発見する前準備として、価格変更テストの結果データを予測モデルに当てはめ、各アイテムに対する予測誤差  $g$  を以下の式 (1) を用いて算出する。

$$g = |h_{price} - \hat{p}_{price}| \quad (1)$$

ここで、 $h_{price}$  は実際の販売価格、 $\hat{p}_{price}$  は予測モデルによる予測販売価格を表している。また、 $\hat{p}_{price}$  は各アイテムの価格変更法に適した出品価格を用いて販売価格を予測したものである。例えば、「20%UP」して販売したアイテムに対しては、「20%UP」した出品価格を入力し、予測販売価格との誤差を算出する。この時、予測誤差が小さいアイテムは高精度で予測可能なアイテムと解釈でき、予測モデルによって出品価格を変更した場合にも販売価格を適切に予測できると考えられる。そのため、これらのアイテムは提案分析モデルにおいて高い精度で販売価格予測可能なアイテムと特定できる。

EC サイト A の各アイテムには多種多様な特徴があり、これらの特徴と予測誤差の関係性を考慮したクラスタリングを行うことが分析上効果的だと考えられる。いま、全  $N$  件の販売履歴データ中の各アイテムには  $M$  種類のカテゴリ区分  $\mathcal{P} = \{p_m : 1 \leq m \leq M\}$ 、 $L$  種類のブランド区分  $\mathcal{B} = \{b_l : 1 \leq l \leq L\}$ 、 $J$  種類のコンディション  $\mathcal{C} = \{c_j : 1 \leq j \leq J\}$ 、 $Q$  種類の性別情報  $\mathcal{S} = \{s_q : 1 \leq q \leq Q\}$ 、各アイテムの出品価格  $x \in R^+$ 、買取価格  $r \in R^+$ 、定価情報  $t \in R^+$  が付与されている。さらに、予測モデルによる各アイテムへの予測誤差を  $g \in R^+$  とする。本研究では、これらの変数が共起する事象  $X = (p_m, b_l, c_j, s_q, x, r, t, g)$  と捉え、これらの間に潜在クラスを仮定する。いま、 $K$  個の潜在クラス集合を  $Z = \{z_k : 1 \leq k \leq K\}$  とすると、確率モデル式は式 (2) で表される。

$$P(X) = \sum_{k=1}^K P(z_k) P(p_m|z_k) P(b_l|z_k) P(c_j|z_k) \cdot P(s_q|z_k) P(x|z_k) P(r|z_k) P(t|z_k) P(g|z_k) \quad (2)$$

各潜在クラス  $z_k$  のもとで、アイテム区分やブランド区分、コンディションや性別情報は多項分布、出品価格や買取価格、定価と予測誤差に関しては正規分布を仮定する。

さらに、潜在クラスのパラメータ推定に関して述べる。いま全  $N$  件のうち  $n$  番目の販売履歴データにおけるアイテム区分を  $i_n \in \mathcal{P}$ 、ブランド区分を  $v_n \in \mathcal{B}$ 、コンディションを  $u_n \in \mathcal{C}$ 、性別情報を  $d_n \in \mathcal{S}$  とする。また、出品価格を  $e_n$ 、買取価格を  $f_n$ 、定価を  $h_n$ 、予測誤差を  $y_n$  とすると、 $n$  番目のデータはこれらの共起  $(i_n, v_n, u_n, d_n, e_n, f_n, h_n, y_n)^T$  で表現できる。この時、全  $N$  件に対する対数尤度関数  $LL$  は式 (3) で表せる。

$$LL = \log \prod_{n=1}^N \sum_{k=1}^K P(z_k) P(i_n|z_k) P(v_n|z_k) \cdot P(u_n|z_k) P(d_n|z_k) P(e_n|z_k) \cdot P(f_n|z_k) P(h_n|z_k) P(y_n|z_k) \quad (3)$$

モデルの各パラメータを推定には、EM アルゴリズム [5] を用いて、対数尤度関数  $LL$  を最大化するように行われる。

### 3.2.4. 分析対象アイテムの価格設定に関する分析

最後に、分析対象アイテムの出品価格を変更した場合の販売価格を予測し、その変動を分析することで出品価格と販売価格の関係性を捉え、出品価格の設定に関して考察する。具体的には、価格変更テストと同様の方法（「0%(変更なし)・10%UP・20%UP」）で出品価格を変更し、それぞれの方法における販売価格を予測する。そして、最も販売価格が高くなる時の出品価格をより良い価格設定法とし、各アイテムに関して分析・考察する。

## 4. 実データ分析

EC サイト A の販売履歴データを活用した販売価格予測モデルの構築や価格変更テストの結果データを統合的に活用した提案分析モデルによる分析結果を以下で述べる。

### 4.1. 分析条件

予測モデル構築には EC サイト A における 2017 年 1 月から 2018 年 11 月の販売履歴データを用いる。データ件数は 1,132,114 件あり、本研究で予測モデルとして用いる Lightgbm のパラメータは、事前に行なったパラメータサーチにより設定した。

一方、価格変更テストの結果データは 2018 年 12 月から 2019 年 1 月に出品したのち、2019 年 5 月に取得したものであり、対象となった 4,500 件のうち実際に販売した 3,257 件を活用する。提案分析モデルを適用可能な分析対象アイテムの発見には潜在クラスモデルを用いており、そのクラス数は、解釈性の観点から  $K = 6$  とした。また、潜在クラスモデルにおいて各アイテムはそれぞれの潜在クラスに確率的に所属するが、本研究では分析効率の観点から、各アイテムの所属確率が最大のクラスにのみ所属するとしている。

### 4.2. 実データ分析結果

#### 4.2.1. 販売価格予測モデルとアイテムクラスタリング

Lightgbm による販売価格予測モデルによる価格変更テストの対象アイテムに対する予測誤差とアイテムの特徴の共起を考慮した潜在クラスモデルによりクラスタリングを行ない、各潜在クラスに対してアイテム区分やブランド区分などの生起確率をもとに各クラスの特徴を分析した。以下の表 3 に各クラスの代表的な特徴を示す。

表 3: 各クラスのアイテムの代表的な特徴

k	特徴
1	トップス/ローブランド/商品状態 C,D
2	トップスやアウター/ローブランド/商品状態 A,B
3	アウター/ミドルブランド/商品状態 C,D
4	トップスやボトムス/ローブランド/商品状態 A,B
5	アウターやトップス/ミドルブランド/商品状態 A,B
6	アウター/ハイブランド/商品状態 C,D

表 3 のように、各潜在クラスでアイテムのカテゴリ区分やブランド区分、コンディションに関して異なる特徴が見られた。特に、クラス間で似たようなブランド区分、コンディションであってもカテゴリが異なるクラスが存在する。さらに、各クラスの特徴の違いを見出すため、各クラスに所属す

るアイテムの平均出品価格や平均予測誤差を表 4 に示す。

表 4: 各クラスのデータ数と平均誤差

k	RMSE	平均出品価格	誤差率 (%)	データ数
1	508	4,630	10.97	930
2	1,276	5,786	22.05	593
3	1,564	8,598	18.19	534
4	2,260	7,426	30.43	389
5	3,664	9,698	37.78	358
6	4,258	10,948	38.89	453
計				3,257

表 4 に示すように各潜在クラスにおけるアイテムの平均出品価格は大きく異なっている。ここで、平均出品価格に対する平均予測誤差の大きさを誤差率として算出し、この誤差率をもとに予測精度を評価する。誤差率を基準とした場合、class1 は予測しやすいクラス、class5 や class6 は予測が困難なクラスであると考えられる。つまり、class1 に属するアイテムは学習した Lightgbm によって予測が可能であると解釈でき、これらのアイテムを分析対象として出品価格を変更した場合の販売価格を予測する。

#### 4.2.2. 出品価格と販売価格の関係性分析

分析対象となったアイテムに対して、その出品価格を様々に変更（「0%、10%UP、20%UP」）した場合のそれぞれの販売価格を予測し、その変動を分析する。以下では分析結果として例を 2 つ示す。まず、「ブランド区分 D のアウター」に着目した分析結果を表 5 に示す。

表 5: 分析結果例 1 (ブランド区分 C のアウター)

状態	平均出品価格 (円)	0%	10%UP	20%UP
A	6,518	3,082	<b>3,544</b>	3,157
B	6,350	3,015	3,259	<b>3,449</b>
C	3,167	1,779	1,780	<b>1,909</b>
D	1,600	1,090	902	<b>1,173</b>

表 5 では、各アイテムの商品状態ごとに出品価格を変更した場合の予測販売価格を表しており、予測販売価格が最も高くなった部分を太字で表している。表 5 より、同一のブランド区分でも商品状態の違いによって予測販売価格が異なることがわかる。特に、状態 A のアイテムは出品価格を「10%UP」した場合に最も販売価格が高くなった。一方、状態 B、C、D のアイテムに関しては出品価格を「20%UP」した場合に販売価格が高くなるアイテムであることがわかった。また、状態 D のアイテムに関しては「10%UP」すると販売価格が下がってしまう結果となった。次に、「ブランド区分 D のアウター」に着目した分析結果を表 6 に示す。

表 6: 分析結果例 2 (ブランド区分 D のアウター)

状態	平均出品価格	0%	10%UP	20%UP
A	10,240	3,391	4,385	<b>4,944</b>
B	2,900	<b>1,631</b>	1,449	1,485
C	3,175	1,621	1,601	<b>1,688</b>

表 6 に示すように、状態 A のアイテムは、出品価格を「20%UP」をした場合に最も販売価格が高くなった。一方、状態 B のアイテムに関しては出品価格を「10%UP・20%UP」した場合に販売価格が下がってしまうため、「0%(変更なし)」

が適切な設定方法であると考えられる。また、状態 C のアイテムは「10%UP」すると販売価格が下がるが、「20%UP」した場合に販売価格が高くなった。このように、アイテムの特徴によって出品価格と販売価格の関係性が様々であり、それぞれで適切な設定法が存在することが明らかとなった。

## 5. 考察

本研究では既存の販売履歴データと、新たに価格変更テストにより追加されたデータの双方を効果的に活用する方法を検討した。そして、提案分析モデルが適切な出品価格に関する分析を可能とすることを示した。一方、本研究は EC サイト A における出品価格設定システムの考案が最終的な目標であるが、提案分析モデルを実務で利用するにはさらに検討が必要である。本研究で設計した価格変更テストは 2018 年 12 月から 2019 年 1 月に実施しており、季節性の強いアイテムが多く含まれている。そのため、本モデルがビジネス的に継続的に利用可能であることを示すには、年間を通じた出品実験を行なう必要がある。その際、新規出品アイテムに対して提案分析モデルを適用し、価格を変更することが適切であるとされたアイテム群の出品価格を変更し、予測結果と実際の販売結果を比較することで、さらに提案モデルの妥当性の検証が可能となる。また、本研究では価格を 3 段階に変更して出品実験を実施したが、より適切な価格設定を探索するためには、価格変更の幅を広げることも有効と考えられる。

## 6. まとめと今後の課題

EC サイト A における膨大な販売履歴データと独自に実施した価格変更テストの結果データを統合的に扱い、出品価格と販売価格の関係性をモデル化する分析モデルを提案した。特に、過去の販売履歴データをもとに構築した予測モデルに価格変更テストの結果データを組み込み、さらにクラスタリングによって、高い精度で予測可能なアイテムの特定を可能とした。提案分析モデルを新規出品予定のアイテムに適用することで、出品価格設定に役立つと考えられる。今後の課題として、アイテムの季節性を考慮したより詳細な分析や分析結果の実応用などがあげられる。

### 参考文献

- [1] 石垣 司, 竹中 毅, 本村陽一, “日常購買行動に関する大規模データの融合による顧客行動予測システム,” 人工知能学会論文誌, Vol. 26, No. 6, pp. 670-681, 2011.
- [2] 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,” 人工知能学会論文誌, Vol. 60, No. 4, pp. 1151-1161, 2019.
- [3] T. Hofman, “Probabilistic latent semantic indexing,” *22th Annual International ACM SIGIR*, pp. 55-57, 1999.
- [4] G. Ke, Q. Meng, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Y. Liu., “Lightgbm: A highly efficient gradient boosting decision tree,” *In Advances in Neural Information Processing Systems*, pp. 3149-3157, 2017.
- [5] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, pp. 1-38, 1977.