混合回帰モデル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の提案と価格設定に関する研究

情報数理応用研究

5215C029-1　仁ノ平将人
指導教員　後藤正幸

Prediction Model of Selling Prices of Second-hand Fashion Items Based on Mixture Regression Model

NINOHIRA Masato

1　研究背景・目的

近年の情報技術の発展により，EC(電子商取引) サイトを通じた商品の購買が普及している．特にアパレル商材に関する EC サイトの市場規模は，ここ数年で大幅に拡大し続けている．このように，ファッションアイテムに関する購買履歴データが大量に蓄積され始めたことを契機に，顧客の購買行動分析や自動画像タグ付与などを目的とした，機械学習手法の活用が模索し始められている．

本研究で対象とする某ファッション EC サイト A では，ユーザから中古ファッションアイテムを買取り，値付けを行い再販売，出品を行なっている．この EC サイトでは，売れ残りを防ぐため，出品アイテムに対し一定のアルゴリズムで自動的に値下げをする仕組みを採用している．このビジネスモデルにおいて，各アイテムに対し，ある価格 (出品価格) で出品をすると，最終的にいくらで販売されるか (販売価格) を予測することは，値付けシステムの構築や経営戦略を考える際に重要である．

一方，EC サイト A の出品アイテムに着目すると，カテゴリや色，素材といった様々な特徴を持ったアイテムが出品されている．このような特徴量をもとに販売価格を予測するモデルとして最も基本的な方法として，重回帰分析が知られている．しかし，出品アイテムが持つ特徴量の種類があまりにも膨大であるために，単純に全データに対して重回帰分析を行なったとしても，精度の高い予測モデルを得ることは難しい．

また，ファッションアイテムは一般に，流行や季節に影響を受けやすいと考えられる．実際，EC サイト A に出品されているアイテムの季節ごとの値下がりの傾向を分析すると，季節により出品から販売までのリードタイムが長く，元の出品価格から大幅に値下げされて販売されるアイテムや，逆に販売までのリードタイムが短く，値下げが生じにくいアイテムが混在していることが明らかになった．これより販売価格の予測を行う際に，あらかじめ季節ごとに値下がり率 (以下，オフ率) が大きいと考えられるアイテムと，そうでないアイテムといった基準でクラスタリングを行い，クラスタごとに異なる回帰モデルを適用することが有効であると考えられる．

他方，データのクラスタリング手法として確率的にクラスタリングを行う潜在クラスモデル [1] が知られている．また，潜在クラスモデルと重回帰分析を組み合わせ，データの構造により異なる複数の回帰式を混合する手法として混合回帰モデル [2] がある．本研究では，この EC サイトにおける出品アイテムの販売価格予測モデルの構築のために，混合回帰モデルの考え方をベースに，アイテムの特徴，季節ごとのオフ率の傾向をもとに潜在クラスモデルを用いてクラスタリングを行なった後に，データの各潜在クラスへの所属確率を用いて潜在クラスごとに回帰式を構築する推定モデルを提案する．さらに，提案モデルが EC サイト A の実購買データにおいて販売価格を予測するモデルとして有効なモデルあることを示し，提案モデルから得られる知見をもとにした出品価格の決定方法についての考察を行う．

2　事前分析

一般に，ファッションアイテムは流行や季節に敏感な商材であることが知られている．EC サイト A の出品アイテムに対し，季節ごとのオフ率の傾向で出品アイテムの分類が可能であるならば，その傾向に応じた異なる回帰式を構築することは，販売価格を予測する上で有効な手段であると考えられる．そこで以下では，季節ごとにオフ率の傾向で出品アイテムを分類可能かという仮説の検証を行う．

まず，各アイテムの一定期間の販売数量のうち，大きく割引が生じた数量の割合を求める．いま，N 種類のアイテムカテゴリ集合を $\mathcal{I} = \{i_n : 1 \leq n \leq N\}$ とし，アイテムカテゴリ i_n に対し，一年間を M 期に区切ったときの m 期 $(1 \leq m \leq M)$ における 50%以上のオフ率で販売された数量の割合を q_{nm} とする．いま，季節ごとのアイテムのオフ率の傾向を分析するために，各アイテム i_n を，この q_{nm} を要素とする M 次元のベクトル $\boldsymbol{q}_n = (q_{n1}, \ldots, q_{nm}, \ldots, q_{nM})^{\mathrm{T}}$ で表し，これらに k-means 法を適用する．ここでは，1 年間を 12ヶ月に区切り $(M = 12)$，クラスタ数を 6 としたときの各クラスタの中心ベクトル $\boldsymbol{\nu}_k = (\nu_{k1}, \ldots, \nu_{km}, \ldots, \nu_{kM})^{\mathrm{T}} (k = 1, 2, \cdots, 6)$ により，各クラスタに所属するアイテムの季節ごとのオフ率の傾向を解釈した結果を表 1 に示す．

表 1: 各クラスタに所属するアイテムのオフ率の傾向

k	傾向
1	年間を通じて低いオフ率
2	年間を通じて一定のオフ率
3	年間を通じて高いオフ率
4	春先に高いオフ率
5	冬に低いオフ率
6	冬に高いオフ率

表 1 より，クラスタごとに所属するアイテムの季節ごとのオフ率の傾向が異なることがわかる．これにより，季節ごとのオフ率の傾向をもとに出品アイテムを分類できることがうかがえる．したがって，出品アイテムに対し上記のような条件でクラスタリングを行い，クラスタ別に回帰式を構築することは，販売価格を推定するために有効であることが示唆される．

3　提案モデル

3.1　提案モデルへの着想

前述の通り，EC サイト A では，様々な特徴を持ったアイテムが出品されている．このために，単一の重回帰モデルを適用しても，高い精度の予測販売価格を得ることは難しい．また，表 1 より，出品アイテムの季節ごと

のオフ率の傾向を分析すると，秋にオフ率が高くなりやすいアイテムや，年間を通じて一定のオフ率が維持されやすいアイテムといったように，季節によるオフ率の傾向の違いにより，アイテムの分類が可能であることが明らかになった．以上より，アイテムのカテゴリ，色，素材といった特徴量に加え，季節ごとのオフ率の傾向をもとにデータをクラスタリングした後に，クラスタごとに回帰式を構築することで，より精度の高い予測販売価格が得られることが期待される．

他方，データのクラスタリング手法として確率的にクラスタリングを行う潜在クラスモデル [1] の有用性が知られている．そこで，本研究では新たな潜在クラスモデルを用いて，アイテムの特徴量と季節ごとのオフ率の傾向でデータを分類し，潜在クラスごとに異なる回帰式を仮定するモデルを提案する．

また，実応用を考えた場合，モデルの学習に用いられた過去の出品データのみではなく，販売価格が未知の新規出品データに対しても高い精度の予測販売価格を得られることが望ましい．そこで，新規出品データに対し，学習で得られたで得られた各潜在クラスへの所属確率と，各潜在クラスにおける回帰式の出力を算出し，これらを混合することで，新規出品データの予測販売価格を推定することを考える．

以上の議論をまとめると，提案モデルはアイテム属性や季節ラベルを用いた潜在クラスモデルによるクラスタリングと潜在クラスによる混合回帰モデルの 2 フェーズから構成される．以降ではこれらの詳細について説明を行う．

3.2 提案モデルの詳細

3.2.1 潜在クラスモデルによるクラスタリング

まず，アイテム属性や季節ラベルを用いた潜在クラスモデルによるクラスタリングについて述べる．いま，全 L 件の出品履歴データに出現する M 種類の季節ラベルを $\mathcal{S} = \{s_m : 1 \leq m \leq M\}$ とする．さらにアイテムの色や素材といった $j(\leq J)$ 番目の補助情報の要素集合を $\mathcal{A}_j = \{a^j_{v_j} : 1 \leq v_j \leq V_j\}$ とする．例えば，ある j において $\mathcal{A}_j$ をアイテムの色の集合とすると，V_j は色の種類数であり，$a^j_{v_j}$ は何色かを表す．出品アイテムの J 種類の補助変数を表すために，J 次元のベクトル $\boldsymbol{o} = (o_1, \ldots, o_j, \ldots, o_J)^{\mathrm{T}} (o_j \in \mathcal{A}_j)$ を定義する．また，各アイテムの出品価格を $b \in \mathbb{R}^+$，オフ率を $c \in \mathbb{R}^+$ とする．提案モデルでは，アイテムを季節ごとのオフ率の傾向とその属性をもとにクラスタリングを行うために，1 つの出品データをこれらの共起 $(i_n, s_m, \boldsymbol{o}, b, c)^{\mathrm{T}}$ と捉え，それらの間に潜在クラスを仮定する．K 個の潜在クラス集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ としたとき，提案モデルの確率モデルは式 (1) で表される．なお，$\delta(x, y)$ は $x = y$ のとき 1，それ以外は 0 を取る指示関数とする．

$$F(\, i_n \,, s_m, \boldsymbol{o}, b, c) = \sum_{k=1}^{K} P(z_k) P(i_n|z_k) P(s_m|z_k) \cdot P\,(b|z_k) P(c|z_k) \prod_{j=1}^{J} \prod_{v_j=1}^{V_J} P(a^j_{v_j}|z_k)^{\delta(o_j, a^j_{v_j})} \quad (1)$$

いま，各潜在クラス z_k のもとでのアイテムの出現確率 $P(i_n|z_k)$, 季節ラベルの出現確率 $P(s_m|z_k)$, j 番目の補助情報の出現確率 $P(a^j_{v_j}|z_k)$ には多項分布，出品価格 b の出現確率密度 $P(b|z_k)$，オフ率 c の出現確率密度 $P(c|z_k)$ にはそれぞれ，正規分布 $N(\mu_k, \sigma^2_k)$, $N(\lambda_k, \varphi^2_k)$ を仮定する．すなわち，μ_k は潜在クラス z_k に所属するデータの出品価格の平均値，λ_k はオフ率の平均値である．

3.2.2 潜在クラスによる混合回帰モデル

次に，潜在クラスによる混合回帰モデルについて述べる．回帰式で用いる出品価格やアイテムカテゴリなどをダミー変数で表した説明変数を $\boldsymbol{x} = (1, x_1, \ldots, x_d, \ldots, x_D)^{\mathrm{T}}$ としたとき，提案モデルでは，各潜在クラス z_k ごとに異なる回帰係数 $\boldsymbol{\beta}_k = (\beta_{0k}, \beta_{1k}, \ldots, \beta_{dk}, \ldots, \beta_{Dk})^{\mathrm{T}}$ を仮定する．さらに，混合回帰モデル [2] の考え方を援用し，各潜在クラスの回帰式の出力 $\boldsymbol{\beta}^{\mathrm{T}}_k \boldsymbol{x}$ をクラスタリングの際に得られるデータの各潜在クラスへの所属確率 $P(z_k|i_n, s_m, \boldsymbol{o}, b, c)$ で重みを付けて混合することで，販売価格 y が生成されるモデルを仮定する．

$$y = \sum_{k=1}^{K} P(z_k|i_n, s_m, \boldsymbol{o}, b, c) \boldsymbol{\beta}^{\mathrm{T}}_k \boldsymbol{x} + \varepsilon \quad (2)$$

ただし，ε は，平均 0 分散 σ^2 の正規分布に従う誤差項である．

3.3 提案モデルのパラメータの学習

本節では，提案モデルにおけるパラメータの学習方法について述べる．

3.3.1 潜在クラスモデルによるクラスタリング

まず，潜在クラスモデルによるクラスタリングのパラメータ推定について述べる．l 番目の出品データにおけるアイテムカテゴリを $t_l(\in \mathcal{I})$，出品日の季節ラベルを $u_l(\in \mathcal{S})$，j 番目の補助情報を $w_{lj}(\in \mathcal{A}_j)$，$\boldsymbol{w}_l = (w_{l1}, \ldots, w_{lj}, \ldots, w_{lJ})^{\mathrm{T}}$ を l 番目の出品データの J 種類の補助情報を表すベクトルとする．さらに，出品価格を g_l，オフ率を h_l(共に連続値) とすると，l 番目の出品データはこれらの共起 $(t_l, u_l, \boldsymbol{w}_l, g_l, h_l)^{\mathrm{T}}$ で表現できる．このとき，全 L 件の出品データに対する対数尤度関数 LL は以下の式 (3) で表される．

$$LL = \log \prod_{l=1}^{L} \sum_{k=1}^{K} P(z_k) P(t_l|z_k) P(u_l|z_k) \cdot P(g_l|z_k) P(h_l|z_k) \prod_{j=1}^{J} P(w_{lj}|z_k) \quad (3)$$

このモデルのパラメータは EM アルゴリズム [3] を用いて対数尤度関数 LL を最大化するように，以下の更新式を収束するまで繰り返すことで推定する．

[E-step]

$$P(z_k|t_l, u_l, \boldsymbol{w}_l, g_l, h_l) \propto P(z_k) P(t_l|z_k) P(u_l|z_k) \cdot P(g_l|z_k) P(h_l|z_k) \prod_{j=1}^{J} P(w_{lj}|z_k) \quad (4)$$

[M-step]

$$P(z_k) \propto \sum_{l=1}^{L} \alpha_{kl} \quad (5)$$

$$P(i_n|z_k) \propto \sum_{l=1}^{L} \alpha_{kl} \delta(t_l = i_n) \quad (6)$$

$$P(s_m|z_k) \propto \sum_{l=1}^{L} \alpha_{kl} \delta(u_l = s_m) \quad (7)$$

$$P(a^j_{v_j}|z_k) \propto \sum_{l=1}^{L} \alpha_{kl} \delta(w_{lj} = a^j_{v_j}) \quad (8)$$

$$\mu_k = \frac{\sum_{l=1}^{L} \alpha_{kl} g_l}{\sum_{l=1}^{L} \alpha_{kl}} \tag{9}$$

$$\sigma_k^2 = \frac{\sum_{l=1}^{L} \alpha_{kl} (g_l - \mu_k)^2}{\sum_{l=1}^{L} \alpha_{kl}} \tag{10}$$

$$\lambda_k = \frac{\sum_{l=1}^{L} \alpha_{kl} h_l}{\sum_{l=1}^{L} \alpha_{kl}} \tag{11}$$

$$\varphi_k^2 = \frac{\sum_{l=1}^{L} \alpha_{kl} (h_l - \lambda_k)^2}{\sum_{l=1}^{L} \alpha_{kl}} \tag{12}$$

なお，数式の簡略化のために，α_{kl} を l 番目の出品データの潜在クラス z_k への所属確率 $P(z_k|t_l, u_l, \boldsymbol{w}_l, g_l, h_l)$ とした．

3.3.2 潜在クラスによる混合回帰モデル

次に，潜在クラスによる混合回帰モデルのパラメータ推定について述べる．l 番目のデータの，回帰式で用いる説明変数を $\boldsymbol{x}_l = (1, x_{l1}, \ldots, x_{ld}, \ldots, x_{lD})^{\mathrm{T}}$，販売価格を y_l としたとき，重み付け重回帰モデル [4] の考え方を援用し，各潜在クラス z_k における回帰式のパラメータ $\boldsymbol{\beta}_k$ は，各データの各潜在クラス z_k への所属確率で重み付けされた二乗誤差を最小にするように，以下の式 (13) で推定する．

$$\hat{\boldsymbol{\beta}}_k = \arg\min_{\boldsymbol{\beta}_k} \sum_{l=1}^{L} \alpha_{kl} (y_l - \boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{x}_l)^2 \tag{13}$$

3.4 新規出品データの販売価格の予測

前述の通り，予測モデルの構築においては，販売価格が未知の新規出品データに対しても高い精度の予測販売価格を得られることが望ましい．そこで，新規出品データに対し，学習により得られた各潜在クラスにおける回帰式の出力を各潜在クラスへの所属確率を用いて混合することで，新規出品データに対しても予測値の算出を行う．いま，新規出品データ数を L' とし，$l'(\leq L')$ 番目の新規データのアイテムのカテゴリを $t_{l'}(\in \mathcal{I})$，季節ラベルを $u_{l'}(\in \mathcal{S})$，j 番目の補助情報を $w_{l'j}(\in \mathcal{A}_j)$，出品価格を $g_{l'}$ とする．このデータに対して，オフ率が未知であることに留意して，学習により得られた各潜在クラスへの所属確率 $P(z_k|t_{l'}, u_{l'}, \boldsymbol{w}_{l'}, g_{l'})$ を以下の式 (14) で求める．

$$\begin{aligned} P(z_k|t_{l'}, u_{l'}, \boldsymbol{w}_{l'}, g_{l'}) \propto \\ P(z_k) P(t_{l'}|z_k) P(u_{l'}|z_k) \cdot \\ P(g_{l'}|z_k) \prod_{j=1}^{J} P(w_{l'j}|z_k) \end{aligned} \tag{14}$$

さらに，予測対象である l' 番目の新規出品データの説明変数を $\boldsymbol{x}_{l'} = (1, x_{l'1}, \ldots, x_{l'd}, \ldots, x_{l'D})^{\mathrm{T}}$ とすると，潜在クラス z_k における回帰式の出力 $\hat{y}_{l'k}$ を $P(z_k|t_{l'}, u_{l'}, \boldsymbol{w}_{l'}, g_{l'})$ で混合することで最終的な予測販売価格 $\hat{y}_{l'}$ が得られる．

$$\hat{y}_{l'k} = \hat{\boldsymbol{\beta}}_k^{\mathrm{T}} \boldsymbol{x}_{l'} \tag{15}$$

$$\hat{y}_{l'} = \sum_{k=1}^{K} P(z_k|t_{l'}, u_{l'}, \boldsymbol{w}_{l'}, g_{l'}) \hat{y}_{l'k} \tag{16}$$

4 実験

以下では，提案モデルの有効性を示すために，EC サイト A に蓄積された実データを用いて，提案モデルの予測精度について評価を行う．また，提案モデルを用いて得られたパラメータの分析についても行い，考察を与える．

4.1 実験概要

実験データとして，2016 年に EC サイト A 上で取引された，某ファッションブランドの出品履歴データを用いる．データの件数は 67,211 件 ($L = 67,211$) であり，販売されているアイテムカテゴリ数は 79 種類 ($N = 79$) である．また，潜在クラスによるクラスタリングを行う際に用いる季節ラベル s_m には，アイテムの出品月を用いる ($M = 12$)．さらに，アイテムの補助情報 A_j には色，素材などの 8 種類 ($J = 8$) とし，回帰式に用いる説明変数 $\boldsymbol{x}$ には，以下の表 2 に示す 175 種類の変数 ($D = 175$) を用いた．

表 2: 回帰式で用いた説明変数一覧

説明変数	ユニーク数	ダミー変数 or 連続値
アイテム	78	ダミー変数
出品月	12	ダミー変数
補助情報	84	ダミー変数
出品価格	1	連続値

評価指標として，テストデータに対する平均二乗誤差 (MSE) と，モデルの当てはまりを評価する R^2 値の 2 つの指標を用いて評価を行う．また，新規出品データへの評価を行なうため，10-fold クロスバリデーションによる実験を行なった．比較手法として，データのクラスタリングを行わない単一の重回帰分析と，回帰問題に対し高い予測精度を示すことで知られているランダムフォレスト回帰 (以下，RF) の 2 つを用いた．なお，RF の木の数に関しては事前実験を行い，最も精度の高かった 150 を採用した．

4.2 結果

対立手法，提案手法の潜在クラス数 K を変えたときに得られる MSE と R^2 値を以下の図 1，2 に示す．

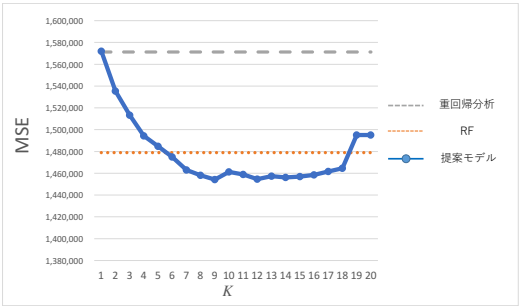

図 1: 潜在クラス数を変化させたときの MSE の結果

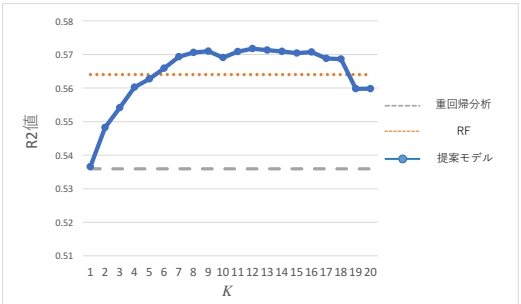

図 2: 潜在クラス数を変化させたときの R^2 値の結果

図 1，2 より，提案モデルは一定の潜在クラス数 K のときに，比較手法よりもよい評価値が得られることがわかる．また，K の値が大きくなった際に低い評価値を示しているのは過学習が生じたためと考えられる．この結果より，潜在クラス数 K の設定に留意すれば，提案モデルは当該 EC サイトにおけるアイテムの販売価格を推定

するモデルとして有効なモデルであることがわかる．以降では，MSE，R^2 値共に最もよい値が得られた潜在クラス数 $K = 9$ について，結果の分析と考察を行う．

4.3 得られた結果の分析，考察

本節では，実験の結果，最も高い精度を示した潜在クラス数 $K = 9$ の提案モデルで得られたパラメータについて考察する．ここでは，各潜在クラスに所属しているアイテムの特徴並びに，各潜在クラスごとに説明変数が目的変数である販売価格に与える影響力の 2 つの観点から分析を行う．

まず，各潜在クラスに所属しているアイテムの特徴を分析するために，各潜在クラスに対し，アイテムカテゴリの生起確率 $P(i_n|z_k)$, 及び色や素材といった補助情報の生起確率 $P(a^j_{v_j}|z_k)$ をもとに，生起確率が高い要素に対し解釈を与えた結果を以下の表 3 に示す．

表 3: 各潜在クラスに所属するアイテムの特徴

k	特徴
1	デニムやスカート
2	メンズのカーディガンなどの上着
3	レディースのパンツ類
4	バッグなどの小物類
5	レディースのサロペット・ジャケット類
6	コート類
7	メンズの T シャツ類
8	レディースの高品質のカットソー
9	カットソーなどの人気商品

また，各潜在クラスで，どの月に出品されたアイテムが出現しやすいかを表す $P(s_m|z_k)$ を以下の図 3 に示す．

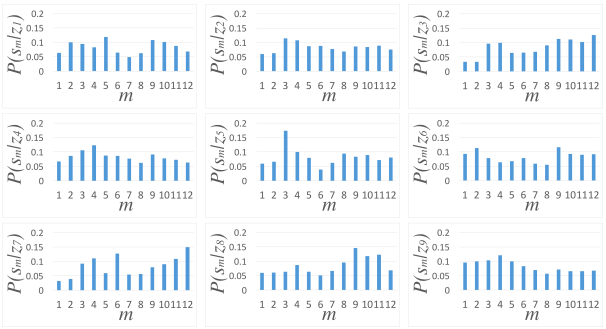

図 3: 各潜在クラスにおける出品月の出現確率

表 3，図 3 より，潜在クラス 1 には春秋に出品されるデニム，スカートが高い確率で所属しているというように，潜在クラスごとに異なる特徴を持ったアイテムが属していることがわかる．

次に，説明変数が目的変数に与える影響力の分析をするために，学習により得られた各潜在クラスの回帰係数 $\boldsymbol{\beta}_k$ についての分析を行う．説明変数の代表例として，出品価格の回帰係数と t 値を表 4 に示す．

表 4: 各潜在クラスの出品価格の回帰係数と t 値

k	出品価格の回帰係数	t 値
1	0.685	9.92
2	1.085	4.32
3	0.376	2.81
4	0.661	15.5
5	0.047	0.05
6	0.704	33.2
7	0.259	1.72
8	0.168	-
9	0.332	-

なお，潜在クラス 8，9 は，同じ出品価格を持つデータのみが所属したため，t 値の算出が不可能であり，-で表記した．

表 4 より，各潜在クラスにおいて異なる出品価格の係数や t 値が得られていることがわかる．例えば，潜在クラス 2 では他クラスに比べ，出品価格の回帰係数が大きいので，出品価格を変化させた際に他の潜在クラスに比べ販売価格の変動が大きくなることが考えられる．このように提案モデルを用いることで，販売価格に対する各説明変数の各潜在クラスごとに異なる影響度の定量化が可能であると言える．

5 出品価格の設定に関する考察

以上の議論により，提案モデルは現行の出品価格のシステムに対し高い精度で販売価格の予測が可能であり，さらに，潜在クラスごとに異なる販売価格に対する各説明変数の影響度を定量化できることが示せた．しかし，本研究の最終目標は，当該ビジネスモデルにおける出品価格の値付けシステムの考案である．そこで，提案モデルにより得られた出品価格の t 値に着目すると，この t 値が低くなっている潜在クラスに所属するアイテムは，販売価格の予測の際に出品価格の重要度が低いと解釈できる．このことから，これらのアイテムでは，出品価格と販売価格の相関が低く，実際に取引されそうな価格を想定した効果的な値付けができていない可能性がある．そこで，現行の出品価格と異なる出品価格で出品させた際に，各潜在クラスに所属するアイテムの販売価格を分析することで，値付けシステムの構築に対し非常に重要な情報を得られることが期待される．このように，現行の出品価格の設定と異なる価格で出品させた際に，提案モデルにより得られた各潜在クラスに所属するアイテムが，どのような価格で取引・販売が行われるかを実証的に評価することが求められる．

6 まとめ

本研究では，ファッション EC サイトにおいて，予めアイテムの基本情報や季節ごとのオフ率の傾向をもとに潜在クラスモデルを用いて確率的にクラスタリングを行い，潜在クラスごとに販売価格を目的関数とする回帰式を構築する予測モデルを提案した．提案モデルは単一の重回帰分析や，一般的に高い精度を持つ機械学習手法である RF よりも，新規出品データに対して高い予測精度を示すことが明らかになった．また，提案モデルにより得られたパラメータを分析することで，それぞれの潜在クラスに所属するデータに対し，販売価格に対する有効な要因の分析が可能であることを示した．今後の課題として，現行の出品価格と異なる価格で出品する実験を行い，その結果をもとに，出品価格の設定ルールについて検討する必要がある．

参考文献

[1] T. Hofmann, “Probabilistic Latent Semantic Indexing,” *22nd Annual International ACM SIGIR*, pp. 55–57, 1999.

[2] S. Faira, G. Soromenho, “Fitting mixtures of linear regressions,” *Statistical Computation and Simulation,* Vol. 80, No. 2, pp. 201–225, 2010.

[3] A. P. Dempster, N. M. Laird, 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.

[4] W. S. Cleveland, S. J. Devlin, “Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting,” *American Statistical Association*, Vol. 83, No. 403, pp. 596–610, 1988.