

# 機械学習による中古車落札価格の要因分析及び異常検知

工藤 大輝<sup>1,a)</sup> 福西 亮介<sup>2,b)</sup> 黛 広樹<sup>2,c)</sup> 鈴木 智也<sup>1,d)</sup>

**概要:** 本研究では、オートオークションで落札された中古車データを対象として、重回帰式ベースのヘドニックアプローチと非線形モデルの XGBoost に基づいて個車価格推定モデルを構築し、正答率と各特徴量の重要度について比較分析した。中古車の各特徴量は非線形関係を有していることから XGBoost の正答率が良好であり、重回帰式のみでは捉えられない特徴量の重要性を確認した。またオートオークション会場は全国各地にあり、地域や規模に応じて落札価格に「割安」や「割高」などの特徴が生じる可能性がある。そこで非線形な個車価格推定モデルを利用することで、落札価格の異常検知や裁定取引に応用できることを確認した。

## Factor Analysis and Abnormality Detection of Used Car Contract Prices by Machine Learning

### 1. はじめに

中古車はユーザーから買取店へ流れた後、主にオートオークションを通じて販売店に売却される。そこで買取店はオートオークションでの落札価格を予想して、ユーザーに対し買取価格を提示する必要がある。その際に合理的な買取価格を算出する査定システムが必要となるが、中古車や住宅のような不動産は同一物が存在しないため価格モデルの構築が難しい。そこで伝統的には、重回帰式に基づくヘドニック法 [1-5] によって不動産を構成する各属性の重要度を推計し、それらの線形結合によって価格モデルを構築する方法が主流である。たとえば、国土交通省が公表する不動産価格指数 [6] は、このヘドニック法をベースに算出している。しかし近年のソフトコンピューティング技術の進歩により、ビッグデータや機械学習を積極的に活用する事例 [7,8] が増えており、関連学会誌において不動産テックの特集が組まれている (人工知能学会誌, 日本機械学会誌, 日本不動産学会誌, など)。しかしいずれも住宅販売を対象

にしており、中古車への応用可能性について検討されていない。そこで本研究では、全国のオートオークションで落札された中古車の長期実績 (約 160 万件) をビッグデータとみなし、これを機械学習することで価格査定モデルの精度を向上できるか検証する。

機械学習を導入する利点は、各属性が有する非線形性への対応である。たとえば排気量が多いほど価格を上昇させる要因であるが、税金や保険料が高くなるため非線形的に作用する [9]。また年式においても古いほど価格を低下させる要因であるが、あえてビンテージ車を好むユーザー層も存在する。これらの価格に与える影響力においても、メーカーや車種のブランドによって複雑に変化すると考えられる。同様に中古住宅においても価格構造の非線形性が報告されている [10]。一方、機械学習を導入する短所として説明力の低下が考えられるが、各属性の重要度 (Feature Importance) を評価できる決定木ベースのアルゴリズムを用いる。さらに従来ヘドニック法による偏回帰係数も参照することで、重要度のポジネガ極性を評価する。

機械学習によって価格査定モデルを強化できれば、異常検知にも活用できる。価格査定モデルで個車状態に応じた適性価格を算出し、実際の落札価格が適性価格より乖離するほど異常とみなす。オートオークションの落札価格は大人気による合意 (集合知) であるため合理性が高いと考えられるが、地域や季節によって割安や割高に偏る可能性がある。たとえば雪国では四駆の需要が高いため、需要過剰

<sup>1</sup> 茨城大学大学院理工学研究科  
Graduate School of Science and Engineering, Ibaraki University, Hitachi, Ibaraki 316-8511, Japan

<sup>2</sup> 株式会社プロトコーポレーション  
PROTO CORPORATION, Shinjuku, Tokyo 160-0023 Japan

a) 19nm4341@vc.ibaraki.ac.jp

b) r-fukunishi@proto-g.co.jp

c) h-mayuzumi@proto-g.co.jp

d) tomoya.suzuki.lab@vc.ibaraki.ac.jp

により割高で取引されるかもしれない。さらに地方のオークション会場ほど規模が小さく市場参加者が少ないため、不合理な価格形成が顕著かもしれない。もしそうならば、割安な地域で現物を仕入れ、割高地域で販売するといったビジネスも考えられる。そこで価格査定モデルの応用として、落札価格の異常割安・割高の検知を試みる。

## 2. 重回帰式による個車落札価格の推定

ヘドニックアプローチは、財の価格をその財の属性を基に回帰して属性の計算価格を推定し、属性の量と計算価格の積和によって、多様な財の価格を評価する方法論である [3]。本研究では、過去に落札された中古車価格  $y_i$  とその属性  $x_{p,i}$  ( $p = 1, 2, \dots, P$ ) に基づいて重回帰分析を行い、推定後の偏回帰係数  $\hat{\alpha}_p$  を各属性の計算価格とみなす。

$$y_i = \hat{\alpha}_1 x_{1,i} + \hat{\alpha}_2 x_{2,i} + \dots + \hat{\alpha}_P x_{P,i} + \epsilon_i \quad (1)$$

ここで  $\epsilon_i$  は残差であり、 $i$  は学習期間における中古車のインデックスである。目的変数  $y$  と説明変数  $x_p$  の量的変数については標準化を施し、最小二乗法により  $\hat{\alpha}_p$  を推定する。

次に、学習期間以降に出品される中古車のインデックスを  $j$  と標記すると、次式のように落札価格を推定できる。

$$\hat{y}_j = \hat{\alpha}_1 x_{1,j} + \hat{\alpha}_2 x_{2,j} + \dots + \hat{\alpha}_P x_{P,j} \quad (2)$$

価格推定においても学習期間と同じ標準化を適用するが、 $\hat{y}$  については標準化の逆変換によって元の価格スケール [円] に戻す。本研究では、日本全国のオートオークション会場で取引された中古車の落札実績データを対象とし、検証車種は利用者の広範な年齢層や用途を考慮し、20 車種とする。なお 2017 年 1 月から 2017 年 12 月を価格推定期間とし、1 か月毎に推定を行う。その際に、推定月前月から過去 2 年間で学習期間とする。なお、車種別に学習 (式 (1)) および推定 (式 (2)) を行う。投入する説明変数を表 1 に示す。

式 (1) および式 (2) の前処理として、説明変数間の多重共線性について配慮する。本研究では、VIF (分散拡大係数) を計算し、VIF が 10 を超える説明変数  $x_p$  は多重共線性の原因となる可能性が高いため、各月の重回帰式から除外した。残された説明変数は車種毎および月毎で異なるため、具体的な記載は割愛する。

データへのあてはめや偏回帰係数の解釈の便宜性から、ヘドニックアプローチでは目的変数のみに対数変換を施す半対数型を用いることが多い [3]。そこで本研究においても説明変数は無変換とし、目的変数について対数変換と無変換の 2 通りで推定を行う。この理由として、4 章で導入する非線形モデルでは目的変数の対数変換を必要としないため、本章の線形モデルと比較可能にするためである。対数変換を適用する場合は、式 (1) では  $y_i$  を  $\log y_i$  とし、式 (2) では  $\hat{y}_j$  を  $\log \hat{y}_j$  とする。その後、 $\exp(\log \hat{y}_j)$  によって  $\hat{y}_j$  を得る。

表 1 説明変数

Table 1 Explanatory variables

量的変数 [単位]	
走行距離	出品時の走行距離 [1000km]
落ち年	出品年と新車登録年の差 [年]
評価点	内装, 外装の総合状態 (0~5) [点]
車検残月	出品時の車検の残り月数 [か月]
排気量	エンジン排気量 [cc]
流通台数	推定月前月の全会場の出品台数 [台数]
新車価格	新車時の価格 [円]
質的変数 (ダミー変数, 該当すれば 1 を付与)	
オプション	ナビ&テレビ有り, ナビのみ レザーシート有り, サンルーフ有り
修復歴	修復歴 (事故歴を含む) の有無
取引月	(1 月を基準 0 とする) 2 月~12 月 (シルバーを基準 0 とする) 白, 真珠, 黒, 黄, 灰, 金, 琥珀, 紫, アイボリー, 青, 青黒 赤, 茶, 鉄, ピンク, 濃灰, 濃緑 薄黄, 薄緑, 葡萄, オレンジ, その他 (関東を基準 0 とする)
車体色	
取引地域	北海道, 東北, 中部, 近畿 中国・四国, 九州・沖縄

推定価格  $\hat{y}_j$  の精度評価として、式 (2) における決定係数  $R^2$  と正答率  $A$  を算出する。決定係数  $R^2$  は 1 に近いほど学習期間以後の中古車  $j$  に対する価格推定モデルの当てはめ精度が高く、正答率  $A$  も同様である。ただし正答率  $A$  を算出する際には実務利用を考慮し、以下の条件を満たす場合に「正答」とみなした。

- if  $y_j \leq 50$  万円,  $y_j - 5$  万円  $\leq \hat{y}_j \leq y_j + 5$  万円
- if  $50$  万円  $< y_j < 150$  万円,  $0.9 \cdot y_j \leq \hat{y}_j \leq 1.1 \cdot y_j$
- if  $150$  万円  $\leq y_j$ ,  $y_j - 15$  万円  $\leq \hat{y}_j \leq y_j + 15$  万円

ここで  $y_j$  は実際の落札価格であり、 $\hat{y}_j$  はその推定値である。正答数を推定台数で割ることで正答率  $A$  を算出する。

式 (2) に基づいて落札価格を推定した様子を図 1, 2 に示す。代表的な事例として、PRIUS の結果を示す。対数変換の有無を比較すると、図 1 の対数変換なしの場合は高価格帯では過小評価される個車が多く、全体的に分布が曲がっている。一方、図 2 のように対数変換を施すことにより、過小評価は改善され、より直線的な分布形状を示している。

次に全対象車種について、正答率  $A$  と決定係数  $R^2$  を集計した結果を表 2 に示す。やはりほとんどの車種において、対数変換を施した方が推定精度が高い。しかし一部の軽自動車や SUV において推定精度が低下する場合もある。

## 3. 落札価格に影響する各属性の非線形性

重回帰式による個車価格推定において、対数変換を施すこ

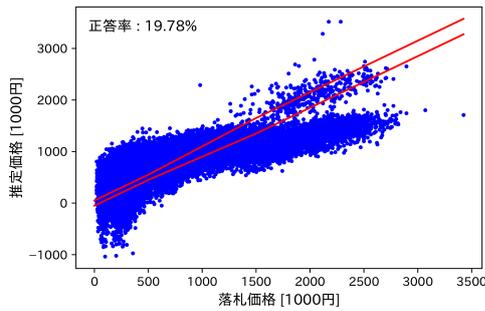


図1 推定価格  $\hat{y}_j$  (縦軸) と 真の落札価格  $y_j$  (横軸) の相関図. ただし目的変数に対数変換を施さない場合. 図中の各点は落札されたPRIUSの個車であり, 上下の実線は誤差の許容範囲である.

Fig. 1 Correlation diagram between the estimated price  $\hat{y}_j$  shown in the vertical axis and its actual price  $y_j$  shown in the horizontal axis without applying the logarithmic transformation to the objective variable. Each dot corresponds to individual car of PRIUS, and two solid lines show the upper and lower limits of the acceptable error range.

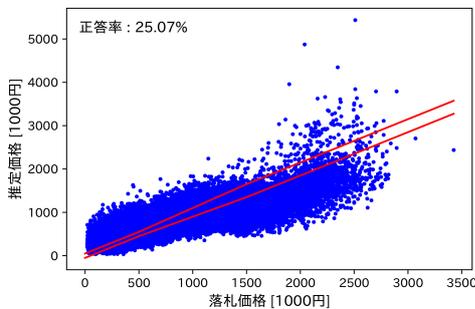


図2 図1と同様. ただし目的変数に対数変換を施した場合.

Fig. 2 Same as Fig.1, but applying the logarithmic transformation to the objective variable.

とにより各評価指標は向上したが, 決定係数  $R^2$  は 0.4~0.8 程度に留まっている. この原因の一つとして, 重回帰式では表現できない非線形構造の存在が考えられる. なお, 中古住宅においても価格形成メカニズムに非線形性の存在が報告されており [10], 中古車においても例外ではないと示唆される. そこで, 学習期間 (2015年1月~2016年12月) に落札された個車の落札価格  $y_i$  と各属性  $x_{p,i}$  の関係性を調べ, 式 (1) のようなシンプルな重回帰式では表現できない非線形性を確認した. その代表的な事例を紹介する.

図3は, PRIUSの落ち年別の走行距離に対する落札価格の分布である. 1年落ちや5年落ちのように新しい個車においては走行距離は大きな減価要因であるが, 10年落ちや15年落ちのように古い個車ではその感応度は低下し, 走行距離による減価率は小さい.

このように, 中古車の属性には落札価格に対して非線形的に影響する可能性がある. そこで次章において機械学習を導入し, 有用な非線形モデルの自動構築を試みる.

#### 4. 非線形モデルによる個車価格推定

重回帰式である式 (1),(2) を XGBoost の関数  $F$  に置き換

表2 推定価格  $\hat{y}_j$  の正答率  $A$  と決定係数  $R^2$

Table 2 The accuracy  $A$  and the coefficient of determination  $R^2$  of the estimated price  $\hat{y}_j$

車種	正答率 $A$ [%]		決定係数 $R^2$	
	無変換	対数変換	無変換	対数変換
LS	4.91	7.56	0.45	0.57
CROWN	10.93	18.72	0.44	0.53
PRIUS	19.78	25.07	0.58	0.69
SERENA	13.95	23.43	0.62	0.58
ESTIMA	10.10	14.36	0.50	0.66
STEP WGN	16.07	40.60	0.46	0.56
TANTO	25.36	32.31	0.74	0.80
WagonR	51.85	50.99	0.89	0.86
ALTO	49.70	54.78	0.86	0.84
MOVE	44.87	51.87	0.87	0.86
FIT	18.70	38.03	0.40	0.46
DEMIO	34.31	60.24	0.72	0.86
AQUA	20.10	21.21	0.33	0.41
NOTE	22.14	26.10	0.60	0.66
Vitz	23.42	49.94	0.71	0.83
Forester	25.98	41.68	0.88	0.66
PAJERO	22.05	37.37	0.84	0.83
PRADO	27.31	24.66	0.90	0.82
OUTLANDER	12.98	17.11	0.37	0.54
Roadstar	20.11	23.12	0.73	0.80

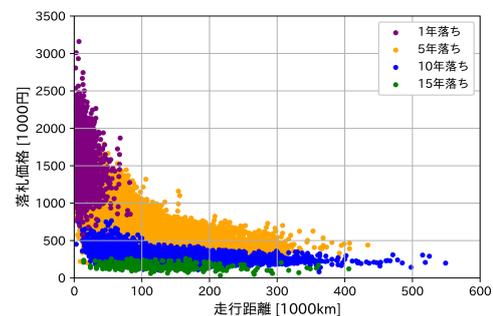


図3 走行距離と落札価格の関係 (PRIUS の例)

Fig. 3 Relationship between mileages and contract prices of PRIUS

えると, それぞれ

$$y_i = F(x_{1,i}, x_{2,i}, \dots, x_{p,i}) + \epsilon_i \quad (3)$$

$$\hat{y}_j = F(x_{1,j}, x_{2,j}, \dots, x_{p,j}) \quad (4)$$

となる. 式 (3) で関数  $F$  を学習し, 式 (4) で未学習データに適用する. 関数  $F$  は非線形性を表現できるため, 目的変数  $y_i$  には対数変換を適用しない. さらに機械学習はモデルの解釈性を重視しないため, 多重共線性を考慮せず全ての説明変数を入力する. その他, 価格推定期間や学習期間などについては, 2章の重回帰分析と同一とする.

図2と同様に, PRIUSの落札価格を推定した様子を図4に示す. 図2の重回帰式 (対数変換あり) と比較して, 高価格帯においても安定的に落札価格を推定できている. 分布形状もより直線的になっていることから, 推定誤差の縮小

表 3 XGBoost による推定価格  $\hat{y}_j$  の正答率  $A$  と決定係数  $R^2$ . カッコ内に表 3 の重回帰式 (対数変換) からの改善量を示す.

**Table 3** The accuracy  $A$  and the coefficient of determination  $R^2$  of  $\hat{y}_j$  estimated by XGBoost. Each figure in parenthesis shows the improvement from the multiple linear regression with logarithmic transformation shown in Table 3.

車種	正答率 $A$ [%]	決定係数 $R^2$
LS	30.81 (+23.25)	0.94 (+0.37)
CROWN	56.87 (+38.15)	0.97 (+0.44)
PRIUS	63.88 (+38.81)	0.96 (+0.27)
SERENA	66.34 (+42.91)	0.98 (+0.39)
ESTIMA	52.06 (+37.70)	0.96 (+0.30)
STEP WGN	69.28 (+28.68)	0.96 (+0.40)
TANTO	68.80 (+36.49)	0.97 (+0.16)
WagonR	75.29 (+24.30)	0.96 (+0.11)
ALTO	74.61 (+19.83)	0.95 (+0.11)
MOVE	74.94 (+23.07)	0.97 (+0.10)
FIT	75.56 (+37.53)	0.95 (+0.48)
DEMIO	73.97 (+13.73)	0.96 (+0.10)
AQUA	72.55 (+51.34)	0.94 (+0.52)
NOTE	71.92 (+45.82)	0.96 (+0.30)
Vitz	72.91 (+22.97)	0.95 (+0.12)
Forester	69.22 (+27.54)	0.98 (+0.32)
PAJERO	60.47 (+23.10)	0.97 (+0.14)
PRADO	51.02 (+26.36)	0.96 (+0.15)
OUTLANDER	52.80 (+35.69)	0.93 (+0.39)
Roadstar	51.59 (+28.47)	0.96 (+0.17)

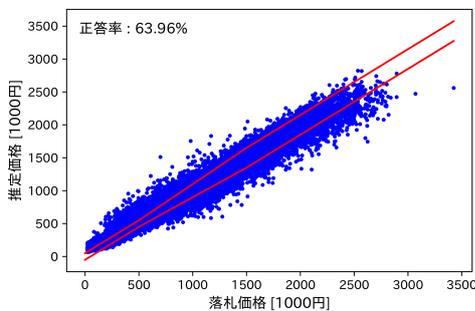


図 4 図 1 と同様. ただし価格推定モデルに XGBoost を適用した場合.

**Fig. 4** Same as Fig.1, but using XGBoost to estimate  $\hat{y}_j$ .

を確認できる.

次に全対象車種について, 正答率  $A$  と決定係数  $R^2$  を集計した結果を表 3 に示す. 全ての車種において推定精度が改善されている. 特に正答率  $A$  は 20 ~ 50% も向上し, 軽自動車やコンパクトカーのような大衆車種において 70% 以上の精度を実現している. このように XGBoost による機械学習モデルの恩恵を確認できる.

## 5. 要因分析

学習後の重回帰式と XGBoost を観察することで, 中古車の各属性がどのように落札価格に寄与するのかを分析

する. 重回帰式においては式 (1) で推定された偏回帰係数  $\hat{\alpha}_p$  を重要度とみなし, 有意水準 5% の  $t$  検定を適用する. XGBoost においてはノード分割時のゲイン  $G_m$  に基づいて以下のように説明変数  $x_p$  の重要度  $I_p$  を算出する.

$$I_p^{(k)} = \sum_{m \in M_p^{(k)}} G_m^* \quad (5)$$

ここで  $M_p^{(k)}$  は,  $k$  番目の決定木において, 説明変数  $x_p$  に基づいて分割されたノード  $m$  の集合を表し,  $G_m^*$  は実際のノード分割で得られた最大ゲインを意味する. これを XGBoost を構成する全ての決定木において平均すると,

$$I_p = \frac{1}{K} \sum_{k=1}^K I_p^{(k)} \quad (6)$$

によって説明変数  $x_p$  の重要度  $I_p$  を算出できる [12]. さらに解釈しやすいように,  $\sum_{p=1}^P I_p = 1$  になるようにスケール調整を施す.

分析結果を図 5, 図 6 に示す. いずれも 2017 年の各月から過去 2 年間で学習期間として, 偏回帰係数  $\hat{\alpha}_p$  と重要度  $I_p$  を算出した. なお誌面の都合により, 車種を厳選して典型的な事例を示す. 図 5 は量的変数の重要度を示しており, 落ち年が落札価格に強く影響することが分かる. XGBoost による重要度  $I_p$  は非線形的かつ間接的な影響も反映できるが, ポジネガ極性を評価できない. しかし重回帰式による偏回帰係数  $\hat{\alpha}_p$  は線形的かつ直接的な影響のみに限定されるが, ポジネガ極性を評価できる. これにより, 落ち年が落札価格を下げるマイナス要因であることが分かる. 一方, 図 6 のレザーシートの解釈においては, 偏回帰係数  $\hat{\alpha}_p$  は統計的に有意な影響力を示していない. しかし XGBoost によればレザーシートの影響力は非常に高いため, 重回帰式では見落としてしまう重要な説明変数を検出できている. この事例においては, レザーシートの有無が評価点に影響し, 評価点を通じて落札価格に影響を与えたと推察される. このように XGBoost は間接的な影響も捉えられ, そして重回帰式はポジネガ極性を評価できるため, それぞれを併用することが肝要である.

## 6. 異常検知

前章までの落札価格の推定は, 主に中古車の買取査定モデルとしての活用を想定している. さらに本章では, 買取査定モデルの応用として, 落札価格の異常割安・割高の検知を試みる.

まず落札価格の推定には, 4 章で有用性を検証した XGBoost を適用する. 毎月において前月から過去 2 年間の落札実績に基づいて式 (3) により関数  $F$  を学習し, 式 (4) により当月の中古車  $j$  の推定落札価格  $\hat{y}_j$  を得る. これを検証期間において毎月繰り返す.

異常度  $\phi$  の算出において, オートオークション会場毎で

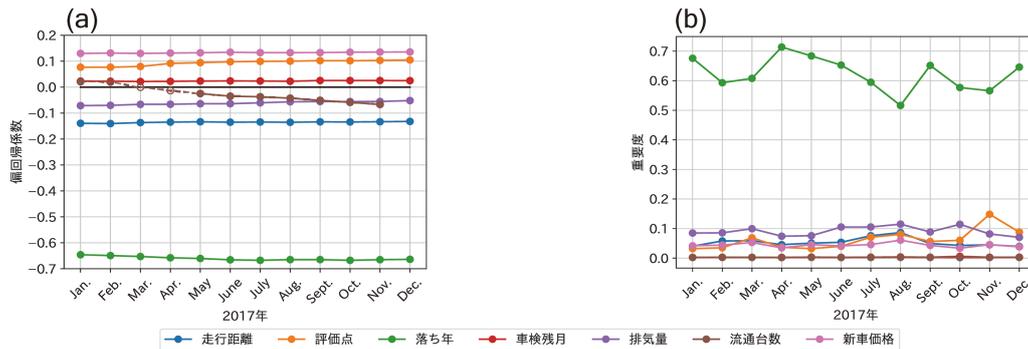


図5 量的変数の重要度 (PAJERO の例): (a) 重回帰式による偏回帰係数  $\hat{a}_p$  (点線は有意水準 5% の  $t$  検定において有意でない場合), (b) XGBoost による重要度  $I_p$

Fig. 5 Importance of quantitative variables in PAJERO: (a) the partial regression coefficients  $\hat{a}_p$  estimated by multiple linear regression where dotted lines mean the cases that  $\hat{a}_p$  cannot pass the  $t$ -test at the 5% significance level, (b) the Importances  $I_p$  estimated by XGBoost

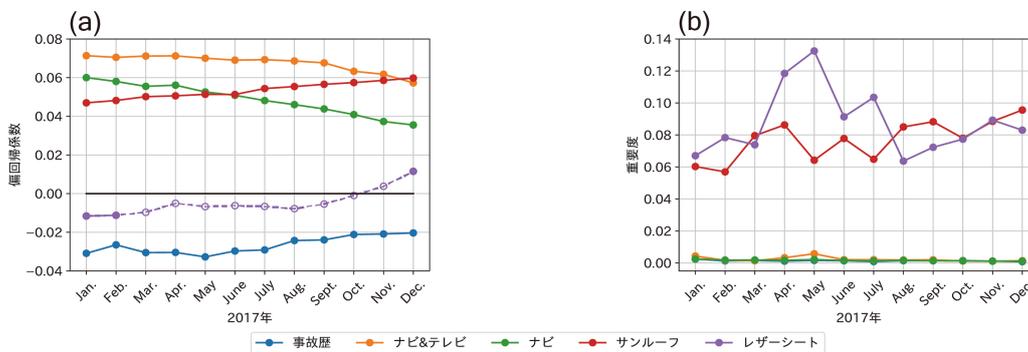


図6 図6と同様, ただしオプション装備の重要度 (PRADO の例)

Fig. 6 Same as Fig.6, but Importance of optional equipments in PRADO

分析する場合は, ある月  $t$  に会場  $v$  で落札された中古車  $j$  の集合を  $J_v(t)$  と表記すると,

$$\phi_v(t) = \frac{1}{|J_v(t)|} \sum_{j \in J_v(t)} (y_j - \hat{y}_j) \quad (7)$$

によって異常度  $\phi_v(t)$  を算出する. 分析対象は日本全国の 79 会場 ( $v = 1 \sim 79$ ) とする.  $\phi_v(t) > 0$  であるほど割高に落札されやすい会場であり,  $\phi_v(t) < 0$  であるほど割安に落札されやすい会場と言える. また  $t$  による変化を観察することで, 異常度の季節性について分析できる.

表 4 に, 会場別の異常度  $\phi_v(t)$  を示す. なお誌面の都合により, 主要な 13 会場に限定する. 会場別に割安月と割高月を調べ, 割安な上位 3 ヶ月を赤色, 割高な上位 3 ヶ月を緑色で示す. 季節性を明確にするために検証期間を 2013 年~2017 年の 5 年間に拡大し, 月  $t$  毎に  $\phi_v(t)$  を統合して平均化した. 結果として, 年度末の 1~3 月には割高で落札されやすく, 新年度を迎えた 4~6 月には割安傾向が強い. これは中古車の新学期需要により, 年度末に買い手が増えるためだと考えられる. さらに, ボーナス支給月の 6 月と 12 月を転換点とし, 割安と割高のサイクルを確認できる. 特に北部の会場ほど標準偏差が大きく, 年間を通して割安と割高のサイクルが大きい.

この理由を考察するため, 図 7 に全 79 会場の異常度  $\phi_v(t)$  の標準偏差と落札台数の関係を示す. なお, 落札台数は会場の規模と同義である. 結果として, 両者は反比例の関係にある. つまり, 規模が小さい会場ほど落札台数が少ないため, 季節性やトレンドの影響を過大に受けやすいと考えられる. 一方, 規模が大きい会場ほど落札台数が多いため, 季節性やトレンドに対しても安定的な価格形成がされやすいと考えられる. なお, 落札台数が少ない小規模会場ほど割安傾向 (赤色) であり, 落札台数が多い大規模会場ほど割高傾向 (緑色) にある. これは需給の観点から, 前者は供給過多, 後者は需要過多と言える. したがって小規模会場で仕入れて, 大規模会場で売却するような自己売買 (ディーリング) が考えられる.

## 7. まとめ

本研究では, 重回帰式ベースのヘドニックアプローチと非線形モデルの XGBoost に基づいて個車価格推定モデルを構築した. 中古車の各属性は落札価格に対して非線形的かつ間接的に影響することにより, XGBoost を適用することで推定精度を向上できた. しかし要因分析においては重回帰式の方が安定しており, さらに偏回帰係数の符号から

表 4 会場別の異常度  $\phi_v(t)$  [1000 円]  
Table 4 Abnormality  $\phi_v(t)$  [1000 yen] in each auto-auction venue and each month

会場名	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	平均値	標準偏差
NAA 札幌	-4.2	9.3	69.7	15.9	-58.0	-3.4	-29.5	-16.1	0.8	-6.5	-34.4	4.9	-4.29	29.92
TAA 東北	-8.1	6.1	-27.0	-3.9	-26.4	-19.6	1.1	-7.4	-9.9	-17.5	-33.6	-22.7	-14.08	11.75
JU 福島	-3.2	-12.1	-8.6	-5.5	-14.2	-27.6	5.2	-4.1	-14.5	15.4	-6.3	-17.2	-7.73	10.55
JU 茨城	-10.8	-3.4	-5.9	-4.7	-7.9	-43.4	-5.3	-2.6	9.6	6.1	9.0	-10.4	-5.80	13.17
JU 埼玉	6.7	-4.2	5.3	-22.9	-22.7	-10.9	-9.9	-2.3	-11.2	2.6	-6.3	-15.2	-7.58	9.43
USS 東京	27.6	23.4	12.9	3.6	8.6	10.2	18.3	18.4	17.1	13.1	11.2	11.0	14.62	6.38
TAA 横浜	-1.7	12.4	-5.3	-8.5	-3.0	4.9	1.4	10.5	-1.2	15.2	-2.2	-12.3	0.85	8.07
USS 名古屋	31.1	29.5	19.3	9.6	11.1	13.1	20.0	17.2	18.3	18.1	13.6	14.0	17.90	6.37
TAA 中部	3.3	-1.7	-3.4	-18.9	-23.0	-9.9	-11.3	-4.0	2.9	-12.0	-12.9	-6.3	-8.10	7.78
JU 岐阜	8.9	20.5	-5.4	-2.4	-8.8	-10.2	5.3	13.5	-3.8	2.1	6.3	6.3	2.69	8.85
TAA 近畿	-2.2	5.9	3.2	-9.1	-2.6	4.2	8.7	-1.0	5.5	-4.5	6.2	0.0	1.19	5.06
JU 広島	-14.5	-18.0	-20.7	-33.5	-60.9	-24.7	30.9	-19.9	13.4	-14.9	-21.5	-33.3	-18.13	21.95
USS 九州	2.1	13.6	1.2	-4.7	-6.1	-1.8	1.3	9.3	4.3	-1.3	-1.1	2.1	1.57	5.30

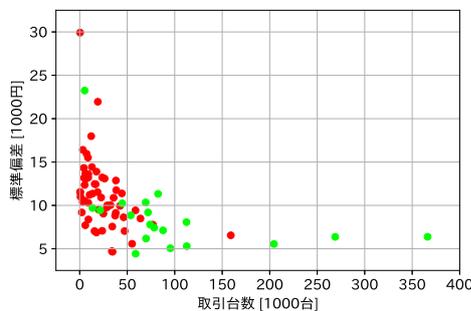


図 7 各会場における落札台数と異常度  $\phi_v(t)$  の標準偏差の関係。赤色は割安な会場 ( $\phi_v(t)$  の平均値が負)、緑色は割高な会場 ( $\phi_v(t)$  の平均値が正) を表す。

Fig. 7 Relationship between the number of contracted cars and the standard deviation of  $\phi_v(t)$  in each auto-auction venue  $v$ . Red color points mean relatively cheap venues because the average of  $\phi_v(t)$  is less than 0. Green color points mean relatively expensive venues because the average of  $\phi_v(t)$  is more than 0.

各属性のポジネガ極性を解釈できた。一方、XGBoostにより非線形かつ間接効果も考慮することで、レザーシートなど重回帰式では棄却してしまう重要な説明変数を検出できた。つまり要因分析においては、双方を併用することが肝要である。異常検知においては中古車相場の割安割高を算出し、新学期需要やボーナス期などの季節性や、オートオークション会場の地域や規模に応じて割安や割高に偏りやすい性質を確認した。これにより本研究の価格推定モデルは、中古車の買取査定のみならず、オートオークション会場や季節を横断するディーリングにも活用できると示唆される。

謝辞 本研究の遂行にあたり有益なご助言を頂いた、株式会社プロトコーポレーションの下山力三氏、長谷川恵理子氏、平野大河氏、並びに鈴木研究室修了生の櫻井大宙氏、速見勇磨氏、山下梨瑛氏に感謝申し上げます。本稿の内容は全て筆者個人の見解であり、株式会社プロトコーポ

レーションの公式見解ではありません。なお本研究の一部はJSPS 科研費 (20K11969) の助成により行われました。

参考文献

- [1] Court, T.A.: Hedonic Price Indexes with Automotive Examples, *The Dynamics of Automobile Demand*, pp.98–119 (1939).
- [2] Griliches, Z.: Hedonic Price Indexes for Automobiles—An Econometric of Quality Change, *The Price Statistics of the Federal Government*, pp.173–196 (1961).
- [3] 太田 誠: ヘドニック・アプローチの理論的基礎, 方法及び日本の乗用車価格への応用, 季刊理論経済学, Vol.29, No.1, pp.31–55 (1978).
- [4] Harrison, D. and Rubinfeld, L.D.: Hedonic Housing Prices and the Demand for Clean Air, *Journal of Environmental Economics and Management*, Vol.5, No.1, pp.81–102 (1978).
- [5] Noland, C.: Assessing Hedonic Indexes for Housing, *Journal of Financial and Quantitative Analysis*, Vol.14, No.4, pp.783–800 (1979).
- [6] 国土交通省: 不動産価格指数の作成方法, 入手先 (<https://www.mlit.go.jp/common/001360416.pdf>) (参照 2020-09-15).
- [7] 清水千弘: ビッグデータで見る不動産価格の決まり方, 日本不動産学会誌, Vol.31, No.1, pp.45–51 (2017).
- [8] 大和祐, 野村眞平: SUUMO でのビッグデータ活用事例, 日本不動産学会誌, Vol.31, No.1, pp.78–83 (2017).
- [9] Meng, M.S., Liu, J.L., Kuritsyn, M., et al.: Price Determinants on Used Car Auction in Taiwan, *International Journal of Asian Social Science*, Vol.9, No.1, pp.48–58 (2019).
- [10] Shimizu, C., Nishimura, G.K. and Karato, K.: Nonlinearity of Housing Price Structure—The Secondhand Condominium Market in Tokyo Metropolitan Area, *International Journal of Housing Market and Analysis*, Vol.7, No.3, pp.459–488 (2014).
- [11] Chen, T. and Guestrin, C.: XGBoost—A Scalable Tree Boosting System, *Proc. The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.785–794 (2016).
- [12] 杉本知之, 下川敏雄, 後藤昌司: 樹木構造接近法と最近の発展, 計算機統計学, Vol.18, No.2, pp.123–164 (2005).