

価格予測モデルに基づく、XAI 手法を用いた中古ブランド品買取価格決定支援 Support for determining the purchase price of second-hand luxury goods using XAI based on a price prediction model

井田 篤志[†] 家入 祐也[‡] 菱山 玲子[†]

1 研究背景と目的

2021 年の国内の中古品市場規模は市場規模は成長を維持し続けており、2020 年の市場規模は 2 億 4169 億円に上る [1]. その中でもブランド品の割合は 10.2% を占め、市場規模は 2,464 億円と非常に大きな市場であるといえる [1]. このように年々拡大を続けている中古ブランド品市場において、商品の買い取りを担当する査定人が適正な買取価格を算出することは、重要な事項であると言える [2].

しかしながら、カラーの違いや傷の度合いなどといった微細で膨大な情報によって価格が変化し中古ブランド品において、適正な査定価格を算出することは困難である. 現状の査定では、これまでの経験や EC サイト上での価格といった不明確な参考情報を元に査定を行うため、査定人によって価格に幅が生じ、不安定な買取価格となってしまう. 特に、査定歴が浅く、経験が少ない若手の査定人にとっては、適正価格を提示することがより難しい.

中古ブランド品の適正買取価格は、メーカーの供給量や直販価格変更などのメーカーによる要因や、為替変動や需要変動などの影響を受けるため、時期によって大きく変動する. 例として、円安や定価改定に伴い、ブランド品買取価格の上昇が見られた事例がある [3]. そのため、過去データを用いて算出した予測価格は、時勢に合致していない可能性があり、そのまま買取価格として用いることは適切ではない.

そこで、適正価格算出の支援としては、より正確で、有効的な参考情報を提示することが最適である. 参考情報の一つとして、膨大な各商品情報が価格に対して、どのような影響を与えているのかが挙げられる. 本研究では機械学習モデルによる買取価格予測根拠の説明が参考情報として有効であるかの検証を行う. 機械学習の利点として、商品の各特徴量と価格の非線形な関係に対応できるため、複雑な関係を説明できる点が挙げられる.

2 XAI

2.1 LIME と SHAP

現在の機械学習技術では、容易に高精度な機械学習モデルを生成することができる. 高精度な機械学習モデルを他分野へ応用する試みも行われている一方で、懸念点として、高精度なモデルであるほど、モデルがブラックボックスである点が挙げられる. 不透明性を解決する技術が、近年、XAI(explainable Artificial Intelligence)と呼ばれ、広く浸透している [4].

本研究では、説明手法として、表形式データの分析に多く使われる LIME(Local Interpretable Model-Agnostic Explanations)[5] と SHAP(SHapley Additive exPlanations)[6] の 2 手法を採用した.

LIME とは、説明対象データから発生させた近傍データを基に、ホワイトボックスなモデルを生成し、解釈する (例: 線形回帰モデルを構築し、モデルの各係数を寄与度とする) 手法である. SHAP では、報酬分配ゲーム理論のシャープレイ値の考え方をゲーム参加者を特徴量に、報酬をモデルの予測の期待値に置き換え、応用している.

2.2 LIME と SHAP の比較

SHAP におけるシャープレイ値は予測値とバイアス項の差分が特徴量間で公平に分解されるが、LIME において、それは保障されない. よって LIME と SHAP によって算出される寄与度は同じ説明対象データ、モデルであったとしても異なるものとなる. SHAP では、単位が予測値と同じ (今回は、円) になるが、LIME では単位が存在しない. 本研究では、LIME と SHAP によって算出された寄与度が異なることによって、査定時の参考情報としての有効度合いに差が生じるかの検証を行う.

3 先行研究

Sakai ら [7] による研究では、胎児心臓超音波スクリーニングへの説明手法の適用が行われた. 超音波検査は医療分野において幅広く利用されているが、検査者間の間に診断技術の差異が大きいことや、画像劣化などの超音波検査特有の問題が存在する [8]. 上記の問題を解決する手段として、機械学習モデルで画像データから異常の有無について学習を行い、予測結果に対して説明手法を適用することで、異常判定時に根拠となる箇所の可視化を行う支援機能を構築した.

また、説明手法の効果検証を検査者に対して行い、異常検知における診断支援機能として有効であることを示した. 機械学習モデルの予測対象が異常の有無といった正誤が存在するものであるため、診断支援機能としての効果検証を、検査者がスクリーニング動画をもとに診断を行い、正誤判定によって説明手法の有無による診断精度の違いについて考察することによって行った. 結果として、効果検証の精度より、説明手法が診断支援として有効であることが示された.

Sakai らの研究 [7] における予測モデルは分類モデルであり、明確な正誤が存在するため、効果検証を機能使用者がテストを行い、得られた精度で行うことができたが、中古ブランド品の買取価格には明確な正誤が存在しないため、本研究では、精度による効果検証を行うことができない. そこで、本研究では効果検証に精度を利用できない事例について、支援機能の利用者へのアンケートを用いて効果検証を行う.

[†]早稲田大学 大学院創造理工学研究所

Department of management engineering, Waseda University

[‡]早稲田大学 大学院情報生産システム研究科

Graduate School of Information Production and Systems, Waseda University

表 1: 前処理内容

特徴量	前処理
商品名	名称の統一化 ダミー変数化
型番	名称の統一化 ダミー変数化
素材	観測数 50 以下のカテゴリ削除
カラー	名称の統一化 ダミー変数化 観測数 50 以下のカテゴリ削除
取引日	UNIX 時間への変換 年, 月を特徴量追加
ランク	平均値を基に数値置き換え

表 2: カテゴリ変数処理の例

ダミー変数化前	ダミー変数化後	合算前寄与度	合算後寄与度
型番	A 型番_A	1	20
	B 型番_B	0	-10
	C 型番_C	0	5
サイズ	サイズ	40	15
保証書	保証書	1	10

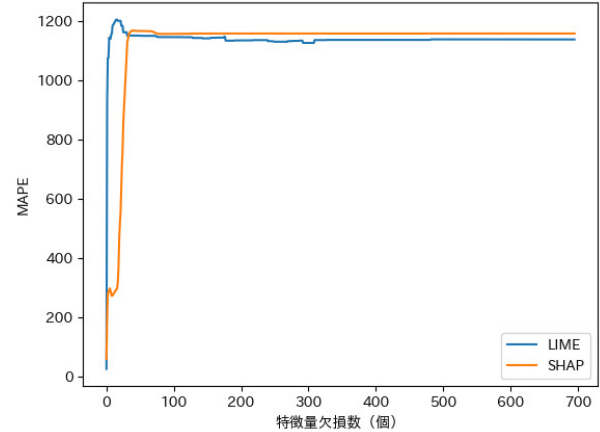


図 1: pixel flipping

4 買取価格予測モデルの構築

4.1 データクレンジング

本研究では、ブランド品の買取販売を行う A 社の 2018 年から 2022 年までのブランド B の POS データ、13,922 件を用いて、モデルを構築する。使用した特徴量は商品情報（商品名、型番、定価、カラー、サイズ、素材）、付属品の有無（箱、保証書、取り扱い説明書、ケース、タグ）、商品の状態ランク、取引日である。行った前処理を表 1 に示す。POS データは A 社の査定人により記録されたものであり、査定人によって略称による記載や半角全角の違いなどの差異があったため、商品名、カラー、素材については名称の統一化を行った。また、商品状態ランクはランクごとの平均買取価格を基に数値への置き換えを行った。

4.2 学習モデル構築

学習モデルは勾配ブースティング決定木 (GBDT) の手法である LightGBM[9] を用いた。POS データは時系列データのため、テストデータは 2021 年度の 1 月ごとのデータ、トレインデータはそれ以前のデータを用いることで計 12 回の交差検証を行った。

本研究では、予測精度の評価指標として、 R^2 , MAE , $MAPE$ の 3 手法を用いた。各指標について、式 1, 2, 3 に表す。

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y} - y_i)^2} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100 \quad (3)$$

計 12 回の交差検証の平均は、 $MAPE$ が 22.3, R^2 が 0.87, MAE が 273139 となった。使用した POS データの平均買取価格は、1,577,249 円であったため、 MAE の値は、その 17% であることから、価格幅が大きい中古ブランド品において、十分な予測ができていると言える。

5 説明手法の実装

5.1 カテゴリ変数の処理

解釈性手法の前処理として、商品名、型番、カラーの 3 つの特徴量に、ダミー変数化されたカテゴリ変数ごと

の合算を行った。例を表 2 に表す。カテゴリ変数は学習のためにダミー変数化しているため、該当しないカテゴリ（ダミー変数化後に 0 をとるカテゴリ、表 2 では、型番 B, C）についても寄与度が全て算出される。これにより、該当する型番の寄与度を読み取ることが困難となる。そのため、カテゴリごとに寄与度を足し合わせたものをカテゴリ変数の寄与度とした。（表 2 における型番は A であり、型番 A の寄与度は表 2 の型番 A, B, C の寄与度の合算値となる。）

5.2 pixel flipping による説明精度評価

説明手法の精度を評価する手法として、pixel flipping[10] がある。主に画像データやテキストデータに用いられる手法であり、説明手法において重要度が高い、もしくは低いと判断された特徴量（画像データの場合はピクセル、自然言語の場合は単語）から順に欠損させていった際に精度がどのように変化するかについて評価する手法である。説明手法が正確に各特徴量の寄与度を算出することができた場合、重要度が高い特徴量ほど、精度の変化は大きくなると言える。

本研究では、LIME, SHAP とともに、誤差指標 $MAPE$ を用いて、重要度が高い順に特徴量を欠損させた際の指標の変化を取得する。欠損させたい特徴量の平均値を平均とする標準偏差 5 の正規分布に従う乱数を当てはめることによって欠損を表現した。高い順から欠損させた結果を図 1 に示す。

図 1 より、SHAP, LIME 共に、欠損させる特徴量数が増えるほど、 $MAPE$ の値の上がり幅が小さくなっていることがわかる。このことより、SHAP, LIME の両手法において、重要であると判断された特徴量ほど、予測に対して大きく寄与しており、両手法共に、正しく寄与度が算出されたといえる。

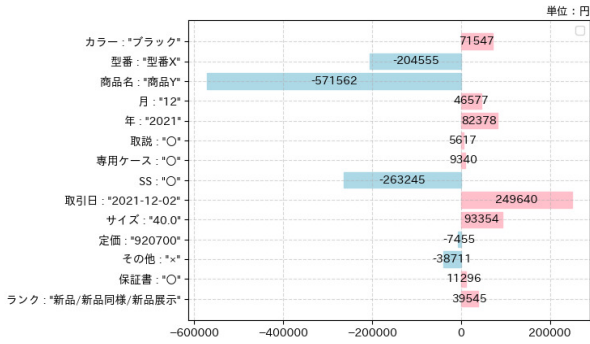


図 2: アンケートに用いた画像例

表 3: 手法別の平均値の比較

	重要な要素の理解	各項目の影響の理解	価格確定割合
LIME	3.14	3.67	51.7%
SHAP	2.90	3.52	48.6%
全体	3.02	3.60	50.1%

6 効果検証実験

本研究では、A 社の買取査定人 21 人を対象にしたアンケートを用いて効果検証実験を行った。アンケートでは、ランダムに選択した商品の寄与度を示した図 (例: 図 2) を SHAP と LIME でそれぞれ提示し、図について質問をおこなった。寄与度を表した図は、各特徴量名とその示す値の寄与度がわかるような棒グラフの図を作成した。図 2 の各棒グラフは商品の特徴量名とその値を表し、棒グラフの数値がの正負が価格に与える影響を表す。例として、図 2 の一番上の棒グラフは商品のカラーがブラックであることにより、価格が 71,547 円上昇することを表している。

アンケートでは、以下の項目の質問を行った。

- 図から価格決定に重要な項目が理解しやすいか
- 図から各項目が価格にどのような影響を与えているかが理解しやすいか
- 普段の査定時の参考情報に図が加わった場合、自身の経験・知識以外 (web などの参考のみ) でどの程度、買取価格を確定できるか

項目 1, 2 は、7 段階 (数字が大きいほど、理解度が高い) の評価。項目 1 では、商品のどの特徴量が重要かどうかを設定したか、項目 2 では、項目の価格に与える影響がどのようなものかについて結果を得ることができる。この 2 項目により、説明のわかりやすさ、どの程度理解できるかについて考察できる。

項目 3 では、0~100% で評価を行い、参考情報のみでどの程度買取価格を確定できるかについて質問する。これにより、説明が買取支援機能としてどの程度有効的かについて測ることができる。

7 結果と考察

7.1 有効度評価

有効度評価で得られた各アンケートの平均値を表 3 に示す。重要な要素の理解、各項目の影響の理解について、両項目の共に、7 段階の中央値である、3.5 に近い値をとり、説明に対する理解度が低いといえる。また、価格確定割合についても、50% に近い値をとった。このことより、説明手法が査定時の参考情報で有効的である

表 4: 各アンケート項目の相関係数

	質問項目	重要な要素の理解	各項目の影響の理解	価格確定割合
LIME	重要な要素の理解	1.000000	0.857626**	0.078936
	各項目の影響の理解	0.857626**	1.000000	0.09259
	価格確定割合	0.078936	0.09259	1.000000
SHAP	重要な要素の理解	1.000000	0.703169**	-0.022304
	各項目の影響の理解	0.703169**	1.000000	0.182886
	価格確定割合	-0.022304	0.182886	1.000000
全体	重要な要素の理解	1.000000	0.775736**	0.035322
	各項目の影響の理解	0.775736**	1.000000	0.141984
	価格確定割合	0.035322	0.141984	1.000000

* p<0.05

** p<0.01

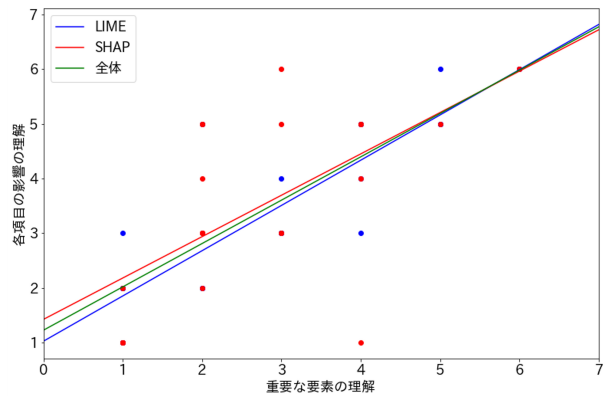


図 3: 重要な要素と各項目の与える影響の比較

とは認められない。

結果より、説明手法のみでは、価格予測に対する寄与について理解することが難しく、知識・経験も十分に必要であることがわかる。ただし、説明に対する理解度が低いことから、長期間にわたる使用などによって理解度が変化することで、価格確定割合の値が変化する可能性も考えられる。

7.2 アンケート項目の比較

アンケートの各項目について、相関係数を用いて、比較を行った。相関係数について、表 4 に示す。表 4 より、重要な要素の理解と各項目の影響の理解には正の相関が見られる。重要な要素の理解と各項目の影響の理解について、無相関検定を行なったところ、LIME, SHAP, 全体の全てにおいて、優位水準 1% で有意となったため、相関があるとと言える。特に、SHAP について、重要な要素の理解と各項目の影響の理解には、強い正の相関がある。重要な要素の理解と各項目の影響の理解をプロットした図を図 3 に示す。このことより、重要な項目について理解できる場合、その影響についても理解することができているとわかる。

7.3 LIME と SHAP の比較

LIME と SHAP の両手法について t 検定を用いて比較を行った。p 値は重要な項目の理解が 0.645、各項目が与える影響の理解が 0.464、価格確定割合が 0.623 となり、p 値の値は 0.05 よりも大きく有意差は認められなかった。

このことより、手法による寄与度の値の違い、単位の違いは、査定支援において、影響がないことがわかる。よって、テーブルデータの査定支援においては、pixel flipping による説明精度がより高いとされた指標を用いることが望ましいと言える。

表 5: 平均値の査定歴別の比較

手法	査定歴	アンケート項目	平均値
LIME	2年以下	重要な要素の理解	3.90
		各要素の影響の理解 価格確定割合	4.40 53.6%
	3年以上	重要な要素の理解	2.45
		各要素の影響の理解 価格確定割合	3.00 50.0%
SHAP	2年以下	重要な要素の理解	3.7
		各要素の影響の理解 価格確定割合	4.8 50.5%
	3年以上	重要な要素の理解	2.18
		各要素の影響の理解 価格確定割合	2.36 46.8%

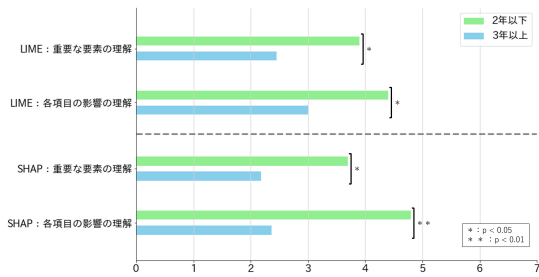


図 4: 重要な要素と各項目の影響についての査定歴ごとの比較

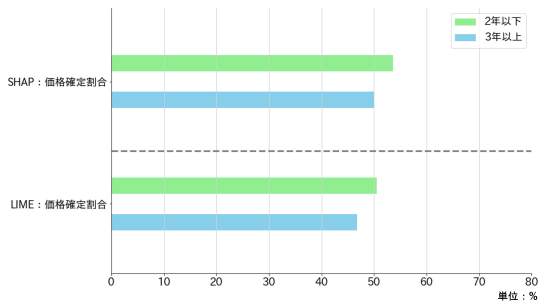


図 5: 価格確定割合についての査定歴ごとの比較

7.4 査定歴別の比較

査定歴が 2 年以下の若手 (10 人) と 3 年以上 (11 人) のベテラン (11 人) の 2 群について t 検定を用いて比較を行った。各アンケート項目についての平均値の結果を表 5, t 検定の結果を図 4, 図 5 に示す。

重要な要素の項目が理解できるかにおける、若手とベテランの平均値は、SHAP ともに有意水準 5% で有意となり、若手の方が高い値を示した。また、各項目の予測に与える影響が理解できるかについては、LIME が有意水準 5% で、SHAP では有意水準 1% で有意となり若手のほうが高い値を示した。ただし、価格の確定割合については、有意差が見られなかった。

分散分析より、2 年以下の若手の方がベテランにくらべて、説明に対する理解度が高いことがわかる。理由として、年齢や経験値の差などが考えられる。このことより、年齢やこれまでの経験などの査定人の特性によって、説明の査定支援としての有効度合いに違いがある可能性が示唆され、追加の検証が必要であると言える。

8 まとめと今後の課題

本研究では、機械学習モデルによる買取価格予測根拠の説明が買取時の参考情報として有効かどうかの検証を

行った。結果として、重要な要素と各項目が価格に与えるの影響の理解度合は、中央値を示し、また、価格確定割合も 50% に近い値となったため、説明手法の有効性は認められなかった。比較をしたところ、LIME と SHAP 間の違いは認められなかった。また、重要な項目の理解と各項目の影響の理解には相関があることが分かった。

加えて、査定歴が 2 年以下の若手と 3 年以上のベテランの 2 群比較により、査定歴が 2 年以下の若手のほうが、説明結果に対する理解度が高い結果となり、査定人によって有効度合いが異なる可能性が示唆された。また、説明手法に対する理解度が低いことから理解度の変化によって有効度が変化する可能性も考えられる。

このことより、今後の課題として、査定人へのインタビューなどの定性的な評価や実際の査定現場に導入した上での長期的な評価などによる査定人ごとの詳細な有効度評価、異なる表現方法を用いてのユーザーインターフェース検証が必要であると言える。これらの検証により、査定支援におけるより最適な説明を実現できると考えられる。

参考文献

- [1] 環境省環境再生・資源循環局総務課リサイクル推進室: 令和 3 年度リユース市場規模調査報告書, 環境省 (オンライン), 入手先 (<https://www.env.go.jp/content/000076424.pdf>) (参照 2023-01-13) .
- [2] 安藤 根八: 中古ブランド品販売企業における鑑定士育成システムの考察, 関西ベンチャー学会誌, Vol.13, pp.77-87 (2021).
- [3] 日本経済新聞: ブランド中古品の売買が活況 背景に円安、定価超えも, 日本経済新聞 (オンライン), 入手先 (<https://www.nikkei.com/article/DGXZQOUB248EA0U2A021C2000000/>) (参照 2023-06-12).
- [4] 亀谷由隆: 説明可能 AI 技術のこれまでとこれから, 電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review, Vol.16, No.2, pp.83-92 (2022).
- [5] Ribeiro, M.T., Singh, S. and Guestrin, C.: “Why Should I Trust You?” Explaining the Predictions of Any Classifier, *Proc. KDD 2016*, pp.1135–1144 (2017).
- [6] Lundberg, S.M. and Lee, S.I.: A Unified Approach to Interpreting Model Predictions, *Proc. Advances in Neural Information Processing Systems (NIPS)*, pp.4768–4777 (2017).
- [7] Sakai A, Komatsu M, Komatsu R, et al.: Medical Professional Enhancement Using Explainable Artificial Intelligence in Fetal Cardiac Ultrasound Screening, *Biomedicine*, Vol.10, No.3, 551 (2022).
- [8] 理化学研究所: 説明可能 AI を用いた超音波画像診断支援 - 胎児心臓超音波スクリーニングへの臨床応用に期待 -, 研究成果 (プレスリリース) 2022, 理化学研究所 (オンライン), 入手先 (https://www.riken.jp/press/2022/20220322_2/index.html) (参照 2023-01-14).
- [9] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu: A Highly Efficient Gradient Boosting Decision Tree, *Proc. Advances in Neural Information Processing Systems (NIPS)*, pp 3149–3157 (2017).
- [10] Grégoire Montavon, Wojciech Samek, Klaus-Robert Müller: Methods for Interpreting and Understanding Deep Neural Networks, *Digital Signal Processing*, Vol.73, pp.1-15 (2018).