

機械学習に基づく中古スマートフォン製品の価格要因分析モデルに関する研究

1X19C123-0 森川卓哉
指導教員 後藤正幸

1. 研究背景・目的

近年スマートフォンの普及に伴い楽天市場やメルカリを始めとする多くのサービスで中古スマートフォンが売買されている。それに付随して多くの取引情報が蓄積されており出品や購入の際に蓄積された情報を活用することが可能となっている。中古スマートフォン市場において“中古買取企業”が“売り手”からスマートフォンを購入し自社の販売サービスを通じて“買い手”に販売するビジネスモデルが存在する。このビジネスモデルにおいて、最適な買取や価格設定を行うためには、市場における中古スマートフォンの価格動向とその価格決定要因を分析することが必要となる。しかし、中古スマートフォンはフリーマーケット形態やオークション形態など様々なサービスで取り扱われているという特徴がある。例えばある販売サービスでは重要となる価格決定要因が他のサービスでは重要とならないケースも多々考えられる。そのような事例に対して勾配ブースティング手法に説明可能 AI を援用したモデルを用いて価格決定要因分析を行うことが考えられるが、これまで多くの中古品販売市場の分析で取り上げられてきた販売サービスの違いを考慮しない単一の価格決定要因分析モデル [1] の適用は望ましくない。

そこで本研究では販売サービス間の価格決定要因が異なる中古スマートフォンに対し説明可能 AI の技術を援用した分析モデルを提案する。具体的には勾配ブースティングを一種である LightGBM [2] によりモデルを構築しその後モデルの特徴量ごとの影響を説明可能な Shapley Additive Explanations (以下 SHAP) [3] を用いて解釈するための方法を提案する。更に実際の中古スマートフォンの販売データを提案手法に基づき分析しどのサービスへの出品が適切かという観点から考察を行う。

2. 準備

2.1. 対象事例のデータの特徴

本研究では各販売サービスにおいてスマートフォン価格がどのように決まっているか把握可能な価格要因分析モデルの構築を試みる。しかし、中古スマートフォンはその種類が膨大であり頻りにアップデートされると同時に機能の有無や傷や汚れなどの状態も 1 台 1 台大きく変わってしまう。またサービスによって状態の値段に対する影響も大きくことなると考えられる。

中古スマートフォン市場はいくつかの EC 販売サービスが主体となって構成されておりサービスによって出品方法や販売方法が異なっている。今回扱うデータの基となる販売形態は大きく 3 つに分けられる。1 つ目は出品者が企業であり自社サービスを通じて独自に価格を決定して販売を行う“企業”形態である。2 つ目は個人が販売サービスを通じて独自に価格を決定して販売できる“フリマ”形態である。3 つ目は個人がサービスを通じて出品し最高額を提示した人に販売する

“オークション”形態である。3 つの形態は販売方法として異なるため出品者は目的に応じたサービスで出品している。そのため各形態では価格決定要因が異なる可能性がある。

2.2. LightGBM とその要因分析方法

LightGBM とは勾配ブースティングを用いた決定木構造のモデルであり計算コストが小さく高精度であるという特徴を持つ。LightGBM ではモデルの出力と正解の誤差を最小化するように逐次的な木の付加を行うことでモデルを構築する。また LightGBM モデルから推定値を特徴量ごとに分析する手法として SHAP が提案されている。SHAP とはそれぞれの特徴量の予測への貢献度を計算することによりあるデータに対する予測を説明する手法である。貢献度は Shapley 値を使って計算される。

3. 提案

3.1. 提案の着想

中古買取企業が適切な買取や価格設定をするためには価格決定要因を分析することが必要である。その方法として LightGBM モデルを構築し SHAP を適用することが考えられる。しかしながら中古スマートフォン市場の持つ以下の特徴から販売サービスの違いを考慮しない単一の分析モデルの適用は難しい。

中古スマートフォン市場は販売形態やサービスが多岐にわたるといえる特徴を持つ。フリマやオークション形式など様々な形態で販売されており販売サービス毎に買い手も異なっていると考えられる。そのためサービスや形態毎に価格決定の構造が異なると考えられる。そこで提案では販売サービス毎のモデル構築と価格決定要因分析を行うことで販売サービス間の価格決定要因の違いを考慮した分析を提案する。

3.2. 販売サービスの違いを考慮した価格要因分析モデル

本研究では中古スマートフォン市場の特性を踏まえ販売サービス間での価格決定要因が異なる製品に対する価格決定要因分析モデルを提案する。具体的には販売サービス毎に個別のモデルを構築し、サービス毎に作った LightGBM モデルに SHAP を適用することでサービスにおいて大きく影響を与える特徴量を確認する。本研究では、全ての販売サービスのデータで学習した“サービス全体”モデルとサービス毎のデータで学習した“サービス個別モデル”の差分による分析を行う。これにより価格決定要因が異なる場合には個別でモデルを作成した方が適切であることを示す。その後、各サービス毎に個別の LightGBM モデルを学習し学習したモデルに SHAP を適用することで各サービスにおける価格決定要因の差異についての分析を行い、各販売サービスで他サービスと比較して高額で販売される製品の特徴を示す。これにより、企業の適切な買取と価格設定を実現する。

4. 実データ分析

4.1. 分析の目的

本節では全ての販売サービスのデータで学習した“サービス全体”モデルとサービス毎のデータで学習した“サービス個別モデル”についての比較を行い個別モデルの優位性を確認する。これにより個別モデルの方が高精度で価格を表現するモデルであることを確認する。また個別モデルに SHAP を適用することでサービス毎の影響の大きい価格決定要因を抽出すると同時にサービスによって高価売られる要因が大きく異なることを示し提案手法の妥当性について確認する。

4.2. 分析条件

実験で用いる中古スマートフォンのデータは 8 つのサービスから集計し iPhone の取引データのみを抽出する。また異常な販売価格データを取り除くために 3 万円以上 20 万円未満のデータのみを抽出した結果データ数は 993165 件となった。またモデル作成のための学習データとテストデータを全てのモデルで 7:3 の割合で分割しテストデータに対する予測精度で評価を行う。個別モデルは 63 種の特徴量から作成し全体モデルはそれに加えて販売サービス名を特徴量として加える。

4.3. サービス全体モデルとサービス個別モデルの差異

サービス全体モデルとサービス個別モデルを構築した。評価指標をそれぞれのテストデータに対する Root Mean Square Error(以下 RMSE) と決定係数(以下 R^2) を表 1 に示す。 R^2 は 0 から 1 の値を取り 1 に近いほどモデルの精度が高いことを示す。個別モデルの R^2 を $R^2(\text{個別})$ で示し、全体モデルの R^2 を $R^2(\text{全体})$ で示す。

表 1: 全体モデルと個別モデルの精度

形態	サービス名	RMSE(個別)	$R^2(\text{個別})$	$R^2(\text{全体})$
-	全体	5328	-	0.964
企業	楽天市場	3145	0.989	0.983
企業	ヤフーショッピング	4015	0.963	0.937
企業	リコレ	4911	0.968	0.956
企業	ムスビー	4372	0.954	0.938
企業	イオシス	3271	0.991	0.978
フリマ	メルカリ	6798	0.923	0.911
フリマ	ラクマ	9884	0.868	0.859
オークション	ヤフオク	7015	0.917	0.902

表 1 の結果から $R^2(\text{個別})$ と $R^2(\text{全体})$ を比べると個別の方が精度が高いことが分かる。サービス全体モデルでは学習データ数が多いにも関わらず精度がサービス個別モデルより悪いことから各サービスにおける価格決定要因が異なる場合には個別でモデルを作成する方が適切といえる。

4.4. SHAP の実装結果

図 1 は楽天市場の個別モデル(楽天モデル)に SHAP を適用した際の特徴量のバイオリンプロットである。楽天市場では 64GB128GB などの低容量の製品が低価格で販売されていることが分かる。また sim 未解除製品が低価格つまり sim 解除済みの製品が高額で販売されていることが分かる。同様に他のサービスの個別モデルに SHAP を適用し他のサービスと比較して価格が高い要因を持つ製品を分析した結果を表 2 に示す。

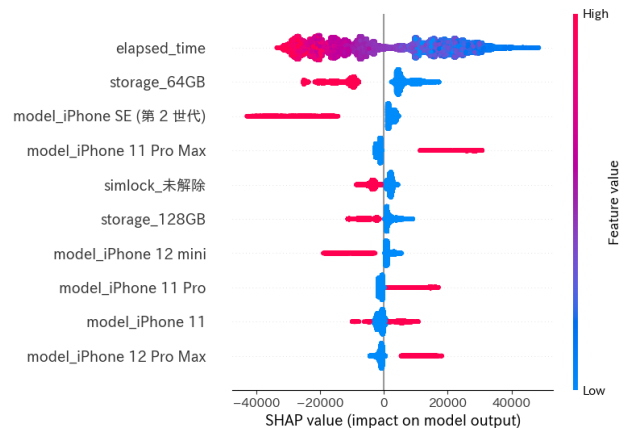


図 1: 楽天モデルにおける SHAP のバイオリンプロット

表 2: サービス毎の高価格製品

形態	サービス	高価格で売られる製品
企業	楽天市場	sim 解除済みの製品
企業	ヤフーショッピング	目立った傷や汚れが無い製品
企業	リコレ	simfree ネットワーク制限が無い製品
企業	ムスビー	新品の製品
企業	イオシス	64GB の製品
フリマ	メルカリ	新品の製品
フリマ	ラクマ	新品の製品
オークション	ヤフオク	新品の製品

企業形態の販売サービスではサービス毎に価格が高い製品が異なる。これは販売サービス毎に出品している企業が異なり各企業らの価格決定の違いによりサービス独自の価格構造が構築されているためと考えられる。またフリマ形態やオークション形態では新品に近い製品が好まれる。これらのフリマやオークションサービスでは出品内容と実物が異なる事例などがありこのようなサービス形態ではユーザが“新品”など確かな製品を好む傾向にあるためと考えられる。

5. 結論と今後の課題

本研究では LightGBM モデルに SHAP を適用することで高い精度を保ちつつ価格要因を分析する手法を提案した。また実データ分析によりスマートフォンの状態をもとに最も高く売れるようなサービスの発見やターゲットに合わせたサービスの選定が可能と考えられる。今後の課題として iPhone と他機種における価格決定要因の差異やサービス間の類似性に関する分析などが挙げられる。

参考文献

- [1] 工藤大輝 他 “機械学習による中古車落札価格の要因分析および割安評価” 情報処理学会論文誌数理モデル化と応用 (TOM) 14 (3) 39-50
- [2] Ke Guolin et al. “LightGBM: A highly efficient gradient boosting decision tree.” Advances in neural information processing systems 30 (2017): 3146-3154.
- [3] Lloyd S. Shapley. “A Value for n-person Games.” In Contributions to the Theory of Games volume II by H.W. Kuhn and A.W. Tucker editors. Annals of Mathematical Studies v. 28 pp. 307-317. Princeton University Press.