

修士論文概要書

Master's Thesis Summary

Date of submission: 01/10/2024 (MM/DD/YYYY)

専攻名 (専門分野) Department	経営システム 工学専攻	氏名 Name	増田 雅樹 Masaki Masuda	指導 教員 Advisor	後藤 正幸 印 Seal
研究指導名 Research guidance	情報数理応用研究	学籍番号 Student ID number	5222C031-7		
研究題目 Title	機械学習に基づく中古スマートフォン端末の将来価格予測モデルに関する研究 A Study on Future Price Prediction Models for Second-hand Smartphone Devices Based on Machine Learning Model				

1. はじめに

近年、携帯電話やスマートフォンの中古品を利用するユーザーや中古端末を対象としたサービスを展開する企業が増加し、それに伴い中古端末の市場規模は拡大している。中古端末の価格は、定価が定まっていないことや市場の流動性の高さから非常に変動しやすいことが特徴として挙げられる。さらに中古端末の価格は、発売からの経過日数や端末の状態といった観測可能な要因だけでなく、世界情勢の変化や自然災害等の予測不可能な要因によって変動すると考えられる。そのため、将来価格を予測することは非常に難しいタスクとなっている。一方で、企業においては将来価格を予測することは自社の利益拡大を図るために重要な課題である。なぜならば、将来価格の高精度な予測により、企業における各サービスの最適な価格設定や在庫管理を行うことが可能になるためである。例えば、レンタルサービスではレンタル料金を設定するための基礎情報として将来の予測価格を活用し、販売サービスでは現在の仕入れや在庫管理のための基礎情報として活用することができる。このとき、企業の意思決定においては短期的な価格変動ではなく数年先の長期的な価格変動の予測が必要となる。

本研究ではネット通販サイトにおける過去の中古端末の売買データを基に、中古端末の将来価格を予測するモデルの構築を目的とする。このとき価格の長期的な傾向を捉え、かつ今後発売される新機種の価格予測も可能なモデルの構築が望まれる。さらに、中古端末は時期がほぼ同じであっても特定の金額で売買されるわけではなく、ある程度のばらつきの中で売買される。このばらつきは端末発売からの経過日数や定価といった端末情報だけでなく予測不可能な事象といった外的要因によって変化することを考慮する必要がある。従って、将来価格のばらつきを予測する必要がある。そこで本研究においては、予測不確実性を考慮したモデルである Natural Gradient Boost (NGBoost)[1]を用いて過去の売買データをモデル化する。このとき、「値上げ」や「新機種発売」などの外的要因を表す情報を特徴量として追

加することで端末情報に依らない価格変動を考慮可能とする。また、本研究で使用するデータセットの期間は約2年間であるため、長期的な予測をするためにデータセットを全て学習データとしてモデルに学習させる。そして、外的要因の発生を仮定した複数のシナリオに対して提案モデルを用いたシミュレーションを行う。最後に、実際に企業が業務を行う上で提案モデルがどのように活用可能であるかについて考察を行う。

2. 基礎分析

2.1. データセットの概要

本研究で使用するデータセットは中古スマートフォンを取り扱うネット通販サイト A における過去の売買データである。データの取得期間は2021年7月から2023年7月となっており、週次で売買履歴が記録されている。データには、データの取得日や機種、ストレージ容量、端末の状態、キャリア等が記録されている。なお、本研究で対象とするスマートフォンの機種は、機種の発売順がわかりやすいよう、iPhone シリーズの中で iPhone5 以降に発売された 33 機種とした。

2.2. 各月における機種ごとの売買台数

図 1 に対象期間における各月の機種ごとの売買台数を示す。なお、本節においては発売日を考慮して iPhone5, iPhone6, iPhone7, iPhone8, iPhoneX, iPhone11, iPhone12, iPhone13, iPhone14 の 9 機種を対象とする。図 1 より、2021 年 8 月以降はどの機種においても売買台数が右肩下がりになっていることが分かる。これは、新型コロナウイルス感染症や世界規模での半導体不足等が影響していると考えられる。一方、2022 年 5 月および 6 月には急激に売買台数が増加しているが、これは急速な円安によって外国人購入者が増加したためと考えられる。さらに、同年 7 月には iPhone の値上げが行われたことにより一時的に売買台数が減少している。

2.3. 機種ごとの売買平均価格と移動平均価格の推移

本節では対象期間における各機種の販売価格変動を確認する。図 2 に売買が行われた日における、機種ごとの平均

価格および前後 7 日移動平均価格を示す。本節で対象とする機種は前節と同様の機種とする。

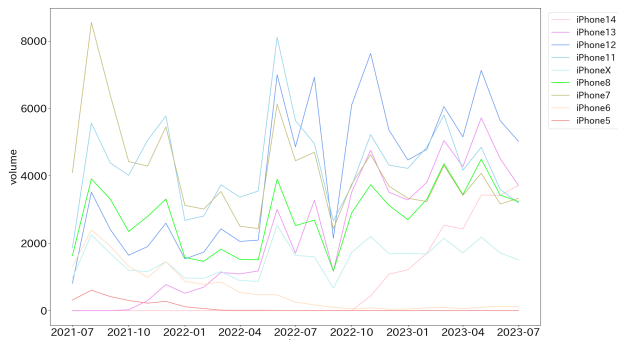


図 1. 対象期間における各月の機種ごとの売買台数

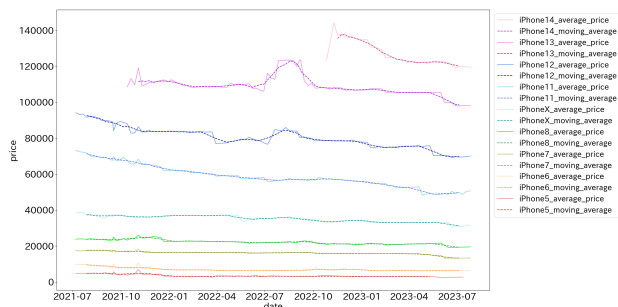


図 2. 対象期間における平均価格と移動平均価格の推移

図 2 より、古い機種ほど価格推移が横ばいとなっており、新しい機種は全体的に右肩下がりな推移していることがわかる。しかし、短期的には減少傾向と異なる変動も見られる。例えば、iPhone13 および iPhone12 は 2022 年 7 月付近で価格が上昇しているが、これは前節で述べた iPhone13 や iPhone12 を含む特定機種の価格の値上げが行われたことに起因するものと考えられる。しかしその後価格が下落していることから、新品価格の値上げは中古端末に対しては短期的な影響であったと言える。さらに、前節では分析対象の全機種が同様の推移をしていたが、売買価格に関しては機種によって異なる推移をしている。このように、価格変動においては単に経過日数による価格の減少だけでなく外的要因による特定の機種への影響による変動も発生することがわかる。

2.4. 検索トレンドと価格変動の関連性

前節までの基礎分析から、価格変動には対象端末の特徴（経過日数や傷の有無等）に依らない外的要因が影響していると考えられる。そのため、本節においては、ニュースなどの報道情報や SNS、検索エンジンのトレンド情報から、外的要因による影響が価格変動にどのように表れているかを確認する。ここでは、Google の検索トレンド [2] に焦点を当て、対象期間内でトレンドワードに「iPhone」を含む検索トレンドを抽出する。抽出したトレンドを大別すると「値上げ」「新機種」「既存機種」の 3 つの関連単語であった。表 1 にデータ期間において上位 50 件以内に出現した検索トレンドの出現回数を示す。また、図 3 に検索トレンドと

価格変動の関係を示す。図 3 において、縦線は「iPhone」を含む検索トレンドが出現した日を表しており、縦線の色が濃くなっている部分は複数の検索トレンドが同時に出現していることを示している。

表 1. 各関連単語の出現回数 (2021 年 7 月～2023 年 7 月)

関連単語	値上げ	新機種	既存機種
出現回数	8	14	4

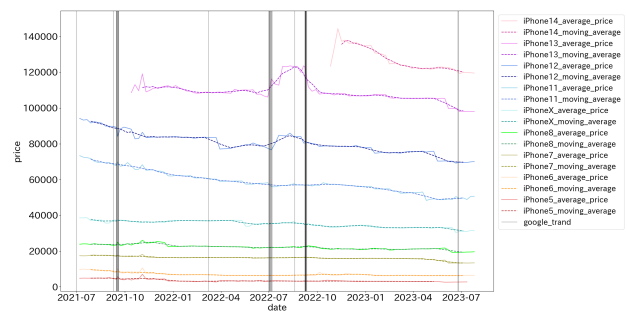


図 3. 検索トレンドと価格変動の関係

図 3 より、検索トレンドが出現したタイミングで価格変動が起こっていることが分かる。このことから、トレンド情報を予測モデルに組み込むことで、これらの外的要因に起因した価格変動を考慮することが可能になると考えられる。

3. 提案への着想

3.1. 従来の点推定を用いた価格予測と本研究の課題

将来価格の予測を対象とする問題には、株価や外国為替の予測が挙げられる。Hu ら [3] の研究では、株価や FX の予測に用いられるディープラーニング手法がいくつか挙げられている。例えば、畳み込みニューラルネットワーク (CNN) や再帰型ニューラルネットワーク (RNN)、Long Short-Term Memory (LSTM) などがある。これらの手法は点推定によって将来価格を予測するモデルとなっている。

上記のように、一般的な回帰問題においては、入力データに対して予測結果を 1 つの値で出力することが多い。しかし本研究で対象とする中古端末においては、予測対象の中古端末は同じ条件であっても特定の金額で売買されるわけではなく、ある程度のばらつきの中で売買される。さらに、企業における価格設定においては、将来の価格がどの程度の価格幅を持つかが意思決定に大きな影響を与える。そのため、本研究が対象とする問題においては実測値が予測価格の範囲の中に収まっていることが重要であり、点推定による予測が実測値と同値であることを追求することはあまり重要ではない。すなわち、本研究においては将来における価格のばらつきを予測するために、予測不確実性を考慮することを考える必要がある。以上より、本研究では点推定ではなく不確実性を考慮可能なモデルの構築を考える。

3.2. NGBoost[1]

NGBoost は予測値を確率分布として推定することによって、予測に対して不確実性を考慮可能なモデルである。例

例えば、確率分布に正規分布を仮定した場合、正規分布のパラメータである平均と分散が予測値として出力される。通常、点推定を行う場合においては、実測値と予測値を比較するために損失関数を用いて評価する。しかしばらつきを考慮するため、NGBoost では損失関数の代わりにスコアリングルールを用いて評価を行う。このとき、スコアリングルール S は推定されたパラメータ θ を持つ確率分布 P と実測値 y のスコア $S(P, y)$ として算出される。推定した確率分布 P が実測値の真の分布 Q に近いほどスコア $S(P, y)$ は低い値をとるため、以下の式 (1) を満たす。

$$E_{y \sim Q}[S(Q, y)] \leq E_{y \sim Q}[S(P, y)] \quad \forall P, Q \quad (1)$$

また、NGBoost ではパラメータ θ を求めるために、スコア $S(P, y)$ の勾配 $\nabla S(P, y)$ に対してフィッシャー情報量の逆行列をかけた自然勾配を用いて最適化を行う。そして、勾配ブースティングを用いて自然勾配を予測するように複数の弱学習器を学習させ、効率的に学習を行うことが可能となっている。最終的に、入力データ x に対するパラメータ θ は式 (2) で表される。このとき、 $\theta^{(0)}$ はパラメータ θ の初期値、 η は学習率、 $f^{(m)}$ と $\rho^{(m)}$ はそれぞれ m 個目 ($m = 1, 2, \dots, M$) の弱学習器とそれに対する重みである、

$$\theta^{(M)} = \theta^{(0)} - \eta \sum_{m=1}^M \rho^{(m)} \cdot f^{(m)}(x) \quad (2)$$

3.3. 本研究への展開

NGBoost を活用した例として、Kavzoglu ら [4] や Zhu ら [5] の研究が挙げられる。Kavzoglu らは地滑りの発生予測に対して NGBoost を用いており、Zhu らは岩盤の深さを予測するために NGBoost と XGBoost を組み合わせたモデル (N-XGBoost) を提案している。これらの結果から、事象に対する要因が不安定である場合や意思決定の信頼性を高めるために、予測の不確実性を考慮することは非常に重要であると言える。本研究で対象としている中古スマートフォンの将来価格においては、発売からの経過日数や対象端末の状態の変化だけでなく、外的要因による売買価格の不確実性が課題である。一方で、不確実性を考慮することで、企業の価格戦略において柔軟で信頼性の高い意思決定が可能になると考えられる。そこで、本研究ではデータセットに対して NGBoost を適用し、中古スマートフォンの将来価格を確率分布の形で予測することとする。このとき、2.4 節のトレンド情報を特徴量として追加することで、外的要因による変動を考慮した価格のばらつきを予測するモデリングが可能となる。

4. 提案モデル

4.1. 提案モデルの概要

本研究においては正規分布を仮定した NGBoost を用いる。企業においては、各サービスにおける価格設定が課題となっている。各サービスにおいては仕入れ～売却、貸出

～返却の期間が複数年に渡るため、短期的な予測ではなく予測時点から数年先のより長期的な予測が必要となる。そこで、提案モデルにおいては目的変数を移動平均価格とすることにより、短期的な価格変動ではなく長期的な傾向を把握することを考える。このとき、今後発売される機種のように学習データにない機種の予測できるよう、提案モデルにおいては機種に関する情報を入力特徴量から除外している。またストレージ容量に関して、機種ごとに発売されるストレージ容量の差異を標準化するために、同一機種におけるストレージ容量が小さいものから順にラベリングしたものをストレージ容量を表す特徴量とする。さらに、トレンド情報を組み込むことで端末情報に依らない外的要因を考慮した予測を可能にする。モデルに入力する特徴量を表 2 に示す。なお、トレンド情報は日次データであるため、?? 節で示した関連単語のいずれかとして週次データに変換した上で入力している。

表 2. モデルへの入力特徴量

特徴量	入力形式	次元数
ストレージ容量	1,2,3,4	1
キャリア	one-hot ベクトル	4
simlock 解除の有無	0,1	1
ネットワーク制限の有無	one-hot ベクトル	3
端末の状態	one-hot ベクトル	6
発売からの経過日数	実測値	1
定価	実測値	1
トレンド情報 (値上げ)	0,1	1
トレンド情報 (新機種)	0,1	5
トレンド情報 (既存機種)	0,1	1

4.2. 提案手法の学習

本研究における問題設定上、予測時点 t までの全てのデータが利用可能であるため、 t までのデータを訓練データとしてモデルに学習させる。ここでは、 t を 2023 年 7 月 12 日とし、データセットの全期間および同期間のトレンド情報を学習データとして使用している。提案手法の学習データに対する精度を表 3 に示す。出力は正規分布のパラメータである平均と分散であるため、平均に対しては R^2 と RMSE、分散に対しては予測された分散を基に算出した 3σ 内に正解データがどの程度含まれているかで評価を行う。

表 3. 予測モデルの訓練データに対する精度

データ数	1,389,068
予測した σ の平均	5170.919
R^2	0.980
RMSE	5854.115
予測した 3σ 以内の割合	0.999

5. シナリオシミュレーションに基づく将来価格予測

本研究の問題設定上、短期的な予測ではなく長期的な予測が必要であるため、データセットを学習データと検証データに分割しても 2, 3 年先の予測精度を検証することは

きない。そこで、データセットを全て学習データとしてモデルに学習させる。そして、トレンドの出現有無を仮定した複数のシナリオに対する人工データを作成し、提案モデルを用いて予測した将来価格のシミュレーションを行う。

5.1. シミュレーション条件

本研究におけるシミュレーション期間は2023年7月12日から2025年7月31日までとし、出現した関連単語は「値上げ」「新機種」を仮定している。このとき、各シナリオにおける価格変動の差異を観察する。表4に本シミュレーションで使用する人工データの作成条件を示す。このとき、定価や経過日数（予測期間）に関してはどちらのシナリオを仮定した場合でも同一であるため割愛している。

表 4. 人工データの作成条件

	シナリオ 1	シナリオ 2
対象端末	iPhone14	iPhone13
端末の状態	新品	新品
network 制限の有無	有	有
キャリア	NTT ドコモ	NTT ドコモ
simlock 解除の有無	解除済	解除済
ストレージ容量	128GB	128GB
	256GB	256GB
	512GB	512GB
トレンド情報	値上げ	新機種
トレンド出現回数	2回	2回

5.2. 実験結果

図4および図5にそれぞれシナリオ1とシナリオ2に対するシミュレーション結果を示す。各図において縦線がトレンドが出現した時点を表している。各シナリオにおいて、トレンドが出現した際に価格変動していることが分かる。

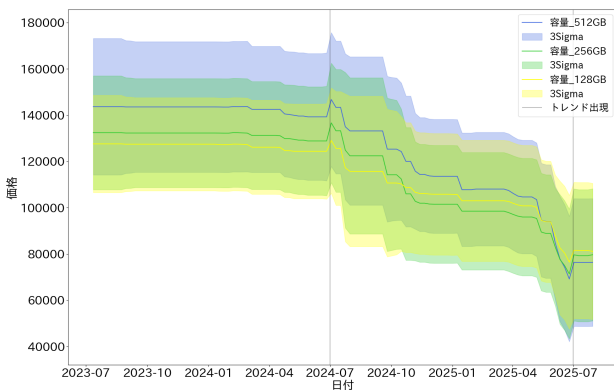


図 4. シナリオ 1 に対するシミュレーション結果

6. 考察

実験結果より、外的要因を考慮した価格のばらつきをモデル化することができた。一方で、図4および図5から、トレンドが出現していないにも関わらず価格が急落していることが分かる。これは、機種ごとの売買台数の偏りによって学習が上手くできていない可能性が考えられる。そのため、偏りが少ないようデータ選択する必要があると考えられる。また、企業においては不確実性を考慮した予測を行うことで対象端末の将来価格がどの程度のばらつきの中で売

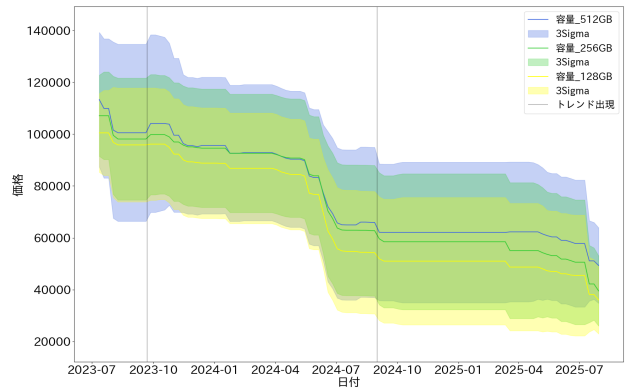


図 5. シナリオ 2 に対するシミュレーション結果

買されるかが予測可能となり、点推定と比較してより柔軟で信頼性の高い意思決定が可能になると考えられる。さらに、将来に対する複数のシナリオを用いてシミュレーションを行うことで、各サービスにおいて適切な価格設定が可能となるため企業的意思決定にとって有益であると考えられる。

7. 結論と今後の課題

本研究では、中古スマートフォンの売買データに NG-Boost を適用し、将来の中古価格のばらつきを予測可能なモデルの提案を行った。さらに、提案モデルはトレンド情報を用いることで外的要因による価格変動を考慮することが可能である。提案モデルを用いてトレンドの出現を仮定したシナリオに対するシミュレーションを行い、その有用性を示した。一方で、提案モデルに対する学習データの選択方法やより正確な外的要因による影響を反映させることが必要であると考えられる。

謝辞

本研究を行うにあたり用いた貴重なデータは株式会社 Belong 様よりご提供いただきました。深く感謝致します。

参考文献

- [1] Tony Duan, Avati Anand, Daisy Yi Ding, Khanh K Thai, Sanjay Basu, Andrew Ng, and Alejandro Schuler. Ngboost: Natural gradient boosting for probabilistic prediction. In *International conference on machine learning*, pp. 2690–2700. PMLR, 2020.
- [2] Trend calendar, <https://jp.trend-calendar.com/>. 最終閲覧日 2023/11/5.
- [3] Zexin Hu, Yiqi Zhao, and Matloob Khushi. A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, Vol. 4, No. 1, p. 9, 2021.
- [4] Taskin Kavzoglu and Alihan Teke. Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (xgboost) and natural gradient boosting (ngboost). *Arabian Journal for Science and Engineering*, Vol. 47, No. 6, pp. 7367–7385, 2022.
- [5] Xing Zhu, Jian Chu, Kangda Wang, Shifan Wu, Wei Yan, and Kiefer Chiam. Prediction of rockhead using a hybrid n-xgboost machine learning framework. *Journal of Rock Mechanics and Geotechnical Engineering*, Vol. 13, No. 6, pp. 1231–1245, 2021.