

中古自動車の査定価格決定支援システムにおける区間推定モデルの評価 Evaluation of an Interval Prediction Model for a used car price determination support system

林 雄太郎¹⁾ 松下 旦²⁾ 岩崎 敦³⁾
Yutaro Hayashi Akira Matsushita Atsushi Iwasaki

1 はじめに

本稿では、中古車価格を予測するための機械学習モデルの仕様変更における評価について述べる。(株)オークネットは BtoB のオンライン中古車売買支援サービスを手掛けており、業務のひとつとして、顧客から持ち込まれた中古車両の価格査定を行うものがある。従来は査定担当者が車両の状態や直近の相場を確認しつつ、手動で査定価格(幅値)を算出していたが、担当者の負担減や査定時間の短縮が望まれていた。そこで、2020 年に車両の情報および直近の取引価格を入力とし、担当者に価格幅を提案する「中古自動車の査定価格決定支援システム」を開発し、実運用を開始した。担当者は当システムから提示された区間値を確認し、必要があれば調整したうえで、顧客に返信する。その背後では、自社で運営している中古自動車オークションでの入札データから構築した機械学習モデル、および機械学習モデルの定期的なバージョン更新システムが運用されている [1]。このシステムにより、担当者ごとの査定価格のばらつきを抑えつつ、査定依頼を受信してから返信までにかかる時間を短縮することに成功した。

これまで運用してきた機械学習モデルは、勾配ブースティング決定木アルゴリズム (XGBoost, LightGBM) により構築した「点推定モデル」であり、未知の入力を受け取って価格の推定値を出力する。その後、機械学習モデルの推定値を受け取ったシステムが、規定のロジックに従い区間推定値(幅値)に変換し、査定担当者に表示している。これまでは、この点推定モデルの性能を向上させる取り組みを研究として行ってきた。[2, 3]

今回は、モデルとして区間推定値を出力する「区間推定モデル」を構築し、従来の点推定モデルと規定のロジックを併用した現行のシステムと比較して、どのようにそのパフォーマンスが変化するのか、評価する研究を行った。その結果、信頼区間を大きく設定した場合、評価指標上では区間推定モデルがよい結果を出したものの、信頼区間が小さくより狭い区間での推定を行わせたとき、現行システムが勝る結果となった。

2 中古自動車価格査定サービス

本節では、オークネットにおける「中古自動車価格査定サービス」および、機械学習モデルを用いた「中古自動車の査定価格決定支援システム」について説明する。

まずは、「中古自動車価格査定サービス」について説明する。街中にあるような中古自動車販売店は通常、商品となる中古自動車について、一般顧客や事業者からの

依頼を受けて価格査定を行う。その後、自社で買い取って販売するか、オークションを通じて他の事業者へ販売する。中古自動車販売店が一般顧客から下取りを行う際には、どのような価格で買い取るべきかを判定する際には、担当者が市場価格の動向や車の状態をもとに、その車両が将来オークションで出品された場合、どれくらいの価格で落札されるかを予想した上で、損をしない程度に査定価格を決定してきた。ただし、中古自動車販売店によっては、自社が詳しくない車種であったり、買い取り価格を正確に判定できない場面も存在する。そのため、オークネットはそのような事業者向けに、中古自動車価格査定サービスを提供してきた [4, 5]。中古自動車価格査定サービスでは、車両と相場価格に熟知したオークネット社員が、販売店から送られてきた車両の写真・車両状態の情報をもとに査定し返信することに加え、査定担当者がいない中古自動車販売店への下取りの代行や、その後のオークションへの出品に至るまでのサービスの提供を行ってきた。このような価格査定業務はとも属人性が強いほか、年間数万件を超える査定依頼を全て人力で処理するのは社員にも負担がかかっており、オークネット内の査定担当者は当サービスの運用継続に苦労していた。これら課題に対して、機械学習モデルを用いた「中古自動車の査定価格決定支援システム」を構築、運用し、属人化の解消と業務効率化を目指した。

次に、このシステムで採用している機械学習モデルについて説明する。この機械学習モデルは、オークネットが所有するデータである車両情報とオークションでの落札価格の関係を学習した回帰モデルであり、車両に関する情報を入力として受け取り、その車両の価格を点推定値として出力する。当システムでは、機械学習モデルに車両情報を入力として送信し、推定値を受信した後、システム内の規定のロジックに従い区間推定値(幅値)に変換し、ユーザーである査定担当者に表示している。査定担当者はこの区間推定値を参考価格とし、一部定性情報を加味し修正を加えたうえで、顧客に査定価格を区間値で返信する。ここでいう定性情報とは、具体的にはペットやタバコの匂いの有無、オプションパーツの有無など、査定時にメモ欄に書いてある情報などを指しており、査定価格に影響を及ぼし得るデータとなる。しかし、現時点では機械学習モデルに対してこれら定性情報を入力として学習させるのは難しく、機械学習モデルの推定値をもとに最終的には査定担当者が確認し、修正した上で顧客に査定価格を返信している。

オークネットでは 2020 年にこのシステムを実装し、2024 年現在まで運用し続けることで、価格予測の精度向上や査定依頼処理時間の短縮といった KPI を改善することに成功した。ここでは、価格予測の精度を「オークションで成約した価格と提案システムによる査定価格

- 1) 株式会社オークネット
- 2) 京都大学
- 3) 電気通信大学

幅との差」として、査定依頼処理時間を「顧客から査定依頼を受けて担当者が作業を始めてから最初の査定価格を返信するまでの時間」として定義する。

3 区間推定モデル構築の背景

本節では、区間推定モデルを構築、評価するに至った背景について説明する。

それは点推定モデルを取り扱う現行のシステムが行う「区間推定値」の出力方法にある。前節でも取り上げたように、現行のシステムは機械学習モデルから受け取った点推定値をあらかじめ規定したロジックに従い、機械的に区間推定値へと修正している。また、査定担当者は、システムが出力した値を確認し、必要に応じて修正し再度区間推定値として算出する。

最終的な正解値を、査定した中古自動車がオークションに出品された場合の落札結果とする。この場合、区間推定値が実際落札価格を包含し、かつより狭い区間で推定できるほど、性能が良い機械学習モデルであるといえる。現行のシステムは、予測値の出力までは機械学習モデルが担うが、その後の区間推定値への修正処理はシステムおよび査定担当者が行っている。これを、そもそもモデルが区間推定値を出力するようにさせるなど、現在の区間推定値の出力方法には改善の余地があるのではないかと考え、今回の研究を行うに至った。結果として、後述の評価指標上では、より大きな信頼区間で比較した場合は区間推定モデルが現行システムに勝る結果となったが、信頼区間を下げた場合は現行システムが勝った。

4 検証内容

本節では、今回の検証内容について説明する。

4.1 データに関する基本情報

本研究では、データとして主に2種類を使用した。現行システムを用いて行われた査定業務に関するデータと、オークネットが保有する中古車オークション市場での落札データとなる。

査定業務に関するデータとは、以下のように3種、計5個の値となる。区間値は、下限・上限の2つを含む。

1. 現行の機械学習が出力した予測値
2. システムが出力した修正「区間値」
3. 査定担当者が確認し、最終的に決定した「区間値」

中古車オークション市場での落札データとは、メーカー、車種、走行距離、経過年数、色、車両のグレードなど車両に関する19個の特徴量、その車両がオークションにて出品された際の落札価格となる。

両データは車両ごとに一意となる「車台番号」を持つため、データの結合条件としてこれを利用した。これにより、車両ごとにどのような特徴を持ち、現在稼働している点推定モデルがどのように推定し、システムや査定担当者がどのような修正を加えたのか、最終的にオークションでどのような応札値がついたのかを追跡できるようなテストデータを作成した。テストデータの期間は、直近のデータとして2024年3月から2024年5月までと抽出している。また、区間推定モデル作成後は、このテストデータに含まれる車両の特徴量を用いて区間推定

値を出力し、区間推定モデルの評価および現行システムとの比較した。

現行システムで用いられる点推定モデルは、相場変動に対応させるため、機械学習に用いるデータ期間を月に2回の頻度で更新している。よって、区間推定モデルの評価は、同時期の学習データを用いて更新された点推定モデルが、実際の査定業務に使われた時期に従って区切りを入れたテストデータを用いて行うものとする。

4.2 評価指標

本研究では区間推定モデルの出力した区間推定値と正解値を用いてパフォーマンス評価を行うため、評価指標には Winkler Score [6] を用いた。

定義 1 推定区間の下限を L 、上限を U 、真の値を Y とし、有意水準を α とする。このとき Winkler Score WS を以下に定義する：

$$WS = \begin{cases} (U-L) + \frac{2}{\alpha}(L-Y) & \text{if } Y < L, \\ U-L & \text{if } L \leq Y \leq U, \\ (U-L) + \frac{2}{\alpha}(Y-U) & \text{if } U < Y. \end{cases}$$

このスコアは、推定区間の幅と真の値がその区間内に含まれているかどうかに基づいて計算される評価指標である。この値が小さいほど、そのシステムやモデルがうまく区間推定していると言える。推定区間が真の値を包含している場合は区間幅がスコアとなり、包含していない場合は区間幅にペナルティが加えられスコアとなる。単に推定区間を拡大すればスコアが大きくなってしまいが、真の値を包含する確率が高まりペナルティを受ける確率を低め、結果的にスコアを小さくすることができるかもしれない。逆に、推定区間を狭めればスコアを小さくできるが、真の値を包含する確率が低まりペナルティを受ける可能性が高くなり、結果的にスコアが大きくなってしまふかもしれない。このように、区間幅とペナルティという2つの条件によって、Winkler Score は算出される。

5 区間推定モデル

本節では、今回の検証にて新規構築する区間推定モデルについて説明する。

5.1 区間推定モデルの構築

区間推定モデルの構築は、勾配ブースティング決定木アルゴリズムのひとつである LightGBM の分位点回帰を用いて行った。モデル構築に使用するハイパーパラメータの調整により、幅を持たせた予測、すなわち区間推定が可能となる。また、区間推定モデルとはいえ、LightGBM で構築されたモデルのそれぞれは点推定値を出力するため、区間下限値、区間上限値を出力する2つのモデルを構築し、両者の出力値をもって区間推定値とする。なお、現行システムで用いられる点推定モデルは、LightGBM の回帰を用いて構築されている。

本検証では、データ期間別に5種類、信頼区間 (confidence interval) の値別に4種類、計20種の区間

表1 信頼区間, テストデータ期間別の Winkler Score

信頼区間	区分	24/03 前	24/03 後	24/04 前	24/04 後	24/05 前
95	現行システム	557.9	398.7	260.3	265.7	314.0
	区間推定モデル	135.0	120.5	76.8	110.9	106.3
90	現行システム	282.1	202.9	133.5	135.8	160.1
	区間推定モデル	109.5	76.9	62.0	76.8	75.8
85	現行システム	190.5	137.6	91.2	92.6	108.8
	区間推定モデル	130.7	77.2	53.1	67.0	69.1
30	現行システム	46.6	35.0	24.8	24.6	28.2
	区間推定モデル	49.6	36.5	31.4	33.2	35.2

推定モデルを構築した。学習に用いたデータの期間については、現行システムが利用する点推定モデルが学習した期間と同等に設定している。信頼区間については、代表的な95%信頼区間から5パーセントごとに下げたものを3つ、実際の査定業務で行われるようなより狭い区間での予測ができるように、30%信頼区間を設定したものを加え、4種類とした。信頼区間の値により、区間推定モデルの構築に使用するハイパーパラメータや、Winkler Score 算出時に使用する値が定まる。例えば、信頼区間が95%のとき、有意水準は5%となり、区間の下限は2.5%、上限は97.5%となる。

また、モデル構築に用いるハイパーパラメータについては、分位点回帰を行うために不可欠なパラメータ (objective, metric, alpha) のみを追加・変更した。これは現行システムで運用されている点推定モデルと条件をできる限り合わせた上で、区間推定モデルのパフォーマンスを測るためである。

6 検証結果と考察

本節では、区間推定モデルの構築と現行システムとの比較を行った結果と考察について記載する。

6.1 検証結果

現行システムと区間推定モデルの両者を評価すべく、共通のテストデータを用いて各期間の Winkler Score を算出した結果を、表1に示す。表内の数値の太字は、Winkler Score という評価基準上で現行システム、区間推定モデルのいずれかが一方より優れたことを指す。例えば、2024年5月前半(表内「24/05前」)の場合を見ると、区間推定モデルで計算した Winkler Score は、信頼区間が95%のとき106.3、90%のとき75.8、85%のとき69.1、30%のとき35.2と、信頼区間が下がるに伴って、Winkler Score も徐々に小さくなっていくことが分かる。

また、Winkler Score とは別に、正解値であるオークション落札結果が推定区間に含まれた確率を信頼区間別に算出した。結果を表2にて示す。区間推定モデルの方は、信頼区間が95%から30%へ下がるに伴い、正解値を包含する確率が90.0%から27.9%へと下がっていることが分かる。これは、信頼区間が下がることによってそれぞれの推定区間が狭まっていることが原因だと考えられる。信頼区間別に、落札結果と現行システム、区間推定モデルの各区間推定値の関係を表したものを図1, 2, 3, 4に示した。これを見ると、区間推定モデルの区間推定値の上限は緑、青の点で示され、信頼区間が下がるにつれて、上限と下限の区間幅が狭まっていることが分かる。なお、現行システムのデータにより算出した値は、推定区間が変わらないため、31.5%で一定となる。

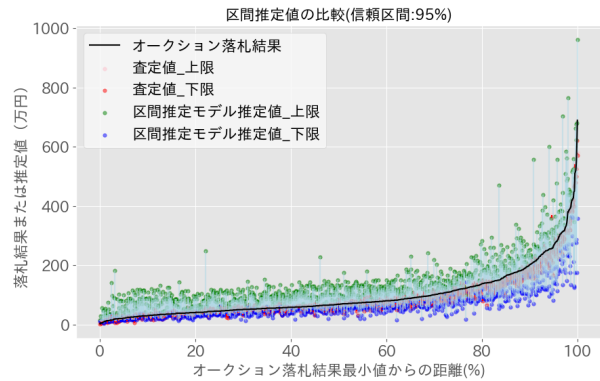


図1 信頼区間95%時の落札結果と各区間推定値の関係

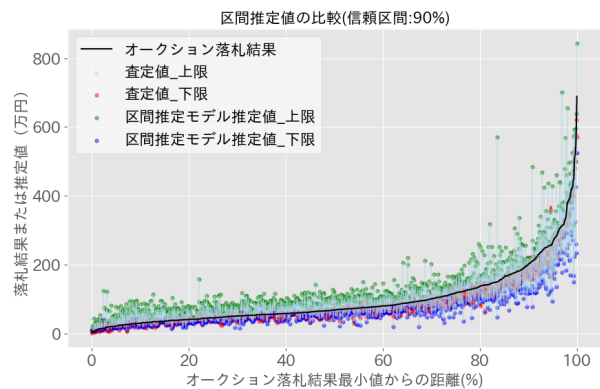


図2 信頼区間90%時の落札結果と各区間推定値の関係

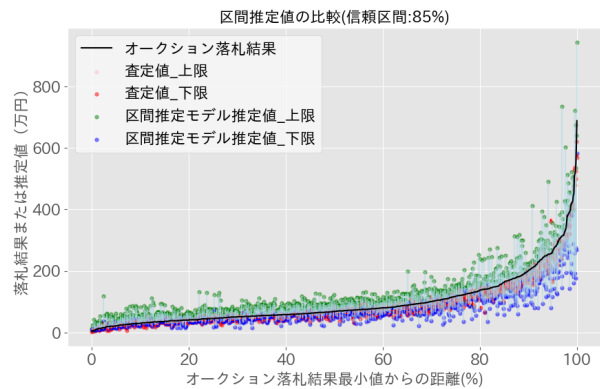


図3 信頼区間85%時の落札結果と各区間推定値の関係

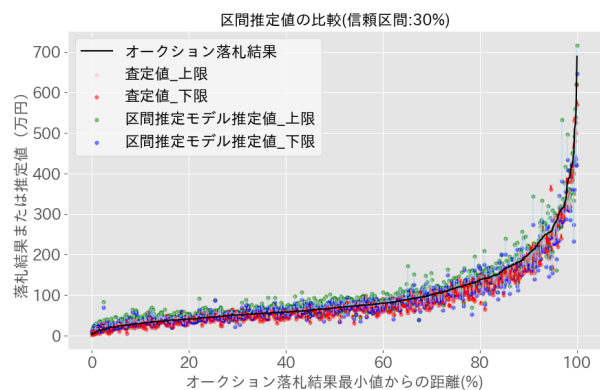


図4 信頼区間30%時の落札結果と各区間推定値の関係

表2 信頼区間ごとの推定区間に包含される確率

信頼区間	現行システムの包含確率	区間推定モデルの包含確率
95	31.5	90.0
90	31.5	83.9
85	31.5	77.6
30	31.5	27.9

6.2 考察

以上の結果を見ると、Winkler Score を評価基準としたとき、信頼区間の値が大きい場合、新規構築した区間推定モデルが現行システムに勝る結果となったが、信頼区間の値が小さい場合、現行システムのほうが勝ることが判明した。

これは、評価基準として扱った Winkler Score が持つ特性によるものと考えられる。Winkler Score は、推定区間が正解値を包含する場合はその推定区間の区間幅を、包含しない場合は区間幅にペナルティが加えられて算出される。ペナルティは信頼区間の大きさによって決定される変数 α によって変動し、一般に信頼区間が大きいほど推定区間の幅が広く、正解値を包含する確率が高いが、包含しない場合に受けるペナルティも大きくなり、Winkler Score が大きく算出される。これは、変数 α は信頼区間が大きいほど小さくなり、Winkler Score 算出時に分母として用いられるためである。例えば信頼区間が 95% のとき、変数 α は 0.05 だが、信頼区間が 30% のとき、変数 α は 0.70 となる。

区間推定モデルは、信頼区間が大きいほどより大きな区間幅で推定するが、現行システムは信頼区間を考慮して推定区間を決定しているわけではなく、査定する中古自動車の価格帯に応じてあらかじめ規定された区間を基準として、査定額となる推定区間を決定している。その区間がかなり狭いものであるため、Winkler Score という評価基準上では悪い結果として表れているのではないかと考えた。また、信頼区間を大きく下げ、30% に設定し検証した場合は、区間推定モデルの推定区間が相応に狭くなり、推定区間に正解値が包含されずペナルティを受ける確率も上がるため、現行システムが優れた結果になったと考えられる。

7 おわりに

本稿では、株式会社オークネットの「中古自動車の査定価格決定支援システム」における、査定価格決定のプ

ロセスに存在する課題を展開したうえで、その改善の取り組みについて、区間推定モデルの構築と評価をテーマに述べた。これまでの研究では、現行システムで用いられている点推定モデルの性能向上をテーマにしてきたが、今研究で初めて区間推定モデルをテーマに扱った。最終的に顧客に返信する、中古自動車の査定額（区間値）をどのように算出するか、というプロセス全体の最適化を目指すうえで、非常に興味深い研究になった。現在使用されている点推定モデルの性能向上とあわせて、区間推定モデルの性能向上にも取り組みを続けていきたい。

また、株式会社オークネットは中古自動車だけでなく、バイクや花卉、電子機器やブランド品といった多様な商材を扱うオークションを運営しており、それぞれの取引情報をデータとして大量に蓄積している。今後はさらに研究領域を広げ、他商材のデータ活用も推進していく。

参考文献

- [1] 段裕之, 林雄太郎, 松下旦, 岩崎敦. 中古自動車の査定価格決定支援システムの実装, 2023. 2023 年度人工知能学会全国大会 (第 37 回) .
- [2] 段裕之, 林雄太郎, 西山佑典, Le Binh Thanh, 松下旦, 岩崎敦. 中古車査定価格支援システムにおける機械学習モデル改善の取り組み, 2023. 2023 年度第情報科学技術フォーラム (第 22 回) .
- [3] 西山佑典, Le Binh Thanh, 段裕之, 林雄太郎, 松下旦, 岩崎敦. 車種グループ別モデル構築による中古自動車価格予測精度の改善, 2024. 2024 年度人工知能学会全国大会 (第 38 回) .
- [4] オークネット. 中古車事業者向け査定サービスをフルリニューアル「サテロク プレミアム」を 8 月 1 日より本格提供開始, 2018. https://www.aucnet.co.jp/nw/20180801_01/.
- [5] オークネット. サテロク プレミアム, 2023. <https://www.aucnet-nyukai.com/sateroku>.
- [6] Robert L. Winkler. A decision-theoretic approach to interval estimation. *Journal of the American Statistical Association*, Vol. 67, No. 337, pp. 187–191, 1972.