

# Causal Treeに基づく選択バイアスを考慮した頑健な条件付き平均処置効果推定手法の提案

坪井 優樹<sup>1,a)</sup> 阪井 優太<sup>1,b)</sup> 鈴木 佐俊<sup>1,c)</sup> 後藤 正幸<sup>1,d)</sup>

受付日 2022年10月11日, 採録日 2023年5月1日

**概要:** 統計的因果推論に基づくビジネス施策の効果検証では、条件付き平均処置効果（以下、CATE）と呼ばれる、処置の有無やアウトカムに関連のある変数が同じ顧客群における処置効果の推定が有用であることが多い。このCATEを推定するための基本的な手法であるCausal Treeは解釈性が高く、観察研究においても適応可能であるため非常に有用である。ただし、Causal Treeを観察研究で用いる際には顧客を人為的に選択することによる系統的な誤差（以下、選択バイアス）を除去する必要がある。傾向スコアを用いた改良手法であるCausal Tree-Transformed Outcome（以下、CT-TO）が提案されている。しかしCT-TOは、傾向スコアを推定するモデルが正しく指定されていない場合には、最終的に誤ったCATEの推定に結び付きやすいことなどの改善点がある。そこで本研究では、Causal Treeをベースとし、高い頑健性を持って選択バイアスが存在する状況に対応したCATE推定手法を提案する。具体的には、Doubly Robust Estimatorという、ある集団全体の処置効果を推定可能な手法を活用したCATEの推定に基づいて、Causal Treeの目的関数の導出を行う。これにより、CT-TOが有する高い解釈性を維持しつつ、選択バイアスが存在する状況において、より精度の高いCATEの推定が可能になる。最後に、人工データと実データに対して提案手法を適用し、その有効性を検証する。

**キーワード:** 統計的因果推論, 条件付き平均処置効果, Causal Tree, 選択バイアス, 効果検証

## A Robust Estimation Method for Conditional Average Treatment Effects Taking Account of Selection Bias Based on Causal Tree

YUKI TSUBOI<sup>1,a)</sup> YUTA SAKAI<sup>1,b)</sup> SATOSHI SUZUKI<sup>1,c)</sup> MASAYUKI GOTO<sup>1,d)</sup>

Received: October 11, 2022, Accepted: May 1, 2023

**Abstract:** In the effectiveness testing of business measures based on statistical causal inference, it is often useful to estimate the conditional average treatment effect (CATE). Causal Tree, which is the basic method for estimating CATE, is highly interpretable and can be adapted to observational studies, making it very useful. However, when Causal Tree is used in observational studies, selection bias needs to be removed, and an improved method using propensity scores, Causal Tree-Transformed Outcome (CT-TO), has been proposed. However, CT-TO has some points to be improved, such as the fact that it tends to lead to incorrect CATE estimation if a model for estimating propensity scores is not correctly specified. Therefore, this study proposes a Causal Tree-based CATE estimation method for situations where selection bias exists with a high degree of robustness. Specifically, the objective function of Causal Tree is derived based on the estimation of CATE using doubly robust estimator, a method that can estimate the treatment effect for an entire population. This enables more accurate CATE estimation in situations where selection bias exists, while maintaining the high interpretability of CT-TO. Finally, the proposed method is applied to artificial and real data to verify its effectiveness.

**Keywords:** causal inference, conditional average treatment effect, causal tree, selection bias, effect verification

<sup>1</sup> 早稲田大学  
Waseda University, Shinjuku, Tokyo 169-8555, Japan

a) yuki-t0814@ruri.waseda.jp

b) y-sakai1003@akane.waseda.jp

c) suzukist@fuji.waseda.jp

d) masagoto@waseda.jp

## 1. はじめに

企業がマーケティング施策などを設計する際、施策の効果検証を適切に行い、正しい意思決定につなげることは、効果の低い施策実施を回避するためにも非常に重要な課題である。一般に、施策対象である個人ごとの施策効果を実験的に評価することは困難であるが、集団に対する平均的な施策効果という意味で、平均処置効果（以下、ATE）や条件付き平均処置効果（以下、CATE）といった指標で効果が測られることが多い[1]。マーケティングの分野においては、顧客の属性に基づいて施策を検討および実施するセグメントマーケティングが行われることが多く、CATEによって、どのような特徴を持つ顧客群に対する施策効果が高いかを把握することは重要である。このような分析によって、効果的なユーザー群を特定し、費用対効果を最大化することができる。また近年では、単に分析者が条件を指定するのではなく、CATE 推定手法を用いて適切に特徴条件の数を大幅に減らすことで、どのような特徴がどのように処置効果に影響を与えているかについての解釈がさかんに行われている[1]。

これらの ATE や CATE の観点から、施策などの処置が結果に与える効果（以下、処置効果）を最も正しく評価する方法としては、ランダム化比較実験（以下、RCT）[2]が望ましい。ビジネス領域では A/B テスト [3] と呼ばれることも多いが、基本的には、ユーザーを施策を実施する群（以下、処置群）と施策を実施しない群（以下、対照群）にランダムに分割し、処置効果を推定することで正しい評価が可能となる。しかし実際には、ビジネスが展開されている中で施策は実施されるものであり、つねに RCT や A/B テストが実施できるとは限らず、過去の施策実施結果のログデータ（観察データ）などから処置効果を推定しなければならないケースがある。また、一般にビジネス施策は「施策効果が高いと推定される顧客」に対して実施されることが多い。このような場合、観察データには、ユーザーを人為的に選択することによる系統的な誤差（以下、選択バイアス）が含まれる可能性があり、この選択バイアスを考慮して処置効果を正しく評価する必要がある。このような枠組みは、統計的因果推論 [4] と呼ばれ、様々な領域で広く研究されている。

一方、CATE を推定するための具体的、かつ基本的な手法として Causal Tree [5] が知られている。この手法は回帰木 [6] のアルゴリズムを用いているため解釈性に優れ、処置効果に影響を与える要因を分析する際に非常に有用であり、観察研究においても適応可能である。ただし、Causal Tree を提案した Athey らは、このモデルを観察研究に用いる際には、処置群と対照群の単純比較による選択バイアスを除去するため、葉内の推定値を修正する必要があるとも言及しており、実際に Causal Tree-Transformed Outcome

（以下、CT-TO）[7] が新たに提案されている。CT-TO は、Causal Tree の処置効果を推定する際に、各ユーザーにおいて処置が実施される確率である傾向スコア [8] を用いて各葉における処置群と対照群内の結果を調整することで、選択バイアスを除去する。ただし、CT-TO で使用する傾向スコアは、処置の有無や結果に関連のある変数（以下、共変量）から処置が実施される確率を予測するモデルが正しく指定されていない場合には誤った結果を与える可能性がある [9]。また、傾向スコアが非常に小さい場合や大きい場合、ユーザーに対する重み付けが極端になるという問題点などがある [10]。

一方、傾向スコアによる処置効果推定手法を拡張した手法として Doubly Robust Estimator [11] が提案されている。Doubly Robust Estimator は、選択バイアスが存在する状況において ATE を推定可能な手法であり、傾向スコアを用いた推定と、共変量と処置の有無を表す変数（以下、処置変数）で結果を説明する回帰分析の 2 つの ATE 推定手法を組み合わせている。そのため、傾向スコアのみを用いた場合よりも頑健性を高めることができることが知られており、近年応用研究で広く利用されている [9]。

そこで本研究では、Causal Tree をベースとした、高い解釈性を維持しつつ、CT-TO より高い頑健性を持って選択バイアスが存在する状況に対応した CATE 推定手法を提案する。具体的には、Doubly Robust Estimator を活用した結果および CATE の算出に基づいて、Causal Tree の目的関数の導出を行う。これにより、高い解釈性を維持しつつも選択バイアスが存在する状況において CATE の適切な推定が可能になる。最後に、人工データを用いた推定精度評価と実データを用いた分析を行い、提案手法の有効性を示す。

## 2. 準備

### 2.1 対象問題

近年、観察データを活用した比較によって施策の効果を求めたいという需要が大きく、効果検証によるデータ駆動型マーケティングへの期待が高まっている。たとえば、マーケティング施策としてクーポン配布を実施するとき、クーポン配布にはコストがかかるため、企業はクーポンを使用してくれるユーザーに対して優先的に効率良く配布を行いたいと考えられる。このように「クーポンの利用率」を効果と考えた場合、最も素直な方法は「クーポン利用率の高いユーザー群を特定し、それらのグループを選択してクーポン配布する」という方法であった。しかし、この方法はビジネスの観点からは適切とはいいきれず、「クーポン配布の処置効果」で施策対象グループを特定すべきであるという考え方が急速に広まっている。すなわち、クーポンを配布しなくても購入するユーザーに対してではなく、クーポンを配布したことによってはじめて購入に至るユーザーに対し

表 1 ルービン因果モデルにおける処置効果  
Table 1 Treatment effects in Rubin causal model.

	処置を実施した場合の結果	処置を実施しなかった場合の結果
処置を実施されたユーザ	観測	欠測
処置を実施されなかったユーザ	欠測	観測

て施策を実施することが望ましいという考え方である。そのため、近年では、与えられたデータから変数間の関係性を明らかにできる統計的因果推論 [4] を用いた施策の効果の推定がさかに行われており、正しい意思決定や改善につなげるために、非常に重要な課題となっている。統計的因果推論には様々なアプローチが存在しており、最も代表的なアプローチとして、Rubin によって提案されたルービン因果モデル [12] がある。ルービン因果モデルとは、実際には観測ができない潜在的結果を仮想的に考えるアプローチである。本研究では、このルービン因果モデルに基づいた統計的因果推論の領域を扱う。

## 2.2 処置効果の定義

因果関係において原因となるものを処置といい、マーケティング上では何らかの施策を指す。N 人の母集団における、i 番目のユーザ ( $i = 1, \dots, N$ ) において、処置の有無を表す変数を処置変数と呼び、 $Z_i \in \{0, 1\}$  とする。ただし、処置を実施する場合を  $Z_i = 1$ 、処置を実施しない場合を  $Z_i = 0$  と表す。また、因果関係において結果を表す変数を結果変数と呼び、 $Y_i(Z_i) \in \mathbb{R}$  で表す。すなわち、処置を実施したときの結果変数は  $Y_i(1)$ 、処置を実施しなかったときの結果変数は  $Y_i(0)$  と記載される。このとき、i 番目のユーザにおける処置効果は式 (1) のように定義する。

$$Y_i(1) - Y_i(0) \quad (1)$$

ルービン因果モデルでは、i 番目のユーザはこれらの値の組  $\{Y_i(1), Y_i(0)\}$  を持つと考え、この値の組を i 番目のユーザの潜在的結果変数という。しかし実際には、表 1 に示すように処置の有無によって、どちらか一方の潜在的結果しか観測できないため、このような i 番目のユーザにおける処置効果を推定することができないという問題（統計的因果推論の根本問題） [13] がある。

### 2.2.1 平均処置効果 (ATE)

ある集団全体に対する処置効果を ATE (平均処置効果) という。ATE は統計的因果推論における代表的な推定対象であり、処置群の結果変数の平均値と対照群の結果変数の平均値の差を ATE として推定することが多い [1]。ATE を式 (2) で表す。なお、式 (2) の等号は、期待値の加法性より、X, Y が確率変数であるとき  $E[X + Y] = E[X] + E[Y]$  であるため成り立つ。

$$ATE = E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)] \quad (2)$$

ただし、ATE の推定をする場合は、交換可能性、正值性、SUTVA (stable unit treatment value assumption) という 3 つの前提条件が成り立つ必要がある [14], [15]。交換可能性とは、 $Z_i$  の値が  $z$  であるすべてのユーザについて、 $Y_i(z) \perp Z_i$  が成り立つことである。つまり、 $Z_i$  の値が 0 か 1 をとるとき、 $E[Y_i(z) | Z_i = 1] = E[Y_i(z) | Z_i = 0]$  が成立する。正值性とは、 $Z_i = 1$  になる確率も  $Z_i = 0$  になる確率も 0 ではないという前提条件である。つまり、 $0 < p(Z_i = 1 | \mathbf{X}_i) < 1$  である。SUTVA とは、i 番目のユーザの潜在的結果変数  $\{Y_i(1), Y_i(0)\}$  は他のユーザの結果に影響することはなく、また、i 番目のユーザに対する処置は一意に決まるという前提条件である。

### 2.2.2 条件付き平均処置効果 (CATE)

異質性を考慮した処置効果として CATE (条件付き平均処置効果) がある。CATE は、共変量と呼ばれる、処置変数や結果変数に関連のある変数を用いて、同じ特徴を持った集団で条件付けたときの ATE を表す [1]。i 番目のユーザにおける K 次元の共変量を  $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{iK})$  とすると、 $\mathbf{X}_i$  の値が  $\mathbf{x}$  である場合の CATE は式 (3) で定義される。

$$CATE = E[Y_i(1) | \mathbf{X}_i = \mathbf{x}] - E[Y_i(0) | \mathbf{X}_i = \mathbf{x}] \quad (3)$$

ただし CATE の推定は、ATE の推定における前提条件に加えて、式 (4) に示す Unconfoundedness という前提条件が成り立つ必要がある。ここで、 $\perp$  は確率変数としての独立性を意味している。

$$\{Y_i(1), Y_i(0)\} \perp Z_i | \mathbf{X}_i \quad (4)$$

## 3. 関連研究

### 3.1 実験研究と観察研究

処置効果を最も正しく評価する方法としては、RCT (A/B テスト) [2], [3] のような実験研究が望ましい。これは、ユーザが処置群と対照群にランダムに分割されていることで、群間の結果の平均値の差を処置の有無による差として扱うことができるためである。しかし、RCT は、コストの大きさや倫理的な観点からなど、様々な理由から実施が難しい場合がある。特に、利益が求められるビジネス活動においては、施策対象者をランダムに選択する RCT は利益を無視した実験であり、つねにこのような実験のみを継続することには無理がある。このように、RCT による厳密な実験の結果ではなく、日々のビジネス活動の中で実施されている施策の結果を用いて、施策効果の評価ができれば、ビジネス現場としては非常に望ましい。

一方、施策対象者がランダムに選択されず、何らかの偏りを持って選択される場合、群間のユーザの特徴の違いに

よって系統的な誤差（選択バイアス）が生じる。通常、過去に蓄積された膨大なデータは平常時のビジネス活動の中で集められており、多くのマーケティング施策は特定のセグメントを対象に実施されている。したがって、RCTを用いた実験的なデータではなく、実運用の中で得られた選択バイアスを含む観察データを用いて処置効果を推定することが多い。この場合、処置効果が高いと見込まれるユーザー群を人為的に選択して施策を打っているため、単純な結果の比較では選択バイアスを含んだ処置効果を推定してしまい、適切な比較が困難になるという問題がある。そのため、ビジネスにおける処置効果の推定を行う際に用いる手法は、選択バイアスに対応可能であることが望ましい。

### 3.2 選択バイアスへの対応

一般に、最も基本的な選択バイアスに対応可能な ATE 推定手法は、回帰分析であるが、選択バイアスの影響を取り除いて ATE を推定するためには、強い前提条件を成立させる必要がある [4]。そのため、回帰分析における問題点に対処するために、各ユーザーにおいて処置が実施される確率である傾向スコア [8] を用いて、処置群と対照群のデータの性質を近くすることで ATE を推定する方法が広く活用されている。Robins ら [16] は傾向スコアをユーザーの結果変数  $Y_i(Z_i)$  の重み付けとして用いて ATE を推定する Inverse Probability Weighted Estimator（以下、IPW 推定量）を提案した。これにより、処置群と対照群の間の偏りをなくすような ATE の推定を可能にした。ただし、傾向スコアは、共変量から処置が実施される確率を予測するモデルが正しく指定されていない場合には誤った結果を与える可能性があり [9]、傾向スコアが非常に小さい場合や大きい場合、ユーザーに対する重み付けが極端になるなどの問題点がある [10]。そのため、Bang ら [11] は回帰分析と IPW 推定量を組み合わせる ATE の推定を行う Doubly Robust Estimator を提案し、IPW 推定量のみを用いる場合よりも頑健に ATE を推定することを可能にした。

### 3.3 CATE 推定手法

近年では、マーケティング施策としてセグメントマーケティングが行われることが多いため、ATE のような集団全体ではなく、ユーザーが持つ特徴による異質性を考慮した CATE の推定がさかんに行われている。CATE の推定値の大きさを比較することにより、どのような特徴を持つ集団に対して処置を実施することが有効であるかについての分析が可能となり、ターゲティング最適化につなげることができる [17]。

Athey らは本研究の従来手法である、回帰木を用いて CATE を推定する Causal Tree [5] を提案した。この手法は通常の回帰木のように結果変数  $Y_i(Z_i)$  を予測するのではなく、直接 CATE を予測する。また、この手法は回帰木に

基づいているため解釈性に優れ、処置効果に影響を与える要因を分析する際に非常に有用である。たとえば、Knittel ら [18] は、エネルギー事業者が個別の家庭にエネルギー使用状況などの情報提供を行うことによる、節電行動への効果検証を行った。また Rana ら [19] は、ヒマラヤ山脈における森林管理政策の実施による植生成長への効果を検証している。Causal Tree を提案した Athey らは後に、選択バイアスの除去のために Causal Tree の CATE 推定時に傾向スコアを用いる Causal Tree-Transformed Outcome (CT-TO) や、Random Forest [20] に基づき Causal Tree を拡張した Causal Forest、一般化モーメント法 [21] を活用して Causal Forest [22] を拡張した Generalized Random Forests（以下、GRF）[23] を提案している。さらに、Oprescu らは、GRF にネイマン直交条件 [24] を組み合わせた Orthogonal Random Forest（以下、ORF）[25] を提案しており、GRF に比べてパラメータ推定量の正則化バイアスを下げている。また、Powers らは、Causal Tree のアルゴリズムに基づいたブースティング決定木である Causal Boosting [26] を提案している。このように、近年では Causal Tree を拡張した手法の提案が行われている。しかし、Causal Tree を拡張した手法は主に木を複数用いているため、回帰木や Causal Tree が持つ単一の木構造による結果の可視化が可能という利点を維持できておらず、解釈性を下げている。

Causal Tree 以外の木構造に基づく CATE 推定手法として、Uplift Tree [27] が提案されている。この手法は、CATE の増加に関連が強い属性ごとにユーザー群を分割する。具体的には、分割後のユーザー群における処置群の結果変数の分布と対照群の結果変数の分布の距離が、分割前に比べて増加するように集団を分割する。Uplift Tree は Causal Tree と比較して、CATE を最大化するように木を分割する観点とは同じであるが、目的関数に群内の結果変数の分散に関する項がない点や学習データを分割しない点で異なる。また、複数の回帰木の和によって予測結果を算出する Bayesian Additive Regression Trees（以下、BART）[28] を活用した CATE 推定手法 [29], [30] も提案されている。この手法の利点としては、事後分布を得られることがあげられる。ただし、加法的回帰木であるため、単一の木を樹形図で可視化して処置効果に影響を与える要因についての分析を行うことは困難である。

さらにほかにも、CATE を直接推定する手法ではないが、説明可能性を考慮可能な手法として、TOM [7], [26] や SDRM [31] があげられる。これらの手法は、処置変数と傾向スコアを活用し、結果変数を CATE として変換することができる。具体的には、TOM は処置変数と傾向スコア、SDRM は処置変数と傾向スコアおよび回帰分析 (Doubly Robust Estimator) を利用して変換する。TOM や SDRM を使用する際、変換後は任意の回帰モデルで変換された CATE を予測するだけでよいため、容易に使用

できる点が利点としてあげられる。ゆえに、仮に回帰木で変換された CATE の推定を行えば、Causal Tree のように樹形図で可視化することができる。ただし、これらの手法は、結果変数から CATE へ変換する際のみ処置変数を用いるため、一般に効率的でないと言及されている [5]。また、Künzel らは Meta-learner [32] という機械学習のアルゴリズムを内部的に用いることで CATE を推定する計算フレームワークを提案した。Meta-learner は処置群と対照群のそれぞれで任意の回帰モデルを用いた推定モデルを作成し、それぞれの推定値を出した後、その差分を CATE とする。この手法は計算フレームワークであるため、その中には様々な種類のアルゴリズムが存在し、T-learner [32]、DR-learner [33]、X-learner [32] などがある。T-learner は処置群と対照群のそれぞれで結果変数  $Y_i(Z_i)$  の予測モデルを構築し、それぞれのモデルで推定した結果変数  $Y_i(Z_i)$  の差分を CATE とする手法である。また、DR-learner は T-learner に Doubly Robust Estimator を活用した手法であり、X-learner は T-learner の 2 つの回帰モデルの重み付けに傾向スコアを用いた手法である。Meta-learner も TOM および SDRM と同様に任意の回帰モデルを用いるが、基本的に 2 つの回帰モデルを用いるため、回帰木を用いても樹形図で可視化することができない点で異なっている。

さらに近年では、ニューラルネットワークを用いた手法の提案がさかんに行われており、Johansson らは Counterfactual Regression [34] を提案し、処置群と対称群が判別できないような特徴表現を得たうえで潜在的結果変数  $\{Y_i(1), Y_i(0)\}$  を予測することで、処置効果の汎化誤差の上界の最小化を行った。また、Louizos らは変分オートエンコーダ [35] を応用した CEVAE [36] を提案し、Yoon らは GAN [37] を応用した GANITE [38] を提案するなど、様々な機械学習アルゴリズムを統計的因果推論の領域に適用する動きが注目を集めている。

いずれにおいても、近年提案されている CATE 推定手法の主な目的はあくまで選択バイアスへの対応や予測精度の追求であり、説明可能な CATE 推定手法に主眼が置かれた研究はあまり行われていない。しかしながらマーケティング施策の効果検証においては、処置効果を推定することはもちろん、どのような特徴がどのように処置効果に影響を与えているのかに関する解釈性についても、原因を特定し、マーケティング施策をより効果的にさせるうえで非常に重要な要素である。したがって、高い予測精度だけでなく、高い解釈性を持つ CATE 推定手法もビジネスにおいて非常に有用であるといえる。

## 4. 従来手法 (Causal Tree)

### 4.1 Causal Tree

Causal Tree とは、回帰木のアルゴリズムをベースとし

た CATE 推定手法であり、解釈性が高く、処置効果に影響を与える要因についての分析に有用である。この手法は、通常回帰木と比較して 2 つの点が異なる。1 つ目は、結果変数  $Y_i(Z_i)$  を推定するモデルではなく、CATE を推定するモデルを作ることである。通常回帰木のように結果変数の二乗和誤差を最小化するように分岐を選ぶのではなく、CATE の二乗和の平均値を最大化しつつ、各群の結果変数の分散を小さくするように分割する。2 つ目は、推定に用いる学習データの使用方法が異なることである。具体的には、学習データを 2 分割し、1 つは木の分割に利用し、もう 1 つはその木構造の下でそれぞれの葉の処置効果を推定することに用いる。

以降は、式 (4) に示した Unconfoundedness を前提として、定式化を行う。まず、Causal Tree における CATE を定義する。共変量空間  $\mathbf{X}$  を分割した領域、すなわち木を  $\Pi$  とする。このとき、木  $\Pi$  は式 (5) で定義される。

$$\Pi = \{l_1, \dots, l_{\#(\Pi)}\}, \text{ with } \bigcup_{j=1}^{\#(\Pi)} l_j = \mathbf{X}. \quad (5)$$

ただし、 $l_j$  は木  $\Pi$  を構成する葉ノード、 $\#(\Pi)$  は共変量空間  $\mathbf{X}$  を分割した数を表す。枝は、分岐後の各葉ノードに所属するサンプルにおける目的関数が小さくなるように分岐される。ここで、 $l(\mathbf{x}; \Pi)$  を  $\mathbf{x} \in l$  である葉ノード  $l \in \Pi$  とする。このとき、Causal Tree における CATE は式 (6) で定義される。

$$\begin{aligned} \tau(\mathbf{X}_i; \Pi) &\equiv E[Y_i(1) \mid \mathbf{X}_i \in l(\mathbf{x}; \Pi)] \\ &\quad - E[Y_i(0) \mid \mathbf{X}_i \in l(\mathbf{x}; \Pi)] \end{aligned} \quad (6)$$

これは、分割された葉ノード内における処置群と対照群の結果変数の平均値の差を表す。Causal Tree は回帰木のアルゴリズムをベースとしているため、与えられたデータに対して、段階的に条件を設けて、枝を分岐する。つまり、同じ葉ノード内に落ちたサンプルは共変量の値が同じであると考えられる。したがって、ATE を処置群と対照群の結果変数の平均値の差として推定すると、CATE を推定することは各葉ノード内に対して ATE を推定することと同等であると見なすことができる。

また、Causal Tree では、与えられた学習データをランダムに 2 分割し、木を作るための訓練データ  $S^{tr}$  と作られた木を用いて推定値を出すための推定データ  $S^{est}$  に分けて学習を行う (図 1, 図 2)。このデータの分割によって過学習を防ぐことができる。加えて、予測した  $\tau(\mathbf{X}_i; \Pi)$  は一貫性や漸近正規性を満たすことが知られている。

さらに、Causal Tree は結果変数  $Y_i(Z_i)$  を推定するのではなく、CATE を推定するため、通常回帰木の目的関数から変更されている。ここで、テストデータを  $S^{te}$ 、 $i$  番目のユーザにおける真の処置効果を  $\tau_i$  とする。このとき Causal Tree の目的関数は、平均二乗誤差を用いて式 (7) で定義できる。しかし、ここには観測不可能な  $\tau_i$  が含まれて

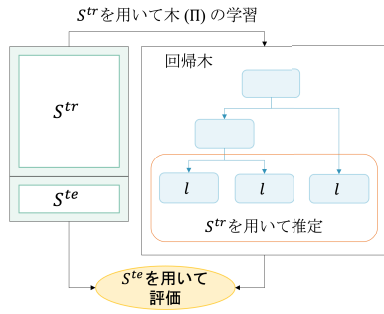


図 1 回帰木の学習データの利用方法

Fig. 1 How to use training data for regression trees.

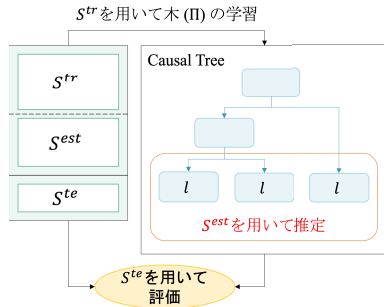


図 2 Causal Tree の学習データの利用方法

Fig. 2 How to use training data for Causal Tree.

おり、そのまま使用することができない。なお、最後の項  $\tau_i^2$  は後の式変形の都合上導入している。

$$\begin{aligned} &MSE_{\tau}(S^{te}, S^{est}, \Pi) \\ &\equiv \frac{1}{N^{ste}} \sum_{i \in S^{te}} \{(\tau_i - \hat{\tau}(\mathbf{X}_i; S^{est}, \Pi))^2 - \tau_i^2\} \end{aligned} \quad (7)$$

そこで Athey らは、式 (8) で定義される MSE の  $S^{te}$  と  $S^{est}$  に対する期待値である  $EMSE_{\tau}(\Pi)$  を用いた式変形を行うことによって、最終的に Causal Tree の目的関数を式 (9) で定義することを可能にしている。ただし、詳細な式変形はスペースの都合で割愛する。

$$EMSE_{\tau}(\Pi) \equiv E_{S^{te}, S^{est}} [MSE_{\tau}(S^{te}, S^{est}, \Pi)] \quad (8)$$

$$\begin{aligned} &\widehat{EMSE}_{\tau}(S^{tr}, N^{est}, \Pi) \equiv -\frac{1}{N^{tr}} \sum_{i \in S^{tr}} \hat{\tau}^2(\mathbf{X}_i; S^{tr}, \Pi) \\ &+ \left( \frac{1}{N^{tr}} + \frac{1}{N^{est}} \right) \cdot \sum_{l \in \Pi} \left( \frac{V_{S_1^{tr}}(l)}{p} + \frac{V_{S_0^{tr}}(l)}{1-p} \right) \end{aligned} \quad (9)$$

ここで、訓練データ数を  $N^{tr}$ 、推定データ数を  $N^{est}$ 、 $Z_i = 1$  における  $S^{tr}$  を  $S_1^{tr}$ 、 $Z_i = 0$  における  $S^{tr}$  を  $S_0^{tr}$ 、 $V_{S_1^{tr}}(l)$  は葉ノード  $l$  に該当する  $S_1^{tr}$  の  $Y_i(Z_i)$  の分散、 $V_{S_0^{tr}}(l)$  は葉ノード  $l$  に該当する  $S_0^{tr}$  の  $Y_i(Z_i)$  の分散、 $p$  は  $S^{tr}$  における  $S_1^{tr}$  の割合を表す。この目的関数の最小化は、各葉ノードにおける、CATE の二乗和の平均値を最大化しつつ、分散は小さくすることを意味する。つまり、各群の結果変数  $Y_i(Z_i)$  の分散がなるべく小さくなることを目指しつつ、各葉ノードにおける処置効果はできるだけ大きくすることになる。

## 4.2 Causal Tree-Transformed Outcome (CT-TO)

CT-TO とは、Causal Tree を提案した Athey らが、観察研究に用いる際に処置群と対照群の単純比較による選択バイアスを除去するために提案した手法である。Causal Tree における CATE  $\tau(\mathbf{X}_i; \Pi)$  では、各葉ノードにおける群間の結果変数の平均値の差を CATE と定義している。そのため、群間の結果変数の平均値の差を単純に比較すると、選択バイアスが含まれるために CATE を適切に推定できない可能性がある。そして、その場合は分岐途中のノードにおいて、誤った CATE に基づいた枝の分割が行われてしまう。そこで CT-TO は、Causal Tree における葉内の処置効果を推定する際に、IPW 推定量に基づき傾向スコアの逆数を用いて重み付けを行うことで、選択バイアスの除去を可能にしている。

ここでまず、式 (10) に傾向スコアを表す。なお、傾向スコアでは、式 (11) で定義される強く無視できる割当て条件 [39] と、SUTVA が成り立つ必要がある。

$$e(\mathbf{X}_i) = p(Z_i = 1 | \mathbf{X}_i) \quad (10)$$

$$\{Y_i(1), Y_i(0)\} \perp Z_i | \mathbf{X}_i, 0 < p(Z_i = 1 | \mathbf{X}_i) < 1 \quad (11)$$

また、実際には各ユーザにおける真の傾向スコアは分からないため、 $\mathbf{X}_i$  を説明変数、 $Z_i$  を結果変数とした任意の回帰モデルから事前に推定する必要がある。回帰モデルとしては、ロジスティック回帰モデルやプロビット回帰モデルが使用されることが多い。

葉ノード  $l$  内のサンプルを  $S^l$  とすると、上記の傾向スコアを用いて、CT-TO における CATE は式 (12) で定義される。

$$\begin{aligned} \tau_{TO}(\mathbf{X}_i, e(\mathbf{X}_i); \Pi) &\equiv \frac{\sum_{i \in S^l: \mathbf{X}_i \in l(\mathbf{x}; \Pi)} Y_i(Z_i) \cdot Z_i / e(\mathbf{X}_i)}{\sum_{i \in S^l: \mathbf{X}_i \in l(\mathbf{x}; \Pi)} Z_i / e(\mathbf{X}_i)} \\ &- \frac{\sum_{i \in S^l: \mathbf{X}_i \in l(\mathbf{x}; \Pi)} Y_i(Z_i) \cdot (1 - Z_i) / (1 - e(\mathbf{X}_i))}{\sum_{i \in S^l: \mathbf{X}_i \in l(\mathbf{x}; \Pi)} (1 - Z_i) / (1 - e(\mathbf{X}_i))} \end{aligned} \quad (12)$$

## 5. 提案手法

### 5.1 概要

Causal Tree は解釈性に優れ、処置効果に影響を与える要因を分析する際に非常に有用である。また、CT-TO は、CATE の推定時に傾向スコアを用いて推定値を修正することで、選択バイアスへの対応を行っている。しかし、CT-TO で使用する傾向スコアは、共変量から処置が実施される確率を予測するモデルが正しく指定されていない場合には誤った結果を与える可能性がある。また、傾向スコアが非常に小さい場合や大きい場合、ユーザに対する重み付けが極端になってしまうという問題点がある。

そこで提案手法では、Doubly Robust Estimator を活用し、CT-TO よりも頑健な CATE 推定手法へと改良する。ゆえに、より適切に葉内の群間の結果変数の平均値の差を推定し、選択バイアスの影響を除去して木を分割するこ

とができるため、結果的に適切な CATE の推定が可能になる。

### 5.1.1 Doubly Robust Estimator

Doubly Robust Estimator は、選択バイアスが存在する状況において ATE を適切に推定可能な手法であり、ATE 推定時に共変量を考慮して選択バイアスを除去できる。また、Doubly Robust Estimator では  $\mathbf{X}_i$  と  $Z_i$  で  $Y_i(Z_i)$  を説明する回帰分析と、IPW 推定量を組み合わせて ATE の推定を行う。Doubly Robust Estimator では、回帰分析と IPW 推定量で用いる傾向スコアを推定する回帰モデルのどちらか一方が正しく指定されていれば、一致推定量が得られ、処置効果を適切に推定できる [9]。そのため、傾向スコアのみを用いる場合よりも頑健性を高めることができる。Doubly Robust Estimator における ATE は、式 (10) の傾向スコアを用いて式 (13) で表される。

$$ATE_{DR} = E \left[ \frac{Y_i(Z_i) \cdot Z_i}{e(\mathbf{X}_i)} - \frac{\hat{Y}_i(1) \cdot (Z_i - e(\mathbf{X}_i))}{e(\mathbf{X}_i)} \right] - E \left[ \frac{Y_i(Z_i) \cdot (1 - Z_i)}{(1 - e(\mathbf{X}_i))} + \frac{\hat{Y}_i(0) \cdot (Z_i - e(\mathbf{X}_i))}{(1 - e(\mathbf{X}_i))} \right] \quad (13)$$

ただし、傾向スコア  $e(\mathbf{X}_i)$  は  $\mathbf{X}_i$  を説明変数、 $Z_i$  を結果変数とした任意の回帰モデルから事前に推定する。また、回帰分析により事前に推定された、 $Z_i = 1$  における  $\hat{Y}_i(Z_i)$  を  $\hat{Y}_i(1)$ 、 $Z_i = 0$  における  $\hat{Y}_i(Z_i)$  を  $\hat{Y}_i(0)$  とする。

## 5.2 Doubly Robust Estimator を活用した Causal Tree の定式化

### 5.2.1 Doubly Robust Estimator を用いた CATE

Doubly Robust Estimator を用いた Causal Tree における CATE を式 (14) で表す。

$$\tau_{DR}(\mathbf{X}_i, e(\mathbf{X}_i), \hat{Y}_i(0), \hat{Y}_i(1); \Pi) \equiv E \left[ \frac{Y_i(Z_i) \cdot Z_i}{e(\mathbf{X}_i)} - \frac{\hat{Y}_i(1) \cdot (Z_i - e(\mathbf{X}_i))}{e(\mathbf{X}_i)} \middle| \mathbf{X}_i \in l(\mathbf{x}; \Pi) \right] - E \left[ \frac{Y_i(Z_i) \cdot (1 - Z_i)}{(1 - e(\mathbf{X}_i))} + \frac{\hat{Y}_i(0) \cdot (Z_i - e(\mathbf{X}_i))}{(1 - e(\mathbf{X}_i))} \middle| \mathbf{X}_i \in l(\mathbf{x}; \Pi) \right] \quad (14)$$

ただし、Doubly Robust Estimator で用いる  $e(\mathbf{X}_i)$ 、 $\hat{Y}_i(0)$ 、 $\hat{Y}_i(1)$  は事前に推定する。この式 (14) は、葉ノード  $l$  内の結果変数  $Y_i(Z_i)$  を Doubly Robust Estimator によって調整後、群間の平均値の差を計算している。CT-TO の CATE ような傾向スコアの逆数による重み付けで推定値を修正する際、傾向スコア推定には対照群の共変量の情報を利用している。しかし、結果変数の周辺分布の母数推定の際には対照群の共変量の情報を利用していない [9]。一方 Doubly Robust Estimator では、単に傾向スコアで重み付けするのではなく、推定方程式に処置群の共変量のデータを用いた項  $\hat{Y}_i(0)$  を追加することで、データ利用の効率を向上させ

ている。これにより、分布についての仮定をしないセミパラメトリックな推定量の中では最も推定量の分散が小さくなる推定量である局所有効なセミパラメトリック推定量を構成できる [40]。したがって、選択バイアスの影響を考慮しつつ、頑健な推定が可能になる。

### 5.2.2 Doubly Robust Estimator を用いた目的関数

Doubly Robust Estimator を用いた Causal Tree における CATE を用いた目的関数の導出を行う。Doubly Robust Estimator による  $V_{S_1^{tr}}(l)$ 、 $V_{S_0^{tr}}(l)$  をそれぞれ式 (15)、(16) としたもとの、提案手法の目的関数を式 (17) と定義する。

$$V_{S_1^{tr}}^{DR}(l) = V \left[ \frac{Y_i(Z_i) \cdot Z_i}{e(\mathbf{X}_i)} - \frac{\hat{Y}_i(1) \cdot (Z_i - e(\mathbf{X}_i))}{e(\mathbf{X}_i)} \right] \quad (15)$$

$$V_{S_0^{tr}}^{DR}(l) = V \left[ \frac{Y_i(Z_i) \cdot (1 - Z_i)}{(1 - e(\mathbf{X}_i))} + \frac{\hat{Y}_i(0) \cdot (Z_i - e(\mathbf{X}_i))}{(1 - e(\mathbf{X}_i))} \right] \quad (16)$$

$$\begin{aligned} & \widehat{EMSE}_{\tau_{DR}}(S^{tr}, N^{est}, \Pi) \\ & \equiv -\frac{1}{N^{tr}} \sum_{i \in S^{tr}} \hat{\tau}_{DR}^2(\mathbf{X}_i, e(\mathbf{X}_i), \hat{Y}_i(0), \hat{Y}_i(1); \Pi) \\ & + \left( \frac{1}{N^{tr}} + \frac{1}{N^{est}} \right) \cdot \sum_{l \in \Pi} \left( \frac{V_{S_1^{tr}}^{DR}(l)}{p} + \frac{V_{S_0^{tr}}^{DR}(l)}{1-p} \right) \quad (17) \end{aligned}$$

式 (17) の提案手法の目的関数の最小化は、各葉ノードにおける Doubly Robust Estimator を用いて推定された CATE の二乗和の平均値を最大化しつつ、分散は小さくすることを意味する。つまり、提案手法の目的関数は選択バイアスを考慮した頑健性を持つ目的関数であり、選択バイアスの影響を除去した CATE に基づいた枝の分割が期待される。

## 6. 人工データを用いた推定精度評価

提案手法の有効性を示すため、推定精度の評価を行う。統計的因果推論の根本問題より、個人レベルの処置効果を推定することができないため、実データセットを用いて正確な精度評価を行うことは不可能である。そこで、まずは個人レベルの処置効果の正解データを持つことができる人工データセットを用いて評価を行う。

### 6.1 実験に用いる人工データセット

本実験で用いる人工データセットは、従来手法 (Causal Tree) が提案された論文 [5] で用いられた人工データセットと Nie ら [41] が用いた人工データセットをもとに作成し、20 次元の共変量  $\mathbf{X}_i$  が存在する。

学習時のデータには、 $i$  番目のユーザにおいて、 $\mathbf{X}_i$ 、 $Y_i(Z_i)$ 、 $Z_i$  が与えられ、テストデータでは、 $\mathbf{X}_i$ 、 $Y_i(Z_i)$  が与えられる。結果変数  $Y_i(Z_i)$  は 8 個の共変量からなる式 (18) で表される関数  $b(\mathbf{X}_i)$  と 4 個の共変量からなる式 (19) で表される関数  $\tau(\mathbf{X}_i)$  を用いて式 (20) に従い生成した。



$$b(\mathbf{X}_i) = \frac{1}{2} \sum_{k=1}^4 X_{ik} + \sum_{k=5}^8 X_{ik} \quad (18)$$

$$\tau(\mathbf{X}_i) = \sum_{k=1}^4 1\{X_{ik} > 0\} \cdot X_{ik} \quad (19)$$

$$Y_i(Z_i) = b(\mathbf{X}_i) + \frac{1}{2} \cdot (2 \cdot Z_i - 1) \cdot \tau(\mathbf{X}_i) + \sigma \cdot \epsilon_i \quad (20)$$

ただし、 $b(\mathbf{X}_i)$  は処置の有無  $Z_i$  に無関係な項であり、 $\tau(\mathbf{X}_i)$  は処置効果を表す。したがって、目的変数には、20 個の共変量のうち 8 個の共変量が影響を及ぼし、残りの 12 個は影響を及ぼさない。また、20 次元の共変量  $\mathbf{X}_i$  は各要素が独立な標準正規分布に従うものとし、ノイズ項  $\epsilon_i$  も標準正規分布に従うものとした。ただし、 $\epsilon_i$  の係数  $\sigma$  によりノイズ項の大きさを変更し、本実験では  $\sigma = 1, 4, 8$  とした。さらに、選択バイアスを存在させるため、式 (19) の  $\tau(\mathbf{X}_i)$  を構築する 4 個の共変量のうち、2 個の共変量からなる式 (21) で表される関数  $e(\mathbf{X}_i)$  で得られるパラメータを持つベルヌーイ分布に従って処置の有無  $Z_i$  を生成する。

$$e(\mathbf{X}_i) = \frac{1}{1 + e^{-X_{i1}} + e^{-X_{i2}}} \quad (21)$$

このとき、処置群と対照群の結果変数の平均値に差があるか否か調べるために、2 母集団の平均値の差に関する対応のない両側 t 検定を行った。その際、帰無仮説は「処置群と対照群の結果変数の平均値は等しい」、対立仮説は「処置群と対照群の結果変数の平均値には差がある」となり、p 値には両側検定における p 値を採用した。すなわち、 $F_T(x)$  を t 分布の分布関数、 $t$  を式 (22) に示す得られた統計量とすると、両側検定における p 値は  $p \text{ 値} = 2F_T(t)$  で与えられる。ただしここでは、 $\bar{x}_1$  は 1 群目の標本平均、 $n_1$  は 1 群目のサンプルサイズ、 $\bar{x}_2$  は 2 群目の標本平均、 $n_2$  は 2 群目のサンプルサイズ、 $s^2$  は不偏分散を表す。

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2(\frac{1}{n_1} + \frac{1}{n_2})}} \quad (22)$$

その結果、両側検定における p 値は、 $\sigma = 1$  のとき  $3.430 \times 10^{-20}$ 、 $\sigma = 4$  のとき  $1.448 \times 10^{-7}$ 、 $\sigma = 8$  のとき  $2.123 \times 10^{-3}$  であった。したがって、 $\sigma = 1, 4, 8$  のすべての設定で、有意水準 1% の両側検定において、帰無仮説は棄却され、対立仮説が採択された\*1。

## 6.2 実験条件

本実験では、学習データ数は 1000 件、テストデータ数は 250 件とした。比較手法には、従来手法である Causal Tree、CT-TO に加えて、TOM、SDRM、T-learner、DR-learner、X-learner、Uplift Tree、ORF を用いる。本実験では、傾向

\*1 式 (20) の第 2 項は非正規項であるため、厳密には t 検定は妥当な設定ではない。しかし、サンプルサイズ  $n_1$ 、 $n_2$  が十分大きければ漸近的に Z 検定または t 検定を行うことは妥当であるため、得られた p 値は漸近的には信用できる値であると考えられる。

表 2 各手法の RMSE

Table 2 RMSE of the proposed and comparison methods.

	$\sigma=1$	$\sigma=4$	$\sigma=8$
TOM	4.596	7.059	12.788
SDRM	1.107	1.707	2.136
T-learner	1.460	3.034	5.563
DR-learner	0.901	1.785	3.339
X-learner	<b>0.803</b>	2.020	3.864
Uplift Tree	1.473	1.842	2.934
ORF	1.191	1.829	3.131
Causal Tree	1.660	2.153	3.795
CT-TO	1.638	2.168	4.018
提案手法	0.996	<b>1.340</b>	<b>1.713</b>

スコアはロジスティック回帰により推定し、Doubly Robust Estimator の回帰分析と、Meta-learner の回帰モデルには XGBoost [42] を使用した。また、提案手法と同様に、木構造での可視化を目的とする際に適用候補となることを想定し、TOM および SDRM による結果変数の変換後の値を予測する回帰モデルは、回帰木とした。Causal Tree と CT-TO と提案手法の木の深さの最大値は、テストデータで RMSE が小さくなるように探索的に決定し、11 とした。同様に、Uplift Tree は 7、TOM および SDRM で用いる回帰木では 8 とした。また、ORF と XGBoost の木の数はそれぞれ 500 とした。評価指標は、各条件に対して 10 回試行した処置効果を表す  $\tau(\mathbf{X}_i)$  の Root Mean Squared Error (RMSE) の平均値とする。なお、今回の実験設定は、提案手法も Causal Tree および CT-TO も現実的には正しく機能しない設定になっている。理由は、処置の有無  $Z_i$  は関数  $e(\mathbf{X}_i)$  で得られるパラメータを持つベルヌーイ分布に従って生成していることから、実際には 1000 件のデータで完璧にモデル化することは難しいと考えられるためである。すなわち、傾向スコアの話については、どちらも完璧にモデリングすることは難しい課題であると思われる。ただし、提案手法は二重に頑健な推定により、Causal Tree および CT-TO よりは正しく機能し、より高い推定精度が期待される。

## 6.3 実験結果

各手法の RMSE を表 2 に示す。まず、すべての条件下で、提案手法は従来手法である Causal Tree や CT-TO よりも高い精度を示している。また、TOM と SDRM、T-learner と DR-learner をそれぞれ比較すると、Doubly Robust Estimator を活用している SDRM、DR-learner の方がすべての条件下で精度が高いことが分かる。したがって、提案手法は Doubly Robust Estimator の活用によって従来手法と比較して高い精度を持ち、選択バイアスが存在する状況において非常に有効な手法であるといえる。ここで、Causal Tree と CT-TO を比較すると、 $\sigma = 1$  のときは



CT-TOの方が推定精度が高く、 $\sigma = 4, 8$ のときは Causal Treeの方が高かった。この理由として、CT-TOは、大きなノイズを含む結果変数に対して傾向スコアで重み付けを行った後にCATEを算出したため、かえって推定精度が悪化してしまったと考えられる。このような点からも、傾向スコアだけでなく、共変量と結果変数でモデリングするDoubly Robust Estimatorを活用している提案手法は、二重に頑健であるため有用であるといえる。

また、提案手法と、TOMとSDRMとUplift Treeを比較すると、提案手法がすべての条件下で最も精度が高いことが分かる。ゆえに、観察研究において木構造での可視化を目的とする際には、提案手法の適用がふさわしいといえる。

最後に、提案手法と、DR-learnerおよびX-learnerと比較すると、提案手法は $\sigma = 4, 8$ のときは推定精度が高かったが、 $\sigma = 1$ のときは低かった。この理由として、比較的ノイズ項 $\sigma$ による影響が小さい状況においては、回帰木をベースとする提案手法よりも、XGBoostを用いているDR-learnerやX-learnerの方が高い表現力を駆使して推定を行えることが考えられる。一方で、ノイズ項 $\sigma$ が大きい状況においては、2つの回帰モデルから推定値を算出するDR-learnerとX-learnerは、提案手法と比較して過学習しやすいと考えられる。実際のビジネスにおいて、効果が見込まれるユーザー群を人為的に選択して施策を打つような場合、選択バイアスが存在することに加え、データのばらつきも存在することが多いと考えられる。さらに、学習データとして、真の施策効果は与えられないため、ばらつきを大きく予想することは比較的難しいと予想される。したがって、一般に効果が見込まれるユーザー群を人為的に選択して施策を打つ際、データのばらつきが小さいと想定

できない場合には、X-learnerやDR-learnerではなく提案手法を選択することも考えられる。加えて、X-learnerやDR-learnerよりも推定精度が低くなると思われる場合でも、分析者の目的次第では木構造で可視化できる提案手法を用いた方が有効であると考えられる。

## 7. 実データを用いた分析

提案手法の有用性を示すため、The MineThatData E-Mail Analytics And Data Mining Challenge (以下、MineThatData) [43] というデータを用いて、通信販売会社のユーザに対して、男性向けメールを配信したことによる、購入額への効果について分析を行う。

### 7.1 分析条件

本実験は、男性向けメールの配信という処置に対して、メール配信後2週間以内に購入した際の購入額（以下、spend）を結果とする。そして、あるユーザ群における処置群と対照群の結果の平均値の差、すなわち処置による購入額の増減をCATEとして分析を行う。ここで、このデータセットにおける変数とその説明を表3に示す。ただし、mens, womens, newbie, history\_segment 0-100, history\_segment 100-200, history\_segment 200-350, history\_segment 350-500, history\_segment 500-750, history\_segment 750-1000, history\_segment 1000+ はダミー変数を表し、それぞれ条件にあてはまる場合は1、条件にあてはまらない場合は0であるとする。MineThatDataはRCTによって得られたデータである。そこで、本実験では、購買傾向が一定以上あるユーザのみに施策を実施した状況を想定し、元々のデータに対して選択バイアスを加えたうえで、分析を行うことで提案手法の有効性を示す。具体的には、最後の購

表 3 データセットにおける変数とその説明

Table 3 Variables and their descriptions.

変数の種類	変数名	説明
共変量	recency	最後の購入からの経過月数
	history	昨年の購入額
	mens	昨年の男物商品の購入有無
	womens	昨年の女物商品の購入有無
	newbie	過去12カ月以内に新しくユーザになったか
	history_segment 0-100	昨年の購入額が0ドル以上100ドル未満であるか
	history_segment 100-200	昨年の購入額が100ドル以上200ドル未満であるか
	history_segment 200-350	昨年の購入額が200ドル以上350ドル未満であるか
	history_segment 350-500	昨年の購入額が350ドル以上500ドル未満であるか
	history_segment 500-750	昨年の購入額が500ドル以上750ドル未満であるか
	history_segment 750-1000	昨年の購入額が750ドル以上1000ドル未満であるか
	history_segment 1000+	昨年の購入額が1000ドル以上であるか
	zipcode	zipcodeをもとに地区を分類したもの (Rural, Suburban, Urban)
channel	昨年どのチャンネルから購入したか (Phone, Web, Multichannel)	
結果変数	spend	メールが配信されてから2週間以内に購入した際の購入額

表 4 各データにおける ATE の推定値  
Table 4 ATE in each dataset.

	RCT が適用されたデータ	選択バイアスが存在するデータ
施策を実施されたユーザの購入額の平均値	1.423	1.951
施策を実施されなかったユーザの購入額の平均値	0.653	0.653
ATE の推定値	<b>0.770</b>	<b>1.298</b>

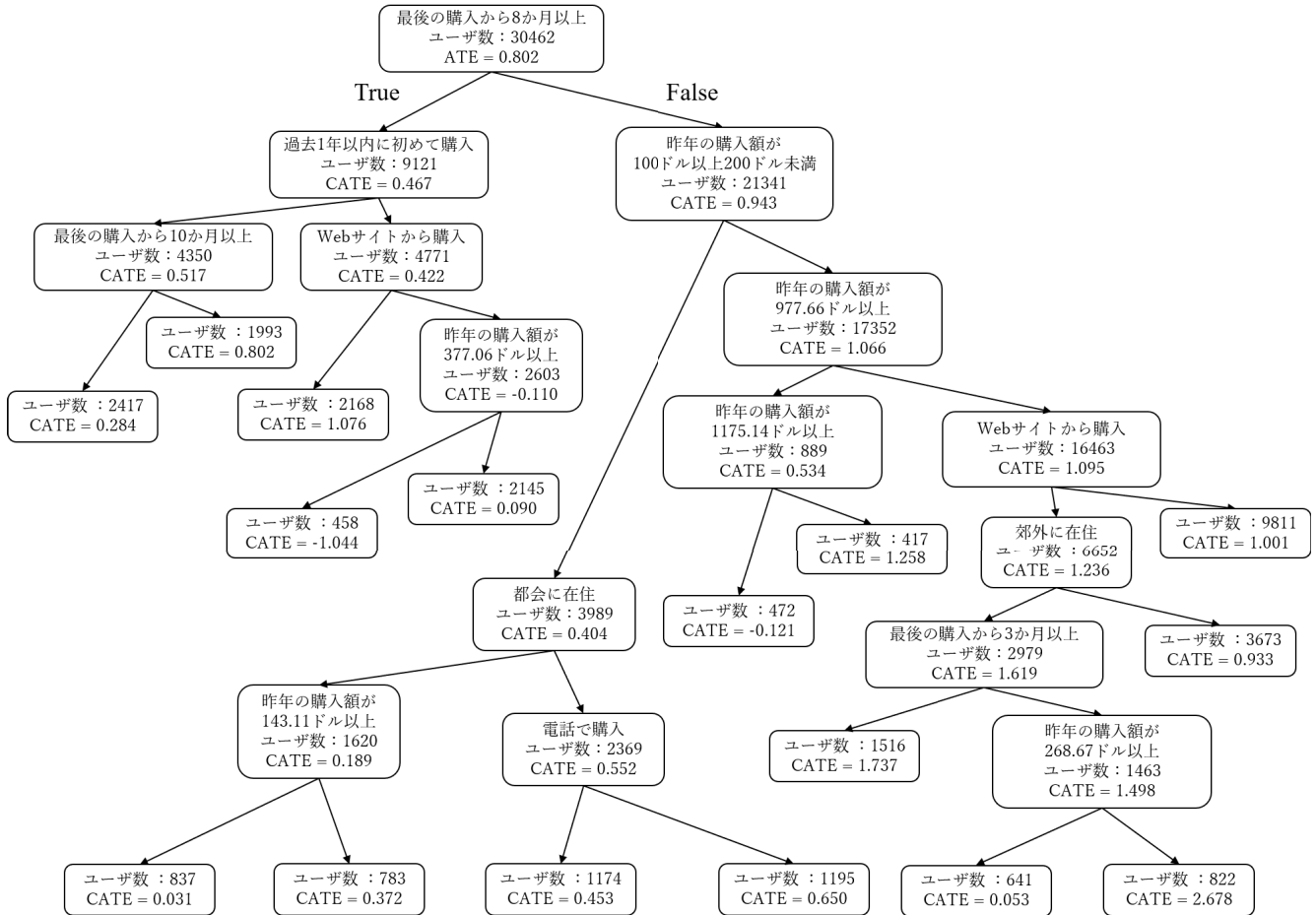


図 3 提案手法を用いて得られた木構造

Fig. 3 Tree structure obtained using the proposed method.

入からの経過月数が「3カ月未満」のユーザ、昨年「300ドル以上」購入したユーザ、昨年「男物商品」を購入したユーザを購買傾向があるユーザと定義し、元のデータの処置群のデータから、購買傾向があるユーザ以外を削減するという操作を行った。表 4 に RCT が適用された元々のデータにおける ATE と、選択バイアスが存在する新データにおける ATE を推定した値を示す。ただし、この場合の ATE の推定は処置群の購入額の平均値と対照群の購入額の平均値の差を利用している。表 4 より、選択バイアスが存在する新たに作成されたデータでは、選択バイアスの影響によって ATE を過剰に推定していることが分かる。新データ作成後の処置群のユーザ数は 9,156 人、対照群のユーザ数は 21,306 人であった。本分析では、提案手法における傾向スコアはロジスティック回帰を用いて推定し、Doubly Robust Estimator の回帰分析は木の数を 500 とし

た XGBoost を用いた。

### 7.2 分析結果

提案手法の可視化を行い、得られた木構造から考察を行う。選択バイアスが存在するデータを提案手法に適用したときに得られた木構造を図 3 に示す。

図 3 より、葉ノードで分けられたユーザ群ごとに CATE (購入額の増分) の推定値が異なり、16 のユーザ群が存在することが分かる。したがって、男性向けメール配信という施策はユーザごとに効果に個人差があるといえる。このように、木構造からどのような基準で分岐されているのが判断しやすいため、提案手法は効果的なユーザ群や、効果に影響を及ぼす要因を容易に行うことが可能である。

ここで、CATE の推定値が大きい上位 3 つのユーザ群の特徴とその群における CATE の推定値を表 5 に示す。表 5

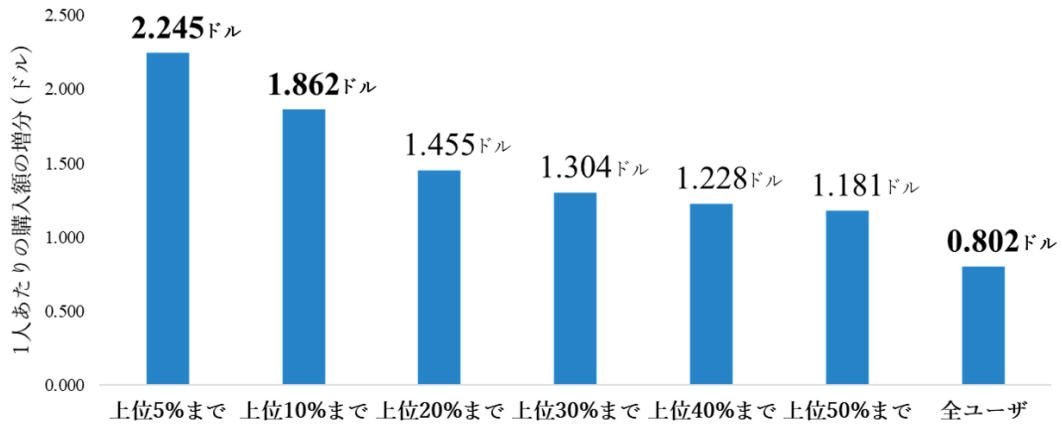


図 4 CATE の推定値が大きいユーザーのみに施策を講じる場合におけるユーザー 1 人あたりの CATE の推定値

Fig. 4 CATE per user in the case where measures are implemented only for users who have large CATE.

表 5 CATE の推定値が大きい上位 3 つのユーザー群の特徴とその群における CATE の推定値

Table 5 The characteristics of the top three user groups with the highest CATE and CATE of these groups.

ユーザー群の特徴	ユーザー数 (人)	CATE (ドル)
・最後の購入から「2 カ月未満」 ・「Web サイト」から購入 ・「郊外」在住 ・昨年の購入額が「268.67 ドル未満」 かつ「100 ドル以上 200 ドル未満でない」	822	2.678
・最後の購入から「3 カ月以上 8 カ月未満」 ・「Web サイト」から購入 ・「郊外」在住 ・昨年の購入額が「977.66 ドル未満」 かつ「100 ドル以上 200 ドル未満でない」	1,516	1.737
・最後の購入から「8 カ月未満」 ・昨年の購入額が「977.66 ドル以上 1175.14 ドル未満」	417	1.258

より、潜在的には昨年の購入額が大きいユーザーや男物商品を購入したユーザーが男性向けメール配信に対する施策効果が見込めそうだが、必ずしも高い施策効果があるとはいえないことが分かる。このことから、元々購入額が大きいユーザーや男物商品を購入していたユーザーは、男性向けメール配信の有無にかかわらず購入するため、施策による効果としては大きくなかったことが考えられる。したがって、単純にヘビーユーザーにこのような施策を講じることは効果的とはいえ、提案手法を用いた分析が有用であったことが分かる。今後は、この分析結果から特定された CATE の推定値が大きいユーザー群に対して、男性向けメールを優先的に配信することが有効である。そこで、CATE (購入額の増分) の推定値が大きいユーザーのみに施策を講じる場合を想定し、考察を行う。図 4 に CATE の推定値が大きいユーザーのみに施策を講じる場合におけるユーザー 1 人あたり

の CATE の推定値を示す。図 4 より、再度施策を講じる際に、施策効果が大きいユーザーを対象として施策を講じることで、高い費用対効果が期待できることが分かる。本分析では、上位 10% のユーザーまでに施策を実施する場合、全ユーザーに実施する場合と比較して、ユーザー 1 人あたり 2 倍以上の効果があると想定されることが分かった。実際に上位何%のユーザーまで施策を講じるかについては、施策における目的や予算によって、検討が可能である。本分析における対象データでの施策はメール配信であり、コストが大きいことが考えられることから、実務上では比較的多くのユーザーに施策を実施しても大きな損失にはつながらないと考えられる。しかし、クーポンなどの施策はコストがかかるため、提案手法で推定された CATE の推定値とコストを比較して判断する必要がある。以上より、提案手法が実用上で有用な分析が行えることを確認できた。

## 8. 考察

### 8.1 選択バイアスへの対応と頑健性

本研究では、Causal Tree をベースとし、Doubly Robust Estimator を活用することによって、選択バイアスが存在する状況に対応した CATE 推定手法を提案した。人工データセットを用いた評価実験により、選択バイアスの存在やノイズ項  $\sigma$  の大きさにかかわらず、提案手法は従来手法より推定精度が高かった。これは Doubly Robust Estimator が持つ、 $X_i$  と  $Z_i$  で  $Y_i(Z_i)$  を説明する回帰分析と、IPW 推定量で用いる傾向スコアを推定する回帰モデルのどちらか一方が正しく指定されていれば処置効果を適切に推定可能であるためだと考えられる。したがって、選択バイアスの有無やノイズ項  $\sigma$  の大きさにかかわらず、提案手法は従来手法よりも有用であるといえる。また、ノイズ項  $\sigma$  が小さいと予想される場合においては、X-learner や DR-learner の方が良い推定精度であったが、解釈性を優先とする場合

には、提案手法の方は木構造から効果が高いユーザ群が判断しやすいため有用であると考えられる。

## 8.2 実応用上における提案手法の適用

本研究においては、提案手法を実際のデータセットに適用することで実応用上で分析可能であることを示した。提案手法は回帰木を用いた Causal Tree をベースにしているため、解釈性が高い。また、実験結果から適用範囲が広いことが分かった。以上より、提案手法は分析者にとって非常に使いやすい手法であり、様々な場面で使用可能であるといえる。したがって、企業はマーケティング施策を講じる際に、提案手法を用いることで、適切な効果検証を行い、正しい意思決定につなげることが可能になると考えられる。ただし、提案手法は回帰木を用いた Causal Tree に基づいているため、線形性のあるデータには適していない点や少しのデータの違いでまったく異なる木になることがある点などの欠点も存在する。そのため、分析者は問題設定に応じて提案手法の適用可否を判断する必要がある。

また、ここでは選択バイアスが存在する状況における効果検証として、企業のマーケティング施策に対する効果を対象としたが、マーケティング施策以外の様々な処置に対しても提案手法は有用性が高いと考えられる。たとえば、医療の現場における薬の効果検証があげられる。薬を与えるという処置は、倫理的な観点から患者にランダムに割り当てることはできないため、データには選択バイアスが存在する。したがって、このようなデータセットに提案手法を適用することによって、どのような特徴を持つ患者に対して薬を与えることが有効であるかの分析や、薬による効果に影響を与える要因についての分析が可能となり、患者に合った治療につなげることが期待される。

## 9. 結論と今後の課題

本研究では、Causal Tree をベースとし、選択バイアスが生じる状況に対応した CATE 推定手法を提案することを目的として、DR 推定法を活用した CATE の算出に基づいた Causal Tree の目的関数の導出を行った。人工データセットを用いた実験では、推定精度を比較することによって、提案手法の有効性を示した。また、実際に EC サイトのユーザに対してメール配信を行ったデータセットに提案手法を適用し、推定結果を解釈し、どのような特徴を持つユーザ群に対して処置を行うことが有効であるかの分析を行うことにより、提案手法の有用性を示した。以上により、提案手法は、Causal Tree の特徴である解釈性を維持しつつ、選択バイアスの有無にかかわらず頑健性を持って CATE の推定が可能であることが示された。したがって、提案手法は、正しい意思決定につなげるために有用かつ実用性の非常に高い手法であるといえ、マーケティングに限らず医療や教育などの様々な分野での活用が期待される。

今後の課題として、処置群と対照群のデータ数が極端に不均衡な状況や共変量が非常に多く存在する状況など、一般的なビジネスの現場で起こりうる様々な状況において提案手法が有用であるかを検証することがあげられる。

謝辞 本研究の一部は、JSPS 科研費 21H04600 の助成を受けたものです。

## 参考文献

- [1] Yao, L., Chu, Z., Li, S., Li, Y., Gao, J. and Zhang, A.: A Survey on Causal Inference, *ACM Trans. Knowledge Discovery from Data*, Vol.15, No.74, pp.1–46 (2021).
- [2] 岩崎 学: データサイエンスと統計的因果推論, *Direct Marketing Review: Journal of the Academic Society of Direct Marketing*, Vol.18, No.1, pp.5–16 (2019).
- [3] 成田悠輔, 粟飯原俊介, 齋藤優太, 松谷 恵, 矢田紘平: すべての機械学習は A/B テストである, *人工知能*, Vol.35, No.4, pp.517–525 (2020).
- [4] 安井翔太: 効果検証入門: 正しい比較のための因果推論/計量経済学の基礎, 技術評論社 (2020).
- [5] Athey, S. and Imbens, G.: Recursive Partitioning for Heterogeneous Causal Effects, *Proc. National Academy of Sciences*, Vol.113, No.27, pp.7353–7360 (2016).
- [6] Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.: *Classification and Regression Trees*, CRC Press (1984).
- [7] Athey, S. and Imbens, G.W.: Machine learning methods for estimating heterogeneous causal effects, *stat*, Vol.1050, No.5, pp.1–26 (2015).
- [8] Rosenbaum, P.R. and Rubin, D.B.: The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, Vol.70, No.1 (1983).
- [9] 星野崇宏: 調査観察データの統計科学—因果推論選択バイアスデータ融合, 岩波書店 (2009).
- [10] Seeger, J.D., Bykov, K., Bartels, D.B., Huybrechts, K. and Schneeweiss, S.: Propensity score weighting compared to matching in a study of dabigatran and swarfatin, *Drug safety*, Vol.40, No.2, pp.169–181 (2017).
- [11] Bang, H. and Robins, J.M.: Doubly robust estimation in missing data and causal inference models, *Biometrics*, Vol.61, No.4, pp.962–973 (2005).
- [12] Rubin, D.B.: Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *Journal of Educational Psychology*, Vol.66, No.5, pp.688–701 (1974).
- [13] Holland, P.W.: Statistics and Causal Inference, *Journal of the American Statistical Association*, Vol.81, No.386, pp.945–960 (1986).
- [14] 大久保将貴: 因果推論の工具箱, 理論と方法, Vol.34, No.1, pp.20–34 (2019).
- [15] Rubin, D.B.: Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment, *Journal of the American Statistical Association*, Vol.75, No.371, pp.591–593 (1980).
- [16] Robins, J.M., Rotnitzky, A. and Zhao, L.P.: Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data, *Journal of the American Statistical Association*, Vol.90, No.429, pp.106–121 (1995).
- [17] Zhang, W., Li, J. and Liu, L.: A Unified Survey of Treatment Effect Heterogeneity Modelling and Uplift Modelling, *ACM Computing Surveys (CSUR)*, Vol.54, No.8, pp.1–36 (2021).
- [18] Knittel, C.R. and Stolper, S.: Using machine learning

- to target treatment: The case of household energy use, Technical report, National Bureau of Economic Research (2019).
- [19] Rana, P. and Miller, D.C.: Machine learning to analyze the social-ecological impacts of natural resource policy: Insights from community forest management in the Indian Himalaya, *Environmental Research Letters*, Vol.14, No.2, p.024008 (2019).
- [20] Breiman, L.: Random Forests, *Machine Learning*, Vol.45, No.1, pp.5–32 (2001).
- [21] Hansen, L.P.: Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica: Journal of the Econometric Society*, pp.1029–1054 (1982).
- [22] Wager, S. and Athey, S.: Estimation and inference of heterogeneous treatment effects using random forests, *Journal of the American Statistical Association*, Vol.113, No.523, pp.1228–1242 (2018).
- [23] Athey, S., Tibshirani, J. and Wager, S.: Generalized random forests, *The Annals of Statistics*, Vol.47, No.2, pp.1148–1178 (2019).
- [24] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C. and Newey, W.: Double/debiased/neyman Machine Learning of Treatment Effects, *American Economic Review*, Vol.107, No.5, pp.261–65 (2017).
- [25] Oprescu, M., Syrgkanis, V. and Wu, Z.S.: Orthogonal Random Forest for Causal Inference, *International Conference on Machine Learning*, PMLR, pp.4932–4941 (2019).
- [26] Powers, S., Qian, J., Jung, K., Schuler, A., Shah, N.H., Hastie, T. and Tibshirani, R.: Some methods for heterogeneous treatment effect estimation in high dimensions, *Statistics in medicine*, Vol.37, No.11, pp.1767–1787 (2018).
- [27] Rzepakowski, P. and Jaroszewicz, S.: Decision trees for uplift modeling with single and multiple treatments, *Knowledge and Information Systems*, Vol.32, No.2, pp.303–327 (2012).
- [28] Chipman, H.A., George, E.I. and McCulloch, R.E.: BART: Bayesian additive regression trees, *The Annals of Applied Statistics*, Vol.4, No.1, pp.266–298 (2010).
- [29] Hill, J.L.: Bayesian nonparametric modeling for causal inference, *Journal of Computational and Graphical Statistics*, Vol.20, No.1, pp.217–240 (2011).
- [30] Hahn, P.R., Murray, J.S. and Carvalho, C.M.: Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion), *Bayesian Analysis*, Vol.15, No.3, pp.965–1056 (2020).
- [31] Saito, Y., Sakata, H. and Nakata, K.: Doubly robust prediction and evaluation methods improve uplift modeling for observational data, *Proc. 2019 SIAM International Conference on Data Mining*, SIAM, pp.468–476 (2019).
- [32] Künzel, S.R., Sekhon, J.S., Bickel, P.J. and Yu, B.: Metalearners for Estimating Heterogeneous Treatment Effects using Machine Learning, *Proc. National Academy of Sciences*, Vol.116, No.10, pp.4156–4165 (2019).
- [33] Kennedy, E.H.: Optimal Doubly Robust Estimation of Heterogeneous Causal Effects, arXiv preprint arXiv:2004.14497 (2020).
- [34] Johansson, F., Shalit, U. and Sontag, D.: Learning Representations for Counterfactual Inference, *International Conference on Machine Learning*, PMLR, pp.3020–3029 (2016).
- [35] Kingma, D.P. and Welling, M.: Auto-encoding Variational Bayes, arXiv preprint arXiv:1312.6114 (2013).
- [36] Louizos, C., Shalit, U., Mooij, J.M., Sontag, D., Zemel, R. and Welling, M.: Causal effect inference with deep latent-variable models, *Advances in neural information processing systems*, Vol.30 (2017).
- [37] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative Adversarial Nets, *Advances in Neural Information Processing Systems*, Vol.27 (2014).
- [38] Yoon, J., Jordon, J. and Van Der Schaar, M.: GANITE: Estimation of Individualized Treatment Effects using Generative Adversarial Nets, *International Conference on Learning Representations* (2018).
- [39] 星野崇宏, 繁榊算男: 傾向スコア解析法による因果効果の推定と調査データの調整について, *行動計量学*, Vol.31, No.1, pp.43–61 (2004).
- [40] Tsiatis, A.A.: Semiparametric theory and missing data (2006).
- [41] Nie, X. and Wager, S.: Quasi-oracle Estimation of Heterogeneous Treatment Effects, arXiv preprint arXiv:1712.04912 (2020).
- [42] Chen, T. and Guestrin, C.: Xgboost: A Scalable Tree Boosting System, *Proc. 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp.785–794 (2016).
- [43] Hillstrom, K.: MineThatData: The MineThatData E-Mail Analytics And Data Mining Challenge, available from <https://blog.minethatdata.com/2008/03/minethatdata-e-mail-analytics-and-data.html> (accessed 2022-06-07).



坪井 優樹

データ分析によるビジネス課題の解決に従事。



阪井 優太 (学生会員)

1996年生。2019年早稲田大学創造理工学部経営システム工学科卒業。2021年同大学大学院修士課程修了。現在、同大学院創造理工学研究科経営システム工学専攻博士後期課程在学、日本経営工学会会員。



鈴木 佐俊 (正会員)

2001年早稲田大学大学院理工学研究科機械工学専攻経営システム工学専門分野修了。2019年同大学院創造理工学研究科経営システム工学専攻博士後期課程。現在、早稲田大学データサイエンス研究所招聘研究員、株式会社富士通総研政策支援グループシニアマネジングコンサルタント。



後藤 正幸 (正会員)

1969年生。1994年武蔵工業大学大学院修士課程修了。2000年早稲田大学大学院理工学研究科博士課程修了。博士(工学)。1997年同大学理工学部助手。2000年東京大学大学院工学研究科助手。2002年武蔵工業大学環境情報学部助教授。2008年早稲田大学創造理工学部経営システム工学科准教授。2011年同大学教授。情報数理応用とデータサイエンス、ならびにビジネスアナリティクスの研究に従事。著書に、『入門パターン認識と機械学習』コロナ社(2014)、『ビジネス統計—統計基礎とエクセル分析』オデッセイコミュニケーションズ(2015)等。IEEE, INFORMS, 電子情報通信学会, 人工知能学会, 日本経営工学会, 経営情報学会等各会員。