RIETI Discussion Paper Series 20-J-045

自然実験としてのアルゴリズム：機械学習・市場設計・公共政策への統一アプローチ

成田 悠輔
経済産業研究所

粟飯原 俊介
ZOZO テクノロジーズ、半熟仮想株式会社

齋藤 優太
東京工業大学、半熟仮想株式会社

松谷 恵
ZOZO テクノロジーズ

矢田 紘平
イェール大学

RIETI
Research Institute of Economy, Trade & Industry, IAA
独立行政法人経済産業研究所
https://www.rieti.go.jp/jp/

RIETI Discussion Paper Series 20-J-045

2020 年 12 月

自然実験としてのアルゴリズム：

機械学習・市場設計・公共政策への統一アプローチ

成田悠輔（経済産業研究所、イェール大学、半熟仮想株式会社）
粟飯原俊介（ZOZO テクノロジーズ、半熟仮想株式会社）
齋藤優太（東京工業大学、半熟仮想株式会社）
松谷恵（ZOZO テクノロジーズ）
矢田紘平（イェール大学）

要旨：公共政策からビジネスまで、機械学習や市場設計などのアルゴリズムを利用した意思決定が広がっている。その際に重要なのが、過去に使われたことのない新しい意思決定アルゴリズムの性能を予測することだ。正確な性能予測は着実なアルゴリズム改善に資する。この論文は、過去に使われたアルゴリズムが自然に蓄積したデータを用いて、未知のアルゴリズムの性能を予測する技法を提案する。この方法は幅広いアルゴリズムに適用可能で、使える場面はウェブ広告配信・価格設定・金融機関の審査のようなビジネスから、裁判の判決、データ駆動教育・医療、そして教育・労働市場設計やオークションなどの公共政策まで多岐にわたる。具体的な応用として、私たち自身が行ったファッション EC サービス ZOZOTOWN 上での実装を紹介する。そこで用いた数千万件のファッション推薦データとコードはオープンソースで GitHub 上で開放中だ。最後に、同じ技術を用いて様々な政策領域の評価・設計・予測も行えるという未来展望を与える。

キーワード：因果推論、機械学習、反実仮想、オフライン政策外評価、ファッション EC、公共政策

JEL classification：D23, L22, L25, M10

I. あらすじ

アルゴリズムによる意思決定が黄金時代を迎えている。その代表が機械学習だ。ソーシャルメディアからEC、金融、裁判、監視にいたるまで、機械学習による予測や分類を用いた意思決定が爆発的に広がっている。監視を例にとれば、監視カメラが捉えた画像データを画像認識することで、人物が犯罪やテロに加担する可能性を予測する。そして危険性が高いと予測された人物を追跡するという意思決定を行う。このプロセス全体が意思決定アルゴリズム・ソフトウェアになる。

機械学習アルゴリズムだけではない。たとえば、世界各地の学校選択・入試制度や労働市場・臓器移植市場などでは割当（マッチング）アルゴリズムが用いられている。国債市場や卸売市場からオンラインの広告や中古品市場まで、オークションアルゴリズムが用いられる場面も枚挙にいとまがない。このようなオークションやマッチングなどの中央集権的な市場設計もまた、アルゴリズムを用いた意思決定である。その他多くの公共政策領域でも、アルゴリズム的ルールを用いて受益資格が決まる場面が多い。

アルゴリズムによる意思決定を行う上で重要なのが、まだ使われたことのない新しい意思決定アルゴリズムの性能を予測することだ。ふたたび監視を例にとれば、新しい人物追跡アルゴリズムを用いた場合の犯罪発生率などがそのアルゴリズムの性能になる。正確な性能予測があれば、着実にアルゴリズムを改善することができる。

すぐに思い浮かぶ性能予測方法は、古いアルゴリズムと新しいアルゴリズムをランダムに人や地域に割り当てて比較するランダム化実験（RCT, A/Bテスト）だろう。だが、RCTは工数も費用もかかる上、被験者に不公平感を与えて炎上しかねないという倫理的問題を抱えている(Narita 20)。RCTに頼ることなく、過去のアルゴリズムが自然に生み出したデータだけで性能予測する方法はないだろうか？

私たちは、過去のアルゴリズムが蓄積したデータを用いて、別の新たなアルゴリズムの性能予測を行う方法を提案する。この方法は以下の観察に基づく。アルゴリズムが意思決定を行なった場合、そこから生成されたデータには、意思決定がランダムに、あたかもサイコロを振ったかのように行われる自然実験がほぼ必ず含まれるという観察である。たとえば、多くの確率的な強化学習・バンディットアルゴリズムは選択（探索）をランダムに行うため、ほとんどRCTそのものである。

また、教師付き学習で予測された何らかの変数がある基準値を上回るかどうかで選択を決めるアルゴリズムを考える。この場合、基準値の近くでは、ほぼ同じ状況であるにも関わらず、基準値をたまたま上回ったかどうかというほとんど偶然の要因で異なった意思決定が行われる。これも局所的な自然実験とみなせる。

こういった自然実験は様々な目的のために使える。意思決定のうちどれが効果的か測るために使え、新たな意思決定アルゴリズムを導入するとどのような性能を発揮しそうか予測するためにも使える。

私たちは、この観察を一般の機械学習アルゴリズムについて定式化し、アルゴリズムが自然に生成したデータを用いてアルゴリズムを改善する手法を開発する。この手法が使える場面は、ビジネスから政策まで幅広い。具体的な応用として、私たち自身が行ったファッション EC サービス ZOZOTOWN 上での配備を紹介する。この応用では、ZOZOTOWN の一部でのクリック率を約 40%高め、さらなる改善の方法を見つけることにも成功した。この実装で用いた数千万件のファッション推薦データ、そしてそのデータを用いて推薦アルゴリズムを開発するためのソフトウェア基盤はオープンソースで GitHub 上で公開中だ。

2. 機械学習アルゴリズムが世界を食い尽くす

機械学習アルゴリズムに基づく意思決定が雨後の筍状態である。たとえば、Amazon、Apple、Facebook、Google、Microsoft、Netflix をはじめとするウェブ企業は、表示するコンテンツ（映画、音楽、ニュース等）や広告の選択、価格や検索結果順位の決定といった問題に、機械学習を応用している（特に、後述するバンディットアルゴリズムや強化学習アルゴリズム）。また、Uber や Lyft、DiDi といった自動車共有サービスの価格は、各時点・場所における需要と供給の情報をもとに、独自のアルゴリズムによって決定されている。

機械学習アルゴリズムを利用した意思決定は、デジタル世界以外にも広がっている。たとえば裁判や保釈判決がその例だ。米国企業 Northpointe（現 Equivant）が開発したソフトウェア COMPAS は、教師あり機械学習を用いて被告人の再犯確率を予測する(Dieterich et al 16, Kleinberg et al 17)。COMPAS の予測した危険指数は、米国の多くの裁判官の判断材料として実際に利用されている。また、多くの国の金融機関は、機械学習により顧客の返済能力を予測・数値化した信用スコアに基づいてクレジットカードやローンの審査を行っている。そのほか、機械学習アルゴリズムを用いた人事採用システムも登場している。表 1 にこれらの例の一部をまとめた。

表 1　機械学習アルゴリズムに基づく意思決定の例

	アルゴリズムが用いる変数（X）	アルゴリズムの意思決定（Z）	結果変数（Y）	アルゴリズム例
ウェブ企業	利用者の閲覧履歴、アクセスの時間・場所	表示コンテンツ	利用者がコンテンツにアクセスしたかどうか	バンディット等の強化学習(本多 16, Sutton and Barto 18)
自動車共有サービス	利用者がアプリを開いた時点における周辺地域の需要と供給の情報	サービスの価格	利用者がサービスを利用したかどうか	価格上昇動的価格決定(Cohen et al 16)
裁判官	被告人の犯罪歴、年齢等の属性	釈放すべきか否か	被告人が再犯したかどうか	教師あり学習(Dieterich et al 16, Kleinberg et al17)

アルゴリズム化した世界が私たちの分析の素材である。2011 年にネットスケープ創業者で投資家の Marc Andreessen は「ソフトウェアが世界を食い尽くす(software is eating the world)」と言った。Andreessen が念頭においていた「ソフトウェア」は、主に人間が完全に書き下した規則にしたがって動く古典的ソフトウェアだろう。その後の 10 年弱で、上にあげた様々な例のように、ソフトウェアにどんどん機械学習・統計モデルが載るようになった。機械学習ソフトウェアを書いて動かすのは、人間というよりデータである。これが現在の機械学習ブームの立役者の一人 Andrej Karpathy の唱える Software2.0 の描像であって、その精神は彼の有名なつぶやきに詰まっている。

「ごめん、君より最急降下法の方がコードを書くのがうまいみたいだ。

(Gradient descent can write code better than you. I'm sorry.)」

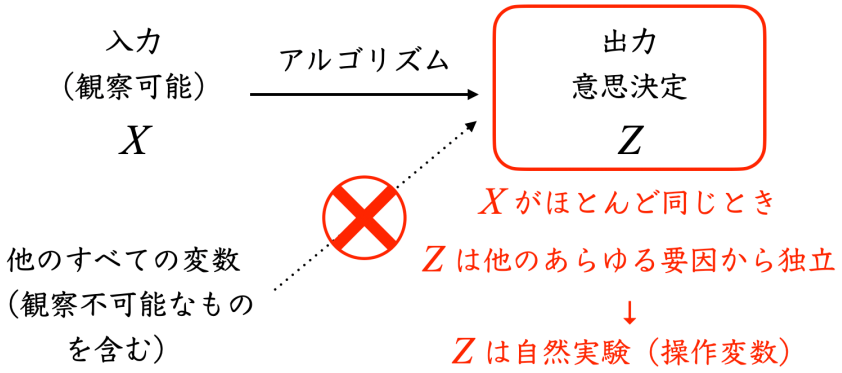
データ駆動の Software2.0 としての機械学習アルゴリズムが人間の脳を置き換え、世界を食い尽くすようになった。ソフトウェアに食い尽くされた世界が吐き出したデータをどう味わえばいいか---それが私たちの問いである。

3. アルゴリズムは自然実験を生む

上の問いに答えるため、私たちはアルゴリズムによる意思決定の隠れた特徴に注目する。まず、非常に一般的に言って、意思決定アルゴリズムとは「各個人iについて観察できる変数X_iを読み込んで、個人iに対して行う意思決定Z_i上の確率分布を決める関数」であると定義できるだろう。この定義から、あまり注目されないが重要な特徴がすぐに見えてくる。意思決定Z_iが観察可能な入力変数X_iのみに基づくことである。たとえば、ウェブ企業のアルゴリズムは、データに含まれる利用者の属性や視聴履歴など、アルゴリズムが読み込むように指示された変数のみに基づき、表示するコンテンツを選択する。

意思決定が観察可能な入力変数のみに基づくことから、次のように言えるだろう（図 1）。意思決定を左右する入力変数X_iの値がほとんど同じ様々な個人に対して、様々な異なる選択Z_iが行われているとしよう。すると、各個人に対する選択Z_iは、あたかも RCT のように、乱数かサイコロによってランダムに決められたものと考えることができる。結果として、アルゴリズムが直接使う入力変数以外の他のどんな要因にもとらわれることなく、意思決定そのものの効果を測ることができる。つまり、機械学習アルゴリズムによる意思決定はほとんど RCT である。統計学や計量経済学の用語を使えば、アルゴリズムによる意思決定は、意図せず自然に発生した実験という意味で自然実験（natural experiment）であり操作変数（instrumental variable）だと言える。以下、2 つの具体例を用いて解説しよう。

図 1 なぜアルゴリズムは自然実験か？



3.1. 確率的アルゴリズム

ウェブサービスの運営者の目的の 1 つは、利用者の好みに合わせたコンテンツや広告を表示し、アクセス数やクリック数を最大化することである。好みや属性の違う利用者ごとに最もよいコンテンツを表示するためには、様々な候補を試してデータを貯める一方で、貯まったデータから良さそうだと推測された候補を実行していく必要がある。このような候補探索と知識利用のバランスを考慮しながら目的関数を最適化する手法が、バンディットアルゴリズム(本多 16)やその上位概念である強化学習アルゴリズム(Sutton and Barto 18)である。

たとえば、バンディットアルゴリズムの代表例であるϵ-貪欲（ϵ-greedy）アルゴリズムは、利用者がページを訪れると、過去のデータからその利用者にとって最良の選択肢（購買などにつながる可能性が最も高いと推測される選択肢）を機械学習で予測する。そして最良と予測された選択肢を$1-\epsilon$の確率で選択する。残りのϵの確率では、ランダムに選択肢を選ぶ。ϵは 0 に近い小さい値に設定され、最良と推測されたものが高確率で選択される。

ここで重要な点は、①どれが最良であるかは利用者の観察できる属性のみに基づいて推測され、②どの選択肢も正の確率でランダムに選ばれることである。この2つの点から、ϵ-貪欲法による選択肢の割り当ては、層別（stratified）RCT（被験者を属性によって分類し、各グループ内では選択肢をランダムに割り当てる実験）とみなすことができる。このように選択肢をランダムに選ぶアルゴリズムは、ほかにもトンプソン抽出（Thompson sampling）などがある。トンプソン抽出は、購買などにつながる効果が高そうだと予測される選択肢ほど高い確率で選択するアルゴリズムだ。

こうしたアルゴリズムの運用で生成されたデータは、層別RCTで集められたデータと同じ手法で分析し、様々な意思決定の効果を推定したり、まだ使われたことのないアルゴリズムの性能を予測することに使える。そのような性能予測ができることは古くから知られ、政策外（off-policy）評価やオフライン（offline）評価などと呼ばれることが多い(Precup 00, Li et al 10)。今オンラインで動いているアルゴリズムの外で独立にできる評価だという意味でオフラインであり、今使われている政策以外の政策の効果を評価しているという意味で政策外だからだ。政策外オフライン評価のよく知られている実例は、Netflixがトップ画面に表示する動画を決めるために用いるバンディットアルゴリズムの改善である(Amat et al 18)。

筆者の一部と株式会社サイバーエージェントの共同研究ももう1つの例である。バンディットアルゴリズムにより生成されたデータの統計学的に最も効率的な分析手法を開発し、同社の広告配信ログデータを用いて、広告画像の選択アルゴリズムの性能評価を行っている(株式会社サイバーエージェント 18, Narita et al 19, Narita et al 20)。また、後ほど詳しく説明するファッションECサイトZOZOTOWNとの共同事業では、似た手法を用いて服装のコーディネートの推薦アルゴリズムなどを作り変えている。

3.2. 非確率的アルゴリズム

確率的なアルゴリズムはほとんどRCTなので話は簡単だった。では、確率的でない、どれかの選択を確率1で行うアルゴリズムはどうだろうか？実は、そんなアルゴリズムにも自然実験が隠れている。

たとえば、自動車共有サービスUberの核をなす非確率的なアルゴリズムを見てみよう。Uberは、乗客と車を運転する運転手を、アルゴリズムを使って組み合わせるアプリである。乗客がアプリを開いて目的地を入力すると、目的

地までの料金と推定到着時刻が表示され、乗客がその条件を受け入れると周辺の運転手と出会うという仕組みである。Uber の通常料金は、目的地までの距離や乗車時間等により決まる。

ただ、需要（乗客の数）が供給（運転手の数）を超過した場合は、「価格上昇（surge pricing）」と呼ばれるアルゴリズムによって、需要と供給が釣り合うように料金を引き上げる。価格上昇アルゴリズムは、その時点・場所における需要と供給の情報をもとに上昇乗数を計算し、通常料金に上昇乗数をかけたものを実際の料金として乗客に提示する。

このアルゴリズムも自然実験を生み出すという観点から興味深い。面白いのは、まず連続的な上昇乗数を計算し、その小数点第 2 位以下を四捨五入したものを実際の料金計算に反映させるという点である。たとえば、もともとの連続的な上昇乗数が 1.249 であれば 1.2 に丸めるが、もともとの値が 1.251 であれば 1.3 に丸められる。つまり、四捨五入によって上昇乗数が非連続的に飛び上がる閾値が生まれる。

閾値の付近においては、需要と供給の状況がほとんど同じであるにもかかわらず、大きな料金の変化が見られることとなる。言いかえれば、閾値付近では「局所的な RCT もどき」とでも呼ぶべき状況が生まれている。閾値付近でたまたま料金が高かった状況と低かった状況を比べれば、料金が乗客のサービス利用などに与える影響を測ることができる。Cohen et al (16)はこのアイデアに基づき、Uber のデータを用いて、料金が乗客のサービス利用に与える影響を推定した。社会科学でよく用いられる「回帰不連続デザイン(regression discontinuity design)」の一種だ(安井 20)。

機械学習アルゴリズムが生成したデータに局所的な自然実験が見られる例は多く存在する。たとえば、先ほど紹介した裁判官の判決や金融機関のローン審査において、教師あり機械学習によって計算された連続的な指数が閾値を超えるかどうかで意思決定を行うとしよう。この場合、閾値付近では、対象となる個人の属性がほとんど同じであるにもかかわらず、異なった意思決定が行われることとなる。非確率的アルゴリズムの中でも、意図せず局所的な自然実験が生まれるのである。

4. 機械学習生成のデータを分析する

ここまで、具体例を用いながら、いくつかの機械学習アルゴリズムが意図せざる自然実験を生成することを見た。他にもアルゴリズムは無数に存在するが、それらについてはどうだろうか？アルゴリズムがどのような条件を満たせば自然実験が存在するのか分析してみよう。技術的な詳細や証明は Narita and Yata (20)に譲り、ここでは骨組みのみ素描する。

4.1. 定式化

すでに述べた通り、機械学習アルゴリズムによる意思決定は、観察可能な入力変数のみに基づくという特徴があった。アルゴリズムが用いる変数群を、p次元の連続型確率変数ベクトル$X_i \in \mathbb{R}^p$で表すとしよう。添字のiでこれが個人iの属性であることを示している。以下の議論は、X_iが離散変数を含む状況へも拡張できる。議論を簡単にするため、意思決定は、ある処置（treatment, arm）を行うかどうかの 2 択であるとし、その選択を 2 値変数$Z_i \in \{0,1\}$で表す。$Z_i = 1$であれば処置を行ったこと、$Z_i = 0$であれば処置を行わなかったことを意味する。そして、機械学習アルゴリズムを次のような既知の関数MLで表す。

$$ML: \mathbb{R}^p \to [0,1]$$

MLは、変数X_iの値を入力すると、処置を行う確率を出力する関数である。これまで紹介したアルゴリズムはすべてこの関数の例とみなせる。たとえば、先に紹介したϵ -貪欲法アルゴリズムで処置するかどうかを選択する状況を考えよう。過去のデータから処置を行うことが最適とされる個人の属性xにおいては$ML(x) = 1 - \epsilon$、処置を行わないことが最適とされるxにおいては$ML(x) = \epsilon$となる。個人の信用スコアが一定以上であれば融資をするというアルゴリズムを金融機関が用いる場合、信用スコアが閾値以上となる個人の属性xにおいては$ML(x) = 1$、信用スコアが閾値未満となるxにおいては$ML(x) = 0$となる。

いま、属性X_iを持つ様々な個人iに対して、MLアルゴリズムを回して意思決定Z_iを行う。そして意思決定Z_iに影響を受けて各個人について何らかの結果Y_iが観察される。たとえば監視カメラによる人物監視の例で言えば、iが監視カメラに映る人物、X_iがその人物の表情や行動に関する監視カメラデータ、Z_iがその人物を追跡するという意思決定、Y_iがその人物による犯罪の発生などとなる。金融機関による審査の例では、iがローン応募者、X_iがローン応募者の属性や過去の行動履歴、Z_iがローン貸付、Y_iが夜逃げなどとなる。

私たちの目的は、結果Y_iを最適化する上で処置を行うべきかどうか学習し、属性X_iに応じて誰に処置を行うかを決めるアルゴリズムを設計することである。そのために、処置を行うことの効果を知りたい。まず、個人iに対して処置が行われた場合の潜在的結果（potential outcome）を$Y_i(1)$で表す。$Y_i(1)$は、個人iが仮に処置を受けたとすると観察されるY_iの値である。一方で、処置が行われなかった場合の潜在的結果を$Y_i(0)$とする。

したがって、その個人iにとって処置が結果に与える因果効果は$Y_i(1) - Y_i(0)$になる。現実には、個人iは処置を受けるか受けないかのどちらか一方しか生じないため、$Y_i(1)$と$Y_i(0)$のどちらか一方しか観察できず、個々人にとっての因果効果はデータからは測れない。この難しさがいわゆる因果推論の根本問題（Fundamental Problem of Causal Inference）と呼ばれる困難である(Imbens and Rubin 15)。データ上で観察される結果変数は$Y_i = Z_i Y_i(1) + (1 - Z_i)Y_i(0)$のみであって、この観察データを用いて処置を行うことの因果効果をどうにか測りたい。

機械学習アルゴリズムMLを用いてデータ(X_i, Z_i, Y_i)が次のように生成されるとしよう。まず、$(X_i, Y_i(1), Y_i(0))$が未知の確率分布にしたがって選ばれる。X_iをアルゴリズムMLに入力し、処置が行われる確率$ML(X_i)$が出力される。その確率に基づき、$(Y_i(1), Y_i(0))$と独立にZ_iの値が選ばれ、結果変数Y_iの値が観察される。

4.2. 因果効果の学習

4.2.1. 疑似傾向スコア

私たちの目的は、現在用いられているアルゴリズムが蓄積したデータを用いて、別の新たなアルゴリズムの性能を予測することだった。この目的のためには、アルゴリズムが行う様々な選択がどのような効果をもたらすか、因果関係としての効果をまず正確に測る必要がある。

どんな現行アルゴリズムから出てきたデータであれば、因果効果を学習できるだろうか？鍵となるのが、「擬似傾向スコア（Quasi Propensity Score）」という概念である。変数X_iの台（support）を$\mathcal{X}$とする。まず、ある$x \in \mathcal{X}$において、$B(x, \delta)$を中心がxで半径がδのp次元の球体とする。つまり、$d(x, x^*)$をxとx^*の間のユークリッド距離として$B(x, \delta) = \{x^* \in \mathbb{R}^p : d(x, x^*) < \delta\}$。ここで

$$p^{ML}(x; \delta) = \int ML(x^*) dU_{B(x,\delta)}(x^*) \tag{1}$$

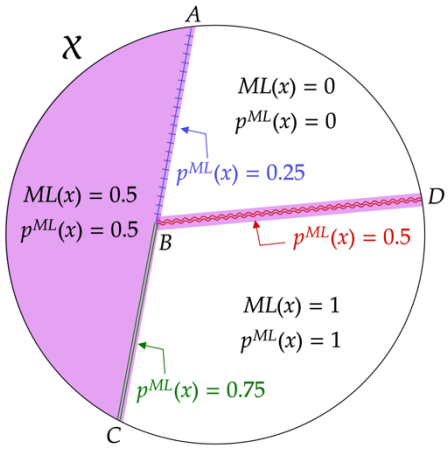
とする。$U_{B(x,\delta)}$は$B(x, \delta)$上の一様分布を表し、$p^{ML}(x; \delta)$は$B(x, \delta)$におけるMLの平均を表す。そして、xにおける擬似傾向スコアを、以下で定義する。

$$p^{ML}(x) = \lim_{\delta \to 0} p^{ML}(x; \delta)$$

擬似傾向スコアは、xの限りなく小さな近傍で処置が行われる平均確率と解釈できる。図２は、X_iが２次元の場合の関数MLと擬似傾向スコアの例を示した。緩い正則条件を満たすMLとxについては、擬似傾向スコア$p^{ML}(x)$が必ず存在すると証明することもできる。

図２ アルゴリズムML と疑似傾向スコア$p^{ML}(x)$の例

アルゴリズムが用いる入力変数の台$\mathcal{X}$が下のような２次元の空間であるとし、MLの値により$\mathcal{X}$は３つの部分に分割されるとする。それぞれの部分の内側にあるxにおいては、$p^{ML}(x)$は$ML(x)$に等しい。２つの部分の境界線上にあるxにおいては、$p^{ML}(x)$は２つの部分のMLの値の平均となる。



4.2.2. 無限データでの学習（識別）

擬似傾向スコアは、どんな自然実験が存在しているかを示すリトマス試験紙になる。xにおける擬似傾向スコアが０と１以外の値であれば、xの付近で、処置が行われる場合と行われない場合のどちらもが正の確率で存在する。彼らはほとんど同じ属性を持つほとんど同じ人たちなので、そのようなxの付近では自然実験が発生していると考えられる。よって、xの付近で処置を受けた人と受けていない人を比べれば、因果効果を学習できそうだ。実際、以下の命題が成り立つ。

命題１　条件付き期待値関数$E[Y_i(1)|X_i]$と$E[Y_i(0)|X_i]$がX_iについて連続であるとし、MLと$\mathcal{X}$にいくつかの技術的条件を置くと、以下が成り立つ。

ⓐ $p^{ML}(x) \in (0,1)$が成り立つすべての$x \in \mathcal{X}$において、因果効果$E[Y_i(1) - Y_i(0)|X_i = x]$を識別できる。

ⓑ Sを$\mathcal{X}$の開部分集合とする。因果効果$E[Y_i(1) - Y_i(0)|X_i \in S]$を識別できるなら、すべての$x \in S$において$p^{ML}(x) \in (0,1)$が成り立つ。

ここである因果効果のパラメーターが識別（identify）できるとは、その因果効果のパラメーターが(X_i, Z_i, Y_i)の同時分布から一意に定まることをさす。つまり、仮に無限大のデータがあり(X_i, Z_i, Y_i)の同時分布がわかれば、その因果効果のパラメーターが学習できることを意味する。

(a)は因果効果が無限データで学習できるための十分条件、(b)は必要条件を示している。(a)から、$p^{ML}(x) \in (0,1)$となるようなxが存在すれば、データから何らかの因果効果を学習できることがわかる。たとえば図2に戻ると、図で色づけされた部分が命題1の条件を満たす。したがって、色づけされた部分の属性を持つ個人にとっての因果効果なら学習できることになる。

命題1で示された因果効果が学習できるための条件は、直感的には何を意味するのだろうか？$p^{ML}(x) \in (0,1)$の条件を満たすxの周辺には、X_iの値がほとんど同じであるにもかかわらず、処置を受ける個人と受けない個人が混在している。彼らはほとんど同じ属性を持ったほとんど同じ人たちなので、処置を受ける個人と受けない個人の間で結果Y_iに差があれば、それは処置の効果と言えるはずだ、という論理である。

教師あり学習、強化学習やバンディットといった一般的な機械学習による意思決定では、(a)の条件が満たされ、何らかの因果効果の学習が可能である。数学的にも「ほとんどすべての」MLアルゴリズムについて、命題1の条件が少なくともあるxについては成り立つことが示せる。つまり、ほとんどすべての機械学習は自然実験を含み、それを使って因果効果を学習できることになる。

4.2.3. 有限データでの学習（推定）

命題1では、仮にアルゴリズムMLが生み出したデータが無限に存在した場合の学習（識別）を扱った。では、実世界の有限なデータをどのように分析すれば因果効果を推定できるだろうか？n人の個人を含むデータ$\{(X_i, Z_i, Y_i)\}_{i=1}^{n}$が与えられたとする。まず、$\delta_n$を小さい値に設定し、それぞれの個人$i$について$p^{ML}(X_i; \delta_n)$を計算しよう。$n$が増えるにつれ、$\delta_n$は0に収束していくと考える。$p^{ML}(X_i; \delta_n)$は、人間の手で解析的に求めるか、(1)式の右辺の積分をシミュレーションで近似すれば計算できる。次に、データのうち$p^{ML}(X_i; \delta_n) \in (0,1)$となるものを使って、以下の回帰式を最小2乗法（Ordinary Least Square）で推定する。

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2 p(X_i; \delta_n) + \epsilon_i$$

この最小２乗法では、擬似傾向スコア$p(X_i;\delta_n)$を制御した上で、結果変数Y_iを処置変数Z_iに回帰している。前節での議論が示唆するように、擬似傾向スコアを共有する個人の間では処置が行われるかどうかがほとんどランダムに決まると考えられる。そのため、上の最小２乗法を用いて同じ擬似傾向スコアを共有した人の中で処置を受けた人と受けなかった人を比べれば、処置の因果効果を測れると期待できる。話を単純にするため、誰にとっても因果効果は定数、つまりすべての個人iについて$Y_i(1)-Y_i(0)=\beta$だと仮定しよう。すると、次の命題が成り立つ。

命題２　いくつかの技術的条件の下で、データが増え$n\to\infty$で$\delta_n\to 0$となるにしたがって$\hat{\beta}_1\to_p\beta$となる。つまり、最小２乗推定量$\hat{\beta}_1$は、因果効果βの一致推定量（consistent estimator）となる。

どんなにX_iが高次元でMLアルゴリズムが複雑でも、上の単純な最小２乗法さえ回せばよいのは嬉しい。どんなに入り組んだデータやアルゴリズムが用いられていても、因果効果の学習は単純明快に済むのである。

4.3. まだ見ぬアルゴリズムの性能予測

命題２では処置Z_iが結果Y_iに与える影響を測った。同じ理屈で、まだ使われたことのない仮想のアルゴリズムの性能を予測することもできる。アルゴリズムの設計・運用者が、いま使われているアルゴリズムMLよりも別のアルゴリズムML'の方がいいかもしれないと悩んでいるとしよう。ML'に移行すべきかどうか決めるために、MLとML'のどちらがよりよい結果を生み出しそうか知りたい。もしML'を使った場合どのような結果Y_iが得られるだろうか？

仮想のアルゴリズムML'を使った場合に達成できるYの期待値（ML'の性能）を

$$V\left(ML'\right)=E[ML'(X_i)E(Y_i(1)-Y_i(0)|X_i)]+E(Y_i(0))$$

と書こう。第２項$E(Y_i(0))$は処置が誰にも行われない場合の結果、第１項$E[ML'(X_i)E(Y_i(1)-Y_i(0)|X_i)]$は$ML'$が行う処置によって生まれる結果の増分だ。この性能$V\left(ML'\right)$も、前節で議論した最小２乗法を使って簡単に学習できる。前節と同じく、誰にとっても因果効果は等しいと仮定しよう。アルゴリズムMLから得られたデータが私たちの手元にある。すると、命題２と同じ条件の下で、次の事実が成り立つ。

命題３　任意の仮想アルゴリズム ML' について、

$$\frac{\sum_{i=1}^{n} Y_i}{n} + \hat{\beta}_1 E\left[ML'(X_i) - ML(X_i)\right]$$

は $V\left(ML'\right)$ の一致推定量である。つまり、データが増え $n \to \infty$ で $\delta_n \to 0$ となるにしたがって

$$\frac{\sum_{i=1}^{n} Y_i}{n} + \hat{\beta}_1 E\left[ML'(X_i) - ML(X_i)\right] \to_p V\left(ML'\right)$$

ここで、$E[ML'(X_i)]$ は $ML'(X_i)$ の X_i についての平均をとったもので、ML' が使われた場合の処置確率の期待値を表す。よって $E\left[ML'(X_i) - ML(X_i)\right]$ は ML から ML' への移行で平均処置確率がどれだけ上がるかを表す。これに処置の効果の推定値 $\hat{\beta}_1$ をかけた $\hat{\beta}_1 E\left[ML'(X_i) - ML(X_i)\right]$ は、ML から ML' への移行で Y_i の平均がどれだけ上がるかを推定したものになる。それを ML の下での Y_i の平均 $\frac{\sum_{i=1}^{n} Y_i}{n}$ に足せば ML' の下での Y_i の平均が得られるというカラクリだ。

たまたま使われていたアルゴリズム ML から出てきたデータをうまく使えば、どんな仮想のアルゴリズム ML' の性能 $V\left(ML'\right)$ も予測できることがわかった。そして、$V\left(ML'\right)$ を最大にする ML' を選べば、工数と費用のかかる RCT をすることなくよりよいアルゴリズムを見つけ出すことができる。この方法を実世界で使ってよりよいアルゴリズムを見つけ出してみよう。

5.　社会実装：ZOZOTOWN のファッションバンディット

5.1.　ZOZOTOWN の挑戦

ZOZOTOWN（https://zozo.jp/）は日本最大級のオンライン・ファッション通販サービスで、1300 以上のショップ、7400 以上のブランドの取り扱いがある（2019 年 12 月末）。ZOZOTOWN が近年直面している課題が顧客層の拡大と多様化だ。これまでは若い女性客が中心だったが、ソフトバンクグループの Yahoo! JAPAN との業務提携による PayPay モールへの出店で新たな顧客層を獲得、2019 年秋にはゴルフ大会を主催し中高年男性層でのシェア拡大を狙ったり、自宅で足の計測を行える新商品 ZOZOMAT で靴市場にも参入したりと、新市場への殴り込みも積極的に行っている。中国市場にも自前のアプリとサイトで進出中だ。

新規顧客の流入はありがたいが，顧客の多様化にどう対応するかという難しさが立ち上がる。ZOZOTOWN は，天性のマーケターである創業者・前澤友作前社長が原宿系と呼ばれるような主に若者を中心とした層に刺さる施策を打って成長してきた会社だった。しかし近年、顧客の裾野や取り扱いブランドが広がるにつれて、次々に繰り出す施策

と経営判断の不確実性が増している。特に中国市場で顕著だが、移り変わりが激しい未知の市場を相手にするときには、経営者の直観だけでなくデータと事実に基づき様々な施策の効果をすばやく予測・実行・評価していくことが求められる。

こうした背景をもとに需要が高まっているのが、よりデータ駆動にアルゴリズムを開発し改善していく枠組みだ。そのような試みの 1 つとして、前節で素描したアルゴリズム性能予測の方法を ZOZOTOWN のサービスデザインに応用した。技術的な詳細は Saito, Aihara, Matsutani, and Narita (20)に譲る。

5.2. メタ A/B テスト

ZOZOTOWN のトップページを生成しているアルゴリズムをよりよいものにしたい。そのために、新たなアルゴリズムを導入した場合の性能を予測して、実装の手間とリスクを抑えたい。その性能予測に必要なデータを構築するため、2019 年 11 月の 7 日間に渡って ZOZOTOWN のトップページ上で実験を行った。実験の対象となったのは、トップページの「身長と体重で選ぶマルチサイズアイテム」と呼ばれるファッションアイテム推薦欄だ（図 3）。何が表示されるかは来訪ユーザーの性別によって異なり、女性（women's）、男性（men's）、性別問わず（all）の 3 つの領域がある。それぞれの領域ごとに異なるアイテム候補群の中からアイテムが推薦される。この各領域を「キャンペーン」と呼ぼう。

図 3 ZOZOTOWN のファッション推薦

実験では、各キャンペーンにユーザーが到来するごとに、まずどのようなアルゴリズムで服を推薦するかランダムに選ぶ。アルゴリズムの候補は 2 つで、(1)ランダムに服を選ぶ素朴 A/B テストアルゴリズムか(2)トンプソン抽出と

呼ばれるバンディットアルゴリズム（2.1 節で軽く触れた）である。この段階で選ばれるのは、服ではなくアルゴリズムであることに注意してほしい。そして、選ばれたアルゴリズムが最終的に推薦する服を選ぶという流れだ。表 2 に実験データの記述統計を示した。合計数百の服に対する合計数千万の訪問・閲覧が起きた巨大な実験データであることがわかる。

表 2 実験データの記述統計

キャンペーンごとに、2 つのアルゴリズムに関するユーザー訪問データ数（#Data）と候補となるファッションアイテムの数（#Items）を示している。TS はトンプソン抽出、Average Age はユーザーの平均年齢、CTR は平均クリック率、Relative-CTR は All キャンペーンにおけるランダムアルゴリズムと比べた場合の相対的平均クリック率を表す。

Campaigns	Behavior Policies	#Data	#Items	Average Age	CTR (V^{π}) ±95% CI	Relative-CTR
ALL	RANDOM	1,374,327	80	37.93	0.35% ±0.010	1.00
	BERNOULLI TS	12,168,084			0.50% ±0.004	1.43
MEN'S	RANDOM	452,949	34	37.68	0.51% ±0.021	1.48
	BERNOULLI TS	4,077,727			0.67% ±0.008	1.94
WOMEN'S	RANDOM	864,585	46	37.99	0.48% ±0.014	1.39
	BERNOULLI TS	7,765,497			0.64% ±0.056	1.84

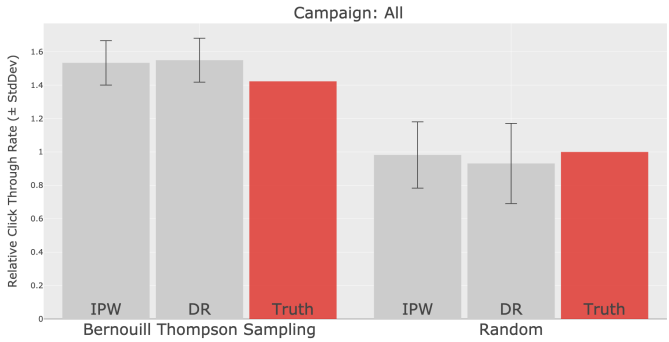
この実験は、ファッション・アイテムをランダムに選ぶアルゴリズム自体をランダムに選ぶ、いわばメタ A/B テストである。メタ A/B テストには大きな利点がある。異なるアルゴリズムが実際にどのような性能を持つかデータ上で確認できるという利点だ。

このメタ A/B テストの特性を生かして、まずは 4 章で導入した分析手法が妥当か確かめよう。そのために、上記の(1)か(2)のどちらかのアルゴリズムが作り出したデータを使って、もう一方のアルゴリズムの性能を予測する。その予測をメタ A/B テストデータと突き合わせ、予測がどれくらい正確かを評価する。性能指標としてはクリック率を用いる。4 章の記号で言えば、iがユーザー、Z_iが推薦される服、Y_iが推薦された服をクリックしたかどうかになる。Z_iが 2 値ではなく数多くある服のうちのいずれかである点のみ 4 章の理論的枠組みから逸脱している。

性能予測の正しさを検証したのが図 4 である。メタ A/B テストデータで実際に観察された各アルゴリズムの性能を表したのが Truth である。そして、命題 3 で導入した性能予測方法のちょっとした拡張を使った 2 種類の性能予測が IPW（Inverse Probability Weighting の略）と DR（Doubly Robust の略）で表されている。まったくランダムに選ぶアルゴリズムと比べ、トンプソン抽出アルゴリズムは 40%ほど高いクリック率を達成していることがわかる。そして、性能予測は実際の性能とほぼ一致している。命題 3 も示唆する通り、理論的な性能予測は正確なようだ。

図 4 新旧アルゴリズムの性能比較

ランダムアルゴリズム（右パネル）とトンプソン抽出アルゴリズム（左パネル）のそれぞれについて、Truth がメタ A/B テストデータで実際に観測されたアルゴリズムの性能を、IPW と DR がそれに対する予測値を表している。予測値は Truth とほとんど同じ値で、予測が正確なことがわかる。



5.3. アルゴリズム改造へ

トンプソン抽出アルゴリズムは素朴すぎるランダムアルゴリズムより高性能なことがわかった。だが、それが最適である保証はない。もっといいアルゴリズムがどこかにあるのではないだろうか？命題3で示した方法に基づいて、「身長と体重で選ぶマルチサイズアイテム」をデザインするためのよりよいアルゴリズムを見つけ出そう。どのような属性X_iを用いるか、そしてどのようなMLアルゴリズムを用いるかによって様々な選択肢が考えられる。ここでは次のような候補を考える。

- 用いるMLアルゴリズムの候補：ロジスティック ε-貪欲で ε の値は3つの値のどれか、ロジスティック・トンプソン抽出(Chapelle 11)、ロジスティック信頼上限（Upper Confidence Bound UCB; Li et al 10)でパラメーターの値は2つの値のどちらか
- 用いる属性の候補: 属性集合1（ユーザーの年齢・性別・会員年数）、属性集合2（属性集合1に加え、過去のクリック履歴に基づくユーザーと服との相性の良さの予測を加えたもの）

用いるMLアルゴリズムの候補が6、属性の候補が2あるので、それらの組み合わせで合計12の仮想アルゴリズム候補を検討する。4.3節で解説した方法（の発展版）を使って、これら12の仮想アルゴリズムの性能を予測した結果が表3である。この表では、現在使われている「属性を考慮しないトンプソン抽出」と比較して12の仮想アルゴリズムがどれくらいの性能を持つと予測されるかを示している。

結果として、属性集合2とロジスティック信頼上限の組み合わせを用いると、特に大きなさらなる改善が見込まれることがわかった。現在のアルゴリズムと比べ、男性向けキャンペーンで64.6%、女性向けキャンペーンで56.8%、

そして性別問わずのキャンペーンでは63.8%のクリック率改善が見込まれる。この結果を踏まえ、現在「属性集合2とロジスティック信頼上限」を融合したアルゴリズムを実際にサービスに組み込み中である。

表3 よりよいアルゴリズムを求めて

数値は各仮想アルゴリズムの予測性能と現在使われているトンプソン抽出アルゴリズムの性能の比を表している。

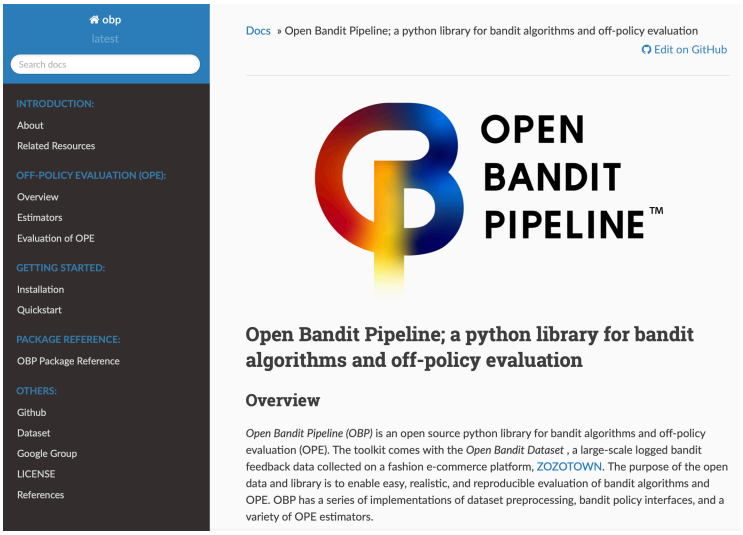
赤の太字は最良のもの、**青の太字**は2番目に最良のものを示している。

Counterfactual Policies		Campaigns		
Algorithms	**Context Sets**	**All**	**Men's**	**Women's**
LOGISTIC ϵ-GREEDY ($\epsilon = 0.01$)	CONTEXT SET 1	0.9124 [0.8617, 0.9617]	0.7230 [0.6788, 0.7643]	0.9451 [0.8961, 0.9965]
	CONTEXT SET 2	0.8355 [0.7867, 0.8840]	1.1587 [1.1058, 1.2128]	1.1929 [1.1425, 1.2438]
LOGISTIC ϵ-GREEDY ($\epsilon = 0.05$)	CONTEXT SET 1	1.1074 [1.0049, 1.2121]	0.8957 [0.8528, 0.9443]	0.9991 [0.8987, 1.0940]
	CONTEXT SET 2	**1.2221 [1.1254, 1.3269]**	1.3359 [1.2324, 1.4345]	1.2195 [1.1367, 1.2979]
LOGISTIC ϵ-GREEDY ($\epsilon = 0.1$)	CONTEXT SET 1	0.8951 [0.8609, 0.9300]	1.0193 [0.9679, 1.0658]	0.7639 [0.7051, 0.8163]
	CONTEXT SET 2	1.0979 [1.0185, 1.1805]	1.2165 [1.1451, 1.2869]	**1.5253 [1.3731, 1.6912]**
LOGISTIC TS	CONTEXT SET 1	1.0162 [0.9595, 1.0767]	1.0094 [0.9419, 1.0822]	1.1996 [1.1271, 1.2945]
	CONTEXT SET 2	1.0933 [0.9575, 1.2091]	**1.4192 [1.3046, 1.5415]**	1.2064 [1.1179, 1.2920]
LOGISTIC UCB ($\alpha = 0.1$)	CONTEXT SET 1	1.2184 [1.0945, 1.3572]	1.1222 [1.0647, 1.1857]	0.8889 [0.8360, 0.9422]
	CONTEXT SET 2	**1.6381 [1.3333, 2.0067]**	**1.6459 [1.5257, 1.7685]**	**1.5676 [1.4307, 1.7049]**
LOGISTIC UCB ($\alpha = 1.0$)	CONTEXT SET 1	0.5000 [0.2873, 0.7243]	0.9535 [0.8659, 1.0401]	0.7823 [0.6471, 0.9043]
	CONTEXT SET 2	0.4763 [0.2667, 0.6957]	1.1306 [0.8896, 1.3026]	0.9141 [0.7386, 1.0664]

5.4. オープンデータとソフトウェア

この性能評価で用いたデータはオープンデータ Open Bandit Dataset として公開した。合わせて、このデータを用いて上記のようなオフライン政策外評価を実行するためのコード群のパイプラインをオープンソースソフトウェア Open Bandit Pipeline として公開した（図5）。GitHub 上の以下の URL にすべての情報がまとまっている：https://github.com/st-tech/zr-obp　ソフトウェアの利用法や ZOZOTOWN 上でのサービス実装のエンジニアリング構成などの詳細は株式会社 ZOZO テクノロジーズ・ZOZO Research Tech Blog(2020a, b)にまとまっている。

図5　オープンソースソフトウェア Open Bandit Pipeline



このようなデータの公開は世界を見渡してもほとんど前例がない。これまでの強化学習やバンディットに関する政策外評価の研究は、人工生成されたシミュレーションデータか非公開の企業私有データで行われ、実世界応用でどのような成果が出るか研究者たちが同じ土俵で競争し評価し合うことは難しかった。バンディット・アルゴリズムを ZOZOTOWN のような大規模サービス上に実装して構築した実データを公開することで、バンディットアルゴリズムの評価・設計・予測に関する企業の壁を超えたオープンイノベーションに貢献したい。

6. 公共政策への応用

最後に、すでに議論した機械学習以外にも、私たちの提案する手法は多くの公共政策分野に応用できる。たとえば多くの公立学校選択制や大学入試制度は、割当アルゴリズムを用いて誰がどの学校への入学権を得るかを決めている。有名な Gale と Shapley の受入保留（Deferred Acceptance）アルゴリズムがその代表例である(安田 10)。 こういったアルゴリズムは、生徒の学校に対する選好順位や、生徒が学校で持つ優先権などの情報を用いて割り当てを決める。たとえば生徒がある学校の近所に住んでいたり、その学校にすでに兄弟姉妹が通っていると優先的に入学できることがある。そういった配慮が典型的な優先権だ。こうした生徒iの選好や優先権に関するデータが理論におけるX_iに相当する。X_iに基づいて、どの学校に入学権を得られるかというZ_iが決まる。このような学校選択・大学入試アルゴリズムも私たちの枠組みで分析でき、様々な学校に入ることが生徒の将来に与える影響を測ることができる。この発想を用いて、たとえば Abdulkadiroğlu et al 17a, 17b, 19, Narita 20 は米国のチャータースクールや有名進学校が将来の成績に与える効果を測定した。

他にもオークション(Kawai and Nakabayashi 14, Chawla et al 17)などの中央集権的資源配分制度も私たちの枠組みのMLアルゴリズムの例と見なせる。さらに、生活保護や米国における医療保険、そして新型コロナウィルス騒動後の経済対策として話題になった雇用調整助成金や持続化給付金のように、一部の家庭・個人・企業のみを対象にした公共政策制度では、観察できる属性に公式を当てはめて各家庭・個人・企業が受益資格を持つかを決めることが多い(Currie and Gruber 96, Cohodes et al 16, Brown et al 17)。こういった受益資格を持つかどうかを決める規則も私たちの枠組みのMLアルゴリズムの例と考えられる。したがって、この論文の手法を用いることで、生活保護、医療保険、雇用調整助成金や持続化給付金が政策受益者の将来の経済状態に与える影響を測ることができる。表４にはこのような政策への応用例をまとめた。

表４　アルゴリズムに基づく公共政策意思決定の例

	アルゴリズムが用いる変数（X）	**アルゴリズムの意思決定（Z）**	**結果変数（Y）**	**アルゴリズム例**
学校選択制・中央集権入試	家庭の学校への選好、学校での優先権	学校への割り当て・入学権	将来の成績や収入など	受入保留アルゴリズムなどの割当アルゴリズム[安田編 10, Abdulkadiroğlu et al 17a, 17b,19, Narita 20]
オークション	入札者の入札額	入札者が落札したか	入札者の将来の経済パフォーマンス	オークション・アルゴリズム[Kawai and Nakabayashi 14, Chawla et al 17]
雇用調整助成金や持続化給付金などの資格判断	企業や家庭の経済状態や家族構成	受益資格があるかどうか	将来の経済・健康状態	受益資格決定規則[Currie and Gruber 96, Cohodes 16, Brown et al 17]

7.　アルゴリズム化する世界は巨大な実験室

古くから、機械学習を含む人工知能の限界は心を持てないことだと言われてきた(Searle 80, Dreyfus 92)。だが、観察できない心を持つ人間と違い、観察できるデータだけに基づいて選択する貧しいアルゴリズムだからこそ、機械学習は自然実験という恵みを生み出すのである。機械学習アルゴリズムを用いた意思決定が行われるようになった現在の

世界には、アルゴリズムが生み出した自然実験が無数に眠っている。アルゴリズム駆動世界は宝が眠る天然の実験室なのだ。

自然なアルゴリズムの運用により自然実験を得られるということは、次のような改善手続きを可能にする。すなわち、「アルゴリズムの運用により生成されたデータを用いて、アルゴリズムの性能評価・改善を行い、更新されたアルゴリズムを運用することでよりよい意思決定をしつつ新たなデータの生成も行う。そして新たなデータを用いてアルゴリズムを再び評価・改善する・・・」といった運用→データ生成→評価・改善の流れである。この「自然実験を用いた 21 世紀型カイゼン」を繰り返せば、費用や時間がかかるうえに性能の悪いアルゴリズムを試してしまう危険がある人工的 RCT（ランダム化実験）に頼らずとも、優れた意思決定をしていくことが可能になる。この論文では、ファッション EC サイト ZOZOTOWN を用いてその可能性を社会実装した。

今日、人々はデータ分析装置としての機械学習を振りまわし持てはやす。だが、機械学習には同じくらい大事なもう 1 つの役割がある。データ「生成」装置としての役割だ。機械学習がどのような意味で価値あるデータ生成装置なのか、そのデータをどう分析すればどんな情報をとりだせるのか、さらなる探究が待たれる。

参考文献

Abdulkadiroğlu, A., Angrist, J. D., Narita, Y. and Pathak, P. A.: Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation. *Econometrica*, 85 (5), pp. 1373–1432 (2017a).

Abdulkadiroglu, A., Angrist, J. D., Narita, Y., and Pathak, P. A.: Breaking Ties: Regression Discontinuity Design Meets Market Design. Working Paper (2019)

Abdulkadiroğlu, A., Angrist, J. D., Narita, Y., Pathak, P. A. and Zarate, R.: Regression Discontinuity in Serial Dictatorship: Achievement Effects at Chicago's Exam Schools. *American Economic Review*, 107 (5), pp. 240–245 (2017b).

Amat, F., Chandrashekar, A., Jebara, T. and Basilico, J.: Artwork Personalization at Netflix. *Conference on Recommender Systems (RecSys)*, pp. 487-488 (2018).

Brown, D., Kowalski, A. E. and Lurie, I. Z.: Long-Term Impacts of Childhood Medicaid Expansions on Outcomes in Adulthood. NBER Working Paper No. 20835 (2017).

Bundorf, K., Polyakova, M. and Tai-Seale, M.: How Do Humans Interact with Algorithms? Experimental Evidence from Health Insurance. NBER Working Paper No. 25976 (2019).

Chapelle, O. and Li, L.: An Empirical Evaluation of Thompson Sampling, *Advances in Neural Information Processing Systems (NIPS)*, pp. 2249-2257 (2011).

Chawla, S., Hartline, J. D., & Nekipelov, D.: Mechanism Redesign. *arXiv preprint arXiv:1708.04699* (2017).

Cohen, P.Z., Hahn, R.W., Hall, J., Levitt, S.D., and Metcalfe, R.: Using Big Data to Estimate Consumer Surplus: The Case of Uber, NBER Working Paper 22627 (2016).

Cohodes, S. R., Grossman, D. S., Kleiner, S. A. and Lovenheim, M. F.: The Effect of Child Health Insurance Access on Schooling: Evidence from Public Insurance Expansions. *Journal of Human Resources*, 51 (3), pp. 727–759 (2016).

Cowgill, B.: The Impact of Algorithms on Judicial Discretion: Evidence from Regression Discontinuities. Working Paper (2018).

Currie, J. and Gruber, J.: Health Insurance Eligibility, Utilization of Medical Care, and Child Health. *Quarterly Journal of Economics*, 111 (2), pp. 431–466 (1996).

Dieterich, W., Mendoza, C., and Brennan, T. Compas Risk Scales: Demonstrating Accuracy Equity and Predictive Parity. Northpoint Inc, (2016).

Dreyfus, H.L.: *What Computers Still Can't Do: A Critique of Artificial Reason*, MIT Press (1992).

株式会社サイバーエージェント・プレスリリース 広告配信 AI のオフライン評価における、不確実性を減少させる手法を提案, https://www.cyberagent.co.jp/news/detail/id=22571 (2018).

Hoffman, M., Kahn, L. B. and Li, D.: Discretion in Hiring. *Quarterly Journal of Economics*, 133 (2), pp. 765–800 (2017).

本多淳也, 中村篤祥: バンディット問題の理論とアルゴリズム, 講談社 (2016).

Imbens, G. W. and Rubin, D. B.: Causal Inference in Statistics, Social, and Biomedical Sciences. Cambridge University Press (2015).

Kato, M., Ishihara, T., Honda, J. and Narita, Y.: Adaptive Experimental Design for Efficient Treatment Effect Estimation: Randomized Allocation via Contextual Bandit Algorithm, *arXiv preprint arXiv:2002.05308* (2020).

Kawai, K., and Nakabayashi, J.: Detecting Large-scale Collusion in Procurement Auctions. Working Paper (2014).

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. Human Decisions and Machine Predictions. *Quarterly Journal of Economics*, 133(1), pp. 237–293 (2017).

Li, L., Chu, W., Langford, J., & Schapire, R. E.: A Contextual-bandit Approach to Personalized News Article Recommendation, *International conference on World Wide Web (WWW)*, pp. 661-670 (2010).

Narita, Y.: A Theory of Quasi-Experimental Evaluation of School Quality. *Management Science* (2020).

Narita. Y.: Incorporating Ethics and Welfare in Randomized Experiments. *Proceedings of the National Academy of Sciences* (2020).

Narita, Y., Yasui, S. and Yata, K: Efficient Counterfactual Learning from Bandit Feedback, *AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 4634-4641 (2019).

Narita, Y. and Yata, K.: Algorithm is Experiment: Machine Learning, Market Design, and Policy Eligibility Rules, Working Paper (2020).

Narita, Y., Yasui, S. and Yata, K.: Off-policy Bandit and Reinforcement Learning, *arXiv preprint arXiv:2002.08536* (2020).

Precup, D.: Eligibility Traces for Off-Policy Policy Evaluation, *International Conference on Machine Learning (ICML)*, pp. 759–766 (2000).

Saito, Y., Aihara, S., Matsutani, M., and Narita, Y.: A Large-scale Open Dataset for Bandit Algorithms. Working Paper (2020).

Searle, J. R.: Minds, Brains, and Programs, *Behavioral and Brain Sciences*, Vol. 3, No. 3, pp. 417-424 (1980).

Sutton, R.S. and Barto, A.G.: *Reinforcement Learning: An Introduction*, Bradford Book (2018).

安井翔太: 効果検証入門, 技術評論社 (2020).

安田洋祐編: 学校選択制のデザイン, NTT 出版 (2010).

株式会社 ZOZO テクノロジーズ・プレスリリース 顧客の真の欲求を見つけ出す 「因果関係も見通す機械学習」に関する共同研究を開始 ～ 開発における経営リスクを最小限に、消費者の隠れたニーズに応える EC デザイン設計を目指す ～, https://press-tech.zozo.com/entry/20191211_zozoreseach (2019)

株式会社 ZOZO テクノロジーズ・ZOZO Research Tech Blog: Off-Policy Evaluation の基礎と ZOZOTOWN 大規模公開実データおよびパッケージ紹介, https://techblog.zozo.com/entry/openbanditproject (2020a)

株式会社 ZOZO テクノロジーズ・ZOZO Research Tech Blog: バンディット アルゴリズムを用いた推薦システムの構成について, https://techblog.zozo.com/entry/zozoresearch-bandit-overviews (2020b)