

# ほとんどすべての機械学習は A/B テストである

## Almost All Machine Learning Algorithms Are A/B Tests

成田 悠輔\*<sub>1</sub>  
Yusuke Narita

矢田 紘平\*<sub>1</sub>  
Kohei Yata

\*<sub>1</sub> イェール大学  
Yale University

Machine learning and other algorithms produce a growing portion of decisions and recommendations. Such algorithmic decisions are unintentional A/B tests since the algorithms make decisions based only on observable input variables. We use this observation to characterize the sources of causal-effect identification for a class of stochastic and deterministic algorithms. Data from almost every algorithm is shown to identify some causal effect. This identification result translates into consistent estimators of causal effects that are easily implemented even with high dimensional data and complex algorithms.

### 1. アルゴリズム化する世界

機械学習アルゴリズムに基づく意思決定が現実世界で数多く行われるようになってきた。たとえば、アマゾン、フェイスブック、グーグル、マイクロソフト、ネットフリックス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(特に、後述するバンディットアルゴリズムなどの強化学習)の手法を応用している。また、ウーバーやリフト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。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。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、アルゴリズム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はこれらの例の一部をまとめたものである。

### 2. 機械学習アルゴリズムは A/B テストを生み出す

機械学習アルゴリズムによる意思決定には、どのような特徴があるだろうか？その一つは、意思決定が観察可能な変数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の変数のみに基づき、表示するコンテンツを選択する。

意思決定が観察可能な変数のみに基づくことから、次のようなことが言えるだろう。意思決定を左右する観察可能な変数の値がほとんど同じ状況において、さまざまな異なる選択が行われているとしよう。すると、各状況における選択は、あたかも実験のように、アルゴリズムが直接使う変数以外の他のあらゆる変数からは独立してランダムに決められたものと考えることができる。つまり、機械学習アルゴリズムによる意思決定は意図せずに発生した A/B テスト(計量経済学の用語を使えば、最終的な意思決定の操作変数)とみなすことができるのである。以下、2 つの具体例を用いて解説しよう。

### 2.1 ランダム化実験を内蔵するアルゴリズム

ウェブサービスの運営者の目的の一つは、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適なコンテンツを表示するためには、さまざま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最適化する手法が、「バンディットアルゴリズム」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「 $\epsilon$ -貪欲法」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対して最良と推測されるもの(クリックする可能性が最も高いと推測されるもの)を  $1 - \epsilon$  の確率で選択する。残りの  $\epsilon$  の確率では、ランダムに選択肢を選ぶ。通常、 $\epsilon$  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属性ごとに推測され、そして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、 $\epsilon$ -貪欲法による選択肢の割り当ては、層別ランダム化実験(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などがある。こういったアルゴリズムを運用することで生成されたデータは、層別ランダム化実験により集められたデータと同じ統計的手法で分析し、因果効果の推定や、まだ使われたことのないアルゴリズムの性能予測をすることが可能である。たとえば、筆者らと株式会社サイバーエージェントの共同研究では、バンディットアルゴリズムにより生成されたデータの分析手法を開発し、同社の広告配信ログデータを用いて、広告画像の選択方法の性能評価を行っている(Narita et al. 2019)。また、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おすすめアルゴリズムなどを作り変えている。

### 2.2 回帰不連続デザインを内蔵するアルゴリズム

ウーバーが提供する UberX は、乗客と自家用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自動車共有サービスである。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。UberX 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が、

連絡先: 成田悠輔, Yale University, 37 Hillhouse Avenue, New Haven, CT, USA, [yusuke.narita@yale.edu](mailto:yusuke.narita@yale.edu)

需要が供給を超過した場合は、「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。

このアルゴリズムが自然実験を生み出すという観点から興味深いのは、まず連続的な上昇乗数を計算し、その小数点以下第二位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変化す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、離散的な料金の変化が見られることとなる。これは、自然実験の一種である回帰不連続デザインとみなすことができる。Cohen et al. (2016) はこのアイデアに基づき、UberX のデータを用いて、料金が乗客のサービス利用に与える効果 (価格弾力性) を推定した。

このように、機械学習アルゴリズムが生成したデータに回帰不連続デザイン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、教師あり機械学習アルゴリズム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。ここから回帰不連続デザインが生まれるのである。

### 3. 機械学習アルゴリズムが生成したデータの分析

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自然実験を作り出すことを見た。その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因果効果を識別・推定できる (つまり、自然実験的な状況が存在する) のか分析してみよう。

#### 3.1 モデル

本文先述の通り、機械学習アルゴリズムによる意思決定の特徴は、観察可能な変数のみに基づくことである。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  $X \in \mathbb{R}^p$  で表すとしよう。議論を簡単にするため、意思決定における選択肢は、ある処置を行うかどうかの 2 択であるとし、その選択を 2 値変数  $Z \in \{0,1\}$  で表す。 $Z = 1$  であれば処置を行ったことを意味し、 $Z = 0$  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関数  $ML$  で表す。

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

$ML$  は、変数  $X$  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  $\epsilon$ -貪欲法の場合、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。金融機関が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。

1 数式で表すと、 $N(x, \delta) = \{x^* \in \mathbb{R}^p: d(x, x^*) < \delta\}$ 。ここで、 $d(x, x^*)$  は  $x$  と  $x^*$  の間のユークリッド距離を表す。

ここで、因果効果を数学的に記述するため、ある個人に対して処置が行われた場合の結果を  $Y(1)$ 、処置が行われなかった場合の結果を  $Y(0)$  とし、その個人にとっての因果効果を  $Y(1) - Y(0)$  で表す。現実では、処置を行うか行わないかのどちらか一方しか生じないため、 $Y(1)$  と  $Y(0)$  のどちらか一方しか観察できない。また、観察される結果変数を  $Y = ZY(1) + (1 - Z)Y(0)$  とする。

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

#### 3.2 因果効果の識別

本文それでは、どのようなアルゴリズムから出てきたデータであれば、因果効果を識別できるだろうか？鍵となるのが、「擬似傾向スコア (Quasi Propensity Score: QPS)」という概念である。変数  $X$  の台 (support) を  $\mathcal{X}$  とする。まず、ある  $x \in \mathcal{X}$  において、 $N(x, \delta)$  を中心が  $x$  で半径が  $\delta$  の  $p$  次元の球体<sup>1</sup>とし、

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

とする。ここで、 $U_{N(x, \delta)}$  は  $N(x, \delta)$  上の一様分布を表す。そして、 $x$  における QPS を、以下で定義される  $p^{ML}(x)$  とする。

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

QPS は  $x$  の付近において処置が行われる平均的な確率と解釈できる。図 1 は、 $X$  が 2 次元 ( $p = 2$ ) の場合の関数  $ML$  と QPS の例を示したものである。

$x$  における QPS が 0 と 1 以外の値であれば、 $x$  の付近において、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し、その選択がほとんどランダムに行われる。結果として、因果効果の識別が可能になる。<sup>2</sup>

**命題 1** 条件付き期待値関数  $E[Y(1)|X]$  と  $E[Y(0)|X]$  が連続であるとし、 $ML$  と  $X$  にいくつかの技術的条件を置くと、以下の 2 つが成り立つ。

- $p^{ML}(x) \in (0,1)$  が成り立つすべての  $x \in \mathcal{X}$  において、因果効果  $E[Y(1) - Y(0)|X = x]$  を識別できる。
- $S$  を  $X$  の部分開集合とする。因果効果  $E[Y(1) - Y(0)|X \in S]$  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。

(a) は因果効果識別のための十分条件、(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 1 つでも存在すれば、データから何らかの因果効果を識別できることがわかる。教師あり学習アルゴリズム、強化学習やバンディットアルゴリズムといった一般的な機械学習アルゴリズムによる意思決定では、この条件が満たされ、因果効果の識別が可能である。

2 ここで、ある因果効果のパラメーターが識別できるとは、その因果効果のパラメーターが  $(X, Z, Y)$  の同時分布から一意に定まることをさす。

3.3 因果効果の学習・推定

最後に、アルゴリズム $ML$ が生み出したデータをどのように分析すれば因果効果を推定できるかを説明しよう。大きさ $n$ のデータ $\{(X_i, Z_i, Y_i)\}_{i=1}^n$ が与えられたとする。まず、 $\delta$ を小さい値に設定し、それぞれの個人 $i$ について $p^{ML}(X_i; \delta)$ を計算する。 $p^{ML}(X_i; \delta)$ は、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 $p^{ML}(X_i; \delta) \in (0,1)$ となるものを使って、以下の回帰式を最小 2 乗法で推定する。

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

**命題 2** こうした得られた推定量  $\hat{\beta}_1$  は、いくつかの技術的条件の下で、因果効果の一致推定量となる。

4. アルゴリズム化する世界＝自然実験の宝庫

古くから、機械学習を含む人工知能の限界は心を持っていないことだと言われてきた(Searle 1980)。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く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はずだ。

アルゴリズムの運用により自然実験的データ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。そしてその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 A/B テスト(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。

参考文献

[Cohen 2016] Cohen, P., Hahn, R., Hall, J., Levitt, S., and Metcalfe, R.: Using Big Data to Estimate Consumer Surplus: The Case of Uber, NBER Working Paper 22627 (2016).  
[Narita 2019] Narita, Y., Yasui, S., and Yata, K.: Efficient Counterfactual Learning from Bandit Feedback, *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, pp.4634 – 4641 (2019).  
[Searle 1980] Searle, J.: Minds, Brains, and Programs, *Behavioral and Brain Sciences*, 3(3), pp. 417 – 457 (1980).

最終意思決定主体	アルゴリズムが用いる変数(X)	アルゴリズムが決定するもの(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
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、上昇乗数	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)
裁判官	被告人の犯罪歴、年齢等の属性	再犯確率スコア、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

表 1 機械学習アルゴリズムに基づく意思決定の例

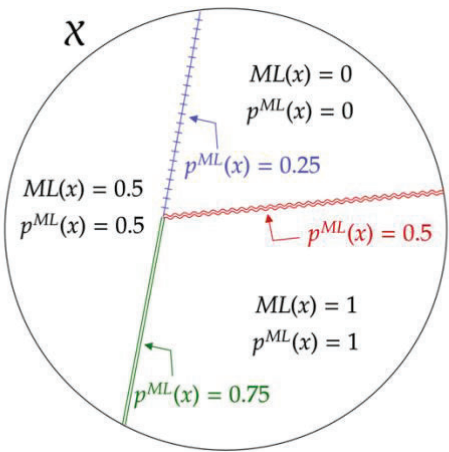


図 1 関数 $ML$ と局所傾向スコア $p^{ML}$ の例

$x$ が $ML$ の値により三つの部分に分割される。それぞれの部分の内側にある $x$ においては、 $ML(x)$ と $p^{ML}(x)$ が一致する。ある二つの部分の境界線上にある $x$ においては、 $p^{ML}(x)$ は二つの部分の $ML$ の値の平均となる。