特　集　DX時代のCAE最新手法　20234569

機械学習は真の理解や発見に寄与できるか*

Can Machine Learning Contribute to True Understanding and Discovery?

瀧川　一学[1)]
Ichigaku Takigawa

With the spread of AI technology in recent years, there has been a growing expectation to utilize machine learning for scientific understanding and discovery. This article aims to first explain that machine learning itself brings neither understanding nor discovery in principle, and then discuss what is important to obtain "understanding" and "discovery" from data.

KEY WORDS
Common Infrastructure, Computer Aided Engineering
Machine Learning, Machine Discovery F2

1　はじめに

近年，多種多様な分野で機械学習の利活用が進んでいる．機械学習は「AI」として一括りにされがちなさまざまな情報技術のごく一つにすぎないが，深層学習を筆頭に現在のAIブームの屋台骨ともいうべき基幹技術となった．顔認識や音声認識，ネットの検索や広告，ネットショッピング，ニュース推薦，機械翻訳・英文校正など，私たちの身近でも機械学習を利活用したサービスは当たり前となっている．

機械学習は「データを予測に変える」技術である．写真に写った顔が誰の顔なのか，録音された音声が何と発話しているのか，私たちはほとんど無意識的にわかるが，その仕組みは私たち自身にもよくわからない．原理と手順を明示化できなければコンピュータプログラムも作れない．経験と勘の領域だと諦めてきたこうした経験則的なタスクは，機械学習が最も得意とするものである．本当の仕組みはよくわからなくても，とにかく大量の見本例データを機械学習にかければ，実用に耐えるレベルで顔認識や音声認識はできる．多くの情報がコンピュータ上に蓄積されさまざまなデータが溢れる現在，こうしたデータの機械学習によって「理解」や「発見」が得られるのでは，という期待は高まっている．

本稿の第一の目的は，私たちの素朴な期待に反して，機械学習そのものは「理解」も「発見」ももたらしてはくれないことを示すことである．本稿の第二の目的は，「理解」や「発見」を妨げる主たる困難さを整理し，データから「理解」や「発見」を得るためには何が重要か，を再確認することである．

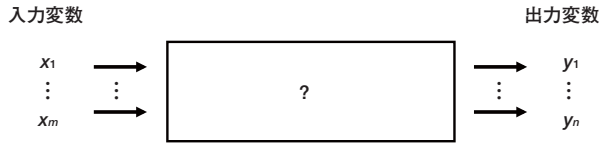
2　予測と理解の非両立性

決定木やランダムフォレスト法の産みの親であるレオ・ブライマンは，晩年に「Statistical Modeling: The Two Cultures」というタイトルの非常に有名なポジションペーパー(1)を発表した．伝統的な統計モデルよりもアルゴリズム的な予測手法を使っていくべきというかなりラディカルな主張であり，いまだに賛否渦巻く議論の種になり続けているが，少なくとも深層学習や決定木アンサンブル法などアルゴリズム的手法が実用領域を席巻する現在を予期させる先見性があった．2021年はこの論文の出版から20周年の記念年であり，Observational Studies誌が現在の視点でこの論文を振り返る特集号(2)を企画した．

ブライマンのかつての共同研究者を筆頭に多数の専門家が寄稿しているが，その中の一人，因果推論の大家ジューディア・パールは次のように述

＊ 2023年6月26日受付

1)【所属1】
京都大学 国際高等教育院データ科学イノベーション教育研究センター
（606−8315　京都市左京区吉田近衛町69　近衛館302号室）
E-mail: takigawa.ichigaku.8s@kyoto-u.ac.jp
【所属2】
北海道大学 化学反応創成研究拠点
（001−0021　札幌市北区北21条西10丁目）



① **予測**：入力変数から出力変数の値を予測する
② **理解**：データの背後の真の入出力関係の情報を得る

図1　統計的モデリングの二つの目的

べている．「ブライマンの主張の要点は次のとおりである．もし予測が目的なら予測そのものに注力すべきで，モデルが背景現象を表現しているなどという幻想は忘れ去るべきである」．ブライマンは統計的モデリングの目的を「予測」と「理解」の二つに明確に分けてみせた（**図1**）．そして前者に全力投入することで，認知や言語の理解なしに音声認識や画像認識や機械翻訳を確かに実用レベルに引き上げることができたのである．高い予測性能を得た代償として「理解」は失われた．これらの機械学習モデルは数千万個やら数千億個やらの天文学的な数のモデルパラメータで制御され，その挙動は「理解できる」とはとてもいえない代物となった．パールが続けて語るとおり，ブライマンが「理解」のために用意した組合せ変数重要度や部分依存度プロットなどの手法は現在もよく使われる手法ではあるが，原因と結果の「理解」の手段としてはあまりに間接的である（理解とは何か？という認知科学上・科学哲学上の長い議論はここでは割愛する）．

端的にいえば，現行の機械学習は「予測」は高精度でできても，なぜ予測できるのか「理解」を与えてはくれない．そして，「予測」ができれば十分な状況は思いの外たくさんあり，その中で十分なデータが取れる問題に機械学習を使うのが基本である．冒頭に挙げた私たちの身近にサービスとして実現している事例は，いずれもこのケースに該当する．というよりも，AIブームを牽引してきた主要なビッグテックが念頭に置いていたタスクはことごとく予測が目的であった．推薦した商品を買ってくれれば理由はどうでもよいし，出した広告をクリックしてくれれば理由は問わないし，ユーザが求めるページを検索結果の最初のほうに出せれば理由は不問である．

一方，「理解」を目的とするならば，明らかに「予測できればOK」ではあり得ない．機械学習でも使ってみようというとき，機械学習の目的

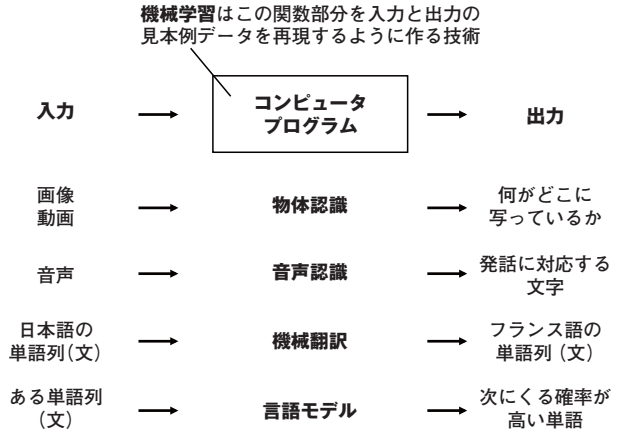


図2　機械学習＝新しいプログラミング

（予測）と私たちが求めるもの（理解）が多くの現場でそもそもずれてしまう．ニュースを賑わす華々しいAI事例で使われるような最先端の機械学習を自分で集めたデータに使ってもほとんどの場合「理解」は得られない．

3　理解とモデルの解釈・説明

3.1　機械学習モデルの解釈・説明

とはいえ，予測が当たるのなら対象現象の一面は捉えているはずであり，その機械学習モデルから何か有益な情報を抽出できないか，と考えるのは自然である．解釈可能な機械学習や説明可能なAI（XAI）として，解釈や説明のためのさまざまな手法が提案されてきたが，これらの手法によって「理解」の問題が解決されるわけではない．その原因は，機械学習や解釈手法の内容云々よりも，機械学習を訓練するデータの取り方や機械学習モデルで用いる入力変数の設計や選別など，手法を適用する前段階にある．

多くの場合，現行の機械学習は**図2**のように，入力変数から出力変数を予測する技術として使われる．入出力の見本例データから挙動を定める新しいプログラミングともいえる（Software 2.0）．

物理学や数理科学の理論モデルなどでは，ある数式が実験データをよく再現できる場合には変数の間に成立する法則として「理解」される．ところが，これは機械学習には通常当てはまらない．機械学習モデルが表す数式は「データをよく再現できるように」作られたものであり，データを再現できるのは当たり前である．逆に数式の内容はデータを与えて初めて定まる．その獲得された数

式において，入力変数と出力変数の間にどのような関係性があるのか，入力変数のうちどれが支配的なのか，どの変数とどの変数の間に関係があるのか，など関心となる入出力関係を「理解」するための手がかりを得る技術がさまざまに考案されてきた．これは，ある特定の入力に対してなぜ機械学習が特定の値を予測として出力するのかの過程の理解を含み，モデルの解釈や説明と呼ばれる．

しかし，「真実はいつも一つ」でも，「解釈」や「説明」は人の数だけ，手法の数だけありえる．事実，私たちはふつう完全な情報を手にすることはできず，不可知な部分は部分的な情報から勝手に補完するものである．その補完分の自由度はデータがないので恣意的になり，データがないゆえに是非の評価もできない．したがって，同一のデータ，同一の機械学習モデルに対して，異なる解釈・説明手法を使えば異なる解釈や説明が出てくる悩ましい事態は自然に予見できるのではないだろうか．

3.2　良い機械学習モデルの多重性（羅生門効果）

さらに「理解」を阻む壁となるのが，用意した検証データで良い予測性能を示す機械学習モデルは無数にあるという事実である．機械学習コンペティションでも上位解法のアプローチは多様で異なるのに，予測性能はほぼ同程度で団子状態になりがちである．ブライマンはこの現象を黒沢映画にちなんで「羅生門効果」と呼んだ．日本人なら芥川の原作小説から「藪の中効果」と呼んだほうがしっくりくるかもしれない．殺人事件の容疑者として連れて来られた複数の被告人は，それぞれが非常に尤もらしく聞こえるまったく違う話を語り，真実は「藪の中」となる．

この本質的な原因は観察できる証拠（データ）が有限個だからである．現行の予測精度の高い機械学習モデルがもつモデル自由度は巨大であり，私たちにとって「ビッグデータ」に思えても，検証用データが有限個である限り，同程度の予測性能を示すモデルは無数にあり得る．点に曲線を当てはめるとき，自由度がとても大きければ有限個の点をすべて通る曲線は無数に存在しうるのである．つまり，図2のコンピュータプログラムは一意に定まらず，むしろ常に同程度に良いものが多数あり得る．

結果として，モデルを「理解」しようとするとき，何重にもこの「多重性」に悩むことになる．手元の訓練データで，深層学習モデル・決定木アンサンブルモデル・カーネル法モデルを初期設定をいろいろ変えながら学習してみたら，だいたい同程度の予測性能となるモデルがいろいろ得られた．そして，さらに各々に対してさまざまなモデル解釈・モデル説明の手法を適用し，多種多様な解釈や説明が得られた．さて，分析者はどれを「理解」として採用すればよいのだろうか．

4　理解と交絡バイアス

4.1　入力されない情報をすべて無視できるか？

図1や図2からもう一つ明確なことがある．それは，機械学習の予測は入力変数として取り上げなかった要因を考慮する仕組みを一切もたないことである．あくまで与えられた1セットの入力変数を関心の出力変数に変換する入出力関数のモデル化が機械学習である．つまり，機械学習は入力変数として取り上げた変数以外のありとあらゆる要因と情報を無視する．

したがって，どのような変数を入力変数として採用するかは，手法のアルゴリズムそのものよりも「理解」に影響を及ぼしうる．少なくとも「解釈」や「説明」の対象にしたい因子は入力変数としてモデルに入れておく必要がある．一方で，そのような因子があらかじめすべてわかっていることは現実的には想定しづらく，また，結果に影響する因子のすべてをデータとして測定できるとも限らない．しかし，もし重要な因子を取りこぼせば，採用した入力変数をなんとかこねくり回して見かけ上の擬似相関を作り出すだけに終わる．アンケート質問10問の回答から血液型を当ててみせる場合のように，入力変数と出力変数の間に因果関係があろうがあるまいが，相関さえあれば予測はできるからである．

このことは，統計学の教科書では「相関関係は必ずしも因果関係を意味しない」という金言に集約される．因果関係とは「原因」と「結果」の関係であり，科学的理解とは通常は因果関係の理解のことである．ところが，因果関係は直接計測できない．データとして手に入るのは相関関係だけである．機械学習は変数間の相関関係に基づく予

測技術であり，それが何らかの因果関係を捉えている保証はまったくない．

因果関係を相関関係から峻別するための鍵となるのが「操作」である．因果推論の分野には「操作なくして因果なし」という格言がある．つまり，変数 X と Y が因果関係，つまり「原因」と「結果」の関係，をもつならば，X の値を変化させればそれに応じて Y の値も変化するはず，ということである．体重が重い人は身長も高い傾向があるが，体重を増やしても身長が伸びるわけではないので，体重と身長の関係は相関関係ではあるが因果関係ではない．統計学，特に，統計的因果推論の分野では，どういう条件下でなら因果関係をデータから推定できるのかについて長らく議論されてきた．なお，「因果推論」というときは想定する因果関係がすでにあって，その効果の大きさをデータから推定することが目的であって，何の仮定もおかず因果関係を大量のデータから自動的に見出したり，与えられた一群の変数の間にどのような因果関係があるのか同定したりできるわけではないことを念のため補足しておく．

4.2　データからの因果関係の推定

ここでは因果推論の難しさの要となる「交絡」と，それを乗り越える奥義「ランダム化比較試験」についてだけ触れておく．簡単のため，「薬を飲む / 飲まない」「病気が治る / 治らない」の 2 値変数の因果関係を考える．まず，「薬を飲んだら病気が治った人が 10 人いた」だけではダメなことに注意したい．対照比較がないからである．薬を飲もうが飲むまいが，別の理由で治った可能性がある．したがって，薬を飲んだグループ（処置群）と薬を飲まなかったグループ（対照群）の間で病気の回復割合を比較することになる．

しかし，この二群の「正しい比較」は予想以上に難しい．例えば，処置群は病院にいた高齢者で，対照群は一般検診に来た大学生だったらどうであろうか．薬にかかわらず高齢者は病気が治りづらく，処置群の回復割合は本当に知りたい回復割合より低くなる．また，医師は目の前に病気の患者がいて有望そうな薬があるのに処方しないでよいのだろうか．病気がひどい人ほど薬を出してしまわないであろうか．この場合，処置群には重症者の割合が高く，やはり回復割合を低く見積もってしまう．

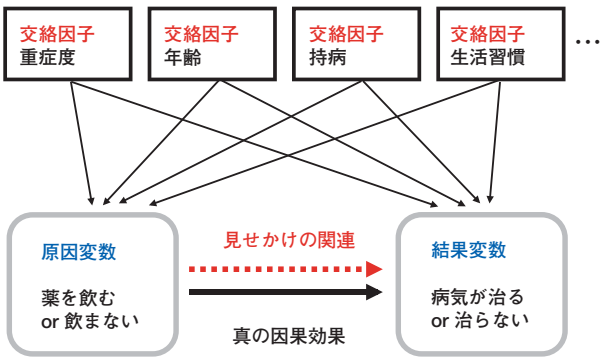


図 3　交絡と見かけ上の相関

このような「年齢」や「重症度」など，原因変数（薬を飲む / 飲まない）と結果変数（病気が治る / 治らない）の両方に影響する要因を交絡因子といい，この状況のことを交絡という（**図 3**）．正しく因果効果を見積もるには，「薬を飲む / 飲まない」以外のすべての背景要因を揃えておく必要がある．しかし，人は一人一人違い，生活習慣，持病，喫煙，飲酒，などなど考えうる交絡因子だけでも無数にある．交絡因子がすべてわかっていると仮定するのは非現実的であり，わかったとしてもデータとして取れるとは限らない．

この無理設定を乗り越える奥義こそが「ランダム化比較試験（RCT）」である．つまり，サイコロを振って偶数なら処置群，奇数なら対照群というように明確な意思をもってこれをランダムに割り付ければ，未知のもの・未計測のものを含むあらゆる交絡因子に悩まされない．原因変数の「薬を飲む / 飲まない」の割り付けとあらゆる他の要因とがランダム化によって独立になるからである．逆にいえば，これができない場合はデータに結果の解釈を歪める恐れのあるバイアスが入っていないことを保証できないため，厳密な因果推論はできない．統計学では，このようなランダム化ができない研究デザインを観察研究，そのデータを観察データという．

4.3　「理解」を得るために

機械学習を活用する場面では，基本的に訓練データは観察データになりやすい．すでに存在するデータをいろいろ集めてきて機械学習の訓練データとするため，変数の値をランダム化することはもうできない．「予測」だけなら問題はないが，「理解」を求めるなら，これは潜在的な交絡が存在し，変数間に見かけ上の連関が生じうるリ

スクを意味する．しかし，ランダム化比較試験は一般にコスト・手間・時間あるいは現実的な制約で難しい場合が多い．放射線の有害性を調べるため，100 人連れてきて，無作為に選んだ 50 人に放射線を当てる人体実験はできない．教育効果を調べるため，100人のうち，50人に良い教育を与え，残り 50 人は教育放棄することもできない．実際の問題解決では取れないデータは多いのである．

材料開発や実機実験など物理・化学・工学では実験研究が主であり，観察研究のようなバイアスはないと思うかもしれないが，機械学習を適用したいような対象はしばしば機序が未知の複雑系であり，すべての背景要因が制御されていると仮定できないことも多い．ここでもう一度図 2 を見ていただきたい．今検討中の系において，入力変数として取り上げていない「すべての」要因について統制されているといえるだろうか．もし一つでもバラバラな要因があるならば，それが交絡因子となり入力変数と出力変数の間に見かけ上の連関が生じるため，どんな手法で獲得された機械学習モデルを解釈・説明しても，それは交絡バイアスによって歪められた見かけ上の関係にすぎないかもしれないのである．

したがって，「機械学習モデルにどの変数を入れるか」を対象現象に関する知識を総動員して注意深くデザインする必要がある．データからの因果推論には事前知識や仮説は必須である．その上で手法の多重性も想定し，複数の分析手法を用いて多角的に情報を積み重ねていくしかない．無頓着に成り行き任せでかき集めたデータから，自動的に関心対象のメカニズムに関する情報を次々と炙り出してくれる魔法のような技術は，理論上存在しえないのである．

5 機械学習から機械発見へ

一方で，交絡因子になりそうな変数はすべて機械学習モデルに入れておく，という発想も自然である．変数に入れておけば交絡因子で説明できるならそれが予測に用いられるため，交絡因子の影響を取り除いて原因変数と結果変数の関係を評価できる見込みが残る．この理屈は正しいのであるが，このアプローチは例えば以下のような別の理由で問題を起こす．

「理解」と並んで科学の目的とされるのが「発見」である．例え，理由がわからなくても，高性能で安全安価な充電池ができたり，治療法がない病気の治療薬ができたり，人工光合成を実現する触媒ができたりすれば万々歳である．「今までにないものを見つける」を法則にまで敷衍すれば「理解」もまた「今までわからなかったこと」の「発見」ともいえる．

ここで再度，図 2 を見ていただきたい．機械学習はこの関数部分を「入出力の見本例データ」から決める．このときに使われるデータを訓練データと呼ぶ．訓練データで予測が合うようにモデルを作るので，その予測がどれだけ当たるかをテストする検証用のデータも別に必要となる．通常の機械学習では訓練データも検証用データも，運用時に実際に入力されるデータ（テストデータ）を模擬したものになっている必要がある．その仮定ですべての手法が作られている．ところが，私たちは多くの場合に「発見」を期待してしまう．機械学習に「今わかっていないこと」，例えば「今存在する材料よりも優れた材料」を期待してしまうのである．機械学習は訓練データに合うよう入出力を決める経験則にすぎない．したがって，過去の延長線上で予測を行うだけであり，「過去にない」ような「発見」をもたらしてはくれないのである．

「発見」を求める場合，もはや「機械学習」ではまったくない．テストデータは訓練データとも検証用データとも違って，文字通り，可能なありとあらゆる入力となり，その空間を効率的に「探索」する問題になる．だとすれば，訓練データも検証用データも「可能となる入力」の全域で万遍なく取得しておく必要がある．したがって，交絡因子になりそうな変数すべてを入れて考えるなら，組み合わせ的に巨大な訓練データが必要だということになる．各変数ごとに 3 水準（大中小）の値を取るにしても，10 変数で 3^{10} ×実験の繰り返し数，100 変数で 3^{100} ×実験の繰り返し数になり，非現実的なデータ規模になる（3^{100} は約 5,154 載）．現行の機械学習は数百万の入力変数を数億個のデータで学習できるが，それは「入力全域での関数近似」など求めない設定だから可能なだけである．

そのうえでなお，個人的には「理解」より「発見」のほうが機械学習を活用する余地が大きいと

考えている．広大な候補の「探索」問題だと割り切れば，まったくのランダムや成り行き任せで探索するより良い方法がありそうである．データが取れたところは機械学習で近似して，有望そうな候補を絞る補助にしたり，まだ不確実性が高くデータを取らなくてはならない領域の同定に活用したりできるからである．新しいことの学習と発見には「今のところ有望そうなものの改良」（知識の活用）と「まだ見ぬ新たなシードの開拓」（知識の探索）の両面が必要である．これは道具は違えど，前世代の AI ブームで追求されたテーマであり，それに倣って私は「機械発見」の問題と呼んでいる．詳細は他の解説[(3)-(5)]に譲るが，データを取りながら学習し，広大な候補空間を効率的に探索する古くて新しい問題である．新しい材料や新しい化学反応の発見だけではなく，囲碁やゲームプレイや行列計算やソートなど，アルゴリズムそのものの探索も機械発見の主題である．材料にせよアルゴリズムにせよ，人力で見つけられてきたさまざまな分野の人智の結晶のような既知の最良解を「機械発見」技術によって効率的に超えていけるのか，今後の研究動向にも注目していただきたい．

フェイス

私は機械学習の研究者ですが，コロナ禍以前の 2019 年 10 月に「AI・IoT 時代のデータ利活用による理解と発見」をテーマに開かれた第 35 回関東 CAE 懇話会で，本原稿のタイトルで講演させていただいたことをきっかけに今回の執筆の機会をいただきました．今では講演時点よりも機械学習・AI は一般の話題にのぼるようになり，CAE や材料開発に限らず，自動車技術全般や周辺業務でも機械学習や AI に接する機会は増えつつあるのではと思います．本稿が機械学習を活用される一助となれば幸いです．

瀧川一学

参 考 文 献
（1） Leo Breiman：Statistical Modeling: The Two Cultures（with comments and a rejoinder by the author）, Statist. Sci., Vol. 16, No. 3, 199−231 (2001)
（2） Special Issue: Commentaries on Breimen's Two Cultures paper, Observational Studies, Vol. 7, No. 1 (2021)，https://muse.jhu.edu/issue/45147
（3） 瀧川一学：機械学習と機械発見：データ中心型の自然科学の教訓と今後，日本結晶成長学会誌，特集：機械学習・AI は結晶成長研究をいかに変えるか？，Vol. 49，No. 1，49-1-01 (2022)
（4） 瀧川一学：表現と介入：機械学習は化学研究の「経験と勘」を合理化できるか？ 化学と教育，ヘッドライン：AI が開く新たな化学領域，Vol. 70，No. 3，122−125 (2022)
（5） 瀧川一学：人工知能基本問題研究会（FPAI），人工知能，Vol. 34，No. 5，603−611 (2019)