

「統計」と「機械学習」の違いが示唆する 保育所保育において目指すべき IT 導入の付加価値

	統計 Statistics	機械学習 Machine Learning
種類 Type	<ul style="list-style-type: none"> ●記述統計 Descriptive Statistics ・データの特徴や傾向の表現 ●推測統計 Inferential Statistics ・一部データから全体を推定 	<ul style="list-style-type: none"> ●予測 Prediction ・将来予測に必要な推論モデルの生成 ●分類・識別・判断 Classification/Identification/Decision ・分類/識別/判断に筆意用な推論モデルの生成
誰 Who	●人が傾向を発見し、人が統計値を用いて説明を支援	●機械が傾向を発見し、機械が予測や分類・識別・判断
志向 Approach	<ul style="list-style-type: none"> ●変動要因・原因の洗い出しや影響度に力点を置く (一般的に理由付けのホワイトボックス化) ●予測より説明 ●信憑性を疑う(性悪説) ●数学中心 ●サンプリングの場合は偏りのないサンプリング ●過学習はない (妥当な全データを使うことが問題にならない) 	<ul style="list-style-type: none"> ●最高の予測が重要で予測結果が良ければ、何が変動要因・原因かや影響度は問わない。 (一般的に理由付けのブラックボックス化(ランダムフォレストや決定木等)) ●説明よりは予測 ●信憑性は関係ない(性善説) ●ロジカルなデータ処理のコンピュータサイエンス中心 ●サンプリングがどうであるかが重要な起点ではない ●過学習がある (妥当な全データを使うことを一般的に問題視する)
良い事例 Example	●社会科学の説明、推定	<ul style="list-style-type: none"> ●自然科学予測(天気予報) ●ロボット処理

(出典)「統計」と「機械学習」の違いの整理で多くの事業会社で「機械学習」が使えない理由が視えてきた！

<https://qiita.com/mnoda/items/ee074d6614cd0b40525d>

AIとか機械学習という言葉がニュースなどを通じて、耳目に達することが多くなっている。とはいえ、AI、人工知能、機械学習というと、グーグルのような最先端の大企業での遠い話、あるいは、むしろさん臭い SF のような話で、眉に唾して話しを聞かないといけないうイメージを持っている方も多いであろう。過去に何度か人工知能ブームというものがあったものの、ものにならなかった歴史があるのも事実であり、そのような感想を持つのは、ある意味では健全であると言えよう。

ただ、この人工知能という言葉は、その意味するところが広すぎる。目下、その発達、進化に注目が集まっているデータ分析の技法については、機械学習、マシーンラーニング(ML)という言葉の方が、よりの確な表現となる。

端的に言ってしまえば、この機械学習という技法は、統計学の応用である。勿論、従来の統計学のままということではないが、データの集まり、多くの場合は、数字に置き換えたデータのパターンから傾向や関係を抽出して、データを分類したり、予測したりするという点では、やっている「計算」は、多くの場合、共通しているのである。

統計手法との「共通性」を意識して、大きく整理すれば、機械学習とは、

- データと「知りたいこと」との統計的、確率的関係を数式化する
- その数式の「当てはまり具合」を、繰り返し計算による調整で、高めていく
- データと繰り返しの数が多いので、そこを機械で実施させる

ということになる。

「この機械学習／統計手法に基づくデータ分析が、人にどのような意味を持つのか」を、行動経済学という別の角度から考えてみたい。

近時の行動経済学の理論的仮説として、人間の思考には、2つのモードがあるとされている。直感的な早い思考である「システム1」と、論理的で遅い思考である「システム2」である。様々な情報に基づいて、論理的、統計的に思考するシステム2は、本来的に人間の思考に備わっている。だからこそ、人間が機械学習を作り出すことができるのである。しかし、このシステム2は高コストな仕組みであり、日常的に動いているのはシステム1の方だ。システム1は、「(印象に残っている)経験と勘」によるスピーディーな判断を可能とするが、その分、ある種のバイアスを持っていることになる(行動経済学に基づく各種の実験によって、これらのバイアス、人間の認知傾向について、様々なものが発見されており、人事評価の評価基準の策定などにも応用されている)。

この行動経済学の知見を踏まえると、機械学習とは、システム1が持つバイアスを補正するための情報を生み出すため、高コストのシステム2の思考の一部を、機械に下請けに出していると解釈できる。

さて、冒頭の表は、機械学習の実務家が私見として、統計と機械学習の違いをまとめたものであり、学術的な厳密さをもって、この2つの概念を比較したものではない。しかし、筆者が日頃から漠然と、統計と機械学習の差異として感じていたものを、明瞭に言語化してくれたと思っている。この統計と機械学習の比較が、あらゆる局面で有効とはできないが、直面している課題の解決のために、機械学習を適用するべきか、それとも統計的な分析を優先的に行うべきかの判断基準の1つとして有効だと思っている。

統計は、目の前にしている事象、出来事の特徴や傾向を把握し、「なぜ、そうなっているのか」を探っていくようにする知的営みであり、原因を探っていくことが、その目的、動機となっている。

一方、機械学習では、インプットがアウトプットに変化していくメカニズム、因果関係を解明することよりも、主に複雑な事象の予測をしたいという欲求、目的を持っている。有名なのは、囲碁やチェスといった複雑なゲームの勝利の方程式だが、天気予報であったり、複雑な化合物の組成であったりといった、かなりの多数の要素が複雑に絡み合っている現象のインプットとアウトプットの関係性を予測することが目的、動機、となっている。逆に、インプットとアウトプットの関係の一つ一つのステップやメカニズムの解明には優先順位が置かれていない。

保育所保育において、子どもの体調や発達過程をデータとして把握するという目的、動機は、とにかく「予測」できれば良いということにはならないだろう。それよりも、直面している「子どもの姿」「子どもの行動」の意味やそこに至る過程を解明することが求められているはずだ。ということであれば、統計的発想が重要で、そこに「予測」を主眼とする機械学習的なものの考え方を加味することになるであろう。

特に、重要なのは、システム1によってなされる保育士の子ども理解におけるバイアスを、どう補正するかということである。人間には、様々な評価バイアス、認知バイアスがあるとされている。これらは、先に紹介した行動経済

学では、ヒューリスティックと呼ばれ、大きく次の3つに整理されている。

①利用可能性ヒューリスティック

関連する情報を全て入念に調べることなく、簡単に手に入る情報に頼る。

⇒子どもの発達状況の理解を、印象に残りやすい年度初めの様子とごく最近の観察だけで行っている。

②代表性ヒューリスティック

類似性がありそうに見えるというだけの理由からもっともらしい関連性を見つける。

⇒ある発達のパターンやエピソードに類似した様子を観察すると、そのまま目の前の子どもにも当てはめてしまう。

③「アンカリングと調整」ヒューリスティック

特定の参照点を判断のアンカー(錨)として、この参照点に合わせて選択を調整、判断する。

⇒今持っている子どもの発達イメージと違う行動や状況を観察しても、それによってイメージを更新しない。

このような保育士という人が持っているかも知れない、子どもの発達過程や体調の理解におけるヒューリスティックを、統計的に説明されるデータ分析によって修正するということが、保育の質の向上のために有効なのではなかろうか。

機械学習やデータ分析から得られた洞察、インテリジェンス、あるいはその情報共有に基づいて、子どもの保育のための環境構成や関わり方、保護者支援として何をすべきかを最終的に判断するのは、人間である保育士でなければならない。そのための保育士の深い思考をサポートするような IT システム、IT 導入でなければならないのである。

IT の導入というと、『業務の効率化』、なかんずく『ペーパーレス化』が目的、ゴールイメージとなっていることが良くある。繰り返しになるが、「深く」はあるが、「高コスト」となってしまう保育士の思考を助ける「子どものデータ分析」という情報生産にこそ、保育所保育における IT 導入の「目指すべき付加価値」があるのだと思われる。是非とも、このような認識を持って、保育所の IT 化が進むことを期待したいところだ。

●当レポートは、信頼できると思われる各種データに基づいて作成されていますが、当社はその正確性、完全性を保証するものではありません。当レポートのご利用に際しては、ご自身の判断にてお願い申し上げます。また、当レポートは執筆者の見解に基づき作成されたものであり、当社の統一した見解を示すものではありません。なお、当レポートに記載された内容は予告なしに変更されることもあります。当レポートは著作物であり、著作権法に基づき保護されています。当レポートの全文又は一部を著作権法の定める範囲を超えて無断で複製、翻訳、翻案、出版、販売、貸与、転載することを禁じます。