人間参加型の AI 活用 (Human-in-the-loop)

第一生命情報システム株式会社 デジタル推進部
桐生佳介

(要旨)

2010 年代初頭に深層学習という人工知能開発の手法が話題になり、それ以来あらゆるビジネスや製品に AI が組み込まれるようになった。AI と聞くと 100%の正解を導いてくれる魔法のような存在だと思ってしまいがちだが、現実はそうではない。AI に完璧を求めることは事実上不可能であり、その特性を前提とした運営が求められる。さらに、AI は一回作ってしまえば良いというものではなく、性能を維持するためには継続的なメンテナンスが必要である。この課題を人間と AI の共存により解決しようとするアイデアが Human-in-the-loop である。

1. AI の普及

昨今、ニュースや新聞で AI という言葉を聞かない日はない。製造業では画像解析 AI を活用した不良品検出システムが当たり前のように使われているし、インターネット広告業界ではユーザーの趣味趣向に合わせ、思わずクリックしたくなるような広告を機械的に選出して掲示している。また、一般家庭においても AI スピーカーが日常生活に溶け込んでいる。音楽をかけて、と問いかければ大手ストリーミングサービスから好みの楽曲を選んでくれるし、スマートリモコンと組み合わせれば声で家電の操作が可能となる。実際に使ってみると、ずいぶんと便利な世の中になったものだと感心する。

2. AI は完璧ではない

そんな AI スピーカーも人間の声を誤って認識することもあるし、問いかけた内容にうまく応答してくれないこともある。これは人間の要求に対して AI の学習が不足していることを意味する。ただし AI スピーカーであれば、間違った答えが返ってきたとしても誰かが深刻な悩みを抱えることもないし、その間違いに愛嬌すら感じることもあるだろう。一方、近い将来実現が期待されている自動運転技術や、ビジネス上の意思決定を導くための AI という話になれば状況は変わってくる。AI が間違えることで人の命が危ぶまれるかもしれないし、企業に大きな損害をもたらすかもしれない。このリスクに対して言えるのは、AI は完璧ではないということだ。AI は人間と同じで 100%の正解を保証することはできない。さらに、AI の性能は時間の経過とともに劣化するというのが一般則である。いつの時代も銀の弾丸というものは存在しないものだ。しかし、AI の間違いに対するリスク軽減や性能の維持は、ある工夫により実現できるとも考えられている。それが Human-in-the-loop（以下 HITL）である。本稿では、人間と AI が共存する方法の一つである HITL について解説する。

3. AI が花開く時代

AI と呼ばれる仕組みには多くの場合、機械学習という技術が使われている。機械学習とは簡単に言えば大量のデータをコンピューターに読み込ませ、その傾向を覚え込ませることでデータの予測や分類ができる技術である。ここ数年、データ分析を通じてビジネス上の課題を解決する「データサイエンティスト」という職種もよく耳にするようになった。機械学習は彼らが課題に立ち向かうための武器の一つでもある。データサイエンティストの市場価値が高まったことも影響しているのか、書店の IT 技術書コーナーに立ち寄ると、機械学習に関連する書籍も多く並んでいる。

10 年ほど前から「ビッグデータ」という単語が世間を賑わせてきたが、大量のデータが存在すること自体に意味があるのではない。機械学習のような仕組みに大量のデータを与えることでビジネス上の価値を生み出すことができる、ということに意味がある。AI が普及してきた背景には、データを大量に保管できる大容量ストレージの低廉化、データを処理するプロセッサの高速化といったハードウェア的な進歩がある。これら技術の成熟と機械学習アルゴリズムの研究が並行して行われた結果、AI が花開く時代が訪れたわけだ。

4. まだまだこれからの AI 運用

現在、日本国内においては機械学習のモデル生成についての議論が活発である。モデルとは、あるデータの集合に対する振る舞いを定式化したものである。人間で例えれば、次の日の天気を予想したい場合、直近数日間の天候や今日の空模様を参考にするだろう。これは人間の経験則、つまりモデルに基づいた予測と言える。機械学習でも人間と同様に過去のデータを入力して学習する。例えば、冒頭に書いた不良品検出であれば製品の大量の画像と、それぞれの画像が示す正常・異常の評価をセットにして学習データとする。そして大量の画像データから、製品に傷や欠けがあれば異常（つまり不良品）なのだという傾向を学ぶ（この手法を「教師あり学習」と呼ぶ）。こうして出来上がったモデルを活用し、製造ラインを流れる製品に対し判定を行うことで不良品の検出ができるというわけだ。

さて、ここで製造工程に変化があったとしよう。顧客ニーズの変化に対応するため、製品のデザインが変更になった。ここで不良品検出システムはどのように振る舞うだろうか。多くの場合、製造工程の変更前よりも検出精度が低下するだろう。不良品を正常と判断してしまう場合もあるだろうし、良品を異常と判断してしまうことも起こりうる。この事象をコンセプトドリフトによるモデルの劣化と呼ぶ。モデルの学習時点で使用したデータから、その内容や傾向が変化してしまうことが原因である。機械学習モデルは「一回作ったら終わり」ではない。継続的なメンテナンスが必要なのだ。

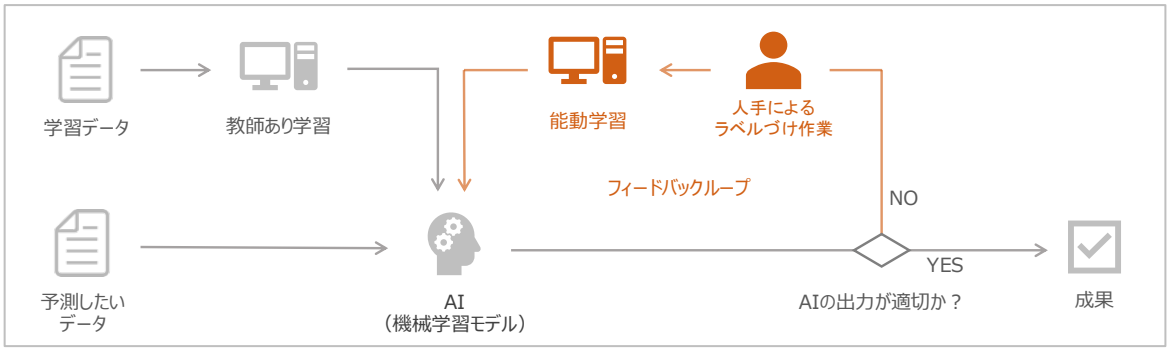
5. HITL とは何か

HITL は、元をたどると航空管制や軍事作戦行動の分野で使われていた言葉だ。航空機や航空管制はシステムと人間の相互作用によって成り立っている。言いかえれば、システム単体で成り立つものでなく実際にそれらを扱う人間を適切にトレーニングしてこそ、緊急事態を含めたあらゆる状況に対応できるようになる、ということだ。人間参加型のシステムという概念ではあるが、どちらかと言えば設計されたシステムに対する人間側のトレーニングという意味合いが強い。

AI 運用の領域で語られる HITL は航空管制のそれとは異なる。AI を運用する現場に人間が介入することで、 AI がより正しい成果を生むことを目的としている。先に述べた通り AI は 100%の正解を導くことはできないし、さらに時間の経過とともにその性能は低下してゆく。HITL は読んで字のごとく、AI を使った処理形態にループを作り、その中に人間を介在させることで運用上の諸問題を解決しようという考え方である。具体的な手法は以下の通りだ。

① 大量の学習データを用い、教師あり学習で予測モデルを生成する
② 学習と性能検証を繰り返し、十分な性能が得られれば本番環境へ移行する
③ 未知のデータに対する出力結果を観測し、問題なければそのまま成果とする
④ AI からの出力に誤りがあれば、人間によるオペレーションで対処する
⑤ AI が誤ったデータに対して人間がラベルづけの作業を行い、AI に再学習させる（能動学習）

図 1 HITL のイメージ（筆者作成）

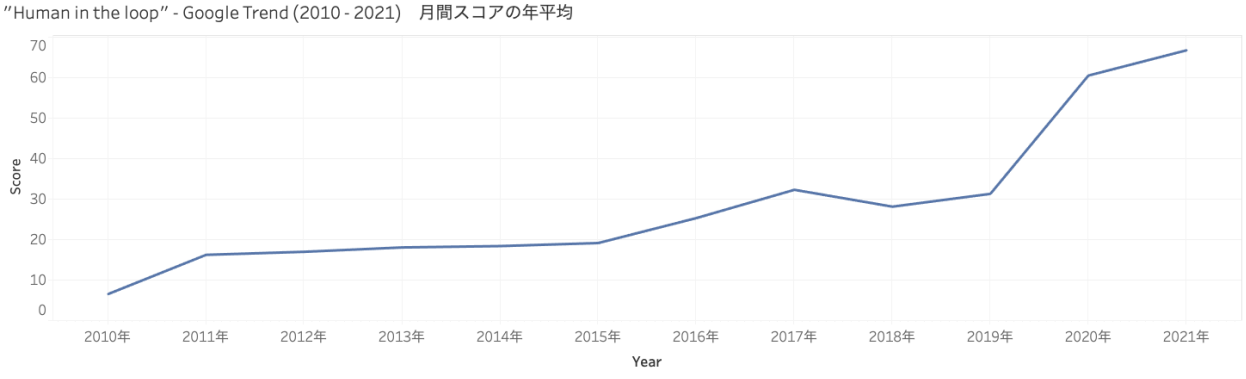


HITL で要となるのは、AI から出力された結果を人間がチェックし、そこから新たな学習データを得るというスキームだ。AI が間違えてしまったデータや全く未知のデータに対して人間がラベルづけを行い、それらのデータセットを用いて集中的に再学習を行う。これを能動学習(Active Learning)と呼ぶ。ちょうど人間がテスト勉強をするときに、間違えた問題や知らない問題に優先的に取り組むという構図と同じだ。効率よく成績を上げるには苦手分野の対策を繰り返すべきというのは人間も AI も変わらない。HITL はこの絶え間ない「ループ」により、持続可能なモデル運用を実現するのだ。

6. HITL に対する関心

AI を運用していくには不可欠な概念である HITL であるが、世間からの関心はどうだろうか？これについて Google Trend（注 1）を用いて可視化することにした。”Human in the loop”という言葉が Google で検索された件数をトレンドスコアとしてプロットしたものが図 2 である。グラフを見るとここ数年で急激に関心が高まっていることが読み取れる。深層学習が火付け役となった第 3 次 AI ブームにより、データ利活用の機運が高まった結果、その運用方法についても注目され始めていると考えられる。

図 2 HITL に対する関心度の変遷（Google Trend より筆者作成）



次に HITL に対する関心の度合いを地域別で見ることにしよう。GAFAM（注 2）をはじめとする大手 IT 企業が軒を連ねる米国は流石のトップとなっている。AI 関連の研究が盛んであるがゆえ、その運用についても関心が高いのは納得である。欧州諸国がトップ 5 に入っていない（英国は 6 位）のは意外であったが、英語以外を母国語とする都合もあるのかもしれない。日本も同様にランク外であり、HITL という言葉自体がまだ一般に浸透していないことがうかがえる。

図 3　HITL に対する関心の高い地域　（Google Trend より）

7. 国内の事例

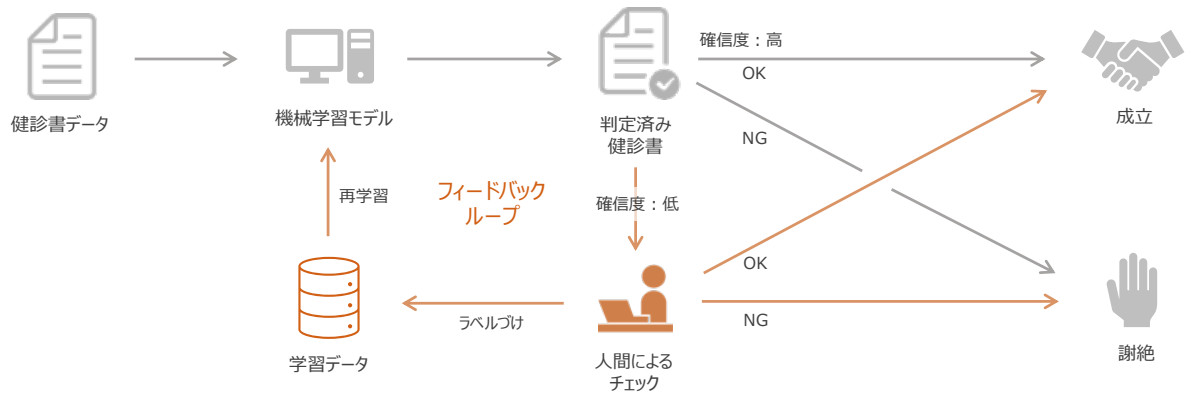
日本国内では大手フリーマーケットサービスが HITL を導入している。同社は危険物や転売品などガイドラインに違反する出品物の検知に機械学習モデルを利用している。当然ではあるが、流行り廃りといった消費者志向の変化や、社会的構造の変遷により出品物の潮流も変化する。例を挙げると、新型コロナウィルスの流行により不織布マスクの需要が増加した際には、転売で利益を上げようと考えたユーザーからマスクの出品が相次いだ。こうした事態が TV ニュースやネット上で話題となったうえ、2020 年 3 月より日本政府がマスク転売に対する罰則の政令を発出したため、同社はマスクを違反出品物として扱わなくてはならなくなった。このようにポリシーの変更によって検知すべき対象が変化した場合、もしくは新たに現れた場合には、すでに学習済みのモデルを再学習する必要がある。

では、どのような仕組みで実現しているのだろうか？同社では機械学習モデルが違反と判定した商品については全て人間の目で確認することになっている。もちろん、モデルの判断に誤りが無いかを確認することが第一義ではあるが、このプロセスを経ることで再学習に使うデータを蓄積することができる。つまり、人間が結果のチェックを行うと同時に本来下すべきであった判断をラベルづけすることができるという意味だ。このデータを用いて能動学習を行えば、新たなポリシーに順応したモデルを生成することができる。同社ではこの HITL が機能しているがゆえに、違反出品物の検知システムが高い性能を維持できているという。

8. 生命保険業界での活用ユースケース

生命保険業では HITL はどのように機能させられるだろうか。例えば、保険の新契約時には、顧客の保険を引き受けるべきか否かを判断する「査定」というプロセスが存在する。この査定内容を機械学習モデルで判断させるシステムがあったとしよう。死亡保険や定期保険などリスクに備えるタイプの保険に加入する際、顧客は健康診断書を保険会社に提出する場合が多い。この健康診断書に記載の内容をもとに機械学習モデルが契約を成立すべきか謝絶すべきか判断することを考える。業務の性質上、高い精度が求められるわけだが、AI に 100%はない。そこで、HITL を導入したイメージが図 4 である。機械学習モデルが出力する結果としては OK もしくは NG の 2 択になるわけだが、多くの場合その裏には何%の確率で OK なのかというスコアを持っている。この確信度が高いケースにおいては機械学習モデルの出力をそのまま成果とし、確信度が低ければ人間のチェックを介して最終判断を行う。このときモデルの出力が誤っていれば人間がデータにラベルづけを行い、再学習に回す。こうすれば、業務全体としての精度を保ったまま、徐々に機械学習モデルの性能を向上させることができるだろう。システムに人を介入させるにはコストがかかるが、その恩恵により機械学習モデルの性能が上がっていけば、だんだんと手がかからなくなってくるはずだ。

図 4　機械学習モデルによる査定業務に HITL を導入したイメージ（筆者作成）



9. HITL の導入に向けて

実用的に AI 運用したいのであれば、まず AI には継続的な再学習が必要であるということを認識しなくてはならない。その前提に立ったうえで次に必要となるのはフィードバックループを実現するための設計である。具体的には次の 4 点であると考える。

① 人間が AI からの出力を観測するためのビジネスプロセス設計
② 人間が効率的にラベルづけ作業を行うためのユーザインタフェース設計
③ 能動学習に用いるデータセットを蓄積するためのシステム設計
④ 再学習および本番環境へ機械学習モデルの再移行を行うためのオペレーション設計

能動学習に用いるデータの選定方法やモデルの最適化手法などは、細かく語れば枚挙にいとまがない。言ってしまえば、すべての人間が理解しておく必要はないだろう。しかし、上記の 4 点については HITL を実現するための原理原則であり、AI 運用に関係するすべてのステークホルダーに認識してもらいたいポ

イントだ。業務プロセスを考えるビジネス企画やシステム開発予算を取る見積もりの時点から、AI 運用は始まっている。地に足がついた AI 運用ができるか否かは、機械学習モデルの本番リリース後に決まるのではない。HITL はプロジェクトの最後に議論されるべきものではなく、むしろ最初から計画に入れておくべきなのだ。

10.おわりに

AI が普及するにつれ「人間の仕事が奪われるかもしれない」という悲観論も見かけるようになった。人間に残されたのはクリエイティブな仕事だけだとか、将来なくなる仕事はこれだとか、なかなかショッキングな議論が繰り広げられている。少なくとも現時点の AI 技術で言えることは「ある作業は無くなっても、仕事自体は無くならない」ということだ。いくら AI による強力なサポートがあったとしてもビジネス上の意思決定を下すのは人間であるし、AI がより良い成果を得られるよう維持するには、人間によるコントロールとメンテナンスが必要となる。

SF の世界での AI はまるで魔法のような振る舞いをするが、現実はそうではない。人間と AI が共存してこそ、人間だけでも AI だけでも導くことのできない大きな成果を生み出すことができるのだ。

[注]

1） あるキーワードについて、Google で検索された回数の時系列的な推移を可視化できるサービス。最も多く検索された時期を 100 として相対的に表現する。本稿では年間平均の値をグラフ化している。

2） Google, Amazon, Facebook, Apple, Microsoft の総称。

[参考文献]

Association for Computing Machinery, Inc / “Human-in-the-Loop Modeling and Simulation”
https://www.acm-sigsim-mskr.org/MSAreas/InTheLoop/humanInTheLoop.htm (参照 2021.6.30)

Robert (Munro) Monarch, Human-in-the-Loop Machine Learning, Manning, 2021

澁井 雄介, AI エンジニアのための機械学習システムデザインパターン, 翔泳社, 2021

※本レポートの発行元は第一生命情報システム株式会社です