Over the AI ―― AIの向こう側に（9）：

へつらう人工知能 ～巧みな質問を繰り返して心の中をのぞき見る

http://eetimes.jp/ee/articles/1703/30/news034.html

今回は、機械学習の中から、帰納学習を行うAI技術である「バージョン空間法」をご紹介しましょう。実は、このAI技術を説明するのにぴったりな事例があります。それが占いです。「江端が占い師に進路を相談する」――。こんなシチュエーションで、バージョン空間法を説明してみたいと思います。

2017年03月30日 11時30分 更新

[江端智一, EE Times Japan]

今、ちまたをにぎわせているAI（人工知能）。しかしAIは、特に新しい話題ではなく、何十年も前から隆盛と衰退を繰り返してきたテーマなのです。にもかかわらず、その実態は曖昧なまま……。本連載では、AIの栄枯盛衰を見てきた著者が、AIについてたっぷりと検証していきます。果たして”AIの彼方（かなた）”には、中堅主任研究員が夢見るような”知能”があるのでしょうか――。⇒連載バックナンバー

25年前、占いに手を出した記憶

25年ほど前の話になります。

当時、大学院の修士課程の2年生になった私は、進路について、随分悩んでいました。それは、博士過程に進むべきか、就職すべきかという単純な二者択一の悩みだけではありませんでした。

今の大学に残れるのか、残れなければどの大学を受験すべきか、学費はどうすれば良いのか、地元に戻るべきか、就職先に合わせるべきか、大学に残り続けるのが良いのか、就職をすべきなのか ―― そこには、無数のオプションがあり、悩み尽くして疲れ果てた私は、ついに、普段であれば絶対に手を出さないであろうものに、手を出しました ―― 占いです。

当時、私は大阪梅田の第三ビルの中にある、中高生向けの通信教育教材会社で、その教材に関する質問にやってくる子どもたちに個別指導をする講師のアルバイトをしていました。そして、そのビルの地下に、数個の占い用のブースがあるのを知っていました。

そこで、生まれて初めて「占い」なるものをした私は、私の人生において最大級の激怒、軽蔑、憎悪、その他の、ありとあらゆるネガティブな感情を共なって、人生の負の遺産として、私の記憶に焼き尽けることになったのです。

そもそも私は、その当時から、血液型や星座による性格判定やら占いやらを、信じない人間でした。占いが、「絶対的にあてにならないこと」を経験として知っていたからです。

例えば、私の星座である、「さそり座」の人間の特性は、

- 無口で一途な正確
- 信念に従い、意志が固く、不言実行型
- 成功するまでは、決して諦めない根性の人
- 理知的で、天才肌

だそうですが、

―― 誰だ、それは？

また、私の血液型である「AB型」の人間の特性は、

- 寂しがり屋
- 引っ込み思案な性格
- プライベートは秘密主義
- 常識を重んじる、ロマンチスト

だそうですが、

―― だから、本当に、その人間は一体どこにいるんだ？

私という人間は、

1. 頼まれればスピーチでも講演でもホイホイと引き受け、ベラベラとしゃべりまくり、
2. 初めての仕事でもさっさと取り掛かり、うまくいかないと分かればとっとと撤収し、
3. 電子レンジから鉄道システムに至るまで、節操なくさまざまな業務の仕事にかかわり、
4. 毎日、自分のWebサイトで、自分だけでなく、自分の家族のことまで公開して、
5. 電車の飛び込み自殺で人体が切り刻まれるシミュレーションのプログラムを、平然とした顔で組み上げる（「1/100秒単位でシミュレーションした「飛び込み」は、想像を絶する苦痛と絶望に満ちていた」）

―― そういう「さそり座」の「AB型」の人間です。

私は、占い師だけでなく、占いの客となる人間も、気持ち悪いと思っています。

占い師：「あなたは、外見的には規律正しく自制的ですが、内心ではくよくよしたり、不安になったりする傾向がありますね」
客：「ええ！　スゴーイ、なんで分かるんですかぁ!?」

意味が分かりません。「誰よりも分かっているはずの自分の性格を、他人に当ててもらって驚いている人」が、一体、「何」を求めて占い師に金を払っているのか、私には全く理解できないのです（バカにしているのではなく、「本当」に分からない）。

そもそも、この手の「占い師による自己能力アピール」の手法は、1956年段階で、心理学的に解明されています。「バナーム効果」というものです。

占い師が、客に対して、「あなたはロマンチストな面を持っています」「あなたは快活に振舞っていても心の中で不安を抱えている事があります」のようなフレーズを与えると、その客は、――ほとんどの客が ―― あたかも、その占い師が『自分のみが有する特殊な性格を言い当てている』と思い込んでしまい、占い師に信頼を寄せるようになる、というものです。

また、客が、占い師に対して、その占いに対する根拠の説明を要求しないことも、気持ち悪いと思っていますし、逆に、占い師が、客に対して、そのような説明をしないことも、信義則に反する不作為であると思っています。

もちろん、私は、その根拠の説明が、科学的や論理的であることを要求しません。占いは、それが、科学的でも、論理的でもなく、その個人の有する特殊な能力（占い師が霊視とか霊感とか呼んでいるもの）に依拠している（ことになっている）点にこそ、価値があるのですから。

ですから、その説明が、うそでも作り話でも、例えばこんな、頭のイカれたような話でも私は受け入れます。

土星と木星の視野が30度以内に入ると、そこから惑星間オーラが地球におよび、この惑星オーラは、平均気温11℃の環境で生まれた「さそり座」の、特に血液に「AB型」の血液に影響を与えるんですよーー。

――と、こんな話であっても、私は、説明義務は果たしたものとして評価します（もちろん、そんな話は信じませんけど）。

全ての占い師は、あの宗教団体の教祖 ―― 著名人に面会してもらえず、「（その著名人の）守護霊インタビュー」などという本を出版している ―― のマインドに学ぶべきです。占い師を名乗るのであれば、客観性絶無の、自分の中だけで閉じたトンデモ……もとい、オリジナル理論を、自力で創作するくらいの気概を示せ、と私は言いたいのです。

私が、「占い」なるものに価値を見いだせるとしたら、その「フレーズの創作性」にあります。

「上から落ちてくるものに気を付ける」
「黄色のTシャツがOK」「水色のハンカチはNG」
「好きな人から告白されるかも」
「人からあらぬ疑いをかけられるかも」


画像はイメージです

同じような内容のフレーズを、微妙に表現を替えながらも、毎日（毎朝）、生産し続ける努力は評価します*)が、この程度のフレーズであれば、簡単なプログラムで作ることができると思います（で、そのプログラムは100年分の、重複のない200万フレーズ（365 日 x 100年 x12（星座） x 4（血液型））を、数分で作成し終えるでしょう）。

*）これも近いうちにテキスト分析してみようと思っています（「我々が求めるAIとは、碁を打ち、猫の写真を探すものではない」）。

それでも、歴史上「占い」というサービスが消滅したことは1度もありません。それは、占いが「正確な未来予測」や、「その予測に基づく適切なアドバイス」ではなく、それらとは全く異なる意義を持っているからなのです。

あきれかえる「AIによる人類の滅亡」説

こんにちは、江端智一です。

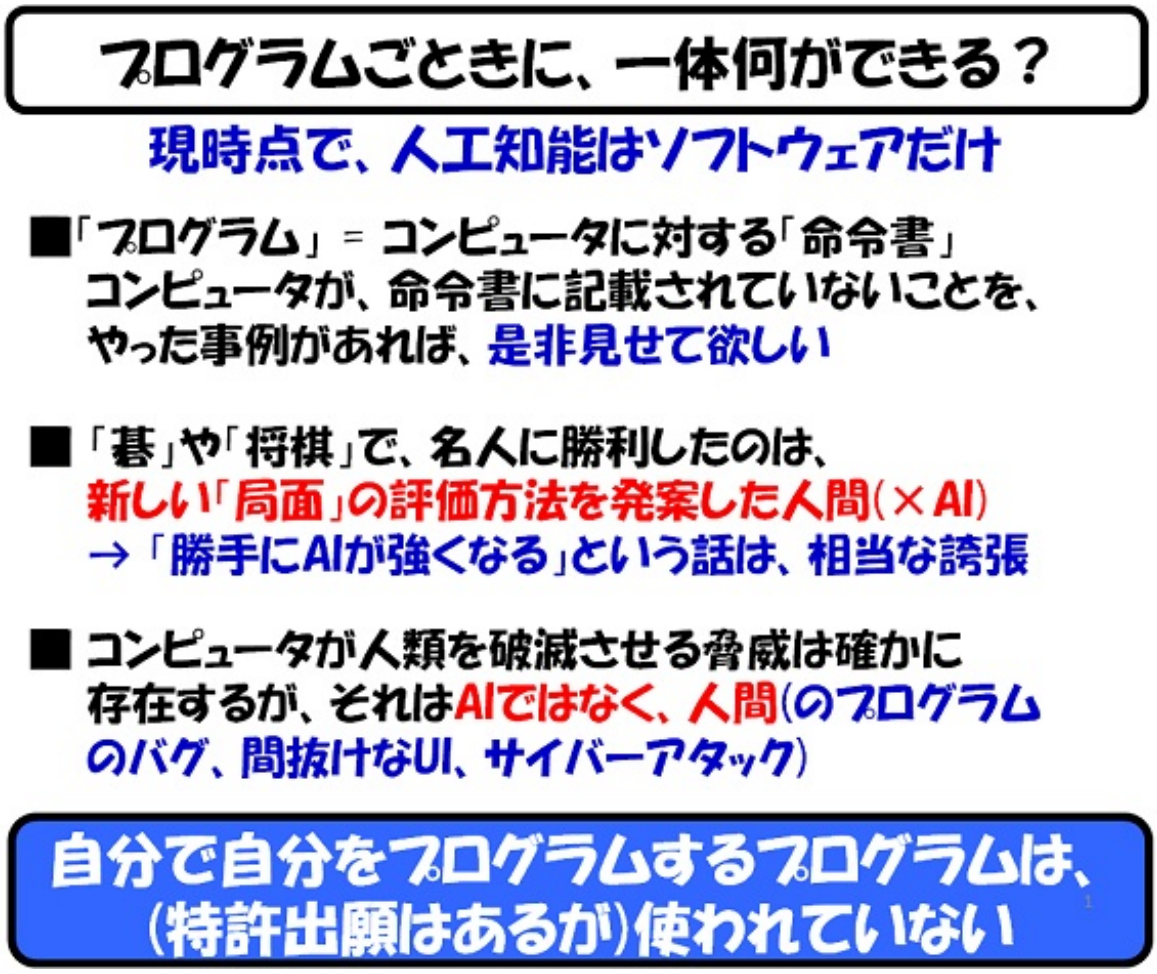
今回は、前半では、前回に引き続きCSAJでの研究講演会の内容（一部加筆）をお話し、後半では、機械学習の1つである「バージョン空間法」について解説します。

私は、今回のAIブームだけでなく、前回（第2次）のAIブームにも立ち会った当事者（兼、被害者&加害者）であり、第1次ブームの話についても、いろいろと調べました。そして、これらのブームの全てに必ず登場するのが以下のフレーズです。

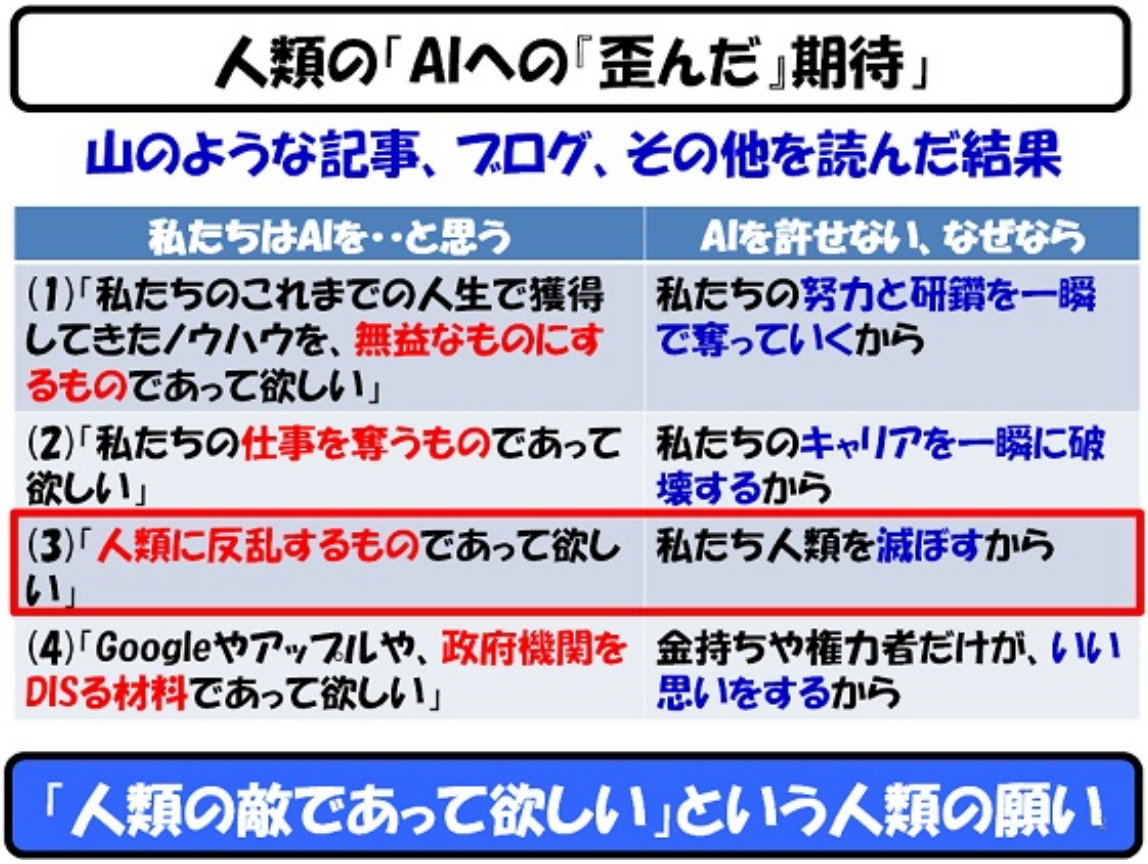
「人工知能が人類を滅亡させる」

これまで、私は、ネットの記事や書籍で、このフレーズが出てきた段階で、「読む価値なし」として読むのを止めていたのですが、今回は、この連載のために、我慢して読んでみました。が ――もう、なんていったらいいのか、バカバカしくてお話にもなりません。

その理由を説明するのも面倒なので、研究講演会の資料を以下に張りつけておきます。



この連載の第1回「中堅研究員はAIの向こう側に“知能”の夢を見るか」で、私たちの中に存在する「AIへの期待」についてお話しましたが、これも、これから何度も使い回すことになりそうなので、このような資料も作りました。



　このような、「人工知能（AI）による人類滅亡アプローチ」を記載している本を何冊か読んだのですが、それらの著書（というか「著者」）に共通していえることは、AI技術に対する、絶望的なまでの「無勉強」です（「不勉強」ですらない）。

　まず、「AI」なる技術があると思っている時点でもうダメです。「AI」というのはコンセプトであって、技術ではありません。

　このシリーズでは、このことを一貫して言い続けているのですが（今回も後述しますが）、この理解なく「AIによる人類滅亡」を憂いている人は、「天が落ちてきたり、大地が崩れたりしないかと、あり得ないことを心配して、夜も眠れず食事も食べられなかったという、『杞』の国の人のようなものだ（『杞憂』の語源となった故事）」と ―― 私は「残念な人」と思っています。

AIに対する理解は、目がくらむほどデタラメだ

　さらに、「AI技術」という概念を理解している人であっても、そのAIの技術内容の理解が目もくらむほどデタラメだったりします。

　例えば、

（1）遺伝的アルゴリズム（GA）が、新しい解を見つけ出して、人類滅亡を……（以下省略）

　GAについては、第7回の「抹殺する人工知能 ～ 生存競争と自然淘汰で、最適解にたどりつく」で紹介しましたが、いくらGAと言えども、解空間の外側まで解探索できる訳がありません（そもそも、私には「『人類滅亡』を包含する解空間というのが観念できない）。

（2）エキスパートシステムが、人類滅亡が合理的な解答であると判断して……（以下省略）

　エキスパートシステム（「笑う人工知能 ～あなたは記事に踊らされている～」参照）が、どんなに努力しても、「人類滅亡」というルールのないシステムで、そのルールを導出することがでる訳がありません。

（3）ディープラーニングの「教師なし学習」によって、人類滅亡という解が……（以下省略）

　……おい、お前、「教師なし学習」の意味、全然理解せずに使っているだろう*）

*）「教師なし学習」については、この連載の終盤で、あなたが「うんざり」するほど説明させて頂く予定です

　前回のコラムで、私は「100冊の人工知能の本を読んでも時間の無駄です。最もてっとり早いのは、100行のプログラムを自分で書くことです」と言いましたが、プログラムの経験のない人に、私はそこまでは要求しません。ですが、少なくとも「人工知能脅威論」を論じるなら、せめて、勉強くらいしてこい、と言いたいです。

　私だって、1つのAI技術を「あ、なんか分かった様な気がする」と感じられるようになるまでに、往復3時間の通勤電車の中で、1週間くらいぶっ続けで勉強するくらいの努力はしていますので、こういう「無勉強」な本を読むと、いつでも、こんな気持ちになります。

　―― 舐めんなよ

　前述の占い師の話ではありませんが、なにより、私が憤慨していることは、―― うそでも、作り話でも、イカれた話でも、一切構わないのですが ―― 一本、筋の通った「人工知能による人類滅亡シナリオ」を、彼ら（著者）が持っていないことです。

　おおむね、彼らの文章の構成はこんな感じになっています。

（A）「〇〇という技術がある」→
（B）「△△大学の××教授によれば、〇〇という使い方が可能であるという」→
（C）「これを発展させていけば、近い未来に、人類を滅ぼせるだけの脅威となろう」

　では、1つずつ突っ込んでみましょう。

- まず（B）の「△△大学の××教授」って誰？　私聞いたことないぞ、そんな奴
- 「〇〇の使い方が可能になる」の根拠の記載がないけど、それは何？
- （C）の「発展」って、機能的な話？　リソース的な話？　それとも、推測とか願望？　そして、なにより、そのAI技術を使って、どういうアプローチで人類を滅ぼせるの？

という観点の記述が、ない、皆無、絶無。

　そこで今回、私は、（本当に面倒でしたが）AI技術の一つ一つについて、検証してみました。結

果は以下の通りです。

　単一のAI技術で無理なら、これらのAI技術を組み合わせれば、人類を滅亡させられるかというと ―― 当然、無理に決まっています。

　情報処理試験を受けたことがある人なら知っていると思いますが、「複数の技術はそれを組み合わせることで、最も懦弱な技術の性能に引っ張られる」ことは、システム論の常識です。

それでも、もう少し頑張って仮説を立ててみた

　それでも、私はもう少し頑張って考えてみました。

■仮説：AI技術を搭載したコンピュータが暴走したら、人類を滅亡させられる

　残念ですが、「暴走する」のは「AI技術」の仕業とはいえず、単なる「故障」です。「故障」と「AI技術」は無関係です。

　それに、もし、AI技術のプログラムが、バグなどで、暴走したのであれば、コンピュータの電源

を落とせばいいのです。そのAI技術が、先回りして送電経路の制御を掌握する恐れがあるというのであれば（そんな都合のよいバグが発生する可能性は絶無でしょうが）、その地域一体を全部停電させれば良いですし、発電所を停止しても良いです。自家発電装置や蓄電池があるのであれば、その装置を破壊するか、ケーブルを切断すれば足ります

　電源を喪失することによって、逆に暴走を始める物騒なものは、世界中探したって「原子炉」くらいなものです（参考記事）。

■仮説：AIによって指揮系統を掌握された軍隊や兵器が、人類を滅亡させる

　これは、かなり無理をすれば考えられなくはないのですが、現実的ではありません。コストが見合わないからです。

　軍隊や兵器を掌握するための、専用のハードウェアと専用のソフトウェアから成る専用システムと、そのための、国家予算レベルの膨大なコストと時間が必要です。

　しかし、どこの世界に、自国を滅亡させるのためのAI技術を作るべく国家予算を費やす国があるか、ということです。

　多分、システムエンジニアでない人の中には、コンピュータシステムというのが、ちょっとした変更で、そのシステムの目的を、別の目的に変更できると考える人がいるのでしょうが ―― 私のところで、私の仕事の手伝い*）を、やってみますか？　3時間もあれば簡単に考えが変わると思います。

*）例えば、数百万オーダのオブジェクトを同時に取り扱う、絶望的に可読性の悪い、数千行のソースコードのデバッグなど

コンピュータによる人類滅亡の危機は、確かにあった

　このように、AIによる人類滅亡論を否定した私ではありますが、コンピュータシステムによる人類滅亡のシナリオ自体を否定している訳ではありません ―― というか、分かっているだけで、人類は10回以上も、このような危機に遭遇しているのです。

誰が人類を滅ぼすのか？

仮説：ITやAIが人類を滅ぼせる

テーマ	きっかけ	原因	結果
全面核戦争	約2000発のソ連のミサイルが飛んできているという核攻撃警告（1980年）	コンピュータチップの誤作動	理由は不明 #2000発というのは、非現実的だったから？ →回避
	本土との連絡が取れなくなった潜水艦艦長が、報攻撃と判断（1962年）	通信システムの断絶	潜水艦の副官の説得→回避
	5発の核ミサイルの警告（1983年）	アメリカ領土から反射した日光が衛星を誤作動	現場の中佐が上「誤警報」であることに賭けた →回避
	ノルウェーのロケット実験（1995年）	ロシアのレーダー運用チームに連絡なし	報復作戦準備中に、「実験」であることが判明→回避
（3大）原発事故	システムの「安全装置」が誤作動（スリーマイル島原子力発電所事故（1979年））	薬品弁が閉鎖→ポンプ、タービン停止→安全弁開放→冷却材喪失	運転員の誤判断により、非常用炉心冷却装置はが手動で停止→炉心溶解（メルトダウン）
	外部電源喪失を想定した実験（チェルノブイリ原子力発電所事故（1986年））	実験性能を上げる為、安全装置を解除して実験を強行	緊急停止ボタン押下にもかかわらず、原子炉内の蒸気圧により、7秒後爆発→原子炉爆発
	外部電源喪失（福島第一原子力発電所事故（2001年））	原子炉内部や核燃料プールへの注水が不可能	冷却水蒸発→炉心溶融（メルトダウン）＋ 原子炉格納容器爆発
大規模停電	北アメリカ大停電（2003年）	送電管理システムダウンの連鎖	計5000万人が被害、被害総額約7000億円

仮説は棄却→人類滅亡の危機を作ったのも防いだのも、結局は人間

ただ、それでも、人類滅亡の危機を発生させたのも、そして回避させたのも、結局のところ、コンピュータではなく人間自身だった、という結論に落ち着きます。

しかし、私は、『あなたが私（江端）の言っていることを、うのみにしてよい』とも思ってはいません。

私は、「AIによる人類滅亡の危機」というものが ―― その内容がどれほど荒唐無稽なものであれ ―― それを叫び続けることには価値があると思っています。

「あなたが感情的に叫び続けることで、私（エンジニアたち）には、それに理性的に説明する機会が生じる」―― これは、意外に重要なことだと思っています。

なぜなら、科学者や研究者やエンジニアは、自分の技術に自信があって、そして、タチの悪いことに、それを心から信じているのです。

だから、自信タップリに、「確率論を基礎にした原子力発電の大規模事故の確率は、原子炉1基当たり『10億年に1回』→だから原子力発電所の事故への心配は全て杞憂」などということを、信念を持って主張してきた訳です（参考記事）。

以下は、6年前に、私が作成した実際の原子力発電の大規模事故の年表です。

私たちエンジニアは、自戒を込めて、この事実を思い出さないといけないと思うのです。

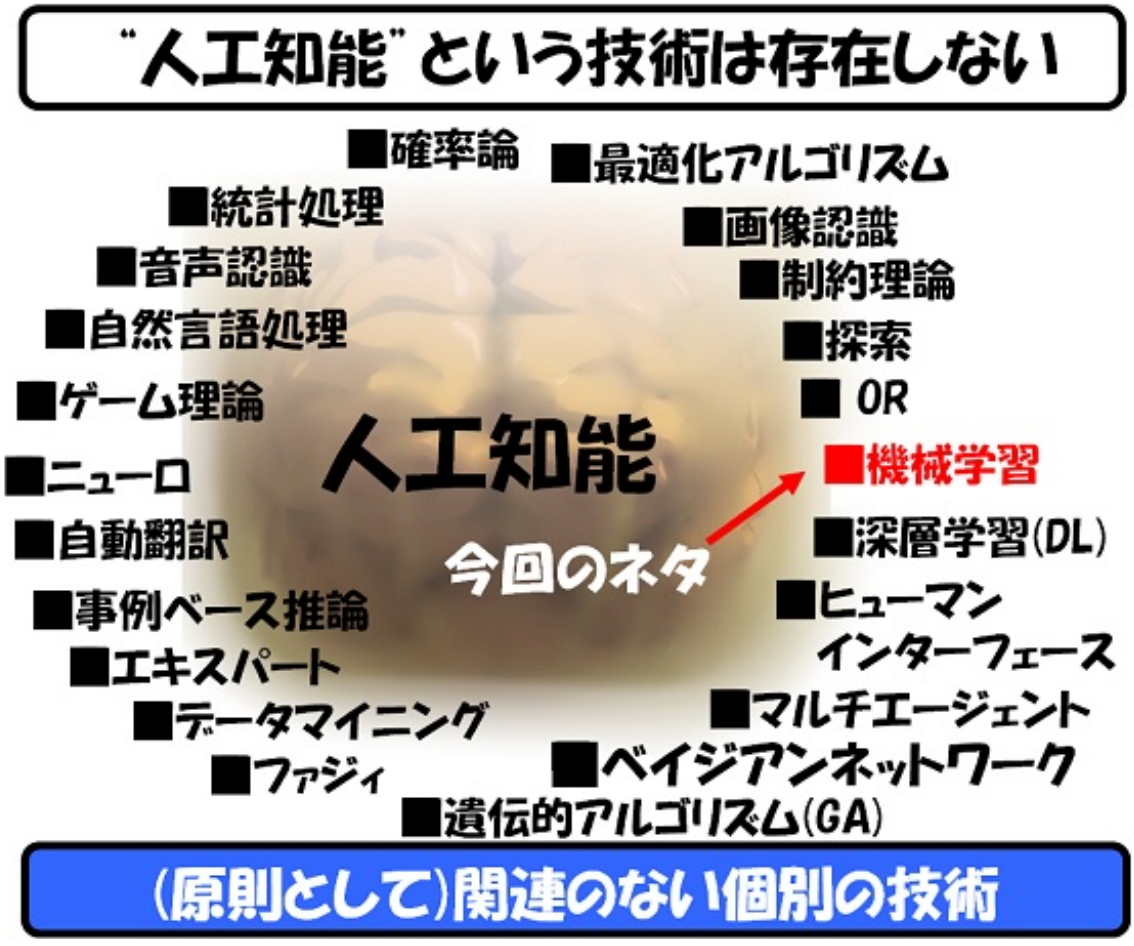
近い未来、私が「AI技術が起こした大規模事故年表」なんてものを作っている可能性だってあるのですから。

帰納学習を使う「バージョン空間法」

ここからは後半になります。この連載の後半は、「私の身の回りの出来事」を使った、「数式ゼロ」のAI解説になります。

今回は、機械学習と、その中の1つである帰納学習法を使ったAI技術である「バージョン空間法」についてお話します。

「『バージョン空間法』が“人工知能”なのかどうか」については、今回も『江端AIドクトリン』に基づいて私が勝手に判定しました



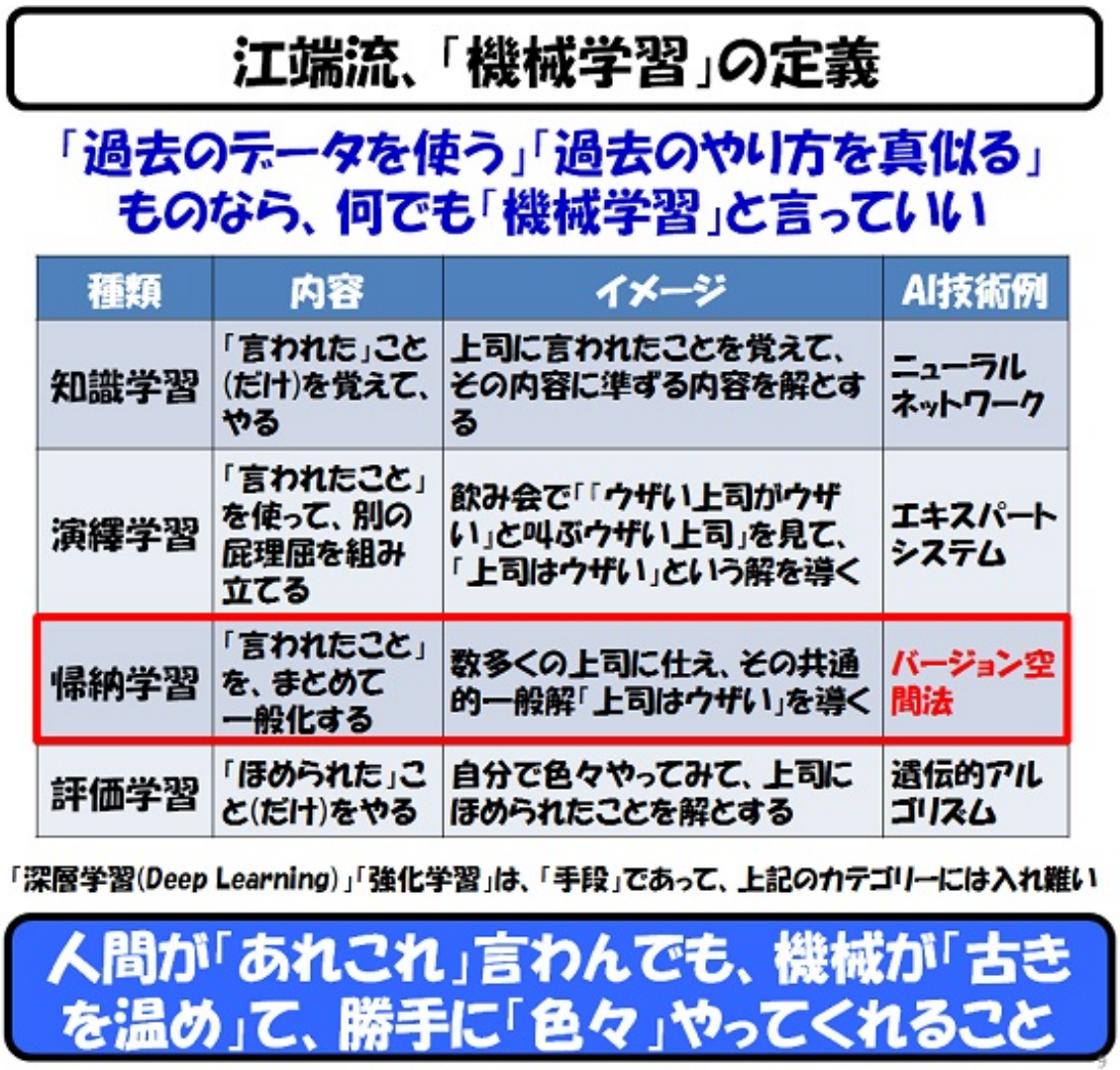
最初に、「機械学習」について解説します。

この「機械学習」なるものも、「人工知能」と同様に、きっちとした定義がある訳ではないようで、『なんかスゴそう』という感じはしますが、結局のところは、『昔の何かを使って、何かをする』という程度の理解で問題ありません。

もう1つ特徴を挙げるとすると、「機械学習」は、その『昔の何か』を後発的に次々と追加していける仕組みが備わっている、ということです。つまり、使えば使うほど、お利口さんになってい

く、ということもいえるかもしれません（逆に、使えば使うほど、使えない奴になる、というケースもあります（何度も目にしました））。

私なりに「機械学習」をまとめてみたものが、以下の表です。



上の表に示した学習方法は、私たち人間が日常的に当たり前に行っている学習なのですが、こうした当たり前のことをコンピュータに実行させるのは恐しく難しいのです。

まず、このような知識をどのように分解するかが難しく、どのように格納しておくのかが難しい。さらに、分解した知識をどのように関連づけるかが難しく、そこからどのように組合わせて結論を出すのかも、絶望的に難しいのです。

コンピュータを「超高速の電卓」のままにしておけばよかったものを、人類は「この『電卓』を使えば、人間と同じものが作れるに違いない」と勝手に信じ込み、先ほど述べた「AIへの“ゆがんだ”期待」も同時に発生した訳です*)。

*）そろばんも電卓もコンピュータもメカニズムは基本的に同じ（10進数の代わりに2進数を使っていることと、そろばんの「願いましては……」で始める読み手の代わりをプログラムがやって

いるだけの違い）なのに、「そろばんが人類を滅ぼす」とは、誰も言いません。

「バージョン空間法」の考え方

　それはさておき。

　上の表に「演繹学習」と「帰納学習」という学習方式が出てきましたが、これ、結構重要なので、簡単にもう一度説明します。

<演繹学習の例>

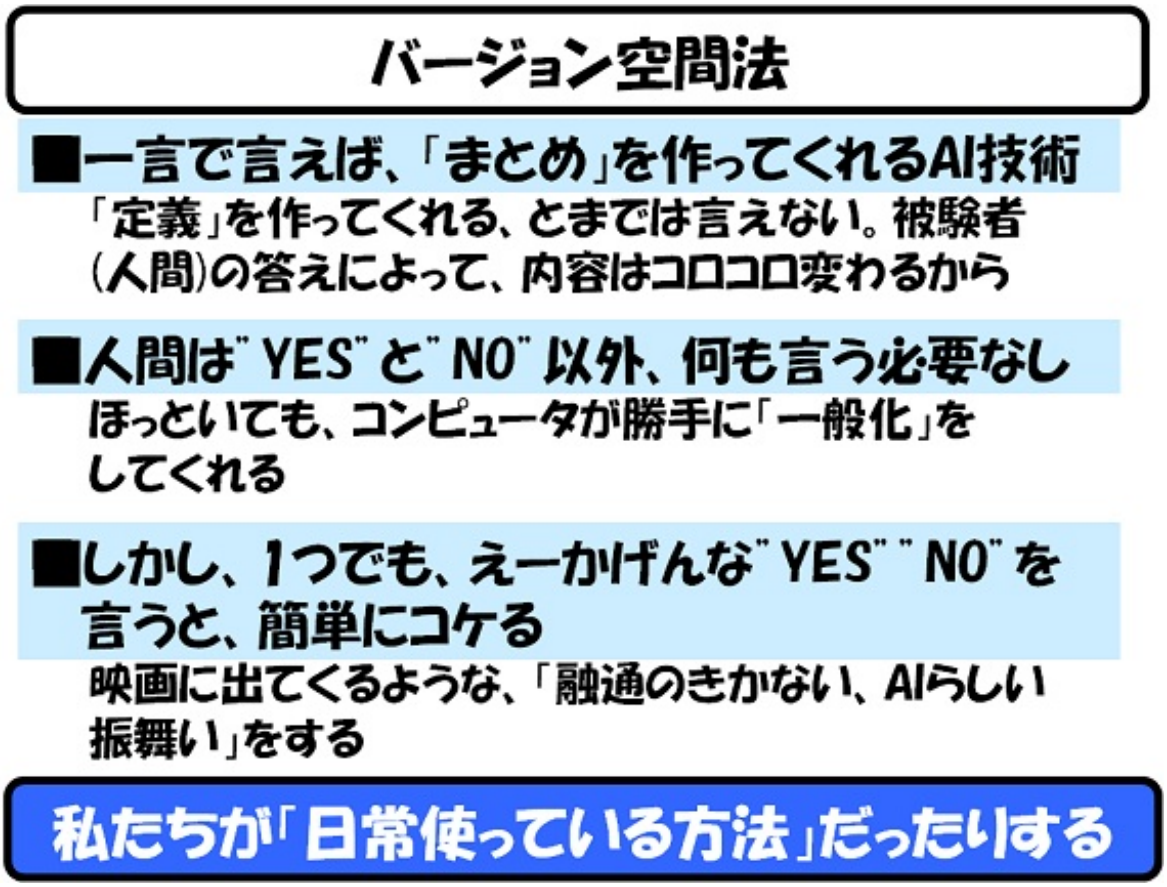
　「上司はウザい。私は部署を異動した→（演繹学習）→ならば、異動後も、その後の異動後も、その後も、死ぬまで全ての上司はウザいであろう」

<帰納学習の例>

　「異動前の上司はウザかった。その異動の前の上司もウザかった→（帰納学習）→ならば、移動後の上司もウザいであろう」

　学習から導かれる結論は同じなのですが、演繹学習は、1つの理屈を複数のケースに当てはめていく感じで、帰納学習は、たくさんのケースから1つの理屈を導き出す、という感じになります。

　さて、本日ご紹介する「バージョン空間法」は、機械学習の中でも、この帰納学習を行うAI技術です。



「バージョン空間法」のアルゴリズムはこんな感じになっています。

バージョン空間法のアルゴリズム

■前提

概念空間を適当に定義する(イメージとしては、N次元空間 H=(α,β,γ)←3次元)

■手続

(Step.1)

(a)最も一般化された概念空間 G=(*,*,*) と、

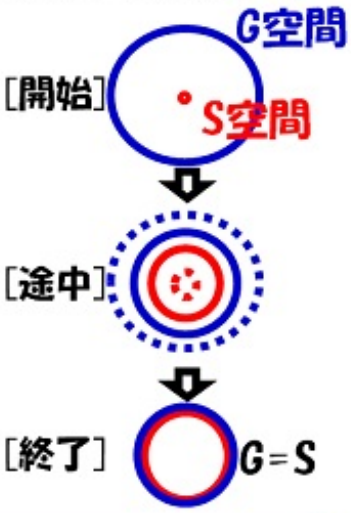
(b)最も特殊化された概念空間 S=(0,0,0)(←例)

の2つを定義

(Step.2)

コンピュータは、人間に色々質問して、以下のことを、ひたすらやり続け、S空間とG空間が一致したら終了する

事例X	Gの更新	Sの更新
人間が"YES"と答えた場合	Xに矛盾する要因を、片っ端から取り除く	Xを追加し、必要に応じて一般化("*"にする)
人間が"NO"と答えた場合	Xを含まないように、最小限に特殊化(定数として特定する)	Xを取り除く



うん、何を言っているか分かりませんよね
(大丈夫、私も分かっていません)

このバージョン空間法を、暴力的なまでに単純化すると、こういう感じです。

コンピュータ:「0〜10までの間の1つの整数を決めてください」

江端:「決めました」

コンピュータ:「それは5以上ですか」

江端:「はい」

コンピュータ:「それは7以下ですか」

江端:「いいえ」

コンピュータ:「それは8以上ですか」

江端:「はい」

コンピュータ:「それは8以下ですか」

江端:「はい」

コンピュータ:「あなたの決めた数字は"8"です」

　こんな感じのやりとりを、もっと複雑な条件下で実施し、できるだけ少ない回数で、広大な解空間の中から、正解に到達するという解探索方式です。

　今回、バージョン空間法について調べてみたのですが、第二次ブームのエキスパートシステムで使用されたという論文があるだけで、具体的なアプリケーションが発見できませんでした(Google検索で287件)。

「YES・NOだけを使う対話型AI技術」というところが、いかにもエキスパートシステムっぽいのですが、そのエキスパートシステム自体が、ほぼ全滅している状況(「陰湿な人工知能 ～「ハズレ」の中から「マシな奴」を選ぶ」)ですから、仕方ないかもしれません。

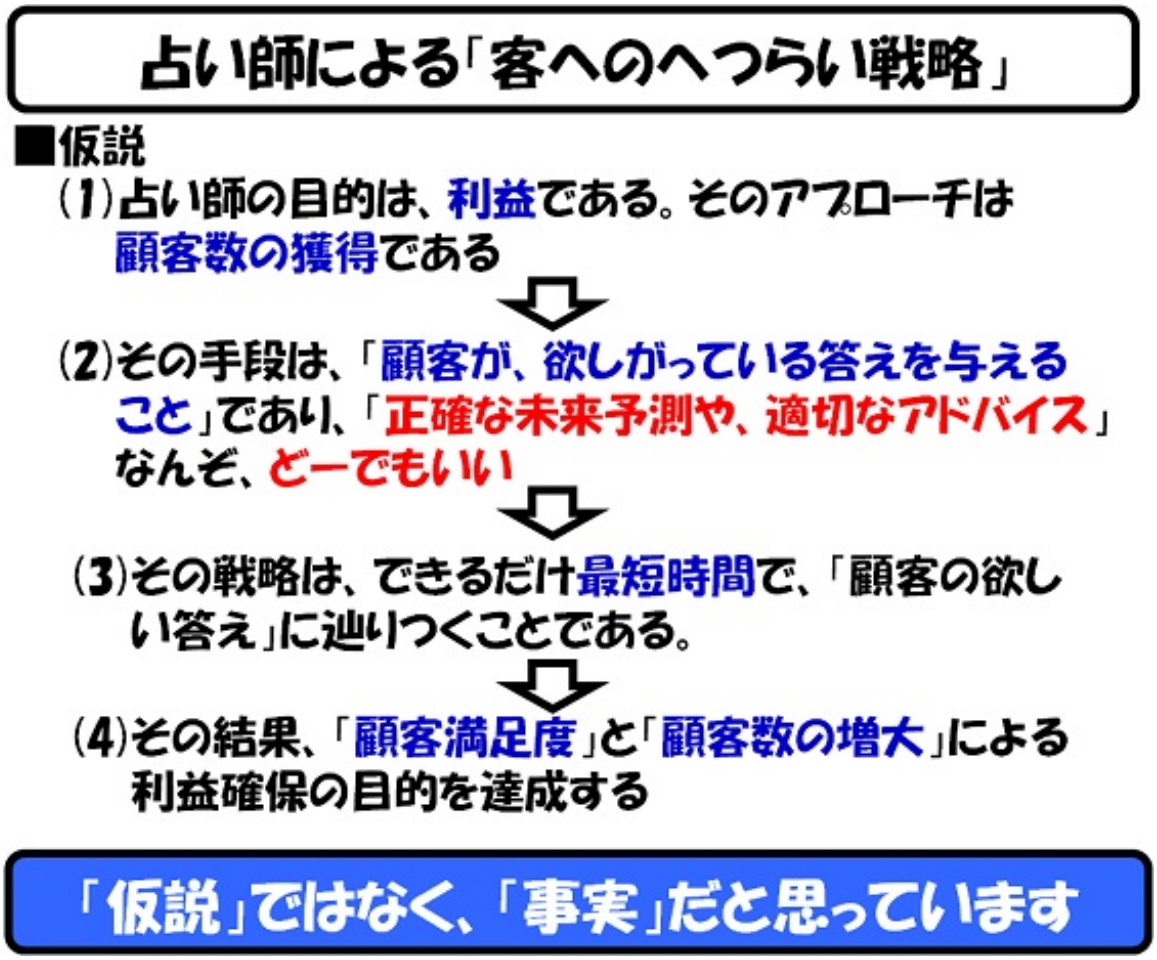
　しかし、この「バージョン空間法」の考え方は、コンピュータの世界ではなく、私たちの現実の世界で役に立っているのです。

占いの、もう1つの意義とは

　さて、ここで「占い」の話に戻ります。

　私は、冒頭で、『占いの目的が「正確な未来予測」や、「その予測に基づく適切なアドバイス」ではなく、それらとは全く異なる意義がある』と申し上げました。

　それは、「客に『へつらう』」という意義です。



なぜ、私が、あれほどまでに「占い師」に対して激怒したかというと、その占い師が、自分の技能や技術に基づいて、未来予測をするわけでもなく、適切なアドバイスをするわけでもなく、「私の気持ちに『へつらう』」ということだけに、30分という時間と3000円という金を、私から奪取したからです。そして、その程度のことで成り立つビジネスに、燃えるような義憤を覚えずにはいられなかったからです。

私は、その占い師の（持っているのであろう）特殊な技能と技術によって、私の考えに対する否定、修正、あるいは逆提案を期待していたのです。「情報量ゼロ」に対して3000円もの大金を払わされたことが、悔しくて仕方がなかったのです。

嫁さんにこの話をしたら、もっとすごい話が返ってきました。

私の嫁さんは、独身のころ、結婚したいと思っている相手（私ではない）がいました。しかし、お義母さんの方は、その人と自分の娘を結婚させたくなかったのです。

そこで、この2人は、それぞれ別の日に、同じ占い師のところに出向いて、相談をしてきました（もし意図してやったとしたらすごい母娘だと思いますが、本当に、偶然だったようです）。

そして、その占い師の回答は、以下の通りだったそうです。

嫁さんへの回答→「その人とは、絶対に添い遂げるべきです。大切なのは相手に対する愛情

です。愛があれば、それ意外のことは、後から付いてくるものです」

お義母さんへの回答→「その人との結婚は、絶対に諦めさせるべきです。結婚というのは、愛だけでは成り立たないものです。若い娘さんには、それが見えていないのです」

占い師というのは、私たち一般人にはない、特殊な能力や機器（霊視・霊感という第六感、タロットカード、易という特殊な道具）を用いて、絶対唯一の未来を予見し、そこから、依頼者に対する最善の選択肢を提示するのが仕事ですよね。なのに、

なんで、完全同一の唯一対象に対して、2つの真逆の未来と、2つの真逆の選択を提示できるんだ？

もうご理解いただけたと思いますが、占い師というのは、良く言えば「背中を押してくれる」存在であり、普通に言えば「普通に何もせず」「いてもいなくても、どーでもよく」「客に『へつらう』」存在であるということです。

「バージョン空間法」を、「客へのへつらい戦略」で説明しよう

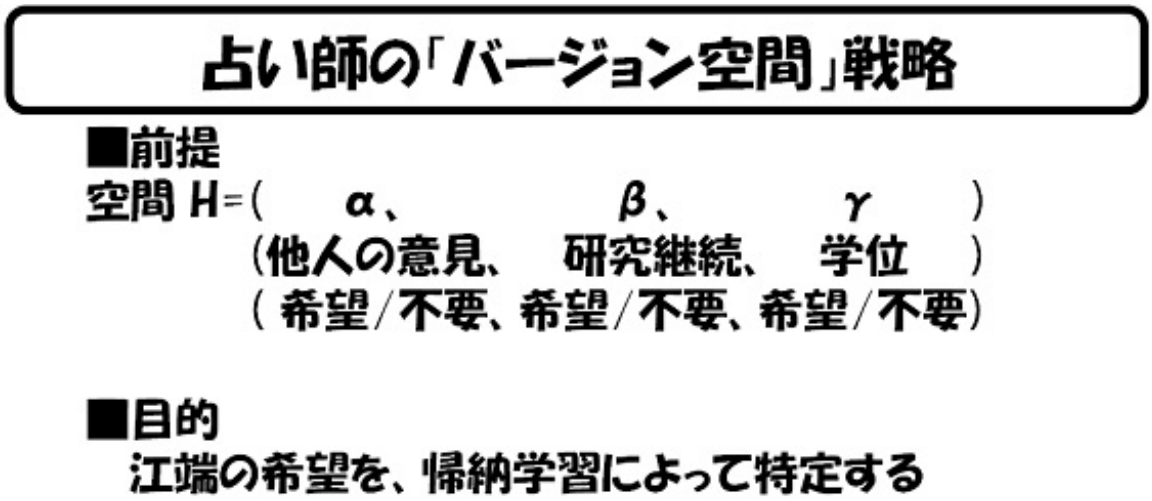
ここからは、「バージョン空間法」を、占い師の「客への『へつらい』戦略」の事例で説明します。

まず、占い師は、冒頭の江端の悩んでいる内容

「現在の大学に残れるのか、残れなければどの大学を受験すべきか、その場合の学費はどうするのか、地元に戻ることをどう考えるべきか、大学に残り続けるのが良いのか、就職をすべきなのか」

をヒアリングして、江端の関心のある事項を、以下の3次元のバージョン空間として把握することにしました。

(α)他人の意見を参考にしたいのか、そうではないのか
(β)今の研究を続けたいのか、そうではないのか
(γ)進学によって得たいもの(学位)があるのか、そうではないのか



(Step.1)

占い師:「あなたは、先生やご両親の意見を聞いて、自分の研究を続けながら、そして卒業資格(学位)を得たいと思っているんですね」

江端:「そうです」

占い師の質問	江端解答	G空間	S空間
(希望、希望、希望)	YES	(*,*,*)	(希望、希望、希望)

*:「どーでもいい」

この時点において、最も小さいS空間が点として確定しました。ここから占い師は、現時点においては無限大であるG空間を少しづつ小さくしつつ、空間ゼロのS空間を少しづつ大きくしていく作業に入ります ―― 江端に気付かれないように。

(Step.2)

占い師:「先生やご両親のご意見は参考になりますものね。研究を続けるだけであれば、例えば

、企業に入って研究所に勤務するということもできますよね」

江端:「それは、ちょっと違うかな、と思います」

占い師の質問	江端解答	G空間	S空間
(希望、希望、不要)	NO	~~(希望,*,*) or~~ (不要,*,*) or ~~(*,希望,*) or~~ (*,不要,*) or (*,*,希望) or ~~(*,*,不要)~~ ↓ (不要,*,*) or (*,不要,*) or (*,*,希望)	(希望、希望、希望)

*:「どーでもいい」

さて、ここで江端は、占い師の質問について否定しましたので、これはG空間を小さくすることができます。具体的には、「他人の意見を希望しない」「研究テーマの継続を希望しない」「学位取得を希望する」という概念記述の空間にまで小さくさせることができます。S空間の方は、それを大きくさせる事例ではないので、変化しません。

(Step.3)

占い師:「先生やご両親のご意見ご意見を参考にしながら、今やられている研究と別の研究を、他の学校で開始するというのも、考えられそうですね」

江端:「そうですね」

占い師の質問	江端解答	G空間	S空間
(希望、不要、希望)	YES	~~(不要,*,*) or~~ (*,不要,*) or (*,*,希望)	(希望、希望、希望) and (希望、不要、希望) ↓ (希望、*、希望)

*:「どーでもいい」

さて、ここで江端は、研究テーマの継続に対する否定的な見解を、否定していません。つまり、江端にとっては、現在やっている研究のテーマの継続性は、問題にしていないことが判明します。つまりS空間における勉学継続のパラメータは、江端の心の中では、配慮されていないことが

分かりました。一方、G空間においては、質問の内容に会わない解空間が一気に消滅することになります。

（Step.4）

占い師：「それでは、今の研究を継続しないで、さらに進学とか転学もしないで、先生やご両親のお勧めになっていることをされてはいいかがでしょか」

江端：「……」

占い師の質問	江端解答	G空間	S空間
(希望、不要、不要)	NO	~~(α,不要,γ)~~ (*,*,希望)	(希望、*、希望)

*：「どーでもいい」

　江端は、この意見に対して、難色を示していますので、G空間の一部の解空間が失われて、さらに小さくなっていきます。

（Step.5）

占い師：「ご自分のことですから、やはりご自分で決意された上で、今とは別の研究を、進学または転学して続けられる、ということもありえますね」

江端：「そうですね」

占い師の質問	江端解答	G空間	S空間
(不要、不要、希望)	YES	(*,*,希望)	(希望、*、希望) and (不要、不要、希望) ↓ (*、*、希望)

*：「どーでもいい」

　この時点をもって、G空間とS空間が一致し、占い師は、（江端すら気が付いていなかった）江端の内心を解読することに成功しました。

占い師：「では、お告げを伝えます。ご両親や先生の意見を参考にしてもしなくても良いでしょうし、また今の研究を続けても、また研究テーマを変更しても良いでしょう。しかし、1つ確実なことは、あなたは、進学または転学をすべきです！」

という、江端にとっては、「プロセスや環境はどーでもよくて、ただ学位が欲しいだけ」という、なんとも幼稚な希望 ーー 幼稚であるが故に自分自身では認められなかった気持ちへの『へつらい』に、この占い師は成功した訳です。

　しかし、率直に言って、こんなやり方、

ーー 面倒くさいな

って思いますよね。

　普通の占い師であれば（というか、普通の人間であれば）、この様な5つのステップも必要なく、2回程度の質問で、この程度の結論に至れそうです。エキスパートシステムが、当たらなかった理由は、この辺りにあるような気がします。

　しかし、そもそもコンピュータの帰納学習は、どうしてもこのような手法（記号化して、論理式に当てはめる）にならざるを得ません。そのようなコンピュータの制約がある中では、「バージョン空間法」は、比較的イケている方法なのです。

　例えば、FIND-Sアルゴリズムというものがあります。これは、特殊な仮説から開始し，事例と矛盾しないように，仮説の最小一般化を繰り返し，最大特殊化仮説を導くものですが、バージョン空間法のように、2つの空間のサンドイッチみたいなことができません。

　ですから、仮説を特殊化しすぎていたり、一般化しすぎていたりしても、それを検知することができませんし、質問を打ち切るタイミングも分かりません。解答に矛盾があっても、その発見が難しいという問題があります。

　とはいえ、やっぱり、このバージョン空間法を、実際のサービスとして適用するのは難しそうです。まず、概念空間を自動的に作り出す手段がありませんし、解空間の広さによっては、質問の回数が、膨大になり、意味のある時間内で終了できないかもしれません（例えば、占いのブースに2年間くらい泊まり込むとか）。

　ともあれ、機械学習の中の帰納学習の1つである「バージョン空間法」が、人間の思考アプローチに近いものであること、そして、過去の情報をうまく組み会わせて動いていることの2つをご理解いただけたのであれば、十分です。

江端の本音「AI技術によって、全ての占い師を消滅させたい」

それでは、今回のコラムの内容をまとめてみたいと思います。

【1】これまでの3回のAIブームの全てで必ず登場してきた「人工知能が人類を滅亡させる」のフレーズについての検討を行いました

【2】さまざまな文献、ネット情報より、私たちが「人工知能は人類に反乱するものであって、人類を滅亡させて欲しい」と願っていることを明らかにしました。また、そのような種類の文献の著

者は、AI技術に対して絶望的なまでに無勉強（×不勉強）であり、AIとAI技術の違いを理解しておらず、仮に理解していても、AIの技術内容の理解が、目もくらむほどデタラメであることを、実例で明らかにしました

【3】「人工知能による人類滅亡論」を唱えている人の、記述レトリックを明らかにして、それが論理破綻していることを明らかにしました。また、実際のAI技術の内容の検証を行い、AI技術単体であっても、それらを複合されたものであっても、人類を滅亡させるAI技術のシナリオは作れないことを明らかにしました

【4】一方、コンピュータの誤動作等による人類滅亡の危機は、これまでも現実に存在していたこと、また、今後もその可能性が十分にあることを、過去の実際の事故を例示して明らかにしました。加えて、「人工知能による人類滅亡論」がどれほどナンセンスであったとしても、エンジニアへの警鐘として意味がある、という江端見解を、「原発事故の発生確率『10億年に1度』という安全神話の理論の破綻」の事実から論述しました

【5】「占い師の目的が『利益』であり、その手段として『顧客の欲しがっている答えを発見して『へつらう』（だけ）』」という仮説を使って、機械学習の1つである、帰納学習を行う「バージョン管理法」について説明しました

【6】また、「バージョン管理法」に基づく占い師の行動をシミュレーションして、その手法のアルゴリズムを説明しました

以上です。

今回は、「人類には『AI技術によって人類を滅亡させて欲しい』という願望がある」、という、江端持論を展開しましたが、もちろん、私は、これが一般的な社会通念であるとは思っていません。

しかし、私（江端）の「AI技術によって、全ての占い師を消滅させたい」という願望は、うそ偽りなく本当です。特に、「情報量ゼロ」の「私の気持ちに『へつらう』」だけの占い師なんぞ、まとめて全滅したいです。その程度のことをするだけなら、占いブースには、PCが1台鎮座しているだけで十分です。

ところで、PCと言えば、学生時代（の学園祭で）、「PC占い」は大人気でした。多くの大学生たちが「コンピュータってすごいね！　すっごく当たっている!!」と叫んでいましたが ―― その占いのプログラムをPC9801に入力していたのは、当時、大学生だった私の嫁さんです。

嫁さんは、PC専門誌に記載されていた、訳の分からない呪文のような文字列を、ただPCに打ち込んでいただけでした。そして、嫁さんは、現在も、プログラミングについて全く理解していません。

それはさておき。

別段、私が熱くならなくても、実際のところ、「占い」を真剣に信じている人は、そんなにはいな

いのかもしれません。

　彼らは、占い師とのコミュニケーションによって、自分の悩みを言語化して第三者に渡すことで気持ちがラクにすることができ、そして、アドバイスを ―― それが情報量ゼロのアドバイスであっても ―― 受け取ることに対して、対価を支払っているのだと思います。

　そして、私たちが、占いによって、自分の心の中を具現化（言語化）したいと思うのと同様に、私たちの「人類を滅亡させたい」という気持ちも具体化したいと思っているのかもしれません。

　しかし、政治も軍事も地政学も分からない私たちは、結局のところ、その方法が分かりません。

　だからこそ、私たちは「人類を破滅させる人工知能」なるものを、考えずにはいれらないのかもしれません。

⇒「Over the AI ――AIの向こう側に」⇒連載バックナンバー

Profile

江端智一（えばた ともいち）

　日本の大手総合電機メーカーの主任研究員。1991年に入社。「サンマとサバ」を2種類のセンサーだけで判別するという電子レンジの食品自動判別アルゴリズムの発明を皮切りに、エンジン制御からネットワーク監視、無線ネットワーク、屋内GPS、鉄道システムまで幅広い分野の研究開発に携わる。

　意外な視点から繰り出される特許発明には定評が高く、特許権に関して強いこだわりを持つ。特に熾烈（しれつ）を極めた海外特許庁との戦いにおいて、審査官を交代させるまで戦い抜いて特許査定を奪取した話は、今なお伝説として「本人」が語り継いでいる。共同研究のために赴任した米国での2年間の生活では、会話の1割の単語だけを拾って残りの9割を推測し、相手の言っている内容を理解しないで会話を強行するという希少な能力を獲得し、凱旋帰国。

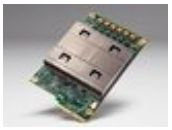
　私生活においては、辛辣（しんらつ）な切り口で語られるエッセイをWebサイト「こぼれネット」で発表し続け、カルト的なファンから圧倒的な支持を得ている。また週末には、LANを敷設するために自宅の庭に穴を掘り、侵入検知センサーを設置し、24時間体制のホームセキュリティシステムを構築することを趣味としている。このシステムは現在も拡張を続けており、その完成形態は「本人」も知らない。

本連載の内容は、個人の意見および見解であり、所属する組織を代表したものではありません。

関連記事

人身事故を「大いなるタブー」にしてはならない

「人身事故」という、公で真正面から議論するには“タブー”にも見えるテーマを取り上げた本シリーズも、いよいよ最終回となります。今回は、「飛び込み」を減らすにはどうすればいいのか、という視点を変え、「飛び込み」さえも構成要素として取り込む鉄道インフラシステムについて考えてみたいと思います。



GoogleからAI用プロセッサ「TPU」が登場

Googleが、人工知能（AI）に向けたアクセラレータチップ「Tensor Processing Unit（TPU）」を独自開発したことを明らかにした。同社が2015年にリリースした、オープンソースのアルゴリズム「TensorFlow」に対応するという。

理研、東芝とNEC、富士通の3社とAI研究で連携へ

理化学研究所（理研）は2017年3月10日、東芝、NEC、富士通の各社と、理研革新知能統合研究センター内に連携センターを開設する。設置期間は、2017年4月1日から2022年3月31日までの予定だ。

近代科学の創始者たちに、研究不正の疑いあり（天動説の「再発見と崩壊の始まり」編）

プトレマイオスの「数学集成（アルマゲスト）」を「再発見」することに大きく貢献したレギオモンタヌスは、「再発見」以降に同書を最初に批判した学者でもあった。「数学集成（アルマゲスト）」の欠点に気付いたレギオモンタヌスは、新しい天文学理論の構築に取り掛かる。しかし、レギオモンタヌスが早逝したことにより、その試みは、ついえる。後を引き継いだのが、地動説への一大転換を果たすことになるコペルニクスであった。

運転者の好みを学習する車載向けAI技術

ニュアンス・コミュニケーションズ・ジャパンは、車載インフォテインメント向け人工知能（AI）技術について、記者説明会を開催した。

AIの“苦悩”――どこまで人間の脳に近づけるのか

人工知能（AI）の研究が始まった1950年代から、AI研究の目的は「人間の大脳における活動をいかにコンピュータ上で実現させるか」だ。大手IT企業や大学の努力によって、AIは少しずつ人間の脳に近づいているのは確かだろう。一方で、自然言語処理の分野では、“人間らしさ”を全面に押し出した「人工無能（人工無脳）」も登場している。