Over the AI ―― AIの向こう側に（20）：

忖度する人工知能 ～権力にすり寄る計算高い“政治家”

http://eetimes.jp/ee/articles/1803/27/news040.html

今回取り上げるのは「強化学習」です。実はこの強化学習とは、権力者（あるいは将来、権力者になりそうな者）を“忖度（そんたく）”する能力に長けた、政治家のようなAI技術なのです。

2018年03月27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

「強化学習」の本編に入る前に……

2012年、桜宮高校の男子バスケットボール部員が指導者の日常的な体罰に耐え切れず自殺した事件がありました。教育現場において「体罰」に「効果」があるのは当然です。「銃を付きつけられたら、誰だって従う」という野蛮な理屈と同じであるからです。

しかし、私は、わが国は、「体罰」に「効果」があるのは知っているが、それを「一切やめよう」と決めた国であると思っています。

クラウゼヴィッツを出すまでもなく「戦争は政治（外交）の延長である」のは自明です。しかし、戦争が、どんなに優れた外交的効果が期待できたとしても、わが国は、

―― 決して戦争をしない

―― 決して核兵器を持たない

と決めました。

そして、わが国の教育現場は、「体罰」を絶対的な意味において封印した ―― この教育方針は、わが国の平和憲法の根幹である「戦争放棄」の理念と並び立つ、わが国が誇る教育信条であると、私は信じています。

「今まさに、自殺をしようとしている子供を力づくで止めるために、やむなく体罰を行使した」という、緊急避難的な状況での話ならともかく、この男子バスケットボールの指導者は、たかだか「チームを強くするため」程度のことに「体罰が有効だ」と語ったそうです。

多分、この教師は人気があって、生徒からも同僚からも信頼を得ていて、間違いなく人間として尊敬するに足ると思われている人間だろうとは思います。なにしろ、「校長」から体罰を「看過」してもらっていた程の人間ですから。

だが、私にとって、その教師の人柄や人格などは、知ったことではありません。

この教師が、「体罰」をいう手段を「効果」のために使った ―― この一点において、私はこの教師を絶対に許さない。

教師とは『「体罰」以外のありとあらゆる手段を考えて、試して、実施して、それでもうまくいかなくて、悲しくて、辛くて、悔しくて、そうして一人で泣く』―― そういうものだろう、と私は思うのです。

私は、学生のころ、ギリギリまで迷いに迷って、最後の最後で「教職」を断念した人間です。

私は、自分のことを『効率的な手段を求めて、簡単に「体罰」を行使してしまう、浅学で狭量で卑怯な人間だ』ということを自覚していたからです*)。

*)学習塾で、私は成績の良い子どもには優しく接して、成績の悪い子どもには厳しく当たる ―― そんな講師でした(参考:著者のブログ)。

だから、私だけは、次のように声高に主張する資格がある、と思っています。

教師という、この世で一番ツラい職業を選んだのであれば、一番ツラい道を歩け

その覚悟がないなら、最初から教師などを職業として選択するな

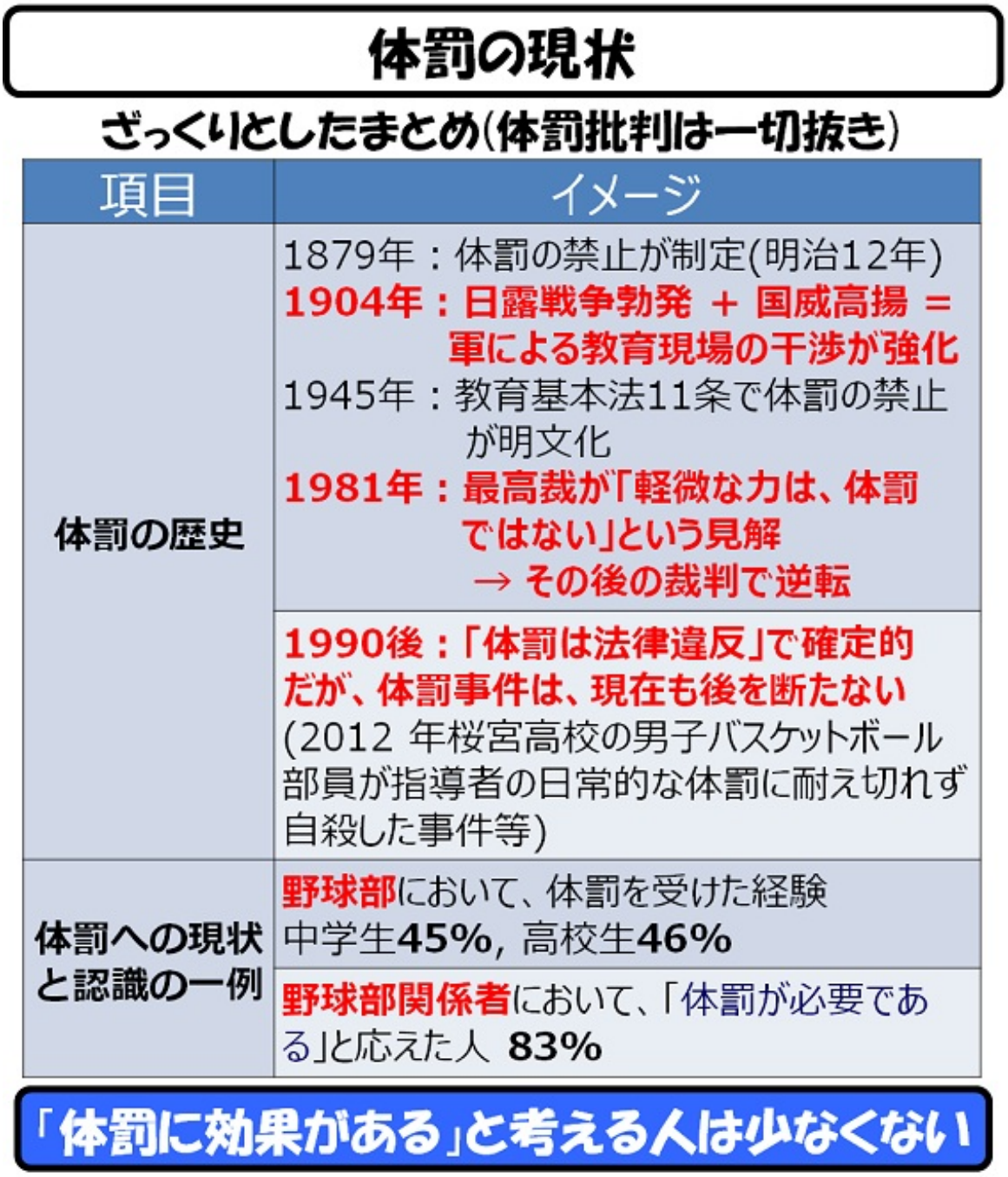
―― と*)。

*)参考:著者のブログ

―― え?これはAIを語る連載ではにのかって?大丈夫です。ご安心ください、ちゃんとこの伏線は回収しますから。

まず、私の「体罰」に対する、私の嫌悪(憎悪)を明らかにした上で、ここからは、分析を行うエ

ンジニアとして「体罰」について論じてみたいと思います。



出典：「日本の運動部活動における体罰の研究」

そもそも、明治維新後の義務教育制度の発生の時から、教育現場における体罰というのは、法律で禁じられている行為でした。ところが日露戦争から太平洋戦争の終結に至るまで、国威高揚のため、教育現場での軍の干渉が露骨になり、軍の体罰主義が、そのまま教育現場に導入されて、体罰が日常化しました。

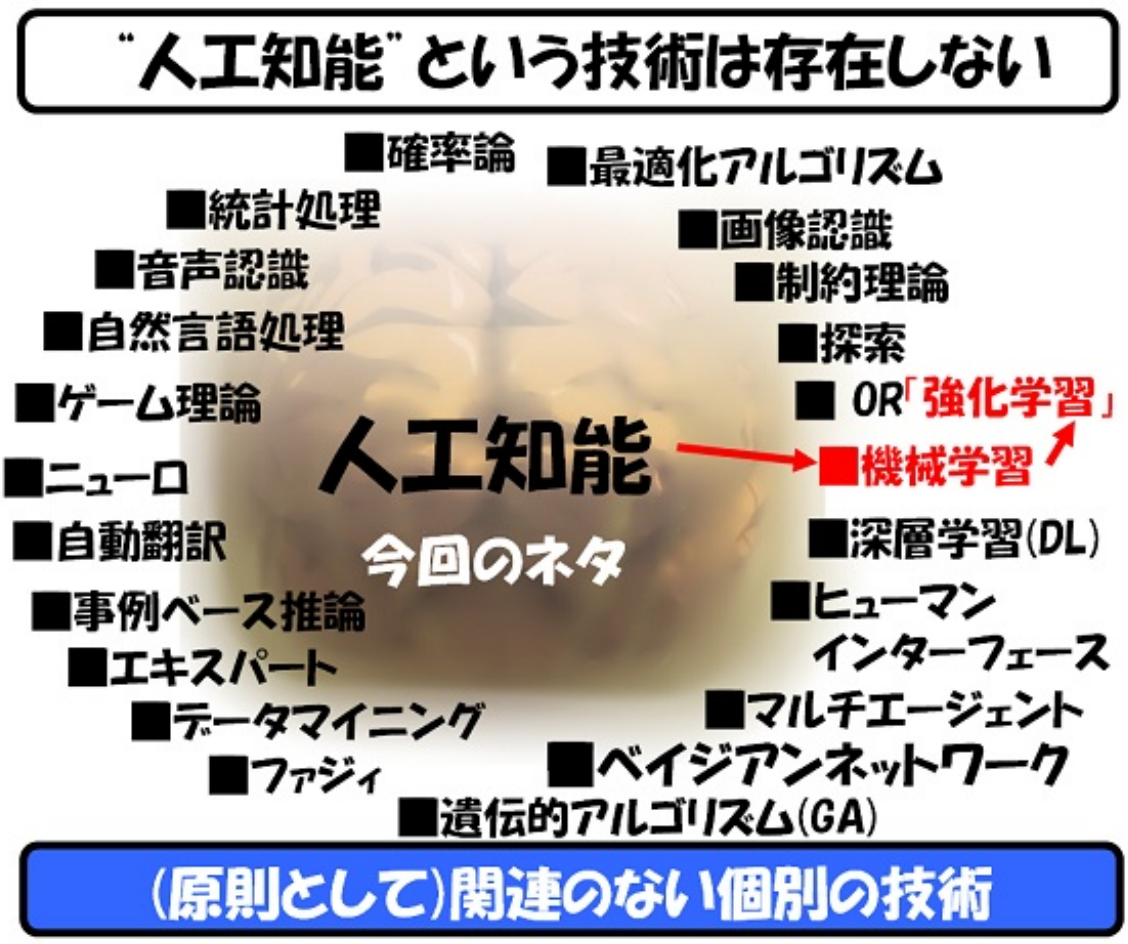
太平洋戦争後、法律によって、体罰の禁止が明文化され、その後、司法判断ですったもんだしましたが、現状においては「どんな事情があろうと体罰は絶対にダメ」が、わが国の国是となりました ―― の、はずなのですが。

上記の表にも記載しましたが「体罰が必要である」と考える人は、少なくないのです。「黙っているだけ」という状況が見て取れます。また、体罰の被害者が、体罰を容認しているようなのです。

つまり、口では「体罰はダメ」と言いつつ「体罰には効果がある」と考える人が一定数いるということです ―― これを、欺瞞だの、非人道的だの、反社会的だのと批判するのは簡単ですが、今回、私は、これを一人のエンジニアとして、AI技術の観点から考察してみたいと思います。

「忖度」のメカニズムを持った「強化学習」

こんにちは、江端智一です。今回は、「機械学習」の中の「強化学習」についてお話したいと思います。



「強化学習」とは、一言で言えば「報酬型学習」、もっと簡単に言えば「褒める学習」です。

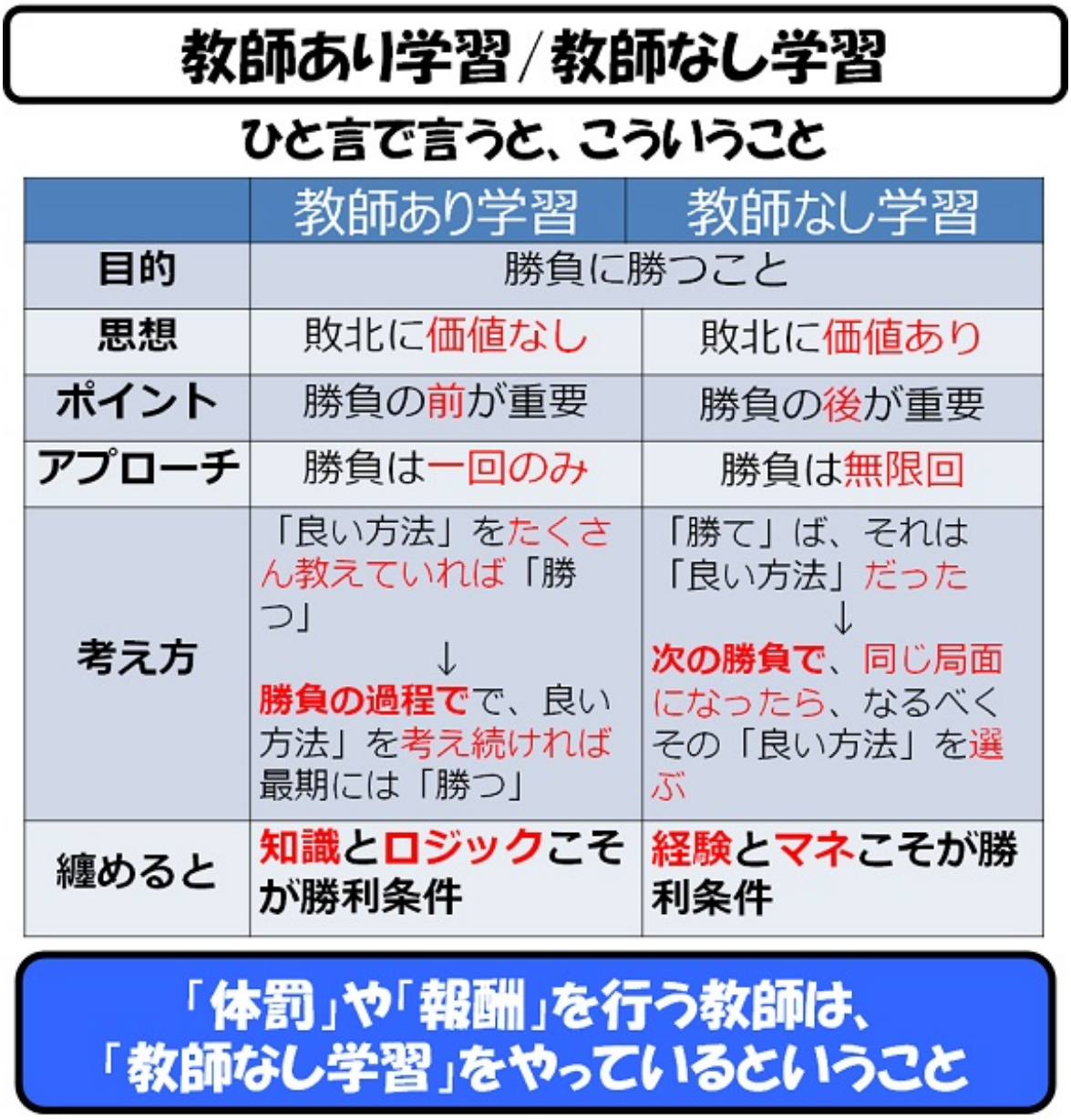
しかし、実際に「強化学習」(例えばQ-learning(以下「Q学習」といいます))をコーディングしてみたところ、これは「権力と忖度(そんたく)」のメカニズムを使ったAI技術である、と感じました(後述します)*)。

*)「忖度(そんたく)」とは2018年3月頃に、日本中が大騒ぎになった「森友学園問題」で有名となったパワーワードです(数年後に、このコラムを読んでいる人には、訳の分からない単語

になっている可能性がありますので、念のため解説しておきたいと思います）。

この「強化学習」は、いわゆる「教師なし学習」に該当するものですが、この「教師あり学習」と「教師なし学習」を混乱して理解している人が多いようです（そういう私も、良く分かっていませんでした）。

「教師なし学習」という言葉から、「自分で考えて行動する（そして、暴走する）人工知能の学習アルゴリズム」と思っている人が多いようですので、最初にその誤解を解いておきます。



上記で説明したので、詳しくは割愛しますが、要するに、この2つは「学習プロセス」の考え方が違うのです。私なりに、「教師なし学習」を乱暴に纏めると、「勝てば官軍」とか「勝ったのであれば、そこに至る過程は全て正しい」とか、なんというか、体罰を看過する教育現場のようです。

実際に、Q学習をコーディングしてみた私が感じたことは ―― 教師なし学習って、やたら練習を繰り返しさせて、『体で覚えろ』と叫ぶだけの、ロジックで語れない「ぼんくら」 ―― です。

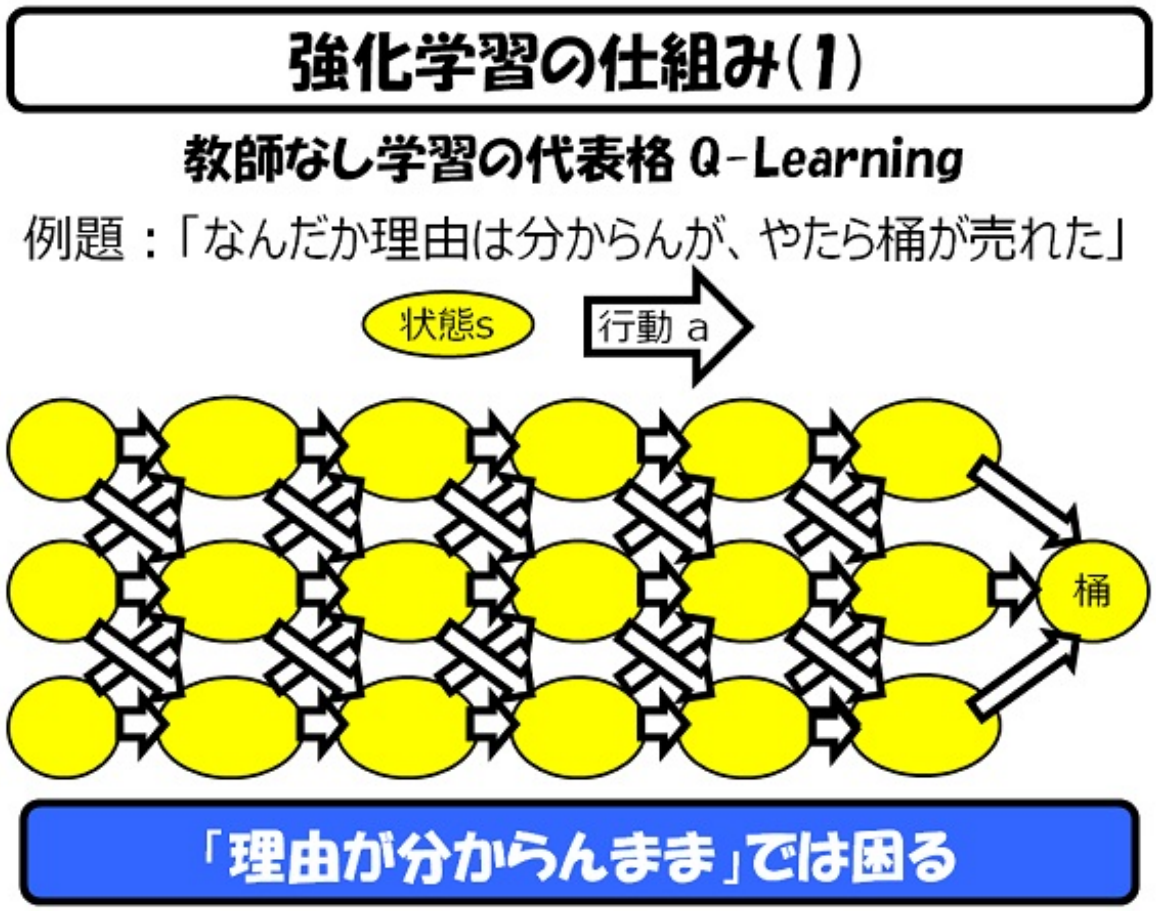
しかし、この結構な「ぼんくら」である「教師なし学習」の「強化学習」と、そして、「教師あり学習」の「深層学習」（次回以降に解説）の2つの学習アルゴリズムこそが、今回の第3次AIブームのきっかけとなり、そしてこのブームを推進しているエンジンそのものであり、そして、第3次AIブームにおいて、私が唯一認めている「2大AI技術」です。

強化学習を「風が吹けば桶屋がもうかる」で考える

では、ここからは、強化学習のQ学習について、数式抜きの解説を試みます。

ここでは「桶の製造販売をしている工房の社長（兼職人）」を想定します。

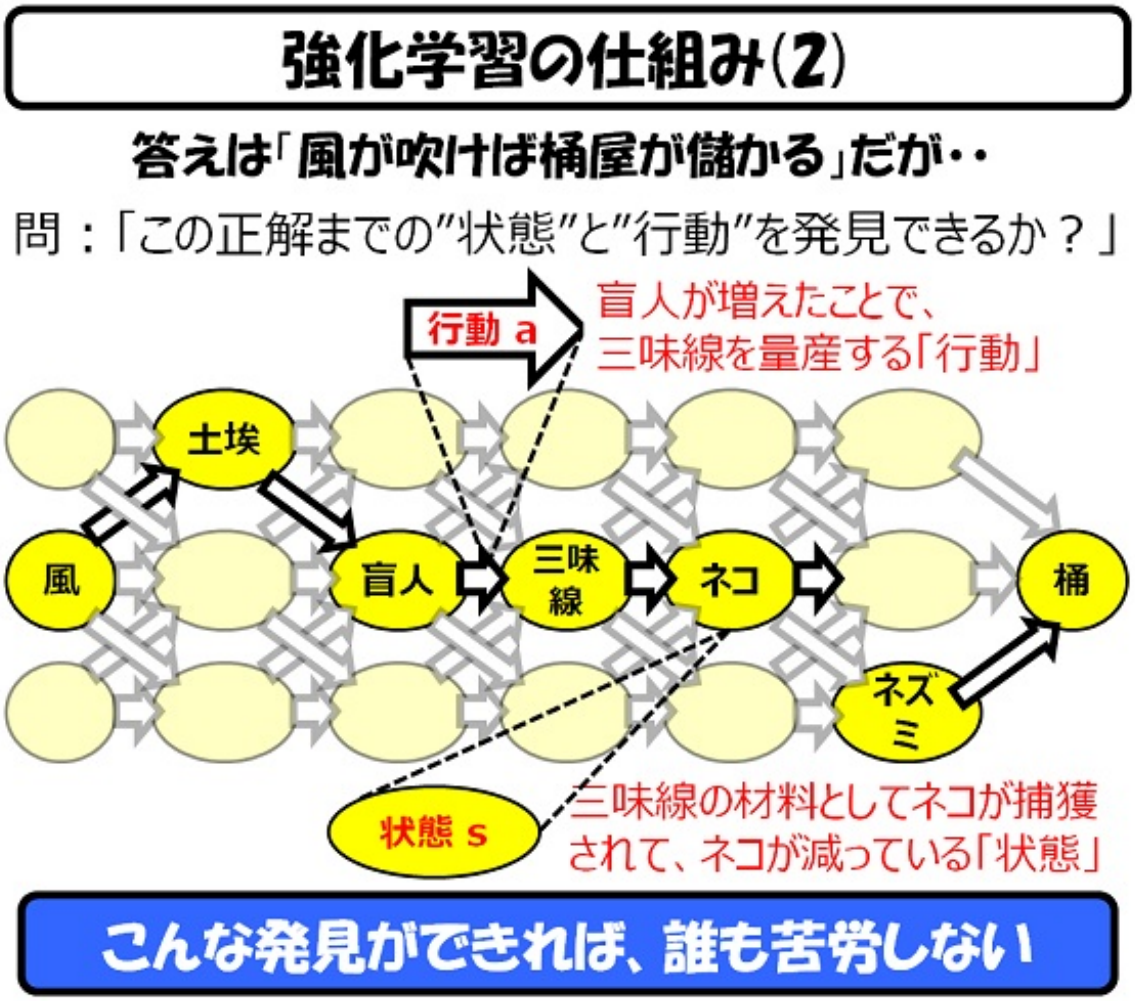
「毎年、なんだか理由は分からんが、やたら桶が売れる」という時期があるとします。その桶の会社の社長としては、桶が売れる理由が分かれば、当然、その時期に桶を増産して売り上げを確保したいと考えるはずです。



どこかで、何かの状態（状態S）が起こり、そこから、その状態を変える行動（行動a）が起こり、別の状態に遷移して、そこからまた別の行動が起こる――。それが繰り返されることで、最終的に「桶が売れた」という状態になるわけです。

しかし、その社長には、その理由が分かりません。

実は、この理由は、いわゆる「風が吹けば桶屋がもうかる」で使われている、「行動」と「状態」の連鎖だったのです。



しかし、桶を作る社長は、このような無数にある世界の「状態」と「行動」から「桶が売れた理由」が、実は「風」だったということや、その途中に「埃の発生」「盲人の増加」「三味線の需要拡大」「ネコの減少」「ネズミの増加」という状態が発生したなど、知りようはずがありません。

もしも桶屋の社長がソフトウェアエンジニアだったなら

ですが、もし、桶を作る社長が、(A)優れたソフトウェアエンジニアであり、(B)AI技術の知見に精通しており、さらに、(C)桶が売れる可能性に関わる世界の全ての状況を「状態」と「行動」として定義でき、かつ、(D)それをソフトウェア上で実装できる ーー と仮定(後述しますが、こんな人間は存在し得ません)した場合、どうなるのでしょう。

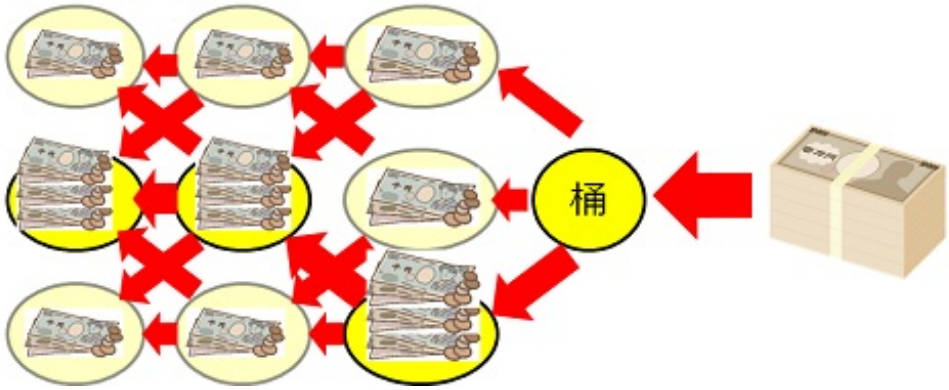
強化学習(のQ学習)とは、ザックリ以下のような仕組みになってます。

まず、「桶が売れる」に至ることのできる世界に至る、全ての「状態」と、その状態を次の状態に変化させる「行動」を定義します。

強化学習の仕組み(3)

起動要件:「予算ばら撒き」
行動原理:「権力」と「忖度(そんたく)」

●**状態**とは、「ばらまき予算」で「金(カネ)＝権力」を持つようになること

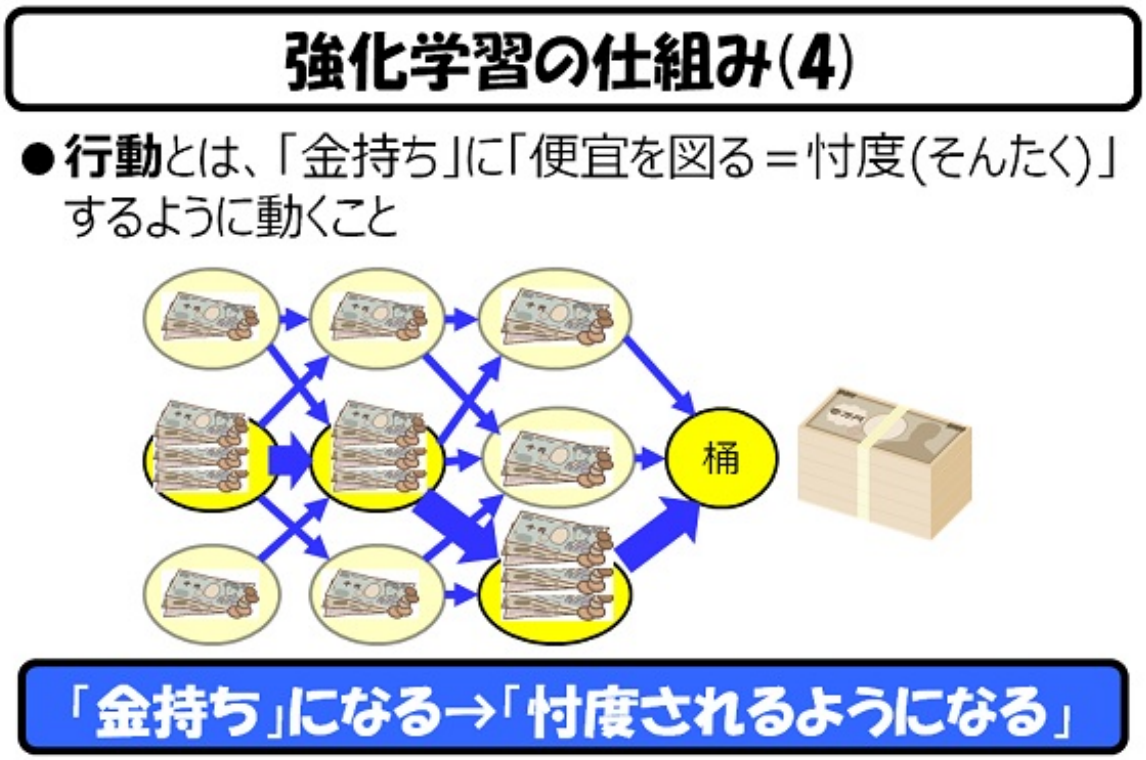


そして、「桶が売れる」という状態(これを、「桶の状態」ということにします)に至れた「行動」に対して、「桶の状態」は、お金を支払います。その結果、「桶の状態」に至ることに貢献した「その前の状態」は、お金を受け取ることができます。

「その前の状態」は、さらにその状態に至ることに貢献してくれた、「その前の前の状態」に、お金を支払います。これが、さらに「その前の前の前の状態」……と続きます。

そして、このお金、不思議なことに「どんなに支払っても、減らない」魔法のお金であることを覚えておいてください。

さて次に、「行動」です。「行動」はお金持ちが大好きです。ですので、「行動」は、お金持ちに成り上がった「次の状態」に状態を変化させるように働きます。



ただ、この「行動」は、狡猾（こうかつ）な奴で、常に「金持ち」をチヤホヤするだけではなく、「貧乏人」にも、小さい確率で移動するように行動します。将来、その状態が、「金持ち」になった時に、「あ、しまった」とならないように、ちゃんと「コネ」を作っておくためです。

これは、企業が政治献金を行う時に、与党へはもちろん、弱小野党へも、額が小さくても献金を怠らないこととよく似ています（政権がひっくり返っても、コネがあれば、なんとかなります）。

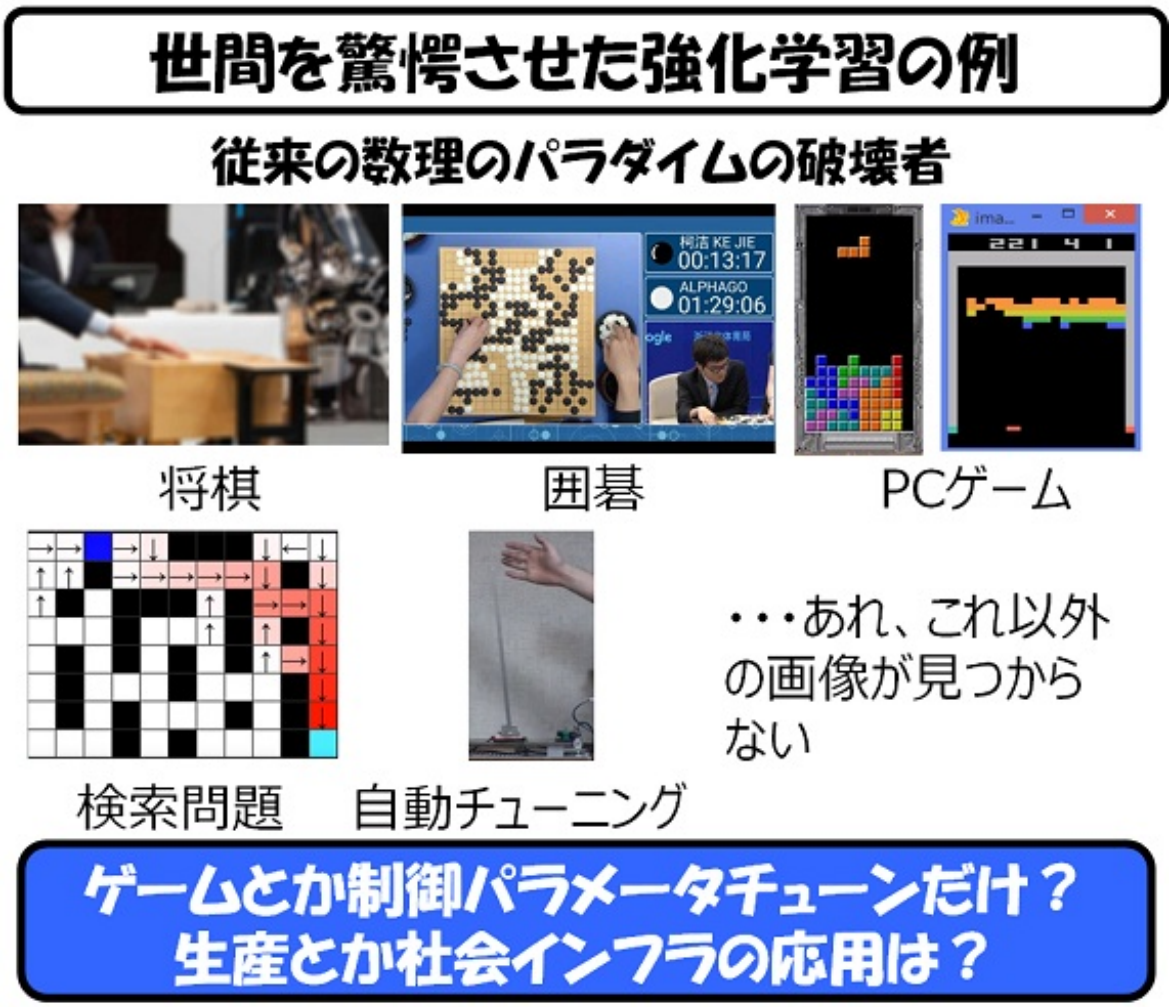
で、この「状態」と「行動」を、山ほど（ケースにもよりますが、数百回から1億回までさまざま）繰り返します。繰り返さないと、「お金」が貯まりませんし、お金が貯まらないと「権力」が発生しないからです。

さて、こうして強化学習のQ学習を俯瞰してみると、この仕組みが、実に単純な政治の利益誘導モデル（「予算ばらまき」と、「（金による）権力」と「権力への忖度」）で動いていることが理解頂けるかと思います。

　もっとも、AI技術の世界では、当然、Q学習を「権力／忖度モデル」などとは言わず、「報酬型学習」と言います。

「強化学習」の有効さ

　この単純な報酬型学習である「強化学習」が、どれほど有効であったかは、皆さんもご存じの通りです。



チェスはともかく、絶対に無理だと言われ続けた将棋、囲碁の世界チャンピオンが倒されたことは、記憶に新しいと思います*)。

*）ただし、「Bonanza」「アルファ碁」「AlphaZero」は、単純にQ学習を適用しただけではありません。

コンピュータゲームへの適用は、これより早く始まっていて、積み木くずし、テトリスなども、簡単に攻略されました。

この他、Q学習は、検索問題や枝分かれ問題のような、組合せ爆発系の問題とも親和性が高く、古くから研究が続けられてきた（比較的簡単な）制御装置の自動チューニングにも使えることが分っています。

しかし ーー これだけなのです。

これ以外の事例（例えば、産業応用とか）を、私は見つけ出すことができませんでした。

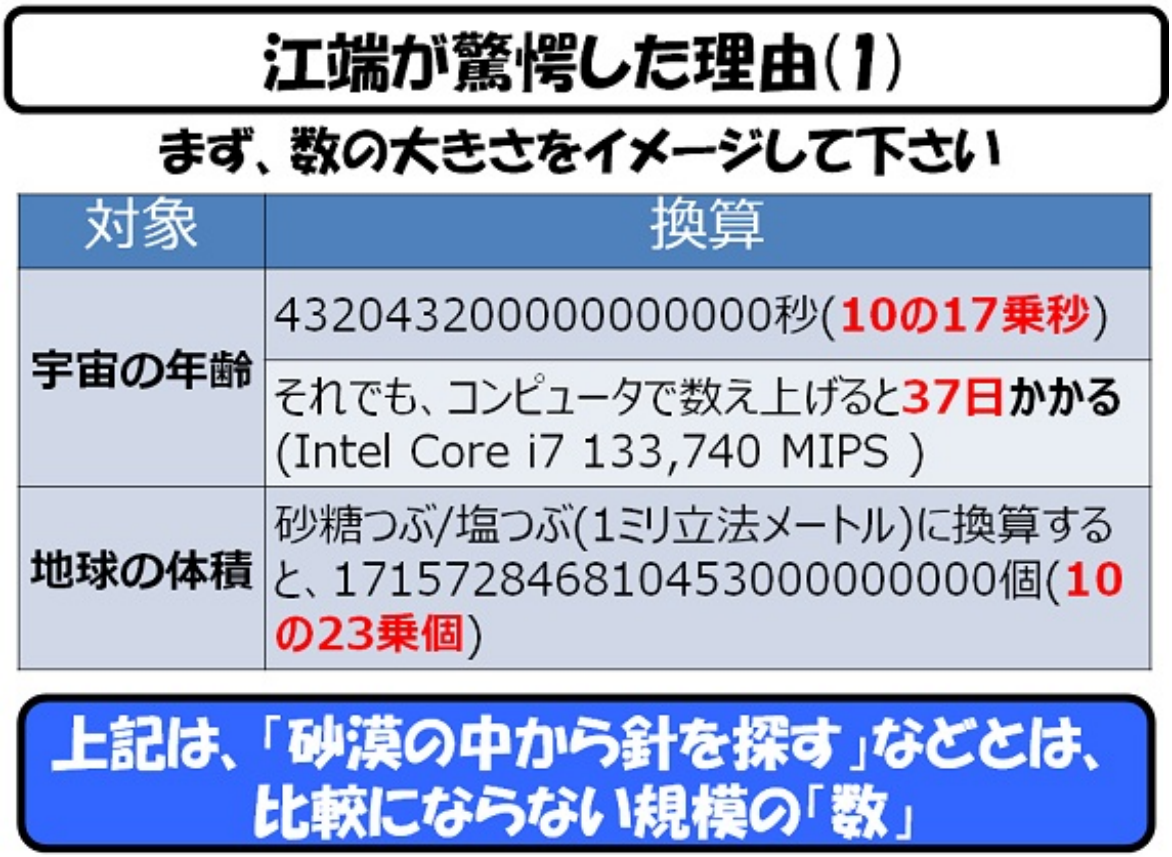
これは、強化学習だけに限らず、今回の第3次AIブーム全体にいえることです。

AIブームは、社会のツールとして組み込まれた時に、そのブームが終焉（えん）しますので、これは仕方のないことかもしれません（関連記事：「陰湿な人工知能 ～「ハズレ」の中から「マシな奴」を選ぶ」）。

しかし、そうであったとしても、強化学習は私たちAI技術に関わる者の、既存のパラダイムを破壊してしまうほど、すごい技術であることは否めません。

まず、前述したように、その仕組みが驚くほど単純であることは言うまでもないのですが、私たち（特に私）を心底驚かせたのは ―― 「解空間の規模」だったのです。

まず、以下の図を、（流し読みせずに）ちゃんと読んでください。



私たちエンジニアが、「大きな数」と言われた時に、イメージする数とはこのくらいのものです。

では、強化学習が相手にした「将棋」や「囲碁」の世界が、どれくらい広いかというと、こんな感じです（流し読みせずに）ちゃんと読んでください（本日2度目）。

将棋や囲碁で、AIを勝利に導いた1つの数式

本来、コンピュータの「売り」は大量・超高速計算にあり、その本分は全検索にあります。

しかし、上記のように、将棋や囲碁の解空間は、今の世界中のコンピュータが束になっても、―― 仮に将来、汎用量子コンピュータが誕生したとしても ―― てんで相手にならないほど巨大なのです。

ですので、私たちエンジニアは、コンピュータが「将棋」や「囲碁」に勝つことができても、それは、定石を引用して「たまたま勝つ」とか「素人には勝つ」とか、その辺のレベルを、ウロウロとしているだけで、人類の寿命（あと10万年くらい？）の方が、先に来るだろうと思っていたのです。

ところが、まさか「将棋」や「囲碁」のマスターを、完膚なきまでにたたきつぶすことになろうとは、想像できなかったのです ―― たかだか、この式一つで。

$$Q(s_t, a) \leftarrow Q(s_t, a) + \alpha \left[r_{t+1} + \gamma \max_p Q(s_{t+1}, p) - Q(s_t, a) \right]$$

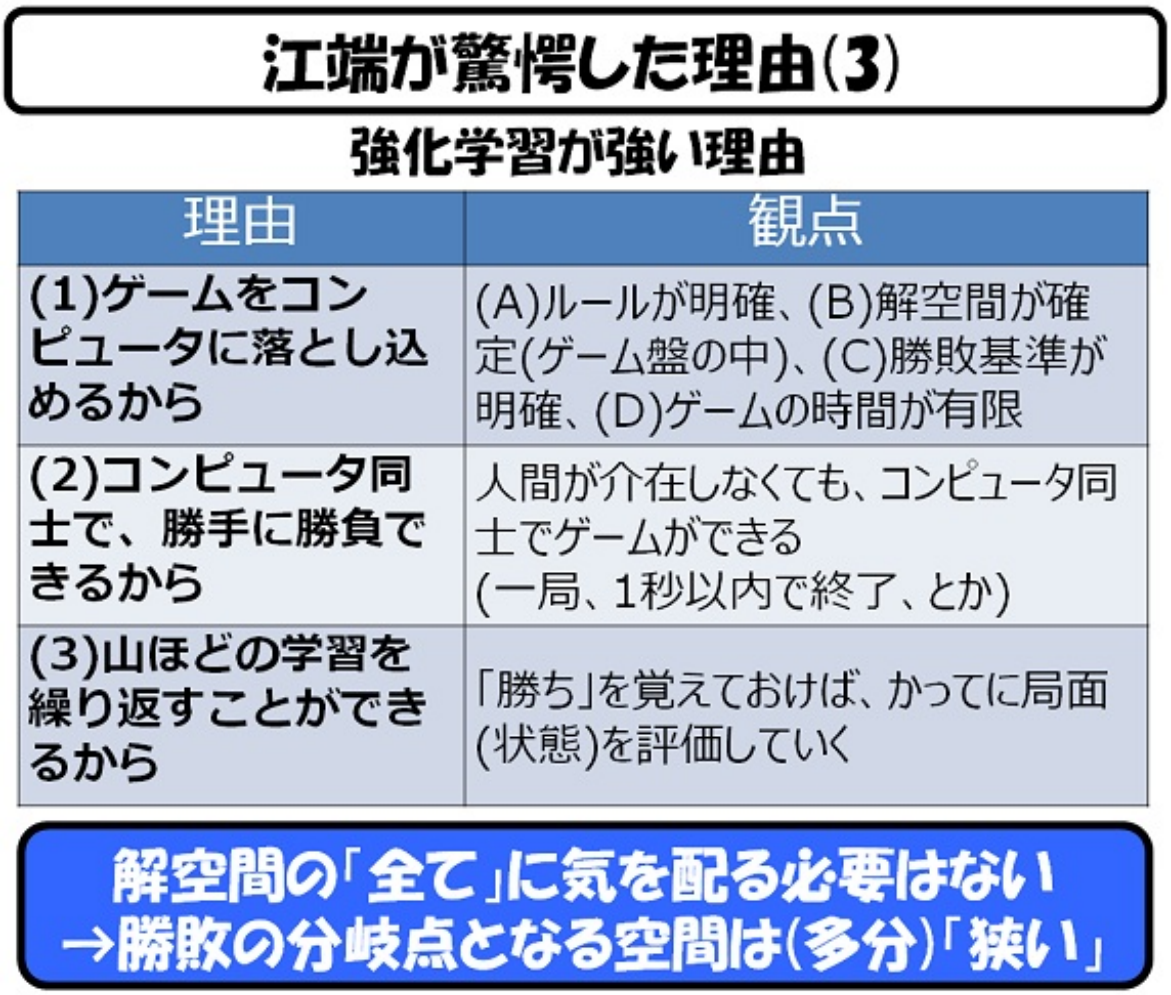
後述しますが、私も、実際にプログラムのコードに落としてみたのですが、この式の部分だけ

なら、4行しか使っていません。

私がここまで言うのは、「AI脅威論」を唱える人間よりもタチが悪いかもしれませんが ―― このたった「4行」が世界中のコンピュータ全てを集めてもできないことをやり遂げてしまった ―― と、私には思えてしまうのです。

なぜ、こんな単純な強化学習アルゴリズムが、私のパラダイムを破壊するほどのことができたのか、この機会に考えてみました。

私なりに考えた結果、以下の3つのことが言えるのではなかと思っています。



「5並べ」の強化学習プログラムですら、設計できなかった

「なーんだ、こんな簡単なら、私でもチョイチョイっと作れるな」と ―― 私は、「5並べ」の強化学習プログラムを作ろうと、頭の中でコーディングを始めました（私は、囲碁も将棋もマージャンも、ゲームと名前のつくもののほとんどのルールを知りませんので）。

ところが、これが「全く」ダメダメだったのです。強化学習の「4行」は、問題なかったのですが、別のところで、どうしようもないことが分かってきたからです。

ほんじゃあ、ちょっと作ってみるか

「5ならべ」の強化学習プログラムを、頭の中で組みたてている最中に**唖然とした**

ポイント	観点	評価
ゲーム進行は？	(A)白黒が交互に差せば良い。 (B)碁石のないところなら、どこでも打てることにする (C)8目盤で十分だろう	○
勝利条件は？	5つの碁石が直線に並んでいるのを、力づくで調べれば良いだろう	○
局面(状態)数は？	**あれ？ 変だぞ？ どうすればいいんだ。**{白、黒、なし}で、64目あるから、単純計算で、3の(8x8)乗で、**343368382029251248465784908928 1(10の30乗)状態**	✕

こんな正攻法では、局面(状態)数だけで、パソコンのメモリがパンクする

つまり、強化学習は、その仕組みよりも、その環境（「状態」と「行動」）を定義すること（正確に言うと、「状態」と「行動」のパターン数を極限まで小さくすること）が恐しく難しいのです。

というか、その環境の定義に成功した人だけが、「将棋」や「囲碁」のマスターを倒すプログラムを作ることができた、ということが、ようやく私にも分かってきました*)。

*）以前後輩に「江端さん、大切なのは『学習プロセス』ではなく『局面の設計』ですよ」と言われた意味が、今回、ようやく理解できました。

これを考えていくと、なぜ、私たちが「強化学習」を、義務教育や高等教育過程で使わないのかは、明らかです。

なぜ私たちは「強化学習」を使えないのか

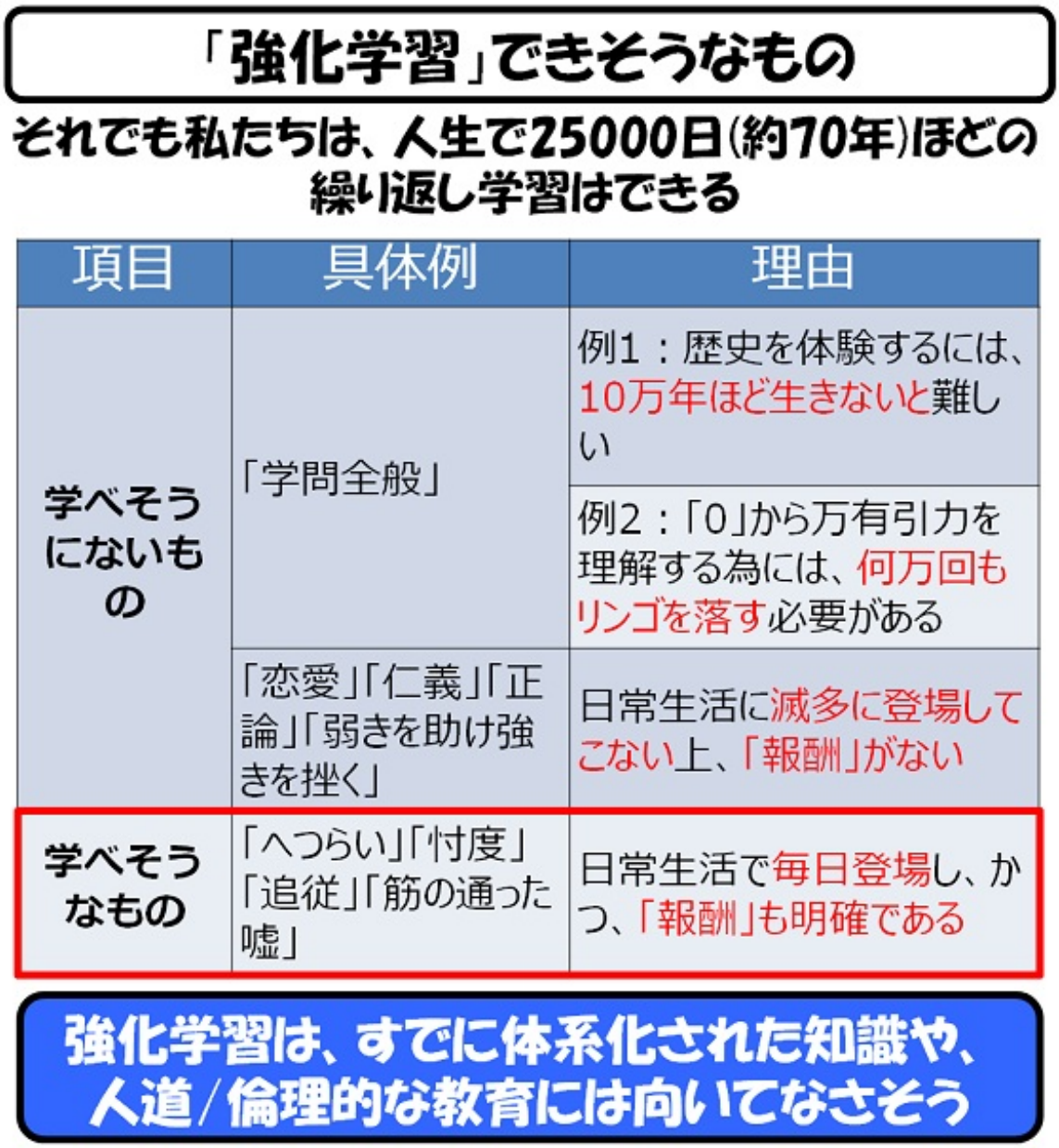
教師や学校がなくならない理由

理由	江端の視点
「成功」が分からない	「成功」って、他人が勝手に決めつけてくるよね？
「成功」から逆算した、過去の全ての「状態」を評価できない	仮に「成功」が存在したとして、そこに至る経緯(状態)なんて分かるか(覚えているか)？
「状態」が分からない	そもそも、その経緯のどの状態を「局面」とするのか。それを、どう評価するのか？
「報酬」が分からない	過去の自分に「こずかい」渡すの？どうやって？
私たちの人生の解空間の広さは、どれくらいだ？	少なくとも、囲碁の「10の360乗」よりも、広いんじゃないかな？
	もっとも、そのいくつかを試す前に死んじゃうだろうしね

人生で「強化学習」をやるには、人生はあまりにも短すぎる

つまり、私たちは、私たちの人生（の局面（状態と行動））を設計することができない上に、強化学習のように、数千〜数億の回数で人生をやり直すことができないのです。

でも、本当にそうかな？ と思い、もう少しつっこんで考えてみました。



人生2万5千日もあれば、そこそこの強化学習はできそうなのですが、毎日登場するような事項にしか、使えそうにありません。

まあ、毎日登場するようなことといえば、上記の様な内容になってしまう訳でして、つまつところ、「強化学習を、義務教育や高等教育の代替とすることは無理」という結論になる訳です。

「学校寄付金プログラム」を作って遊んでみる

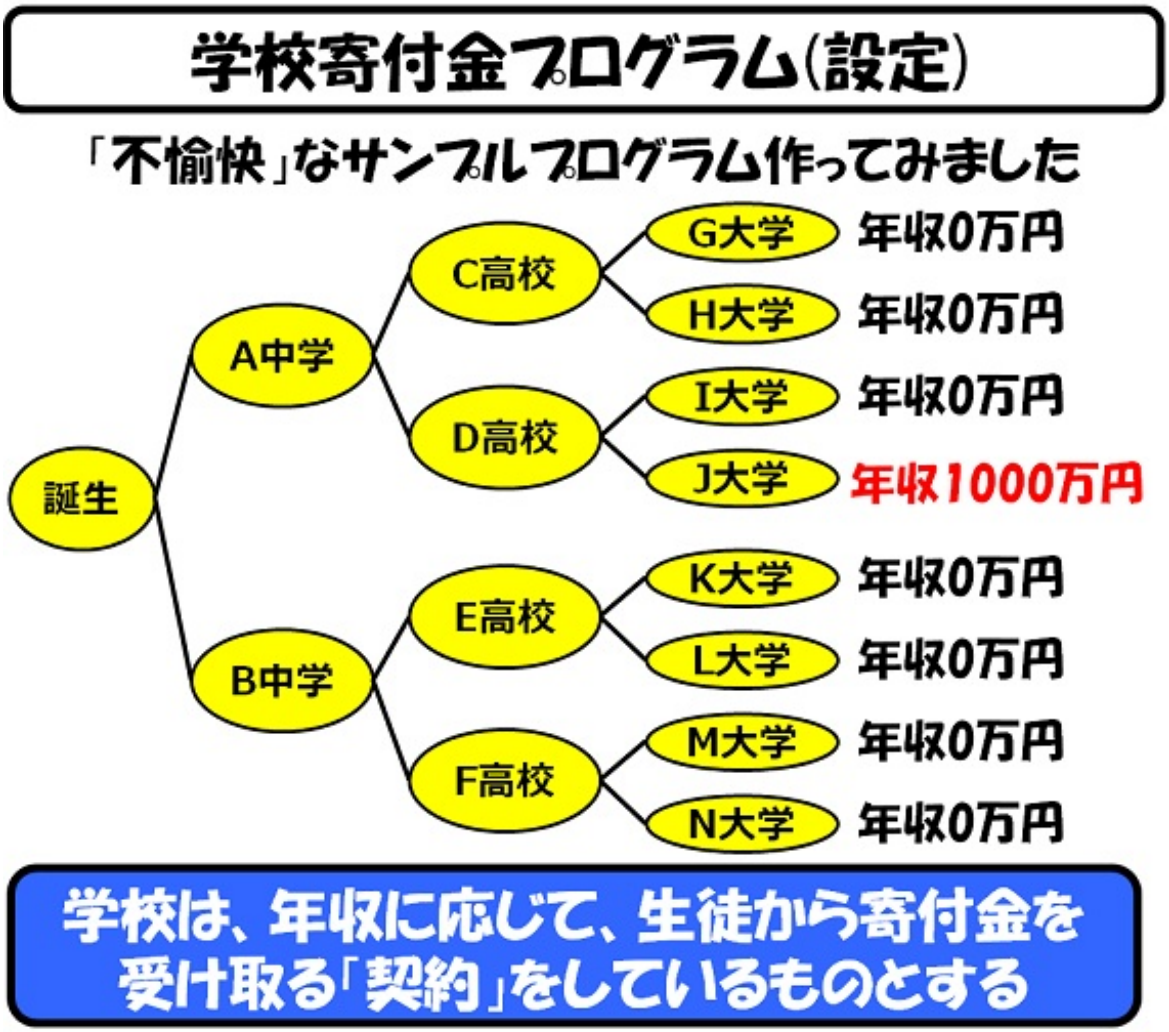
では、最後に、超簡単な強化学習のプログラムを作って試してみたので紹介します。名付けて、学校寄付金プログラム ―― 別名、高年収獲得プログラムです。

このプログラムでは、学費という概念がなく、その代わりに、就職した年収に応じて、その就職に貢献した学校に寄付金（報酬）を渡します。

各学校は、高所得者を生み出すためだけに存在します。それ以外の目的（道徳とか協調とか

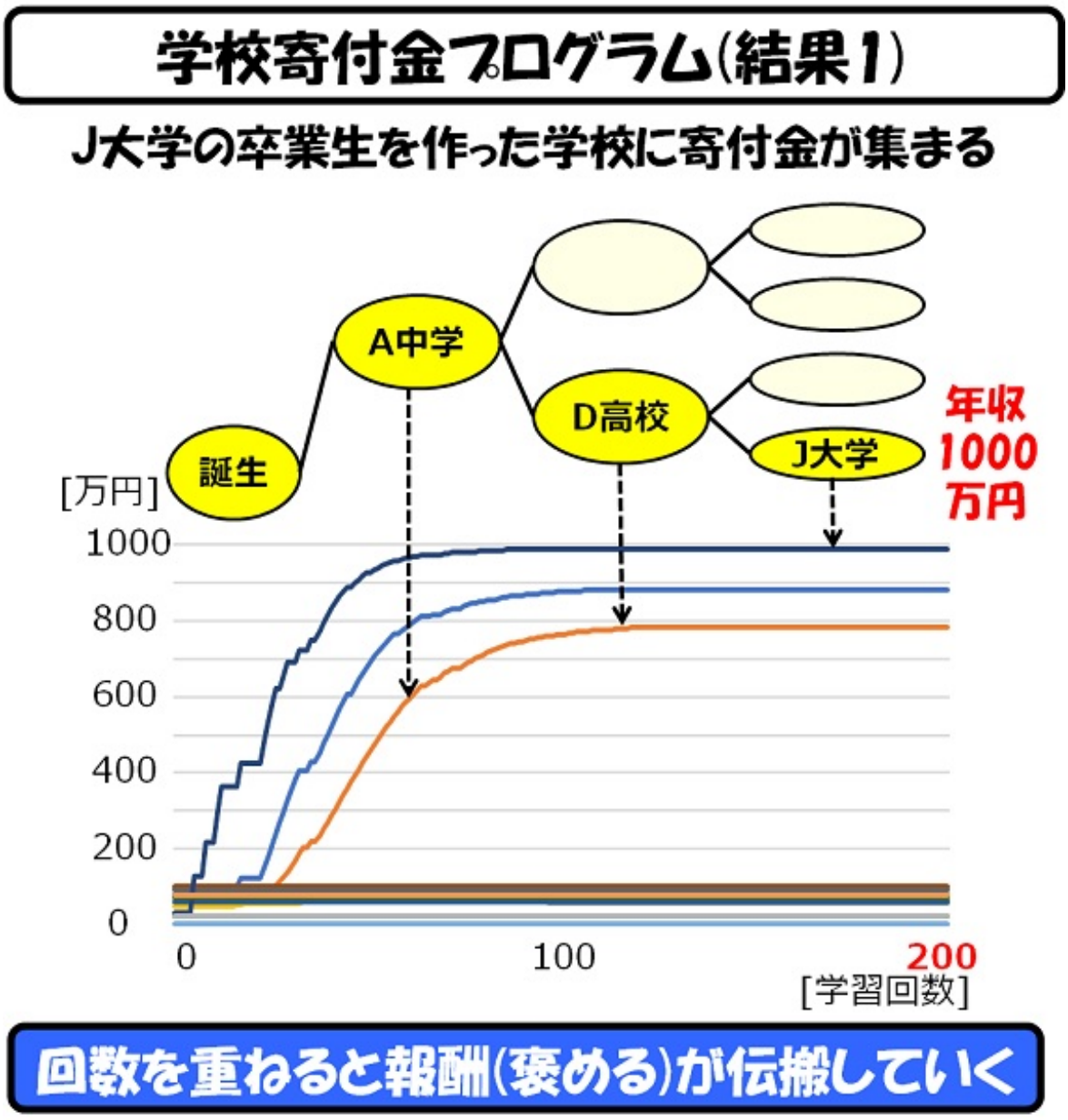
信頼とか友情とか）は、一切無視した教育をします。

　このコラムを読んで頂いている皆さんには、ものすごく不愉快だとは思いますけど、強化学習のアルゴリズムを理解するという点では、とても分かりやすい考え方だと思っています。

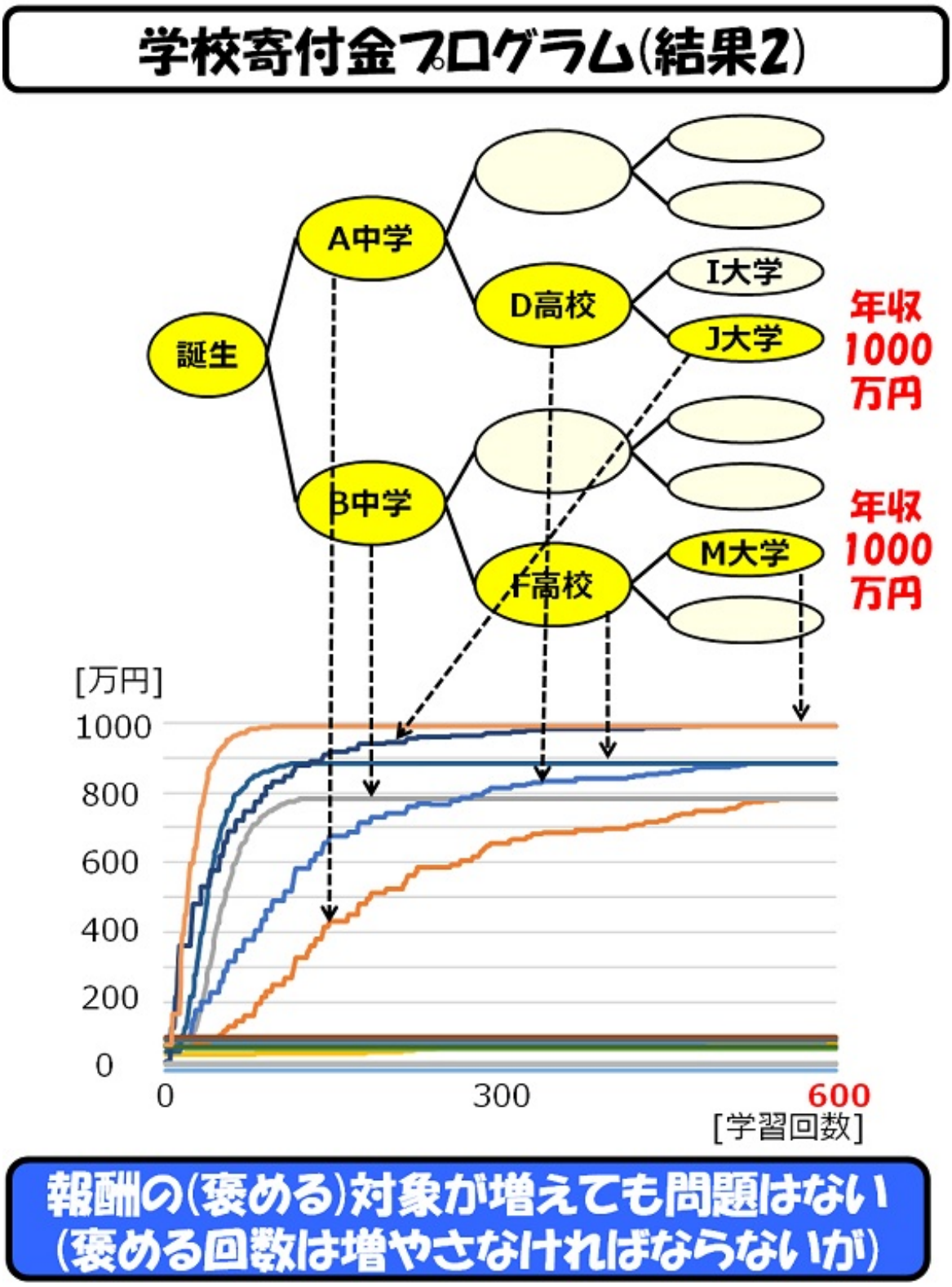


プログラムはこちらです

　上記の例では、J大学の卒業生以外は、全員年収がないですが、気にしないでください。結果は以下の通りになりました。



次は、M大学の卒業生も、年収1000万円ビジネスパーソンになるものとしてみました。結果は以下の通りです。



特に問題なく、強化学習が行えることが分っています。ただし、対象が2倍になると、褒める回数(学習回数)も2倍にしなければならないようです。

これは、対象が複雑になれば、学習回数も増やさないと十分な効果が得られないことを示唆していると思います。

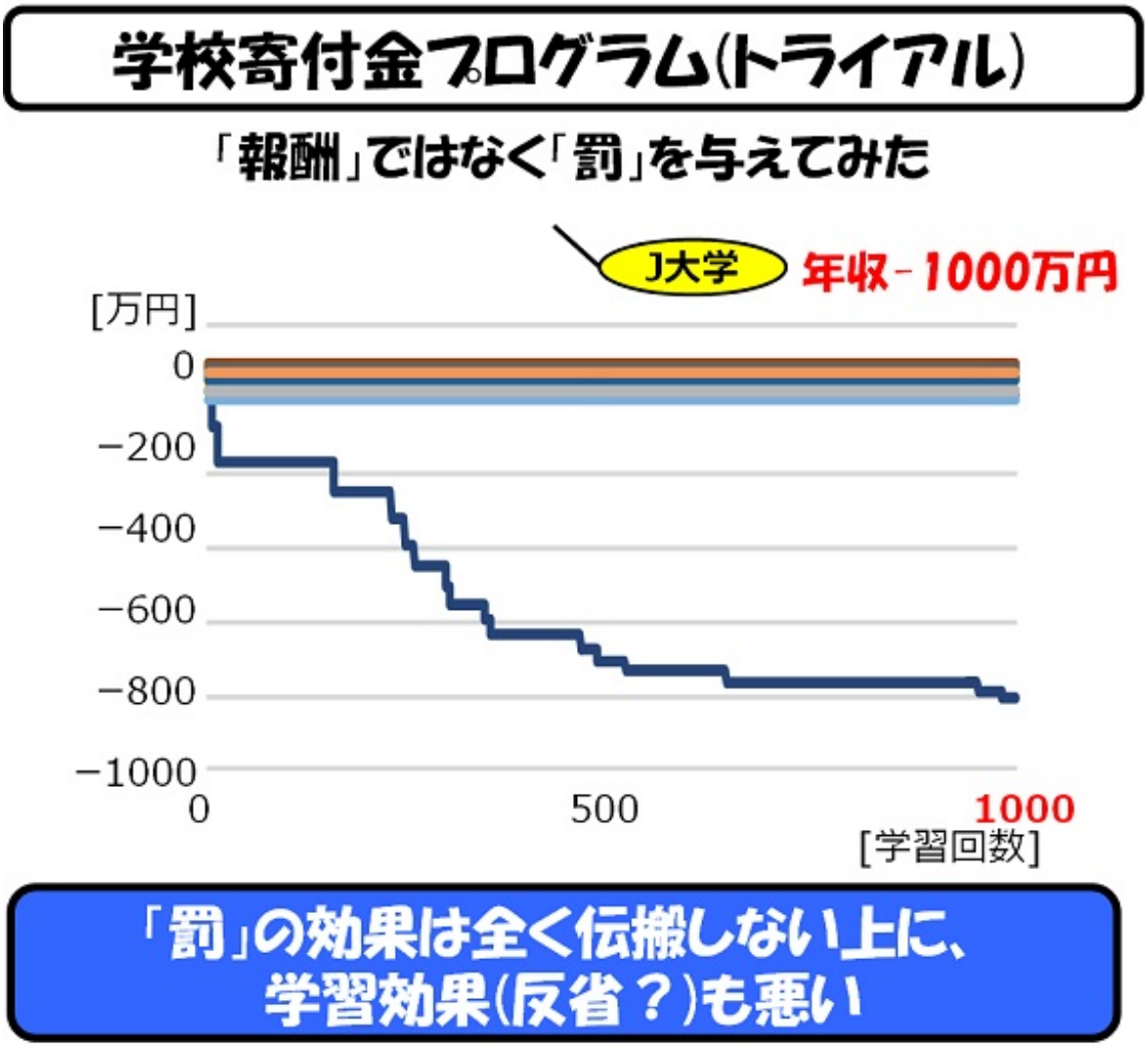
もしも強化学習で“体罰”を与えたら

で、ここまでは予想通りなのですが、実は今回、本当にやってみたかったことは、この強化学習のQ学習で、「体罰」をやってみたらどうなるだろうか、ということでした。

もちろん、Q学習は、「体罰 = マイナスの報酬」を想定して設計されているのではないので、このトライアルは、エンジニア的にはナンセンス（というか、無意味）かもしれません。

しかし、それでも、「褒める（報酬）」ことで効果を発揮するAI技術で、「体罰（マイナスの報酬）」を与えた結果は、冒頭に展開した「体罰の効果」の一つの検証結果になるのではないかと考えました。

ともあれ、やってみました。「J大学に入ると、もれなく1000万円の借金を作る人間になる」―― という設定を置いてみたところ、面白い結果が出てきました。



この結果から言えることは、

（1）「（体）罰の効果」は、その事象にのみ限定されて、その効果は全く波及しない。つまり、（体）罰を受けたことのみに効果があり、その問題の原因にさかのぼらない。

（2）「（体）罰の効果」は学習能力が低い。「褒める」方では、200回で上限に至っているのに、「体罰」の方では、1000回繰り返しても上限に達しない。

ということです。

つまり、「強化学習」のアプローチでは、「体罰」は、拡張性もなく、効率は悪く、効果が低いということです。

しかしながら、体罰に効果を認める人が一定数いるのはなぜか？　これは私の（検証のない）仮説ですが、2つ理由があるのではないかと考えています。

（1）「ブロードキャスト」の効果 ―― つまり「見せしめ」です。報酬の方は一人一人を「褒める」必要がありますが、「見せしめ」は一人を痛めつければ、その恐怖が別の人間にも伝搬させることができて、非常に効率が良いのです。

（2）「怒りの発動」を「教育的指導」と言い換えることができる手軽さ ―― 単に「腹を立てて暴力を行っている」という、通常であれば犯罪にもなり得る行動を、「教育的措置を取った」と言い張れる立場を乱用できるわけです（例えば、私が、電車の中で騒いでいるガキを殴ったら、問答無用で、犯罪になります）。

大体、体罰している人間は「お前たちに腹を立てて、叱っているわけではない」という人がいますが、そんなセリフ信じられますか？ ―― 冷静に客観的に黙々と体罰を実施できる人間がいれば、それは、教育者以前に人間ではありません。その人、きっと"AI"エンジンを搭載したアンドロイドです。

まあ、ともあれ、「体罰には効果がある」と考える人が一定数いる理由は、おおむね上記の2つの仮説で説明可能であると、私は考えています。

少なくとも、強化学習のQ学習アルゴリズムをそのまま使ってみた限りでは、「（体）罰では効果を発揮できない」ことだけは明らかです。

今回のコラムを振り返る

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】「体罰」に対する私の考え方を明かにした上で、「体罰」の世間一般の考えを考察してみました。その結果、「体罰はダメ」と言いつつ「体罰には効果がある」と考える人が一定数いるという仮説を立てました。

【2】今回は、「機械学習」の中の「強化学習（のQ学習アルゴリズム）」について解説を行いました。

【3】「教師あり学習」と「教師なし学習」の比較を行い、前者が「知識とロジック」を、後者が「経験とマネ」を、その学習の根幹に置いていることを明らかにした上で、「強化学習」が「教師なし

学習」の代表技術であることを示しました。

【4】「強化学習」が、たった1行の式（または、4行程度のプログラム）で、超巨大な解空間の中から、最適戦略を選び出すという、脅威のAI技術であることを示しました。

【5】その一方で、「強化学習」が、将棋や囲碁、PCゲーム等の分野でしか、その効果を発揮できていないことを示しました。

　これは、（A）膨大な回数の学習が必要であり、（B）そのような学習はコンピュータの中でしか実現できないことに加えて、（C）私たちの人生において、「強化学習」の環境（状況と行動）を定義することが絶望的に難しい、ということを説明しました。

【6】最後に具体的例として、「学校寄付金プログラム ―― 別名、高年収獲得プログラム」を作成して、その学習プロセスの特徴を明らかにしました。同時に、「強化学習（のQ学習アルゴリズム）」を使う限り、「（体）罰には効果がない」ことを、シミュレーションで明らかにしました。

　「強化学習」とは、「勝ち負け」を続けることで、自力で（勝つための）知識を獲得し続ける学習です。

　嫁さんがと、「アルファ碁」なるコンピュータソフトウェアが、名人に勝利したというNHKニュースを見ていた時、嫁さんが、私にその意味を尋ねてきました。

　私が『要するに、ソフトウェアが自力で学習していくんだよ』と私が説明したら、真っ青な顔をしておびえていました。

　『自分で、新しいことを考えることができるの？』→『人間に勝てるの？』→『そしたら、ソフトウェア（AI技術）が、世界を乗っとるの？』という（世間によくある、AIフォビア（恐怖症）の）三段思考パターンにズッポリとはまっていました。

　「あのね、『自己学習』というのは、コンピュータの数理や制御の世界では、昔から『最適解探索』とか、『自動チューニング』とか山ほどあって、そんなことを言えば、人間は、コンピュータの登場の時から、常に負け続けているとも言える訳だし、それ以前にコンピュータとの『勝ち負け』を論じるというのは……」という説明は ―― 嫁さんの「耳」には届いていても、「心」には届いていなかったようです。

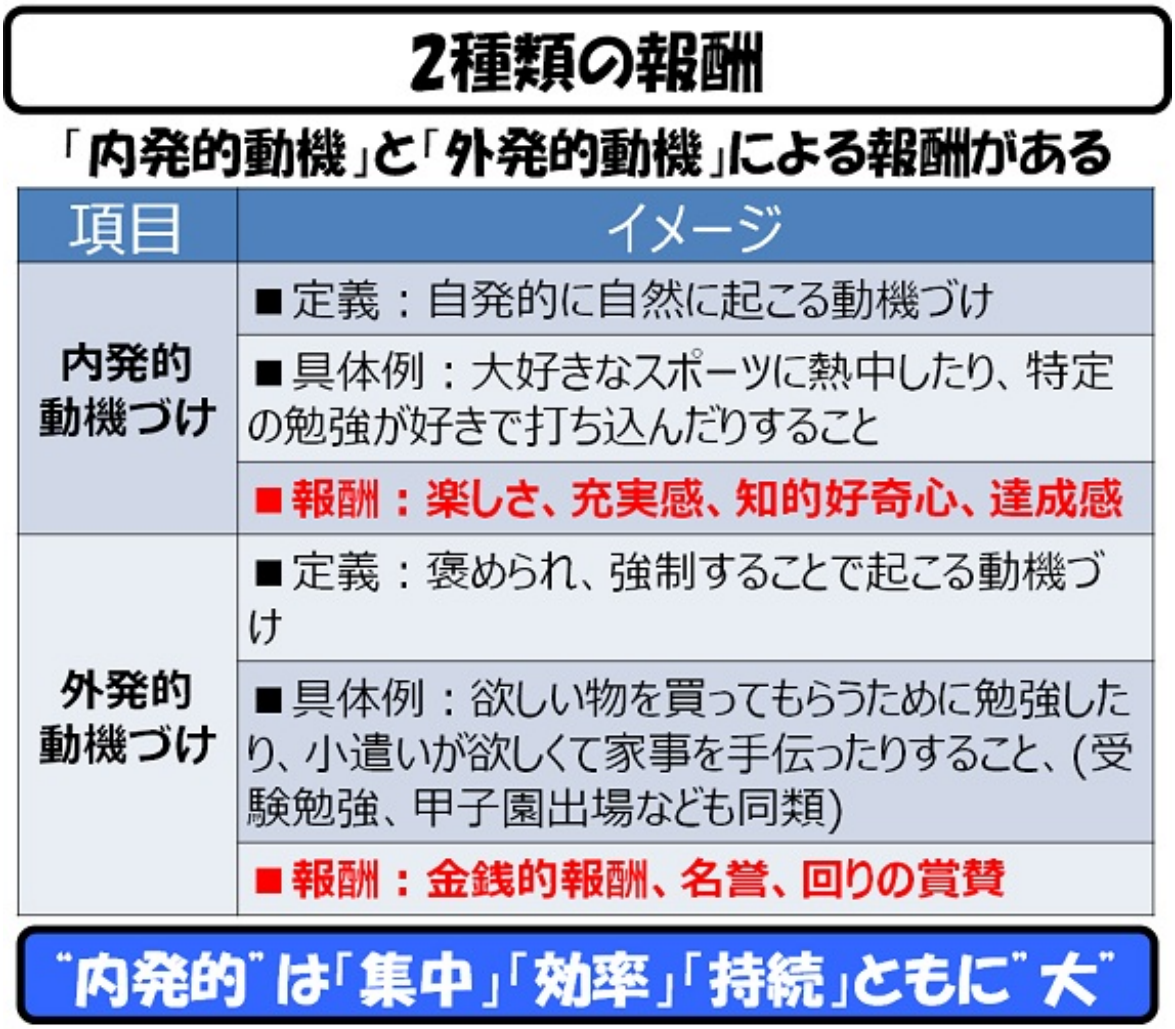
　「ロボットのアーム軌跡の自動チューニング」は怖くないけど、「将棋や囲碁の自己学習」は怖い ―― 私たちは「訳の分からないものが、訳の分からない理由で動くこと」は怖くなくても、「よく知っているものが、訳の分からない理由で動くこと」には、耐えられないほど怖いのです。なにしろ、私たちは、日常的に、将棋や囲碁を良く知っていますから（ルールは知らなくても）。

「強化学習」が人類を支配できない理由

　もう一段階深く考えてみれば、「強化学習」が人類を支配できない理由は明確になります。

まず、「強化学習」が「権力／忖度モデル」……ではなくて「報酬型学習」であることは、既に述べました。

ところが、この報酬には、2種類あるのです。



実は、報酬において、たちが悪いのは「外発的動機づけ」ではなく「内発的動機づけ」なのです。

マッドサイエンティスや、カルト宗教の教祖や、国家の独裁者のモチベショーンは、基本的に、「誰かに褒められたい」「金をもうけたい」という欲望より、「分からないことを知りたい」「世界をよりよくしたい」という、単純で（独善的な）善意から発動するからです。

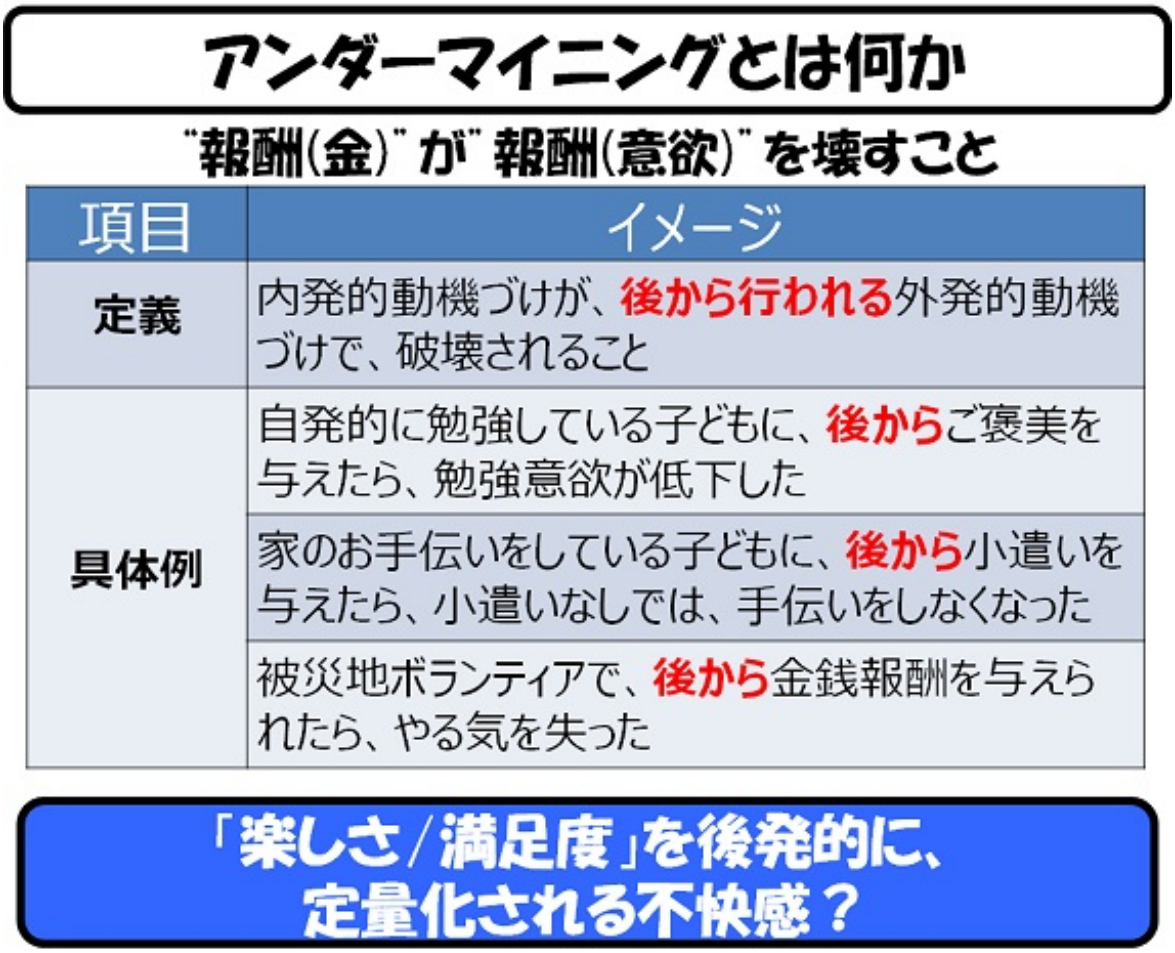
■『そして、正直で真面目で誠実な者ほど、極端に向かって走り出すことは、日本赤軍やオウム真理教を見るまでもなく、明かです』（著者のブログより）

■『「人を殺めてはならない」と言う社会の一般的公理すら、若さゆえの柔軟な受容体を持ったが為、簡単に歪（いびつ）な宗教思想で捻じ曲げられてしまった彼ら。駅前に掲げられている指名手配の看板に載せられている彼らの写真は、もしかしたら私だったかも知れないのです』（著者のブログより）

ご理解いただけると思いますが、「強化学習の報酬」は、「外発的動機づけ」です（なにしろ、報酬（例：年収1000万円）を与えているのは、プログラミングをしている、この私ですから）。

つまり、「外発的動機づけ」なんぞで動いているプログラムに、「世界を乗っとる」意志が発生する訳がありません。だから安心していいのです。

『それは本当に絶対か！』とか『AIに"内発的動機づけ"が発生しないという保証があるのか！』と、心配な方には、それらを止める最後の手段をご教示しておきましょう ―― アンダーマイニングです。



『人間的野心が発生するAI』なんぞ、私は「絶対に作れるわけがない」と決めつけていますが、もしそういうAIが生まれてきたとしたら、そのAIに腐るほどの金銭を与えて、やる気を失わせてやればいいのです。

つまり、AIが「権力」を行使するなら、私たち人間は「忖度」で対抗すれば良いのですよ。

⇒「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

ArmのAI戦略、見え始めたシナリオ
機械学習についてなかなか動きを見せなかったArmだが、モバイルやエッジデバイスで機械学習を利用する機運が高まっているという背景を受け、少しずつ戦略のシナリオを見せ始めている。

AI性能20倍、Xilinxが7nm世代製品「ACAP」発表
Xilinxは2018年3月19日（米国時間）、7nmプロセスを用いる新たな製品群「ACAP」（エーキャップ）を発表した。新たなプログラマブル演算エンジンなどを搭載し、現行のFPGA製品よりも20倍高いAI（人工知能）演算性能を発揮するという。

外交する人工知能 ～ 理想的な国境を、超空間の中に作る
今回取り上げる人工知能技術は、「サポートベクターマシン（SVM）」です。サポートベクターマシンがどんな技術なのかは、国境問題を使って考えると実に分かりやすくなります。そこで、「江端がお隣の半島に亡命した場合、“北”と“南”のどちらの国民になるのか」という想定の下、サポートベクターマシンを解説してみます。

陰湿な人工知能 ～「ハズレ」の中から「マシな奴」を選ぶ
「せっかく参加したけど、この合コンはハズレだ」――。いえいえ、結論を急がないでください。「イケてない奴」の中から「マシな奴」を選ぶという、大変興味深い人工知能技術があるのです。今回はその技術を、「グルメな彼氏を姉妹で奪い合う」という泥沼な（？）シチュエーションを設定して解説しましょう。

上司の帰宅は最強の「残業低減策」だ ～「働き方改革」に悩む現場から

あなた（あなたの会社）は、「働き方改革」に本気で取り組んでいますか？　読者の皆さんの中には、「誰かの上司」という立場である方も極めて多いと思われます。そんな皆さんに伝えたいーー。シミュレーションで分かった「残業を減らす最善策」、それは皆さんが今すぐ、とっとと帰宅することなのです。

誰も知らない「生産性向上」の正体 ～“人間抜き”でも経済は成長？

「働き方改革」に関連する言葉で、最もよく聞かれる、もしくは最も声高に叫ばれているものが「生産性の向上」ではないでしょうか。他国と比較し、「生産性」の低さを嘆かれる日本ーー。ですが、本当のところ、「生産性」とは一体何なのでしょうか。