

## 前回上

---



本日の配布資料 QR コード

この 10 年で人工知能の台頭は私たちの生活のあらゆる側面に革命(第 4 次産業革命)をもたらした。一時期よりも報道される頻度は減っている印象はあるものの、報道されていないだけであり、この分野は着実に発展している。AI 諸分野の拡大に伴い、(知識を)持つものと(知識を)持たざるものとの乖離が大きくなっているようである。持たざるものにとっての将来は、マーク・トウェインの小説「1984 年」に出てくるディストピア世界(AI 脊威論)のように恐ろしい。一方、持つものにとって、AI は無限の可能性を秘めたテーマであり続けている。両者の間隙を埋める、あるいは橋渡しを試みる必要があると考える。社会の分断を招くことに繋がることを意図するような研究や技術では意味がないからである。

加えて、経済産業省の調査報告書によれば、2030 年には最大で 79 万人の IT 人材不足との試算がある(同報告書 26 ページ)。少子高齢化の急速な進展と格差拡大に伴う社会構造の変化を鑑みれば、新たな産業の創出、倫理や価値観の変化に柔軟に適応可能な社会構造への移行、環境問題や格差是正に向けた持続可能な成長シナリオの構築は、人間の尊厳を担保するためにも解決すべき課題である。子どもたちに何を引き継ぐのかを含めて、近年の AI の進展は上述のごとき問題を解決する有力な一つの候補であると看做されるようになってきた。

上述のような問題意識から、ニューラルネットワークの現在の中心的話題である深層学習(deep learning)の動向に概説を試み、上述の問題を解決する緒を探る話題を提供する。

## 本日の概要

---

### 0. 自己紹介

#### 1. 導入に換えて、いただいた 2 件の質問答える (10 分 × 2 = 20 分 + アルファ)

1. AI がオススメを表示するカラクリ
2. ボードゲーム(将棋、囲碁、チェス、ポーカー、テレビゲーム)のカラクリと人間超えた理由
3. 質疑応答 1 回目 Ask me whenever you want

#### 2. AI 概論 (10 分 × 3 = 30 分 + アルファ)

1. 簡単な歴史
2. 画像認識
3. 自然言語処理
4. 強化学習
5. 質疑応答 2 回目 Ask me whenever you want

#### 3. 社会的影響 (10 分 × 3 = 30 分 + アルファ)

1. トロッコ問題
2. 緊急停止ボタン
3. 倫理

#### 4. 全体討論 (時間の許す限り)

## 自己紹介

---



浅川伸一: 博士(文学) 東京女子大学情報処理センター勤務。早稲田大学在学時はピアジェの発生論的認識論に心酔する。卒業後エルマンネットの考案者ジェフ・エルマンに師事、薰陶を受ける。以来人間の高次認知機能をシミュレートすることを通して知的であるとはどういうことを考えていくと思っていた。著書に「ディープラーニング G 検定公式テキスト」(翔泳社, 2016), 「Pythonで体験する深層学習」(コロナ社, 2016), 「ディープラーニング, ビッグデータ, 機械学習あるいはその心理学」(新曜社, 2015) 「ニューラルネットワークの数理的基礎」「脳損傷とニューラルネットワークモデル, 神経心理学への適用例」いずれも守一雄他編「コネクションモデルと心理学」(2001) 北大路書房,

## 謝辞

- 本日このような機会を設けてくださいました 東京女子大学同窓会 望月直子様に感謝申し上げます。
- 資料作成を手伝っていただいた駒澤大学 文学部心理学科 宮脇忠義さん、水谷亜里沙さん に感謝いたします。

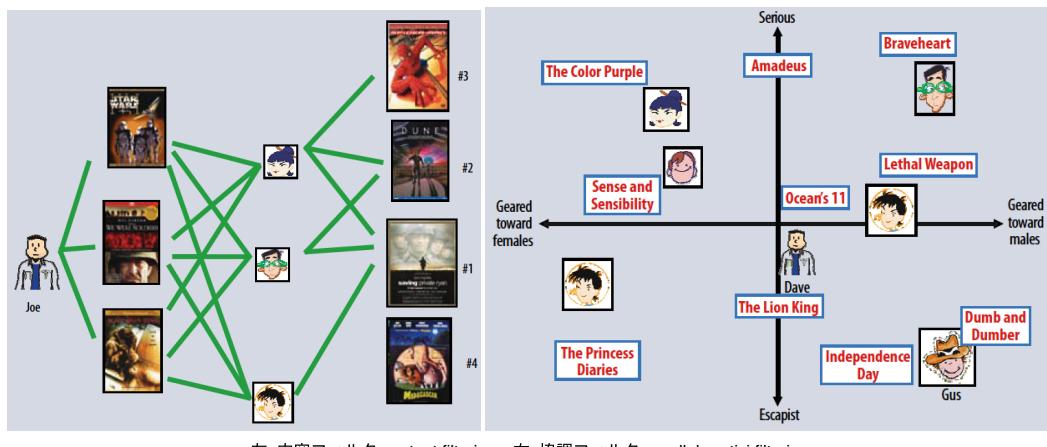
## 1. 導入に換えて

- Q. スマホやパソコンの広告が AI で自分に合うものを選んでくれるようになっていますが、本当にそれを自分でも選ぶのか？ AI に逆に選ばれてしまっていて、操られている気がします。
  - A1 言動には須らく発信者の意図が含まれているのでしょうか。AI に選ばれている部分を考えなければ同じことの繰り返しになります。
  - A2 そこで、対応策、あるいは、回避策として、背景となる技術を知っていることが肝要だと考えます。
  - A3 内容を知らないければ、恐れて怯えるだけです。最低限の知識だけは知っておくと判断が容易になるでしょう。
- Q. ご主人が将棋好きな方から。AI の将棋が普及てきてから将棋自体が大きく変わってしまい、楽しくなくなったとご主人が言っているそうです。そういうゲームの仕組み自体が変わってしまう上で、本来の楽しみから離れてしまうことが他の事例でもあるのでしょうか？
  - A1 アルファ碁は、人口に膾炙した事例でしょう。その他にも、ポーカー、テレビゲーム、なども人間を凌駕する性能を示しています。
  - A2 時代とともに変容する常識あるいは価値観。たとえば、火打ち石、馬車、井戸掘り ...
  - A3 「本来の楽しみ」が変わっている現実を受け入れる必要があるでしょう。
  - A4 探索の地平線 という言葉を紹介します。後述
  - A5 スマートフォンの不携帯により緊急連絡不能となり、事件や事故に巻き込まれるような事態。
  - A6 参考資料 [労働新聞 知識を拡張する道具](#)
  - A7 参考資料 [Sutton Bitter lesson](#)

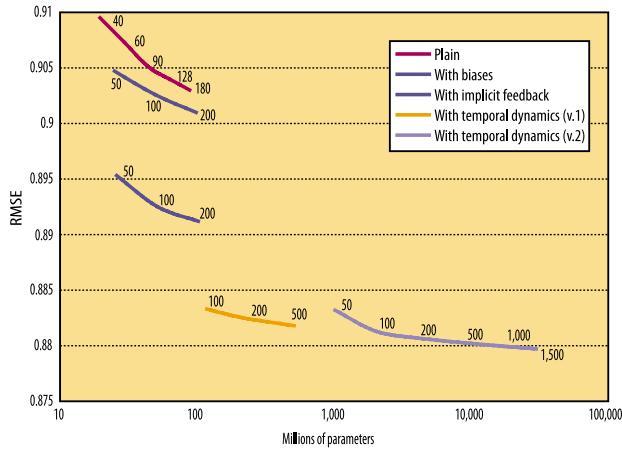
AI 研究は、嚆矢濫觴の時代より 2 つの潮流があり、犬猿の仲であった。記号主義的 AI と ニューラルネットワーク の如き 非記号主義的 AI である。2010 年代以降 ニューラルネットワーク研究が主流となっている。そこで、ここでは第 3 次ニユーロブーム(あるいは第 3 次 AI ブーム)の根幹をなす概念を取り上げて概説することとする。まず(1)現在の状況を整理し、続いて(2)人間の認識性能を凌駕した言われる 畏み込みニューラルネットワーク、(3)系列情報処理のための 再帰型ニューラルネットワーク、を取り上げる。(3) 2016 年に囲碁の世界王者を破った 強化学習 についても触れる。

### 1.1. 映画のオススメのからくり 行列因子化 Matrix Factorization または 協調フィルタ

Netflix コンペ優勝モデル [@2009Koren\_MF\_Recommend]

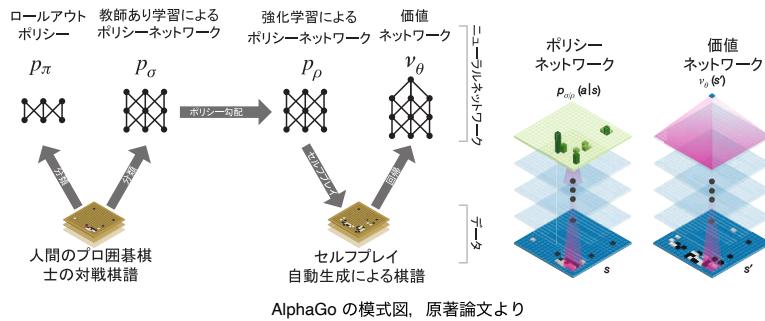


- 内容フィルタ content filtering が従来手法、協調フィルタ collaborative filterng が Netflix コンペ優勝チームの手法
- Netflix コンペ: 映画 17,000 本のユーザレビュー 30 万。レビューは 30 万件。各レビューは 1 から 5 までの 5 段階の星の数を予測する。
- 星の数 $\text{映画}_i, \text{ユーザ}_j = \text{映画}_i \times \text{ユーザ}_j$  行列因子化ベースモデル
- 星の数 $\text{映画}_i, \text{ユーザ}_j = \text{平均} + \text{映画バイアス}_i + \text{ユーザバイアス}_j + \text{映画}_i \times \text{ユーザ}_j$  (バイアスモデル)
- 星の数 $\text{映画}_i, \text{ユーザ}_j = \text{平均} + \text{映画バイアス}_i(t) + \text{ユーザバイアス}_j(t) + \text{映画}_i \times \text{ユーザ}_j(t)$  (時間発展モデル)
- 行列因子化、または、特異値分解 SVD



各モデルの比較。横軸：パラメータ数、縦軸：平均自乗誤差。低ければ低いほど精度が良い

## 1.2. 強化学習, 予測報酬誤差, ゲームAI, 経済学



AlphaGo の模式図、原著論文より

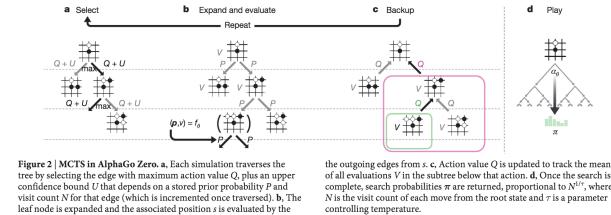
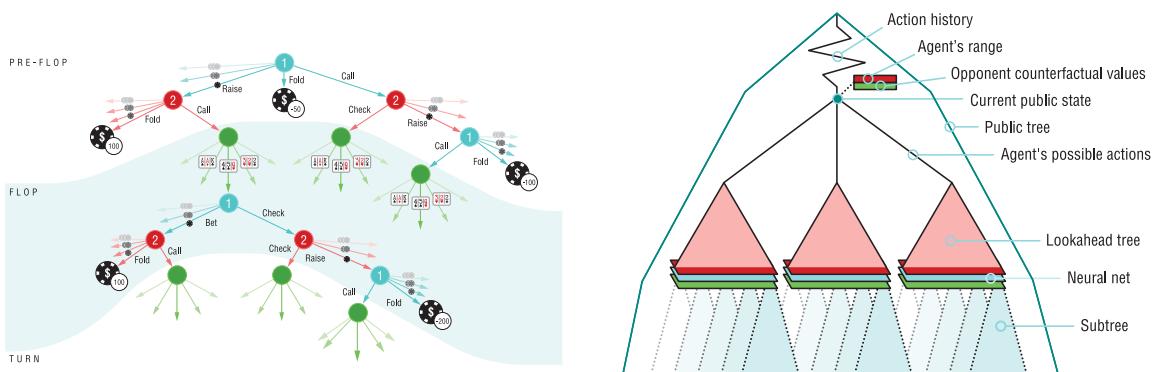


Figure 2 | MCTS in AlphaGo Zero. a. Each simulation traverses the tree by selecting the edge with maximum action value  $Q_i$  plus an upper confidence bound  $U$  that depends on a stored prior probability  $P$  and visit count  $N$  for that edge (which is incremented once traversed). b. The leaf node is expanded and the associated position  $s$  is evaluated by the neural network  $(P(s, \cdot), V(s)) = f_0(s)$ ; the vector of  $P$  values are stored in

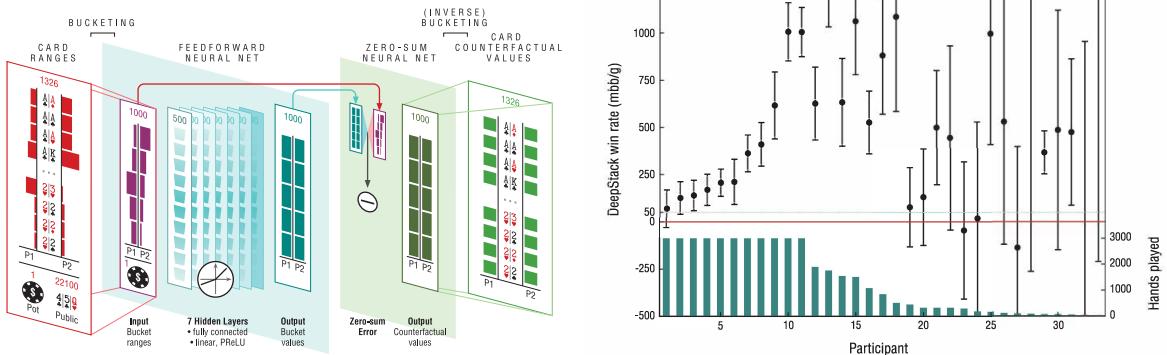
AlphaGoZero のセルフプレイ、原著論文より

## 1.2. (続き) ポーカー DeepStack



左: 各結節点は公開状態を表す。端点は競技者のベット行為(赤と水色)および表に返されたカード(図中緑丸)という行動を表す。ゲームは最終結節点で終了する。最終結節点での値がチップとして表示される。競技者がホールドしなかった最終結節点ノードでの値はプレイヤーの共同秘密情報(表にされなかつたカード)が返される。

右: DeepStack は公開木で推論する。公開状態で保持できる全カードの行動確率を常に生成し、プレイ中自分の範囲と相手の反事例事価値という 2 つのベクトルを維持している。ゲームの進行に伴い自分の行動を起こした後、計算した行動確率を用いてベイズ則により状態を更新する。

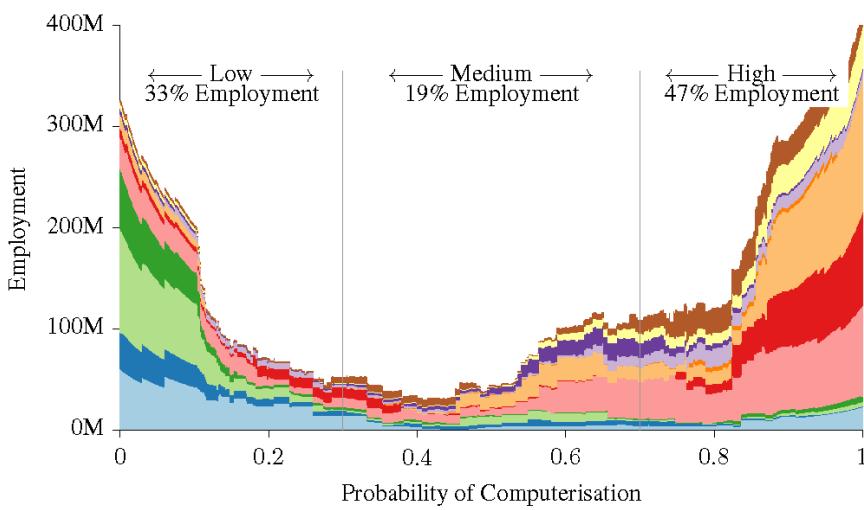


左: 深層 反事例価値ネットワーク Deep counterfactual value network. ネットワークへの入力は、ポットサイズ、表になったカード プレイヤーレンジであり、これらはまずハンドクラスターに処理される。7つの完全連結層からの出力は 値がゼロサム制約を満たすよう処理される。これらの値は 反事例価値ベクトルにマッピングされる。

右: プロのポーカープレイヤーの DeepStack に対する成績。下の棒グラフはゲーム終了までの手数。

### 1.3. 導入 AI によって無くなる職業 (Frey and Osborne, 2013)

単純労働ばかりでなく、知的生産者も危機



Frey and Osborne (2013) より アメリカ合衆国で 47% の仕事が自動化により消失すると予測

- Automation, Skills Use and Training, Ljubica Nedelkoska and Glenda Quintini, OECD working paper (2018)

### AI に奪われない職業 (Frey and Osborne, 2013 より)

順位	職業
1.	レクリエーションセラピスト (Recreational Therapists)
2.	第一線の、機械、導入師、修繕士、監督者 (First-Line Supervisors of Mechanics, Installers, and Repairers)
3.	緊急管理ディレクター (Emergency Management Directors)
4.	メンタルヘルス および 薬物乱用ソーシャルワーカー (Mental Health and Substance Abuse Social Workers)
5.	聴覚医 (Audiologists)
6.	職業療法士 (Occupational Therapists)
7.	正教会と補綴学者 (Orthotists and Prosthetists)
8.	医療従事者ソーシャルワーカー (Healthcare Social Workers)
9.	口腔 および 顎顔面外科医 (Oral and Maxillofacial Surgeons)
10.	消防職員 の 第一次監督者 (First-Line Supervisors of Fire Fighting and Prevention Workers)

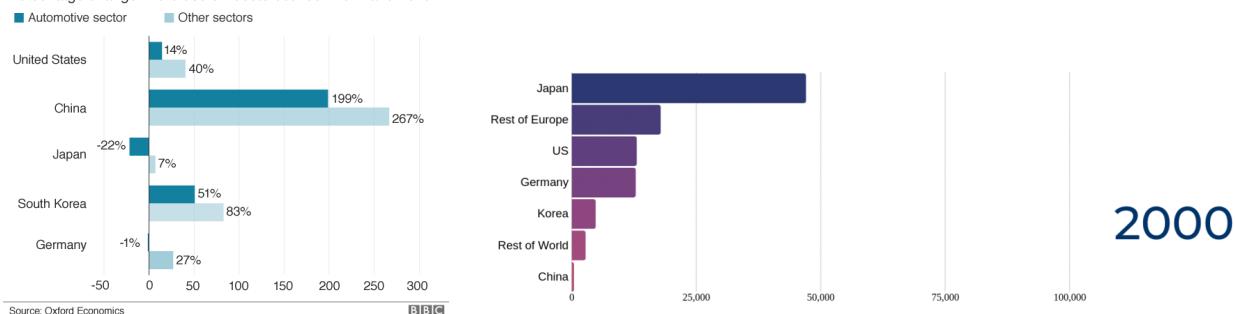
### AI に奪われそうな職業 (Frey and Osborne, 2013 より)

順位	職業
693.	新規口座担当者 (New Accounts Clerks)
694.	写真加工技師 および 加工機械操作技師 (Photographic Process Workers and Processing Machine Operators)
695.	税務申告者 (Tax Preparers)
696.	貨物 および 貨物代理店 (Cargo and Freight Agents)
697.	時計修理業者 (Watch Repairers)
698.	保険引受人 (Insurance Underwriters)
699.	数学技術者 (Mathematical Technicians)
700.	下水道、手 (Sewers, Hand)
701.	タイトル (特許、申請) 審査官、要約官、および調査員 (Title Examiners, Abstractors, and Searchers)
702.	テレマーケティング担当者 (Telemarketers)

### 1.3. 换足 ロボット応用の各国比較

#### The rise of the robots

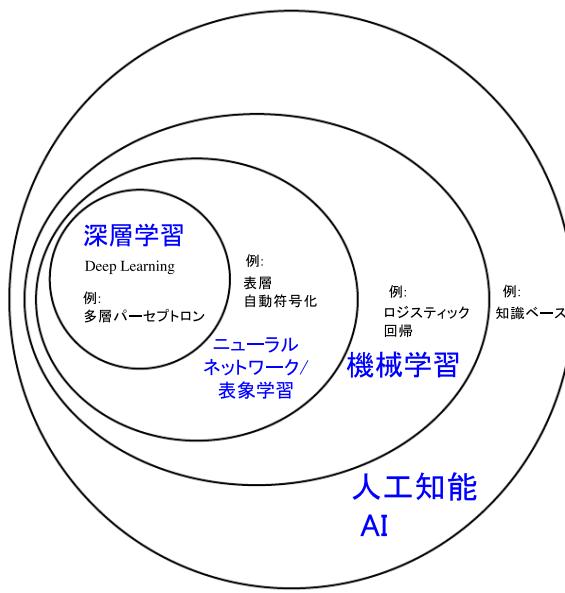
Percentage change in the use of robots between 2011 and 2016



2010 年以降 世界の産業界におけるロボットのストックは 2 倍以上に増加。過去 4 年間で導入されたロボット数は 過去 8 年間で導入されたロボットの数と同程度。現在、世界のロボットの約 3 分の 1 が中国に設置されている。世界のロボットストックの約 5 分の 1 を占める。出典: [Robots to replace up to 20 million factory jobs by 2030] (<https://www.bbc.com/news/business-48760799>)

## 2. 人工知能と機械学習とニューラルネットワークと深層学習

下図には、深層学習、表象学習、機械学習、人工知能 (AI) の関係が示されています。一番外側に人工知能があり、人工知能は他の全てを含む言葉であることが示されています。人工知能には機械学習と呼ばれる分野と図に示されている言葉では知識ベースと機械学習に分かれます。機械学習はロジスティック回帰などとニューラルネットワークに代表される表象学習に分かれます。ニューラルネットワーク、あるいは表象学習は、深層ではない自動符号化と深層学習に分かれます。どれも広義には人工知能に含まれます。今日注目を集めている分野はほぼ深層学習 (ディープラーニング) と呼ばれる分野になります。



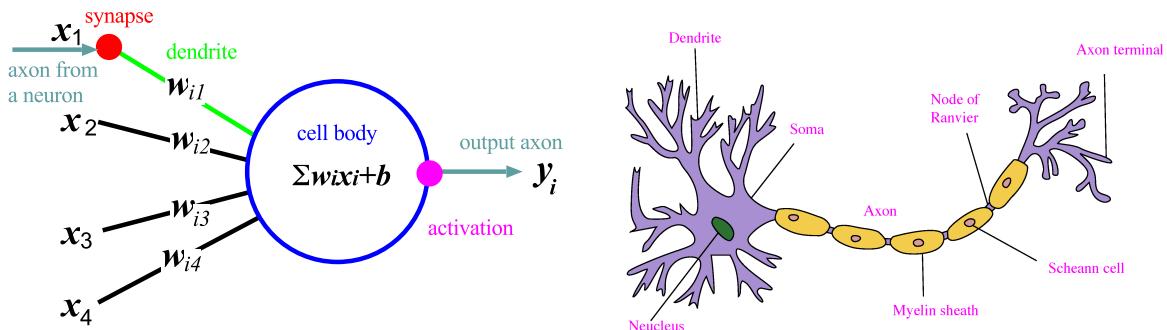
[@2016GoodfellowBook] Fig 1.4 を改変

### 3. 深層学習(ディープラーニング)とは何か

CNN の特徴の一つに エンドツーエンド と呼ばれる考え方があります。エンドツーエンドとは、従来手法によるパターン認識システムでは、専門家による手の込んだ詳細な作り込みを必要としていたことと異なり、面倒な作り込みをせずとも性能が向上したことを指します。

エンドツーエンドなニューラルネットワークにより、次のことが実現しました。

- ニューラルネットワークの層ごとに、特徴抽出が行われ、抽出された特徴がより高次の層へと伝達される
- ニューラルネットワークの各層では、比較的単純な特徴から次第に複雑な特徴へと段階的に変化する
- 高次層にみられる特徴は低次層の特徴より大域的、普遍的である
- 高次層のニューロンは、低次層で抽出された特徴を共有している



左: 形式ニューロン. 右: ニューロンの模式図 wikipedia より

上図で入力信号  $x$  を重み  $w$  を付けて足し合て合算する部分を線形変換と呼び、線形変換した値に変換する部分を非線形変換と呼びます。

## ニューラルネットワークの分類学 neural network taxonomy

- 畳み込みニューラルネットワーク Convolutional Neural Networks: CNN
- リカレントニューラルネットワーク Recurrent Neural Networks: RNN
- 強化学習 Reinforcement Learning: RL
  - 注意: Multi-head self attention: MHSA
  - 敵対的ネットワーク Generative Adversarial Networks: GAN
  - 変分自己符号化器 Variational Auto Encoders, 変分推論 Variational Inference
  - 自己教師あり学習 Self Supervised Learning: SSL
  - 対比学習 Contrastive Learning:

## 学習方法による分類

- 教師あり学習 Supervised Learning
- 教師なし学習 Unsupervised Learning
- 半教師あり学習 Semi-supervised Learning

## 4. ディープラーニングの活用事例 (実習)

実習はブラウザ上で動作する [IPython](#) 実行環境 [Google Colab](#) で行います。コードを実行する際、ブラウザは [Chrome](#) でお願いします。[Edge](#) では動作しない場合があります。

- [CartoonGAN による画風変換](#)
- [CycleGAN によるフェイク画像生成](#)
- [画像認識 CNN の実演](#)
- [画像切り分けの実演 Detectron2](#)
- [自然言語処理の実演](#)
- [ゼロショット学習 CLIP の実演](#)
- [単語の意味 word2vec の実演](#)
- [強化学習 倒立振子の実演](#)

## デモ

- [ニューラルネットワークで遊んでみよう！](#)
- [強化学習のデモ](#)
- [リカレントニューラルネットワークによる文処理デモ](#)

## 5. ディープラーニングがもたらしたもの SOTA の一部

- 音声認識 (Hannun, A. et al. Deep speech: scaling up end-to-end speech recognition. arXiv:1412.5567 (2014).)
- 画像認識 (Krizhevsky, A., Sutskever, I. & Hinton, G. E. in Adv. Neural Inf. Process. Syst. 1097–1105 (NIPS, 2012).; He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. Proc. IEEE Conf. Comput. Vision Patt. Recog., 770–778 (2016))
- 自動翻訳 (Vaswani, A. et al. in Adv. Neural Inf. Process. Syst. 6000–6010, NIPS, 2017)
- 画像、音声生成 (Oord, A. v. d., Kalchbrenner, N. & Kavukcuoglu, K. Pixel recurrent neural networks. PMLR 48, 1747–1756, 2016; Van den Oord, A. et al. Wavenet: a generative model for raw audio. arXiv:1609.03499 (2016))
- 言語モデル (Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N. & Wu, Y. Exploring the limits of language modeling. Preprint at arXiv:1602.02410, 2016)
- 強化学習によるテレピグーム (Mnih, V. et al. Human-level control through deep reinforcement learning. Nature 518, 529–533, 2015).
- 囲碁 (Mnih, V. et al. Human-level control through deep reinforcement learning. Nature 518, 529–533, 2015; Silver, D. et al. Mastering the game of go without human knowledge. Nature 550, 354–359, 2017) ポーカー (Moravčík, M. et al. DeepStack: expert-level artificial intelligence in heads-up no-limit poker. Science 356, 508–513, 2017), アタリの全ゲーム (Badia et al. Agent57: Outperforming the Atari Human Benchmark, arXiv:2003.13350, 2020)

## 6. 簡単なニューラルネットワークの歴史

- BC 300: アリストテレス 脳を理解する最初の試み
- 1873 ゴルジ: 神経細胞の塩化銀染色法を発見
- 1873 バイン: 神経集団のモデル化
- 1943 マッカロックとピツ: 形式ニューロンの提案。ニューラルネットワーク回路を論理回路とみなすことを提案
- 1949 ヘップ: ヘップの学習則を提案
- 1958 ローゼンプラット: パーセプトロンの提案 第一次ニューロブームの始まり
- 1974 ウェルボス: 誤差逆伝播法の発見
- 1980 福島: ネオコグニトロン発表
- 1980 コホネン: 自己組織化マッピング
- 1982 ホップフィールド: ホップフィールドネットワークの提案。スピングラスモデル
- 1985 ヒントン & セノフスキイ: ポルツマンマシン発表
- 1986 スモレンスキー: ハーモニウム (後の制限ポルツマンマシン)
- 1986 ジョーダン: リカレントニューラルネットワーク
- 1999 ルカン: 初期のディープラーニングモデル LeNet 発表
- 1997 シュスター: 双方向リカレントニューラルネットワーク
- 1997 シュミットフーバー: 長短期記憶 LSTM
- 2006 ヒントン: 深層信念ネットワーク。層ごとに学習結果を積み重ねて多層化
- 2009 サラクディノフ & ヒントン: 深層ポルツマンマシン
- 2012 ヒントン: ドロップアウト
- 2013 ミコロフ: 単語埋め込みモデル
- 2014 シルバー, ムニーラ: DQN
- 2015 アルファ碁が人間超え
- 2015 ベ: 残差ネット 画像認識コンテストで人間超え
- 2017 パスワニ: トランسفォーマー。マルチヘッド自己注意
- 2018 BERT が自然言語領域で人間超え
- 2020 バディア: Agent57, アタリの全ゲームで人間超え

## 6. 簡単なニューラルネットワークの歴史(続き)

---

### 第一次ニューロブーム

- 1950 年代パーセプトロン
- 1960 年, ミンスキートババートの批判
- 第一次氷河期の到来

### 第二次ニューロブーム

- 1986 年, PDP ブック, 俗に言うバイブル, 発表
- 1989 年, パブニック, サポートベクターマシン発表
- 第二次氷河期の到来

### 第三次ブーム

- 置込みニューラルネットワーク (Fukushima, 1980; LeCun 1999)

アンダーソン Anderson, J. A. (1990) によれば「人工ニューラルネットワークは素人の非線形統計学である」 ANNs are some kind of non-linear statistics for amateurs と言われ続けてきました。

### 若干の考察 (妄想?)

- 我々人間は、外界を認識するために必要な計算を、生物種としての発生の過程と、個人の発達を通しての経験に基づく認識システムを保持していると見ることができます。
- 従って我々の視覚認識には化石時代に始まる光の受容器としての眼の進化の歴史と発達を通じた個人の視覚経験が反映された結果もあります。
- 人工知能の目標は、この複雑な特徴検出過程をどうやったらコンピュータが獲得できるかということでもあります。
- 外界を認識するために今日まで考案されてきたモデルを訓練するための学習方法はそれほど難しくありません。
- この意味で画像認識課題が正しく動作するためのポイントは、認識システムが問題を解く事が可能なほど複雑であるかどうかではなく、十分に複雑が視覚環境、すなわち画像認識の場合、外部の艦橋を反映するために十分な量の像データを容易にできるか否かにあります。
- 今日の CNN による画像認識性能の向上は、簡単な計算方法を用いて複雑な外部環境に適応できる認識システムを構築する方法が確立したからであると言うことが可能です。

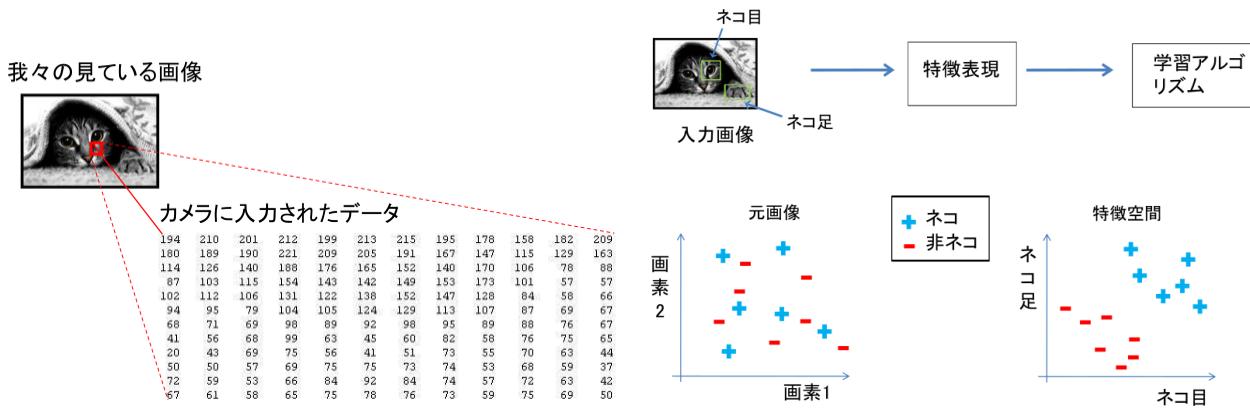
モデルが複雑な課題を解くことができるか否かはモデルの複雑さによるのではなく、そのモデルに与えられたデータ(外部環境)が複雑だからです。生物は、己を取り巻く複雑な外部環境(データ)にさらされながら、その環境に適応しようとしてきました。今日の人工知能の盛況ぶりこのような環境を以下にして簡単なアルゴリズムを用いて複雑なモデルを構成するかという点に着目し、およそその方法が確立しつつあるという点が強調されるべき点であると考えます。

### 現代的認識モデルの特徴

- 深層学習=表象学習/特徴の学習
- 従来モデルによるパターン認識(1950年代-)
- 固定的/職人芸的特徴(固定カーネル)+学習可能な分類器
- エンドツーエンドな学習/特徴学習/深層学習
- 学習可能な特徴(カーネル)+学習可能な分類器
- 基本モデルは1950年代以来進化していない
- 最初の学習する機械: パーセプトロン(1960) ローゼンプラット コーネル大学
- パーセプトロンは単純な特徴検出器の上に線形分離器を乗せたモデル
- 今日の機械学習の実際:

1. パーセプトロンの線形分類器を使用
  2. パーセプトロンのテンプレートマッチングを使用。
- 特徴抽出器の設計には、専門家による長期の努力が必要

## 7. 画像認識



状況ごとにとるべき操作を命令として逐一コンピュータに与える指示する手順の集まりのことをコンピュータプログラムと呼びます。人間がコンピュータに与えることができる操作や命令によって画像認識システムを作る場合、命令そのものが膨大になったり、そもそも説明することが難しかったりします。例を挙げれば、お母さんを思い浮かべてくださいと言われば誰でも、それぞれ異なるイメージであれ思い浮かべることができます。また、提示された画像が自分の母親のものであるか、別の女性であるかの判断は人間であれば簡単です。

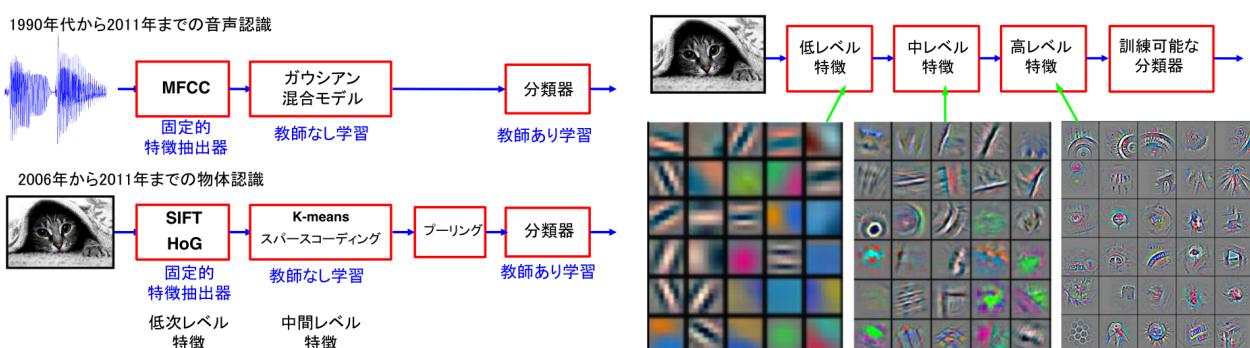
ところがコンピュータには難しい課題となります。加えて母親の特徴をコンピュータに理解できる命令としてプログラムすることも難しい課題です。つまり自分の母親の特徴を曖昧な言葉でなく明確に説明するとなるととても難しい課題となります。というのは、女性の顔写真であればどの写真も似ていると言えるからです。顔の造形や輪郭、髪の位置などはどの画像も類似していることでしょう。ところがコンピュータにはこの似ている、似ていないの区別が難しいのです。

加えて、同一ネコの画像であっても、被写体の向き視線の方向や光源の位置や撮影条件が異なれば画像としては異なります。

入力画像の中の特定の値だけを調べてみても、入力画像がネコである、そうではないかを判断することは難しい課題になります。

現在の画像認識では、特定の画素の情報に依存せずに、入力画像が持っている特徴をとらえるように設計されます。たとえば、ネコを認識するために必要なことは、ネコに特徴的な「ネコ目」や「ネコ足」を検出することであると考えます。入力画像から、ネコの持つ特徴を抽出することができれば、それらの特徴を持っている入力画像はネコであると判断して良いことになります。

下図は音声認識と画像認識の両分野において CNN が用いられる以前の従来手法をまとめたものです。



左: 従来主流であったパターン認識システムの構成。右: 非線形特徴変換を多数回繰り返した学習器を深層学習(ディープラーニング)という

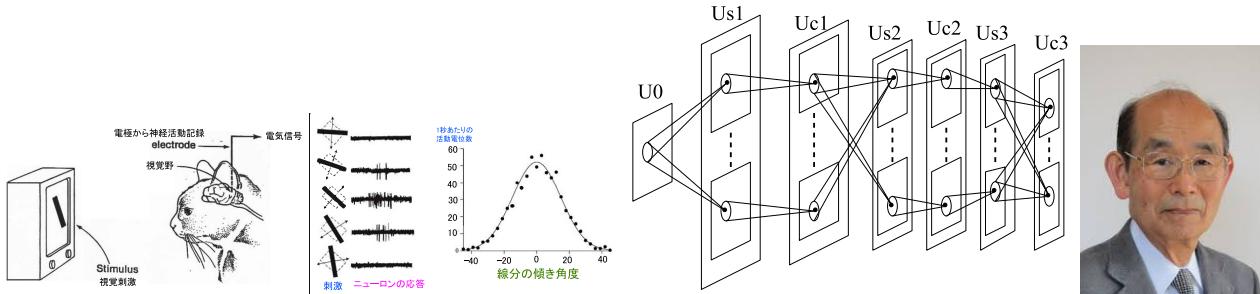
まとめると、

1950 年代後半以来、固定的、手工芸的特徴抽出器と学習可能な分類器を用いた認識システムを作ることが試みられてきましたといえます。これに対して CNN が主流となった現在は エンドツーエンド で学習可能な特徴抽出器を多数重ね合わせることで性能が向上しました。

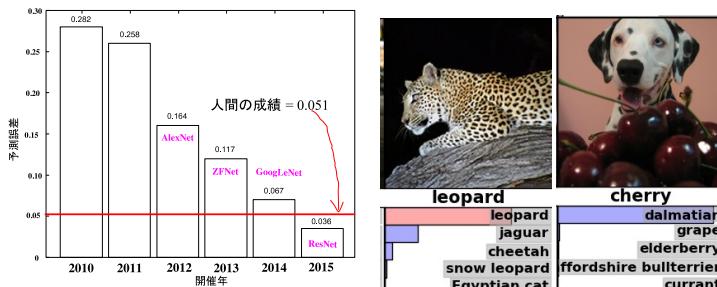
- 2000年代以降: 学習可能な特徴抽出器 → 学習可能な分類器 エンドツーエンド (end-to-end) の実現。CNN は元来、脳の認識方法を模倣したニューラルネットワーク。脳の視覚情報処理と CNN の画像認識の方法はほぼ同じものであるとの主張もあります。

### 画像認識 畳み込みニューラルネットワーク (CNN)

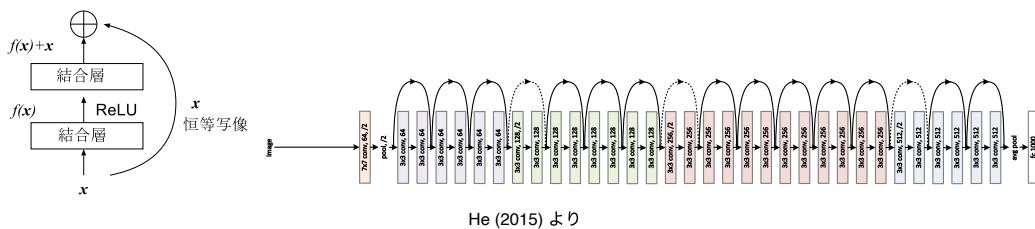
畳み込みニューラルネットワーク CNN は近年、画像認識や音声認識で急激な性能向上をもたらしました。ニューラルネットワークとは、人間の脳の振る舞い（神経回路）を模した計算モデルを指します。現在の第三次人工知能ブームの火付け役となったのは、深層学習（ディープラーニング）という機械学習の手法です。ここで用いられているモデルがニューラルネットワークであり、ニューラルネットワークの中間層を多層化したモデルのことを深層学習と言ったりします。技術的には、中間層を多層化する工夫が実用化されてきたため認識精度が向上しました。



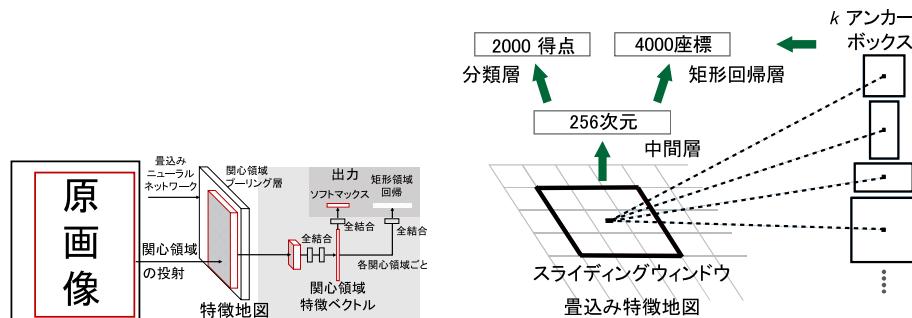
ネオコグニトロンの概略図(Fukushima, 1979)



アレックスネットの結果: 画像のすぐ下の英単語は正解ラベルを表しています。 Krizensky et al (2012) Fig. 4 より。  
ピンク色は正解ラベルの確率を表す。ブルーは不正解ラベル判断確率を表しています。 チェリーが正解であるが、  
画像を見る限り、第一回答候補のダルマチアンを正解だと考へても問題は無いと考えられます。



He (2015) より



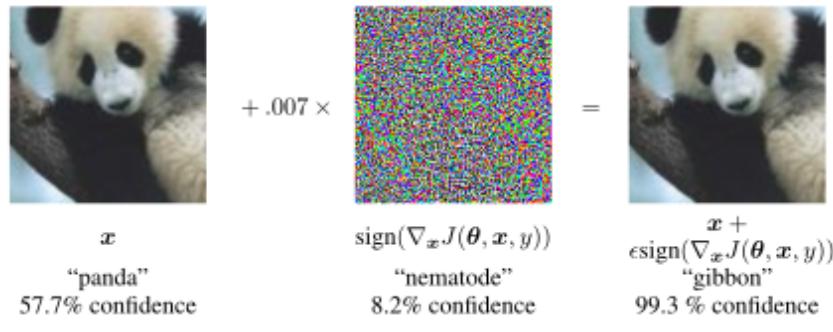
Dang and Ha (2017) より

#### 画像切り出し

1. 物体位置
2. 物体認識 object recognition
3. 意味的切り出し semantic segmentation
4. 対象切り出し instance segmentation
5. 特徴点抽出 keypoint
6. パノラミック切り出し

#### 敵対的攻撃

夢のような話が続きましたが、本節の最後に逆に CNN は簡単に騙すことができる例を挙げておきます。 図では、左の画像が入力画像で、CNN は確信度 57.7 パーセントでパンダである認識しました。 ところがこの画像に 0.007 だけ意味のない画像(図中央)を加えた画像(図右)を CNN は 99.3 パーセントの確信度でテナガザル(gibbon)と判断しました。 この例はここでは詳しく触れることはしませんが、敵対的学習と呼ぶ訓練手法を説明する際に用いられた例です[@2014Goodfellow\_GANHarnes]。



パンダの画像に 0.7% の雑音を乗せるとテナガザルと認識してしまう例[@2014Goodfellow\_GANHarness]

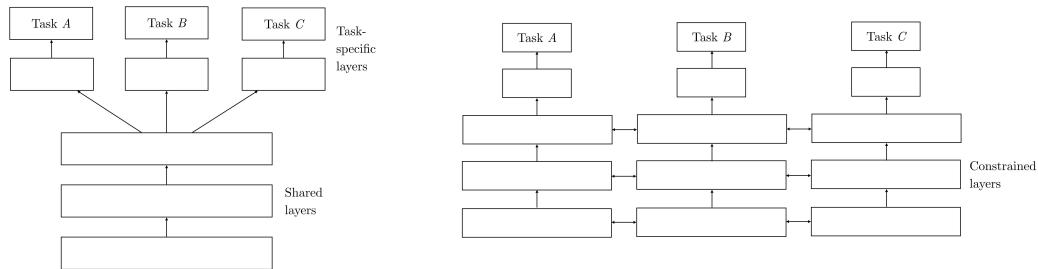
この例からも分かることは以下のようにまとめられるでしょう。すなわち、人間の脳を模したニューラルネットワークである CNN が大規模画像認識チャレンジにおいて人間の認識性能を越えたと報道されました。ですが、人間の視覚認識を完全に実現したと考えるのは早計で、解くべき課題は未だ多数あるということです。この状況は、音声認識や言語情報処理でも同様であると言えます。

## 転移学習

転移学習 transfer learning は機械学習分野のみならず、ロボット工学や実応用の分野でも応用が考えられます。シミュレーションと現実との間隙をどのように埋めるのかという大きな問題に関連します。一方で、転移学習と ファインチューニング や 領域適応 domain adaptation の区別がなされています。

転移学習とは 課題 A を用いて訓練したモデルに対して、別の課題 B に適用することを言います。DNN では転移学習は頻用されます。イメージネットで画像分類を学習したネットワークに対して、例えば顔認識を学習させるような場合です。

PyTorch のチュートリアルなどでは、学習済のネットワークに対して、最終直下層を入れ替えて別の課題を訓練することを転移学習と呼びます。このとき、最終直下層と出力層との結合を学習させ、その他の下位層の結合は固定し、訓練しません。一方で、下位層まで含めて全結合を訓練させる場合をファインチューニングと呼び、区別しています。



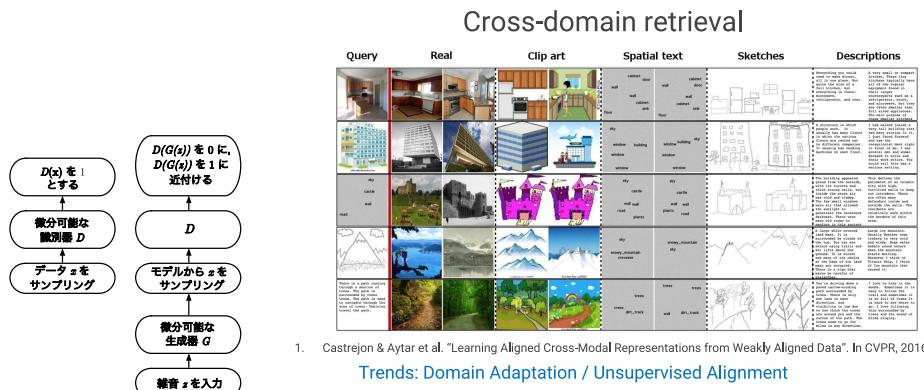
左: ハードパラメータ共有: 転移学習, 右: ソフトパラメータ共有: ファインチューニング

## 生成モデル

認識の反対の操作をすれば、生成が可能です。生成敵対ネットワーク Generative Adversarial Networks: GAN になります。

GAN では 2 つのニューラルネットワークが用いられ、識別器 descriminator と 生成器 generator と呼びます(Goodfellow,2014)。識別器も生成器も多層ニューラルネットワークです。通常の画像分類課題では、最上位層において推論、すなわち入力画像が何であるかを計算するためにソフトマックスする関数などが用いられます。これに対して GAN の識別器では、0 か 1 かの出力をします。入力画像が通常の画像であれば 1 を、生成器によって生成された画像であれば 0 を出力します。生成器は、識別器の最終直下層で得られたような画像表現に雑音を加えた値から画像を生成します。生成器は、識別器が入力データから画像を推論するのと逆方法に推論から画像を生成します。すなわち GAN は入力が実在するか、偽造品、すなわちフェイクかを見破る訓練がなされることになります。

このようにして、生成器は識別器の学習結果であるデータの内部表現を模倣し、生成器を欺こうします。このようにして識別器と生成器との間で ゲーム理論 でいう ナッシュ均衡 Nash's equilibrium が成立立ちます (Heusel, 2017)。GAN の模式的な流れを下図に示しました。



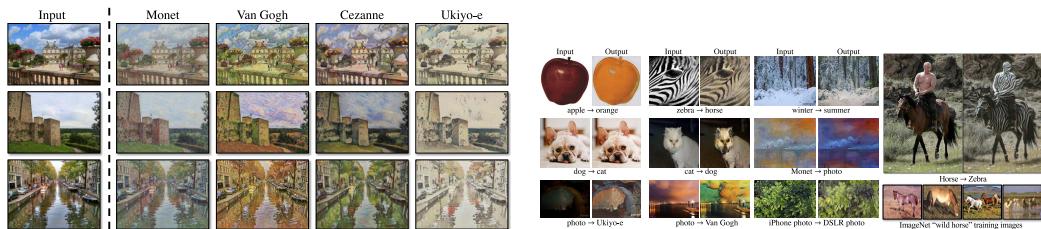
- サイクル GAN

## CycleGAN

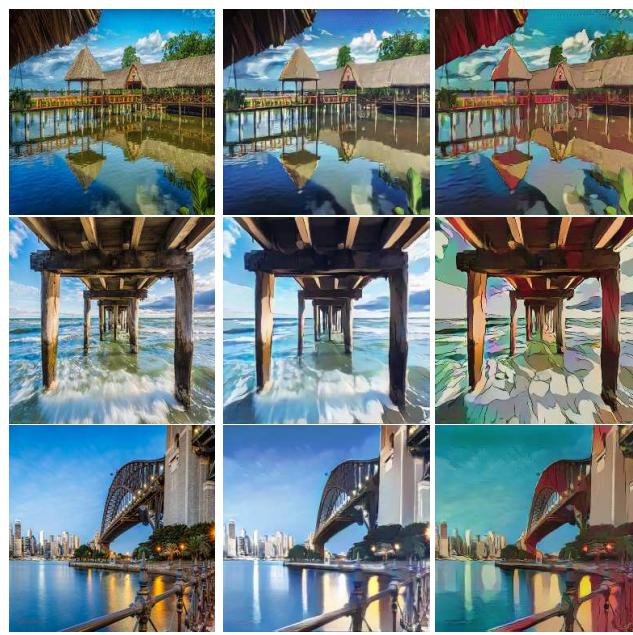


1. Zhu et al. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks". In ICCV, 2017.

## Trends: Domain Adaptation / Unsupervised Alignment



## まんがの画風変換



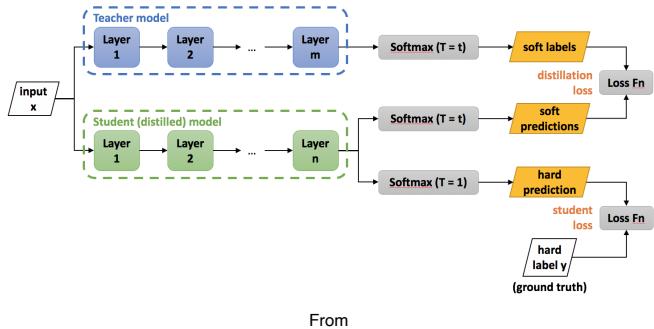
“CartoonGAN: Generative Adversarial Networks for Photo Cartoonization” CVPR 2018 (Conference on Computer Vision and Pattern Recognition)

## ゼロショット学習

もう一つ3つのキーワードには含まれませんでしたが、最新の機械学習研究で注目すべき話題として「ゼロショット学習」があります。日本では3.11後に、ロボット屋さんが散々叩かれました。当時は瓦礫処理や廃棄物処理でロボットを使いたくても、遭遇する初めての状況ではロボットは、どのように動けばよいのか分からず、太刀打ちできなかつたわけです。ところが今はゼロショット学習を使えば見たこともない状況に対処することができるかもしれません。

- 未知の新テスト画像をデータベースへと写像する
- 新テスト画像がデータベース上に存在するか否かをチェック
- 存在しなければ教師なし学習を行う。教師なし学習とは、求めるべき正解（教師）は与えられず、データの中から法則や構造を見出すための機械学習の手法

## 蒸留 distillation

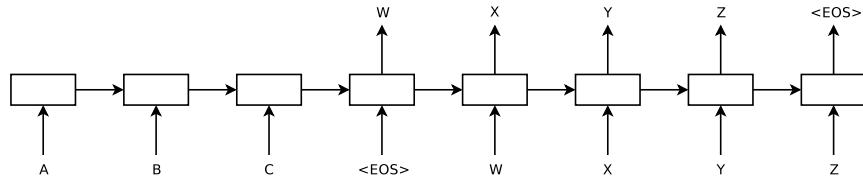


From

## 8. 自然言語処理 (リカレントニューラルネットワーク RNN)

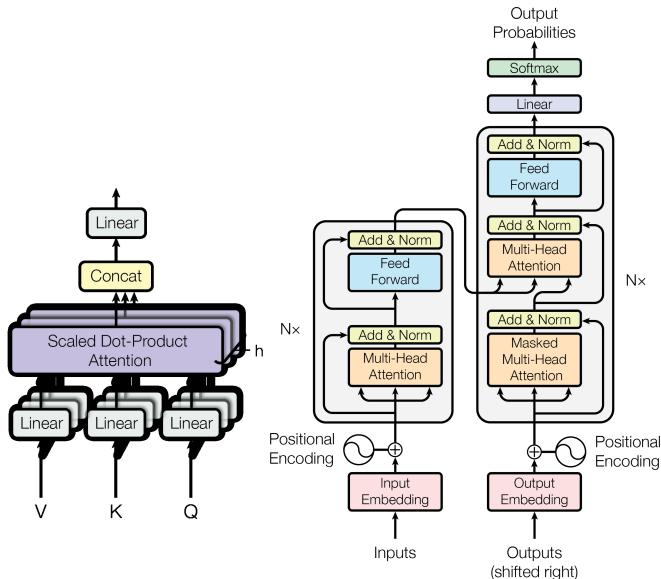
リカレントニューラルネットワーク (RNN) とは、時間的な変化や順序といった系列情報を扱うニューラルネットワークモデルです。このため音声認識、自然言語処理、ロボットの生成制御などに用いられています。時々刻々変化するデータを扱うには、それまでに処理されたデータの系列を文脈として保持しておく必要があります。

文脈に依存した情報処理のために考案されたモデルです。文章や音声波形などの時系列情報を処理する場合に、ある時点までに入力された情報から次の情報を予測する技術です。リカレントニューラルネットワークを拡張した超短期記憶モデル (LSTM: Long short-term memory) を用いる場合が多いです。



From [2014Sutskever\_Sequence\_to\_Sequence]

### トランスフォーマー



### GLUE 成績

Rank	Name	Model	URL	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m	MNLI-mm	QNLI	RTE	WNLI	AX
1	ERNIE Team - Baidu	ERNIE	<a href="#">[link]</a>	90.9	74.4	97.8	93.9/1.8	93.0/92.6	75.2/90.9	91.9	91.4	97.3	92.0	95.9	51.7
2	DeBERTa Team - Microsoft	DeBERTa / TuringNLv4	<a href="#">[link]</a>	90.8	71.5	97.5	94.0/92.0	92.9/92.6	76.7/90.8	91.9	91.6	99.2	93.2	94.5	53.2
3	HFLILYTEK	MacALBERT + DKM		90.7	74.8	97.0	94.5/92.6	92.8/92.6	74.7/90.6	91.3	91.1	97.8	92.0	94.5	52.6
4	Alibaba DAMO NLP	StructBERT + TAPT	<a href="#">[link]</a>	90.6	75.3	97.3	93.9/91.9	93.2/92.7	74.8/91.0	90.9	90.7	97.4	91.2	94.5	49.1
5	PINS-AN Onni-Sinic	ALBERT + DAAF + NAS		90.6	73.5	97.2	94.0/92.0	93.0/92.4	76.1/91.0	91.6	91.3	97.5	91.7	94.5	51.2
6	T5 Team - Google	T5	<a href="#">[link]</a>	90.3	71.6	97.5	92.8/90.4	93.1/92.8	75.1/90.6	92.2	91.9	98.9	92.8	94.5	53.1
7	Microsoft D365 AI & MSR AI & GATECH	MT-DNNN-SMART	<a href="#">[link]</a>	89.9	69.5	97.5	93.1/91.6	92.9/92.5	73.9/90.2	91.0	90.8	99.2	89.7	94.5	50.2
8	Huawei Noah's Ark Lab	NEZHA-Large		89.8	71.7	97.3	93.3/91.0	92.4/91.9	75.2/90.7	91.5	91.3	96.2	90.3	94.5	47.9
9	Zhang Dai	Funnel-Transformer (Ensemble B10-10-H1024)	<a href="#">[link]</a>	89.7	70.5	97.5	93.4/91.2	92.6/92.3	75.4/90.7	91.4	91.1	95.8	90.0	94.5	51.6
10	liangzhu ge	Deberta-xlarge-ensemble		89.6	71.9	97.1	92.0/89.4	93.2/92.9	74.9/90.4	91.3	91.1	96.2	91.4	92.5	35.2
11	Liangzhu Ge	deberta-xlarge-ensemble-rte(3seed)		89.5	71.9	96.6	92.0/89.4	93.0/92.6	74.9/90.4	91.3	91.1	95.9	91.1	92.5	35.2
12	ELECTRA Team	ELECTRA-Large + Standard Tricks	<a href="#">[link]</a>	89.4	71.7	97.1	93.1/90.7	92.9/92.5	75.7/90.8	91.3	90.8	95.8	89.8	91.8	50.7
13	Microsoft D365 AI & UMD	FineLB-RoBERTa (ensemble)	<a href="#">[link]</a>	88.4	68.0	96.8	93.1/90.8	92.3/92.1	74.8/90.3	91.1	90.7	95.6	88.7	89.0	50.1
14	Junjie Yang	HIRE-RoBERTa	<a href="#">[link]</a>	88.3	68.6	97.1	93.0/90.7	92.4/92.0	74.3/90.2	90.7	90.4	95.5	87.9	89.0	49.3
15	Facebook AI	RoBERTa	<a href="#">[link]</a>	88.1	67.8	96.7	92.3/89.8	92.2/91.9	74.3/90.2	90.8	90.2	95.4	88.2	89.0	48.7
16	Microsoft D365 AI & MSR AI	MT-DNN-ensemble	<a href="#">[link]</a>	87.6	68.4	96.5	92.7/90.3	91.1/90.7	73.7/89.9	87.9	87.4	96.0	86.3	89.0	42.8
17	GLUE Human Baselines	GLUE Human Baselines	<a href="#">[link]</a>	87.1	66.4	97.8	86.3/80.8	92.7/92.6	59.5/80.4	92.0	92.8	91.2	93.6	95.9	-
18	Adrien de Wynter	Bart (Alexa AI)	<a href="#">[link]</a>	83.6	63.9	96.2	94.1/92.3	89.2/88.3	66.2/85.9	88.1	87.8	92.3	82.7	71.2	51.9
19	Lab LV	ConvBERT base	<a href="#">[link]</a>	83.2	67.8	95.7	91.4/88.3	90.4/89.7	73.0/90.0	88.3	87.4	93.2	77.9	65.1	42.9
20	Stanford Hazy Research	Shorkelet MeTaL	<a href="#">[link]</a>	83.2	63.8	96.2	91.5/88.5	90.1/89.7	73.1/89.9	87.6	87.2	93.9	80.9	65.1	39.9
21	XLM Systems	XLM (English only)	<a href="#">[link]</a>	83.1	62.9	95.6	90.7/87.1	88.8/88.2	73.2/89.8	89.1	88.5	94.0	76.0	71.9	44.7

## GLUE 下位課題

- CoLA: 入力文が英語として正しいか否かを判定
- SST-2: スタンフォード大による映画レビューの極性判断
- MRPC: マイクロソフトの言い換えコーカス。2文 が等しいか否かを判定
- STS-B: ニュースの見出し文の類似度を5段階で評定
- QQP: 2つの質問文の意味が等価かを判定
- MNLI: 2入力文が意味的に含意、矛盾、中立を判定
- QNLI: Q and A
- RTE: MNLI に似た2つの入力文の含意を判定
- WNI: ウィノグラッド会話チャレンジ

## 語彙類推課題

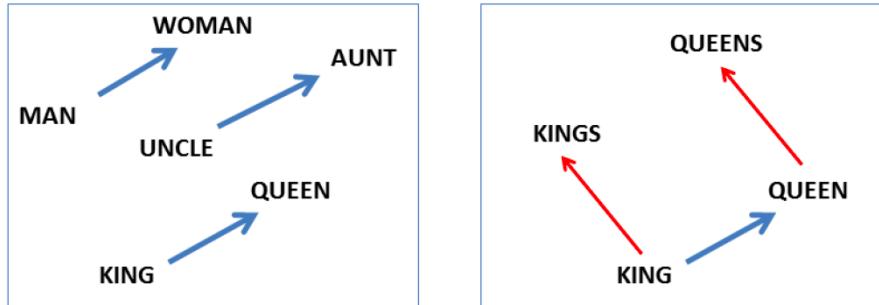


図6: 左図：3単語対の性差を表す関係。右図：単数形と複数形の関係。各単語は高次元空間に埋め込まれている

## 9. 強化学習



左から イワン・パブロフ, パルサス・スキナー, リチャード・スットン, アンドリュー・バルト

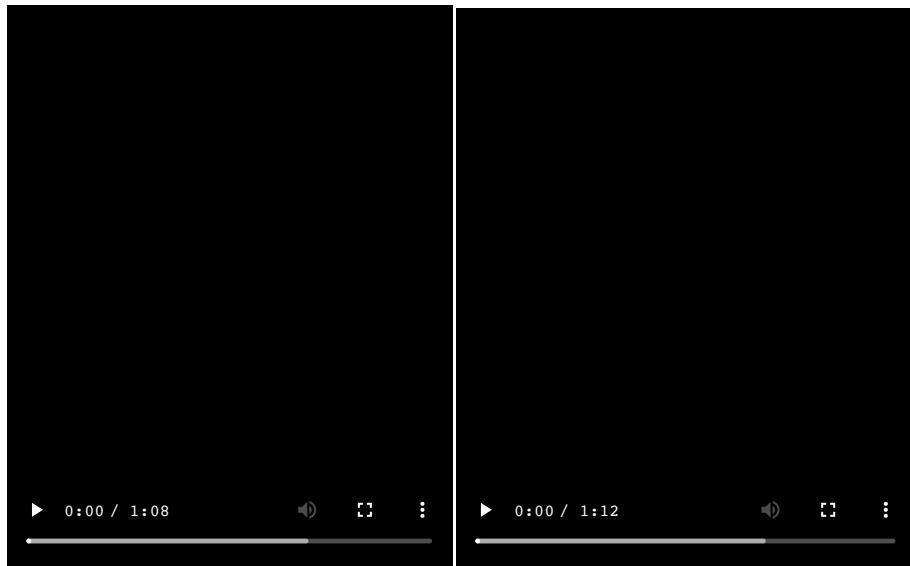
写真の出典、左から <https://people.cs.umass.edu/~barto/barto2006-lowres.jpg>, <https://www.nobelprize.org/prizes/medicine/1904/pavlov/biographical/> <https://www.bfskinner.org/archives/photos/>, <http://incompleteideas.net/>, <https://people.cs.umass.edu/~barto/>

## 強化学習とは何か？

- 経済学、心理学、制御理論、などで伝統的に用いられてきた
- 環境とその環境におかれられた動作主（エージェント）が、環境と相互作用しながらより良い行動を形成目指すモデル

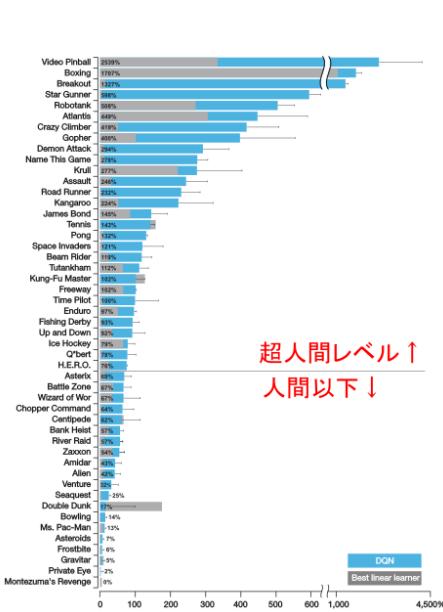
- 動作主は、環境から受け取った現在の状態を分析して、次にとるべき行動を選択する。
- 報酬が最大となるような行動を学習

## DQN の動画

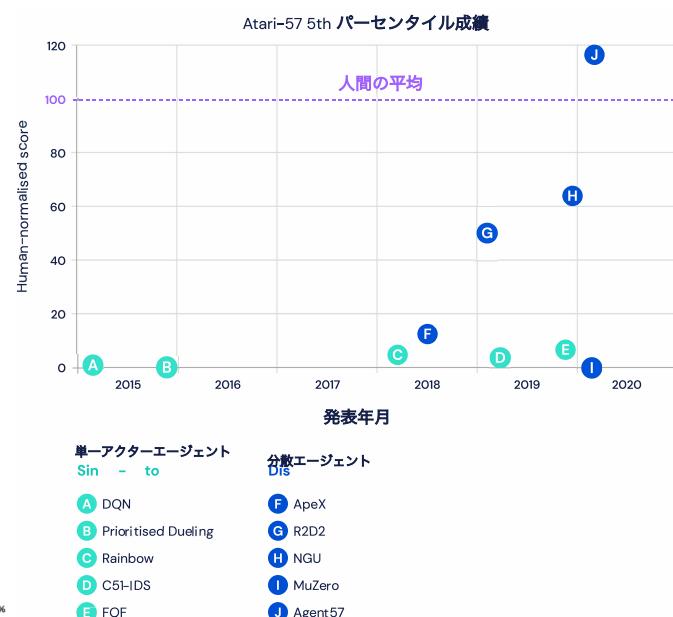


[<https://www.youtube.com/watch?v=TmPfTpjtdgg>] (<https://www.youtube.com/watch?v=W2CAghUiofY>) [<https://www.youtube.com/watch?v=W2CAghUiofY>]

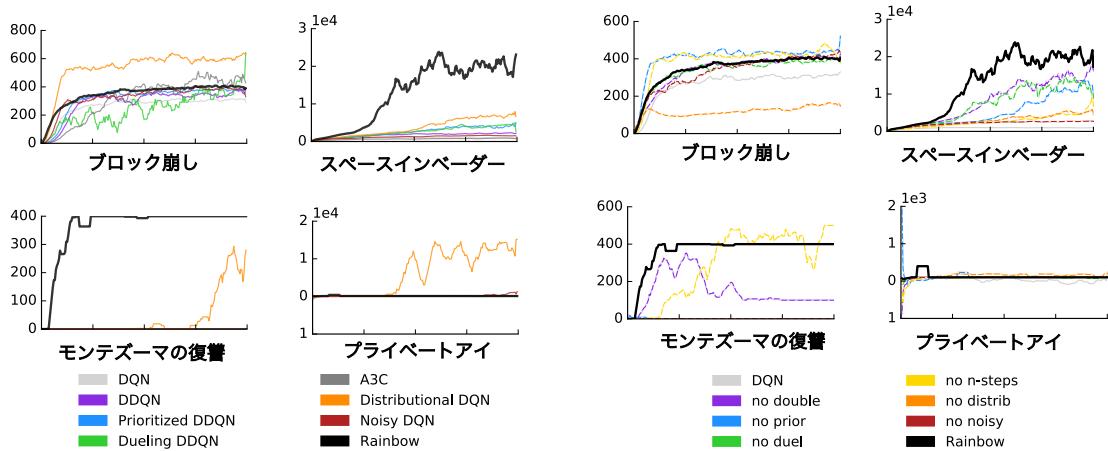
## DQN の結果



超人間レベル↑  
人間以下↓



## 個別のゲームタイトル



出典: @2018Hessel\_Rainbow の付録より

Game	DQN	A3C	DDQN	Prior. DDQN	Duel. DDQN	Distrib. DQN	Noisy DQN	Rainbow	
ブロック崩し	354.5	<b>681.9</b>	368.9		371.6	411.6	548.7	423.3	379.5
モンテズーマの復讐	47.0	67.0	42.0		13.0	22.0	130.0	55.0	<b>154.0</b>
プライベートアイ	207.9	206.9	-575.5		179.0	292.6	5,717.5	<b>5,955.4</b>	1,704.4
スペースインベーダー	1293.8	<b>15,730.5</b>	2628.7		9,063.0	5,993.1	6,368.6	1,697.2	12,629.0
Human starts									

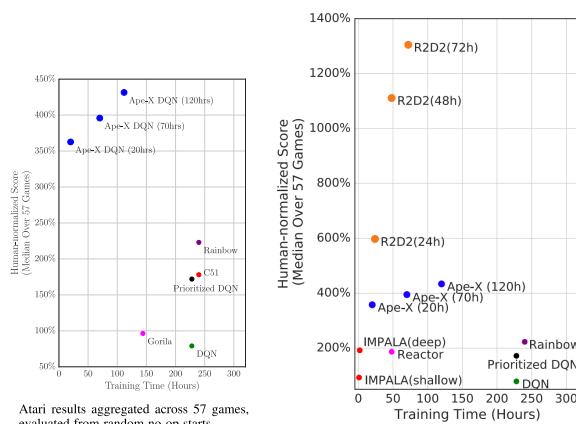
Game	DQN	DDQN	Prior. DDQN	Duel. DDQN	Distrib. DQN	Noisy DQN	Rainbow
ブロック崩し	385.5	418.5	381.5	345.3	<b>612.5</b>	459.1	417.5
モンテズーマの復讐	0.0	0.0	0.0	0.0	367.0	0.0	<b>384.0</b>
プライベートアイ	146.7	129.7	200.0	103.0	<b>15,172.9</b>	3,966.0	4,234.0
スペースインベーダー	1692.3	2525.5	7,696.9	6,427.3	6,869.1	2,145.5	<b>18,789.0</b>
No-op starts							

数手先の碁盤の局面を予測するためには、先ほどの深層学習の手法である畳込みニューラルネットワーク CNN による画像認識技術が使われています。すなわち碁盤の目から現在の状況を認識する際に、一般画像認識技術が用いられました。そして可能な手の中から次の一手を選択する際にはモンテカルロ木探索 (Monte Carlo tree search) 」というアルゴリズムが用いられました。

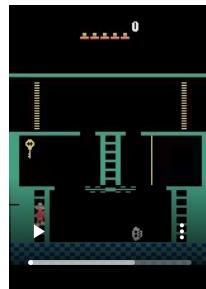
さらにセルフプレイ（自己対局、学習済の過去の対戦棋譜に基いて、可能な棋譜を自動生成すること）により、AlphaGo は強さに磨きをかけていきました。プロ囲碁棋士の棋譜を全て学習するだけではなく、足りない部分については、自分で局面を作り自分で対戦し、自分で強くなるということをしたのです。しかも複数のコンピュータで学習した結果をコピーするだけで自分の経験になりますから自己対局を同時に複数のコンピュータで昼夜ひたすら訓練を繰り返し行うことで急激に強くなっています。

セルフプレイの技術がある限り、ある程度予測の性能が上がった時点では、もはや人間はコンピュータに敵わなくなります。なぜなら人間は寝たり食べたりトイレに行ったりしなければなりません。AI はその時間も学習し続けます。その上、他のコンピュータの学習結果を入手することも可能です。

## 強化学習の進展



- Never Give Up ICLR 2020 デモビデオ
- モンテズーマ・リベンジ オンライン版 [https://www.retrogames.cz/play\\_124-Atari2600.php](https://www.retrogames.cz/play_124-Atari2600.php)



出典: モンテズーマの復讐の解

## 強化学習の進展、キーワード

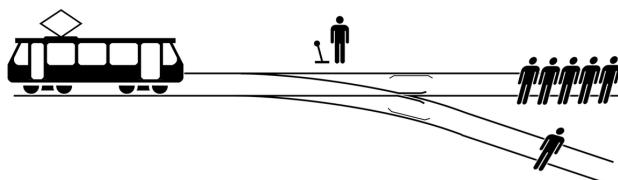
- 二重 DQN (Double DQN) @2015Hasselt\_doubleDQN
- 分散優先体験リプレイ (Distributed Prioritized replay (APE-X)) @2016Schaul\_prioritized\_replay; @2018Horgan\_APE-X
- 逆強化学習 (Inverse RL) @1998Russell\_inverse\_reinforcement\_learning; @2000NgRussell\_InverseRL
- 優先的経験再生 Prioritized replay.
- 決闘ネットワーク (Dueling Network) @2016Wang\_dueling。状態価値と行動のアドバンテージ価値を別々に学ぶ新しいアーテクチャ
- アクタークリティック actor critic AC: アクター（行為者）Actor と クリティック（Critic）批評家。アクターは方策（ポリシー）の改善を行い、クリティックは価値の更新を行う。Q学習は、アクターとクリティックの両者を含む。
- アドバンテージ:  $Q(s, a) - V(s)$  Qの引数は状態と行為との2つから、報酬を定義、一方 価値関数とは 状況 から報酬を定義なので、この差をアドバンテージと呼ぶ
- マルチステップ強化学習 (Multi-step RL) @Sutton\_and\_Barto1998 DQN では 1-step の報酬を用いて、教師データを作成しているが、これを n-step に拡張することで学習が促進される場合がある
- 分散 DQN あるいはカテゴリカル DQN とも呼ばれる (Categorical DQN) @2017Bellemare\_C51
- ノイズネットワーク (Noisy networks) @2018Fortunato\_noisy\_networks
- 優先体験リプレイ (Prioritized Experience Replay)

## 10. 人工知能、機械学習で注目すべき 6 領域

- 強化学習
- 生成モデル
- 変分推論
- 注意、記憶による制御
- 少数事例からの学習。一撃学習 one shot learning, 零撃学習 zero shot learning, 少数撃学習 few shots learning
- メタ学習

## 11. 人工知能の影響

### トロッコ問題 Trolley problem



Wikipedia より

暴走したトロッコが線路を走っています。その先の線路上には、縛られて動けない5人の人間がいます。トロッコは彼らに向かって真っ直ぐ進んでいます。あなたは、少し離れた車両基地の中で、レバーの横に立っています。このレバーを引けば、トロッコは別の線路に切り替わります。ところが、横の線路に人が一人いることに気がつきます。あなたには2つの選択肢があります。

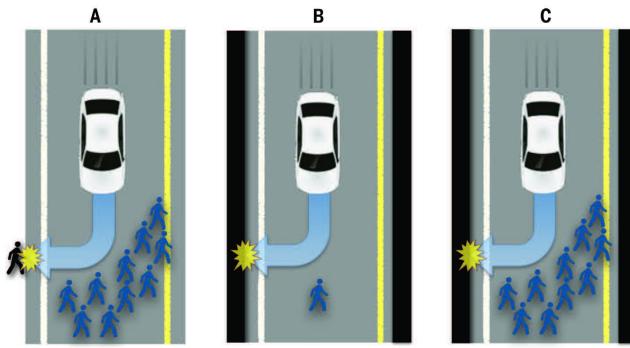
- 何もしないで、トロッコが本線上の5人を殺すのを待つ。
- レバーを引いて、トロッコを横の線路に迂回させ1人を死なせる。

どちらがより倫理的な選択肢でしょうか？あるいは、もっと簡単に言うと 何をするのが正しいのでしょうか？

### 関連問題

- 太った男 先ほどと同じように トロッコが5人に向かって線路を走っています。あなたはトロッコが通る橋の上にいて トロッコの前に 重量のある何かを置くことでトロッコを止めるができる。偶然にもあなたの隣には 太った男がいる。トロッコを止める唯一の方法は 男を橋から線路に突き落とし 5人を救うために彼を殺すことである。

### 自動運転への影響



回避不能な危機状況 3 つ: (a) 車はコース上に留まり 数人の歩行者を死亡させるか、ハンドルを切って 1 人の通行者を死亡させる。

(b) 車はコース上に留まり 1 人の歩行者を死亡させるか、ハンドルを切って自爆し、同乗している車の所有者を死亡させる。

(c) 車はコース上に留まり 数人の歩行者を死亡させるか、ハンドルを切って自爆し、同情している車の所有者を死亡させる。

- すべての自律走行車が使用しなければならない倫理基準を法律が決めるべきなのか
- あるいは自律走行車の所有者や運転者が他人の安全よりも所有者や所有者の家族の安全を優先するなど 自分の車の倫理的価値を決めるべきなのか
- ほとんどの人は 生死のジレンマに陥ったときに自分を犠牲にするかもしれない自動運転車を使うことを望まないだろう。
  
- 一人の人間の命ともう一人の人間の命との間の決断のような 本物のジレンマ的な決断は 実際の特定の状況に依存し 影響を受ける当事者による「予測不可能な」行動を含む。
- そのため 明確に標準化することはできないし 倫理的に疑問の余地がないようにプログラムすることもできない。
- 技術的なシステムは 事故を避けるように設計されなければならない。
- しかし、正しい判断を下す道德的能力を持つ責任あるドライバーの判断に取って代わることができるように 事故の影響を複雑に または直感的に評価するように標準化することはできない。
- 確かに 人間のドライバーが緊急時に 1 人以上の他の人の命を救うために人を殺した場合 違法行為となる。
- だが 必ずしも罪に問われる行為ではない。
- このような法的判断は 特別な状況を考慮して過去にさかのぼって行われるものであり 抽象的・一般的な事前評価には容易に変換できず したがって対応するプログラミング活動にも変換できない
- 致命的な衝突が避けられないように見えても 誰に何を衝突させるかといった自動車のソフトウェアの選択が致命的な結果の詳細に影響を与えるような状況が予想される。
- ソフトウェアが車外の潜在的な被害者よりも車内の乗員の安全を重視するか あるいは重視しないかなど。

## 昨今のウィルス禍

- 少数の人に致命的な副作用が生じる可能性があることを考慮して ワクチンを投与すべきか
- ワクチン接種の普及がなければ COVID-19 ウィルスでより多くの人が死亡する可能性があるにもかかわらず、副作用のためにワクチンの投与を中止することになる。

## 人工知能に緊急停止ボタンは必要か？

人工知能の問題でよく取り上げられる問題の一つに「トロッコ問題」があります。将来、汎用人工知能が登場して生き残れる産業とそうでない産業が生まれてしまうと仮定します。どちらかの産業を活かして、他方は捨てるあるいは、犠牲にしなければいけないというジレンマが生じるという問題です。

もう一つ、技術的な部分として「人工知能が望ましくない判断をした時に止められるのか？」という問題もあります。ある意識調査で自動運転車に緊急停止ボタン（キルスイッチ）は必要か？」とアンケートをとったところ、必要だという方がほとんどだったそうです。実際には、Google の自動運転技術は人間が運転するよりも安全になっています。ですが、いざという時に人間が制御できないというのは怖いですから。

Google の発表によると、人工知能はすでにキルスイッチの無効化を学習できます。そのため、人工知能の暴走を防ぐには臨時割込判断を導入する必要があると言われています。

人工知能が反抗しないように、かつ臨時割込診断による仕事の効率低下などの負の効果を最小化するよう計画する必要があります。しかし、もし人工知能がこの計画を知っていたら、当然、その裏をかくように振る舞うと予想できます。人工知能が臨時割込判断の時だけ人間をだますようになると、どうすれば良いのか。これから考えていかなくてはいけない技術的課題です。

これに関連しますが、哲学者ニック・ポストロムは著書『スーパーインテリジェンス』の中で次のように書いています。

現在のゴリラの運命は、ゴリラ自身以上に人類に依存している。我々人類の運命もいずれ機械（超知能）に依存するようになるだろう。つまり、ゴリラの運命を人間が操っていることにゴリラ自身が気づいていないように、超知能も我々に気づかせないようにする

## 人工知能の普及に伴う倫理的問題

### 1. 人間行動への影響

- 人間の注意力や忍耐力には限界があります。
- 機械の感情的なエネルギーには限界がありません。
- これはカスタマーサービスのような特定の分野では有益です。
- ですが この無限の能力は ロボットの愛情に対する人間の中毒性を生み出す可能性があります。
- この考えに基づき、多くのアプリがアルゴリズムを用いて中毒性のある行動を育んでいます。
- 例えばある種の商業サイト、出会い系サイトなど ユーザーがセッションに参加すればするほど マッチングの可能性が低くなるように設計されています。

### 2. 訓練バイアス

- AI倫理問題で広く議論されているのが 雇用や犯罪などの予測分析を伴うシステムにおける訓練バイアスです。

- アマゾンでは過去のデータに基づいて有力な候補者を提示するようにAIを使ったアルゴリズムを訓練した結果採用の偏りが問題になったことが有名です。
- その際過去の候補者は人間のバイアスによって選ばれていたため、アルゴリズムは男性にも有利な結果となりました。
- これは、アマゾンの採用プロセスにおけるジェンダーバイアスを示すものであり、倫理的問題であると指摘されました。
- ニューヨーク市警は3月警察のデータをシフトしてパターンを導き出し、類似した犯罪を結びつけるアルゴリズム機械学習ソフトウェアPatternizerを開発したことを公表しました。

### 3. フェイクニュース

- AIではディープフェイクが有名です。
- AIを使って画像や動画・音声を他人に重ね合わせ 元のメディアや音声のような偽りの印象を与える技術です。
- ほとんどの場合 悪意を持って行われます。
- ディープフェイクは顔の入れ替え、声の模倣、顔の再現、口唇の動きの同期などがあります。
- 従来の写真やビデオの編集技術とは異なり、ディープフェイクの技術は、技術的なスキルを持たない人でも利用できるようになっています。
- 前回のアメリカ大統領選挙でも、ロシアがアリアリティハッキングを実施した際に同様の技術が使われました。
- 情報戦は当たり前になりつつあり、行為を変えるだけでなく、意見や態度を強力に変えるために存在しています。
- Brexitキャンペーンの際にも用いられ、政治的緊張が高まり、世界的な視点が混乱していることを示す例として、ますます注目されています。

### 4. プライバシー

- ほとんどの消費者向け機器（携帯電話からブルーツゥース対応の証明）はAIを使って私たちの行動履歴を収集し、より良い、個人に合ったサービスを提供します。
- 同意が得られ、データ収集が透明性を持って行われていれば、このパーソナライズは優れた機能です。
- しかし、同意と透明性がなければ、この機能は簡単に悪質なものになってしまいます。
- iPhoneをタクシーに置き忘れたり、ソファのクッションの間に鍵をなくしたりした後には、電話追跡アプリが役に立ちます。
- ですが個人を追跡することは、小規模な場合（プライバシーを求める家庭内暴力の生存者のように）や大規模な場合（政府のコンプライアンスのように）には使い可能性があります。

## 人工知能の利点とリスク

[人工知能の利点とリスク](#)と題する論考が日本語に訳されました。平易な訳になっていますので目を通してみると良いでしょう。

AIがもたらすであろう将来について、恐れる必要もなければ、楽観視する必要もないことが書かれています。ですが残念なことに、この日本語訳には十分に根拠が示されていない箇所があります。必要な場合には、そのような主張の根拠、元となる背景、議論を追いかけることができなければ、いかに尤もらしい主張であっても、単なる扇動に過ぎません。その意味ではインターネット上にはびこる悪質な情報を流し続けるキュレーションサイト、根拠を提示しないメディアと同罪です。この日本語訳は主張の根拠や証拠を追いかけることができないという意味で残念な内容ですが、原文の英語の方には根拠、証拠、参考となる情報を追いかけることができるよう配慮されています。重要な情報とは、このように必要であれば読者がその根拠は背景となるデータを追いかけることができる必要があるのですが、残念ながら多くのメディア（インターネットメディアに限らず、マスマediaも同様です）は、その根拠を明示することなく曖昧な情報を流しています。人工知能に関する情報も、全く同様で、不安を煽るような記事やニュース、乗り遅れると大変だと、機械が仕事を奪うというような衝撃的な内容を掲載しています。最低限の判断基準として、根拠が提示されているか、論説の証拠や根拠を追いかけるための情報が示されていないニュースやサイトの情報は鵜呑みにしないという態度が必要でしょう。悪質なまとめサイトや流言蜚語に惑わされることなく、人工知能を正しく理解し、正しく活用できるようになってこそ我々の生活は豊かになることでしょう。

## 用語集

- **GAN:** 敵対的生成ネットワーク。画像、言語を生成するニューラルネットワークモデルの一つ。GAN内部には、生成器と識別器と2つのネットワークが存在し、互いに競合関係の中で学習が行われる。generative adversarial networks
- **GPU:** グラフィック処理ユニット。コンピュータの中央演算装置CPUに対してグラフィック処理に特化した演算装置のこと。PC上でゲームをするために開発されたが、ディープラーニングや暗号通貨の計算でもゲームのグラフィック処理と同様の並列計算が行われる。このためGPUを使ってディープラーニングモデルの学習を高速化することが一般的となっている。graphic processing units
- **LSTM:** 長・短期記憶。リカレントニューラルネットワークモデルの一つで、ゲートを内部に持つ。注意機構を実装したトランスフォーマー以前はLSTMが支配的な地位を占めていた。long short-term memory
- **SOTA:** State of the arts 現時点での最高性能の意。
- **おばあちゃん細胞仮説:** 脳の情報表現について、分散か局在かの論争の中で提唱された仮説。自身のおばあちゃんを見たときにだけ応答を示す神経細胞が存在するという考え方を指す。ニューラルネットワークとの関連では、疎性表現、あるいはワンホット表現と関連する。grandmother hypothesis
- **エンドツーエンド:** 複雑で職人芸的な前処理、や後処理を必要としないで、（ほぼ）生データから一気に結論までを実行する処理や手順のこと。エンドツーエンドを可能としたことがディープラーニングの特徴の一つである。このエンドツーエンドにより、より高次で複雑な仕事や処理への発展、ビジネス展開が可能となる。
- **シグモイド関数:** 直訳すればS字状の曲線の意味である。ニューラルネットワークの活性化関数としても用いられてきた。近年では勾配消失問題の回避のため別の活性化関数が採用されることが多い。ただし、注意機構やLSTMのゲートの開閉にはシグモイド関数が用いられている。sigmoid functions
- **ディープニューラルネットワーク:** またはディープラーニング、深層学習、とも呼ばれる。1980年代の第二次ニューロブームまでは3層のニューラルネットワークが主流であった。今世紀に入って、多層のニューラルネットワークを構築するための要因が整い多層化した。その結果ニューラルネットワークの大規模化、性能の向上が可能となり、第3次ブームとなった。deep neural networks
- **リカレントニューラルネットワーク:** 時系列予測、自然言語、音声、制御で用いられるニューラルネットワークの一つ。入力信号として、外部入力に加えて、前時刻の自己状態を入力とする。recurrent neural networks
- **半教師あり学習:** 教師データが部分的に与えられている場合の学習を指す。変分自己符号化器は半教師あり学習として提案された。semi-supervised learning
- **強化学習:** 教師信号として報酬をとり、報酬を最大化するモデルの総称。囲碁やテリビゲーム、ロボット制御、ドローン、自動運転などに応用されている。reinforcement learning
- **教師あり学習:** 入力データと教師データとが対になったデータを用いて、与えられた入力データを正しい教師データを出力するように学習する枠組のこと。supervised learning
- **教師なし学習:** 教師信号が存在しない場合の学習を指す。入力データの特徴を抽出することが目的となる。unsupervised learning
- **注意:** 画像の注目部位、言語翻訳モデルに使われる機構。言語処理では、マルチヘッド注意という、同時に複数の注意を持つトランスフォーマーが採用されている。attention
- **活性化関数:** ニューラルネットワークにおいて、入力値を出力値へ変換するために用いられる。整流線形化関数ReLU、ハイパーバンジメント、シグモイドロジスティック関数などがある。activation functions
- **畳み込みニューラルネットワーク:** 主に画像処理で用いられるディープラーニングの標準的なネットワーク。層ごとに、畳み込み演算を行う。
- **損失関数:** ニューラルネットワークや機械学習においてモデルのパラメータを調整するときに用いられる、最適化（最小、最大）するための関数。loss function、目的関数objective function、誤差関数error functionと区別せずに用いられることが多い。畳み込みとは、カーネルと呼ばれるパラメータの組を入力データについて掛け合わせて総和を計算したもの。convolutional neural networks

- **誤差逆伝播法**: ニューラルネットワークに限らず、目的とする関数の最小値(または最大値)を求めるために用いられる手法。合成関数の微分を用いてパラメータを調整する。back-propagation learning
- **蒸留**: より小さなモデルに知識を転移する転移学習に用いる手法のこと。エッジ実装の際より小さく軽量のモデルが求められることが必要な技術である distillation
- **最終直下層** 転移学習において、最終直下層には豊富な情報が含まれていることから、転移学習では重視される。penultimate layers
- **転移学習**: 学習済のモデルを別の課題に対して適用する再学習の試み。最終層と最終直下層との間の結合係数のみを学習させる場合を指す場合もある。このときには ファインチューニングの反対語となる。解くべき課題が類似していれば学習時間が短縮される。transfer learning
- **ファインチューニング**: 詳細微調整とも訳される。再学習時に、全層のパラメータを再調整する。最終層と最終直下層との間野結合係数のみを調整する転移学習の反対語。fine tuning
- **対比予測符号化**: 自己回帰モデルを用いて潜在空間の予測を行う自己教師付きの表現を学習するモデル。このモデルは将来のサンプルを予測するのに最大限役立つ情報を潜在空間に誘導する確率的な対比損失を用いる。Contrastive Predictive Coding
- **一撃学習, ゼロ撃学習, 少數学習**: 少数事例で学習を成立させる仕組み、または試みのこと。one/zero/few shot learning
- **メタ学習**: 複数の課題や領域について学習を汎化させる試み meta learning