

子どもと AI 第 25 回子ども健康科学会

浅川伸一

2024 03 月 02 日 駒澤大学

自己紹介



浅川伸一(あさかわ しんいち) asakawa@ieee.org
東京女子大学 情報処理センター

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エフ・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かを考えていると思っていた。著書に「AI白書2019, 2018」(2019年, アスキー出版, 共著), 「深層学習教科書ディープラーニングG検定(ジェネラリスト)公式テキスト」(2018年, 翔泳社, 共著), 「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「コネクショニストモデルと心理学」(2001)北大路書房など。

謝辞

第25回日本子ども健康科学会にて、講演の機会をお与えいただきました大会長 永田陽子先生、司会の労をとっていただきました堀内 雅彦先生にお礼申し上げます。

目次

1. はじめに
2. 心理学と人工知能の現在
3. 世界モデル，他者モデル
4. 転移学習，ファインチューニング
5. まとめ

1. はじめに

1.1 大規模言語モデル (LLM) の性能向上

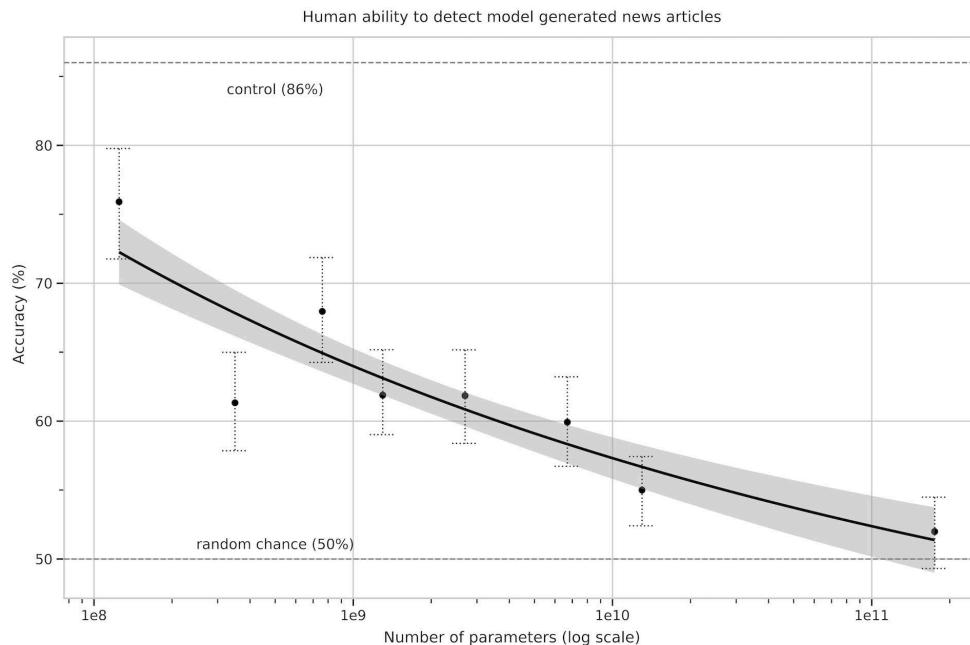
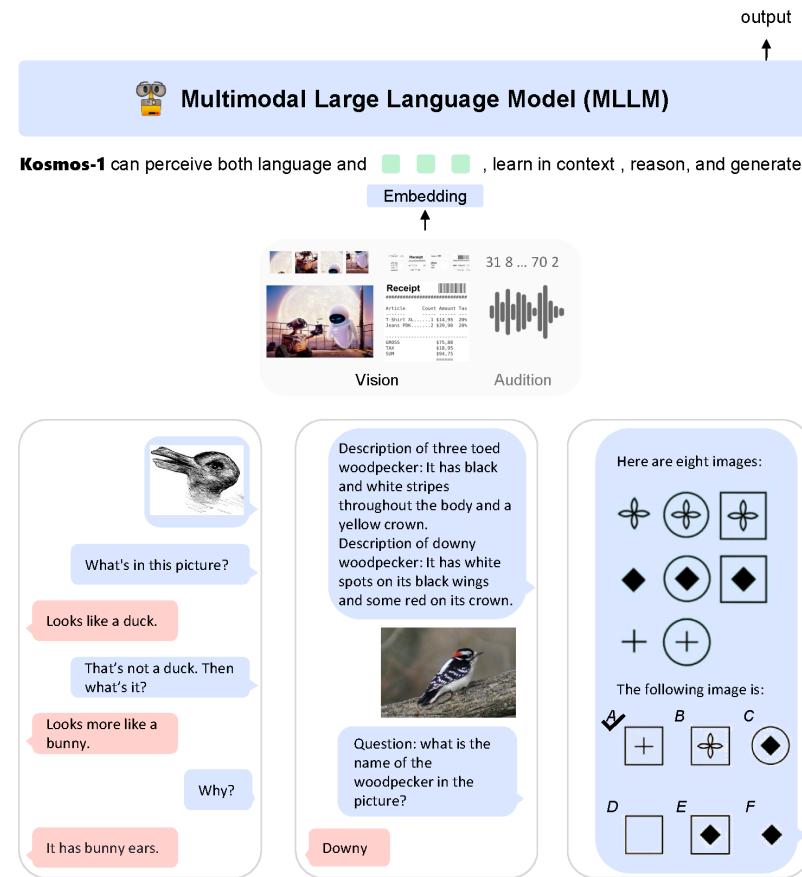


図 3.13: ニュース記事がモデルによって生成されたものであるかどうかを識別する人間の能力 (正しい割り当てと中立でない割り当ての比率で測定) は、モデルサイズが大きくなるほど低下する。意図的に悪い対照モデル (出力のランダム性が高い無条件 GPT-3 小型モデル) の出力に対する精度を上部の破線で示し、ランダムな確率 (50 %) を下部の破線で示す。ベストフィットの線は 95 %信頼区間を持つべき乗則である。ニュース記事がモデルによって生成されたものであるかどうかを識別する人間の能力 (正しい割り当てと中立でない割り当ての比率で測定) は、モデルサイズが大きくなるほど低下する。意図的に悪い対照モデル (出力のランダム性が高い無条件 GPT-3 小型モデル) の出力に対する精度を上部の破線で示し、ランダムな確率 (50 %) を下部の破線で示す。ベストフィットの線は 95 %信頼区間を持つべき乗則である。Brown+2021, arXiv:2005.14165 Fig. 3

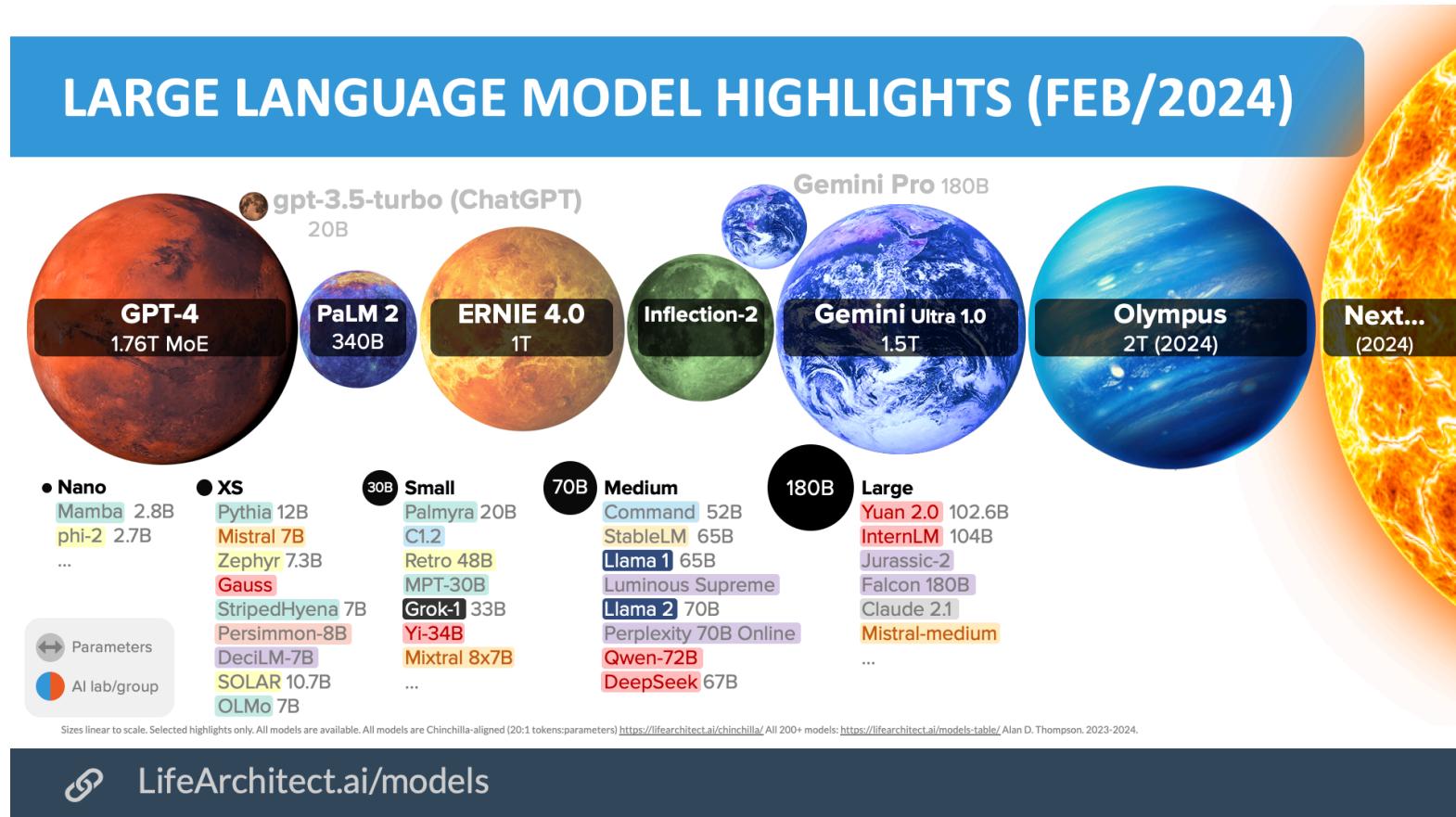
1.2 GPT-4

chatGPT の後継モデルである GPT-4 では、マルチモーダル、すなわち、視覚と言語の統合が進んだ。



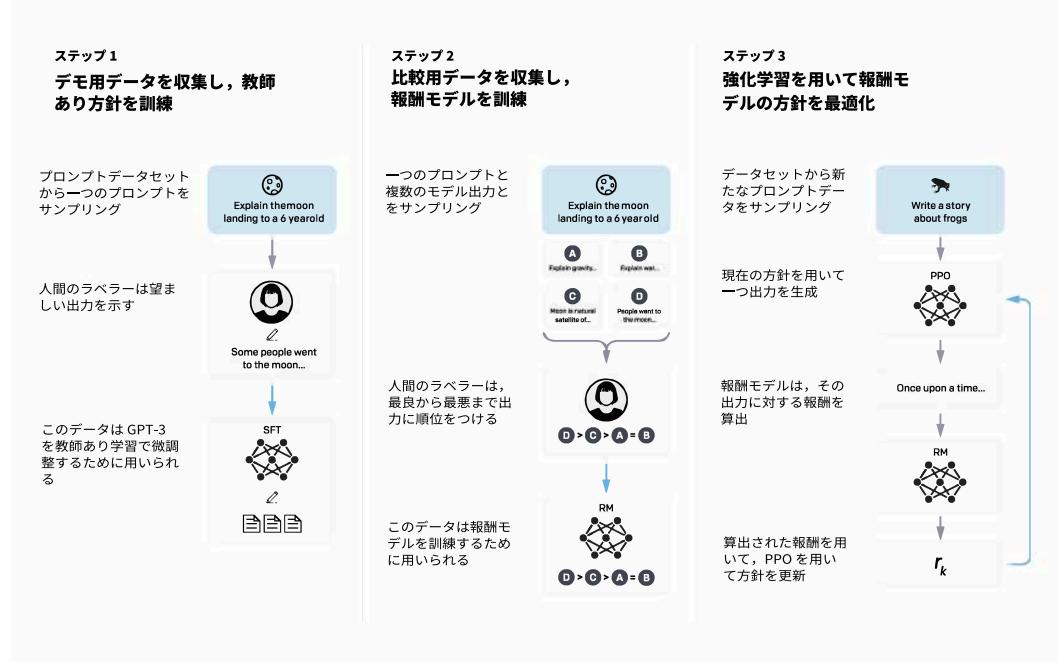
Kosmos-1 の概念図

1.3 LLM のコーパスサイズ



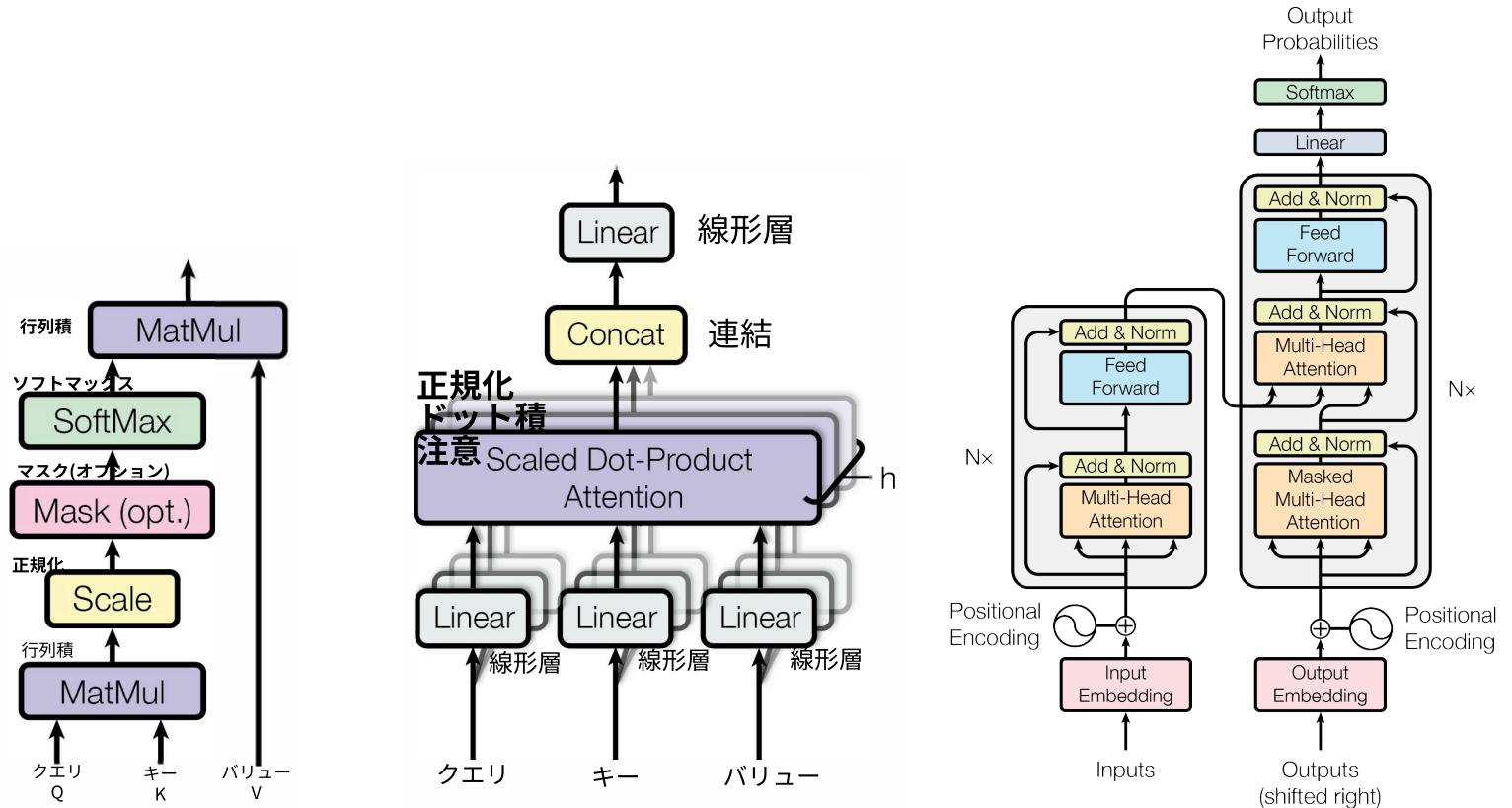
<https://s10251.pcdn.co/wp-content/uploads/2024/02/2024-Alan-D-Thompson-AI-Bubbles-Planets-Rev-1.png>

1.4 chatGPT の訓練



- 参考: 岩下, 吉原, 浅川 (2023) 自然言語処理を用いた例文生成とその妥当性—日本語教師の支援を目的とした BERT・T5 を用いた文生成シミュレーション—, 2023 年度 日本語教育学会春季大会

1.5 Transformer の概略図



Transformer 2017 Vaswani++ Fig.2 を改変

matmul は行列の積, scale は、平均 0 分散 1 への標準化, mask は 0 と 1 とで、データを制限すること, softmax はソフトマックス関数

Transformer における注意 = ソフトマックス関数。

2. 心理学と人工知能の現在

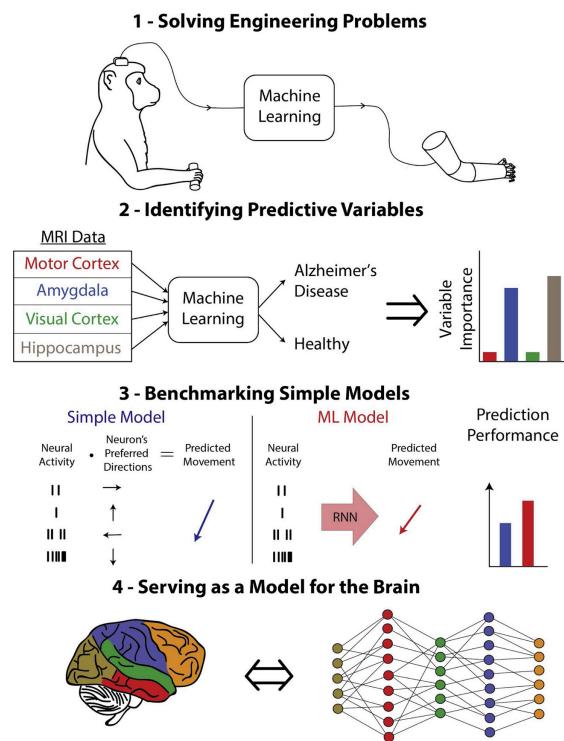
「学習」の違い

- 機械学習: 教師あり学習, 教師なし学習, 強化学習。データを用いてデータの性質を記述するための, モデルのパラメータを調整する。
- 心理学: 行動主義(強化学習), 認知心理学, 発達心理学などにおける学習では, 観察学習, 社会的, 文化的要因を含む意味で用いられる, 広義の行動変容。

関連すると思われる心理学理論

1. 行動主義 Watson, Skinner
2. Piaget の発達理論
3. Bandura の観察学習
4. Vygotsky の社会文化学習理論

2.1 Glaser(2019) の 教師あり機械学習の 4 つのレベル



1. 工学的な問題の解決: 機械学習は、医療診断、ブレインコンピュータインターフェース、研究ツールなど、神経科学者が使用する手法の予測性能を向上させることができる。
2. 予測可能な変数の特定: 機械学習により、脳や外界に関連する変数がお互いを予測しているかどうかをより正確に判断することができる。
3. 単純なモデルのベンチマーク: 解釈可能な簡易モデルと精度の高い ML モデルの性能を比較することで、簡易モデルの良し悪しを判断するのに役立つ。
4. 脳のモデルとしての役割: 脳が機械学習システム、例えばディープニューラルネットワークと同様の方法で問題を解決しているかどうかを論じることができる。

Glaser+2019 Fig. 2 より

2.1.1 視覚野と視覚モデルとの対応関係

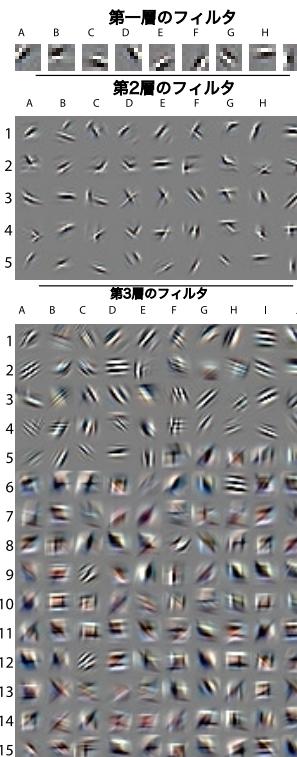
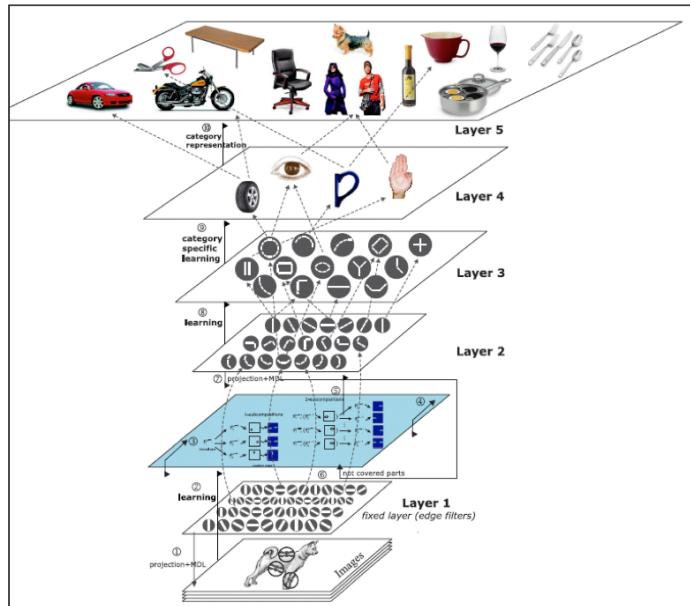


Figure 7. 食べ物の情景画像で訓練された各層のフィルタ。フィルタ多様性と、各層で増す複雑性に注目。図8のフィルタとは対照的にフィルターの方位選択性は均等に分布している。

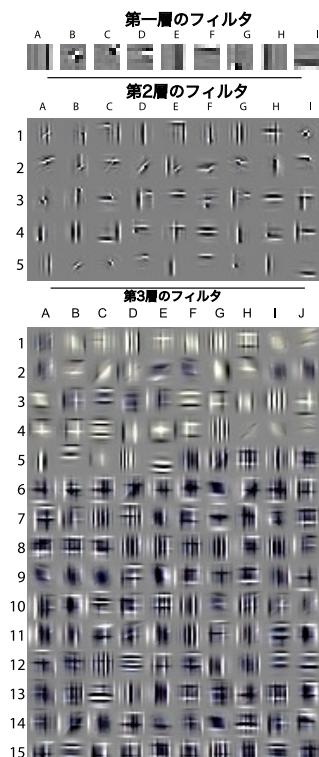
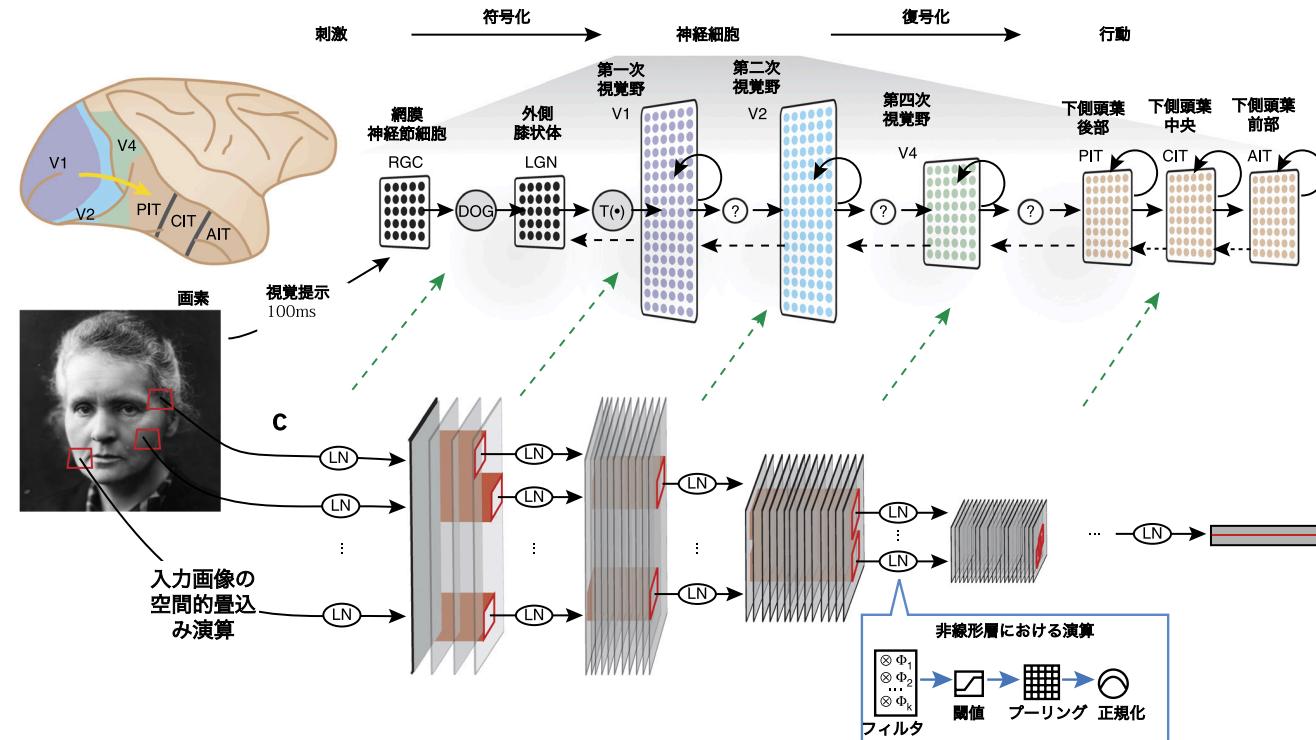


Figure 8. 都市データセットで訓練されたモデル各層のフィルタ。水平、および垂直方向の優位性に注目。

出典: 左 Zeiller2012, 右 Zeiller+2010, Fig. 7, and 8

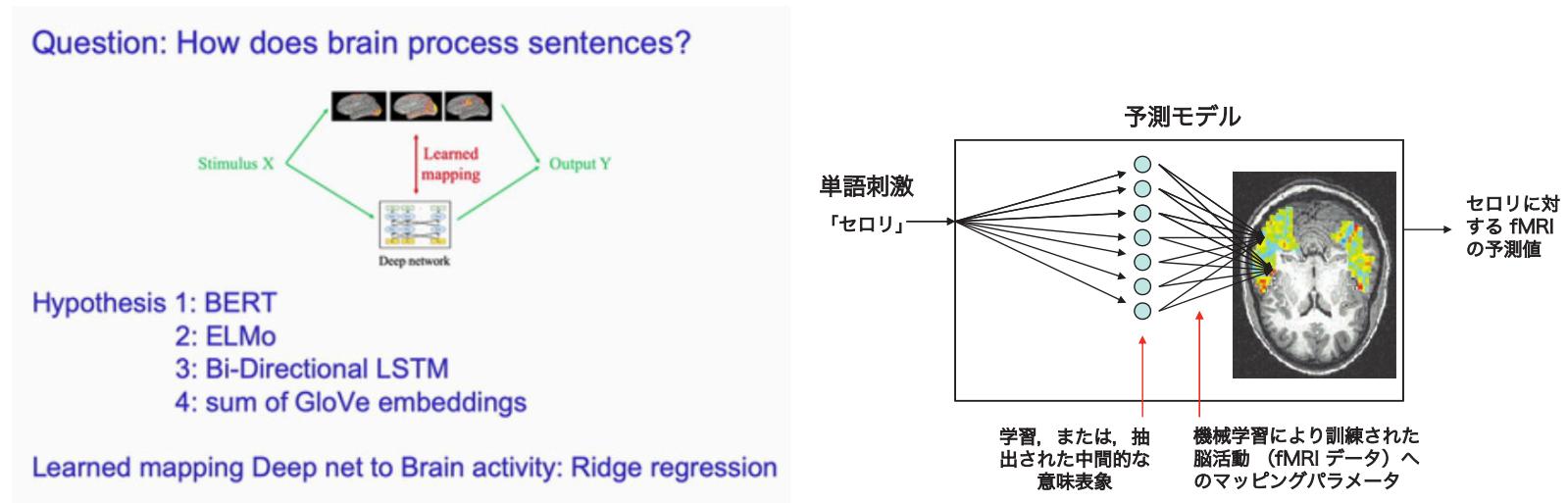
2.1.1 視覚野と視覚モデルとの対応関係 (2)



出典: Yamins+2016, Fig.1

2.1.2 単語埋め込み表現を用いた脳活動の予測

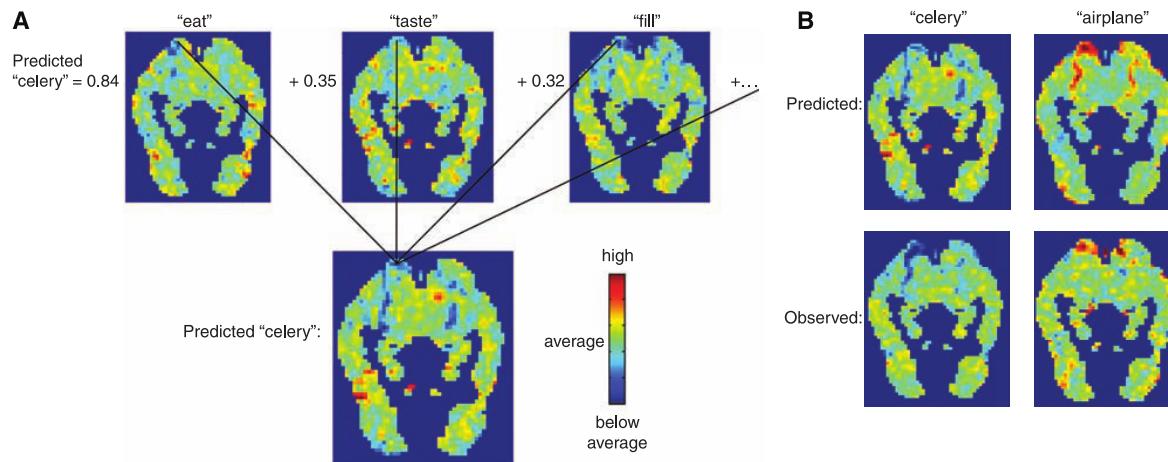
- 名詞の意味に関連した人間の脳活動の予測, Mitchell, 2018, Predicting Human Brain Activity Associated with the Meanings of Nouns



Mitchell (2008) 図 1. 任意の名詞刺激に対する fMRI 活性化を予測するモデルの形式。左のように「セロリ」から右の脳画像を予測するために、中間表現として、兆単位の言語コーパス（言語研究では訓練や検証に用いる言語データをコーパスと呼ぶ）から得られた意味特徴を用いる。

fMRI の活性化は 2 段階の処理から予測される。第 1 段階では、入力刺激語の意味を、典型的な単語使用を示す大規模なテキストコーパスから値を抽出した中間的な意味的特徴の観点から符号化する。第 2 段階では、これらの中間的な意味的特徴のそれぞれに関連する fMRI シグネチャの線形結合として、fMRI 画像を予測する。

2.1.2 単語埋め込み表現を用いた脳活動の予測 (2)



Mitchell+2008, 図 2. 与えられた刺激語に対する fMRI 画像の予測。

他の単語(下図左) eat, taste, fillなどの単語からセロリを予測する回帰モデルを使って予測する。(A) 参加者 P1 が「セロリ」刺激語に対して、他の 58 の単語で学習した後に予測を行う。25 個の意味的特徴のうち 3 つの特徴量のベクトルを単位長にスケーリングすることである。(食べる, 味わう, 満たす)について学習した c_{vi} 係数は、パネル上部の 3 つの画像のボクセルの色で示されている。刺激語「セロリ」に対する各特徴量の共起値は、それぞれの画像の左側に表示されている(例えば「食べる(セロリ)」の共起値は 0.84)。刺激語の活性化予測値((A) の下部に表示)は 25 個の意味的 fMRI シグネチャを線形結合し、その共起値で重み付けしたものである。この図は予測された三次元画像の1つの水平方向のスライス [z=-12 mm in Montreal Neurological Institute (MNI) space] を示している。(B) 「セロリ」と「飛行機」について、他の 58 個の単語を使った訓練後に予測された fMRI 画像と観察された fMRI 画像。予測画像と観測画像の上部(後方領域)付近にある赤と青の2本の長い縦筋は、左右の楔状回である。

2.4 Breiman2001 による 2 つの文化

データモデル:

- データからパラメータの値を推定し、そのモデルを情報収集や予測に利用
- 伝統的な統計学 p 値崇拜主義
- データモデルの限界
 - データモデルにこだわるあまり、統計学における多変量解析のツールは、分類では判別分析とロジスティック回帰、回帰では重回帰に固まってしまっている。
 - 多変量データが多変量正規分布であると本気で信じている人はいないが、多変量統計解析に関するすべての大学院の教科書では、このデータモデルが多くのページを占めている。
 - 未知の物理的、化学的、生物学的機序を含む複雑な系の制御されていない観察から収集されたデータでは、統計学者が選択したパラメトリックモデルによって自然がデータを生成するという先驗的な仮定は、適合度検定や残差分析に訴えても立証できない疑わしい結論になることがある。通常、複雑な系によって生成されたデータ、例えば医療データや金融データに単純なパラメトリックモデルを適用すると、アルゴリズムモデルと比較して精度と情報が損なわれる。

アルゴリズムモデル

暗箱の中は複雑で未知であると考える。アプローチは、関数 $f(x)$ 、すなわち x を操作して応答 y を予測するアルゴリズムを見出すこと。

- モデルの検証方法：予測精度によって測定する機械学習的手法

2.5 P 値廃絶宣言

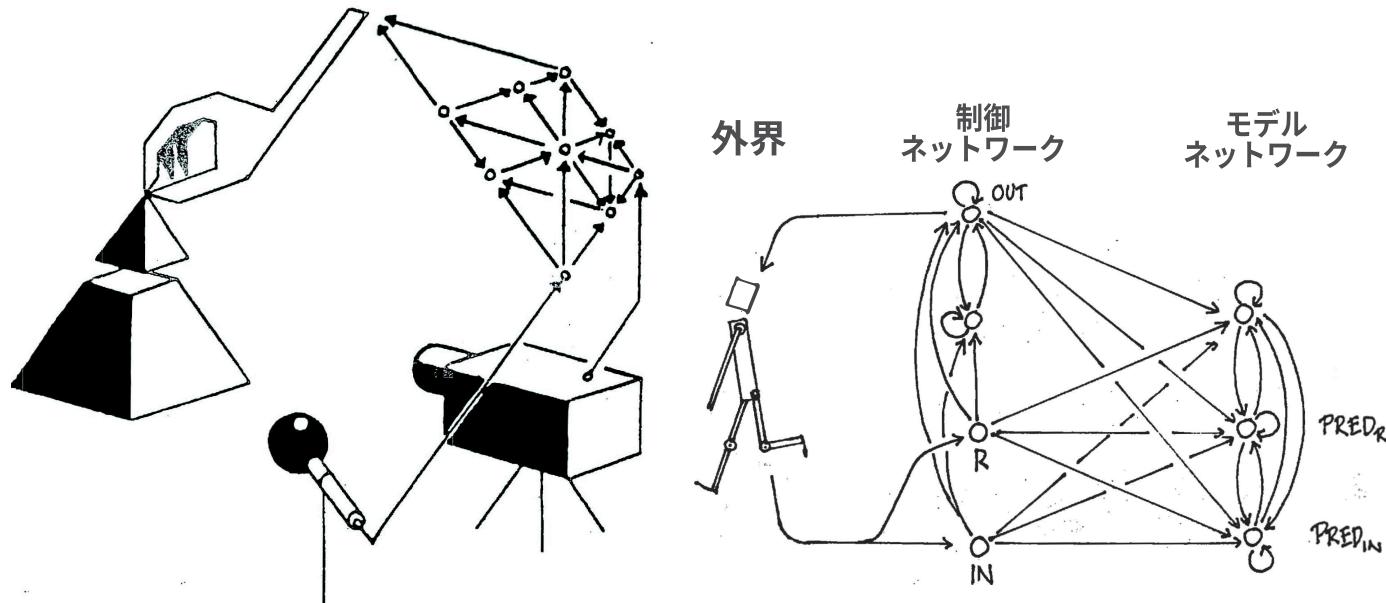
- ASA アメリカ統計学会の声明

1. *P* 値は、データが指定された統計モデルとの程度相性が悪いかを示すことができる
2. *P* 値は、研究された仮説が真である確率を測定するものではない。そうではなく、データがランダムな偶然だけから、生成された確率を測定するものである
3. 科学的な結論やビジネスや政策の決定は、*p* 値が特定の閾値を超えたかどうかだけに基づくべきではない
4. 適切な推論を行うには、完全な報告と透明性が必要である
5. *P* 値や統計的有意性は、効果の大きさや結果の重要性を測定するものではない
6. それ自体では、*p* 値はモデルや仮説に関する証拠の良い尺度を提供しない。

参考資料

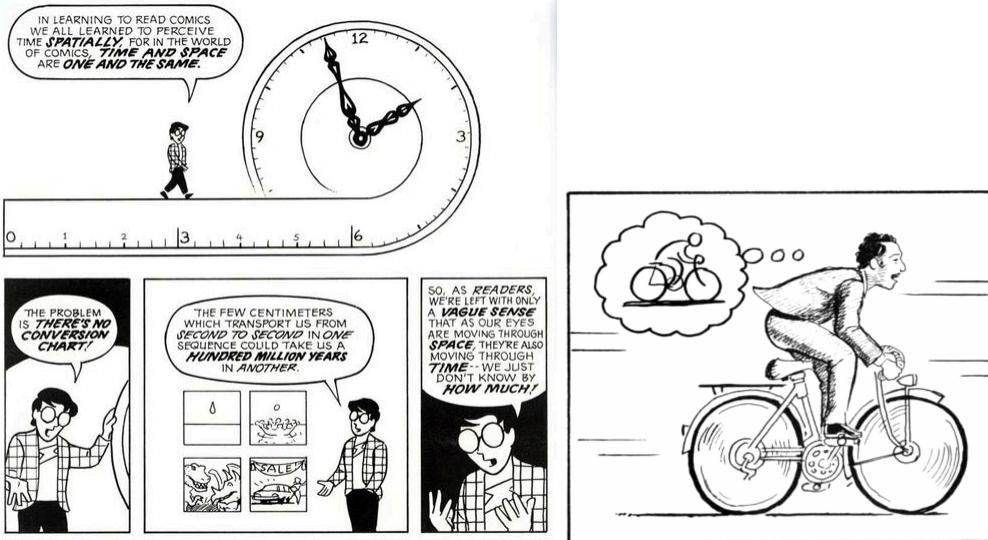
- 基礎と応用社会心理学 (BASP) 編集方針 (2014,2015)
- アメリカ統計学会の声明 2014, 2015
- 統計学の誤り：統計的妥当性の「ゴールドスタンダード」である P 値は多くの科学者が想定しているほど信頼できるものではない
- 統計的有意性を引退せよ

3. 世界モデル



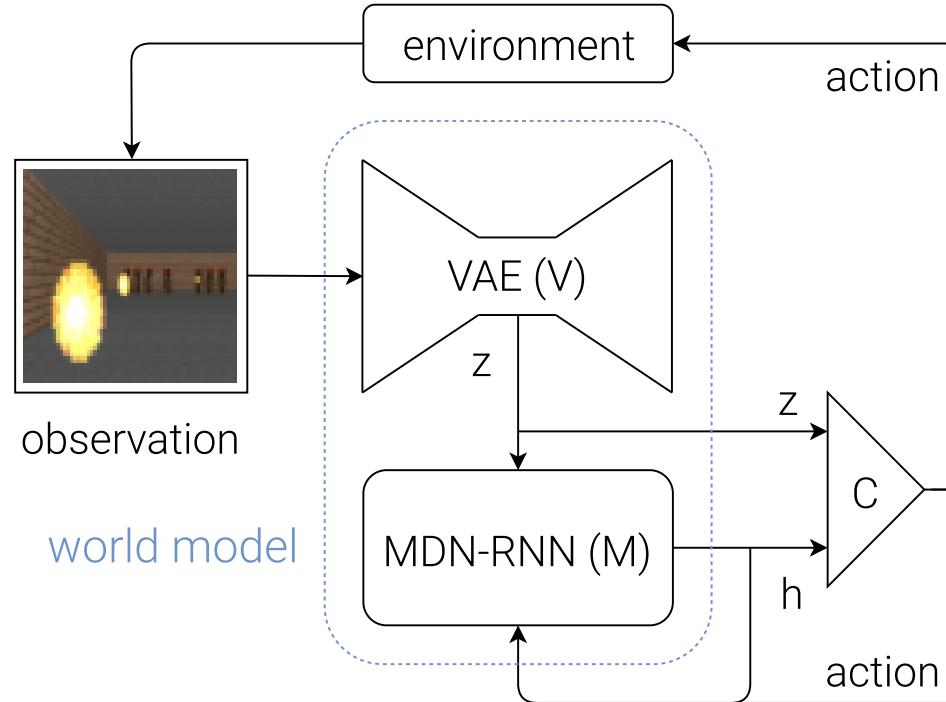
Schmithuber (1990) **Making in World Differentiable: On Using Self-Supervised Fully Recurrent Neural Networks for Dynamic Reinforcement Learning and Planning in Non-Stationary Environments**, Fig. 1 and 2.

3.1 世界モデル (2)



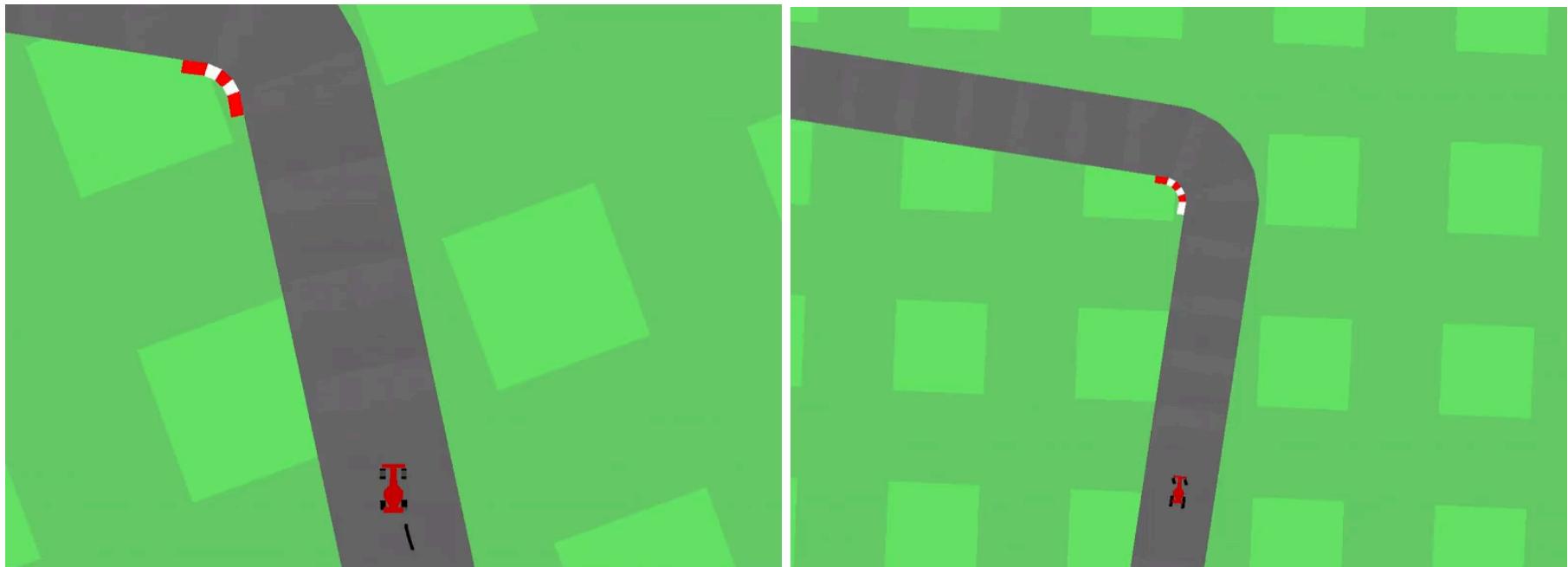
A World Model, from Scott McCloud's. From *Understanding Comics*.

3.2 世界モデル アーキテクチャ



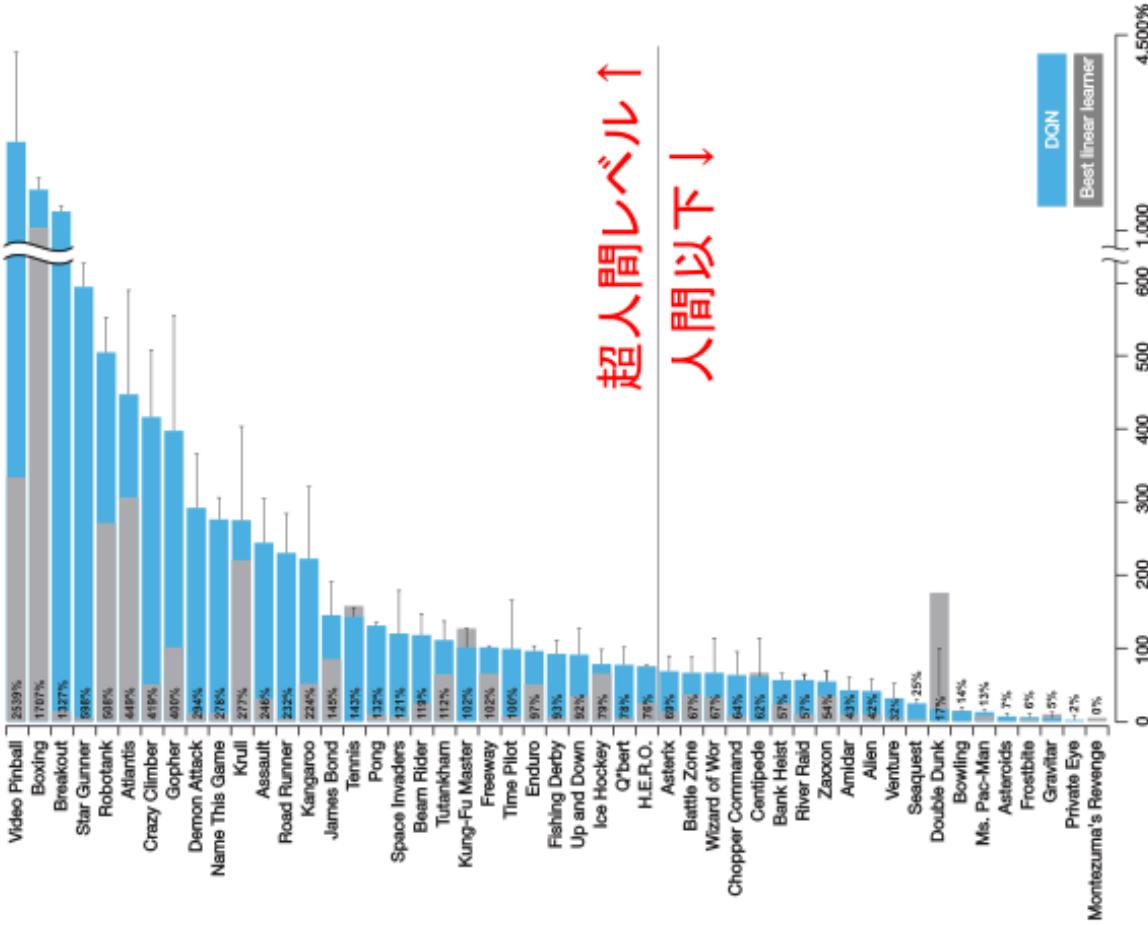
モデルのフロー図。観測データは、まず各時間ステップ t で視覚処理器 V によって処理され、潜在表現 z_t が生成される。コントローラ C への入力はこの潜在ベクトル z_t と各時間ステップでの、内部モデル M の隠れ状態 h_t が結合されたもの。 C は次に、運動制御のための行動ベクトル a_t を出力する。 M は現在の z_t と行動 a_t を入力として、自身の隠れ状態を更新し、時間 $t + 1$ で使用する h_{t+1} を生成。

3.3 世界モデル カーレース

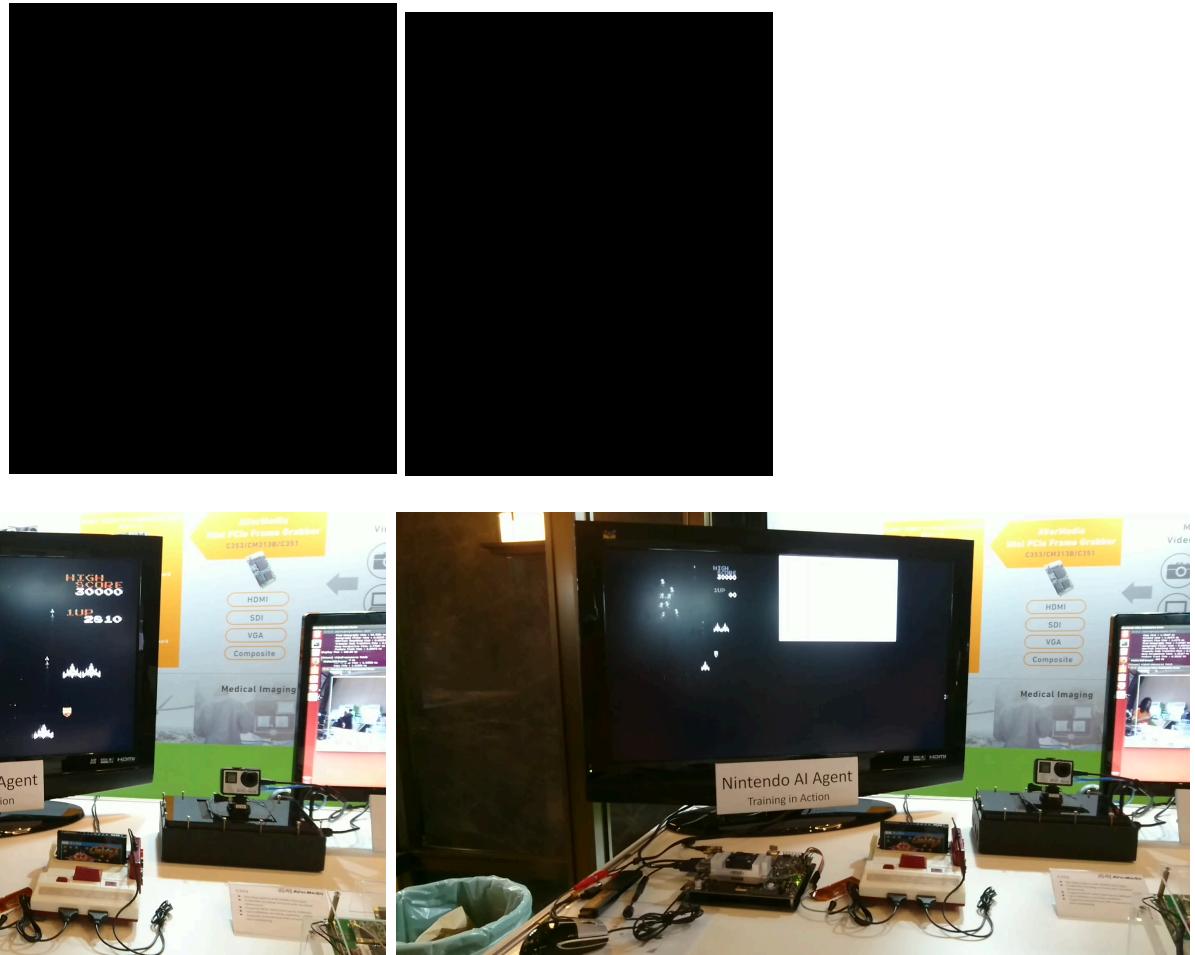


左: 外界入力の圧縮表現 z_t のみを用いた場合。右: 外界入力の圧縮表現 $z - t$ と内部モデルの中間層表現 h_t を使った場合。左図では、ふらついた不安定な行動となる

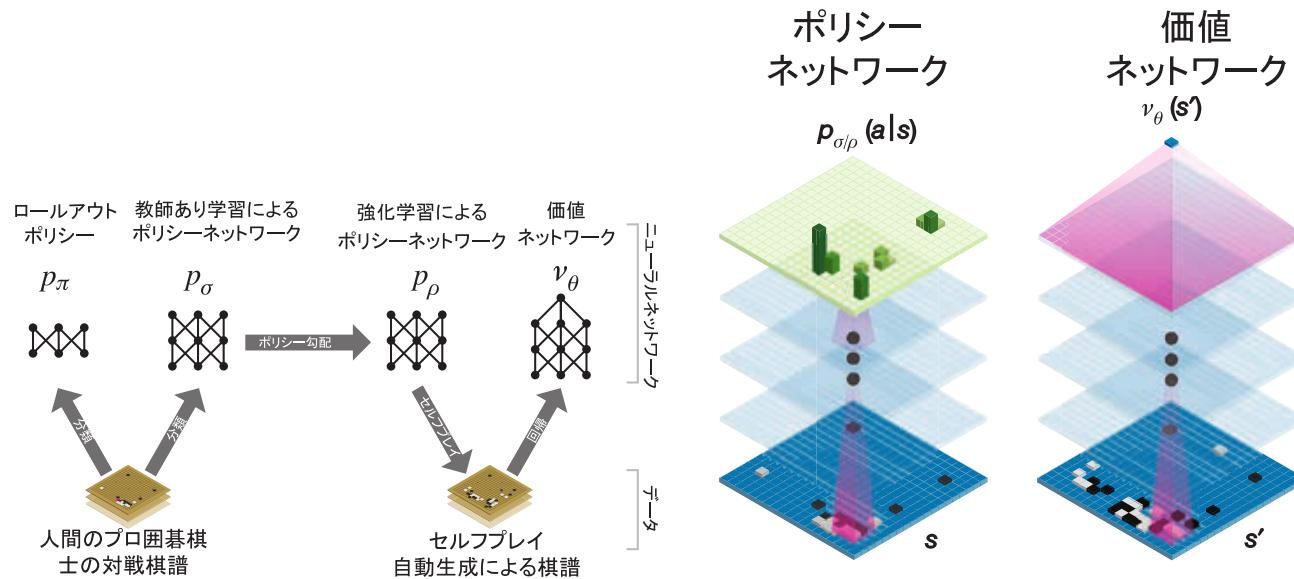
3.4 Agent57, DQN



3.4.2 DQN (2)

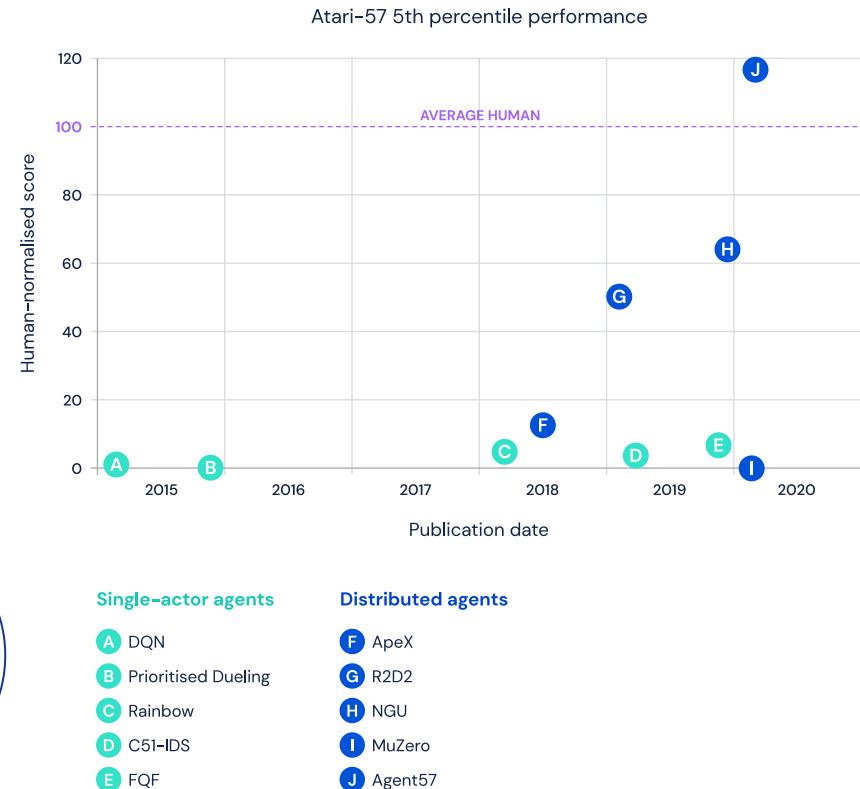
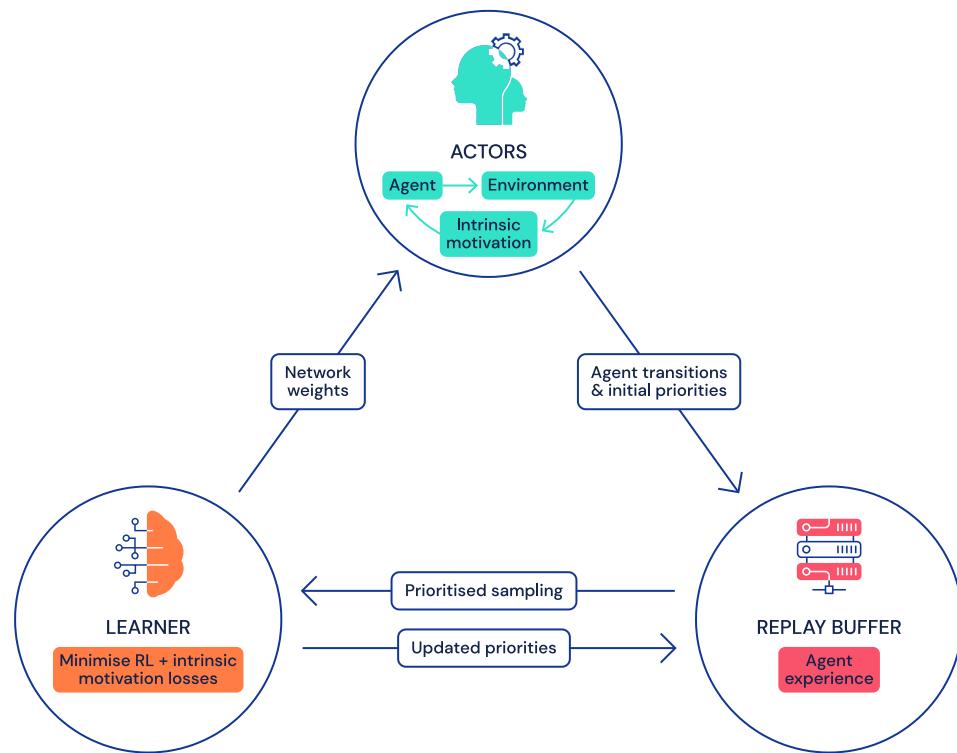


3.4.3 AlphaGo



Silver+(2016) Fig. 1 より

3.4.4 Agent57



3.4.5 DQN の改善

DQN の初期の改良では、二重 DQN、経験の優先再生、決闘アーキテクチャなど、学習効率と安定性が向上した。これらの変更により、エージェントは経験をより効率的かつ効果的に利用できるようになった。

- 分散エージェント
- 短期記憶
- エピソード記憶
- 直接的な探索を促すための本能的動機づけの方法
- 長い時間軸での新規性



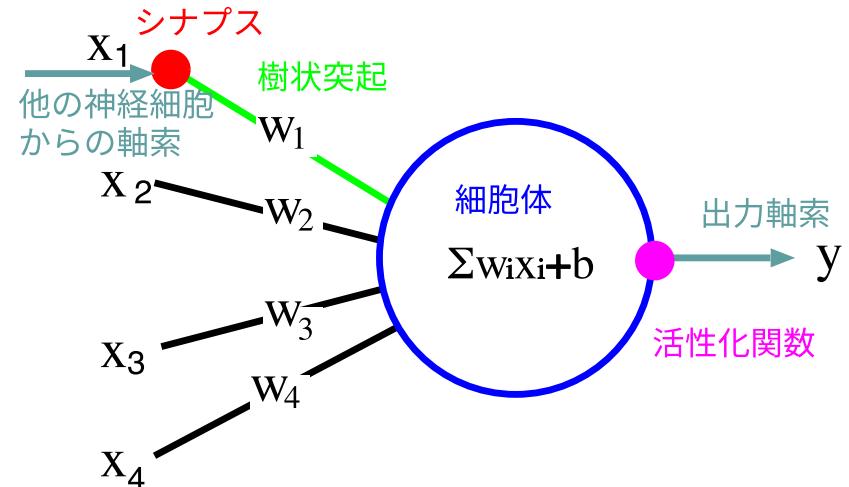
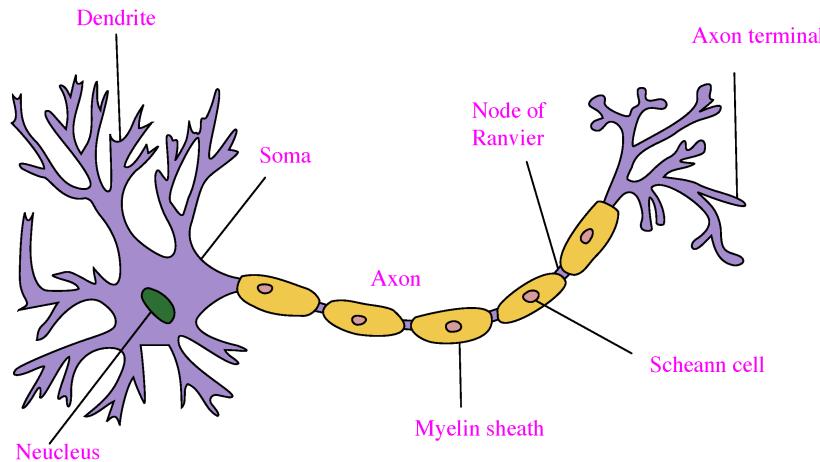
左から、ピットフォール、スキー、ソラリス、モンテズーマの復讐

- 短期間での新規性
- メタコントローラ：探索と利用のバランス
- Agent57: すべてをまとめる

5. まとめ

付録

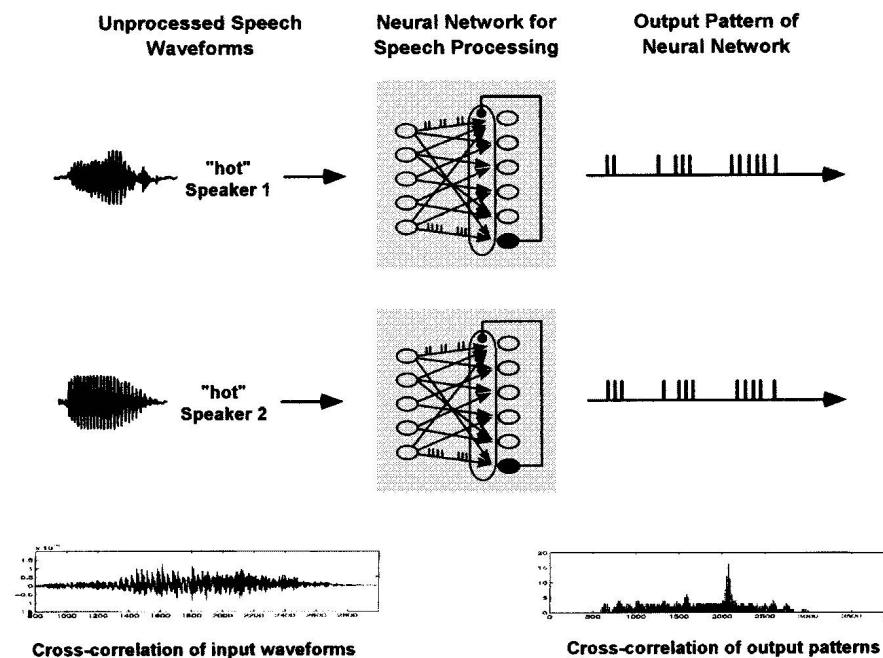
A.1 ニューラルネットワーク



左: ニューロンの模式図, wikipedia より。右: ニューロンの模式図に対応するニューラルネットワーク表記。

ニューロンは、多入力 (multi inputs), 一出力 (single output) と見做しうる。各入力 x_i に対して、対応するシナプス結合強度 w_i を乗じて、総てを足し合わせる (\sum)。さらにバイアス項 b を加えた後、活性化関数 (activation function) を用いて変換した値が出力 y となる。

実際のニューロンを抽象化したニューラルネットワークモデルでは、ニューロンを丸で表し、軸索を矢印で表すことが行われる。実際のニューロン電位変化とニューラルネットワークとの対応関係を模式的に表現した図が下図である。活性化関数とは、例えば任意の時間範囲内のスパイク頻度を表したり、スパイク頻度の割合を表したりすると解釈できる。



Berger+2001, Fig. 3(b)

ニューラルネットワーク (2)

一つのニューロンを丸で描き，ニューロンの群を長方形で表現する。ニューロン群は層 layer と呼ばれる。

入力ニューロンの信号 x_i を変数とみなせば，一つのニューロンは，線形回帰，すなわち重回帰分析を行っていると考えることができる：

$$y = \sum_{i=1}^N w_i x_i + b \quad (\text{重回帰分析の定義式})$$

上式は，重回帰分析の定義式でもある。加えて，上式の出力に非線形変換を施すこと考える。例えば， $y \in [0, 1]$ なる変換を行うロジスティックシグモイド関数 $f(x) = [1 + \exp(-x)]^{-1}$ を行えば，ロジスティック回帰分析となる：

$$\begin{aligned} y &= \sigma \left(\sum_{i=1}^N w_i x_i + b \right) \quad (\text{ロジスティック回帰の定義式}) \\ \sigma(x) &= \frac{1}{1 + e^{-x}} \end{aligned}$$

最初期のニューラルネットワークモデル、たとえばパーセプトロンや ADALINE は、単層のニューラルネットワークモデルとみなすことができる。すなわち、1950 年代のニューラルネットワークモデルは、ロジスティック回帰と同じと言っても過言ではない。

ニューラルネットワーク (3)

1980 年代になると、上記の非線形変換を 2 回繰り返す、3 層のニューラルネットワークが提案された。これには、一般化デルタ則、現在では、**誤差逆伝播法 back-propagation** と呼ぶ学習アルゴリズム、すなわち、最適化計算手法が提案されたため、3 層のニューラルネットワークが可能となった。

誤差逆伝播法は、3 層のみならず、多層ニューラルネットワークにも適用可能である。

$$y = \prod_{\ell=1}^N \sigma \left(\sum_{i \in \mathcal{I}_\ell} w_{\ell,i} x_{\ell-1,i} + b_\ell \right)$$

1980 年代当時は、計算機の能力や記憶容量の制限などの理由により、ニューラルネットワーク研究は下火になる。

実際に、多層ニューラルネットワークでは、**勾配消失問題 gradient vanishing problems**、**勾配爆発問題 gradient exploding problems** や **信用割当問題 credit assignment problems** が指摘され、実用的な解を得ることが難しいとされてきた。

- 勾配消失問題とは、誤差逆伝播を多層に渡って繰り返すと、層を経るたびに、学習に必要な微分の値 (勾配) が 0 に近づく (消失) することを指す
- 勾配爆発問題とは、誤差逆伝播法を、再帰的な結合を持つ層内、層間での繰り返した場合、再帰結合層内、相関、での学習に必要な微分の値 (勾配) が、再帰的な処理のために指数関数

的に増大(爆発)することを指す。

- 信用割当問題とは、多層に渡る全結合層(fully connected layers)においては、すべてのニューロンが意思決定に関与することになるため、多層ニューラルネットワークの識別に、どの入力特徴が関与するのかが、不明、あるいは曖昧になることを指す。

ニューラルネットワーク (4)

上述のような問題を解決する努力が積年に渡って継続し、今日のような流行が齎された。これら努力の中には、以下のような技法が挙げられる:

- 置み込みニューラルネットワーク (CNN: Convolutional Neural Networks),
- 確率的勾配降下法 (SGD: Stochastic Gradient Decent methods),
- 整流線形化ユニット (ReLU: Rectified Linear Unit), LSTM (Long Short-Term Memory modeling) や Transfomer で採用されている、ゲート機構 (gating mechanisms), あるいは、注意機構 (attention mechanisms),
- 残差ネット (ResNet) や U-Net で用いられている スキップ結合 (skip-connections)
- 自然勾配法 (Natural Gradient methods) や Adam などの 最適化手法 (optimization techniques)

A.2 人工知能の歴史

	人工知能の置かれた状況	主な技術等	人工知能に関する出来事
1950年代			チューリングテストの提唱（1950年）
1960年代	第一次人工知能ブーム (探索と推論)	・探索、推論 ・自然言語処理 ・ニューラルネットワーク ・遺伝的アルゴリズム	ダートマス会議にて「人工知能」という言葉が登場（1956年） ニューラルネットワークのパーセプトロン開発（1958年） 人工対話システムELIZA開発（1964年）
1970年代	冬の時代	・エキスパートシステム	初のエキスパートシステムMYCIN開発（1972年） MYCINの知識表現と推論を一般化したEMYCIN開発（1979年）
1980年代	第二次人工知能ブーム (知識表現)	・知識ベース ・音声認識 ・データマイニング ・オントロジー ・統計的自然言語処理	第五世代コンピュータプロジェクト（1982～92年） 知識記述のサイクプロジェクト開始（1984年） 誤差逆伝播法の発表（1986年）
1990年代	冬の時代	・ディープラーニング	
2000年代	第三次人工知能ブーム (機械学習)		ディープラーニングの提唱（2006年）
2010年代			ディープラーニング技術を画像認識コンテストに適用（2012年）

出典: 総務省 第1部 特集 IoT・ビッグデータ・AI～ネットワークとデータが創造する新たな価値

- 2014 AlphaGo
- 2015 ImageNet 画像認識コンテストで、ResNet が人間の性能を超える
- 2018 Transformer 発表
- 2022 chatGPT 発表

A.3 機械学習の定義，古典的定義と現代的定義

- アーサー・サミュエル Arthur Samuel (1959): 機械学習とは，明示的にプログラムで指示せずにコンピュータに学習させる能力を研究する分野。
- トム・ミッチャエル Tom Mitchell (1999): ある課題 T とその成績 P の評価からなる経験 E をとおして学習するコンピュータプログラムを機械学習という。



左: Arthr L. Samuel(1901-1990) from <http://www.i-programmer.info/history/people/669-a-l-samuel-ai-and-games-pioneer.html>

右: Tom Mitchell from <http://wamc.org/post/dr-tom-mitchell-carnegie-mellon-university-language-learning-computer>

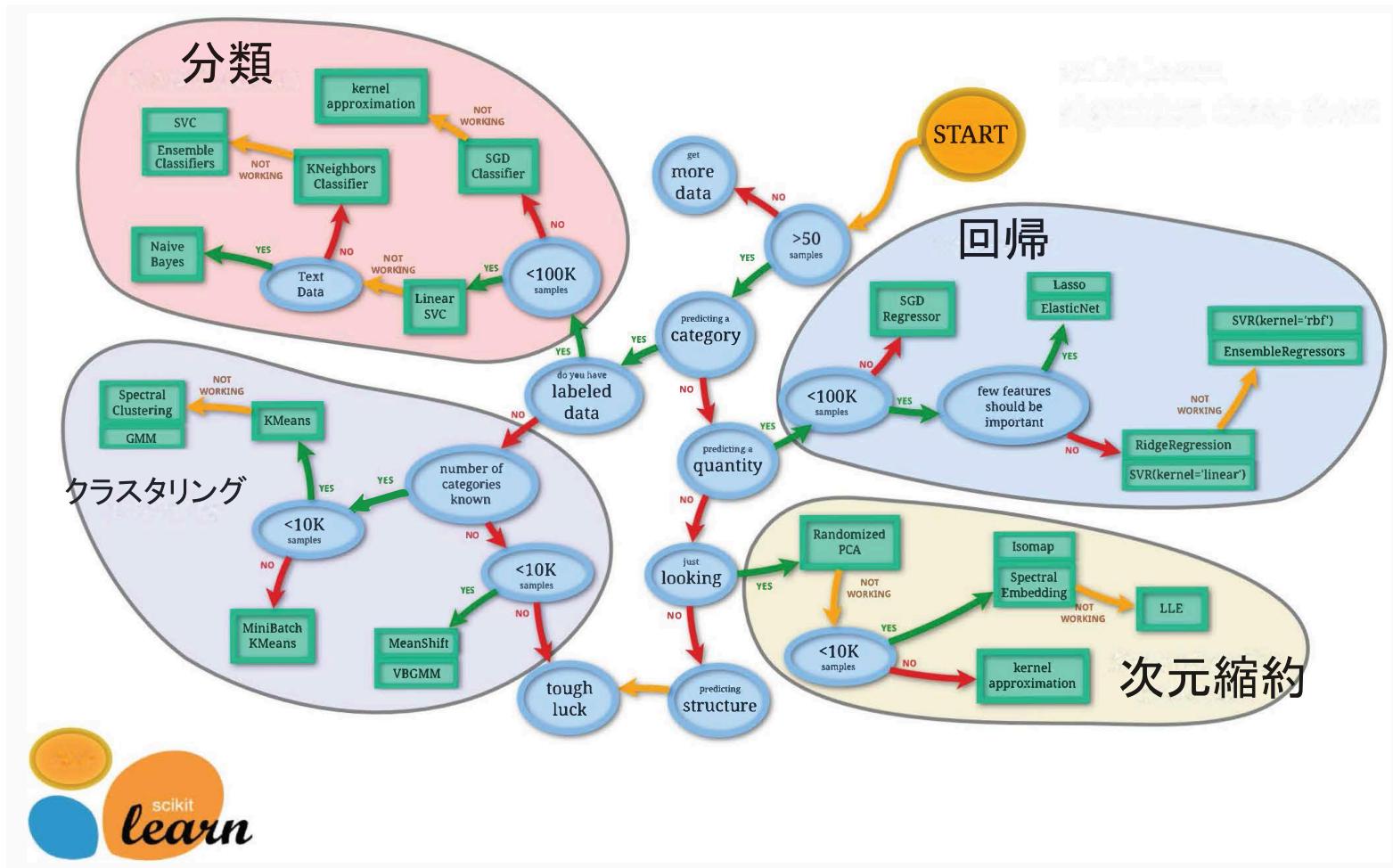
A.3 機械学習の分類

機械学習は、実用上 教師あり学習 **supervised learning** と 教師なし学習 **unsupervised learning** の 2 種類に大別される。

- 教師あり学習: データの特徴量とラベルの関係を何らかの方法でモデル化する。モデルが決まれば、未知のデータにラベルを適用することができる。教師あり学習は、分類 **classification** と 回帰 **regression** に分けられる。
 - 分類: ラベル、すなわち教師信号は離散的なカテゴリ。典型的には、真か偽かの 2 値。あたえられたデータが、基準に当てはまるか、当てはまらないか、という問題。2 値分類より分類するグループが多い分類問題を、多クラス分類 **multi-class classification** と呼ぶ。
 - 回帰: ラベルは連続的量。データとラベルが、それぞれ一つづつ、一対一対応であれば、 $x \mapsto y : y = f(x)$ 一般には、单回帰と呼ばれる。
- 教師なし学習 **unsupervised Learning**: ラベルを参照せずにデータセットの特徴をモデル化する手法を指す。データセットに語らせる と表現されることもある。教師なし学習のモデルには クラスタリング **clustering** や 次元削減 **dimension reduction** などが含まれる。クラスタリングアルゴリズムは、データの異なるグループを識別し、次元削減アルゴリズムでは、データのより単純な表現を探し出す。

さらに、教師付き学習と教師なし学習の中間に位置する、いわゆる **半教師付き学習 semi-supervised learning** と呼ばれる手法もある。半教師付き学習法は、不完全なラベルしか得られない場合に有効な場合がある。さらに、完全なラベルが与えられるデータであっても、半教師あり学習の手法を用いて訓練する場合もある。

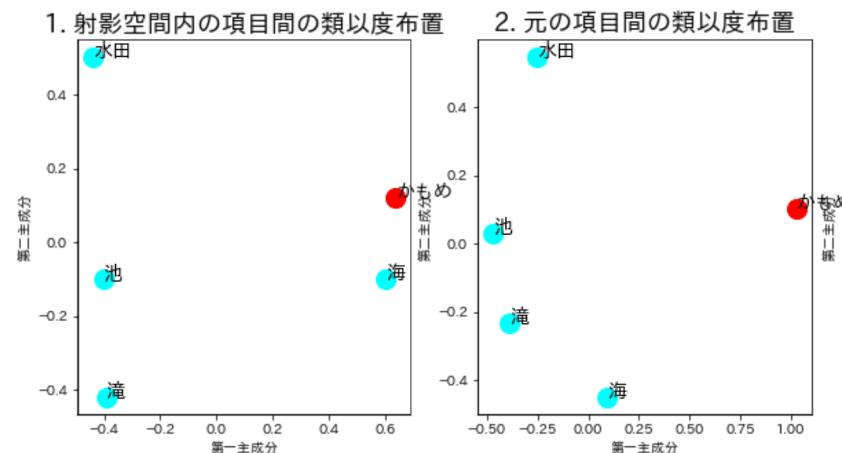
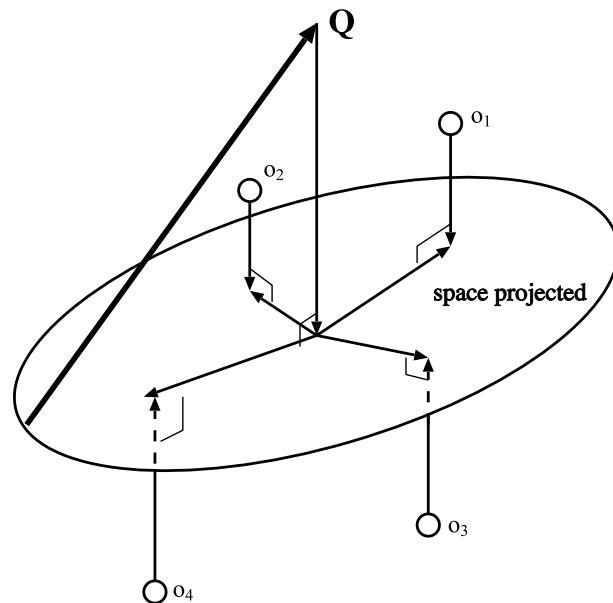
scikit-learn による機械学習の分類図



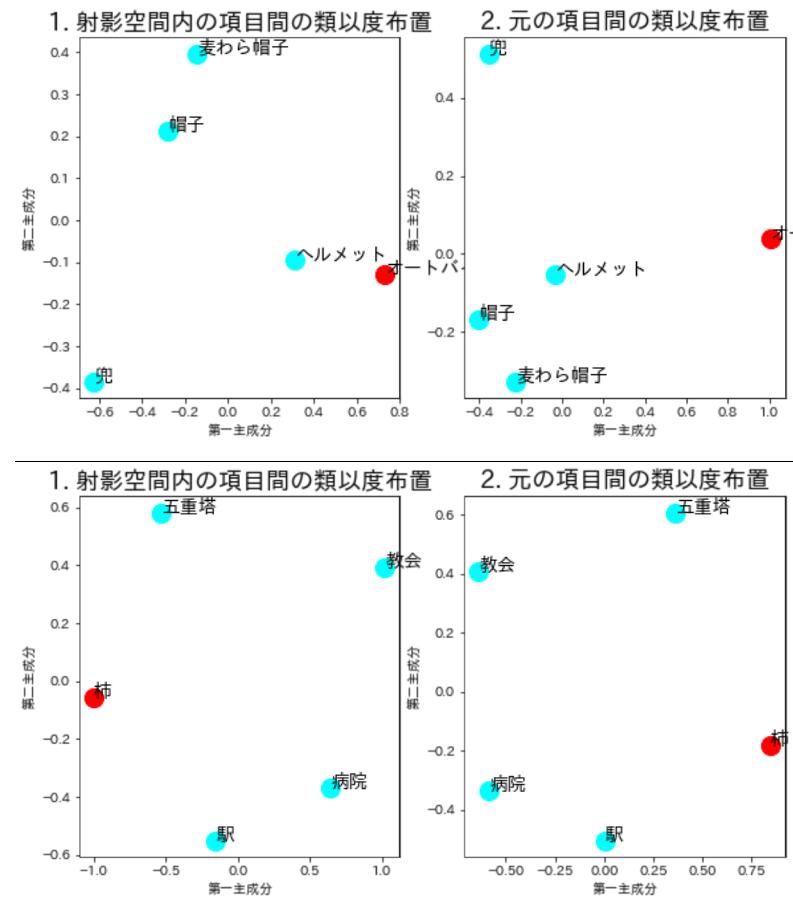
出典: http://scikit-learn.org/stable/tutorial/machine_learning_map/ を改変

A.4 Word2vec を用いた単語の意味空間

- ピラミッド・パームツリー・テスト: 認知症検査 (意味連合検査, 佐藤(2022) {:target=" '_blank' "})
- ターゲットと最も関連のあると考えられる選択肢を一つ選べ。
 1. ターゲット: オートバイ, 選択肢: 麦わら帽子, 帽子, ヘルメット, 兜
 2. ターゲット: かもめ, 選択肢: 水田, 池, 瀑, 海
 3. ターゲット: 柿, 選択肢: 五重塔, 教会, 病院, 駅

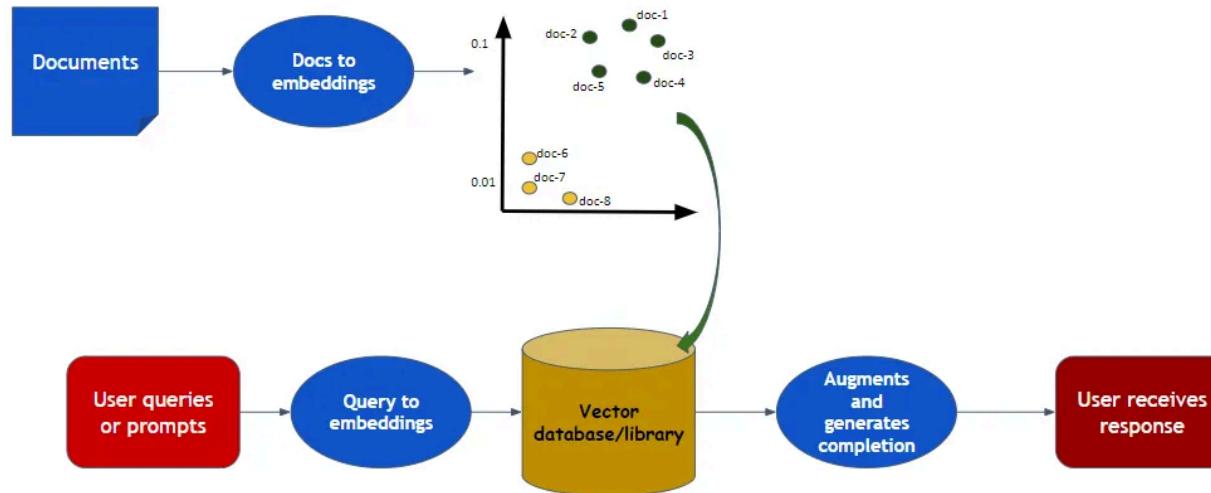


A.4 Word2vec を用いた単語の意味空間(2)



近藤・浅川(2020) より

A.5 RAG (Retrieval Augmented Generation)



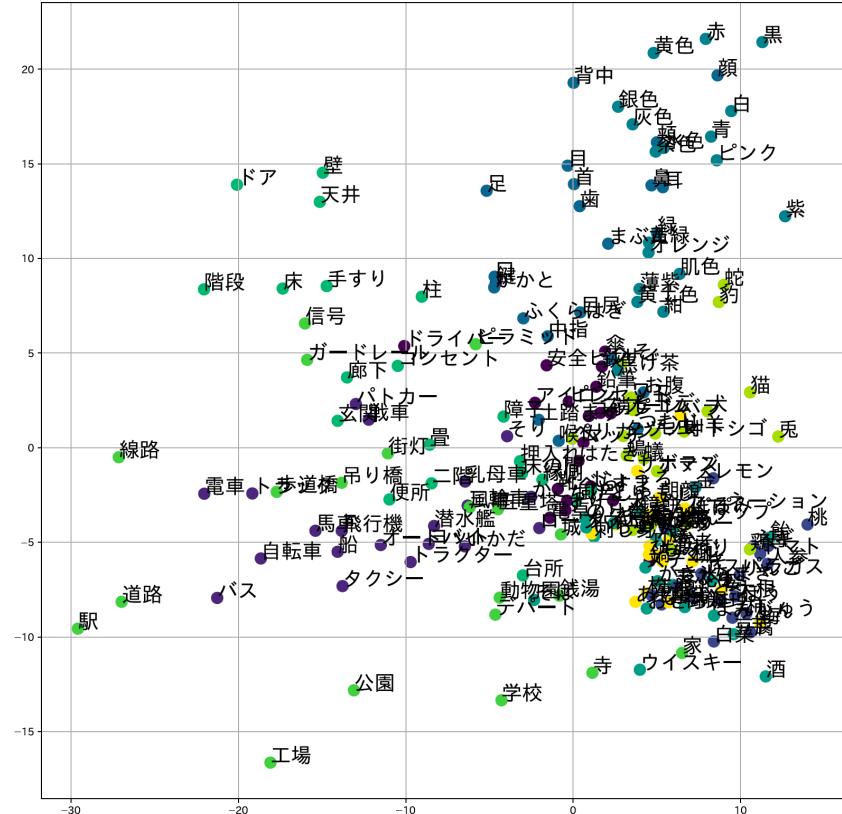
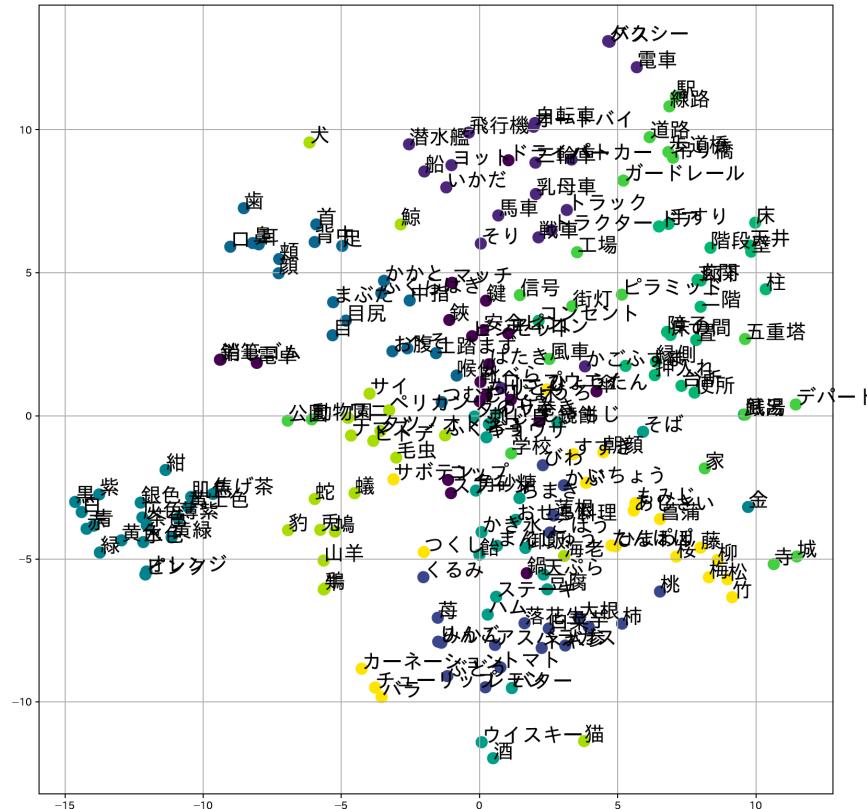
From [Implementing RAG with Langchain and Hugging Face](#)

文献

- Mikolov+2013, word2vec オリジナル論文
- 浅川+2018, *Analogy comprehension between psychological experiments and word embedding models*, Asakawa+2018
- 近藤・浅川 (2017) 日本語 Wikipedia の word2vec 表現と語彙特性の関係

A.5 tSNE を用いた概念の可視化

TLPA 失語症検査の呼称課題の布置を tSNE を用いて可視化。



左: tSNE, 右: 主成分分析 結果。

- tSNE を用いた TLPA 200語の word2vec 視覚化 [Open in Colab](#)
- 効率よく t-SNE を使う方法

