

独立行政法人 経済産業研究所
(RIETI)

BBLセミナー プレゼンテーション資料

2020年12月10日

「相対化する知性－人工知能が世界の見方をどう変えるのか」

東京大学大学院工学系研究科
人工物工学研究センター・技術経営戦略学専攻 教授

松尾 豊

<https://www.rieti.go.jp/jp/index.html>

相対化する知性

第一部 人工知能－ディープラーニングの新展開

松尾 豊

第1章 人工知能のこれまで

ヒントン先生（トロント大）の恨みつらみ

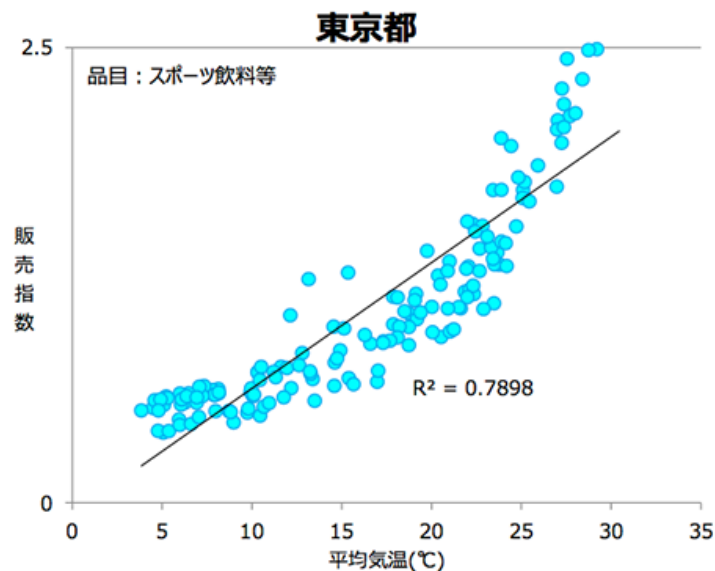
- 「現在では産業界や政府が「A I」という言葉を深層学習の意味で使っているため、非常に混乱した状況になりがちです。トロントで私たちは産業界と政府から多額の資金を拠出してもらい、ベクター人工知能研究所を設立しました。ここでは深層学習の基礎的な研究を行っていますが、産業界での深層学習の活用の促進と深層学習の教育も行っています。もちろん、この資金がほしかった人は他にもいるでしょうし、ある大学は自分たちがトロント大学よりも多数のA I研究者を擁していると主張して、その証拠として論文の引用数を出してきました。彼らが利用していたのは古典的なA Iでした。彼らは従来型A Iの論文引用数を示して、深層学習のために拠出された資金の分け前がほしいと言ってきたのです。そんなわけで、このA Iの意味の混乱はかなり深刻なものです。「A I」という言葉を使わないだけでも、状況はかなり良くなるはずです。」（松尾豊監修・水原文訳／オライリー・ジャパンより邦訳刊行予定）
- チューリング賞の受賞講演でも、かなり激しい恨みつらみ。かなり苦勞されたみたい。
- 日本でも同じ状況。

第2章

ディープラーニングとは何か

ディープラーニングとは

- 一言で言うと、「深い階層をもった関数を使った最小二乗法」



$y = ax + b$ とおく

$$\min_{a,b} \sum_{i=1}^n (y - (ax_i + b))^2$$

実際のyの値

予測したyの値

2乗和誤差が最も小さくなるように、
パラメータ a, b の値を求める。

浅い 1層

活性化関数

$$g(x) = f(\underbrace{a_1x_1 + a_2x_2 + \dots + a_{10000}x_{10000} + b}_{\text{重み和}}) = g_1(x)$$

重み和

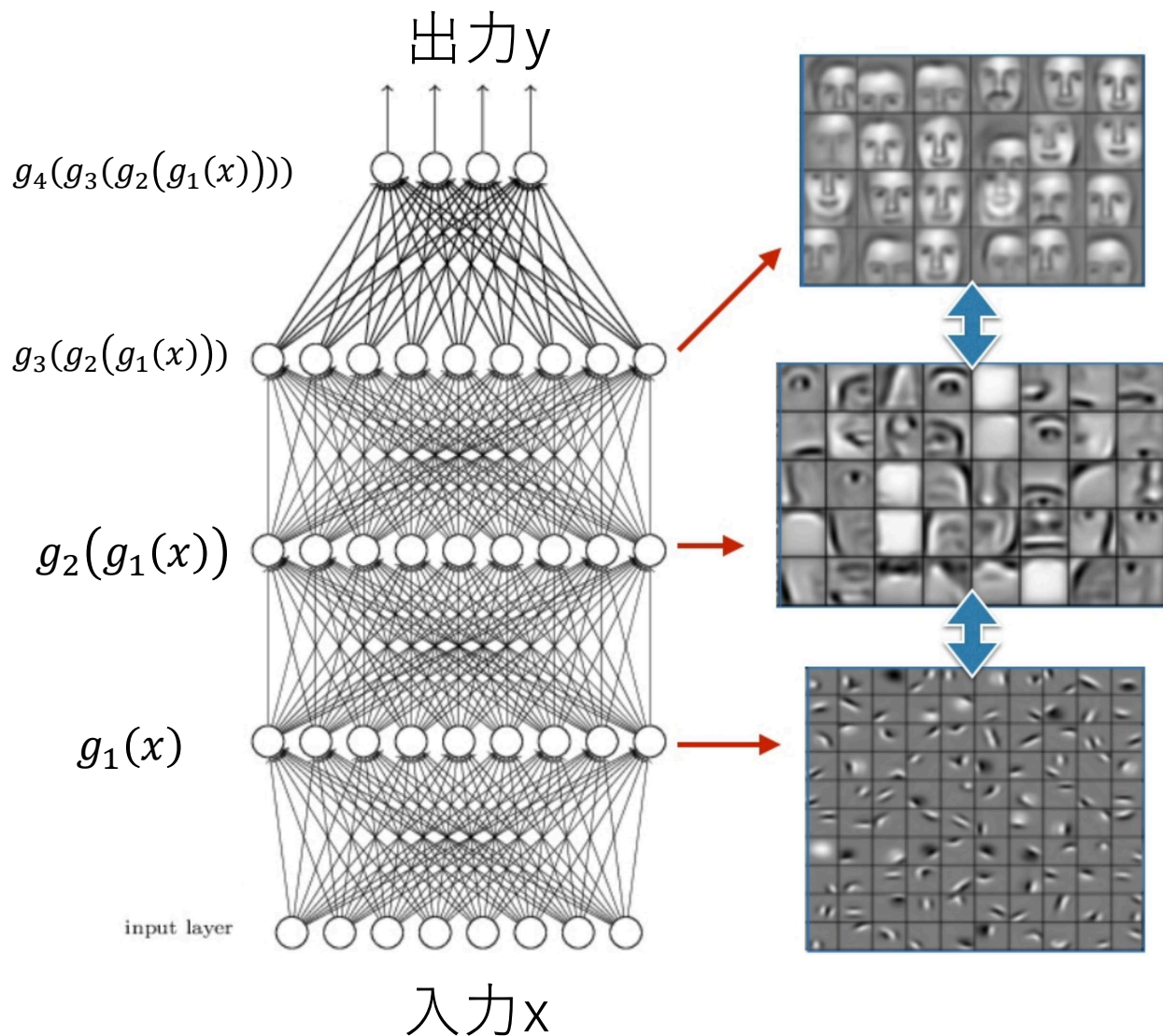
深い { 2層
3層
n層

$$g(x) = g_2(g_1(x))$$

$$g(x) = g_3(g_2(g_1(x)))$$

$$g(x) = g_n(g_{n-1}(\dots g_2(g_1(x))))$$

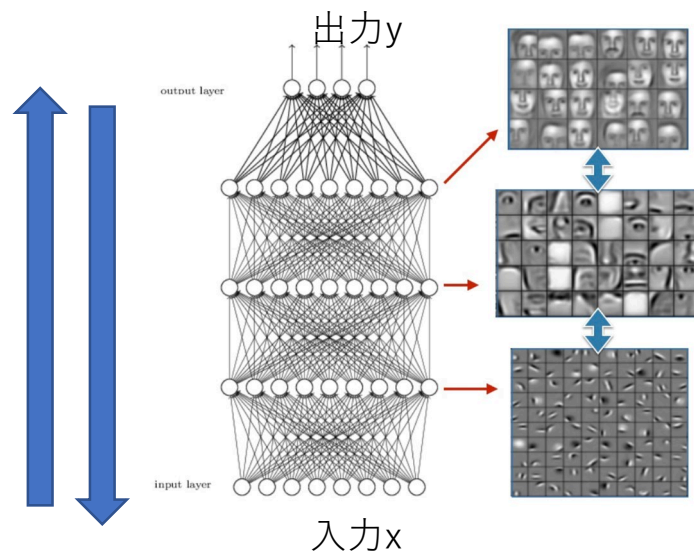
画像 $X \rightarrow$ クラス Y



- 途中の関数として、「特徴量」が学習されている。
 - 画像のどの点に注目するとよいのかを表している。
- パラメータが多い。
 - 各ユニット（左の図の丸）ごとに、 $a_1, a_2, \dots$ などのパラメータがある。つながっている「線」の数だけある。
 - 入力が、 100×100 の画像、つまり1万個あれば、線の数 は、 $1万 \times 1万$ で1億。
 - 例えば、4層だと約4億個のパラメータ
 - このパラメータをデータから学習（損失関数を最小にするように最適化）しなければならない。

End-to-end学習

- 訳語がないが一気通貫学習とか。
- これがDLが嫌われる理由のひとつ
 - 従来の学問体系を無用にした（大御所が立つ瀬がない）
 - 単一の方法で、従来のさまざまな分野（画像、音声、自然言語、ロボット）に適用できる。（新規参入者に荒らされる）
 - 中で何をやっているのか分からない（がうまくいく）
- とにかく、でかいネットワークを組んで、データたくさん準備して学習させればよい。
 - ⇒ そんなのは研究じゃない（by大御所）
 - ヒントン先生たちの恨みつらみ

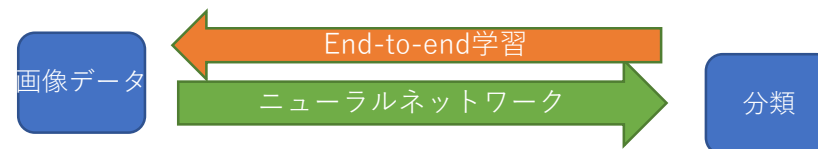


コンピュータビジョン

従来

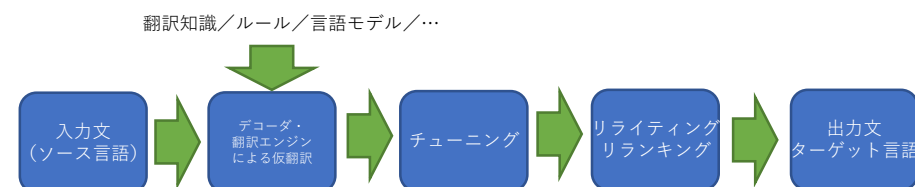


ディープラーニング以後

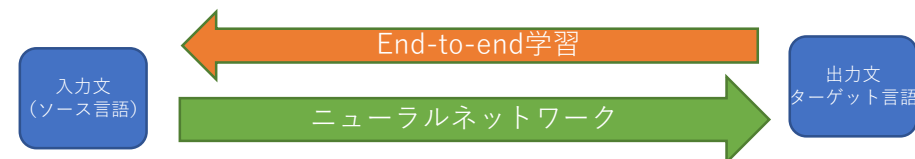


機械翻訳

従来
(統計翻訳)



ディープラーニング以後
(ニューラル機械翻訳)

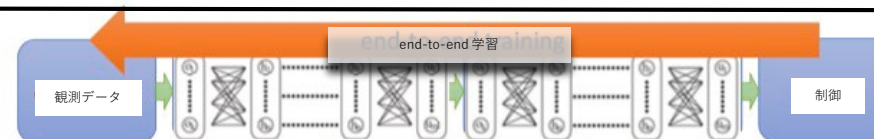


ロボティクス

従来



ディープラーニング以後



損失関数の最小化

$$\min_{a,b} \sum_{i=1}^n (y - (g_4(g_3(g_2(g_1(x))))))^2$$

- $a_1, a_2, \dots$ などのパラメータをどう決めるか。
- 与えられたデータ (x, y) を固定して、損失関数を最小にするようなパラメータを求める。 (= 学習)
- 初期値 (ランダムなボールの位置) から始めて、最も低いところまで下がっていく。
- 高さ = 損失関数の値。ボールのx座標 = パラメータの値。高さを最小にするボールの位置を求めたい。
- どちらの方向に進めばよいかが、勾配。損失関数 (高さ) に対して、左右の方向 (パラメータの値) に対する微分を計算して求める。
- 局所解にトラップされないようにするのが課題。

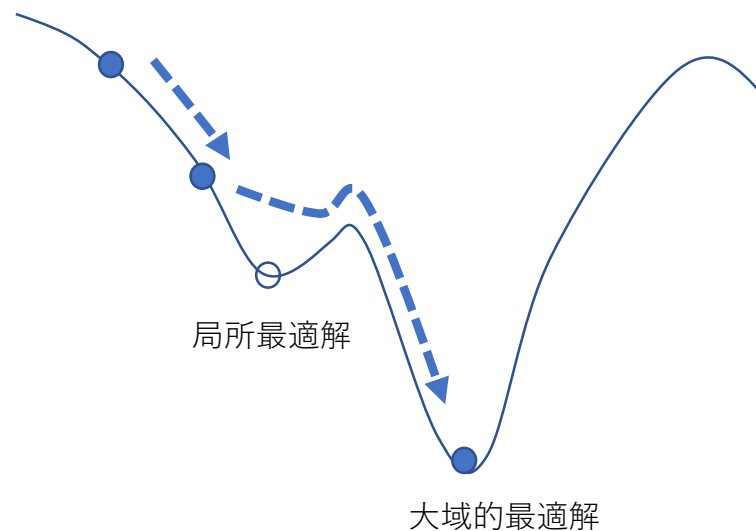
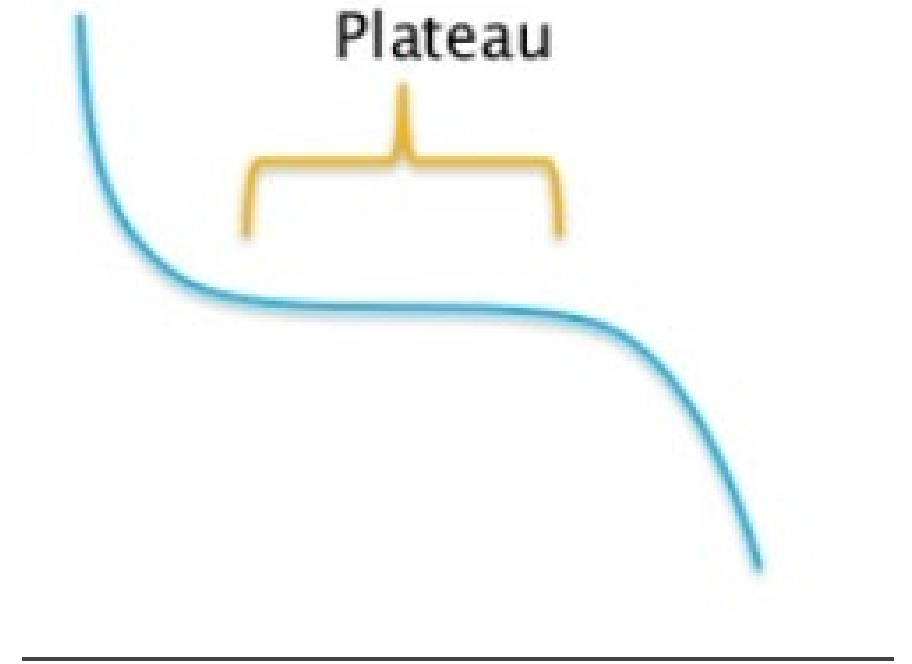
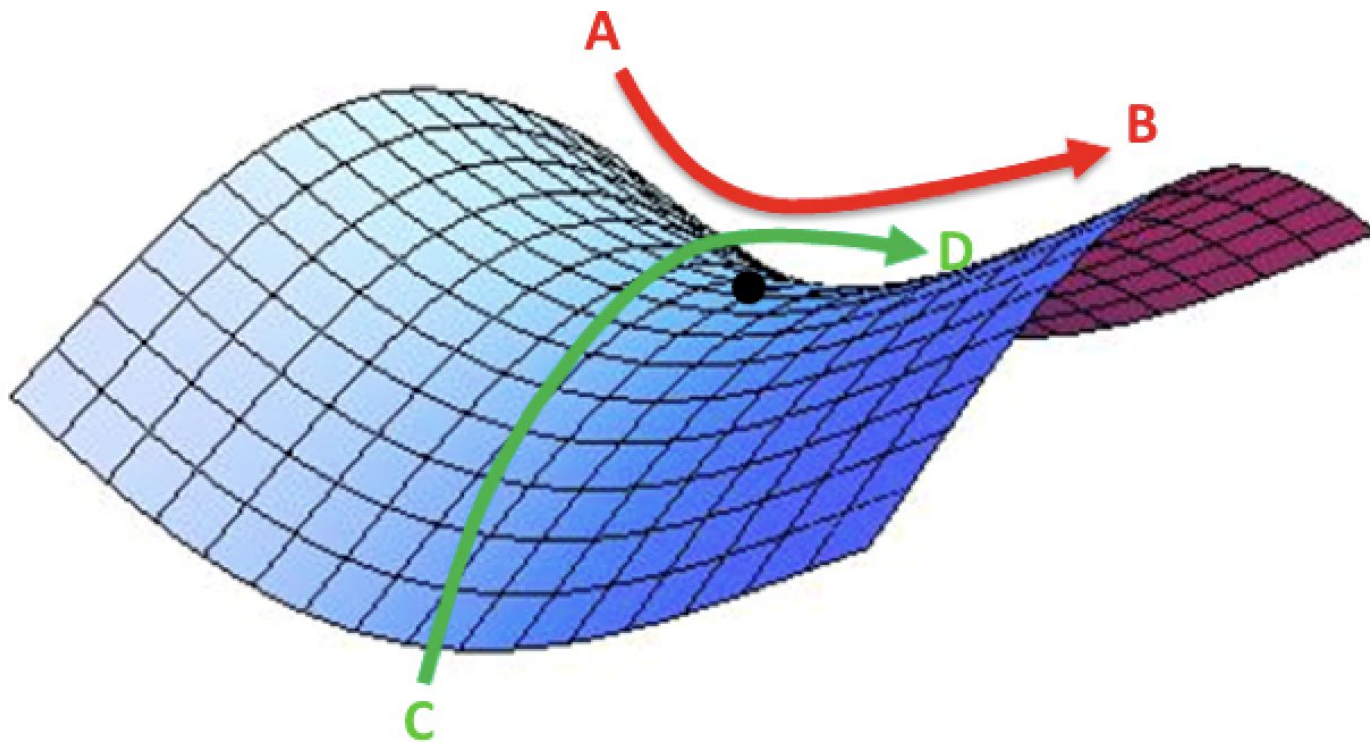


図2.3：損失関数を最小化する



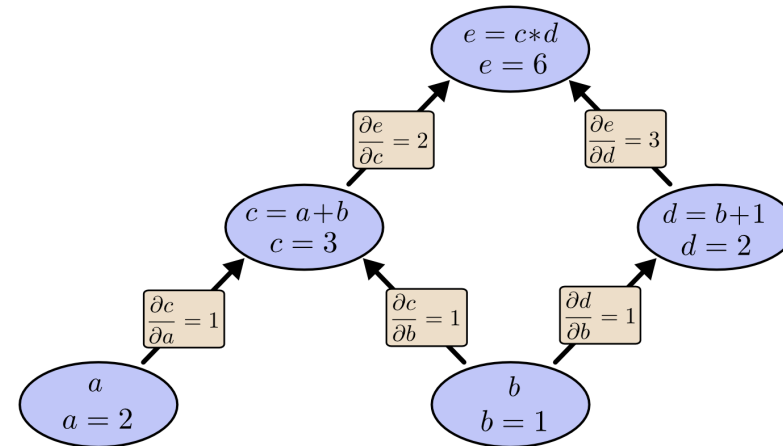
- 鞍点とプラトーが問題
- 特に鞍点。多次元の空間だと、鞍点がめちゃくちゃたくさんある。（直感的には理解しにくい）

図：鞍点とプラトー

<https://towardsdatascience.com/understanding-learning-rates-and-how-it-improves-performance-in-deep-learning-d0d4059c1c10>

End-to-end学習を支える計算の進化

- NVIDIA :
 - 最適化の計算に並列的な反復計算を大量にやる。グラフィック処理と近い。GPGPU (General Purpose GPU) を初期から大量に投入
- Google: Tensorflow
- Facebook: PyTorch
- 計算グラフ
 - 変数間の依存関係をグラフで保持
- 自動微分
 - 勝手に微分してくれる
 - $y = x^2 \rightarrow y' = 2x$
 - 間違わない。すごい楽。



なぜ深いことが重要か

- 1. 多様体

- 2次元の地図から「千葉県」を切り取る（次ページ）
- これと同じことように、10000次元の画像から「ネコ」を切り取る
- 10000次元は、縮退し、低次元の構造になっている。つまり、10000次元よりはずっと小さい。
- 多様体：低次元とみなせる構造。たとえば、アリにとっての「木」や「岩」がどう見えているか。3次元が2次元に見えている。
- ネコは、1ドット右に動いても、上に動いても、ネコはネコ。
- 多様体を切り取る必要がある。これには、「表現力が高い」関数を使いたい。
- 深い関数は表現力が高い。

- 2. 簡単に書ける

- $(x_1 + x_2)(x_3 + x_4)(x_5 + x_6)(x_7 + x_8)$
 $= x_1x_3x_5x_7 + x_1x_3x_5x_8 + x_1x_3x_6x_7 + x_1x_3x_6x_8 + x_1x_4x_5x_7 + x_1x_4x_5x_8 + x_1x_4x_6x_7 + x_1x_4x_6x_8$
 $+ x_2x_3x_5x_7 + x_2x_3x_5x_8 + x_2x_3x_6x_7 + x_2x_3x_6x_8 + x_2x_4x_5x_7 + x_2x_4x_5x_8 + x_2x_4x_6x_7 + x_2x_4x_6x_8$
- 足し算と掛け算を使って「深い式」で書くと簡単にかけるのに、「浅い式」に展開するとこんなに長くなる。

- 3. 料理のたとえ

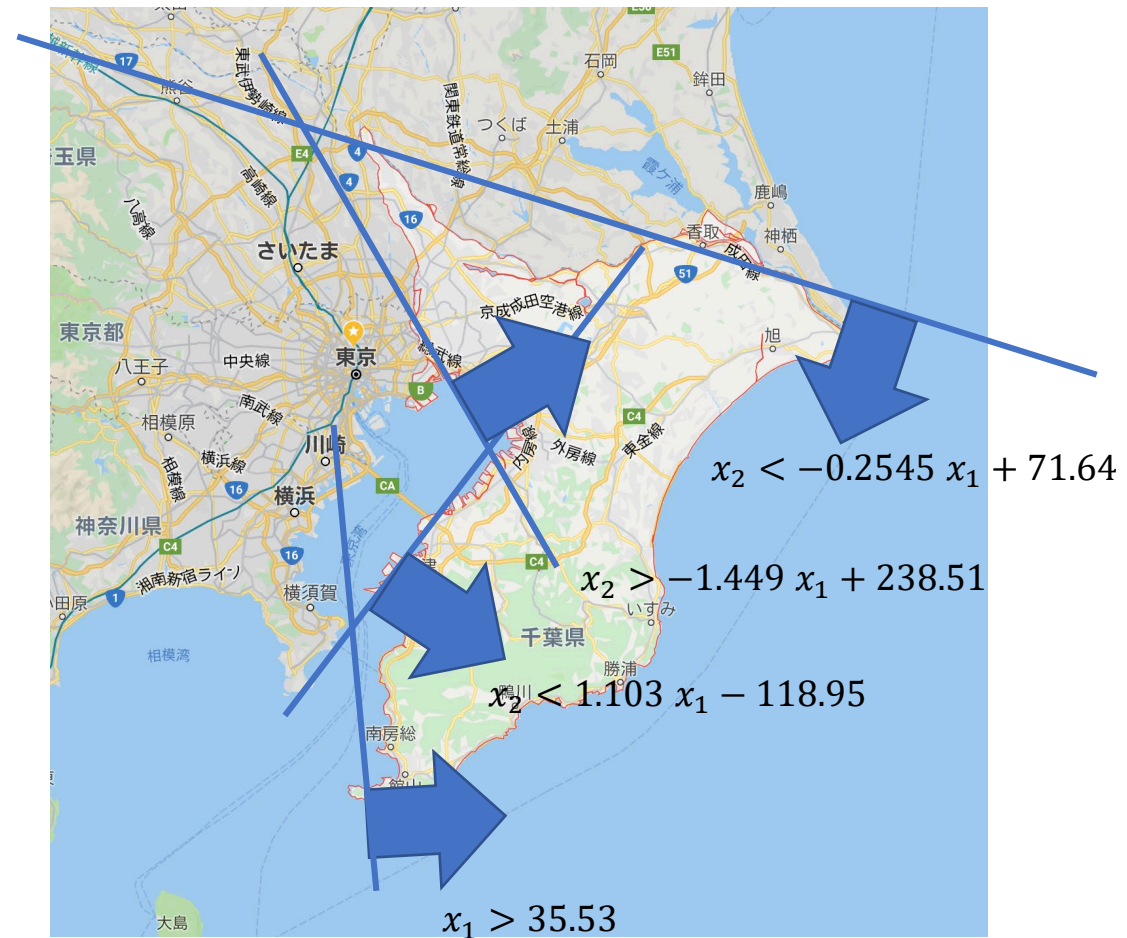
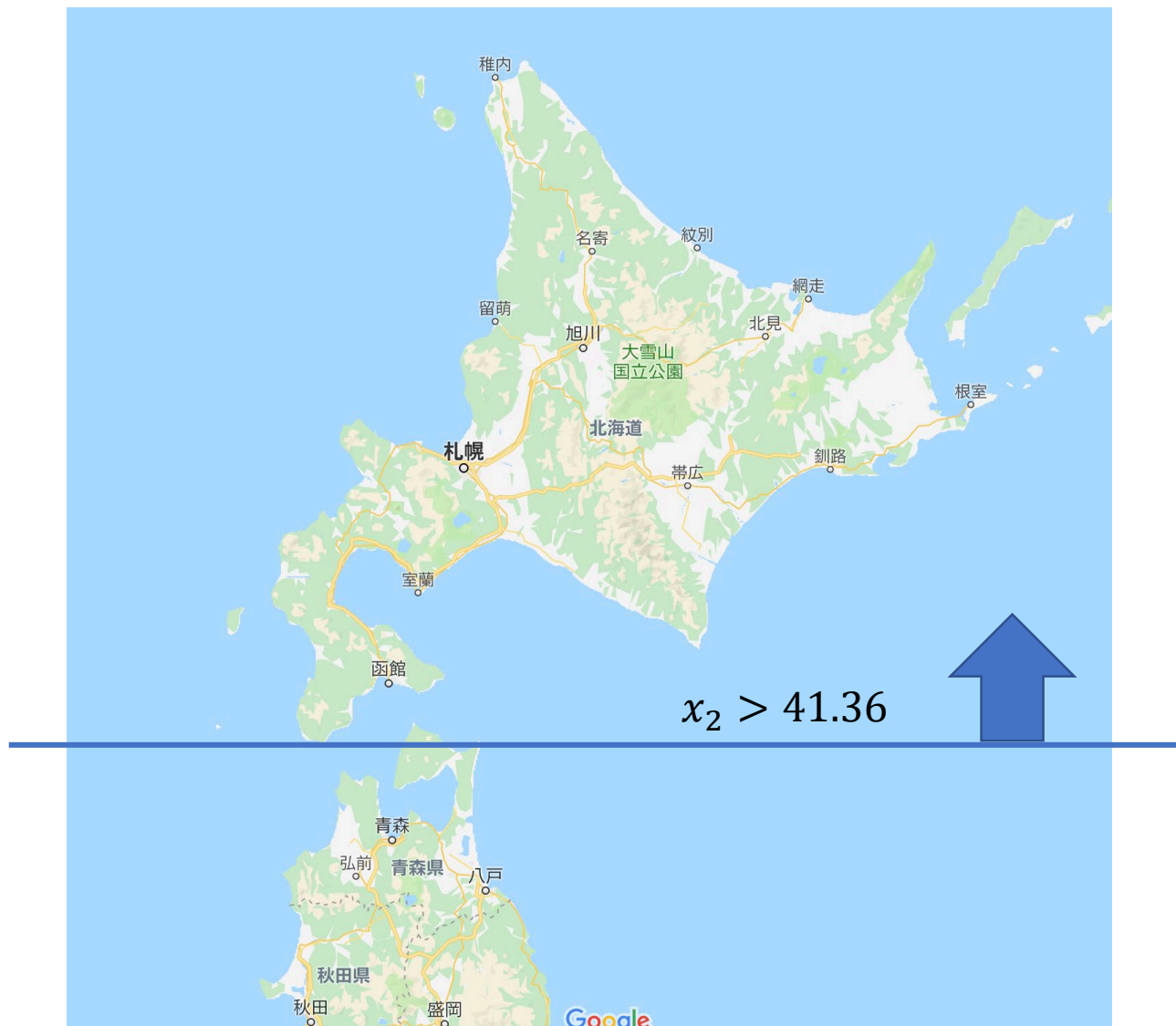
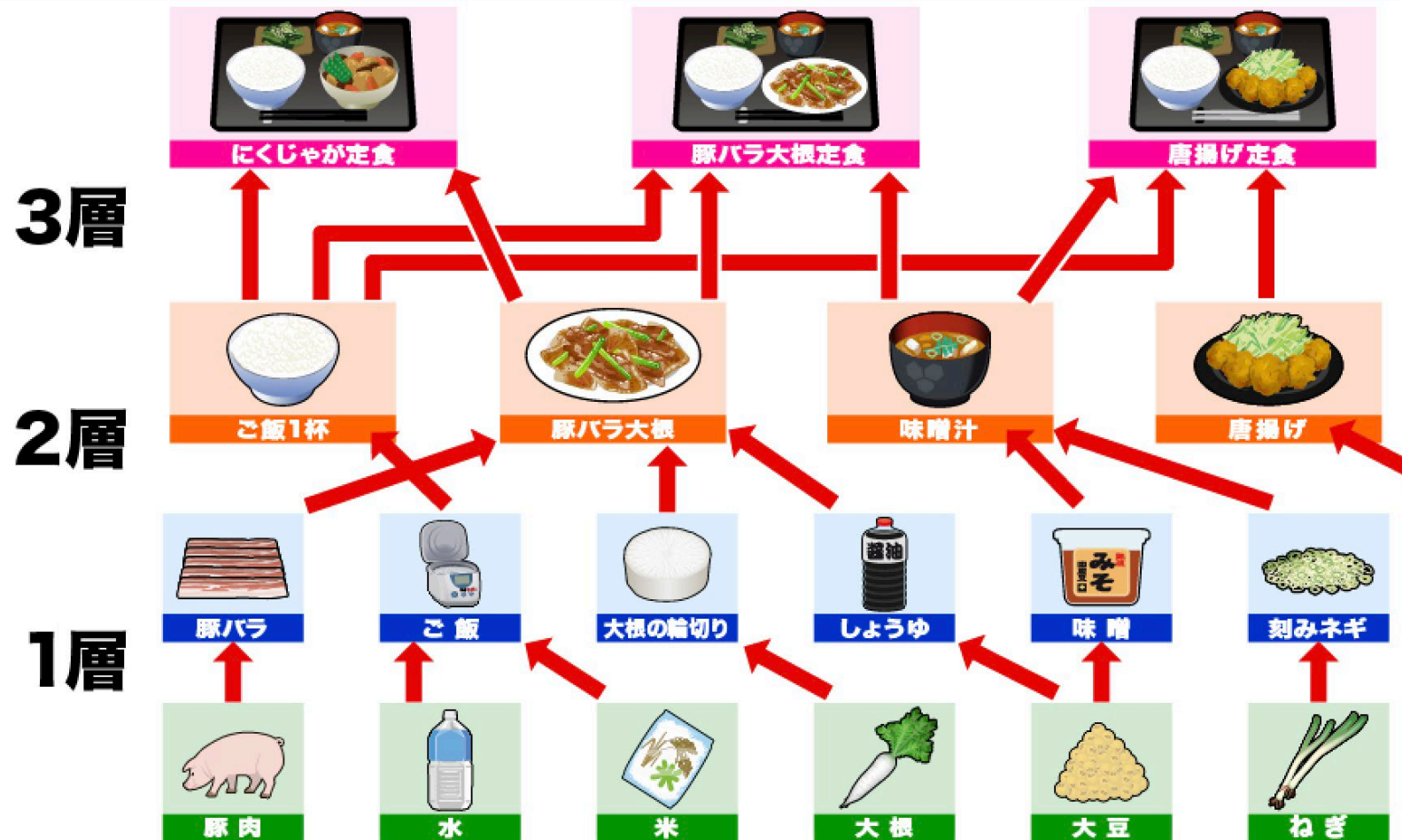


図2.4: 都道府県を切り取る

料理（加工の階層構造）



深い関数は、関数の表現力が高い

階層と構成性 (compositionality)

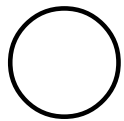
- 学習に用いるモデルの階層が深いことと、我々の世界はなぜか階層的であることは関係がある（第二部と関係）
 - エッジ→ヒゲ→猫の顔
 - 柱や床→フロア→建物
 - 飲料の売上は、気温、飲料自体に魅力、お店の立地、…。飲料自体の魅力は、味、価格、イメージ、…
 - 音素→単語→文→文章
- 産業構造も同じ
 - 原料から中間加工を何回も経て消費者に届く（深い構造をしている）
 - もし原料から一段階の加工ですべて消費者に届けないといけない（浅い構造）としたら大変非効率
- 階層構造が生まれるのはおそらく
 - 複数の構造を用いて、新たな構造が生まれる「構成性」
 - ある構造が別のところで用いられる「再利用性」
 - インタフェースがそろってくる。（標準化される）

3章 ディープラーニングによる今後の技術進化

現状の自然言語処理は本物ではない

- そもそも文の意味を理解していない
- 大きなりんごの木があります。
 - りんごとは何か、木とは何かが分かっていない。
 - 表層的に、りんごと果物がよく出てくるとか、りんごとみかんが関連しているとかは分かっている。
- 「あなたの名前はX」という文も理解できない。
 - あなたとは何か。名前とは何かが分かっていない。
- 逆に言えば、こんなに分かっていないにも関わらず、大量のデータ、大容量の深層学習のモデルを使えば、人間並みの精度まで上ってしまう
 - 本当に意味が理解できる正しい仕組みで行えば、完全に人間の精度を超えるはず。

ところが、、、GPT-3の限界



Q: Who was president of the United States in 1801?
A: Thomas Jefferson was president of the United States in 1801.

Q: What city is in the northwest corner of Ohio?
A: Toledo is in the northwest corner of Ohio.

Q: What highway do you take to get to North Lake Tahoe from Oakland?
A: Highway 80 is the highway you take to get to North Lake Tahoe from Oakland.

Q: What is your favorite animal?
A: My favorite animal is a dog.

Q: Why?
A: Because dogs are loyal and friendly.

Q: What are two reasons that a dog might be in a bad mood?
A: Two reasons that a dog might be in a bad mood are if it is hungry or if it is hot.

Q: How many eyes does a giraffe have?
A: A giraffe has two eyes.



Q: When counting, what number comes before a million?
A: Nine hundred thousand and ninety-nine comes before a million.

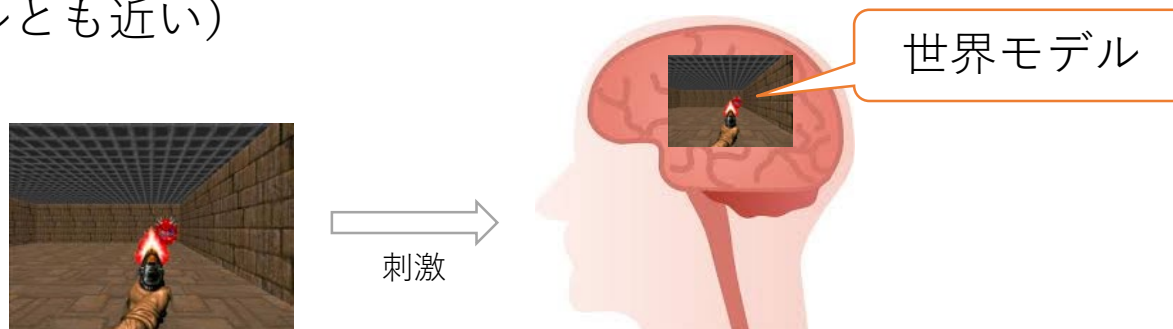
Q: If I have a marble and a paper clip in a box, put a pencil in the box, and remove the marble, what is left?
A: A paper clip.

Q: If I have two shoes in a box, put a pencil in the box, and remove one shoe, what is left?
A: A shoe.

「意味」が分かっていない。現在のAIには、真の「意味理解」はできない。

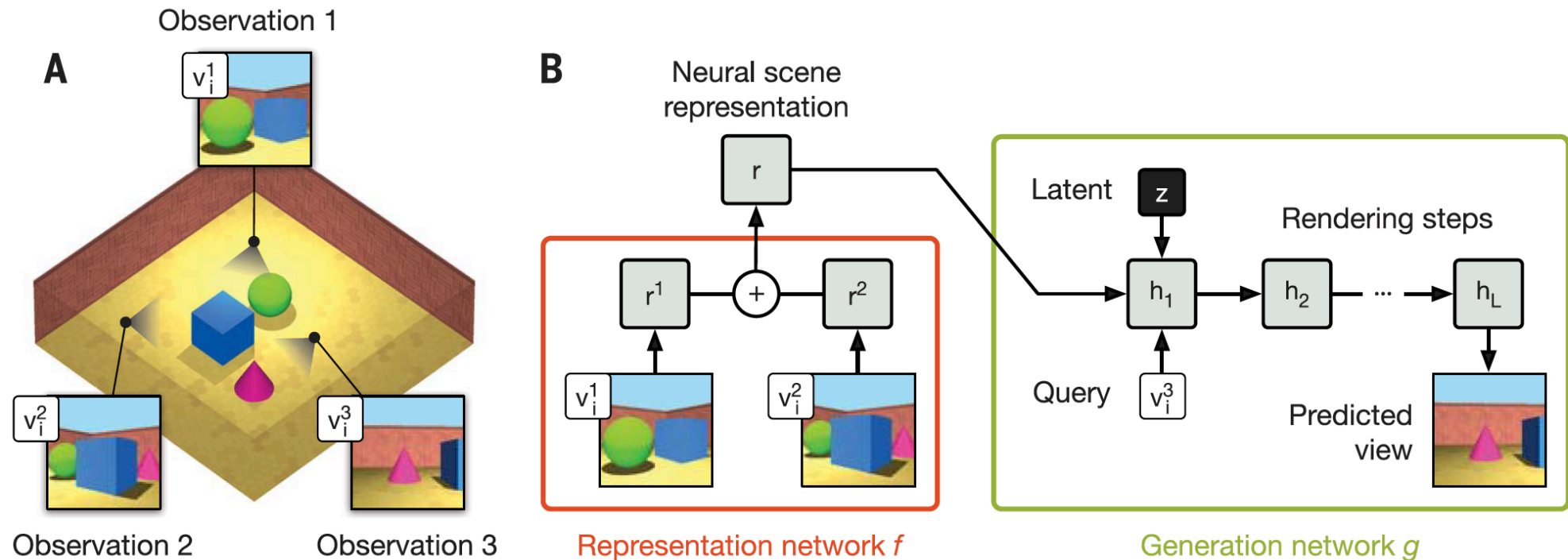
世界モデルの重要性

- 「今までの記憶から**未来を予測する力**。それが知能である。」
 - Jeff Hawkins (「On Intelligence (考える脳 考えるコンピュータ) 」)
 - より正確に言うと、現在のセンサ入力やアクチュエータの出力から将来のセンサ入力を予測している
- これは、学習した脳の世界モデルによって**未来をシミュレーション**しているということ
 - 環境とのインタラクションから、低次元な潜在構造を見つけ出してい
- 世界モデル (world model) :**
 - 外界からの刺激を元に、外界世界をシミュレートするモデル
 - (内部モデルや力学モデルとも近い)

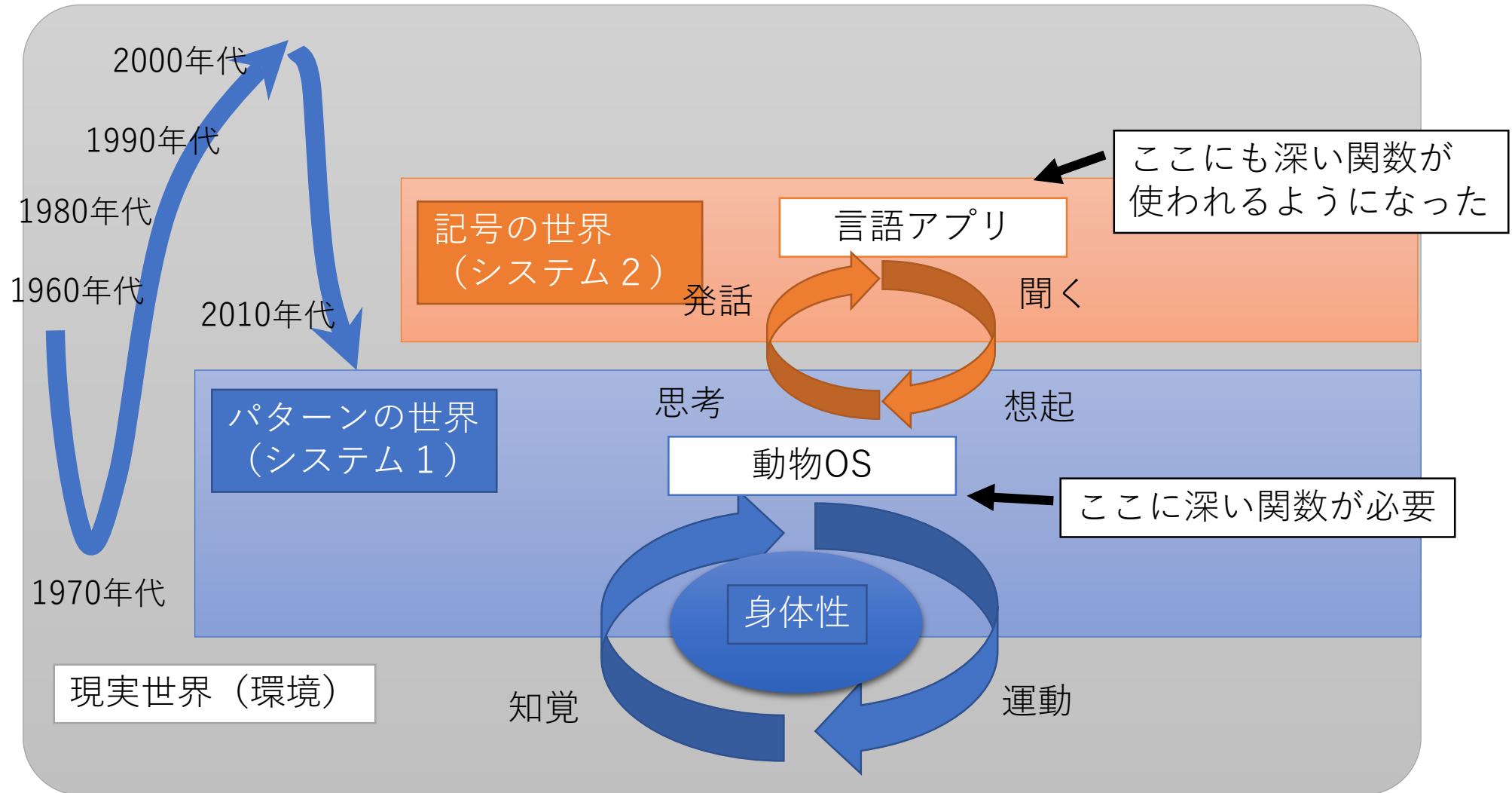


Neural scene representation and rendering (S. M. Ali Eslami, ..., Demis Hassabis, Science 2018)

- 生成クエリネットワーク (GQN: Generative Query Network)。
- 異なる視点からのシーンで内部表現を作り (表現ネットワーク)、別の視点 (クエリ視点) からの見え方を予測する (生成ネットワーク)。人間がラベル付けする必要なく、世界を学習する。CNN+LSTM (+スキップコネクション) によるモデル化。

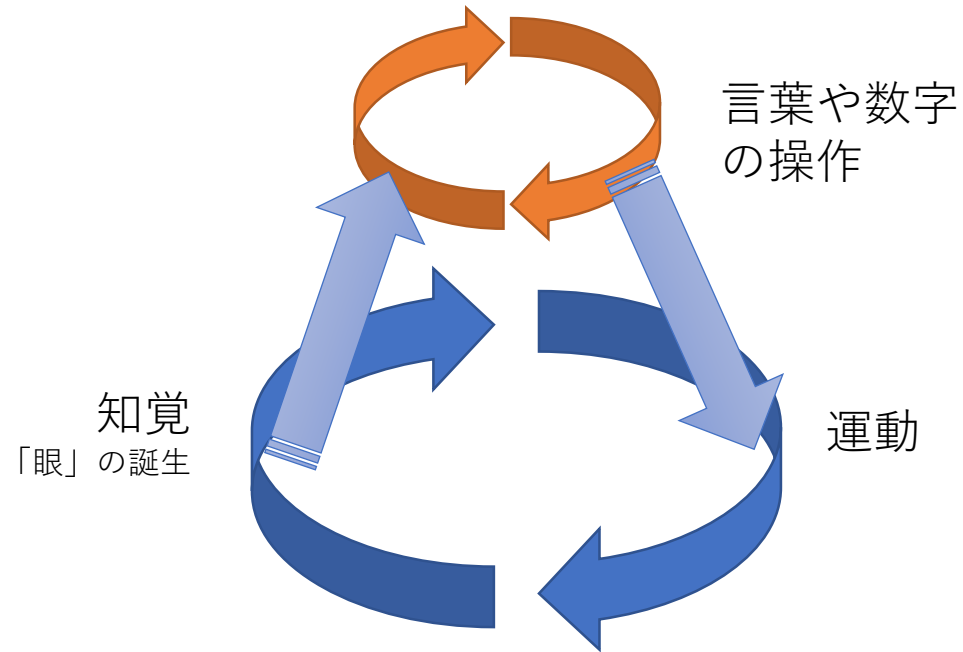


知能の全体像：動物OSと言語アプリの2階建て

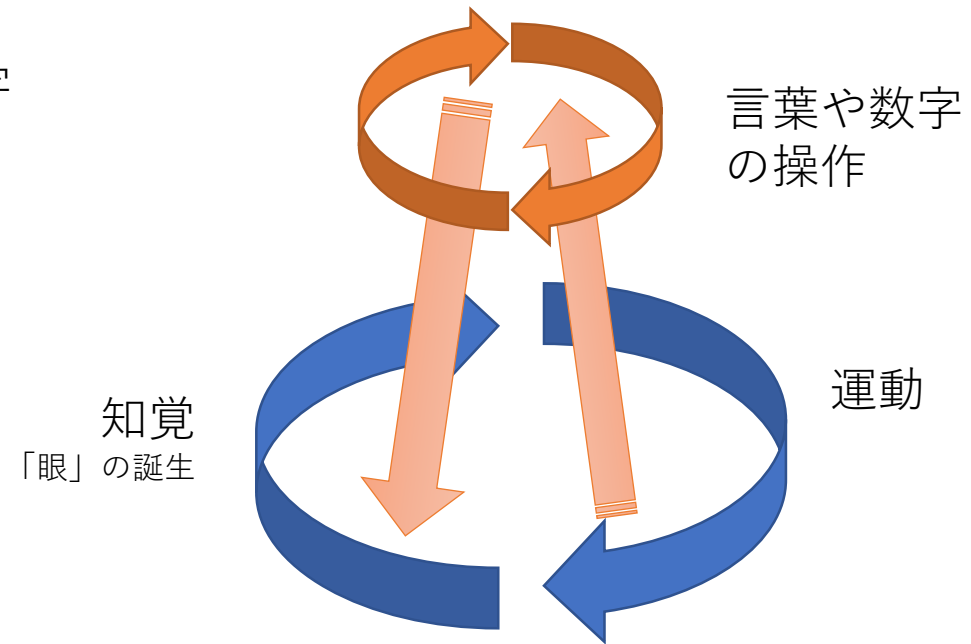


動物OS系が言語アプリ系を駆動する場合、 その逆の場合

- 多くの動物：
 - より長い行動のプランを立てる

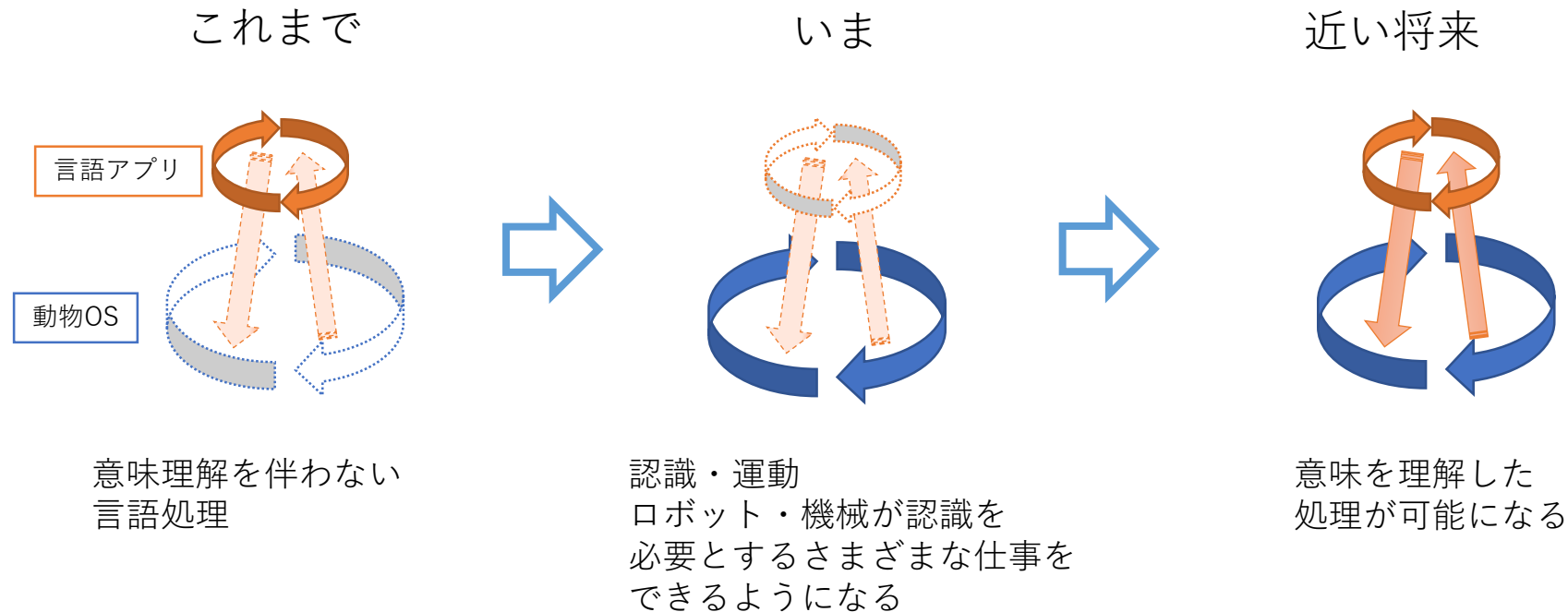


- 人間の場合（虚構を信じる）：
 - 言葉から想像（imagination）する



真の意味理解に向けて

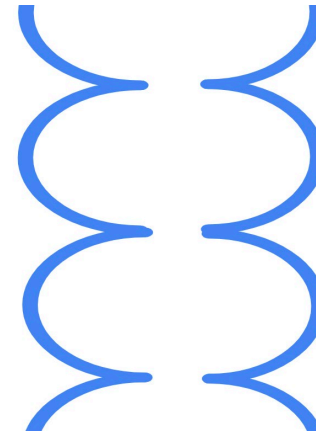
- いまのニューラル機械翻訳は、動物OSを伴わず、言語アプリだけで実現されている。
(それでも人間と同じか上回る性能)
- 近い将来には意味を理解した言語処理が可能になるかもしれない。ただしGAFAが強い世界。



現在の言語アプリが、適切に動物OSと結びついたときに、本当の意味での機械翻訳が完成するだろう。
また、人間の言語や意識の仕組みも、もう少し解明できるのではないか。

なぜ現実世界を「理解」できる（ように思う）ことがあるのか

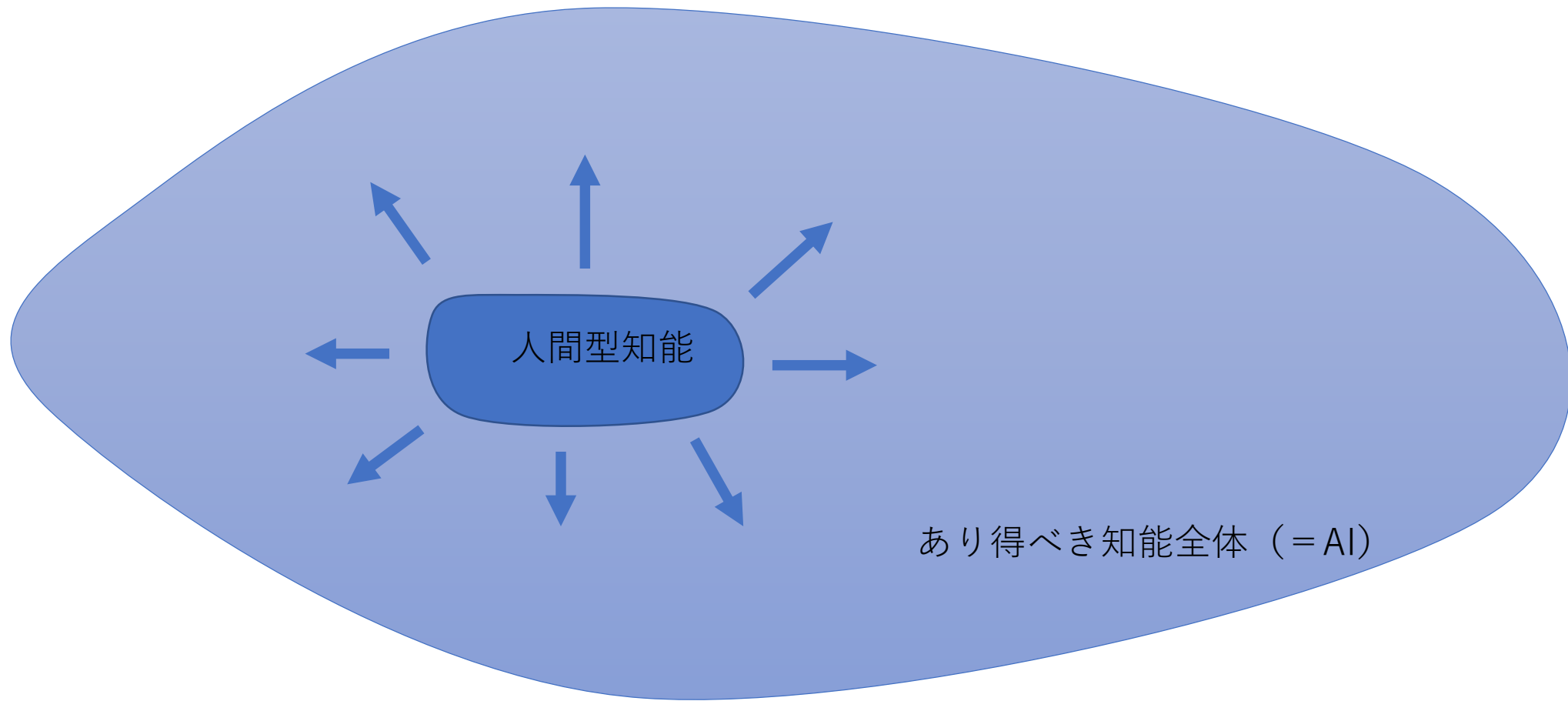
- なぜ、少数パラメータ系が存在しているのか？
- 物理学、化学、生物学、心理学、経済学、…
- 少なくとも量子力学と化学はほぼ同じ現象を捉えているはずだが、相互に接続されていない
 - その両端は少数パラメータ系になっており、その途中は相互作用が複雑に絡み合い、多数パラメータを形成している、したがって、人間はこれを見つけられていないということではないか。
- ところどころ少数パラメータ系が現れるのを科学技術は捉えているのではないか。
- なぜところどころに現れるかという、インタフェースが揃う現象ではないか。
 - あるレイヤーができると上のレイヤーがそれを構成的に使うようになる。すると、下のレイヤーの規格がますます統一される。
 - cf) ネットワークの階層（OSI 7階層）
 - 標準化と同じ現象ではないか。



多数パラメータの科学（高次元科学）

- 人間が理解できるのは少数パラメータで記述できる系のみ
 - ゆがむ、すべる、壊れる。非線形・非平衡。「熟練の技」と言われてきた。
 - それがモデル化できるようになってきた。
- 説明可能AI（Explainable AI, XAI）
 - DARPAをはじめ、研究は進んでいる
 - ただし、本質的には理解できない
- 現実世界には、人間が理解できないけれど、モデル化できる、予測できることが相当たくさんあるようである。
 - 深層学習による画像認識
 - 深層強化学習による物体把持
 - アルファ碁
 - 深層学習による翻訳
 - いずれも数千万～数億のパラメータを使ってモデル化
- むしろ、少数のパラメータで記述できる系のみを科学技術として体系化してきたのではないか。

第5章：人間を超える人工知能



あり得べき知能の全体集合と人間型知能

人間を超える知能

- 生命のための知能
 - 身体や感情、教育や文化が、もともとは生存のためにできている
 - 自己保存や自己複製
 - つまり、人間の知能は、その生存という目的と分かちがたく結びついてしまっているが、本来、知能は目的非依存でいいはず。
- 多数パラメータの科学（前述）
 - 人間の「理解」が言語OSにとっての再現可能性ということに規定されている。多数パラメータを必要とするような初期条件の設定や遷移を扱うことが苦手。（個体としては熟練の業として学習できるが、伝わらない。）
- ハードウェアの限界
 - 自然界は情報の保存が苦手。離散化（デジタル化）、フリップフロップなどでループさせることで情報を保存。人間の場合は、海馬と大脳皮質でループさせることで短期記憶を保存（おそらく）。
 - 伝達のバンド幅。早口でしゃべっても、Wifiの方が早い。高速な通信は自然界は生み出さなかった。（情報の保存が苦手なことと関連）
 - 我々の知能は、我々のすむ3次元空間で訓練された「動物OS」を基盤にしている。もっと性能のよい（多次元空間で訓練された）OSに変われば、直感的に理解できることがもっと増えるはず。