

2023年11月版

本当にわかりやすいAI入門

今さら聞けない「AIとは？」を、数式や専門用語なしにイチから解説します！

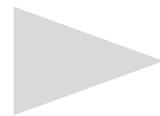


@segavvy

はじめに

今

AIって何？



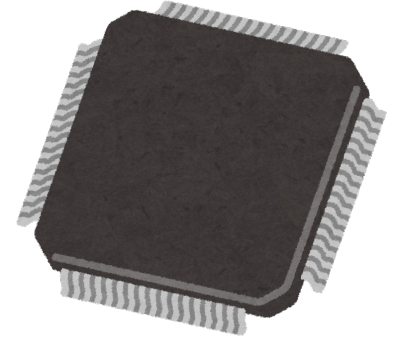
1時間後

AIが何なのか
ざっくりだけど
わかった！

**7章にわけてお話します
よろしくお願いします！**

注意

- わかりやすくなるように極端に簡略化している部分や個人的な見解でまとめている部分があります
- 数式や専門用語は極力使わないようにがんばりましたがページ下部の注釈では専門用語を使用しています
- 2023年11月時点の情報に基づきまとめています
AI界隈は動きが早いので 常に最新の情報で確認してください



第1章

AIはなぜ人間みたいなことができるのか？

そもそもAIとはなにか？

そもそもAIとはなにか？

- AIとは「artificial intelligence」の略
- 日本語にすると「人工知能」
- コンピューターで人間みたいなことをする仕組みのこと※

**AIはなぜ人間みたいなことができるのか？
→ 人間のマネをするものをAIと呼ぶから**

マネの仕方についてもう少し踏み込みます

※「人工知能」という言葉は、厳密にはそのための技術や研究分野を指す言葉としても使われます。

どうやって人間のマネをさせるのか？

コンピューターにマネさせる方法は大きく2つ

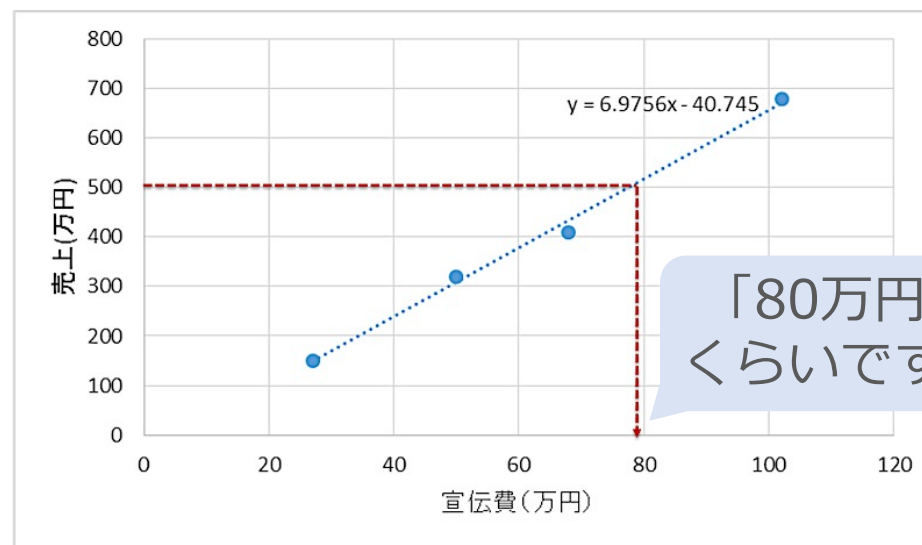
1. 「人間が考えたルール」を実行させる
2. 「ルール自体を考えること」まで実行させる

「人間が考えたルール」を実行させる

(例) Excelで必要な宣伝費を予測する

売上	宣伝費
320万円	50万円
150万円	27万円
680万円	102万円
410万円	68万円
500万円	?万円

500万の売上を狙うには
宣伝費はいくら？

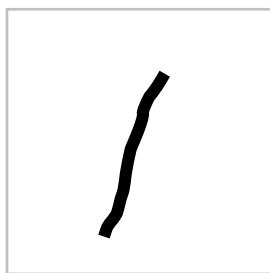


Excelには人間が考えたルール※が
実装されているので それで簡単に予測できる

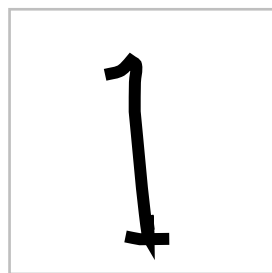
※「線形回帰」というルールがExcelの分析ツールに実装されており、この例ではそれを使っています。

「ルール自体を考えること」まで実行させる

(例) 手書きの文字を認識する



この数字は？



こっちは？

手書き文字の認識のルールを人間が考えるのはすごく大変

- 縦方向の棒が1本しかなければ「1」
- 上の部分から少し左に伸びた横棒があっても「1」
- 縦棒の下部分に短い横棒があっても「1」
- 棒が多少斜めになっていても「1」
- いずれかの棒が多少離れていても「1」
- 「7」との違いは… etc.



**ルールを考えるのが大変なものは
「ルールを考える作業」までコンピューターにやらせるのが主流**

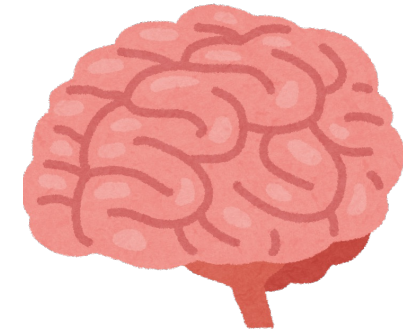
第1章 まとめ

- コンピューターで人間のマネをする仕組みが「AI」
- 人間のマネをさせる2つの方法
 1. 「人間が考えたルール」をコンピューターに実行させること
 2. 「ルール自体を考えること」までコンピューターに実行させること

第2章以降は後者の「ルール自体を考えること」まで
コンピューターに実行させる方法について解説

(休憩) 「AI」という言葉の混乱

- 「人間が考えたルール」をコンピューターに実行させる方式までAIと呼ぶことに違和感を覚えた方も多いかと思います。
Excelの例もただの分析機能ですから、AIとは呼べないと思われる方も多いでしょう。
- でも、たとえば家電では、事前のルール通りに動作するだけのものも「AI搭載！」と宣伝されていたりします。AIの定義は非常にあいまいで混沌としています。
- ちなみに数年前のAIブームでは「AI」という言葉に魅了されてExcelでできる処理まで最先端のAIを使おうとして失敗した事例が多くあり、「PoC (Proof of Concept: 概念実証) 疲れ」といった言葉が生まれたりしました。
- 目的が達成できれば「その手段がAIかどうか？」は本来どうでもよいことです。目的と手段の混同に注意して、AIという言葉に踊らされないようにしましょう。



第2章 脳はすごい

突然ですが 脳の仕組みのお話

脳はすごい

- 計算できる
- 読める・書ける
- しゃべれる
- 絵だって描ける
- 知識や経験で判断できる
- 将来の予想もできる
- 経験や訓練で効率が上がる
- できることが勉強で増やせる

コンピューターでは
それぞれのプログラムが必要

できることを増やすには
プログラムの追加・修正が必要

脳は 生まれた時にできた1つの
仕組みで なんでもできる

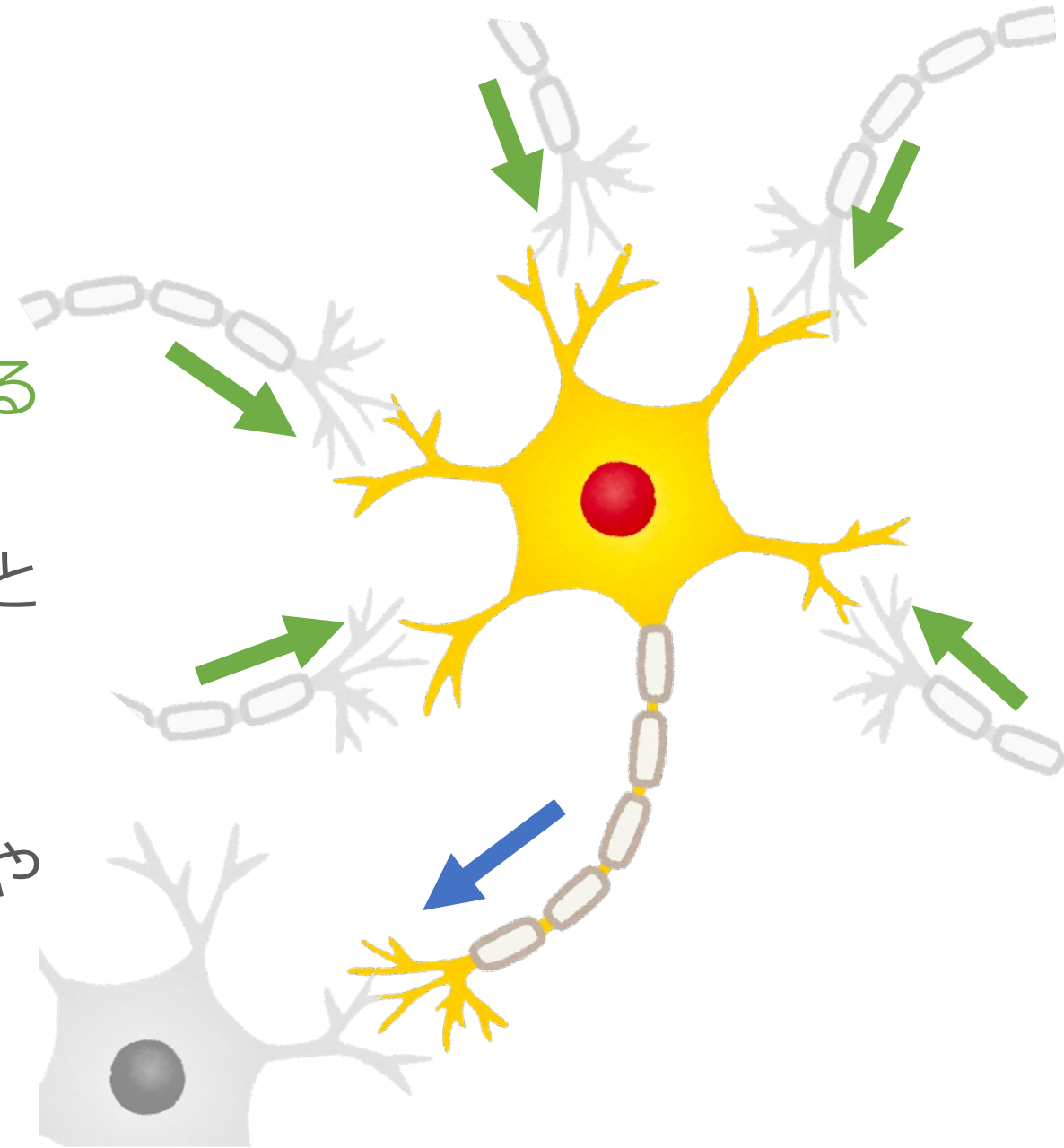
脳はすごい！

脳の仕組み

神経細胞がつながりあっていて
他の複数の細胞から信号を受け取る

信号の強さが一定の境界を超えると
次の神経細胞へ信号を伝える

次の細胞への信号の伝わりやすさや
伝えるかどうかの境界は まちまち



細胞をコンピューターでマネしたら？

- 1943年に神経生理学者・外科医のマカロックさんと論理学者・数学者のピッツさんがタッグを組み脳の神経細胞をコンピューターでマネした仕組みを考案※
- 脳の仕組みを調べてコンピューターでマネすれば人間のように思考できるAIが完成するはず！という奇想天外なアプローチ

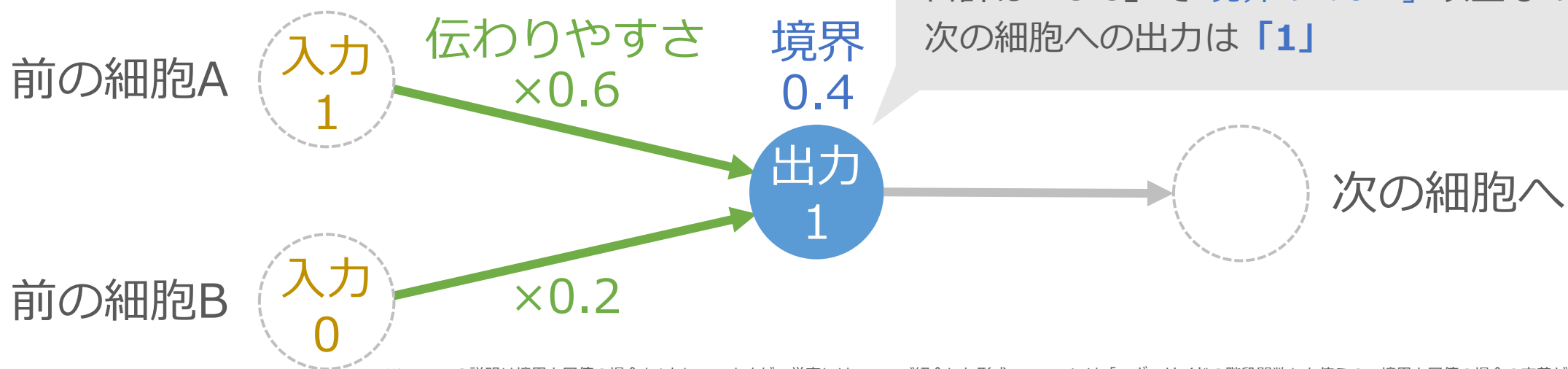


※当時、ここで考案された仕組みには呼称がなく、その後「形式ニューロン」などの名前で呼ばれるようになりました。

考案された仕組み

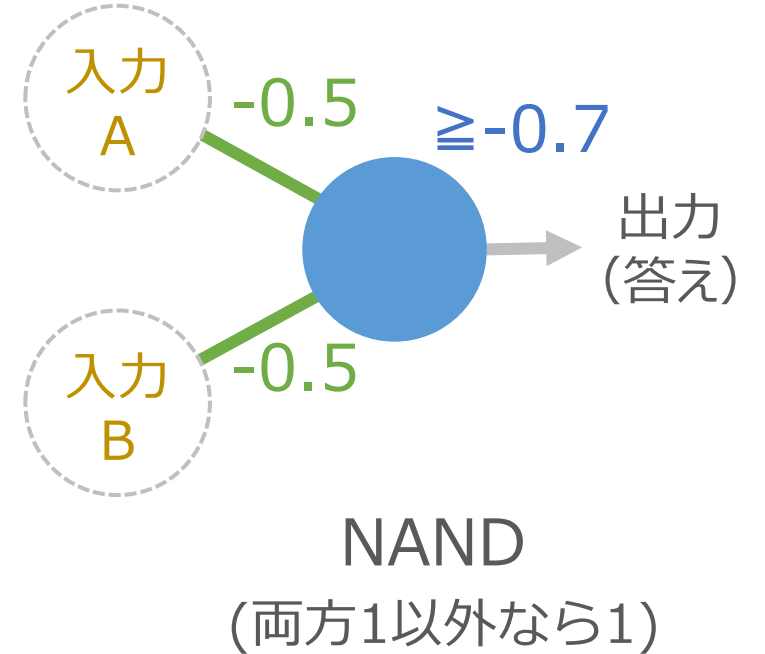
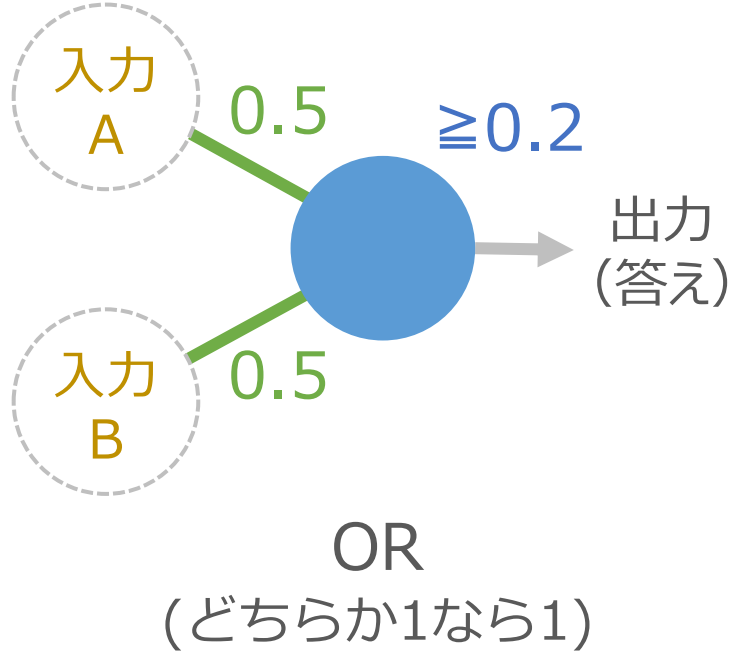
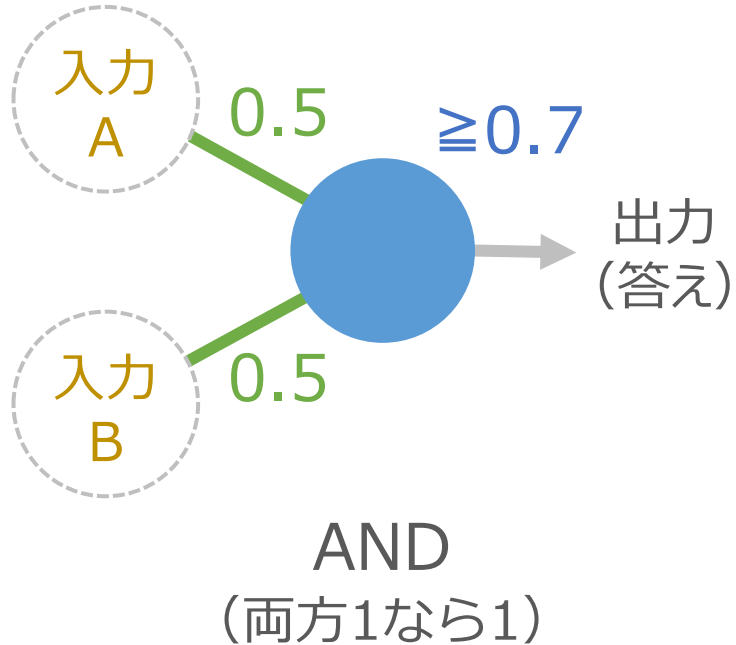
細胞をマネたシンプルな仕組み

1. 前の細胞の出力（0か1）を**入力**として受け取る
2. それぞれの**伝わりやすさ**を掛け算
3. その合計が**境界**以上なら1を出力※



※ ここでの説明は境界と同値の場合も1していますが、厳密にはここでご紹介した形式ニューロンは「ヘヴィサイドの階段関数」を使うので境界と同値の場合の定義がありません。

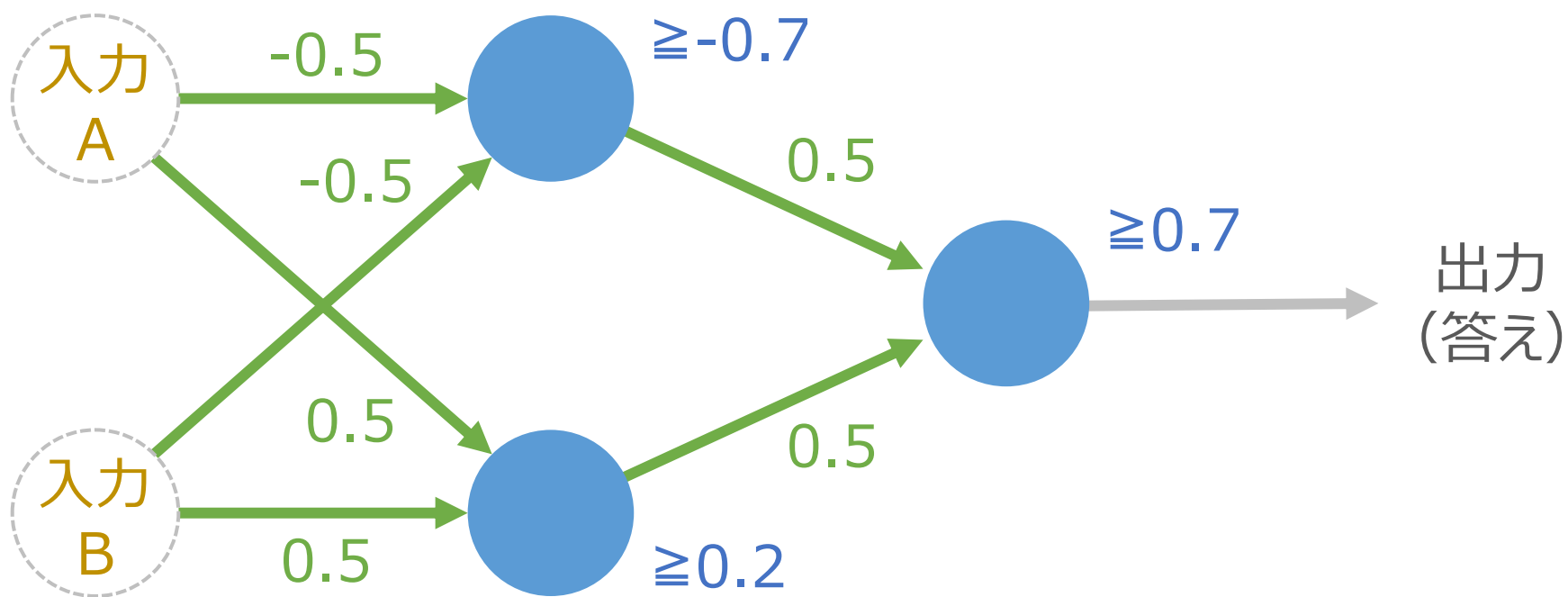
細胞1つでAND・OR・NANDができた！



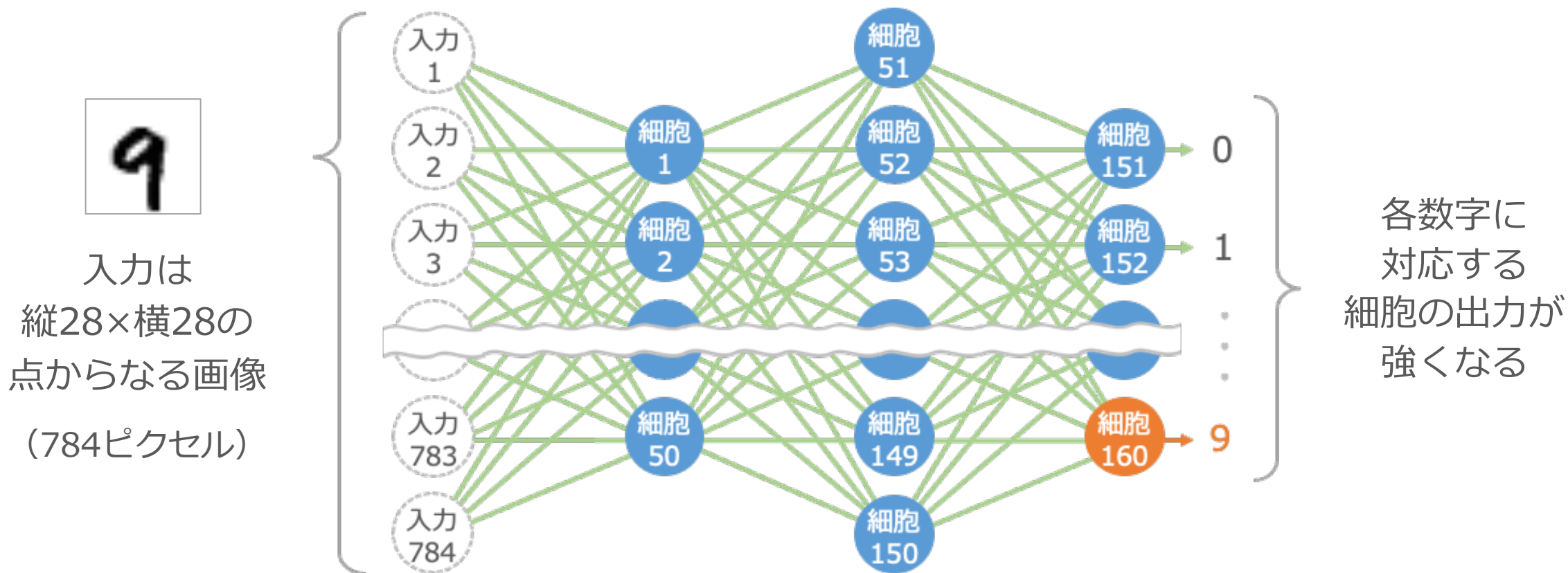
伝わりやすさと境界の値を変えるだけで
仕組み自体は変えずに3種類の機能が実現できる！

細胞が3つあればXORもできる！

細胞1つではXOR（どちらかのみ1なら1）は実現できないが
3つ使うと実現できる



160個あれば手書き数字も認識できる！※



※この例は書籍「ゼロから作るDeep Learning —Pythonで学ぶディープラーニングの理論と実装」で紹介されているもので、このような脳をマネした仕組みを「ニューラルネットワーク」(neural network)と呼びます。

なお、この図の仕組みでも手書き数字画像セットの「MNISTデータベース」で90%を越える認識ができますが、入力784ピクセルのデータを細胞1から50へ「全結合」(affine combination)しているためピクセル間の上下左右の位置関係が失われてしまい、精度を上げにくい構造になっています。そのため実際の画像認識では、ピクセルの近いもの同士だけを次の細胞につなげる「畳み込みニューラルネットワーク」(convolutional neural network)や、最終の出力と入力寄りの細胞との関連を覚えるように細胞のつながり方を工夫した「注意機構」(アテンション: attention)などが使われます。

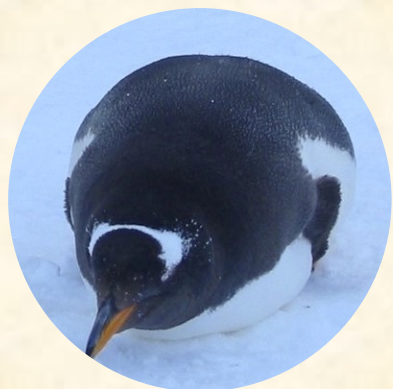
第2章 まとめ

- 脳をマネした仕組みでは
細胞間の伝わりやすさと境界値を変えるだけで
できることが変えられる
- 細胞の数を増やせば 文字認識のような複雑なこともできる

でも 伝わりやすさと境界値はどうやって決める？

第3章では細胞間の伝わりやすさと境界値の決め方を解説

(休憩) 自己紹介



えがしら たかし

江頭 貴史 @segavvy (セガビ)

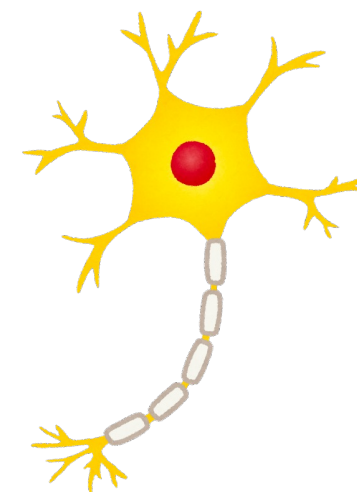


株式会社アイアクト

人工知能・コグニティブソリューション部

テクニカルディレクター兼カスタマーサクセスチーム リーダー

ものづくりや仕組みづくりが大好き。AIの社会実装に邁進中！



第3章

伝わりやすさと境界の決め方

脳のマネするのはいいけど 伝わりやすさと境界値はどうやって決める？

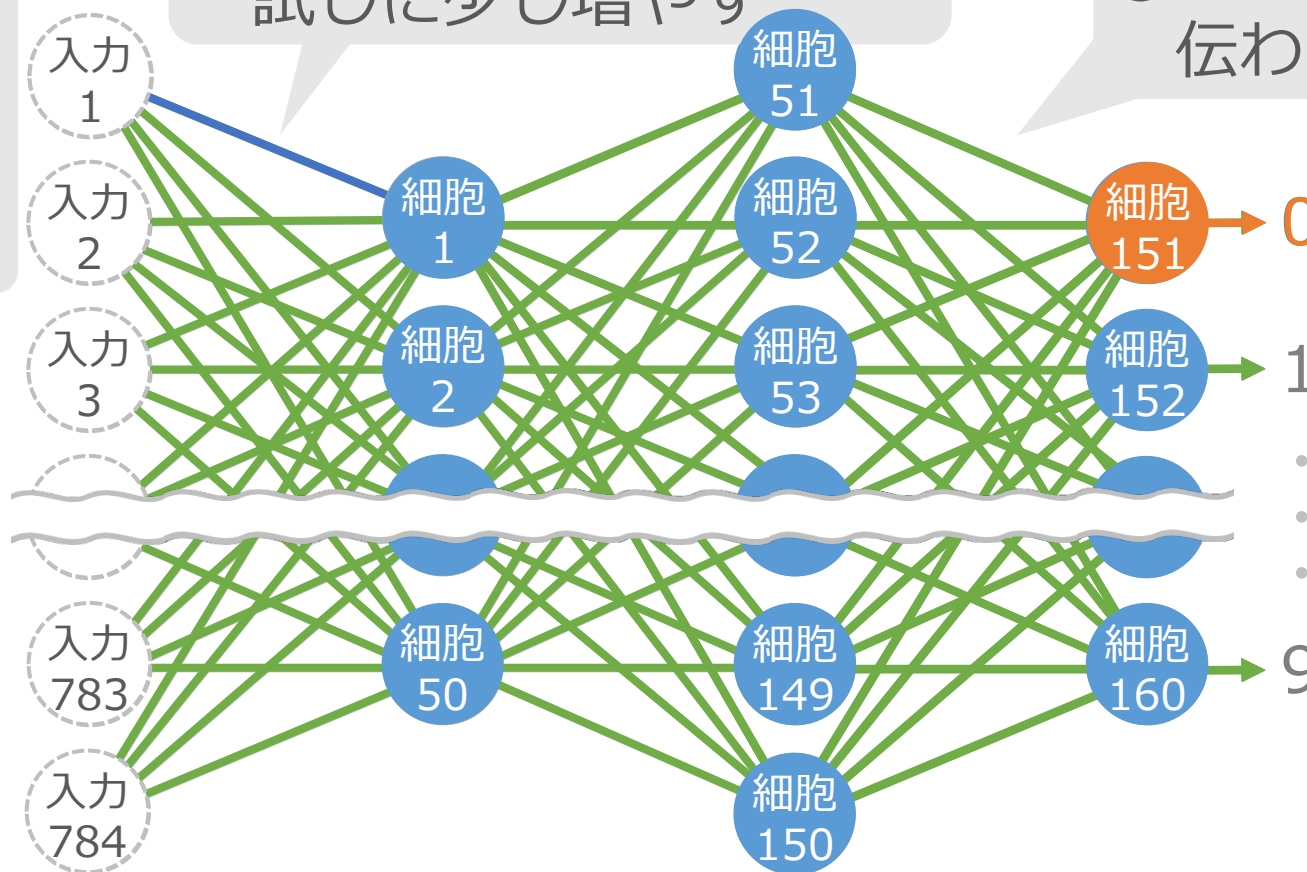
まず1か所だけ調整

①伝わりやすさと境界の値はランダムにしておいて画像を1枚入力



入力は
28×28の点

②1か所の伝わりやすさを試しに少し増やす



③後続の細胞に伝わる出力を再計算※

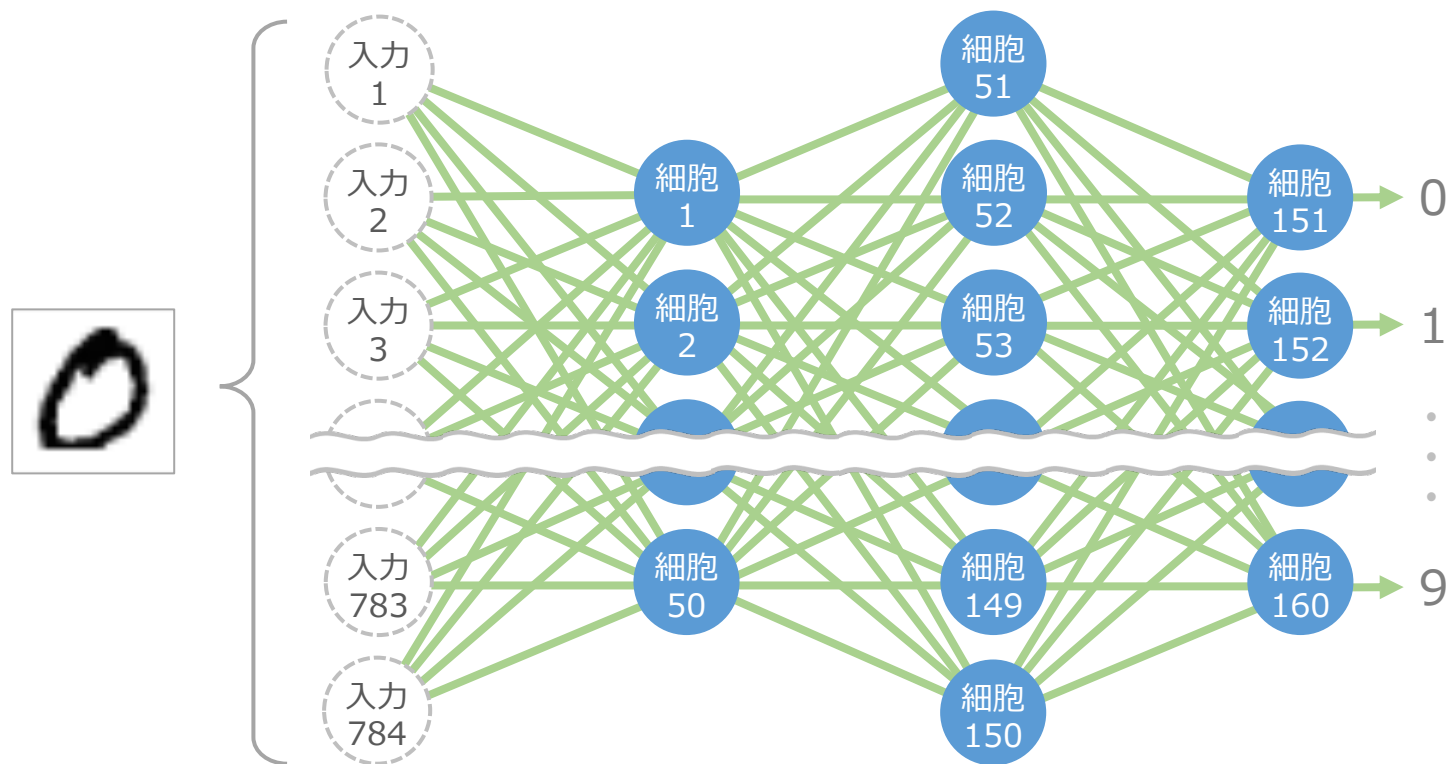
④「0」の出力が最初よりも強くなれば正解なので変更を確定

弱くなった時は逆の調整（少し減らす）

※前章では、各細胞は入力の合計が境界値以上なら1を出力するとご説明しましたが、実際には0か1のどちらかではなく、「シグモイド関数」や「正規化線形関数」（ReLU）と呼ばれる仕組みで、0から1の間のなだらかな出力にします。0か1のどちらかにしてしまうと、調整によって入力の合計が変化しても境界をまたがない限り出力が変化しないため調整の正否が判定できず、調整効率が悪くなってしまいます。

すべての値を調整

他の伝わりやすさと境界も順番に変えて「0」が強くなる方向へ



調整する値の数：

伝わりやすさ

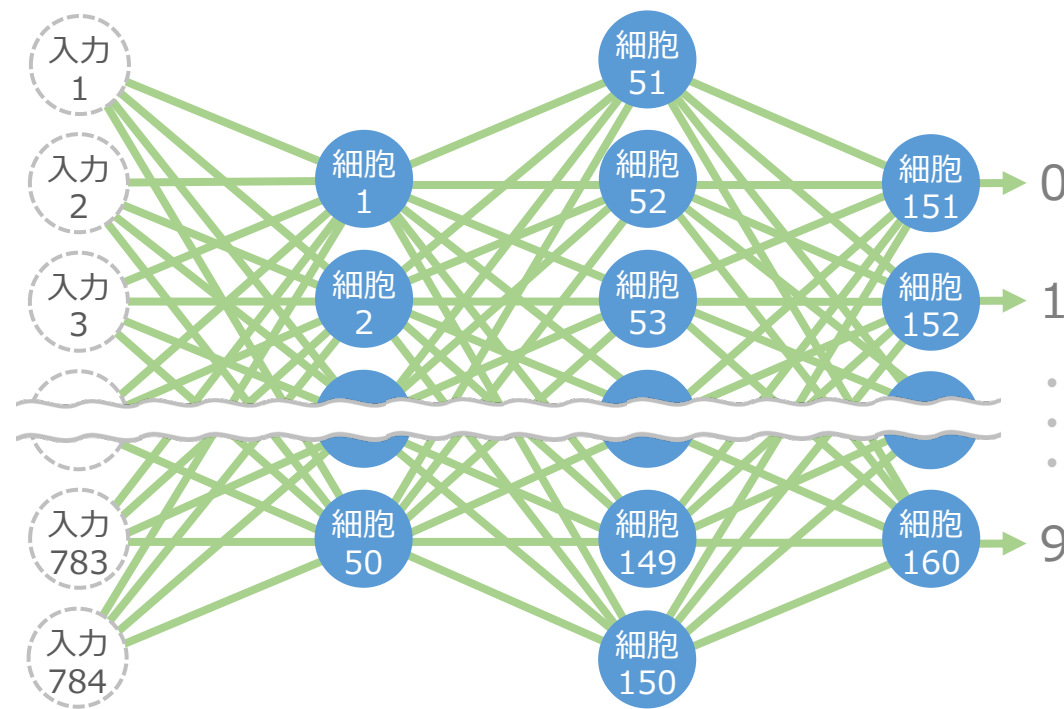
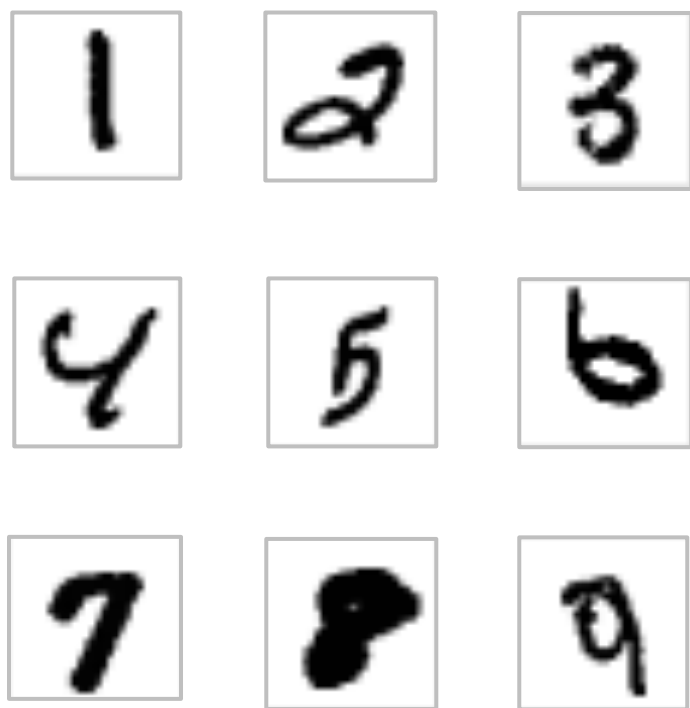
45,200個 (=線の数)

次に伝える境界

160個 (=細胞の数)

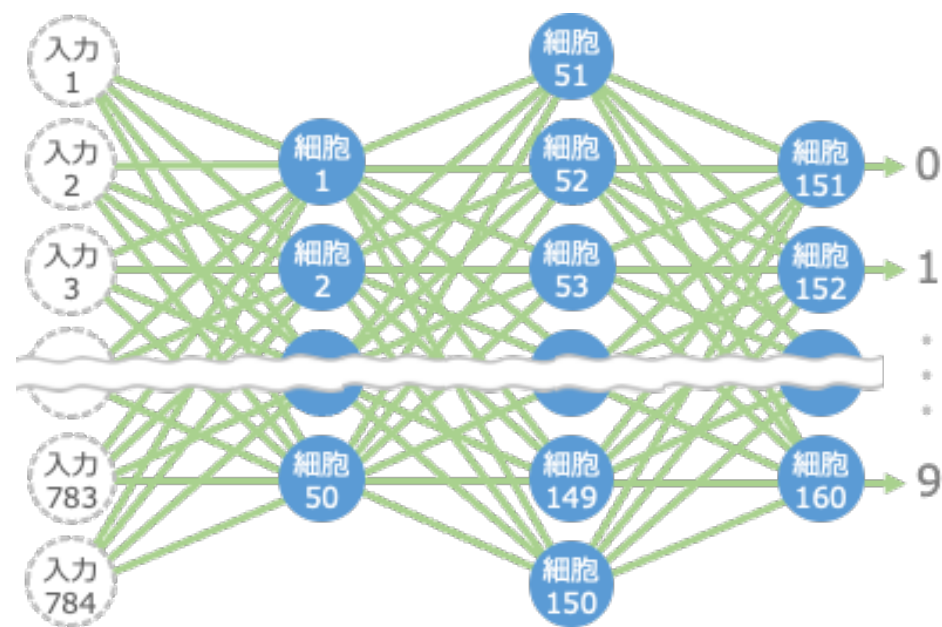
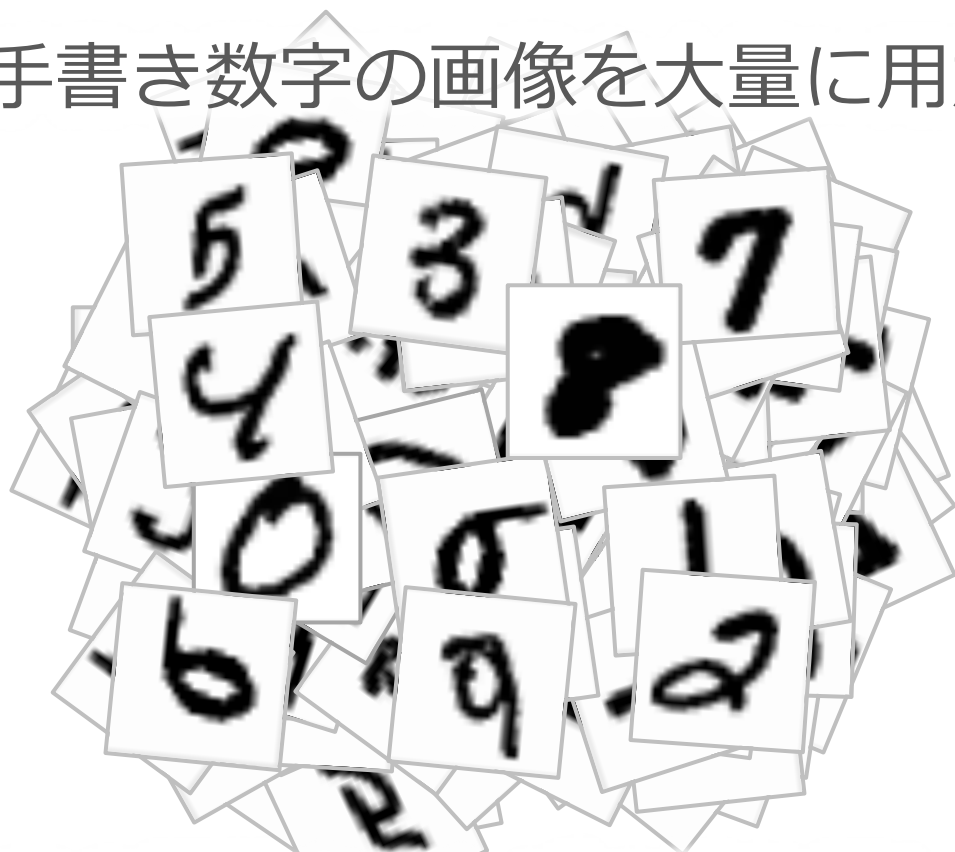
他の数字の画像でも調整

「0」以外の数字の画像でも同じように調整



いろいろな画像でさらに調整

手書き数字の画像を大量に用意してさらに調整



この調整を少しずつ繰り返して※ほぼ目的の出力になったら完了

※値の調整は最終の出力がどう変化するかを確認しながらおこなうため、最終出力までの計算を繰り返す必要があります。しかし、細胞が多いと計算量が膨大になってしまいます。そこで実際には「偏微分」を利用します。偏微分は式の中のある値が変わった時に式の最終結果がどう変化するかを調べるためのもので、偏微分ができれば再計算をせずに最終出力がわかるので各値の調整が効率化できます。実用化されている仕組みは、最終結果が偏微分できるように設計することで調整作業を大きく効率化しています。

この調整作業はコンピューターに任せる

- 調整はプログラムを開発してコンピューターにお任せできる※
- ただし 調整用の大量のデータは人間が用意しないといけない
例では「手書き数字の画像」と「正解の数字」のペアが必要

第1章より

手書き文字の認識のルールを人間が考えるのはすごく大変

- 縦方向の棒が1本しかなければ「1」
- 上の部分から少し左に伸びた横棒があっても「1」
- 縦棒の下の部分に短い横棒があっても「1」
- 棒が多少斜めになっていても「1」
- いずれかの棒が多少離れていても「1」
- 「7」との違いは… etc.



各細胞間の伝わりやすさと境界をコンピューターで調整すれば「手書き数字を認識するルール」ができあがる

どうやって人間のマネをさせるのか？

コンピューターにマネさせる方法は大きく2つ

1. 「人間が考えたルール」を実行させる
2. 「ルール自体を考えること」まで実行させる

※このように、用意した大量のデータを使って各値の調整作業をコンピューターで実行することを「機械学習」(machine learning)と呼びます。なお、機械学習という言葉は脳をマネしていない仕組みでも使われる用語です。また、脳をマネした仕組みの中で、特に大量の細胞を何層も重ねたような構造を機械学習することを「深層学習」(ディープラーニング: deep learning)と呼びます。今回ご説明したの手書き数字を認識する仕組みも深層学習の一例です。

第3章 まとめ

- （第2章より）細胞の数を増やせば 複雑なこともできる
- 各細胞間の伝わりやすさや 各細胞の境界の値は
いろいろな入力データを使い 少しずつ調整を繰り返して決める
- 調整は大量のデータを用意してコンピュータで実行
これでルール自体を考えることもコンピュータに任せられる

細胞増やして調整をコンピュータに任せたらなんでもできる!?

そうは問屋が卸さない 現実について第4章で

(休憩) DXとAIはどう違う？

- DXはデジタルトランスフォーメーションの略で、デジタル技術を活用してビジネスや顧客体験を変革することを指します。
- このデジタル技術としてAIが採用されることが多いのですが、AIを使わなくてもDXは実現できますし、DXに関係しない部分でもAIは使われます。
- DXやAIという言葉に囚われすぎず、本来の目的にあった手段を検討することが大切です。
- よろしければ拙作のブログ記事 [「本当にわかりやすいDX入門～そもそもDXとはなにか？～」](#) もどうぞ！





第4章 細胞増やすだけではダメだった

課題を乗り越えるための取り組み その1

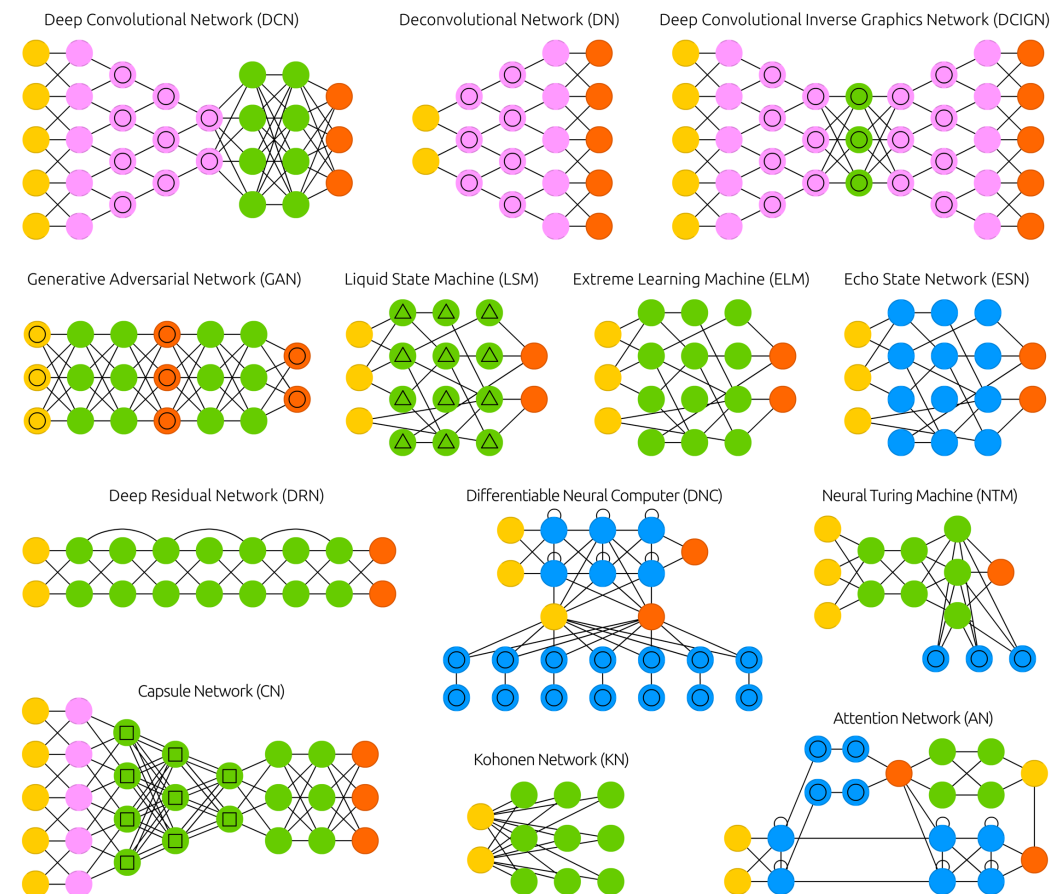
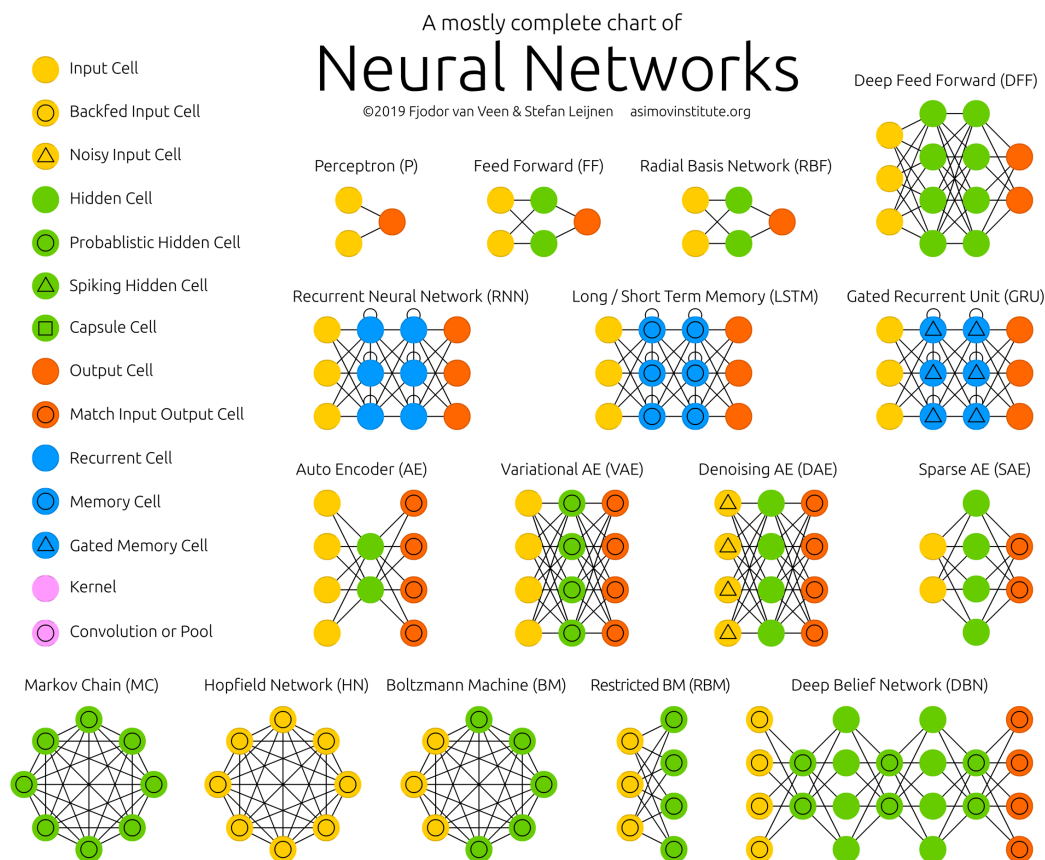
細胞増やすだけではダメだった

- 細胞間の伝わりやすさと境界の値の調整はさじ加減が難しい
- コンピューターに調整させ続けても 失敗することがある
- 細胞を増やすだけでは解決できないことも多い

仕組みの改良が必要

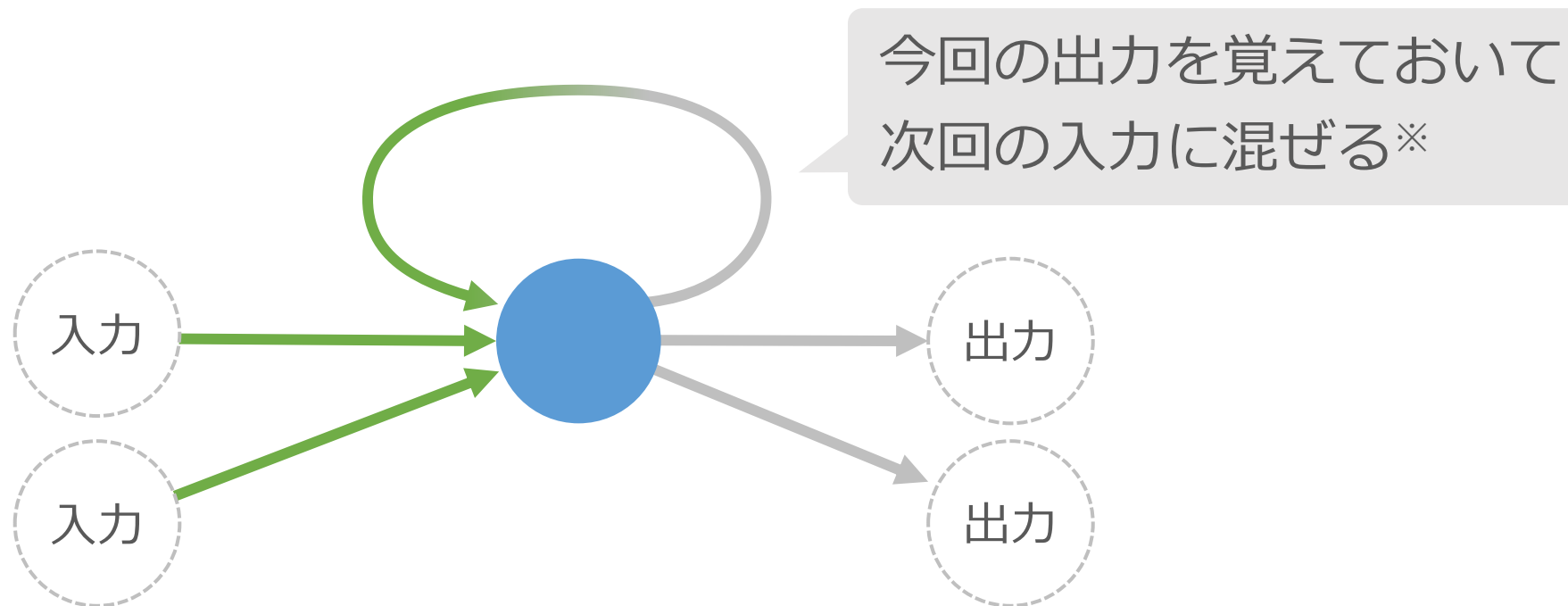


さまざまな細胞の機能とつなぎ方



(出典) THE ASIMOV INSTITUTE. 「THE NEURAL NETWORK ZOO」 . <https://www.asimovinstitute.org/neural-network-zoo/>

つなぎ方の工夫の例



このような工夫をすると 以前の入力の考慮ができる

※細胞の出力を次の自身の入力に使うつなぎ方を「回帰型ニューラルネットワーク」（リカレントニューラルネットワーク：recurrent neural network）と呼びます。順番に意味のあるデータで使われる仕組みの1つです。

出力を次回の入力に混ぜる例：文章生成※1

小説を使って 続く単語が
出力になるよう調整する※3

調整	入力※2	理想の出力
1回目	吾輩	は
2回目	は	猫
3回目	猫	で
4回目	で	ある
5回目	ある	。
:	:	:

文章の書き出しを与えると
その著者風の続きを生成してくれる

生成	入力※2	出力
1回目	吾輩	(未使用)
2回目	は	(未使用)
3回目	犬	で
4回目	で	ある
5回目	ある	。
:	:	:

「吾輩は犬」だけ
入力する

続きは 得られた
出力をそのまま入力

これで続きが
生成されていく

※1 この例は書籍「ゼロから作るDeep Learning ② 自然言語処理編」で紹介されているものをベースにしていますが、書籍では「コーパス」が英文だったので、私は「青空文庫」の小説で試しました。当時のメモを「ゼロから作るDeep Learning ②で素人がつまづいたことメモ: まとめ」にまとめているので、よろしければご参照ください。なお、ここでは回帰型ニューラルネットワークのわかりやすい例として文章生成を取り上げましたが、文章生成や音声認識などの時系列データを取り扱う場合は、改良型である「長・短期記憶」(LSTM: Long Short-Term Memory)や「ゲート付き回帰型ユニット」(GRU: Gated Recurrent Unit)などの方が精度が上がります。また、これらの仕組みは時系列順にデータを処理する関係で並列処理がしにくいため、大規模なデータセットを処理する場合は「トランスフォーマー」(transformer)がよく使われています。

※2 単語は第3章でご紹介した手書き数字画像のような固定数の数値の羅列ではないので、実際には単語を入力する前に「単語の埋め込み」(word embedding)という仕組みで固定数の数値の羅列に変換します。

※3 このように大量のテキストデータを使って文章を生成できるようにした仕組みを「言語モデル」(language model)と呼びます。

第4章 まとめ

- 細胞の数を増やすだけでは限界がある
- 細胞自身の機能やつなぎ方にはさまざまな工夫がある
- たとえば 出力を次回の入力にまぜると文章の生成もできる

細胞の工夫でなんでもできる!?

もう1つの大きな課題について 第5章で

（休憩） AIでなぜグラボが人気になる？

- グラボ（グラフィックボード）は画像処理のための行列計算に特化した演算装置です。画像処理では大量の点の情報を一気に計算するために行列の計算を多用します。グラボはその行列の計算専用に設計されているため、画像処理が高速化できます。
- ここまでにご紹介した手書き数字の認識や文章の生成では、その調整や認識の作業で各細胞の出力と伝わりやすさの掛け算、そしてその結果の足し算を大量に繰り返します。この計算は、少し工夫するとグラボが得意な行列の計算に置き換えられます。そのため、グラボを流用するとAIの調整や認識が高速化できるのです。
- 最近ではAI専用の演算装置もどんどん開発されています。
- 行列は一時期高校の数学から外れていたのですが、AIの普及に伴い重要度が増しており、文科省は2022年に数学Cで復活させています。



第5章

時間も手間もお金もかかる

課題を乗り越えるための取り組み その2

時間と手間とお金がかかる問題

- 細胞が増えると調整対象が増えて 必要なマシンパワーも増大
- マシンパワーを使うにはお金がかかる
- 調整のためのデータの準備も大変
- できれば調整作業はやりたくない

調整済みのものの流用や組み合わせで節約
今回は3つの例をご紹介します



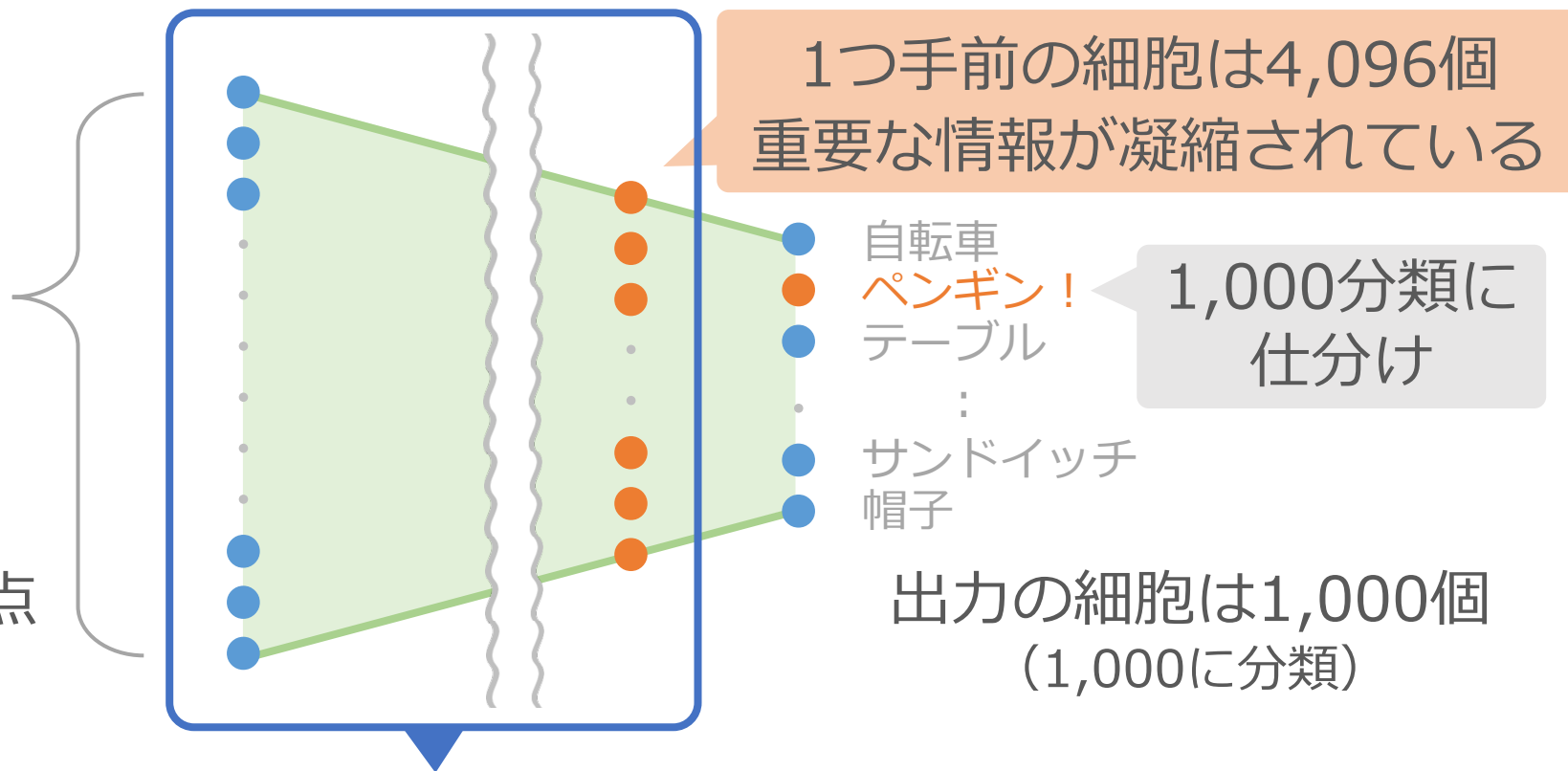
調整済み!

(例1) 画像分類の仕組みと……

使い回すのは 画像を1,000分類に仕分ける調整済みの仕組み※



入力は150,528個の点
($224 \times 224 \times 3$)



画像の情報を4,096個の数字の羅列に凝縮する仕組みとして流用

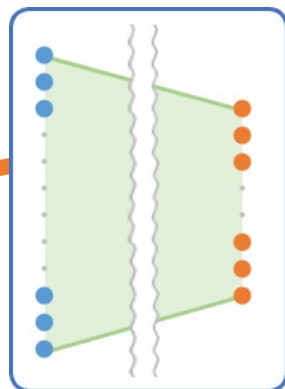
※この例は「VGG16」という仕組みです。ここでの説明には「AI人工知能テクノロジー」さんのブログ「[VGG16モデルを使用してオリジナル写真の画像認識を行ってみる](#)」を参考にいたしました。

(例1) 文章生成を組み合わせると……



入力は
150,528個の点
(224 x 224 x 3)

調整済みの画像分類の
仕組みを流用して
画像の重要情報を凝縮



調整	入力	理想の出力
1回目	凝縮された 画像の情報	ペンギン
2回目	ペンギン	が
3回目	が	雪
4回目	雪	の
5回目	の	上
6回目	上	に
7回目	に	立って
⋮	⋮	⋮

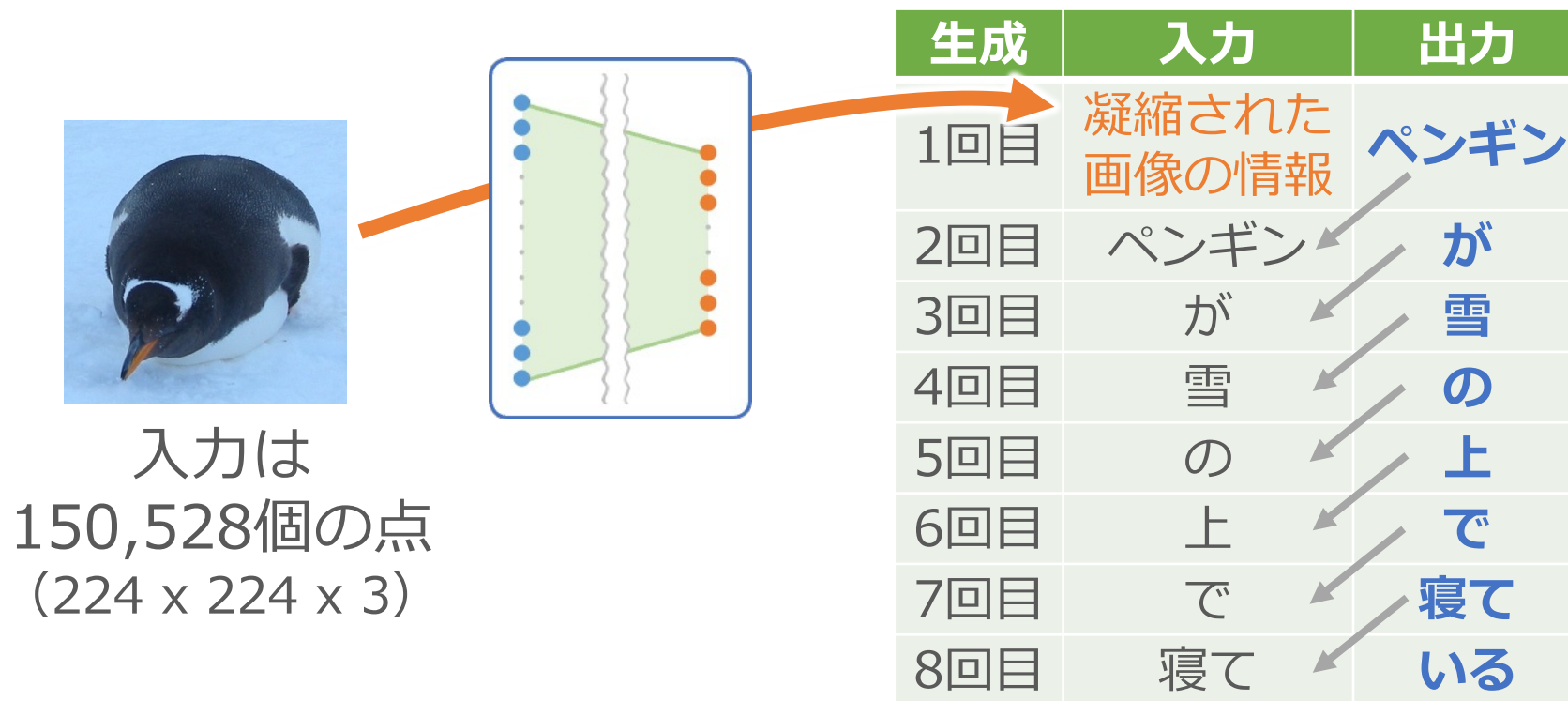
画像の情報を
最初に入力※

以降は画像の
説明文に
なるように調整

このように
画像と説明文のペアで
調整を繰り返すと……

※ここではわかりやすくなるように、凝縮した画像の情報と続く単語が同じ形で入力される表にしています。実際の画像の情報は、ニューラル ネットワークの中間部分に直接入力する形になっています。

(例1) 画像の説明文が作れる！

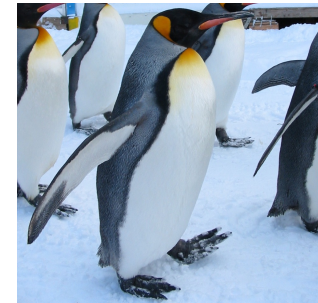
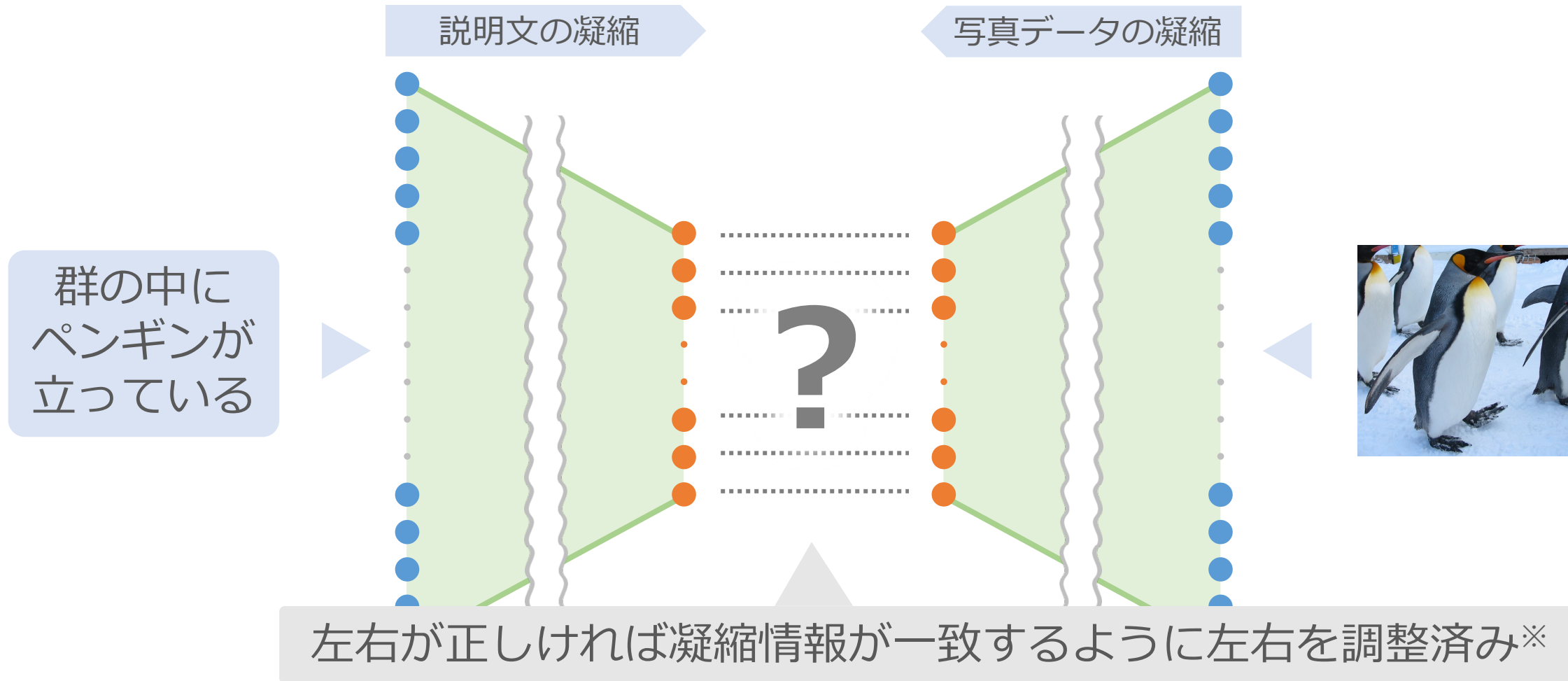


画像の説明文が
生成できる！

既存の仕組みを流用して**重要な情報**を画像から凝縮※
これでデータ準備や調整などのコストが省ける

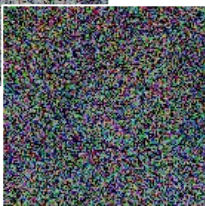
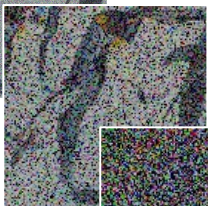
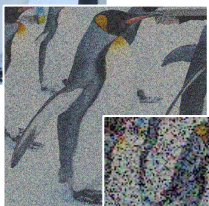
※この例ではVGG16を流用することで画像を4,096個の情報に凝縮しましたが、このように情報を凝縮させる仕組みを「エンコーダー」(encoder)と呼びます。また、それを利用して文章を生成させましたが、このように凝縮した情報から別の情報を生成する仕組みを「デコーダー」(decoder)と呼びます。このようなエンコーダーとデコーダーの組み合わせ例は多くあります。たとえば、日本語の文章をエンコーダーで凝縮し、それを英文生成のデコーダーで文章生成することで日英翻訳が実現できます。

(例2) 画像と説明文の正誤判定の仕組み※と…



※この例は「CLIP」(Contrastive Language-Image Pre-training) という仕組みです。ここでの説明は「TRAIL (Tokyo Robotics and AI Lab)」さんの記事「[CLIP：言語と画像のマルチモーダル基盤モデル](#)」を参考にいたしました。

(例2) 画像を繊細化する仕組み^{※1}を……



画像を何度も汚して^{※2}
調整用の画像を作成

調整	入力	理想の出力
ペア1		
ペア2		
:	:	:
ペアn		

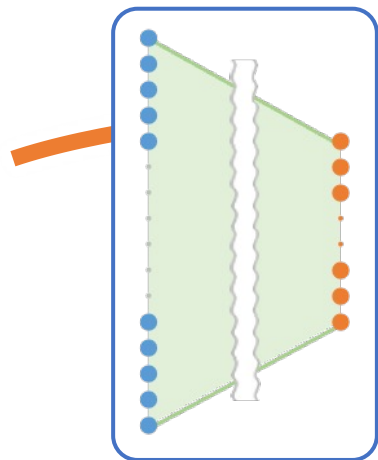
汚した画像と
汚す1つ前のペアを
入力と理想の出力にして
大量の画像で調整

画像の汚れを取り除き
繊細な画像が
生成できる！^{※3}

※1 この例は「Stable Diffusion」という仕組みの一部です。ここでの説明は「Gigazine」さんの記事「画像生成AI「Stable Diffusion」がどのような仕組みでテキストから画像を生成するのかを詳しく図解」を参考にいたしました。
※2 Stable Diffusionでは画像を直接汚すのではなく、実際にはまず凝縮して情報量を減らし、それに対してノイズを加えて汚します。そのため、実際には目で汚れていく様子が確認できるようなデータとしては処理されていません。
※3 ノイズを加えて汚したデータを用意し、それを元に戻せるように調整していく仕組みを「拡散モデル」(diffusion model)と呼びます。

(例2) 組み合わせで調整していくと……

群れの中に
ペンギンが
立っている

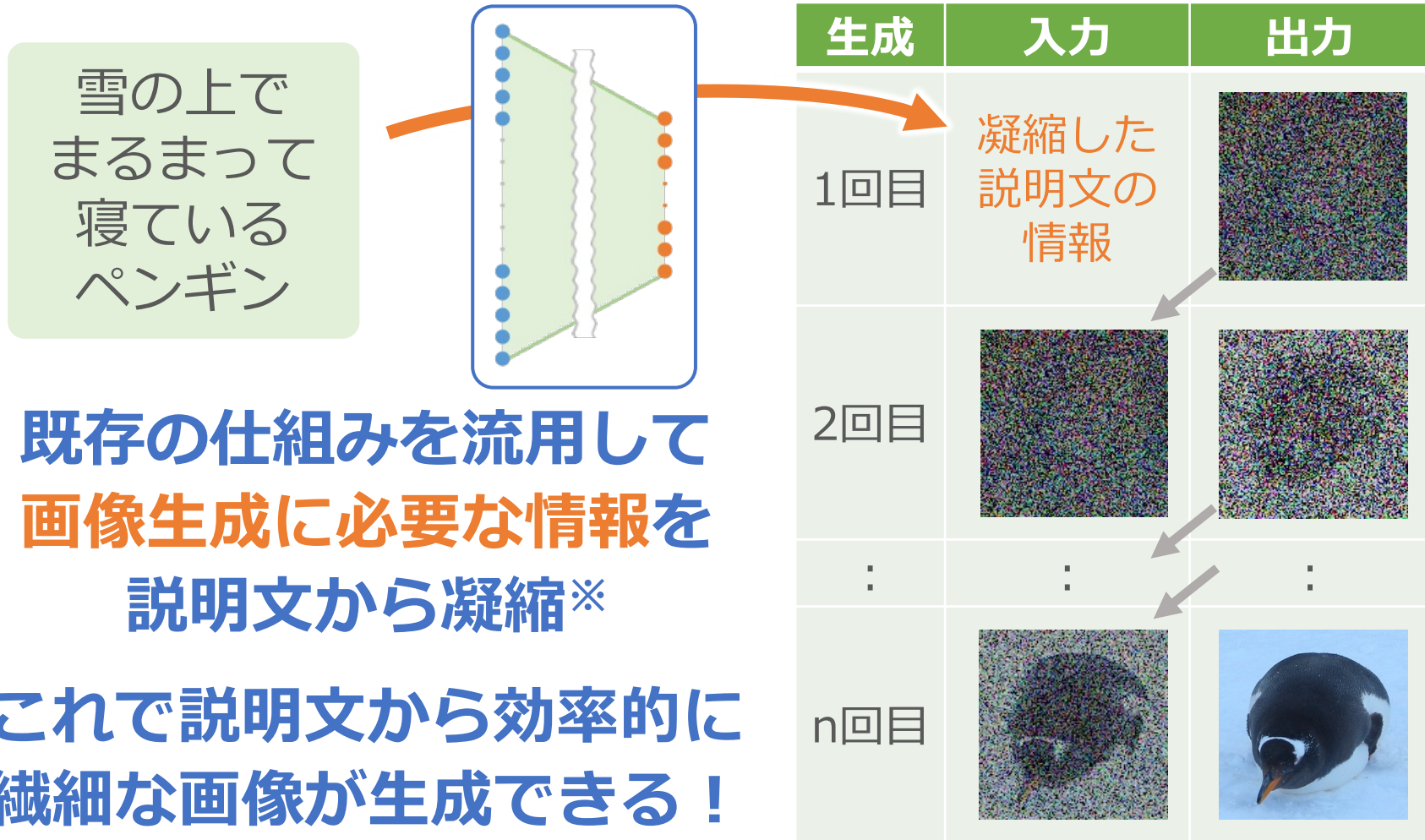


画像と説明文の
正誤判定における
説明文の凝縮部分（左半分）を
流用して説明文の情報を凝縮

調整	入力	理想の出力
1回目	凝縮した 説明文の 情報	
2回目		
:	:	:
n回目		

理想の出力は
正解の画像を
何度も汚して
作成

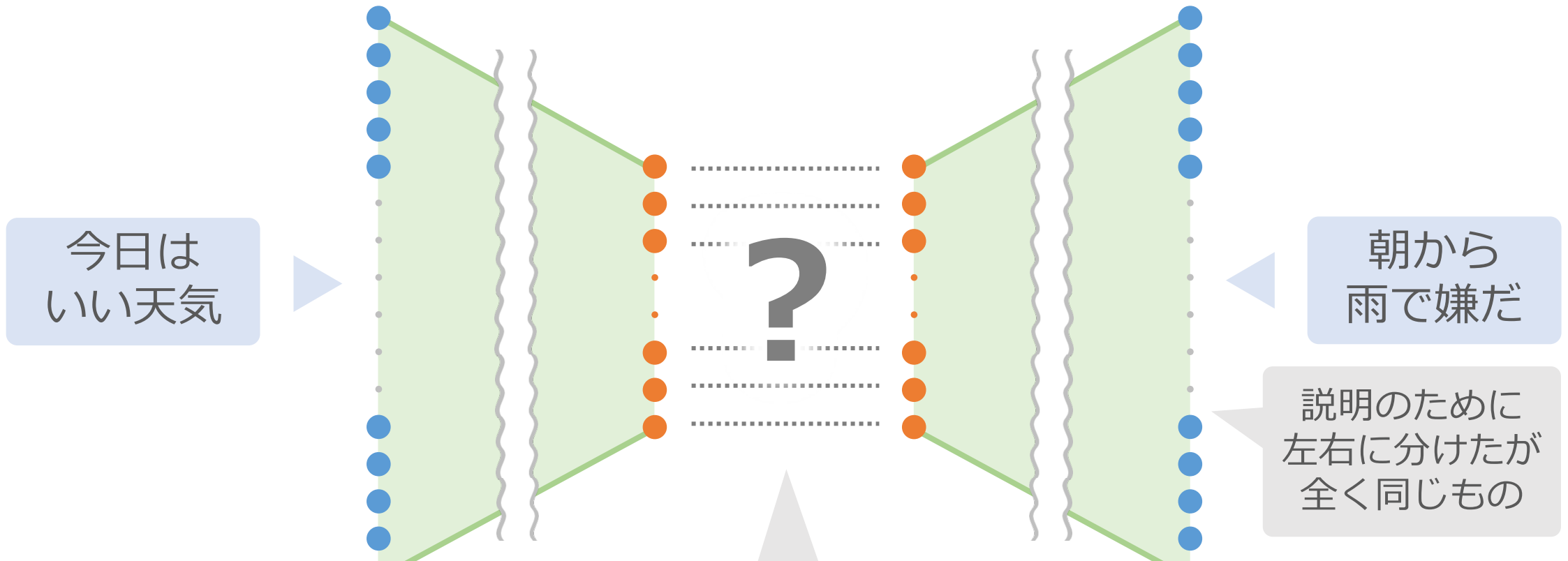
(例2) 説明文から繊細な絵が描ける！



説明文に
あった
繊細な画像が
生成できる！

※ここで説明したStable Diffusionでは、調整済みのCLIPの仕組みを、画像を生成するという別の目的に流用しています。また、1つ前の(例1)では、調整済みのVGG16を、文章を生成するという別の目的に流用しました。このように、調整済みの仕組みを別の目的に流用することを「転移学習」(transfer learning)と呼びます。

(例3) 文を数字の羅列^{※1}に凝縮する仕組みと…



同じ仕組みに2つの文を入力して凝縮情報を比較
 同じ意味の文章は似た数字になるように 矛盾する場合は似ないように調整済み^{※2}

※1 数字の羅列は「ベクトル」と呼ばれ、ベクトル同士の距離や類似度を計算することができます。ベクトルの距離や類似度の計算方法にもいくつか種類があり、「ユークリッド距離」や「コサイン類似度」などが有名です。

※2 この例は「教師ありSimCSE」(supervised SimCSE)という仕組みです。ここでの説明は書籍「大規模言語モデル入門」を参考にいたしました。実際には2つの文章とその関係(含意/中立/矛盾)のセットを大量に用意して調整して、含意の場合は凝縮したベクトルが近くなるように、矛盾の場合は遠くなるように調整しています。

(例3) 数字の羅列が検索できる仕組み※で……

数値の羅列
3, 1, 4, 1, 5, 9, …
2, 6, 5, 3, 5, 8, …
9, 7, 9, 3, 2, 3, …
8, 4, 6, 2, 6, 4, …
:

数字の羅列に
特化した
検索の仕組み

「2, 2, 5, 0, 4, 9, …」で
検索

「3, 1, 4, 1, 5, 9, …」と
「1, 3, 6, 0, 4, 8, …」が
近いです!

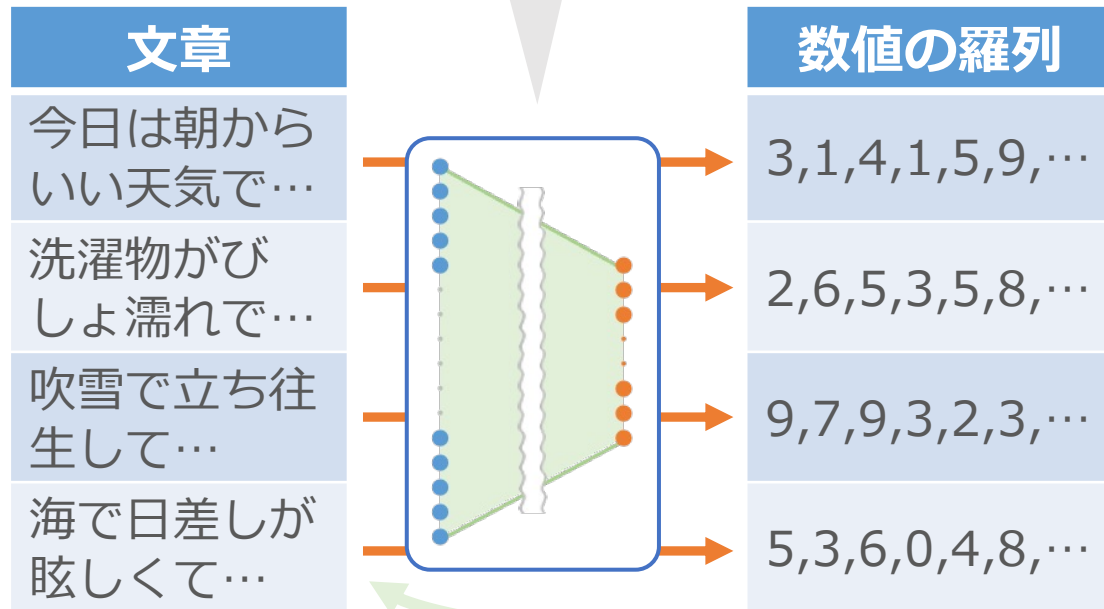
検索対象の数字の羅列を
事前に登録しておけば……

数字の羅列を条件にして
似たような羅列を検索できる

※数字の羅列である「ベクトル」を使って情報を検索する仕組みは「ベクトル空間モデル」や「ベクトル検索」と呼ばれ、「ベクトルストア」(vector store)や「ベクトル検索エンジン」(vector search engine)、「埋め込みデータベース」(Embedding Database)と呼ばれるシステムに実装されています。通常の検索の仕組みは条件に一致するものを探しますが、ベクトル検索の場合は、「ユークリッド距離」や「コサイン類似度」などにより似たようなベクトルを検索できることが特徴です。

(例3) 意味の似ている文章が検索できる！

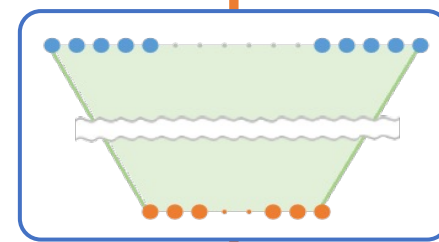
文を数字の羅列に凝縮する仕組みを流用して変換※1



数字の羅列に
特化した
検索の仕組み



「晴れの日の話」で検索



「4,2,5,0,4,8,…」で検索

「3,1,4,1,5,9,…」と
「5,3,6,0,4,8,…」が近い

対応する元の文章を確認

**意味の把握に必要な情報の凝縮も
その検索も 既存の流用で実現！※2**

「今日は朝からいい天気で…」と
「海で日差しが眩しくて…」がヒット！

※1 文章を凝縮してベクトルに変換することを、「文章の埋め込み」(text embedding)と呼びます。多言語に対応する仕組みの場合は、同じ意味なら日本語のベクトルも英語のベクトルも中国語のベクトルも近い関係になります。

※2 これまでの文章の検索は「全文検索」の利用が一般的でしたが、字面が違いすぎたり表現が違いすぎたりすると意味の似た文章を検索できませんでした。文章の埋め込みとベクトル検索の組み合わせはこの弱点を補えるものとして非常に期待されていますが、文章の埋め込みに使う仕組みが情報を適切に凝縮できない文章は検索できないので、決して万能ではありません。そのため、複数の検索の仕組みを組み合わせたり、検索結果の並びを調整する仕組みと組み合わせたりする工夫が重要になっています。

第5章 まとめ

- 調整済みの仕組みを流用すれば コストが抑えられる
- 調整済みの仕組みで情報を凝縮して他の用途に使うと
効率的にできることが増やせる
(例) 画像の説明文を生成、説明文から絵を描く、意味の似た文の検索
- 用途にあった凝縮の仕組みを使わないと 精度はあがらない

ここまでが2021年くらいまでのお話
第6章からは最近のお話を

(休憩) エンジニアにオススメの本

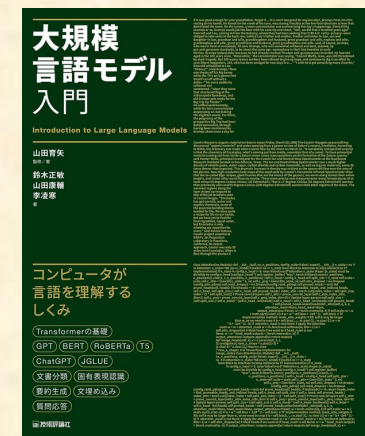
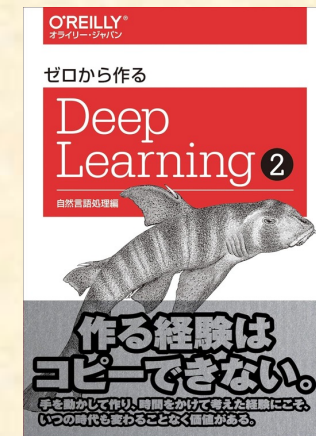
最新のAIを使ったアプリ開発を 実際に動かしながら体験してみたい！

- [ChatGPT/LangChainによるチャットシステム構築 \[実践\] 入門](#)



今回ご紹介しているような仕組みの詳細を 実際に動かしながら学びたい！

1. [ゼロから作るDeep Learning](#)
—Pythonで学ぶディープラーニングの理論と実装
2. [ゼロから作るDeep Learning ②](#)
—自然言語処理編
3. [大規模言語モデル入門](#)





第6章 文章生成の大規模化による進化

文章生成の仕組みが驚くべき進化を遂げる

文章生成の仕組み※は質問にも答えられる

ある程度

第4章より

文章の書き出しを与えると
その著者風の続きを生成してくれる

生成	入力※2	出力
1回目	吾輩	(未使用)
2回目	は	(未使用)
3回目	犬	で
4回目	で	ある
5回目	ある	。
:	:	:

「吾輩は犬」だけ
入力する

続きは 得られた
出力をそのまま入力

これで続きが
生成されていく

調整データには
インタビューや掲示板の
やり取りなどの
質疑応答もある

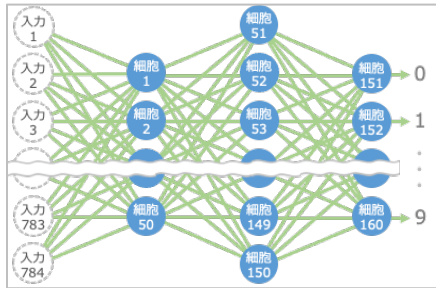
夏目漱石の小説で調整したら
夏目漱石風の文章が生成できる

大量の文章で調整したら
大量の知識に基づく文章が生成でき
質問にもある程度は回答できた！

※ ここでの説明は2020年に発表されたGPT3（Generative Pre-trained Transformer 3）と呼ばれる一昔前の仕組みの特性です。特定の事実に関する直接的で明確な質問の場合は、ある程度の回答ができます。ただし、流暢な会話や長い会話のやり取りは苦手です。

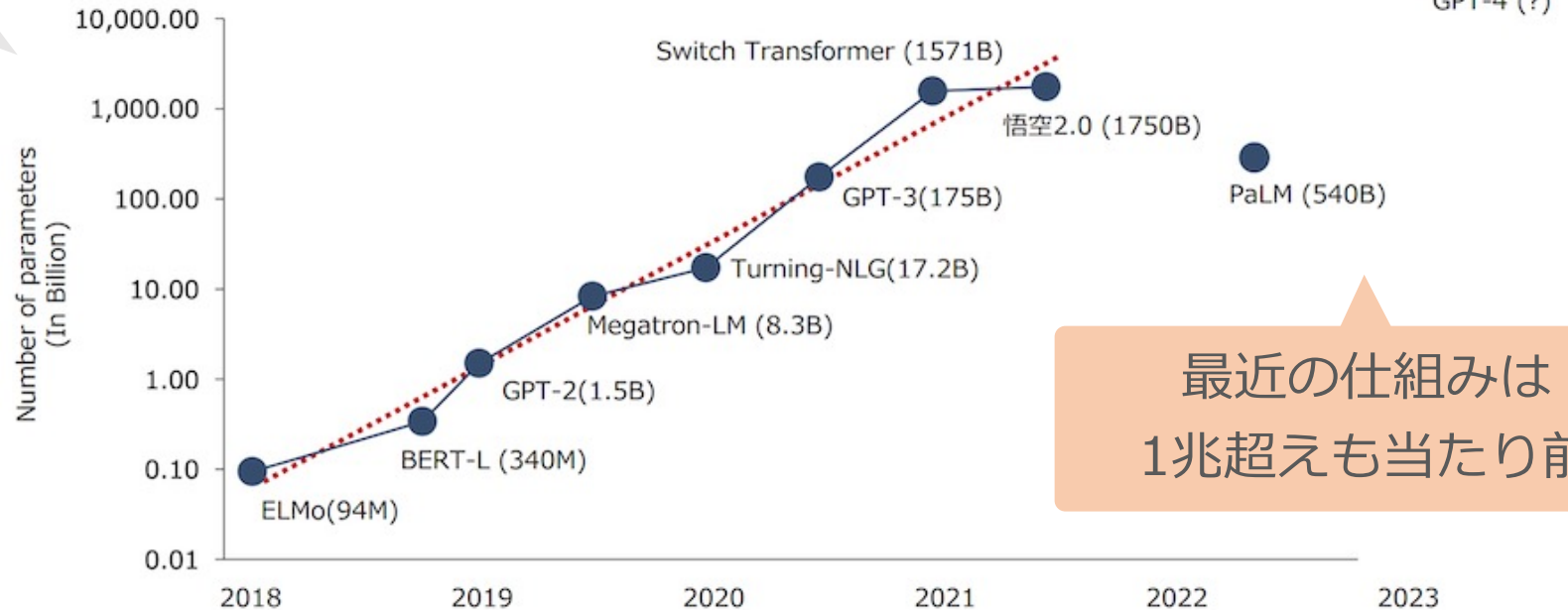
さらに仕組みを大規模化したら……

縦軸は調整する値の数
単位は10億 (B)



ちなみに第2章の
手書き数字認識の例は
たったの45,360個

大規模言語モデル (Large Language Models; LLMs) より大きなモデルに



最近の仕組みは
1兆超えも当たり前

(出典) 東京大学松尾研究室 [「AIの進化と日本の戦略 23/2/17」](#)

※厳密な定義はないのですが、言語モデルに対して調整する値の数と調整に使うデータ量、そして調整のための計算量を大規模化したものを「大規模言語モデル」(LLM: Large Language Model)と呼びます。言語モデルの大規模化が進む理由の1つは、「自然言語モデルのスケーリング則」(Scaling Laws for Neural Language Models)で、大規模化すればするほど性能が上がるという法則が見えてきたためです。なお、このような大規模言語モデルはいろいろな用途で活用できるため「基盤モデル」とも呼ばれます。

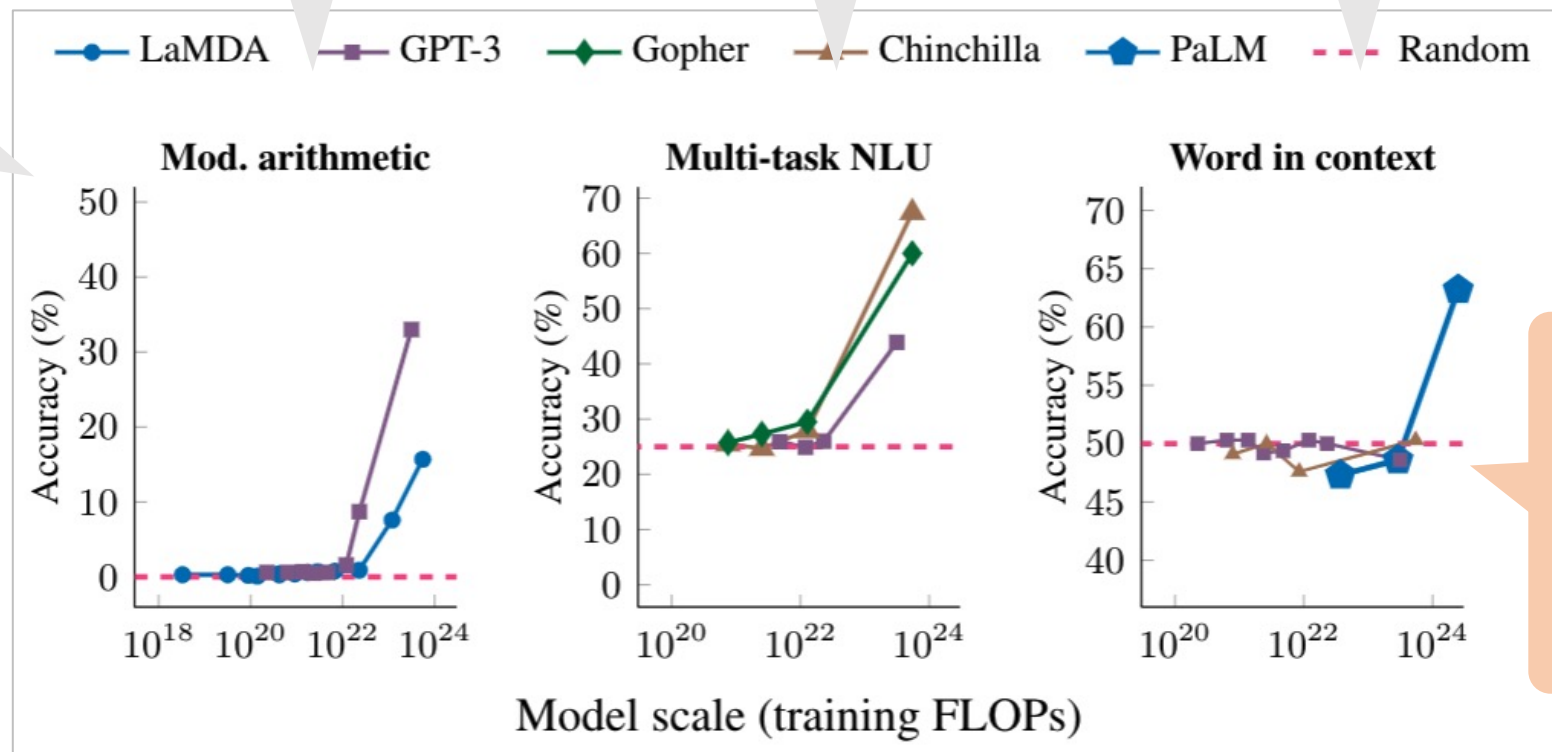
性能が飛躍的に向上することも判明！

複雑な数学の問題

大学の受験問題

単語の意味の特定

縦軸は
回答の正確さ
横軸は
調整の計算量



調整の計算量があるポイントを超えると性能が飛躍的に向上する

(出典) [「Characterizing Emergent Phenomena in Large Language Models」](#)

難しい問題に答えることができる！

追加の調整で会話もできるようになった！



①大量の文章で調整

文章生成の
仕組み※1

②対話の文章を大量に用意して追加で調整

追加で調整	入力	理想の出力
1回目	〇〇を教えて	それは□□です。
2回目	△△の方法は？	まず最初に…
3回目	この◇◇はどう？	次の方法が…
:	:	:

③回答の優劣を別の仕組みに判断させて
その結果に基づきさらに調整を繰り返す

回答の良し悪しを判定する仕組み※2

人間らしい会話や 会話の往復もできるようになった！

※1 ここでの説明は「InstructGPT」という仕組みの調整方法です。①で調整したGPT3に対して、②の調整によって人間らしい会話や会話の往復をできるようにします。このように、既存の仕組みに対して追加のデータで振る舞いを調整することを「ファインチューニング」（fine-tuning）と呼びます。また、②のファインチューニング後の仕組みに対して、さらに③の仕組みで回答の良し悪しを判定させて、その結果を基に調整を繰り返します。この調整では、これまでの例のように理想の出力を使うのではなく、結果の良し悪しの判定結果しか使いません。このようにに結果の良し悪しから正しい結果を出力できるように調整する方法を「強化学習」（RL : reinforcement learning）と呼びます。

※2 回答の良し悪しを判定する仕組みは、②の調整済みの仕組みに対して出力が良し悪しを返すように改造したものを用意して、人間が結果の良し悪しを判定したデータを理想の出力として調整することで作り上げています。この③のように、人間の判定結果を使って強化学習する方法は「人間のフィードバックによる強化学習」（RLHF : Reinforcement Learning from Human Feedback）と呼ばれます。

専門家ではなくても使えるAIが爆誕！

これまで

- **専門家**が専用の仕組みを設計・開発して試行錯誤
- **専門家**が大量のデータを整備してプログラムで調整
- でき上がった仕組みの調整は**専門家**によるコツが必要

これから

- 文章で指示できるので**誰でも**試行錯誤できる
- 仕組みの調整もある程度は**文章**で指示できる※1
- 追加の情報も**文章**で与えればそれに基づいた処理ができる※2

文章で指示できるので 誰でもAIを活用できるように！

※1 機械学習やファインチューニングでLLMを調整するのではなく、文章で指示して調整することを「インコンテクストラニング」（In-Context Learning）と呼びます。

※2 LLMの性能向上により、LLMの機械学習やファインチューニングで使っていない情報（たとえば最新の情報や社内情報）も、インコンテクストラニングで与えるとそれに基づく回答ができるようになりました。インコンテクストラニングは機械学習やファインチューニングよりも簡単なので、外部のデータベースから情報検索してインコンテクストラニングでLLMに与えて回答生成させる「RAG」（Retrieval-Augmented Generation）とよばれる手法が注目されています。

第6章 まとめ

- 文章生成の仕組みが大きく進化した
- 大量の情報から得られた知識に基づき 人間の質問に答えられる
- 公開されている情報は知っているので 汎用的に使える
- 会話の往復ができるようになったので 関連情報を伝えたり 複雑な指示も出せる

ひと昔前のSFみたいなことができるようになった！

最終章では進化したAIのこれからについて

(休憩) 一緒に働く仲間を探しています！

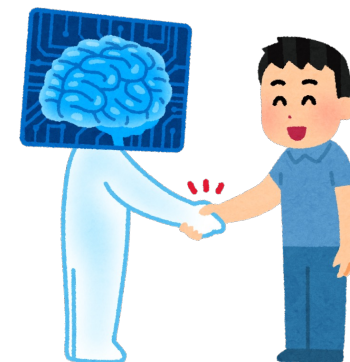
株式会社アイアクトのAIの事業部では、事業拡大のため仲間を探しています！

- 企業向けパッケージソフトの企画やマネジメントや開発リーダーの経験があり、クラウドやAIの知識・経験はないけどこれから学びたいと考えている方
- カスタマーサクセス/サポートの経験があり、お客様のご支援が好きで、AIやプログラミングの知識・経験はないけどこれから学びたいと思ってる方
- AIそのものの研究開発よりも、実用的になったAIの要素技術を組み合わせて実際に世の中で役立つ形にすることに興味のあるエンジニアの方



X@segavvy

ご興味がありましたら、まずは会社に関係なく軽くお話ししましょう。
個人のX(Twitter)のDMかメール segavvy@yahoo.co.jp でお声がけください！



第7章 AIのこれから

AIの動向や活用方法、課題など

できることが増えていく

1年前

- 質問に答える
- 要約する
- 翻訳する
- プログラムを書く

他の仕組みと連携

- 画像や動画を作る
- 聞く・しゃべる
- 画像や動画を観る
- 資料を読む・作る
- プログラムを書いて動かして集計・分析

他のシステム・サービスと連携※

- ネットの検索
- 評判のよいお店の調査や予約
電車や飛行機の予約
なくなった日用品の注文
家電の制御 etc.
- 社内資料の検索
チャット履歴を要約
Web会議の議事録作成
売上データの集計
必要資材の発注 etc.

※ この例のように、AI以外のものと連携することでもAIのできることが増やせます。AIの提供側は自社のAIでできることを社外の開発者に増やしてもらえよう、外部のシステム・サービスと連携するための接続口となる「API」（Application Programming Interface）を設けています。開発者はこれを使うことでAIと自分のシステム・サービスを連携させることができます。

また、AIの提供側は、そのような連携の仕組みがたくさん市場に流通するように、配布・販売するためのお店（マーケットプレイス、ストア）の整備を進めていくものと思われます。これにより、開発者は作った連携の仕組みをお店で「プラグイン」（Microsoft社の場合は「Copilot」という呼び名になる模様）として配布・販売できるようになり、AIの利用者はお店に必要なプラグインを選ぶことで自分が使うAIのできることを簡単に増やせるようになるのではないかと私は想像しています。

判断も任せられる※ようになっていく

人間の依頼の例	AIの内部の処理の例
わかりやすいAIの入門記事を教えて	1. 保有知識に情報がない → ネット検索が必要と判断 2. ネットを検索して ヒットしたページを取捨選択して要約して回答
社内の勤怠システムがエラーで使えないから直して	1. 保有知識に情報がない → 勤怠システムのマニュアルが必要と判断 2. 社内の文書管理システムを検索して マニュアルを確認し原因候補を抽出 3. 原因に該当するか判断できない → システムのログの分析が必要と判断 4. 該当システムのログを取得して分析 5. 原因が特定できたらマニュアルの対処方法を回答
事業部の来年度の予算案を作って	1. 保有知識に予算策定の流れはあるが情報がない → 情報取得が必要と判断 2. 中期経営計画の資料を検索して取得し 内容や来年度の位置付けを把握 3. 今年度の実績資料を検索して取得し 年度末の着地を予想 4. 今年度の達成・未達要因の分析 → 失敗案件の原因分析が必要と判断 5. 主要な案件資料を検索して分析 → 外部環境の変化の情報が必要と判断 6. ネットで市場概況を検索して分析 → ……

※AIが判断して次の行動を決め、さらにその結果を判断して次の行動を決めていくようなアプローチを「ReAct」(REasoning and ACTing)と呼びます。これまでのITの自動化のほとんどは、その処理の途中に高度な判断はありません。そのような高度な判断は人間による手作業になっていました。今後はこの部分もAIに任せられるようになっていくものと思われます。

AIのオススメの使い方



まだの方は「便利な人」として付き合い始めてみましょう！

日常的に使えば コツもわかり 苦手なこともわかり
新機能にもすぐ気づき 自然にAIの動向が把握できる※

※個人の意見です！

現在のAIの課題

知識はあるが丸暗記はしていない

知識はあっても一字一句を覚えている訳ではない

重要なものは出典元を確認するなど裏取りが欠かせない

誤り・偏見などの排除は不完全

それらしく答えるが基本的にはネットやSNSと同じ

内容が正しいかそのまま使えるかは従来通りの判断が必要

特殊用途には個別調整が必要

特殊な知識に基づく判断などは従来同様の調整が必要

仕組みの調整には質の良いデータとAIの専門家が必須

膨大なエネルギー消費

利用や調整に使う膨大な電力消費は環境破壊につながる

仕組みの小型化や専用のハード開発の取り組みが急務

他にも課題は山積：生成物の著作権の扱い、品質・信頼性の維持の方法、悪用の防止、AIの判断の透明性の確保、AI提供側への力の集中、etc.

利用側はネットやSNSを使うのと同じ警戒感を！
提供側は技術だけに踊らされない社会実装を！

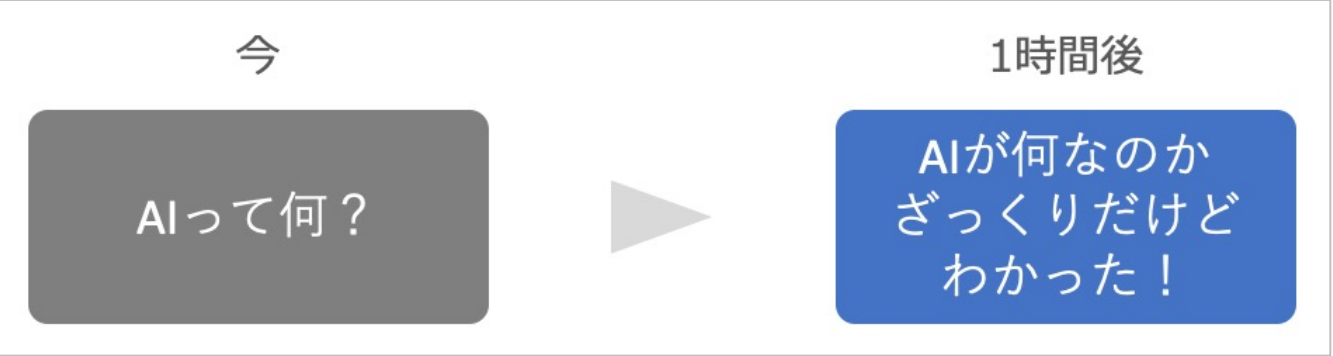
第7章 まとめ

- 文章で指示できるようになったAIは
音声・動画・画像なども理解できるようになり
さまざまなシステム・サービスとも連携できるようになる
- これまで人間がやっていた判断も任せられるようになっていく
- 「便利な人」として付き合いはじめることをオススメ
- ただし 多くの課題があることを理解した上での活用を！

おわりに

目次	超まとめ
第1章 AIはなぜ人間みたいなことができるのか？	「AI」は人間のマネをする仕組みのこと
第2章 脳はすごい	脳をマネした仕組みは手書き文字の認識などができる
第3章 伝わりやすさと境界の決め方	仕組みの調整作業はコンピューターに任せられる
第4章 細胞増やすだけではダメだった	脳の仕組みを工夫することでできることが増やせる
第5章 時間も手間もお金もかかる	仕組みの流用や組み合わせで効率的にできることが増やせる
第6章 文章生成の大規模化による進化	AIへ文章で指示が出せるようになった
第7章 AIのこれから	課題を把握しつつ まずは使ってみましょう！

「はじめに」より



**このお話がAIの理解の
一助になれば幸いです**

**お時間をいただき
ありがとうございました！**