



第63回

# AI(人工知能)

AI (Artificial Intelligence: 人工知能) は、2022年11月にChatGPTをアメリカの企業であるOpenAIが発表したことで、今まで以上に社会的な注目を集めることとなりました。AIに興味・関心を抱き、大学で学んでみたいと考える高校生も増えたのではないのでしょうか。ただ、AIに関する研究は非常に多岐にわたり、全体像が捉えにくいものとなっています。

そこで、今回の記事では、「AI研究の現在」というテーマについて、人工知能学会所属の先生方に「AIマップβ 2.0」を基に解説いただきます。そのうえで、「レコメンド機能」「音声アシスタント」「ChatGPT」という、3つのキーワードと、AIの研究との関わりについて紹介していきます。

## Contents

- ◆概説 AI研究の現在  
人工知能学会 ..... p62  
環境を知覚し環境に働きかける仕組みを持つAI  
今後は技術だけでなく  
人間との関係も研究課題に
- ◆レコメンド機能  
東京都立大学 高間康史 教授 ..... p66  
本人が気づいていない「好み」を予測する研究や  
譲れない点を踏まえて行動変容を促す  
推薦(レコメンド)の研究へ
- ◆音声アシスタント  
東京工業大学 篠田浩一 教授 ..... p68  
機械学習の成果で実用化が進む音声アシスタント  
耐雑音性の向上やフェイク防止の研究に期待
- ◆ChatGPT  
東北大学 鈴木潤 教授 ..... p70  
産業革命に匹敵する社会変革を生み出す生成AI  
基盤技術、今後の学びや生き方との関わりを解説

概説

AI研究の現在

# 環境を知覚し環境に働きかける仕組みを持つAI 今後は技術だけでなく人間との関係も研究課題に

## 人工知能学会

Point

- AIの概念は時代と共に変化
- 知的活動を分析し技術を連動させAIを実現
- 基礎から社会実装までAI研究は広範囲



市瀬 龍太郎 教授   砂川 英一 参事   本村 陽一 首席研究員

今回は、人工知能学会に所属し、AIマップの作成と普及に携わっている、

- ・東京工業大学 工学院 市瀬 龍太郎 教授
- ・(株)東芝 研究開発センター 研究企画統括部 砂川 英一 参事
- ・産業技術総合研究所 人工知能研究センター 本村 陽一 首席研究員

の3名に、「AI研究の現在」をテーマにお話をうかがった。

### 行動主体として「エージェント」を設定し 環境からの情報を基に行動を決定するのがAI

「AI (Artificial Intelligence : 人工知能)」に明確な定義が存在しているわけではないが、東京工業大学工学院の市瀬龍太郎教授は、「人間と同じ知的作業をする機械を工学的に実現したもの」と表現する。人間の知的作業の中身を扱う学問には、脳神経科学や認知心理学などがあるが、AIはそうした知見も生かしつつ、機械で同様な作業を実現させることをめざしている。

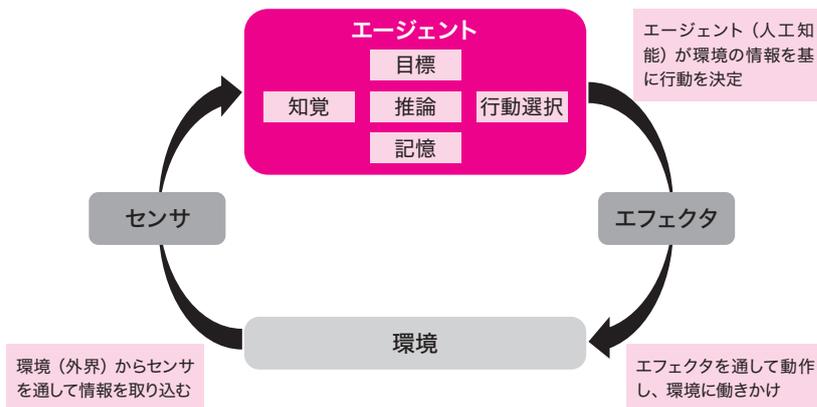
知的作業を行うためには、外界を認識し、行動を決定する仕組みが必要だ。人間は自分のいる環境を目や耳、鼻、皮膚などによって知覚し、口や手、足などを通して環境に働きかける。この一連の動作が知的作業であり、それを機械で実現するために考えられたのが「エージェント」という概念だ<図表1>。

「エージェントは、各種センサを使って環境から情報を取得し、その情報を基に行動を決定し、エフェクタを通じて環

境に働きかけます。人間という行動主体が身体と心理からできているように、行動主体としてのエージェント＝AIは、ハードウェアとソフトウェアがセットになったものといえます」(市瀬教授)

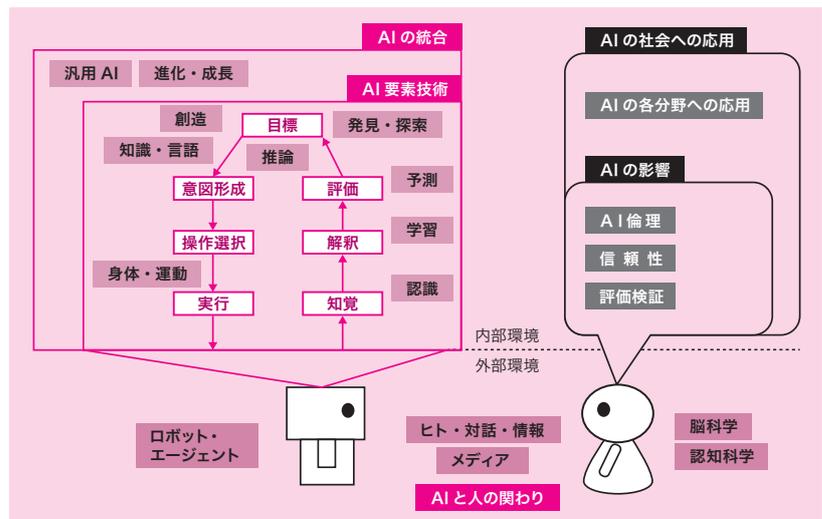
環境からの情報に応じて環境に何らかの働きかけを行う機械は多いが、温度を感知して電流を遮断するサーモスタットや、光量を感知して電灯を点灯・消灯する機械はもはやAIとは呼ばれない。AIと呼ぶ以上は、知的作業が必要だ。

図表1 エージェントとは



※市瀬教授提供資料を基に河合塾作成

図表2 知的作業(知能活動)のフロー



※人工知能学会「AIマップβ 2.0 (2023年5月版)」([https://www.ai-gakkai.or.jp/pdf/aimap/AIMap\\_JP\\_20230510.pdf](https://www.ai-gakkai.or.jp/pdf/aimap/AIMap_JP_20230510.pdf))より

## 人間の知的活動を7段階に分け ステップごとに異なる技術で対応

では、エージェントの行う知的作業とはどのようなものを指すのか。市瀬教授によれば、何を知的と捉えるかは時代と共に変わってきているのだという。

「かつては計算という高度な知的作業を行う電子計算機もAIの1つくらいに思われていましたが、現在では電子計算機そのものをAIと呼ぶ人はいません。同様に現在注目されているChatGPTな

どの生成AIも、やがて知的作業を行っているわけではないと思われるようになるかもしれません」(市瀬教授)

人間の知的作業の一連の流れに関しては、アメリカの認知科学者D.A.ノーマン氏が1988年に提唱した「行為の7段階モデル」がある。環境を①「知覚」し、②「解釈」・③「評価」を行った上で、④「目標」に照らして⑤「意図形成」を行い、⑥「操作選択」して、⑦「実行」するというものだ<図表2>。

たとえば、知り合いに会って挨拶するのは、人間にとっては単純な行為のように思えるが、AIに行わせようとするとさまざまな技術を総動員しなければならない。

知り合いだと①「知覚」するには、画像認識や音声認識などの「認識」技術が必要であり、相手の行動や状態を②「解釈」し③「評価」するには、その意図や背景を理解するなど、「認識」とは異なる要素技術が必要だ。④「目標」は人間が設定するにしても、それに沿って⑤「意図形成」や⑥「操作選択」を行うには、学習データなどから抽象化した知識などを基に「推論」する技術が必要だし、実際に握手や会釈をする⑦「実行」段階では、音声や動作などに関するハードウェアを制御する技術が欠かせない。

「これらがすべてうまく噛み合っはじめて、知り合いに挨拶することができるAIができるわけです。ですからAIの実現には、広範囲にわたる細かな要素技術が必要で、それぞれの技術に磨きをかける研究を進める一方で、最近では、AIと人間の関わり、AIが社会に及ぼす影響などについても研究すべき段階に来ているといっているといいでしょう」(市瀬教授)

## 基礎研究から応用研究へと連続性を持ち 実世界からサイバー空間までカバーするAI研究

このようにAI研究は、要素技術の研究から、それを社会に実装するための研究、さらには社会とAIの関係を解明するための研究にいたるまで多岐にわたる。そこで、人工知能学会作成の「AIマップβ 2.0 (2023年5月版)」にある「AI技術マップE『AI研究の現在』」を参考にAI研究の広がりを見てみたい<図表3>。

このマップは、AI研究におけるキーワードの関連性を示したもので、横軸方向は、基礎研究から応用研究への流れを示し、最も左側の「基礎・理論」から、右側方向に広がっていく。また縦軸方向は、上に位置している研究キーワードほど実世界に近く、下に位置している研究キーワードほどサイバー空間に近いことを示している。

以下、各層の研究についてのあらましと、近年の研究動向などを紹介する。

### ■第1層 基礎・理論

AIを実現するための基本的な理論の研究がここにマッピングされている。コンピュータのアルゴリズムだけでなく、その基礎となる数理的な研究や、認知科学などの研究がここに属している。

「AIの基礎・理論の研究には長い歴史がありますが、最近になって新しい流れも見えてきました。これまでのAI研究は、将棋を強くするAIや、自動車の自動運転を実現するAIなど、特定の分野で力を発揮できるような方向に発展してきました。深層学習(ディープラーニング)の成功を契機に、いろいろな分野で活躍できる『汎

用人工知能 (Artificial General Intelligence)』に期待が集まっていますが、従来の特化型AIと同じアプローチでは上手くいかないこともわかってきました。そこで、どのようなアプローチなら汎用人工知能が可能なのかという研究がはじまっています」(市瀬教授)

■第2層 ～ AI研究に共通する技術・概念～

この層には、基礎・理論ではないものの、AI研究で共通している技術や概念がマッピングされている。

「画像音声メディア処理」の中心には、画像や音声の認識技術があり、深層学習によってかなり認識精度が上がったため、実用化の幅も広がった。最近では、人間が普通に行っている、画像や音声、文字などを組み合わせながら行う知的活動をAIで実現しようとする取り組みもある。「マルチモーダル処理」はそうした複数のメディアの利用を考える研究として注目を集めている分野だ。

「知識の利用と共有」は、蓄えられた知識の活用方法を模索する研究領域で、膨大なデータから価値のある情報を見出す「データマイニング」などはすでにさまざま

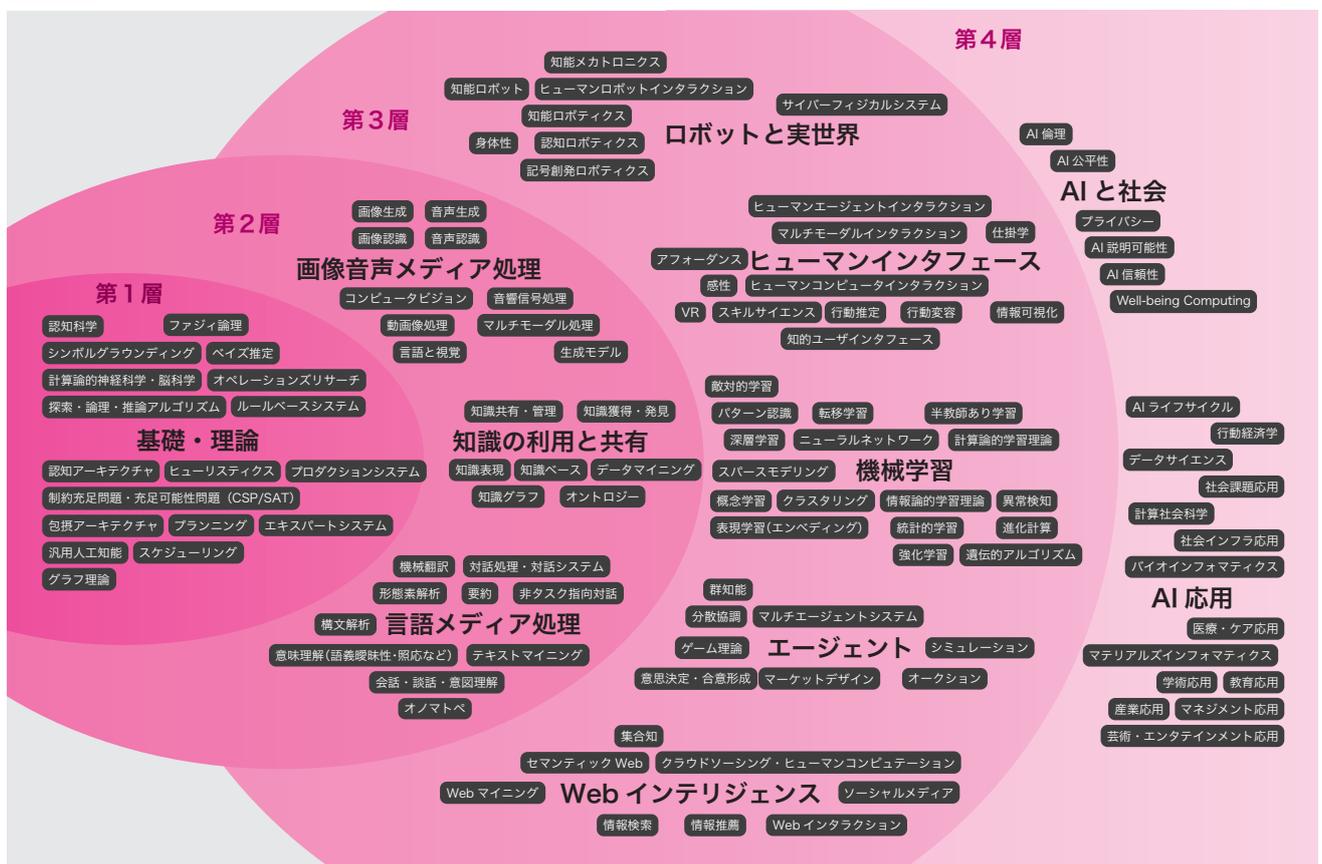
な分野で実用化されている。

(株)東芝研究開発センター研究企画統括部の砂川英一参事は、「ここは、意味や知識の“つながり”を扱う分野と考えていいと思います。人間は、新しく得た知識を他と合わせて考えることができますが、その要素となる概念も含めてコンピュータが知識を扱うことは難しく、そうした課題を解決するのに必要な技術の研究開発が盛んに行われています。ChatGPTでは、この領域や『言語メディア処理』『機械学習』の研究成果が組み合わせられて使われています」と語る。

「言語メディア処理」は、音声ではなくテキストになった言語を扱う技術領域だ。音声を変換するのは「画像音声メディア処理」の技術だが、テキストになってからは「言語メディア処理」の研究対象になる。

「たとえば『対話処理・対話システム』の研究は、空腹な人にレストランを紹介するなど、対話の目的や流れに沿った応答ができるようなAIを実現することをめざした研究といえます」(砂川参事)

図表3 AI研究の現在



※人工知能学会「AIマップβ 2.0 (2023年5月版)」(https://www.ai-gakkai.or.jp/pdf/aimap/AIMap\_JP\_20230510.pdf) を基に河合塾作成

### ■第3層 現在最もホットな研究領域

この層は、現在最も盛んに研究が行われている領域で、深層学習に代表される「機械学習」を中心にして、実世界との関わりが強い領域から、サイバー空間との関わりが強い領域までがマッピングされている。

「ロボットと実世界」は、産業ロボットのような精密な制御だけではなく、AIによって人間と同じように考えて行動する機械の実現が大きな目標の1つだ。

産業技術総合研究所人工知能研究センターの本村陽一首席研究員は、「産業ロボットに深層学習を導入した結果、人間の動きを三次元的に学習することが可能になり、AI技術によって、従来以上に産業ロボットの有用性が高まりつつあります」と話す。

「単純作業を繰り返すロボットは、開発コストを低くおさえられます。一方で、ものづくりの現場は複雑化しており、複数のタスクをこなせるロボットであれば、どの製造工程でも使えます。そのため、現在は高度なタスクを後から学んでこなせるロボットの開発や、その延長線上として工場自体を知的にする動きも見られます」(砂川参事)

「ヒューマンインタフェース」は、機械と人間との接点を研究する領域であり、ヒューマンエラーが出にくい動作の実現などに貢献している。

「機械学習」の領域では、深層学習の研究が特に盛んである。深層学習の基になる技術は、1940年代から技術開発が行われてきたが、画像認識の分野で非常に高いパフォーマンスが出たことから一気に注目され、現在ではさまざまな分野で深層学習が使われるようになってきている。

「深層学習の発展には、大量のデータが蓄積されてアルゴリズムの検証に使えるようになったことや、多階層の並列計算が可能なGPU（Graphics Processing Unit：画像処理装置）というハードウェアが開発されたことなどが寄与しています。現在は、こうしていろいろな研究成果が上手く噛み合ってきた、非常にいい時代だといえます」(本村首席研究員)

「エージェント」の領域は、冒頭で説明した行動主体としてのAIを、実際にコンピュータ上で振る舞わせてみて、どのような結果になるのかを考察する研究領域といつてよい。たとえば株価動向を調べる際に、人間のモデルとして仮想のエージェントを大量に作って投資行動を行わせ、株価がどう動くかを検証するといったような

研究が行われている。

「Webインテリジェンス」の領域は、サイバー空間を扱う技術で、購入履歴からお勧め商品を提示するような技術などが代表的だ。

### ■第4層 重要になり得る研究領域

「AIと社会」の研究領域は、AIが社会実装されはじめたからこそ立ち上がってきた領域といえる。

「倫理的なAIをどう作るか、社会性を持ったAIをどう作るかなどが盛んに議論されています。偏ったデータで学習させれば、偏った回答しか出てきません。公平な社会の実現や、プライバシーにどう配慮するかなど、今後考えなければならない研究テーマが数多く含まれています」(市瀬教授)

「AI応用」は、医療や教育など、特定の領域で力を発揮するAI開発に取り組む領域で、実用化が進む一方で、より高いパフォーマンスを発揮するための研究開発に力が入れている。

### ==== 複数分野を横断的に見ることがポイント AI研究は大学院からでも遅くはない!?

このような広がりを見せるAIの研究に携わるには、どのような学部・学科を選択すればよいのだろうか。

AIの中身を深く知ろうとするならば、プログラミングの基礎や数理統計などを学ぶことが必要で、そういった科目が充実している工学系、情報系の学部・学科がおすすだ。しかしAIを応用することを考えれば、応用する分野に関する基礎知識が必要になってくる。

「AI研究のポイントは分野横断型の視点にあり、課題を俯瞰したり、客観化したりするために自分の思考をリフレームする必要があります。個人的には学部で何らかの専門を修めた後、大学院からAIの勉強をはじめても十分に間に合うと思っています」(本村首席研究員)

人間の持つ知能を機械で実現することを夢見るAI研究は、究極的には人間とは何かという根源的な疑問に遭遇する。その意味で哲学や心理学、生物学などとの関連も深い。工学が強く関与する確率・統計やデータサイエンス、それらの基礎をなす数理学などは、いずれもAI研究の基盤となる。こうした分野から入るものよし、応用分野から入ってもよいということなのだろう。

## レコメンド機能

# 本人が気づいていない「好み」を予測する研究や 譲れない点を踏まえて行動変容を促す推薦の研究へ

東京都立大学 システムデザイン学部 情報科学科 高間 康史 教授

## Point

- 情報過多の時代ゆえに必要なになった情報推薦
- 「好き」を予測する2つのアプローチ
- 他のAI技術との連携やプライバシー配慮が次のテーマ



## 必要な情報を手に入れるための労力を 計算機に担わせる技術がレコメンド機能

レコメンドは「推薦する」という意味の英語ですが、AIの世界でレコメンド機能といえば、基本的にはユーザーが好むもの、関心を持つものを計算機が推測して推薦する機能を意味します。かつては情報を入手すること自体に大変な労力が必要でした。現在はインターネットの普及により、誰でも手軽に情報を入手できるようになりましたが、逆に、今度は膨大な情報の中から必要なものを選び出すことが難しくなりました。その労力を機械に代替させようと生まれたのがレコメンド機能です。

Webページに表示される広告の多くはレコメンド機能の産物です。ほかにもAmazonなどのオンラインショッピング、YouTubeなどの動画サイト、音楽のサブスク、ニュースアプリなど、幅広いサービスでレコメンド機能が使われています。大まかに条件を指定すれば、自分で探さなくても表示してくれる仕組み、人によって異なる表示がされる部分には、基本的にレコメンド機能の技術が使われているといいでしょう。

レコメンド機能の研究は、1990年代くらいから本格化しました。母体になったのは、1950年代から研究されてきた情報検索です。当初は膨大な情報の中から、必要なものを探し出すことが目的で、その技術はインターネットのサーチエンジンとして開花します。やがて不要な情報を排除してから必要な情報だけを提示する「情報フィルタリング」の考え方が生まれ、オンラインショッピング

の普及もあってレコメンド機能の研究が盛んに行われるようになりました。現在では、テキストだけでなく、画像や音楽、動画などあらゆるものがネットに上がっているため、レコメンドの対象も多岐にわたっています。

## 「内容ベースフィルタリング」から ロコミを模した「協調フィルタリング」へ

レコメンド技術のコアは、ユーザーの「好き」を予測することであり、大きく2つのアプローチがあります<図表>。

1つは、「内容ベースフィルタリング」といい、対象となるアイテム（商品や映画、曲など推薦するもの）の属性や特徴に基づいて、推薦するアイテムを決める方法です。あるアイテムが好きなら同じような特徴を持ったアイテムも好きだろうという発想です。扱っているテーマや著書などから文献を探してくる情報検索の要素技術がそのまま適用されているといいでしょう。

ユーザーが好きな属性や特徴を知る方法は、①ユーザーに「関心のあるものを教えてください」と直接聞く方法、②そのユーザーの閲覧履歴、購入履歴から共通する属性や特徴を推定する方法の2種類があります。計算機では、その人の関心をベクトルで表現します。それは「ユーザープロファイル」と呼ばれます。アイテムの属性や特徴もベクトルで表現できるため、両方のベクトルを比較して、ユーザープロファイルに似たベクトルを持ったアイテムを推薦するわけです。

もう1つは、レコメンド機能特有の「協調フィルタリ

ング」という方法です。ユーザーの関心やアイテムの特徴などを知らなくても、趣味が似ている人は同じようなものを気に入るだろうという発想が出発点で、好みや似ているユーザーが好むアイテムを推薦します。Amazonの「この商品を買った人は、こんな商品も買っています」という機能はこれに当たります。ネット上で他の人が何を購入（選択）したかという情報さえあれば、気に入るアイテムを推薦することができます。

技術的には、「評価値行列」を入力として使います。評価値行列はユーザーを行、扱っているアイテムを列とする行列で、購入すればどこかの行と列の交点に評価値

が入ります。評価値が入っていないアイテムを、もしそのユーザーが評価するとしたらどんな評価値が入るだろうかという評価値予測を行うのが、協調フィルタリングの基本的な機能です。最初は口コミがそうであるように、同じアイテムにつけた評価値が近ければ趣味が同じユーザーだと考え、同じ趣味を持つユーザーが高評価をつけている別のアイテムを推薦するという方法でした。

しかし1人のユーザーが購入するアイテムは限られており、行列はサイズの割に情報が乏しいのが実状です。そこで予測精度を高めるため、評価値行列を圧縮することを考えました。行列分解という線型代数の計算手法や、ニューラルネットワークを使った非線形の計算などを駆使して、予測精度を高める研究が続けられています。

### 「好き」ではないが行動変容を促す推薦や ユーザーの情報を使わない推薦の研究へ

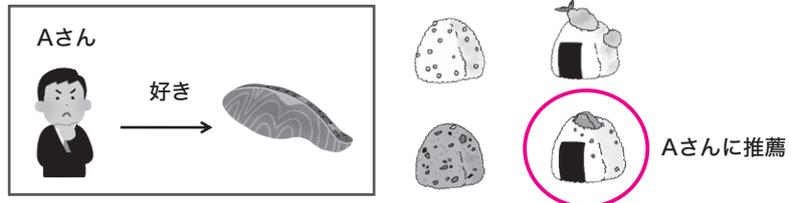
レコメンド機能の研究は、ユーザーの好みのアイテムを推薦する部分では、ほとんど差がなくなっており、さまざまな場面で実用化されています。今後は、以下のような方向に発展させる研究が模索されています。

たとえば、行動変容を促す推薦です。感染リスクを避ける行動や、ダイエット中の食事などを推薦する際には、

#### 図表 情報推薦 2つの考え方

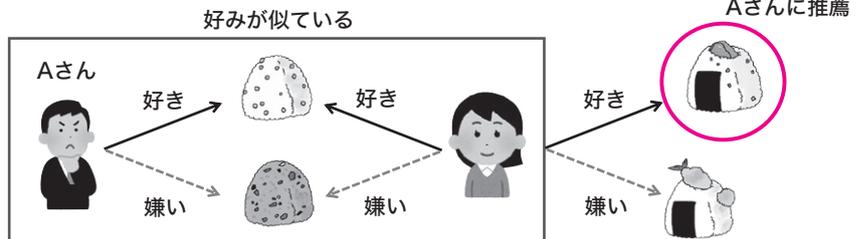
##### (1) 内容ベースフィルタリング

鮭が好き → 「鮭のおむすび」を推薦



##### (2) 協調フィルタリング

「好みや似ている人が好きなもの」を推薦



※高間教授提供資料を基に河合塾作成

その人が好むこと（もの）を勧めても意味がないですが、嫌いなものを勧めても拒否されます。そのため、そのユーザーが譲れない点を探り、妥協点を見つけて適切に行動してもらえるようなレコメンド機能を実現できないかという研究が必要になってきます。

また、自分自身では絶対に選択しないが、レコメンドに従ったら気に入ったといった、自分自身が気づいていない好みを探して推薦する「セレンディピティ」のある推薦を研究している人もいます。さらに、「そろそろ雨が降るので洗濯物を取り込んで」「薬を飲む時間だよ」など、日常生活の中で対話や行動を共にするロボットに推薦させるような、推薦タイミングも含めて他のAI技術と連携したレコメンド機能の研究も期待されています。

なお、レコメンド機能は、実際のユーザーの好みの情報や検索履歴、購入履歴などを使いますから、プライバシーの問題が発生します。その懸念を払拭することは、レコメンド機能の研究でも重要になってきました。情報を秘匿したままデータを扱うセキュアな計算方法や、できるだけ情報を集めないで推薦する方法の研究が行われているほか、私の研究室でも、実ユーザーの評価値行列と同じような意味を持つ、人工的な評価値行列を作ることができないかといった研究を始めています。

## 音声アシスタント

# 機械学習の成果で実用化が進む音声アシスタント 耐雑音性の向上やフェイク防止の研究に期待

東京工業大学 情報理工学院 篠田 浩一 教授

### Point

- 音声アシスタントは複数の技術の集合体
- 雑音の中での会話を認識できる技術の開発をめざす
- 嘘を見破る技術の開発が重要になってきている



### 別々の研究分野が音声アシスタントに結実

SiriやAlexaに代表される、スマートフォンやスマートスピーカーに向かって話しかけると、質問に答えてくれたり、機器を適切に操作してくれたりする音声アシスタント。近年になって技術の実用化が進み、さまざまな製品に搭載されるようになってきました。

この技術は、①環境中にあふれている音の中から人の音声を認識してテキスト化し（音声認識）、②テキストの内容を解析して、適切な対応を行い（自然言語処理）、③音声での応答が必要な場合に音声を出力（音声合成）するという3段階のステップに分けることができます<図表>。

それぞれの技術は、元々別の研究分野でした。音声認識は空気中の音波（音響信号）からどうやって人の音声を切り出して文字にするかという技術、自然言語処理は人が読み書きできる自然言語（テキスト）をコンピュータで扱う技術、音声合成は人の声聞こえる音声を機械に出させる技術として発展してきました。しかし、ニューラルネットワークを使った機械学習の技術がハードウェア・ソフトウェアの進歩によって高性能になったことで、どの技術もそれぞれ機械学習を取り入れるようになり、研究に必要な技術という点では①～③の間の差がなくなってきました。そのため、3つを区別せずにまとめて機械学習を行う研究も出てきています。

ただ、それぞれの技術がめざす方向は異なります。ここでは「音声」に関わる①音声認識と③音声合成につい

て、もう少し詳しく紹介します。

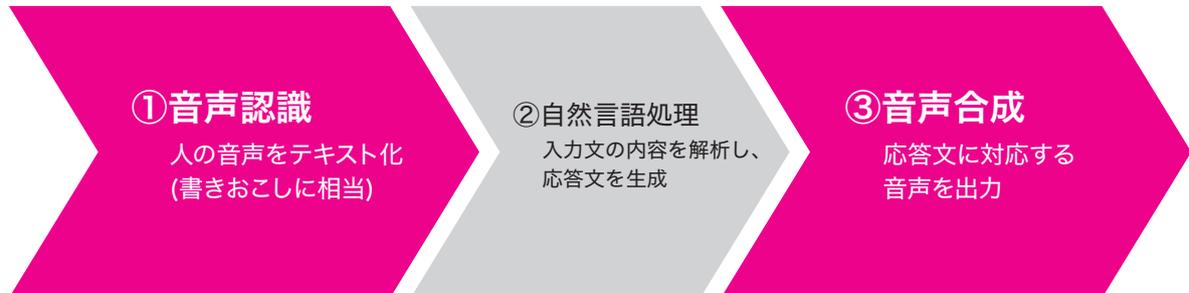
### 音声認識では雑音との切り分けに課題

音声認識技術は、音から、人の発する音声、つまり文字列に相当する音素列を取り出す技術といえます。そもそもマイクに入ってくる音は人の声も雑音も混ざった状態の音波として入ってきます。しかも、咳払いや笑い声、泣き声など文字にならない情報がたくさん含まれていますし、テキストでは同じように表記される音声も、人によって、状況によって、感情的になったり、抑揚が違ったりさまざまな波形として入ってきます。その中から書き起こしに相当する部分だけを正確に認識しなければなりません。

そのためのアプローチの1つが機械学習です。実際に人が話した音声データと、それを書き起こしたテキストを用意して学習させ、音声データから自動的にテキストに書き起こす処理を行う機械を作ろうというわけです。実験室レベルでも700～1,000時間分の音声データ（と書き起こしデータ）が必要ですが、実用化にはさらに膨大なデータが必要です。それでも画像や動画データに比べれば少ない計算処理ですむ利点はあります。

深層学習の導入でかなり認識率は高くなりましたが、それでも課題はあります。データが多ければ多いほど認識率は高くなりますが、話者が少ない言語の場合は、音声データの収集には限界があります。また、書き起こしデータはどうしても多くの人手に頼るためミスを排除できず、データの質をどう担保するかが問題です。

図表 音声アシスタントの3ステップ



本記事では、①・③を中心に紹介

※取材内容を基に河合塾作成

耐雑音性も大きな課題といえます。人は雑音の中でも人の声だけを認識できますが、機械にとっては困難です。雑音込みで学習させるにしても、雑音は多種多様ですべてを学習させることは不可能です。

さらに、話した瞬間にリアルタイムで文字に変換できなければ使い物になりません。現在の音声アシスタントは、手元のデバイスである程度処理してから、さらにクラウドで処理したものが出力されている場合が多いですが、小さな端末だけで処理から出力まで行うためにはよりデバイスへの負担が軽いアルゴリズムを開発する必要があります。

### 音声合成では真贋を見分ける技術が重要に

音声合成技術の研究は、発声器官を模した機械づくりから出発しました。具体的には、どの筋肉をどう動かせばどんな声が出るかといった研究が行われていました。しかし、現在では音声データを使ったデータ駆動型のアプローチが主流です。つまり、ニューラルネットワークを使って、テキストデータから音声データを得るという操作です。

ただし、テキストになった段階で、話者の性別や感情、抑揚などの情報がすべて抜け落ちています。ですから、人の音声として出力する場合は、それらの情報を適切に付与する技術が不可欠です。現在はモノログや丁寧なやり取りなら、違和感なく聞けるくらいになっていますが、まだカジュアルな会話を楽しむレベルにはありません。もっとも、「機械にはカジュアルに話してほしくない」と感じる人もいるかもしれません。

最近出てきた課題は、合成された音声の真贋を見分けるFake Detectionと呼ばれる技術の開発です。音声合成も含め、テキストや画像、動画を生成するAIを使えば、偽物を簡単に作ることができます。世の中に悪影響を与えないためにも、この分野の技術開発がますます重要になっています。

### 音声は知能を考える入口!?

現在、実験室のような理想的な環境下であれば、音声認識も音声合成も、ほぼ「人間並み」の性能が出せるようになっていきます。人も聞き間違いや言い間違いをしたり、アクセントが変になったりしますから、その程度の水準に達したという意味です。

しかし、まだまだ課題はあり、音声認識の研究でいえば、耐雑音性を高めること、とりわけパーティーでの会話のように、さまざまな音声が行き交う中で複数人の音声の認識率を上げる研究や、音声信号に含まれる情報の中から、文字にする部分、感情を表す部分、話者の特徴を表す部分などを分解して、より少ないパラメータで音声認識ができるような技術開発に向けた研究が期待されています。

現在はChatGPTの登場で文字ベースのAIに注目が集まっています。しかし、人類は文字を使う前から音声でコミュニケーションをして知的な活動を行い、文字さえも発明したわけですから、AI研究の目標である知能の把握には、知能の入口・出口としての音声を含めて探求することが重要で、そこに知能を解明する鍵が隠されているのだと考えています。

## ChatGPT

# 産業革命に匹敵する社会変革を生み出す生成AI 基盤技術、今後の学びや生き方との関わりを解説

東北大学 大学院情報科学研究科 鈴木 潤 教授

## Point

- ChatGPTの基盤技術は単語の出現確率を予測する言語モデル
- 人間のニューロンネットワークを模倣したモデルが貢献
- 人間が生み出したものを模倣し、生成している



## 言葉の意味を理解しているのではなく 言語モデルによる確率計算で文章を生成

ChatGPTをはじめとする対話型文章生成AIは、対話形式の指示を受けつけ、その指示に適した文章を生成する文書生成器と捉えることができます。今回、ChatGPTを「個別のモデル」として例に挙げながら、この生成AIについてお話しします。

まず、ChatGPTは、人間と話すように自然に会話できるだけでなく、翻訳や要約、レポート・スピーチ原稿の作成、アイデア出しなど、人間が作成するのと変わらないレベルの文章を生成してくれるのですが、それを可能にしている技術について、簡単に紹介します。

ChatGPTをはじめとする対話型文書生成AIの基盤技術は、言語モデルと呼ばれるAI技術です。言語モデルは30～40年前から主に機械翻訳や音声認識などに活用するために研究されてきました。基本的な考え方は、ある言葉の次に出てくる単語の出現確率を予測するというものです<図表>。大量の文章を単語単位に分解し、2つ以上の単語で構成される文脈の後に来る、最も高い出現確率の単語を予測し、今度はその単語も加えた文脈の後に来る単語を予測して…ということを繰り返すことで文章の形にしていくわけです。

言語モデルは、単に出現確率の高い単語を計算して順に出力しているだけですから、文章や単語の中身を人間のように理解しているわけではありません。にもかかわらず、ChatGPTはまるで意味がわかっているかのよう

に確かな文章を生成することができます。それを可能にしたのがニューラルネットワークです。

## 確率精度を高めるために活用した ニューラルネットワークに意外な効果

ニューラルネットワークは、人間の脳内のニューロンのネットワークを模倣した数学モデルです。機械学習や深層学習に利用されており、ざっくりいえば、関数を近似するための装置として機能します。言語モデルは次の単語の確率を予測するわけですから、予測関数というものがあると仮定できます。真に求めたい関数がわからなくても、ニューラルネットワークを使えば、その関数を非常に高い精度で近似することができます。

かつては大量の文書のデータから次に出てくる単語の数を数え、そこから単純に確率を計算していましたが、優秀な関数近似法であるニューラルネットワークを使って大量の文章を学習させれば、次の単語の確率予測の精度が大幅に向上します。現在は、2017年に考案されたTransformerと呼ばれるニューラルネットワークが広く用いられており、ChatGPTはもちろん類似の多くの対話型文章生成AIでも採用されています。また、画像処理、音声処理、信号処理などでも広く使われています。

こうして言語モデルの確率精度を高めるためにニューラルネットワークを使ったわけですが、ChatGPTが成功した大きな要因は、実は高い近似精度ではなく、ニューラルネットワークが本来持っている特性にありました。

その特性とは、似ている、似ていないを判断して、似

ているものを勝手に学習してくれるという性質です。人間がそう設計したわけではなく、ニューラルネットワーク自体に最初から備わっている特性です。コンピュータ内部ではデータは数字の塊として処理されますが、その塊が似ているものを自動的に学習してくれるのです。

言語においては、字面ではまったく違うのに、何となく似たような意味の文はたくさんあります。その後

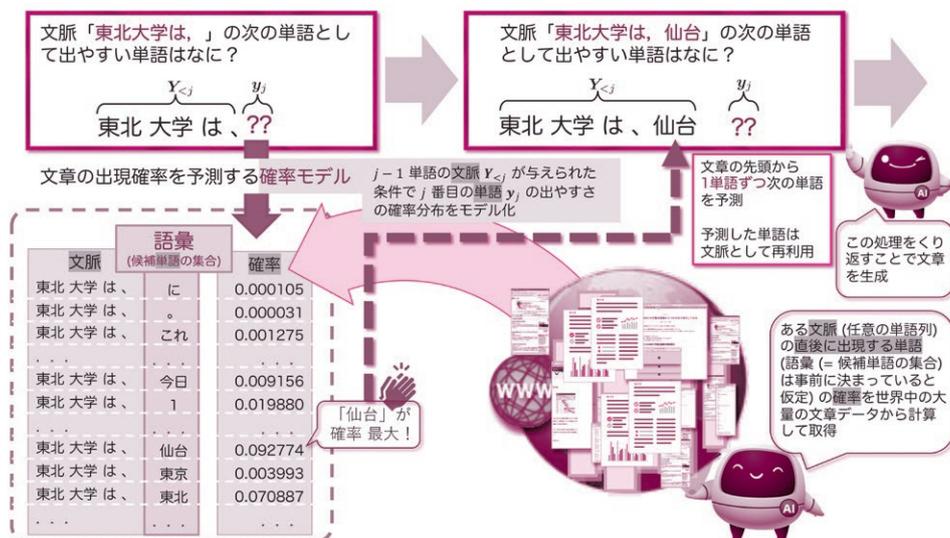
に続く言葉が同じであれば、その前の文脈もきっと同じような意味だろうというように、勝手に学習してくれるのです。全然違う文章が似ているとか似ていないとか、構文の形が同じとか違うとかを、言葉の意味ではなく、統計的な観点で勝手に学習していくわけです。

しかも学習データは、Web上のある時点までのすべてのテキスト、それも専門的な論文から日常会話に至るまであらゆる領域を網羅し、人間が一生かかっても読み切れない膨大な量のテキストです。ChatGPTに何を聞いても、ある程度答えてくれるのは、このニューラルネットワークの特性が非常に効果的に機能した結果だと考えることができます。

### 実用化には人間によるトレーニングも必要 社会や生活のあらゆる局面での活用を期待

もともと、こうしたAI技術だけでは、自然に対話しているような受け答えはできません。対話を可能にするには、こう聞かれたら、こう答えるというようなお手本が必要です。そのお手本の「こう聞かれたら」の部分をプロンプト(指示文)といい、ある程度の数のプロンプトとそのプロンプトに適切な回答文のペアを用意すれば、あとはAIが勝手に学習してくれます。自然な受け答えができるかどうかは、プロンプトとその回答文のペアを適切に学習させたかどうかにかかっており、それが成功し

図表 言語モデルの概要



※鈴木教授提供

たからこそChatGPTの能力に注目が集まったわけです。

もう1つ重要なことは、悪意や偏見に満ちた文章など、社会通念上不適切な文章を生成しないようにすることです。ここにも人間の関与が必要で、生成した文章が適切かどうかを点数づけることで、不適切な文を生成しないように学習させる必要があります。

このようにChatGPTも、人間がデータを与えてはじめて実用的になるということは覚えておく必要があります。大元のデータも元々は人間が作ったものです。その莫大な知識データをベースにしているために、一見賢そうに見えるものの、ChatGPTが生成する文章は、あくまでも人間の模倣という基本原理は押さえておく必要があります。

とはいえ、人間生活のなかで言葉が介在しない世界はほとんどありません。現在PCブラウザ版はテキストベースの受け答えしかないChatGPTですが、スマートフォン対応版は音声入力を受けつけるようになりました。さまざまな広がりを見せるChatGPTは、今後もありとあらゆる仕事や教育、生活に影響を与えるでしょう。産業革命以来の、革命的な変化が社会に起きることは間違いありません。ChatGPTを使わないという選択肢はなく、ChatGPTあるいはそれ以後の生成AIを使うことを前提とした上で、働き方や学び方、生活のあり方を新しくデザインしていく必要があるでしょう。