

数学的なアプローチで人工知能に挑む

情報工学研究院知能情報工学研究系 教授

本田 あおい(情知H3)



1 はじめに

人工知能の始まりと言われるチューリングテストをチューニング

が提唱した1950年から人工知能は日々進化し続けてきましたが、近年の深層学習をコア技術とする人工知能技術の発展の速度は驚異的といえます。特に2022年から2023年にかけての進化は著しく、Stable Diffusionを用いたAI画像はプロのイラストレーターに匹敵する精度で描かれ、ChatGPTはまるで人間のような対話を展開するなど、仕組みを理解していても、その性能には驚愕せざるを得ないという状況です。

このように高度に進化した人工知能技術の研究には、情報工学だけでなく様々な知見を集結してアプローチする必要があります。中でも数学の基盤に基づく数学的なアプローチは特に不可欠なものであると言えます。すでに、機械学習の分野では微積分や線形代数、統計学のみならず、確率微分方程式や関数空間論等の高度な数学がモデルや手法の開発に用いられています。

申し遅れましたが、私は情報工知能情報工学研究系に所属しています。本田あおいといいます。私は情報工知能情報工学学科の一期生で、古巣で教育に携われることに喜びを感じる毎日です。研究室では、数理的なアプローチによる人工知能の研究を行っています。深層学習モデルは優れた性能を持つ反面、その答えを出した理由や答えの導出プロセスを説明できないという弱点があります。これは、モデルの内部ロジック

がわからないためであり、このような「出力の導出プロセスが説明できない問題」はAIのブラックボックス問題と呼ばれています。私の研究室では、数学の道具を使って色々な方向からこのブラックボックス問題の解決に挑んでいます。

2 機械学習の解釈性

機械学習とは、人工知能(Artificial Intelligence)を実現するための技術(特にコンピュータを用いるもの)の総称です。現在の高度なAIはほとんどが深層学習ニューラルネットワークという機械学習技術が用いられており、そのため、人工知能、機械学習、深層学習、ニューラルネットワークは正確には意味に違いがあるものの、ほぼ同義で使われています。さて、多くの機械学習モデルはブラックボックスモデルであり、予測はうまくできても、なぜその予測をしたのかを説明することができません。これが問題につながる場合があります。有名な例では、米アマゾンが、開発を進めていたAI人事採用システムに女性の評価を不当に低くしていることが判明し、運用を打

ち切らざるを得なかったというものがあります。原因は学習に用いた過去10年分のデータが男性ばかりであったためとされています。ブラックボックスである限り、よい予測をしているように見えても、望ましいプロセスで導出している保証は



図1 説明性を持つAI

ないわけです。AI利用の安全性は社会的な課題にもなっており、最近では政府のAI安全性検証機関が設置されました。機械学習の解釈性に関する研究はAI利用の安全性にも直接リンクする現在ホットなトピックの一つです。解釈性へのアプローチはモデル依存型と、非依存型の2つに大別されます。モデル非依存型は、学習済みのAIに対して大量のデータの入力を行い、入出力の関係

なっています。モデル非依存型アプローチはどのようなモデルに対しても適用可能ですが、短所として大量の入出力を調べるために計算量が膨大であること、調べるために用いた入力データに偏りがあるなどの場合には解釈ミスが起こってしまうという欠点があります。これに対して、モデル依存型では、モデル自体をホワイトボックスにする、あるいは特定のモデルに対して学習済みモデルのパラメータの値などから直接推論

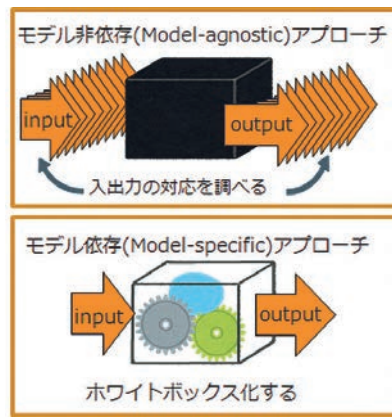


図2 解釈性への2つのアプローチ

から導出プロセスを読み取るようにするものです。多くの手法が考案されており、中でも2017年に発表されたSHAP値は代表的ともいえる手法で、この論文は被引用件数19775件(2024年3月Google Scholarより)の大ヒット論文と

プロセスを読み取るものです。解釈ミスは起こりにくく、計算量も比較的少なく済みますが、一般に難しく、またモデルの解釈性を高めることは、往々にして予測精度を下げることに繋がります。

3 研究室の研究テーマ

このような人工知能の技術発展の中、研究室では次のような研究を行っています。

非線形積分ニューラルネットワークモデルの構築

加法性を仮定しない測度による非線形積分を用いてホワイトボックス型のニューラルネットワークを設計するものです。非加法的な測度とは、 $1+1$ が 2 にならない測度です。数学的には加法性を持たない測度の理論体系は現在構築の途上ですが、非加法的な測度は古くから、工学、社会科学、心理経済学といった分野で人間の意思決定の表現として精力的な研究がなされてきました。このような加法性を仮定しない測度による積分はいくつか考案されていますが、私自身も離散数学の重要概念である包除原理に基づく「包除積分」を考

案しました。この積分を利用した数理モデルでニューラルネットワークを構築すると、パラメータや演算が、積分に用いる測度や相互作用演算に対応づけできるので、学習済みのモデルから情報を読み取ることが可能になります。新しいネットワークモデルであるため、自分たちでプログラムを作る必要があり、当初はそれぞれが苦労してコーディングして研究を進めていましたが、昨年ようやく標準的なライブラリを完成させることができ、研究の効率化が進んでいるところでは、比較的単純なデータに対しては、他の機械学習手法と遜色のない精度が実現でき、なおかつ推論過程に関する情報抽出を実現できるようにになりました。現在は、様々な種類のより複雑な問題に対応できるような改良に取り組んでいます。また、今、手計算で行っている情報抽出を、手動ではなく直観的に理解しやすいグラフや理解しやすい文章による記述などを自動出力するような手法の開発も行っています。この解釈性を持つニューラルネットワークのプログラムの部品を集めたライブ

ラリはプログラムコードの保存・公開サイトGitHubにて使用例とともに公開しています。(GitHub AoiHonda Lab/ELNeuralNetwork)

モデル非依存型解釈手法の開発

先述のSHAP値は入力する特徴量が出力に与える影響を評価する手法で、協力ゲーム理論における最も重要な概念であるシャープレイ値を利用したものです。SHAP値は比較の入力次元数、特徴量の数が20程度の少ない問題に対して適用できる手法であり、高次元数の入力や入力特徴量間に相互作用や階層構造、あるいはネットワーク構造のようなものがあっても、出力への貢献は一律に計算されるものになっています。そこで、SHAPでは表現できない特徴量間の構造を反映するような指標の開発を行っています。また、複雑で精度の高いブラックボックスモデルに対して、解釈可能なモデルで局所的に近似する局所代理モデルの開発にも取り組んでいます。

テキストマイニングを用いた顧客サービスの向上

研究室では、テキストマイニングを活用して顧客サービスの向上に取

り組んでいます。企業との連携プロジェクトにおいて、大量のお客様からのフィードバックやクレーム、質問などのテキストデータを分析し、有益な情報を抽出する手法を開発しています。前の2テーマとは違い、

既にある人工知能技術をどのように活用するかを考える研究ともいえませんが、これら技術を実際に現場で使う際には、必要なデータを得るための「データマネージメント」やデータを使える状態に整理する「データクレンジング」が必要で、これは非常に手間のかかる作業であり、かつフェーズ毎にさまざまな難しさを内包しています。これらの管理や自動化を考案し、そして最新の言語モデルを用いて効果的に情報抽出する手法の開発に取り組んでいます。

以上のテーマを中心に、学部4年生から修士2年までのメンバーが研究に取り組んでいます。毎週の研究室ゼミでは活発に議論が行われ、思いもよらない視点からの斬新なアイデアが飛び出しますし、ChatGPTにプログラムをコーディングさせるなど、どんどん新しい技術を吸収し

ている学生たちの柔軟で意欲的な姿勢には私自身が学ぶことばかりです。

4 私自身の研究テーマ

私自身は本学を卒業後、長らく数学分野の研究に従事してきました。ヒルベルト空間、ルベীগ空間に代表される完備なノルム空間、つまりバナッハ空間の空間としての線形性や位相的性質の解析に加え、バナッハ空間で成り立つ不等式の研究を行い、その後、ルベীগ測度の加法性を単調性に緩めたファジィ測度の研究にも携わるようになりました。シャノンエントロピーを非加法的な測度に適用できるような拡張や、シャープレイ値の必要十分条件の発見なども行っており、先述した包除積分は、これらの一連の測度・積分論研究の中で考案したものです。これらの理論的な研究は「精密な」といえば聞こえはよいですが、その応用先が何であるかは特に意識していませんでした。発見した定理たちが実問題に役立つのは私が死んだあとの遠い未来のことだろうと思っていました。しかし、コンピュータの性能と大量のデータが取得できるよう

になったことで、これらの理論を自らの研究で応用できるようになったことは非常に感慨深いものです。

これらの理論研究は、現在も単独で進める他、国内外の数学仲間や関連する研究者たちとの共同研究も進んでいます。最新の成果としては、積分に双対の概念を導入し、包除積分に用いる演算の双対性との対応を明らかにしたものがあります。この成果は、学生たちがデータ分析の研究で色々な演算を試みる中でインスパイアされたものです。また、研究室の学生たちも「応用」で使っているうちに「理論」についての理解が深まっており、数学の研究も研究室テーマにできそうです。

5 女性教員として

私が知能情報工学科に入学したときクラスには女子が1割強いました。先生方が「女子が多い」と喜んでおられたのが印象的でした。このまま女子の比率は増え、将来は男女半々になる日が来るだろうと思ったものですが、実際には30年以上たった今もほとんど増加していません。こんな未来は全く予測していませんでした。

た。もちろん、これは九州工大だけの問題ではなく、近年のOECD(経済協力開発機構)の調査によると日本は理系大学の女子学生割合で加盟38カ国中最下位となっています。これにはいろいろな要因が考えられますが、ロールモデルの不在もその一因と言われています。私は2023年7月に教授に昇任いたしました。情報工学科に所属する女性の教授は27年ぶりとのことでした。つまり30年近く女性教授が不在であったという現状があります。国内の他大学の理系学部や工学系学部と比較しても極端な偏りがあると言えます。

女子中高生や大学低学年の女子学生たちに理系の魅力を自然な形で伝えるためには、教職員、学生を含め大学に在籍する女性全員が生き生きと活動していること、活躍する女性研究者との交流の場を設けるなどの取り組みが必要です。本学の教育・研究の質を向上し、大学の価値を高めるために、多様な人材が切磋琢磨する学びの環境づくりのためにも、今後より一層の努力をしたいと思っています。皆様、どうぞよろしくお願ひします。