九州大学
KYUSHU UNIVERSITY

PRESS RELEASE（2024/12/12）

AI の思考を解き明かす！

～ニューラルネットワークの隠れたパターンを解明～

ポイント

① 従来の視覚化ツールではニューラルネットワーク内のデータ処理を正確に理解することが困難だったため、新しい手法が求められていた。
② 本研究グループは、新しい視覚化手法「k* Distribution（k *分布）」を開発し、3 つのデータ整理パターンを発見した。
③ これらの方法は AI の思考を理解するための革新的な手段を提供する。

概要

これまで、研究者たちはディープニューラルネットワーク（※１）がデータをどのように整理しているかを、t-SNE や UMAP のようなツールを使って視覚化していました。しかしこれらのツールは重要な詳細を歪めることがあり、特定のデータグループがどのように扱われているかを把握することが困難でした。

九州大学大学院システム情報科学研究院のヴァスコンセロス ヴァルガス ダニロ准教授らの研究グループは、ニューラルネットワークの隠れた層の中でデータがどのように整理されているかをより明確に見ることができる新しい方法「k*分布」を提案しました。k*分布という新しい方法を使用することで、研究者はこれらのパターンを正確に視覚化し、分析することができました。この方法により、以前のツールよりもデータの構造をしっかりと保持し、異なるグループを簡単に比較できるようになりました。

この研究は、AI の「思考」を理解するための革新的な手段を提供します。さらに AI の改善だけでなく、画像の処理方法に応用することで、より優れた診断ツールの開発につながる可能性もあり、将来の研究や実世界での応用に役立てることができると期待しています。

本研究成果は国際雑誌「IEEE Transactions on Neural Networks and Learning Systems」に 2024 年 9 月 16 日（月）（日本時間）にオンライン版で早期公開されました。

研究者からひとこと：

ニューラルネットワークにとって最も重要なことは、新しい手法の開発ではなく、それらを新たな視点で理解する方法であると言えるでしょう。私たちの可視化は、AI に対する化学元素のスペクトルのように機能します。それは、ニューラルネットワークを改善する方法や、実際に何をしているのかに新たな光を当てます。したがって、今回の発見は、可視化が人工知能の進歩の主要な支柱であるため、あらゆる応用 AI 分野でますます活用されるでしょう。

【研究の背景と経緯】

シェフが多くの材料を使って料理を作る場面を想像してください。その材料がどのように混ざり合い、作用し合うかを理解するためには、シェフは細かく観察する必要があります。同様に、ディープニューラルネットワークも複雑なデータを処理する方法を学びますが、具体的にどのように情報を処理しているかを理解するのは難しいです。研究者たちは、t-SNE や UMAP のようなツールを使ってニューラルネットワークがデータをどのように整理しているかを視覚化しますが、これらのツールは重要な詳細を歪めることがあり、特定のデータグループがどのように扱われているかを把握することが困難です。この研究では、ニューラルネットワークの隠れた層の中でデータがどのように整理されているかをより明確に見ることができる新しい方法「k*分布」を提案しています。これにより、異なる情報がどのように処理・整理されているかを研究者がより深く理解できるようになります。

【研究の内容と成果】

研究者たちは、ニューラルネットワークがデータを整理する主なパターンとして、3 つのパターンを発見しました。それは「分断型」、「重複型」、そして「クラスター型」です。分断型はデータが小さなグループに分かれ、重複型は異なるデータが混じり合い、クラスター型は似た種類のデータが密集していることを意味します。k*分布という新しい方法を使用することで、研究者はこれらのパターンを正確に視覚化し、分析することができました。この方法は、以前のツールよりもデータの構造をしっかりと保持し、異なるグループを簡単に比較できるようにします。この方法はさまざまなニューラルネットワークで機能し、入力データに対する変換処理にどのように対応するかも評価できるようになりました。これにより、ネットワークが情報を処理し、グループ化する方法について、より深い理解が得られるようになります。

【今後の展開】

この研究は、AI の「思考」を理解するための革新的な手段を提供します。ニューラルネットワークがデータをどのように整理し解釈しているかをより明確に示すことで、よりスマートで効率的な AI システムの設計に貢献します。翻訳者が突然、言語の微妙な構造すべてを見通せるようになるのと同じように、このツールは AI 研究者に大きな優位性をもたらします。例えば、画像認識の分野では、物体が部分的に隠れている状況でも、システムが物体をより正確に区別できるようになります。AI の改善だけでなく、画像の処理方法に応用することで、より優れた診断ツールの開発につながる可能性もあり、将来の研究や実世界での応用に大きな影響を与えることが期待されています。

【参考図】

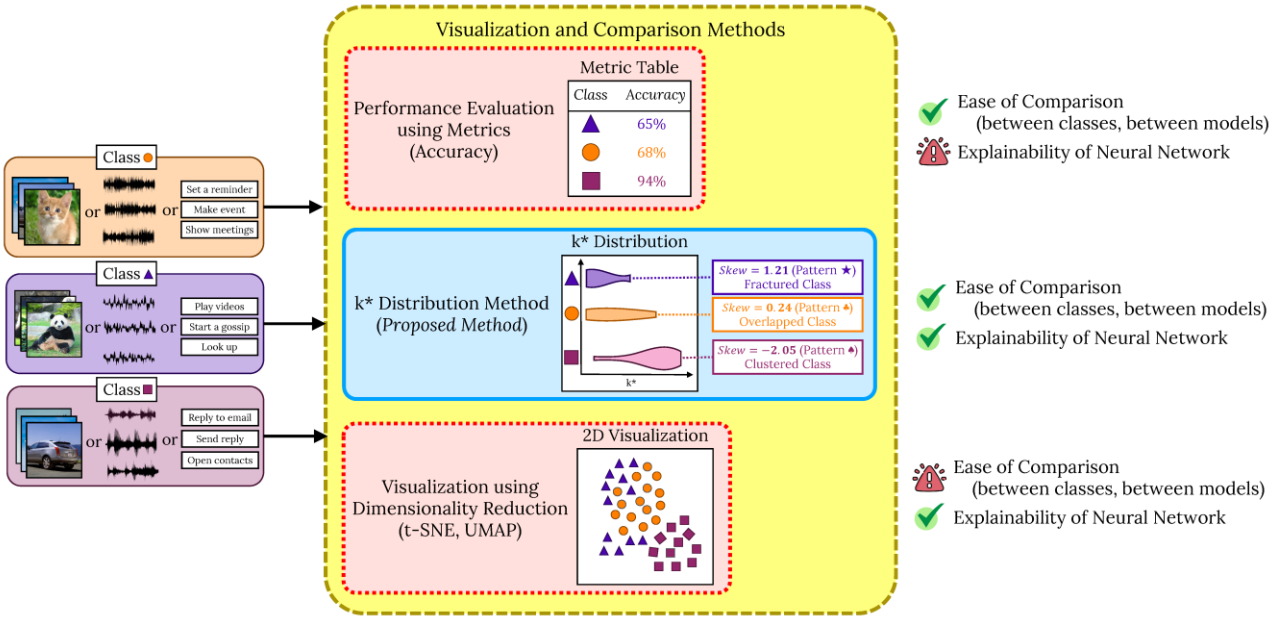

図１　ニューラルネットワークの理解を深めるための新しい可視化手法

この図は、新しい「k*分布」手法が、従来のメトリック表や 2D の視覚化よりも、クラス間およびモデル間の違いをより分かりやすくし、ニューラルネットワークがどのようにデータを解釈しているかを明確にする様子を示しています。

【用語解説】

（※１）　ディープニューラルネットワーク

人間の脳のように情報を学習し、認識するために多層の計算構造を持つ人工知能の仕組みです。

【謝辞】

本研究は、日本学術振興会（JSPS）挑戦的萌芽研究助成金（JP22K19814）、日本科学技術振興機構（JST）戦略的基礎研究推進プログラム（先進的知能システムプログラム（AIP）加速研究）（JP22584686）、および JSPS 学術変革領域研究（A）（JP22H05194）の助成を一部受けて実施されました。

【論文情報】

掲載誌：IEEE Transactions on Neural Networks and Learning Systems

タイトル：k* Distribution: Evaluating the Latent Space of Deep Neural Networks Using Local Neighborhood Analysis

著者名：Shashank Kotyan; Tatsuya Ueda; Danilo Vasconcellos Vargas

D　O　I　：10.1109/TNNLS.2024.3446509

【お問合せ先】

<研究に関すること>

九州大学大学院システム情報科学研究院情報学部門 准教授
VASCONCELLOS VARGAS DANILO（ヴァスコンセロス ヴァルガス ダニロ）
TEL：092-802-3599
Mail：vargas@inf.kyushu-u.ac.jp

<報道に関すること>

九州大学　広報課
TEL：092-802-2130　FAX：092-802-2139
Mail：koho@jimu.kyushu-u.ac.jp

Kyushu University VISION 2030
総合知で社会変革を牽引する大学へ