

Deep Tensorとナレッジグラフを融合した説明可能なAI

Explainable AI Through a Combination of Deep Tensor and Knowledge Graph

● 富士 秀 ● 森田 一 ● 後藤 啓介 ● 丸橋 弘治
● 穴井 宏和 ● 井形 伸之

あらまし

近年のAI(人工知能)の最も大きな成果の一つは、深層学習による機械学習精度の大幅な向上である。しかし、深層学習は膨大なデータから巨大なニューラルネットワークを学習するため、たとえ正しい推定ができたとしてもその理由や根拠を示すことが難しい。そのため、金融や医療などの信頼性が求められるビジネス分野へのAI適用を妨げる原因となっている。筆者らはこの問題を解決するため、独自の深層学習を発展させた機械学習技術であるDeep Tensorと、ナレッジグラフと呼ばれる過去の文献やデータベースから構築したグラフ型の知識ベースを融合し、Deep Tensorの推定結果に対する理由や根拠を論理的に説明するAI技術を開発した。

本稿では、この説明可能なAIを実現する技術をネットワーク侵入検知とゲノム医療に適用した事例に基づいて紹介する。

Abstract

One of the most significant advancements made in AI (artificial intelligence) in recent years is the greatly enhanced accuracy of machine learning through deep learning. However, because deep learning deals with huge volumes of data and involves vast neural networks in the learning process, it is often difficult to explain how or why an output was reached even if the estimation was correct. This point has been an impediment to applying AI technology in such business areas as finance and medicine, which demand absolute reliability. As an attempt to address this issue, we have developed an AI Technology that combines Deep Tensor, Fujitsu's unique learning technology based on enhanced deep learning, and Knowledge Graph, a knowledge base presenting graph data taken from past documents and databases. This has enabled us to logically explain the reasons and basis in which Deep Tensor reaches its estimation output. This paper explains the technology that makes explainable AI possible in terms of application cases in network intrusion detection and in genomic medicine.

ま え が き

近年のAI（人工知能）の発展は目覚ましいものがあり、一部の領域では人間を凌駕する成果を挙げている。毎年開催される画像認識の精度を競うコンペティションILSVRCでは、2012年から深層学習による大きな精度向上が続き、2015年にはついにAIの精度が人間の精度を上回った。また、これまで複雑で機械が人間に勝つのは難しいとされた将棋や囲碁についても、2013年に将棋で⁽¹⁾、2016年に囲碁で⁽²⁾ 現役プロ棋士に勝利しており、これまで人間が発見できなかった新たな定石をAIが生み出すまでに至っている。

近年ではただ精度向上を目指すのではなく、より人間にとって使いやすく、また協調して動作するAI技術やサービスが登場してきている。例えば、チャットボットはユーザーとの対話を通じてユーザーの求める答えを返す。これは、ユーザーがあらかじめ明確な目的や検索方法を指定する必要があった従来の手法とは大きく異なり、AIがより人間に寄り添って価値を提供している。

AIの重要な要素である機械学習においても、単純に推定するだけでなく、なぜそのような推定をしたのか、その根拠を示すような技術の開発が始まっている。従来の機械学習技術は、どのような推定を行っているのか人間には理解できないブラックボックスであることが多かった。これでは、生命に関わる場合などの重要な意思決定には適用しにくい。推定の根拠を示すことができれば、こ

のような課題を解決し、より幅広い分野への適用が期待できる。

このような背景の中、富士通研究所ではDeep Tensor⁽³⁾ およびナレッジグラフ技術と呼ばれる機械学習技術を開発してきた。

ナレッジグラフ技術は、各分野で扱われる多種多様な知識データを統一的なグラフ形式のデータに変換し、統合的な一つの知識ベース（ナレッジグラフ）として扱えるようにする技術である。従来は個別に扱われてきた知識を統合することで、今まで記述が困難であった複雑な関係性などを記述できるようになった。

一方Deep Tensorは、グラフデータの学習が可能な、深層学習を発展させた富士通独自の機械学習技術である。従来は、グラフデータの学習に用いる特徴量の設計に専門家の手を加える必要があり、精度に限界があった。Deep Tensorはグラフデータをテンソルと呼ばれる数学的な形式で表現し、それをテンソル分解⁽⁴⁾を用いて統一的なテンソル表現に変換することにより、グラフデータの特徴量を抽出する。更に、従来のニューラルネットワークの学習技術を拡張した拡張誤差逆伝搬法を用いてテンソル分解そのものを最適化し、特徴量の設計を自動化する（図-1）。これにより、高精度な推定が可能となる。

今回、筆者らはDeep Tensorとナレッジグラフ技術を組み合わせることで、機械学習の推定理由や根拠を論理的に説明する技術を開発した。本稿では、まず従来の深層学習の課題であるブラックボッ

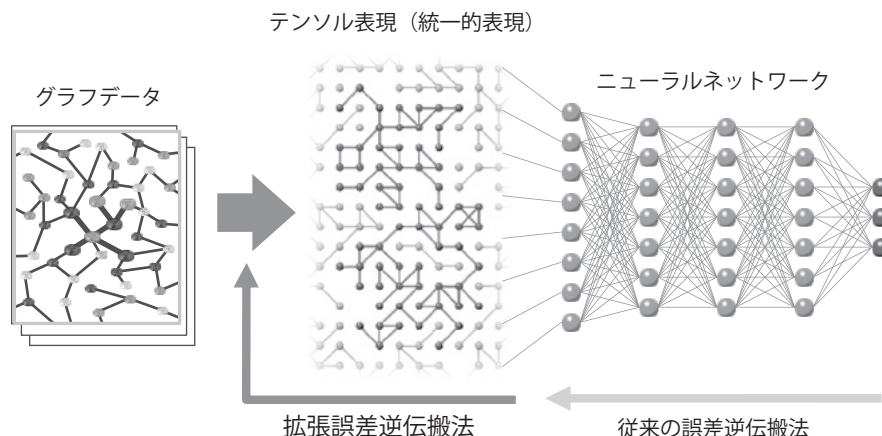


図-1 Deep Tensorの学習プロセス

クス型機械学習について述べる。次に、このブラックボックス性を解決するため、筆者らが開発した二つの技術について述べる。一つ目は推定結果に大きく寄与した入力データ中の因子を特定する技術であり、この技術の概要と有効性についてセキュリティ分野への適用例をとおして述べる。二つ目は、その因子とナレッジグラフ中の知識をつなぎ合わせて根拠を構成する技術であり、ゲノム医療分野への適用例をとおして技術の概要と有効性について述べる。

ブラックボックス型機械学習の課題

専門家がAIと協調して問題解決を行うためには、機械学習の推定結果を専門家が理解し、それを踏まえて意思決定する必要がある。

深層学習による機械学習は、高い性能が得られる一方で、大規模なニューラルネットワークが学習された結果、入力と出力の関係が非常に複雑になる。そのため、なぜその結果が出てきたのかをユーザーに説明できない。

その根拠を説明するためには、推定結果から専門家が改めて文献などを調べて検証する必要がある。特に、関係性が部分的にしか知られていないような事象に関しては、専門家が裏付けとなる根拠を文献やデータベースから探し出し、つなぎ合わせて検討する必要があった。

この検討を行う際、例えば後に述べるゲノム医療への適用例では、結果に関係する医学・医療文献を調べる必要がある。医学・医療の分野では、1か月に10万件以上の文献が公開される。この中から、関連する文献を挙げて検証するために、これまでは多大な労力が必要であった。

近年、入力データの中で推定結果に大きく寄与する部分を特定する研究が世界中で行われている。しかし、画像認識においては、画像のどの部分が推定結果に寄与したかを説明できる程度にとどまっている。また、推定結果に大きく寄与した入力データ中の部分を特定できたとしても、その推定に至った根拠を示すことはできなかった。

開発した技術

富士通研究所では、Deep Tensorと過去の文献やデータベースから構築したナレッジグラフの活用

技術を融合することで、Deep Tensorによる推定理由や根拠を提示する技術を開発した。この技術では、第1段階として、Deep Tensorにおいて推定結果に大きく寄与した推定因子（部分グラフ）を特定する。第2段階として因子をナレッジグラフと対応付け、これらをナレッジグラフ上でつないで、一連の情報を推定根拠として構成し、ユーザーに提示する。これにより、Deep Tensorによる高い推定精度を維持したまま、ブラックボックス型機械学習の課題を解決できる。

以下では、これら2段階のステップについてより詳しく述べる。

● Deep Tensorによる推定因子特定技術

Deep Tensorは、テンソル分解によってグラフデータを統一的なテンソル表現へと変換し、そのテンソル表現をニューラルネットワークに入力することで学習を行う。ここで注目すべきは、テンソル表現は推定精度が高くなるように学習されるため、入力グラフの中から特に推定に貢献する特徴的な因子を抽出した構造になっている点である。また、テンソル表現は入力グラフから線形変換で得られるため、変換後のテンソル表現から入力グラフを線形逆変換で容易に得ることができる。

筆者らは、この性質を利用してテンソル表現の中から推定結果に大きく寄与した因子を特定し、それらを線形逆変換することにより、推定結果に大きく寄与した入力グラフの推定因子を特定する技術の開発に成功した。

このように、Deep Tensorは高精度な推定だけでなく、推定因子を特定することにも適しており、高度な分析や人間の判断支援など様々な応用が期待できる技術と言える。

● ナレッジグラフによる根拠構成技術

グラフデータを学習・推定に用いるDeep Tensorは、グラフデータを構築するナレッジグラフ技術と親和性が高い。ナレッジグラフは、多種多様な知識データが含まれる巨大なグラフデータである。ナレッジグラフから学習・推定に必要な部分グラフを切り出すことで、Deep Tensorの学習・推定に用いるグラフデータを作成できる。

ナレッジグラフを基にしたグラフデータを使用して、Deep Tensorで推定および推定因子を特定することで、推定に大きく寄与した因子をナレッジ

グラフ上の知識に読み替えることができるようになる。

ナレッジグラフによる根拠構成技術は、この推定に影響した知識を互いに適切に関連付けし、推定する対象と推定結果を結び付ける、脈絡のある一連の知識として根拠を構成する技術である。

ナレッジグラフは多種多様な知識データを含むため、ある知識に対してグラフ上の関連をたどることにより、無数の知識を提示できる。推定理由と推定結果を結び付ける際には、それらの関連をたどって根拠を構成する。しかし、ナレッジグラフ上には関連をたどってその二つを結び付ける無数のパスが存在し得る。そのため、根拠を構成する際に、無関係な情報を構成要素としてしまえば、推定結果を説明する根拠として不適切なものとなる。

本技術では、推定因子を手掛かりとしてグラフ構造を探索することにより、特定された推定因子に関連性の高い知識だけを抽出して、根拠として構成する（図-2）。

評価・適用事例

今回開発した技術の実用可能性を評価した。まず、第1段階である推定因子特定技術について、ネットワーク侵入検知を対象として評価した。そして、

第1段階と第2段階を統合した説明可能なAI全体について、ゲノム医療を対象として評価した。以下にそれぞれの評価について述べる。

● 推定因子特定技術の評価

米国DARPAが公開している、ネットワーク侵入検知用ベンチマークデータセット⁽⁵⁾を使用して評価した。データセットでは、侵入検知システムが検知したネットワーク通信ログに対し、誤検知ではなく実際の侵入であることが分かっているログが正解因子として記載されている。

まず、各データについて、Deep Tensorの推定因子特定技術によって推定に大きく寄与した上位3点の因子を抽出した。推定因子の抽出例を図-3に示す。ここで、図中の「★」は実際の侵入に関与した正解因子を示している。そして、抽出した推定因子が正解因子と一致する正答率を既存手法と比較することで評価した。

既存手法の正答率が10%から28%であったことに比べ、Deep Tensorの推定因子技術の正答率は23%から52%となり、既存の手法よりも高い正答率が得られることが確認できた。このことから、Deep Tensorは推定因子の特定に適した説明能力の高い機械学習技術と言える。

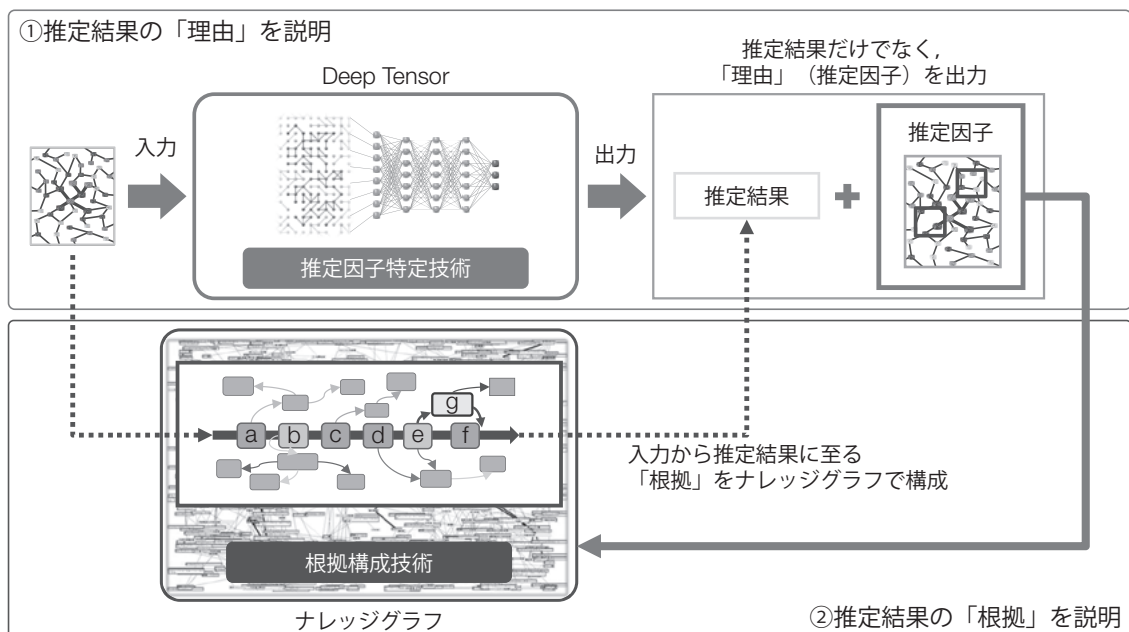
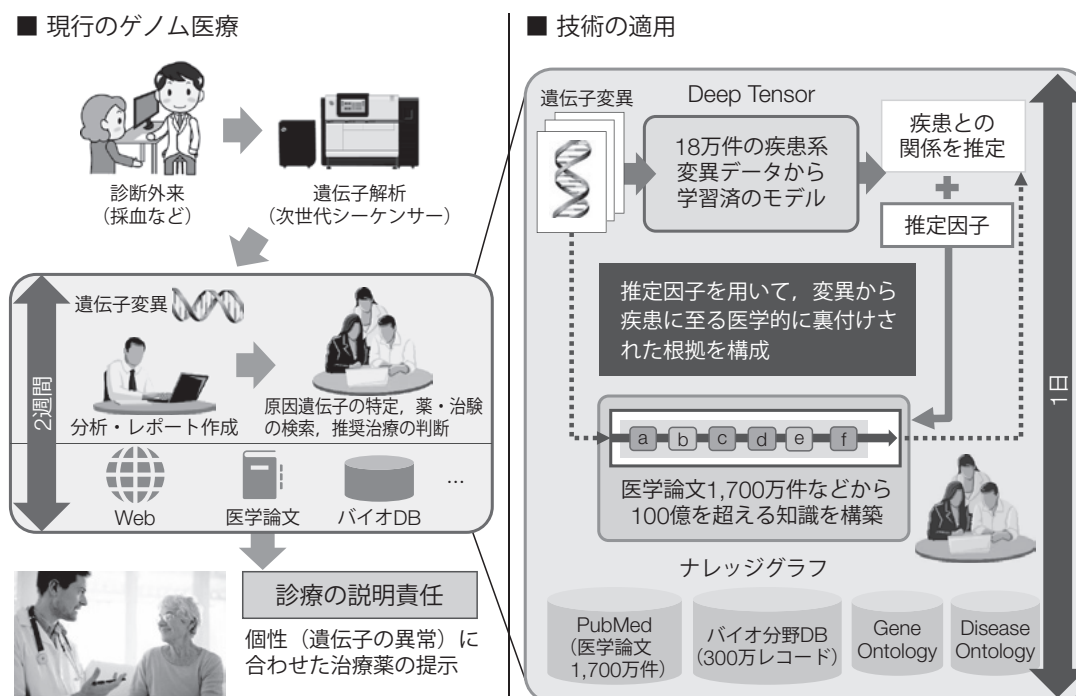


図-2 Deep Tensorとナレッジグラフの組み合わせによる根拠構成技術

Jul 29 18:50-19:00 (全612ログ)							
送信IP (10種類)				受信IP (15種類)			
		ログ数	寄与度			ログ数	寄与度
★	207.253.084.013	6	0.636	★	172.016.118.020	6	0.636
	172.016.114.207	66	0.207		207.037.252.205	66	0.207
	207.230.054.203	117	0.001		206.132.025.051	27	0.005
送信Port (371種類)				受信Port (246種類)			
		ログ数	寄与度			ログ数	寄与度
★	49724	6	0.636	★	584	3	0.410
	8739	1	0.003		80 (http)	130	0.186
	8328	1	0.003	★	410	2	0.181

図-3 推定因子の抽出例



北海道大学病院のWebサイト (<http://www.huhp.hokudai.ac.jp/hotnews/detail/00001144.html>) を参考に作成

図-4 ゲノム医療への応用による効果検証

● 説明可能なAI(推定因子特定技術および根拠構成技術)の評価

ナレッジグラフ技術による根拠構成技術では、ゲノム医療における専門家の調査作業の効率化を想定した模擬実験を行った。生物情報学分野における公開データベースや医療文献データベース⁽⁶⁾から構築したナレッジグラフを利用して、関係性が部分的にしか知られていないような事象に関して裏付けとなる知識を探し出し、ひも付けが可能であるかを検証した(図-4)。

まず、公開データベースから構築した遺伝子変異と病因性の関係について、Deep Tensorを用いて学習した。次に、推定因子特定技術で求めた因子に関する情報や学術論文を抽出し、根拠を構成した。図-5の根拠構成例では、推定対象の遺伝子変異を赤(五角形のノード)、推定結果に大きく寄与した因子を青(丸のノード)、医療文献などから抽出した学術的な裏付けとなる知識を黄色(四角形のノード)、疾患の候補を紫(三角形のノード)として表している。ノード間をつなぐエッジ(実線)

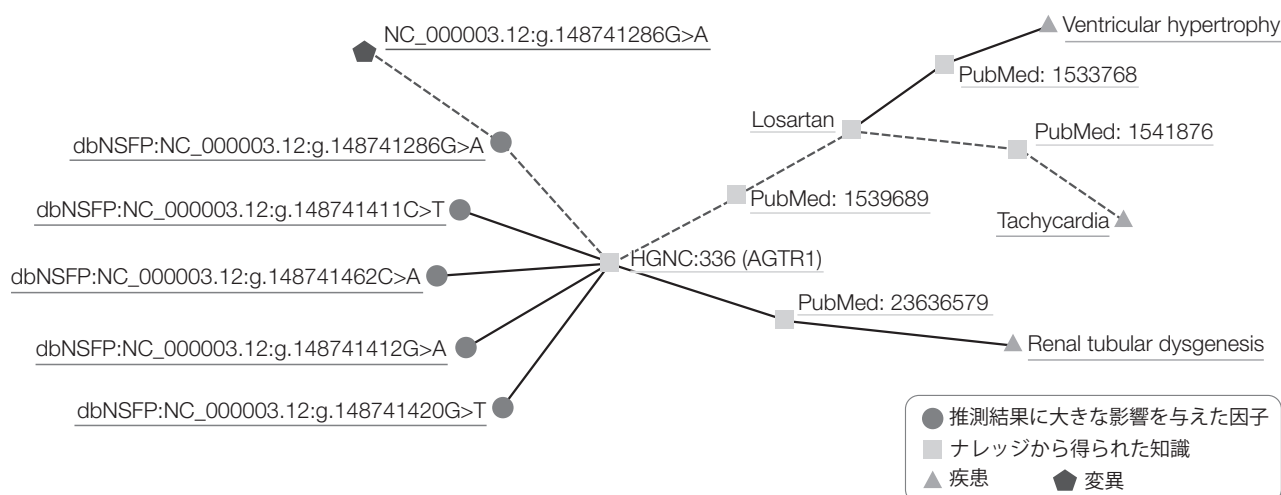


図-5 根拠の構成例

は、それらの知識がナレッジグラフ上で関連を持っていることを表す。グラフ中の破線は、遺伝子変異とその遺伝子、遺伝子に関わる薬（Losartan）、薬に関係する疾患をつないでおり、遺伝子変異と疾患の関連を示している。遺伝子変異から、ナレッジグラフ上で遺伝子と薬品間の関係と、薬品と疾患間の関係をたどることによって、疾患の候補までを互いに関連付けて一覧できる根拠となる知識のグラフを構成した。

図-5の例では、遺伝子変異NC_000003.12:g.148741286G>Aに関する論文34件を全て調査する代わりに、グラフ上で遺伝子変異と疾患（Tachicardia）を結ぶ線上にある2件の論文を検証することで、遺伝子変異と疾患の関係を判断できるようになる。これにより、遺伝子変異に関わる疾患を特定・検証する労力を大きく削減できた。

む す び

本稿では、Deep Tensorとナレッジグラフ技術の融合により、Deep Tensorが行った推定の理由や根拠を論理的に説明するAI技術について述べた。

なお、ゲノム医療分野での効果検証に用いたデータの一部は、国立研究開発法人日本医療研究開発機構（AMED）「臨床ゲノム情報統合データベース整備事業」における「ゲノム医療を促進する臨床ゲノム情報知識基盤の構築」に関し、国立大学法人京都大学との共同開発によって得られた成果によるものである。

今後は、推定した根拠の説明に関する定量的評価や有用性の評価についてゲノム医療および他分野への応用を計画している。ゲノム医療分野では、医療に関わる研究機関の協力を得て、今回の技術によって示された根拠が専門家に納得されるものであるか、十分に分かりやすいかという観点で検証していく。また新たな応用先として、金融分野において、融資先の自動推定を学習させた場合に規制や規則の知識を用いて推定の妥当性を確認するなど、本技術の他分野への適用を進めていく。本技術は、様々な分野のナレッジグラフの拡充やPoC（Proof of Concept）を進め、2018年度中に富士通のFUJITSU Human Centric AI Zinraiの関連サービスとして製品化する予定である。

参考文献

- (1) 株式会社ドワンゴ, 日本将棋連盟: 第2回将棋電王戦. 2013年3月.
- (2) Google : AlphaGo: Mastering the ancient game of Go with Machine Learning, January 2016.
- (3) K. Maruhashi et al. : Learning Multi-way Relations via Tensor Decomposition with Neural Networks, Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), February 2018.
- (4) T. G. Kolda et al. : Tensor decompositions and applications. SIAM Review, Vol.51, No.3, p.455-500, 2009.
- (5) MIT Lincoln Library : DARPA Intrusion Detection

Data Sets.

<http://www.ll.mit.edu/ideval/data/>

(6) US National Library of Medicine : PubMed.

<https://www.ncbi.nlm.nih.gov/pubmed/>



井形 伸之 (いがた のぶゆき)

(株) 富士通研究所
人工知能研究所
兼 富士通(株) テクニカルコンピューティング・ソリューション事業本部
情報検索, 知識処理, 自然言語処理などの研究に従事。

著者紹介



富士 秀 (ふじ まさる)

(株) 富士通研究所
人工知能研究所
兼 富士通(株) テクニカルコンピューティング・ソリューション事業本部
ナレッジグラフ, 自然言語処理などの人工知能関連の研究に従事。



森田 一 (もりた はじめ)

(株) 富士通研究所
人工知能研究所
兼 富士通(株) テクニカルコンピューティング・ソリューション事業本部
自然言語処理の研究に従事。



後藤 啓介 (ごとう けいすけ)

(株) 富士通研究所
人工知能研究所
文字列索引, グラフデータに関する機械学習の研究開発に従事。



丸橋 弘治 (まるはし こうじ)

(株) 富士通研究所
人工知能研究所
グラフデータに関するデータマイニング, 機械学習の研究開発に従事。



穴井 宏和 (あない ひろかず)

(株) 富士通研究所
人工知能研究所
数式処理, 数値最適化, 人工知能の研究開発に従事。