

Deep Tensorの推定結果を説明する 推定因子特定技術

Inference Factor Identification Technology to Explain Results for Deep Tensor Inferences

松尾 達 大木 雄介 丸橋 弘治

あらまし

人工知能（AI）の推定結果に対して、推定の理由や根拠を説明することが求められている。特に、医療や金融などの説明責任が伴う分野でAIが受け入れられるためには、推定の理由や根拠を示して十分な信頼を得る必要がある。しかし、Deep Learningなどの高精度な推定を行うAIの多くは、推定の理由や根拠を説明することが困難である。富士通研究所が開発した、人やモノのつながりを表現するグラフデータを高精度に解析可能なAIであるDeep Tensorにおいても、推定理由の説明は重要な課題である。そこで富士通研究所では、この課題を解決する技術として推定因子特定技術を開発した。これは、Deep Tensorが生成したコアテンソルと呼ばれる特徴量を使用し、推定結果に寄与したグラフデータの要素を提示することによって説明を行うものである。

本稿では、推定因子特定技術について述べるとともに、医療および金融分野への適用事例を紹介する。

Abstract

Inference results provided by AI require accountability in terms of the reasons and basis behind the inference. For AI to be accepted in areas where accountability is needed such as the medical and financial sectors, in particular, the reasons and basis for an inference must be shown to earn sufficient trust. However, with many AI methods providing high-accuracy inferences such as deep learning, explaining the reasons or basis for an inference is difficult. For Deep Tensor, an AI technology developed by Fujitsu Laboratories capable of high-accuracy analysis of graph data that represent connections between people and things, accountability for reasons of an inference is an important issue. Accordingly, Fujitsu Laboratories has developed inference factor identification technology as a technology to resolve this issue. This uses feature values called core tensors generated by Deep Tensor to present elements of graph data that contributed to the results of an inference, thereby providing explanation. This paper describes the inference factor identification technology and presents examples of its application in the medical and financial sectors.

1. まえがき

近年の人工知能（AI）技術の進展によって、AIの適用分野は拡大している。特に、大量のデータを学習させることで高精度な判断を実現するDeep Learningの進展は目覚ましく、画像データを扱う分野を中心に普及が進んでいる。しかし、このような高度なAIの多くはブラックボックスであるため、AIの判断結果の理由を説明することが課題となっている。富士通研究所では人やモノのつながりによって表現されるグラフデータを高精度に解析できるDeep Tensor⁽¹⁾⁻⁽³⁾を開発した。しかし、Deep Tensorにおいても判断結果の理由を説明することが課題であった。

そこで、富士通研究所ではDeep Tensorの推定結果を説明する推定因子特定技術を開発した。本技術によって、Deep Tensorの推定結果に寄与したグラフデータの要素を提示できる。これによって、対象データの専門家がDeep Tensorの推定結果を合理的に解釈できるか、ひいてはDeep Tensorを信頼できるかの判断が容易になる。推定結果に合理的な解釈があり信頼を得られれば、特に医療や金融などの説明責任が求められる分野においても、グラフデータを対象としたAIを活用できるようになる。

本稿では、推定因子特定技術について述べ、医療および金融分野への適用事例を紹介する。

2. Deep Tensorの推定結果説明における課題

ブラックボックスとなっているAIの推定結果を

説明可能にする従来技術は存在していた。一方、Deep Tensorにはグラフデータ特有の課題があるため、その技術を適用できなかった。

本章では、Deep Tensorの推定結果を説明する際の課題について述べる。

2.1 Deep Tensor

Deep Tensorは、富士通研究所が開発したグラフデータを対象としたDeep Learningである。グラフデータとは、人やモノのつながりを表すデータである。つながりの主体となる人やモノは「ノード」、それらのつながりは「エッジ」として表現される。グラフデータの学習においては、グラフデータから重要な特徴を抽出することが課題となるが、Deep Tensorはその課題を解決する技術である。

Deep Tensorの概要を図-1に示す。Deep Tensorでは、グラフデータをテンソルと呼ばれる数学上の概念で表現する。更に、独自技術である構造制約テンソル分解を用いて、抽出したグラフデータの重要な特徴をニューラルネットワークに入力することで高精度な推定（分類や回帰）を実現する。テンソル表現においてグラフデータは、ノード間の組み合わせに対してエッジが存在するか否かを示すデータとして表現される。このテンソルデータを、コアテンソルと呼ばれるテンソル、および要素行列と呼ばれる行列に分解する。これは一般的にテンソル分解と呼ばれるが、Deep Tensorではコアテンソルにグラフデータの重要な特徴が含まれるように分解する。具体的には、ターゲットコアテンソルと呼ぶテンソル分解の基準となる指標を導入し、コアテンソルがターゲットコアテンソルに類似するように分解する。

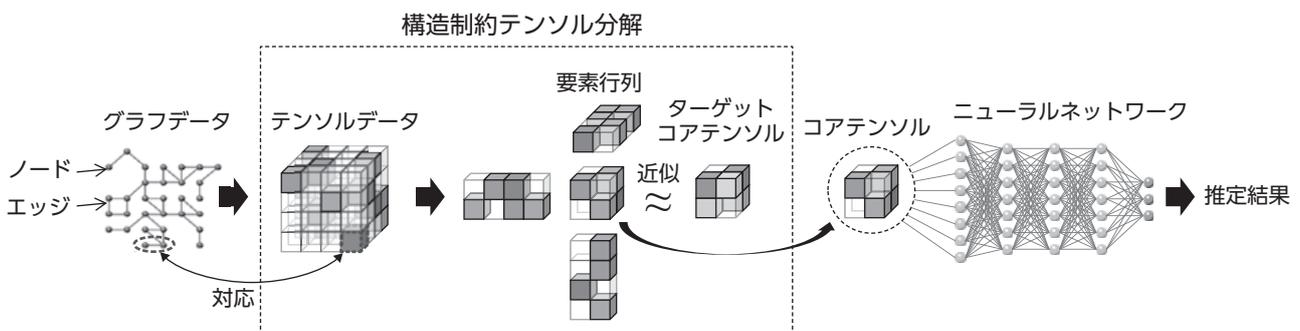


図-1 Deep Tensorの概要

このターゲットコアテンソルはあらかじめ与えられるものではなく、独自の学習技術によってニューラルネットワークと一体となって学習される。つまり、推定精度が高くなるように学習されるため、学習後のターゲットコアテンソルに類似するように分解されたコアテンソルには、推定結果に寄与するグラフデータの重要な特徴が含まれていることになる。重要な特徴が何であるかについてはデータに依存するが、例えばグラフの部分構造や全体構造が挙げられる。

このように、Deep Tensorはグラフデータに対して高精度な判断ができる。しかし、ブラックボックスとなっているニューラルネットワークを含んでい

るため、そのままでは推定結果を説明することが困難である。そこで、ブラックボックスとなっているAIの推定結果を説明可能にする技術を応用して、Deep Tensorに適用する。

2.2 推定結果を説明可能にする従来技術LIME

ブラックボックスとなっているAIの推定結果を説明可能にする従来技術としてLIME (Local Interpretable Model-Agnostic Explanations)⁽⁴⁾がある。LIMEでは、説明に適した変数を入力とする線形回帰モデルを使用して元のモデルを近似することによって、推定結果を説明する。

図-2に、入力された画像がニワトリであることを推

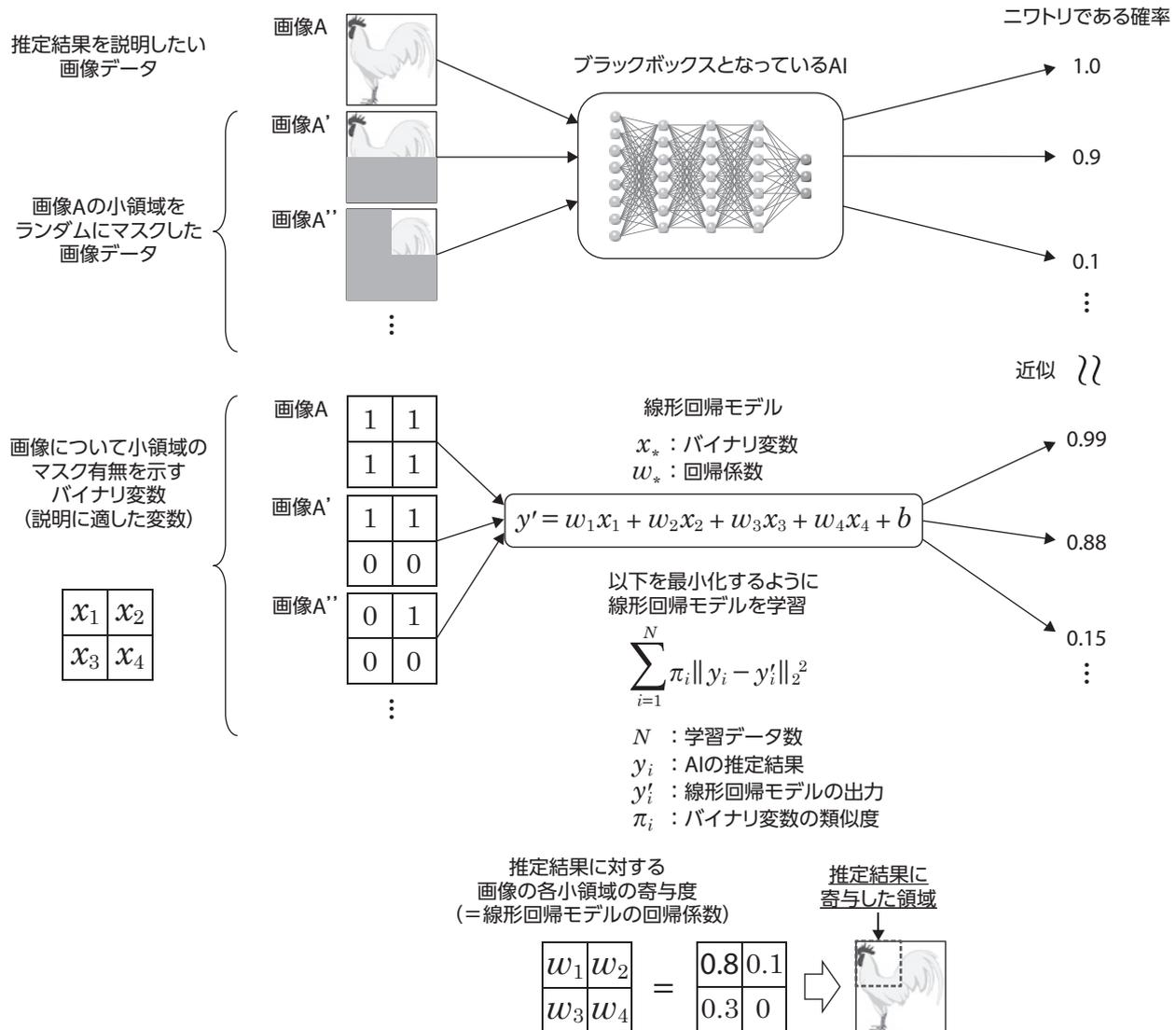


図-2 画像データ向けAIに対するLIMEの適用イメージ

定するAIに対して、LIMEを適用する場合のイメージを示す。まず、推定結果を説明したい画像データを小領域に分割し、各小領域をランダムにマスクした画像データを多数生成して、AIでニワトリである確率を推定する。次に、小領域のマスク内容を示すバイナリ変数（0：マスクした、1：マスクしていない）を入力とし、その入力に対応した画像に対するAIの推定結果を出力とする線形回帰モデルを学習する。

このとき、推定結果を説明したい画像データに対応するバイナリ変数と、各画像データに対応するバイナリ変数との類似度に基づいて、重み付けした学習を行う。これによって、線形回帰モデルはその画像データに類似した画像データについて、AIの振る舞いを近似するモデルとなる。

このようにして得られた線形回帰モデルの各変数の回帰係数を、推定結果への寄与度とみなすことによって、回帰係数の大きい変数が推定結果に大きく寄与しているという説明ができる。すなわち、画像のどの小領域に注目してニワトリと推定したかを提示できる。

2.3 グラフデータに対する推定結果説明の課題

Deep TensorにLIMEを適用しようとする場合、グラフデータについて説明に適した変数が用意できないという課題が生じる。上述した画像データと同様に、何らかの部分グラフに分割するという方法が考えられるが、自明かつ適切な分割方法はない。そのため、グラフデータに応じて専門家の知見に基づいた分割方法を検討することになる。しかし、専門家であっても適切な分割方法を導出できるとは限らない。特に膨大なデータから適切に導出することは極めて困難である。もし分割方法が不適切であれば、推定結果の説明も不適切なものになってしまう。

3. 推定因子特定技術

本章では、Deep Tensorの推定結果を説明する推定因子特定技術について述べる。

本技術ではLIMEと同様に線形回帰モデルを学習することによって、Deep Tensorの推定結果に寄与したグラフデータの要素を提示する。その際に、

Deep Tensorの仕組みを利用して前章で述べた課題を解決する。

前述したように、Deep Tensorが生成するコアテンソルにはグラフデータの重要な特徴が含まれている。そのため、コアテンソルを使用すればグラフデータの分割方法の検討は不要になる。しかし、コアテンソルそのものを解釈することは困難であることが多く、そのような場合はコアテンソルについて得られた寄与度で推定結果を説明することも困難である。そこで、この寄与度を元のテンソルデータについての寄与度、すなわちグラフデータのエッジ単位の寄与度に変換する。これによって、グラフデータのどの部分が推定結果に大きく寄与したかを提示できる。

推定因子特定技術の全体像を、図-3に示す。まず、コアテンソルの各要素を入力、Deep Tensorの推定結果を出力とする線形回帰モデルを、LIMEと同様の方法で学習する。図中の例では、コアテンソルの要素数が8個であるため、線形回帰モデルの入力変数の数も8個となる。この線形回帰モデルの学習データは、推定結果を説明したいデータとそのほかのデータのコアテンソル、およびこれらのデータに対するDeep Tensorの推定結果である。ここで「そのほかのデータ」とは、例えばDeep Tensorの学習に使用したデータである。

線形回帰モデルの学習では、コアテンソルが推定結果を説明したいデータと類似しているデータについて、Deep Tensorの振る舞いを近似するように学習する。具体的には、推定結果を説明したいデータと各学習データのコアテンソルの類似度に基づいて、重み付けして学習する。

得られた線形回帰モデルの回帰係数はコアテンソルの各要素に対応しており、これをコアテンソルの各要素の推定結果に対する寄与度とみなす。この寄与度に構造制約テンソル分解の際に得られる要素行列を掛け合わせると、元のテンソルデータの各要素の寄与度、すなわちグラフデータの各エッジの寄与度が得られる。これが、本技術のアウトプットである。

このようにして得られた各エッジの寄与度の大小関係に注目すれば、グラフデータのどの部分が推定結果に大きく寄与したかわかる。そのため、対象

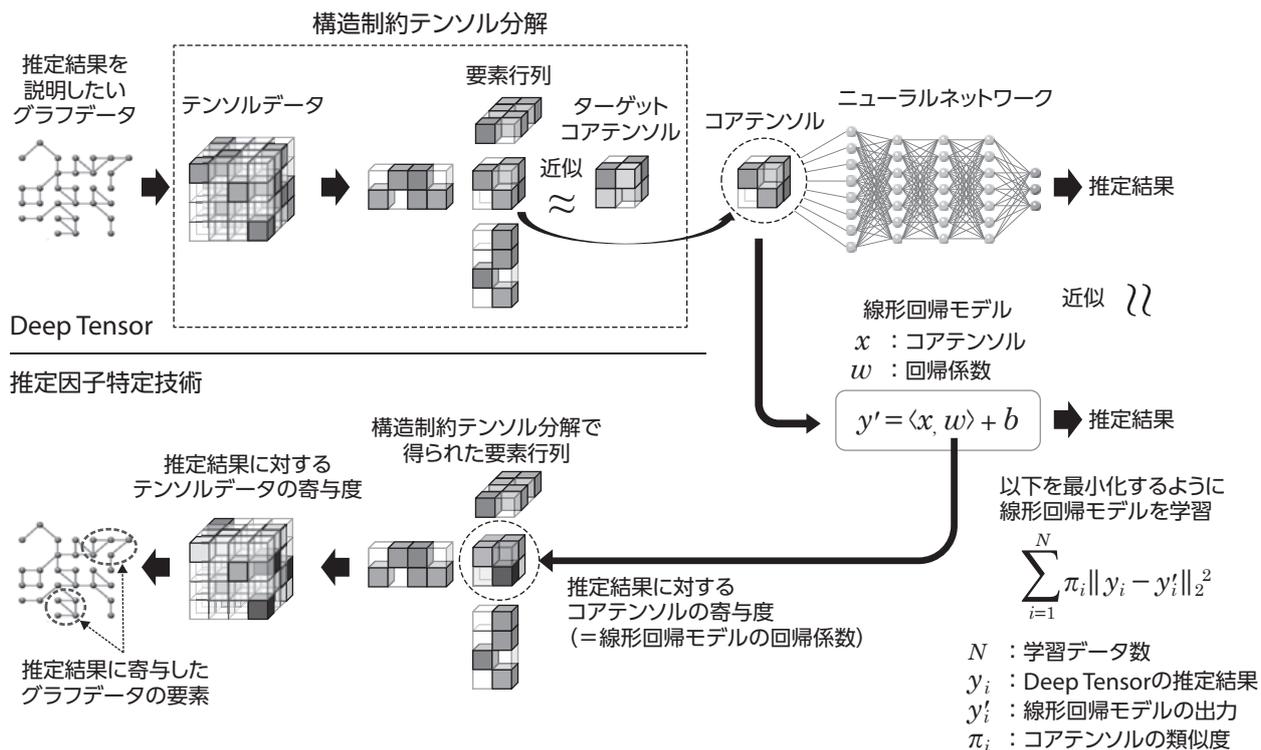


図-3 推定因子特定技術の概要

データの専門家が推定結果を合理的に解釈することが容易になる。

4. 推定因子特定技術の適用事例

本章では、医療および金融分野への推定因子特定技術の適用事例を紹介する。

4.1 医療分野への適用事例

本節では、創薬における化合物の毒性推定への適用事例を紹介する。本事例では、オープンデータであるTox21データセット⁽⁵⁾を用いて、化合物の化学構造から毒性の有無を推定する。

まず、データセット中の各化合物について、化合物中の原子をノード、原子間の結合をエッジとしてグラフデータを構築し、データセットを学習用と評価用に分割する。次に、Deep Tensorで学習用データセットを学習し、学習後のDeep Tensorで評価用データセットを推定する。この推定結果に推定因子特定技術を適用し、化合物中の各結合について推定結果への寄与度を得る。最後に、寄与度の高い結合

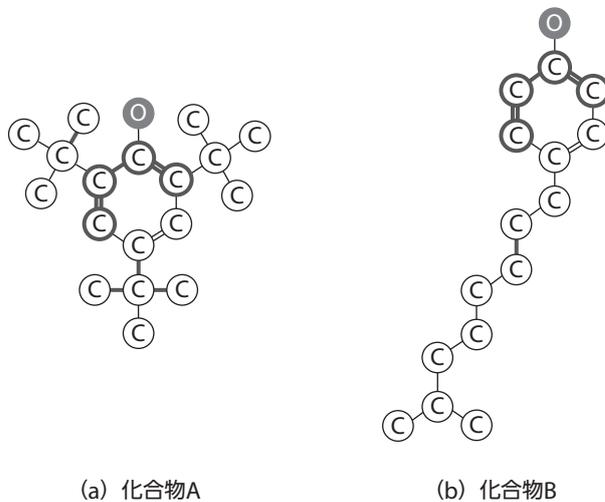


図-4 毒性ありの化合物に対する適用結果

を視覚的に確認する。

Deep Tensorが毒性ありと推定した二つの化合物の化学構造を、図-4に示す。この図において、推定因子特定技術によって得られた推定結果への寄与度が相対的に高い結合を、太線で示している。化合物A、Bの寄与度が高い結合を比較すると、共通する

化学構造（太丸枠の原子による構造）が存在することがわかる。

ここで、Deep Tensorの推定においてこの二つの化合物のコアテンソルは類似していた。これは、推定に寄与する化学構造に何らかの類似点があることを意味しており、この共通する化学構造がその類似点であると考えられる。毒性の推定結果に寄与する化学構造として共通に現れたものであるため、この化学構造は毒性推定において注目すべき構造である可能性が示唆された。

以上のように、推定因子特定技術を使用することで、推定結果に寄与した化学構造がわかった。また、コアテンソルが類似したデータを比較することによって、推定結果に寄与する共通の化学構造も明らかになった。これらの結果に合理的な解釈があれば、Deep Tensorの推定が妥当であると判断できる。

4.2 金融分野への適用事例

本節では、金融機関における与信審査への適用事例を紹介する。本事例では、金融機関が保有する企業間の取引履歴を用いて、企業への投融資のリスクを推定する。具体的には、取引履歴から審査対象企

業に関する企業間の取引関係グラフを構築し、このグラフの特徴から投融資のリスクが高いまたは低いどちらであるかをDeep Tensorで推定する。なお、基本的な手順は前述した毒性推定の事例と同様である。

取引関係グラフの一例を、図-5に示す。これらは、審査対象企業であるA社およびB社がどの企業と取引したかという関係性を、取引時期も考慮してグラフ化したものである。丸ノードは取引先企業、四角ノードは取引時期を表している。また、図中には明示していないが、取引関係グラフのエッジには取引金額の情報が含まれている。

A社とB社の取引規模（取引企業数や取引金額）は同程度であるが、取引関係グラフの特徴は異なる。A社の取引先には、ある一時期にのみ取引した企業（以下、一見客）が多い。一方、B社では、複数時期にわたって取引した企業（以下、常連客）が多い。このような特徴から、Deep TensorはA社を高リスク、B社を低リスクと推定した。

図-5において、推定因子特定技術によって得られた推定結果への寄与度が相対的に高かった取引関係を、点線枠で示す。A社では一見客との取引関係の

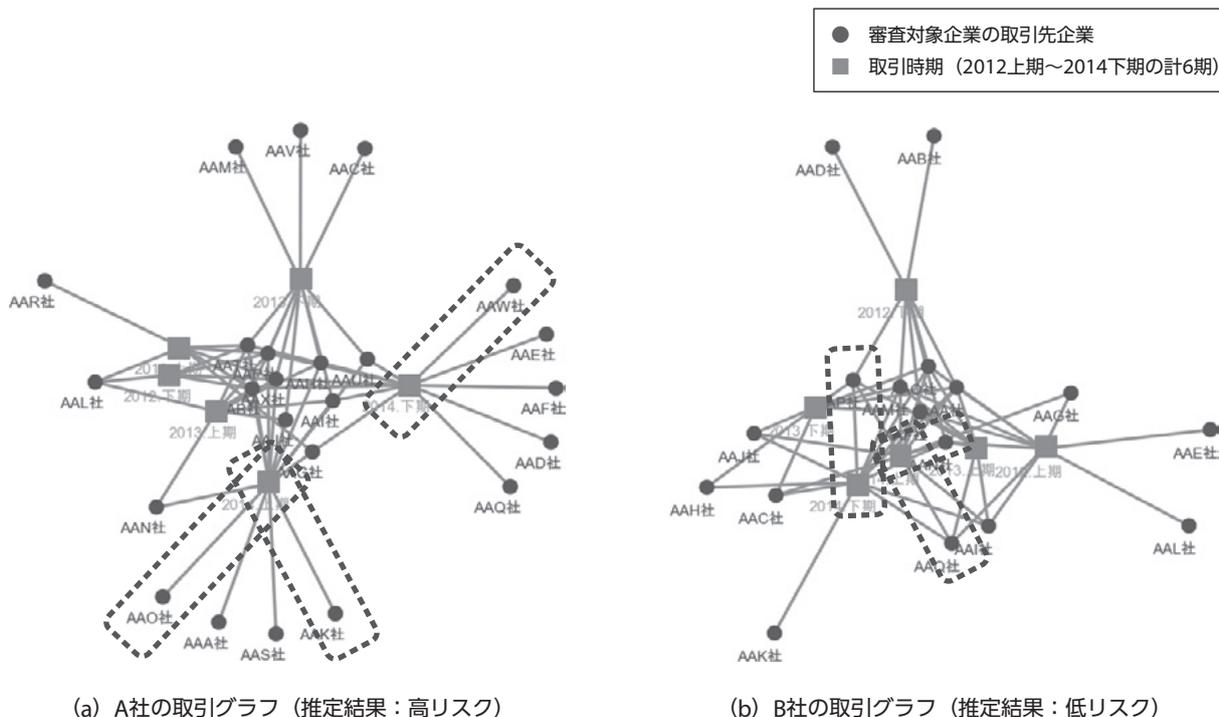


図-5 取引関係グラフに対する適用結果

寄与度が高いため、一見客に依存した取引関係を懸念していると考えられる。一方、B社では常連客との取引関係の寄与度が高いため、常連客との安定した取引関係を評価していると考えられる。

このように、推定因子特定技術を用いることで、同規模の取引グラフを有するが投融資のリスクは異なる企業に対して、Deep Tensorの推定結果を左右する取引関係が明らかになった。これらの取引関係の特徴に注目することで、合理的と思われる解釈が得られた。

5. むすび

本稿では、Deep Tensorの推定結果を説明する推定因子特定技術とその適用事例について述べた。推定因子特定技術では、コアテンソルに基づいてDeep Tensorの推定結果を近似する線形回帰モデルを学習する。この学習で得られた推定結果に対するコアテンソルの寄与度をグラフデータの寄与度に変換し、グラフデータの各エッジに対する推定結果への寄与度を得る。本技術を医療および金融分野に適用し、Deep Tensorの推定結果に寄与したグラフデータの要素を提示できることを示した。

本稿で述べた技術による説明とは、推定に寄与した要素を提示することであるが、これだけでは十分な説明とは言えない。より良い説明とするためには、それらの説明がデータの背後にある現象と結びつけて理解できるものでなければならない。それにより初めて、法や倫理を考慮に入れた判断や、新たな発見を促すことなどが可能となる。富士通研究所は、この課題を解決するための方向性の一つとして、推定因子特定技術とナレッジグラフを組み合わせた「説明可能なAI」を発表した。⁽⁶⁾ 今後は、我々の「説明可能なAI」をより実用的なものとして発展させ、AIプラットフォームサービスFUJITSU Human Centric AI Zinraiにおいて活用していく予定である。⁽⁷⁾

本稿に掲載されている会社名・製品名は、各社所有の商標もしくは登録商標を含みます。

参考文献

(1) 富士通研究所：人やモノのつながりを表すグラフ構造

のデータから新たな知見を導く新技術「Deep Tensor」を開発。

<https://pr.fujitsu.com/jp/news/2016/10/20.html>

(2) 丸橋弘治：人やモノのつながりを表すグラフデータから新たな知見を導く新技術 Deep Tensor. FUJITSU, Vol.68, No.5, p.29-35 (2017).

<https://www.fujitsu.com/jp/documents/about/resources/publications/magazine/backnumber/vol68-5/paper05.pdf>

(3) K. Maruhashi et al. : Learning Multi-way Relations via Tensor Decomposition with Neural Networks. Proc. of Thirty-Second AAAI Conference on Artificial Intelligence (AAAI 2018), 2018.

<https://pdfs.semanticscholar.org/bab9/54548db21034436ee4def167664a1790ba86.pdf>

(4) M. T. Ribeiro et al. : “Why Should I Trust You?” :

Explaining the Predictions of Any Classifier. Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016), p.1135-1144, 2016.

<http://www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf>

(5) U.S. Department of Health & Human Services : Tox21 Data Challenge 2014.

<https://tripod.nih.gov/tox21/challenge/data.jsp>

(6) 富士通：AIの推定理由や根拠を説明する技術を開発。

<https://pr.fujitsu.com/jp/news/2017/09/20-1.html>

(7) 富士通：FUJITSU Human Centric AI Zinrai.

<https://www.fujitsu.com/jp/solutions/business-technology/ai/ai-zinrai/>

著者紹介



松尾 達 (まつお たつる)

(株) 富士通研究所

人工知能研究所

機械学習における説明可能性の研究開発に従事。



大木 雄介 (おおき ゆうすけ)

(株) 富士通研究所
人工知能研究所
機械学習における説明可能性の研究開発に従事。



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
機械学習における説明可能性の研究開発に従事。