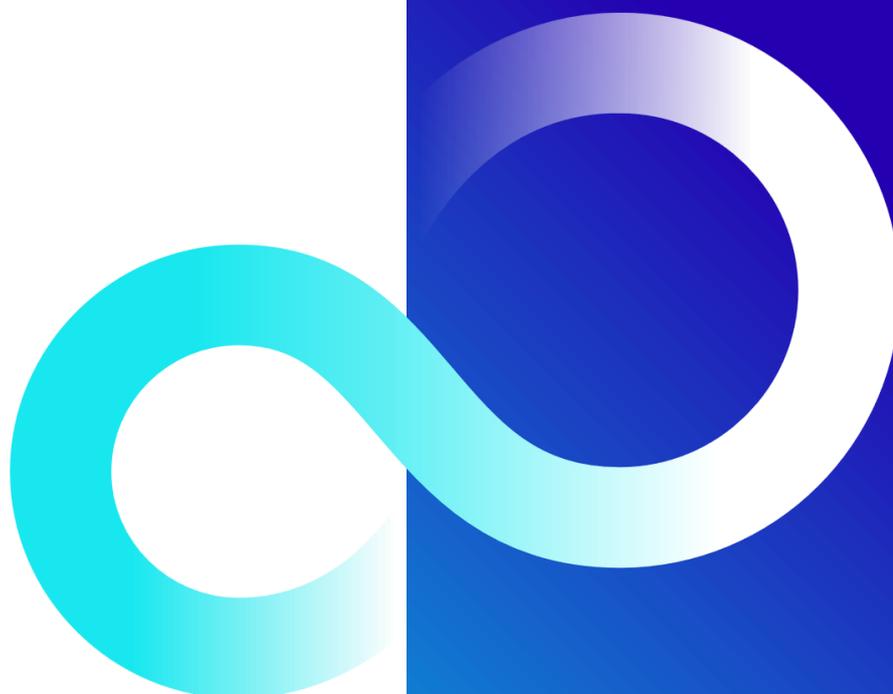


Fujitsu AI White Paper

FUJITSU

# Fujitsu Causal Knowledge Graph

データドリブンな意思決定を変革する  
因果ナレッジグラフ技術



富士通株式会社  
富士通研究所 人工知能研究所  
第1版：2024年10月発行

# 目次

1. はじめに .....	3
2. 因果分析とナレッジグラフの技術動向.....	4
2.1. 因果分析の技術動向 .....	4
2.2. ナレッジグラフの技術動向.....	6
3. 富士通が目指す因果ナレッジグラフ.....	10
3.1. 因果ナレッジグラフの基本コンセプト：なぜ因果ナレッジグラフが必要なのか .....	10
3.2. 因果ナレッジグラフの定義：因果ナレッジグラフの中身は.....	11
3.3. ユースケース：因果ナレッジグラフが提供する価値とは.....	13
4. 富士通の因果ナレッジグラフ関連技術.....	16
4.1. 各技術の位置付け.....	16
4.2. 因果ナレッジグラフ構築技術 .....	18
4.3. 因果ナレッジグラフ活用技術 .....	21
5. 因果ナレッジグラフ事例集.....	24
5.1. 事例 1：ネットワークの障害原因分析 .....	24
5.2. 事例 2：商談データとエンゲージメントサーベイを統合した因果分析.....	26
5.3. 事例 3：睡眠とライフスタイルに関するデータの因果関係の推定.....	28
6. おわりに.....	31
A. 参考文献.....	32
B. これまでの富士通の実績.....	34

# 1. はじめに

富士通では、不確実な状況下でも多様なデータに基づいて効率的に意思決定できる世界を目指している。VUCA（Volatility・Uncertainty・Complexity・Ambiguity）とも呼ばれる変化の激しい時代で効率的な意思決定を行うためには、自組織が持つデータだけでなく、周辺組織や地域、他分野の動向も含めた複数のデータを統合的に分析し、データ項目間の因果関係に基づいた論理的な推論を行うことが必要不可欠である。

近年登場した大規模言語モデル（LLM：Large Language Model）は、様々な種類の膨大な文書によって学習されており、幅広い知識を自然言語のインターフェースで活用可能で、意思決定の支援においても有用なツールとなりつつある。ただし、LLMの判断は、内包する一般的な知識に基づくものであるため、特定の専門的分野での最新のデータに基づく意思決定には不向きである。例えば、ある企業の利益率向上を目指すユースケースで、その企業の売上に重要な要素である、従業員満足度、店舗数、気候、口コミといったデータをLLMに与えたとしても、その要素同士の因果関係が明確でなければ有効な支援は期待できない。

本課題の解決に向けて「因果ナレッジグラフ」を提唱する。これは、データ項目間の関係を視覚化して情報の探索や推論を支援するナレッジグラフ（KG：Knowledge Graph）と、データに内在する因果関係を推定する因果分析という、富士通が長年培ってきた二つの技術を融合させた概念である。因果ナレッジグラフにはデータに含まれる因果関係とそれに関連した知識が統合されて格納されており、これを用いることでデータに基づいた効率的な意思決定が実現可能となる。

さらに、様々な分野の因果ナレッジグラフを蓄積することで、多岐にわたる分野における深い分析と意思決定をサポートできる。例えば、経営、医療、スポーツ、製造業など、各分野で様々に関連するデータの相互の関係性や因果関係を明確にし、よりの確かな分析と予測を行うことが可能となる。これにより、異なる分野のお客様のニーズに応じた、カスタマイズされたデータ分析と意思決定支援を実現する。

本ホワイトペーパーでは、データに基づいた意思決定に必要な不可欠な「因果ナレッジグラフ」のコンセプトに加え、背景となる技術動向や、因果ナレッジグラフの構築と利用のための技術、そのユースケースについて紹介する。

## 各章の内容

2章では、因果ナレッジグラフの重要な技術要素である「因果分析」と「ナレッジグラフ（KG）」の技術動向について説明する。因果分析は主にデータマイニング・統計分析の最新技術として、KGはインターネット時代の知識表現形式として、それぞれ別々に発展してきた。これら2つの技術に関して、基礎、研究動向、また最先端の取り組みであるLLMとの相互利用例を解説する。

3章では、データドリブンな意思決定を変革するための富士通が考える因果ナレッジグラフに関する、基本コンセプト、因果ナレッジグラフの定義、ユースケースについて説明する。これにより、なぜ因果ナレッジグラフが必要なのか、因果ナレッジグラフの内部構造、因果ナレッジグラフが提供する価値を解説する。

4章では、因果ナレッジグラフを構築・活用する富士通の技術群について解説する。大きくは既存の文献から既知の因果関係を抽出する技術（文献からの因果抽出）、数値データから未知の因果関係を推定する技術（統計的因果探索）、さらに複数の数値データから因果関係を推定する技術（統合因果探索）、既知の因果関係を前提知識として利用して因果関係を推定する技術（因果知識転用）、既知の因果関係と未知の因果関係を因果ナレッジグラフとして統合する技術（因果知識化）、因果ナレッジグラフを活用する技術（根本原因分析・因果意思決定・因果ナレッジグラフ推論）から構成されている。

5章では、4章で紹介した技術群を「ICTサポート」「人事」「健康・ヘルスケア」の3分野に適用した事例を紹介する。

## 2. 因果分析とナレッジグラフの技術動向

富士通が目指す因果ナレッジグラフの説明に入る前に、本章では、因果ナレッジグラフの重要な技術要素である「因果分析」と「ナレッジグラフ（KG：Knowledge Graph）」の技術動向について説明する。因果分析は主にデータマイニング・統計分析の最新技術として、KGはインターネット時代の知識表現形式として、それぞれ別々に発展してきた。以下の節では、因果分析とKGにおける基礎、研究動向、また最先端の取り組みである大規模言語モデル（LLM：Large Language Model）との相互利用例を解説する。

### 2.1. 因果分析の技術動向

#### ■ 因果分析の基礎

因果分析とは、原因と結果の関係を明確にするための手法であり、特定の原因が結果にどのような影響を与えるかを理解し、予測することを目的とした技術である。「モノ」や「コト」の因果関係を解明することで、複雑な構造や仕組みを紐解くことができることから、科学、医療、経済学、社会科学など幅広い分野で応用されており、政策立案やビジネス戦略の策定などにも使われる技術である。

既知の因果を分析するときは、予め分析者が事前に因果関係を設定し、「因果の方向性（どちらが原因でどちらが結果か）」や、「隠れた共通因子」が現れないように因果関係を定義した上で、実際の効果の有無や程度を検証する。しかし、未知の因果関係を調べるためには、このような分析者による事前の設定は不可能であるため、実験的なアプローチや統計的なアプローチを用いる必要がある。

統計的なアプローチの代表例として、「統計的因果探索」がある。これは、因果関係が自明でない事象群に対し、統計的な処理を用いて、データから因果関係を推定する技術である。これにより、分析者は因果関係が自明でないモノ・コトに対しても因果分析することが可能となる。

因果分析における最も重要な課題は、「正しい因果関係」を精査することにある。上記のような統計的因果探索を用いても、常に正しい因果関係が推定できる保証はなく、従来は、専門家による手作業での確認が必要だった。最近ではLLMの登場により、専門家の確認作業をLLMに代替させる試みが登場している。以下では、代表的な手法である統計的因果探索の技術動向と、因果分析におけるLLMの利用例について解説する。

#### ■ 統計的因果探索の技術動向

統計的因果探索とは、ある観測された数値データから原因と結果の関係を統計学的なアプローチを用いて推定する手法である。この手法では、単なる相関関係ではなく、特定の変数が他の変数にどのような影響を与えるかを明らかにすることにより、経営戦略や治療方針の策定などにおいて、施策の影響や介入の効果を科学的に検証することが可能となる。

統計的因果探索は、データ内の各項目（例：食生活、運動、健康など）をノードとし、推定された原因と結果に対応するノードをエッジで接続した因果グラフを出力する（図 1 参照）。因果グラフは有向グラフであり、矢印の元になるノードが原因、矢印の先になるノードが結果を表す。また、エッジには原因が結果に与える影響度を表した「因果効果」が付与される。通常、因果効果は正負の符号が付いた数値で表され、正の値ならば原因が増加するに従って結果も増加、負の値ならば原因が増加するに従って結果は減少することを意味する。

統計的因果探索の代表的な手法として、Pearlが導入した構造方程式モデル（Structural Equation Model） [1]とRubinが導入した潜在的結果理論 [2]がよく知られている。構造方程式モデルは原因から結果への因果効果を方程式として記述することで、因果関係の流れを視覚的に捉え厳密に定式化することを可能にした。

様々な仮定の下で正当化される 構造方程式モデルに従う観測データの因果探索アルゴリズムは大きく以下の3種類に分けられる。

- **Constraint-based methods** : 因果関係の問題を条件付き独立の問題に置き換え、マルコフ同値類（同じ条件付き独立構造を共有するクラスを推定する。PCアルゴリズム [3]などが有名。
- **Score-based methods** : 因果グラフに対し定義されるスコアを最適化して観測データの背後にある因果グラフを推定する。GESアルゴリズム [4]など。
- **Function-based methods** : 構造方程式モデルを記述する方程式を直接推定する。線形加法的かつ誤差変数が非ガウスのモデルはLiNGAMと呼ばれ、DirectLiNGAM [5]などが有名。近年は非線形 [6]など多方面への拡張が知られている。



図 1：統計的因果探索の入出力例

## 因果分析の実用例およびサービス事例

現在、多くの企業が統計的因果探索を利用した因果分析サービスを提供しているが、各企業がサポートする因果探索技術は様々である。すなわち、非線形モデルにも適用可能とする企業や、因果探索アルゴリズムとしてconstraint-basedなアプローチを採用する企業、Pearlの構造方程式モデルではなくRubinの潜在的結果理論に基づいた因果探索を行う企業などがある。因果探索よりも介入効果推定や最適介入値推定に重点を置く傾向が見られる企業もある。

また、因果探索支援プロダクトも提供されており、例えば、多様な探索アルゴリズムを顧客が自分で選んで実行できるよう初心者でも使いやすいインターフェースが提供され、さらにユーザが特定のエッジに関する有無を指定することで専門的知識を活用できるようになる機能などが提供される場合もある。このように、統計的因果分析をサービスとして提供する複数のプレイヤーが存在する状況であるが、基本的にはユーザが保有する個々の観測データから因果構造を推定する機能、あるいはコンサルを提供するのが一般的である。

## 因果分析における大規模言語モデルLLMの利用例

一般に因果分析では、推定されたデータ項目間の因果構造を専門知識に基づいて解釈していく必要がある。また、因果構造を推定するにあたり、専門知識として確定された因果関係がある場合は、前提知識として組み入れる必要がある。従来はこのような作業は専門家が労力をかけて行っていたが、近年のLLMの急速な発展により、このような作業をLLMにサポートさせる研究がいくつか登場している。

例えば、LLMに対して妥当な因果関係であるかを問い合わせながら統計的因果探索を繰り返すことで、専門的知識と矛盾しない因果構造を推定できる可能性が報告されている [7]。ただし、統計的因果探索において最も重要な課題の一つである相関関係と因果関係を論理的に区別することに関しては、現状のLLMは著しく性能が低いという報告もある [8]。

企業が提供するサービスとしては、現時点ではLLMを用いて統計的因果探索の入力や出力をサポートする程度であり、LLMにより因果探索アルゴリズムを飛躍的に高度化させるような事例はまだ確認することはできない。例えば、LLMをインターフェースとして用い因果グラフを根拠として「解約率の原因は何？」のような質問に回答する、LLMに統計的因果探索の事前知識や未観測共通原因の候補をサジェストさせる、因果分析の結果をLLMに説明させる、というような因果探索自体を高度化させるものではなく、因果分析のユーザビリティを向上させるものがほとんどである。LLMとの組み合わせによる統計的因果探索の高度化については、まだまだ多くの課題を解決していく必要がある。

## 2.2. ナレッジグラフの技術動向

### ナレッジグラフの基礎

ナレッジグラフ（以下、KG）は、1972年頃から用語として使われ始め、1980年代から継続して研究開発されている知識表現である。さまざまな定義があるが、包括的には、実世界の知識を蓄積して伝達することを目的としたグラフ構造のデータといえる。ここで、KGのノードは関心のあるエンティティ（例えば、人、場所、事物、概念など）を表し、エッジはこれらのエンティティ間の関係（例えば、「XはYの一部である(is\_part\_of)」、「XがZを所有する(own)」など、図 2参照）を表す。

情報やデータをKGとして表すことの本来的な利点は以下である。

- **知識の統合と共有と更新**：複数のデータソースからの情報を、統一されたグラフ構造で表現することで、形式や構造などが異なるデータも互換性を保ったまま統合と共有が容易であり、新規のデータの追加更新も柔軟に対応できる。
- **複雑な関係の表現**：関係データベース（RDB）では表現が難しい多対多の関係や循環的な関係、階層的な関係を含め、エンティティ間の複雑な関係を容易に表現できる。
- **効率的かつ柔軟な検索**：RDBでは複数のテーブルをまたがる複雑なクエリや、エンティティに関する関係性を含んだクエリを効率的に実行できる。
- **推論と知識発見**：既存の知識から関係性などを用いて新たな知識を導出する演繹的推論とグラフ構造の規則性を見つけて新たな知識を抽出する帰納的推論の両方を適用でき、知識発見や未知の情報の推測が可能になる。

以下では、KGの研究動向と、因果分析におけるLLMの利用例について解説する。

### ナレッジグラフの研究動向

ひとくちにKGと言ってもその技術領域は幅広く、研究分野も大きく知識表現、知識獲得、推論の3つの分野に分けられる。

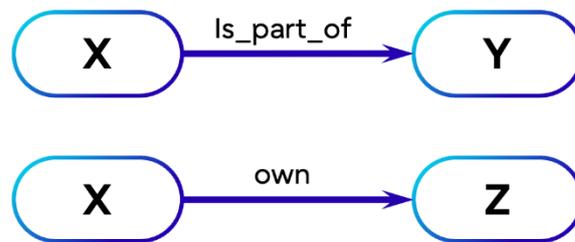


図 2：簡単なナレッジグラフの例

知識表現とは、計算機が人の知識を理解し推論できるように、情報や概念を計算機上で表現する技術である。KGの代表的な知識表現には、RDF（Resource Description Framework） [9]とプロパティグラフがある。RDFはW3Cによって標準化されており、データを「主語-述語-目的語」のトリプルで表す。例えば、「XはYの一部である」は、「<X>-<一部>-<Y>」と表される。また、URIでリソースを識別可能とし、オントロジーなどの標準的な語彙を用いることで、データの意味を規定し、相互運用性や再利用性を高めている。プロパティグラフは、データを構成するノードとエッジに対して、プロパティと呼ばれるキーと値からなる複数の属性を付与したグラフデータである。プロパティを柔軟に追加でき、グラフデータベースとの親和性が高い。なお、近年はRDFの拡張として、W3Cでの標準化はまだされていないが、トリプルに対してメタ情報を定義でき、ハイパーグラフの表現力を持つRDF-starが提案されている。

知識獲得は、テキストやデータベースなどから、エンティティとその関係を抽出して、KGを構築、拡張する技術である。関係抽出、エンティティ発見とエンティティリンク、KG補完とリンク予測の3つの観点で研究されている。関係抽出はテキストからエンティティに関する関係性を抽出する技術であり、KGの自動構築に使われる。エンティティ発見、エンティティリンクは、テキストからエンティティを抽出してKGに結びつける技術であり、この技術によりエンティティの曖昧性が解消される。

KGにおける推論は、KGから新たな知識を導き出したり、未知の情報を推測したりする推論技術である。演繹的推論は、「ソクラテスは人間である」「すべての人間はいつか死ぬ」という既存の知識から、演繹的に「ソクラテスはいつか死ぬ」といった新たな知識を導出する技術である。帰納的推論は、機械学習技術を適用して、KG上でパターンや規則性を発見することで新たな知識を生成する技術である。また、近年、深層学習を用いてエンティティと関係をベクトル空間に埋め込み、その類似性に基づいて推論するグラフ埋め込み技術が研究、応用されている。これらの推論結果をエンティティ間の関係性などを用いて人が理解可能な形で説明する技術も研究されている。

LLMが飛躍的に発達した近年は、KGとLLMを組み合わせた研究や応用が数多く発表されている。

## ナレッジグラフの実用例およびサービス事例

KGを用いた実用例としては、検索サービスや質問応答システム、推薦システムなどが挙げられる。検索サービスでは検索クエリに含まれるエンティティに関する情報を提示するのに使われている。質問応答システムでは、KG上でエンティティ間の関係をたどることで、より適切な回答を返す効果がある。推薦システムでは、ユーザが過去に購入した商品のジャンルなどの情報から関連する別の商品を提示するのに商品情報のKGを用いている。

また、一般的な情報や特定分野に関するKGが既にいくつも公開されている。一般的な情報としては、Wikipediaから抽出したデータを基にしたDBpedia [10]、WikipediaとWordNetを統合したYAGO [11]、DBpediaと人の寄稿を基にしたFreeBaseを含めて構築されたGoogle Knowledge Graph [12] がある。特定分野で公開されている例としては、生物医学分野では、薬物に関する情報をまとめたDrugBank [13]、DrugBankにPubMedの論文情報やその他の情報を統合したROBOKOP [14]やBio2RDF [15]がある。他には、金融業界の標準的な共通言語を提供している Financial Industry Business Ontology [16]、イギリスの政府系情報を提供しているGOV.UK [17]などがある。これらインターネット上で公開されているKGの多くは、リソースの識別可能性があり、また相互運用性が高いRDFで記述されている。

KGを格納するデータベースは、大きくは知識表現に対応してRDFストアとグラフデータベースの二種類が挙げられる。代表的なRDFストアとして、Virtuoso [18]やGraphDB [19]が挙げられる。代表的なグラフデータベースは、Neo4j [20]やAmazon Neptune [21]がある。問い合わせ言語は、RDFストアはトリプルに対する検索が可能なSPARQLに対応していて、Neo4jなどはグラフデータベース向けに設計された Cypherに対応している。

## ナレッジグラフにおける大規模言語モデルLLMの利用例

GPT4などのLLMの問題点として、LLMは単純な質問への回答は優れているが、関連する情報が連なった多段階の推論が必要となる複雑な質問への回答精度は著しく低下すること、さらに、LLMから得られた回答が正しいものであるか判断するため、回答が導かれた過程や根拠を調べるのが困難であることなどが指摘されている。

一方、KGは、本質的にその構築が難しく、特に既存の方法では新しい事実の生成や未知の知識の表現は課題である。そこで、KGとLLMを統合し、各々の長所短所を相互に補完して活用する研究開発が進んでいる。

ここで、LLMとKGの統合技術に関する最新のサーベイ論文 [22]によると、その統合形態は、以下の3つに分類される(図 3)。

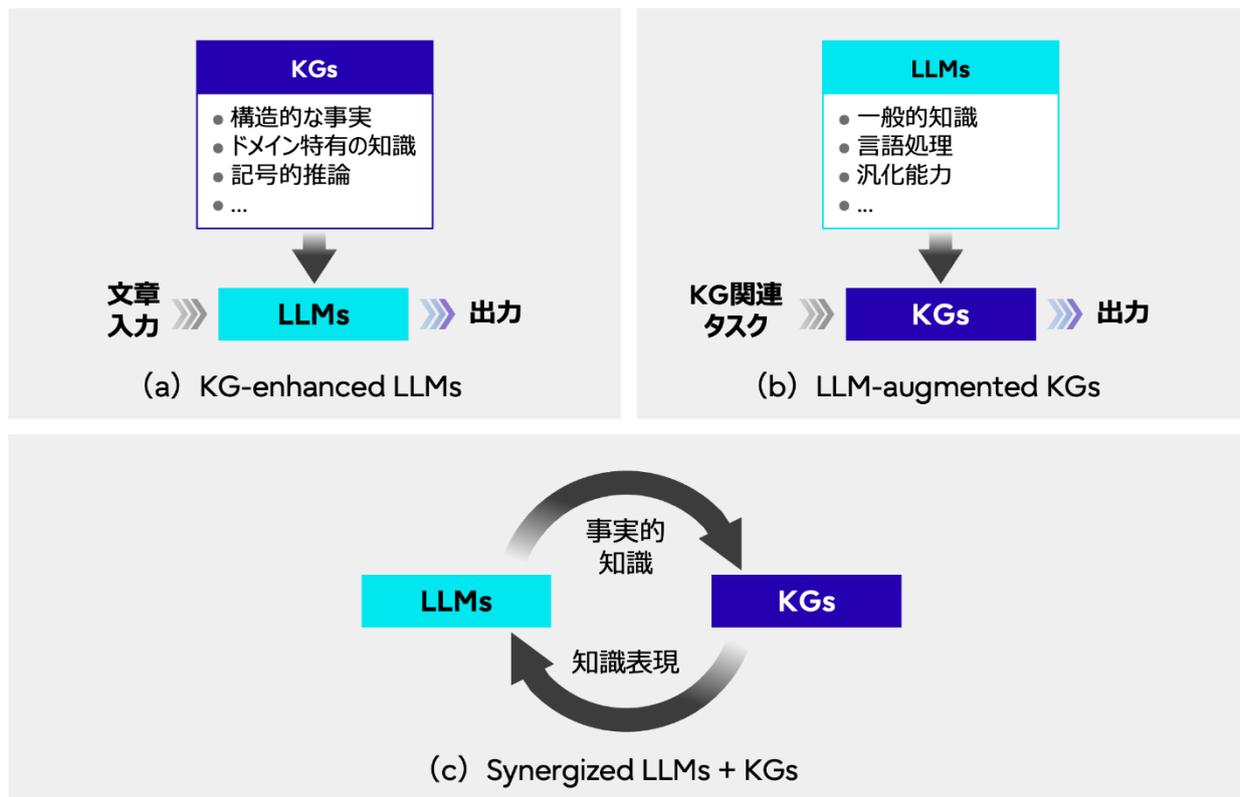


図 3 : LLMとKGの統合形態  
出典 : [22]のFig.6を元に作成

(a)KG-enhanced LLMs : KGでLLMの性能・解釈可能性を向上

- 事前学習でのKG適用、KGを用いたLLM推論、LLMが学習した知識の理解・推論プロセス解釈にKGを利用

(b)LLM-augmented KGs : LLMでKGタスクを補強

- LLMによるKG埋め込み、LLMによるKG補完、LLMによるKG生成、LLMによるKGからのテキスト生成、LLMを利用したKG質問応答

(c)Synergized LLMs + KGs : LLMとKGの相乗効果

- LLMとKGを統一されたフレームワークに統合し、相互に強化

上記のように、基礎研究の分野では、KGとLLMを統合して相互に補完しながら活用する技術開発が注目を集めている。一方、未だ数少ないものの、ビジネス分野においても両者を統合して活用する事例がいくつか存在する。例えば、企業内検

素などのLLMを用いた対話型AIアシスタントへの適用はその代表的な事例であり、KGを活用することで、複数のエビデンスから回答生成が必要な質問に対しても適切な回答を導くことができる [23] [24]。

また別の活用事例として、LLMから出力された回答に対してKGを用いて検証する技術も開発されており、例えば医療用LLMの回答精度向上にオープンな臨床データで作成したKGが活用されている [25]。これらの活用事例は図 3の「a. KG-enhanced LLMs」に該当し、LLMの課題解決にKGを活用している事例である。

同様に図 3の「b. LLM-augmented KGs」として、KGの課題解決にLLMを活用している事例も存在する。特に最近では、LLMを含む生成AIを用いることで、PDF、ウェブページ、ドキュメントなどの非構造化データからKGを迅速に作成するサービスが提供されている [26]。これらの事例の他にも、LLMとKGを統合して活用することにより、金融データのリスク評価や、法律文書の要約、交通パターン・ソーシャルメディアの分析、さらにはEC・小売業における商品・コンテンツのリコメンデーションシステムに加え、教育・ヘルスケア分野におけるパーソナライズされた教育・医療の提供などへの適用も期待されている [27]。

### 3. 富士通が目指す因果ナレッジグラフ

本章では、データドリブンな意思決定を変革するための、富士通が考える因果ナレッジグラフに関する、基本コンセプト、因果ナレッジグラフの中身、ユースケース（提供価値）について説明する。

#### 3.1. 因果ナレッジグラフの基本コンセプト：なぜ因果ナレッジグラフが必要なのか

企業経営においてデータドリブンな意思決定を実現するためには、通常のデータマイニング・統計分析で抽出される「相関」だけでは不十分である。相関には、疑似的相関や潜在的な共通因子が存在することが多く、これに基づいて論理的な意思決定を行うことは不可能と言える。例えば、「ノーベル賞受賞者の数」と「チョコレート消費量」に相関があることが発見された場合（図 4参照）、その結果を元に、社内からノーベル賞受賞者を輩出する目的で、社員にチョコレートを無料配布するような施策を打つことは無意味であろう。

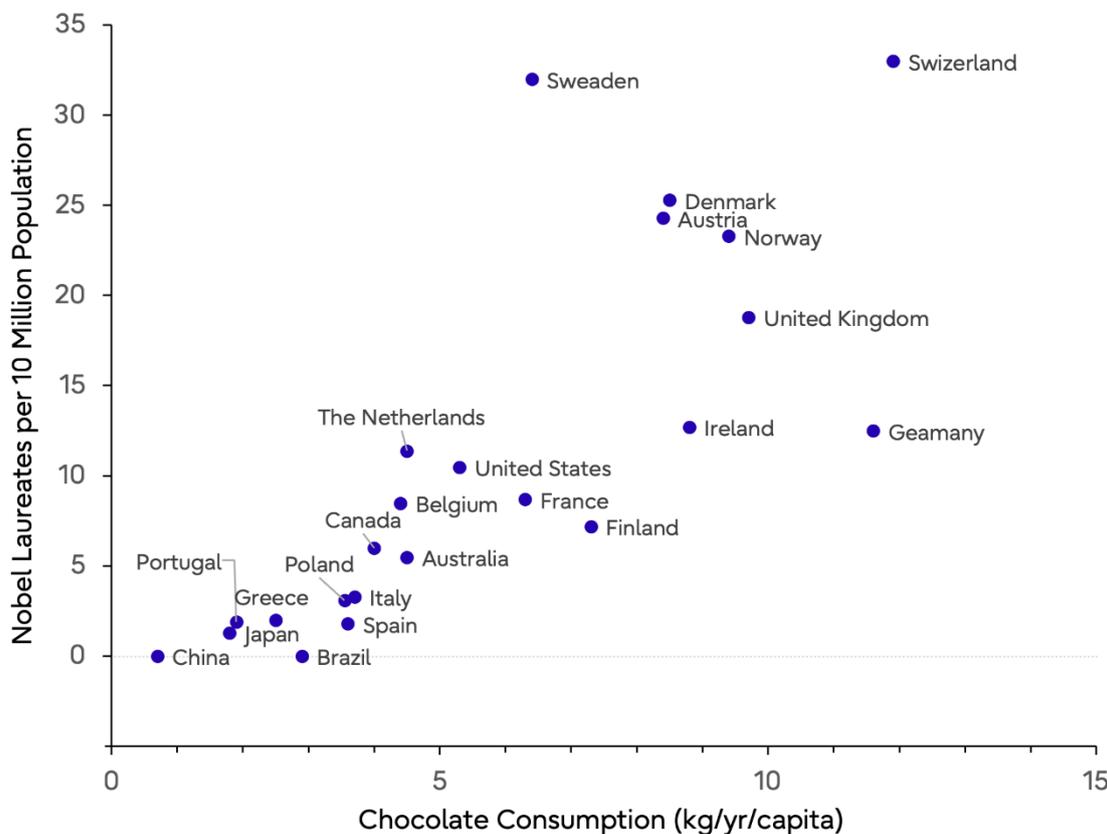


図 4：チョコレート消費量とノーベル賞受賞者の散布図

出典： [28]のFig.1を元に作成

このような施策は、因果関係を無視した表面的な相関に基づいているため、実際の成果を導く可能性は低い。したがって、企業が真にデータドリブンな意思決定を行うためには、相関を超えて因果関係を正確に捉えることが重要となる。正しい因果関係を把握することで、データに基づいた論理的かつ有効な意思決定が可能となり、ビジネスの成功に繋がると言える。

ただし、因果関係を捉えるには重大な課題がある。正確な因果関係を整備するためには、通常、有識者による膨大な確認作業が必要となり、多くのコストと時間が掛かる。その理由は2つある。1つ目は、既知の因果関係の多くは自然言語で記述（非構造化データと呼ばれる）されているため、計算機で容易に処理できる形式になっていないこと。計算機が理解できるように、有識者が情報や概念を構造化する必要がある。2つ目は、そもそも明白な因果関係が分かっていない場合が多いため、因果関係を見つけ出すために、有識者による追加の分析と検証が必要であること。これらの課題は、データドリブンな意思決定を支える因果関係の正確な把握を困難にしている。

上記2つの理由に対し、2章で見てきたように、これまで以下の2つの技術が開発されてきた。既知の因果関係に対しては、因果関係が書かれた文献からLLMなどを用いて自動的に因果関係を抽出する技術である。そして、未知の因果関係に対しては、統計的因果探索を用いてデータの項目間の因果関係を自動的に推定する技術である。

しかしながら、現行の2つの技術にはそれぞれ以下のような限界がある。まず、文献からの因果関係の抽出において、抽出される因果関係はすべて定性的なものであり、そのため、ある事象（原因）が発生した際に引き起こされる別の事象（結果）の発生確率や影響度を予測することができない（限界1）。次に、データからの統計的因果探索では、すべての因果関係をデータから推定するため、データ内のバイアスやデータ量の少なさに起因して、本来の原因と結果が逆転した因果関係を導き出してしまう可能性がある（限界2）。さらに、現行の統計的因果探索では、与えられたデータの範囲内でしか因果関係を導き出せないため、観測条件が異なる複数のデータ群（観測した対象物が異なる、観測項目が異なる、観測時の測定条件が異なる、など）を結合して因果関係を導くことができない（限界3）。こうした限界を克服するためには、より高度な技術とアプローチが必要となる。

これらの限界を突破し、データドリブンな意思決定の変革を実現する試みとして、富士通は以下の技術開発に取り組む。

- **限界1を突破する技術**：文献から抽出した因果関係とデータから推定した統計的な因果関係を統合することで、これまで定性的にしか分かっていなかった因果関係においても、因果効果や影響度を測定できるようにする。これにより、例えば、従業員エンゲージメントにおいて、組織行動論の教科書からエンゲージメントの観点と施策の間の因果関係は把握でき、それに実際の自社のエンゲージメント調査結果から推定される因果効果を組み合わせることで、定量的な期待効果も含めて施策を策定することが可能となる。
- **限界2を突破する技術**：文献から抽出した既知の因果関係を蓄積し、数値データに対する統計的因果探索を実行する際に、既知の因果関係を予め設定した上で、未知の因果関係を推定する。これにより、統計的因果探索で導かれる因果関係と既知の因果関係の矛盾を解消することが期待できる。
- **限界3を突破する技術**：異なるドメインに共通するデータ分布を自動的に判断する技術を開発することにより、観測条件や粒度が異なる数値データでも結合して統計的因果探索を実行できるようにする。これにより、大量のデータセットから学習し成長し続ける因果探索を実現する。

上記を実現するためのデータ構造として、「ナレッジグラフ(KG)」を採用する。KGは、知識表現のひとつであり、特に因果関係を自然に表現できるため、今回の目的に最適といえる。また、KGの応用分野としてデータ統合があり、文献や数値データなど複数のデータソースから抽出した因果関係を統合しやすい利点がある。

以上、これまで説明した技術群とデータ構造を有し、企業経営におけるデータドリブンな意思決定の変革を実現するのが「富士通が目指す因果ナレッジグラフ」である。

## 3.2. 因果ナレッジグラフの定義：因果ナレッジグラフの中身は

図 5に「富士通が目指す因果ナレッジグラフ」のイメージ図を示す。

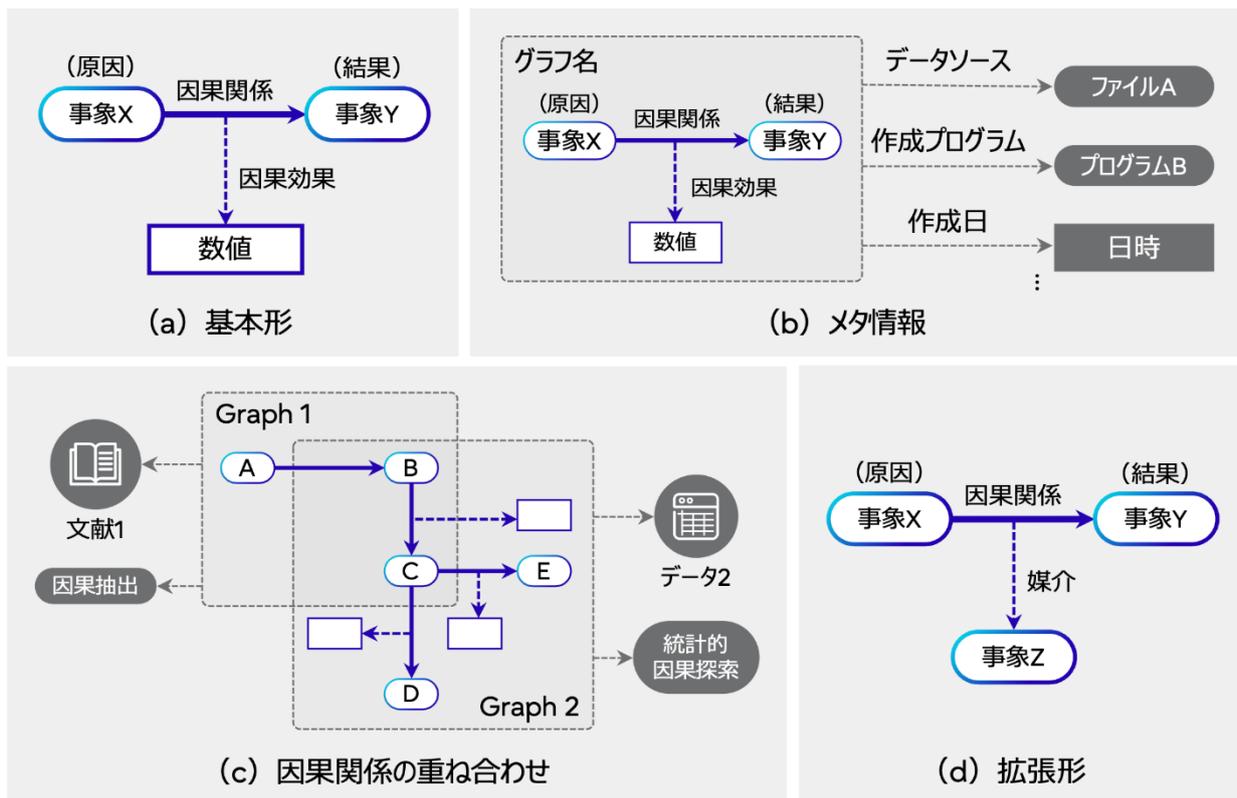


図 5 : 富士通の因果ナレッジグラフ (イメージ図)

図 5 (a) に示すように、KGを基本形態とするため、2つのノード (楕円) を1本のエッジ (矢印) でつないだ3つ組が基本形となる。ノードはある事象を表し、矢印の元になる事象が原因、矢印の先になる事象が結果を示している。よって、図 5 (a) の実線部分は、「事象X (原因) が、事象Y (結果) を引き起こす (因果関係がある)」ことを表しており、これが基本形となる。

図 5 における点線はオプションとして追加可能な情報を示しており、図 5 (a) では「因果効果」を追加した例を示している。因果効果は通常数値で表されるため、事象などのエンティティ型 (楕円) ではなく、文字列型 (四角) で示している。

図 5 (b) はメタ情報を追加した例となっており、因果関係に対し名前 (グラフ名) を付けたり、その因果関係を抽出したデータソースや作成プログラムなどの情報を追加できる。また、数値データであるならば、観測時の条件や平均・分散といった統計値を持たせてもよい。

図 5 (c) は、いくつかの因果関係を重ね合わせた例である。ここではメタ情報として「Graph1は文献1から因果抽出プログラムで作成した」こと、因果関係として「事象A (原因) →事象B (結果)」「事象B (原因) →事象C (結果)」の2つが抽出されたことを示している。同様に、Graph2では「データ2から統計的因果探索プログラムで作成した」こと、因果関係として「事象B (原因) →事象C (結果)」「事象C (原因) →事象D (結果)」「事象C (原因) →事象E (結果)」が因果効果付きで推定されたことを示している。

ここで、「事象B (原因) →事象C (結果)」は、文献1とデータ2の双方から抽出・推定されており、これを重ね合わせることで、複数のデータソースから抽出した因果関係を統合できる。尚、文献1から抽出した「事象B (原因) →事象C (結果)」は、前節の限界2である「統計的因果探索における矛盾解消」として働き、データ2から推定した因果効果付きの「事象B (原因) →事象C (結果)」は、前節の限界1である「文献から抽出した因果関係とデータから推定した統計的な因果関係の統合」に相当する。

図 5 (d) はオプションとして、因果関係自体に因果効果以外の追加情報を付与した例である。ある因果関係が起きる際の媒介や副作用などを記述できる。

### 3.3. ユースケース：因果ナレッジグラフが提供する価値とは

本節では、2つの未来シナリオを通して、因果ナレッジグラフが提供する価値について説明する。

#### 未来シナリオ1 人事部の田中さんの事例

人事部の田中さんは、職場のエンゲージメント向上を目指したアクションテイキングの支援を行っている。各職場に対し、エンゲージメントサーベイ（ES：Engagement Survey）の結果の見方や、施策の相談にのっているが、各職場に適した施策が提案できていないか心配でもある。今日も事業部長の佐藤さんから「エンゲージメントを上げるために、『やりがい』や『充実感』を上げると言われても、実際、何すればよいか分かんないよ」との相談を受けた。

そこで、田中さんは富士通の因果ナレッジグラフを試してみることにした。因果ナレッジグラフには、予め、組織マネジメントの教科書を解析し、「『スキルアップ支援』や『成長分野への人員配置』が『キャリア形成のチャンス』を増やす」ことや、「『ビジョンの明確化』や『成長分野への人員配置』が『会社の将来性』を高める」などの因果関係が格納されている（図 6 の上部）。田中さんは、これに加え、佐藤さんの部署のエンゲージメントサーベイ結果を分析させてみた。

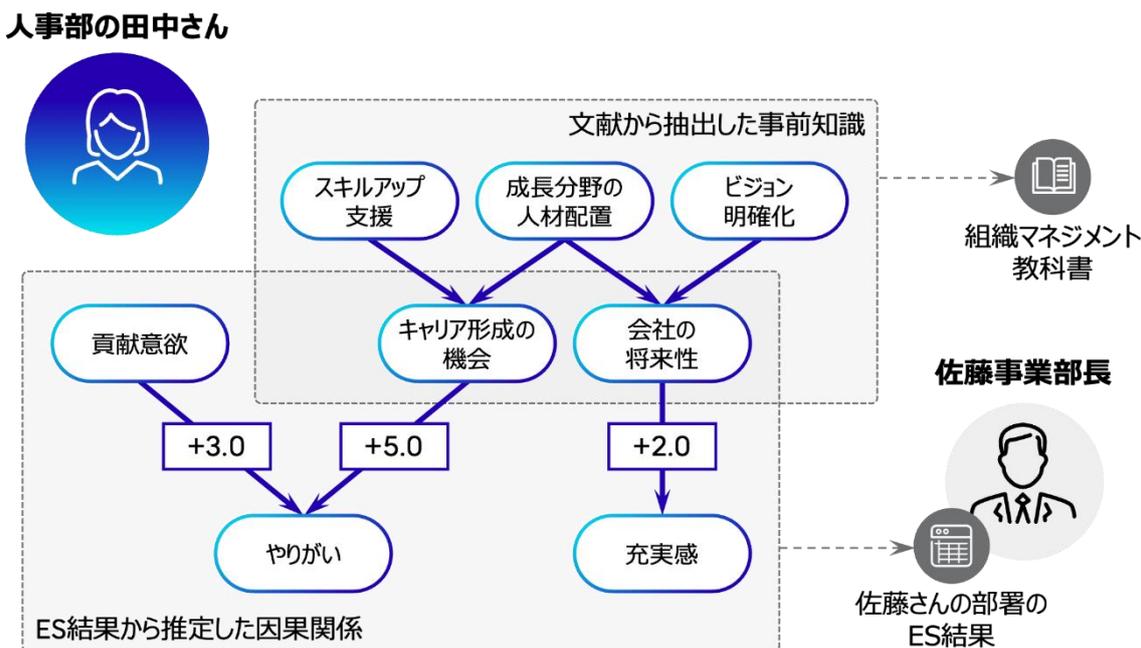


図 6：マネジメント教科書とエンゲージメントサーベイから生成された因果ナレッジグラフ

富士通の因果ナレッジグラフでは、エンゲージメントサーベイにおける各項目間の因果関係を推定することができる。そこで、「『やりがい』を上げるには、『貢献意欲』や『キャリア形成の機会』が原因」となっていること、「『充実感』を上げるには『会社の将来性』が原因」となっていること、などが分かった（図 6 の下部）。ここで、『キャリア形成の機会』や『会社の将来性』がキーとなり、既知の因果関係と推定した因果関係が結合することで、「『やりがい』や『充実感』を上げるためには、『スキルアップ支援』『成長分野への人員配置』『ビジョンの明確化』といった3つの施策が有効」であることが分かった。

そこで、田中さんは佐藤さんに『成長分野への人員配置』『ビジョンの明確化』の2つの施策を助言すると共に、他の部署のエンゲージメントサーベイ結果を分析させてみたところ、どの部署でも『やりがい』を上げるには、『キャリア形成のチャンス』が原因』であることが分かった。そのため、全社共通の『スキルアップ支援』の施策として、e-learning教育の充実を人事部長に提言することにした。

人事部長からは、「なぜ『スキルアップ支援』が重要なのか？」「その効果はどのくらいか？」を問われたが、田中さんは因果ナレッジグラフを見せながら、「『やりがい』を高めるための重要な施策であること」「『キャリア形成の機会』を増やすことで、『やりがい』が5ポイント向上すること」などを説明することができた。

## 未来シナリオ 2 営業の鈴木さんの事例

営業部の鈴木さんは、マネージャーとして毎日忙しく働いている。特に、最近は取り扱い製品への問い合わせが増え、商談件数も以前の倍に増えている。しかしながら、比較的新しい分野の製品であるため、商談の成功率もそれほど高くない。部長からは「商談の勝ち負けの要因を分析しろ」と言われている。一方で、12月のエンゲージメントサーベイ結果が返ってきており、人事部からは「エンゲージメントサーベイ結果を分析して、それぞれのチームで施策を考えろ」と催促されている。「正直、毎日の業務も忙しく、分析している暇もない」のが、鈴木さんの本音である。

ある日、鈴木さんは、とある講演で「従業員エンゲージメントを高めることが、ビジネス成功の鍵」との言葉を耳にした。内心、「ほんまかいな？」と思いつつも、富士通の因果ナレッジグラフに、鈴木さんの部署のエンゲージメントサーベイ結果と商談データを同時に分析させてみることにした（図 7右側）。

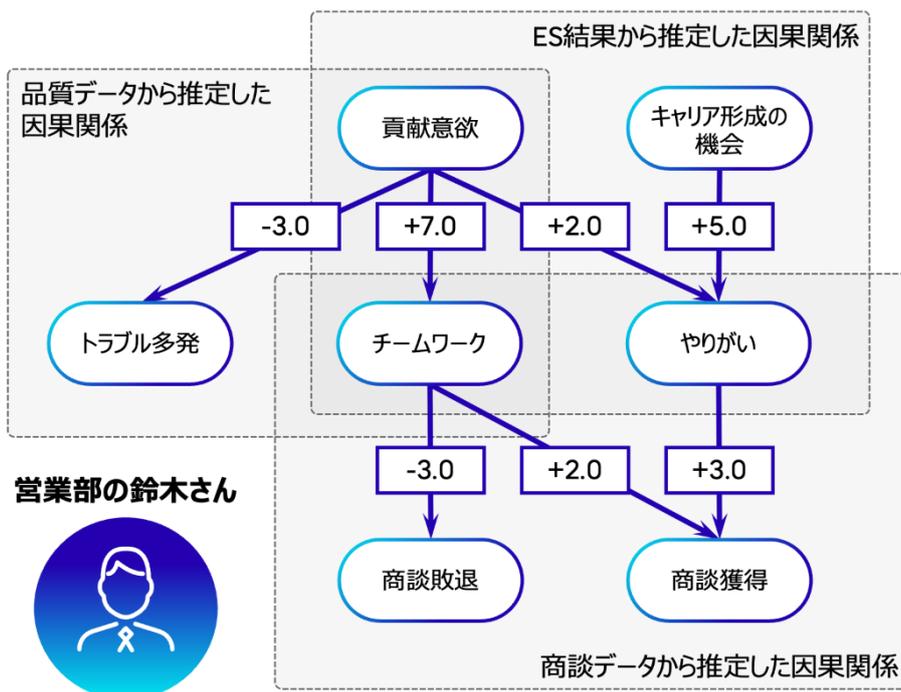


図 7：3つのデータから推定した因果関係

その結果、「商談の『獲得』や『敗退』の要因として、『チームワーク』や『やりがい』が影響」していること、「『チームワーク』や『やりがい』を上げるには、『貢献意欲』や『キャリア形成の機会』を上げる」ことなどが見えてきた。確かに、「従業員エンゲージメントを高めることが、ビジネス成功の鍵」は一理ありそうだ。

さらに、鈴木さんはプロジェクトの課題・リスクを管理している品質データも加えて、分析させてみたところ、「『貢献意欲』の低下が『トラブル多発』を引き起こしている」ことも分かった（図 7左側）。

ここで、富士通の因果ナレッジグラフの一機能である因果意思決定機能（4章で後述）を使い、『商談獲得』数の増加と『トラブル多発』の減少を同時に解決する要因を分析させてみたところ、最も効果的な施策として「『貢献意欲』の向上」であることが分かった。

鈴木さんはこれらの分析結果を部長と共有すると共に、鈴木さんチームの最重要課題を「『貢献意欲』の向上」と定め、早速、『貢献意欲』を上げるための施策について、人事部の田中さんに相談しに行った。

## 4. 富士通の因果ナレッジグラフ関連技術

本章では、因果ナレッジグラフを構築・活用する富士通の技術群について解説する。大きくは既存の文献から既知の因果関係を抽出する技術（文献からの因果抽出）、数値データから未知の因果関係を推定する技術（統計的因果探索）、複数の数値データから推定した因果関係を統合する技術（統合因果探索）、既知の因果関係を前提知識として利用して因果関係を推定する技術（因果知識転用）、既知の因果関係と未知の因果関係を因果ナレッジグラフとして統合する技術（因果知識化）、因果ナレッジグラフを活用する技術（根本原因分析・因果意思決定・因果ナレッジグラフ推論）から構成されている。尚、ここで紹介する因果ナレッジグラフ関連技術は、現在構想段階にあり、現時点では基本機能の一部が実装されている。

### 4.1. 各技術の位置付け

図 8 に因果ナレッジグラフの関連技術と全体構成を示す。はじめに、各種文献から既知の因果関係を抽出し、グラフDBに格納する。これは、因果ナレッジグラフの一部を構成すると共に、統計的因果探索の事前知識として使用される（限界2の突破）。つづいて、単体の数値データに対しては統計的因果探索を用いて未知の因果関係を推定し、同様にグラフDBに格納する。複数の数値データに対しては統合因果探索を用いて、それぞれの前提条件や平均・分散などの統計値を合わせた上で因果関係を推定し（限界3の突破）、グラフDBに格納する。最後に文献から抽出した因果関係と数値データから推定した因果関係に対し、それぞれのノードを「事象」としてエンティティ化することで、因果ナレッジグラフを構築する。この時、文献・数値データのどちらにも共通して現れる事象を同一の事象として表すことで、因果ナレッジグラフとして統合する（限界1の突破）。構築した因果ナレッジグラフは、後述する活用技術だけでなく、既知の因果関係として、統計的因果探索や統合因果探索の前提知識として利用する（限界2の突破）。

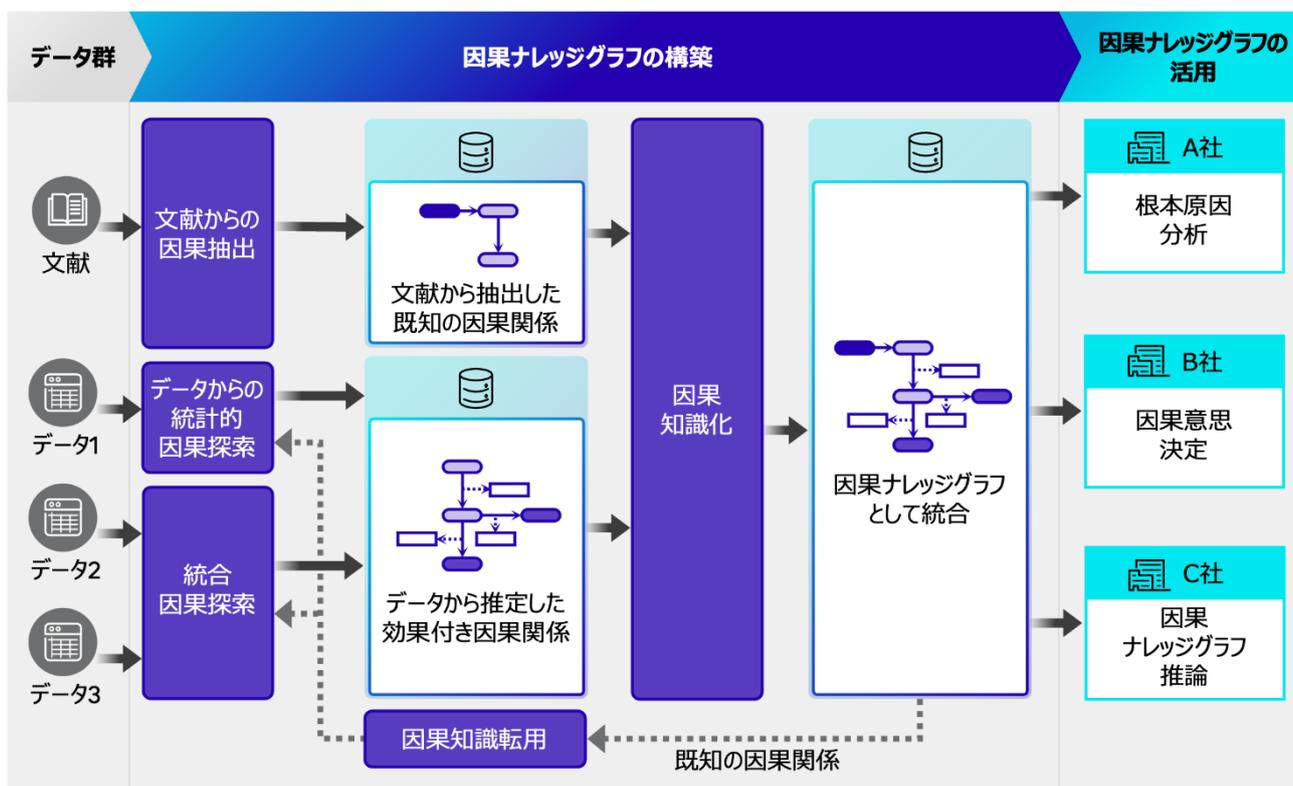


図 8 : 因果ナレッジグラフ関連技術の全体構成

続いて、因果ナレッジグラフの活用技術である。因果の連鎖を逆向きに辿ることで根本的な原因を究明する根本原因分析（RCA：Root Cause Analysis）、ある事象の発生による副作用や悪影響をも考慮した因果意思決定、複数の因果関係から新たな因果を導出する因果ナレッジグラフ推論など、幅広い応用がある。また、顧客先のデータも同様に因果ナレッジグラフ化することで、顧客先のデータを反映したアプリケーションの開発も可能である。

表 1に各技術の概要を示す。

表 1：因果ナレッジグラフ関連技術の概要

	技術	概要	入出力
因果ナレッジグラフの構築	文献からの因果抽出 (限界2を突破する技術)	ナレッジグラフの構造の型を事前に定義したナレッジグラフスキーマに基づいてナレッジグラフを生成するための処理フローとプロンプトを自動的に変換する技術により、文書から自動的に因果関係を抽出し因果ナレッジグラフを作成することを可能にする技術。	入力：文書データ 出力：因果ナレッジグラフ
	データからの統計的因果探索	データ全体にあてはまる一般的な因果関係ではなく、特定の条件のデータにのみあてはまる因果関係を知りたいニーズは高い。そのため、因果関係のあらゆるバリエーションを網羅するように全ての条件を高速に列挙する技術により、未知の因果関係が現れる条件を効率的に探索する。	入力：多変数の数値データ 出力：因果グラフ
	統合因果探索 (限界3を突破する技術)	観測時の測定対象の粒度や項目が異なるデータを結び付けて分析をすることで、統合的な因果グラフを構築する。これにより大量のデータセットに対する因果推定を可能にする技術。	入力：多変数の数値データ 出力：因果グラフ
	因果知識転用 (限界2を突破する技術)	既に因果ナレッジグラフに格納されている因果関係のなかから、与えられた数値データの前提知識として活用できる因果関係を自動的に判断することにより、より妥当性の高い因果グラフを構築する。	入力：多変数の数値データ、前提知識として用いる可能性のある因果関係 出力：因果グラフ
	因果知識化 (限界1を突破する技術)	文献から抽出される因果関係と数値データから推定される統計的因果関係から、それぞれに出現する関連事象をエンティティ化し、関係付けて定義することで、統合された因果ナレッジグラフを構築する技術。構築された因果ナレッジグラフは、新たに入力される文書データ、数値データを補完して根本原因分析、因果意思決定を強化するほか、ナレッジグラフ上で新たな因果関係を推論することにも使われる。	入力：文献から抽出される因果関係、および数値データから推定される因果関係 出力：因果ナレッジグラフ
因果ナレッジグラフの活用	根本原因分析	因果ナレッジグラフを用いて施策の根拠を論理的に説明する。RAGでは検索に失敗してしまうような複合的な事象についても、高精度に回答を生成する事が可能。	入力：因果ナレッジグラフ 出力：根本となる原因（事象エンティティ）
	因果意思決定	データから発見された可能性のある全ての因果関係に基づいて、目標を達成するために最も効果の高い施策（事象）を推薦する技術。本技術によって推薦される施策は、目標を達成できる施策の中で最小のコストで、なおかつ、目標以外の項目に対する影響は抑制されるように最適化される。	入力：因果ナレッジグラフ、変更したい属性とその目標値 出力：最も効果的な施策（事象エンティティ）
	因果ナレッジグラフ推論 (LLMを用いた質問応答)	LLMを活用して、ユーザの質問に対して因果ナレッジグラフに含まれる因果関係に沿った回答を与える技術。因果ナレッジグラフをサブグラフ群に分割してRAGの検索対象とすることで生成される回答の精度を向上させている。	入力：因果ナレッジグラフ、質問文 出力：質問に沿った回答文

## 4.2. 因果ナレッジグラフ構築技術

### 文献からの因果抽出

文書からトリプル（2つの単語/事象/情報と、その関係性）を抽出してナレッジグラフを作成しておくことで、文書に関連するクエリに対して、短時間で高精度な回答を生成する事が出来る。例えば、複数文書を参照する必要があるような複雑なクエリの場合、LLMにおける通常の検索拡張生成（RAG：Retrieval Augmented Generation）では複数の文書を正しく参照する事が困難なため回答精度が低下するが、複数文書の内容を反映したナレッジグラフを用いる事で回答精度の低下を防ぐ事ができる。富士通では、これを「ナレッジグラフ拡張RAG」と呼んでいる。

ナレッジグラフは有用である一方、人手で構築するには専門家が膨大な工数をかける必要があるという課題があり、自動生成技術が広く検討されてきた。汎用的なナレッジグラフを作成するアプローチとしては、LLMにプロンプトエンジニアリング技術（in-context learningなど）を適用する事で、文書内の名詞を網羅的に抽出してそれらの関係性を抽出する手法がある。しかし、汎用的なナレッジグラフは一問一答のようなタスクには適用可能であるが、因果とは関係ない大量のトリプルが生成されたり事象の言い換えなどによって因果関係が分離されたりするため、専門家の膨大な抽出・変換作業が依然として必要であった。また、ある特定の目的に沿ったナレッジグラフを生成したい場合は、生成したいナレッジグラフと同様の特徴を持つデータセットを用いて言語モデル（BERTなど）を学習するアプローチが考えられるが、一般的にデータセットの準備作業は膨大な工数が必要である。以上のように、従来手法では専門家の作業が依然として必要であるため、実ビジネスにおいて企業の意思決定を迅速化することは出来ていなかった。

そこで我々は、学習データ不要で、文献から因果関係を自動抽出する因果ナレッジグラフ抽出技術を開発した。本技術はLLMを活用した技術である。LLMは単語抽出など文法構造を理解した処理には優れているが、文書全体から事象間の因果関係を抽出するためには、LLMの着目点を適切に誘導する必要があった。そこで図 9に示すように、因果関係のナレッジグラフの構造の型をナレッジグラフスキーマとして事前に定義しておき、そのナレッジグラフスキーマに沿って入力文書を分析するフレームワークを考案し、因果関係を示すナレッジグラフを出力可能とした。本フレームワークでは、LLMを活用したドキュメント分析を処理フローに従って繰り返し実施し、順々にナレッジグラフを構築していく。これは、因果関係のナレッジグラフスキーマを処理フローに自動変換するフロー変換技術と、具体的なLLM指示プロンプトに自動変換するプロンプト変換技術によって実現している。

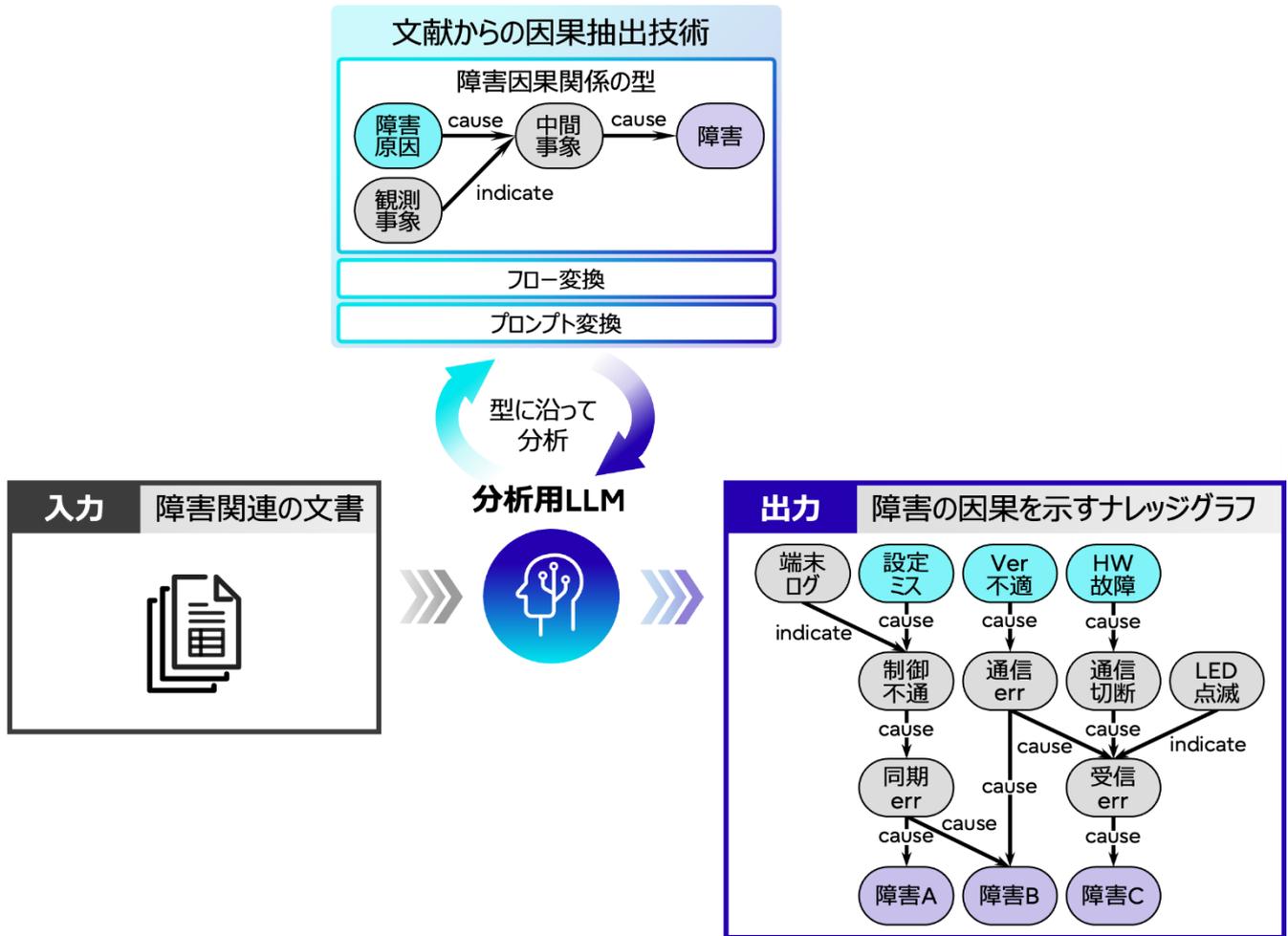


図 9 : 文献からの因果抽出技術の概要

## データからの統計的因果探索

ここでは、富士通が開発した統計的因果分析の実用性を高めるいくつかの技術を紹介する。

データ全体にあてはまる一般的な因果関係ではなく、特定の条件のデータにのみあてはまる因果関係を知りたいニーズは高い。医療領域では、ある特徴を持つ患者特有の薬剤反応を知ることで、知見の加速や既存薬再配置などが期待できる。小売領域では、販売促進効果の人による違いを知ることで、きめこまやかな販売戦略の立案が期待できる。

しかし、そのような因果関係は特定の条件下のデータを集めないと知ることができないが、その条件を事前には知ることができないというジレンマがある。そこで我々は、因果関係のあらゆるバリエーションを網羅するように全ての条件を高速に列挙し、未知の因果関係が現れる条件を効率的に探索する技術を開発した。本技術は大腸がんと健康な大腸の組織における遺伝子発現データに適用し、大腸がんの分類の際に重要とされている遺伝子を自動で特定することに成功した。またさらなる課題として、変数の数が数千・数万になると、列挙や因果探索に膨大な計算量が必要となる。我々は本技術を富士通上で実装することで、従来4,000年かかっていた計算を1日以内に完了させた。本技術をヒトの約20,000遺伝子に適用し、肺がんの薬剤耐性に関する新たな知見を導くことに成功した。

他にも、異なる観点から実用性を向上させる取り組みとして、より複雑な因果関係を捉える因果探索技術の開発も行っている。従来技術では主に原因と結果がシンプルな線形関係を持つという想定で探索が行われていた。しかし現実の複雑な

システムや現象では、線形関係では表現しきれない複雑な因果関係が現れうる。そこでニューラルネットワークを用いた非線形回帰を活用することで、多変数の間のより複雑な因果関係を推定する技術を開発した [6]。

また、統計的因果探索に必要な十分な量のデータが揃わないという課題に対し、富士通はエージェントベースモデリング (ABM) をベースとした統計的因果探索の融合技術を提案した。ABMを用いて個々の行動と相互作用をモデル化することで、社会現象の理解や「To-Be」シナリオの実験が可能となる。本技術は、ABMがシミュレーションで生成する個人レベルのデータとマクロレベルの反事実データから、個人の目的、行動、環境要因、および分析対象の間の因果関係を特定することができ、現実的な施策策定と評価を支援できる。これまでに、空港での乗客体験向上につながる間接的な施策提案とその影響提示 [29]、店舗における顧客体験と売上を最適化する説明可能な店舗製品配置パターンを生成することに成功した [30]。

## 統合因果探索

現状の統計的因果探索では、複数のデータセットを統合して因果関係を推定することは困難であるが、現実には様々な状況で観測された多数のデータセットから総合的に因果関係を判断することが求められる。例えば、各病院では様々な検査項目に対してデータを収集しているが、1つの病院ですべての検査を行うことは非現実的である。通常、血圧や身長、体重といったどの病院でも計測する項目に加えて、その病院でしか実施していない検査項目が存在する。各病院では測定環境などが異なるが、もしこれらのデータを統合して因果探索することができれば、それぞれの病院の検査項目を横断する分析が可能となる。これを実現するための技術として、我々は「統合因果探索」と呼ばれる、複数データセットに跨る因果探索技術を開発している。

因果関係の発見に向けた多くのアプローチは、隠れた交絡因子が存在しないことを前提として、単一のデータセットから固定的な因果関係モデルを学習するよう設計されている。しかしながら、データの取得にコストがかかるため、解析に使用する変数の一部しか測定しない場合がある。そこで異なる病院が測定した健診データや、異なる企業が収集している人事データなど、異なる変数セットを持つ複数のデータセットから因果関係を発見する方法が必要になる。富士通の統合因果探索は、各データセット間で共通する変数を鍵にその背後にあるデータの分布を推定し、異なるドメインのデータセットを統合した因果ナレッジグラフを構築することが可能である。

## 因果知識転用

現状の統計的因果探索技術では、与えられたデータセットに基づく因果関係は導かれるが、その因果関係を他のデータセットの因果探索に役立てることは困難である。それは、異なるデータセットが必ずしも同じ因果構造に従って生成されるとは限らないため、因果関係をどこまで流用できるかを判断することが難しいということが理由の一つである。因果探索において特定の因果関係を前提知識として用いることも行われるが、現状は前提知識としての妥当性は専門家が人手で判断している。しかし、変数の数が数十を超える場合には専門家の判断にも限界が生じる。富士通の因果知識転用技術は、因果構造を推定したいデータセットに最も矛盾の少ない因果関係を、因果ナレッジグラフから前提知識として自動的に取得することにより、与えられたデータセットに閉じた現状の因果探索の限界を突破し、大量のデータセットの学習により成長し続ける因果AIを実現する技術である。

## 4.3. 因果ナレッジグラフ活用技術

### ■ 根本原因分析

因果ナレッジグラフの活用技術として、ネットワーク障害時の根本原因を自動分析する技術を開発している。

図 10は、図 9で抽出した因果ナレッジグラフを根本原因分析で活用した事例である。ナレッジグラフを探索・参照することで適切な因果関係の抽出を行い、多様な事例を網羅的に考慮した障害原因候補を列举できる。これにより、通常のRAGでは検索に失敗してしまうような複合的な事象についても、高精度に回答を生成することが可能となる。さらに、各障害原因の影響や確認手順など、絞り込み及び復旧に必要な知識を提示することも可能である。

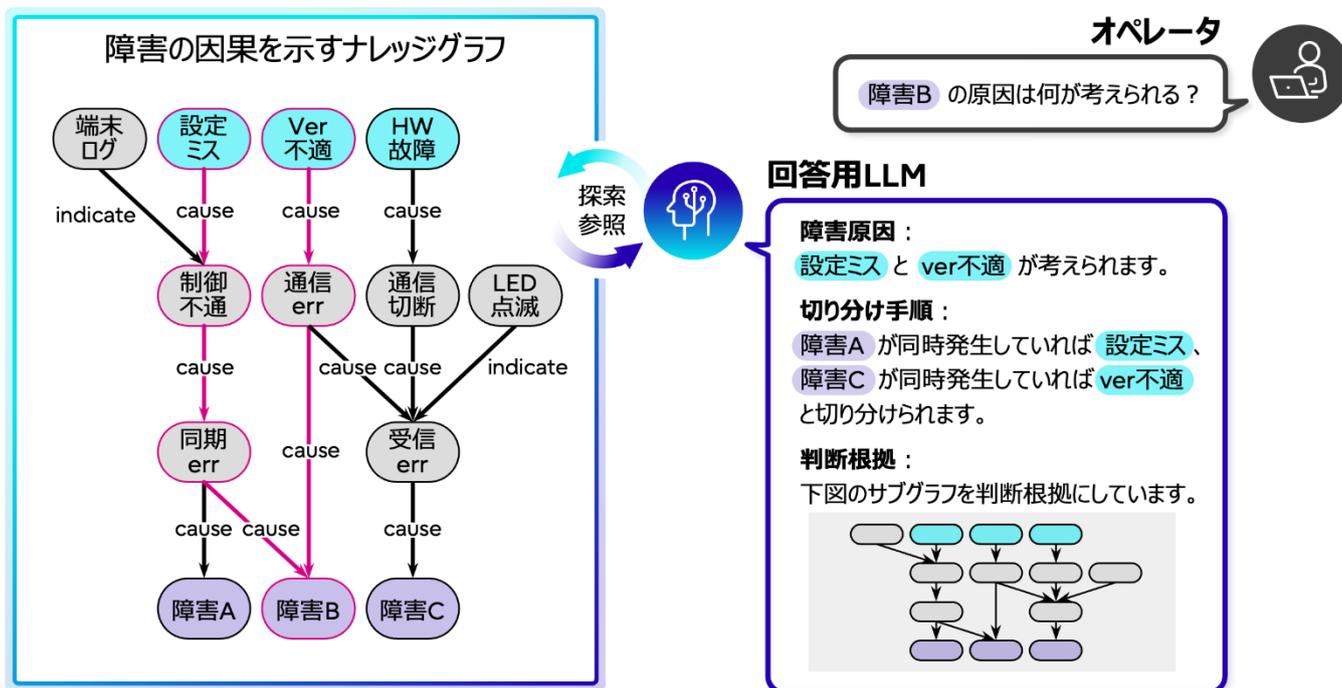
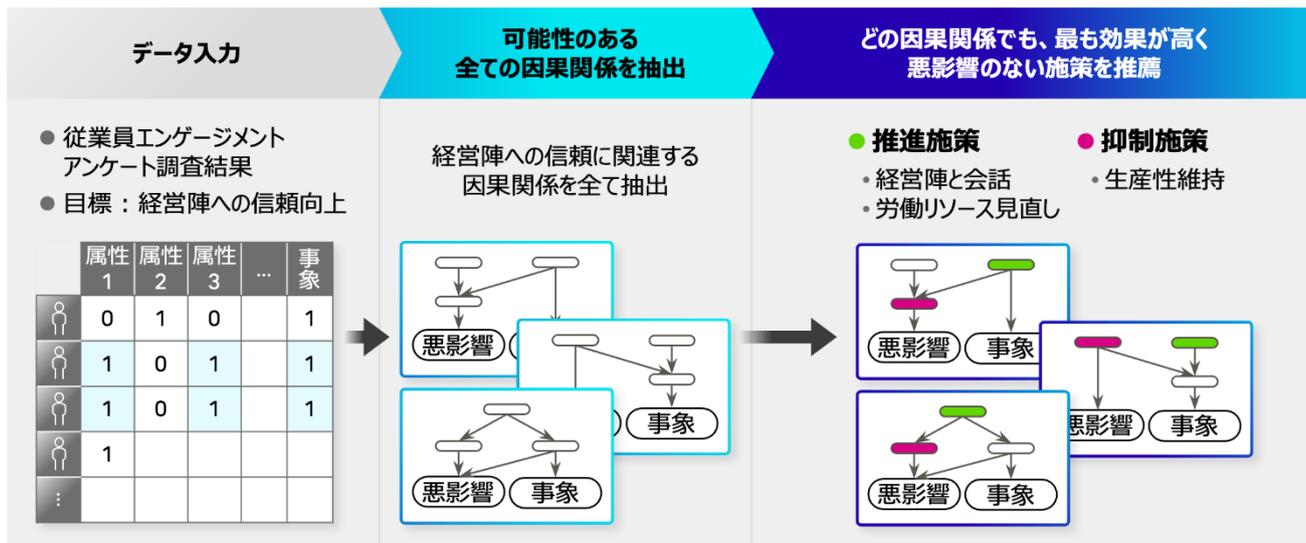


図 10 : 因果ナレッジグラフを用いた根本原因分析の概要

### ■ 因果意思決定

与えられたデータにおける各項目間の因果関係が特定できると、所望の目標を達成するためにどの項目に介入すべきかを決定すること、すなわち施策の策定が可能となる。例えば人事領域では、従業員に対するエンゲージメント調査の結果から項目間の因果関係を特定することで、従業員の生産性を維持したまま経営陣への信頼を向上させるための施策を策定できる。しかし、項目間の因果関係が特定できたとしても、そこから目標を達成するために効果的な施策を人手で導くことは難しい。富士通の因果意思決定は、データから推定された可能性のある全ての因果関係に基づいて、目標を達成するために最も効果の高い施策を推薦する。本技術によって推薦される施策は、目標を達成できる施策の中で最小のコストで、なおかつ、目標以外の項目に対する悪影響を抑制されるように最適化されている。本技術をエンゲージメントサーベイの結果に適用し、従業員の生産性や経営陣への信頼を向上させるための施策について新たな知見を導くことに成功した。



例.従業員エンゲージメント調査結果から経営陣への信頼を向上する施策を導く

図 11：因果意思決定の実施例

## 因果ナレッジグラフ推論（LLMを用いた質問応答）

膨大な量の文献や数値データから構築された因果ナレッジグラフは、科学的発見や重要な意思決定のための推論の基盤になることが期待される。しかし、因果ナレッジグラフの規模が大きくなり項目間の関係性が複雑化すると、その内容を解釈して推論に活用することは困難になっていく。

そこで我々は、LLMを用いた大規模因果ナレッジグラフの推論技術を開発している。本技術は、因果ナレッジグラフをサブグラフ群に分割してLLMのRAGに格納し、質問への回答時に検索参照することで大規模グラフに対応したうえ、推論への利用に適した分割によってLLMの生成する回答の精度を向上させるものである。これは、図 3で示したLLMとKGの統合技術における「a. KG-enhanced LLMs」に該当し、富士通ではこの形式を「ナレッジグラフ拡張RAG」と呼んでいる。

図 12は健康診断の結果とアンケート項目のデータから生成された因果ナレッジグラフに基づいて推論を行った事例である。「肥満を防ぐにはどうすればよいか」という質問に対し、「便通の回数を一定に保つ」といった一般的知識には含まれていない対処方法を因果ナレッジグラフに基づいて回答している。

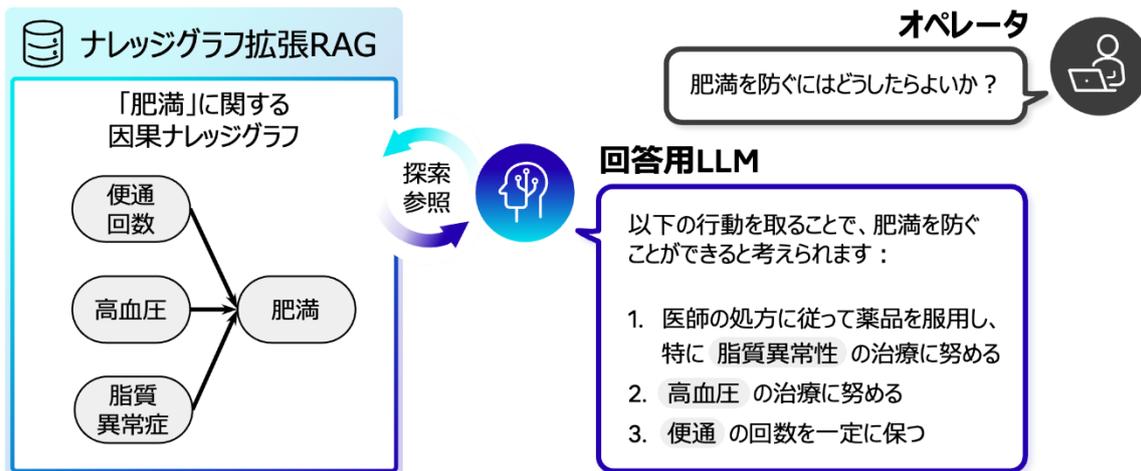


図 12 : 因果ナレッジグラフとLLMを用いた質問応答の例

## 5. 因果ナレッジグラフ事例集

本章では、因果ナレッジグラフを「ICTサポート」「人事」「健康・ヘルスケア」の3分野に適用した事例を紹介する。ICTサポート分野では、複数の通信機器の仕様書やトラブルシューティングガイドからネットワーク障害の因果ナレッジグラフを構築し、根本原因を分析する。人事分野では、商談データと従業員のエンゲージメントサーベイデータを跨いだ統合因果探索を実施し、エンゲージメント項目が商談獲得に与える影響を分析している。健康・ヘルスケア分野では、弘前大学で推進されている健診健康ビッグデータから導いた因果関係を前提知識として使用することで、別の健診データに対しても妥当な因果関係を推定できるようにした事例である。

### 5.1. 事例1：ネットワークの障害原因分析

#### 概要

ネットワーク導入および運用では、複雑なトポロジーで繋がった様々な機器を扱う必要があるため、送受信エラーや遅延などといった障害が発生した際にその原因を突き止める事は困難である。一方で、一度の障害で数十億円の損害が発生するなどその影響は大きく、いかに障害を予防するか、または障害復旧を迅速にするかがネットワーク事業者にとっては重大課題となっている。現状は一部自動化の取組みが行われているものの、設定ミス・バージョン(ver)不適・ハードウェア(HW)劣化など多岐にわたる障害原因の分析・調査は専門家が多大な工数をかけており、分析・調査の自動化および支援技術に対する要望は大きい。

そこで我々は、ネットワークにおける障害の原因分析を実現するため、因果ナレッジグラフを適用して障害の複雑な因果関係の分析を自動化した。今後は、クラウドシステム、工場機器運用など、あらゆる分野の障害原因分析へも適用先を拡張していく予定である。

#### 使用するデータ

トラブルシューティングガイド、仕様書、マニュアル類、などの文献

#### 使用する技術

文献からの因果抽出、根本原因分析

#### 実際の出力の詳細説明

前節で説明した図 10では、ネットワークにおける障害の原因分析に適用した事例における動作概要を示している。図 10の左側が、トラブルシューティングガイドや仕様書から自動抽出した因果ナレッジグラフである。この因果ナレッジグラフの探索・参照によって適切な知識抽出を行う事で、多様な事例を網羅的に考慮した障害原因候補を列挙できる。RAGでは検索に失敗してしまうような複合的な事象についても、高精度に回答を生成する事が可能となる。更に、各障害原因の影響や確認手順など、絞り込み及び復旧に必要な知識を提示する事も可能となる。

本技術は、富士通が提供しているPoC環境である「Kozuchi」に「Fujitsu ナレッジグラフ拡張RAG for Root Cause Analysis」というWebアプリケーションとして公開しており、図 13では、本Webアプリケーションの画面例を示す。図 13上段のように障害内容をクエリとして入力すると、図 13下段のように判断根拠となるナレッジグラフとともに障害原因分析結果を得る事が出来る。

**クエリ入力画面**

クエリ

Test fileをダウンロードしていたら、断続的に通信が切断する現象が起きました。

**回答サマリ**

KG

We identified 2 root cause paths related to incidents similar to yours, where the primary issue is "Disconnected intermittently." The root causes include a switch configured to send PTP on a 'non-forwardable' MAC and a specific software version on the FCC radio. Each root cause has specific conditions and log indicators that can help in identifying the issue.

Path ID	Incident registered in DB	Root Cause(s)	Conditions	Intermediate event	References
1	Disconnected intermittently	Switch configured to send PTP on 'non-forwardable' MAC	RU log shows "S-plane packet not arrived"	PTP not getting locked, S-plane packet not accepted by RU	MOP 3 in TSG
2	Disconnected intermittently	Software version v5304 on FCC radio	No RU found on the mobile devices, No packet received on radio, Fronthaul interface LED blinking	Continuous transmission of packets from DU to radio over an extended idle period	MOP 8 in TSG

**障害原因**

「MACアドレス設定ミス(原因1)」と「ソフトウェアバージョン設定ミス(原因2)」の可能性

**切り分け手順**

「S-planeパケットのエラー」が観測されていれば原因1、「LED点滅など」が観測されていれば原因2

**判断根拠となるナレッジグラフ**

**回答閲覧画面**

図 13 : 因果ナレッジグラフを用いた根本原因分析のWebアプリケーション画面例 (上段 : クエリ入力画面、下段 : 回答閲覧画面)

これによって、専門家が大量の文献を読んで分析・調査をする代わりに即座に調査結果を得られるようになるうえ、より網羅的で手戻りのない復旧手順を提示できるようになる事から、障害復旧までの時間を大幅に短縮する事ができる。

## 5.2. 事例2：商談データとエンゲージメントサーベイを統合した因果分析

### 概要

「従業員のエンゲージメント向上が商談獲得にどのように影響しているか」や、「商談の勝敗がエンゲージメントサーベイにどのように影響しているか」は、人事部門の関心事である。また、営業現場では、実際に商談の成功率を獲得するための施策が求められている。

しかしながら、単純な相関分析などを用いて商談成功率向上のための施策を立案した場合、成功率の原因ではなく、単なる疑似相関である要因を見つけてしまうリスクがある。そのため、リスクを極力少なくするために要因間の原因と結果の関係を明らかにし、各要因がどれだけの影響を与えているか確認可能な因果分析技術の適用が重要である。

異なる領域である商談データとエンゲージメントサーベイデータを横断し、両者を統合した因果グラフを構築することは従来の手法では難しい。そこで、富士通の統合因果探索技術を使い、両者のデータを横断した因果グラフを出力することにした。これは両者に共通の項目をヒントにし、両データセットが共通の因果構造から生成されていると仮定したうえで因果グラフを推定する技術である。これにより、商談とエンゲージメントサーベイの間にどのような関係があるか判明し、さらにこの因果グラフに対して因果意思決定を適用し、商談成功率向上の施策も立案することができた。

### 使用するデータ

商談の詳細と勝敗を記録したデータ、従業員のエンゲージメントサーベイデータ

### 使用する技術

統計的因果探索, 統合因果探索, 因果意思決定

### 実際の出力の詳細説明

図 14は商談データとエンゲージメントサーベイデータの横断分析結果の例である（モデルケースであり、実態を表すものではない）。赤のノードは商談データでの項目、青のノードはエンゲージメントサーベイデータの項目である。この2つの項目間における因果関係を観察することで両者にどのような関係があるかがわかる。

また、図 15は商談成功率を10%上げるための施策を因果意思決定で出力した例である。これは直接の原因であるソリューション認知率・担当面識率を制約することで、エンゲージメントサーベイの項目を動かすアクションを出力している。

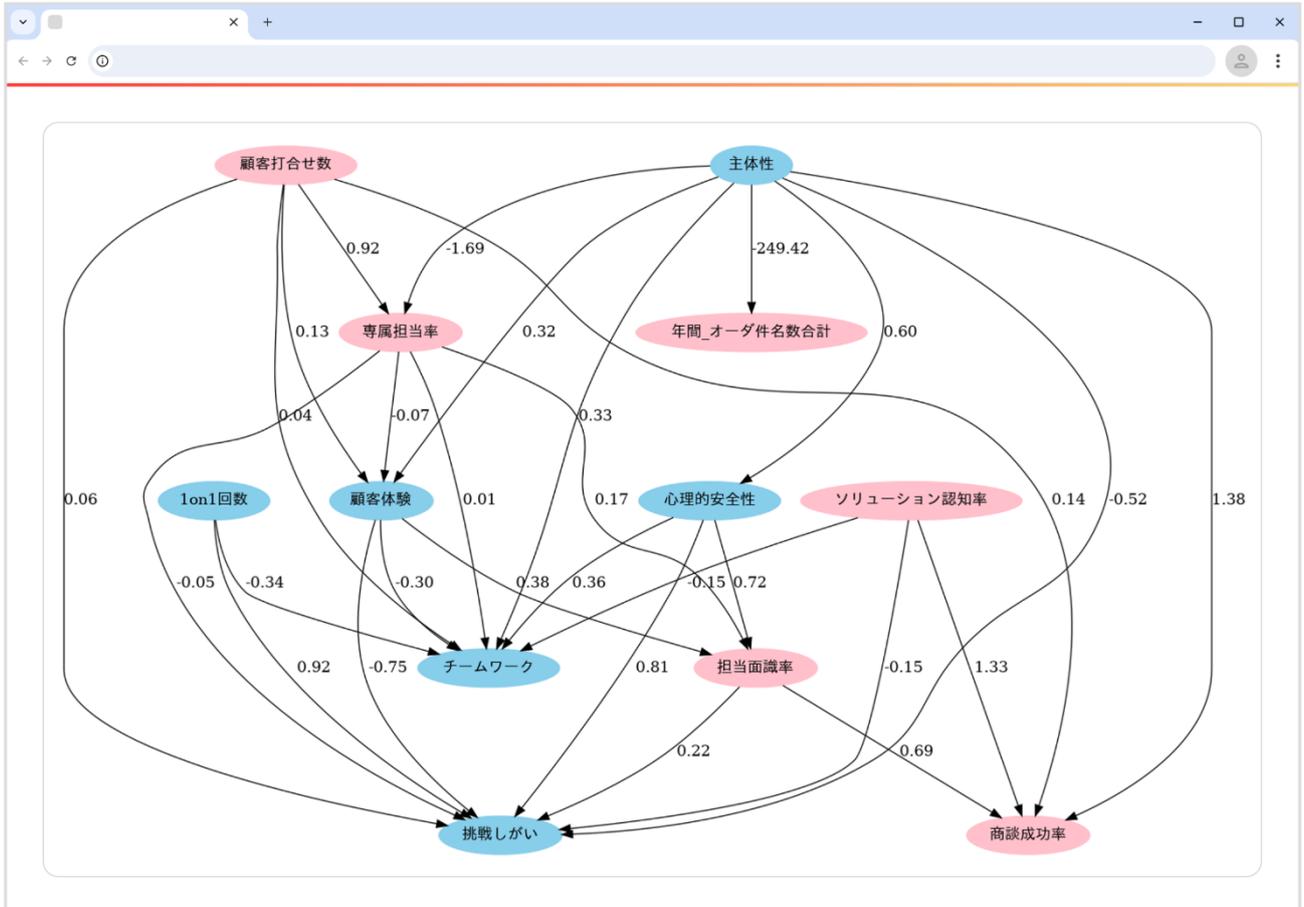


図 14 : エンゲージメントと商談データの統合因果グラフ

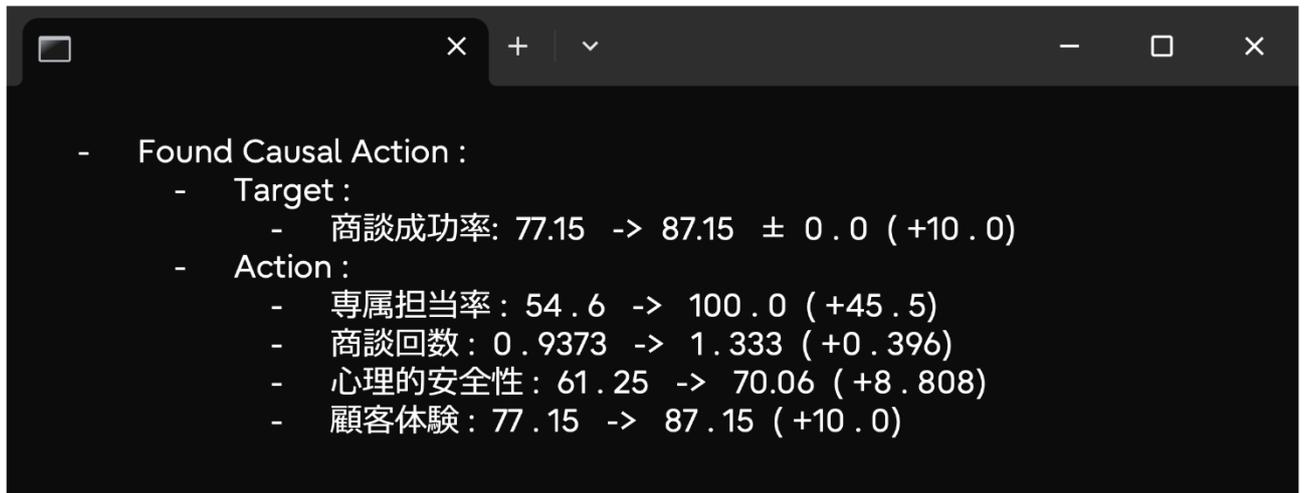


図 15 : 因果意思決定支援による出力

## 5.3. 事例3：睡眠とライフスタイルに関するデータの因果関係の推定

### I 概要

良質な睡眠は、高血圧や心疾患リスクの低減、精神的ストレスの緩和、ホルモンバランスの調整、過食の抑制など、健康的な生活を支える重要な要素であることが広く知られている。では、睡眠の質を向上させるには具体的に何をすべきか。この問いに答えるためには、単なる相関関係を超え、睡眠の質に影響を及ぼす要因との因果関係に着目したデータ分析が不可欠である。どのようなライフスタイルが睡眠の質を左右する“原因”となっているのかを明確にすることで、より効果的な改善策を見出すことができるだろう。

しかしながら、単一のデータセットに基づく因果関係の特定には、サンプル数の不足、交絡因子（データセットに含まれていない未知の因子）の影響、データ収集過程におけるバイアスなど、さまざまな問題が伴うことが多い。これらの問題は、分析の結果の信頼度を低下させる要因となってしまっている。富士通の因果知識転用技術では高い信頼性を有する既知の因果ネットワークを利用し、それを未知のデータセット上の因果関係に関する情報に変換することで、未知のデータセットに対する因果探索の精度を向上させることを可能にしている。

今回、弘前大学が主導する大規模健診データ「弘前大学COI-NEXT 岩木健康増進プロジェクト健診健康ビッグデータ」[31]に対して、京都大学の研究グループが独自のベイジアンネットワーク技術を適用し、信頼性の高い「弘前（岩木）健診因果ネットワーク」を構築した。この因果ネットワークを富士通の因果知識転用を使って睡眠とライフスタイルに関するデータセットの因果探索に活用したところ、弘前健診因果ネットワークを活用しなかった場合と比較して、より妥当性の高い因果関係を導き出すことができた。

### 使用するデータ

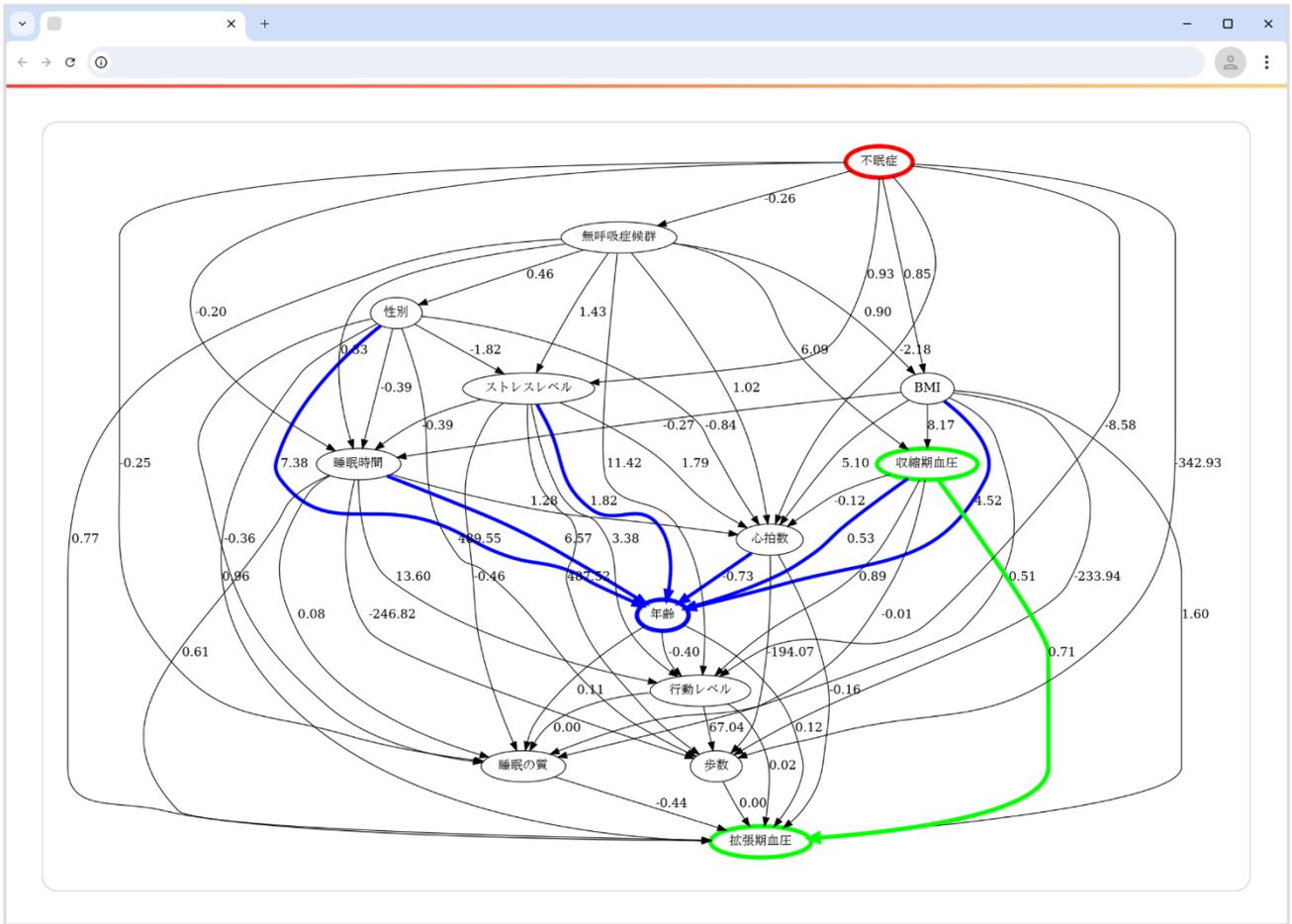
弘前健診因果ネットワーク、Sleep Health and Lifestyle Dataset [32]

### 使用する技術

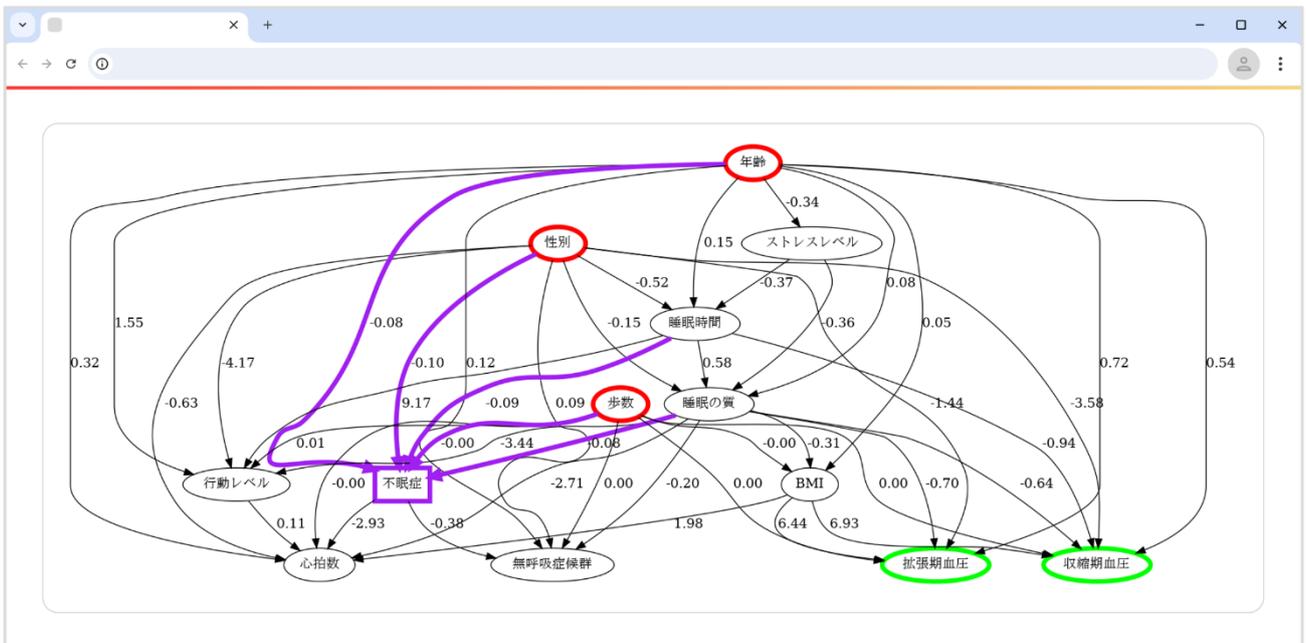
統計的因果探索技術、因果知識転用技術

### 実際の出力の詳細説明

今回は、睡眠とライフスタイルに関するオープンデータである「Sleep Health and Lifestyle Dataset」（以下SHデータと呼ぶ）における因果関係を推定する際に、因果知識転用を用いた場合と用いない場合の2パターンを実行する。図 16に実行結果を示す。図 16の（1）はSHデータのみを用いた因果関係の推定結果であり、図 16の（2）が因果知識転用により「弘前健診因果ネットワーク」を活用したSHデータの因果関係の推定結果である。それぞれの因果グラフを比較すると、以下を観察することができる。



(1) 弘前健診因果ネットワークを活用しない因果グラフ



(2) 弘前健診因果ネットワークを活用した因果グラフ

図 16 : 弘前因果ネットワーク活用の有無による、睡眠とライフスタイルに関する因果関係の推定の違い

- 赤の太線で囲まれたノードはそれぞれの因果グラフにおいて親ノード（該当ノードの直接の原因になっているノード）を持たないノードである。弘前健診因果ネットワークを活用した因果グラフ（図 16（2））では“年齢”、“性別”、“歩数”の三つが親ノードを持たない外生因子と推定されており、これらの因子が他のノードから影響を受けないという直観と相違していない。その一方弘前健診因果ネットワークを活用しない因果グラフ（図 16（1））では“不眠症”のみが外生因子として推定されており、活用した因果グラフにおいては外生因子であった“年齢”や“性別”が親ノードを持ってしまっている。
- 活用なしの因果グラフ（図 16（1））において、“年齢”が直接の原因となっている因果関係を青色で記載している。例えば“睡眠時間”が“年齢”の原因になっているという明らかに妥当ではない因果関係が推定されてしまっている。
- 活用した因果グラフ（図 16（2））において“不眠症”が直接の原因となっている因果関係を紫色で記載している。“睡眠時間”や“睡眠の質”が“不眠症”の直接影響しているという妥当な結果を推定することができている。
- 緑色で囲まれた“収縮期血圧”と“拡張期血圧”は様々な健康状態の結果として検出される数値であり、これらが互いに因果関係にあることは常識に反する。活用なし因果グラフ（図 16（1））では“収縮期血圧”から“拡張期血圧”に因果が検出されてしまっているが、活用した因果グラフ（図 16（2））ではこれらは直接の因果関係に無い。

このように、因果知識転用を用いて信頼性の高い既知の「弘前健診因果ネットワーク」の因果関係に関する情報を活用することで、SHデータのようなそのままの因果探索では信頼の低い結果しか得られないようなデータセットに対しても、より妥当性の高い因果関係の推定が可能になっている。

## 6. おわりに

本ホワイトペーパーでは、LLMが抱える課題と、その解決策として期待される因果ナレッジグラフの可能性について論じた。LLMは自然言語処理において革新的な技術であるが、真に信頼できる意思決定支援には、データの相関関係だけでなく因果関係に基づいた論理的推論が不可欠である。因果ナレッジグラフは、この課題を克服し、AIを単なる情報提供ツールから、より高度な意思決定支援ツールへと進化させる可能性を秘めている。

2章、3章では、因果ナレッジグラフを構成する「因果分析」と「ナレッジグラフ」の技術的背景、そして富士通が考える因果ナレッジグラフの定義、ユースケース、必要性について詳述した。4章では、富士通が取り組む因果ナレッジグラフ構築と活用のための技術群を紹介し、特に文献からの因果抽出、統計的因果探索、統合因果探索といった基盤技術について解説した。5章では、「ICTサポート」「人事」「健康・ヘルスケア」の3分野で因果ナレッジグラフを構築・活用した事例を紹介した。

富士通では、これらの技術を通じて、様々な分野における因果ナレッジグラフの構築と応用を進めていく。これは、ビジネスにおける戦略的意思決定、医療における治療計画、スポーツにおける育成計画など、多岐にわたる分野において、より精密で信頼性の高い意思決定を支援することに繋がる。富士通はクロスインダストリーな因果関係を結合した因果ナレッジグラフによって、複雑な社会課題をデータに基づく論理的な意思決定によって解決し、より良い未来社会の実現に貢献していく。

---

### 執筆責任者：

コグニティブグラフ コアプロジェクト

横尾郁 シニアプロジェクトディレクター

### 執筆者：

小川雅俊 リサーチディレクター

丸橋弘治 リサーチディレクター

太田唯子 シニアリサーチマネージャー

高木拓也 シニアリサーチマネージャー

上村健人、鶴飼孝典、大川佳寛、岡嶋成司、尾関諒介、金森憲太郎、菊月達也、鈴木浩史、

Chang Shuang、二宮照尚、築島瞬、和田章宏

(役職&あいうえお順)

### 監修：

井形伸之 シニアリサーチディレクター

## A. 参考文献

- [1] J. Pearl, Causality, Cambridge university press, 2009.
- [2] D. B. Rubin, "Causal inference using potential outcomes: Design, modeling, decisions.," *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 322-331, 2005.
- [3] P. Spirtes, C. Glymour and R. Scheines, Causation, prediction, and search, MIT press, 2001.
- [4] D. M. Chickering, "Optimal structure identification with greedy search," *Journal of machine learning research*, vol. 3, no. Nov, pp. 507-554, 2002.
- [5] S. Shimizu, "DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model," *Journal of Machine Learning Research-JMLR*, vol. 2, no. Apr, pp. 1225-1248, 2011.
- [6] K. Uemura, "A multivariate causal discovery based on post-nonlinear model," Conference on Causal Learning and Reasoning, 2022.
- [7] M. Takayama, T. Okuda, T. Pham, T. Ikenoue, S. Fukuma, S. Shimizu and A. Sannai, "Integrating Large Language Models in Causal Discovery: A Statistical Causal Approach.," CoRR abs/2402.01454, 2024.
- [8] Z. Jin, J. Liu, Z. Lyu, S. Poff, M. Sachan, R. Mihalcea, M. T. Diab and B. Schölkopf, "Can Large Language Models Infer Causation from Correlation?," CoRR abs/2306.05836, 2023.
- [9] W3C, "RDF (Resource Description Framework)," [Online]. Available: <https://www.w3.org/RDF/>.
- [10] dbpedia, 2024. [Online]. Available: <https://wiki.dbpedia.org/>.
- [11] yago, 2024. [Online]. Available: <https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/>.
- [12] googlekg, 2024. [Online]. Available: <https://developers.google.com/knowledge-graph/>.
- [13] drugbank, 2024. [Online]. Available: <https://www.drugbank.ca/>.
- [14] robokop, 2024. [Online]. Available: <https://robokop.renci.org/>.
- [15] bio2rdf, 2024. [Online]. Available: <http://bio2rdf.org/>.
- [16] fibo, 2024. [Online]. Available: <https://www.omg.org/spec/FIBO/>.
- [17] govuk, 2024. [Online]. Available: <https://www.gov.uk/>.
- [18] Openlink, "Virtuoso," [Online]. Available: <https://virtuoso.openlinksw.com/>.
- [19] OntoText, "GraphDB," [Online]. Available: <https://graphdb.ontotext.com/>.
- [20] Neo4j, "Neo4j Database," [Online]. Available: <https://neo4j.com/>.
- [21] Amazon, "Amazon Neptune," [Online]. Available: <https://aws.amazon.com/neptune/>.
- [22] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang and X. Wu, "Unifying Large Language Models and Knowledge Graphs: A Roadmap," 2024. [Online]. Available: <https://arxiv.org/abs/2306.08302>.
- [23] Glean Technologies, Inc., "How Glean's knowledge graph powers great enterprise search and generative AI," [Online]. Available: [https://cdn.prod.website-files.com/6127a84dfe068e153ef20572/65f9da1c15f95fbf6866c112\\_Glean%20Knowledge%20Graph.pdf](https://cdn.prod.website-files.com/6127a84dfe068e153ef20572/65f9da1c15f95fbf6866c112_Glean%20Knowledge%20Graph.pdf).
- [24] T. Bratanič, "Knowledge Graphs & LLMs: Multi-Hop Question Answering," 2024. [Online]. Available: <https://neo4j.com/developer-blog/knowledge-graphs-llms-multi-hop-question-answering/>.

- [25] C. Bharadwaj and S. Devi, "Building a Clinical Intelligence Engine using MedLM augmented Clinical Knowledge Graphs," 2024. [Online]. Available: <https://cloud.google.com/blog/topics/healthcare-life-sciences/building-a-clinical-intelligence-engine-using-medlm?hl=en>.
- [26] Neo4j, Inc., "Neo4j Partners with Google Cloud to Launch New GraphRAG Capabilities for GenAI Applications," 2024. [Online]. Available: <https://neo4j.com/press-releases/neo4j-google-cloud-graphrag-for-genai/>.
- [27] N. Maryam, "Synergizing Knowledge Graphs with Large Language Models (LLMs): A Path to Semantically Enhanced Intelligence," 2024. [Online]. Available: <https://enterprise-knowledge.com/synergizing-knowledge-graphs-with-large-language-models-llms/>.
- [28] F. H. Messerl, "Chocolate Consumption, Cognitive Function, and Nobel Laureates," the New England Journal of Medicine, 367, 1562-1564., 2012.
- [29] S. Chang, T. Kato, Y. Koyanagi, K. Uemura and K. Maruhashi, "An Iterative Analysis Method Using Causal Discovery Algorithms to Enhance ABM as a Policy Tool," 2023 Winter Simulation Conference (WSC), pp. 138-149, 2023.
- [30] S. Chang, S. Yamane and K. Maruhashi, "Causal-based Rack Layout Optimization in Retail: Incorporating ABM and Causal Discovery," Winter Simulation Conference 2024, Florida, U.S., Dec. 2024, 2024.
- [31] 弘前大学 CO-NEXT, "岩木健康増進プロジェクト健診健康ビッグデータ," 2024. [Online]. Available: <https://coi.hirosaki-u.ac.jp/>.
- [32] Kaggle, "Sleep Health and Lifestyle Dataset," 2024. [Online]. Available: <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>.

## B. これまでの富士通の実績

### I 因果分析に関する富士通の研究実績

- 2017年3月：[「理研AIP-富士通連携センター」、「理研AIP-東芝連携センター」、および「理研AIP-NEC連携センター」を開設](#)：富士通PR
- 2020年12月：[ヒトやモノなどのデータの一つひとつが持つ特徴的な因果関係を発見する技術を開発](#)：富士通PR
- 2021年2月：[望む結果までの手順を導くことができる「説明可能なAI」を世界で初めて開発](#)：富士通PR
- 2021年2月：[反実仮想説明に基づくアクションの最適順序を推薦する新たなAI技術を開発](#)：富士通テックブログ
- 2022年10月：[「富士通×東北大学 発見知能共創研究所」の設置について](#)：富士通PR
- 2022年3月：[スーパーコンピュータ「富岳」と「発見するAI」で、がんの薬剤耐性に関わる未知の因果メカニズムを高速に発見する新技術を開発](#)：富士通PR
- 2022年4月：[富士通とAtmonia、カーボンフリーな物質であるアンモニアの触媒探索をHPCとAIで加速する共同研究を開始し、ゼロエミッションに貢献](#)：富士通PR
- 2022年4月：[多変数データを用いた非線形因果探索技術の開発](#)：理研PR
- 2023年2月：[HPCとAIを活用した量子化学シミュレーション高速化技術を開発し、クリーンなアンモニア合成に向けた触媒候補探索期間の大幅な削減に成功](#)：富士通PR
- 2023年5月：[京都大学、Chordia Therapeutics、富士通、因果関係を発見するAI技術を用いてがんの新薬に対するバイオマーカーを発見する実証実験を開始](#)：富士通PR
- 2023年10月：[クリーンなアンモニア合成に向けて、量子化学シミュレーションを活用し、高価な貴金属以外の触媒材料候補を発見](#)：富士通PR

### I ナレッジグラフに関する富士通の研究実績

- 2013年4月：[オープンデータの活用革新! リンクが張られた公開データ（LOD: Linked Open Data）向け大規模データ格納・検索技術を開発 LOD: Linked Open Data）向け大規模データ格納・検索技術を開発](#)：富士通PR
- 2014年1月：[世界中で公開されているオープンデータへのリンクを自動的に付与する技術を開発](#)：富士通PR
- 2014年2月：[業界初！自治体の特性を見える化する評価ツールを開発](#)：富士通PR
- 2014年12月：[Linked Open Data技術を適用した地域の特性を発見するツールを公開](#)：富士通PR
- 2015年1月：[川崎市で子育て支援アプリの実証実験を開始](#)：富士通PR
- 2017年3月：[富士通とInsight、システム生物学研究機関のSystems Biology Irelandと共同研究を開始](#)：富士通PR

- 2017 年 6 月 : [オープンデータを活用し企業の営業活動を支援する企業情報分析Webアプリの無償公開を開始](#) : 富士通PR
- 2017 年 9 月 : [AIの推定理由や根拠を説明する技術を開発](#) : 富士通PR
- 2017 年 9 月 : [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](#) : 雑誌Fujitsu
- 2018 年 7 月 : [Deep Tensorとナレッジグラフを融合した説明可能なAI](#) : 雑誌Fujitsu
- 2018 年 9 月 : [人に信頼され社会を発展させる人工知能](#) : 雑誌Fujitsu
- 2018 年 10 月 : [信用リスク判定など金融業務のデジタル化を加速する「説明可能なAI」](#) : Fujitsu Journal
- 2018 年 10 月 : [富士通とInsight、システム生物学研究機関のSystems Biology Irelandと共同研究を開始](#) : 富士通PR
- 2019 年 9 月 : [ナレッジグラフで実現する「トラストで説明可能なAI」と社会実装](#) : 雑誌Fujitsu
- 2019 年 11 月 : [新しいAIによるがんゲノム医療の効率化を東大医科研との共同研究で実現](#) : 富士通PR
- 2020 年 1 月 : [AIを使った「がんゲノム医療」で、がん治療を加速する](#) : Fujitsu Journal
- 2021 年 10 月 : [愛知県がんセンターと富士通、がん患者ごとの遺伝子変異に基づき治療薬の選択をAIにより支援するシステムを開発](#) : 富士通PR
- 2024 年 5 月 : [がんのタイプ分けなどのゲノム医療分野の課題を世界最高精度で解く、説明可能なAI技術を開発](#) : 富士通PR
- 2024 年 5 月 : [ナレッジグラフと大規模言語モデルを融合する研究開発を強化することで、生成AIの信頼性を高め業務活用を加速](#) : 富士通 News

