

AI ビッグデータ応用における仮説解釈説明の構造探求と関連情報通信技術の体系化

代表研究者	石川 博	東京都立大学 システムデザイン研究科 特別先導教授
共同研究者	遠藤 雅樹	職業能力開発総合大学校基盤ものづくり系情報通信准教授

1 研究調査の概要

仮説（モデル）の解釈説明の詳細構造を確立し、それに基づいて説明生成のためのレファレンスアーキテクチャを提案する。また仮説の解釈説明に適した仮説の記述手段を探求し、提案する。さらに大規模言語モデルに基づく生成AIと適切な回答を得るためのprompt engineeringを用いて、自然言語により説明を提示する方法を探求する。以上の提案を検証するために科学や行政など多様なユースケースを収集し分析する。

2 研究調査の詳細

2-1 研究調査の目的

現在活発に推進されているAI（人工知能）や機械学習の実用化、ビッグデータ応用システムの受容において仮説（モデル）の解釈説明が必要とされている。

【リサーチクエスチョン】

将来のSociety 5.0（参考1）の実現や社会的課題の解決の上でカギとなる、AIやビッグデータが広く受け入れられ利用されるためには、その応用サービスの利用者（分野専門家や一般ユーザ）に対して、仮説生成の説明とモデルの特徴と個別の決定に対する理由の説明及び仮説生成のためのデータ生成を含む応用サービスの全プロセスの説明の両方を行う必要がある。そのために初年度の成果を踏まえて、説明はどのような詳細構造を持ち、その構造に基づいてどのように説明を生成すべきか、また説明の記述の手段として何が適切か、どのように説明を提示すれば利用者が納得できるかというリサーチクエスチョン（学術的問い）を解決する必要がある。

【研究目標】

そのために以下のようなサブゴール（課題）に分けて基盤的研究を行う。

A. 説明の構造化アプローチの確立：

応用サービスがデータ管理とデータ分析からなるハイブリッドなエコシステムである点と説明の構成要素に関する具体性の程度を考慮し、理論的枠組の出発点として説明の詳細構造（構造化アプローチ）を明確化する。

B. 説明生成のためのアーキテクチャの提案：

理論的側面の検討を促進するため、構造化アプローチに基づき、説明を生成し提示するためのレファレンスアーキテクチャを提案する。

C. 多様な説明の記述手段・開発環境の探求と生成AIを用いた自然言語による説明提示：

多様な記述手段・開発環境に基づき記述された説明を、大規模言語モデルに基づく生成AIおよびprompt engineeringを用いて自然言語で提示する手法を検討する。

D. ユースケースの検討深化：

行政ユースケースとして健康福祉の推進（参考2）などSDGs関連のテーマにも挑戦し、科学ユースケースとして宇宙探査機の異常検知における仮説生成と解釈の具体化も検討する。

(参考)

1. 内閣府、Society5.0、www8.cao.go.jp/cstp/society5_0/index.html
2. 国連広報センター、「SDGs 目標 3：あらゆる年齢のすべての人々の健康的な生活を確保し、福祉を推進する」 https://www.unic.or.jp/files/Goal_03.pdf

2-2 研究調査の動機、情報通信との関係、これまでの経緯

【研究調査の動機】

そもそもなぜ仮説の解釈が必要なのであるだろうか。データ分析応用（AI・機械学習やビッグデータ応用システム）の当事者は、分析者、分野専門家、およびエンドユーザである。一言でいえば、データ分析応用の当事者に仮説を受容してもらうためである。

つまり分析者や専門家は、仮説が中心的な役割を果たす分析応用全体が純粋に技術的に信頼に足るものかどうか判断する必要がある。さらに、サービスの提供者はそのような応用をサービスの享受者に理解させて、利用してもらうためにその説明責任を果たす必要がある。言い換えればサービスの享受者は、サービスの個別の決定を含む応用全般に関して説明を受ける権利（right to explanation）を有する【参考文献4】（項目11参照）。そのためには基本として仮説を解釈することが必須となってくる。

例えば科学分野の応用では、分析者と専門家はデータエンジニアと科学者からなるグループである。また社会基盤的応用では、分析者はデータエンジニアであり、専門家は意思決定者であり、エンドユーザはサービスを受ける一般人である場合が多い。

【情報通信との関係】

実用化に向けて研究開発が進む情報通信として説明可能なAIや解釈可能な機械学習の研究（【参考文献5、6、7、8】など）は活発である。しかしその大半は個別的な技術の開発が中心であり、解釈説明の要素とその関係からなる解釈説明の全体的構造が明らかではないために、その中で位置付けや体系化が十分になされていない。さらに現状の仮説生成の記述にはプログラミング言語が用いられ、記述のレベルが低いために、このギャップを解決するための仮説の解釈説明に適する手段の探求と提案が必要である。さらに科学研究手法を研究の対象とする科学哲学においても解釈説明の研究【参考文献9】は行われてきたが、科学哲学の成果と現状の情報通信技術は十分には関連付けられていない。

【これまでの経緯】

東京都福祉保健局と協働でCOVID-19の新規陽性者数予測モデルの構築【参考文献15】やJAXAと協働した、月周回衛星かぐやデータを用いた月探査の研究【参考文献14】によって事業担当者や科学者から仮説・モデルの構築方法や特徴、判断理由の説明をしばしば求められた。こうした経験から解釈説明が仮説を受け入れるためには、必須であり、特に医療や科学の分野では仮説の精度より重視されることを痛感した。

2-3 研究調査に関する内外の研究の動向

説明可能な人工知能（eXplainable AI、XAI）や解釈可能な機械学習（Interpretable Machine Learning、IML）の分野でミクロな説明機能に相当する研究が活発化し始めた【参考文献5】。XAIやIMLに関する個別研究は(1)特徴量に基づく説明、および(2)解釈可能なモデル、(3)説明モデルの導出に分類される。(1)には説明ベクトルの導入や弱分類器の利用により重要な属性を発見する研究【参考文献6】や【参考文献13】がある。(2)には文字の生成過程を考慮することで解釈できるモデルを作る研究【参考文献7】がある。(3)には任意の分類器から、結果に局所的な近似モデルを用いてその説明を導出する研究【参考文献8】や可視化する研究【参考文献12】がある。すでに解釈説明の概念化の検討【参考文献14】には着手したが、本研究では、現在はまだ十分に研究されていない解釈説明構造を明らかにする。次に本研究はミクロな説明技術として、生成モデルの説明と個別決定の理由の説明にそれぞれ上記(1)と(3)カテゴリのアプローチを発展させつつ、従来十分に研究されていないマクロな説明技術の研究を推進し、両者を統合発展させた画期的な解釈説明基盤の構築を目指す。その際に構造に欠落する要素技術を探求する。さらに生成モデルに依存しない高水準の言語としてデータベース操作言語SQL【参考文献10】も検討対象のひとつになる。

3 研究調査の成果と今後の計画

3-1 研究調査の成果

初年度の研究成果を踏まえ、解釈説明基盤形成を目指して、研究開発を推進した。

A. 説明の構造化アプローチの確立

説明の対象は、以下のような基本的な要素から構成され则认为、説明の詳細構造（構造化アプローチ）を確立した。

- ・ データを用いて仮説を作る方法
- ・ 個別の入力データには依存しないような仮説（モデル）の全体的な特徴
- ・ 個別の出力結果を生み出すに至った入力事実（個別データ）と仮説のうち適用（実体化、instantiation）された部分
- ・ 仮説生成に使用するデータ生成を含む手順全体の記述
- ・ データと仮説生成の実行方法

昨年度に説明の対象として提案した、以下のような構成要素からなる解釈説明の構造を深化させた。

- ・ データの作り方 (how_d) : How to generate Data
- ・ 仮説の作り方 (how_h) : How to generate Hypothesis
- ・ 仮説の特徴 (what_h) : What features of Hypothesis
- ・ 仮説結果の理由 (why_h) : Why (What Reasons) for Hypothesis decision
- ・ 生成方法の実行計画 (how_e) : How to Execute data/hypothesis

特に**機械論的説明**（項目 D 参照）の観点から、構成要素のうち**データと仮説の作り方および実行計画**に着目し、説明を生成する方式を考案した。

さらに入力としての仮説・データは、以下の形式をとる：

- ・ テキスト（pdf を含む）
- ・ 表
- ・ チャート（グラフ）
- ・ コード
- ・ 画像

また出力としての仮説・説明も同様の形式をとる。

これらの形式の実施可能性については、生成 AI や大規模言語モデルを用いて、仮説生成・説明の複数の実験によって確認した。

B. 説明生成のためのアーキテクチャの提案

理論的研究の出発点となるように、構造化アプローチに基づいた説明生成のレファレンスアーキテクチャを提案した。

本研究調査で提案する、生成 AI の発展を利用した「**仮説・説明生成のためのレファレンスアーキテクチャ**」を下図に示す。

各モジュールの役割を説明する。

データの入出力管理はデータベースが行い、データの作り方は主にツール管理が担当する。

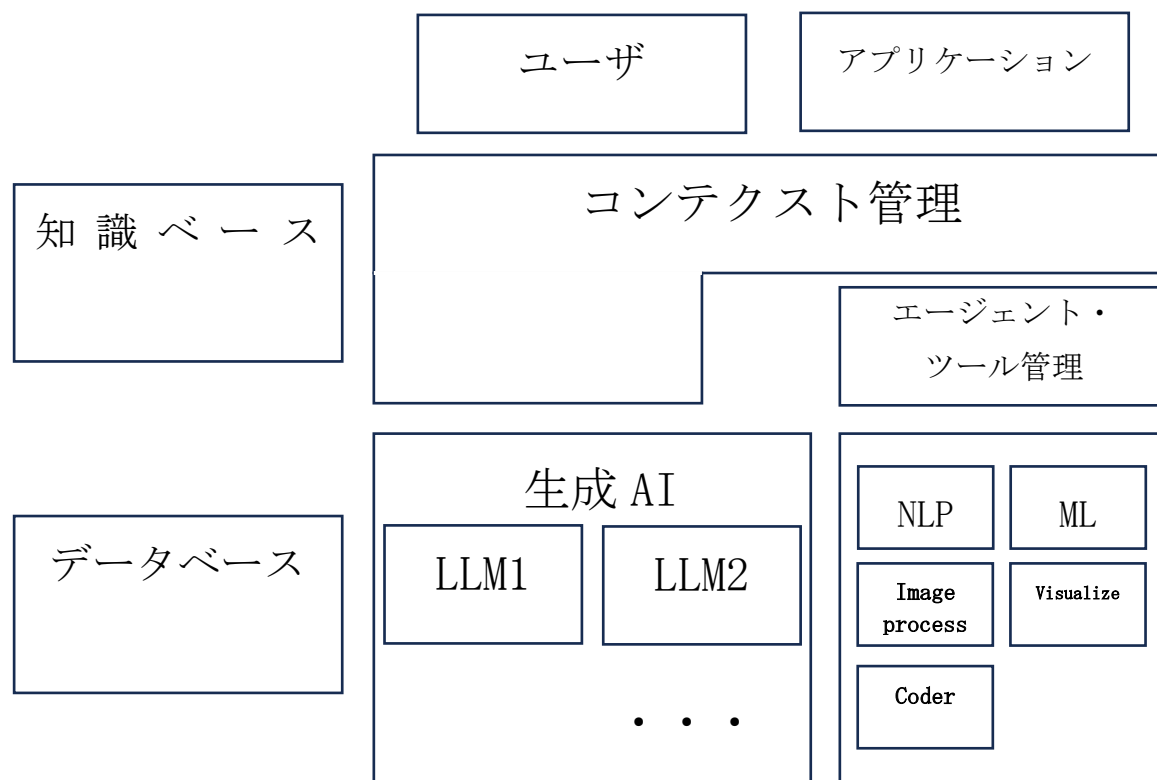
仮説の作り方には、prompt engineering を発展させたコンテキスト管理を通して**生成 AI**（ChatGPT や Gemini）や**大規模言語モデル LLM**（GPT や Gemma）を利用するが、高度な推論には永続性のある知識ベースを用い、エージェント管理やツール管理（主に機械学習 ML）も利用できるようにする。

仮説の特徴管理については、コンテキスト管理がツール管理を用いて行う。

仮説の結果については、コンテキスト管理が生成 AI や大規模言語モデルを用いるか、ツール管理（機械学習 ML）を用いて生成する。

仮説結果の理由と実行計画の生成提示には、コンテキスト管理が、生成 AI や大規模言語モデルを用いて行う。

アーキテクチャの実現は基本的にオンプレミス（構内コンピュータシステム）でもクラウドサービスでもよい。本研究では、各種実験にクラウドサービス（Google Colaboratory）を利用した。



仮説・説明生成のためのレファレンスアーキテクチャ

C. 多様な説明の記述手段・開発環境の探求と生成 AI を用いた自然言語による説明提示

仮説の説明などの説明対象の記述にプログラミング言語と分析用ライブラリが利用でき、SQL 言語と専用ライブラリによって、データの生成と仮説の生成の両方を一度に記述できる。そこで説明記述の手段として SQL とプログラミング言語を検討対象にする。さらに多様なメディア（テキスト、画像、グラフ等）の利用を許す統合開発環境の普及も考慮しながら、これらの組み合わせによる説明記述に対して、**大規模言語モデル**（Large Language Model）に基づく生成 AI に適切な説明を引き出すための prompt engineering を用いて、自然言語により説明を提示するという試みを検討した。

具体的には現在の**生成 AI**の発展に基づいて、オープンな**大規模言語モデル**（Google Gemma など）を用いて**仮説の生成方法**と**仮説の説明方法**の実験を、クラウド上の統合開発環境（Google Colaboratory）を利用して複数種類用意して実行し、実施可能性を確認した。特に**仮説生成**では、大規模言語モデルと外部**データベース**や**知識ベース**を組み合わせる方式や関係抽出による**知識グラフ**の構築・利用を行う方式も実験に加えて実施した。

従来は python に代表されるプログラミング言語が仮説生成の手段として使われてきたため、仮説の解釈説明には低水準という問題点があった（資料5）（資料6）。

そこで、その問題点を解決するために、prompt engineering に基づく**生成 AI**による、仮説の解釈説明に適した仮説の記述手段を探索し、それを実験した。具体的には SQL に加えて、python で記述された実行計画を**生成 AI**（Google Gemini など）に対して prompt として入力し、**自然言語**（英語・日本語）による**説明**を生成する方式を実験し、実施可能性を確認した。

D. ユースケースの検討深化

開発した技術を進んだ社会基盤的応用に適用した場合の妥当性を検討した。

具体的には分析の対象として、昨年度に収集したユースケースである地域観光（資料2）（資料3）（資料4）（資料8）、スマートシティ、月惑星科学、ソフトウェア品質管理（資料1）などの異なる分野に、新たに医療分野（資料7）も加えて、解釈説明の観点からユースケース分析を行った。

特に情報科学だけでなく**科学哲学**や**認知心理学**のような周辺技術の調査もあわせて行った。その結果、説明には複数の種類があることが分かった。

科学哲学においては、以下のような説明モデルを提案する。

- ・ **被覆法則モデル**
説明は、法則と初期条件を前提とする論理的に有効な演繹的推論の結論として提示される。
- ・ **反事実依存モデル**
現実世界で実際に起こった事象だけでなく、起こり得た別の可能性（反事実的状況）を考慮することで、因果関係の強固さを評価する。
- ・ **統計的関連性モデル**
条件Cが事象Eの発生確率に統計的に有意な影響を与えるかどうかを評価する。
- ・ **機能的・統合的モデル**
少数の基本的な原理から多様な現象を演繹的に導き出すことができる理論は、高い説明力を持つと評価される。

さらに認知心理学においては、以下のような2種類の説明の枠組みを提案する。

- ・ **目的論的説明**
ある現象や行動を「何のために存在するか」「どのような目的を果たすか」という観点から理解しようとする説明である。
- ・ **機械論的説明**
ある現象を構成要素と因果的メカニズムに基づいて記述するものであり、「どのように？」(How?)という問いに答える。

本研究調査で対象とするユースケースでは、上記の説明のモデルのうち、**被覆法則モデル**、**反事実依存モデル**、**統計的関連性モデル**および**機械論的説明**を重視するが、その他も補完的に考慮することが肝要であることが分かった。

なお研究開発は、共同研究者に加えて仏ポー大（L'Université de Pau et des Pays de l'Adour）のR. Chbeir 教授やJAXA 山本幸生准教授他からなる多様で国際的な研究協力者ネットワークを活かして実施した。

3-2 研究調査の将来計画

本研究期間の終了後は、構築した統合的仮説生成・説明方式と実装アルゴリズムを複数のユースケース（スマートシティや月惑星科学など）に適用して、その妥当性の検証を行い、ソーシャルビッグデータの利活用と流通を促進すると同時に情報通信分野において、言わば**生成AI**を利用した「**解釈説明工学**」のような新領域の構築を目指す。

将来的には、本研究の成果により、ビッグデータを媒介として、これまで縁遠かった異業種セクタが協業できるようなダイナミックな知的社会 Society5.0 を実現することを目指す。その意味で、本研究は社会的意義も十分に高いと考えられる。

以下のように成果を公表するとともに、研究終了後も研究成果を発展させる。

- ・ 著作や論文発表を通して、研究成果を公表する。
- ・ 国際会議・研究フォーラムにおいて仮説解釈説明に関連するスペシャルトラックなどを開催して研究成果の発表をおこなうと同時に、関連する研究者の交流する機会を設けて、この分野の強化を行う。
- ・ 所属する大学において国際研究拠点の構築を推進する。
- ・ HPで研究成果を周知するとともに所属する大学において社会人向け授業を行う。

新しいユースケースへの適用による提案の妥当性を確認していくとともに、新領域確立に向けて新たに発掘された研究課題に取り組んでいく。

【参考文献】

1. 内閣府、Society 5.0. https://www8.cao.go.jp/cstp/society5_0/index.html
2. NRI、「情報銀行」は消費者に受け入れられるか？. NRI Journal、2018.
3. EU GDPR. <https://gdpr.eu/>
4. M. E. Kaminski、The Right to Explanation、Explained. University of Colorado Law School、2019.
5. C . Molnar 、 Interpretable Machine Learning 、
<https://christophm.github.io/interpretable-ml-book>
6. D. Baehrens et al.、How to explain individual classification decisions. JMLR、11、2010.
7. B. M. Lake et al.、Human-level concept learning through probabilistic program induction. Science 350(6266):1332-1338、2015.
8. M. T. Ribeiro et al.、“Why Should I Trust You?” Explaining the Predictions of Any Classifier. <https://arxiv.org/abs/1602.04938>
9. The Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/>
10. U. Malik et al.、SQL for Data Analytics、Packt. 2019.
11. 石川博、仮説のつくりかた：多様なデータから新たな発想をつかめ。320 頁、共立出版、2021.
12. Hiroshi Ishikawa、Yasushi Miyata、Social Big Data: Case Studies. Trans. Large Scale Data Knowl. Centered Syst. 47、pp. 80-111、Springer Nature、2021.
13. Hiroshi Ishikawa、Yukio Yamamoto、Masaharu Hirota、Masaki Endo、An Explanation Framework for Whole Processes of Data Analysis Applications: Concepts and Use Cases. Intl Journal on Advances in Software、vol. 13、No. 1&2、pp. 1-15、2020.
14. 加藤広大、山田竜平、山本幸生、廣田雅春、横山昌平、石川博、惑星の位置関係に基づく深発月震分類のための特徴量の検討. Journal of Space Science Informatics Japan、Vol. 7、pp. 43-52. 2018.
15. Yusuke Takamori、Riku Watanabe、Daiju Kato、Masaki Endo、Hiroshi Ishikawa: Numerical estimation of new COVID-19 positive cases using time series analysis by machine learning. Proc. Intl. Conf. ACM MEDES、2021.

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
[1] Quality Control Methods Using Quality Characteristics in Development and Operations.	Digital 2024, 4 https://doi.org/10.3390/digital4010012	2024.3
[2] Time-Series Estimation of the Best Time to View Seasonal Organisms Using Geotagged Tweets and Co-occurring Words.	Communications in Computer and Information Science, vol 2022. Springer Nature	2024.2
[3] Estimating the best time to view cherry blossoms using Natural Language Processing tasks	INFSOC IWIN 2023	2023.9
[4] Relation between time-series forecasting methods and estimation of best times to see cherry blossoms	INFSOC IWIN 2023	2023.9
[5] Hypothesis Generation and Interpretation: Design Principles and Patterns for Big Data Applications	Springer Nature ISBN 9783031435393	2024.2
[6] モダン SQL データ管理から分析へ	共立出版 ISBN 9784320125698	2023.11
[7] Temporal fractal nature of human behavior in personal computer operations	Physica A: Statistical Mechanics and its Applications Elsevier https://doi.org/10.1016/j.physa.2024.129765	2024.6
[8] Proposal for a System to Estimate the Best Time to See Yellow Leaves Using IoT Devices for Tourists	MMEDIA 2025: The Seventeenth International Conference on Advances in Multimedia	2025.5