

令和 6 年 5 月 28 日現在

機関番号：82111

研究種目：基盤研究(B)（一般）

研究期間：2019～2023

課題番号：19H04168

研究課題名（和文）解釈可能なAIシステムの実現に向けたナレッジグラフに基づく推論・推定技術の体系化

研究課題名（英文）Classification of inference and estimation techniques based on knowledge graphs for realizing Explainable AI systems

研究代表者

川村 隆浩（KAWAMURA, Takahiro）

国立研究開発法人農業・食品産業技術総合研究機構・農業情報研究センター・副センター長

研究者番号：10426653

交付決定額（研究期間全体）：（直接経費） 13,100,000円

研究成果の概要（和文）：本課題では、解釈可能なAIの実現を目指し推理小説をナレッジグラフ化したデータセットを公開し、理由付きで犯人を推理することをタスクとし、オープンサイエンス形式にて論理推論や機械学習技術の深耕・体系化を実施した。結果として、国内外で計8回の技術コンテストを実施、延べ約50件の技術提案をいただき、それらを分類、整理したものを以下の学術論文等にて発表した。

Contextualized Scene Knowledge Graphs for XAI Benchmarking, T. Kawamura, et al.,
11th Int'l J. Conf. on Knowledge Graphs 2022

研究成果の学術的意義や社会的意義

ナレッジグラフとは、近年、DX化を推進するツールとして注目されているデータの整理・活用方法である。データ群をナレッジグラフとして統合することで、AIでの活用が容易になるという利点がある。本課題では、AIの社会実装に当たって必ず問題となるAIシステムの解釈可能性（何故、AIはそういう結果を出したのか）にフォーカスし、ナレッジグラフとAIシステムの在り方を整理し、広く社会に浸透させることを図ってきた。また、2023年からは生成AIによってAI分野に大きな変革が起きたことを踏まえ、生成AIを説明するためにナレッジグラフを如何に用いるかにもフォーカスして技術を整理し、社会実装への貢献を図った。

研究成果の概要（英文）：In this project, with the aim of realizing Explainable AI, a dataset of a mystery novel converted into a knowledge graph was made public, and the task was to deduce the culprit with a reason. As a result, we held 8 technology contests in Japan and abroad, and received a total of about 50 technology proposals, which we classified and organized, and published them in the following academic papers.

Contextualized Scene Knowledge Graphs for XAI Benchmarking, T. Kawamura, et al.,
11th Int'l J. Conf. on Knowledge Graphs 2022

研究分野：知識処理

キーワード：ナレッジグラフ 機械学習 論理推論 解釈可能性 説明可能性 オントロジー セマンティックWeb 人工知能

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

本研究におけるナレッジグラフ推論チャレンジは、2017年ごろから人工知能学会セマンティックウェブとオントロジー研究会に参加していた研究者ら有志で企画し、2018年に第1回を開催した。当時、深層学習を起爆剤とした人工知能(AI)技術への関心が高まり、様々な社会システムにAI技術が利用されるようになりつつある中で、安全・安心にAIを活用していくためにはシステムが判断に至った理由を説明できる(解釈可能性を有する)AI技術が必須となるという研究背景の下、解釈可能性にフォーカスした推論(推定)技術に関するチャレンジ(技術コンテスト)を開催し、オープンサイエンスに則って認識の共有と必要な技術の開発・促進を図りたいという趣旨で研究を開始した。

2. 研究の目的

しかし当時、帰納的な機械学習技術と演繹的な知識活用技術を適切に評価するためのデータセットは存在しなかった。機械学習による関係性の推定等で用いられるベンチマーク用データセットの多くは、比較的単純な関係性しか含んでおらず、複数のサブタスクを組み合わせると全体のゴールを達成する問題には使えない。一方で、ルールの知識の多くはドメインに依存しており、機械学習にも適用できる大規模で汎用的なテストセットとして使えるものは少ない。推論・推定技術を用いて解釈可能なAIを実現するためのタスクに適したデータセットとしては、2項関係の推定で足りるような比較的単純な関係だけでなく、現実社会を反映したより複雑な、例えば時間的、因果関係的、確率的関係性を含んでいることが必要である。

そこで、本研究ではまず上記の要件を満たす推論・推定技術共通の評価用データセットとして、現実の社会問題や人間関係など複雑で構造的な関係性を含んだシーンまたはイベント中心のナレッジグラフを構築し、それらをオープンデータとして公開し、幅広くIT技術者・研究者から推論・推定に関する手法を結集して、提案技術の分類を行うことを目的とした。

本タスクは、ナレッジグラフの補完(Knowledge Graph Completion)の一種として一般化することができるためナレッジグラフを含むさまざまな知識ベースの構築、情報抽出や関係抽出、知識更新や保守などに適用できる汎用的な問題設定として位置づけることができる。また、現実の社会問題にフォーカスし、結果に対する説明性を重視していることに加え、以下の点に研究目的の独自性が存在する。

- ・ 現実問題は全てが個別ケースであり類似のシーンが必ずしも複数回出現しないため、知識またはデータが必ずしもビックデータとはなりにくく学習が困難であること、
- ・ 単に1つの関係性をベクトル空間における近似性などで説明するのではなく、それらを組み立てて、または連鎖させて全体としてのゴールを導く必要があること、
- ・ ナレッジグラフには登場人物による虚偽発言なども含まれていること、などである。

3. 研究の方法

2018年に開催した第1回ナレッジグラフ推論チャレンジでは、解釈可能性をはかる対象タスクとしてシャーロックホームズの推理小説を題材とし、小説で描かれた事件の犯人を同定し、その結論に至った理由を説明することとした。チャレンジ開催にあたり、対象とする小説の内容を元に構築したナレッジグラフを公開し、参加者には公開されたナレッジグラフを利用して、犯人の推定とその理由を生成するシステムの開発を促した。その際、提供された推理小説のナレッジグラフと合わせて、参加者が独自に要した外部知識などを利用することも推奨した。

提供する推理小説のナレッジグラフについては、第1回チャレンジでは「まだらの紐」1編のみを対象としていたが、第2回以降、対象小説を増やしていき、現在は8つの小説を対象としたナレッジグラフを公開している。その間、推理小説のナレッジグラフを表すデータモデルの改良や語彙の統制、ナレッジグラフ構築のためのガイドラインの整備など、提供するナレッジグラフの品質の改善に取り組んだ。

これらの取組みを通し、推理小説を題材としたナレッジグラフ推論チャレンジを2018年の第1回から毎年開催を続け、国内版を4回、国際ワークショップの形で開催した国際版の推論チャレンジ(The 1st International Knowledge Graph Reasoning Challenge 2023)を合わせて計5回開催した。応募作品の総数は、対象小説1つ以上のタスクを解くシステムを開発する本部門が計15作品、いずれかのタスクを部分的に解くツールを開発するツール部門が9作品、実現方法のアイデアを対象としたアイデア部門が10作品に上った。

さらに2022年からは、より実社会のニーズに近いタスクを設定することを目指したナレッジグラフ推論チャレンジ【実社会版】を開催している。実社会版の推論チャレンジでは、高齢者の家庭内における危険な状況を検出して説明するシステムの開発をタスクとして設定した。この

タスクを実施するために提供したデータは、シナリオデータから得られる仮想環境内での日常的な行動をエミュレートした動画、動画の内容に対応するナレッジグラフ、動画の元としたシナリオの3種類のデータで構成される。これらのデータを用いることにより、動画とナレッジグラフという異なる形式のデータを統合的に利用した推論および説明生成を行う技術の開発が行われることを企図した。第1回のナレッジグラフ推論チャレンジ【実社会版】には6作品が応募され、ナレッジグラフと動画を組み合わせた作品も見られた。また、推論チャレンジ【実社会版】の第2回は2023年から2024年にかけてThe Second International Knowledge Graph Reasoning Challenge (IKGRC 2024)として実施され、4作品の応募があった。現在、第3回目以降の継続開催について検討を進めている。

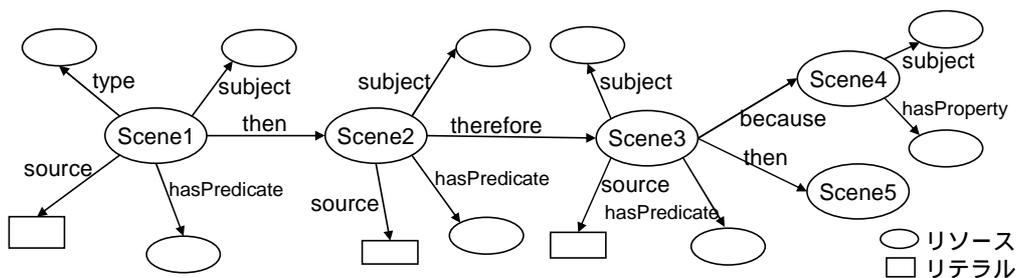
一方、最終年となったナレッジグラフ推論チャレンジ2023では、近年、大規模言語モデルを用いた生成AIが注目され、AIにおける説明生成技術が大きな転換点を迎えたことを踏まえ、チャレンジが対象とするタスクの再検討を行った。そして、これまでの推論チャレンジで取り組んできた説明可能なAI技術の開発・共有と、大規模言語モデルによる生成AIで問題とされている品質保証に関する課題の共通点に注目し、大規模言語モデルを用いたナレッジグラフの構築というタスクを新たに設定した。

4. 研究成果

(1) ナレッジグラフのスキーマ

ナレッジグラフの設計にあたっては、2017~2018年に日本国内でオープンワークショップを5回開催し、スキーマの基本設計と詳細設計、ナレッジグラフ構築方法の検討と具体的な構築作業を行った。参加者の延べ人数は110名程度であった。ワークショップでは、まずナレッジグラフの試験的な記述を通して、推論・推定に必要な知識(ナレッジグラフで記述すべき内容)とその表現方法について検討し、参加者からのフィードバックを受けて、小説内で描かれている場面(シーン)とシーン間の関係を中心に、各シーンに関与した人、もの、場所などを記述するという基本方針を決定した。

また、スキーマ設計にあたっては、題材とする小説を表すための表現力(expressiveness)に加えて、ナレッジグラフとしての構築のし易さや、推論処理にデータとして与える際の利便性を考慮し、シーンを中心に主に5W1Hをエッジとして持つスキーマとした。したがって、推理小説のストーリーが各シーンと、シーン間の関係によって表現されている。推理小説の各シーンには一意のIRIが割り当てられており、そのIRIを主語に人や組織、場所に関する情報を目的語として追加することで小説内の1場面を説明する。シーン間の関係では、そのIRIを参照することで時系列的な行動や事象の因果関係を説明する。これによって一連のストーリーを表現する。なお、常識を公理として記述したり、時刻表のような情報を表すため、ルールやテーブル形式のデータもリンク可能である。また、自然言語処理のため、小説の内容をリテラル値として保持している。下図にナレッジグラフの全体イメージを示す。



(2) チャレンジ応募作品の一覧(一部抜粋)

本チャレンジへのアプローチは大きく以下の2つに大別された。

- ・一階述語論理によるルールやオントロジーの定義などに基づいて推論する知識处理的なアプローチ
- ・提供したナレッジグラフや他の事件、小説を教師データとして学習することにより、犯人に結び付く情報を見つける機械学習的なアプローチ

そこで、提案技術を分類するための1つ目の観点として、(1)知識処理/機械学習のいずれか、もしくはその両方を中心とした手法であるかを比較した。また、本チャレンジで提供する推理小説のナレッジグラフは小説の内容から主要な部分を抽出し、小説の記載内容に沿って作成したものである。そのため、小説の読者にとっては常識であるために明示的には記述されない知識(例:ナイフで心臓を指されると死ぬ)が多く含まれる。そのため、本チャレンジでは推論に必要な外部知識を応募者が補完することを認めており、有用な外部知識を導入ことも重要な評価観点としている。そこで2つ目の観点としては、(2)各手法で用いられている外部知識を比較し

た。下表に上記2つの観点を中心に提案技術を比較した結果を示す。

(1)のアプローチの別については、第1回ではほとんどの作品が知識処理を中心としていたが、第2回より機械学習を用いたアプローチが大幅に増加していることが分かる。また、双方のアプローチを相互補完的に用いる手法も見られるようになり、本チャレンジのタスクが知識処理と機械学習の融合研究の対象タスクの1つとなり得ることが示唆された。一方、(2)については、第1回では推論に必要な知識を独自に作成したオントロジーやルール記述として用いる手法が多く見られたが、第2回以降はWordNetやWikipedia、Wikidataなどの既存のリソースを活用する手法の提案が多くなされている。この背景には、独自に知識を記述する際に必要なコストを軽減したいという課題があるだろう。今後、既存のリソースが必要な外部知識としてどの程度まで活用できるかの評価が重要な課題の1つになるとと思われる。さらなる詳細は発表論文、またはWebサイトを参照していただきたい。

開催回	応募部門	応募者	(1)中心とするアプローチ	(2)利用した外部知識	手法の概要
第1回	本部門1	田村光太郎, 外園康智 (株式会社野村総合研究所)	知識処理	独自に作成したルール等	テンソル分解とSAT問題を軸とした証言内容を満たす状況の探索
	本部門2	上小田中411 (株式会社富士通研究所)	知識処理	独自に作成したオントロジーおよび推論規則	動機, 機会, 手段に関する知識を記述したオントロジー, および推論規則 (SHACL) を利用した推論
	本部門3	FLL-ML (株式会社富士通研究所)	機械学習	シャーロック・ホームズの他の小説本文	他の小説で機械学習した内容をもとに犯人を推定する
	本部門4	teamOIF (立命館大学)	知識処理	独自ルール (犯人推定) およびオントロジー (動機)	独自ルールとオントロジーを用いて, 犯人となる登場人物を推定
	本部門5	塚越雄登 (電気通信大学)	知識処理	独自記述のルール	必要な独自ルールをトリプルストア (Stardog) が推論可能な形式で記述し犯人の推定に利用
	アイデア部門1	白松研 feat. 59 (名古屋工業大学)	-	なし	誰が犯人かを複数のエージェントで議論することにより, 説明性をもって推理するというアイデアの提案
	アイデア部門3	橋本一成, 他 (富士ゼロックス株式会社)	知識処理	独自に作成したオントロジー	事件に関わるイベントや状況を独自作成の推論オントロジーに基づいて記述し, 推論および犯行根拠の生成
第2回	本部門1	田村光太郎, 角田充弘, 外園康智 (株式会社野村総合研究所)	知識処理 + 機械学習	小説の本文および, 独自に作成したルール, 外部情報等	犯行フレームとBERTの類似文判定によるファクト抽出, 述語論理を用いた仮設推論などの組み合わせ
	本部門2	黒川茂莉 (株式会社KDDI総合研究所)	機械学習	なし	ナレッジグラフ埋め込み技術を用いた犯人推論
	本部門3	伊鍋貴宏, 石映沂 (株式会社サキヨミラボ)	機械学習	なし	ナレッジグラフの双曲空間の埋め込み, ALBERTモデルを用いた推論過程の可視化
	本部門4	上小田中411 (株式会社富士通研究所)	知識処理 (+ 機械学習)	独自作成のオントロジー, WordNet, Wikipedia	第1回のオントロジー (動機, 手段) を汎用知識 (WordNetとWikipedia) を用いた拡大し, ナレッジグラフ埋め込みを学習して予測
	ツール部門1	勝島修平 (東京都市大学)	機械学習	なし	グラフニューラルネットワークを用いた未知エンティティ問題として推定
	ツール部門2	中川豪 (法政大学)	-	なし	tfidfによる重要語抽出ツールによる犯人に結び付く単語の抽出
	ツール部門3	佐藤社, 西山慶一郎 (大阪電気通信大学)	知識処理	NRC Emotion / Affect Intensity Lexicon	ナレッジグラフを単語の感情強度の辞書とマッピングし, シーンごとの登場人物の感情の変化を推論
	アイデア部門1	村上勝彦 (富士通研究所), 高松邦彦 (神戸常盤大学), 杉浦あおい (神戸市立西神戸医療センター)	機械学習	Wikipedia	「踊る人形」の暗号解読の数理的方法を踏まえた検討
	アイデア部門2	村嶋 義隆	機械学習	なし	自然言語処理を用いた犯人推測方法の検討
第3回	本部門1	勝島修平, 穴田一 (東京都市大学大学院)	機械学習	なし	単語の共起のCBoW (Continuous Bag-of-Words) による学習と, 主成分回帰分析
	本部門2	黒川茂莉 (KDDI総合研究所)	機械学習	ConceptNet	ナレッジグラフの欠損したリンクを予測による犯人推定
	ツール部門1	加藤敦丈	知識処理	なし	
	ツール部門2	鶴飼孝典 (富士通研究所)	知識処理	Wikidata	Wikidataのクラス階層を利用した小説のオントロジーを簡易構築
	アイデア部門1	荒川博, 黄瀬輝, 佐々木裕之, 辻寿嗣, 大坪紹二 (パナソニック株式会社)	機械学習	WordNet	WordNetの動詞との類似度 (word2vec, BERTを利用) を用いた多義性の自動判別
	アイデア部門2	和泉幸樹 (大阪電気通信大学)	知識処理	Wikidata, ICD-10	WikidataおよびICD-10とのマッピング情報を用いた推論
	アイデア部門3	伊鍋貴宏, 石映沂 (サキヨミラボ)	機械学習	なし	ナレッジグラフの双曲空間埋め込みを用いたナレッジグラフの作成者の癖を考慮した推定

最後に、大規模言語モデルを用いたナレッジグラフの構築をタスクとして実施したナレッジグラフ推論チャレンジ2023では、5つの作品に応募され、それぞれ異なる観点からの大規模言語モデルの活用が発表された。今後、これまでの取り組んできたナレッジグラフを用いた説明可能なAI技術の流れを汲みつつ、大規模規模言語モデルによる生成AI技術を組み合わせることで、安全・安心なAIに新たな潮流が生み出されることを期待したい。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 Kawamura Takahiro, Egami Shusaku, Matsushita Kyoumoto, Ugai Takanori, Fukuda Ken, Kozaki Kouji	4. 巻 0
2. 論文標題 Contextualized Scene Knowledge Graphs for XAI Benchmarking	5. 発行年 2022年
3. 雑誌名 IJCKG '22: Proceedings of the 11th International Joint Conference on Knowledge Graphs	6. 最初と最後の頁 64-72
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3579051.3579061	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kawamura Takahiro, Egami Shusaku, Tamura Koutarou, Hokazono Yasunori, Ugai Takanori, Koyanagi Yusuke, Nishino Fumihito, Okajima Seiji, Murakami Katsuhiko, Takamatsu Kunihiko, Sugiura Aoi, Shiramatsu Shun, Zhang Xiangyu, Kozaki Kouji	4. 巻 12032
2. 論文標題 Report on the First Knowledge Graph Reasoning Challenge 2018: Toward the eXplainable AI System	5. 発行年 2020年
3. 雑誌名 Semantic Technology, LNCS, Springer	6. 最初と最後の頁 18~34
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-41407-8_2	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計9件（うち招待講演 0件／うち国際学会 1件）

1. 発表者名 古崎 晃司, 江上 周作, 松下 京群, 鶴飼 孝典, 川村 隆浩
2. 発表標題 説明生成のための知識グラフ構築ガイドラインの考察 - ナレッジグラフ推論チャレンジを例にして
3. 学会等名 第36回人工知能学会全国大会論文集 2E6-GS-3-02
4. 発表年 2022年

1. 発表者名 福田 賢一郎, 江上 周作, 鶴飼 孝典, 森田 武史, 大野 美喜子, 北村 光司, QIU YUE, 原 健翔, 古崎 晃司, 川村 隆浩
2. 発表標題 イベント中心知識グラフによる人間生活を含む環境のサイバー空間への転写にむけて
3. 学会等名 第36回人工知能学会全国大会論文集 3L4-GS-8
4. 発表年 2022年

1. 発表者名 江上周作, 鵜飼孝典, 太田雅輝, 川村隆浩, 松下京群, 古崎晃司, 福田賢一郎
2. 発表標題 イベント中心ナレッジグラフ埋め込みにおける メタデータ表現モデルの分析
3. 学会等名 人工知能学会セマンティックWebとオントロジー研究会(SIG SWO), 人工知能学会研究会資料 SIG-SWO-057-05
4. 発表年 2022年

1. 発表者名 鵜飼 孝典, 江上 周作, 大野 美喜子, 福田 賢一郎, 川村 隆浩, 古崎 晃司, 松下 京群
2. 発表標題 コンペティションによる協創：安心安全を守るAIの開発に向けて
3. 学会等名 第191回ヒューマンインタフェース学会研究会「社会のデザイン・市民のデザイン」 SIG-UXSD-15
4. 発表年 2022年

1. 発表者名 Kouji Kozaki, Shusaku Egami, Kyoumoto Matsushita, Takanori Ugai, Takahiro Kawamura
2. 発表標題 Knowledge Graph Reasoning Techniques through Studies on Mystery Stories -Report on the Knowledge Graph Reasoning Challenge 2018 to 2020
3. 学会等名 Proceedings of 1st International Workshop on Knowledge Graph Reasoning for Explainable Artificial Intelligence (KGR4XAI) (国際学会)
4. 発表年 2021年

1. 発表者名 古崎 晃司, 江上 周作, 松下 京群, 鵜飼 孝典, 川村 隆浩
2. 発表標題 第3回ナレッジグラフ推論チャレンジ2020開催報告～説明性のある人工知能システムを目指して～
3. 学会等名 第35回人工知能学会全国大会論文集 (2H1-GS-3a-01)
4. 発表年 2021年

1. 発表者名 鶴飼 孝典, 江上 周作, 大野 美喜子, 窪田 文也, 福田 賢一郎, 川村 隆浩, 古崎 晃司, 松下 京群
2. 発表標題 高齢者の家庭内事故予防に役立つAIシステムの開発 -産業版ナレッジグラフ推論チャレンジに向けて-
3. 学会等名 人工知能学会セマンティックWebとオントロジー研究会(SIG SWO), 人工知能学会研究会資料 SIG-SWO-056-15
4. 発表年 2022年

1. 発表者名 川村 隆浩, 江上 周作, 松下 京群, 田村 光太郎, 角田 充弘, 外園 康智, 黒川 茂莉, 鶴飼 孝典, 古崎 晃司
2. 発表標題 第2回ナレッジグラフ推論チャレンジ2019開催報告 ~ 説明性のある人工知能システムを目指して ~
3. 学会等名 第34回人工知能学会全国大会 (104-GS-4-01)
4. 発表年 2020年

1. 発表者名 川村 隆浩, 江上 周作, 田村 光太郎, 外園 康智, 鶴飼 孝典, 小柳 佑介, 西野 文人, 岡嶋 成司, 村上 勝彦, 高松 邦彦, 杉浦 あおい, 白松 俊, 張 翔宇, 古崎 晃司
2. 発表標題 第1回ナレッジグラフ推論チャレンジ2018開催報告 - 説明性のある人工知能システムを目指して -
3. 学会等名 第33回人工知能学会全国大会, 1K2-J-4-04 2019年6月
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

1st Int'l Knowledge Graph Reasoning Challenge 2023 https://ikgrc.org/2023/ ナレッジグラフ推論チャレンジ https://challenge.knowledge-graph.jp/2021/ ナレッジグラフ推論チャレンジ https://challenge.knowledge-graph.jp/2020/ ナレッジグラフ推論チャレンジ https://challenge.knowledge-graph.jp/2020/

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	古崎 晃司 (Kozaki Koji) (00362624)	大阪電気通信大学・情報通信工学部・教授 (34412)	
研究 分 担 者	江上 周作 (Egami Shusaku) (20846000)	国立研究開発法人産業技術総合研究所・情報・人間工学領 域・研究員 (82626)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計2件

国際研究集会 The 1st International Knowledge Graph Reasoning Challenge 2023	開催年 2022年～2023年
国際研究集会 The 1st International Workshop on Knowledge Graph Reasoning for Explainable Artificial Intelligence (KGR4XAI)	開催年 2021年～2021年

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関