

事象概念を用いた推理知識表現方法

橋本 一成, 鷹合 基行, 外池 昌嗣, 辰巳 大祐, 平出 聡

富士ゼロックス株式会社

〒 220-8668 神奈川県横浜市西区みなとみらい 6-1

{Kazunari.Hashimoto, Motoyuki.Takaai, Masatsugu.Tonoike,
Tatsumi.Daisuke, Satoshi.Hirade}@fujixerox.co.jp

Abstract. 我々は, 世の中に存在する状況や行動を機械可読性を備えた情報として定式化し, それらの情報から解釈性を備えた形で情報を抽出することを可能とした知識表現手法の研究を進めている. 知識表現手法を具体化した推理オントロジーによって, 不正情報を抽出しその不正根拠を論理的に説明する仕組みを提案した. ナレッジ推論チャレンジのデータを用いることにより犯行事象の抽出及び犯行根拠の説明にも対応できることがわかった.

Keywords: 事象, 推理オントロジー, 知識表現

1 はじめに

現代社会において, さまざまな状況や行為が様々な基準や制度等によって規制されている. これらの基準や制度等に反している状況や行為を意味する不正な状況や行為は存在している. しかし, これらの不正な状況や行為は全てが明らかにされておらず潜在化しているものが存在する. 設計製造分野を例にすると, 1つの製品は多くの設計者によって設計されている. 各設計者の設計間に相対的に矛盾が生じると製品仕様に対する不正な状況となる. しかし, これらの不正な状況全てを設計者が把握できない場合が存在する [1]

我々は, 世の中に存在する状況や行動を機械可読性を備えた情報として定式化し, それらの情報から解釈性を備えた形で不正な状況を表す情報を抽出することを可能とした知識表現手法の研究を進めている. 本チャレンジに向けて, 知識表現手法を犯行情報に対応した形へと具体化した.

本論文では, 2章で我々の研究である事象概念を用いた不正根拠抽出について述べ, 3章では本研究を応用した本チャレンジへのアプローチについて述べ, 4章でまとめる.

2 事象概念を用いた知識表現方法

2.1 事象概念

人や集団または物体は, 時間や空間に基づく環境の中でさまざまな状態や動作を形成している. 我々はこれらの状態や動作を総称して事象と呼ぶ. 事象を形成する人や集団・物体・時間・空間を実体と呼ぶ.

事象は, 対象世界において存在が明らかにされている事象と存在が明らかにされていない事象に分けられる. 前者を顕在事象, 後者を潜在事象と呼ぶ. ある基準

に反する事象を「ある基準に対する不正事象」と呼ぶ。本研究の目的はこの不正事象及び不正事象が成立する根拠（以下、不正根拠と記す）を機械的に求めることにある。不正事象は顕在事象または潜在事象として存在するため、潜在事象を明らかにすることが必要となる。我々は潜在事象の大部分を顕在事象から求める（生成する）ことができるという仮説を立てた。この既知の事象を入力とし、未知の事象を出力とするプロセスを事象推理または推理と呼ぶ。

不正根拠抽出を実現するためには、事象推理を可能とする事象の表現方法が求められる。事象間の関係は「事象における実体の意味的關係」によって説明することができる。

例えば、「事象 X:A さんがPC を落とした」「事象 Y:PC が壊れた」という2つの事象の間に因果関係が存在するとする。「事象 X における PC の意味的關係:落下物」と「事象 Y に対する PC の実体意味的關係:破損物」はいずれも実体である PC の意味的關係を表している。2つの事象に存在する意味的關係は同じ実体に基づくことにより2つの事象間の関係を表現しているといえる。我々は「事象における実体の意味的關係」を、事象実体意味概念と呼ぶ。

以上のように、事象概念 \rightarrow 事象実体意味概念 \rightarrow 実体概念という概念間を用いて、事象のコンテキストをナレッジグラフで表す。

2.2 提案手法の概要

本研究では、文書や構造化情報等の顕在化している情報から事象概念を用いて不正根拠を抽出する。事象概念に基づく不正根拠抽出フローを Fig.1 に示す。

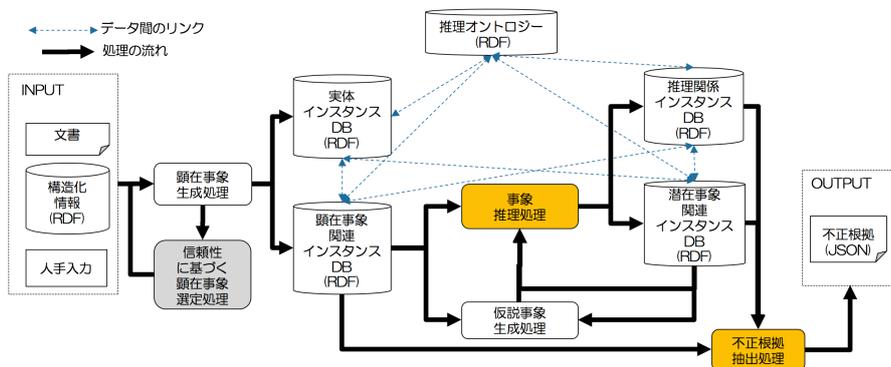


Fig. 1. 事象概念に基づく不正根拠抽出フロー

事象のナレッジグラフは事象インスタンス・事象実体意味インスタンス・実体インスタンスによって成り立つ。

顕在事象関連インスタンス DB・潜在事象関連インスタンス DB とはそれぞれの事象の事象関連インスタンスを格納したものである。事象関連インスタンスとは事象のナレッジグラフの一部であり、事象インスタンス、事象実体意味インスタンス及び前記2種のインスタンス間のリンクから成り立つ。

実体インスタンス DB は各事象のナレッジグラフで出現した実体インスタンスを格納したものである。推理関係 DB とは、事象推理によって生成された推理関係を表すインスタンスを格納したものである。上記 DB 及び DB 間のインスタンスやリンクの意味的なスキーマは推理オントロジーで定義する。

顕在事象生成処理とは、文書や既存のナレッジグラフから顕在事象に関するナレッジグラフ (事象インスタンス・事象実体意味インスタンス・実体インスタンス) を生成する処理である。信頼性に基づく顕在事象選定処理とは、顕在事象の中から事象の信頼性が高いものみに絞り込む処理である。

仮説事象生成処理とは、実体間に存在する制約条件の範囲内で、確率的に実体インスタンスを組み合わせて、事象のナレッジグラフ (事象インスタンス・事象実体意味インスタンス・実体インスタンス) を生成する処理である。

事象推理処理とは、顕在事象・潜在事象または仮説事象から新たに潜在事象を生成する処理である。

不正根拠抽出処理とは、顕在事象または潜在事象から不正事象を抽出し、不正事象と不正事象の根拠となる事象から不正根拠を生成する処理である。

本提案を犯罪推理への応用として推理オントロジー (2.3)・事象推理処理 (2.4)・不正根拠抽出 (2.5) 3点を中心に具体化した。

2.3 推理オントロジー

推理オントロジーとは推理知識に関する概念を体系的に表したものであり、事象クラス・事象実体意味クラス・実体クラス・事象推理クラスや事象推理に関するクラスを定義する。実体クラスとは Player(人や集団)・物体クラス・空間クラス・時間クラス・属性クラスを下位概念とする。事象クラスは各種実体を主体としたクラスや不正事象クラスを下位概念とする。事象実体意味クラスは各種実体に対し事象における意味関係クラスを下位概念とする。事象推理に関するクラスとして、事象推理クラスと事象推理条件クラスを下位概念とする。

事象のコンテキストや事象推理における意味的關係はクラス間のリンクで定義し、各種インスタンス生成時におけるインスタンス間のリンク生成に用いられる。推理オントロジーを具体化したものについて、3.1 で述べる。

2.4 事象推理

事象推理では、推理オントロジーに基づき複数の事象のナレッジグラフを入力とし、事象推理インスタンス・事象推理条件インスタンスを中間データとして経由し、事象コンテキストを出力とする。入出力の事象のナレッジグラフは同じ実体インスタンスを含む。

推理オントロジーにおける事象推理の概念例を Fig.2 に示す。事象推理クラス P は事象クラス $I_1 \dots I_m$ を入力であることを意味するリンクをもち、事象クラス O を出力であることを意味するリンクをもつ。事象推理条件 $C_1 \dots C_n$ は入出力事象間の事象実体意味クラスの対応関係を示す。例えば、事象推理条件 C_n は事象クラス I_1 の事象実体意味クラス IR_{1k} と事象クラス I_m の事象実体意味 IR_{ms} と事象 O の事象実体意味 OR_n が対応関係 (同じ実体にリンクする) にあることを意味する。

事象推理は以下の STEP から成り立つ。

STEP1 入力事象インスタンス集合 (候補) の検出

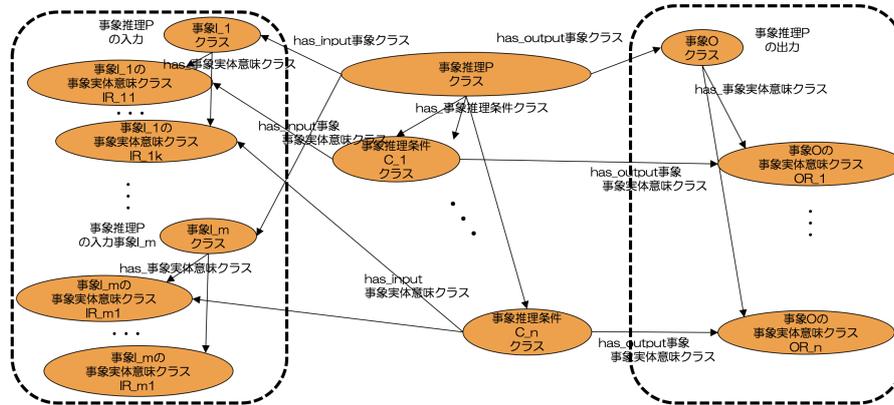


Fig. 2. 推理オントロジーにおける事象推理の概念例

STEP2 推理条件インスタンスの生成

STEP3 入力事象インスタンス集合の確定

STEP4 推理インスタンスの生成

STEP5 出力事象インスタンス・出力事象実体意味インスタンスの生成

事象推理処理の手順を Fig.2 の事象推理クラス P を用いて説明する。(具体例は 3.3 にて述べる).

STEP1 では, I_1, \dots, I_m に属するインスタンスを収集し, インスタンス集合を生成する. 1つの事象クラスに対し, 複数のインスタンスが存在する場合があるため, 事象インスタンス集合が候補として複数存在する場合がある.

STEP2 では, STEP1 の入力事象にリンクした事象実体意味インスタンスが推理条件クラスの条件を満たすとき, 推理条件インスタンスを生成する.

STEP3 では, STEP2 の推理条件インスタンスが $C_1 \dots C_n$ の全てのクラスで生成されたとき, 入力事象インスタンス集合として確定する

STEP4 では, 推理インスタンスを生成し入力事象インスタンス集合と推理条件インスタンス集合とのリンクを生成する.

STEP5 では, 推理インスタンス及び推理条件インスタンスにそれぞれ対応した出力事象インスタンス及び出力事象実体意味インスタンスを生成する. 推理条件インスタンスを経由して対応した入力事象と出力事象の事象実体意味インスタンスが同じ実体インスタンスへのリンクをもつように出力事象の事象実体意味インスタンスと実体インスタンスとの間にリンクを生成する.

事象推理手順を SPARQL によって実装し, ナレッジグラフを更新する. Fig.1 の事象推理処理とは, 推理オントロジーで定義した全ての事象推理概念に対して, 事象推理を実施することである.

2.5 事象概念に基づく不正根拠抽出

事象推理処理による不正根拠抽出の概念図 Fig.3 を用いて, 顕在事象から不正根拠を抽出するまでの流れを説明する.

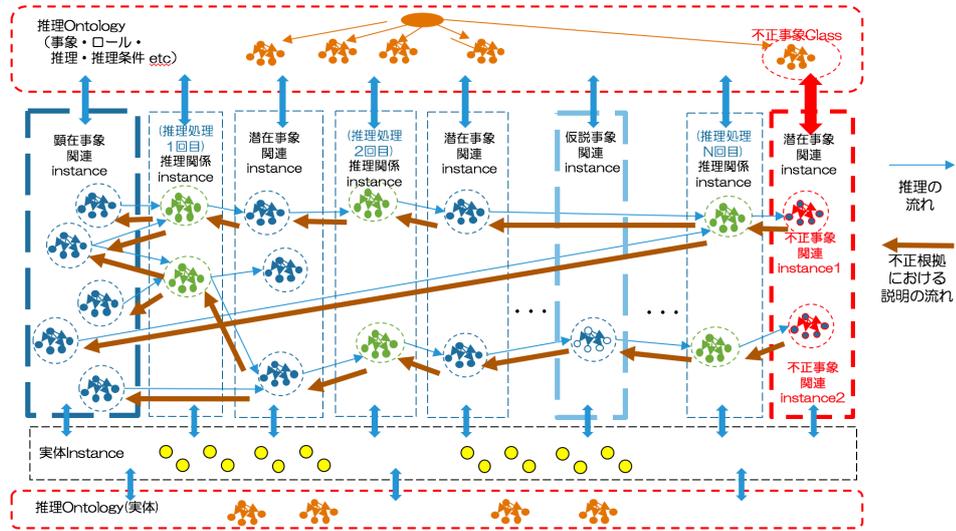


Fig. 3. 事象推理処理による不正根拠抽出の概念図

Fig.3 では、顕在事象 (左端) から事象推理処理を N 回繰り返した結果生成された潜在事象に不正事象 (右端) が含まれていることを示している。顕在事象関連インスタンスから不正事象に関するインスタンスまでの間には、推理の流れ (Fig.3 の右方向青色矢印) に基づき生成されたインスタンスやリンクが存在する。不正根拠が論理的なものであるためには、不正事象から顕在事象までの各事象の内容と事象の推理根拠を段階的に説明する必要がある。この推理の方向とは逆方向 (Fig.3 の左方向茶色矢印) の順序に基づきインスタンスやリンク及び周辺のクラスのリテラル情報を収集することにより各事象の内容と事象の推理根拠を構造的に記述した JSON ファイルを生成する。JSON の階層構造の概要を以下に示す。

事象 不正事象 1 の内容

推理根拠 事象 1A と事象 1B から不正事象 1 への推理説明

事象 1A 事象 1A の内容

... 事象 1A から顕在事象までの事象及び推理根拠を記述

事象 1B 事象 1B の内容

... 事象 1B から顕在事象までの事象及び推理根拠を記述

事象 不正事象 2 の内容

推理根拠 事象 2A と事象 2B と事象 2C から不正事象 2 への推理説明

事象 2A 事象 2A の内容

... 事象 2A から顕在事象までの事象及び推理根拠を記述

事象 2B 事象 2B の内容

... 事象 2B から顕在事象までの事象及び推理根拠を記述

6 事象概念を用いた推理知識表現方法の提案

事象 2C 事象 2C の内容

... 事象 2C から顕在事象までの事象及び推理根拠を記述

以上のようにある事象の情報はその事象の推理根拠を内部情報としてもち、推理根拠では、推理の入力となった事象をさらに内部情報としてもつ入れ子構造で格納する。

3 ナレッジグラフ推論チャレンジへの応用

3.1 推理オントロジーの具体化

推理小説等で発生するイベントや状態を想定し、2.3 で述べた推理オントロジーの下位概念を拡張した。

推理オントロジーの一部を Fig.4 に示す。



Fig. 4. 推理オントロジー

死傷や消息不明、紛失等を下位概念とする被害事象クラスや不正事象の下位クラスとして犯行事象クラスを追加した。

犯行事象クラス『PHN_11010_加害者が被害者に殺傷能力を使って害を与える』では加害者・被害者・犯行時刻・犯行現場の事象実体意味クラスとのリンク (has_事象実体意味 Class) により犯行事象のコンテキストを表現するのに必要な概念間の関係を定義している。

事象推理の入出力事象はできるだけ上位の概念にて定義することにより本チャレンジ以外の犯罪案件にも対応できることを目指す。

3.2 提供データの事象表現

提供データ (Speckled Band.xlsx) の ID = 0 ~ 401 のデータを推理オントロジーに基づくナレッジグラフで表す。ナレッジグラフの例として, ID = 85 「事件当夜, 村の医師がジュリアを診た」 ID = 86 「事件当夜, ジュリアは死んだ」のナレッジグラフを Fig.5 に示す。

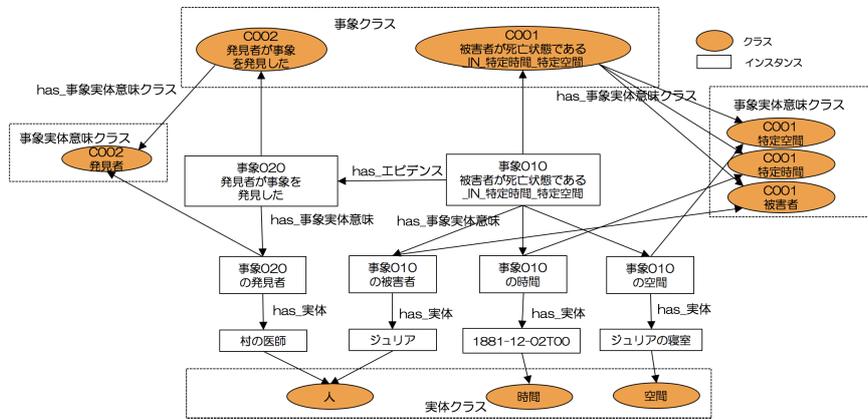


Fig. 5. Speckled Band データのナレッジグラフ例 (ID=85, 86)

ID=85 のデータは推理オントロジーの C001 クラス及び C001 の事象実体意味クラスにより, 事象 010 インスタンス, 事象 010 の事象実体意味である被害者・時間・空間のインスタンス, 実体インスタンスであるジュリア・1881-12-02T00・ジュリアの寝室インスタンスによって表現することができる。ID = 86 のデータも同様に表現することができる。ID = 86 のデータは ID = 85 のデータの根拠であることから, 事象間のリンク (has_エビデンス) により表現する。

3.3 事象推論例

事例推理のナレッジグラフ例を Fig.6 を示す。推理オントロジーでは, 推理クラス「P005」は「C002 犯人が被害者を殺害する」と「C003 主人と従者が主従関係にある」を入力事象とし「C004 教唆者の指示により非教唆者が被害者を殺害する」を出力事象とする。P005 は出力事象の生成条件となる入力となる事象の事象実体

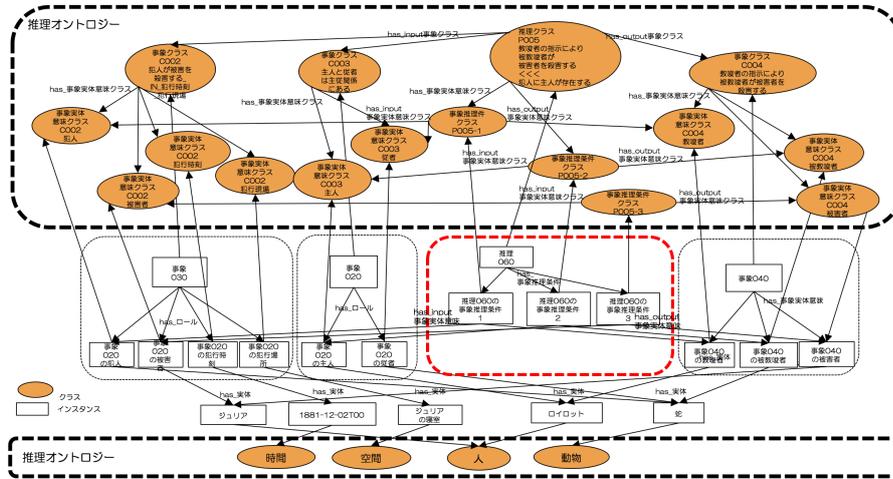


Fig. 6. 事例推理のナレッジグラフ例

意味間の関係と入出力事象間の事象実意味対応関係として以下のように定義されている。

2.4の事象推理手順に基づき、P005に関する事象推理は以下のように行われる。

STEP1 入力事象インスタンス集合 (候補) の検出

事象 020 インスタンス 事象 030 インスタンスを検出する。

STEP2 推理条件インスタンスの生成

事象推理条件クラス P005-1 ~ 3に基づき事象推理条件インスタンスが生成される。

STEP3 入力事象インスタンス集合の確定

事象推理条件インスタンスは、事象推理条件クラス P005-1 ~ 3全てを満たすため、事象 020 インスタンスと事象 030 インスタンスが入力事象として確定する

STEP4 推理インスタンスの生成

STEP2の3つの事象推理条件インスタンスとリンクした事象推理インスタンスを生成する。

STEP5 出力事象インスタンス・出力事象実意味インスタンスの生成

STEP2の3つの事象推理条件インスタンスに基づき、出力事象実意味インスタンスとして、事象 040の教唆者・被教唆者・被害者を作成する。事象 040の教唆者・被教唆者・被害者は事象推理条件インスタンスに対応して、実体インスタンスであるロイロット・蛇・ジュリアとのリンクを生成する。また、出力事象インスタンスとして事象 040 インスタンスを生成し、出力事象実意味インスタンスとのリンクを生成する。

これらの手順は下記3点を意味する。

- 1 C002の犯人とC003の従者の実体が一致する。そのとき、その実体はC004の被教唆者に相当する。
- 2 C003の主人の実体が存在する。そのとき、その実体はC004の教唆者に相当する。
- 3 C002の被害者の実体が存在する。そのとき、その実体はC004の被害者に相当する。

よって、これらのオントロジー上の関係を用いて、C002のインスタンスである事象030「蛇が1881-12-02T00に寝室でジュリアを殺害する」とC003のインスタンスである事象020「ロイロットと蛇は主従関係にある」からC004のインスタンスである事象040「ロイロットの指示により蛇がジュリアを殺害する」を推理することができる。

3.4 仮説事象生成

2.2で述べた仮説事象を生成するためには、事象実体意味インスタンスとして生成する実体の組み合わせを事実上存在する可能性の範囲で生成しなければならない。そこで、事実上成立する可能性がある時間・空間・Player(人・集団)の組み合わせ(アリバイ)を生成するために時空間という概念を定義した。

空間は時間により状態が変化するため、人物がそこに滞在することがありえない状況が存在する。そこで時間と空間により制約されたインスタンスとその関係によりアリバイを検討する。Fig.??に時空間概念によるアリバイ検討例を示す。ジュリアの寝室へのドアがロックされているかどうかにより時空間インスタンスT1_ドア経路2は存在しないが時空間インスタンスT2_ドア経路2は存在することを示している。これにより、ジュリアとロイロットが存在しうる時空間(青色点線のリンク)に異なる。これらのリンクにより生成された時空間のナレッジグラフ(時間・空間・Player・物体)から仮説事象を生成する。

3.5 犯行根拠抽出

犯行事象から死亡事象を含む顕在事象までの事象・事象実体意味・実体のクラスやインスタンス及びリンクを抽出し、犯行根拠のJSONファイルを生成する。

Fig.6に相当する犯行根拠の部分を図.7を示す。事象040の事象keyに対してC004の概要(教唆者の指示により非教唆者が被害者を殺害する)をvalueとする事象040の教唆者key・被教唆者key・被害者keyに対して、事象040の各事象実体意味インスタンスにあるロイロット・蛇・ジュリアをvalueとする。事象040の推理根拠keyのvalueとして事象040を出力とする事象推理060の情報が記載する。さらに推理プロセス060の根拠事象keyのvalueとして推理プロセス060の入力事象である事象020、事象030の情報を記載する。事象040と同様、事象020・事象030も推理根拠keyを所有し、内部に各事象の根拠となる情報を段階的に記載する。

これらの事象や事象推理のJSON情報は、ナレッジグラフにおける事象インスタンスや事象推理インスタンスの周辺のクラスやインスタンスのリテラルから収集する。Fig.7の事象040と事象推理060に関するナレッジグラフとJSON情報の対応関係例を図.8を示す。事象040のJSON情報は事象040インスタンスや事象040の事象実体意味インスタンスではなくリンクしたクラスや実体インスタンスから収集するため、JSON情報を自動的に生成することが可能になる。

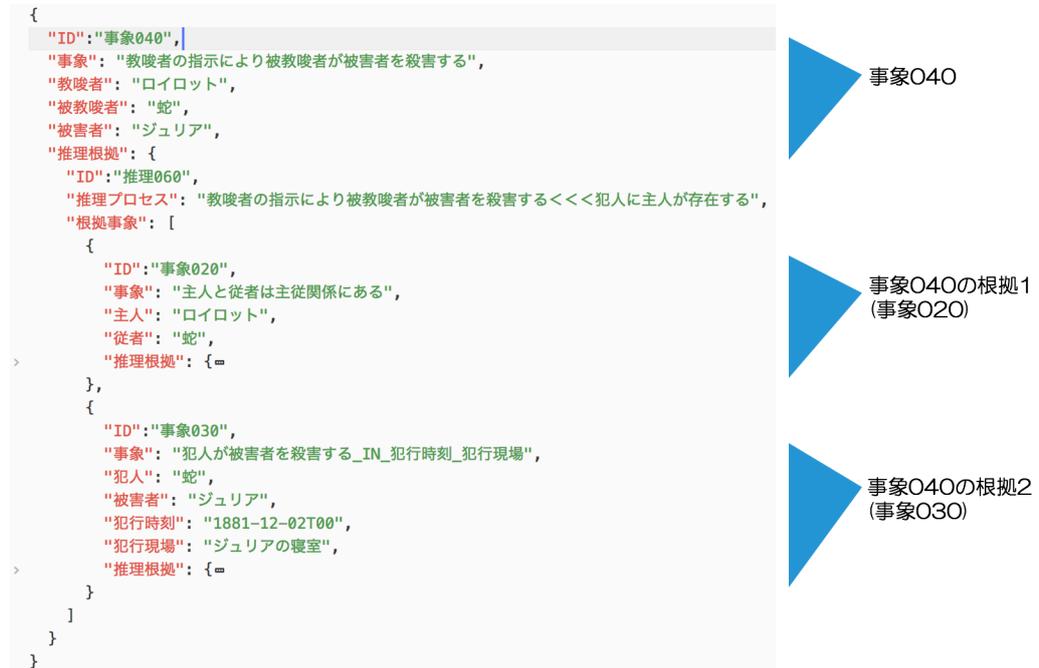


Fig. 7. 犯行根拠例

4 まとめ

本論文では、事象概念を用いた知識表現手法を具体化した推理オントロジーを用いて、不正事象を検出し、その根拠を機械的に説明する仕組みを提案した。推理オントロジーを拡張し、ナレッジ推論チャレンジのデータを用いることにより犯行事象の抽出及び犯行根拠の説明にも対応できることがわかった。

今後は本チャレンジデータを活用し、推理オントロジーを拡張することにより実用性を高めていく予定である。

References

1. Kazunari Hashimoto, Motoyuki Takaai Seiji Suzuki, Yohei Yamane, Masao Watanabe, and Hiroshi Umemoto. An Ontology-based Validation Approach to Resolve Conflicts in the Manufacturing Design Process, the 4th Workshop on Linked Data Quality (LDQ 2017)
2. Protege : <http://protege.stanford.edu/>

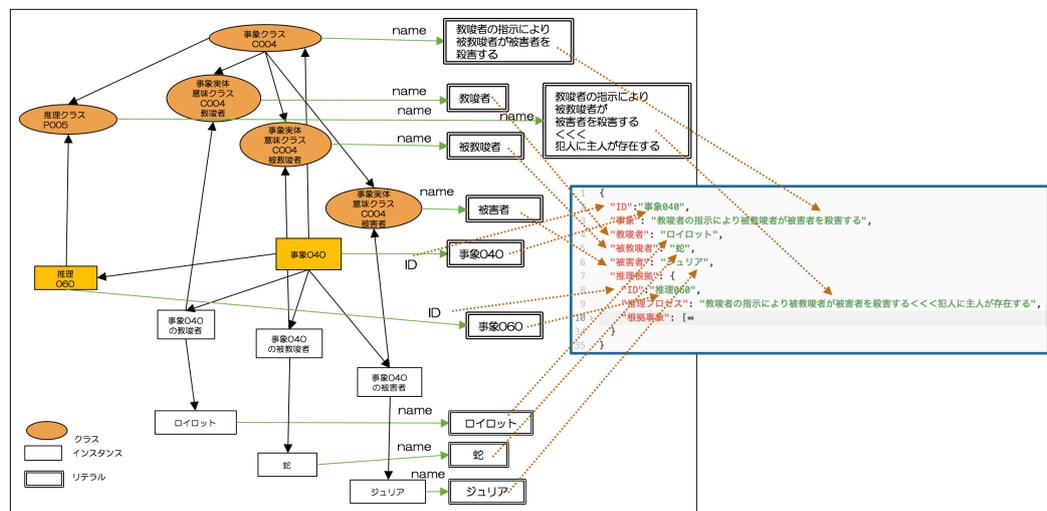


Fig. 8. ナレッジグラフと犯行根拠の対応関係例