

知識の機械的理解・活用 に向けて

知識データ工学研究室

教授 三輪誠

井田龍希 (D1), 山田晃士 (D1)

略歴

～2008年 東京大学大学院 新領域創成科学研究科 基盤情報学
専攻 博士後期課程 博士 (科学)

2008年～2011年 東京大学大学院 情報理工学系研究科 コン
ピュータ科学専攻 辻井研究室 特任研究員

2011年～2014年 マンチェスター大学 Research Associate

2014年～2023年 豊田工業大学准教授

2015年～ 産業技術総合研究所人工知能研究センター 招聘研究員

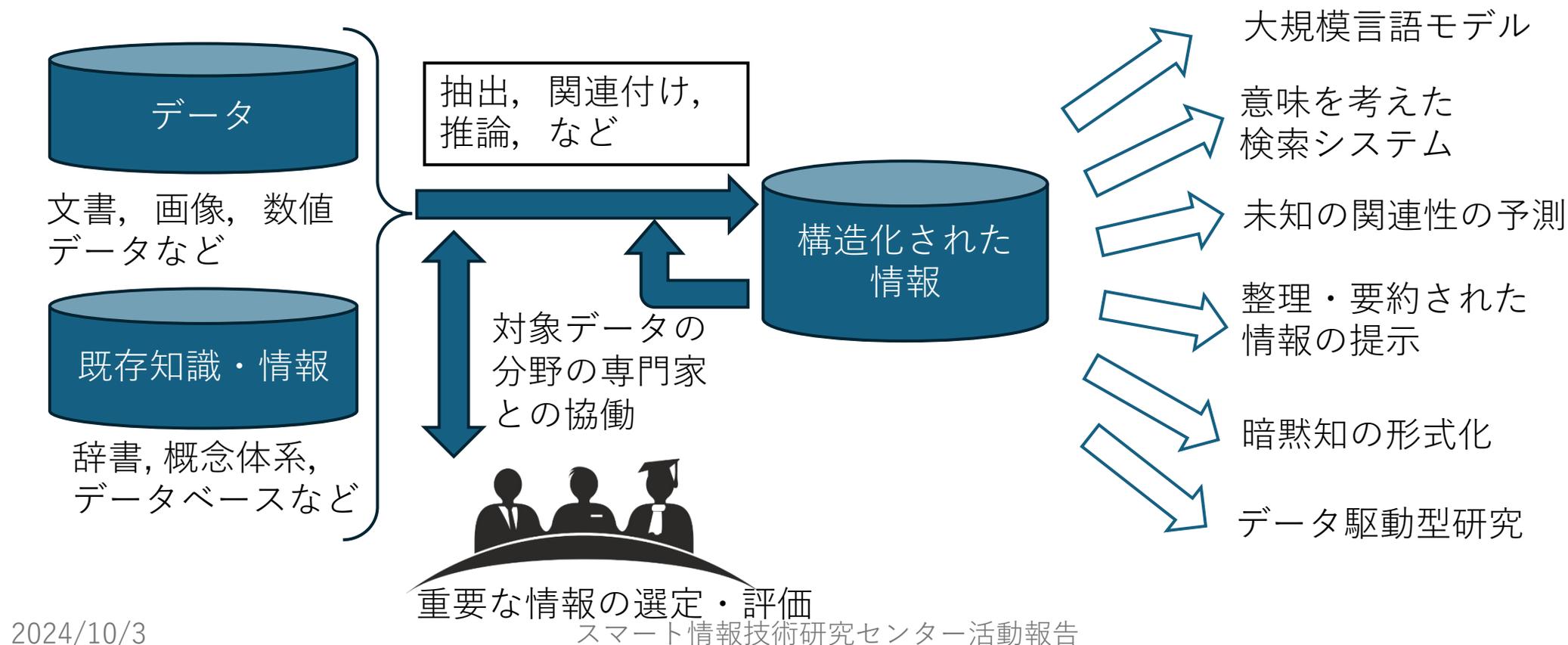
2024年～ 豊田工業大学教授

研究分野・テーマ

- 研究分野
 - 自然言語処理
 - 深層学習・表現学習
 - バイオメディカルインフォマティクス
- 研究テーマ
 - 情報抽出・知識発見
 - **探索と機械学習・深層学習**による効率的な重要情報の発見
 - 知識利用
 - 既存知識や情報の**統合的な利用**による対象問題の知識に基づいた解析

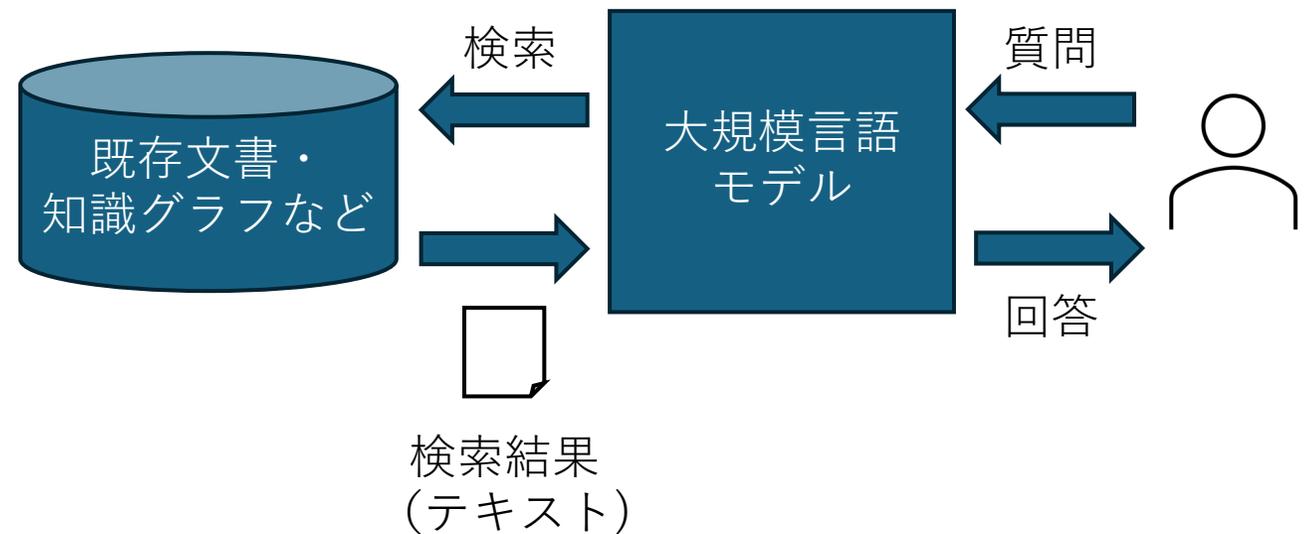
研究概要

データに埋もれた経験や情報を抽出し，知識と関連付け，理解し，利用する，表現・抽出・関連付け・推論・利用方法の研究



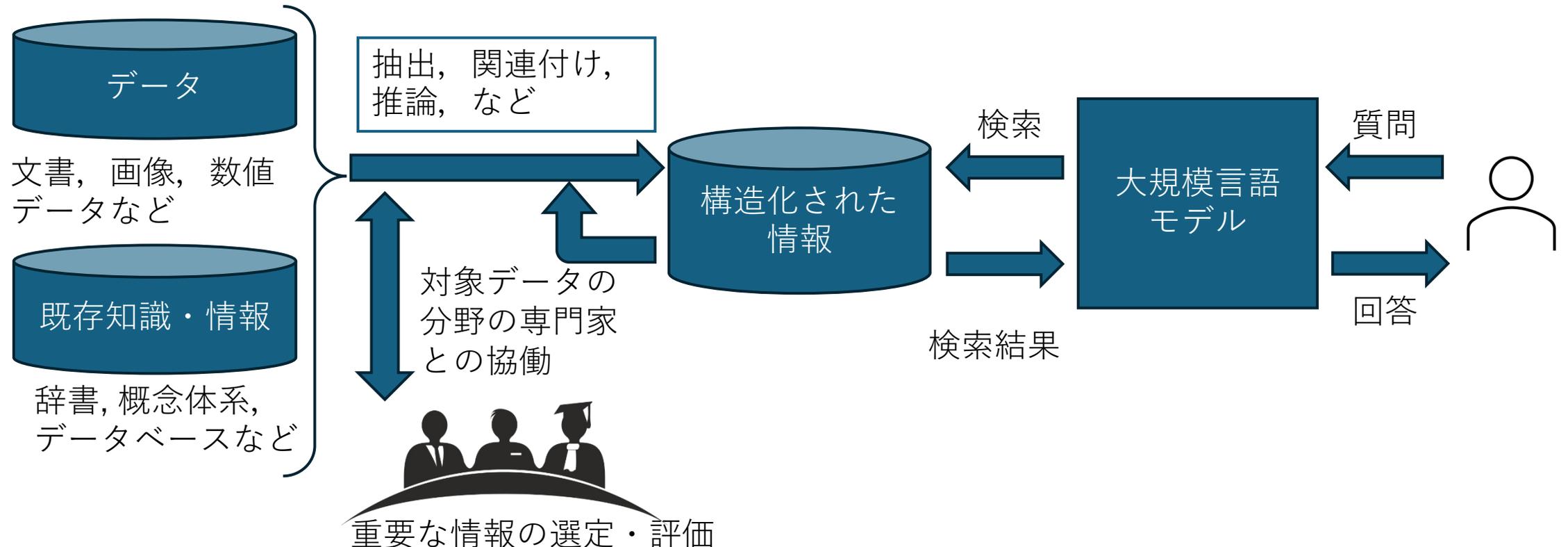
大規模言語モデルの外部情報利用例

- 検索拡張生成 (RAG; Retrieval Augmented Generation) による検索結果を利用した生成
 - 外部情報や文書を検索して、結果をその場で使う
 - 文書・概念の背景にある情報（数値・画像・分野データ）などは利用されない



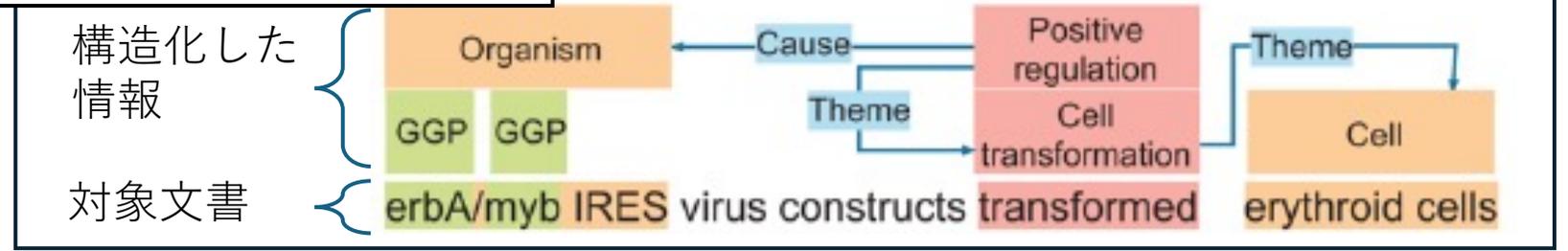
目標とする知識利用

- 抽出・関連付け・推論して構造化された情報を利用



研究課題

テキストからの情報抽出



• テキストからの情報抽出

- テキストに記述された情報をどのように構造化して抽出するか？

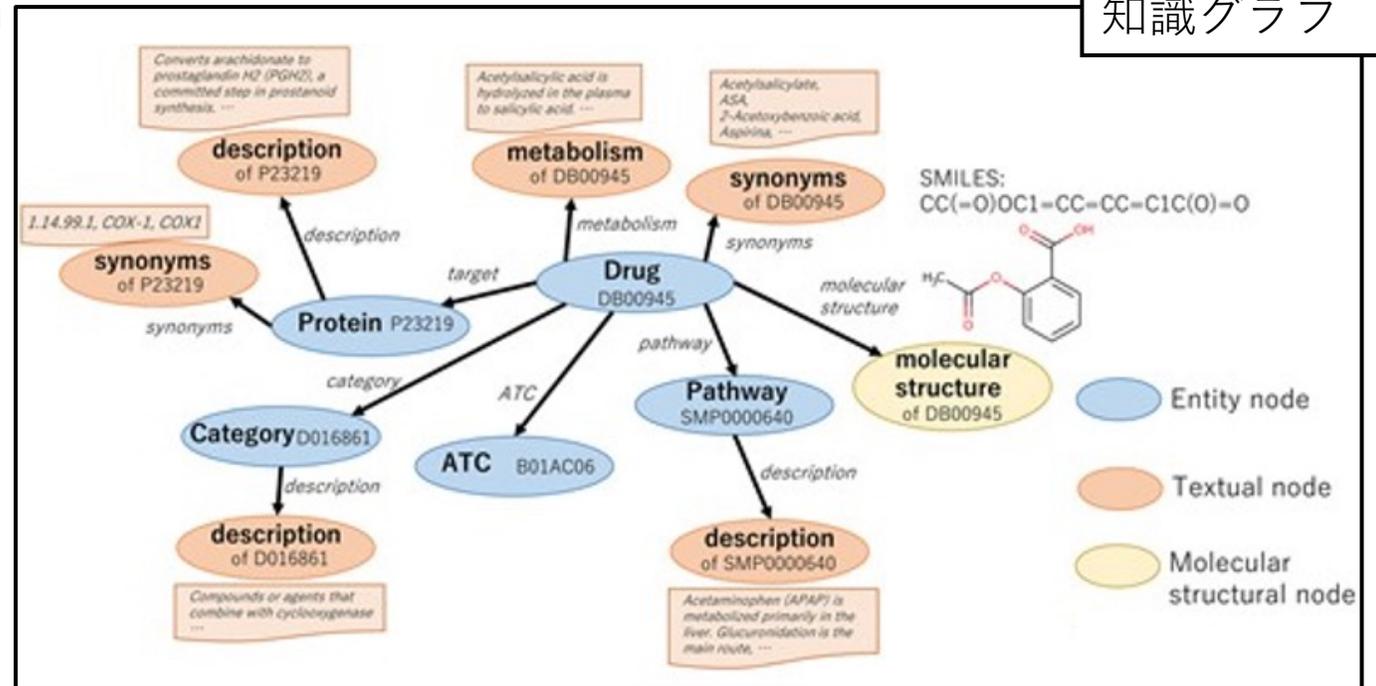
➔ 用語・関係・イベント抽出

• 知識グラフ上でのマルチモーダル表現学習

- 概念, テキスト, 画像, 分野情報 (化学構造, 数値など) をどう関連づけ表現するか？

➔ グラフ表現学習

知識グラフ



研究事例の紹介

- (1) 複数質問を用いた関係抽出
- (2) 一対多関係を用いたグラフ表現学習

研究事例の紹介

- (1) 複数質問を用いた関係抽出**
- (2) 一対多関係を用いたグラフ表現学習

研究事例 (1):複数質問を用いた関係抽出

関係抽出とは

- 関係抽出:テキストから対象とする用語間の関係を汎化して抽出

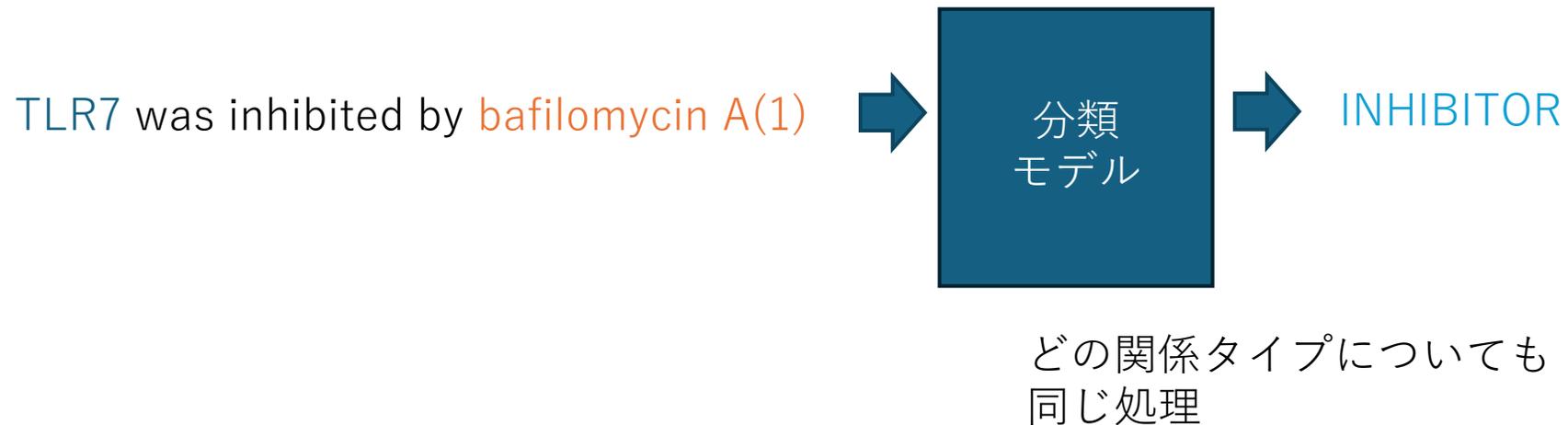
タンパク質-薬物間相互作用の例



- "was inhibited by"・"was deactivated by"などの表現や、能動態・受動態などの構文によらず、同じ意味の関係であれば、同じ関係タイプとして抽出
- ➔ テキストのエビデンスが見ついた構造化された情報を整備するための基盤技術

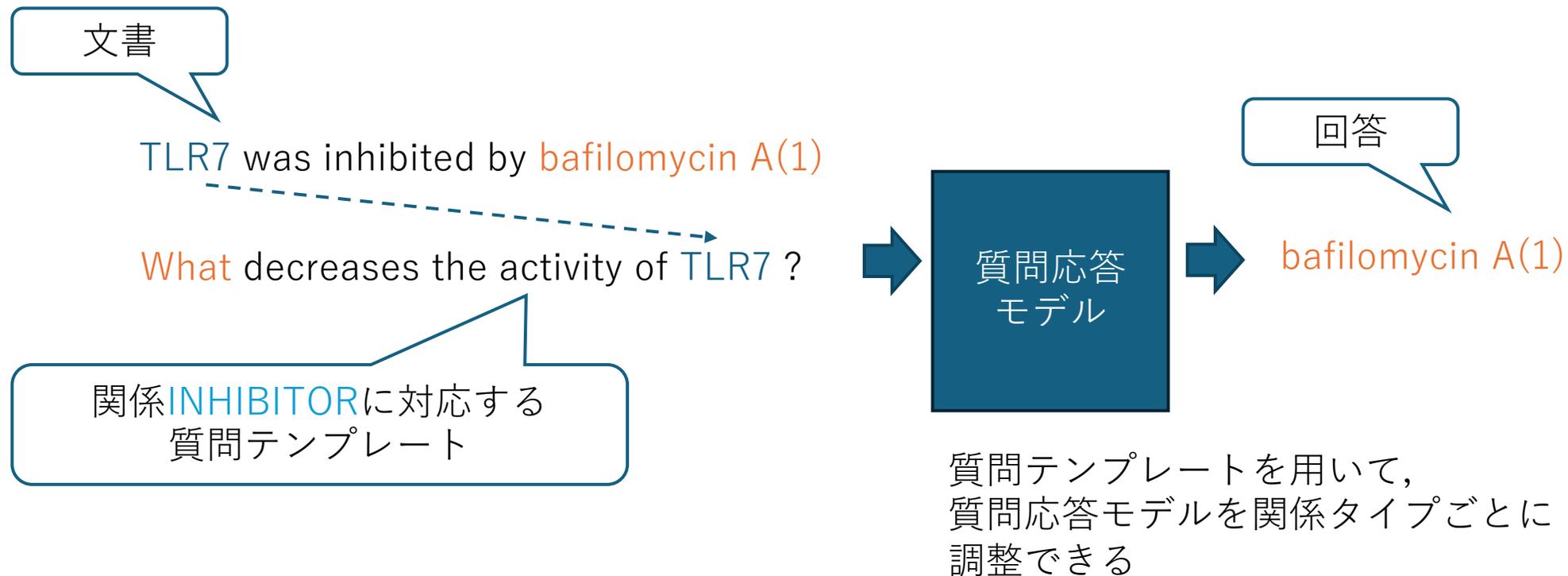
研究事例 (1):複数質問を用いた関係抽出 従来の分類による関係抽出

- 文書と用語を入力し，関係タイプに分類



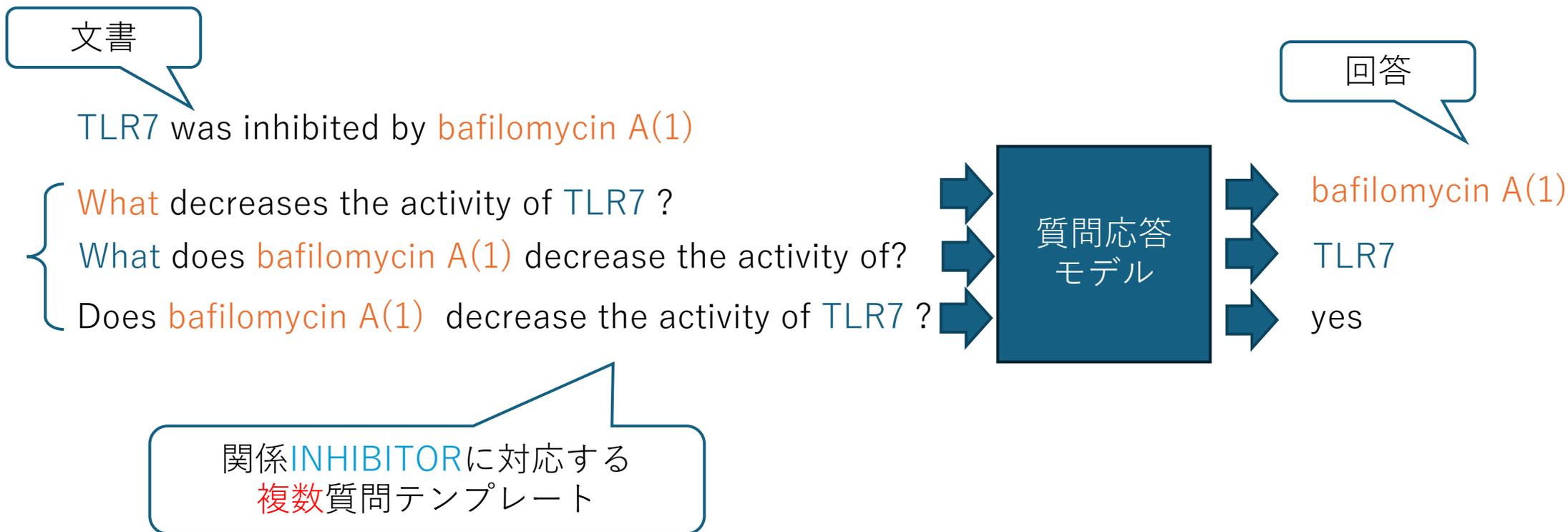
研究事例 (1):複数質問を用いた関係抽出 質問応答を用いた関係抽出

- 関係タイプに対応する質問に回答できたら、関係があると判定



研究事例 (1):複数質問を用いた関係抽出 複数質問を用いた関係抽出

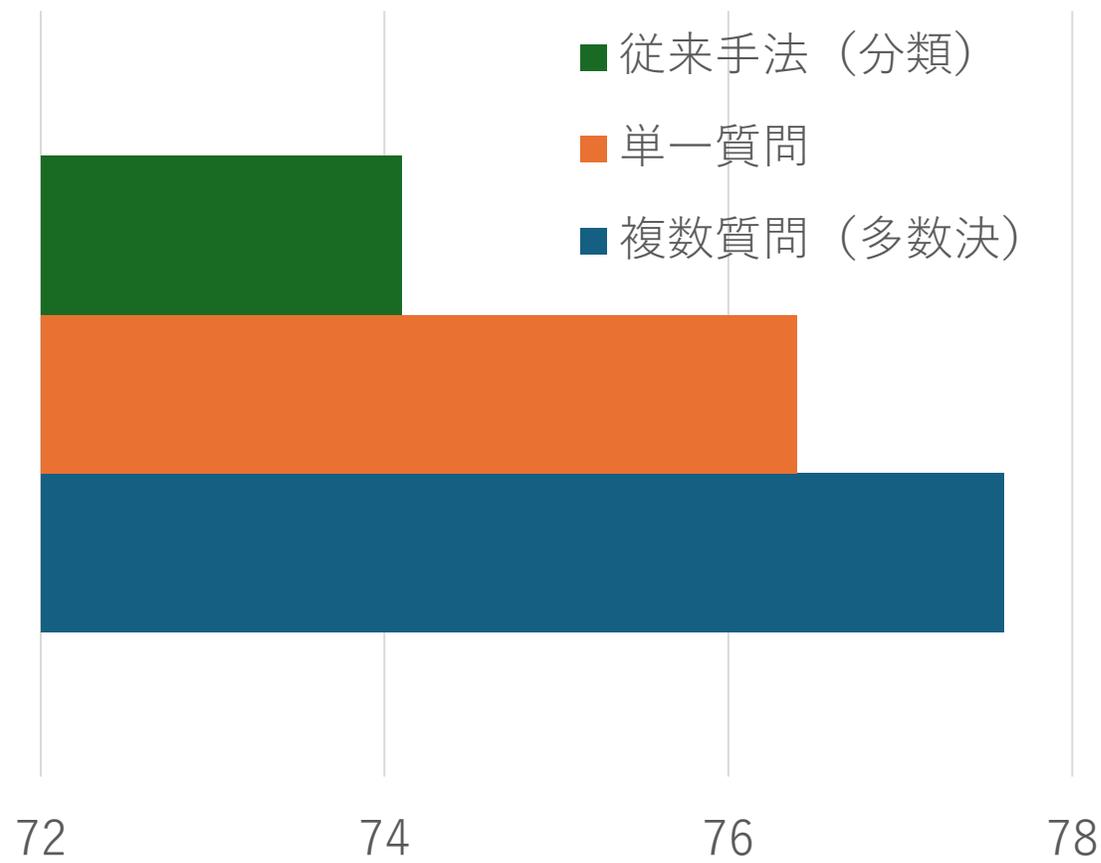
- 関係タイプに対応する複数質問の回答を元に関係を判定



研究事例 (1):複数質問を用いた関係抽出

複数質問を用いた関係抽出の評価

- 質問応答・複数質問の利用により，性能向上
→ 質問の追加・利用，回答の統合におけるさらなる工夫の可能性
- 本研究ではBERTを利用した質問応答モデルを用いたが，今後はLLMでの利用を検討



DrugProtコーパスで評価 (F値)

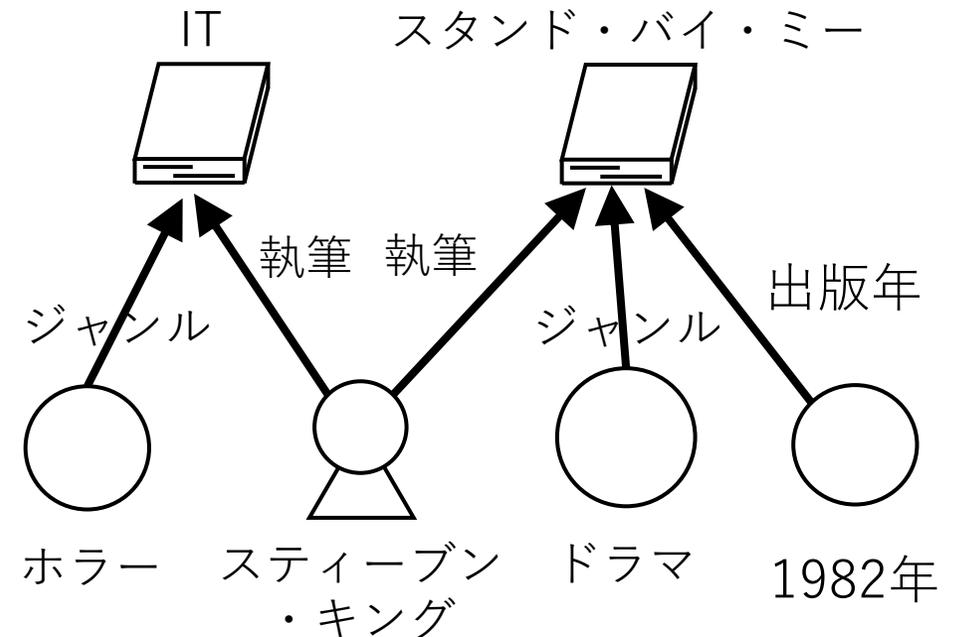
研究事例の紹介

(1) 複数質問を用いた関係抽出

(2) 一対多関係を用いたグラフ表現学習

研究事例 (2): 一対多関係を用いたグラフ表現学習 知識グラフ

- 知識グラフ: 知識を二項間関係 (トリプル) の集合で表現
 - (スティーブン・キング, 執筆, IT)
 - (スティーブン・キング, 執筆, スタンド・バイ・ミー)
 - …
- トリプルストアやRDF, SPARQLなど大量の知識を効率的に管理・利用する技術が発展



研究事例 (2): 一対多関係を用いたグラフ表現学習

グラフ表現学習

- グラフ表現学習: 複数の二項間関係 (トリプル) をニューラルネットワークで表現・学習
 - (スティーブン・キング, 執筆, IT)

s

v

o

ニューラル
ネットワーク

高次元空間上での
位置ベクトル

o に対応する
表現

$$v_o \approx f(v_s, v_v)$$

⇔ f , 項目 v_s, v_o , 関係 v_v を調整

- 関係する項目が似た項目同士は表現が近づく → 似た項目の検索
- f によるデータにない関係を予測 → データ補完

研究事例 (2): 一対多関係を用いたグラフ表現学習 グラフ表現学習の問題

- 二項間のみだと同じ関係にある複数項目が同じになる

- $v_{IT} \approx f(v_{\text{スティーブン・キング}}, v_{\text{執筆}})$

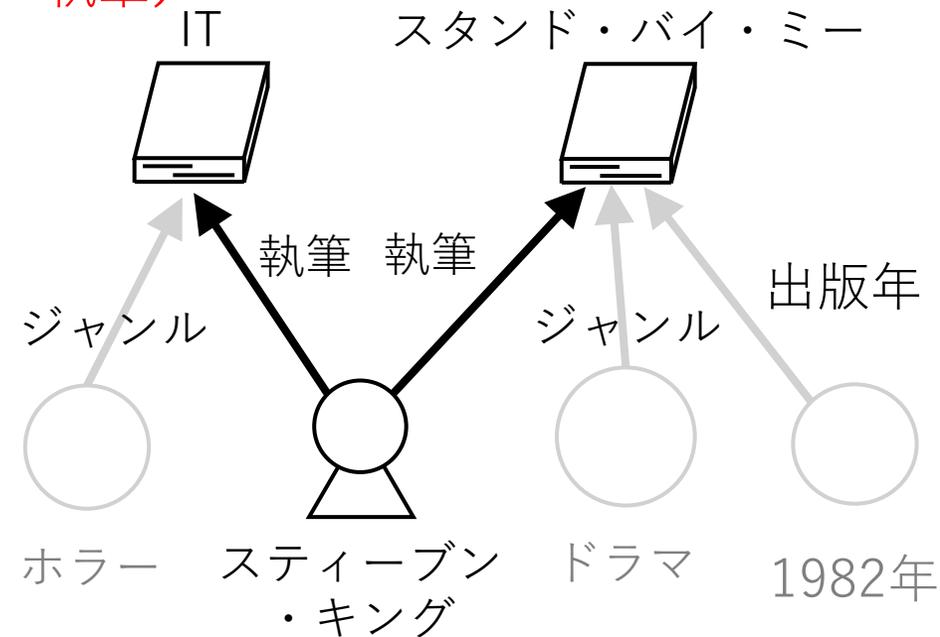
- $v_{\text{スタンド・バイ・ミー}} \approx f(v_{\text{スティーブン・キング}}, v_{\text{執筆}})$

- 他の関係も含めて学習することで違いが学習されることを期待

- $v_{IT} \approx f(v_{\text{ホラー}}, v_{\text{ジャンル}})$

- $v_{\text{スタンド・バイ・ミー}} \approx f(v_{\text{ドラマ}}, v_{\text{ジャンル}})$

➔ 上手く学習できるとは限らない



研究事例 (2): 一対多関係を用いたグラフ表現学習 一対多関係の学習

- 同じ項目に関係する複数の項目（一対多関係）を直接表現するように学習

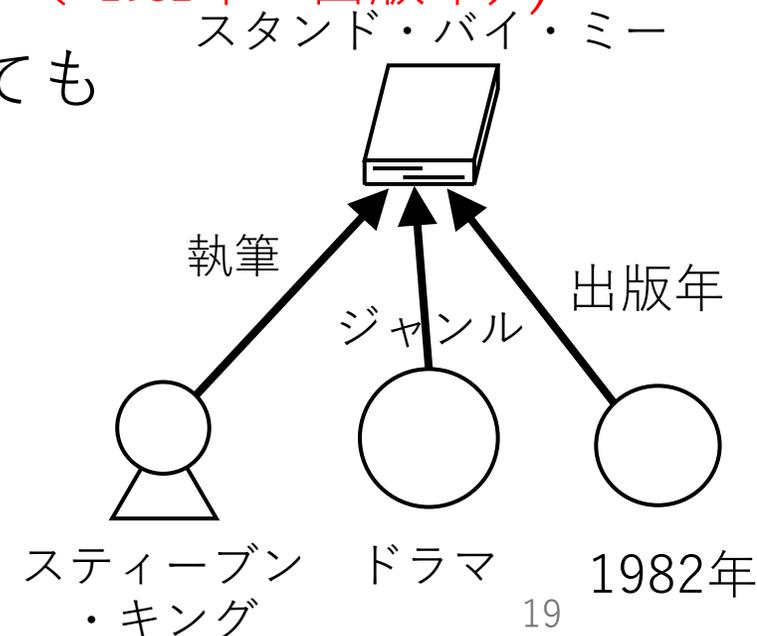
$v_{\text{スタンド・バイ・ミー}} \approx$

$f\left(\left(v_{\text{スティーブン・キング}}, v_{\text{執筆}}\right), \left(v_{\text{ドラマ}}, v_{\text{ジャンル}}\right), \left(v_{1982\text{年}}, v_{\text{出版年}}\right)\right)$

- 関係する項目が一部が欠けても，項目数が変わっても表現できるように学習

- 可能な組み合わせは膨大 (n 項目だと $2^n - 1$ の組合せ)

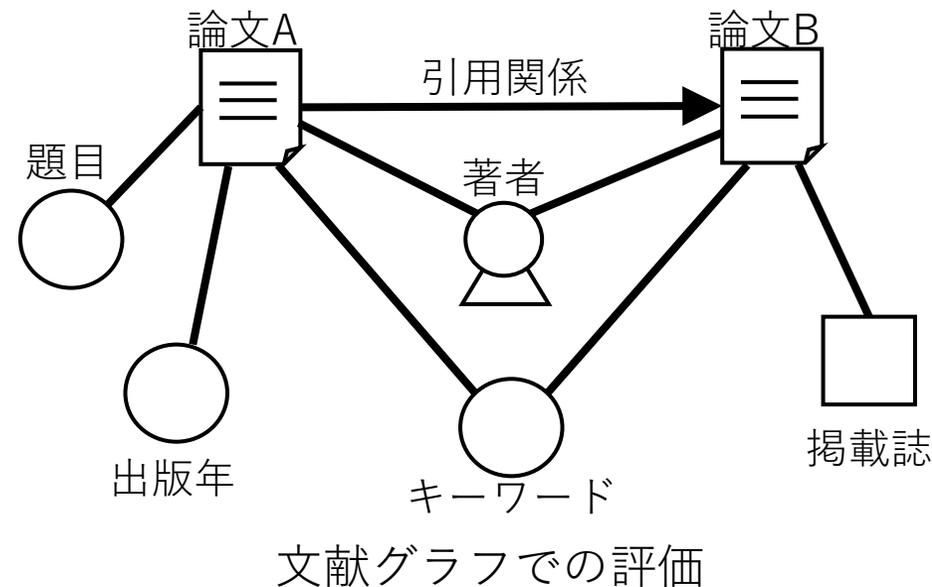
- ランダムに抽出しながら学習



研究事例 (2): 一対多関係を用いたグラフ表現学習 一対多関係の学習の評価

- 文献グラフでの学習
 - 論文はテキスト・出版年は数値など、マルチモーダルなデータ
- 一対多関係をモデル化することで、複数関係の組合せを表現
 - 二項間関係のモデル化では、二項間関係で学習して予測を平均

➔ 組合せの表現や学習における工夫，獲得した表現の利用が今後の課題



手法	1位一致率
二項間関係のモデル化	0.4143
一対多関係のモデル化	0.5770

まとめ

- データに埋もれた経験や情報を抽出し，知識と関連付け，理解し，利用する，表現・抽出・関連付け・推論・利用方法の研究
 - データ（テキスト）から情報を取り出す情報抽出
 - 知識をグラフの形で表現するグラフ表現学習
- 知識の機械的理解・活用に向けて，知識をより深く活用するにはどのような理解が必要か，機械はどのような知識の表現をしているのかを探っていく