

ナレッジグラフ推論チャレンジ2019技術勉強会

推論チャレンジ2018における FLL-MLチームの取り組みについて

2019年8月27日
岡嶋成司

■ FLL-ML

- 岡嶋成司、鵜飼孝典、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）

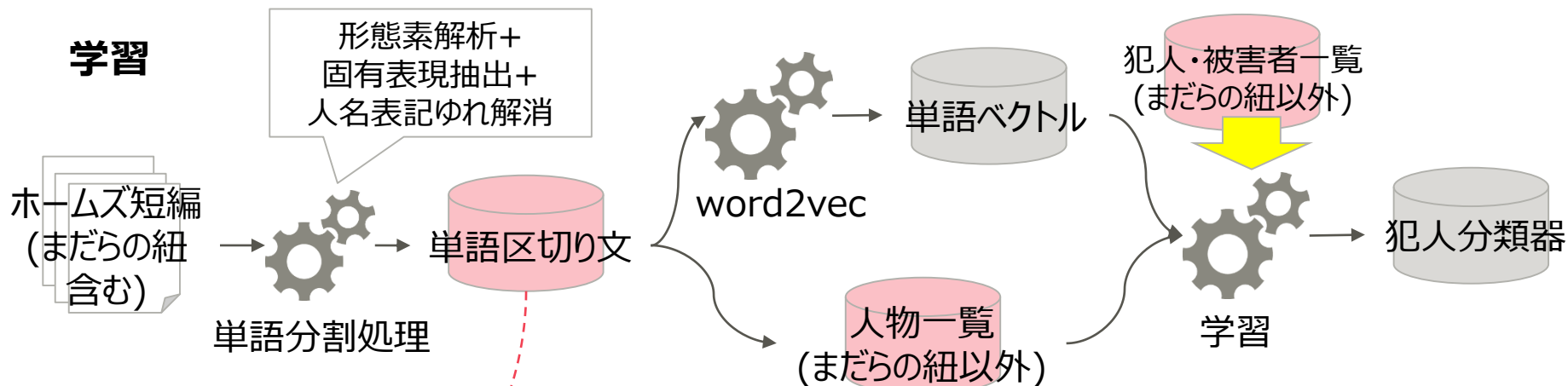
- 機械学習により、文脈から犯人の予測を行い、予測された犯人に紐づく文から動機と手段を示す文を理由として出力する

- 犯人の予測
 - 青空文庫にある「まだらの紐」以外の22のシャーロックホームズシリーズの短編の登場人物について、犯人の出現文脈を学習して予測する

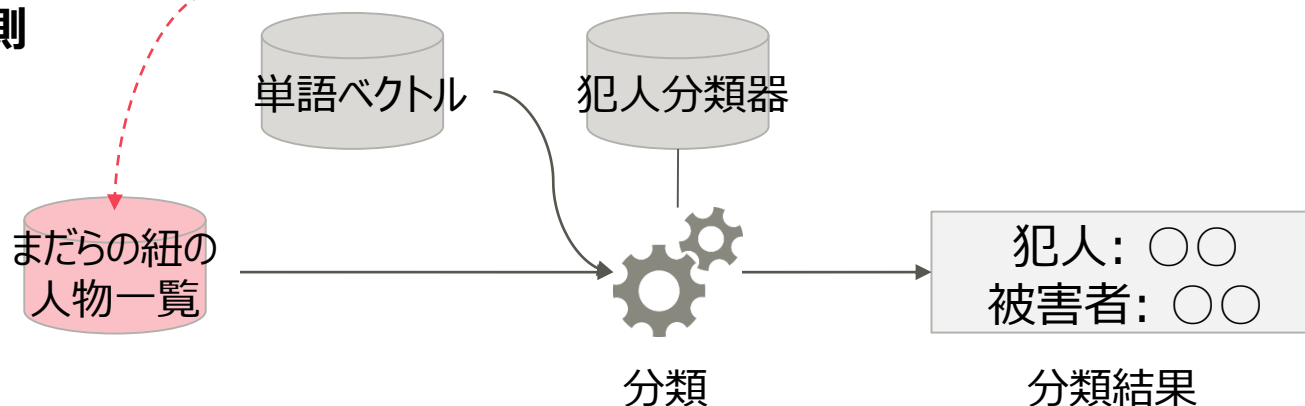
- 理由の導出
 - 犯人と意味的に近い文の中から、動機と手段に関する文を提示して理由とする
 - 動機と手段を示す文は、「まだらの紐」以外の短編の文から学習する

- ホームズ短編に出現する登場人物について、対応する単語ベクトルを特徴として「犯人」「被害者」「その他」の分類モデルを学習し、予測

学習



犯人の予測

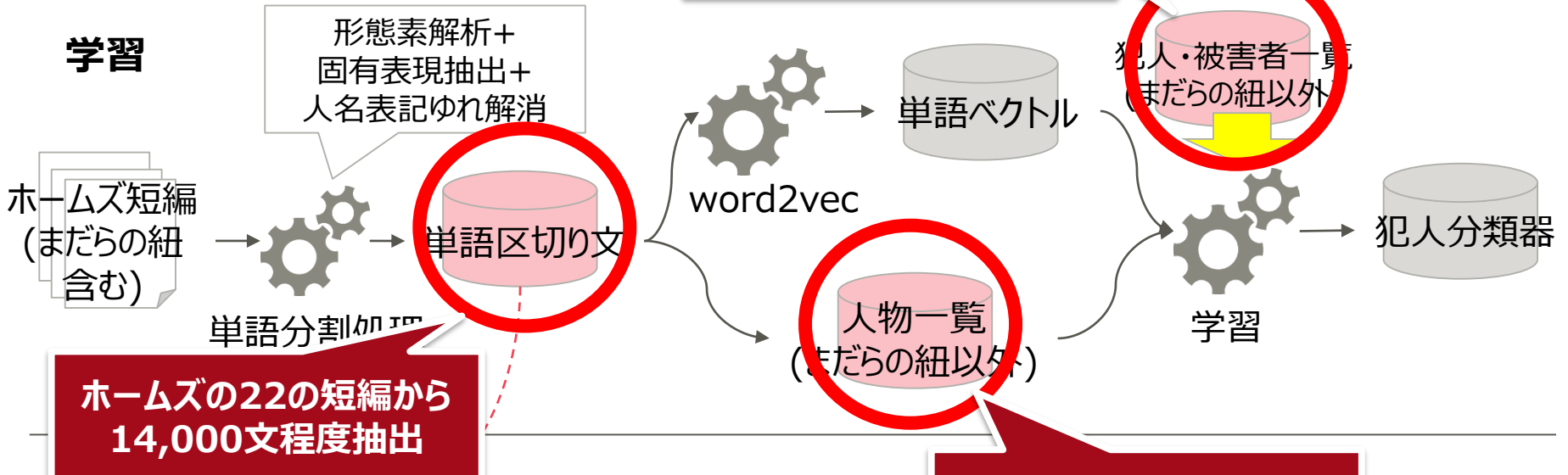


犯人の予測

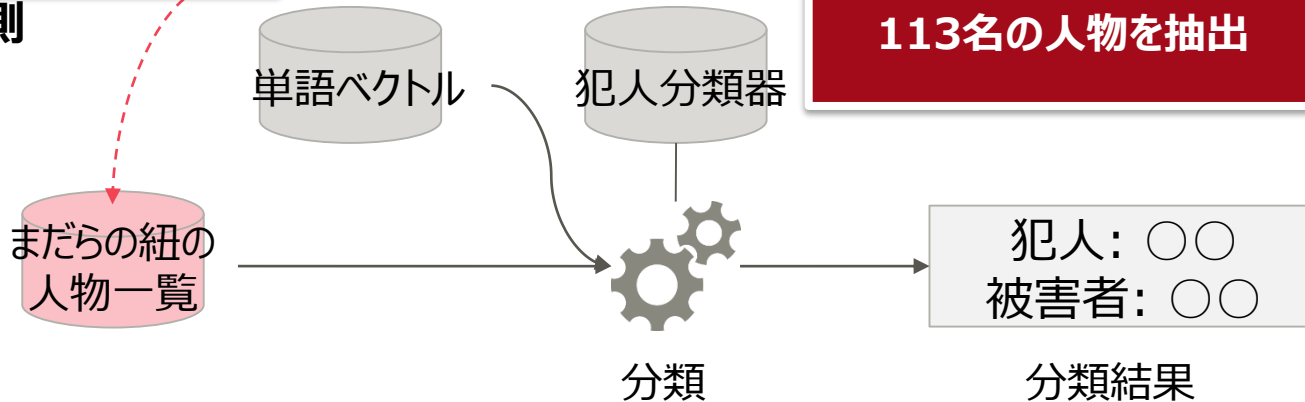
- ホームズ短編に出現する登場人物について 対応する単語ベクトルを特徴として「犯人」「被害者」を学習し、予測

ホームズ短編から
犯人・被害者を抽出

学習



犯人の予測

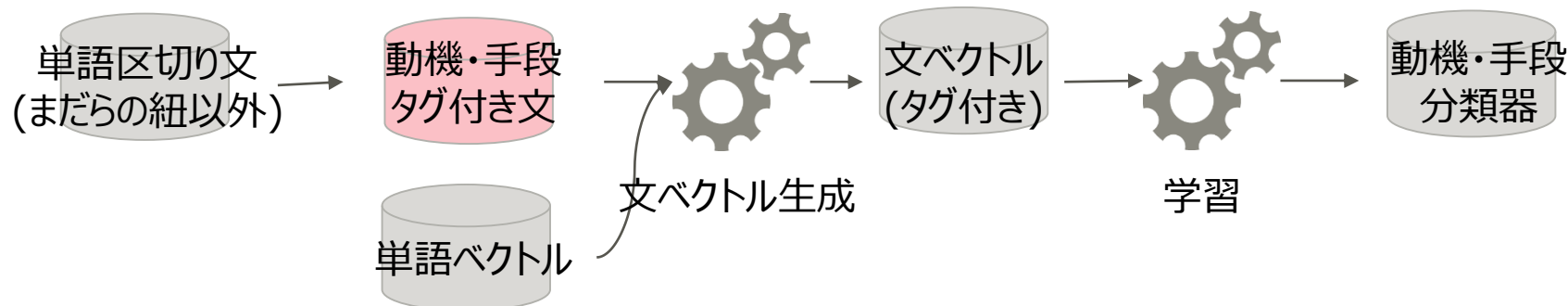


113名の人物を抽出

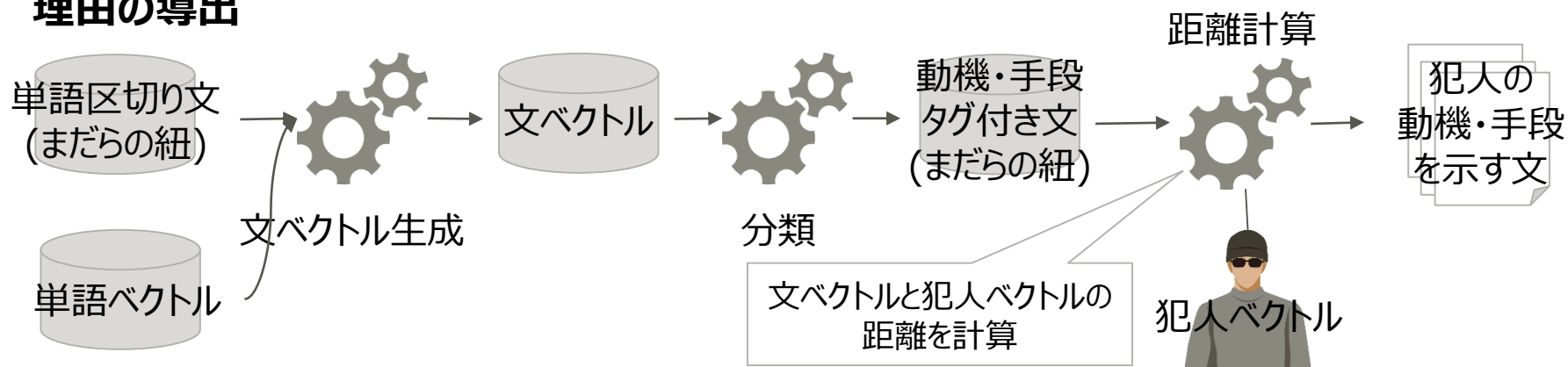
- 予測された犯人と意味的に近い文の中から、動機と手段を示している文を理由として提示する

- 動機と手段を示す文の分類器は、「まだらの紐」以外の短編の文から学習

学習



理由の導出



- 予測された犯人と意味的に近い文の中から、動機と手段を示している文を理由として提示する

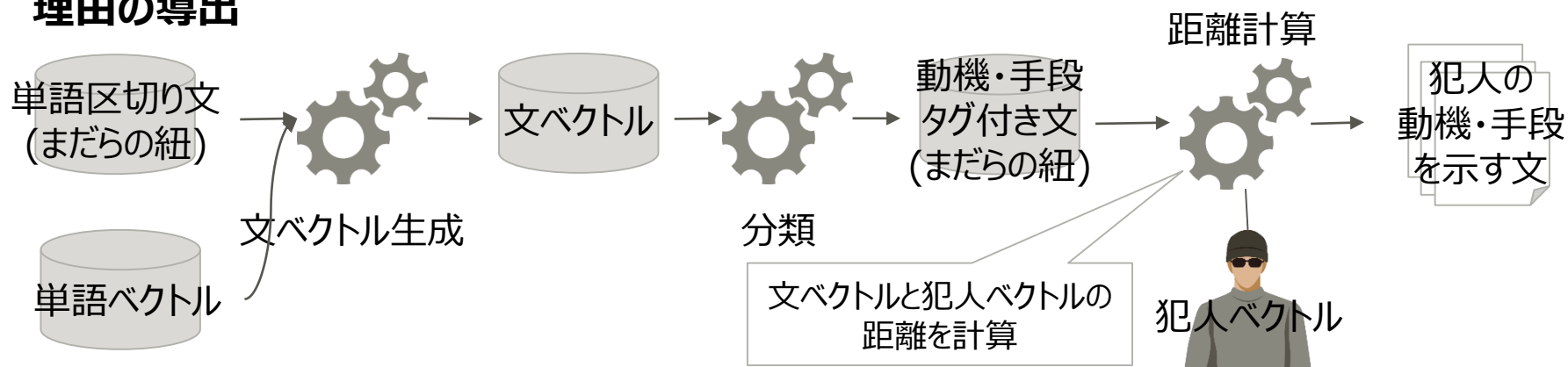
- 動機と手段を示す文の分類

目視および関連単語を利用した検索により抽出
約3,000文

以外の短編の文から学習



理由の導出



■ 「まだらの紐」の犯人予測

■ 犯人・被害者の分類スコアが最大の人物を選択

- 犯人: グリムズビ・ロイロット
- 被害者: ジュリア

■ 理由

■ 犯行の動機: 癪癢あるいは金銭問題

● 根拠

- もっとも、長い禁固刑に苦しめられたせいか、その後、内地へ戻ってからは、すれた気むずかしい人になってしまいました
- あるとき、家内で窃盗事件が立て続けに起こりまして、かんしゃくを起こし、現地で雇った執事を殴り殺してしまったそうです
- もっともそれは、わたくしども姉妹が、継父と一緒に暮らしているあいだのことだけで、もしわたくしどもが結婚いたしますと、それから後はわたくしどもにも毎年ある決まった金額の金を分けてくれる決まりになっているのでございます
- ただ数エーカーの土地と、築二〇〇年にもなる屋敷だけは残りましたが、それも抵当に入っている始末でございます
- 母はかなりの金持ちでございまして、一年に一〇〇〇ポンド以上の収入がございましたが、その金はそっくりロイロット博士に遺言で譲ってしまいました
- そして娘が結婚すると、ひとりにつき年収二五〇ポンドずつ受け取れるようになってある

■ 犯行の手段: 特定できず

- 得られた理由は、機械学習のモデルの説明ではない
 - 推定結果に近いデータを理由として提示したもの
- 機械学習の解釈性研究の分類

1. 大域的な説明	複雑なモデルを可読性の高いモデルで表現することで説明
2. 局所的な説明	特定の入力に対するブラックボックスモデルの予測の根拠を提示することで説明
3. 説明可能なモデルの設計	最初から可読性の高い解釈可能なモデルを作る
4. 深層学習モデルの説明	深層学習モデル、特に画像認識モデルの説明法。モデルが画像内のどの部分を認識しているかを特定してハイライトすることで説明

私のブックマーク「機械学習における解釈性 (Interpretability in Machine Learning)」から作成
https://www.ai-gakkai.or.jp/my-bookmark_vol33-no3/

- 今回の説明は2にあたると思われる

■ 推論との違い

- 理由として、癪癢など幅広い可能性をとらえられる ⇒ 全部ナレッジで網羅するのは難しい
- まだらの紐の犯行手段のような複雑な手続きは、単純に近いデータを示すだけでは説明できない ⇒ 複雑な手続きの説明は推論のほうが向いている

■ 問題点

- 犯人の予測精度、動機・手段文の分類精度がイマイチ
 - 考えられる要因
 - 学習に利用する文書数が少ない
 - ホームズ短編には特殊なものも多い（馬が犯人、犯人がいないなど）
 - 学習データのタグ付け基準の揺れ

■ 改善点

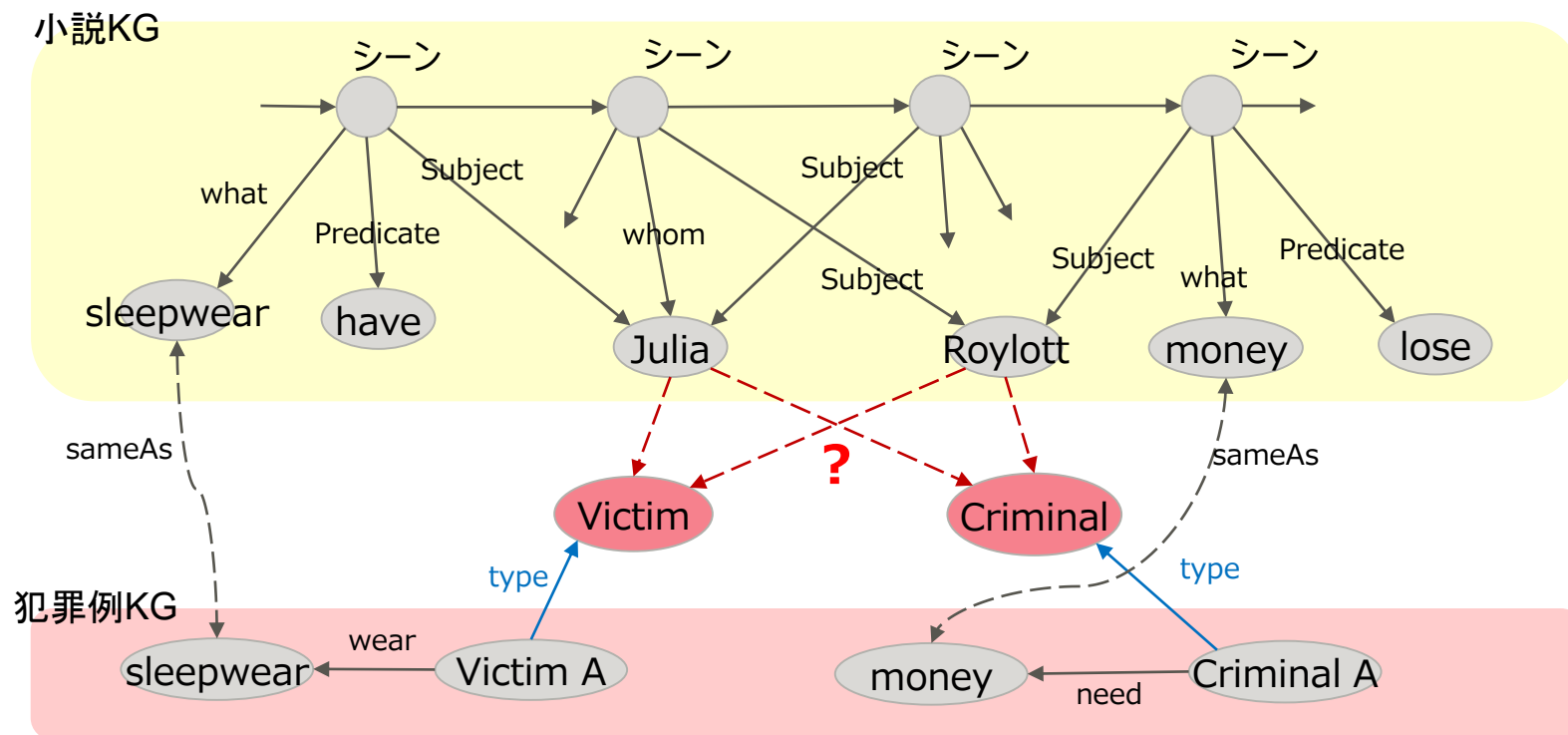
- 利用文書を増やす
 - ホームズの長編小説、ホームズ以外の推理小説など
- 動機・手段の分類手法の改善
 - 文ベクトルの計算方法を変える
 - ニューラルネットワークを利用

第 2 回に向けて

第2回で取り組みたいこと

■ Graph Embedding等の技術の活用

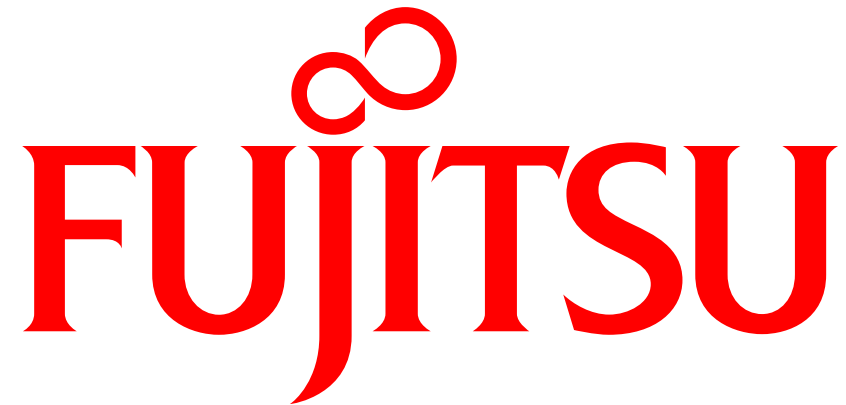
■ 複数知識グラフの緩やかな統合と犯人予測への適用



■ やりたいこと

- Graph Embeddingを利用した犯人の予測、予測理由の説明
- 常識・事例等、広範な予測に活用可能な知識グラフの構築

- もし一緒に取り組みたいという方がおられましたら、勉強会終了後にでもお声がけいただけると嬉しいです
- 進め方や知財等は要相談ですが、ハッカソンのようなかたちで、数回集まって一緒に作業できたらいいなと思っています



shaping tomorrow with you