

第1回ナレッジグラフ推論チャレンジ

# 機械学習によるアプローチ

FLL-MLチーム

岡嶋成司、鵜飼孝典、村上勝彦、高松邦彦、岸田あおい

## ■ チーム名

■ FLL-ML

## ■ メンバー

■ 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）

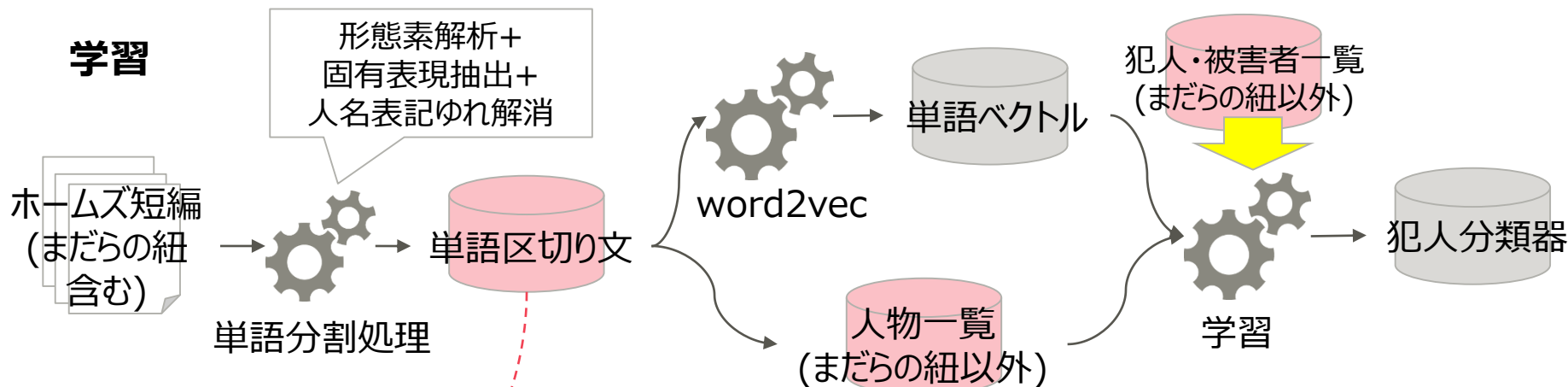
■ 岸田あおい（神戸市立西神戸医療センター）

- 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

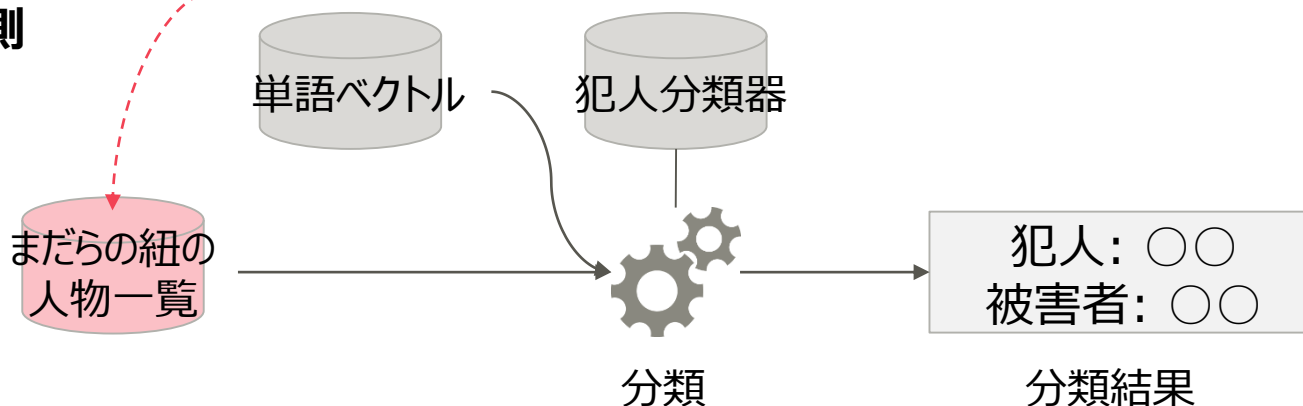
# 犯人の予測

- 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」「その他」の分類モデルを学習し、予測

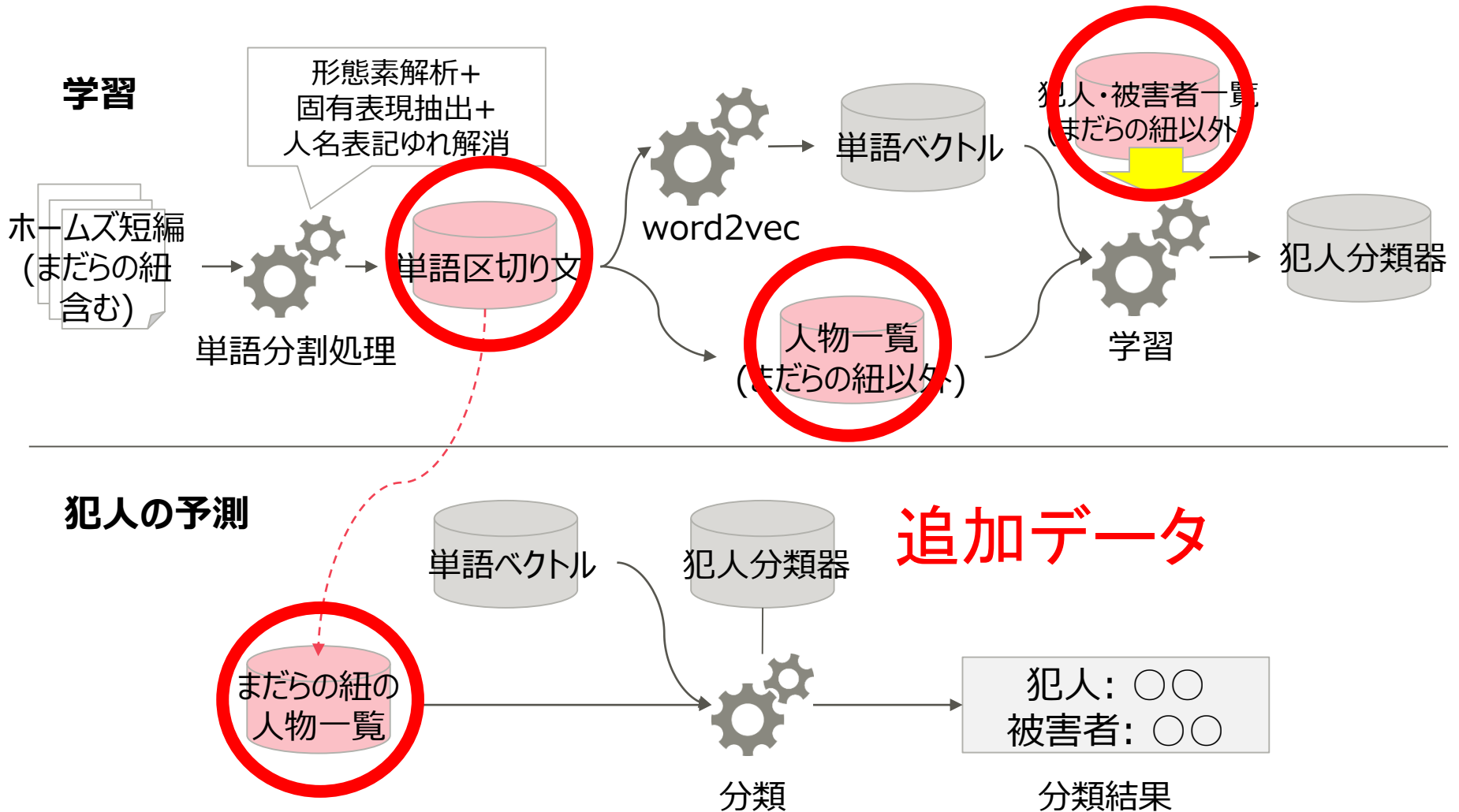
## 学習



## 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」「その他」の分類モデルを学習し、予測



## ■ 単語区切り文の作成

### 1. 形態素解析を利用して単語区切りの文を作成

そこでホズマ・エインジェルさんに会いましたの  
いつまでもホズマを信じております



そこで ホズマ・エインジェル さん に 会う ます た の  
いつ ま で も ホズマ を 信じる て おる ます

### 2. 固有表現抽出を適用し人物を抽出

そこで <PERSON>ホズマ・エインジェル</PERSON> さん に 会う ます た の  
いつ ま で も <PERSON>ホズマ</PERSON> を 信じる て おる ます

### 3. 人物の表記ゆれの解消

そこで <PERSON>ホズマ・エインジェル</PERSON> さん に 会う ます た の  
いつ ま で も <PERSON>ホズマ・エインジェル</PERSON> を 信じる て おる ます

## ■ 14,619文生成

## ■ 単語区切り文から、人物の一覧を作成

### ■ 113名の人物を抽出

- 機械的に抽出しているため、誤りあり

"アイリーン・アドラー","アクトン","アッパ・スウォンダム","アップルドア・タワーズ","アレク・カニンガム","アーサー・ピナー","イライアス・ホイットニ","イレーナ・アードラー","ウィリアムソン","ウィリアム・カーワン","ウィンディゲイト","ウェセクス","エイブ・スレイニ","エフィ","エルシ・パトリック","エヴァンズ","オウクショット","オーガスタス・ミルヴァートン","オールダショット","カルヴァトン・スミス","キュービット","キング","グラント・マンロウ","グリムズビ・ロイロット","グレゴリ","グローリア・スコット","ゴドフリ・ノートン","サイラス・ブラウン","シャーロック・ホームズ","ジェイベス・ウィルソン","ジェイベス・ギルクリスト","ジェイムズ・ウィンディバンク","ジェイムズ・バークリ","ジェイムズ・ライダー","ジャッキー","ジャック","ジャック・ウDDRリ","ジュリア","ジョン・オープン","ジョン・クレイ","ジョン・ストレーカ","ジョン・ヘア","ジョン・ヘクター・マクファーレン","ジョン・ホーナ","ジョージ","スタンデール","セバスチャン・モラン","ゼームス・モリアーティ","ソウムズ","ソヴリン","ダンカン・ロス","ダーヴィシャ","チャリング・クロス","デスボロ","トリダニック・ワーサ","トレヴァ","ナンシ・ディヴォイ","ニール","ネッド・ハンタ","ネヴィル・シンクレア","ハドソン","ハリス","ハリー・ピナー","バニスタ","パーカー","パーシ・トリヴェリアン","ヒュー・ブーン","ヒルトン","ピータソン","ピーター・ジョーンズ","フィッロイ・シムソン","フィツロイ・シムソン","フォードム","ブラッドストリート","ブルック","ブレッキンリッジ","ブレッシントン","ブレンダ・トリジェニス","ブレンダギャスト","ヘイタ","ヘレン・ストーン","ヘンリ・ウッド","ヘンリ・ベイカー","ベドウズ","ベーカー","ホズマ・エインジェル","ホール・パイクロフト","ボブ・カラザズ","ボブ・ファーガソン","マーティン","マーフィ","ミスター・コーネリアス","ミスター・トリジェニス","ミスター・マクファーレン","ミスター・モーティマー・トリジェニス","ミスター・ラウンドヘイ","ミスター・レストレイド","メアリ・サザランド","メイソン","メリウエザー","モウソン","モーカー","ラルフ・スミス","レストレード","ロイヤル・マロウズ","ロス","ロナルド・アディア","ロバート・ファーガソン","ワトソン","ヴァイオレット・スミス","ヴィクタ","ヴィクタ・サヴィッジ","ヴィンセント・スポールディング"



# 追加データ:犯人・被害者一覧

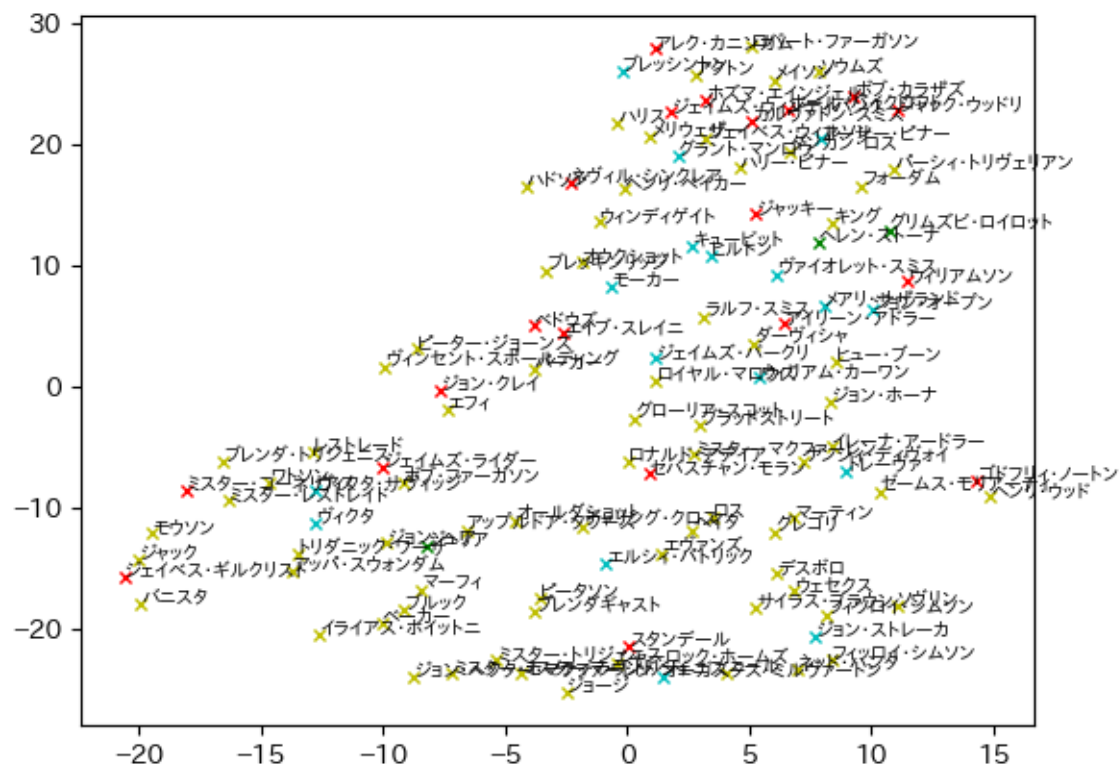
## ■ 青空文庫のホームズの22の短編小説について、犯人・被害者情報を付与

ID	タイトル	URL	犯人	犯人の性別	犯人の年齢	被害者	被害者の性別	被害者の年齢
SH01	ボヘミアの醜聞	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	アイリーン・アイリーン・アドラー, アイリーン・アドラー嬢, アイリーン嬢	女		フォン・クラーム伯爵, ヴィルヘルム・ゴッツライヒ・ジギースモント・フォン・オルムシュタイン, カッセル＝フェルシュタイン大公ボヘミア王	男	30
SH02	赤毛連盟	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	ジョン・クレイ	男	15～29			
SH03	同一事件	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	ホズマ・エインジェル, ジェイムズ・ウィンディバンク, ウィンディバンク, エインジェル, ホズマ	男		サザランド, メアリ・サザランド	女	
SH04	ボスコム谷の惨劇							
SH05	橙の種五粒	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	KKK, クー・クラックス・克蘭カルフーン船長, ジェームズ・カルフーン船長			ジョン・オープンショウ, オープンショウ 青年, イライアス, ジョセフ	男	
SH06	唇のねじれた男	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	ネヴィル・シンクレア	男				
SH07	蒼炎石	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	ジェム・ライダー, ジェイムズ・ライダー, ジョン・ロビンソン, ジェイムズ・ライダー客室係長	男		モーカー伯爵夫人	女	
SH08	まだらのひも	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	グリムズビー・ロイロット, ロイロット 博士, グリムズビー・ロイロット 博士	男		ジュリア・ストーナ	女	
SH09	牧師の親指							
SH10	花嫁失踪事件							
SH11	緑柱石の宝冠							
SH12	柵屋敷							
SH13	白銀号事件	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	馬(白銀号)			ジョン・ストレーカー	男	30ちょっとすぎ
SH14	土色の顔	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	エフィ・ヒープロン	女		グラント・マンロウ, ジャック	男	
SH15	株式仲間人	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	ホール・バイクロフト, バイクロフト	男		アーサー・ビナー, ビナー, ベディントン	男	
SH16	グローリア・スコット号	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	ベドウズ	男		トレーヴァ(ジェイムズ・アーミティッジ)	男	
SH17	マスグレーヴ家の儀式							
SH18	ライギット・パズル	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	アレク・カニンガム, アレク	男		ウィリアム・カーワン, ウィリアム	男	
SH19	曲れる者	<a href="https://www.aozora.gr.jp/cards/">https://www.aozora.gr.jp/cards/</a>	不定	不定		ジェイムズ・パークリ, パークリ 大佐	男	
			ビドゥル, 初老の男, 老人, ロシア人 親子ヘイワド 付き添い					

## ■ ホームズ短編の出現人物の単語ベクトルを可視化

### ■ t-SNEで次元削減してプロット

- 教師なしアプローチでは「犯人」「被害者」「その他」の分類は困難そう



赤 : 犯人  
青 : 被害者  
黄 : その他

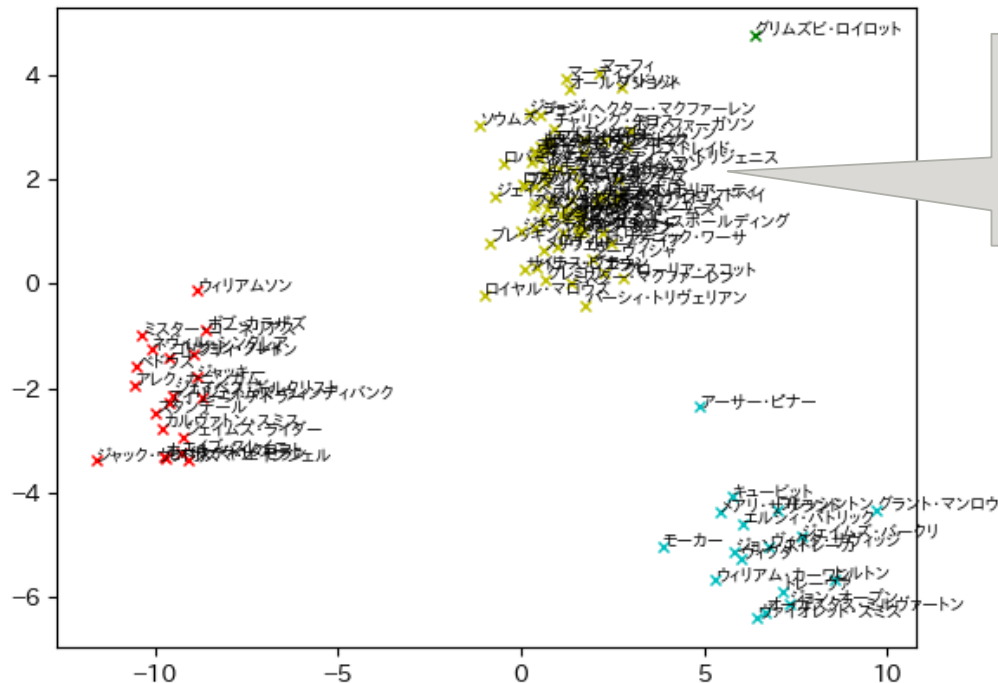
緑 : まだらの紐の人物

## ■ 教師ありの場合

- 線形判別分析でまだらの紐の登場人物以外から判別関数を学習し、ヘレン、ジュリア、ロイロットをプロット

- 全員「その他」に分類される、ロイロットはやや外れた値を持っている

⇒ 犯人・被害者が必ず一人ずついるという仮定を導入



確認しづらいが  
ヘレン、ジュリアは  
ここに埋まっている

赤 : 犯人  
青 : 被害者  
黄 : その他

緑 : まだらの紐の人物

## ■ 学習アルゴリズム

- Random Forest

## ■ 分類精度

- まだらの紐以外に登場する人物で、10分割交差検定
  - 「犯人」「被害者」「その他」の3クラス分類

Accuracy	macro F1	macro Precision	macro Recall
0.699	0.417	0.438	0.436

## ■ 「まだらの紐」の犯人予測

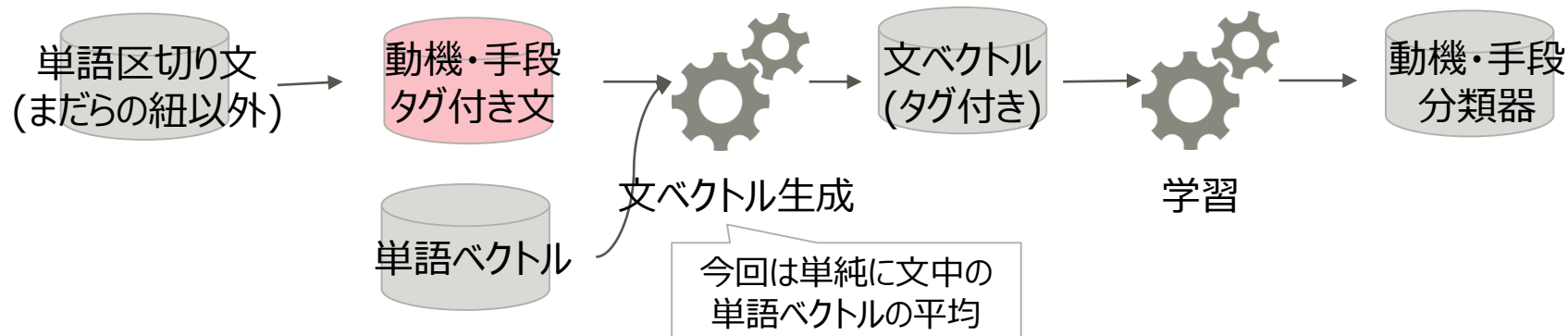
- ヒューリスティックなルールを導入
  - ホームズ、ワトソン、ハドソンなど犯人・被害者でないことが明らかな人物は「その他」として学習
  - 犯人・被害者は必ず一人ずついると仮定し、それぞれ分類スコア最大の人物を選択
- 固有表現抽出による「まだらの紐」の登場人物の抽出結果
  - ホームズ、ワトソン、ハドソン、ヘレン、ジュリア、ロイロット
- 犯人・被害者の分類スコアが最大の人物を選択
  - **犯人: グリムズビ・ロイロット**
  - **被害者: ジュリア**

# 理由の導出

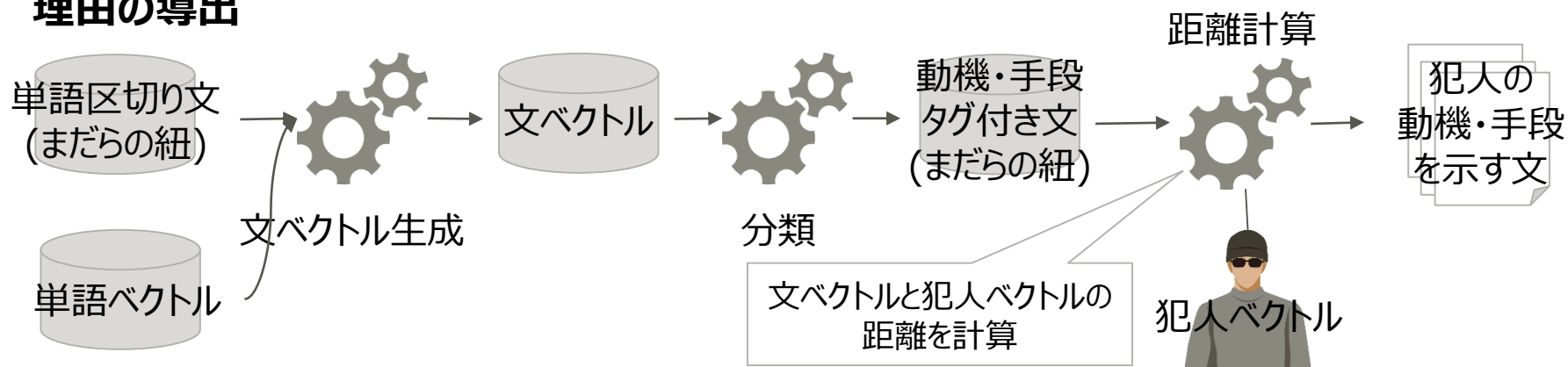
- 予測された犯人と意味的に近い文の中から、動機と手段を示している文を理由として提示する

- 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

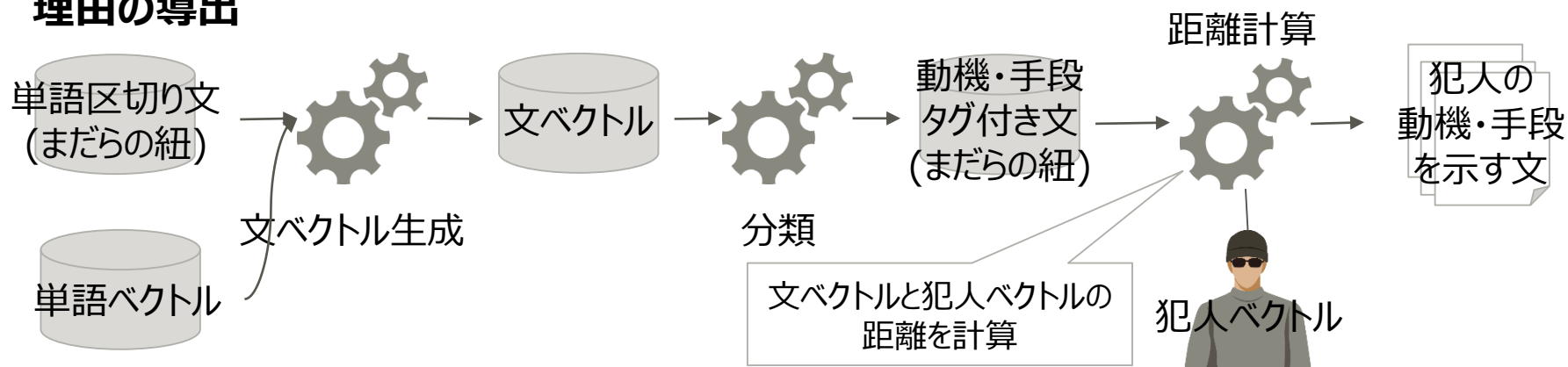


- 予測された犯人と意味的に近い文の中から、動機と手段を示している文を理由として提示する

- 動機と手段を示す文の分類器は、「まだらの紐」以外の短編の文から学習



## 理由の導出



# 追加データ: 目視による動機・手段タグ付け

## ■ 149件の動機・手段・機会等を表す文を抽出し、タグ付けしたデータを準備

24	SH03	opportunity	父が一週間ほどして、またフランスへ出向く予定でしたので、それでホズマは手紙で、あの人が行ってしまうまではお互いにあわない方が無難だと。
25	SH03	opportunity	そのあいだもわたくしどもは手紙は書けましたので、あの方も毎日書いてくださいました。
26	SH03	opportunity	手紙を受けとるのは朝でしたので、父は知る由もありません。
27	SH05	motive	活動目的は政治にあり、殊に黒人有権者の脅迫、その主義に反対する者の殺害ないし国外追放にある。
28	SH05	motive	これでわかったろう、この一覧と日記は南部の顔役の幾人かを結びつけるもので、これが見つかるまで夜もおちおち眠れない連中が大勢いるのだ。
29	SH05	action	その凶行に先だって通例、標的の人物へ奇抜な通知を送ることでよく知られる——榎葉の小枝、または甘瓜の種ないし櫚の種など。
30	SH05	action	記憶が確かなら本文は『A、BおよびCに種を送る』——つまり結社からの通告だ。
31	SH05	action	それから続く記述、AとBとが片づいたというのは国外へ出たということ、最後のCを訪ねるとは、おそらくCの不幸な結果のこと。
32	SH05	opportunity	青年をそこまでおびき出した手口は？
33	SH05	situation	争った形跡も、足跡もなく、強盗の線もなく、不審者の目撃証言もなし。
34	SH05	situation	結社の突然の解体とオープンショウが書類を持ってアメリカから消えた年とが一致する。

## ■ 149件のデータを用いて学習した分類器で、ホームズの短編を分類し、スコアを確認しながら2,288件にタグ付け

### ■ 動機・手段以外のタグは文単位では、判断が難しかったため除外

888	none	887	15.555436	0.256163904	0.208250485	これが事の始まりだったのです
889	none	888	16.046757	0.454416569	0.056860447	自分自身でご覧になって下さい、ホームズさん
890	none	889	16.628782	0.454635995	0.034350123	私は新聞を受け取り、次のように読み上げた
891	motive	890	16.762392	0.387619353	0.024724051	赤毛連盟に告ぐ?米国ペンシルヴァニア州レバノンの故イブリーキア・ホプキンズ氏の遺志に基づき、今、ただ名目上の尽力をするだけで週四米
892	motive	891	16.630323	0.183789045	0.064491845	赤髪にして心身ともに健全な二十一歳以上の男性は誰でも資格あり
893	mean	892	16.436518	0.07783891	0.071732001	月曜日、十一時、フリート街、ボープス・コート七番地、当連盟事務所内のダンカン・ロスに直接申し込まれたし
894	none	893	16.514875	0.739515889	0.029447875	私は、この奇怪極まる広告を二度読み返した
895	not sure	894	16.746943	0.492519085	0.067132273	……意味がさっぱりわからん
896	none	895	16.206215	0.192047887	0.024903496	口をついて出たのは、こんな叫びだった



## ■ 動機・手段に関わりそうな単語で短編を検索してタグ付け

### ■ 動機関連単語：56単語, 294文

- 遊び, 楽しみ, いたずら, 悪戯, 気晴らし, 暇つぶし, 戯れ, ケンカ, 悪ふざけ, 困窮, 窮乏, 貧窮, 窮迫, 逼迫, 貧困, 生活苦, 鬱状態, 病苦, 利欲, 激情, 衝動, 情念, 怨恨, 敵意, 憎悪, 憎しみ, 私怨, 敵愾心, 嫉妬心, 罪悪感, 恨み, 怨み, 妬み, 悪感情, 敵対心, 嫌悪感, 復讐心, 偏見, 劣等感, 猜疑心, 疑心, 猜疑, 義憤, 報復, 制裁, 仕返し, 共犯, 金, 営利, 自己顕示, 防衛, 護身, 服従, 迎合, 信条, 信念

### ■ 手段関連単語：23単語, 348文

- 斬, 刃物, 撲, 鈍器, 絞, 紐, 刺, 殴, 毒, 薬, 扼, 轢, 爆, 圧, 焼, 溺, 射, 噛, 銃, 電, 惨, 鑒, 誅

## ■ 学習アルゴリズム

- Random Forest

## ■ 動機、手段の分類精度

- 「動機」「手段」「その他」の3クラス分類

Accuracy	macro F1	macro Precision	macro Recall
0.503	0.490	0.497	0.494

## ■ 犯人の動機・手段の推定

- 人物と文の距離を人物ベクトルと文ベクトルの距離とする
  - 他の人物と比較して犯人に近い文を犯人関連文とする
- 犯人関連文のうち、動機・手段のスコアが高いものを理由として出力

## ■ グリムズビ・ロイロットに近い文の内、動機のスコアが高い文top30

- ずいぶん愛嬌のある人物だ
- **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、家内で窃盗事件が追っ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**
- だが用心すべきは向こうの方です
- どうかわたくしをお助けください
- そのロイロット博士とわたくしの母が結婚したのは、インドにいた頃で、当時わたくしの母は、ベンガル砲兵隊にいたストーナ少将の若き未亡人でした
- あとは、僕らにお任せください
- **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**
- 母は内地に帰ってまもなく亡くなりました
- この住所を教えてください
- ロイロット家は、イングランドでも古いサクソンの家系の末裔で、サリー州の西の果て、ストーク・モランにございます
- 願わくは、詳しいお話を
- 準備次第、辻馬車呼んでウォータルー駅へ急ごう
- たとえば、大工はいったい何を考えて、隣の部屋へ穴をあけるのか
- ですから、亡くなったとき姉ひとりだったことには間違いございません
- 死に際のひもという言葉の謎
- そんな寝台どこにある
- **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**
- もっと先を読む人間が自分を追っているのですから
- すると、お姉さまがつけさせた
- ところがこの頃から、にわかに父の様子が変わってまいりました
- 今朝はみな同じ具合なのだ
- じゅうぶん見せていただきました、ストーナさん
- ええ、針金につながってもいい
- ひも
- そこで一緒になって再び姉のところへ戻ってみますと、姉は気を失っておりました
- **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**
- さて、できるだけ時間を有効に使いましょう
- 寝間着のままでございました

——— 癪癢持ち？

金銭問題？

**癪癢**、あるいは**金銭問題**によって殺害か？

## ■ グリムズビ・ロイロットに近い文の内、動機のスコアが高い文top30

- 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

癪癪、あるいは金銭問題によって殺害か？

## ■ グリムズビ・ロイロットに近い文の内、手段のスコアが高い文top30

- 年の若いご婦人が興奮気味に來られたらしく、僕に面会したい、と
- わたくしは飛び起きてランプをつけましたが、部屋には何も見あたりません
- それに続いて、がちゃんと何か重い金物でも墮ちたような音がしました
- 我々も肝を冷やし、私はホームズの顔を、ホームズは私を見つめたまま立ちすくんだ
- なんと
- 実に面白い
- と、黒い服に身を包み、顔には厚い面紗をかけた一人の夫人が、窓際の椅子から立ち上がった
- あのお嬢さんが、あの獣に後をつけられたことで、困ったことにならねばよいのだが
- それからインドのカルカッタへ参りまして、そこで、技術と持ち前の性格もあってか、医院を開業致しました
- ええ、そのほかにも少し直したところがいくつもございます
- 私は息を呑む
- 床にはいつくばりながら前後に動いて、床板の隙間まで細かく調べた
- 私は息を呑んだ
- ごらんなさい、引き綱は換気のための小さな穴の上のあたりで、かぎ針に結ばれて動かない
- 老医師はつつかと進むと、暖炉の火掻き棒をつかみ上げ、大きな手で折り曲げてみせた
- だがロムたちが何をしようと
- 外から夜鳥の啼く声が時折、まさに目の前の窓のところで猫に似た長く引きつった鳴き声が一度
- 失礼します、気を強くお持ちに
- とホームズが口を挟む
- と、ホームズはささやいた
- 私には見えなかった
- そのようで
- 一度、何年か前に
- その通りです
- 卓上には、半分窓を開けた角灯が置かれてあり、それから流れ出る一条の光の中に、戸の開いたままの金庫が見える
- どこにもない
- ご婦人は寝台を動かさなかった
- 早すぎて当てずっぽうにも思えるが、その実いつも論拠があり、結果、持ち込まれた問題を解き明かしてしまう
- かすかな物音が計画を台無しにする
- そのほか、小さな籐椅子二つを加えたのがこの部屋の家具のすべてであった

犯行手段らしき文は抽出できず

# 結論と考察

■ 犯人は、「グリズムビ・ロイロット」、被害者は「ジュリア」

■ 理由

■ 犯行の動機

- 癩癩あるいは金銭問題
- 根拠は以下の文
  - 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/](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

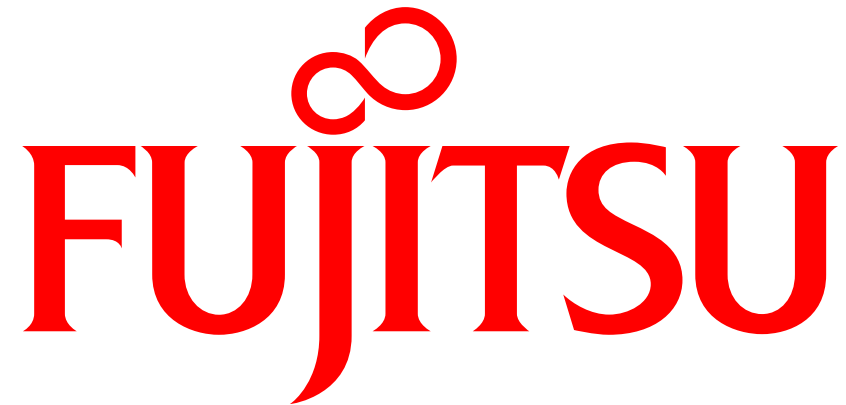
- 理由として、癪癢など幅広い可能性をとらえられる ⇒ 全部ナレッジで網羅するのは難しい
- まだらの紐の犯行手段のような複雑な手続きは、単純に近いデータを示すだけでは説明できない ⇒ 複雑な手続きの説明は推論のほうが向いている

## ■ 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
    - 単語ベクトルの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（単純平均でなく加重平均にするなど）
  - ニューラルネットワークを利用



shaping tomorrow with you