

実行プログラム

■ 実行環境

- python3系
 - 必要ライブラリ: numpy, scipy, scikit-learn, gensim

■ スクリプト

- classify.py: 犯人分類器を学習して、人物を分類するスクリプト
- reasoning.py: 犯人を入力として、犯行理由を推定するスクリプト
- word2vec.py: 単語区切り文から単語ベクトルを学習するスクリプト

■ 設定ファイル

- settings.json: スクリプトの設定ファイル
 - 犯人・被害者、短編小説に登場する全人物、まだらの紐の登場人物、も内部に記載

■ データファイル、モデル

- data/all_ss.txt: すべての単語区切り文
- data/train.tsv: 動機・手段タグ付け文
- data/madara.tsv: まだらの紐の単語区切り文+対応する原文
- model/w2v_cbow_200_10_3_128_ex.model: word2vecの学習済みモデル

■ その他

- misc/reasoning_ex.xlsx: reasoning.pyの出力tsvを見やすくしたもの

■ classify.py

- まだらの紐の登場人物を「犯人」「被害者」「その他」に分類する

```
python3 classify.py -s settings.json
```

- 出力例: それぞれの人物について、「犯人」「被害者」「その他」のスコアを順に表示する

```
ジュリア: [0.05685335 0.12233468 0.82081198]  
グリムズビ・ロイロット: [0.15172414 0.30344828 0.54482759]  
ヘレン・ストーナ: [0.1030303 0.10647858 0.79049112]
```

- 「犯人」「被害者」「その他」の分類精度の測定

```
python3 classify.py -s settings.json -e
```

- 出力例: accuracy, f1, precision, recallなどを順に表示

```
test_f1_macro: 0.4173675610595115  
fit_time: 0.01661520004272461  
score_time: 0.005752277374267578  
test_recall_macro: 0.4363095238095237  
test_accuracy: 0.6993939393939393  
test_precision_macro: 0.4382239057239056
```

■ reasoning.py

- classify.pyで推定した犯人を入力として、犯行理由を推定

```
python3 reasoning.py -s settings.json -t “グリムズビ・ロイロット” > out.tsv
```

- 出力例: 以下のようなtsvを出力
 - 各カラムの意味は左から順に、「犯人と文の距離」「動機スコア」「手段スコア」「その他スコア」「対応する文」

0.3285151720046997	0.42825474956440884	0.1553407734319695	0.41640447700362165	その穴が致命的なのかどうか、これで説明可能なのかどうか、確かめたい
0.47623592615127563	0.42340236698549855	0.15629971546572055	0.420297917548781	それと歯ブラシ、それだけで足りるだろう
0.48528987169265747	0.38894574653633507	0.24732516339388605	0.36372909006977894	ですから、わたくしに近づく危険をどう避ければよいのか教えてくださると思って……
0.5406539440155029	0.28458849247678064	0.3130553979312505	0.4023561095919688	深刻な面持ちでランプをつけ、ホームズは先に立って廊下を進んだ
...				

■ 「動機」「手段」の分類精度の測定

```
python3 reasoning.py -s settings.json -e
```

- 出力例: accuracy, f1, precision, recallなどを順に表示

```
test_precision_macro: 0.49744254019794454
fit_time: 0.052347421646118164
test_accuracy: 0.5031888653981678
score_time: 0.005284595489501953
test_recall_macro: 0.49405311355311354
test_f1_macro: 0.4903909251248568
```

■ word2vec.py

- 単語区切り文からword2vecを学習する

```
cat data/all_ss.txt | python3 word2vec.py -s settings.json
```

- settings.jsonの設定に基づいて学習
- modelディレクトリ以下にword2vecのモデルが生成される

設定ファイルの中身

■ settings.json

classify.py, reasoning.pyで使用する
word2vecのモデルを指定

reasoning.pyで使用する動機・手段
タグ付け文とまだらの紐の単語区切り文を指定

word2vec学習の設定

まだらの紐の登場人物

犯罪者一覧

被害者一覧

ホームズ短編の登場人物一覧

```
{
  "path": {
    "word2vec": "model/w2v_cbow_200_10_3_128_ex.model",
    "train": "data/train.tsv",
    "test": "data/madara.tsv"
  },
  "word2vec": {
    "dim": 200,
    "iter": 128,
    "window": 10,
    "min_count": 3,
    "skip_gram": false
  },
  "data": {
    "label_map": {"motive": 0, "mean": 1, "none": 2},
    "names": ["ジュリア", "グリムズビ・ロイロット", "ヘレン・ストーナ"],
    "criminals": ["アイリーン・アドラー", "ジョン・クレイ", "ホズマ・エインジェル", "ジェームズ・ライダー", "ホール・パイクロフト", "ペドゥズ", "アレク・カニンガム", "ジ", "ブ・スレイニ", "ボブ・カラザズ", "ジャック・ウDDR", "ウィリアムソン", "ミスター・コーネリアス", "セバスチャン・モラン", "ゴドフリイ・ノートン"],
    "victims": ["メアリ・サザランド", "グラント・マンロウ", "アーサー・ピナー", "トル", "ルトン", "キュービット", "エルシィ・パトリック", "ブレッシントン", "ジョン・ストレーカ", "ルヴァートン"],
    "persons": ["アイリーン・アドラー", "アクトン", "アッパ・スウオンダム", "アッブルドア・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", "イレナ・アドラー", "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", "エルシィ・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", "ア・スコット", "ゴドフリイ・ノートン", "サイラス・ブラウン", "シャーロック・ホームズ", "ジョ", "イムズ・パークリ", "ジ>エイムズ・ライダー", "ジャッキー", "ジャック", "ジャック・ウDDR", "フレック", "フロン・オ", "フロン・クレイ", "フロン・ストレーカ", "フロン・ベア", "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", "ダーヴィンシャ", "チャリング・クロス", "デスポロ", "トリダニック・ワーサ", "トレヴァ", "ナンシィ・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", "パニスタ", "パーカー", "パーシィ・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", "フォーダム", "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・ストーナ", "ヘンリ・ウッド", "ヘンリ・ベイカー", "ペドゥズ", "ペーカー", "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"]
  }
}
```