

競馬予想システム

依田倫太郎

更新情報

- 2023/01/25
 - my_modulesの階層変えた
- 2023/02/02
 - my_modulese/prepare 実装
 - my_modulese/constants 実装
 - resultsテーブル -> race_resultsに変数名変更
 - race_resultsテーブルにtrainer_id, owner_id を追加 (過去のcsvとは一致しないので, 古いcsv消して, 新たにデータをダウンロードする必要あり)
 - 詳しくはmain.ipynbを見よ
 - 学習まではもしかしたら動かないかもしれない (未テスト)

更新情報

- 2023/02/05
 - preprocess 実装
 - 過去データで, 訓練シミュレーション可能

ロードマップ

次やること

当日データで予測するコード書く(わかりやすく)

目次

更新情報

ロードマップ

1. 目的
2. ディレクトリ構成
3. ライブラリと使用法

1. 目的

- めざせ不労所得
- ラインから60秒で始められる高収入バイト
- 万馬券を当てよう

2. ディレクトリ構成

データベース

```
├── Data
│   ├── horse_results.csv
│   ├── peds.csv
│   ├── results.csv
│   └── return.csv
```

2. ディレクトリ構成

```
my_library
├── __pycache__
│   ├── horse.cpython-37.pyc
│   └── horse.cpython-39.pyc
├── horse.py
├── my_modules
│   ├── funcs
│   │   ├── __init__.py
│   │   ├── __pycache__
│   │   │   ├── __init__.cpython-37.pyc
│   │   │   ├── __init__.cpython-38.pyc
│   │   │   ├── _calc.cpython-37.pyc
│   │   │   ├── _calc.cpython-38.pyc
│   │   │   ├── _plot_importance.cpython-37.pyc
│   │   │   ├── _plot_importance.cpython-38.pyc
│   │   │   ├── _process_data.cpython-37.pyc
│   │   │   ├── _process_data.cpython-38.pyc
│   │   │   └── _show_best_kaime.cpython-38.pyc
│   │   ├── _calc.py
│   │   ├── _plot_importance.py
│   │   ├── _process_data.py
│   │   └── _show_best_kaime.py
│   ├── prepare
│   ├── preprocessing
│   └── simulator
│       ├── __init__.py
│       ├── __pycache__
│       │   ├── __init__.cpython-37.pyc
│       │   ├── __init__.cpython-38.pyc
│       │   ├── _simulator.cpython-37.pyc
│       │   └── _simulator.cpython-38.pyc
│       └── _simulator.py
└── training
    ├── __init__.py
    ├── __pycache__
    │   ├── __init__.cpython-37.pyc
    │   ├── __init__.cpython-38.pyc
    │   └── _learn.cpython-38.pyc
    └── _learn.py
```

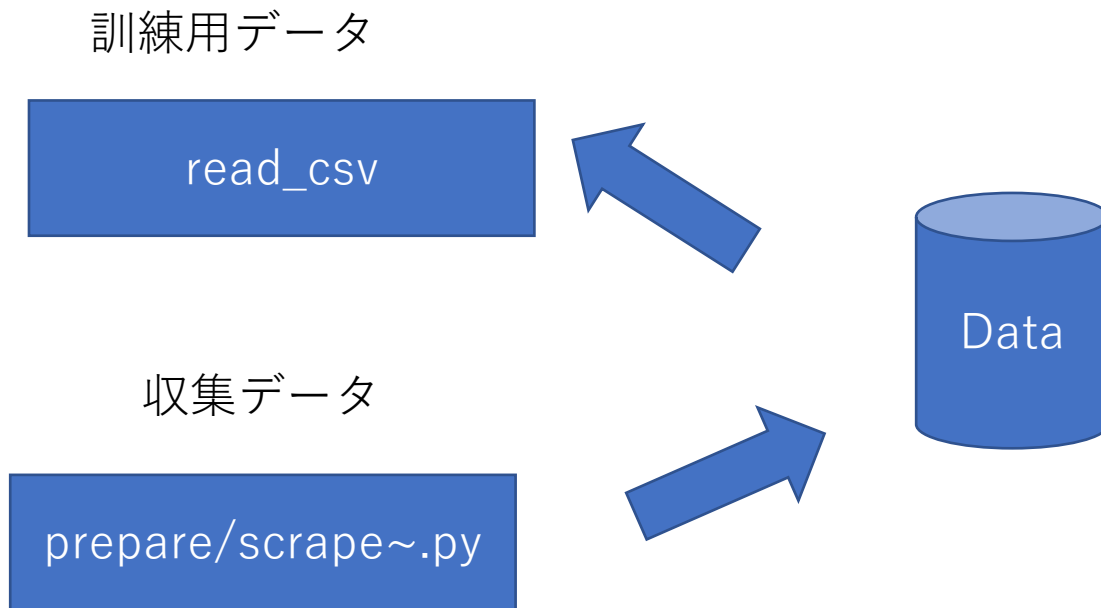
2023/02/04

my_library はもう使わない予定

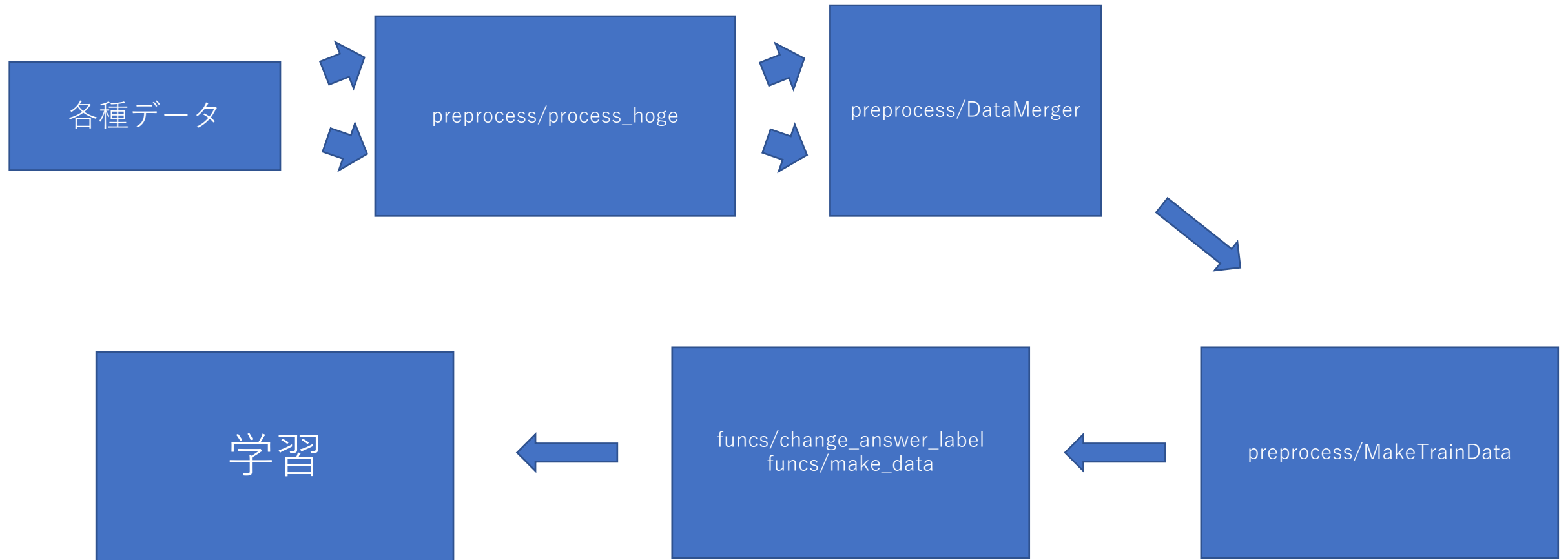
自作ライブラリはmy_modulesで完結させる

ライブラリと使用法

- ※詳しくは, main.ipynbを見よ
- 基本的にmainのセルを頭から順次実行



ライブラリと使用法



※ fasttext

- 競馬の血統データのベクトル化に, 自然言語処理で用いられるfasttextが良いと聞いた
- ただインストール方法がクソだるい割に精度の向上に貢献したかは謎
- 多分エラーのオンパレードなので, fasttextは当分使わない方針
- ただ, 血統データのカテゴリ数がデータ多いと5万とかなるので, 精度の観点からは後々, カテゴリ変数ではなく, ベクトル化していきたい
- エラー出る場合は無視して該当箇所を適宜コメントアウトして!