

说明文档

环境要求

- Python 3.10+

虚拟环境配置

创建虚拟环境

```
1 | python -m venv .venv
```

激活虚拟环境

```
1 | # windows 系统
2 | .venv\Scripts\activate
3 |
4 | # Linux/Mac 系统
5 | source .venv/bin/activate
```

安装依赖包

```
1 | pip install -r requirements.txt
```

运行项目代码

```
1 | python main.py
```

项目参数配置

项目参数配置位于 `configs/custom.yaml` 和 `configs/default.yaml` 其中 `default.yaml` 包含所有参数配置，而 `custom.yaml` 会覆盖其配置。

参数说明如下

```
1 | init_data: [0.5, 0.5, 0.5, 0.5, 0.5]    # 初始向量
2 | filter_threshold: 0.6    # 余弦过滤阈值
3 | datasets:
4 |   path: "data/决策算法测试数据.xlsx"    # 数据文件相对路径
5 |   sence_columns: ["Cdtotal", "CECsoil", "ECsoil", "pHsoil", "SOM"] # 场景列列名
6 |   process_columns: ["TPV", "DosageBC", "Time"] # 工艺列列名
7 |   type_columns: ["Type"]    # 类型列列名
8 |   result_columns: ["Cdreduction rate"] # 结果列列名
9 | train:
10 |   test_size: 0.2    # 验证集比例
11 |   random_state: 42  # 随机数参数
12 |
```

```
13 search_params_threshold: 10 # 参数搜索阈值
14 search_params_num_iter: 4 # 参数搜索迭代次数
15 select_variables_threshold: 0.5 # 变量选择阈值
16 select_variables_max_num: 5 # 变量选择最大数目
17
```

程序运行

初始运行

控制台运行输出

```
Windows PowerShell
版权所有 (C) Microsoft Corporation. 保留所有权利。

安装最新的 PowerShell, 了解新功能和改进! https://aka.ms/PSWindows

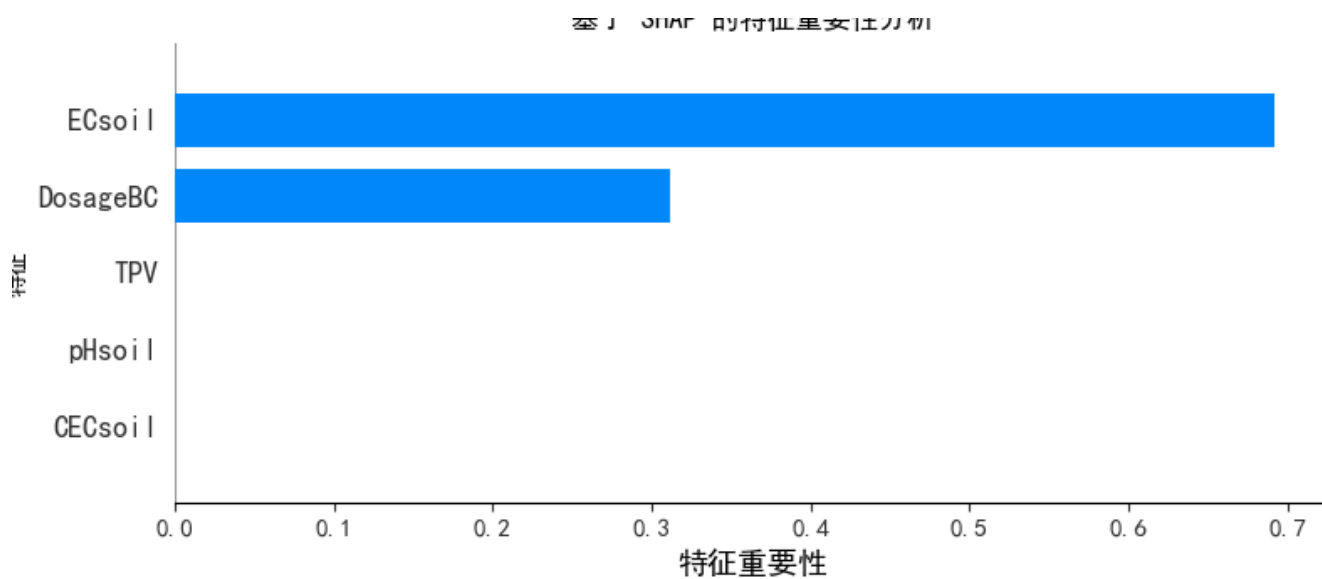
PS C:\Codes\agricultural-decision-system-v2> .\.venv\Scripts\activate
(.venv) PS C:\Codes\agricultural-decision-system-v2> python .\main.py
各 Type 的数据量:
- B: 39 条
- A: 6 条
所有 Type 均满足数据量要求
过滤后的有效 Type: ['A', 'B']
类型: A
  Models    RMSE    MAE    R2
0      Ridge  2.3577  2.2362 -1.9916
1      Lasso  2.1884  1.7120 -1.5774
2       SVR   2.4812  2.2461 -2.3134
3 DecisionTree 2.4474  2.2827 -2.2236
4 RandomForest 2.3621  2.1338 -2.0029
5      XGBoost 2.7058  2.4804 -2.9404
6 LinearRegression 1.9521  1.8859 -1.0509
最好的模型是: LinearRegression
类型: B
  Models    RMSE    MAE    R2
0      Ridge  0.7769  0.6264 -0.0373
1      Lasso  0.7722  0.6495 -0.0249
2       SVR   0.5075  0.3480  0.5573
3 DecisionTree 0.4515  0.2618  0.6497
4 RandomForest 0.4540  0.3071  0.6458
5      XGBoost 0.3817  0.2475  0.7496
6 LinearRegression 0.8609  0.7047 -0.2736
最好的模型是: XGBoost

当前选择矩阵:
  Cdtotal CECsoil ECsoil pHsoil SOM TPV DosageBC Time
A      0      1      1      1      0      1      1      0
B      0      1      1      1      0      1      1      0

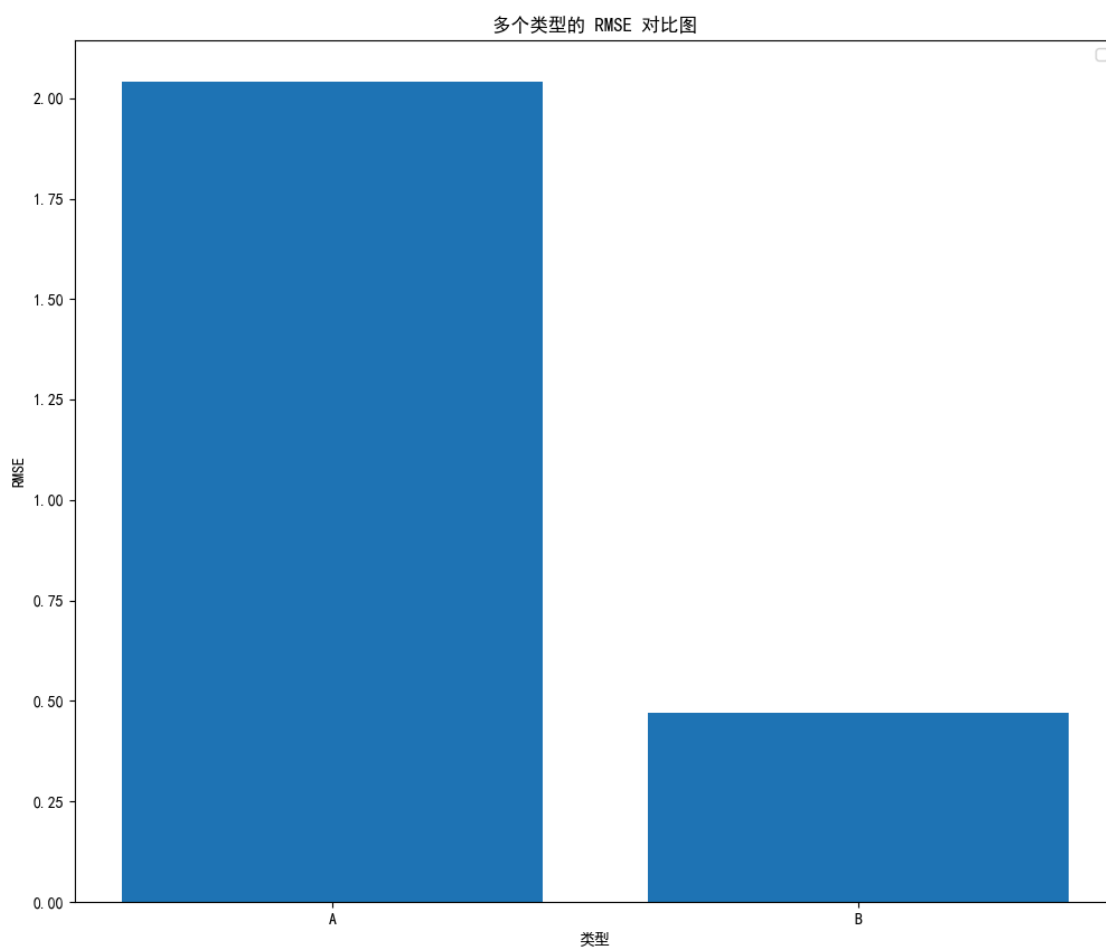
当前 Type 列表: ['A', 'B']
当前所有列: ['Cdtotal', 'CECsoil', 'ECsoil', 'pHsoil', 'SOM', 'TPV', 'DosageBC', 'Time']

请输入修改指令, 格式如: A 0 0 (将 Type A 的第 0 列设置为 0)
输入 -1 退出编辑
>>> |
```

Type A 的 shap 值输出



Type A & B 的 RMSE 对比



模型特征选择人工介入

控制台控制演示

```
Windows PowerShell
6 LinearRegression 0.8609 0.7047 -0.2736
最好的模型是：XGBoost

当前选择矩阵：
  Cdttotal CECsoil ECsoil pHsoil SOM TPV DosageBC Time
A         0         1         1         1         0         1         1         0
B         0         1         1         1         0         1         1         0

当前 Type 列表：['A', 'B']
当前所有列：['Cdttotal', 'CECsoil', 'ECsoil', 'pHsoil', 'SOM', 'TPV', 'DosageBC', 'Time']

请输入修改指令，格式如：A 0 0 （将 Type A 的第 0 列设置为 0）
输入 -1 退出编辑
>>> A 0 1
✅ 已修改：Type 'A' 的第 0 列 -> 1

当前选择矩阵：
  Cdttotal CECsoil ECsoil pHsoil SOM TPV DosageBC Time
A         1         1         1         1         0         1         1         0
B         0         1         1         1         0         1         1         0

当前 Type 列表：['A', 'B']
当前所有列：['Cdttotal', 'CECsoil', 'ECsoil', 'pHsoil', 'SOM', 'TPV', 'DosageBC', 'Time']

请输入修改指令，格式如：A 0 0 （将 Type A 的第 0 列设置为 0）
输入 -1 退出编辑
>>> A 0 1
❗ Type 'A' 的第 0 列已经是 1，未修改。

当前选择矩阵：
  Cdttotal CECsoil ECsoil pHsoil SOM TPV DosageBC Time
A         1         1         1         1         0         1         1         0
B         0         1         1         1         0         1         1         0

当前 Type 列表：['A', 'B']
当前所有列：['Cdttotal', 'CECsoil', 'ECsoil', 'pHsoil', 'SOM', 'TPV', 'DosageBC', 'Time']

请输入修改指令，格式如：A 0 0 （将 Type A 的第 0 列设置为 0）
输入 -1 退出编辑
>>> -1
❌ Type 'A' 的选中列数不为 5，无法退出编辑。

当前选择矩阵：
  Cdttotal CECsoil ECsoil pHsoil SOM TPV DosageBC Time
A         1         1         1         1         0         1         1         0
B         0         1         1         1         0         1         1         0

当前 Type 列表：['A', 'B']
当前所有列：['Cdttotal', 'CECsoil', 'ECsoil', 'pHsoil', 'SOM', 'TPV', 'DosageBC', 'Time']

请输入修改指令，格式如：A 0 0 （将 Type A 的第 0 列设置为 0）
输入 -1 退出编辑
>>> A 1 0
✅ 已修改：Type 'A' 的第 1 列 -> 0

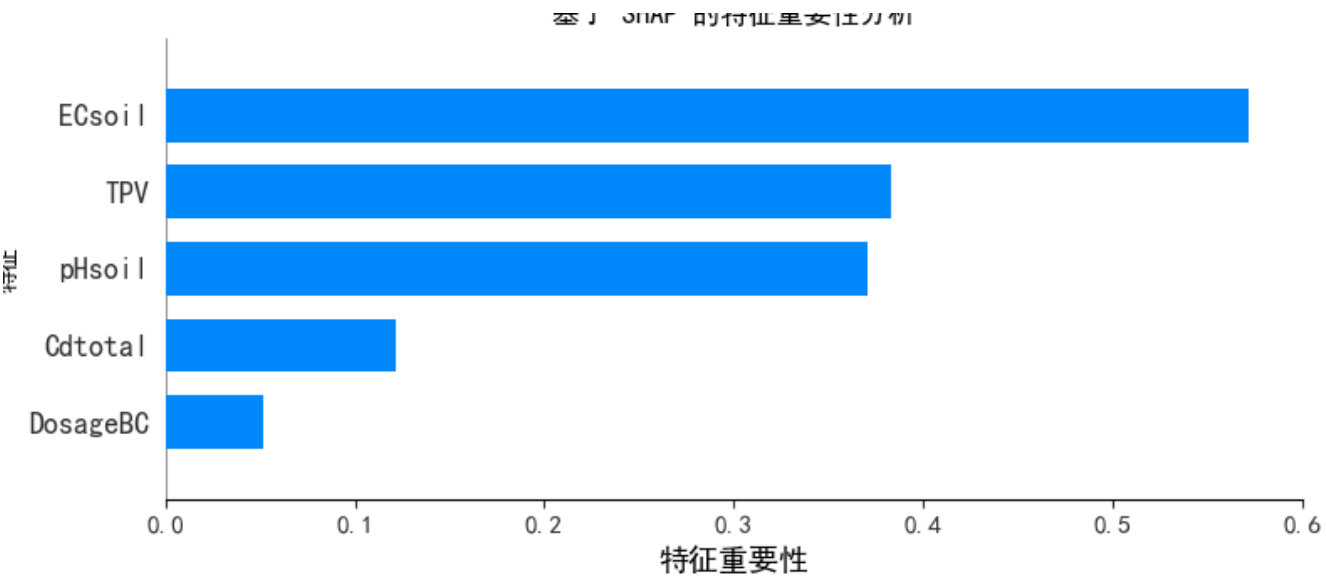
当前选择矩阵：
  Cdttotal CECsoil ECsoil pHsoil SOM TPV DosageBC Time
A         1         0         1         1         0         1         1         0
B         0         1         1         1         0         1         1         0

当前 Type 列表：['A', 'B']
当前所有列：['Cdttotal', 'CECsoil', 'ECsoil', 'pHsoil', 'SOM', 'TPV', 'DosageBC', 'Time']

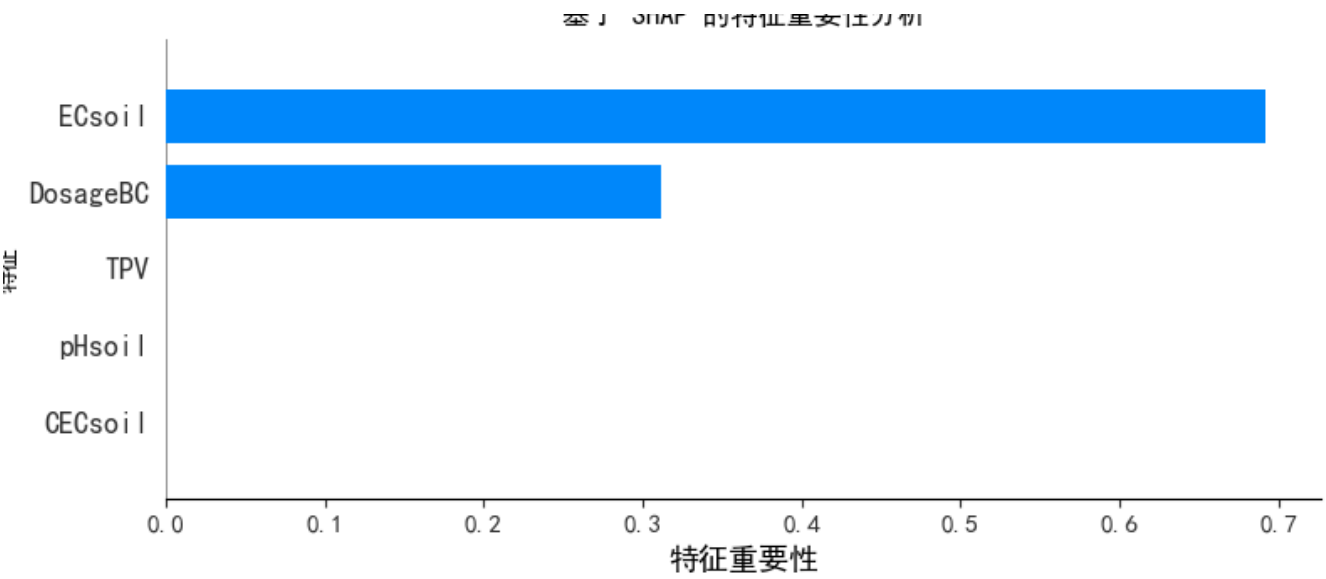
请输入修改指令，格式如：A 0 0 （将 Type A 的第 0 列设置为 0）
输入 -1 退出编辑
>>> -1
退出编辑。
|
```

修改后的 Type A 的 shap 值初始 shap 值对比

修改后



修改前



模型最终结果输出

控制台在未继续修改的情况下退出编辑进入最终结果输出

```
Windows PowerShell
输入 -1 退出编辑
>>> -1
退出编辑。

当前选择矩阵:
  Cdtotal  CECsoil  ECsoil  pHsoil  SOM  TPV  DosageBC  Time
A         1         0         1         1     0         1         0
B         0         1         1         1     0         1         0

当前 Type 列表: ['A', 'B']
当前所有列: ['Cdtotal', 'CECsoil', 'ECsoil', 'pHsoil', 'SOM', 'TPV', 'DosageBC', 'Time']

请输入修改指令, 格式如: A 0 0 (将 Type A 的第 0 列设置为 0)
输入 -1 退出编辑
>>> -1
退出编辑。
最优的类型是: B
将要拓展的列: ['TPV', 'DosageBC']
最优的参数组合及结果是:
  TPV  DosageBC  Time  prediction
0.075097 1629.0125    3  42.806049
0.075097 857.3750    3  42.806049
0.075097  85.7375    3  42.806049
0.075097 23.7500    3  42.806049
0.075097 451.2500    3  42.806049
(.venv) PS C:\Codes\agricultural-decision-system-v2> |
```