

# 1225消融实验存档

---

## 1225消融实验存档

### 0.实验目的

#### 0.1.特殊注意

0.2.简要总结

0.3.工作流程-->

    0.3.1检查5折交叉验证并且获取结果

    0.3.2迁移学习效果比较\*\*

#### 4.工作内容总结

##### 1. 迁移学习的定量分析

    1.1 SGF任务上进行CV验证的性能分析

    1.2 SGF任务上进行CV验证的稳定性分析

    1.3 SIF任务上进行CV验证的性能分析

    1.4 SIF任务上进行CV验证的稳定性分析

#### 1.5.CV验证总结

##### 2. 迁移学习分析

    2.1. SIF-LR上进行迁移

    2.2. SIF-RF上进行迁移

    2.3. SIF-XGB上进行迁移

    2.4. SGF-LR上进行迁移

    2.5. SGF-RF上进行迁移

    2.6. SGF-XGB上进行迁移

#### 2.7.迁移分析

##### 3. 消融实验总结

##### 4. 异常情况

## 0.实验目的

---

验证在：

- 去掉Morgan
- 去掉Avalon
- 去掉所有分子指纹

上述情况下，进行性能的定量分析，以及迁移学习的分析

将结果结合特征分析进行研究，验证前期假设

## 0.1.特殊注意

- 该文件夹中的所有文件（即为该记录所运行的代码）本应属于上一级目录/notebook/
- 如果需要重新使用，请将py文件移动到上一层文件夹中
- 20251220-20251225：消融实验：验证去掉Morgan或者Avalon分子指纹后（或者全部去除），平均的性能/数据稳定性以及迁移效果的变化
- 在某一页ppt中，迁移学习的效果，原本是All但是title上写了Morgan，这里是笔误，忽略不计即可，后期有机会

## 0.2. 简要总结

本文件夹用于存放 2025-12-25 期间的 Ablation (消融) 实验相关资料。包含用于清洗与汇总交叉验证结果的脚本 (如 `clean_cv_json.py`、`aggregate_cv_json.py`)、用于生成对比图和可视化的脚本 (如 `generate_cv_figure.py`)，以及若干用于复现与稳定运行的实验脚本 (如 `Phase3_1_stable.py`、`Phase3_2_stable.py`、`scr.py`)。目录还包含实验产生的 JSON 数据和图像目录 (`cv_figure/`、`trans_figures/`)，总体目的是集中管理消融实验的数据、脚本与图表，方便结果复现、对比与可视化分析。

## 0.3. 工作流程-->

### 0.3.1 检查5折交叉验证并且获取结果

- `Phase3_1_stable.py` 生成 `data.json` (记录每一次的交叉验证结果)
- `clean_cv_json.py` 生成 `data_cleaned.json` (清除 null 值和 NaN)
- `aggregate_cv_json.py` 生成 `data_average.json` (计算 100 次的平均值)
- `generate_cv_figure.py`, 根据四次平均值计算的情况, 计算热力图

### 0.3.2 迁移学习效果比较\*\*

- 直接运行 `Phase3_2_stable.py` 直接生成对应的迁移学习热力图
- 如果需要检查消融 (即同一个数据位置, 四种消融的区别), 可以使用 `scr.py` 绘制折线图

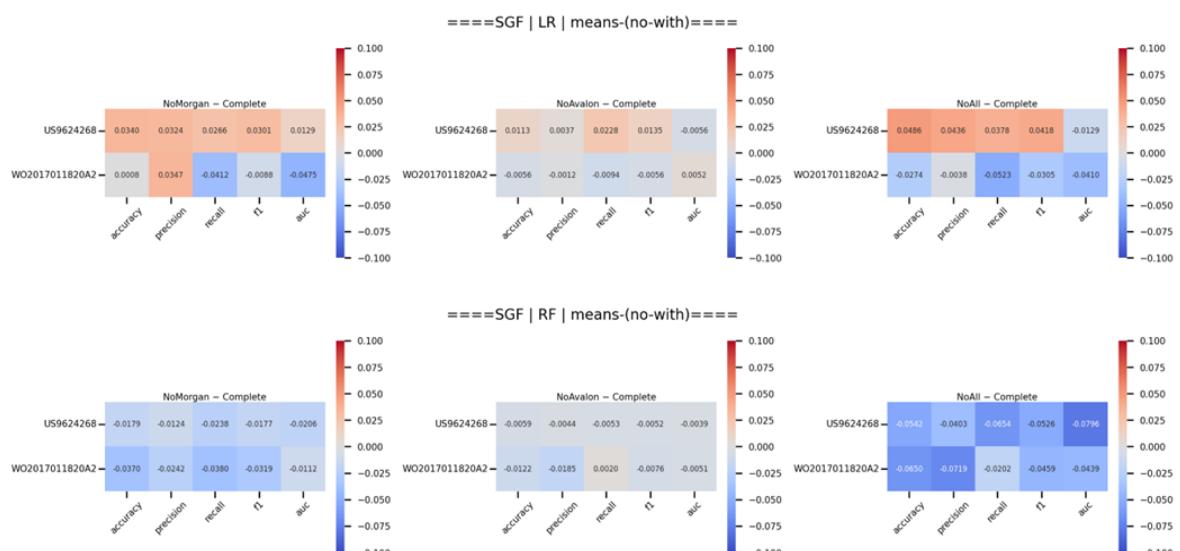
## 4. 工作内容总结

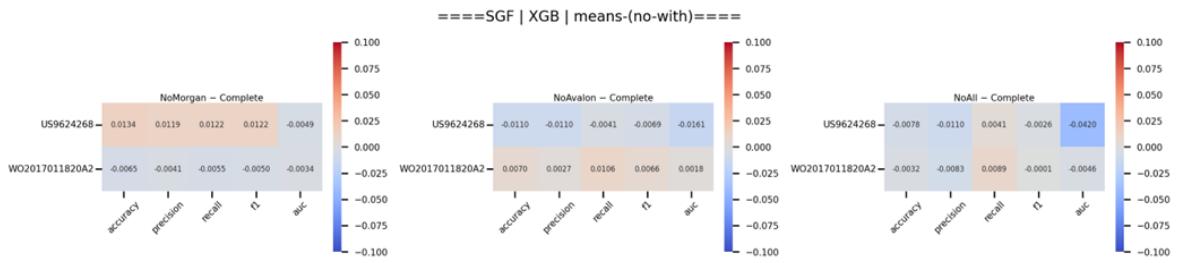
- 报告保存在 `pptx` 文件中, 在该目录下
- `pdf` 文件同步留档

## 1. 迁移学习的定量分析

为了避免随机性影响, 迁移学习进行了 100 次重复预测, 每次使用不同的随机种子, 并将结果取平衡值

### 1.1 SGF 任务上进行 CV 验证的性能分析



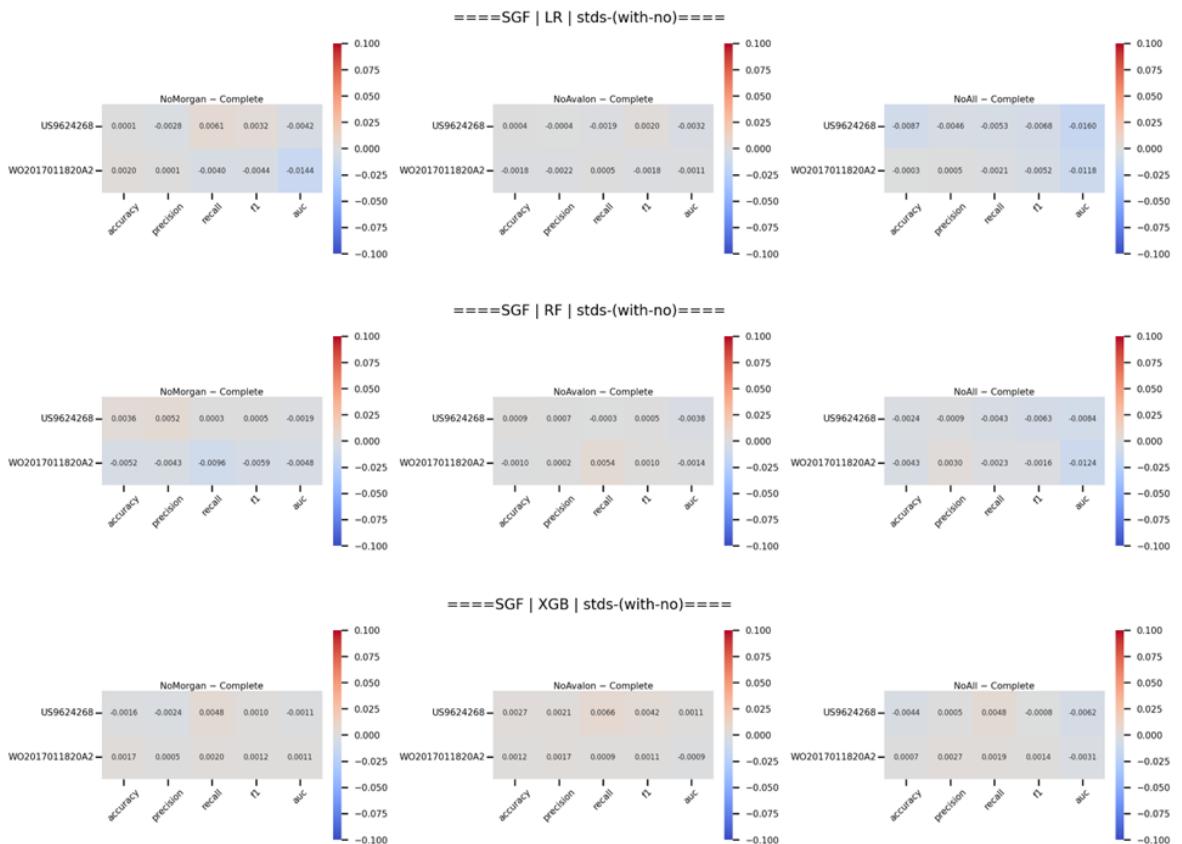


结论：

在SGF任务上：

- 去掉Morgan分子指纹倾向于下降
- 去掉Avalon基本不变
- 均去掉后呈现下降趋势

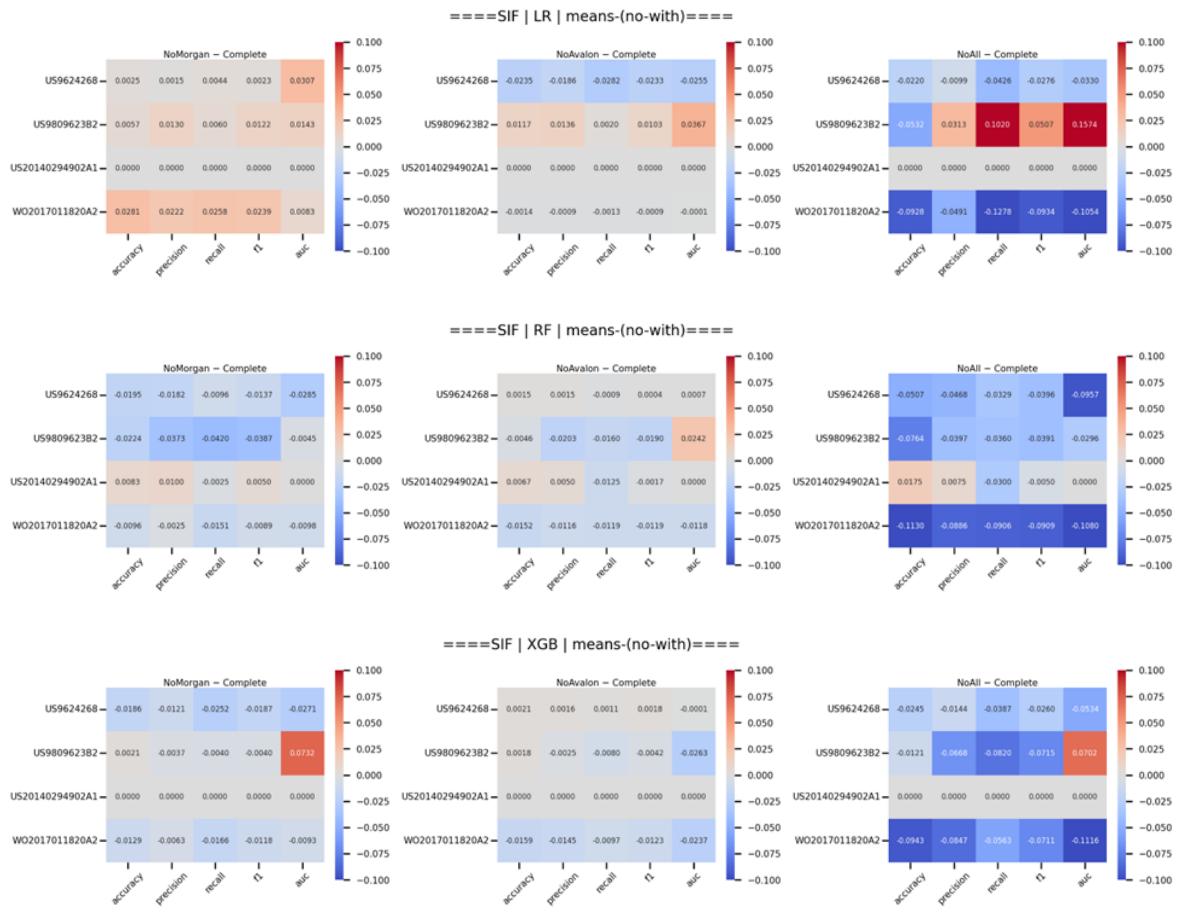
## 1.2 SGF任务上进行CV验证的稳定性分析



在SGF任务上：

- 去掉分子指纹，对于模型本身的预测稳定性影响不大

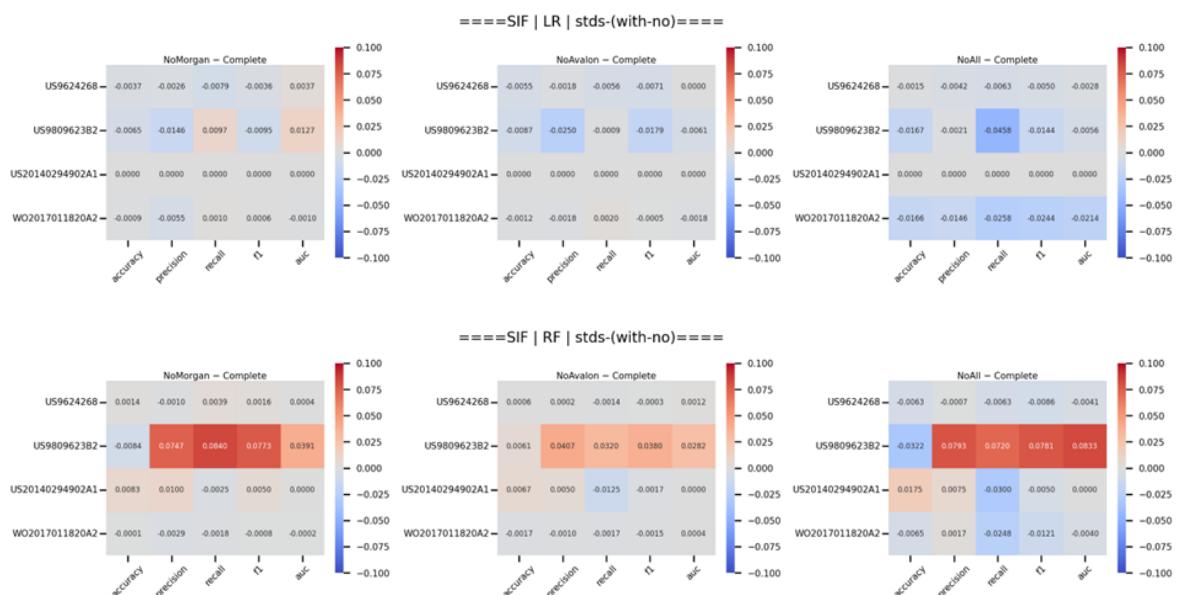
## 1.3 SIF任务上进行CV验证的性能分析

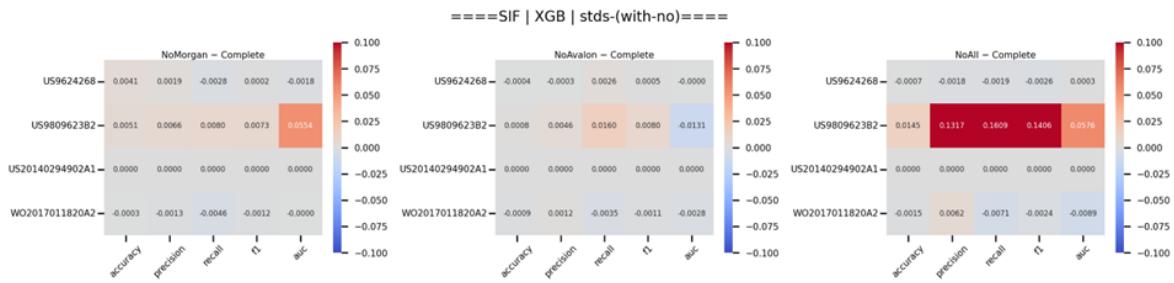


在SIF任务上：

- 去掉Morgan分子指纹,倾向于下降
- 去掉Avalon后呈现下降,但是下降程度较小
- 均去掉后呈现明显下降趋势

## 1.4 SIF任务上进行CV验证的稳定性分析





## 1.5.CV验证总结

### 性能

- 去掉Morgan分子指纹后呈现下降趋势（在LR模型上有一定提升）
- 去掉Avalon分子指纹后也呈现下降趋势（下降趋势小于前者）
- 去掉所有分子指纹的内容后，下降趋势十分明显

### 稳定性

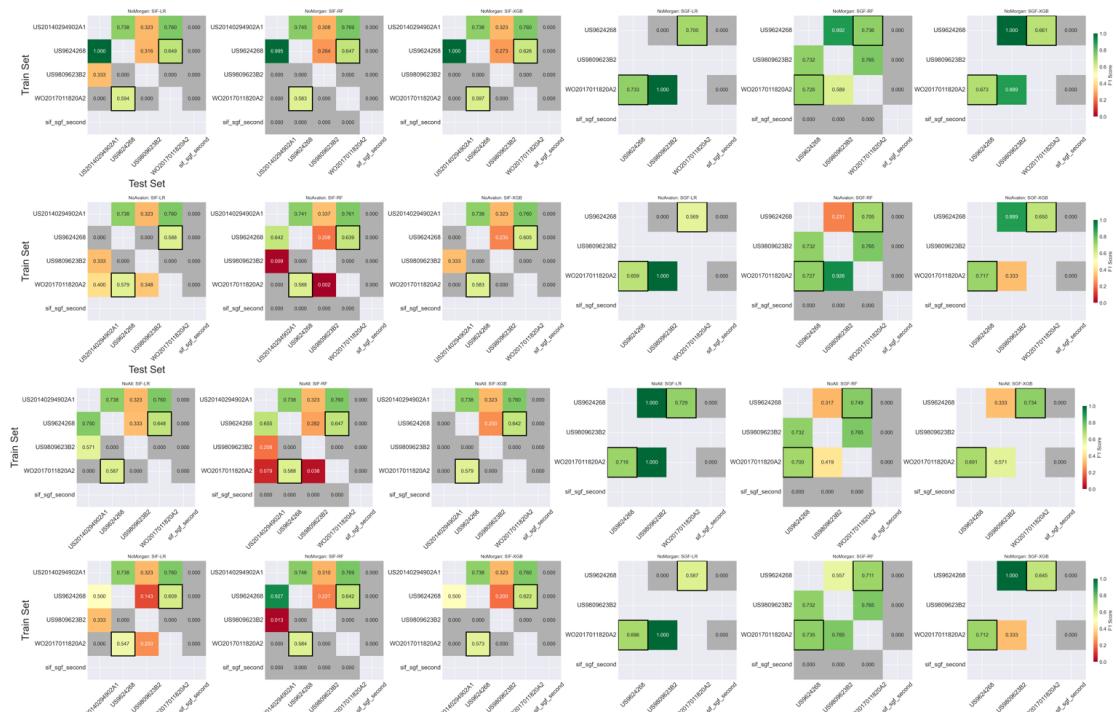
- 去掉分子指纹后稳定性几乎无变化，数据特征没有影响模型自身预测的稳定性
- 除了在US98096数据集上变化奇怪

### 从任务角度分析：

- 在SIF任务上，上述任务变化程度更明显

## 2. 迁移学习分析

有个圣诞树配色也算过了一次圣诞节了

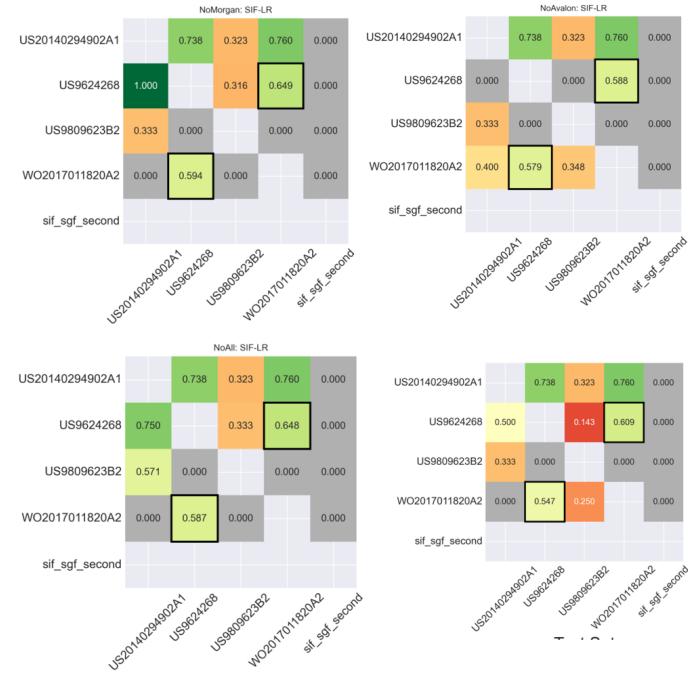
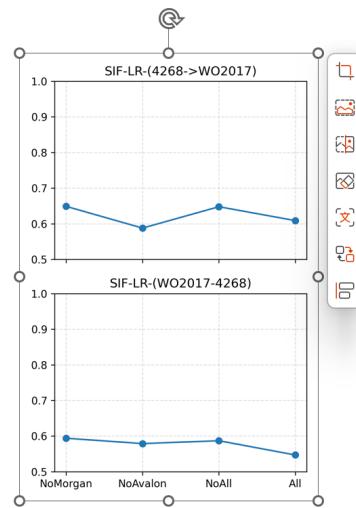


和上述的CV相同，迁移学习部分为了避免偶然性，也是多次迁移取平均值

## 2.1. SIF-LR上进行迁移

SIF-LR

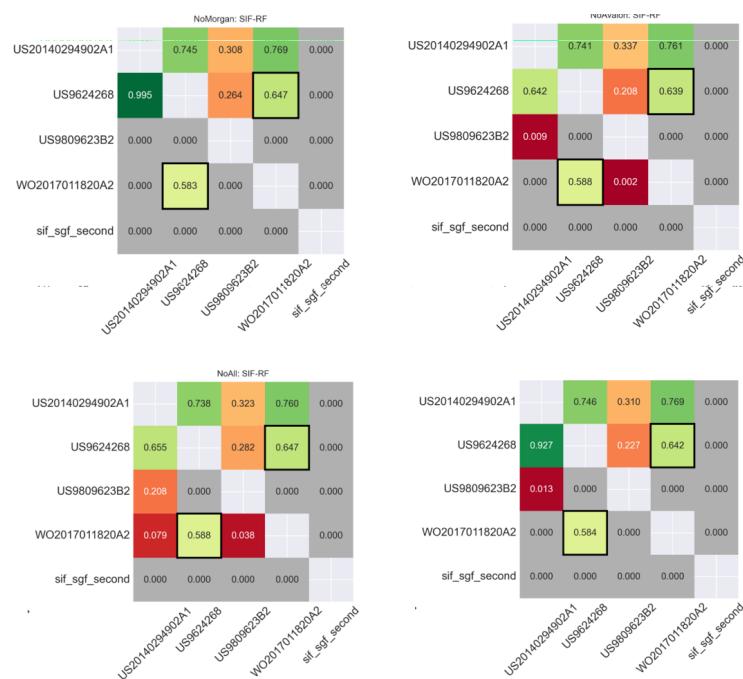
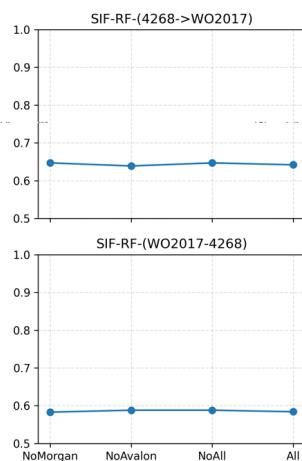
- 去掉Morgan后迁移效果上升
- 去掉Avalon后迁移效果有提升或者维持



## 2.2. SIF-RF上进行迁移

SIF-RF

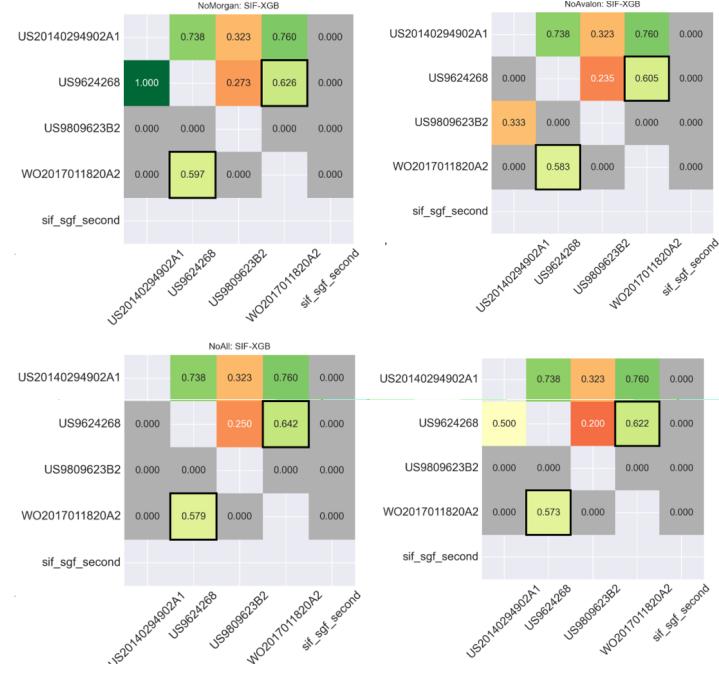
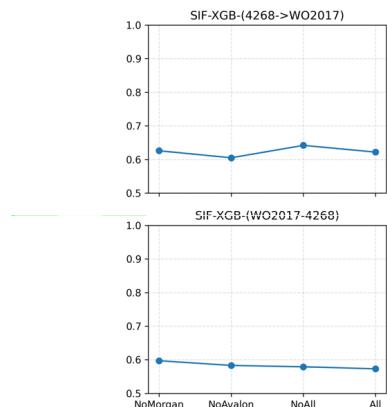
- 去掉分子指纹影响较小



## 2.3. SIF-XGB上进行迁移

### SIF-XGB

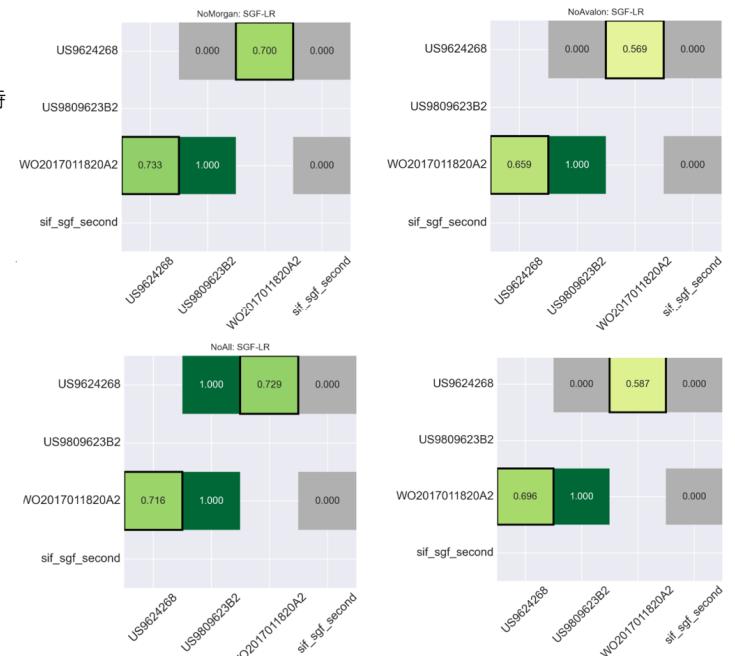
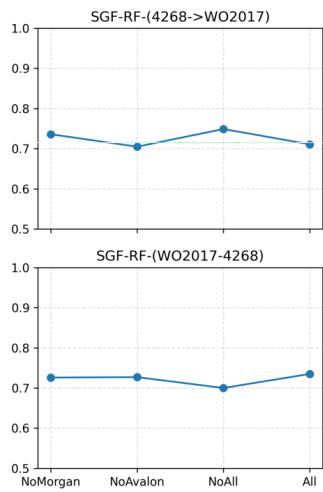
- 去掉分子指纹后变化较小



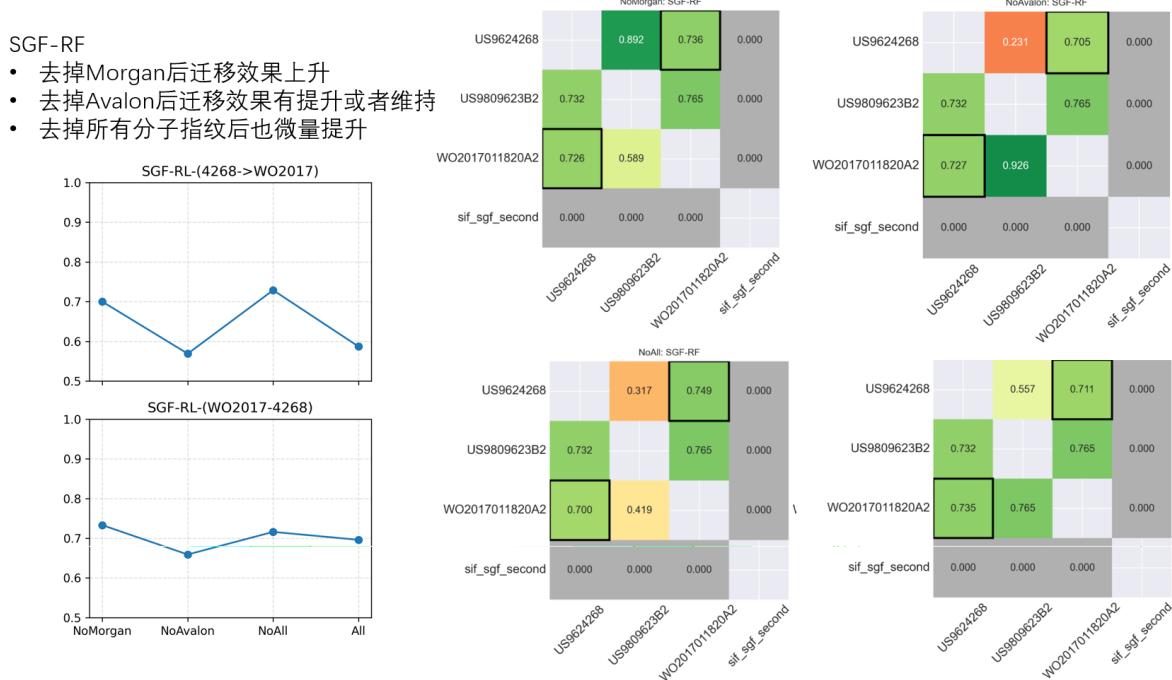
## 2.4. SGF-LR上进行迁移

### SGF-LR

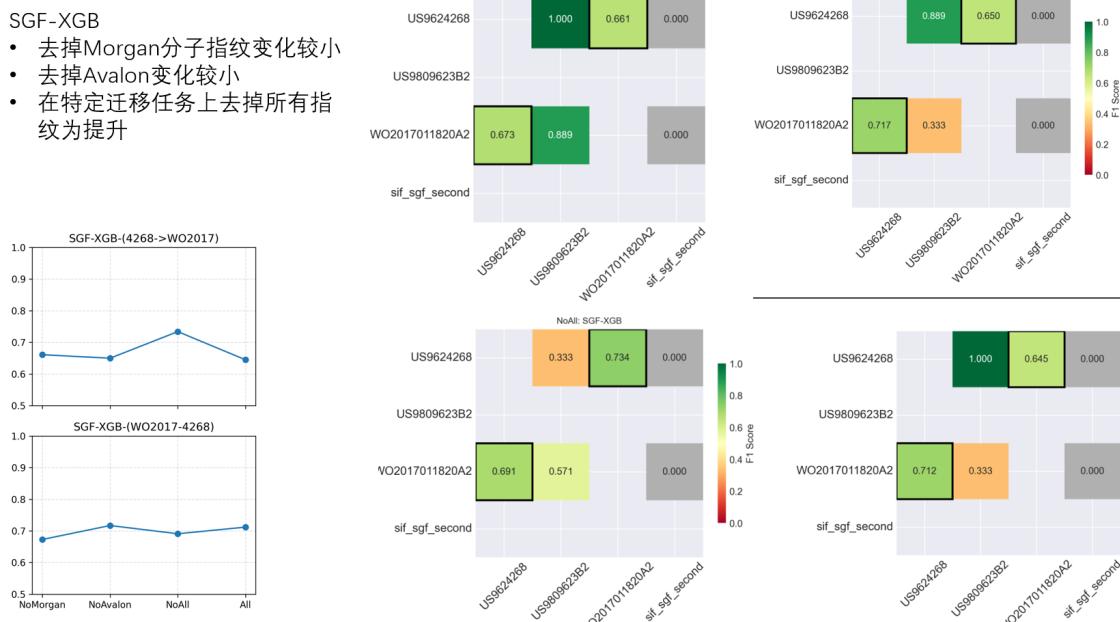
- 去掉Morgan后迁移效果上升
- 去掉Avalon后迁移效果有提升或者维持
- (上述变化数值较小, 1-2%)



## 2.5. SGF-RF上进行迁移



## 2.6. SGF-XGB上进行迁移



## 2.7. 迁移分析

- 去掉Morgan: 整体趋势为上升
- 去掉Avalon: 整体趋势为保持稳定
- 去掉全部分子指纹: 整体趋势是提高, 并且大多和去掉Morgan的迁移结果趋同
- 保留全部分子指纹: 情况与去掉Avalon趋同

## 3. 消融实验总结

- 在性能角度来说, **去掉Morgan影响较大**, 去掉Avalon影响趋近于稳定, 全部去掉后影响显著
- 从迁移角度来看, **去掉Morgan可以提升迁移效果** (并且该效果与“**同时去掉Morgan和Avalon**”趋同)

- 在迁移实验之中，去掉Avalon对迁移效果的影响较小，与不去掉分子指纹的情况趋同
- 综合以上三点，Morgan对整体表现的影响（无论是否正面）要大于Avalon带来的影响

## 4. 异常情况

---

US98096数据集敏感度过高