**Python脚本说明文档**

1. Python版本>=3.6。
2. 依赖包见requirements.txt。使用pip install -r requirements.txt安装依赖包。
3. 脚本按以下顺序执行

|  |  |  |
| --- | --- | --- |
| 序号 | 脚本名称 | 功能说明 |
| 1 | proc\_raw\_data.py | 数据预处理脚本。原始数据经预处理后在路径./data/proc/下生成预处数据，已proc为前缀标记 |
| 2 | run\_defusion.py | 执行算法主文件。计算在不同参数设置下的矩阵分解。结果保存在./results/文件夹下。 |
| 3 | check\_convergence.py | 绘制算法不同参数设置下的收敛曲线。结果保存为png文件到./results/文件夹下，以convergence为前缀标识。 |
| 4 | consistency\_eval.py | 根据矩阵分解结果将样本聚类。结果保存为csv在./results/文件夹下，以clustering为后缀标记 |
| 5 | survival\_eval.R | 由聚类结果计算生存曲线p值。各参数设置下的p值保存在./record\_log\_rank\_test\_pvalue.txt文件下，对应的生存曲线以png格式保存在./data/surv\_curve/下。 |
| 6 | select\_topk.py | 由矩阵分解的权重选择各类型数据的top K特征。结果保存在./data/topk/文件夹下。之后需要将p值最小的clustering文件放到topk 和 proc文件夹下。 |
| 7 | find\_best\_RFparams.py | 使用随机森林计算混合特征权重，结果保存在./FeatureSelction/文件夹中，以importance为后缀标识，并根据留一准确率（leave-one-out）评价top K特征的分类能力，结果保存在eval\_RF\_topk\_features.txt文件中。 |