

## 流程

- 1 训练集数据准备
- 2 开始训练
- 3 测试集数据准备
- 4 验证测试集

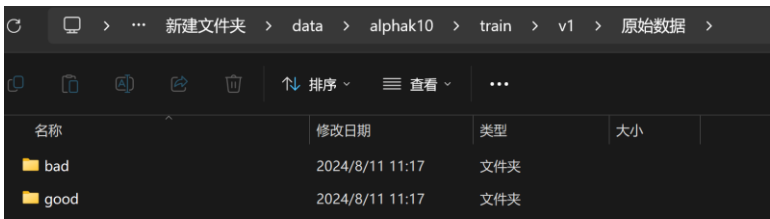
# 1 训练集数据准备

1 创建目录结构如下

mlkl



alphak10



2 把好曲线放入 good 文件夹，把坏曲线放入 bad 文件夹

3 执行脚本文件 生成训练集数据和 csv.py

```
19 DATA = 'mlkl'
20 NUM = '1'
21 NUM_dimension = 1024
22 SIGMA = 1
23 randomSeed = 42
```

超参数		
DATA	目前支持 2 种数据 mlkl alphak10	训练集是 mlkl 就写入 mlkl 训练集是 alphak10 就写入 alphak10
NUM	训练集第一版本就 NUM=1， 如果训练集有第二种版本就 NUM=2，训练集还有第三种版本就 NUM=3。	目前不需要修改
NUM_dimension	是一条曲线有多少个数据点，原始数据有时候是 500 个点，有时候是 1000 个点，那就统一插值到 1024 个点。	目前不需要修改
SIGMA	高斯平滑的超参数设置，选为 1 是比较好的，可以平滑曲线去掉部分噪声。	目前不需要需改
randomSeed	5 折交叉验证划分数据集的超参数，默认使用 42	目前不需要需改

4 训练集结构

执行 **生成训练集数据**和 **csv.py** 脚本文件之后，在 D:\1DeepSIFA\data\mlkl\train\v1 目录文件下就会生成如下结构的文件，4 个文件夹和一个 csv 文件。



名称	修改日期	类型	大小
原始数据	2024/8/25 15:32	文件夹	
归一化后npz	2024/8/25 15:32	文件夹	
归一化插值后npz_1024_高斯平滑1	2024/8/25 15:32	文件夹	
5折交叉验证_42	2024/8/25 15:32	文件夹	
1090.csv	2024/8/25 15:32	Comma Separat...	22 KB

## 2 开始训练

1 把 DeepSIFA 文件放到这个目录下



2 DeepSIFA 包含文件如下

名称	修改日期	类型	大小
model_hub	2024/8/11 16:40	文件夹	
models	2024/8/24 10:58	文件夹	
utils	2024/8/24 10:58	文件夹	
其他功能	2024/8/24 10:58	文件夹	
111.py	2024/8/25 13:57	Python 源文件	3 KB
eval_test.py	2024/8/25 11:32	Python 源文件	18 KB
eval_train.py	2024/8/24 16:45	Python 源文件	18 KB
eval_val.py	2024/8/25 11:31	Python 源文件	18 KB
train.py	2024/8/25 11:28	Python 源文件	17 KB

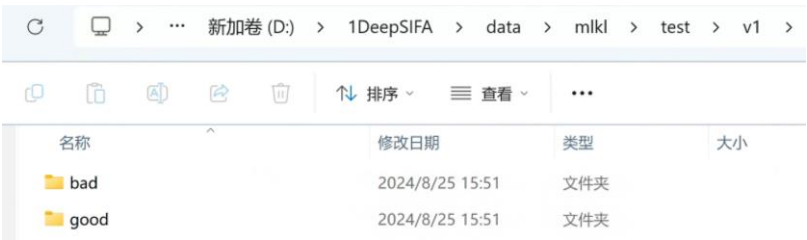
3 如果要训练的话，执行 train.py 文件就可以

```
Terminal Help
train.py
1 import os, argparse
2 os.environ['CUDA_VISIBLE_DEVICES'] = '0'
3 import logging
4 from tqdm import tqdm
5 import sys
6 import gc
7 sys.path.insert(0, os.path.join(os.path.dirname(__file__), '..'))
8 import torch as nn
9 import torch.utils.data
10 from torch.utils.tensorboard import SummaryWriter
11 import time
12 import matplotlib
13 matplotlib.use('Agg')
14 import matplotlib.pyplot as plt
15 from utils.xrayloader import XrayDataset_train, XrayDataset_val
16 from timm.scheduler import create_scheduler
17 from timm.optim import create_optimizer_v2, optimizer_kwargs
18 from utils.metrics_xcls import *
19 from collections import OrderedDict
20 from models.vit import vit_base_patch16_224
21
22 def get_cfg(fold):
23     parser = argparse.ArgumentParser()
24     parser.add_argument('--model_name', type=str, default='LSTM')
25     parser.add_argument('--n_epochs', type=int, default=100)
26     parser.add_argument('--bf_size', type=int, default=128)
27     parser.add_argument('--lr', type=float, default=1e-4, metavar='LR',
28                         help='learning rate (default: 1e-4)')
29     parser.add_argument('--warmup-lr', type=float, default=1e-6, metavar='LR',
30                         help='warmup learning rate (default: 1e-6)')
31     parser.add_argument('--min-lr', type=float, default=1e-6, metavar='LR',
32                         help='lower lr bound for cyclic schedulers that hit 0 (1e-5)')
33     parser.add_argument('--weight-decay', type=float, default=0.05,
34                         help='weight decay (default: 0.05)')
35
36     parser.add_argument('--data_dir', type=str, default='D:/1DeepSIFA/data/mlkl/train/v1/归一化插值后 npz_1024_高斯平滑1')
37     parser.add_argument('--train_csv_dir', type=str, default='D:/1DeepSIFA/data/mlkl/train/v1/折交叉验证_42/train_fold{}_多标签.csv'.format(fold))
38     parser.add_argument('--val_csv_dir', type=str, default='D:/1DeepSIFA/data/mlkl/train/v1/折交叉验证_42/val_fold{}_多标签.csv'.format(fold))
39     parser.add_argument('--weight_path', type=str, default='./model_hub/')
```

data_dir 设置为	'D:/1DeepSIFA/data/mlkl/train/v1/归一化插值后 npz_1024_高斯平滑 1'
train_csv_dir 设置为	'D:/1DeepSIFA/data/mlkl/train/v1/5 折交叉验证_42/train_fold{}_多标签.csv'.format(fold)
val_csv_dir 设置为	'D:/1DeepSIFA/data/mlkl/train/v1/5 折交叉验证_42/val_fold{}_多标签.csv'.format(fold)

### 3 测试集数据准备

1 创建目录



2 把好曲线放入 good 文件夹，把坏曲线放入 bad 文件夹

3 执行脚本文件 生成测试集数据和 csv.py

```
生成训练集数据和csv.py 生成测试集数据和csv.py ×
生成测试集数据和csv.py > ...
1 import os
2 import shutil
3 import pandas as pd
4 import numpy as np
5 import re
6 from scipy.ndimage import gaussian_filter1d
7 from scipy.interpolate import interp1d
8 import matplotlib.pyplot as plt
9 from tqdm import tqdm
10 import matplotlib
11 import gc
12
13 DATA = 'mlkl'
14 NUM = '1'
15 NUM_dimension = 1024
16 SIGMA = 1
17
```

超参数		
DATA	目前支持 2 种数据 mlkl alphak10	训练集是 mlkl 就写入 mlkl 训练集是 alphak10 就写入 alphak10
NUM	训练集第一版本就 NUM=1， 如果训练集有第二种版本就 NUM=2，训练集还有第三种 版本就 NUM=3。	目前不需要修改
NUM_dimension	是一条曲线有多少个数据 点，原始数据有时候是 500 个点，有时候是 1000 个点， 那就统一插值到 1024 个点。	目前不需要修改
SIGMA	高斯平滑的超参数设置，选 为 1 是比较好的，可以平滑 曲线去掉部分噪声。	目前不需要需改

4 测试集结构

执行 **生成测试集数据和 csv.py** 脚本文件之后, 在 D:\1DeepSIFA\data\mlkl\test\v1 目录文件下就会生成如下结构的文件, 3 个文件夹和一个 csv 文件。

🔄

🖥️

>

...

新加卷 (D:)

>

1DeepSIFA

>

data

>

mlkl

>

test

>

v1

>

📁

📄

🔍

🔗

🗑️

↕️ 排序 ▾

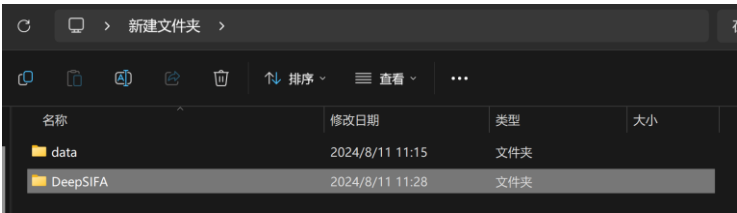
☰ 查看 ▾

...

名称	修改日期	类型	大小
📁 归一化插值后npz_1024_高斯平滑1	2024/8/25 15:59	文件夹	
📁 归一化后npz	2024/8/25 15:59	文件夹	
📁 原始数据	2024/8/25 15:59	文件夹	
📄 337.csv	2024/8/25 15:59	Comma Separat...	8 KB

## 4 验证测试集

1 把 DeepSIFA 文件放到这个目录下



2 DeepSIFA 包含文件如下

名称	修改日期	类型	大小
model_hub	2024/8/11 16:40	文件夹	
models	2024/8/24 10:58	文件夹	
utils	2024/8/24 10:58	文件夹	
其他功能	2024/8/24 10:58	文件夹	
111.py	2024/8/25 13:57	Python 源文件	3 KB
eval_test.py	2024/8/25 11:32	Python 源文件	18 KB
eval_train.py	2024/8/24 16:45	Python 源文件	18 KB
eval_val.py	2024/8/25 11:31	Python 源文件	18 KB
train.py	2024/8/25 11:28	Python 源文件	17 KB

3 如果要验证测试集的话，执行 eval\_test.py 文件就可以

```
train.py  eval_test.py x
DeepSIFA > eval_test.py > ...
1 import os, argparse, math
2 os.environ['CUDA_VISIBLE_DEVICES'] = '0'
3 import logging
4 import numpy as np
5 from tqdm import tqdm
6 import sys
7 import gc
8 sys.path.insert(0, os.path.join(os.path.dirname(__file__), '..'))
9 import torch.nn as nn
10 import torch.utils.data
11 from scipy import stats
12 from torch.utils.tensorboard import SummaryWriter
13 import time
14
15 from utils.xrayloader import XrayDataset_val
16 from utils.metrics_2cls import *
17 from collections import OrderedDict
18 import csv
19 import shutil
20
21
22 def get_cfg(fold):
23     parser = argparse.ArgumentParser()
24     parser.add_argument('--fold', type=str)
25     parser.add_argument('--n_epochs', type=int, default=1)
26     parser.add_argument('--bt_size', type=int, default=1)
27
28     # github上路径
29     parser.add_argument('--weight_path', type=str, default='./checkpoints/fold{}/best_acc.pth'.format(fold))
30     parser.add_argument('--data_dir', type=str, default='/home/node01/linchen/data/F_修改后/第五批/归一化插值后 npz_1024_高斯平滑1')
31     parser.add_argument('--val_csv_dir', type=str, default='/home/node01/linchen/data/F_修改后/第五批/162.csv')
32     parser.add_argument('--results_dir', default='./logs_test_5/', type=str, metavar='FILENAME',
33                          help='Output csv file for validation results (summary)')
```

data_dir 设置为	'D:/1DeepSIFA/data/mlkl/test/v1/归一化插值后 npz_1024_高斯平滑 1'
val_csv_dir 设置为	'D:/1DeepSIFA/data/mlkl/test/v1/337.csv'.format(fold)