**小样本学习-迁移学习-大鼠/人MRI图像分类任务**

1）前置基础知识：

迁移学习、机器学习

CDAN、sklearn下的各种分类器

CDAN训练的pipeline：

源域 S 目标域 T 两个域特征分布有差异；（很常见，源域的数据可能来自于一种环境，而目标域的数据来自于一个完全不同的环境）

1）预训练：在源域（s）上预训练一个深度神经网络，有监督

2）特征提取：使用预训练的网络来提取源域和目标域数据的特征。这些特征是后续适应步骤的基础。

3）领域适应：DAN的核心在于通过减少源域和目标域特征的分布差异来实现领域适应。这通常通过一种称为最大均值差异（Maximum Mean Discrepancy, MMD）的技术来实现。MMD是一种测量两个不同数据分布差异的方法。在DAN中，通过最小化源域和目标域特征的MMD，来使得网络学习到更加一般化的、不特定于任一域的特征表示。

4）微调：在适应了目标域特征之后，可能需要对网络进行微调，以更好地适应目标域的特定任务。这个步骤可能会涉及到使用少量目标域的标记数据，或者使用一些无监督或半监督的学习方法。

5）最后，在目标域的测试集上评估模型的性能。

注意：这里关键是通过对loss的调整，使得特征提取器可以提取域不变的特征，从而“欺骗”域分类器，让其无法产生正确的判断。（先让特征提取器学习，再引入域分类器，通过对抗，让特征提取器更关注域不变特征。）

2）数据集使用说明：

human、mice、human\_extra（参数信息）

其中，

mice作为源域，仅在train阶段出现，

数量：五种不同切片 1045，外加aug的90 和 Fourier(各1045)

human 作为目标域，通过五折交叉验证，分为train，valid和test

数量：五种不同切片 共445，外加aug的90 和 Fourier(各445)，实际按照人的样本为89，其中0.2，即18个样本分别为valid和test（均分）

human\_extra 为样本的对应参数信息，为分类器提供有标签学习，样本数量为89

3）模型整体构建流程：

模型由基于传统机器学习的分类器以及基于深度迁移学习的cdan构成，在读取mice、human、human\_extra数据后，会通过sklearn中的kfold进行五折交叉验证，划分train valid 和 test。其中，human\_extra中对应的参数信息数据，train和valid部分会被用来fit若干种分类器；mice以及human中的train部分会被用来进行cdan的对抗迁移学习；这里，特征提取器以及分类器模型参数的保存会根据在valid上的表现进行；最后，会通过在human的test部分，进行目标域的测试，此时，会将之前训练的分类器与已训练好的cdan中的分类模块和特征提取模块联合进行最终的预测。预测方案为：对不同分类器以及cdan中的分类模块的预测概率进行加权和后求平均，最终预测分数作为01类别的确定。

4）脚本使用说明：

首先可以通过配置json文件

"searchSeed": 1,  
"seed": 7,  
"acc": 0.0

searchSeed为1则表明启动搜索，seed 和 acc 用来保存最佳的随机种子和acc（acc仅方便打印用，正式训练的时候忽略即可）

注意：这里推荐对集成的分类器做一下处理，推荐使用3-4种类。

搜索方式：直接运行main.py即可

然后得到最佳的训练seed后，开始进行模型的训练和推理 （searchSeed为0）

训练方式：直接运行main.py即可

补充：

Trick：

1) test 集观测 (test\_save 置为1即可 )

2）acc 计算 更换阈值 调为0.4 (需要改动代码)

数据模型存储在model\_log下，每次运行main函数会执行删除

数据集正负样本比 1:2

分类器效果不错 加权平均融合 基本 在 0.7-0.8

cadn 效果通过test集上测试 也很OK 平均0.7

mcc loss 加入 需要设计权重 默认不加权 和其他loss 上两个数量级，强行使用100的beta 效果好像会变差 动态调整是最好的