



上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

## 研究生课程作业 2

### GRADUATE COURSE PROJECT REPORT

学生姓名: \_\_\_\_\_ 胡起 \_\_\_\_\_  
学生学号: \_\_\_\_\_ 121039910045 \_\_\_\_\_  
任课教师: \_\_\_\_\_ 吕宝粮 \_\_\_\_\_  
学院(系): \_\_\_\_\_ 电子信息与电气工程学院 \_\_\_\_\_  
开课学期: \_\_\_\_\_ 2022 年 (春季) \_\_\_\_\_

# 1 域对抗网络 (DANN)

## 1.1 DANN 思想

域适应的目的是把具有不同分布的源域 (Source Domain) 和目标域 (Target Domain) 中的数据, 映射到同一个特征空间, 寻找某一种度量准则, 使其在这个空间上的“距离”尽可能近。然后, 我们在源域 (带标签) 上训练好分类器

(Classifier), 就可以直接用于目标域数据的分类。DANN 的适应行为是通过对抗的方式增强前馈模型来实现的。

第一层是特征提取器, 如图 1.1 绿色部分所示 (source: PPT), 对应低阶的特征向高阶特征  $f$  的迁移, 即将数据映射到特定的特征空间, 使标签预测器能够分辨出来自源域数据的类别的同时, 域判别器无法区分数据来自哪个域。而标签分离器对应图 1.1 所示蓝色部分, 对来自源域的数据进行分类, 即把  $f$  作为输入, 输出标签预测。这两部分就是常规的多层神经网络。

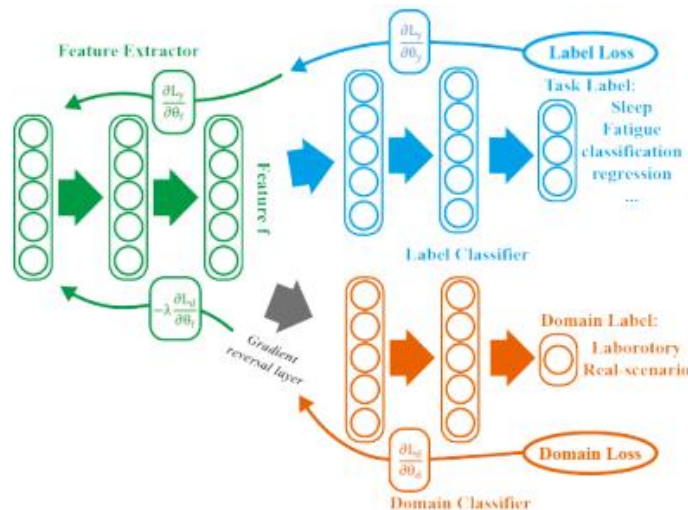


图 1.1

在特征提取器后面, 我们加上一个域判别器 (对应图 1.1 中橙色部分), 中间通过一个梯度反转层 (gradient reversal layer, GRL) 连接。在训练的过程中, 对来自源域的带标签数据, 网络不断最小化标签预测器的损失 (loss)。对来自源域和目标域的全部数据, 网络不断最小化域判别器的损失。

两个分类器都是通过反向传播来更新参数, 但是相比于传统的反向传播, Domain Classifier 采用了一种叫做梯度反转的方式, 即将梯度乘上一个负的  $\lambda$  值, 让 Domain Classifier 和 Feature Extractor 产生对抗行为, 原先的梯度下降变成了

梯度上升，这个做法导致 Feature Extractor 和 Domain Classifier 的参数会反向上升，导致越来越无法区分数据来自 Source Domain 还是 Target Domain，最终无法分辨数据来自哪个 Domain。而 Feature Extractor 和 Label Classifier 结合则采用了正常的梯度下降，当 Domain Classifier 已无法判断数据来源时，即称完成了域适应。

## 2 代码实现

(代码见.py 文件，数据见 data.pkl)

本作业中提供的数据集来自 SEED 数据集，该数据集收集自 5 名受试者，情绪标签为 1 表示积极，0 表示中性，-1 表示消极。

DANN 结构按照上文中基本思想的结构进行设置。Feature Extractor 有两层，两层的节点数都是 128，Label Classifier 和 Domain Classifier 都是三层，前两层各 64 个节点，Label Classifier 的输出层节点数为 3 个，Domain Classifier 的输出层节点 5 个，Feature Extractor、Label Classifier 和 Domain Classifier 的输出分别由 softmax 激活，隐藏层激活函数为 ReLU。根据五个受试者的源数据进行了五个批次的训练，每次训练 50 个 epoch，源代码中将输出每次 epoch 中 source domain 和 target domain 的 loss，并按照每个 batch 输出域准确率和测试准确率 (accuracy) 下文中将报告个体识别准确率和平均识别准确率。模型使用 Python 3.9.7v 和 Torch1.1v 实现。

## 3 结果分析

在图 3.1 中可以看到 5 个 batch 后域准确率可达 96%，test 准确率为 68%，增大 epoch 可能会增加域准确率和测试准确率。

```
epoch: 50, error_s_label: 2.286385, error_s_domain: 3.466798, error_t_domain: 3.429487
test acc is 0.42714159552546366 the acc_domain is 0.6552840741831027

epoch: 50, error_s_label: 2.262347, error_s_domain: 3.455400, error_t_domain: 3.414654
test acc is 0.5428319105092729 the acc_domain is 0.7344715925816897

epoch: 50, error_s_label: 2.260672, error_s_domain: 3.449156, error_t_domain: 3.408555
test acc is 0.5569620253164557 the acc_domain is 0.8737120989108037

epoch: 50, error_s_label: 2.257071, error_s_domain: 3.446409, error_t_domain: 3.405801
test acc is 0.6282013541360023 the acc_domain is 0.9290550485722696

epoch: 50, error_s_label: 2.268615, error_s_domain: 3.445282, error_t_domain: 3.404128
test acc is 0.6826611716220194 the acc_domain is 0.9634972034147777
```

图 3.1

每个 epoch 的 loss 输出如图所示，图中举例 batch\_index=5 时，图 3.2 对应的源域和目标域的 loss。

```
epoch: 48, error_s_label: 2.272233, error_s_domain: 3.446316, error_t_domain: 3.403556  
accurate rate is 0.4485  
epoch: 49, error_s_label: 2.274611, error_s_domain: 3.445990, error_t_domain: 3.403509  
accurate rate is 0.4356  
epoch: 50, error_s_label: 2.273976, error_s_domain: 3.446301, error_t_domain: 3.403719  
accurate rate is 0.4459
```

图 3.2

## 4 小结

在本次作业中，实现了一个 DANN 模型，在基于 EEG 的情感场景中进行域自适应迁移学习。该模型在 SEED 数据集上进行了训练和测试。