

工程实践与科技创新课程： 超并行机器学习与海量数据挖掘 大作业

2022 年 3 月 11 日

一、大作业背景

情感脑-机接口 (Affective Brain-Computer Interface) 研究的目的是使机器具有识别, 理解和反馈用户情绪状态的能力, 是当前情感智能 (Emotion AI) 研究的热点问题之一, 其研究内容涉及多学科交叉, 包括脑科学, 神经科学, 信息科学, 心理学和医学等领域。基于脑电的情绪识别是一个典型的模式分类问题。不同于传统的图像和语音信号, 脑电信号具有非平稳特性, 信噪比低以及个体差异性大的特点, 因此是一个很有挑战的研究课题。本作业从实际应用的角度出发, 需要同学们利用迁移学习相关算法构建具有跨被试能力的情绪识别模型, 利用已知多名被试的脑电情绪数据训练模型, 对未知被试的脑电数据进行情绪识别。

本作业要解决的是一个基于脑电情绪识别的三分类问题。使用的数据集是上海交大计算机系类脑计算与机器智能研究中心 (BCMI) 的脑电情感数据集 (SJTU Emotion EEG Dataset, SEED)¹ 的一个子集。在实验中, 我们选取了 3 类情绪 (高兴、悲伤、中性) 共 15 段电影片段作为情绪诱发刺激素材。每段视频时长约三分钟, 每类情绪有 5 段电影片段。15 名被试每名被试均在实验过程中观看 15 段视频素材, 以诱发相应情绪。

选择该问题的目的是希望通过求解该问题, 让同学们对当前情感脑-机接口研究的问题有初步认识, 加深对相关神经网络模型和机器学习算法的理解。

二、大作业数据描述

本次大作业使用的是 SEED 数据集的子集, 包括 15 名被试, 每个被试 1 次实验, 共 15 组实验脑电数据。脑电特征为微分熵 differential entropy (DE) 特征, 共 62 个导联, 特征提取时间窗口为 1 秒, 特征维度为 $62 \times 5 = 310$ 维。数据标号为三类情绪: 平静、悲伤和高兴, 对应 0、-1、1。

¹ <http://bcmi.sjtu.edu.cn/home/seed/index.html>

数据集：

SEED-III 文件夹下包含 ‘EEG_X.mat’ 与 ‘EEG_Y.mat’ 两个文件，分别存储 15 名被试的脑电特征与情绪类别。可使用 scipy 在 python 中加载 mat 文件如下：

```
=====
import scipy.io as sio
EEG_X = sio.loadmat('EEG_X.mat')['X'][0]
# EEG_X 包含 15 名被试的数据，使用 EEG_X[i]可访问第 i 名被试的数据
EEG_X_0 = EEG_X[0]
# EEG_X_0 为第 1 名被试的脑电数据大小为 3394*310
# 3394 表示视频总时长，310 表示脑电特征维度
# 获取对应的情绪类别方法同上
=====
```

数据获取请访问：<https://cloud.bcmi.sjtu.edu.cn/sharing/fBDlg2VqE>

三、大作业任务

15 名被试中，取 14 名被试的数据作为训练数据，剩余 1 名被试的数据作为测试数据，做 15 折交叉验证。

1、将训练集 14 名被试的数据直接进行拼接，训练一个 SVM 分类器，在剩余 1 名被试数据上进行测试，计算 15 折交叉验证的平均准确率，作为 baseline。

2. 按照上述相同的数据划分方法，自行设计迁移学习算法，以减小被试间差异，提升测试集平均准确率。

四、实验报告

1、实验报告格式需按照国际论文撰写标准，包括题目、摘要、简介、方法描述、程序设计描述（文字或伪代码）、实验数据描述、实验结果分析、结论、参考文献等。报告正文采用五号字体、页数 8-12 页，中、英文均可。

2、提交的文档中还必须包括可执行代码。

五、截止时间

2022 年 4 月 30 日