

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

2019 年 3 月 22 日

一、大作业的目的

情感脑-机交互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，是当前情感智能研究的热点问题，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很具有代表性的模式分类问题。不同于传统的图像和语音信号，脑电信号具有非平稳性，信噪比低以及个体差异性大的特点，因此是个很有挑战的研究课题。

本作业要解决的是一个基于脑电情绪识别的四分类问题。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-4）的一个子集。在实验中，我们选取了 4 类情绪（高兴、悲伤、中性和恐惧）共 24 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，每类情绪有六段电影片段，数据集包含 13 名被试的脑电数据。

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-机交互研究的热点有初步认识，加深对相关神经网络模型和学习算法的理解。

二、问题描述

本次大作业使用的是 SEED-4 数据集的子集，包括 13 个被试，每个被试 1 次实验，共 13 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 4 秒，特征维度为 $62 \times 5 = 310$ 维。数据标号为四类情绪：平静、悲伤、恐惧和高兴，对应 0、1、2、3。

训练集与测试集：

`train_de` 是训练用数据集，含 13 人每人 18 段实验数据，共包含 7293 个样本，`train_label_eeg` 是相应的四类情绪标号数据。

`test_de` 是测试用数据集，含 13 人每人 6 段实验数据，共包含 3770 个样本，`test_label_eeg` 是相应的四类情绪标号数据。

要求用实验数据中的 `train_de` 训练模型，在 `test_de` 上进行测试。

三、任务

1、使用 tensorflow 构建全连接网络，直接学习上述四分类问题。

2、用最小最大模块化网络解决上述问题，用随机方式分解原问题，每个子问题用 1 中全连接网络学习，使用多进程实现，多进程编程可参考使用 python 多进程编程包 multiprocessing。

3、用最小最大模块化网络解决上述问题，基于情绪类别先验分解原问题，每个子问题用 1 中全连接网络学习，使用多进程实现，多进程编程可参考使用 python 多进程编程包 multiprocessing。

4、给出上述分类器的分类混淆矩阵，F1 值以及测试准确率随迭代次数变化的曲线，并说明哪种分类器分类性能最好？四类情绪识别精度如何？哪类情绪容易区分？哪类情绪不容易区分？

5、比较全连接网络与最小最大模块化全连接网络的训练时间和分类时间。

6、重点加分项（可选）：利用最新的深度神经网络如 CNN，RNN，DBN，LSTM，GAN 等解决该问题，并系统讨论和比较深度网络与传统浅层模型。如果有比较好的结果和发现，可以联系助教。

四、实验报告

1、实验报告格式：题目、摘要、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。报告正文五号字体、10 页，中、英文均可。

2、提交的文档中还必须包括程序源代码和执行代码。

五、截止时间

2019 年 04 月 30 日