2022脑机接口算法挑战赛：脑纹识别

——竞赛总结

1. 赛题分析

多维时间序列的分类和异常检测问题。通过脑电进行身份认证，实现ID预测和入侵者检测。两个实验批次，第一轮实验95个人，第二轮实验106个人（其中11个人是新增的入侵者）。

注册集： 95个人第一次实验的EEG数据

供参赛者提取/训练个体特质的模板/分类器。

校准集：20个人第二次实验的EEG数据

让参与者观察脑电图数据，分析不同受试者、不同天、不同范式之间的共性和区别，并对不同预测模型进行离线比较。

校准集中仅包含了这20名受试者第二次实验的部分EEG数据，另外有一部分部分EEG被放在测试集中，且未给出受试者ID。

测试集：106个人第二次实验的EEG数据

A榜：约占测试集总试次数的10%

B榜：约占测试集总试次数的90%

其中11个被试是未在注册集中出现的入侵者。

哪些试次属于A榜或B榜在测试集中并未给出。

数据：[batch\_size, 65, 1000]

第65个通道的数据是任务编码。

1. 官方baseline

卷积，haar小波分解，卷积，卷积，卷积，全连接

1. 整体研究思路

脑电领域目前并没有像CV和NLP那样成熟的解决方案，而且数据采集的质量和数量也尚处于类比CV的90年代的情况。得益于深度学习在近几年的快速发展，脑电算法建模也在快速发展。

基于脑电数据的特殊性，以及缺少相关专业知识，所以采用传统机器学习手动构造特征会比较困难。在搜索了相关文献之后，发现脑电领域的论文大多没有提供代码，而近几年深度学习在多维时间序列方向的研究也才刚刚兴起，所以脑电这个细分领域的基于深度学习的研究也比较有限。在缺少类似于ImageNet这样具有公信力的数据集以及对应排行榜参考的情况下，决定从基础的模型开始尝试。

模型结构是比较重要的，对于算法的预测效果有很大的影响，所以主要在探索模型结构上进行了许多尝试。将数字信号处理和深度学习相结合也是一个比较重要的思路。另外由于样本量很少，所以数据增强也很重要。

1. 模型提升关键点
2. 使用CNN构建基础的模型是一个简单有效的方法。第一名的基础模型也是一个6层的CNN。
3. 使用带通滤波并选择适当的阶数可以大幅提升验证集的准确率。
4. 剔除高频噪声会有显著提升。
5. 使用全局平均池化会有显著提升。
6. 由于参赛时间较短且算力有限，数据增强并未来得及尝试，但根据颁奖会上获奖团队的方案分享，可以带来显著的提升。
7. 不足

因为时间关系（仅参赛不到20天，只使用了百度飞桨的免费算力，且一天只能训练出一两个版本），许多想法并未来得及尝试：

1. 探索更多的模型结构，比如参考今年发表的ConvNeXt模型设计针对脑电时序数据的1DCNN模型。
2. 尝试多种时序数据增强方法。

由于批次效应，导致线上和线下的准确率差异很大，第一名采用了伪标签的方法取得了较大的提升。虽然在技术层面可以做出一些提升，但是根本的解决办法还是提升数据量以及探索时间衰减较弱的刺激方法，这也是未来的研究中推动产品实际落地应用时所要解决的重要问题。

六、总结

总的来说，脑机接口的前景让人充满期待，无论是硬件设备的研发还是算法的研发都在快速发展。但参考CV领域的发展历程，脑电算法还有很长的路要走。数据量的提升是最容易也是最迫切的。从算法角度来看，GCN可能会成为一个重要的方向，但其实纯CNN也能学习出不同脑电通道间的交互关系。