

EEG数据处理的基本思路 (V0.1)

目前，针对SSVEP-BCI数据，一般有两种类型：1) 公开数据集的数据。2) 自己做实验采集的数据。清华大学公开数据集的数据作为目前SSVEP算法的benchmark被广泛应用。而自己做实验采集的数据，经过处理后，可以得到与公开数据集数据存储格式相同的结果。这里，首先介绍两种数据存储类型；然后，介绍数据处理的一般过程。

1 两种数据存储类型

1.1 公开数据集

SSVEP的实验一般包括若干个轮次 (Block)，每一个轮次又包含若干个试次 (Trial)，习惯而言，一个轮次中包含的试次数与刺激目标的个数相等，相当于遍历一遍所有的刺激。对于一个试次，一般包含 n 个导联和 m 个采样点的数据片段，以数据矩阵形式呈现，形如 $n \times m$ 。如果考虑不同 Trial 和 Block，则数据会变为三维或四维的矩阵，形如：

$$DATA = N_{chan}, N_{sample}, N_{trial}, N_{block}$$

如此，通过索引矩阵，即可得到任意试次的数据。

1.2 自己实验得到的数据

自己实验得到的数据跟据实验设备不同，可能会有不同的数据格式，如Neuroscan的格式为.cnt，博睿康设备的格式为.bdf。一般使用基于matlab的工具包EEGLab对数据进行读取，注意EEGLab中有不同的数据读取插件，以解析不同格式的数据。解析后的数据在matlab中被存为一个结构体的matlab数据包中，包括以下几个重要的字段：

data: 存储数据点信息，大小为 N_{chan}, N_{sample}

event: 存储同步标签信息：与实验过程中打标签的数量一致，每一行形如 '标签值': 采样点位置

srate: 采样率（注意，低频SSVEP实验时采集数据采样率一般为1000Hz，处理过程中要降采样到250Hz。）

*通过event里的标签信息，和已知的单个试次的长度，可截取每个试次的信号，并按照block不同，存储为与公开数据集相同的数据格式。

2 脑电数据处理的一般流程：

脑电信号输入微弱电信号，且由于背景脑电的存在，脑电信号的信噪比较低。与其他信号的处理过程一致，脑电信号的处理一般也遵循以下五个步骤：

- 降采样
- 50Hz陷波（如果后续时域滤波频带范围窄，通带上限小于50Hz，则不一定要进行陷波）
- 时域滤波
- 空间滤波（特征提取，我们常用的TRCA、CCA的训练出的向量均为空间滤波器）
- 分类（多采用皮尔逊相关后利用 $\text{argmax}()$ ）

先讨论前三个步骤：

2.1 降采样

为减少数据量，避免部分信号细节影响特征提取的过程，需要进行降采样操作，matlab中，降采样一般用EEGLab的 `pop_resample()` 实现，而python中，降采样一般用Scipy的 `resample()` 函数实现。

2.2 50Hz陷波

在需要利用高频频带的SSVEP信号时（如使用滤波器组相关算法时），**一定要考虑50Hz陷波的问题**，即使是采集设备已经用进行过50Hz陷波，仍要考虑加入。比较好用的陷波器为 `iirnotch()`。

2.3 带通滤波器

带通滤波器的设计至关重要，从经验来看，切比雪夫滤波器的效果要优于巴特沃斯滤波器。设计出滤波器后，尽量查看滤波器的幅频响应曲线，以确定截止频率处是否存在抖动的情况。

* 注意，设计滤波器请小心采样率的设置，避免出现信号采样率和滤波器设计采样率不一致的情况。