BCI Competition 2008 – Graz data set A

C. Brunner[1], R. Leeb[1], G. R. Müller-Putz[1], A. Schlögl[2], and G. Pfurtscheller[1]

[1]Institute for Knowledge Discovery, Graz University of Technology, Austria
[2]Institute for Human-Computer Interfaces, Graz University of Technology, Austria

Experimental paradigm

This data set consists of EEG data from 9 subjects. The cue-based BCI paradigm consisted of four different motor imagery tasks, namely the imagination of movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). Two sessions on different days were recorded for each subject. Each session is comprised of 6 runs separated by short breaks. One run consists of 48 trials (12 for each of the four possible classes), yielding a total of 288 trials per session.

At the beginning of each session, a recording of approximately 5 minutes was performed to estimate the EOG influence. The recording was divided into 3 blocks: (1) two minutes with eyes open (looking at a fixation cross on the screen), (2) one minute with eyes closed, and (3) one minute with eye movements. The timing scheme of one session is illustrated in Figure 1. Note that due to technical problems the EOG block is shorter for subject `A04T` and contains only the eye movement condition (see Table 1 for a list of all subjects).

The subjects were sitting in a comfortable armchair in front of a computer screen. At the beginning of a trial ($t = 0\,\mathrm{s}$), a fixation cross appeared on the black screen. In addition, a short acoustic warning tone was presented. After two seconds ($t = 2\,\mathrm{s}$), a cue in the form of an arrow pointing either to the left, right, down or up (corresponding to one of the four classes

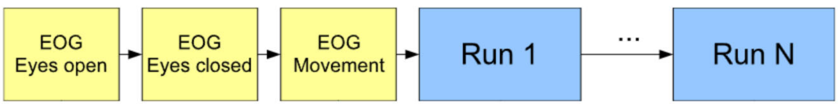

Figure 1: Timing scheme of one session.

2008 年 BCI 竞赛 – 格拉茨数据集 A

C. Brunner、R. Leeb、GR M¨uller-Putz、A. Schl¨ogl 和 G. Pfurtscheller

[1] 奥地利格拉茨技术大学知识发现研究所

[2] 奥地利格拉茨技术大学人机界面研究所

实验范式

该数据集包含来自 9 位受试者的脑电图数据。基于线索的脑机接口 (BCI) 范式包含四种不同的运动想象任务，分别是想象左手（类别 1）、右手（类别 2）、双脚（类别 3）和舌头（类别 4）的运动。每位受试者在不同日期记录两次训练。每次训练包含 6 次运行，中间间隔短暂休息。每次运行包含 48 次试次（四个可能的类别各 12 次），因此每次训练总共包含 288 次试次。

每次实验开始时，进行约 5 分钟的眼电图（EOG）记录，以评估眼电图（EOG）的影响。记录分为 3 个区块：（1）睁眼 2 分钟（注视屏幕上的注视十字）；（2）闭眼 1 分钟；（3）眼动 1 分钟。一个实验的时序如图 1 所示。需要注意的是，由于技术原因，受试者 A04T 的眼电图（EOG）区块较短，且仅包含眼动条件（所有受试者的列表见表 1）。

受试者坐在舒适的扶手椅上，面前是电脑屏幕。在试验开始时（t = 0 秒），黑色屏幕上会出现一个注视十字。此外，还会播放一段简短的警告音。两秒后（t = 2 秒），屏幕上会显示一个箭头提示，该箭头指向左、右、下或上（分别对应四个类别之一）。

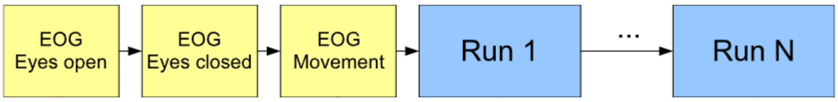

图 1：一个会话的时间安排。

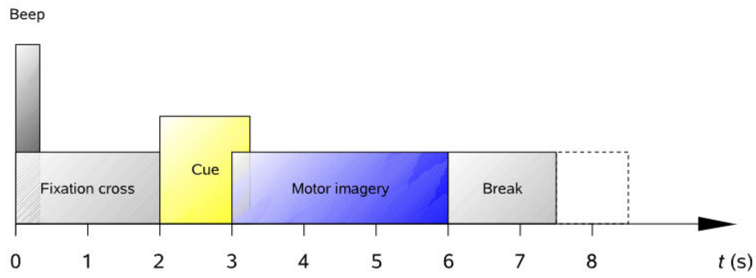

Figure 2: Timing scheme of the paradigm.

left hand, right hand, foot or tongue) appeared and stayed on the screen for 1.25 s. This prompted the subjects to perform the desired motor imagery task. No feedback was provided. The subjects were ask to carry out the motor imagery task until the fixation cross disappeared from the screen at $t = 6$ s. A short break followed where the screen was black again. The paradigm is illustrated in Figure 2.

Data recording

Twenty-two Ag/AgCl electrodes (with inter-electrode distances of 3.5 cm) were used to record the EEG; the montage is shown in Figure 3 left. All signals were recorded monopolarly with the left mastoid serving as reference and the right mastoid as ground. The signals were sampled with 250 Hz and bandpass-filtered between 0.5 Hz and 100 Hz. The sensitivity of the amplifier was set to 100 µV. An additional 50 Hz notch filter was enabled to suppress line noise.

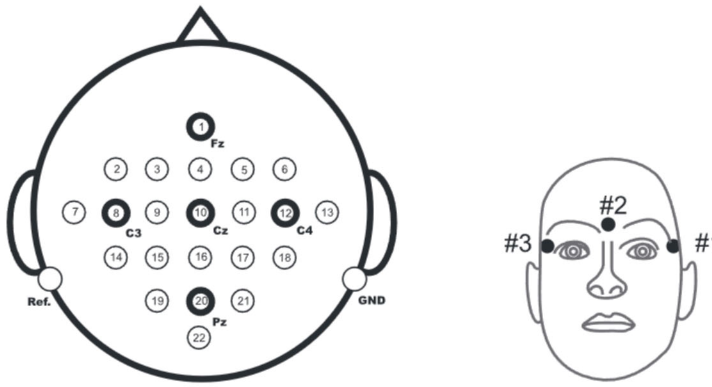

Figure 3: Left: Electrode montage corresponding to the international 10-20 system. Right: Electrode montage of the three monopolar EOG channels.

In addition to the 22 EEG channels, 3 monopolar EOG channels were

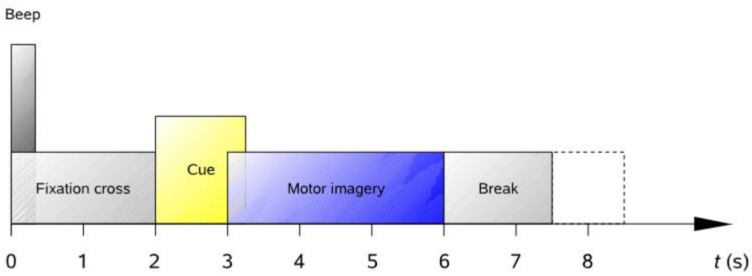

图 2：范式的时间安排。

左手、右手、脚或舌头）的注视点出现在屏幕上并停留 1.25 秒。这促使受试者执行所需的运动想象任务。没有提供任何反馈。受试者被要求执行运动想象任务，直到注视点十字在 t = 6 秒时从屏幕上消失。随后短暂休息，屏幕再次变黑。该范例如图 2 所示。

数据记录

使用 22 个 Ag/AgCl 电极（电极间距 3.5 cm）记录脑电图；蒙太奇图如图 3 左图所示。所有信号均以单极记录，以左乳突为参考点，右乳突为接地点。信号采样率为 250 Hz，并在 0.5 Hz 至 100 Hz 之间进行带通滤波。放大器灵敏度设置为 100 µV。启用一个 50 Hz 陷波滤波器以抑制线路噪声。

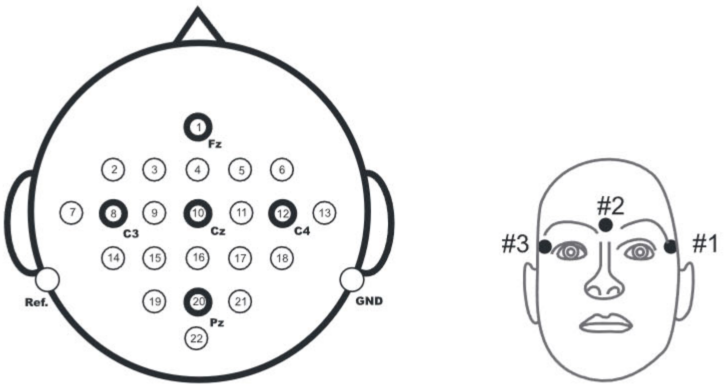

图 3：左图：符合国际 10-20 系统的电极组图。右图：三个单极眼电图通道的电极组图。

除了 22 个 EEG 通道外，还增加了 3 个单极 EOG 通道

recorded and also sampled with 250 Hz (see Figure 3 right). They were bandpass filtered between 0.5 Hz and 100 Hz (with the 50 Hz notch filter enabled), and the sensitivity of the amplifier was set to 1 mV. The EOG channels are provided for the subsequent application of artifact processing methods [1] and must not be used for classification.

A visual inspection of all data sets was carried out by an expert and trials containing artifacts were marked. Eight out of the total of nine data sets were analyzed in [2, 3].

Data file description

All data sets are stored in the General Data Format for biomedical signals (GDF), one file per subject and session. However, only one session contains the class labels for all trials, whereas the other session will be used to test the classifier and hence to evaluate the performance. All files are listed in Table 1. Note that the evaluation sets will be made available after the deadline of the competition (except for one file from subject `A01` which serves as an example). The GDF files can be loaded using the open-source toolbox BioSig, available for free at http://biosig.sourceforge.net/. There are versions for Octave[1]/FreeMat[2]/MATLAB[3] as well as a library for C/C++.

ID	Training	Evaluation
1	A01T.gdf	A01E.gdf
2	A02T.gdf	~~A02E.gdf~~
3	A03T.gdf	~~A03E.gdf~~
4	A04T.gdf	~~A04E.gdf~~
5	A05T.gdf	~~A05E.gdf~~
6	A06T.gdf	~~A06E.gdf~~
7	A07T.gdf	~~A07E.gdf~~
8	A08T.gdf	~~A08E.gdf~~
9	A09T.gdf	~~A09E.gdf~~

Table 1: List of all files contained in the data set, the striked out evaluation data sets will be provided after the deadline of the competition. Note that due to technical problems the EOG block is shorter for subject `A04T` and contains only the eye movement condition.

A GDF file can be loaded with the BioSig toolbox with the following command in Octave/FreeMat/MATLAB (for C/C++, the corresponding function `HDRTYPE* sopen` and `size_t sread` must be called):

```
[s, h] = sload('A01T.gdf');
```

[1] http://www.gnu.org/software/octave/
[2] http://freemat.sourceforge.net/
[3] The MathWorks, Inc., Natick, MA, USA

记录并以 250 Hz 的频率采样（见图 3 右图）。这些通道在 0.5 Hz 至 100 Hz 之间进行带通滤波（启用 50 Hz 陷波滤波器），放大器灵敏度设置为 1 mV。眼电图通道用于后续伪影处理方法 [1] 的应用，不得用于分类。

所有数据集均由专家进行目视检查，并标记包含伪影的试验。在[2, 3]中分析了 9 个数据集中的 8 个。

数据文件描述

所有数据集均以生物医学信号通用数据格式 (GDF) 存储，每个受试者和每个会话对应一个文件。但只有一个会话包含所有试验的类别标签，而另一个会话将用于测试分类器，从而评估性能。所有文件列于表 1。请注意，评估集将在比赛截止日期后开放（除受试者 A01 的一个文件外，该文件用作示例）。GDF 文件可以使用开源工具箱 BioSig 加载，该工具箱可在 http://biosig.sourceforge.net/ 免费获取。BioSig 提供 Octave/FreeMat/MATLAB 版本以及 C/C++ 库。

ID	训练	评估
1	A01T.gdf	A01E.gdf
2	A02T.gdf	A02E.gdf
3	A03T.gdf	A03E.gdf
4	A04T.gdf	A04E.gdf
5	A05T.gdf	A05E.gdf
6	A06T.gdf	A06E.gdf
7	A07T.gdf	A07E.gdf
8	A08T.gdf	A08E.gdf
9	A09T.gdf	A09E.gdf

表 1：数据集包含的所有文件列表，划掉的评估数据集将在比赛截止日期后提供。

请注意，由于技术问题，受试者 A04T 的眼电图 (EOG) 数据块较短，且仅包含眼动状态。

可以使用 Octave/FreeMat/MATLAB 中的以下命令通过 BioSig 工具箱加载 GDF 文件（对于 C/C++，必须调用相应的函数 HDRTYPE* soopen 和 size t sread）：

```
[s，h] = sload（'A01T.gdf'）；
```

[1] http://www.gnu.org/software/octave/

[2] http://freemat.sourceforge.net/

[3] MathWorks 公司，美国马萨诸塞州纳蒂克

Event type		Description
276	0x0114	Idling EEG (eyes open)
277	0x0115	Idling EEG (eyes closed)
768	0x0300	Start of a trial
769	0x0301	Cue onset left (class 1)
770	0x0302	Cue onset right (class 2)
771	0x0303	Cue onset foot (class 3)
772	0x0304	Cue onset tongue (class 4)
783	0x030F	Cue unknown
1023	0x03FF	Rejected trial
1072	0x0430	Eye movements
32766	0x7FFE	Start of a new run

Table 2: List of event types (the first column contains decimal values and the second hexadecimal values).

Note that the runs are separated by 100 missing values, which are encoded as not-a-numbers (NaN) by default. Alternatively, this behavior can be turned off and the missing values will be encoded as the negative maximum values as stored in the file with:

```
[s, h] = sload('A01T.gdf', 0, 'OVERFLOWDETECTION:OFF');
```

The workspace will then contain two variables, namely the signals `s` and a header structure `h`. The signal variable contains 25 channels (the first 22 are EEG and the last 3 are EOG signals). The header structure contains event information that describes the structure of the data over time. The following fields provide important information for the evaluation of this data set:

```
h.EVENT.TYP
h.EVENT.POS
h.EVENT.DUR
```

The position of an event in samples is contained in `h.EVENT.POS`. The corresponding type can be found in `h.EVENT.TYP`, and the duration of that particular event is stored in `h.EVENT.DUR`. The types used in this data set are described in Table 2 (hexadecimal values, decimal notation in parentheses). Note that the class labels (i. e., 1, 2, 3, 4 corresponding to event types 769, 770, 771, 772) are only provided for the training data and not for the testing data.

The trials containing artifacts as scored by experts are marked as events with the type 1023. In addition, `h.ArtifactSelection` contains a list of all trials, with `0` corresponding to a clean trial and `1` corresponding to a trial containing an artifact.

事件类型		描述
276	0x0114	空转脑电图（睁眼）
277	0x0115	空转脑电图（闭眼）
768	0x0300	开始试用
769	0x0301	提示起始于左侧（第 1 类）
770	0x0302	提示开始于右（类 2）
771	0x0303	提示起始脚（第 3 类）
772	0x0304	提示起始舌（第 4 类）
783	0x030F	提示未知
1023	0x03FF	拒绝试用
1072	0x0430	眼球运动
32766	0x7FFE	开始新运行

表 2：事件类型列表（第一列包含十进制值，第二列包含十六进制值）。

请注意，每次运行之间间隔 100 个缺失值，这些值默认编码为非数字 (NaN)。或者，也可以关闭此行为，将缺失值编码为文件中存储的负最大值：

```
[s，h] = sload（'A01T.gdf'，0，'溢出检测：关闭'）；
```

工作区将包含两个变量，即信号 s 和标头结构 h。信号变量包含 25 个通道（前 22 个为脑电图 (EEG) 信号，后 3 个为眼电图 (EOG) 信号）。标头结构包含事件信息，用于描述数据随时间的结构。以下字段为评估此数据集提供了重要信息：

事件类型
事件位置
事件持续时间

事件在样本中的位置存储在 h.EVENT.POS 中。相应的类型存储在 h.EVENT.TYP 中，特定事件的持续时间存储在 h.EVENT.DUR 中。此数据集中使用的类型如表 2 所示（十六进制值，括号中为十进制表示法）。请注意，类标签（即 1、2、3、4 分别对应事件类型 769、770、771、772）仅适用于训练数据，不适用于测试数据。包含专家评分的伪影的试验被标记为类型为 1023 的事件。此外，h.ArtifactSelection 包含所有试验的列表，其中 0 对应干净试验，1 对应包含伪影的试验。

In order to view the GDF files, the viewing and scoring application SigViewer v0.2 or higher (part of BioSig) can be used.

Evaluation

Participants should provide a continuous classification output for each sample in the form of class labels (1, 2, 3, 4), including labeled trials and trials marked as artifact. A confusion matrix will then be built from all artifact-free trials for each time point. From these confusion matrices, the time course of the accuracy as well as the kappa coefficient will be obtained [5]. The algorithm used for this evaluation will be provided in BioSig. The winner is the algorithm with the largest kappa value `X.KAP00`.

Due to the fact that the evaluation data sets will not be distributed until the end of the competition, the submissions must be programs that accept EEG data (the structure of this data must be the same as used in all training sets[4]) as input and produce the aforementioned class label vector.

Since three EOG channels are provided, it is required to remove EOG artifacts before the subsequent data processing using artifact removal techniques such as highpass filtering or linear regression [4]. In order to enable the application of other correction methods, we have opted for a maximum transparency approach and provided the EOG channels; at the same time we request that artifacts do not influence the classification result.

All algorithms must be causal, meaning that the classification output at time k may only depend on the current and past samples $x_k, x_{k-1}, \ldots, x_0$. In order to check whether the causality criterion and the artifact processing requirements are fulfilled, all submissions must be open source, including all additional libraries, compilers, programming languages, and so on (for example, Octave/FreeMat, C++, Python, ...). Note that submissions can also be written in the closed-source development environment MATLAB as long as the code is executable in Octave. Similarily, C++ programs can be written and compiled with a Microsoft or Intel compiler, but the code must also compile with g++.

References

[1] M. Fatourechi, A. Bashashati, R. K. Ward, G. E. Birch. EMG and EOG artifacts in brain computer interface systems: a survey. Clinical Neurophysiology 118, 480–494, 2007.

[4] One evaluation data set is distributed from the beginning of the competition to enable participants to test their program and to ensure that it produces the desired output.

为了查看 GDF 文件，可以使用查看和评分应用程序 SigViewer v0.2 或更高版本（BioSig 的一部分）。

评估

参赛者应为每个样本提供以类别标签 (1, 2, 3, 4) 形式呈现的连续分类输出，包括已标记的试验和标记为伪影的试验。然后，我们将根据每个时间点的所有无伪影试验构建一个混淆矩阵。通过这些混淆矩阵，我们将获得准确率随时间的变化以及 Kappa 系数 [5]。本次评估所用的算法将在 BioSig 中提供。Kappa 值最大的算法 X.KAP00 将获胜。

由于评估数据集要到比赛结束才会分发，因此提交的程序必须是接受 EEG 数据（该数据的结构必须与所有训练集中使用的相同）作为输入并生成上述类标签向量的程序。

由于提供了三个眼电图通道，因此需要在后续数据处理之前使用高通滤波或线性回归等伪影消除技术去除眼电图伪影[4]。为了能够应用其他校正方法，我们选择了最大透明度方法，并在提供眼电图通道的同时，要求伪影不会影响分类结果。

所有算法都必须具有因果关系，这意味着时间 k 的分类输出可能仅取决于当前和过去的样本 x, x, . . . , x。为了检查因果关系标准和工件处理要求是否得到满足，所有提交的内容都必须是开源的，包括所有附加库、编译器、编程语言等（例如，Octave/FreeMat、C++、Python、. . . ）。请注意，只要代码可以在 Octave 中执行，提交的内容也可以在闭源开发环境 MATLAB 中编写。类似地，可以使用 Microsoft 或 Intel 编译器编写和编译 C++ 程序，但代码也必须使用 g++ 进行编译。

参考

[1] M. Fatourechi, A. Bashashati, RK Ward, GE Birch. 脑机接口系统中的肌电图（EMG）和眼电图（EOG）伪影：一项调查. 临床神经生理学 118, 480–494, 2007.

从比赛开始就分发一个评估数据集，以便参与者测试他们的程序并确保它产生所需的输出。

[2] M. Naeem, C. Brunner, R. Leeb, B. Graimann, G. Pfurtscheller. Seperability of four-class motor imagery data using independent components analysis. Journal of Neural Engineering 3, 208–216, 2006.

[3] C. Brunner, M. Naeem, R. Leeb, B. Graimann, G. Pfurtscheller. Spatial filtering and selection of optimized components in four class motor imagery data using independent components analysis. Pattern Recognition Letters 28, 957–964, 2007.

[4] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, G. Pfurtscheller. A fully automated correction method of EOG artifacts in EEG recordings. Clinical Neurophysiology 118, 98–104, 2007.

[5] A. Schlögl, J. Kronegg, J. E. Huggins, S. G. Mason. Evaluation criteria in BCI research. In: G. Dornhege, J. del R. Millán, T. Hinterberger, D. J. McFarland, K.-R. Müller (Eds.). Toward brain-computer interfacing, MIT Press, 327–342, 2007.

[2] M. Naeem, C. Brunner, R. Leeb, B. Graimann, G. Pfurtscheller. 基于独立成分分析的四类运动想象数据分离性研究。《神经工程杂志》3, 208–216, 2006。
[3] C. Brunner, M. Naeem, R. Leeb, B. Graimann, G. Pfurtscheller. 基于独立成分分析的四类运动图像数据的空间滤波及优化成分选择。《模式识别快报》28, 957–964, 2007。
[4] A. Schlöogl、C. Keinrath、D. Zimmermann、R. Scherer、R. Leeb、G.

Pfurtscheller. 一种全自动校正脑电图记录中眼电图伪影的方法。《临床神经生理学》118, 98–104, 2007。

[5] A. Schl¨ogl, J. Kronegg, JE Huggins, SG Mason. 脑机接口研究中的评估标准。收录于：G. Dornhege, J. del R. Mill´an, T. Hinterberger, DJ McFarland, K.-R. M¨uller (编)。《迈向脑机接口》，麻省理工学院出版社，327–342，2007 年。