**75.EEG Database**

1. 数据库网址

https://archive.ics.uci.edu/ml/datasets/EEG+Database

2. 数据库描述

【1.[数据集名称]数据集由[机构名或人名]采集；】The data used in our experiments were collected by E. Alpaydin, C. Kaynak, from Department of Computer Engineering,Bogazici University at July,1998.【2.用于[什么实验目的]】We used preprocessing programs made available by NIST to extract normalized bitmaps of handwritten digits from a preprinted form.【3】

【4】The database has 5620 samples, respectively belong to optdigits.tra with 3823 samples and optidigits.tes with 1797 samples. The categories of network system include seven categories, as shown in Table 1.

Table 1 Category Distribution of Network System [根据数据库绘制]

|  |  |  |  |
| --- | --- | --- | --- |
| Invasion Categories | optdigits.tra | optdigits.tes | Total Number of Samples |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| Total number of samples in total |  |  | 122 |

Henri Begleiter   
Neurodynamics Laboratory,   
State University of New York Health Center   
Brooklyn, New York 

**Data Set Information:**

This data arises from a large study to examine EEG correlates of genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on subject's scalps which were sampled at 256 Hz (3.9-msec epoch) for 1 second.   
  
There were two groups of subjects: alcoholic and control. Each subject was exposed to either a single stimulus (S1) or to two stimuli (S1 and S2) which were pictures of objects chosen from the 1980 Snodgrass and Vanderwart picture set. When two stimuli were shown, they were presented in either a matched condition where S1 was identical to S2 or in a non-matched condition where S1 differed from S2.   
  
Shown here are example plots of a control ([[Web Link]](http://kdd.ics.uci.edu/databases/eeg/control.gif)) and alcoholic ([[Web Link]](http://kdd.ics.uci.edu/databases/eeg/alcoholic.gif)) subject. The plots indicate voltage, time, and channel and are averaged over 10 trials for the single stimulus condition.   
  
There were 122 subjects and each subject completed 120 trials where different stimuli were shown. The electrode positions were located at standard sites (Standard Electrode Position Nomenclature, American Electroencephalographic Association 1990). Zhang et al. (1995) describes in detail the data collection process.   
  
There are three versions of the EEG data set.   
  
1. The Small Data Set   
The small data set (smni97\_eeg\_data.tar.gz) contains data for the 2 subjects, alcoholic a\_co2a0000364 and control c\_co2c0000337. For each of the 3 matching paradigms, c\_1 (one presentation only), c\_m (match to previous presentation) and c\_n (no-match to previous presentation), 10 runs are shown.   
  
2. The Large Data Set   
The large data set (SMNI\_CMI\_TRAIN.tar.gz and SMNI\_CMI\_TEST.tar.gz) contains data for 10 alcoholic and 10 control subjects, with 10 runs per subject per paradigm. The test data used the same 10 alcoholic and 10 control subjects as with the training data, but with 10 out-of-sample runs per subject per paradigm.   
  
3. The Full Data Set   
This data set contains all 120 trials for 122 subjects. The entire set of data is about 700 MBytes.   
  
NOTE: There are 17 trials with empty files in co2c1000367. Some trials have "err" notices, e.g., search/grep for "err" and see "S2 match err" or "S2 nomatch err" etc.

**Attribute Information:**

Each trial is stored in its own file and will appear in the following format.   
  
# co2a0000364.rd   
# 120 trials, 64 chans, 416 samples 368 post\_stim samples   
# 3.906000 msecs uV   
# S1 obj , trial 0   
# FP1 chan 0   
0 FP1 0 -8.921   
0 FP1 1 -8.433   
0 FP1 2 -2.574   
0 FP1 3 5.239   
0 FP1 4 11.587   
0 FP1 5 14.028   
...   
  
The first four lines are header information. Line 1 contains the subject identifier and indicates if the subject was an alcholic (a) or control (c) subject by the fourth letter. Line 4 identifies the matching conditions: a single object shown (S1 obj), object 2 shown in a matching condition (S2 match), and object 2 shown in a non matching condition (S2 nomatch).   
  
Line 5 identifies the start of the data from sensor FP1. The four columns of data are: the trial number, sensor position, sample number (0-255), and sensor value (in micro volts).

This data arises from a large study to examine EEG correlates of genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on the scalp sampled at 256 Hz (3.9-msec epoch) for 1 second.

This data arises from a large study to examine EEG correlates of genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on subject's scalps which were sampled at 256 Hz (3.9-msec epoch) for 1 second.

There were two groups of subjects: alcoholic and control. Each subject was exposed to either a single stimulus (S1) or to two stimuli (S1 and S2) which were pictures of objects chosen from the 1980 Snodgrass and Vanderwart picture set. When two stimuli were shown, they were presented in either a matched condition where S1 was identical to S2 or in a non-matched condition where S1 differed from S2.

Shown here are example plots of a [control](https://archive.ics.uci.edu/ml/machine-learning-databases/eeg-mld/control.gif) and [alcoholic](https://archive.ics.uci.edu/ml/machine-learning-databases/eeg-mld/alcoholic.gif) subject. The plots indicate voltage, time, and channel and are averaged over 10 trials for the single stimulus condition.

There were 122 subjects and each subject completed 120 trials where different stimuli were shown. The electrode positions were located at standard sites (Standard Electrode Position Nomenclature, American Electroencephalographic Association 1990). Zhang et al. (1995) describes in detail the data collection process.

## Data Format

There are three versions of the EEG data set.

#### The Small Data Set

The small data set (smni97\_eeg\_data.tar.gz) contains data for the 2 subjects, alcoholic a\_co2a0000364 and control c\_co2c0000337. For each of the 3 matching paradigms, c\_1 (one presentation only), c\_m (match to previous presentation) and c\_n (no-match to previous presentation), 10 runs are shown.

#### The Large Data Set

The large data set (SMNI\_CMI\_TRAIN.tar.gz and SMNI\_CMI\_TEST.tar.gz) contains data for 10 alcoholic and 10 control subjects, with 10 runs per subject per paradigm. The test data used the same 10 alcoholic and 10 control subjects as with the training data, but with 10 out-of-sample runs per subject per paradigm.

#### The Full Data Set

This data set contains all 120 trials for 122 subjects. The entire set of data is about 700 MBytes.

NOTE: There are 17 trials with empty files in co2c1000367. Some trials have "err" notices, e.g., search/grep for "err" and see "S2 match err" or "S2 nomatch err" etc.

#### File Format

Each trial is stored in its own file and will appear in the following format.

# co2a0000364.rd

# 120 trials, 64 chans, 416 samples 368 post\_stim samples

# 3.906000 msecs uV

# S1 obj , trial 0

# FP1 chan 0

0 FP1 0 -8.921

0 FP1 1 -8.433

0 FP1 2 -2.574

0 FP1 3 5.239

0 FP1 4 11.587

0 FP1 5 14.028

...

The first four lines are header information. Line 1 contains the subject identifier and indicates if the subject was an alcholic (a) or control (c) subject by the fourth letter. Line 4 identifies the matching conditions: a single object shown (S1 obj), object 2 shown in a matching condition (S2 match), and object 2 shown in a non matching condition (S2 nomatch).

Line 5 identifies the start of the data from sensor FP1. The four columns of data are: the trial number, sensor position, sample number (0-255), and sensor value (in micro volts).