**68.Gas Sensor Array Drift Dataset at Different Concentrations**

1. 数据库网址

https://archive.ics.uci.edu/ml/datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations

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 Alexander Vergara, from BioCircutis Institute, University of California San Diego.【2.用于[什么实验目的]】This data set contains 13,910 measurements from 16 chemical sensors exposed to 6 gases at different concentration levels. This dataset is an extension of the Gas Sensor Array Drift Dataset, providing now the information about the concentration level at which the sensors were exposed for each measurement.【3】For processing purposes, the dataset is organized into ten batches, each containing the number of measurements per class and month indicated in the tables below. This reorganization of data was done to ensure having a sufficient and as uniformly distributed as possible number of experiments in each batch. The dataset is organized in files, each representing a different batch. Within the files, each line represents a measurement. The first character (1-6) codes the analyte, followed by the concentration level: 1: Ethanol; 2: Ethylene; 3: Ammonia; 4: Acetaldehyde; 5: Acetone; 6: Toluene. And the categories can be named as Batch ID Month IDs.【4】The database has 13910 samples, respectively belong to training with 11128 samples and testing with 2782 samples. As shown in Table 1.

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

|  |  |  |  |
| --- | --- | --- | --- |
| Categories | Training | Testing | Total Number of Samples |
| Ethanol | 1313 | 328 | 1641 |
| Ethylene | 1549 | 387 | 1936 |
| Ammonia | 2407 | 602 | 3009 |
| Acetaldehyde | 2341 | 585 | 2926 |
| Acetone | 2052 | 513 | 2565 |
| Toluene | 1466 | 367 | 1833 |
| Total number of samples in total | 11128 | 2782 | 13910 |

3. 精简描述

The Data in our experiment were collected by Alexander Vergara, from BioCircutis Institute, University of California San Diego. The dataset includes have 13910 samples, which used to classify the drift data. Through which, we divided the dataset into two part, training data set with 11128 samples and forecasting data set with 2782 samples.