**57.Pseudo Periodic Synthetic Time Series**

1. 数据库网址

https://archive.ics.uci.edu/ml/datasets/Pseudo+Periodic+Synthetic+Time+Series

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 Eamonn J. Keogh and Michael J. Pazzani, from department of Information and Computer Science, University of California.【2.用于[什么实验目的]】We used the database to test the pseudo periodic synthetic. This data set is designed for testing indexing schemes in time series databases. The data appears highly periodic, but never exactly repeats itself.【3】It is a much larger dataset than has been used in any published study (That we are currently aware of). It contains one million data points. The data has been split into 10 sections to facilitate testing (see below). We recommend building the index with 9 of the 100,000-datapoint sections, and randomly extracting a query shape from the 10th section. (Some previously published work seems to have used queries that were also used to build the indexing structure. This will produce optimistic results) The data are interesting because they have structure at different resolutions. The data appears highly periodic, but never exactly repeats itself. This feature is designed to challenge the indexing structure. The data is stored in one ASCII file. There are 10 columns, 100,000 rows. All data points are in the range -0.5 to +0.5. Rows are separated by carriage returns, columns by spaces.【4】The database has 100000 samples, respectively belong to training with 80000 samples and testing with 20000 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 |
| Financial | 20000 | 5000 | 25000 |
| Photon | 40000 | 5000 | 45000 |
| Synthetic | 10000 | 5000 | 15000 |
| Shuttle | 10000 | 5000 | 15000 |
| Total number of samples in total | 80000 | 20000 | 100000 |

3. 精简描述

The Data in our experiment were collected by Eamonn J. Keogh and Michael J. Pazzani. The dataset includes have 100000 samples, which used to testing indexing schemes in time series databases. Through which, we divided the dataset into two part, training data set with 80000 samples and forecasting data set with 20000 samples.