**102.Coil 1999 Competition**

1. 数据库网址

http://archive.ics.uci.edu/ml/datasets/Coil+1999+Competition+Data

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 |  |  | 340 |

|  |  |
| --- | --- |
| **Abstract**: This data set is from the 1999 Computational Intelligence and Learning (COIL) competition. The data contains measurements of river chemical concentrations and algae densities. |  |

**Source:**

Original Owner:   
  
ERUDIT   
European Network for Fuzzy Logic and Uncertainty Modelling in Information Technology   
<http://www.erudit.de/> 

**Data Set Information:**

This data comes from a water quality study where samples were taken from sites on different European rivers of a period of approximately one year. These samples were analyzed for various chemical substances including: nitrogen in the form of nitrates, nitrites and ammonia, phosphate, pH, oxygen, chloride. In parallel, algae samples were collected to determine the algae population distributions.   
  
The competition involved the prediction of algal frequency distributions on the basis of the measured concentrations of the chemical substances and the global information concerning the season when the sample was taken, the river size and its flow velocity. The competition instructions contain additional information on the prediction task: [[Web Link]](http://kdd.ics.uci.edu/databases/coil/instructions.txt)

**Attribute Information:**

There are a total of 340 examples each containing 17 values. The first 11 values of each data set are the season, the river size, the fluid velocity and 8 chemical concentrations which should be relevant for the algae population distribution. The last 8 values of each example are the distribution of different kinds of algae. These 8 kinds are only a very small part of the whole community, but for the competition we limited the number to 7. The value 0.0 means that the frequency is very low. The data set also contains some empty fields which are labeled with the string XXXXX.   
  
The training data are saved in the file: analysis.data (ASCII format).   
  
Table 1: Structure of the file analysis.data   
  
A ... K a ... g   
CC1,1 ... CC1,11 AG1,1 ... AG1,7   
...   
CC200,1 ... CC200,11 AG200,1 ... AG200,7   
  
  
Explanation:   
CCi,j: Chemical concentration or river characteristic   
AGi,j: Algal frequency   
  
The chemical parameters are labeled as A, ..., K. The columns of the algaes are labeled as a, ..,g.