**65.UNIX User Data**

1. 数据库网址

http://archive.ics.uci.edu/ml/datasets/UNIX+User+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 Terran Lane.【2.用于[什么实验目的]】The data is made available under conditions of anonymity for the contributing users and may be used for research purposes only.【3】This file contains 9 sets of sanitized user data drawn from the command histories of 8 UNIX computer users at Purdue over the course of up to 2 years (USER0 and USER1 were generated by the same person, working on different platforms and different projects). The data is drawn from tcsh(1) history files and has been parsed and sanitized to remove filenames, user names, directory structures, web addresses, host names, and other possibly identifying items. Command names, flags, and shell metacharacters have been preserved. Additionally, \*\*SOF\*\* and \*\*EOF\*\* tokens have been inserted at the start and end of shell sessions, respectively. Sessions are concatenated by date order and tokens appear in the order issued within the shell session, but no timestamps are included in this data.【4】The database has 420 samples, respectively belong to training with 300 samples and testing with 120 samples. The categories can be as shown in Table 1.

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

|  |  |  |
| --- | --- | --- |
| Traning | Testing | Total Number of Samples |
| 300 | 120 | 420 |

3. 精简描述

The Data in our experiment were collected by Terran Lane. The dataset includes have 420 samples, which used to classify the UNIX user. Through which, we divided the dataset into two part, training data set with 300 samples and forecasting data set with 120 samples.