**49.UJI Pen Characters (Version 2)**

1. 数据库网址

http://archive.ics.uci.edu/ml/datasets/UJI+Pen+Characters+(Version+2)

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 F. Prat, M. J. Castro(+) etc., from departamento de Lenguajes y Sistemas Informáticos, Universitat Jaume I (UJI), at december 2008.【2.用于[什么实验目的]】We have created the UJI penchars2 character database by collecting samples from 60 writers at two different sites in two phases: 1st phase, 11 writers, carried out at UJI. 2nd phase, 49 writers, carried out at UPV (44 writers) and UJI (5). Classification of test samples into 35 classes: 26 case-independent classes for letters. 9 additional clases for non-zero digits. Zero included in the same class as o's.【3】the total number of samples in this database is 11640: 60 writers x (66+10+21) characters x 2 repetitions. UJIpenchars is a subset of UJIpenchars2 with only 1364 samples: the ASCII letters and digits collected at UJI during the 1st acquisition phase. We have not defined a standard task for UJIpenchars2, but divided the writer set into two disjoint subsets in order to ease the definition of writer independent tasks:40 'trn' writers:The 11 1st phase UJI writers. And 29 UPV writers. 20 'tst' writers:The 5 2nd phase UJI writers. And 15 UPV writers.tidigits.tes with 1797 samples.【4】The database has 11640 samples, there are 60 writers and 194 instances per writer file, totalling 11640 instances: 7760 correspond to the 40 "trn" writers. 3880 correspond to the 20 "tst" writers. As shown in Table 1.

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

|  |  |  |  |
| --- | --- | --- | --- |
| Invasion Categories | 40 “trn” | 20 “tst” | Total Number of Samples |
| Trainning | 3307 | 1665 | 4960 |
| test | 4453 | 2215 | 2480 |
| Total number of samples in total | 7760 | 3880 | 11640 |