**50. Character Trajectories**

1. 数据库网址

http://archive.ics.uci.edu/ml/datasets/Character+Trajectories

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 Ben H Williams, from School of Informatics,University of Edinburgh.【2.用于[什么实验目的]】We used to classify the character trajectory.【3】The characters here were used for a PhD study on primitive extraction using HMM based models. The data consists of 2858 character samples, contained in the cell array 'mixout'. The struct variable 'consts' contains a field consts.charlabels which provides ennummerated labels for the characters. consts.key provides the key for each label. The data was captured using a WACOM tablet. 3 Dimensions were kept - x, y, and pen tip force. The data has been numerically differentiated and Gaussian smoothed, with a sigma value of 2. Data was captured at 200Hz. The data was normalised with consts.datanorm. Only characters with a single 'PEN-DOWN' segment were considered. Character segmentation was performed using a pen tip force cut-off point. The characters have also been shifted so that their velocity profiles best match the mean of the set. Each character sample is a 3-dimensional pen tip velocity trajectory. This is contained in matrix format, with 3 rows and T columns where T is the length of the character sample.【4】The database has 2858 samples, respectively belong to training with 2450 samples and testing with 408 samples. The categories of dataset include six categories, 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 |
| x-position | 407 | 68 | 475 |
| y-position | 405 | 68 | 473 |
| pen tip | 406 | 68 | 474 |
| pressure | 410 | 68 | 478 |
| pen tilt angle | 402 | 68 | 470 |
| pen orientation | 420 | 68 | 488 |
| Total number of samples in total | 2450 | 408 | 2858 |

3. 精简描述

The Data in our experiment were collected by Ben H Williams, from School of Informatics, University of Edinburgh. The dataset includes have 2858 samples, which used to classify the character trajectory. Through which, we divided the dataset into two part, training data set with 2450 samples and forecasting data set with 408 samples.