**58.YearPredictionMSD（TODO：补充简介）**

1. 数据库网址

http://archive.ics.uci.edu/ml/datasets/YearPredictionMSD

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 from the website “http://labrosa.ee.columbia.edu/millionsong/”.【2.用于[什么实验目的]】We used classify the MSD.【3】The database has 90 attributes, 12 = timbre average, 78 = timbre covariance. The first value is the year (target), ranging from 1922 to 2011. Features extracted from the 'timbre' features from The Echo Nest API. We take the average and covariance over all 'segments', each segment being described by a 12-dimensional timbre vector.【4】The database has 515346 samples, respectively belong to training with 463715 samples and testing with 51630 samples. The categories can be as shown in Table 1.

Table 1 Category Distribution of MSD dataset [根据数据库绘制]

|  |  |  |  |
| --- | --- | --- | --- |
| Categories | Training | Testing | Total Number of Samples |
| SHS | 12341 | 5854 | 18196 |
| musiXmatch | 198202 | 39460 | 237662 |
| Last.fm | 135254 | 4682 | 139936 |
| Taste profile | 117918 | 1634 | 119552 |
| Total number of samples in total | 463715 | 51630 | 515346 |

3. 精简描述

The Data in our experiment were collected from the website “http://labrosa.ee.columbia.edu/millionsong/”. The dataset includes have 515346 samples, which used to classify the MSD. Through which, we divided the dataset into two part, training data set with 463715 samples and forecasting data set with 51630 samples.