采用的数据集是movielens

举例说明数据集的参数特征：该数据集保存的是6040个用户对3952部电影的1000209个评分记录。具体可以参考其README文件。

**需要注意的是不同的数据集大小的内容规范也是不一样的**

用户表的数据格式：UserID::Gender::Age::Occupation::Zip-code

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| USerID | Gender | Age | Occupation | Zip-code |
| 用户名 | 性别 | 年龄 | 职业 | 邮政编码 |
| 1 | F | 1 | 10 | 48067 |

有关年龄一栏的详细解释：

**\* 1: "Under 18"**

**\* 18: "18-24"**

**\* 25: "25-34"**

**\* 35: "35-44"**

**\* 45: "45-49"**

**\* 50: "50-55"**

**\* 56: "56+"**

职业一览的详细解释：

\* 0: "other" or not specified

\* 1: "academic/educator"

\* 2: "artist"

\* 3: "clerical/admin"

\* 4: "college/grad student"

\* 5: "customer service"

\* 6: "doctor/health care"

\* 7: "executive/managerial"

\* 8: "farmer"

\* 9: "homemaker"

\* 10: "K-12 student"

\* 11: "lawyer"

\* 12: "programmer"

\* 13: "retired"

\* 14: "sales/marketing"

\* 15: "scientist"

\* 16: "self-employed"

\* 17: "technician/engineer"

\* 18: "tradesman/craftsman"

\* 19: "unemployed" 事业人员

\* 20: "writer" 作家

评价表的数据的格式

UserID::MovieID::Rating::Timestamp