**人工智能实验--第一节特征工程**

**本实验采用python来写代码，主要引用 xlrd，xlwt，numpy，skleaarn，**

**通过 xlrd，xlet进行excl文件的读取和写入 (读取samplesubmission.csv文件)**

**split\_age\_range(user\_id)将特征值 用户id 进行离散化为8个特征值（以id为例）**

**if user\_id >0 and user\_id<=50000:**

**return [1,0,0,0,0,0,0,0,0]**

**elif user\_id >50000 and user\_id<=100000:**

**return [0,1,0,0,0,0,0,0,0]**

**elif user\_id >100000 and user\_id<=150000:**

**return [0,0,1,0,0,0,0,0,0]**

**elif user\_id >150000 and user\_id<=200000:**

**return [0,0,0,1,0,0,0,0,0]**

**elif user\_id >200000 and user\_id<=250000:**

**return [0,0,0,0,1,0,0,0,0]**

**elif user\_id >250000 and user\_id<=300000:**

**return [0,0,0,0,0,1,0,0,0]**

**elif user\_id >300000 and user\_id<=350000:**

**return [0,0,0,0,0,0,1,0,0]**

**elif user\_id >350000 :**

**return [0,0,0,0,0,0,0,1,0]**

**split\_gender(merchant\_id)将特征值merchant\_id进行离散化**

**if merchant\_id >0 and merchant\_id<=3000:**

**return [1,0,0]**

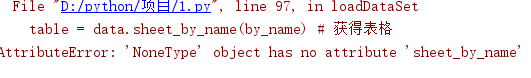
**elif merchant\_id >3000:**

**return [0,1,0]**

**def split\_log():分割数据文件中数据**

**但是在读取表格时总有这个错误 各种百度都解决不出 应该不是函数问题**

**已经放弃了 ┭┮﹏┭┮**



171491406 徐启睿