



北京邮电大学
Beijing University of Posts and Telecommunications

《机器学习》作业说明



■ 期末作业--->提供分类数据集 (50%)

赛道一：电表度数识别 (CV)

赛道二：QoE指标 (网络流量)

***难度相当，任选赛道，3人一组，现场测试**

赛道一：电表读数识别

■ 电表照片数据

手动标注

图像基本处理：去照光、畸变矫正
可选思路：

1. 定位读数区域

（检测任务，选择一个模型）

2. 识别数字（分类任务）



数据下载地址，百度网盘：链接：

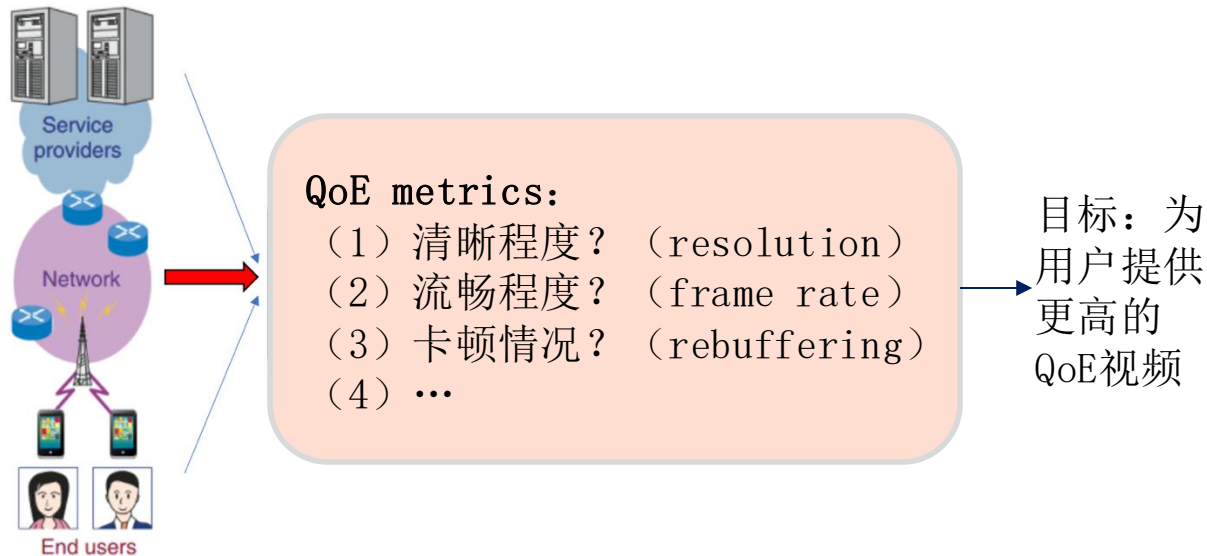
https://pan.baidu.com/s/1fwo0xo_p7r5_VMNCiBHK_g?pwd=ML25 提取码: ML25

赛道一验收方式

- 验收当日下发测试集
- 在给定时间内提交模型的预测结果，并将结果保存至文件中。格式为：第一列是测试数据的id，第二列是电表读数，中间以英文逗号分隔，例：“0,8430.8”。文件名命名为“赛道（1/2/3）-组员（姓名+学号）.csv”
- 同时一起提交设计思路（**最长一页纸**）及代码。
- 评分标准：
 - 15分：程序正常运行，读取数据，输出结果；
 - 15分：准确率达到80%以上；
 - 10分：模型设计思路（**最长一页纸**）及代码；
 - 10分：更高的准确率：90%以上加5分，95%以上加10分。

视频QoE (Quality of Experience)：用户对视频质量的体验

论文复现与优化：Requet Real-Time QoE Metric Detection for Encrypted YouTube Traffic



根据Youtube加密流量，对QoE指标进行预测，包括3个指标：

- (1) **BufferWarning**: BfW、NBfW (二分类)
- (2) **VideoState**: Stall、Buf.decy、Buf.increase、Steady (四分类)
- (3) **Resolution**: 144p、240p、360p、480p、720p、1080p (六分类)

将原始数据集整理了一下，生成了一个新的数据集，名称为RequetDatasetNew，3个文件夹A0/A1/A2，便于做交叉验证，下载地址为：



https://pan.baidu.com/s/1DhS25t_3cu3SQmZdgdVpcQ?pwd=ML25 提取码：ML25

现场验收：**验收当日下发测试集；规定时间内提交结果，超时提交扣分**
提交作业：

- 最终提交的作业需要包含：
 - i. 完整项目文件 (可直接运行)、训练好的模型
 - ii. 运行结果 (3个指标，分别放在3个文件中)
 - iii. 模型设计报告 (格式不限，**最长一页**)
- 命名方式：赛道(1/2/3)-组员1(姓名+学号)-组员2(姓名+学号)-....zip

评分标准，考虑分类的准确率

15分：程序正常运行，读取数据，输出结果；

13分：三个指标的准确率：达到低于原文10-15个百分点；

12分：模型设计思路 (**最长一页纸**) 及代码 (清晰可读) ；

10分：更高的准确率，3个指标均达到论文结果-5点，3个指标均超过论文结果加5分。