

某工业产品结果参数预测

1 数据说明

数据为某工业产品的生产记录，按生产批次顺序记录，每次生产时间间隔是一致的，每隔5小时生产一次，即生产一次产品需要花费5个小时。某批次生产过程当中控制设备的参数不发生变化。不同批次生产过程中的参数可能不同。一次批次的产品生产完毕，无需等待检测完毕，即能够进行下一次生产，检测时间需要5个小时。

attr_1到attr_10是生产过程参数，target是结果检测参数。

训练数据地址：<https://obs-mlsclass7.obs.cn-north-1.myhuaweicloud.com/trainData.csv>

测试数据地址：<https://obs-mlsclass7.obs.cn-north-1.myhuaweicloud.com/testData.csv>

1) 训练数据 **trainData.csv**：共437条训练数据，生产过程+检测过程需要10个小时：

- a.前5个小时attr_1到attr_10不发生变化，进行一个批次(例如批次号为A)产品的生产。
- b.后5个小时attr_1到attr_10可能发生变化，进行下一个批次(例如批次号为B)的产品的生产，同时前一个批次A的产品进行为期5个小时的检测，并得到A的检测结果target。

即：批次A的检测和批次B的生产是同时进行的（且批次A的检测和批次B的生产同时结束），批次B的检测和批次C的生产同时进行。

其中：前5个小时的attr_1到attr_10和A的target构成一条数据记录。后5个小时的attr_1到attr_10和B的target构成下一条数据记录。

2) 测试数据 **testData.csv**：共110条测试数据，全部为生产过程参数。

2 题目

需要对测试数据 **testData.csv**中所有的行（生产过程参数attr_1到attr_10）对应的检测参数进行预测。即当批次A的生产进行完毕后，无需检测，即可知道批次A的检测结果。

3 评分标准

采用MAPE（平均绝对百分比误差）作为评分标准，
$$MAPE = \frac{\sum_{i=1}^n \left| \frac{target_i - predict_i}{target_i} \right|}{n}$$

，即每一条测试数据的真实值 ($target_i$) 减去预测值 ($predict_i$)，然后除以真实值，这个比值求绝对值，然后所有测试数据的这个绝对值求和，然后除以测试数据条数 n 。

最终的MAPE是一个百分比数字，表示着误差百分比，所以采用1-MAPE作为评分，比如MAPE=20%，则得分为1-MAPE=80分。

4 答案格式

1) 提交答案的文件为result_XXXX.csv，包含1维数据<预测值>，列名为<prediction>，请注意最终的预测结果的行顺序请与原始测试数据的行顺序保持一致。

2) XXXX是指华为云用户名

5 题目有效期

本题提交有效期为7天（2019年3月4日-2019年3月10日），将完成结果发送至

chuxia@huawei.com邮箱，邮件标题格式为：ML晋级篇附加题+华为云用户名+微信昵称，逾期无效。

专家将结果进行评分，评选出最终获胜者一名，获得华为20000mAh移动电源。