

第六届“泰迪杯”C题测试结果评估指标说明

C题的测试集包含约2万个问题，每个问题有10段左右的材料。跟训练集的任务一致，参赛队伍需要判断每段材料与问题的相关性，按照 `submit_sample.txt` 中的格式给出每段材料的标签。

评估指标为F1，计算公式为：

“预测标签为1且真实标签也为1”的样本数 $\times 2 \div$ (“预测标签为1”的样本数 + “真实标签为1”的样本数)

注：

- 1、如提交的是浮点数，那么默认大于0.5为1，小于等于0.5为0；
- 2、测试集评估指标只是对模型有效性的一个检验，并不等价于最终获奖排名；
- 3、比赛结束后，会公开测试集标签，供大家自行对比检验。