Stage1 数据集，该数据集只发布400 个样本作为训练集，

组织方保留100 个样本作为测试集。Stage1 数据集信息如表1 所示，该数据集

属性均为数值型。

表1 Stage1 数据集信息

训练集样本数测试集样本数属性数量标签数量

400 100 72 6

竞赛第一阶段数据集文件说明：

1. Stage1\_train\_feature.CSV 文件存有训练样本的特征信息，每一行对应一个

样本的特征向量，第一列为样本id，每个特征值用“,”隔开；

2. Stage1\_train\_label.CSV 文件为训练样本的标签信息，每一行对应一个样本

的标签向量，第一列为样本id，每个标签用“,”隔开，其中1 表示该对应

标签与样本相关，0 表示对应标签与样本无关；

3. Stage1\_test\_feature.CSV 文件存有测试样本的特征信息，每一行对应一个

样本的特征向量，第一列为样本id，每个特征值用“,”隔开；

参赛者需要提交两个CSV 格式的文件，分别命名为

“Stage1\_参赛者编号\_label”和“Stage1\_参赛者编号\_confidence”（例如

Stage1\_G001\_label.CSV 和Stage1\_G001\_confidence.CSV）。

1. Stage1\_参赛者编号\_label.CSV 文件为测试集预测标签结果，每一行对应一

个样本的预测结果，第一列为样本id，标签之间用“,”隔开（预测标签结

果为一个取值范围为{0,1}的标签向量，例如“20,1,0,0,0,1,1”表示id 为

20 的测试样本与标签1,5,6 相关）；

2. Stage1\_参赛者编号\_confidence.CSV 文件为为测试集与每个标签的关联置

信度结果，每一行对应一个样本的预测结果，第一列为样本id，标签之间用

“,”隔开（预测关联置信度结果为一个取值范围为[0,1]的向量，关联置信

度值越高，对应标签与样本相关度越大，例如“20,0.9,0.1,0.3,0.2,0.8,0.8”

表示id 为20 的测试样本与标签1 的相关度最大，与标签2 的相关度最低；

预测关联置信度为标签排序的依据）。





