1.训练和预测代码再trains\_util中的run.py中，如图1所示

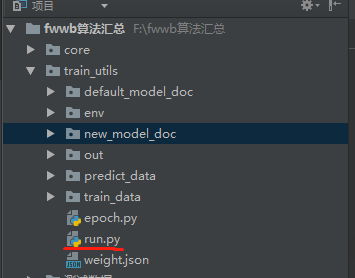


图1：目录结构

2.每次使用先创建一个Epoch对象

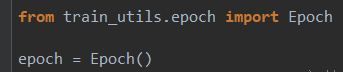


图2：新建Epoch对象

3.使用epoch.init\_enviroment()初始化环境，将之前的训练预测模型数据清空

4.epoch.train\_all(train\_data\_doc)，训练数据，参数是训练数据所在目录，可以直接是出题方给出的数据目录

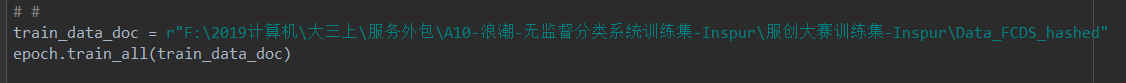
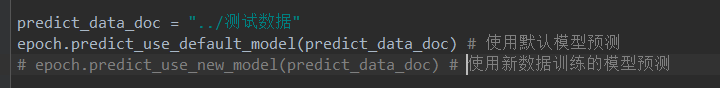


图3：训练模型

5.epoch.predict\_all(predict\_data\_doc)，预测数据，参数是预测数据所在目录，这里是将训练数据每个随机选5000条（不足5000则全部选入）作测试集，可以选择用默认模型（系统用的模型，predict\_use\_default\_model），或者是刚刚训练生成的模型预测(use\_new\_model)。



6.输出结果在/trains\_utils/out目录下，有8个csv，其中ent\_module.csv解释见图5



图4：out目录下的结果



图5：ent\_module.csv

ent\_module.csv是企业最终总的所有模块的数据的汇总，想针对性的看某个模块，例如风险模块的信息，可以进入risk\_module.csv中，里面包含了风险模块每个属性的等级，以及加权总分，风险等级。



图6：pic目录内容

由于将企业分了15个簇，每个簇的描述，用一张图片来表示，它反映了该簇内企业的7个模块和总分总体趋势，也便于分析。

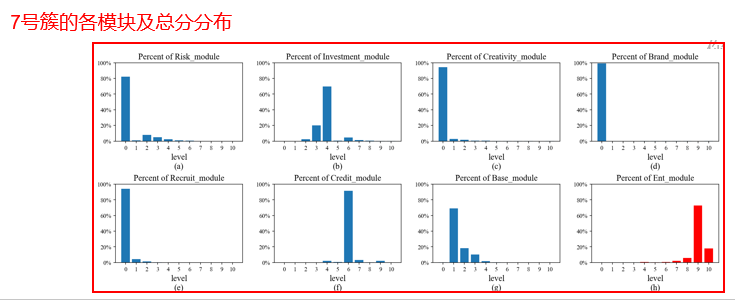


图8：7号簇分析.（a）风险分布；（b）投资分布；（c）创新分布；（d）品牌分布；（e）招聘分布；（f）信用分布；（g）资产分布；（h）总体等级分布

图8（a）反映了风险，可以看出总体企业风险为0-4，风险水平低；没有风险记录；同理b-h。则可以得出7号簇标签：低风险、中等投资、高信用、低资产水平。总体企业等级高。