失效日志预测RNN LSTM结构说明

1.数据处理：

将MySQL表中locality 的 logid整型数据转化为字符串，之后合并为一整个字符数据集；训练集数据和测试集数据分别来自不同的表格，比例为10：1；

2.数据映射：

构建char2id 和 id2char 函数，分别将logid字符串转化为整型数字和数字转化为字符串；

3.数据批处理：

构建BatchGenerator批处理函数，批处理字符串，batch\_size为每批字符串数目，num\_unrollings为每个字符串长度；

4.结果转换：

构造characters 函数将one-hot encoding结果softmax转化为int数字，再用id2char函数将数字转变成字符串；

5.构建LSTM模型：

利用Tensorflow框架定义变量；然后定义lstm\_cell，LSTM Cell有input、output、forget三个gate，saved\_output 是向上输出，saved\_state 保存当前状态。之后定义输入接口，按batch\*logid矩阵Input data，循环执行LSTM Cell，定义loss函数，使用标准 Gradient Descent定义优化，定义预测（建立初始 state 和 output，经过同样的 LSTM Cell，得到下一个预测的字符 sample\_prediction）；num\_nodes 是代表神经网络中LSTM Cell层的Cell个数；

6.训练数据：

批处理字符串长度为num\_unrollings + 1，将模型定义里 train\_inputs 和 train\_labels 错开1个字符。每summary\_frequency 整数倍步，输出平均 loss 值和 learning\_rate；每summary\_frequency \* 10 整数倍步，输出一些字符串结果；

7.预测比较：

prediction sample是输出预测的部分结果，利用accuracy、precision、recall、f1score分别对预测结果与实际结果进行评估。