### 基本框架

构建基于深度学习和神经网络的水质预测模型，采用历史水质、气象等数据进行监督学习训练。经过技术选型，本模型拟采用Keras和Tensorflow。

### 预测模型程序结构约定

整个模型训练、预测的计算模块最终要求能够打包成JAVA的jar包，以方便不同的项目引入并使用。

程序顶层API以JAVA接口的方式暴露，底层实现采用Python语言编写，约定JDK支持1.8，Python解释器采用3.6版本，这将要求预测模型程序运行的宿主机具备以上环境。

### 神经网络层选择

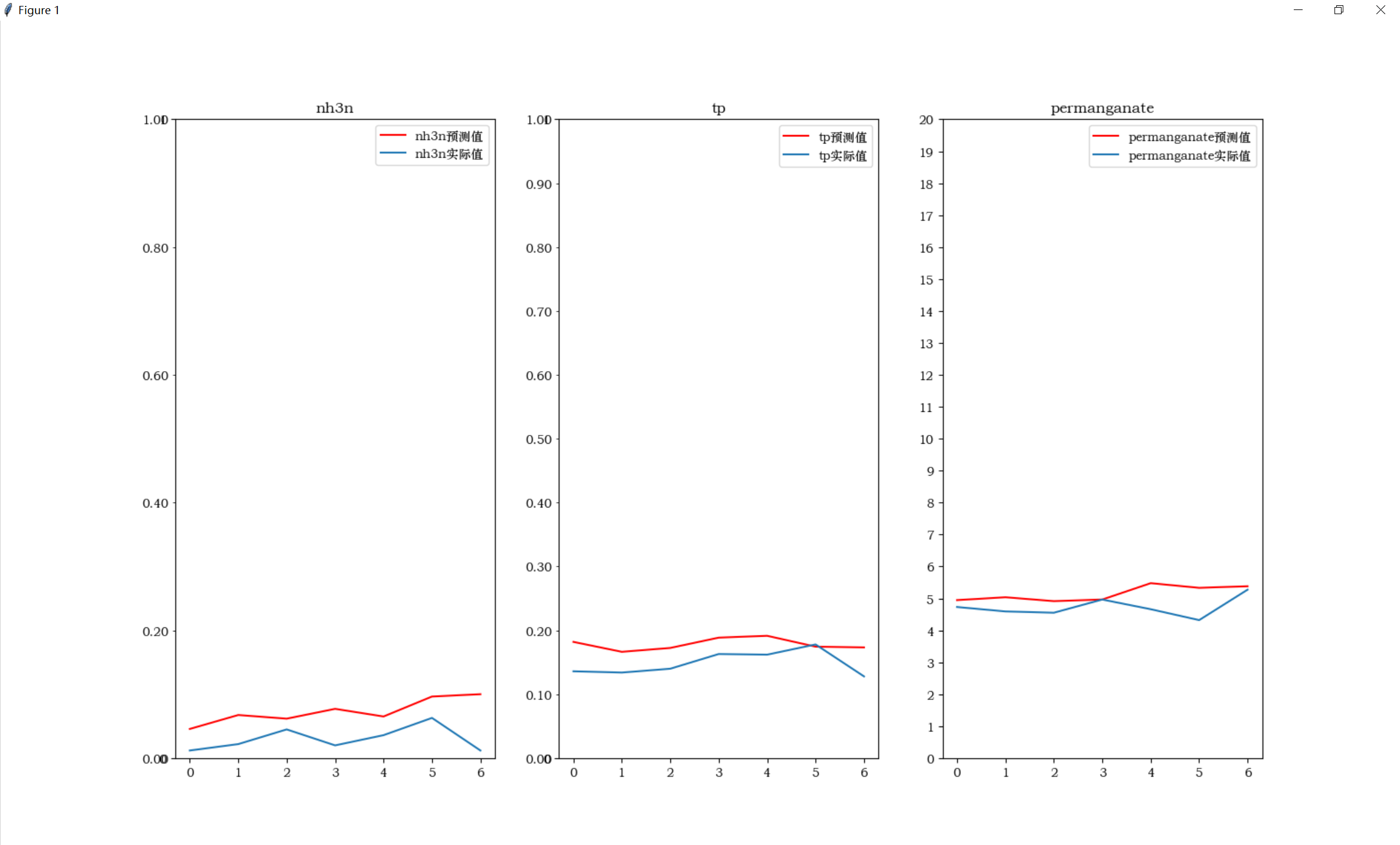
水质监测、水文和气象数据是一种按照时间排序的数据，那么RNN神经网络非常适合这一数据类型的训练和预测。Keras已提供了LSTM、GRU和SimpleRNN等常见的循环神经网络实现，目前演示示例采用LSTM神经网络，最终的神经网络结构需要经过反复调优来确定。

### 机器学习模型训练

目前已有的数据集包括: 各因子监测值、监测时间、水文数据(水位和流量)和气象数据(降雨量和气温)。

监督学习的训练需要训练数据集 TRAIN\_X, TRAIN\_Y。约定以监测时间、水文数据和气象数据作为 TRAN\_X，以各个因子监测值作为 TRAIN\_Y。

以下给出测试模型，测试模型使用了璧山两路口在线监测数据作为训练集，两个监测数据间的时间间隔为4小时，最终模型将能够根据前7个历史数据预测未来的7个数据。

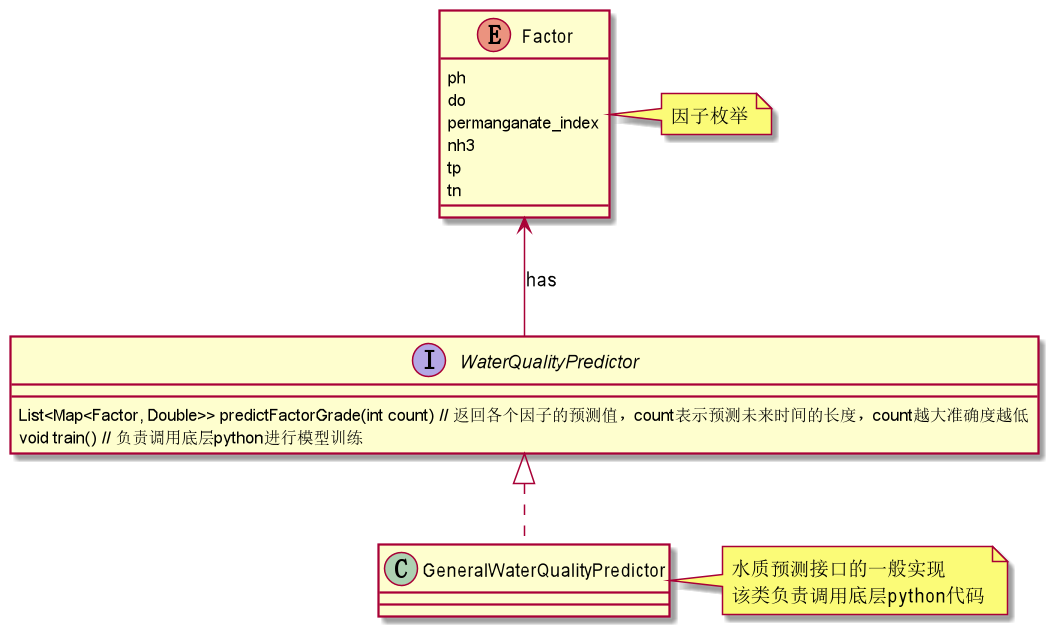


[图]预测氨氮、总磷和高猛酸盐指数未来的7个监测数据（红色曲线为预测值，蓝色曲线为真实值。左一至左三分别为氨氮、总磷、高猛酸盐指数）



[数据]两河口水质监测数据，可用数据量约5000条

### 模型调用接口设计示意



[图] JAVA接口设计

### 其他问题

1. 后续可将气象预报数据作为模型输入数据之一，以提高预测准确度；
2. 目前受限于数据仅含历史数据，以后可设想使用水面图像作为模型输入参数进行水质预测，参考文献: <https://www.mdpi.com/2072-4292/11/6/617>

### 参考文献

1. <https://www.zjujournals.com/eng/article/2021/1008-973X/202104001.shtml>
2. <https://www.mdpi.com/2072-4292/11/6/617>
3. <http://html.rhhz.net/JLDXXBDQKXB/html/1612253480466-787180513.htm>