**天气预测模型设计报告**

小组成员：侯传伟 2023210558 信研2306

张书奇 2023210575 信研2306

张天雨 2023210576 信研2306

## 一、小组成员分工

**张天雨：线性回归模型、RNN模型代码、大部分Web页面代码、报告撰写**

**张书奇：RFR模型代码、小部分Web网页代码、报告撰写**

**侯传伟：LSTM模型代码、小部分Web网页代码、报告撰写**

## 二、数据来源

我们小组的数据来源是2345天气（https://tianqi.2345.com），数据获取方式：使用Python爬虫爬取北京市三年（2019-2021）的每日天气数据（部分数据请见图1）。再将其转换为csv文件方便后续使用。



图1 部分数据示例

在实际使用时，使用了**日期、最高温度、最低温度**三个数据。

## 三、所选用模型和训练过程

我们小组一共训练了四个模型，分别为线性回归模型（Linear）、循环神经网络（RNN）、随机森林模型（RFR）和长短期记忆模型（LSTM）。

**其中，Linear、RNN、LSTM使用Pytorch完成，RFR使用Numpy完成**

这些模型在训练前共有的步骤有：

①处理数据；

②定义所使用的模型；

③使用该模型训练所给的训练集；

④使用训练完成的模型在测试集中进行训练，并将预测的结果打印出来。

在以下各模型单独的章节中除非有特殊情况，否则对步骤①和步骤④不再赘述，着重叙述步骤②和步骤③。

### 3.1 线性回归模型

（1）线性回归模型的定义如图2所示，该模型有三个全连接层（FC）和两个ReLU层。

模型定义：

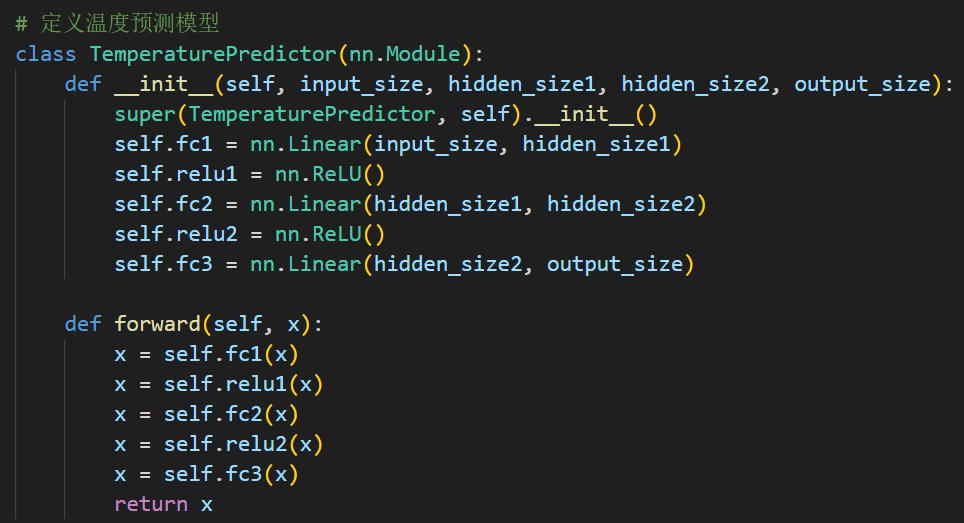


图2 线性回归模型定义

（2）使用该模型进行训练的步骤如图3所示，我们一共训练了600轮，并在每轮打印了损失（Loss）。

模型训练：

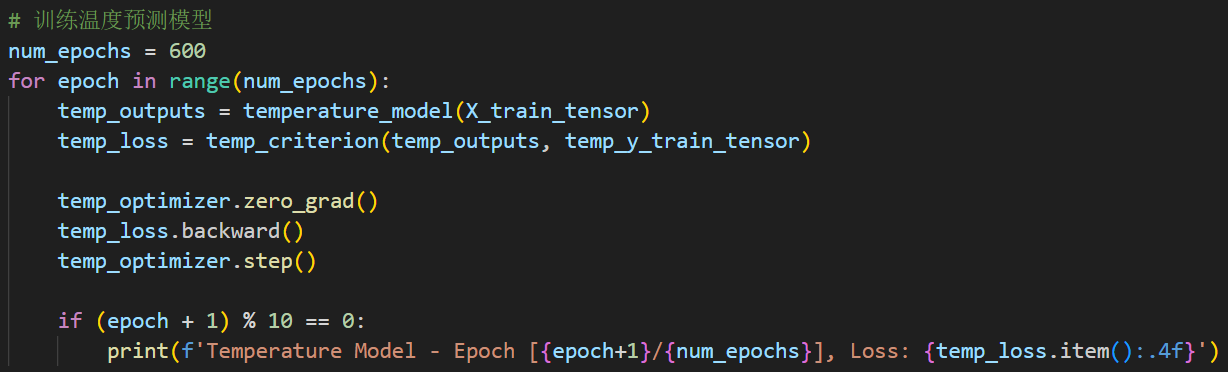
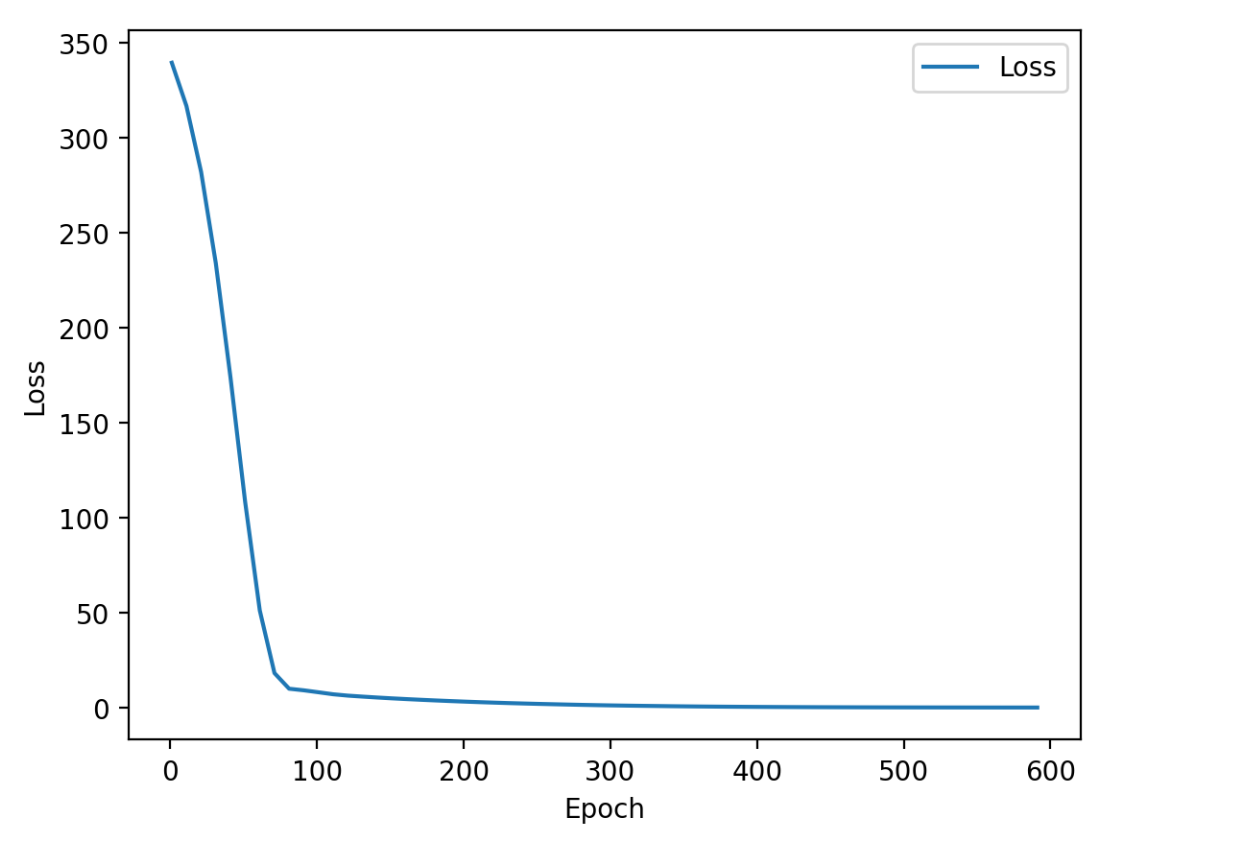
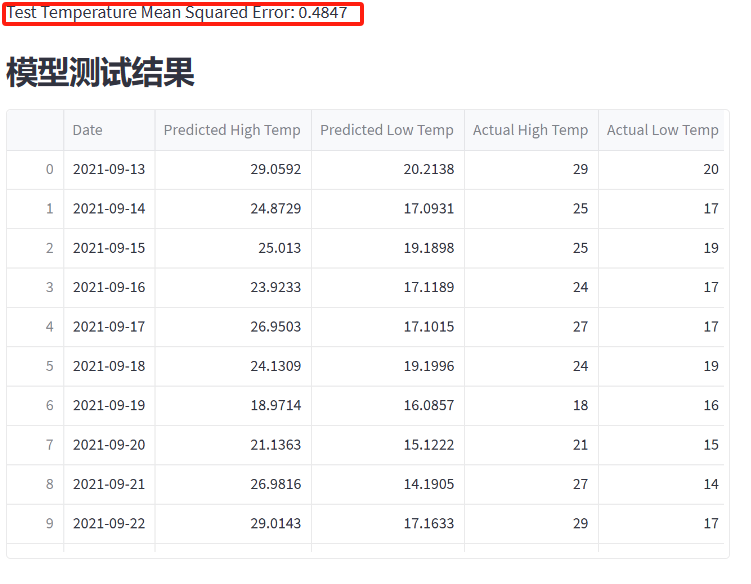


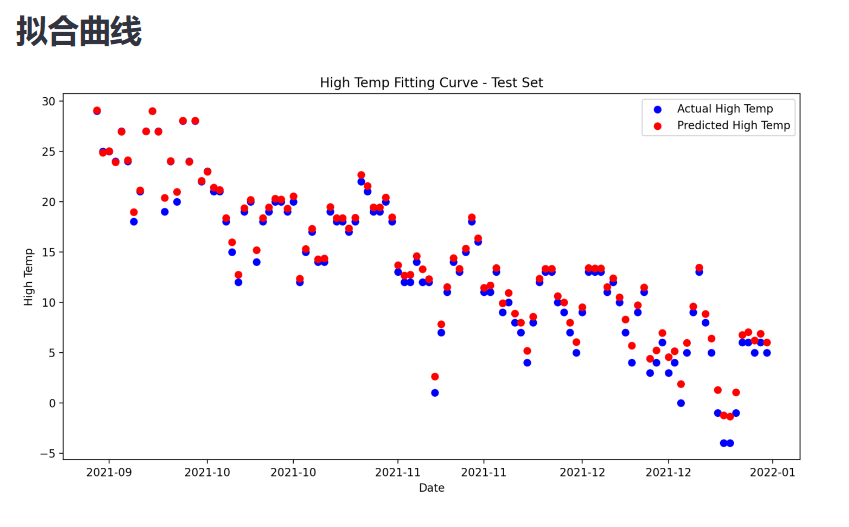
图3 使用线性回归模型训练所给数据集

（3）模型训练Loss：



（4）测试结果：





**可以看出效果不错**

### 3.2 循环神经网络（RNN）

（1）循环神经网络需要对数据做特殊的处理，即将数据转换为模型可用的序列数据。步骤如图4所示。

数据预处理：



图4 数据序列化步骤

在将数据序列化之后定义RNN模型，其定义如图5所示，我们定义了一个3层，具有128个隐藏单元的RNN。

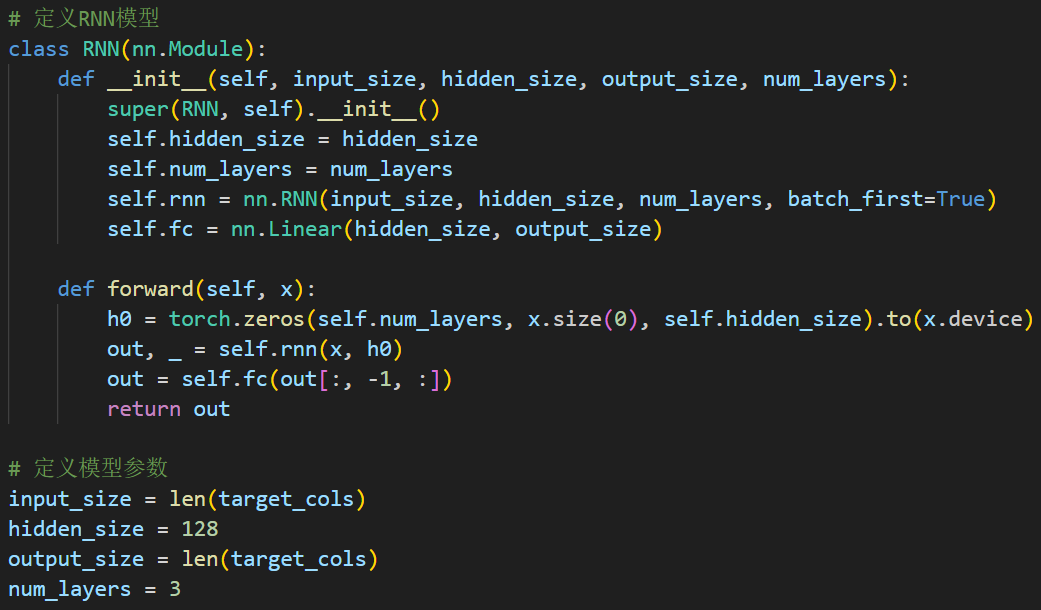


图5 RNN定义

（2）而训练RNN的过程见图6，我们共训练了300轮，并打印了每一轮的损失函数。

模型训练：

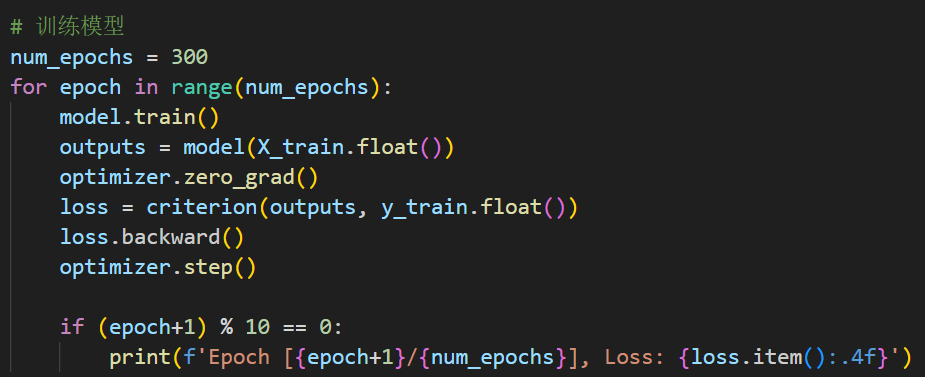
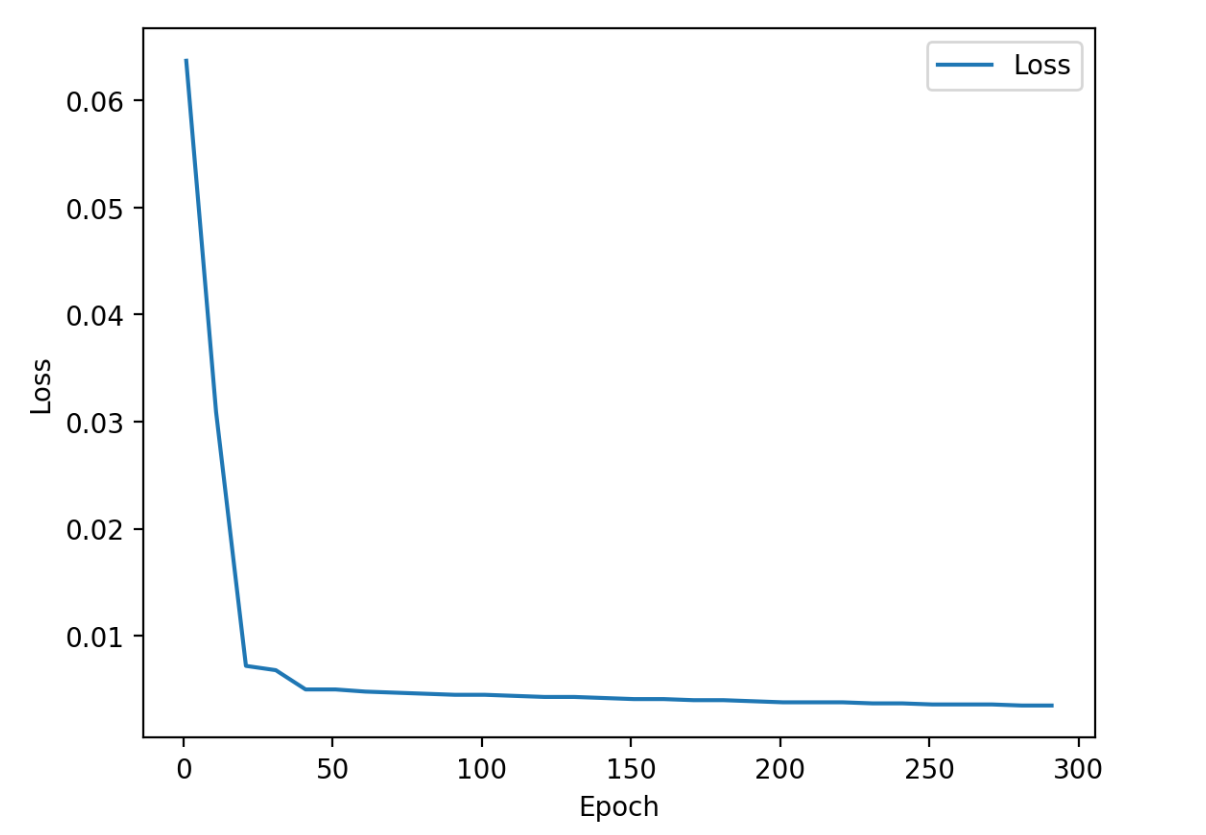
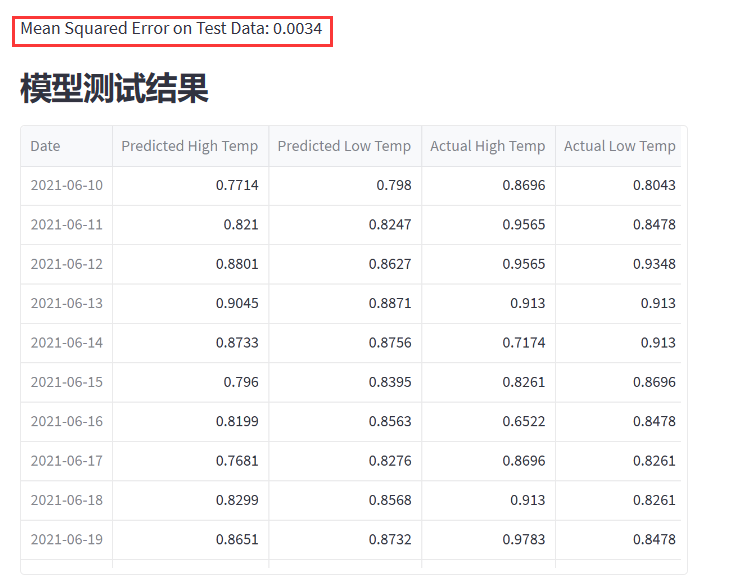


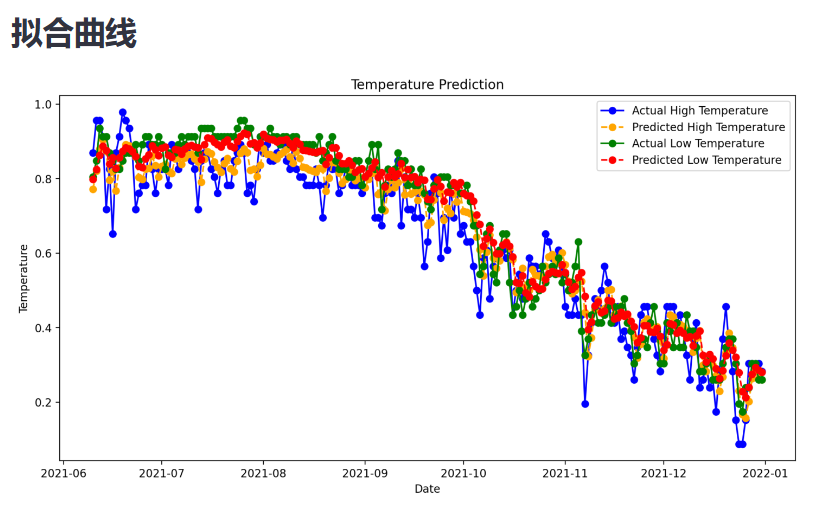
图6 RNN训练过程

（3）模型训练Loss：



（4）测试结果：





可以看出效果也不错

### 3.3 随机森林模型（RFR）

（1）随机森林模型的定义见图7-图9，我们定义了一个大小为10的随机森林，并定义了训练函数和预测函数。在训练时对每个树创建一个决策树回归器，并在预测时取每个回归器的预测值进行平均得到预测数组

模型定义：

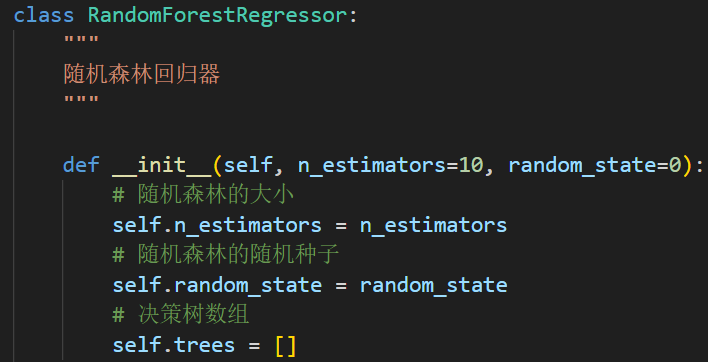


图7 RFR的初始化函数

（2）模型训练：

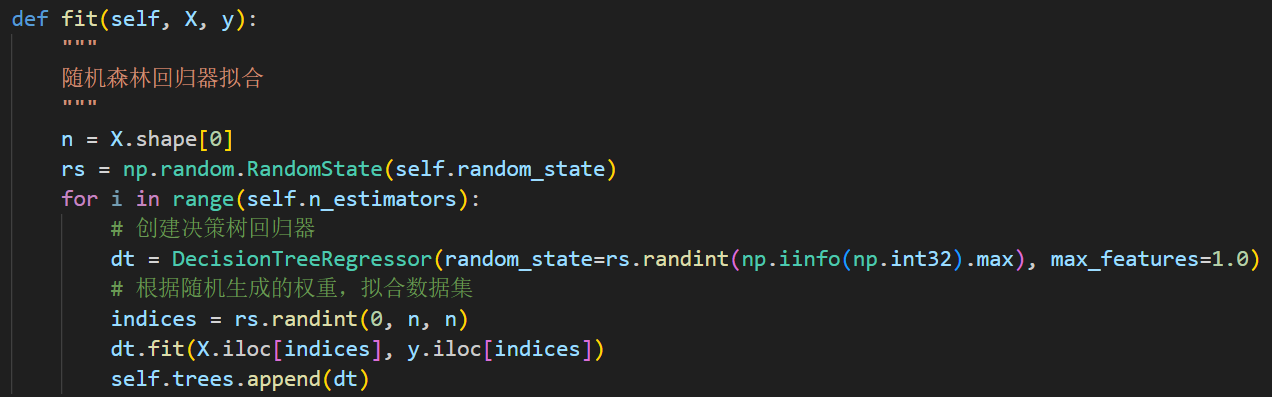


图8 RFR的训练函数

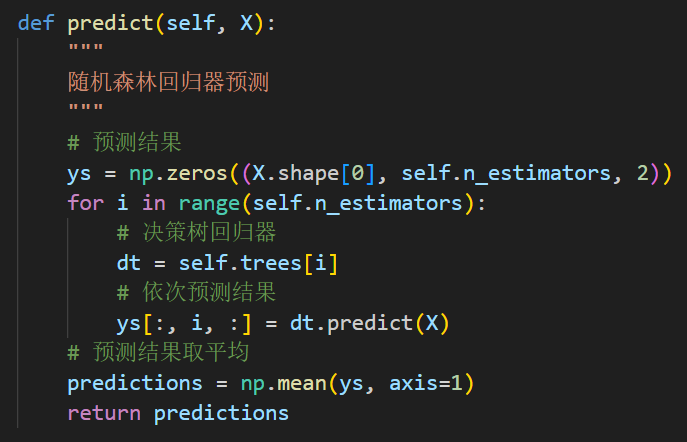
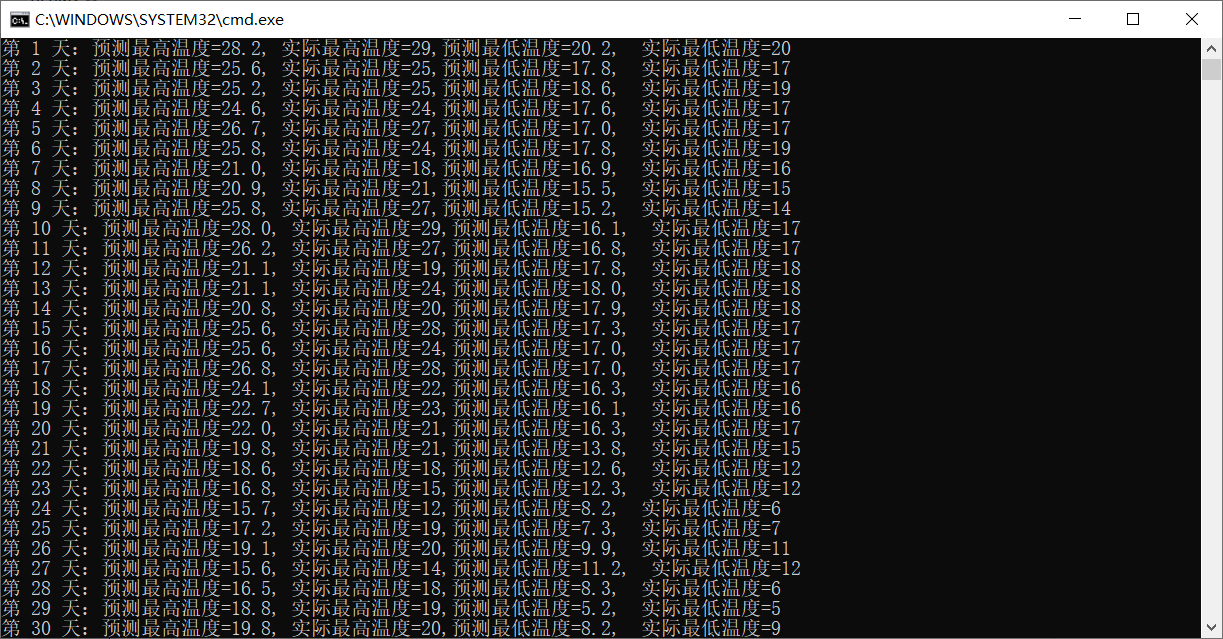
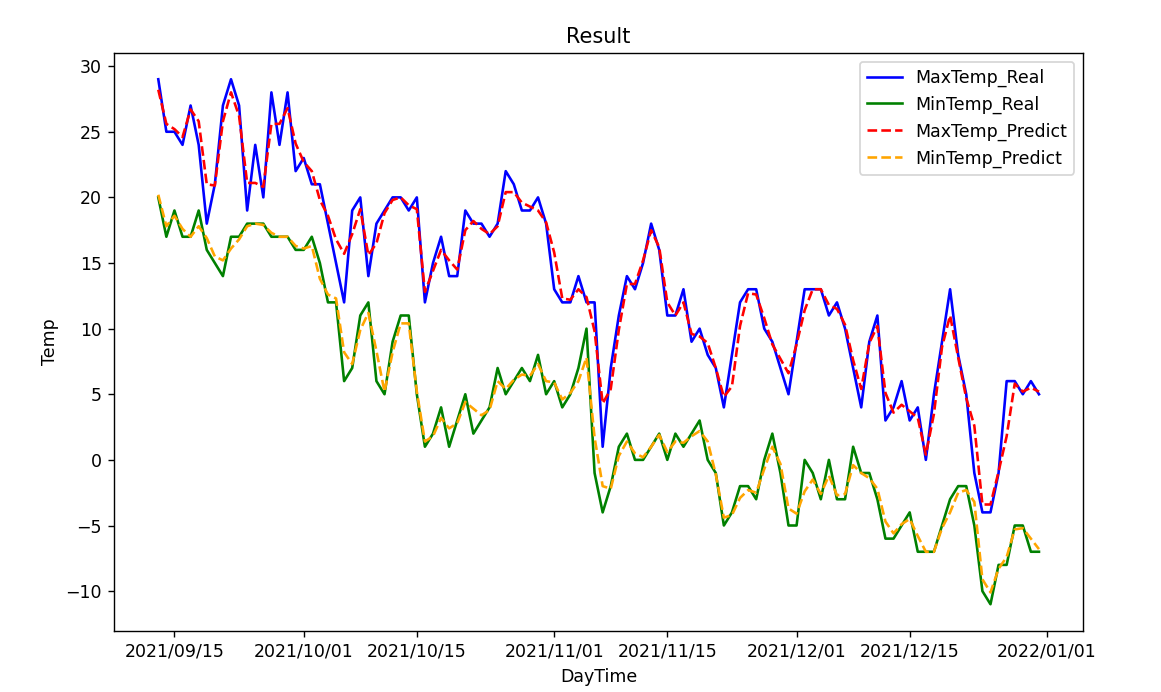


图9 RFR的预测函数

（3）测试结果：





可以看出效果很好

### 3.4 长短期记忆模型（LSTM）

（1）长短期记忆模型的定义见图10，我们定义了一个有两个LSTM层，一个ReLU层和两个全连接层的LSTM模型。隐藏神经元个数为150。

模型定义：

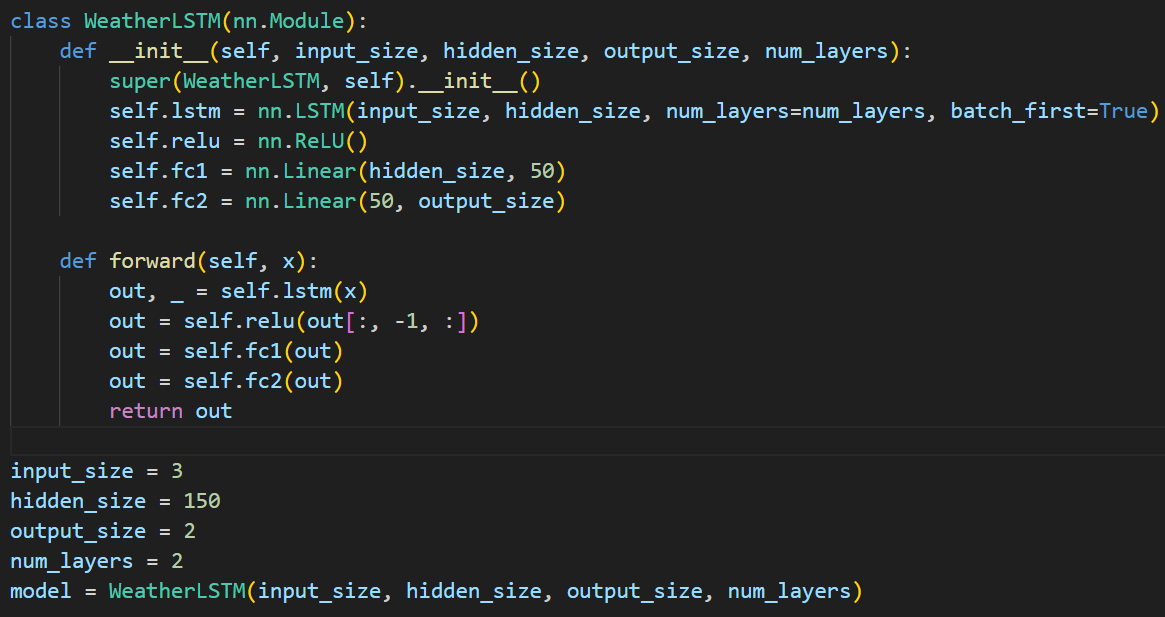


图10 LSTM模型定义

（2）LSTM模型的训练过程如图11所示，我们一共训练了150轮，并打印了每轮的损失。

模型训练：

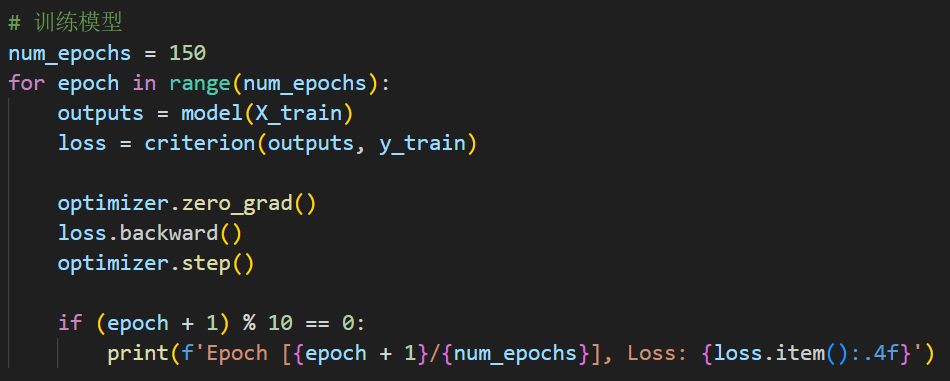
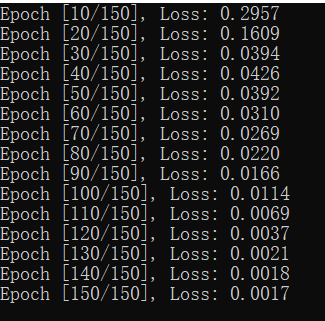
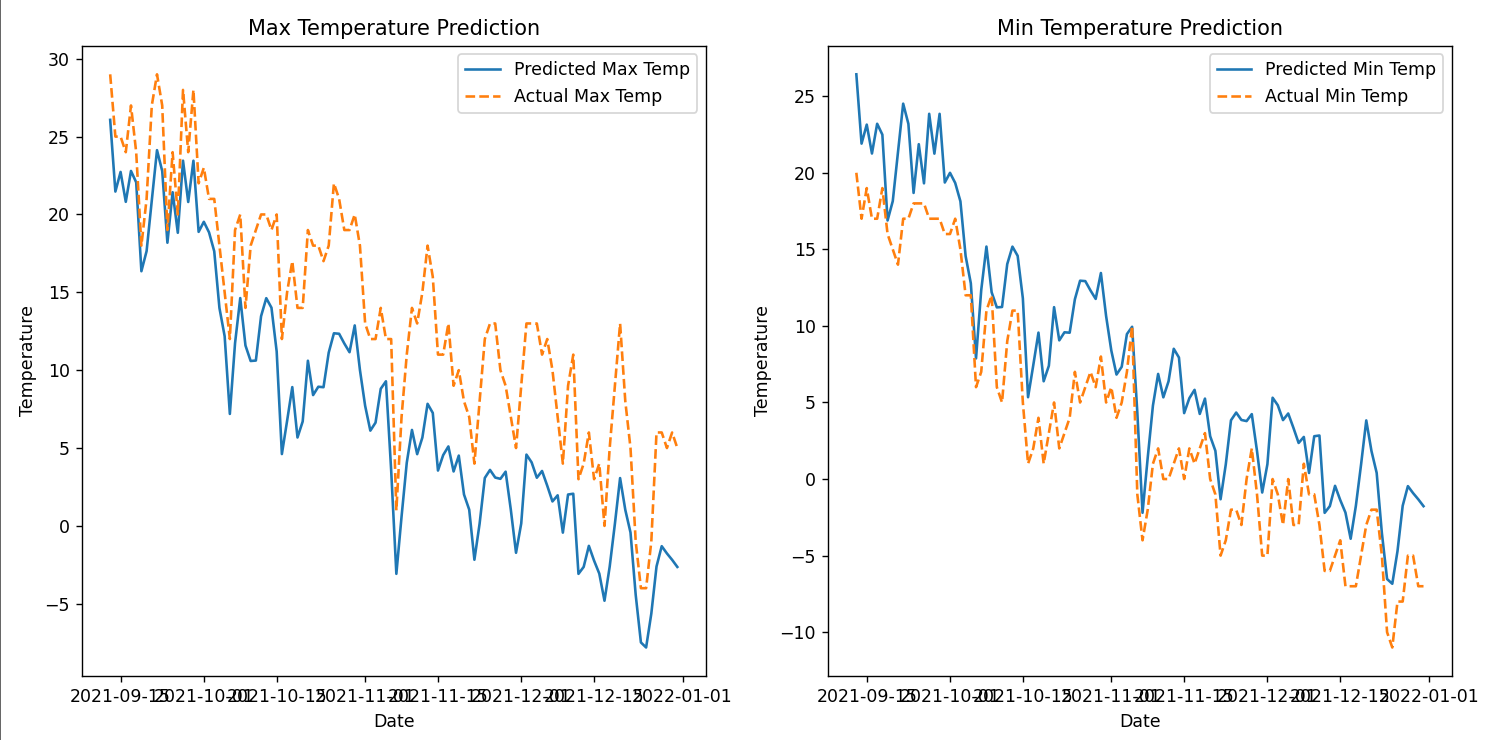


图11 LSTM模型训练过程

（3）模型训练Loss：



（4）测试结果：



## 四、结果与比较

在四种训练完成的模型上对测试集进行预测，所得到的结果如图12-图18所示。在这些图中可以得到如下结论：

（1）四个模型对所给测试集给出的预测结果均与实际值较为贴合，有较好的预测效果。

（2）LSTM模型与其他模型相比预测的误差较大，可能是该数据集不适合使用该模型。

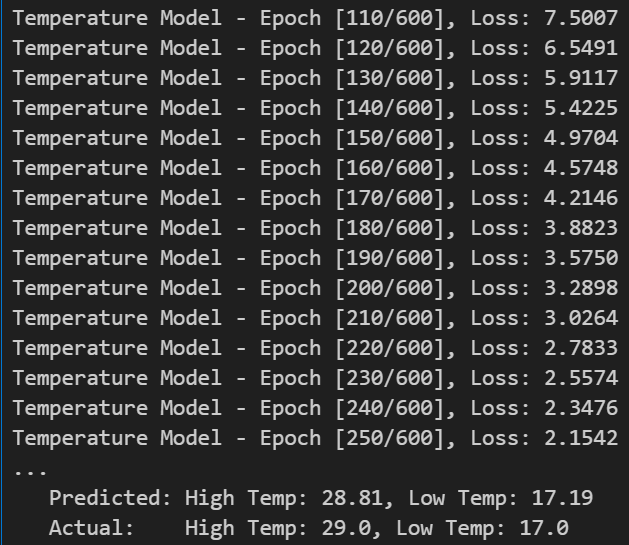


图12 线性回归模型训练过程中的损失

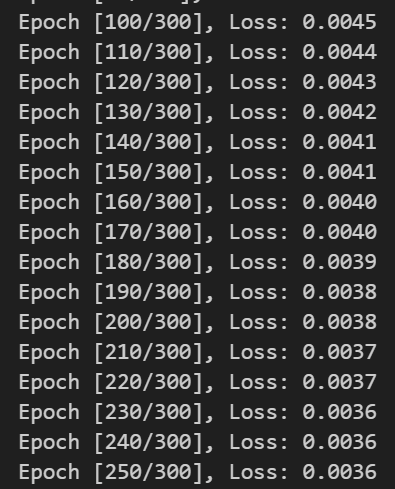


图13 RNN训练过程中的损失

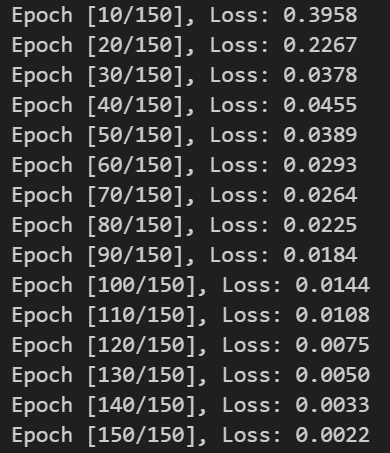


图14 LSTM训练过程中的损失

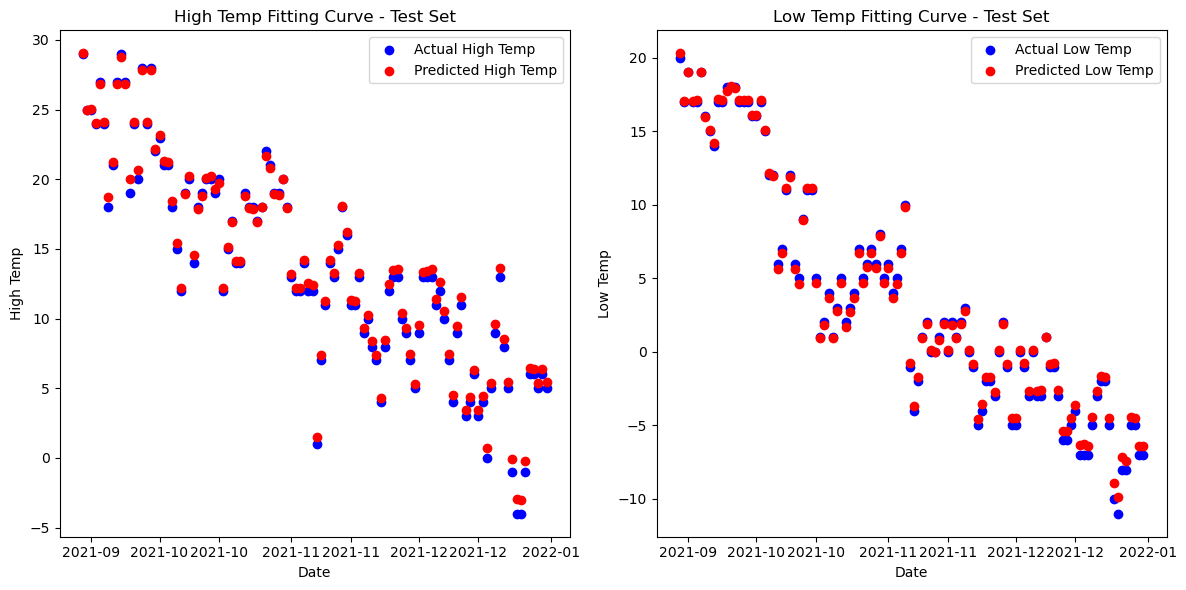


图15 线性回归模型的预测值与真实值的比较

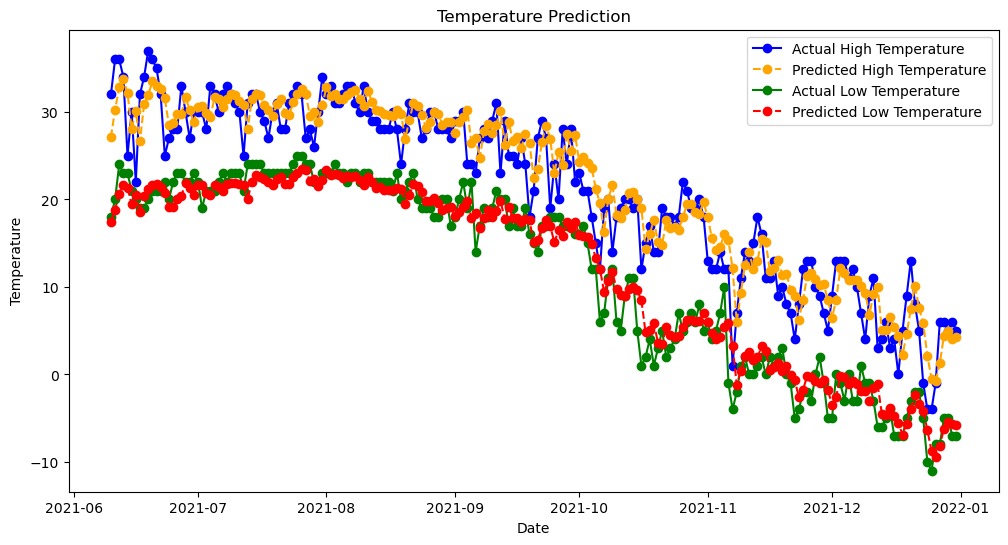


图16 RNN模型的预测值与真实值的比较

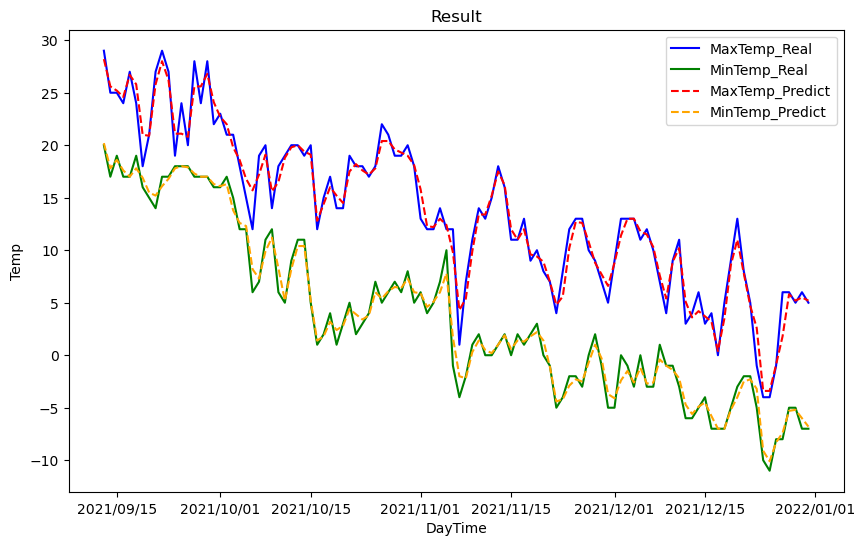


图17 RFR模型的预测值与真实值的比较

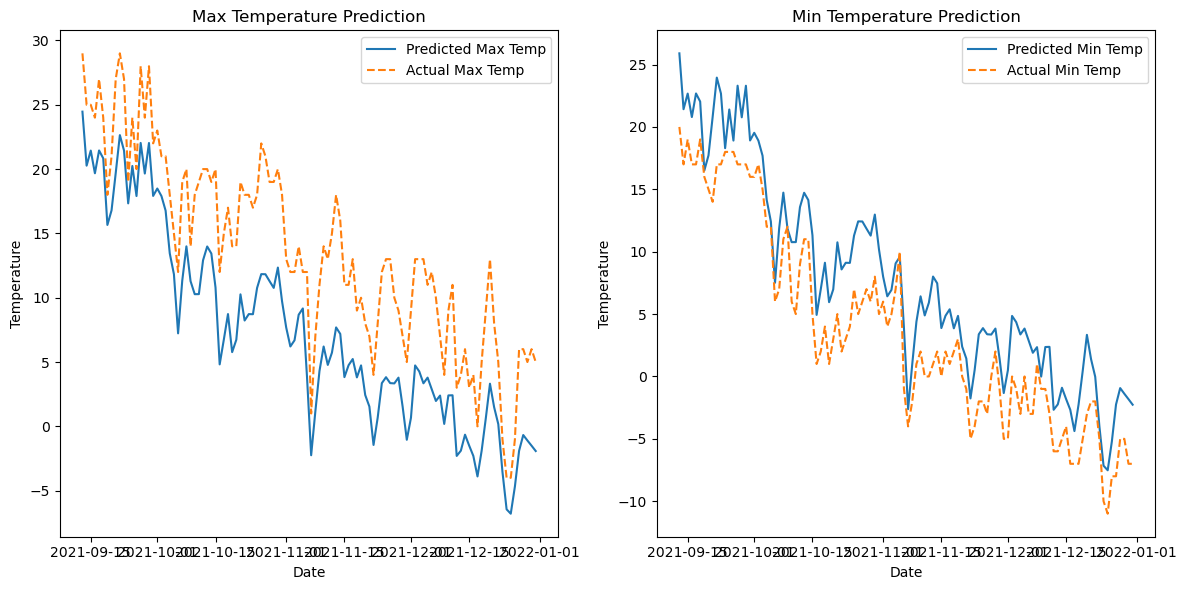


图18 LSTM模型的预测值与真实值的比较

我们还将线性回归模型和RNN模型集成在**网页**上，通过模型训练和模型调用可以得到模型训练的结果和预测的结果，以RNN模型为例，详见图19和图20。模型训练结果中有预测值与实际值的对比图，已在图16中展示。

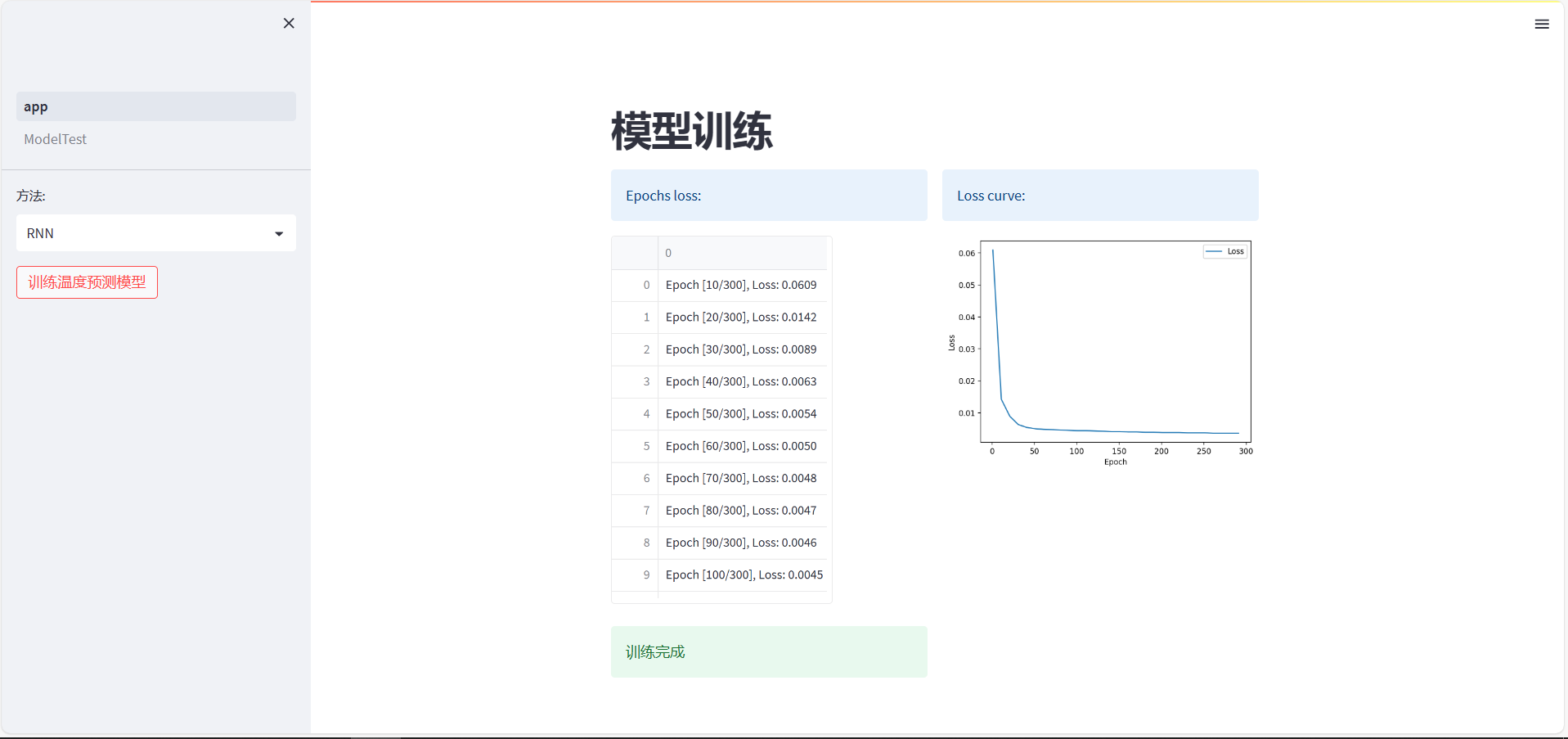


图19 RNN模型训练结果



图20 RNN模型预测部分结果

## 五、总结

对于气温这样的参数来说，主要影响因素多种多样，仅靠日期来进行预测在实际应用时误差较大。如果要进一步开展工作，需要加入其他的参数（如所在地经纬度、气压、湿度等）重新训练模型并进行预测。而在模型选取上，我们选择的四种模型虽然已经可以代表几类典型的神经网络，但是还有补充的空间。也可以再选取几种别的模型进行比较。