# 数据说明

原始数据：

E:\Data\_Lab\20170104-233型-66666次-各车号运行当日数据

实验数据：

E:\linux\_download\20170410-splitedAxis\splitedAxis

处理方式：

采样间隔10s

# 单文件训练

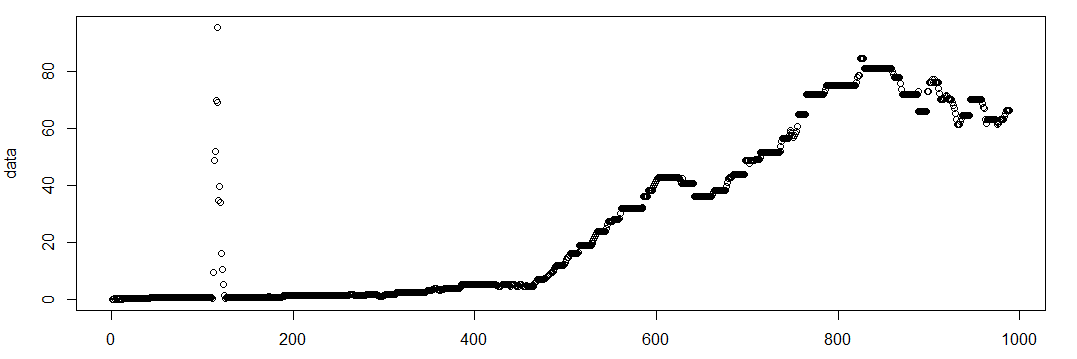
|  |  |
| --- | --- |
| 训练集 | train\_data = data[0:4000,:] |
| 训练说明 | 每批次数据是上一批次在原始数据上，向后滑动一个数据得到 |
| 测试集 | test\_data = data[4000:5000,:] |
| 测试说明 | 每批次只喂一组数据，下一次向后滑动一个数据  相当于给n\_step个数据，预测一个数据 |
| 预测效力 | 给n\_step个数据，预测一个数据 |
| LSTM参数 | # learning\_rate = 0.001 #学习率，当循环次数超过max\_epoch后会递减  # hidden\_size = 200 #隐藏层单元个数  # batch\_size = 10 #一批次处理batch\_size个时间序列，序列长度num\_steps  # num\_steps = 10 #单次处理的时间序列长度  # num\_layers = 2 #LSTM的层数  # learning\_rate = 0.001 #学习率，当循环次数超过max\_epoch后会递减  # hidden\_size = 200 #隐藏层单元个数  # batch\_size = 10 #一批次处理batch\_size个时间序列，序列长度num\_steps  # num\_steps = 10 #单次处理的时间序列长度  # num\_layers = 2 #LSTM的层数  #  # #固定参数  # init\_scale = 0.1 #初始化值的范围，初始值服从随机均匀分布  # max\_grad\_norm = 5 #用于控制梯度膨胀，如果梯度向量的L2模超过max\_grad\_norm，则等比例缩小  # max\_epoch = 4 #epoch<max\_epoch时，lr\_decay=1，epoch>max\_epoch时，lr\_decay逐渐减小  # max\_max\_epoch = 13 #整个数据循环次数  # lr\_decay = 0.5 #学习速率衰减  # n\_input = 13 #输入维度  # n\_output = 6 #输出维度  # display\_size = 100 #每喂给网络display\_size批次，输出 |
| 输出文件夹 | E:\linux\_download |
| 操作文件 | E:\linux\_download \plot\_result.R |
| 输出文件名称说明 | cost\_\*\_test\_axis\_\*  第一个\* = 数据文件开头的标号  第二个\* = 轴号  Cost表示的是输出cost，这里是mse  Test表示输出的是测试文件对应数据 |

## 13\_233\_0394\_CC-66666\_2016-05-20\_merge\_r\_ri\_10

### plot\_cost\_13\_test\_axis\_1

#### 预测差异

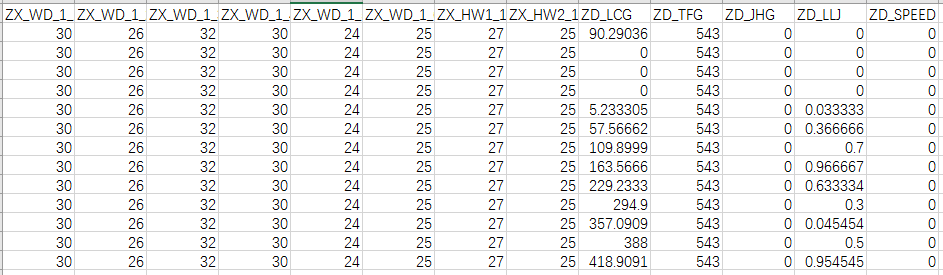
纵坐标：MSE



|  |
| --- |
|  |

#### 预测异常段对应的真实值

文件中位置：4000+10+112 到 4000+10+124

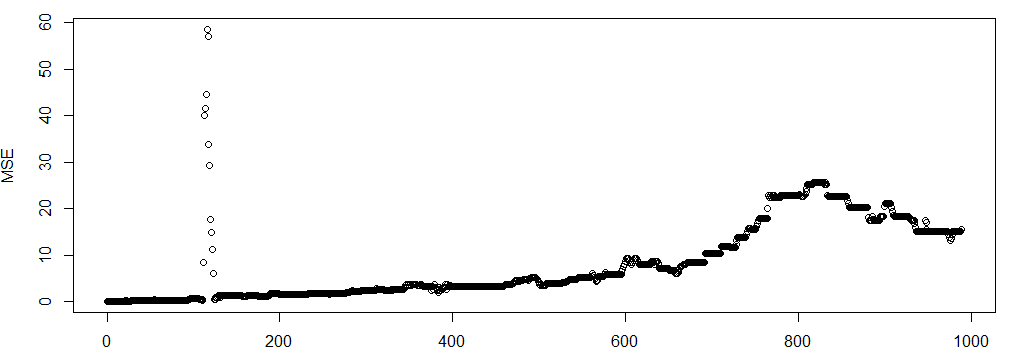


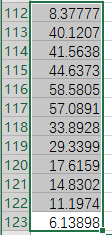
#### 分析

整体的误差趋势是合情理的

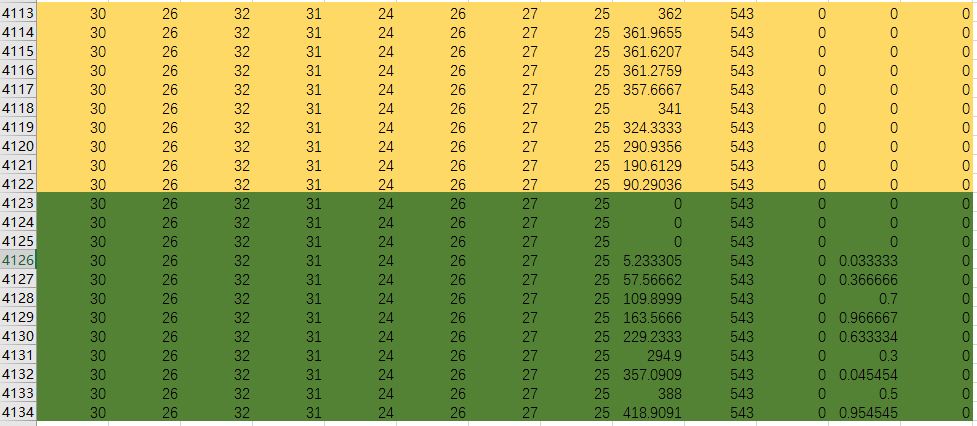
但是突然出现的预测误差原因不明

### plot\_cost\_13\_test\_axis\_2

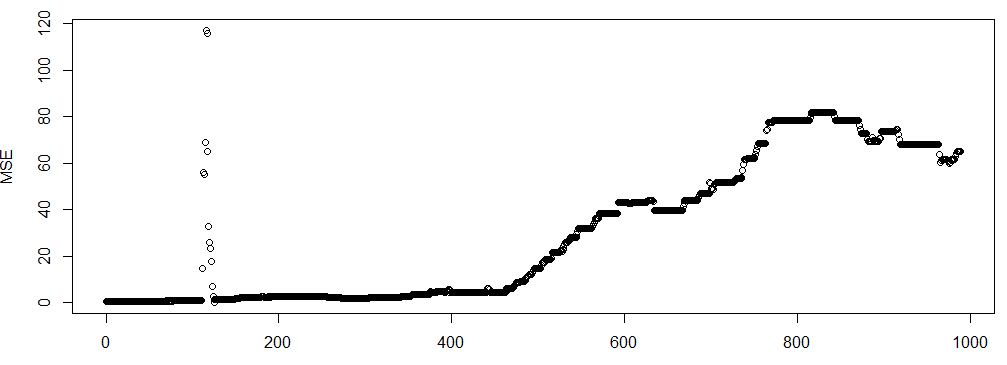


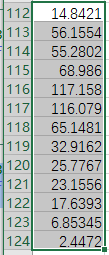




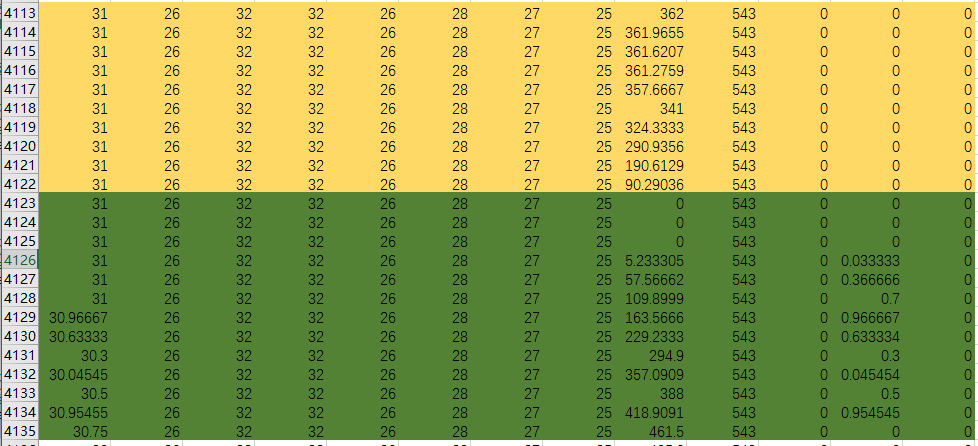


### plot\_cost\_13\_test\_axis\_3

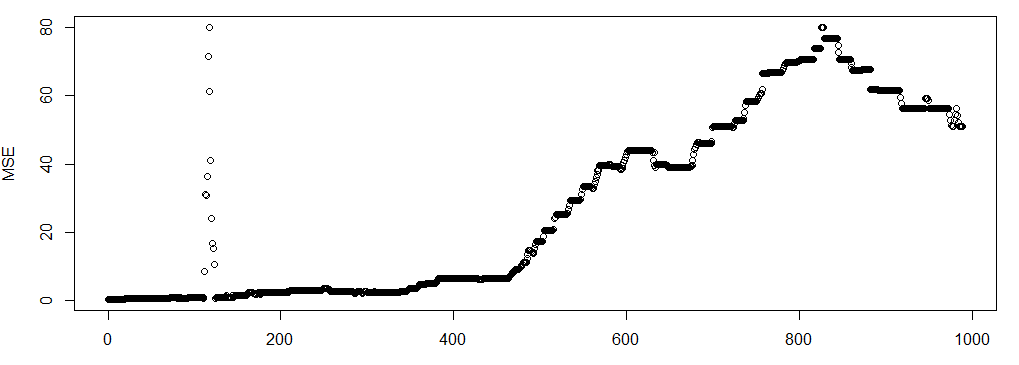


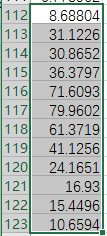




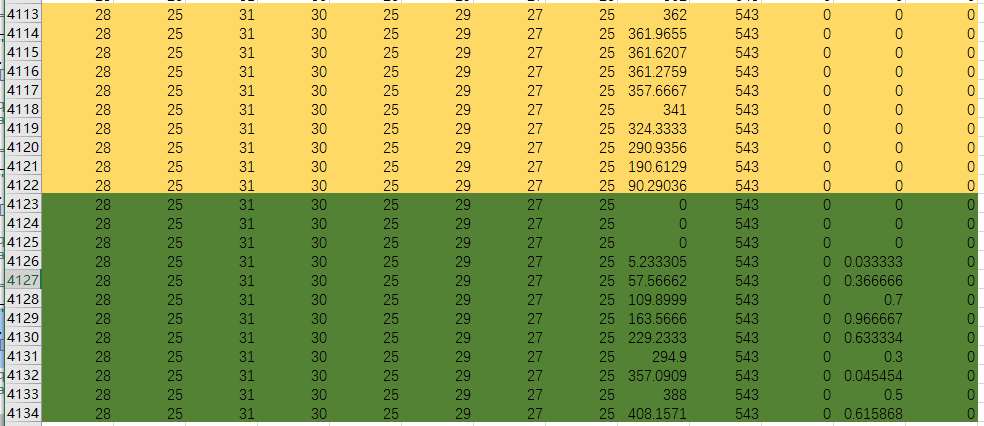


### plot\_cost\_13\_test\_axis\_4

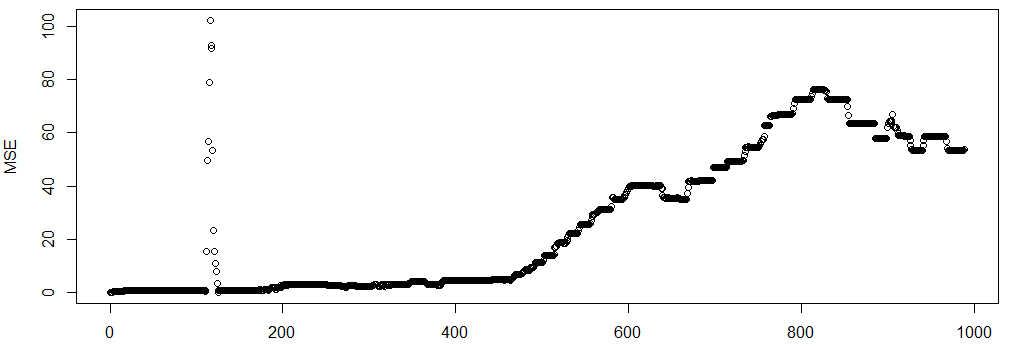






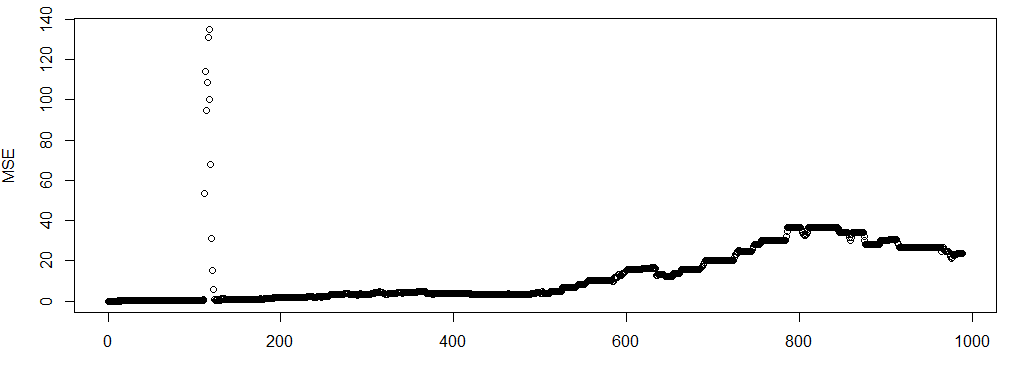


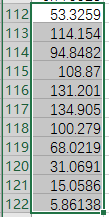
### plot\_cost\_13\_test\_axis\_5





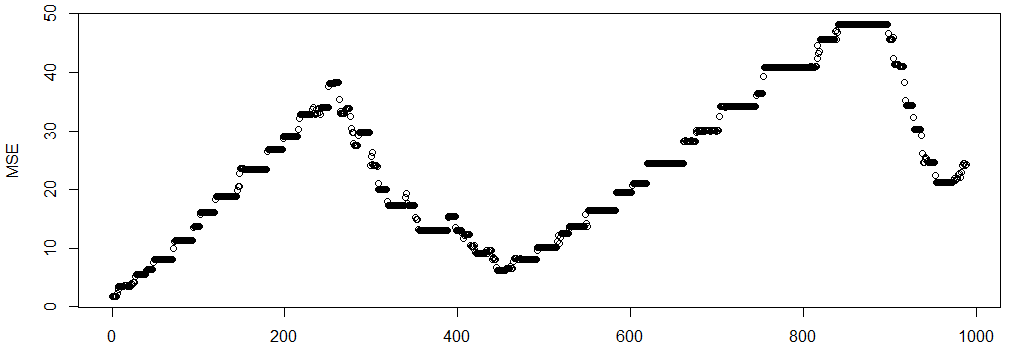
### plot\_cost\_13\_test\_axis\_6



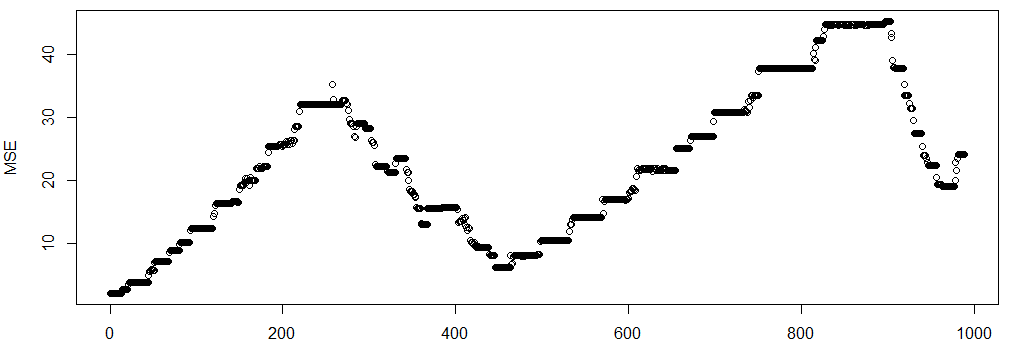


## 31\_233\_0790\_CC-66666\_2016-05-20\_merge\_r\_ri\_10.csv

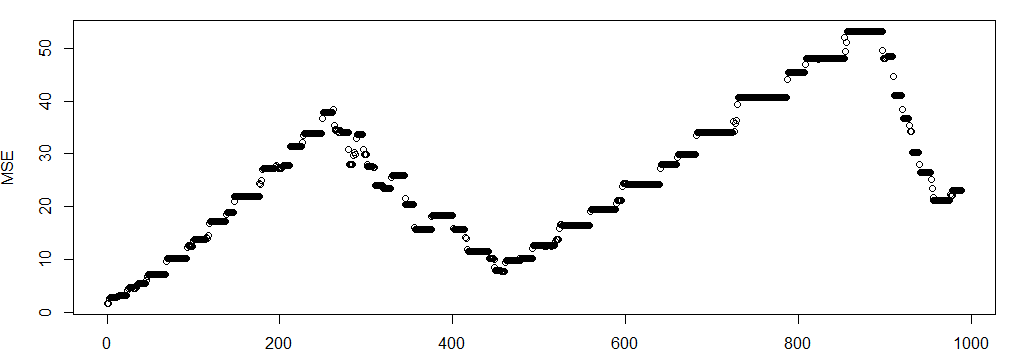
### plot\_cost\_31\_test\_axis\_1



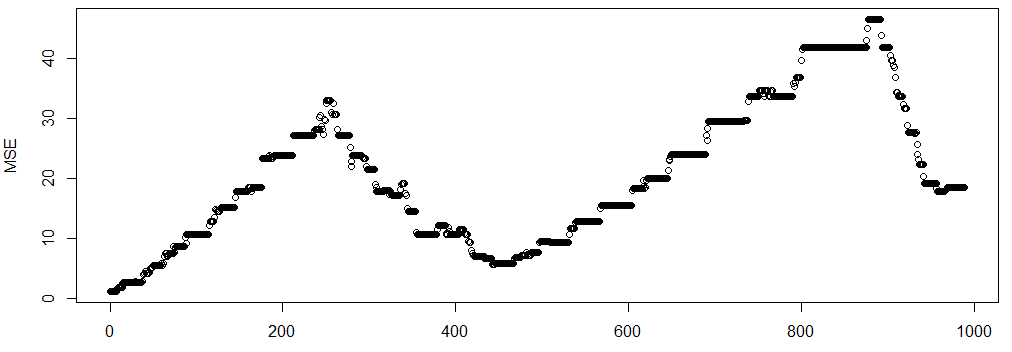
### plot\_cost\_31\_test\_axis\_2



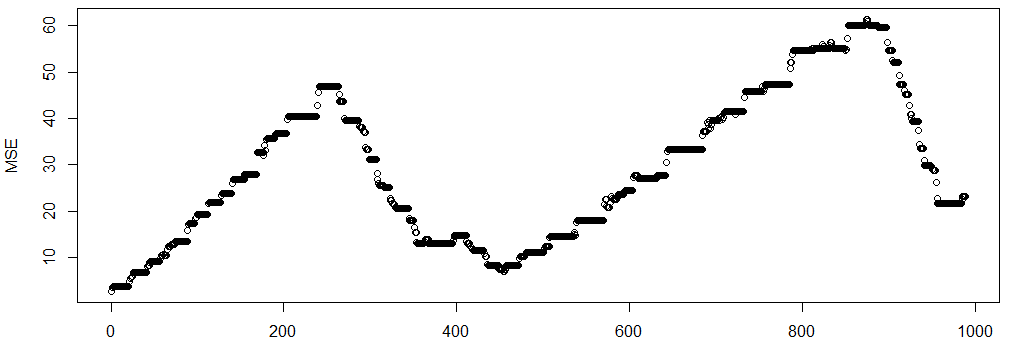
### plot\_cost\_31\_test\_axis\_3



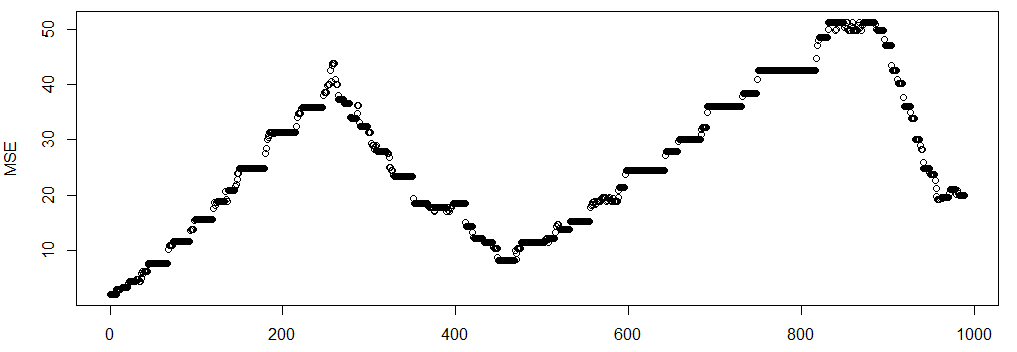
### plot\_cost\_31\_test\_axis\_4



### plot\_cost\_31\_test\_axis\_5

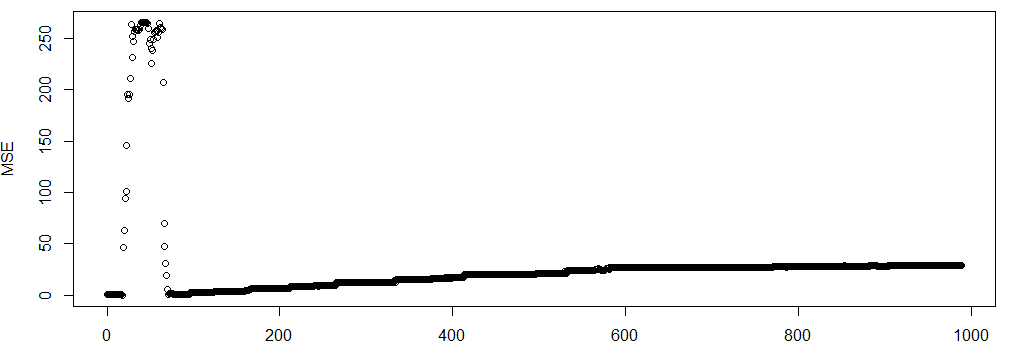


### plot\_cost\_31\_test\_axis\_6

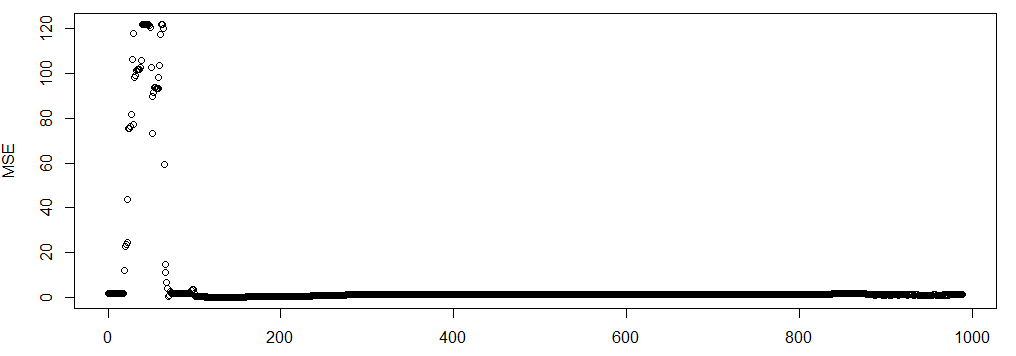


## 4\_233\_0134\_CC-66666\_2016-05-20\_merge\_r\_ri\_10\_axis

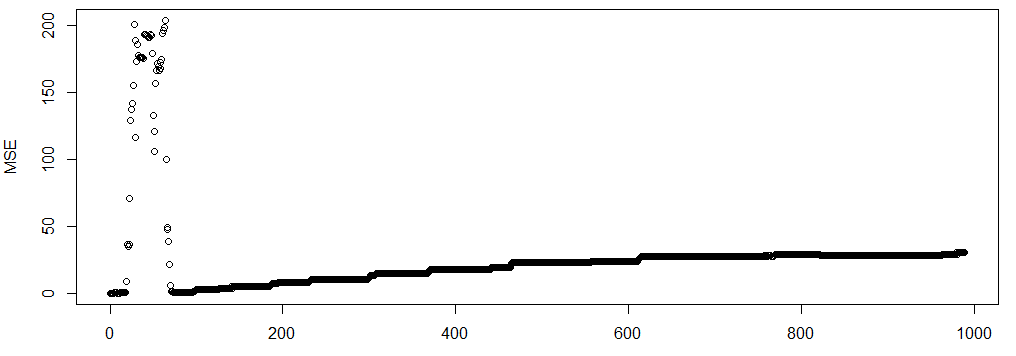
### cost\_test\_4\_axis\_1



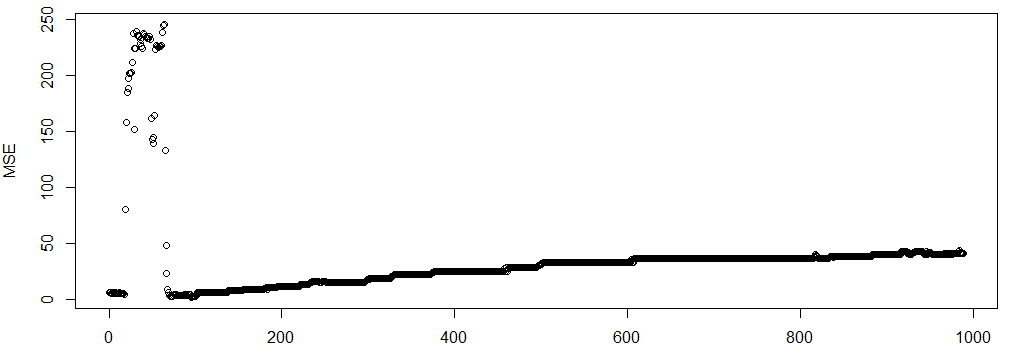
### cost\_test\_4\_axis\_2



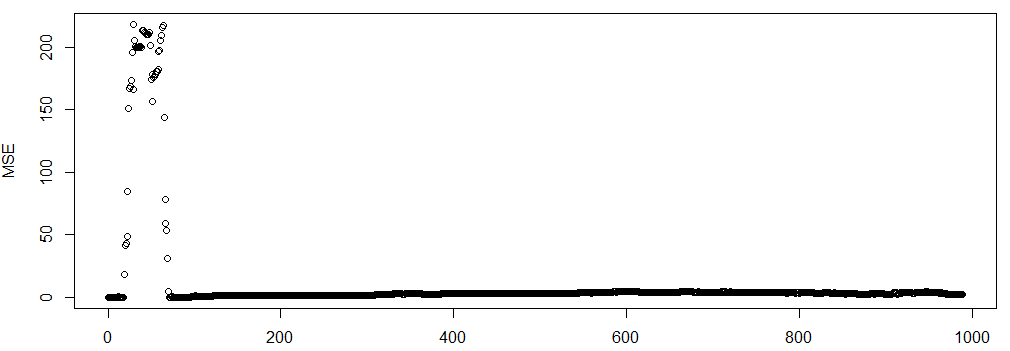
### cost\_test\_4\_axis\_3



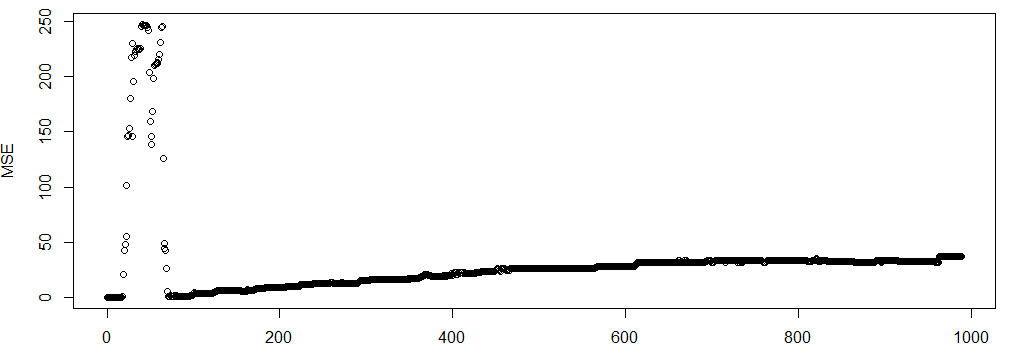
### cost\_test\_4\_axis\_4



### cost\_test\_4\_axis\_5



### cost\_test\_4\_axis\_6



## 分析

13和4号的出现明显的预测误差

检查出现误差附近的真实值，没有明显的数据变化，说明不是用于预测的数据造成的

# 单文件训练+远距预测

远距预测：

因为给定n\_step个数据，只能预测一个数据

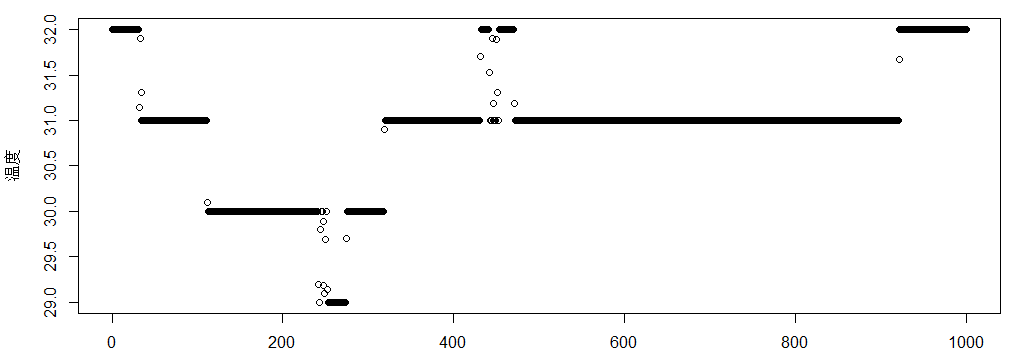
要预测很远的未来的值，就需要记录已经预测的数据，再根据预测的数据，再来预测下一个，以此类推直到到达设定的预测长度

|  |  |
| --- | --- |
| Python文件 | axismodel\_single\_lstm\_v6.py |
| 训练集 | train\_data = data[0:4000,:] |
| 训练说明 | 多个训练文件，循环处理  每批次数据是上一批次在原始数据上，向后滑动一个数据得到 |
| 测试集 | test\_data = data[4000:5000,:] |
| 测试说明 | 每批次只喂一组数据，下一次向后滑动一个数据  相当于给n\_step个数据，预测一个数据 |
| 预测效力 | 给n\_step个数据，预测一个数据  记录预测数据，加入历史数据矩阵  滑动一个数据，用预测数据，预测未来 |
| LSTM参数 | # learning\_rate = 0.001 #学习率，当循环次数超过max\_epoch后会递减  # hidden\_size = 200 #隐藏层单元个数  # batch\_size = 10 #一批次处理batch\_size个时间序列，序列长度num\_steps  # num\_steps = 10 #单次处理的时间序列长度  # num\_layers = 2 #LSTM的层数  # learning\_rate = 0.001 #学习率，当循环次数超过max\_epoch后会递减  # hidden\_size = 200 #隐藏层单元个数  # batch\_size = 10 #一批次处理batch\_size个时间序列，序列长度num\_steps  # num\_steps = 10 #单次处理的时间序列长度  # num\_layers = 2 #LSTM的层数  #  # #固定参数  # init\_scale = 0.1 #初始化值的范围，初始值服从随机均匀分布  # max\_grad\_norm = 5 #用于控制梯度膨胀，如果梯度向量的L2模超过max\_grad\_norm，则等比例缩小  # max\_epoch = 4 #epoch<max\_epoch时，lr\_decay=1，epoch>max\_epoch时，lr\_decay逐渐减小  # max\_max\_epoch = 13 #整个数据循环次数  # lr\_decay = 0.5 #学习速率衰减  # n\_input = 13 #输入维度  # n\_output = 6 #输出维度  # display\_size = 100 #每喂给网络display\_size批次，输出 |
| 输出文件夹 | E:\linux\_download |
| 操作文件 | E:\linux\_download \plot\_result.R |
| 输出文件名称说明 | cost\_test\_\*\_dif\_axis\_\*.csv  cost\_test\_\*\_obs\_axis\_\*.csv  cost\_test\_\*\_pre\_axis\_\*.csv  第一个\* = 多文件处理的实验标号  第二个\* = 轴号  Test表示输出的是测试文件对应数据  Obs表示测试数据  Pre表示预测数据  Dif表示pre-obs |

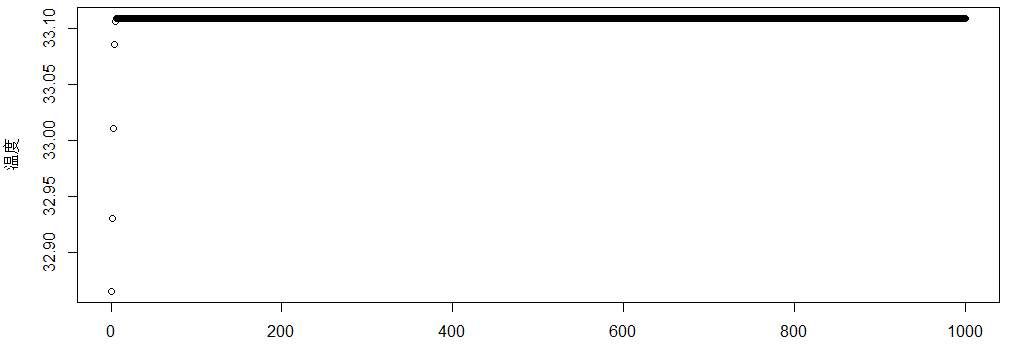
## 31\_233\_0790\_CC-66666\_2016-05-20\_merge\_r\_ri\_10.csv

### 输出图

测试数据



预测数据



### 分析

由于是多维的

要想预测轴温，就必须也预测其他所有变量

再把预测的值当做已经知道的值，再预测下一个

问题：

（1）但是有一个问题，发现以上开始预测区域，速度是0，自然的之后所有的速度都会是0

也就导致数据没有变化趋势的预测，形成水平的预测线

（2）其实轴温的变化与速度的变化是最直接相关的，但是，两者基本是同步变化的，本来远距离预测速度就会带来误差，在利用预测有误差的速度来预测温度，误差叠加，更加大

其他相关变量也是如此

思考：

应该将速度静止段剔除，因为他对预测未来的意义不大

预测时间长度的选择，应该是根据铁总那边对严重故障的反应时间

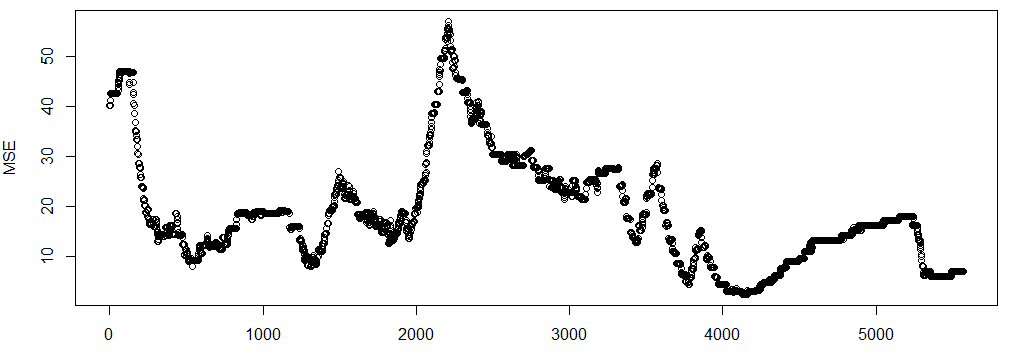
# 多文件训练

|  |  |
| --- | --- |
| 训练集 |  |
| 训练说明 | 多个训练文件，循环处理  每批次数据是上一批次在原始数据上，向后滑动一个数据得到 |
| 测试集 |  |
| 测试说明 | 每批次只喂一组数据，下一次向后滑动一个数据  相当于给n\_step个数据，预测一个数据 |
| 预测效力 | 给n\_step个数据，预测一个数据 |
| LSTM参数 | # learning\_rate = 0.001 #学习率，当循环次数超过max\_epoch后会递减  # hidden\_size = 200 #隐藏层单元个数  # batch\_size = 10 #一批次处理batch\_size个时间序列，序列长度num\_steps  # num\_steps = 10 #单次处理的时间序列长度  # num\_layers = 2 #LSTM的层数  # learning\_rate = 0.001 #学习率，当循环次数超过max\_epoch后会递减  # hidden\_size = 200 #隐藏层单元个数  # batch\_size = 10 #一批次处理batch\_size个时间序列，序列长度num\_steps  # num\_steps = 10 #单次处理的时间序列长度  # num\_layers = 2 #LSTM的层数  #  # #固定参数  # init\_scale = 0.1 #初始化值的范围，初始值服从随机均匀分布  # max\_grad\_norm = 5 #用于控制梯度膨胀，如果梯度向量的L2模超过max\_grad\_norm，则等比例缩小  # max\_epoch = 4 #epoch<max\_epoch时，lr\_decay=1，epoch>max\_epoch时，lr\_decay逐渐减小  # max\_max\_epoch = 13 #整个数据循环次数  # lr\_decay = 0.5 #学习速率衰减  # n\_input = 13 #输入维度  # n\_output = 6 #输出维度  # display\_size = 100 #每喂给网络display\_size批次，输出 |
| 输出文件夹 | E:\linux\_download |
| 操作文件 | E:\linux\_download \plot\_result.R |
| 输出文件名称说明 | cost\_test\_mulfile\_\*\_axis\_\*  第一个\* = 多文件处理的实验标号  第二个\* = 轴号  Cost表示的是输出cost，这里是mse  Test表示输出的是测试文件对应数据  Mulfile表示是多文件训练输出 |

## 13+31

|  |  |
| --- | --- |
| 训练集 | #31\_233\_0790\_CC-66666\_2016-05-20\_merge\_r\_ri\_10\_axis\_1  #13\_233\_0394\_CC-66666\_2016-05-20\_merge\_r\_ri\_10\_axis\_1 |
| 测试集 | #4\_233\_0134\_CC-66666\_2016-05-20\_merge\_r\_ri\_10\_axis\_1 |

### plot\_cost\_test\_mulfile\_1\_axis\_1



# end