

开发进程

➤ 2021.09-至今: TimeToolv2、TimeToolv3、TimeToolv4

(Web端: <http://120.26.89.97:8501/>)

- 采用Streamlit框架, 实现网页端操作
- 自动计算时间间隔
- 时序数据预处理
- 时序数据插值
- 时序数据预测



可选的预测算法:

- 1) Actual: 真实值;
- 2) Moving Average: 移动平均;
- 3) Linear Regression: 线性回归;
- 4) K-Nearest Neighbours: 最近邻;
- 5) ARIMA;
- 6) LSTM

➤ 应用: 清理云冈石窟、锁阳城的时序数据

操作面板

时序数据列选择如下:

value

数据预处理结果

初始数据个数: 1249

“Value”异常个数: 1

“Date”重复个数: 4

“Date”异常个数: 2

有效数据总数: 1243

插值算法选择如下:

Actual

真实值: 输出清洗过异常数据和重复数据, 同时按照最大频次时间间隔切片处理的真实数据结果。

☐ 下载 ?



TimeToolv4

Updated: Oct 13, 2021

选择时序数据处理功能 🤖

选择如下:

插值 (单值缺失)

您选择了: 插值 (单值缺失)

Step 1: 上传文件 🤖

←2. 上传文件 (.csv/.xlsx)



Drag and drop file here

Limit 200MB per file • CSV, XLSX

Browse files



aa温度-data-2021-08-13 16_32_39.csv 26.2KB



您选择对哪一列时序数据进行分析? 🤖

←3. 选择操作的数据列

←4. 数据预处理结果

←1. 插值/预测

value

数据预处理结果

初始数据个数: 1249

“Value”异常个数: 1

“Date”重复个数: 4

“Date”异常个数: 2

有效数据总数: 1243

插值算法选择如下:

knn_mean

最近邻插值: 当前缺失值按照前4个时刻和后4个时刻已知值的平均值进行填充。如果前4个时刻和后4个时刻都为空值, 则插值后仍为空值。

☐ 下载 ?

开始

Made by ruru@zju



aa温度-data-2021-08-13 16_32_39.csv 26.2KB



您选择对哪一列时序数据进行分析? 🤖

您选择的数据对象是: value

Step 2: 选择时序范围 🤖 ← 5. 选择时序范围进行操作

时序范围:

2021-08-11 16:26:00

2021-08-13 10:14:00

2021-08-11 16:26:00

2021-08-13 10:14:00

开始时间: 2021-08-11 16:26:00

结束时间: 2021-08-13 10:14:00

← 6. 选择插值算法

value

2021-08-11T16:26:00	<NA>
2021-08-11T16:34:00	30.8000
2021-08-11T16:36:00	30.7000
2021-08-11T16:42:00	31.0000
2021-08-11T16:46:00	30.9000
2021-08-11T16:48:00	30.8000
2021-08-11T16:50:00	30.8000
2021-08-11T16:56:00	30.5000
2021-08-11T16:58:00	30.4000
2021-08-11T17:00:00	30.4000

← 是否生成结果下载链接

数据预处理结果

初始数据个数: 1249

“Value”异常个数: 1

“Date”重复个数: 4

“Date”异常个数: 2

有效数据总数: 1243

插值算法选择如下:

knn_mean

最近邻插值: 当前缺失值按照前4个时刻和后4个时刻已知值的平均值进行填充。如果前4个时刻和后4个时刻都为空值, 则插值后仍为空值。

☒ 下载

开始

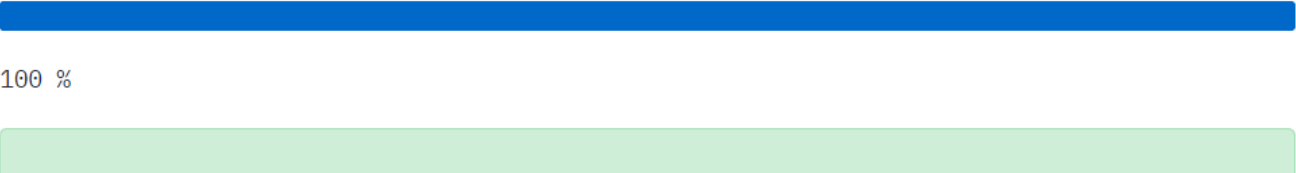
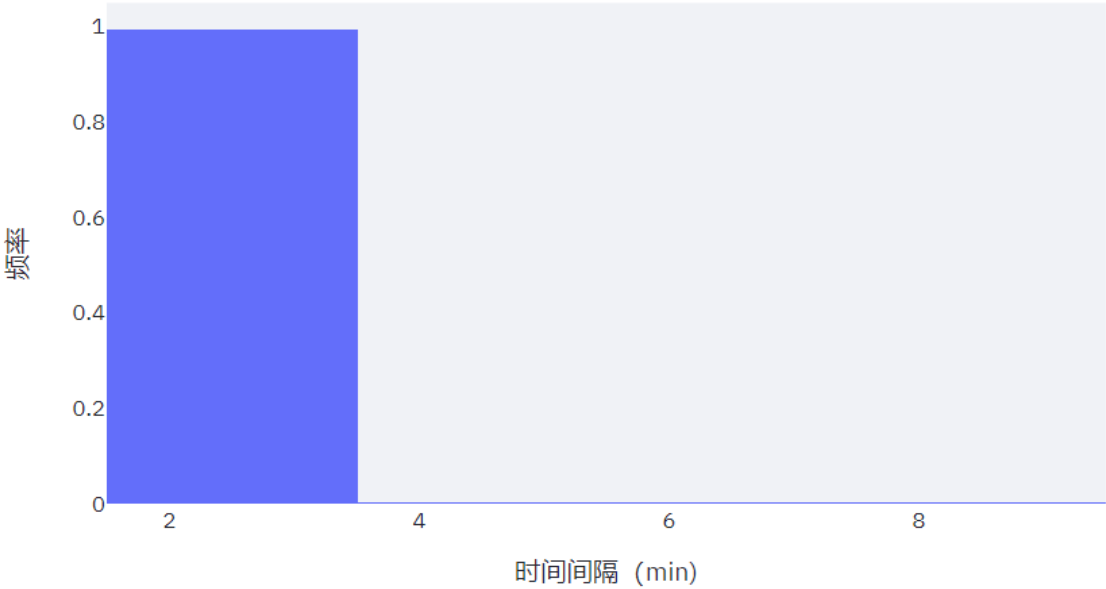
输出信息:

Step 3: 您选择哪种插值方法? 🙌

您选择了: knn_mean

时间间隔分布

← 自动输出时间间隔



数据预处理结果

初始数据个数: 1249

“Value”异常个数: 1

“Date”重复个数: 4

“Date”异常个数: 2

有效数据总数: 1243

插值算法选择如下:

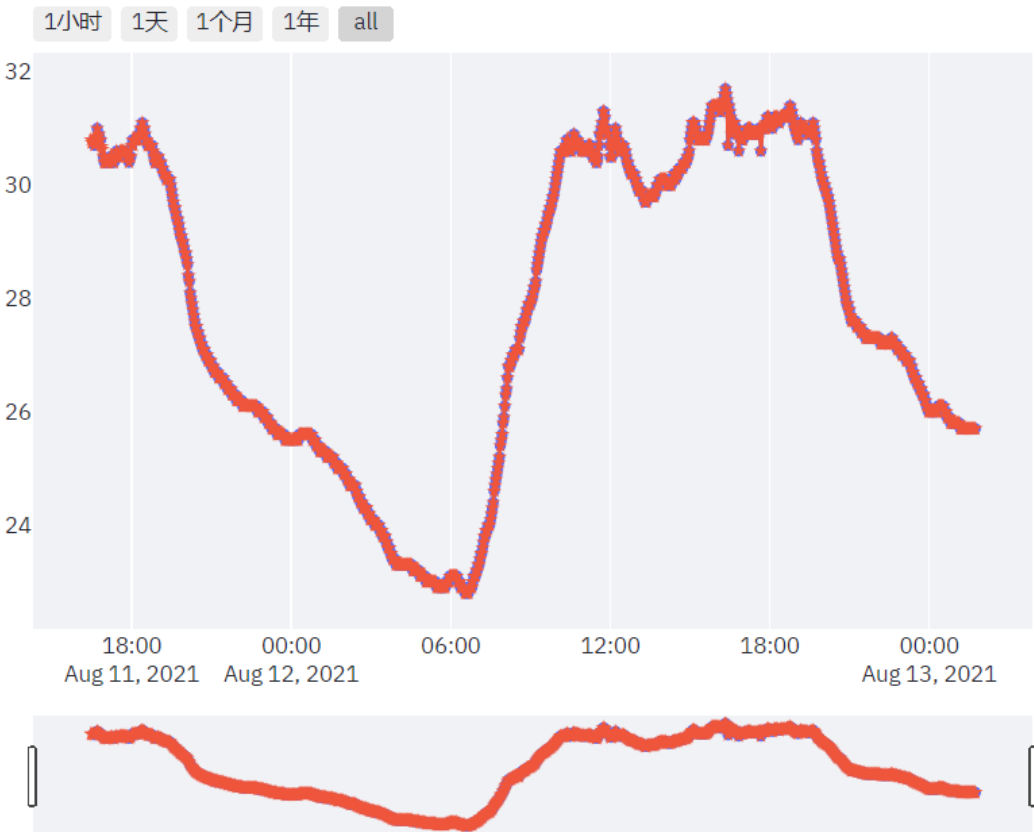
knn_mean

最近邻插值: 当前缺失值按照前4个时刻和后4个时刻已知值的平均值进行填充。如果前4个时刻和后4个时刻都为空值, 则插值后仍为空值。

✓ 下载 ?

开始

输出信息:



← 结果图

查看完整图像 (注意:这可能会花费一些时间...)

+

knn_mean

最近邻插值：当前缺失值按照前4个时刻和后4个时刻已知值的平均值进行填充。如果前4个时刻和后4个时刻都为空值，则插值后仍为空值。

✓ 下载

开始

输出信息:

缺失值个数: 13

时间间隔: 2 Min

最近邻插值 (knn_mean) 均方差:0.00518560。

http://120.26.89.97:8501/downloads/20220106205846knn_mean.csv

生成文件的下载链接

Made by ruru@zju

查看完整图像 (注意:这可能会花费一些时间...)

Step 4: 输出文件

插值后文件表格

	value	missing	knn_mean	"Value"异常个数: 1: "Date"...
2021-08-11T16:26:00	<NA>	连续缺失	<NA>	
2021-08-11T16:28:00	<NA>	缺失	30.8000	<NA>
2021-08-11T16:30:00	<NA>	缺失	30.7500	<NA>
2021-08-11T16:32:00	<NA>	缺失	30.7500	<NA>
2021-08-11T16:34:00	30.8000	未缺失	30.8000	<NA>
2021-08-11T16:36:00	30.7000	未缺失	30.7000	<NA>
2021-08-11T16:38:00	<NA>	缺失	30.8333	<NA>
2021-08-11T16:40:00	<NA>	缺失	30.8500	<NA>
2021-08-11T16:42:00	31.0000	未缺失	31.0000	<NA>
2021-08-11T16:44:00	<NA>	缺失	30.8400	<NA>
2021-08-11T16:46:00	30.9000	未缺失	30.9000	<NA>

See the entire output chart

Step 5: 文件下载链接

插值后文件表格下载

http://120.26.89.97:8501/downloads/20220106205846knn_mean.csv

Done !



“Date”重复个数: 2

“Date”异常个数: 2

有效数据总数: 1245

插值算法选择如下:

Linear_Interpolation

线性插值: 通过连接相邻两个已知值的直线, 以对在这两个已知值之间的缺失值进行填充。

预测算法选择如下:

K_Nearest_Neighbours

最近邻: 通过衡量时间相关变量, 例如: 年, 月, 日, 星期, 时间戳等与因变量的关系, 对因变量进行聚类, 从而预测因变量的值。

☐ 下载 ?

开始



TimeToolv4

Updated: Oct 13, 2021

选择时序数据处理功能 🤝

选择如下:

插值 (连续缺失)

您选择了: 插值 (连续缺失)

Step 1: 上传文件 🤝

Drag and drop file here
Limit 200MB per file • CSV, XLSX

Browse files

aa温度-data-2021-08-13 16_32_39.xlsx 30.7KB

×

有效数据总数: 1245

插值算法选择如下:

Linear_Interpolation

线性插值: 通过连接相邻两个已知值的直线, 以对在这两个已知值之间的缺失值进行填充。

预测算法选择如下:

K_Nearest_Neighbours

最近邻: 通过衡量时间相关变量, 例如: 年, 月, 日, 星期, 时间戳等与因变量的关系, 对因变量进行聚类, 从而预测因变量的值。

☐ 下载

开始

输出信息:

缺失值个数: 548

时间间隔: 2 Min

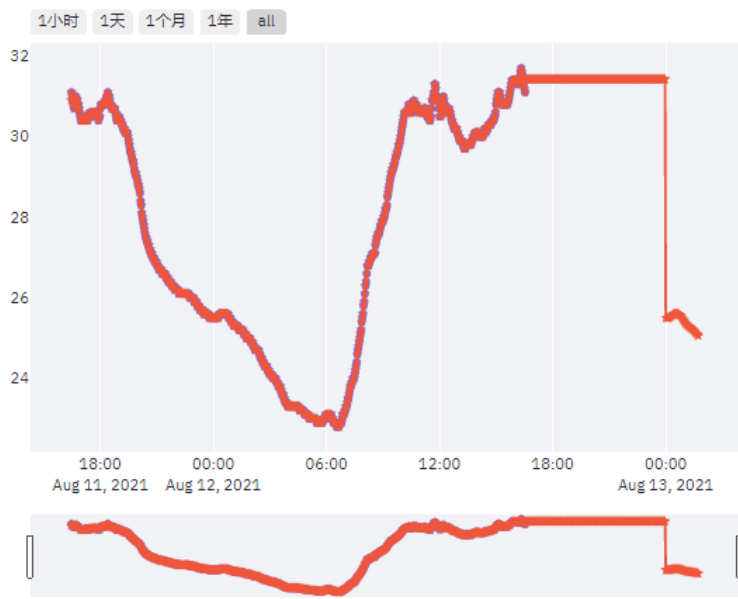
线性插值 (Linear Interpolation) 均方差: 0.00167778。

预测值个数: 532

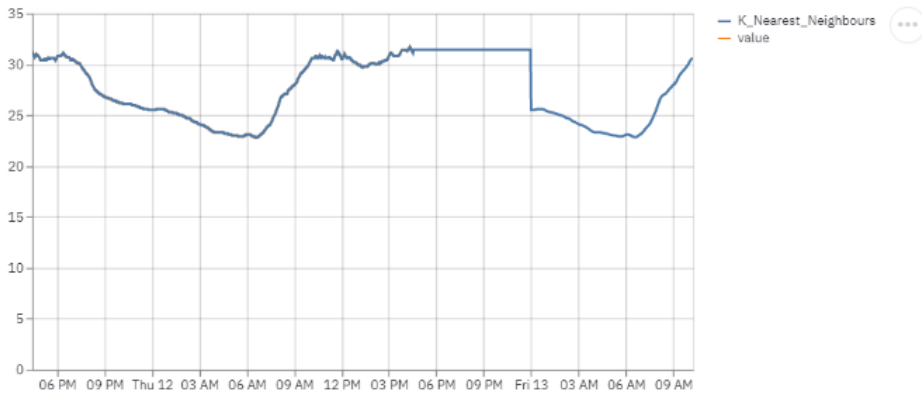
K Nearest Neighbours均方根误差(RMSE):

0.05556823276290885

Made by ruru@zju



← 结果图



有效数据总数: 1245

插值算法选择如下:

Linear Interpolation

线性插值: 通过连接相邻两个已知值的直线, 以对在这两个已知值之间的缺失值进行填充。

预测算法选择如下:

K_Nearest_Neighbours

最近邻: 通过衡量时间相关变量, 例如: 年, 月, 日, 星期, 时间戳等与因变量的关系, 对因变量进行聚类, 从而预测因变量的值。

下载

开始

输出信息:

缺失值个数: 548

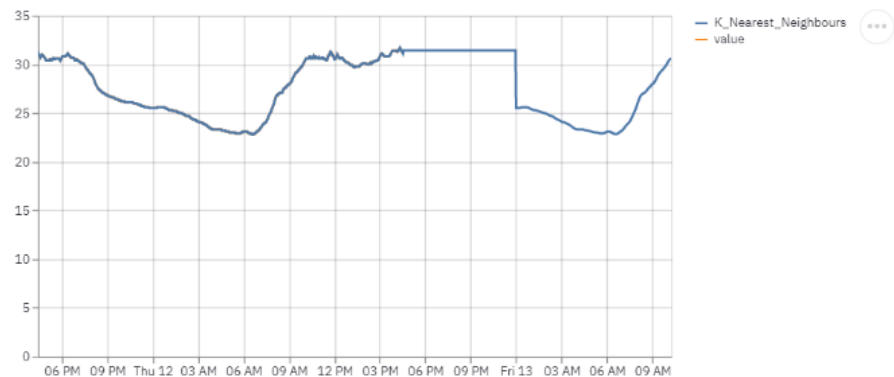
时间间隔: 2 Min

线性插值 (Linear Interpolation) 均方差: 0.00167778。

预测值个数: 532

K Nearest Neighbours均方根误差(RMSE): 0.05556823276290885

Made by ruru@zju



查看完整图像 (注意:这可能会花费一些时间...)

Step 5: 输出预测后的文件 📁

	value	predict	K_Nearest_Neighbours	"Value"异常个数: 1: "Dat
2021-08-11T16:26:00	<NA>	预测	30.9000	
2021-08-11T16:28:00	31.1000	未缺失	31.1000	<1
2021-08-11T16:30:00	31.0000	未缺失	31.0000	<1
2021-08-11T16:32:00	30.9000	未缺失	30.9000	<1
2021-08-11T16:34:00	30.8000	未缺失	30.8000	<1
2021-08-11T16:36:00	30.7000	未缺失	30.7000	<1
2021-08-11T16:38:00	30.8000	未缺失	30.8000	<1
2021-08-11T16:40:00	30.9000	未缺失	30.9000	<1
2021-08-11T16:42:00	31.0000	未缺失	31.0000	<1
2021-08-11T16:44:00	30.9500	未缺失	30.9500	<1
2021-08-11T16:46:00	30.9000	未缺失	30.9000	<1

Done !

← 预测后文件表格