Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting （2021）

Abstract:

降雨量预测（每小时）

研究三种 LSTM 网络结构的适用性：LSTM, Stacked-LSTM and Bidirectional-LSTM

对比 LSTMs，XGBoost 和基于 AutoML 的模型的性能

提出了一种基于双向 LSTM 网络的模型，用于使用英国五个主要城市的时间序列数据按小时预测降雨量

整理了神经网络在降雨量适用的基本原理：不确定性可以设计模型中的最后一层（通常是 softmax）来产生概率，时空依赖性可以用神经网络来学习输入与目标相关联的潜在特征，离散分布

背景知识：介绍前人关于机器学习降雨量的研究，RNN 对梯度消失很敏感，LSTM 优化了这一点。LSTM 可以学习顺序数据，用双向 RNN 的 LSTM 单元可以产生双向 LSTM，可以学习先前和未来的依赖关系

数据集：Openweather，预处理，删除一些无效特征比如海面和地面测量值，构成 5 个数据集，每个有 43 个特征（用 Pearson 删除高相关特征）

预测方法：1. XGBoost 作为一个基准线，Perform a non-exhaustive hyperparameter grid search to obtain the best hyperparameter values to train each model

2. AutoML 对每个数据集推荐一个模型

3. 对于每个数据集训练 2 个 LSTM 和一个 stack-LSTM

4. 找出对于不同数据集性能和准确性最好的那个, 来开发 2 个 stacked LSTM Network 和一个 Bidirectional LSTM Network. (在基于 Stacked-LSTM 网络的两个建议模型中，一个将包含两个隐藏层，另一个将包含三个隐藏层。拟议的双向 LSTM 网络模型将仅包含一个隐藏层。)

（本研究不找参数和超参数的最佳值, XGBoost 进行了朝着缩小 RMSE 的方向进行 20 次迭代，LSTM 用了一个叫 Adam 的优化器和随机梯度下降（SGD）方法，找到了参数）

疑问：选择参数方法不同，Adam 优化器暂且不知，SGD XGBoost 也可以用。

XGBoost 和 AutoML: 训练集和测试集的百分比分别为 85% 和 15%

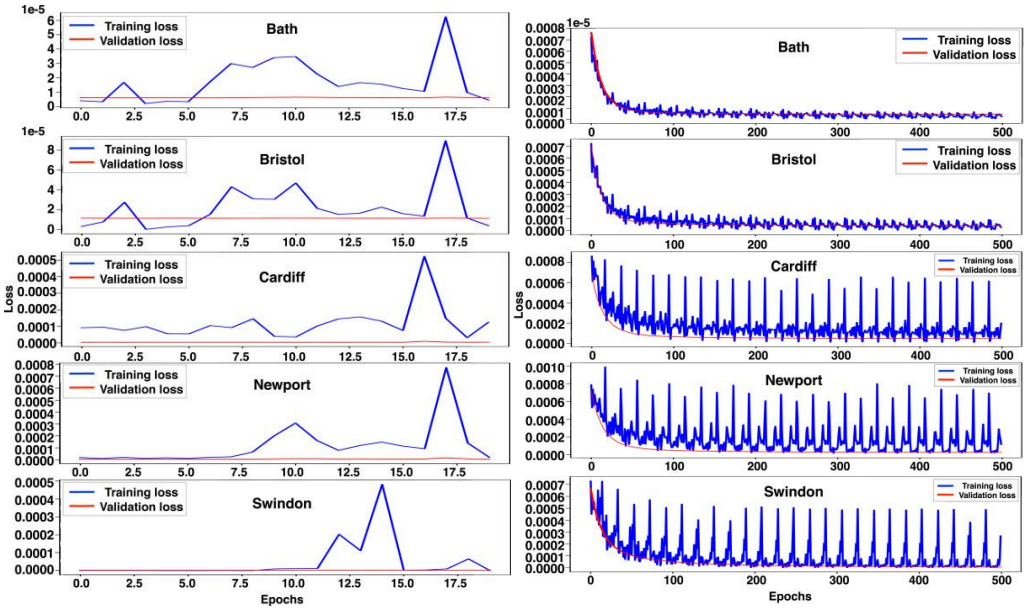
评价指标：loss function, RNSE， MAE， RMSLE

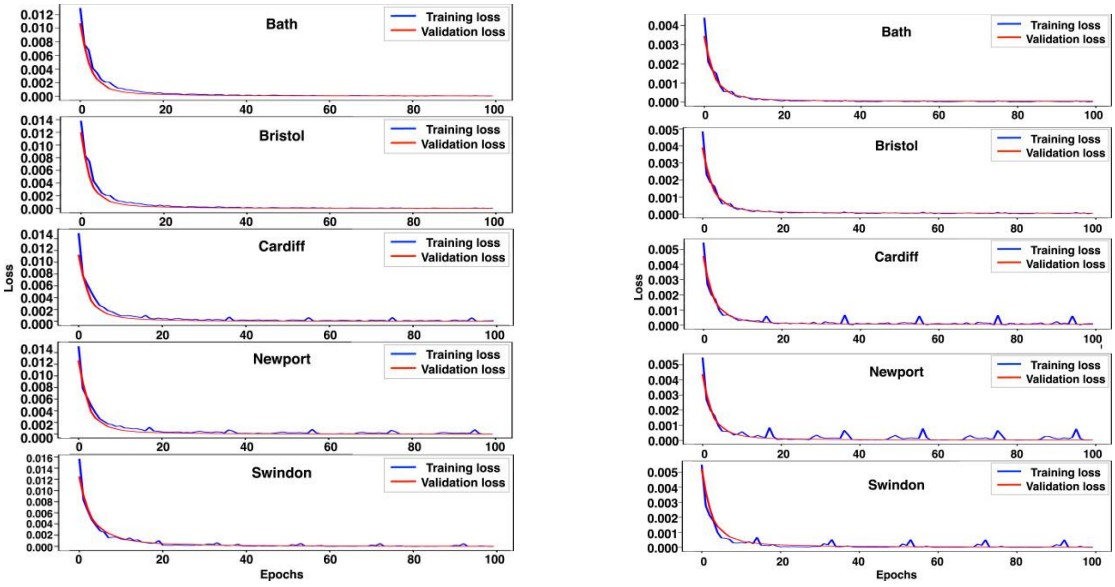
实验：选用了 Model3, LSTMs 表现远优于 XGBoost 或 AutoML(一个 2 位小数，一个 4 位小数)， 然后再训练了一下，Model4,6 > 3

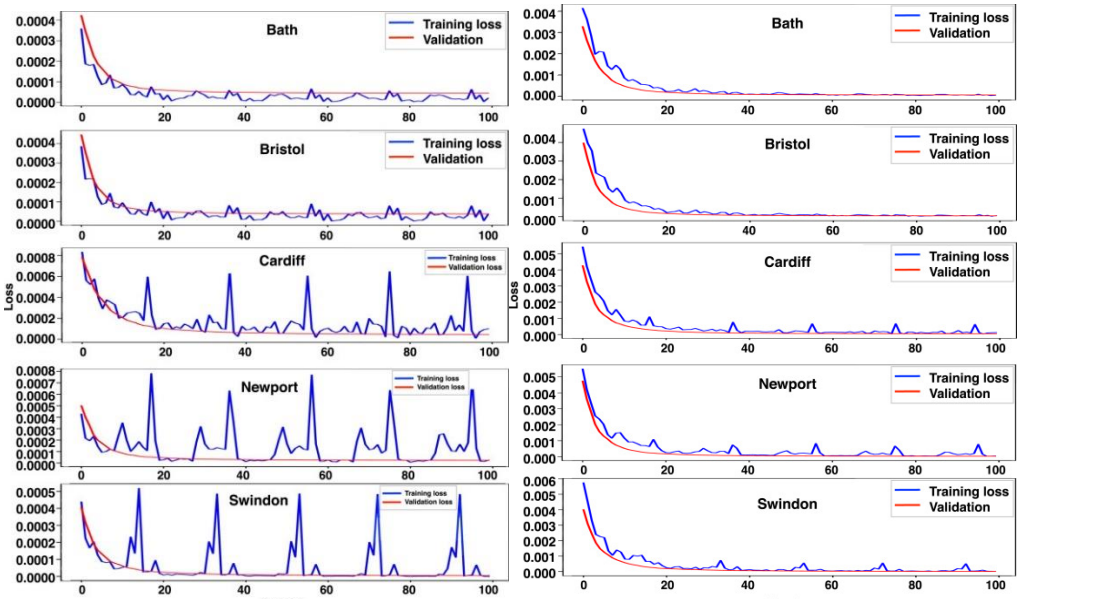
结论：双向 LSTM 网络可用作降雨预报模型，其性能与具有两个隐藏层的堆叠式 LSTM 网络相当

具有大量隐藏层的 LSTM 神经网络不太适合学习天气时间序列的奇异性来预测每小时降雨量值。

与使用基于 LSTM 网络的模型的其他方法类似，双向 LSTM 主要缺点是无法充分泛化。大多数情况下，模型会过度拟合训练数据


Bath
Bristol
Cardiff
Newport
Swindon
Training loss
Validation loss
Loss
Epochs


Bath
Bristol
Cardiff
Newport
Swindon
Training loss
Validation loss
Loss
Epochs


Bath
Bristol
Cardiff
Newport
Swindon
Training loss
Validation
Loss
Epochs

Time series predicting of COVID-19 based on deep learning（2022）

Abstract:

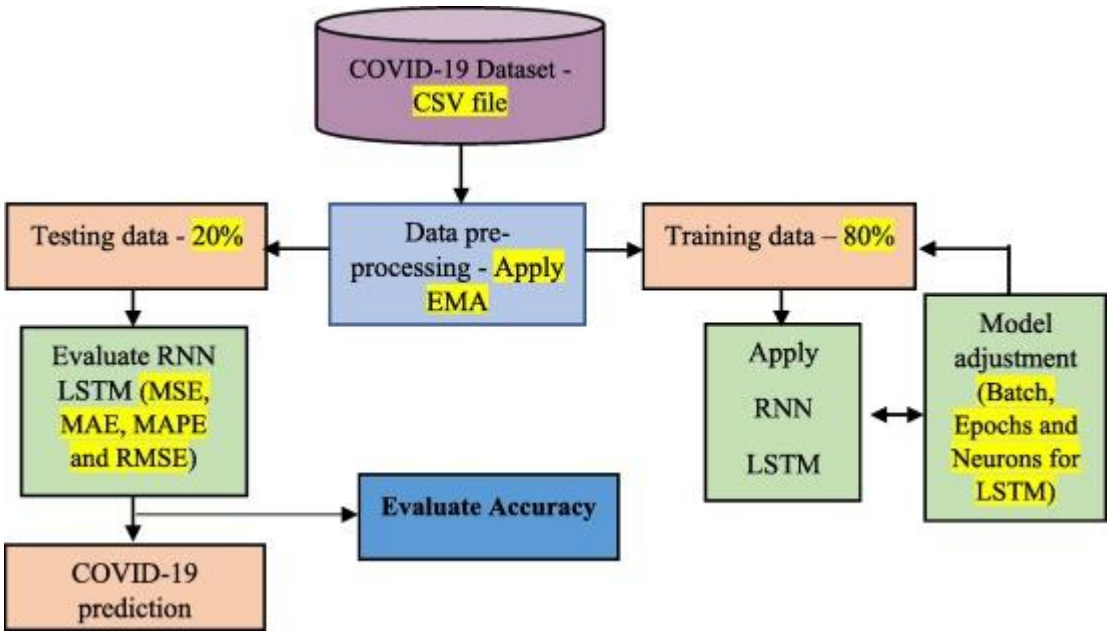
用公共数据开发了一个预测模型，来预测 COVID 在马来西亚，摩洛哥，沙特的传播情况

提出一个基于 RNN 和 LSTM 的深度学习方法

比较和评估 RNN 和 LSTM 的性能

使用了能得到最佳性能的激活函数（LSTM-ReLu, RNN-tanh,sigmoid）

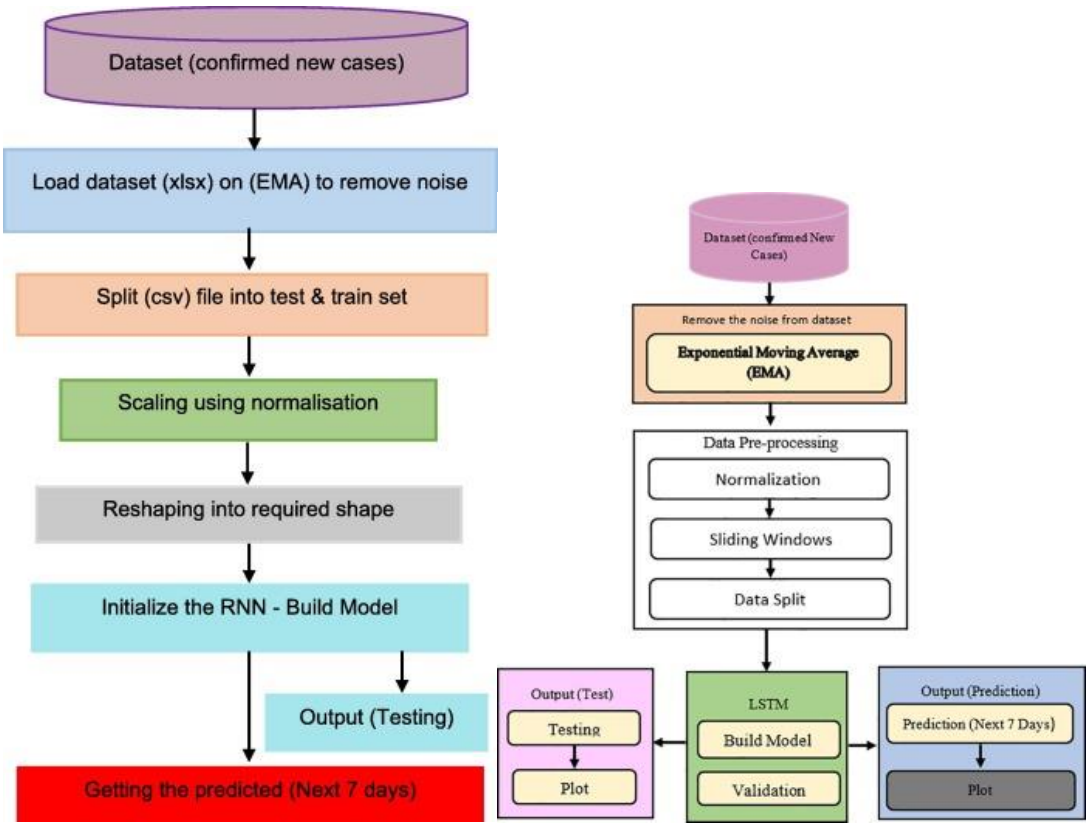
方　　法　　：



讲了 RNN 和 LSTM 的具体流程

结论：

基于 LSTM 的预测模型显示出 98.53% 的准确率，与 RNN 模型显示的准确率相比提高了 5.13%。



确定题目：在 XX 数据集下，

核心思路：以 Towards understanding the importance of time-series features in automated

algorithm performance prediction 这篇论文为主， Feature Importance Explanations for Temporal Black-Box Models 为辅。

方法：1. 数据处理：XX 数据集，LSMT 预测，计算 MSE

用切窗口法精简数据，tsfresh 提取特征

后面同 Towards

选择数据集：UCL 数据集，挑选 time series 和 regression，在 50 个候选中选择

较新的数据集：Metro Interstate Traffic Volume Data Set， 2019

AI4I 2020 Predictive Maintenance Dataset Data Set，2020

Pedal Me Bicycle Deliveries Data Set，2021

Image Recognition Task Execution Times in Mobile Edge Computing Data Set，2020

Power consumption of Tetouan city Data Set，2021

Beijing Multi-Site Air-Quality Data Data Set，2019