

卷积 LSTM 网络：降水临近预报的机器学习方法

一、问题

降水临近预报一直是天气预报领域的一个重要问题，这项任务的目标是在相对较短的时间内（例如 0-6 小时）准确及时地预测当地未来某个时间段的降雨强度。这项任务的意义重大，对于某个地区的精确预测可以有效避免因降雨引发的泥石流、洪水等自然灾害，同时还可以帮助人们做好必要的防范措施，但由于降水临近预报所需的预测分辨率和时间精度远高于周平均气温预测等其他传统预测任务，因此这个问题相当具有挑战性，已成为气象界的热门研究课题。

以前很少有人从机器学习的角度来研究这个至关重要且具有挑战性的天气预报问题。而作者开创性的将机器学习引入降水临近预报问题中，将降水临近预报问题描述为一个时空序列预测问题，其中输入是过去的雷达图像，输出是未来雷达图像。用公式表达，即：

$$\tilde{\mathcal{X}}_{t+1}, \dots, \tilde{\mathcal{X}}_{t+K} = \arg \max_{\mathcal{X}_{t+1}, \dots, \mathcal{X}_{t+K}} p(\mathcal{X}_{t+1}, \dots, \mathcal{X}_{t+K} | \hat{\mathcal{X}}_{t-J+1}, \hat{\mathcal{X}}_{t-J+2}, \dots, \hat{\mathcal{X}}_t) \quad (1)$$

通过 t 时刻之前 J 张雷达回波图像，在 t 时刻进行对未来 K 张雷达回波图像进行预测，选一组预测最准的图像当作预测值，其中每个 X 都有三个维度 (P, M, N)， M 和 N 对应的是图像的宽度和长度， P 则对应该地区雷达回波的强度，因此，可以将 X 转换为一个 3D 张量 $X \in \mathbb{R}^{P \times M \times N}$ ，如图：

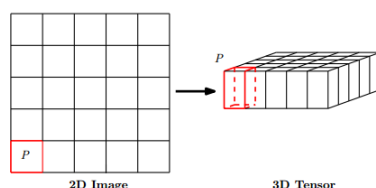


Figure 1: Transforming 2D image into 3D tensor

我们的问题又变成了利用 t 时刻之前的 J 个 X 张量预测未来 K 个 X 张量，又因为雷达回波图像的强度 (dbz) 和降水有一个 $Z-R$ 关系，所以有了未来的回波强度，就可以知道未来的降雨量。因此如何解决这个时空序列预测问题便成为了作者集中讨论的问题。

二、前人的工作

因降水临近预报本身的重要性，研究人员已经对它做了相当多的研究，提出了许多方法。目前，现代的降水临近预报方法大致分为两类：数值天气预报 (NWP) 和雷达外推方法。

数值天气预报 (NWP) 是一个根据大气实况，在一定的初值和边界条件下，通过大型计算机作数值计算，求解描写天气演变过程的流体力学和热力学的方程组，预测未来一定时间段的大气运动状态和天气现象的方法。数值天气预报与经典的以天气学方法作天气预报不同，它是一种定量的和客观的预报，正因为如此，数值天气预报首先要求建立一个较好的反映预报时段的（短期的、中期的）数值预报模式和误差较小、计算稳定并相对运算较快的计算方法。其次，由于数值天气预报要利用各种手段（常规的观测，雷达观测，船舶观测，卫星观测等）获取气象资料，因此，必须恰当地作气象资料的调整、处理和客观分析。第三，由于数值天气预报的计算数据非常之多，很难用手工或小型计算机去完成，因此，必须要有大型的计算机。所以，目前最先进的操作降水预报系统通常采用更快、更准确的基于雷达回波外推的方法。

雷达回波外推方法利用雷达探测到的回波数据，确定回波的强度分布及回波体移动速度和方向，通过对回波体进行线性或非线性的外推，预报一定时间段后的雷达回波状态。具体而言，一些计算机视觉技术，特别是基于光流的方法，已被证明可用于雷达地图的精确外推，沿着这条路径的一个最新进展是香港天文台（HKO）提出的雷达回波（ROVER）算法实时光流变分方法，用于局域系统（SWIRLS）强暴雨短期警报。ROVER 使用中的算法计算连续雷达图的光流，并对假设为静止的流场执行半拉格朗日平流，以完成预测。然而，这些基于光流的方法的成功是有限的，因为流估计步骤和雷达回波外推步骤是分开的，并且确定模型参数以给出良好的预测性能是非常有挑战的。

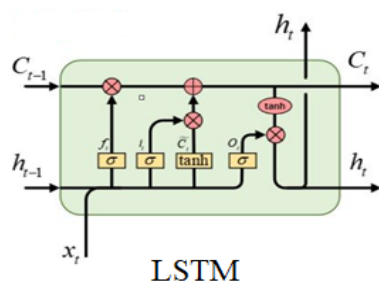
在机器学习上对于时空序列的预测问题也已经有了许多模型，如递归神经网络（RNN）模型和长短期记忆网络（LSTM）模型以及它的全连接（FC-LSTM）模型，这些模型都可以解决一般的时空序列预测问题，但这些模型在具有空间的相关性的问题处理上均表现不佳，也因此无法很好预测一个地区内各处的未来降水，无法很好的解决降水临近预报问题。

三、方法

作者在考虑到 RNN 和 LSTM、FC-LSTM 网络模型在处理具有空间相关性的数据时的缺陷后，开创性地将 FC-LSTM 的思想推广到卷积结构的卷积 LSTM（ConvLSTM）网络模型，把 FC-LSTM 中的输入到状态和状态到状态拓展为卷积结构，并将数据类型从传统的 1D 张量推广到 3D 张量。同时，通过叠加多个卷积 LSTM 层，形成编码-预测结构（Encoding-Forecasting Structure），通过将编码网络的最后状态（学习所有已有数据后的结果）复制到预测网络的初始状态（开始对未来进行预测），并将预测网络中从初始状态开始的所有状态串联起来（所有预测的结果），并将它们送入一个 1×1 的卷积层来生成最终的预测（最有可能的预测），从而建立降水临近预报的端到端可训练模型，对未来降水进行预测。

3.1 卷积 LSTM

在介绍卷积 LSTM 前，应先了解普通的 LSTM，示意图和公式如下（‘ \circ ’ 同下文表示为 Hadamard 乘积）：

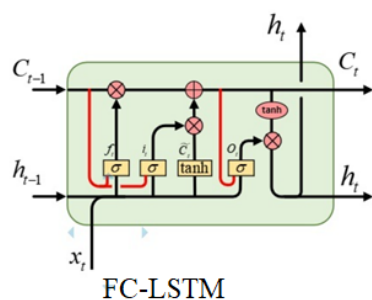


$$\begin{aligned} i^{(t)} &= \sigma(W^{(i)}x^{(t)} + U^{(i)}h^{(t-1)} + b^{(i)}) \\ f^{(t)} &= \sigma(W^{(f)}x^{(t)} + U^{(f)}h^{(t-1)} + b^{(f)}) \\ o^{(t)} &= \sigma(W^{(o)}x^{(t)} + U^{(o)}h^{(t-1)} + b^{(o)}) \\ \tilde{c}^{(t)} &= \tanh(W^{(c)}x^{(t)} + U^{(c)}h^{(t-1)} + b^{(c)}) \\ c^{(t)} &= f^{(t)} \circ c^{(t-1)} + i^{(t)} \circ \tilde{c}^{(t)} \\ h^{(t)} &= o^{(t)} \circ \tanh(c^{(t)}) \end{aligned}$$

https://blog.csdn.net/weixin_44699689

其中整个矩形方框被称作一个记忆块，它主要包含了三个门（遗忘门、输入门、输出门）与一个记忆单元，遗忘门会根据上一时刻的输出 h_{t-1} 和当前输入 x_t 来产生一个 0 到 1 的 $f^{(t)}$ 值，来决定是否让上一时刻学到的信息 C_{t-1} 通过或部分通过。输入门通过 \tanh 层用来生成新的候选值 $\tilde{C}^{(t)}$ 并通过 σ 来决定哪些值用来更新，最后通过把两部分产生的值结合起来对经过遗忘后的 C_{t-1} 进行更新。最后输出门决定了模型的输出，它首先通过 σ 层来得到一个初始输出，然后使用 \tanh 将 C_t 值缩放到 -1 到 1 间，再与 σ 得到的输出逐对相乘，从而得到模型的输出，同时更新 h_{t-1} 为 h_t 。而时空序列的预测正是通过一次次的遗忘和更新修正来实现对未来的预测。

在此基础上，又提出了全连接 LSTM：

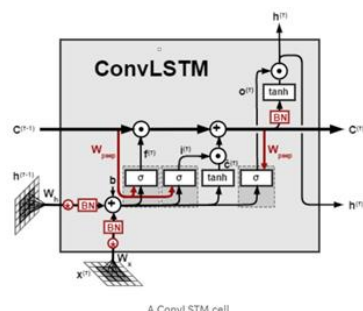


$$\begin{aligned}
 i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\
 f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \circ C_{t-1} + b_f) \\
 c_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\
 o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \circ C_t + b_o) \\
 h_t &= o_t \circ \tanh(c_t)
 \end{aligned}$$

https://blog.csdn.net/weixin_44699689

该模型是 LSTM 的变体，加入了窥探机制，每一个门的输入由三部分决定，即本层的输入，上一层的输出以及记忆线 C 上的信息，同时让原先输入的时间序列向量换成矩阵，以表示空间上的连接关系，但是包含的数据过多，有大量的冗余。

而作者提出的则是全新的卷积 LSTM (*表示卷积操作)：



$$\begin{aligned}
 i_t &= \sigma(W_{xi} * \mathcal{X}_t + W_{hi} * \mathcal{H}_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\
 f_t &= \sigma(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1} + W_{cf} \circ C_{t-1} + b_f) \\
 C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * \mathcal{X}_t + W_{hc} * \mathcal{H}_{t-1} + b_c) \\
 o_t &= \sigma(W_{xo} * \mathcal{X}_t + W_{ho} * \mathcal{H}_{t-1} + W_{co} \circ C_t + b_o) \\
 \mathcal{H}_t &= o_t \circ \tanh(C_t)
 \end{aligned}$$

https://blog.csdn.net/weixin_44699689

它的连接结构同 FC-LSTM，不同的地方在于将中间的全连接换成了卷积，将输入到状态和状态到状态拓展为卷积结构，它可以更好的表示空间中的关系，没有太多的冗余信息。将前后状态连接在一起，得到：

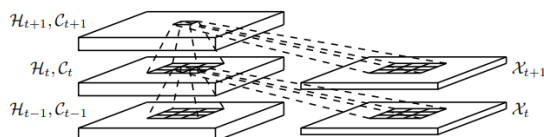


Figure 2: Inner structure of ConvLSTM

在这里，为了确保状态具有与输入相同的行数和列数，在应用卷积操作之前需要填充。作者使用了 Zero-padding 来进行填充 C 和 H，达到和 X 的行数和列数是一样的，这实际上是将外部世界的状态设置为零，并且假设没有关于外部的先验知识，对应于对未来的“完全无知”状态，以更好应对降水情况的错综复杂。

3.2 编码-预测结构

编码-预测结构由两个网络组成，一个编码网络和一个预测网络。这两个网络的连接方式是预测网络的初始状态是从编码网络的最后一个状态复制过来，同时这两个网络都是由许多层 ConvLSTM 堆叠起来的，其结构如下：

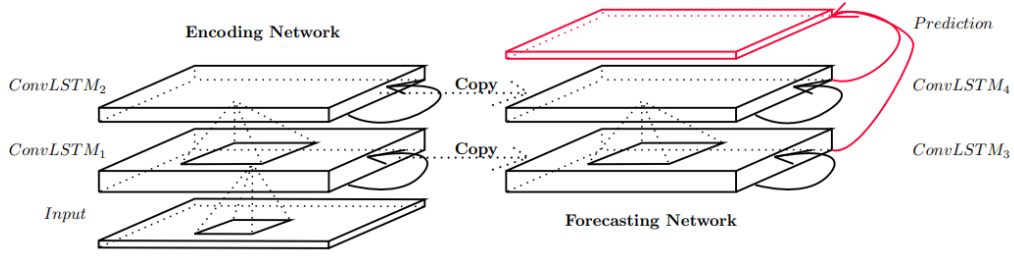


Figure 3: Encoding-forecasting ConvLSTM network for precipitation nowcasting

编码网络将整个输入序列压缩成一个隐藏状态张量，主要作用就是提取序列图像的时间特征和空间特征，预测网络则复制编码网络的最后一个状态作为其初始状态，然后将预测网络中的所有状态连接起来，并将它们馈送到 1×1 卷积层，以生成最终的预测（即展开该隐藏状态以给出最终预测）。对于多个时间步的预测，则可以往上继续堆叠预测端的卷积 LSTM，将水平层上的进行 1×1 卷积之后得出某个时刻的结果，垂直方向进行堆叠，来预测多个时刻的结果。最终的公式即为以下：

$$\begin{aligned} \hat{\mathcal{X}}_{t+1}, \dots, \hat{\mathcal{X}}_{t+K} &= \arg \max_{\mathcal{X}_{t+1}, \dots, \mathcal{X}_{t+K}} p(\mathcal{X}_{t+1}, \dots, \mathcal{X}_{t+K} | \hat{\mathcal{X}}_{t-J+1}, \hat{\mathcal{X}}_{t-J+2}, \dots, \hat{\mathcal{X}}_t) \\ &\approx \arg \max_{\mathcal{X}_{t+1}, \dots, \mathcal{X}_{t+K}} p(\mathcal{X}_{t+1}, \dots, \mathcal{X}_{t+K} | f_{\text{encoding}}(\hat{\mathcal{X}}_{t-J+1}, \hat{\mathcal{X}}_{t-J+2}, \dots, \hat{\mathcal{X}}_t)) \quad (4) \\ &\approx g_{\text{forecasting}}(f_{\text{encoding}}(\hat{\mathcal{X}}_{t-J+1}, \hat{\mathcal{X}}_{t-J+2}, \dots, \hat{\mathcal{X}}_t)) \end{aligned}$$

通过以上呈现的结构，可以很好地对 3 维张量进行预测。同时，由于其具有多个堆叠的卷积 LSTM 层，它具有很强的代表性，适用降水临近预报等复杂动力系统的预报。

四、实验与分析

作者主要在两个数据集上对模型进行实验：Moving-MINIST 数据集和雷达回波数据集。作者首先用 Moving-MINIST 数据集对卷积 LSTM 网络和 FC-LSTM 网络进行了对比，以获得对我们模型的行为的一些基本理解。接着利用雷达回波数据集对卷积 LSTM 网络和最新的 ROVER 算法进行了比较，以验证卷积 LSTM 模型在更具挑战性的降水临近预报问题上的高效性和准确性。

4.1 Moving-MINIST 数据集

作者采用 cross-entropy 交叉熵作为损失函数，以 RMSProp 为优化器，学习率为 0.001 并且有 0.9 的延迟率，并且在验证集上执行了提前停止，分别对两个模型进行训练，最终结果如下图：

Model	Number of parameters	Cross entropy
FC-LSTM-2048-2048	142,667,776	4832.49
ConvLSTM(5x5)-5x5-256	13,524,496	3887.94
ConvLSTM(5x5)-5x5-128-5x5-128	10,042,896	3733.56
ConvLSTM(5x5)-5x5-128-5x5-64-5x5-64	7,585,296	3670.85
ConvLSTM(9x9)-1x1-128-1x1-128	11,550,224	4782.84
ConvLSTM(9x9)-1x1-128-1x1-64-1x1-64	8,830,480	4231.50

结果可以看出，卷积 LSTM 模型在对 Moving-MINIST 数据集的效果要优于 FC-LSTM 模型。

4.2 雷达回波数据集

作者在这里使用的雷达回波数据集是香港 2011-2013 年三年天气雷达强度的子集，选择

了其中前 97 个雨天形成数据集，通过设置 $P = \frac{Z - \min\{Z\}}{\max\{Z\} - \min\{Z\}}$ 将强度值 Z 转化为灰度像素 P ，并在中心裁剪雷达图。在通过 K-Means 去除噪声后将数据集分为不相交的训练集、测试集和验证集。接下来使用降水均方误差（Rainfall-MSE）、临界成功指数（CSI）、虚警率（FAR）、探测概率（POD）和相关性（Correlation）来对卷积 LSTM 和 FC-LSTM 以及 Rover 算法进行测试和评估。得出的结果如下图所示。

Table 2: Comparison of the average scores of different models over 15 prediction steps.

Model	Rainfall-MSE	CSI	FAR	POD	Correlation
ConvLSTM(3x3)-3x3-64-3x3-64	1.420	0.577	0.195	0.660	0.908
Rover1	1.712	0.516	0.308	0.636	0.843
Rover2	1.684	0.522	0.301	0.642	0.850
Rover3	1.685	0.522	0.301	0.642	0.849
FC-LSTM-2000-2000	1.865	0.286	0.335	0.351	0.774

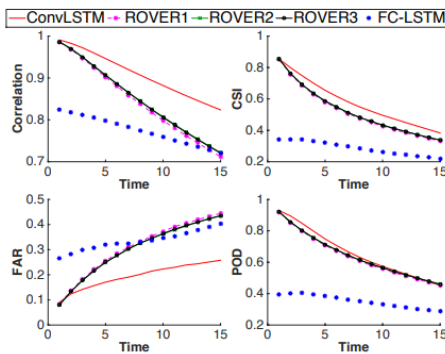


Figure 5: Comparison of different models based on four precipitation nowcasting metrics over time.

从结果可以看出：卷积 LSTM 在雷达回波数据集上优于 FC-LSTM 和 Rover 算法。

对以上实验进行分析，可以发现：1. 卷积 LSTM 优于 FC-LSTM 的主要原因是：雷达图中的强空间相关性，即云的运动在局部区域高度一致。全连接结构有太多的冗余连接，使得优化不太可能捕获这些局部一致性。2. 卷积 LSTM 优于 Rover 算法的主要原因是卷积 LSTM 能够很好地处理边界的状况。在现实生活中的临近预报中，有许多情况是云突然聚集在边界上，这表明一些云是从外面来的。如果卷积 LSTM 网络在训练期间看到了类似的模式，它可以在编码网络中发现这种类型的突然变化，并在预测网络中给出合理的预测。然而，这很难通过光流和基于半拉格朗日平流的方法来实现。另一个原因是，卷积 LSTM 是为此任务进行端到端训练的，并且数据集中的一些复杂时空模式可以通过网络的非线性和卷积结构来学习。对于基于光流的方法，很难找到一种合理的方法来更新未来的流场和端到端地训练一切。

五、我的思考

卷积 LSTM 在处理空间相关性上对比其他网络具有优势的一大原因就在于它采用了卷积操作来替代原先的矩阵乘法操作，而 FC-LSTM 在处理空间相关的问题时会包含太多的冗余，这因该和它本身的结构以及不同于卷积 LSTM 的非卷积算法有关，它的从状态到状态和输入到状态的转换方式的全连接形式造成了 FC-LSTM 处理数据时会出现太多的冗余。

卷积 LSTM 在降水临近预报的处理优于 Rover 算法的一个优势在于卷积 LSTM 对边界情况的准确判断。由于 Rover 算法对于一块地区的降水预测需要周边地区的天气信息进行计算，所以对于缺少周边地区天气信息的边界地区，Rover 算法难以给出准确的预测。而卷积 LSTM

从训练开始就在缺少周边信息的情况下对周边地区的天气进行学习，在实际预测周边地区时，当实际情况在训练中出现过时，卷积 LSTM 便能很好地做出预测。

卷积 LSTM 除了运用在降水临近预报上，还能运用在很多与空间相关或多维的问题上，除了对雷达回波数据集进行分析，可能还可以对于它类似的图像、视频识别上进行运用，例如识别视频中出现的物体、行为等，或者对一个地区内除降水外的其他性质进行预测：如人口、经济等。