

卷积 LSTM 网络:降水临近预报的机器学习方法

施行健陈周荣王浩迪·杨燕香港科技大学计算机科学与工程系

{xshiab, zchenbb, hwangaz dyeeung}
@cse.ust.hk

Wong Wai-kin Woo 汪春

香港天文台中国香港

{wkwong, wwoo}
@hko.gov.hk

摘要

降水临近预报的目标是在相对较短的时间内预测局部区域未来的降雨强度。之前很少有研究从机器学习的角度来考察这个至关重要且具有挑战性的天气预报问题。在本文中,我们将降水临近预报表述为一个时空序列预测问题,其中输入和预测目标都是时空序列。通过扩展全连接 LSTM (FC-LSTM),使其在输入到状态和状态到状态的转换中都具有卷积结构,我们提出了卷积 LSTM(ConvLSTM),并用它来构建一个端到端的可训练模型来解决降水临近预测问题。实验表明,我们的 ConvLSTM 网络能够更好地捕获时空相关性,并且始终优于 FC-LSTM 和最先进的操作 ROVER 算法用于降水临近预报。

1 介绍

临近预报对流降水一直是天气预报领域的一个重要问题。这一任务的目标是在相对较短的时间内(如 0-6 小时)对局部区域的降雨强度进行精确、及时的预测。它对于及时采取诸如生成社会层面的紧急降雨警报、为机场生成天气指导以及和较长期数值天气预报(NWP)模型无缝集成等行动至关重要。由于所需的预报分辨率和时间精度远高于周平均温度预测等其他传统预报任务,因此降水临近预报问题颇具挑战性,已成为气象学界的热门研究课题[22]。

现有的降水临近预报方法大致可分为两类[22],即基于 NWP 的方法和基于雷达回波¹外推的方法。对于 NWP 方法,在临近预报时间尺度上进行预测需要对大气模式中的物理方程进行复杂而细致的模拟。因此,目前最先进的操作降水临近预报系统[19,6]往往采用更快、更准确的基于外推的方法。具体来说,一些计算机视觉技术,特别是基于光流的方法,已被证明对雷达图进行精确外推是有用的[10,6,20]。这条道路上最近的一项进展是雷达回波的变分方法实时光流(*Real-time Optical flow by Variational methods for Echoes of Radar*, ROVER)

¹ 在现实系统中,雷达回波图往往是恒定高度平面位置指示器(CAPPI)图像[9]。

香港天文台(HKO)为其局域系统强暴雨短程预警(SWIRLS)[15]提出的算法[25]。ROVER 使用[5]中的算法计算连续雷达图的光流，并对假定为静止的流场进行半拉格朗日平流[4]，以完成预测。然而，这些基于光流的方法的成功是有限的，因为流量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具有挑战性。

这些技术问题可以通过从机器学习的角度来解决。从本质上讲，降水临近预报是一个以过去雷达图序列为输入，以固定数量(通常大于 1)未来雷达图序列为输出的时空序列预测问题。²然而，这类学习问题，无论其具体应用如何，首先都不是微不足道的，因为时空序列的高维性，尤其是在必须进行多步预测的情况下，除非预测模型很好地捕捉到数据的时空结构。此外，由于大气的混沌特性，为雷达回波数据建立有效的预测模型甚至更具挑战性。

深度学习的最新进展，特别是循环神经网络(RNN)和长短期记忆(LSTM)模型[12,11,7,8,23,13,18,21,26]，为如何解决这个问题提供了一些有用的见解。根据深度学习方法背后的哲学，如果我们有一个合理的端到端模型，并有足够的数据对其进行训练，我们就接近解决问题了。降水临近预报问题满足了数据需求，因为它很容易连续收集海量的雷达回波数据。所需要的是一个适合端到端学习的模型。[23]中提出的开拓性 LSTM 编码器-解码器框架，通过训练时间级联的 LSTM，一个用于输入序列，另一个用于输出序列，为序列到序列学习问题提供了一个通用框架。在[18]中，表明可以通过在量化图像补丁获得的可视单词上构建基于 RNN 的语言模型来完成下一视频帧的预测和中间帧的插值。他们提出了一种循环卷积神经网络来对空间关系建模，但该模型只能提前预测一帧，并且用于状态到状态转换的卷积核的大小被限制为 1。他们的工作随后在[21]中得到了跟进，指出了多步预测在学习有用表示中的重要性。他们构建了一个 LSTM 编码器-解码器-预测器模型，该模型可以重建输入序列并同时预测未来序列。虽然他们的方法也可以用来解决我们的时空序列预测问题，但他们模型采用的全连接 LSTM(FC-LSTM)层并没有考虑到空间相关性。

在本文中，我们提出了一种用于降水临近预报的新型卷积 LSTM(ConvLSTM)网络。我们将降水临近预报表述为一个时空序列预测问题，该问题可以在[23]中提出的一般序列到序列学习框架下解决。为了更好地建模时空关系，我们将 FC-LSTM 的思想扩展到在输入到状态和状态到状态转换中都具有卷积结构的 ConvLSTM。通过堆叠多个 ConvLSTM 层并形成编码-预测结构，我们可以构建一个端到端可训练的降水临近预报模型。为了进行评估，我们创建了一个新的现实生活中的雷达回波数据集，这可以促进进一步的研究，特别是在为该问题设计机器学习算法方面。当在合成的 Moving-MNIST 数据集[21]和雷达回波数据集上进行评估时，我们的 ConvLSTM 模型始终优于 FC-LSTM 和最先进的操作 ROVER 算法。

2 预赛

2.1 降水临近预报问题的公式

降水临近预测的目标是利用之前观测到的雷达回波序列，预测局部地区(如香港、纽约或东京)的一段固定长度的未来雷达图。在实际应用中，雷达图通常每隔 6-10 分钟从天气雷达中获取，并在接下来的 1-6 小时内进行临近预报，即预测前方的 6-60 帧。从 $ma-$

² 值得注意的是，我们的降水临近预报问题与[14]中研究的问题不同，后者的目的只是预测下一帧的中心区域。

从中文学习的角度来看，这个问题可以看作是一个时空序列预测问题。

假设我们在一个由 $M \times N$ 网格表示的空间区域上观察一个动力系统，该网格由 M 行和 N 列组成。网格中的每个单元格内部，有 P 个随时间变化的测量值。因此，任何时候的观测都可以用一个张量 $X \in \mathbb{R}^{P \times M \times N}$ 表示，其中 R 表示观测到的特征的域。如果我们周期性地记录观测结果，我们将得到一个张量序列 X_1, X_2, \dots, X_t 。时空序列预测问题是在给定之前的 J 个观测值(包括当前观测值)的情况下，预测未来最有可能的长度- k 序列：

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

对于降水临近预报，每个时间戳的观测为二维雷达回波图。如果我们将地图划分为平铺的不重叠的斑块，并将斑块内的像素视为其测量值(见图 1)，那么临近预报问题自然就变成了时空序列预测问题。

我们注意到，我们的时空序列预测问题不同于一步时间序列预测问题，因为我们的问题的预测目标是一个同时包含空间和时间结构的序列。尽管长度为 k 的序列中自由变量的数量可以达到 $O(M^k N^k P^k)$ ，但在实践中，我们可以利用可能预测空间的结构来降低维数，从而使问题易于处理。

2.2 用于序列建模的长短期记忆

对于通用序列建模，LSTM 作为一种特殊的 RNN 结构，在之前的各种研究中已被证明是稳定和强大的远程依赖建模方法[12,11,17,23]。LSTM 的主要创新是它的存储单元 c_t ，它本质上是作为状态信息的累加器。该单元由几个自参数化的控制门访问、写入和清除。每次有新的输入来，如果输入门 i_t 被激活，它的信息就会被累积到单元格中。同样，如果遗忘门打开，过去的单元格状态 c_{t-1} 可能在这个过程被“遗忘”。最新的单元输出 c_t 是否会传播到最终状态，由输出门 o_t 进一步控制。使用记忆单元和门来控制信息流的一个好处是，梯度将被锁定在单元中(也称为常数误差旋转木马[12])，并防止其过快消失，这是 vanilla RNN 模型的关键问题[12,17,2]。FC-LSTM 可以被视为 LSTM 的多元版本，其中输入、单元输出和状态都是 1D 向量。在本文中，我们遵循 FC-LSTM 的公式，如[11]。关键方程如下图(2)所示，其中 \circ 表示 Hadamard 积：

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \\ c_t &= f_t \circ c_{t-1} + i_t \circ \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \circ c_t + b_o) \\ h_t &= o_t \circ \tanh(c_t) \end{aligned} \quad (2)$$

多个 lstm 可以堆叠和临时连接，形成更复杂的结构。这样的模型已经被应用于解决许多现实生活中的序列建模问题[23,26]。

3 模型

我们现在展示我们的 ConvLSTM 网络。尽管 FC-LSTM 层在处理时间相关性方面已经被证明是强大的，但它包含了太多空间数据的冗余。为了解决这个问题，我们提出了 FC-LSTM 的扩展，它在输入到状态和状态到状态的转换中都具有卷积结构。通过堆叠多个 ConvLSTM 层并形成编码-预测结构，我们不仅能够为降水临近预报问题构建网络模型，而且能够为更一般的时空序列预测问题构建网络模型。

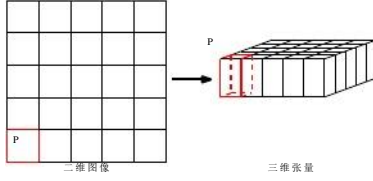


图 1:将 2D 图像转换为 3D 张量

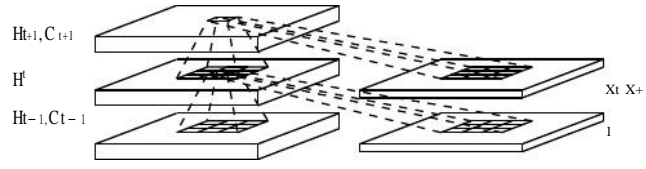


图 2:ConvLSTM 的内部结构

3.1 卷积 LSTM

FC-LSTM 在处理时空数据方面的主要缺点是它在没有编码空间信息的输入到状态和状态到状态转换中使用了完全连接。为了克服这个问题，我们设计的一个显著特征是，所有的输入 X_1, \dots, X_t , 单元格输出 C, \dots, C_t , 隐藏状态 H_1, \dots, H_t , 和 ConvLSTM 的 i_t, f_t, o_t 门是 3D 张量，其最后两个维度是空间维度(行和列)。为了更好地理解输入和状态，我们可以把它们想象成站在空间网格上的向量。ConvLSTM 通过其局部邻居的输入和过去的状态来确定网格中某个单元格的未来状态。这可以通过在状态到状态和输入到状态转换中使用卷积算子轻松实现(见图 2)。ConvLSTM 的关键方程如下(3)所示，其中 $*$ 表示卷积算子， \circ 与前面一样，表示 Hadamard 积：

$$\begin{aligned} i_t &= \sigma(W_{xi} * \mathcal{X}_t + W_{hi} * \mathcal{H}_{t-1} + W_{ci} \circ \mathcal{C}_{t-1} + b_i) \\ f_t &= \sigma(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1} + W_{cf} \circ \mathcal{C}_{t-1} + b_f) \\ \mathcal{C}_t &= f_t \circ \mathcal{C}_{t-1} + i_t \circ \tanh(W_{xc} * \mathcal{X}_t + W_{hc} * \mathcal{H}_{t-1} + b_c) \\ o_t &= \sigma(W_{xo} * \mathcal{X}_t + W_{ho} * \mathcal{H}_{t-1} + W_{co} \circ \mathcal{C}_t + b_o) \\ \mathcal{H}_t &= o_t \circ \tanh(\mathcal{C}_t) \end{aligned} \quad (3)$$

如果我们将状态视为运动对象的隐藏表示，具有较大过渡核的 ConvLSTM 应该能够捕获更快的运动，而具有较小过渡核的 ConvLSTM 可以捕获较慢的运动。同样，如果我们采用与[16]类似的观点，(2)表示的传统 FC-LSTM 的输入、单元输出和隐藏状态也可以被视为最后两个维度为 1 的 3D 张量。从这个意义上讲，FC-LSTM 实际上是 ConvLSTM 的一种特例，所有特征都站在单个单元上。

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。在这里，对边界点上的隐藏状态进行填充，可以看作是利用外部世界的状态进行计算。通常，在第一个输入到来之前，我们将 LSTM 的所有状态初始化为零，这对应于对未来的“完全无知”。类似地，如果我们对隐藏状态执行零填充(本文中使用的)，我们实际上是将外部世界的状态设置为零，并假设对外部世界没有先验知识。通过对状态进行填充，我们可以以不同的方式对待边界点，这在很多情况下是有帮助的。例如，想象我们正在观察的系统是一个被墙壁包围的移动的球。虽然我们看不到这些墙，但我们可以通过发现球一次又一次地在它们上面弹跳来推断它们的存在，如果边界点与内部点具有相同的状态转换动力学，这几乎是不可能做到的。

3.2 编码-预测结构(coding-forecasting Structure)

与 FC-LSTM 一样，ConvLSTM 也可以作为更复杂结构的构建块。对于我们的时空序列预测问题，我们使用了如图 3 所示的结构，它由两个网络组成，一个编码网络和一个预测网络。像在[21]中一样，预测网络的初始状态和单元输出是从编码网络的最后一个状态复制而来的。这两个网络都是通过堆叠几个 ConvLSTM 层形成的。由于我们的预测目标具有与输入相同的维数，我们将预测网络中的所有状态连接起来，并将它们馈送到 1×1 卷积层中，以生成最终的预测。

我们可以使用类似于[23]的观点来解释这个结构。编码 LSTM 将整个输入序列压缩成一个隐藏状态张量，而预测 LSTM 将这个隐藏展开

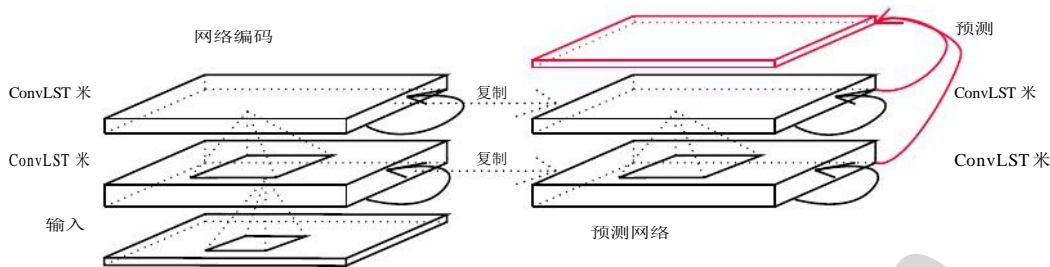


图 3:降水临近预报的编码-预测 ConvLSTM 网络

国家给出最后的预测:

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

这个结构也类似于[21]中的 LSTM 未来预测器模型，只是我们的输入和输出元素都是 3D 张量，保留了所有的空间信息。由于网络具有多个堆叠的 ConvLSTM 层，它具有强大的表示力，这使得它适合于在复杂的动力系统中给出预测，比如我们在这里研究的降水临近预报问题。

4 实验

我们首先将我们的 ConvLSTM 网络与 FC-LSTM 网络在一个合成的 Moving-MNIST 数据集上进行比较，以获得对我们模型行为的一些基本理解。我们用不同的层数和内核大小运行我们的模型，也研究了一些“域外”的情况，如[21]。为了验证我们的模型在更具有挑战性的降水临近预报问题上的有效性，我们建立了一个新的雷达回波数据集，并将我们的模型与基于几种常用降水临近预报指标的最先进的 ROVER 算法进行了比较。在这两个数据集上进行的实验结果导致了以下发现：

- ConvLSTM 在处理时空相关性方面优于 FC-LSTM。
- 使 state-to-state 卷积核的大小大于 1 对于捕获时空运动模式至关重要。
- 更深入的模型可以用更少的参数产生更好的结果。
- ConvLSTM 在降水临近预报方面比 ROVER 表现更好。

在 Theano[3,1]的帮助下，我们在 Python 中实现了模型。我们在一台具有单个 NVIDIA K20 GPU 的计算机上运行所有的实验。此外，附录中还包含了更多说明性的“gif”示例。

4.1 Moving-MNIST 数据集

对于这个合成数据集，我们使用了类似于[21]中描述的生成过程。数据集中的所有数据实例都是 20 帧长(10 帧为输入帧，10 帧为预测帧)，并包含两个在 64×64 补丁内弹跳的手写数字。移动的数字是从 MNIST 数据集中 500 个数字的子集中随机选择的。³ 起始位置和速度方向是均匀随机选择的，速度幅值是在[3,5]中随机选择的，这个生成过程重复 15000 次，得到一个包含 10000 个训练序列、2000 个验证序列和 3000 个测试序列的数据集。我们通过使用时间反向传播(BPTT)[2]和最小化交叉熵损失⁴来训练所有的 LSTM 模型

³ MNIST 数据集:<http://yann.lecun.com/exdb/mnist/>

⁴ 将预测框架 P 与真实框架 T 的交叉熵损失定义为

$$-\sum_{i,j,k} T_{i,j,k} \log P_{i,j,k} + (1 - T_{i,j,k}) \log(1 - P_{i,j,k}).$$

有道文档翻译
pdf.youdao.com

表 1:在 Moving-MNIST数据集上 ConvLSTM 网络与 FC-LSTM 网络的比较。'-5x5'和'-1x1'表示对应的状态到状态的内核大小，要么是 5×5，要么是 1×1。'256'、'128'和'64'表示 ConvLSTM 层中隐藏状态的数量。'(5x5)'和'(9x9)'表示输入到状态的内核大小。

模型	参数个数	交叉熵
fc - lstm - 2048 - 2048	142667776	4832.49
ConvLSTM (5 x5) 5 x5 - 256	13524496	3887.94
ConvLSTM (5 x5) 5 x5 - 128 - 5 x5 - 128	10042896	3733.56
ConvLSTM (5 x5) 5 x5 - 128 - 5 x5 - 64 - 5 x5 - 64	7585296	3670.85
ConvLSTM (9 x9) 1 x1 - 128 - 1 - x1 - 128	11550224	4782.84
ConvLSTM (9 x9) 1 x1 - 128 - 1 - x1 - 64 - 1 x1 - 64	8830480	4231.50



图 4:显示 “域外” 运行的示例。从左到右:输入帧;地面实况;3 层网络预测。

RMSProp[24], 学习率为 10^{-3} , 衰减率为 0.9。此外，我们在验证集上执行早期停止。

尽管生成过程简单，但生成的数据集中存在很强的非线性，因为移动的数字可以表现出复杂的外观，并且在移动过程中会遮挡和反弹。在不了解系统内部动态的情况下，模型很难在测试集上给出准确的预测。

对于 FC-LSTM 网络，我们使用了与[21]中具有两个 2048 节点 LSTM 层的无条件未来预测器模型相同的结构。对于我们的 ConvLSTM 网络，我们将 patch 大小设置为 4×4 ，以便每个 64×64 帧由一个 $16 \times 16 \times 16$ 张量表示。我们用不同的层数测试了我们模型的三种变体。1 层网络包含一个含有 256 个隐藏状态的 ConvLSTM 层，2 层网络有两个含有 128 个隐藏状态的 ConvLSTM 层，3 层网络在三个 ConvLSTM 层中分别有 128 个、64 个和 64 个隐藏状态。所有输入到状态和状态到状态的核都是 5×5 的大小。我们的实验表明，ConvLSTM 网络的性能始终优于 FC-LSTM 网络。此外，更深层次的模型可以给出更好的结果，尽管在 2 层网络和 3 层网络之间的改善不是那么显著。此外，我们还尝试了其他网络配置，将 2 层和 3 层网络的 state-to-state 和 input-to-state 内核分别改为 1×1 和 9×9 。虽然新的 2 层网络的参数数量接近于原始网络，但由于仅用 1×1 状态到状态转换很难捕捉时空运动模式，结果变得糟糕得多。同时，新的 3 层网络比新的 2 层网络表现更好，因为更高的一层可以看到更广泛的输入范围。尽管如此，其性能还是不如具有更大的 state-to-state 内核大小的网络。这为更大的状态到状态内核更适合捕捉时空相关性提供了证据。事实上，对于 1×1 核来说，状态的接受场不会随着时间的推移而增长。但对于更大的核，之后的状态会有更大的接受场，并且与更大范围的输入相关。每个算法在测试集上的平均交叉熵损失(每个序列的交叉熵损失)如表 1 所示。我们需要指出的是，我们的实验设置与[21]不同，[21]假设有无限个训练数据可用。选择当前的离线设置是为了了解不同模型在没有那么多数据可用的场合下的表现。在线设置下的 3 层 ConvLSTM 和 FC-LSTM 的对比在附录中。

接下来，我们在一些“域外”输入上测试我们的模型。我们生成另外 3000 个由三个移动数字组成的序列，其中的数字是从 500 个 MNIST 数字的不同子集中随机抽取的，这些数字与训练集不重叠。由于模型从未见过任何三位数的系统，这样的“出域”运行很好地检验了模型[21]的泛化能力。3 层模型在这个数据集上的平均交叉熵误差为 6379.42。通过观察部分预测结果，我们发现该模型可以成功分离重叠的数字，并预测整体运动，尽管预测的数字相当模糊。图 10 展示了一个“域外”预测实例。

4.2 雷达回波数据集

本文使用的雷达回波数据集是 2011 年至 2013 年在香港收集的三年天气雷达强度的子集。由于不是每天都下雨，我们的临近预报目标是降水，所以我们选择了前 97 个雨天来形成我们的数据集。为了进行预处理，我们首先通过设置 $P = \frac{\max\{Z\} - \min\{Z\}}{Z - \min\{Z\}}$ 将强度值 Z 转换为灰度像素 P

中部 330×330 区域的雷达图。之后，我们应用半径为 10 的磁盘滤波器⁵，并将雷达图大小调整为 100×100 。为了减少测量仪器造成的噪声，我们进一步去除一些噪声区域的像素值，这些像素值是通过对月像素平均值应用 K-means 聚类确定的。天气雷达数据每 6 分钟记录一次，因此每天有 240 帧。为了获得不相交的子集用于训练、测试和验证，我们将每个每日序列划分为 40 个不重叠的帧块，并随机分配 4 个块用于训练，1 个块用于测试，1 个块用于验证。使用 20 帧宽的滑动窗口从这些块中切片出数据实例。因此，我们的雷达回波数据集包含 8148 个训练序列，2037 个测试序列和 2037 个验证序列，所有序列都是 20 帧长(5 帧为输入，15 帧为预测)。尽管从同一天切片的训练和测试实例可能存在一些依赖关系，但这种分割策略仍然是合理的，因为在现实生活中的临近预测中，我们确实可以访问所有之前的数据，包括同一天的数据，这使得我们可以应用模型的在线微调。这样的数据分割可能被视为这个应用程序在现实生活中“启用微调”设置的近似值。

我们将补丁大小设置为 2，并训练一个 2 层 ConvLSTM 网络，每层包含 64 个隐藏状态和 3×3 个内核。对于 ROVER 算法，我们在验证集上调优光流估计器⁶的参数，并使用最佳参数(如附录所示)报告测试结果。此外，我们为 ROVER 尝试了三种不同的初始化方案:ROVER1 计算最后两个观测帧的光流，然后执行半拉格朗日平流;ROVER2 通过最后两个流场的平均值来初始化速度;ROVER3 通过后三个流场的加权平均(权重为 0.7、0.2 和 0.1)进行初始化。此外，我们训练了一个具有两个 2000 节点 LSTM 层的 FC-LSTM 网络。ConvLSTM 网络和 FC-LSTM 网络都优化了 15 个预测的交叉熵误差。

我们使用几种常用的降水临近预报指标来评估这些方法，即降雨均方误差(rain - mse)、临界成功指数(CSI)、虚警率(FAR)、探测概率(POD)和相关性。降雨 - mse 指标定义为预测降雨量与地面真实值之间的平均平方误差。由于我们的预测是在像素级完成的，我们将其投影回雷达回波强度，并使用 Z-R 关系[15]: $Z = 10 \log a + 10b \log R$ 计算网格每个单元格的降雨量，其中 Z 是雷达回波强度，单位为 dB， R 是降雨率，单位为 mm/h， a 、 b 是两个常数， $a = 118.239$ ， $b = 1.5241$ 。CSI、FAR 和 POD 是机器学习研究人员常用的类似于精度和召回率的技能分数。我们使用 0.5 mm/h 降雨率(指示是否下雨)的阈值将预测和地面真相转换为 0/1 矩阵，并计算命中(预测 = 1，真相 = 1)、失手(预测 = 0，真相 = 1)和假警报(预测 = 1，真相 = 0)，这三个技能得分定义为 $CSI = \frac{\text{命中}}{\text{命中} + \text{失手} + \text{假警报}}$ ， $FAR = \frac{\text{假警报}}{\text{命中} + \text{假警报}}$ ， $POD = \frac{\text{命中}}{\text{命中} + \text{失手}}$ 。一个预测帧 P 和一个地真帧 T 的相关性定义为 $\sqrt{\frac{1}{P} \sum_{i,j} P_{i,j} T_{i,j}}$

式中， $\epsilon =$

⁵ 使用 MATLAB 函数 fspecial(' disk ', 10)应用磁盘过滤器。

⁶ 我们使用一个开源项目来计算光流:<http://sourceforge.net/projects/varflow/>

表 2:不同模型在 15 个预测步骤上的平均分比较。

模型	Rainfall-MSE	CSI	远	圆荚体	相关
ConvLSTM (3 x3) 3 x3 - x3 - 64 - 64 - 3	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

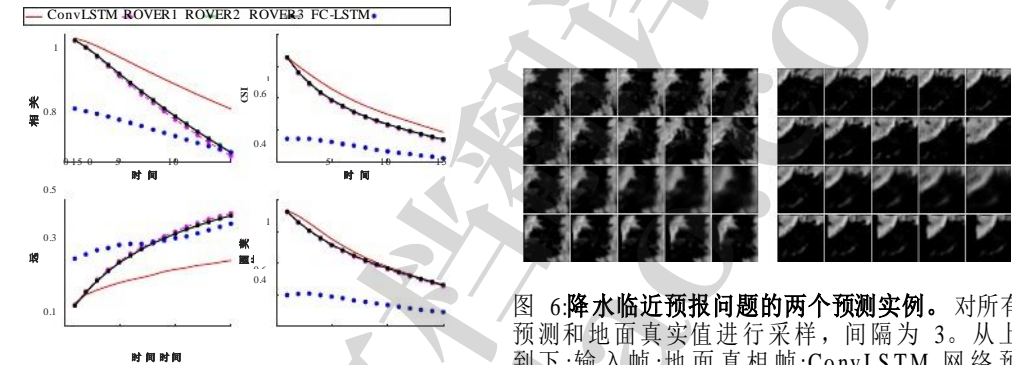


图 5:基于四种降水临近预报指标随时间变化的不同模型比较。

所有结果如表 2 和图 5 所示。我们可以发现 FC-LSTM 网络对于这项任务的性能并不是很好，这主要是由于雷达图中存在较强的空间相关性，即云的运动在局部区域具有高度的一致性。全连接结构有太多的冗余连接，使得优化非常不可能捕捉到这些局部一致性。另外，可以看出 ConvLSTM 优于基于光流的 ROVER 算法，这主要是由于两个原因。首先，ConvLSTM 能够很好地处理边界条件。在现实中的临近预报中，边界处出现突然的云层聚集现象的情况很多，这说明有一些云是从外部来的。如果 ConvLSTM 网络在训练过程中看到过类似的模式，就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，ConvLSTM 针对这项任务是端到端的训练，数据集中一些复杂的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式来更新未来的流场，并端到端地训练一切。ROVER2 和 ConvLSTM 的部分预测结果如图 6 所示。我们可以发现，ConvLSTM 能够更准确地预测未来的降雨等值线，尤其是在边界上。虽然 ROVER2 可以给出比 ConvLSTM 更清晰的预测，但它触发的假警报更多，总体上不如 ConvLSTM 准确。而且，ConvLSTM 的模糊效应可能是由于任务固有的不确定性造成的，即在较长期的预测中，几乎不可能对整个雷达图给出锐利而准确的预测。我们只能对预测进行模糊处理，以减轻这类不确定性带来的误差。

5 结论和未来的工作

在这篇论文中，我们成功地将机器学习方法，特别是深度学习，应用于具有挑战性的降水临近预报问题，到目前为止，这个问题还没有从复杂的机器学习技术中受益。我们将降水临近预报制定为一个时空序列预测问题，并提出了 LSTM 的一个新的扩展，称为 ConvLSTM 来解决这个问题。ConvLSTM 层不仅保留了 FC-LSTM 的优势，而且由于其固有的卷积结构，也适用于时空数据。通过将 ConvLSTM 纳入编码-预测结构，我们构建了一个端到端可训练的降水临近预报模型。对于未来的工作，我们将研究如何将 ConvLSTM 应用于基于视频的动作识别。一个想法是将 ConvLSTM 添加到卷积神经网络生成的空间特征图之上，并使用 ConvLSTM 的隐藏状态进行最终分类。

参考文献

- [1] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N. Bouchard, D. Ward - Farley, and Y. Bengio. Theano: 新功能和速度提升。深度学习和无监督特征学习 NIPS 2012 研讨会, 2012 年。
- [2] Y. Bengio, I. Goodfellow, 和 A. Courville. 深度学习。为 MIT 出版社准备的, 2015 年。
- [3] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Ward - Farley, Y. Bengio. Theano: 一种 CPU 和 GPU 数学表达式编译器。在 *Scipy*, 第 4 卷, 第 3 页。得克萨斯州奥斯汀, 2010 年。
- [4] R. 布里森. *计算机图形学流体模拟*。Ak Peters 系列。Taylor & Francis, 2008 年。
- [5] T. Brox, A. Bruhn, N. Papenberger, J. Weickert. 基于翘曲理论的高精度光流估计。在 *ECCV*, 第 25-36 页。2004。
- [6] P. 张和杨鸿毅. 光流技术在香港终端区重要对流近流预报中的应用。《*第三届 WMO 临近预报和极近程预报国际研讨会(WSN12)*》, 2012 年第 6-10 页。
- [7] K. Cho, B. van Merriënboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio. 使用 RNN 编码器-解码器学习短语表示, 用于统计机器翻译。在 *EMNLP* 中, 2014 年第 1724 - 1734 页。
- [8] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko 和 T. Darrell. 用于视觉识别和描述的长期递归卷积网络。 *CVPR*, 2015。
- [9] R. H. 道格拉斯. 暴风雨天气组(加拿大)。《*气象学中的雷达*》, 61-68 页。1990。
- 乌尔斯·格尔曼和伊兹塔·扎瓦兹基. 大陆雷达图像降水可预测性的尺度依赖。第一部分: 方法学描述。《*每月天气评论*》, 130(12):2859-2873, 2002。
- [11] A. 格雷夫斯. 用循环神经网络生成序列。 *arXiv 预打印 arXiv:1308.0850*, 2013。
- [12] S. Hochreiter 和 J. Schmidhuber. 长短期记忆。 *神经计算*, 9(8):1735 - 1780, 1997。
- [13] A. Karpathy 和 L. 飞. 用于生成图像描述的深度视觉-语义对齐。 *CVPR*, 2015。
- [14] B. Klein, L. Wolf, Y. Afek. 一种用于短距离天气预报的动态卷积层。 *CVPR*, 2015。
- [15] 李佩文、王 w.k.、陈桂英、黎世泰. *swirls* - 一个不断进化的临近预报系统。香港特别行政区政府, 2000 年。
- [16] J. Long, E. Shelhamer, T. Darrell. 用于语义分割的全卷积网络。 *CVPR*, 2015。
- [17] R. Pascanu, T. Mikolov, Y. Bengio. 关于训练循环神经网络的难度。 *ICML*, 1310-1318 页, 2013。
- [18] M. Ranzato, A. Szlam, J. Bruna, M. Mathieu, R. Collobert 和 S. Chopra. 视频(语言)建模: 自然视频生成模型的基线。 *arXiv 预印本 arXiv:1412.6604*, 2014。
- [19] M. Reyniers. 《*基于雷达观测的定量降水预报: 原理、算法和操作系统*》。Institut Royal M'et'eorologique de Belgique, 2008。
- [20] H. Sakano. 基于时变光流物理模型的时空图像模式预测方法。 *遥感学报*, 51(5-2):3023-3036, 2013。
- [21] N. Srivastava, E. Mansimov, 和 R. Salakhutdinov. 使用 lstm 的视频表示的无监督学习。在 *ICML*, 2015。
- [22] J. Sun, M. Xue, J. W. Wilson, I. Zawadzki, S. P. Ballard, J. Onvlee-Hooimeyer, P. Joe, D. M. Barker, P. W. Li, B. Golding, M. Xu, J. Pinto. NWP 用于临近预报对流降水: 近期进展与挑战。 *美国气象通报*, 95(3):409-426, 2014。
- [23] I. Sutskever, O. Vinyals, Q. V. Le. 用神经网络进行序列对序列学习。在 *NIPS*, 第 3104-3112 页, 2014。
- [24] T. Tieleman 和 G. Hinton. 第 6.5 讲- RMSProp: 将梯度除以其最近量级的运行平均值。 *Coursera 课程: 机器学习的神经网络*, 2012 年第 4 期。
- [25] W.C. Woo 和 W.K. Wong. 光流技术在降雨临近预报中的应用。2014 年第 27 届强局地风暴会议上。
- [26] 徐凯, 巴杰, R. Kiros, A. Courville, R. Salakhutdinov, R. Zemel, Y. Bengio. Show, attend and tell: 带有视觉注意力的神经图像字幕生成。在 *ICML*, 2015。

附录

表 3:ROVER 中光流估计器的最佳参数。

参数	意义	价值
L_{max}	最粗糙的空间尺度水平	6
L_{start}	Finest 空间尺度水平	0
n_{pre}	预平滑步骤数	2
n_{post}	后平滑步骤数	2
ρ	局部向量场平滑的高斯卷积参数	1.5
α	能量函数中的正则化参数	2000
σ	用于图像平滑的高斯卷积参数	4.5

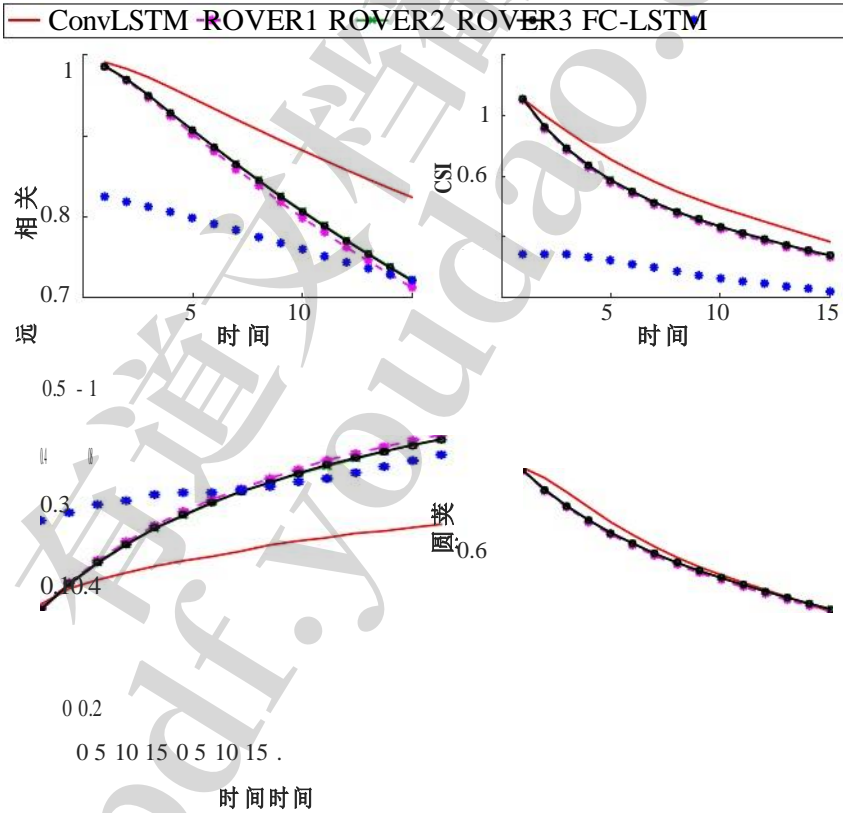


图 7:(放大版)基于四种降水临近预报指标随时间变化的不同模型比较。

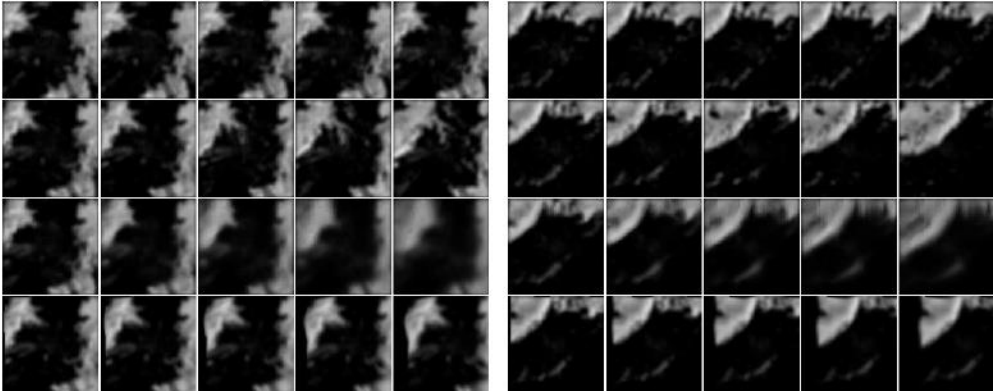


图 8:(大版本 降水临近预报问题的两个预测实例。对所有预测和地面真实值进行采样，间隔为 3。
从上到下:输入帧;地面实况;ConvLSTM 网络预测;ROVER2 预测。

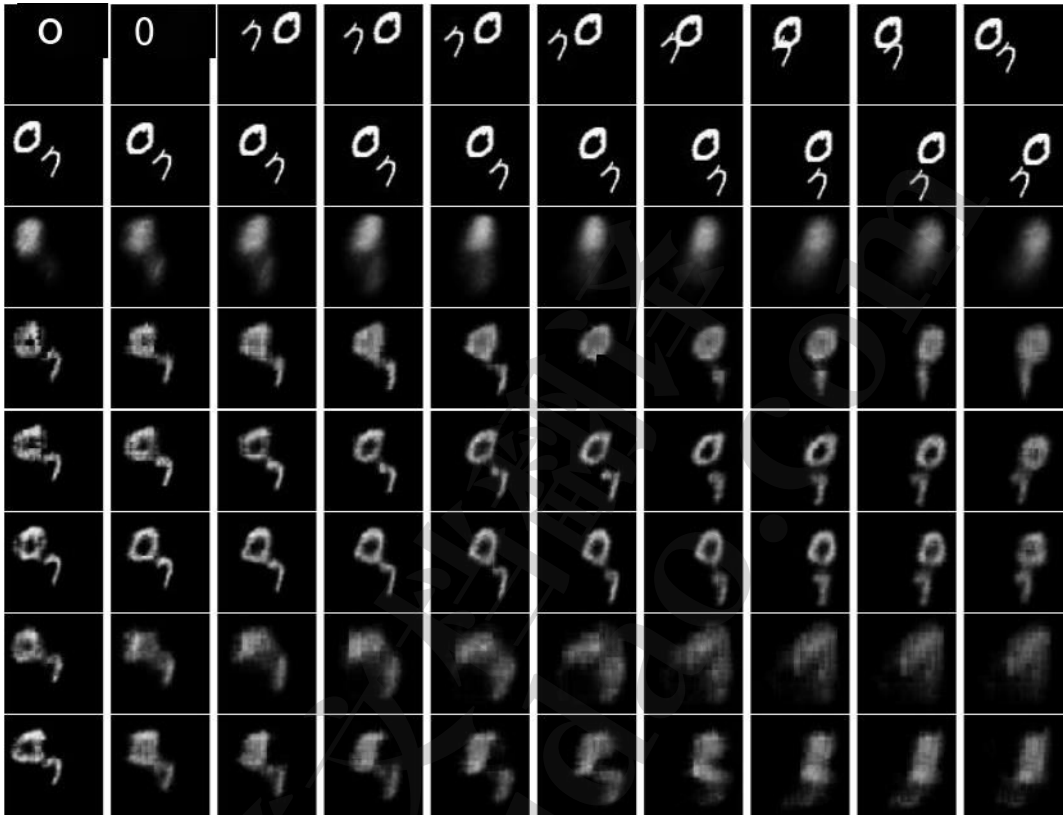


图 9:展示不同模型域内预测结果的说明性示例。从上到下:输入帧;地面实况;FC-LSTM;ConvLSTM-5X5-5X5 - 1-layer;ConvLSTM-5X5-5X5-2-layer;ConvLSTM-5X5-5X5-3-layer;ConvLSTM-9X9-1X1-2 -层;ConvLSTM-9X9-1X1-3-layer。

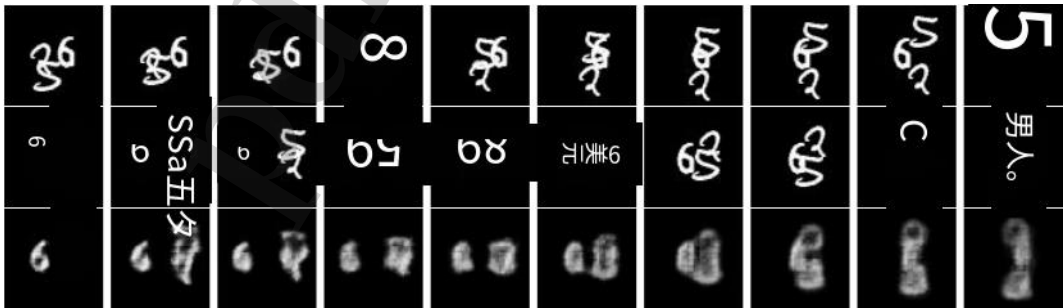


图 10:(大版本)显示出域外运行的说明性示例。从上到下:输入帧;地面实况;3 层网络的预测。

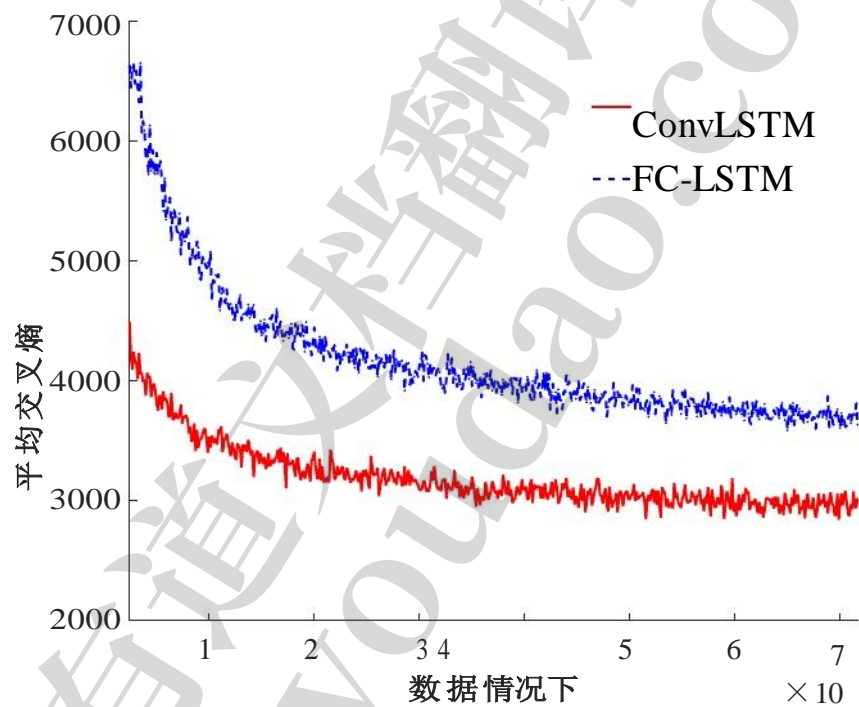


图 11:在线设置下的 3 层 ConvLSTM 和 FC-LSTM 的对比 在每次迭代中，我们生成一组新的训练样本，并记录该迷你批的平均交叉熵。x 轴是数据案例的数量(从 25600 开始)，y 轴是小批的平均交叉熵。我们可以发现，ConvLSTM 的损失比 FC-LSTM 下降得更快。