

PredRNN: A Recurrent Neural Network for Spatiotemporal Predictive Learning

摘要

时空序列的预测学习旨在通过从历史背景中学习来生成未来的图像，其中视觉动态被认为具有模块化结构，可以通过组成子系统学习。本文通过提出一种新的循环网络 PredRNN 对这些结构进行建模，其中一对记忆单元显式解耦，以几乎独立的过渡方式运行，最终形成复杂环境的统一表示。具体来说，除了 LSTM 的原始记忆单元外，该网络的特点是在所有层上以自下而上和自上而下的方向传播的锯齿形记忆流，使得 rnn 在不同层次上学习到的视觉动态能够进行交流。它还利用记忆解耦损失来防止记忆单元学习冗余特征。我们进一步提出了一种新的课程学习策略，迫使 PredRNN 从上下文框架中学习长期动态，这可以推广到大多数序列到序列模型。我们提供详细的消融研究，以验证每个组件的有效性。我们的方法在无动作和有动作条件的预测学习场景的五个数据集上获得了极具竞争力的结果。

关键词：预测学习；时空建模；循环神经网络

1 引言

作为预测学习的一个关键应用，从历史连续帧中生成未来帧已经引起了机器学习和计算机视觉社区越来越多的兴趣，它有许多实际应用和下游任务如降水预报 [2,3]、交通流量预测、深度强化学习和医学图像预测等。对于未来帧预测，时间和空间维度的特征都被需要，同样时空数据的低级细节和高级语义对于生成未来的框架也很重要，而传统的多层 RNN 网络的缺陷在于层与层之间是独立的，其忽略了顶层单元对底层单元的影响，此论文联合建模不同层次 RNN 的空间相关性和时间动态，提出了一种新的记忆预测框架——预测递归神经网络 (PredRNN) 和一个新的 RNN 单元 ST-LSTM，创建了一个统一的时空记忆池，将 LSTM 中记忆状态的内层转换扩展到整个时空记忆流。

2 相关工作

2.1 传统时空序列预测

传统时空序列预测起源于循环神经网络 RNN [1]，它将网络的输出作为下一个时间步新的输入，这样的方法导致其存在梯度消失和梯度爆炸。长短期记忆 LSTM 通过引入记忆细胞，选择性地对信息进行记忆和遗忘。卷积长短期记忆 ConvLSTM [2] 进一步加入了卷积操作，用于提取空间特征，也是论文所提出方法的基础。

2.2 PredRNN

此论文提出了一种称为时空记忆流的之字形循环过渡机制，一种具有一对解耦记忆细胞的新型卷积循环单元，以及一种用于序列到序列预测学习的新训练过程。

3 本文方法

3.1 本文方法概述

本文主要提出了三个方法，用于改进之前的工作：

1. **双流记忆 M、C**：时空数据的低级细节和高级语义对未来框架的生成都具有重要意义，M 引入了一个更深层次的转换路径，提高了短期动态建模能力，C 在较慢的时间尺度上运行，它在遥远的隐藏状态之间提供了更短的梯度路径，从而有利于学习长期信息。
2. **反向调度采样**：它迫使 PredRNN 从上下文框架中学习长期动态。
3. **记忆解耦损失**：为防止 C 和 M 学习冗余特征，使得不同的记忆状态 C、M 被训练成专注于事物的不同方面。

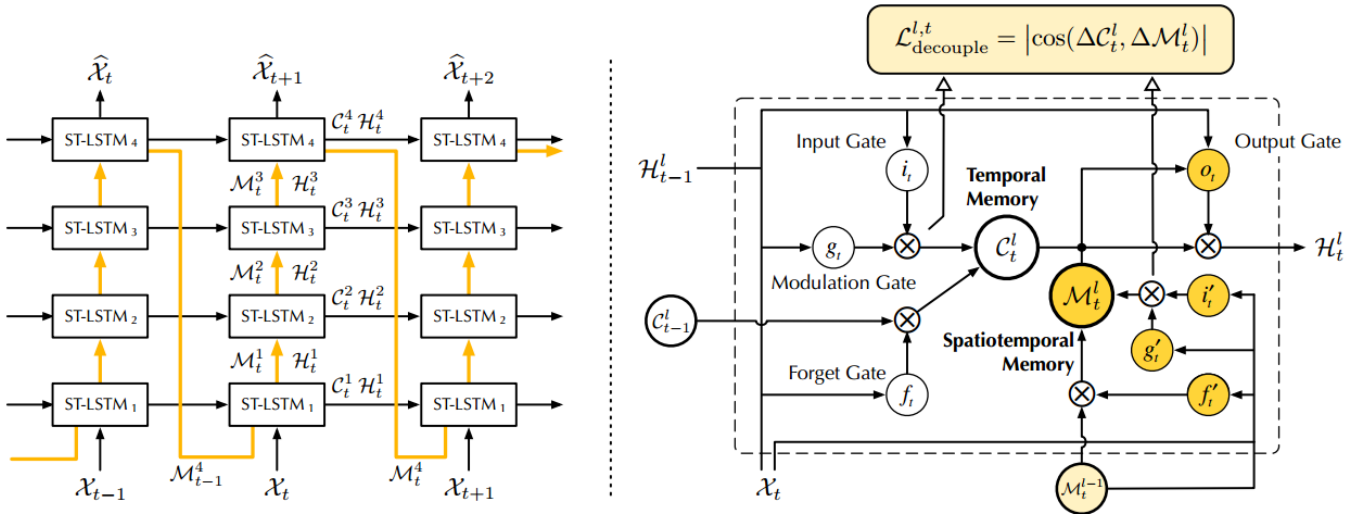


图 1. PredRNN 架构

3.2 时空记忆流

上图中橙色箭头表示的是时空记忆流，论文通过在之字形方向上更新记忆状态来改进以往的水平记忆流，这样不同 LSTM 层的记忆状态不再是独立的，t 时间输入的记忆单元 M 包含过去时间所有层次的信息，整个循环网络中的所有节点共同保持一个记忆。

3.3 解耦记忆细胞

PredRNN 包含不必要的冗余特征，因此提出了一种记忆解耦损失，以鼓励在任何时间步长下两个记忆状态的增量相互正交，释放了 C 和 M 在长期和短期动态建模中的各自优势。整

体记忆解耦方法表述如下：

$$\Delta C_t^l = W_{\text{decouple}} * (i_t \odot g_t), \quad \Delta M_t^l = W_{\text{decouple}} * (i'_t \odot g'_t)$$

$$\mathcal{L}_{\text{decouple}} = \sum_t \sum_l \sum_c \frac{|\langle \Delta C_t^l, \Delta M_t^l \rangle_c|}{\|\Delta C_t^l\|_c \cdot \|\Delta M_t^l\|_c}$$

说明：

- W_{decouple} 表示记忆解耦层使用的 1×1 卷积核（共享于所有 ST-LSTM 单元）。
- $(i_t \odot g_t)$ 和 $(i'_t \odot g'_t)$ 表示逐元素相乘操作，其中 \odot 是元素级乘法符号。
- $\langle \cdot, \cdot \rangle_c$ 表示在第 c 个通道上的点积运算（点积在该通道对应的展平特征图上进行）。
- $\|\cdot\|_c$ 表示在通道 c 上展平特征图后的 ℓ_2 范数。

3.4 反向调度采样

由于当前输入 X_t 比过去信息的信息量更大，所以 PredRNN 会过于依赖当前真实输入，而无法深入挖掘过去的历史信息，为了迫使循环模型从历史观察中学习长期动态，提出了反向调度采样：当前的输入随机选取真实的输入或者上一个时间步输出的预测结果。公式表达如下：

$$\hat{X}_{t+1} = \begin{cases} f_{\theta}(\hat{X}_t^{RSS} \rightarrow X_t, H_{t-1}, S_{t-1}) & \text{if } t \leq T, \\ f_{\theta}(X_t^{SS} \rightarrow \hat{X}_t, H_{t-1}, S_{t-1}) & \text{otherwise.} \end{cases}$$

4 复现细节

4.1 与已有开源代码对比

我利用了原论文的代码，并结合了一个用于图像预测的开源框架，其提供了一个基准，方便不同方法在同一实验条件下进行对比，并且能方便地对方法进行修改或扩展。

4.2 实验环境搭建

准备数据集，安装相关依赖，调整配置参数。

5 实验结果分析

本实验所用的数据集是：Moving MNIST，在此数据集上进行时空序列预测。首先动态生成训练数据，对模型进行训练，然后将连续几帧图像作为模型的输入，得到预测的未来连续帧，实验结果如下：



输入



真实结果



预测结果

6 总结与展望

总结：

- 本文介绍了一种用于时空预测学习的递归网络 PredRNN。PredRNN 采用一种新的时空 LSTM 单元，同时对空间外观的短期变形和多帧的长期动态进行建模。时空 LSTM 的核心是一个之字形的记忆流，它垂直地跨越堆叠的循环层，横向地通过所有的时间状态传播，这使得 PredRNN 在不同层次上的分层记忆表示能够相互作用。PredRNN 还使用了一种新的方法来解耦沿水平和之字形路径的记忆，以及提出了一种新的学习策略，称为反向调度采样，它迫使 PredRNN 的编码部分从更长的上下文框架中学习时间动态。
- 对论文中的模型 PredRNN 进行复现，实现了对移动数字集的未来帧预测，并且得到了与论文相近的效果。

展望：

- PredRNN 计算开销较大，未来可以通过改进模型的计算效率，减少训练和预测时的资源消耗，使得模型更具实际应用价值。
- 论文中引入动作元素，将动作与隐藏状态融合，从而获取过去动作信息，预测未来动作；因此可以考虑将胚胎图像时期信息融入，提升预测准确度。

参考文献

- [1] Marc'Aurelio Ranzato, Arthur Szlam, Joan Bruna, Michaël Mathieu, Ronan Collobert, and Sumit Chopra. Video (language) modeling: A baseline for generative models of natural videos. *arXiv preprint arXiv:1412.6604*, 2014.
- [2] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 802–810, 2015.
- [3] Yunbo Wang, Mingsheng Long, Jianmin Wang, Zhifeng Gao, and S. Yu Philip. PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 879–888, 2017.