

- 1.0 前言
 - 1.0.1 课程目标
 - 1.0.2 课程概览
- 1.1 序列预测问题
 - 1.1.1 序列
 - 1.1.2 序列预测
 - 1.1.3 序列分类
 - 1.1.4 序列生成
 - 1.1.5 序列到序列预测
- 1.2 感知机的局限性
- 1.3 序列模型的保证
- 1.4 LSTM网络
 - 1.4.1 LSTM权重
 - 1.4.2 LSTM门
- 1.5 LSTMs的应用
 - 1.5.1 自动图片标题生成
 - 1.5.2 文本自动翻译
 - 1.5.3 自动手写体生成
- 1.6 LSTMs的限制
- 1.7 扩展阅读
 - 1.7.1 序列预测问题
 - 1.7.2 MLPs用于序列预测
 - 1.7.3 RNNs的保障
 - 1.7.4 LSTMs
 - 1.7.5 LSTM应用
- 1.8 扩展
- 1.9 总结

1.0 前言

1.0.1 课程目标

本课程的目的是让你对LSTM有一个更高层次的理解，这样你就可以解释他们是什么以及它们是怎样工作的。

完成这一课之后，你会知道：

- 序列预测是什么，以及它们和一般的预测建模问题有什么样的区别；

- 多层感知器（Multilayer Perceptrons）在序列预测上的局限性、循环神经网络（Recurrent Neural Networks）在序列预测上的保障以及LSTM是怎样传递那种保障的；
- 一些令人影响深刻的LSTMs挑战序列预测问题的应用以及一些LSTMs局限性的警告

1.0.2 课程概览

本课程被划分为6节，它们是：

1. 序列预测问题；
2. 多层感知机的局限性；
3. 循环神经网络的保障；
4. LSTM网络；
5. LSTM网络的应用；
6. LSTM网络的局限性。

让我们开始吧！

1.1 序列预测问题

序列预测和其他类型的监督学习问题来说是不同的。序列强调观察值的顺序，当训练模型和做出预测的时候这个顺序必须被保存。总的来说，设计序列数据的预测问题被称为序列预测问题，尽管基于输入和输出顺序的不同，这种称法还有很多问题。本节，我们将会看四种类型的序列预测问题：

1. 序列预测（Sequence Prediction）；
2. 序列分类（Sequence Classification）；
3. 序列生成（Sequence Generation）；
4. 序列到序列的预测（Sequence-to-Sequence Prediction）。

但是，首先，让我们一起确定一个集合和一个序列之间的区别。

1.1.1 序列

在应用机器学习中我们经常用到集合，例如训练集和测试集。集合中的每个样本可以被认为是来自其范围内的观察值。在一个集合中，观察的顺序并不重要。一个序列则不同，序列强调观察值的详细的顺序。这里顺序非常重要！在使用序列数据作为模型的输入和输出时，必须在预测问题的定义中特别地慎重对待！

1.1.2 序列预测

序列预测包括预测给定输入序列的下一个值。例如：

输入序列: 1,2,3,4,5
输出序列: 6

列表 1.1: 序列预测问题的例子

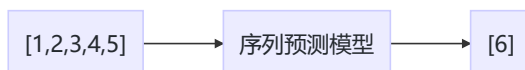


图1.1 序列预测问题的描述

序列预测问题也通常被称为序列学习。从技术上讲，我们可以将所有下面的问题都认为是序列预测问题的其中一种类型。这可能会让初学者感到困惑。

连续数据学习仍然是模式识别和机器学习中的一项基本任务和挑战。包含数据顺序的应用可能需要对新事件的预测、新序列的生成或者诸如序列或者子序列分类的决策。

— On Prediction Using Variable Order Markov Models, 2004.

一般来说，本书中，我们将使用“序列预测”来指代具有序列数据预测问题的这一类别。然而，在这部分，我们将会把序列预测和其他形式的具有序列数据的预测进行区分，将其定义为作为预测下一个时间步长的预测。

序列预测尝试去根据前面的元素来预测序列的元素。

— Sequence Learning: From Recognition and Prediction to Sequential Decision Making, 2001.

一些序列预测问题的问题包括：

- **天气预测。** 给定一个序列的基于时间的天气的观察值，预测明天的天气。
- **股票预测。** 给定一个基于时间的有价证券序列波动值，预测明天有价证券的波动。
- **产品推荐。** 给定一个客户曾经的购物情况，预测下一个阶段客户的购物。

1.1.3 序列分类

序列分类包括预测给定输入序列的分类标签。例如：

输入序列: 1,2,3,4,5
输出序列: "good"

列表 1.2: 序列分类问题的例子

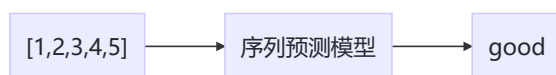


图1.2 序列分类问题的描述

序列分类问题的目的是利用标记数据集[...]建立一个分类模型，以便该模型可以用来预测一个看不见的序列的分类标签。

— Discrete Sequence Classification, Data Classification: Algorithms and Applications, 2015.

输入序列可以由实际值或者离散值组成。在后一种情况下，这些问题可以被称为离散序列分类问题。序列分类问题的一些例子包括：

- **DNA序列分类。** 给定DNA序列值A,C,G和T，预测序列是编码区域还是非编码区域。
- **自动检测。** 给定一个序列的观察值，预测序列是否是反常的。
- **情感分析。** 给定一个文本的序列，例如综述或者是推特，预测这段文本的情感是积极的还是消极的。

1.1.4 序列生成

序列生成包含一个新的输出序列，该序列和语料集里面序列有着相同的特征。例如：

输入序列：[1,3,5],[7,9,11]
输出序列：[3,5,7]

列表 1.3：序列生成问题的例子

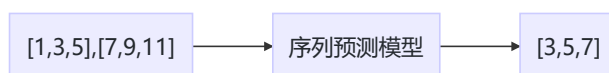


图1.3 序列生成问题的描述

循环神经网络（RNN）可以通过一步一步地处理真实数据序列来预测序列生成，并预测接下来会发生什么。[递归神经网络]可以通过一步一步地处理真实数据序列来预测序列生成，并预测接下来会发生什么。假设预测是概率的，可以通过从网络的输出分布迭代采样来从训练网络生成新序列，然后在下一步骤中将样本喂给输入。换句话说，让网络把他们的发明视为真实的，就像一个人做梦一样。相对新的，seq2seq方法在[结果]实现了state-of-the-art机器翻译。

— Generating Sequences With Recurrent Neural Networks, 2013.

序列生成问题的一些例子包括：

- **文本生成：** 给定一个语料集的文本，例如莎士比亚的文学作品，生成新的句子或者段落的文本，它们可以从语料集中提取出来。

- **手写体预测**：给定一个手写体的语料库，生成具有在语料库中具有手写属性的新的短语的手写体。
- **音乐生成**：给定音乐实例的语料库，生成具有语料库属性的新音乐片段。序列模型可以指以单次观测作为输入的序列生成。一个例子是图像的自动文本描述。
- **图像字幕的生成**：给定图像作为输入，生成描述该图像的单词序列。

例如：

输入序列：[图像像素]
输出序列：["一个人骑着自行车"]

列表 1.4：序列生成问题的例子

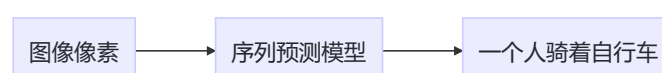


图1.4 序列生成问题的描述

使用适当的句子能够自动描述图像内容是一项非常具有挑战性的任务，但是它可能会产生很大的影响[...]。确实，一个描述不仅必须捕获图像中所包含的对象，而且还必须表示这些对象是如何关联的，以及它们的属性和它们所涉及的活动。

— Show and Tell: A Neural Image Caption Generator, 2015.

1.1.5 序列到序列预测

序列到序列的预测涉及给定一个输入序列并预测一个输出序列，例如：

输入序列：1, 2, 3, 4, 5
输出序列：6, 7, 8, 9, 10

列表 1.5：序列到序列问题的例子

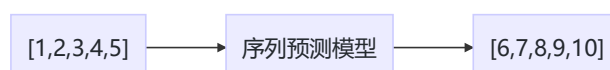


图1.5 序列到序列问题的描述

尽管深度神经网络具有灵活性，能力强大，但是它只能应用于这类问题：即输入和目标可以很明显地被编码为一个固定维度向量。这是一个很明显的限制，因为很多重要问题的输入最好用一个长度未知的序列来表示，并没有一个可知的先验。例如，原因识别和机器翻译是连续的问题。同样地，问答也可以被看做是将代表问题的单词序列映射到表示答案的单词序列。

— Sequence to Sequence Learning with Neural Networks, 2014.

序列到序列预测是序列预测一个很微妙而又很具有挑战性的扩展。不是在在序列中预测单一的下一个值，一个新的序列被预测出来，该序列可能会或者可能不会和输入序列有着相同的长度或者同一时间。这类问题最近在文本自动翻译领域出现了很多的研究（如翻译英语到语法）可以被缩写为seq2seq。seq2seq学习，其核心是，使用循环神经网络来将变长的输入序列映射成为一个可变长度输出序列。

— Multi-task Sequence to Sequence Learning, 2016.

如果输入和输出的序列都是时间序列，那么该问题可以被称为多步时间序列预测。序列到序列问题的一些例子包括：

- **多步时间序列预测。** 给定一系列时间观察值，预测一系列未来时间步长的观测序列。
- **文本摘要。** 给定文本文档，预测描述源原档的突出部分的较短文本序列。
- **程序执行。** 给定文本描述程序或数学方程，预描述正确输出的字符序列。

1.2 感知机的局限性

传统的神经网络叫做感知机（Multilayer Perceptrons），或者简称为MLP，可以被用于序列模型的预测问题。MLP近似从输入变量到输出变量的映射函数。由于一系列的原因，其总的能力对于序列预测问题是很有用的（特别是时间序列预测）。

- **对噪声很健壮。** 神经网络对输入数据和映射函数中的噪声具有很强的鲁棒性，甚至可以在缺失值的存在下支持学习和预测。
- **非线性。** 神经网络对映射函数不做强假设，容易学习线性和非线性关系。

另外，在映射函数中，MLP可以被配置为支持任意数量的输入和输出，但有一定数量的输入和输出。这意味着：

- **多变量输入。** 可以指定任意数量的输入，为多变量预测提供提供直接的支持。
- **多步输出。** 可以指定任意数量的输出，为多步甚至多变量预测提供直接支持。

其能力克服了使用传统线性方法的限制（如用于时间序列预测的ARIMA）。仅凭这些能力，前向神经网络被广泛应用于时间序列预测。

ARIMA模型全称为自回归积分滑动平均模型(Autoregressive Integrated Moving Average Model, 简记ARIMA)。ARIMA模型的基本思想是：将预测对象随时间推移而形成的数据序列视为一个随机序列，用一定的数学模型来近似描述这个序列。这个模型一旦被识别后就可以从时间序列的过去值及现在值来预测未来值。

...神经网络的一个重要的贡献——即它们优雅地逼近任意非线性函数的能力。这种性质在时间序列处理中具有很高的价值，并保证了更强大的应用，特别是在预测的子领域中...

— Neural Networks for Time Series Processing, 1996.

MLP在序列预测上的应用需要输入序列被划分成较小的重叠子序列，这些子序列被显示成给网络用来生成一个预测。输入序列的时间步长成为网络的输入特征。子序列是重叠的，以模拟沿该序列滑动的窗口，以便生成所需的输出。这在一些问题上效果很好，但是它有5个关键的限制：

- **无状态。**MLP学习一个固定函数的近似。在输入序列的上下文中有条件的任何输出必须被泛化和固定到网络权重中。
- **不知道时间结构。**时间步长被建模为输入特征，这意味着网络对观测之间的时间结构或者顺序没有明确的处理或者理解。
- **尺度混乱。**对于需要建模多个并行输入序列的问题，输入特征的数量作为滑动窗口的大小的额外一个元素而增加，而没有任何时间序列的显示分离。
- **固定大小的输入。**滑动窗口的大小是固定的，必须强加到所有的输入中。
- **固定大小的输出。**输出的大小也是固定的，任何不符合的输出都必须强制成为固定的大小。

MLP确实为序列预测提供了很大的能力，但是仍然受到整个关键限制的限制，即必须在模型设计中明确地预先说明观测之间的时间依赖的范围。

序列到序列对“神经网络”提出了挑战，因为它们要求输入和输出的位数是已知的并且是固定的。

— Sequence to Sequence Learning with Neural Networks, 2014

MLP是建模序列到序列模型问题的好的起点，但是我们现在有更好的选择。

1.3 序列模型的保证

长短时记忆，或者LSTM网络是一种类型的循环神经网络。循环神经网络（简称RNN）是一种特殊类型的用于序列问题的神经网络。给定一个标准的前馈MLP网络，RNN可以被认为是添加到体系结构的环路。例如，在给定的层中，除了正向下一层之外，每个神经元可以逐渐地（侧向）传递其信号。网络的输出可以用下一个输入向量作为输入反馈给网络，等等。

循环连接将状态或存储器添加到网络中，并允许它学习和利用输入序列中观测的有序性。

...循环神经网络包含环，这个环将上一个阶段的网络激活作为输入喂给网络，来影响当前阶段的预测。这些激活被存储在网络的内部状态中，其原则上可以保持长期时间上下文信息。这种机制允许RNN在输入序列历史上利用动态变化的上下文窗口。

— Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling, 2014

序列的增加是被逼近的函数的一个新的维度。网络可以学习一个基于时间的从输入到输出的映射函数，而不是单纯将输入映射到输出。内部存储器可以表示输出时基于输入序列的中的最近上下文，而不是刚被呈现为网络的输入。从某种意义上说，这种能力解锁了神经网络的时间序列。

长短时记忆（LSTM）可以解决很多不能被反馈网络使用固定大小时间窗口解决的任务。

— Applying LSTM to Time Series Predictable through Time-Window Approaches, 2001

除了使用神经网络进行序列预测的一般方法外，RNN还可以学习和利用数据的时间依赖性。也就是说，在最简单的情况下，网络显示一个序列中的一个时间的一次观察，并且可以了解先前观察到的观测结果是相关的，以及它们是如何预测相关的。

由于这种学习序列中长期相关性的能力，LSTM网络避免了预先指定时间窗口的需要，并且能够精确地模拟复杂的多变量序列。

— Long Short Term Memory Networks for Anomaly Detection in Time Series, 2015

循环神经网络的保证是输入数据中的时间依赖和上下文信息可以被学习到。

循环神经网络的输入不是固定的，而是构成一个序列，可以用来将输入序列转换成输出序列，同时以可理解的方式考虑上下文信息。

— Learning Long-Term Dependencies with Gradient Descent is Dicult, 1994.

有许多种的RNN，但是LSTM在序列预测中传递RNN的保证。这就是为什么LSTM现在有这么多的声音和应用的原因。

LSTM具有内部状态，它们明确地知道输入中的时间结构，能够分别对多个并行输入序列进行建模，并且可以通过不同长度的输入序列来产生可变长度的输出序列，每一时间一次观察。

接下来，让我们仔细看看LSTM网络。

1.4 LSTM网络

LSTM网络与传统MLP是不同的。像MLP，网络由神经元层组成。输入数据通过网络传播以进行预测。与RNN一样，LSTM具有递归连接，使得来自先前时间步的神经元的先前激活状态被用作形成输出的上下文。

和其他的RNN不一样，LSTM具有一个独特的公式，使其避免防止出现阻止和缩放其他RNN的问题。这，以及令人影响深刻的结果是可以实现的，这也是这项技术得以普及的原因。

RNNs一直以来所面临的一个关键问题是怎么样有效地训练它们。实验表明，权重更新过程导致权重变化，权重很快变成了如此之小，小到没有效果（梯度消失）或者权重变得如此之大，导致非常大的变化或者溢出（梯度爆炸），这问题是非常的困难的。LSTM通过设计而克服了这一困难。

不幸的是，标准的RNN可以访问的上下文信息的范围实际上是非常有限的。问题是隐藏层上给定输入，以及因此在网络上的输出，当它围绕网络的经常性连接循环时，要么指数衰减，要么指数上升。这个缺点...在文献中被称为梯度消失问题...长短时记忆（LSTM）是一种设计用于解决梯度消失问题的RNN体系结构。

— A Novel Connectionist System for Unconstrained Handwriting Recognition, 2009

LSTM网络的计算单元被称为存储单元（memory cell），存储器块（memory block）或者简称单元（cell）。当描述MLPs时，术语“神经元”作为计算单元是根深蒂固的，因此它经常被用来指LSTM存储单元。LSTM单元由权重和门组成。

长短时结构是通过对现有RNN中的误差进行分析，发现长时间滞后对现有体系结构是不可访问的，因为反向传播误差要么指数上升要么衰减。LSTM层由一组递归连接的块组成，称为存储块（memory blocks）。这些块可以被认为是数字计算机中存储芯片的可微版本。每一个包含一

个或者多个递归连接的存储单元和单个乘法单元——输入门、输出门和遗忘门，它们为单元提供写、读和重置操作的连续模拟。...网络只能通过门与细胞相互作用。

— Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures, 2005.

1.4.1 LSTM权重

一个记忆单元具有输入、输出的权重参数，以及通过暴露于输入时间步长而建立的内部状态。

- **输入权重。**用于对当前时间步长的输入进行加权。
- **输出权重。**用于对上次步骤的输出进行加权。
- **内部状态。**在这个时间步长的输出计算中使用内部状态。

1.4.2 LSTM门

记忆单元的关键是门。这些也是加权函数，它们进一步控制单元中的信息。有三个门：

- **遗忘门。**决定什么样的信息需要从单元中丢弃。
- **输入门。**决定输入中哪些值来更新记忆状态。
- **输出门。**根据输入和单元的内存决定输出什么。

在内部状态的更新中使用了遗忘门和输入门。输入门是单元实际输出什么的最后限制。正是这些门和一致的数据流被称为CEC（constant error carousel），它保持每个单元稳定（既不爆炸或者消失）。

每个存储单元的内部结构保证CEC（its constant error carousel）。这表示桥接的基础滞后很长的时间。两个门单元学习在每个存储单元CEC中的打开和关闭对错误的访问。乘法输入门提供保护的CEC从扰动无关的输入。同样地，乘法输出门保护其他单元不受当前不相关存储器内容的干扰。

— Long Short-Term Memory, 1997.

不像传统的MLP神经元，很难画出一个干净的LSTM存储单元。到处都有线、权重和门。看看本章末尾的一些资源，如果你认为基于图片或者基于等式的LSTM内部描述将有助于进一步研究。我们可以将LSTM的3个关键术语归纳为：

- 克服了训练RNN的技术问题，即梯度消失和梯度爆炸问题。
- 拥有记忆来克服与输入序列相关的长期时间依赖问题。
- 一个个时间步长地处理输入序列和输出序列，允许可变长度的输入和输出。

接下来，让我们看看一些例子，其中LSTMs解决了一些具有挑战性的问题。

1.5 LSTMs的应用

我们感兴趣的是LSTMs为解决序列预测问题提供的高雅的解决方案。本节提供3个示例，为您提供LSTM能够实现的结果快照。

1.5.1 自动图片标题生成

自动图片标题的任务是在给定图像的情况下，系统必须生成一个标题，描述图像的内容。在2014年，有深度学习算法实现实现了突破性的进展，利用顶部模型在图片中进行物体分类和对象检测，在这个问题上取得了非常令人印象深刻的结果。

一旦你可以在照片中检测出对象并未这些对象生成标签，你就可以看到下一步就是将这些标签转换成连贯的句子描述。该系统涉及使用非常大的卷积神经网络在照片中检测物体，然后LSTM将标签转换成连贯的句子。

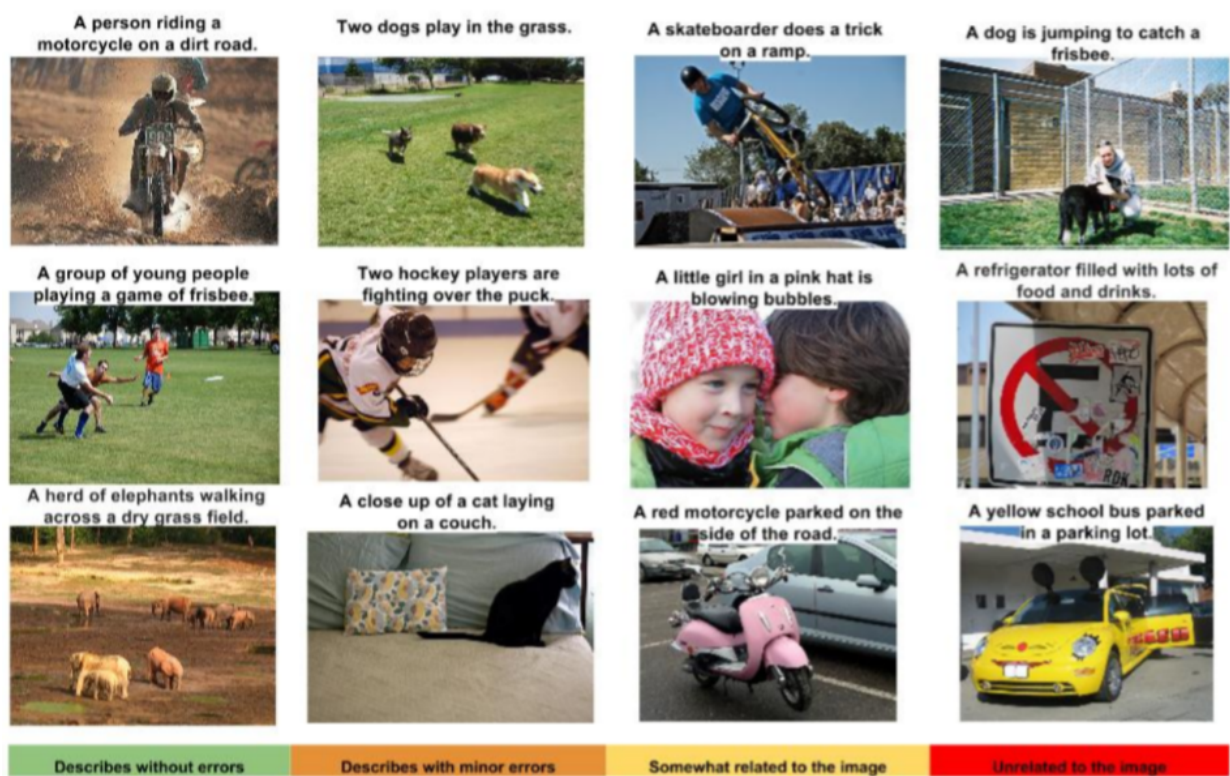


图1.6 LSTM生成标题的示例，取自《显示并告知：一个神经网络标题生成器》，2014

1.5.2 文本自动翻译

自动文本翻译的认为是在一个语言中给出文本的句子，并且必须把它们翻译成为另外一种语言的文本。例如，英语句子作为输入，语法句子作为输出。该模型必须学习单词的翻译，翻译的上下文被修改，并支持输入和输出序列的长度可能总体上相互变化。

图1.7 将英文文本翻译成法语的例子，从预测到预期的翻译，出自神经网络的序列到序列的学习，2014

1.5.3 自动手写体生成

这个任务中，在给定手写体语料库的情况下，生成给定单词或者短语的新的手写体。当手写样本被创建时，手写被提供为笔所使用的坐标序列。在这个语料库中，学习了笔运动与字母之间的关系，并产生了新的例子。有趣的是，不同的风格可以学习，然后模仿。我希望看到这项工作结合一些法医笔记分析的专业知识。图1.8 LSTM生成字幕的例子，摘自《循环神经网络的生成序列》，2014

1.6 LSTMs的限制

LSTMs给人留下了深刻的印象。网络的设计克服了RNNs的技术挑战，用神经网络实现了对序列预测的保证。LSTM的应用在一系列的复杂问题上取得了令人印象深刻的结果。但是LSTMs对于所有的序列预测问题可能不是理想的。

例如，在时间序列预测中，通常用于预测信息在过去观察的一个小窗口内。通常，具有窗口或线性模型的MLP可能是一个不太复杂和更合适的模型。

文献中发现的时间序列基准问题...通常概念上比LSTM已经解决的许多任务更简单。它们通常不需要RNNs，因为有关下一个事件的所有相关信息都是由一些在小时间窗口中包含的最近事件传达的。

— Applying LSTM to Time Series Predictable through Time-Window Approaches, 2001

LSTM的一个重要的局限是记忆。或者更准确地说，记忆是如何被滥用的。有可能迫使LSTM模型在很长的输入时间步长上记住单个观察。这是LSTM的不良使用，并且需要LSTM模型记住多个观察将失败。当将LSTM应用于时间序列预测时，可以看出，该问题表述为回归，要求删除是输入序列中的多个遥远时间步长的函数。一个LSTM可能被迫在这个问题上执行，但是通常比一个精心设计的自回归模型或重新考虑问题一般少一些。

假设任何动态模型都需要 $t-\tau$ ，...，我们注意到[autoregression]-RNN必须将 $t-\tau$ 的所有输入存储到 t ，并在适当的时间覆盖它们。这需要一个循环的缓存，这是一个对RNN进行模拟很困难的结构。

— Applying LSTM to Time Series Predictable through Time-Window Approaches, 2001

警告，LSTM并不是一种令人期待的新技术，而且需要仔细的考虑你问题的框架。把LSTMs的内部状态看做是一个方便的内部变量来捕捉和提供预测的背景。如果你的问题看起来像一个传统的自回归问题，在一个小窗口内具有最相关的滞后观察，那么在考虑LSTM之前，也许使用MLP和滑动窗口开发性能的基线。

基于时间窗口的MLP在某些时间序列预测基准上优于LSTM的pure-[autoregression]方法，仅通过查看最近的一些输入来解决。因此，LSTM的特殊长处，即学会记住很长的、位置的时间段的单个事件，是不必要的。

— Applying LSTM to Time Series Predictable through Time-Window Approaches, 2001

1.7 扩展阅读

如果你想深入研究算法的技术细节，下面是一些关于LSTM的文章。

1.7.1 序列预测问题

- [Sequence on Wikipedia](#).
- On Prediction Using Variable Order Markov Models, 2004.
- Sequence Learning: From Recognition and Prediction to Sequential Decision Making, 2001.
- [Chapter 14, Discrete Sequence Classification, Data Classification: Algorithms and Applications, 2015](#).
- [Generating Sequences With Recurrent Neural Networks, 2013](#).
- [Show and Tell: A Neural Image Caption Generator, 2015](#).
- [Multi-task Sequence to Sequence Learning, 2016](#).
- [Sequence to Sequence Learning with Neural Networks, 2014](#).
- Recursive and direct multi-step forecasting: the best of both worlds, 2012.

1.7.2 MLPs用于序列预测

- Neural Networks for Time Series Processing, 1996.
- [Sequence to Sequence Learning with Neural Networks, 2014](#).

1.7.3 RNNs的保障

- Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling, 2014.
- Applying LSTM to Time Series Predictable through Time-Window Approaches, 2001.
- Long Short Term Memory Networks for Anomaly Detection in Time Series, 2015.
- Learning Long-Term Dependencies with Gradient Descent is Dicult, 1994.
- [On the diculty of training Recurrent Neural Networks, 2013](#).

1.7.4 LSTMs

- Long Short-Term Memory, 1997. Learning to forget: Continual prediction with LSTM, 2000.
- A Novel Connectionist System for Unconstrained Handwriting Recognition, 2009.
- Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures, 2005.

1.7.5 LSTM应用

- [Show and Tell: A Neural Image Caption Generator, 2014](#).
- [Sequence to Sequence Learning with Neural Networks, 2014](#).

- [Generating Sequences With Recurrent Neural Networks, 2014.](#)

1.8 扩展

你想加深对LSTMs的理解吗？本节列出了一些您可能需要考虑的具有挑战性的扩展。

- 为新手学习者写一段LSTM的描述。
- 列出10个你能够想到的序列预测问题的例子。
- 研究并绘制（或者重绘）具有连接、权重和门的LSTM存储单元的图片。
- 研究并列出10个有趣的LSTMs的应用实例，并清楚地描述预测序列问题，包括输入和输出。
- 研究并在电子表格或者Python中用单个的LSTM存储单元进行预测编程。

在网上张贴你的扩展，和我分享这个链接，我很想看到你是怎么样想的。

1.9 总结

在本课中，您发现了用于序列预测的长短时记忆循环神经网络。你知道吗？

- 序列预测是什么？它们是如何不同于一般预测建模问题的。
- 多层感知机用于序列预测的局限性，循环神经网络用于序列预测的保障，以及LSTM是如何实现该承诺的。
- 令人深刻的应用LSTM挑战序列预测问题，以及LSTM局限性的一些警告。接下来，您将发现如何使用反向传播通过时间训练算法来训练LSTM。