

# 神经网络序列到序列的学习

March 8, 2018

## Abstract

深度神经网络（DNNs）是强大的模型，在困难的学习任务上取得了出色的表现。尽管DNNs在很大的标签训练集上的方法，对序列结构做出最小假设。我们的方法使用一个多层长短期记忆网络（LSTM）来把输入序列映射到一个固定维度的向量，然后另一个深度LSTM网络从这个向量中解码出目标序列。我们的主要结果是基于WMT-14数据集做的英语到法语的翻译，这个由LSTM产生的结果在整个测试数据集上达到了34.8的BLEU得分，BLEU得分是衡量机器翻译质量的一个标准。作为比较，一个基于短语的SMT系统在相同的数据集上获得了33.3的BLEU得分。当我们使用LSTM来重跑由之前SMT系统产生的1000个假设，BLEU得分增加到了36.5，很接近先前的最新状态。LSTM也能学习合理的短语和句子表述，对词序很敏感，对主动和被动语态相对不变（不太敏感）。我们发现把源句子（ $\rightarrow d, c, b, a$ ）输入，LSTM的表现明显提高了，因为这样做引入了源和目标语句之间的许多短期依赖关系，从而使优化

## 1 介绍

深度神经网络（DNNs）是非常强大的机器学习模型，注意在诸如语音识别[13, 7]和视觉对象识别[19, 6, 21, 20]等难题上实现卓越的性能。DNNs功能强大，因为对于少量步骤，它们可以执行任意并行计算。DNNs需要足够的信息来指定网络的参数，就可以用监督反向传播来训练。因此，如果存在获得良好结果的大数据集的参数，尽管它们具有灵活性和强大的功能，但DNNs只能用于输入和目标可以用维数固定的矢量进行合理编码的问题。机器翻译是序列问题。同样，问题回答也可以被看作是将表示问题的单词序列映射到表示答案的单词序列。

序列对DNNs提出了挑战，因为它们要求输入和输出的维度是已知的并且是固定的。在本文中，我们展示了一个简单长短期记忆（LSTM）体系结构[16]的直接应用可以解决序列问题的一般序列问题。这个想法是使用LSTM来读取输入序列，一次一个时间步，以获得大的固定维向量表示，然后使用另一个从该向量提取输出序列。LSTM本质上是一个递归的神经网络语言模型[28, 23, 30]，除了它是以输入序列为条件的。由于输入和它们相应输出成功处理长时间依赖关系数据的能力使其成为该应用的自然选择。

已经有许多相关的尝试来解决用神经网络对序列学习问题进行排序的一般序列。我们的方法与Kalchbrenner和Blunsom密切相关[18]，他们首先将整个输入句子映射为向量，并且与Cho等人非常相似[5]。Graves[10]引入了一种新颖的可区分关注机制，它允许神经网络将注意力集中在输入的部分，并且这一想法的一个优雅变体被Bahdanau等人成功应用于机器翻译[2]。连接主义序列分类是将序列映射到具有神经网络的序列的另一种流行技术，尽管它假设输入和输出之间是单调对齐的[11]。

这项工作的主要成果如下。在WMT'14英语到法语翻译任务中，我们使用简单的从左到右的波束搜索解码器，通过从5个深层（全部380m参数）的集合直接提取翻译来获得34.81的BLEU分数。这是迄今为止通过大模型达到的最高分数。这个34.81 BLEU得分是由一个词汇为80k词的词条实现的，因此只要参考翻译包含一个未被这80k涵盖的单词，我们使用LSTM在相同的任务中重新公开了SMT基线的1000个最佳列表[29]。通过这样做，我们获得了令人惊讶的是，尽管最近有其他研究人员使用相关体系结构[26]，但LSTM并没有遭受很长的句子。我们能够使用LSTM的一个有用属性是它学习将可变长度的输入句子映射为固定维向量表示。由于翻译往往是对源句的解释，LSTM找到能够捕捉其含义的句子表示法，因为具有相似含义的句子彼此接近，而不同的句子含义将会很远。定性

## 2 模型

递归神经网络 (rnn) [31, 28] 是前馈神经网络对序列的自然泛化。给定输入序列  $(x_1, \dots, x_t)$ ，标准rnn通过迭代以下等式来计算输出序列  $(y_1, \dots, y_t)$ ：

$$h_t = \text{sigm}(W^{hx}x_t + W^{hh}h_{t-1})$$

$$y_t = W^{yh}h_t$$

只要提前知道输出之间的对齐，rnn就可以轻松地将序列映射到序列。然而，对于复杂和非单调关系的输入和输出序列长度不同的问题，还不清楚如何解决。

一个简单的通用序列学习策略是将输入序列映射到一个固定大小的载体上，然后用另一个rnn将载体映射到目标序列（Cho等人[5]也采用了这种方法）。尽管原则上它可以工作，因为rnn提供了所有的相关信息，但由于长期的相关性[14, 4]（图1）[16, 15]，难以训练rnns。然而，长时间的短期m可能会成功。

## References

[1] The lyx Tutorial,