**Sequence to Sequence Learning with Neural Networks**

**Ilya Sutskever** Google ilyasu@google.com

**Oriol Vinyals**

Google

vinyals@google.com

**Quoc V. Le**

Google

qvl@google.com

**Abstract**

Deep Neural Networks (DNNs) are powerful models that have achieved excel- lent performance on difficult learning tasks. Although DNNs work well whenever large labeled training sets are available, they cannot be used to map sequences to sequences. In this paper, we present a general end-to-end approach to sequence learning that makes minimal assumptions on the sequence structure. Our method uses a multilayered Long Short-Term Memory (LSTM) to map the input sequence to a vector of a fixed dimensionality, and then another deep LSTM to decode the target sequence from the vector. Our main result is that on an English to French translation task from the WMT-14 dataset, the translations produced by the LSTM achieve a BLEU score of 34.8 on the entire test set, where the LSTM’s BLEU score was penalized on out-of-vocabulary words. Additionally, the LSTM did not have difficulty on long sentences. For comparison, a phrase-based SMT system achieves a BLEU score of 33.3 on the same dataset. When we used the LSTM to rerank the 1000 hypotheses produced by the aforementioned SMT system, its BLEU score increases to 36.5, which is close to the previous state of the art. The LSTM also learned sensible phrase and sentence

representations that are sensitive to word order and are relatively invariant to the active and the passive voice. Finally, we found that reversing the order of the words in all source sentences (but not target sentences) improved the LSTM’s performance markedly, because doing so introduced many short term dependencies between the source and the target sentence which made the optimization problem easier.

**摘要**

深度神经网络（DNN）是一种功能强大的模型，在艰难的学习任务中可以取得卓越的性能。虽然DNN在大型标记训练集可用时效果很好，但它们不能用于将序列映射到序列。在本文中，我们提出了一种通用的端到端序列学习方法，它对序列结构做出了最小的假设。我们的方法使用多层长短期记忆（LSTM）将输入序列映射到固定维度的矢量，然后使用另一个深LSTM来解码来自矢量的目标序列。我们的主要结果是，在WMT-14数据集的英语到法语翻译任务中，LSTM产生的翻译在整个测试集上获得了34.8的BLEU分数，其中LSTM的BLEU分数在词汇量不足时受到惩罚话。此外，LSTM在长句上没有困难。为了比较，基于短语的SMT系统在同一数据集上实现了33.3的BLEU分数。当我们使用LSTM重新调整上述SMT系统产生的1000个假设时，其BLEU分数增加到36.5，这接近于先前的技术水平。 LSTM还学习了对词序有敏感且对主动和被动语音相对不变的合理的短语和句子表示。最后，我们发现，逆转所有源语句（但不是目标句子）中单词的顺序可以显着提高LSTM的性能，因为这样做会在源语句和目标语句之间引入许多短期依赖关系，从而使优化问题更容易。