# 介绍：

最近提出的神经翻译机是一种新兴的机器翻译方法，由Kalchbrenner和Blunsom（2013），Sutskever（2014）和Cho（2014B）等人。不同于传统的基于短语的翻译系统（参见，例如Koehn等人，2003），其由许多翻译子组件调谐地组合成的系统，神经机器翻译尝试构建和训练单个大型神经网络，读取句子并输出正确的翻译。

大多数提出的神经翻译机模型属于编码器 - 解码器（Sutskever等人，2014; Cho等人，2014a），具有用于每种语言的编码器和解码器，或者涉及一种特定语言的编码器，它将被应用于每一个句子，它的输出也将被进行对比（Hermann and Blunsom，2014）。神经网络编码器读取源语句，并将之编码成为一个固定长度的向量。然后，解码器从编码向量输出转换。整个编码器 - 解码器系统由编码器和用于一对语言的解码器组成。整个系统将合起来一起参与神经网络的训练，以最大限度地提升源语句正确翻译的概率。

这种编码器 - 解码器方法的潜在问题是神经网络需要能够将源语句的所有必要信息压缩为固定长度向量。这可能使神经网络难以应付长句子，特别是比训练语料库中句子更长的源语句。 Cho等人（2014b）表示，此系统确实表现出了随着输入句子的长度增加而编码器 - 解码器效果迅速恶化的现象。

为了解决这个问题，我们向编码器 - 解码器模型引入了一致的扩展协调和平移。每次提出的模型在翻译中产生一个单词，并在源语句中搜索一组位置，这些位置是最相关信息最集中的位置。然后，该模型基于与这些源位置和所有先前生成的目标词相关联的上下文向量来预测目标词。

这种方法与基本的编码器 - 解码器最重要的区别在于它不会将整个输入句子编码成单个固定长度的矢量。相反，它将输入句子编码为一系列向量，并在解码翻译时自适应地选择这些矢量的子集。这释放了神经翻译模型的限制，不必将源语句的所有信息压缩成固定长度的矢量，而不管其长度如何。我们将展示这将允许模型能够更好地处理长句。

在本文中，我们表明，提出的一致扩展协调和平移的方法可以显着提高基本编码器 - 解码器方法的翻译性能。，可以观察任何长度的句子在本模型中的翻译结果，尤其是较长的句子，改善情况更加明显。在英语到法语翻译的任务中，提出的方法通过单一模型实现了与传统的基于短语的系统相当或接近的翻译性能。此外，定性分析显示，所提出的模型在源句与相应的目标句之间找到语言上合理的（软）对齐。

# 背景：神经翻译机

从概率的角度来看，翻译相当于找到一个目标句子y，它使给定句子x的条件概率最大化，即。在神经机器翻译中，我们拟合一个参数化的模型，以使用并行训练语料库来最大化一对句子的条件概率。一旦条件分布被翻译模型学习，给定一个源语句，可以通过搜索最大化条件概率的句子来生成相应的翻译。

最近，一些论文提出使用神经网络来直接学习这种条件概率分布（参见，例如，Kalchbrenner和Blunsom，2013; Cho等，2014a; Sutskever等，2014; Cho等，2014b ; Forcada和Neco，1997）。这种神经机器翻译方法通常由两个组件组成，第一个组件对源语句x进行编码，第二个组件将之解码成目标句子y。例如，（Cho等人，2014a）和（Sutskever等人，2014）使用两个循环神经网络（RNN）来将可变长的源语句编码为固定长度的向量并将此向量解码为一个可变长度的目标句子。

尽管这是一种相当新的方法，神经机器翻译已经显示出有期望的结果。 Sutskever等（2014）报道说，在英语到法语翻译任务上，基于具有长期短期记忆单元（LSTM）的RNN神经机器翻译具有接近于常规的基于短语的机器翻译系统的最先进的性能。为现有的翻译系统添加神经元素，例如，为短语表中的一对一对短语进行估分，或者重新排列候选翻译，已经允许神经机器翻译超过以前传统的极其翻译系统的最先进的表现水平。

## RNN ENCODER–DECODER

在这里，我们简要介绍由Cho和Sutskever等人提出的底层框架，称为RNN 编码器 – 解码器，我们构建了一种学习同步对齐和平移的新型架构