神经机器翻译

1. 概述

机器翻译研究如何利用计算机自动地实现不同语言之间的相互转换，是自然语言处理和人工智能重要研究领域，也是目前互联网常用服务之一。如Google翻译、百度翻译、微软Bing翻译等，都提供了多种语言之间的在线翻译服务．虽然机器翻译译文质量与专业译员相比仍有较大差距，但是在一些对译文质量要求不太高的场景下，或者是在特定领域翻译任务上，机器翻译在翻译速度上具有明显优势，仍然得到广泛应用．鉴于机器翻译的复杂性和应用前景，学术界和产业界都把该领域作为重点研究方向，成为当前自然语言处理最活跃的研究领域之一。

1957年，Rosenblatt提出了感知机算法，这是一种最简单的神经网络。早期的感知机，因其结构简单，不能处理线性不可分问题，造成了该研究长期的低潮期。20世纪80年代以后，反向传播算法被引入到多层感知机(MLP)，也叫前馈神经网络(FFN)。此后，在Hinton、LeCun、Bengio等人推动下，神经网络重新引起人们关注。2006年，Hinton等人通过逐层预训练方法解决了神经网络训练难题，随后由于计算能力提高，如并行计算、图形处理器的广泛应用，神经网络在学术界和产业界都得到高度重视．近年来，神经网络在图像识别、语音识别等领域取得巨大成功，同时学者们也将该技术应用在自然语言处理任务上，如语言模型、词语表示、序列标注等任务，并取得了令人鼓舞的成绩。

机器翻译相关研究，在多种语言对上，神经机器翻译已经逐渐超过短语统计机器翻译．JunczysDowmunt等人采用联合国语料库，在30个语言对上对神经机器翻译和短语统计机器翻译进行对比，神经机器翻译在27个语言对上超过了短语统计机器翻译方法。与汉语相关的，如中英、中俄、中法之间翻译任务上，神经机器翻译高出6～9个BLEU值。另外，在2016年机器翻译研讨会(WMT)上，爱丁堡大学开发的神经机器翻译系统在英语到德语翻译任务上，超过基于短语、基于句法的统计机器翻译。在大规模计算能力支持下，百度公司采用深层次神经网络架构，在WMT2014英语到法语翻译任务上，首次超过统计机器翻译方法，取得了最好的成绩。在产业界，Google翻译在部分语言上已采用神经机器翻译代替统计机器翻译对外提供服务。著名的商用机器翻译公司Systran同样开发出相应的神经机器翻译系统，涵盖了12种语言32个语言对。在国内，搜狗公司、小牛翻译也在积极开发神经机器翻译系统。目前，神经机器翻译不仅在学术界得到广泛关注，产业界也积极地探索该方法的商用价值。

本文第2节概述在神经机器翻译中常用的神经网络及其特点；第3节展望神经机器翻译未来研究方向。

1. 神经网络在机器翻译中的应用

神经网络依据拓扑结构特点可以分成多种类型，如前馈神经网络、卷积神经网络、循环神经网络等．本文只介绍一些在机器翻译、句法分析、序列标注等自然语言处理任务上常用的神经网络，并对其在机器翻译上的应用作简要概述。

2.1 循环神经网络

循环神经网络主要用于处理序列数据，特别是对变长序列数据有着较好的处理能力，神经机器翻译多数采用循环神经网络实现，如图1所示。

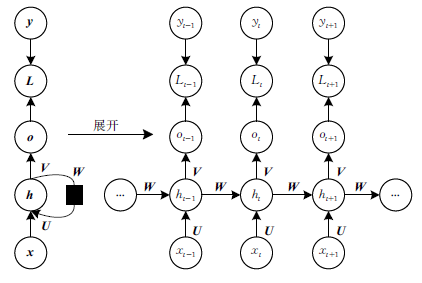


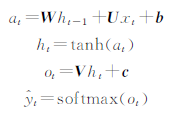
图1 循环神经网络

x表示变长序列数据，在每个时间点t上，隐藏状态ht由以下公式进行更新：

ht = f(ht-1, xt)

f是非线性函数。

通过循环神经网络将输入x映射到输出o。y是模型所要达到的目标序列（通常由训练语料给出），L是损失函数，U为输入到隐藏层的权重矩阵，W为隐藏层到隐藏层的权重矩阵，V是隐藏层到输出的权重矩阵，时间序列t范围为［１，T］，整个网络通过如下进行更新：



循环神经网络使得不同长度的输入序列，其输入向量维数都相同，并且在每个时间点上可以采用相同的变换函数和参数，更适合处理变长序列数据。另外，循环结构在理论上能够捕捉到所有前驱状态，这在一定程度上解决了长距离依赖问题。

2.2 循环神经网络的变形结构

将循环神经网络展开后可以采用反向传播算法训练，称为时间反向传播，在实际应用中会产生梯度消失问题。长短时记忆神经网络是循环神经网络的变形结构，采用了更加高效的遗忘和更新机制，具有与循环神经网络相似的结构和优点，且性能更好。

门限循环单元将长短时记忆循环单元的输入门和遗忘门合并成更新门，又引入了重置门，用更新门控制当前状态需要遗忘的历史信息和接受的新信息，用重置门控制候选状态中有多少信息是从历史信息中得到．该结构是对长短时记忆神经网络的简化，效果与后者相近，并降低了计算量。

递归神经网络是循环神经网络的变形结构，以树形结构进行组织，用于结构化预测和表示，适合表示自然语言句法结构。

2.3 带记忆的神经网络

神经网络没有外部记忆，对变量和数据长时间存储能力很弱，与外部信息交互很困难。Graves等人将循环神经网络与外部记忆耦合，称为神经图灵机。这种模型类似图灵机，并具有神经网络的优势，能够采用梯度下降法训练。除此之外，Weston等人提出了记忆网络，包含一个长时记忆组件，能够读取和写入，在具体任务中可以作为知识库使用。

这些带外部记忆的神经网络能够方便地利用外部资源，增加了神经网络与外部资源交互能力，同时也提高了可解释性和记忆能力。

2.4 Transformer模型

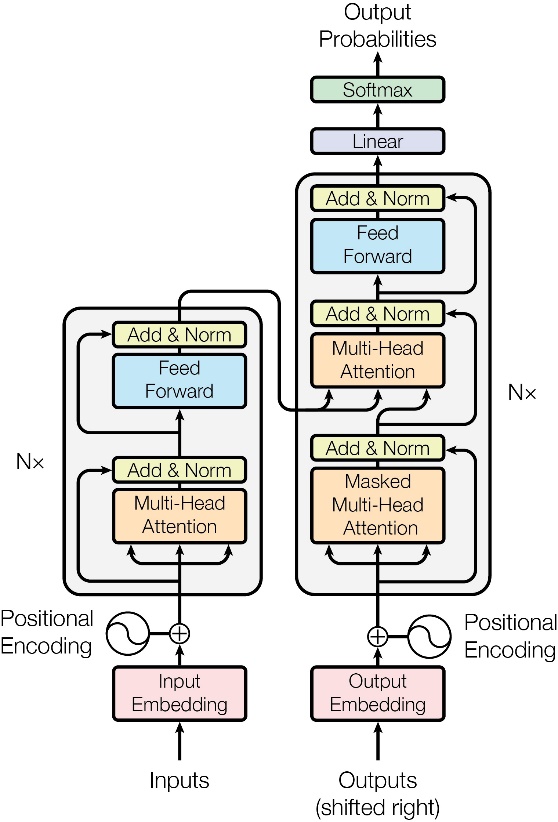


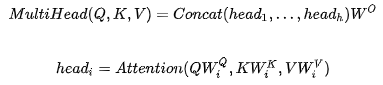
图2 Transformer模型结构

和大多数seq2seq模型一样，Transformer的结构也是由Encoder和Decoder组成。

Encoder由6个相同的层组成，层指的就是图2左侧的单元。每个层由两个子层组成，分别是多头自注意力机制和全连接网络。其中每个子层都加了残差连接和归一化，因此可以将子层的输出表示为：



其中，多头自注意力机制是通过h个不同的线性变换对Q，K，V进行投影，最后将不同的attention结果拼接起来。



自注意力机制就是Q，K，V取相同的值。

Decoder和Encoder的结构差不多，但是多了一个注意力子层。它的输是：Encoder的输出和对应i-1位置Decoder的输出。所以中间的注意力子层不是自注意力，它的K和V来自Encoder，Q来自上一位置Decoder的输出。

1. 未来研究方向

目前，神经机器翻译取得巨大成功，新的研究成果不断涌现出来，可以称作统计机器翻译之后一种全新的机器翻译方法。严格来讲，从2014年开始，神经机器翻译得到人们的广泛关注，随后大量相关成果发表出来。由于研究时间较短，该翻译模型仍然存在许多值得更加深入探索的问题，以下几点有可能成为未来研究集中方向：

(1)提高语言学解释性。基于编码器解码器的神经机器翻译，实现了源语言到目标语言的直接翻译，但是翻译过程很难得到充分的语言学解释。已有工作证明，可以从词语级神经机器翻译编码器中抽取出隐含的句法结构信息，以及在一定程度上对神经机器翻译的翻译过程进行解释和分析。从神经机器翻译模型中抽取出相应的语言学知识来解释翻译过程，以此改进翻译模型，是神经机器翻译未来重要的研究方向。

(2)融合外部先验知识。以离散符号表示的外部资源，如句法标注、词性标注、双语词典等是非常重要的先验知识，在神经机器翻译中难以得到充分利用．融合更加丰富的先验知识是神经机器翻译重要研究内容，也是提高翻译效果的重要方法，有待深入研究。

(3)基于句法的神经机器翻译。神经机器翻译大都是词语级的序列到序列模型，所包含的句法信息较少。句法是重要的关于句子结构的理论，将序列到序列翻译模型扩展至基于句法的翻译模型，如树到序列、序列到树、树到树等，是神经机器翻译模型架构创新的重要体现。

(4)多语言机器翻译。连续空间表示法是有效的多语语义表示方法，注意力机制经实验证明能够在不同语言之间共享，这些为多语言机器翻译研究提供了良好的基础。在多语平行语料，或者多语可比语料基础上研究基于神经网络的多语言机器翻译，不仅具有学术价值同样具有很高的实用价值，是未来重要的发展方向。

(5)多模态翻译。神经网络能够以统一的形式对文字、图像、语音等不同模态数据进行表示。目前，文字与图像之间实现端到端的直接翻译，并且图像信息也被应用到神经机器翻译。高效利用文字本身以外的信息，如语音、图像、位置场景等，以此构建多模态翻译是机器翻译真正实用化的必经之路。

1. 课程总结

本次语言分析和机器翻译课程，我有很多的收获。不但学习到了很多自然语言处理的知识，还听了朱老师的讲座报告，了解了很多课程之外的信息，这对我而言十分有用。在课程实践部分，跟着学长敲代码，使用实验室开发的NiuTensor也让我有很大的收获。总结一下，我觉得本堂课的课程设置很好，作业量不大，而且课程内容完善，谢谢肖老师的教导和诸位学长的指导。