

NMT发展脉络 发展综述

笔记本： 自然语言处理

创建时间： 2019/7/2 22:16

更新时间： 2019/7/5 9:51

作者： beyourselfwb@163.com

参考：2017年 神经机器翻译综述 作者：李亚超，熊德意，张民 刊物：计算机学报

<http://cjc.ict.ac.cn/online/bfpub/lyc-20171229152034.pdf>

刚开始大家都在研究什么，后来研究的人多了，大家又是从哪些角度发散研究的，目前还有哪些没有被研究的方向可以试试的？

经典神经机器翻译模型及其问题与挑战

目前NMT翻译模型主要分为两大类：

- 1、Google的端到端模型？
- 2、蒙特利尔大学的Encoder-Decoder模型？

两者本质上有啥区别？？？

统计机器翻译和神经机器翻译的核心思想是一致的：
把机器翻译看作是求解概率问题！

不同之处：

1、词对齐建模：SMT的关键部分，而NMT中是用attention动态地获得对齐信息的。

2、翻译效果：NMT的译文流利度比SMT好，但其漏翻、重复翻译（过翻译）问题比SMT严重。

表 1 NMT 与 SMT 差异

评价指标	NMT	SMT
表示方法	连续	离散
模型	非线性	对数线性
模型大小	小	大
训练时间	长	短
模型可解释性	弱	强
内存占用	小	大
GPU	必须	非必须
增量式训练	支持	不支持

NMT存在的问题和挑战：

（1）受限于词典的大小和句子长度。词典大小在训练好的模型上就是固定的了，后期fine tuning 或者增量训练

都不能更改词表大小，造成OOV问题，这是很多NLP的共性问题。

句子长度，这个问题现在不是用Attention改善了很多了么？况且如果句子实在很复杂，很长，别说机器了，人

翻译起来都很困难。

（2）难以高效利用外部先验知识：目前通用的NMT只需要双语平行语料对，没有充分利用上额外的先验知识，

如：大规模单语语料，标注语料，双语词典等。不过目前已经有这方面的研究出现了，比如多语言、低资源翻译[1-3]，

还有多任务学习的翻译方式[4]，都可以利用上大规模单语数据。

(3) Attention机制不够完善：没考虑到历史注意力机制？？？还有在生成一些虚词，如"the"的时候，不用过多关注源语言，关注一下目标语言更好。过翻译和漏翻译这两个大问题也和Attention有关。

(4) 神经网络架构单一：这个？？？

参考文献：

[1] SACHAN D S, NEUBIG G. Parameter Sharing Methods for Multilingual Self-Attentional Translation Models[J]. arXiv:1809.00252 [cs], 2018.

[2] TAN X, REN Y, HE D, 等. Multilingual Neural Machine Translation with Knowledge Distillation[J]. arXiv:1902.10461 [cs], 2

[3] WANG Y, ZHANG J, ZHAI F, 等. Three Strategies to Improve One-to-Many Multilingual

[4] DONG D, WU H, HE W, 等. Multi-Task Learning for Multiple Language Translation[C]//ACL 2015: 1723–1732.