

学界 | 机器翻译新突破：谷歌实现完全基于attention的翻译架构

机器之心 2017-06-14

选自arXiv

机器之心编译

参与：吴攀、黄小天、李亚洲

近两年来，机器翻译的突破让人目不暇接，从去年谷歌的颠覆性突破开始到一个月前 Facebook 的全新 CNN 翻译技术，人工智能的巴别塔正越建越高。近日，谷歌再次宣布又在机器翻译上更进了一步，实现了完全基于 attention 的 Transformer 机器翻译网络架构，并且还在 WMT 2014 的多种语言对的翻译任务上超越了之前 Facebook 的成绩，实现了新的最佳水平。机器之心对这篇研究论文进行了摘要介绍。

- 重磅 | 谷歌翻译整合神经网络：机器翻译实现颠覆性突破（附论文）
- 重磅 | Facebook 提出全新 CNN 机器翻译：准确度超越谷歌而且还快九倍（已开源）

论文：Attention Is All You Need

论文链接：<https://arxiv.org/abs/1706.03762>

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez*[†]
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaizer@google.com

Illia Polosukhin*
illia.polosukhin@gmail.com

在编码器-解码器配置中，显性序列显性转导模型（dominant sequence transduction model）基于复杂的 RNN 或 CNN。表现最佳的模型也需通过注意力机制（attention mechanism）连接编码器和解码器。我们提出了一种新型的简单网络架构——Transformer，它完全基于注意力机制，彻底放弃了循环和卷积。两项机器翻译任务的实验表明，这些模型的翻译质量更优，同时更并行，所需训练时间也大大减少。我们的模型在 WMT 2014 英语转德语的翻译任务中取得了 BLEU 得分 28.4 的成绩，领先当前现有的最佳结果（包括集成模型）超过 2 个 BLEU 分值。WMT 2014 英语转法语翻译任务中，在 8 块 GPU 上训练了 3.5 天之后，我们的模型获得了新的单模型顶级 BLEU 得分 41.0，只是目前文献中最佳模型训练成本的一小部分。我们表明 Transformer 在其他任务上也泛化很好，把它成功应用到了有大量训练数据和有限训练数据的英语组别分析上。



图 1：转换器-模型架构

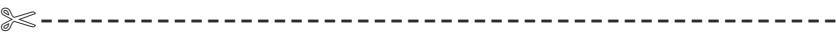


图 2：（左）可延展的 Dot-Product 注意，（右）Multi-Head 注意保护数并行运行的注意层



图 3：在 5、6 层编码器自注意中遵循长期依存关系的注意机制的例子。

本文为机器之心编译，转载请联系本公众号获得授权。



加入机器之心（全职记者/实习生）：hr@jiqizhixin.com

投稿或寻求报道：editor@jiqizhixin.com

广告&商务合作：bd@jiqizhixin.com

点击阅读原文，查看机器之心官网↓↓↓

阅读原文