

韦阳

✉ godweiyang@gmail.com · 📞 15221856016 · 🌐 <https://godweiyang.com>

教育背景

华东师范大学

2018 – 至今

硕士 计算机科学与技术 均分：91.5/100

华东师范大学

2014 – 2018

本科 计算机科学与技术 排名：1/110

研究方向

主要研究机器翻译、模型压缩与加速、成分句法分析、依存句法分析等方向。

实习经历

字节跳动 AI Lab NLP 算法工程师

2020.05 – 至今

主要研究机器翻译模型的压缩与加速。利用参数共享、词向量分解、量化、蒸馏等模型压缩技术，减小模型的体积，最终部署到移动端上。最终实现模型体积压缩 20 倍，同时效果基本无损。

学术成果

A Span-based Linearization for Constituent Trees

ACL 2020

Yang Wei, Yuanbin Wu and Man Lan

我们提出了一种成分句法树的序列化表示方法。基于该表示方法，模型的解码复杂度从 $O(n^3)$ 降低到了 $O(n \log n)$ ，实际解码速度从 30 句/秒提高到了 150 句/秒，同时效果没有任何损失。

LightSeq: A High Performance Inference Library for Transformers NAACL 2021 Industry Track

Xiaohui Wang, Ying Xiong, Yang Wei, Mingxuan Wang, Lei Li

我们提出了 LightSeq，一个高效的 Transformer 系列模型推理库。LightSeq 包括一系列 GPU 优化技术，以简化 Transformer 层的计算并减少显存占用。LightSeq 支持使用 PyTorch 和 Tensorflow 训练的模型。标准机器翻译基准上的实验结果表明，与 TensorFlow 相比，LightSeq 实现了高达 14 倍的加速；与 FasterTransformer 相比，实现了 1.4 倍的加速。

获奖情况

上海市优秀毕业生	2021
国家奖学金（硕士）	2020
金牌 ACM-ICPC 全国邀请赛（西安站）	2017
特等奖学金	2017
银牌 ACM-ICPC 亚洲区域赛（青岛站）	2016
一等奖学金	2016
铜牌 ACM-ICPC 亚洲区域赛（上海站）	2015
国家奖学金	2015

编程技能

- 编程语言：熟悉 Python、C++、C、CUDA，了解 Java、SQL 等。
- 深度学习框架：TensorFlow、PyTorch、DyNet

社交链接

- 技术博客：<https://godweiyang.com>
- GitHub：<https://github.com/godweiyang>
- 知乎：<https://www.zhihu.com/people/godweiyang>，15000+ 赞同，9200+ 粉丝。
- 公众号：算法码上来，4600+ 粉丝。