韦阳

godweiyang@gmail.com · 15221856016 · https://godweiyang.com

教育经历

华东师范大学 2018.9 – 2021.6

硕士　计算机科学与技术　1/105

华东师范大学 2014.9 – 2018.6

本科　计算机科学与技术　1/110

技能

- 研究方向：机器翻译、模型优化、成分句法分析。
- 编程语言：Python、C++、C、CUDA。
- 深度学习框架：PyTorch、TensorFlow。

工作经历

字节跳动　AI Lab NLP 算法工程师 2021.6 – 至今

- **LightSeq 训练推理引擎**
  LightSeq 核心开发者之一，这是业界首个集训练、推理于一体的 Transformer 系列模型加速引擎，GitHub 获得 2100+ star。主要应用于火山翻译的模型训练与部署中，并在公司内外业务中获得广泛应用。
  项目地址：`https://github.com/bytedance/lightseq`
  - *LightSeq* 推理引擎：采用算子融合、显存复用、层级解码等技术，将推理速度提高最多 14 倍。支持 Transformer、BERT、GPT 和 ViT 等模型结构，支持主流代码库的模型导出与推理。
  - *LightSeq* 训练引擎：采用算子融合、显存复用等技术，将训练速度提高最多 3.5 倍。提供了丰富的 C++ 和 Python 接口，支持 Transformer、BERT、GPT 和 ViT 等模型结构，支持 Pytorch 和 TensorFlow 接入，深度融合 Fairseq 和 Hugging Face 等主流代码库。
  - *LightSeq* 量化推理引擎：通过 int8 矩阵乘法与算子，将推理速度在 FP16 基础上进一步提升了最高 1.6 倍，显存降低 40%。支持主流代码库的量化感知训练与模型导出，支持 Transformer、BERT 和 GPT 等模型结构的量化推理，性能基本无损。
- **文言文翻译**
  利用开源文言文-现代文平行语料，训练了文言文-现代文双向翻译模型。并利用线上中-英模型生成出文言文-英文伪平行语料，训练出文言文-英文双向翻译模型。利用英文作为桥接，实现了文言文和近百种语言的互译。最终结合 LightSeq 量化推理引擎，上线火山翻译官网。
  官网地址：`https://translate.volcengine.com`
- **Transformer 移动端部署**
  将 Transformer 模型导出为 Torch Script，并进一步导出为 ONNX 模型，利用 ByteNN 部署在移动端上，实现离线翻译。

字节跳动　AI Lab NLP 算法工程师（实习） 2020.5 – 2021.6

- **Transformer 压缩与量化**
  研究 Transformer 压缩与量化方法，利用层间参数共享、词表分解等方法，结合模型量化，将 Transformer 模型参数量压缩至原来的 1/20，在机器翻译任务上效果几乎无损。
- **iOS 离线翻译**
  将 Transformer 模型导出为 TFLite，并利用 Swift 编写了一个 iOS APP，实现移动端离线翻译。

学术成果

LightSeq2: Accelerated Training for Transformer-based Models on GPUs

第二作者　SC 2022

提出了 LightSeq2 训练加速引擎，支持 Transformer、BERT、GPT 和 ViT 等模型结构，支持 PyTorch 和 TensorFlow，相比 PyTorch 最高提速 3.5 倍。

论文地址：https://arxiv.org/abs/2110.05722

LightSeq: A High Performance Inference Library for Transformers

第三作者　NAACL 2021 Industry Track

提出了 LightSeq 推理加速引擎，支持 Transformer、BERT、GPT 和 ViT 等模型结构，相比 TensorFlow 最高提速 14 倍。

论文地址：https://aclanthology.org/2021.naacl-industry.15

A Span-based Linearization for Constituent Trees

第一作者　ACL 2020

提出了一种成分句法树的序列化表示方法，将解码复杂度从 $O(n^3)$ 降低到了 $O(n \log n)$，解码速度从 30 句/秒提高到了 150 句/秒，并且效果无损。

论文地址：https://aclanthology.org/2020.acl-main.299

获奖荣誉

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

社交链接

- 技术博客：https://godweiyang.com
- GitHub：https://github.com/godweiyang
- 知乎（14000+ 关注）：https://www.zhihu.com/people/godweiyang
- 公众号（9000+ 关注）：算法码上来