Du Junyi

杜骏毅　　NLP 算法工程师

Email **junyiducn@gmail.com**
Phone **+86 180 1181 7053**
Github **cnrpman**
Website **https://junyidu.com**
Google Scholar **bit.ly/3E3NNBV**

SUMMARY

我有数年自然语言处理和知识抽取研究经验，目前头衔为 NLP 算法工程师，目标成为一名懂 NLP 的全栈。当下主要工作内容为 NLP 模型算法的开发部署和搭建模型测评自动化管线，以及结合 NLP 的语音识别研究。略懂前端，爱好探索未知。

WORK

NLP 算法工程师 / Apr 2021 - Present

深声科技 - 广州

TTS 前端

负责 TTS（语音合成）前端模型（字到音标，G2P）开发和部署对接，将原有的零散训练脚本整合成训练框架，加入了多任务训练和跨语言训练等最佳实践，整合了 Dynamic Batching、Sampling ratio scheduling、LR Scheduling、Gradient Accumulation 等优化和 Tricks，使错误率下降约一半，粤语 G2P 在人工测评中大幅领先竞品；目前发展方向为从 Common Crawl 清洗语料，在 BERT 等预训练模型上进行 Domain Specific 的二次预训练，以及开发基于模型可解释性的样本诊断工具。

ASR 预研

基于开源的 Wenet ASR（语音识别）框架搭建中文 ASR 训练框架，并开发自有的 ASR 测评管线；基于开源数据进行探索，对结构进行调整以及加入 MLM Rescoring 后模型指标对比原 Wenet 最好设计大幅提升，在推理算力预算内实现了字错误率下降约三成，部分修改成果已经合并到 Wenet 主干；目前正在进行 ASR 推理流程优化和纠错算法的研发工作。

Python　C++　Tensorflow　PyTorch　ONNX　TorchScript　NLP　ASR　BERT　Joint CTC

研究程序员 / Jan 2020 - Mar 2021

美国南加州大学 - 洛杉矶

在任翔教授资助下完成了 GAILA 项目，构建了一个通过专家采访来补全任务流程图的对话系统，包括 Slot filling 设计，意图模型的实现，以及开发了一个对话前端；完成了同实验室一些 NLP 顶会论文的实验环节。

JavaScript　react-md　Python　PyTorch　NLP　多模态学习　Rasa　Amazon Lex

研究助理 / Mar 2018 - Jan 2020

美国南加州大学 - 洛杉矶

加入了任翔教授的实验室并负责 GAILA 项目的开发，发展出一个能从对话文本中抽取结构化流程知识的框架；提出了一个关系分类模型，相关成果发表在了 NLP 顶会 ACL19 (oral presentation)

Python　PyTorch　NLP　BERT　OpenIE　BiLSTM-CRF

实习研究员 / Aug 2017 - Dec 2017

商汤科技 - 深圳

计算机视觉方向，进行半监督学习相关研究，探索从肮脏的网络爬取多模态数据 (如: Webvision) 中训练出干净模型。

Python　Tensorflow　CV　Curriculum Learning

SKILLS

熟练的语言: Python

会写的语言: C++, Bash

写过的语言: JavaScript (<=ES6), Swift, Java (1.8)

想学的语言: Lisp, CUDA C++

精通的算法领域: NLP

熟悉的算法领域: ASR, 知识抽取, 多模态

用过的前端技术: React, Angular 1

偏好的软件: VS Code Remote Dev, iTerm2, Firefox, Git CLI, Notion, Excalidraw

ACTIVITIES / INTERESTS

爬山, 剧本杀

Selected Publications

- Eliciting Knowledge from Experts: Automatic Transcript Parsing for Cognitive Task Analysis
- **Junyi Du**, He Jiang, Jiaming Shen and Xiang Ren
- **ACL19** (oral presentation)

- Towards Hierarchical Importance Attribution: Explaining Compositional Semantics for Neural Sequence Models
- Xisen Jin, **Junyi Du**, Zhongyu Wei, Xiangyang Xue, Xiang Ren
- **ICLR 2020** (spotlight paper).

- Visually Grounded Continual Learning of Compositional Semantics
- Xisen Jin, **Junyi Du**, Xiang Ren
- **EMNLP 2020**

EDUCATION

美国南加州大学 / Dec 2017 - Dec 2019

计算机科学硕士 GPA: 3.93/4.0

通过的课程 (部分): 算法分析, 机器学习, 自然语言处理 (博士级), 游戏制作, 网页技术

华南农业大学 / Sep 2012 - Jun 2016

计算机科学技术本科 GPA: 93/100

拿过一些奖学金; 打过一点 ACM(省赛银)