

陈英发

188 0117 9013 | chenyingfa1999@qq.com

github.com/donny-chan

24岁 | 男

在校生 | 自然语言处理 | 暑期实习



教育经历

| | |
|---------------------------|---------------------|
| 清华大学 硕士 | 2022年08月 - 2024年07月 |
| 计算机科学与技术 自然语言处理与社会人文计算实验室 | 北京 |
| GPA : 4.0/4.0 | |
| 清华大学 本科 | 2018年08月 - 2022年07月 |
| 计算机科学与技术 计算机科学与技术系 | 北京 |
| GPA : 3.4/4.0 | |
| 二学位 : 数字媒体设计 | |

项目经历

| | |
|--|---------------------|
| Sub-Character Tokenization for Chinese Pretrained Language Models (TACL 2023) | 2021年03月 - 2022年10月 |
| Yingfa Chen, Chenglei Si, Zhengyan Zhang | |
| <ul style="list-style-type: none">将汉字根据字形或者发音转换为 sub-character 序列。此分词器可以获得更短的编码从而提高训练速度,以及对同音字导致的噪声更鲁棒,且它没有牺牲准确率。 | |
| READIN: A Chinese Multi-Task Benchmark with Realistic and Diverse Input Noises | 2022年01月 - 2022年11月 |
| Yingfa Chen, Chenglei Si, Zhengyan Zhang | |
| <ul style="list-style-type: none">构造了新的中文数据集,其含有中文输入(通过输入法或者语音输入)过程中常见的噪声。通过实验验证了主流模型在此数据集上的性能有明显下降,并提出了简单的改进方案。 | |
| BMCook: A Task-agnostic Compression Toolkit for Big Models (EMNLP 2022 Demo) | 2021年12月 - 2022年06月 |
| 框架开发 | |
| <ul style="list-style-type: none">负责开发并测试知识蒸馏部分,仅需改动几行代码即可实现主流的知识蒸馏方法。已提交专利申请,审查中。 | |

实习经历

| | |
|--|---------------------|
| 美团 (北京三快科技有限公司) | 2021年07月 - 2021年09月 |
| 系统开发 自动驾驶模拟系统开发部门 | |
| <ul style="list-style-type: none">使用 perf 和火焰图来记录自动驾驶算法模拟系统的运行时间。通过优化,算法模拟过程所需时间减少了24%。 | |

荣誉奖项

| | |
|--------------------------|----------|
| 北欧数学比赛 NMC - 第 23 名 | 2018年03月 |
| 挪威数学奥赛 - 全国第 8 名 | 2018年03月 |
| 挪威信息学奥赛 - 全国第 10 名 | 2017年03月 |
| "搜狗杯" 第 24 届智能体大赛 - 16 强 | 2020年05月 |

技能/证书及其他

- 技能**: Python (熟练), C++ (熟练), PyTorch
- 语言**: 普通话 (流利), 英语 (托福111/120), 粤语 (母语), 挪威语 (近乎母语)
- 其他成绩**: SAT (1470/1600), SAT Math II (1600/1600), SAT Physics (1600/1600)
- 兴趣爱好**: 羽毛球 (系队)