

纪焄 (Tao Ji)

华东师范大学
计算机科学与软件工程系
理科大楼 B906 实验室,
上海市, 中山北路 3663 号

手机: (+86)15316561072
邮箱: taoji.cs@gmail.com
Github: <https://github.com/JT-Ushio>
主页: <http://www.taoji.site>

个人技能

研究方向: 自然语言处理和深度学习, 目前专注于使用深度学习技术改进依存句法分析. 我还对迁移学习 (跨语言) 和强化学习感兴趣.

课程学习: 平均分 89.6, 排名前 10%. 其中《文本挖掘》98, 《高级机器学习》97 两门课程最突出.

编程技能: 从高中 (NOIP) 到本科 (ACM) 一直参加编程竞赛, 精通大部分算法和数据结构, 擅长 C++, Python, 了解 JAVA.

教育经历

2017 - 现在: 硕士, 华东师范大学

2013 - 2017: 本科, 华东师范大学

发表论文

AntNLP at CoNLL 2018 Shared Task: A Graph-based Parser for Universal Dependency Parsing
Tao Ji, Yufang Liu, Yijun Wang, Yuanbin Wu, Man Lan
In CoNLL 2018 shared task.

ECNU at EPE 2017: Universal Dependencies Representations Parser
Tao Ji, Yuekun Yao, Qi Zheng, Yuanbin Wu, Man Lan
In IWPT 2017.

A Fast and Lightweight System for Multilingual Dependency Parsing
Tao Ji, Yuanbin Wu, Man Lan
In CoNLL 2017 shared task.

项目经历

Antu (<https://github.com/AntNLP/antu>)

Antu 是一个通用 NLP 工具库, 包含了独立于框架的数据处理模块和部分 Dynet、Pytorch 上常用的神经网络组件. 我是该工具库主要的设计者与开发者之一.

模块: NLP 任务上通用的数据 IO 模块

NLP 任务上通用的辅助模块 (Logger、Sampler 等)

神经网络编码器组件 (LSTM、Transformer 等)

神经网络解码器组件 (Tree-LSTM、Softmax 等)

神经网络常用结构组件 (MLP、Attention 等)

将来的工作: 持续优化模块的抽象形式, 提高工具库的丰富程度, 并且依据该库开发更多的 NLP 应用。

ECNU-Parser (<https://github.com/JT-Ushio/ECNU-Parser/tree/conll18>)

ECNU-Parser 是一个基于 Dynet 的多语依存句法分析工具库, 实现并改进了最先进的 Deep Biaffine 依存句法分析器。我们基于 UD 树库训练了支持 57 种语言的多语模型, 在资源丰富的语言上分别训练单语模型和去词化的多语模型, 在资源稀缺的语言上不训练模型而是使用算法寻找迁移效果最好的去词化多语模型。我是该工具库的主要开发者。

性能: 在 UD 树库的 82 个测试集上获得了 70.90 的平均 LAS、在 CoNLL2018 共享任务上排名 9/26(国内排名第 2)。

将来的工作: 继续优化定义语言间距离的算法, 提升跨语言的效率和性能并且提升工具库在多语上的通用性。

曾获奖项

国家奖学金, 2018

优秀毕业论文, 华东师范大学, 2017

全国一等奖, 第 8 届全国大学生计算机设计大赛, 2015

银牌, ACM-ICPC 亚洲东大陆子赛区决赛, 2015

银牌, ACM-ICPC 亚洲区域赛西安站, 2014