

李俊毅

性别：男

主页：<https://dukeenglish.github.io/>

电话：13603542466

方向：自然语言处理，视频生成处理

籍贯：山西

出生年月：1994 年 9 月 1 日

邮箱：41jy@163.com

学习经历与技能

硕士：爱丁堡大学(人工智能，导师：Adam Lopez)

(2016.09-2017.11)

本科：湖南大学(信息安全)

(2012.09-2016.06)

技能：Python, Pytorch, Tensorflow, NLP 算法与视频生成

工作经验

字节跳动商业化-中国区-智能优化

(2020.08-至今)

负责广告视频自动化生成与投放。

电子科学研究院-自然语言处理

(2019.01-2020.07)

负责 NLP 平台与问答系统。对 NLP 相关工作做全面负责，并撰写相关需求文档。协助培养一名研究生（知识图谱方向，主要辅助信息抽取）。

美团点评-美团平台-智能技术中心-自然语言处理

(2018.01-2019.01)

负责 NLP 平台设计工作，并承担 python 框架的主要研发。曾负责推荐内容的 E&E，评论分析，标签体系建立及 MMU 相关研发，并实现视频过滤与检索功能。

美团点评 2018 届算法通道新员工评定 TOP2（不区分一二名）。

部分项目经验

➤ 美团 NLP 平台建设【项目负责人，6 人】

- 负责美团用户平台下 NLP 平台规划设计，负责项目推进。在开发中负责 NLP 平台 python 框架设计实现、服务封装和部分算法设计实现。
- 平台支持酒旅评论分析、景点问题测评等 UGC 处理，跑腿标签生成，POI 相似度、营业预测以及证店一致性等相关业务，业务响应时间缩短 50%以上。
- 设计实现基于 BiLSTM+CRF 和 BERT 的词法分析服务，在内部测试集准确率和召回率以及 f1 值均由 94%左右提升至 96%左右。
- 设计实现 poi 营业状态预测算法，准召均在 90%以上，并就此申请专利。
- 针对美团跑腿业务基于历史搜索内容实现标签提取与推荐，在不同城市的 CTR 和提单率均有所提升，CTR 提升最大达 3.41%，提单率提升最大达 1.76%。
- 设计实现基于 UGC 的标签关键词生成算法并设计标签体系。

➤ 基于 NLP 平台的问答系统【项目负责人，1 人】

- 负责设计研究院对话逻辑与开发规范，并基于自然语言处理平台实现问答系统。本系统作为重要演示成果，参与多项目竞标。

- ◆ 实现基于关键词的召回策略，召回率在 top20 中为 97%。
- ◆ 在 TPU 上实现基于 bert 的模型预训练，使用业务相关的语料与通用预料一同混合训练，并使用小的词表，以降低参数量（使用 bert-base 结构可以降低 9.8%，RoBERTa-tiny 结构可以降低 92.6%）。
- ◆ 实现基于 bert 的文本相似度计算，加载新的预训练模型可以对文本相似度的准确率由 78.5%提升至 88%左右，top5 的召回率由 83%提升到 93%左右。
- 其他项目
 - ◆ **机器翻译**：英语到因纽特语的翻译系统。针对因纽特语由一个根词和多个前后缀组成以及未知词的问题，采用 subword 处理源数据，使用 CNN 解决过翻译和欠翻译的问题（N 对 N 的翻译问题）。得到将近 2.0 的 BLEU 效果提升。
 - ◆ **多模态理解**：设计实现基于关键帧的重复视频内容过滤功能，准确率达 99%。
 - ◆ **E&E**：利用多模态特征实现 item 质量预估，冷启动内容 CTR 整体提升 10%。

社会实践与学术活动

- 论文
 - ◆ CLUE: A Chinese Language Understanding Evaluation Benchmark (COLING2020)
 - ◆ Light Pre-Trained Chinese Language Model for NLP Tasks (NLPC2020)
- CLUE 中文语言理解测评基准成员
 - ◆ NLPC2020 会议 Shared Task 轻量级模型竞赛负责人，负责比赛评估标准制定，比赛推进，结果测评并负责最终总结论文报告，前三名分别为华为诺亚，腾讯 TEG 与小米 AILab。
 - ◆ 参与项目 CLUE、CLUEPretrainedModels 以及 CLUEDatasetSearch 的创建和维护，社区核心维护者之一。
 - ◆ AACL2020 审稿人，Track: Resource and Evaluation。
- 其他工作
 - ◆ 在美团申请有两项算法相关专利（配送优化和 poi 相关算法）。
 - ◆ Coursera 课程 translator&reviewer，翻译出版《Python 忍者秘籍》。。
 - ◆ 曾在 GPU 上通过 CUDAC++的编程实现 GPU 上的语言模型，实现对机器翻译模型的加速。

其他

- 知道自己在每个时间点想要什么。
- 兴趣特长：喜欢火影忍者，漫威系列。爱好篮球、吉他，主持经验丰富。
- 曾担任机器之心海外兼职分析师（2017.01-2017.09，全英，[外网链接](#)），在助理来也实习（2017.10-2017.12）。
- 曾获评湖南大学优秀毕业生，优秀学生干部，大学辩论赛亚军。曾任学生会主席，辩论队领队，湖南大学图书馆学生馆长。