

# 李俊毅

性别：男

主页：<https://dukeenglish.github.io/>

电话：13603542466

方向：自然语言处理

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(Artificial Intelligence) (2016.09-2017.11)

本科：湖南大学(信息安全) (2012.09-2016.06)

技能：Python, Pytorch, Tensorflow, 机器翻译, 问答系统, Bert 及模型压缩。

## 工作经验

电子科学研究院-自然语言处理 (2019.01-至今)

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，主要辅导信息抽取）。

美团点评-美团平台-智能技术中心-自然语言处理 (2018.01-2019.01)

负责公司级别 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）。

助理来也-实习 (2017.10-2017.12)

负责 python 框架升级与维护，W2V，文本相似度计算、爬虫以及数据清洗的工作。

机器之心海外兼职分析师 (2017.01-2017.09)

承担英国 AI 生态调研工作，负责论文，视频及会议分析解读（全英，[外网链接](#)）。

## 部分项目经验

### ➤ NLP 平台建设【项目负责人】

◆ 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，服务框架的设计实现，以及平台内功能的实现。

◆ 平台功能：

✧ NLP 基础功能：分词、词性标注、命名实体识别、句法分析、关系分类

✧ NLP 功能模块：情感分析、评论分类、简繁转换、拼音输入法

✧ NLP 逻辑模块：通用数据预处理、关键词抽取、短文本相似度分析

✧ NLP 深度学习相关模块：词向量计算，文本生成、机器翻译、Bert-NER

◆ 接入相应业务线，并开发业务解决流程 pipeline【累计接入七条业务线】，业务响应时间缩短 50%以上。

◆ 典型业务处理为美团点评酒旅评论分析、景点问题测评等 UGC 处理，美团跑腿标签生成，美团点评内部 POI 处理等相关业务。

◆ 在研究院接入搜索，知识图谱，智能平台等相关产品。

## ➤ 基于 NLP 平台的问答系统【项目负责人】

- ◆ 负责对问答开源平台的调研，并设计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%， robert-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%。
- ◆ 实现基于 BM25 的文档召回与基于 bert 的 QA 系统。

## ➤ 基于 NLP 平台的文本资料库的建设【项目负责人，管理前后端各一名】

- ◆ 实现对 doc, docx, pdf, excle, txt 等多种类型的文档读取与处理，生成关键词与摘要以辅助实现内部文档的自动化标签，形成索引关键字供搜索使用。
- ◆ 关键词抽取部分，采用 TextRank 比使用 tf-idf 抽取的关键词在抽取样（100）本中召回率提高 8% 至 95%，辅以规则过滤后，准确率提高 3% 至 86%。由于辅助人工标记，故召回率更为重要。
- ◆ 根据实际文档需求，采用截断式前 400 字辅以规则过滤作为摘要效果良好。
- ◆ 管理收获：对不同的人以及在不同的进度时要采取不同的管理和跟进措施。

## ➤ 美团发现页面推荐系统 E&E 探索试验【项目负责人】

- ◆ 基于 ASR/CV/NLP 以及 item 基本信息提供组合特征，整体 pvctr 提升接近 10%。
- ◆ 实现对重复视频内容的过滤与基于图片的视频检索功能。

---

## 社会实践与学术活动

- 入职美团当月即主导申请一项专利，随后参与申请专利一项。
- 翻译陈丹琦论文《Neural Reading Comprehension and Beyond》，翻译一本 python 相关的书籍（受疫情影响待出版）
- Coursera 课程翻译及审阅：Neural Networks and Deep Learning。
- 在中科院/北京理工大学等按周组织/参与科研讨沙龙
- 一篇 ACL 论文在投，负责部分实验以及论文撰写与校对。
- CLUE 成员，参与 CLUEbenchmark 的制定，项目贡献以及编码规范，负责 CLUE、CLUEPretrainedModels 以及 CLUEDatasetSearch，clue 社区主要维护者之一。
- NLPCC-2020 会议 Shared Tasks 轻量级语言模型评测比赛负责人，全面负责任务组织，评估标准，结果评测等事项，并将在最后负责撰写评估总结论文。

---

## 其他信息

- 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。
- 湖南大学优秀毕业生，优秀学生干部，大学辩论赛亚军。在校期间担任学生会主席，辩论队领队，湖南大学图书馆学生馆长。
- 现担任所在党总支青年委员，所在工会青年委员。