

李俊毅

性别：男

主页：<https://dukeenglish.github.io/>

电话：13241800726

方向：自然语言处理，视频生成，广告投放与用户增长

邮箱：4lijv@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)

工作经验

字节跳动

(2020.08-至今)

负责直播高光，从0启动项目并上线；负责增长算法方向，从0搭建方向并获得收益。
内部技术论坛创作者/讲师。

电子科学研究院

(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。

美团点评

(2018.01-2019.01)

负责 NLP 平台设计开发工作。算法通道新员工评定 TOP2（不区分一二名）

部分项目经验

➤ 增长算法方向负责人（方向负责人，3人，统筹多团队协作）

- ◆ 以技术身份切入，贯穿产运研角色以推动获取整体项目收益。**AB 实验收益在 3%-30%不等**
- ◆ 自动化投放：针对以关键词作为定向条件的搜索广告，搭建关键词自动更新全链路，从召回到排序分模块构建优化
- ◆ 创意优化：协调多团队，探索实验新的创意样式，获取显著差异化优势
- ◆ SEO：针对搜索引擎进行网站优化获量，打通前后端产运研协作优化
- ◆ RTA 投放：挖掘人群包实现人群定向，建立用户价值预估模型，精细化增长

➤ 直播引流视频生成与投放（项目负责人，4人，统筹多团队协作）

- ◆ 负责直播视频生成的架构设计、模块分工以及整体统筹协调工作，包括：直播视频数据流截取，计划与创意创编模块、视频加工与剪辑生成模块、其他视频元素生成（文案、视频组件、模版等）
- ◆ 负责直播视频生成，实验获得 **20%** 的收益，当前已基于此上线产品
- ◆ 设计并沉淀直播视频生成方法论，将广告创意拆解为业务产运经验和创意各组件要素进行分析，并从时空两个维度对视频进行理解与生成
- ◆ 负责视频生成平台从 0-1 的设计编码，支持业务包括：广告诊断、直播广告视频生成、派生和自助产品
- ◆ 负责设计直播引流视频优选模块：基于视频理解和基于后验数据的优化模型

➤ 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，电子科学研究院）

- ◆ 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。

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 (NLPCC2020)
- ◆ Few-shot Learning for Chinese NLP tasks (NLPCC2021)

➤ **CLUE 中文语言理解测评基准执行委员会**

- ◆ **NLPCC2020 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
- ◆ 参与项目中文自然语言处理评测基础建立（**CLUE**），社区核心维护者
- ◆ ACL/EMNLP/NLPCC/AACL 等 NLP 相关会议审稿人

➤ **应用实践**

- ◆ 累计申请六项专利，五项第一作者
- ◆ 组建 7 人学习小组，设计开发开源工具，定期进行学习交流
- ◆ 组建 11 人团队，参与公司内比赛获奖，成员属性横跨产运研设计与战略
- ◆ Coursera 课程 translator&reviewer，翻译出版《Python 忍者秘籍》

多点分享

- 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。