

# 李俊毅

性别：男

籍贯：山西

政治面貌：中共党员

出生年月：1994 年 9 月 1 日

电话：13603542466

邮箱：41jy@163.com

方向：机器学习，机器翻译，自然语言处理

---

## 学习经历与技能：

**硕士：**爱丁堡大学(Artificial Intelligence) 【2016.9-2017.11】

**本科：**湖南大学(信息安全) 【2012.9-2016.6】

**技能：**Python, CUDA 并行计算, Hadoop 编程, Tensorflow, Chainer

**自然语言处理：**文本分类, 机器翻译, 文本相似度计算

**机器学习算法：**主流机器学习算法如逻辑回归, 线性回归, 神经网络 (CNN, RNN) 及其优化 (数据增强, 正则化, 网络结构优化等)。概率图模型的基本知识及应用 (如贝叶斯网络和马尔科夫随机场)。

---

## 工作经验：

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

主要负责 NLP 平台设计开发, 承担 python 框架的主要开发工作。曾负责推荐内容的 E&E 问题, 以及 MMU 相关研发 (视频与音频处理)

**助理来也-实习 (2017.10-2017.12)**

主要负责 W2V, 文本相似度计算以及爬虫数据清洗

**机器之心海外兼职分析师 (2017.01-2017.08)**

机器之心兼职技术分析师: 负责机器学习前沿论文 review, 视频以及会议分析解读 (全英工作并公开在外网), 熟悉协同工作工具: slack, quip 等 (例如: <https://syncedreview.com/2017/02/24/david-silver-google-deepmind-deep-reinforcement-learning/>) 在英承担苏格兰地区人工智能生态调研工作。

---

## 项目经验：

- NLP 平台建设: (2018.10-至今)

- 负责 NLP 平台规划设计, 并承担 Python 语言框架下的技术模块开发
- 将 NLP 平台接入相应业务线, 并开发业务解决流程 pipeline
- 负责实现 NLP 平台中分词和机器翻译模块

- GPU 上的语言模型设计与实现[github] (2017.01-2017.08)

- 基于 gLM (第一个 GPU 语言模型) 实现一个服务于 MODLMs (Neubig & Dyer 一个语言模型框架, 旨在提升语言模型的性能) 语言模型 giLM。

- 该模型基于 gLM 使用 C++ 实现了前向 trie 数据结构，输入数据分析以及 GPU 上的并行 query 算法。
  - 编码环境：Ubuntu 16.04, NVIDIA CUDA8.0, C++, Tesla K80
  - 相比于 gLM, giLM 更加适应 MODLMs 的需求。在数据集相同的条件下，每秒的返回值 giLM 可以达到 278.96（大数据集）-1539.26（小数据集）倍于 gLM。
  - 机器翻译系统（因纽特语到英语）（2017.03-2017.05）
    - 基于 encoder-decoder 模型与 attention 机制的 RNN 神经网络实现了机器翻译系统
    - 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
    - 在普通编码器之前添加了 CNN 解决过翻译和欠翻译的问题（一对多，多对一，多对多的翻译）
    - 编码环境：Chainer 框架，python 语言
    - Subword（BPE）可以得到将近 2.0 的 BLEU 效果提升，0.1 的 CHEF 提升；添加 CNN 之后 CHEF 有 0.002 的提升，但是 BLEU 略有下降。
  - 基于卷积神经网络的图像识别（2017.02-2017.05）
    - 在不借助框架的情况下完成基础性神经网络编写与应用（MNIST, CIFAR）
    - 基于 Tensorflow 框架下探究网络框架优化、结果优化等可能性。
  - 拥有 Hadoop 编程经验，熟悉基础 Map, Reduce 编程
    - 例如：布隆过滤器，蓄水池算法等
- 

## 社会实践

- 在美团主导申请一项专利，参与申请一项
  - Coursera 课程翻译及审阅：Neural Networks and Deep Learning
- 

## 其他信息

- 本科在校期间曾担任学生会主席，辩论队领队，湖南大学图书馆学生馆长，并曾获湖南大学辩论赛亚军，湖南大学优秀学生干部，湖南大学优秀毕业生等奖励