

# 戴剑波

13.4/1 Meadow Court 13-29 Sciennes ◊ Edinburgh EH9 1NJ ◊ (+44) 07913109707 ◊ (+86) 13486963560

[jianbodai819@gmail.com](mailto:jianbodai819@gmail.com) ◊ [s1792542@sms.ed.ac.uk](mailto:s1792542@sms.ed.ac.uk) ◊ <https://www.linkedin.com/in/jianbo-dai-9085a2161/>

## 教育经历

### 爱丁堡大学, 英国

2019.09 - 至今 (预计毕业日期 2020.11)

人工智能硕士, 预计 GPA: 3.85/4.00 (74.0/100), 排名约前 10%

相关课程: 机器学习与模式识别, 加速自然语言处理, 深度学习, 概率建模与推断, 自然语言理解、生成与机器翻译, 博弈论算法及应用, 编程技巧。

### 曼彻斯特大学, 英国

2017.09 - 2019.06

化学工程学士, GPA: 3.82/4.00 (75.9/100), 排名约前 7%

相关课程: 微分方程数值方法, 离散数学, 过程控制, 过程优化。

### 大连理工大学, 中国

2015.09 - 2019.06

化学工程学士, GPA: 3.83/4.00 (88.3/100), 排名约前 3%

相关课程: 高等数学, C 语言程序设计, 概率与统计, 线性代数, 大学计算机基础, 电工技术, 电子技术。

获奖情况: 2017 年中国大学生 Chem-E-Car 竞赛全国第 8; 2016 年全国大学生数学竞赛市一等奖; 2017 年美国数学建模竞赛三等奖, 2019 年大连理工大学优秀毕业生。

## 个人技能

**编程语言** 熟悉 Python, C, LaTeX, 有 HTML, SQL, C++, Qt 编程经验

**框架工具** 熟悉 Pytorch, Numpy, Matplotlib, SciPy, Pandas

**自然语言** 中文 (母语), 英语 (无障碍商务沟通)

## 项目经历

### 语义分析: 用户问题分类与建议

暑期研究, 耶鲁大学语言、信息与学习实验室

2020.06 - 现在

- 基于 Spider, SparC 和 CoSQL 等关于自然语言与 SQL 的语义分析的数据集格式化输入数据, 收集正样本与负采样负样本并建立新的 SQL 模版与自然语言-SQL 模版库。
- 在格式化的数据上微调与重新训练一个叫 RatSQL 的以关系敏感的自注意力机制增强的 seq2seq 模型, 计算了召回率。
- 根据三种对“部分问题”的不同定义建立三个收集了部分问题与 SQL 模版相对应的数据集。

### 语境化的神经词形还原

硕士毕业设计, 爱丁堡大学

2020.05 - 现在

- 训练了一个叫 Lematus 的神经编码解码模型作为基准模型, 训练结果以原词准确率与莱文斯坦距离为指标进行了评估。
- 在 20 多种不同语言上训练了一个叫 COPYNET 的用复制机制增强的 seq2seq 模型, 使用了 OpenNMT 这一开源框架。
- 用编辑距离算法为每一对曲折词-原词对创立中间模版, 将其用作一个多层模型的辅助输入, 该模型有一个对结构敏感的将语义分析解码过程分解为两段的神经网络结构。

### 神经机器翻译

自然语言理解, 生成与机器翻译课程项目, 爱丁堡大学

2020.02 - 2020.03

- 训练基于带双向 LSTM 的 encoder-decoder 架构的神经网络机器翻译模型, 实现束搜索并增加网络层数来增强模型并比较在不同数据集上的损失, 困惑度及 BLEU 分数。
- 基于论文实现 lexical attention, 输出注意力的热图来研究其对翻译效果的影响。
- 分析 Transformer 体系的实现方法, 并在其基础上实现多注意力机制。

### 使用推特数据探索分布性假设及单词的相似度

自然语言处理课程项目, 爱丁堡大学

2019.11

- 实现多种矢量语义模型来表示单词, 包括基本的 PPMI 及带有 Laplace 平滑和上下文分布平滑的变体。
- 实现不同的单词相似度测量方法如余弦相似度与 Jaccard 相似度并研究其对模型的影响。
- 使用推特数据计算国家名称的相关性并发现当地地理位置更近时, 除非新闻引起国家之间的重大联系, 否则相关性更高。

### 基于字符的三元语言模型

自然语言处理课程项目, 爱丁堡大学

2019.09 - 2019.10

- 通过最大似然估计预测三元字符组的概率, 并使用多种平滑方法 (如加  $\alpha$  平滑) 修正概率, 从而建立三元语言模型。
- 使用建立的模型随机生成多种语言的句子并计算生成文档的困惑度, 以进行语言识别任务。
- 收集不同体裁及大小的数据集并研究其对优化后的模型效果的影响。