



## 刘前

共产党员 (政治面貌)

17888833508 (手机)

www.github.com/charlesliucn (GitHub)

22 岁 (年龄)

汉族 (民族)

charlesliutop@163.com (邮箱)

北京市海淀区清华大学 (地址)

## 教育经历

- 清华大学** 工学学士  
信息科学技术学院 · 电子工程 2014.08 至今
  - 基础课程: 微积分 (94), 线性代数 (92), 复变函数与数理方程 (100)
  - 专业课程: 网络技术基础 (98), 数据结构与算法, 操作系统 (85), 统计信号处理 (87), 随机过程 (91), 媒体与认知 (90), 数字图像处理 (86), 数字逻辑与处理器基础 (86)
  - 实验课程: MATLAB 工程应用 (97), 电子电路课程设计 (93), 数字逻辑与处理器基础实验 (89)
  - 兴趣课程: 数据库 (96), 信息科学技术概论 (95), 文献检索与利用 (93), 计算机文化基础 (93), 高级科技英语 (90), 科学技术史系列讲座 (87)
  - 网络课程: Coursera - CS229(机器学习), Stanford - CS224n(深度学习与自然语言处理)
- 清华大学** 辅修学位  
统计学研究中心 · 统计学 2016.08 至今
  - 专业课程: 数据科学导论 (91), 金融统计 (91), 初等概率论 (88), 线性回归分析, 统计推断, 统计计算, 数据挖掘, 多元统计分析
- 清华大学** 硕士 (保研)  
信息科学技术学院 · 电子工程 2018.08 起

## 获奖情况

- 国家励志奖学金** 校级 2016.10
- 清华大学新生二等奖学金** 校级 2014.09
- 清华大学学业进步优秀奖学金** 校级 2015.10
- 清华大学公益活动优秀奖学金** 校级 2015.10
- 清华大学三星级紫荆志愿者** 校级 2015.12
- 全国部分地区大学生物理竞赛** A 类二等奖 2015.12
- 全国高中数学联赛省级一等奖** 省级 2013.10

## 个人技能

- 操作系统: Windows, Linux
- 编程语言: R, Python, Shell, MATLAB, C/C++, Java, Verilog, SAS,  $\text{\LaTeX}$
- 工具包: Kaldi, Gensim, Tensorflow
- 英语能力: 全国大学英语: 四级 (604), 六级 (556), 清华大学英语水平: 96.75/120

## 学校/社会工作

- 清华大学水木信箱志愿项目 笔友志愿组 2014.11 – 2016.08
- 清华大学电子系无 47 班 党课小组组长 2015.09 – 2016.08
- 清华大学电子系学生会 学习部副部长 2015.09 – 2017.03

## 实习经历

- 清华大学天津电子信息研究院 智能家居研发平台 2017.06 至今
  - 智能家居行业调研分析: 开展实习项目前, 对智能家居行业技术和发展进行调研分析。
  - 蓝牙技术原理及射频测量方案: 学习蓝牙无线通信协议重点内容, 针对协议调研并设计测量方案, 对产品或仪器的性能进行测量与评价。
  - 远程实验开关控制系统: 利用电子测量仪器平台, 完成系统设计、电路搭建、软件控制、界面优化等流程, 实现远程控制的实验系统。

## 项目经历

- RNN/LSTM 语言模型优化 *Python* 2018.03 至今
  - 环境配置: Ubuntu, Python, Tensorflow
  - 项目简介: 基于现有的循环神经网络 (RNN/LSTM 等) 语言模型, 使用不同的 word2vec 方法, 改进词表示 (Representation) 的效果, 降低语言模型的困惑度 (Perplexity); 并且, 尝试将 Attention 机制引入神经网络语言模型。
  - 个人工作: 复现 Skip-Gram、CBOW、Clove 等 word2vec 算法, 将训练得到的词向量作为 RNN 语言模型的输入, 比较三种方法对语言模型性能的影响。同时, 将 NLP 中广泛使用的 Attention 机制引入语言模型。
- 低资源语音识别系统语言模型优化 *Kaldi* 2017.11 – 2018.03
  - 环境配置: Ubuntu, Kaldi, Shell
  - 项目简介: 基于美国国家标准与技术研究院 (NIST)2016 年的 OpenKWS 评测, 使用低资源语种 (格鲁吉亚语) 进行低资源语音识别系统的优化。使用发音字典生成、词表优化、数据筛选、模型插值融合等方法对语言模型的性能进行改进。
  - 个人工作: 对格鲁吉亚语网络数据进行清洗、预处理, 统计词频对词表进行扩充, 减少开发集上的集外词比例; 使用 Morfessor 工具生成现有词表的发音字典; 按照一定原则, 从网络数据中筛选出与训练数据类似的文本作为扩充数据, 筛选方法包括: 困惑度 (Perplexity) 排序、交叉熵距离排序、TF-IDF 文本相似度比较、LSI 模型分析、基于 doc2vec 的文本相似度比较等。经过以上优化, 最终在声学模型不变的基础上, 低资源语音识别系统的词错误率 (WER) 有 3.9% 的提升。

- **基于上证 50 指数成分股的金融数据分析** *R* 2017.09 – 2017.10
  - 环境配置: Windows, R
  - 项目简介: 使用金融统计课程所学的统计方法, 对上证 50 指数 10 余年的股价尝试基本的数据分析, 从中得出一些具有参考价值的结论。
  - 个人工作: 基于上证 50 指数的数据, 分别使用线性时间序列、资产波动率模型、因子模型进行初步分析, 并且进行了投资组合分析与优化, 用资本资产定价模型 (CAPM) 刻画了金融市场中的均衡问题。
- **基于深度学习的移动大数据挖掘及用户行为建模** *Python* 2017.03 – 2017.07
  - 环境配置: Ubuntu, Python, Tensorflow
  - 项目简介: 运用传统马尔科夫链方法和深度学习方法 (RNN 等), 对从时间和空间两个维度对用户的移动数据进行建模分析及预测, 从而根据用户现有的轨迹数据对未来某一时刻的位置进行预测。
  - 个人工作: 设计实现基于马尔科夫的轨迹预测算法, 并基于 Tensorflow 框架尝试对论文的深度学习方法进行复现与评估。
- **中文连续语音识别** *Python* 2017.03 – 2017.05
  - 环境配置: Ubuntu, Kaldi, PyQt
  - 项目简介: 基于 Ubuntu 平台, 安装编译 Kaldi 环境, 分别基于 timit(英文) 和 thchs30(中文) 数据集训练得到语音识别模型, 并且基于训练得到的模型实现实时语音识别。
  - 个人工作: 基于 thchs30 中文数据集, 使用 Kaldi 工具包训练得到语音识别系统; 并且使用 PyQt 设计了图形用户界面, 实现用户点击按键后实时进行语音识别的交互式程序。
- **旧金山湾区公共自行车数据分析与可视化** *R* 2017.01 – 2017.02
  - 环境配置: Windows, R, R Shiny App
  - 项目简介: 使用旧金山湾区 2014 至 2016 年间公共自行车的使用数据, 进行可视化分析, 并使用基本统计方法发现内在规律。
  - 个人工作: 从空间和时间两个角度, 使用基础的机器学习方法 (线性回归、LASOO 回归等) 分析天气等因素对公共自行车使用情况的影响, 并基于分析结果, 对公共自行车的分配提出了可行的建议, 实现了一款 Shiny App, 向公共自行车使用者提供了路线查询等功能。
- **MIPS 处理器设计与实现** *Verilog* 2016.06 – 2016.07
  - 环境配置: Windows, Vivado, Modelsim, Basys 3.0 开发板
  - 项目简介: 组建 3 人小组, 使用 Verilog 编程实现 32 位 MIPS 处理器, 并实现定时器和 UART 外设的相关功能。
  - 个人工作: 基于单周期的模块化原理, 独立实现了各单元模块的设计和调试, 将各模块综合并加入外设功能, 基于 Basys 3.0 FPGA 开发板实现了单周期处理器, 提高了个人硬件设计调试以及团队合作的能力。
- **烟草图像杂质识别** *MATLAB* 2016.03 – 2016.04
  - 环境配置: Windows, MATLAB
  - 项目简介: 使用统计信号的处理方法, 对烟草图像中的杂质进行识别和剔除。
  - 个人工作: 基于烟草和杂质 RGB 和 HSV 的分布进行高斯拟合, 利用训练数据得到分离效果较好的阈值, 在测试数据上, 杂质的召回率达到 90% 以上。