



李显明

生日：1997/08/28

现居：广东省/深圳市

籍贯：广西/北海市

学校：山西大学/本科/计算机科学与技术

成绩：前20%

毕业时间：2019

意向工作地点：深广上北杭成均可

热爱计算机，热爱新技术，热爱coding，希望通过不断地学习成为一名合格的全栈工程师

(+86) 18601697828

xmlee97@gmail.com

技能

博客

GitHub：<https://github.com/SeanLee97>

技术博客：<https://seanlee97.github.io/archives/>

技术栈

操作系统：

- Linux：日常使用的系统，掌握常用命令如：top、netstat、ps、ipcs、kill等，熟悉Linux下的编程环境
- Windows

编程语言：

- 熟悉Python，包括web相关、数据科学相关；了解Python调用C++扩展；熟悉Flask框架，了解多进程/多线程/网络编程（select, poll, epoll）等库，Python web后台开发常用技术栈：flask + gunicorn + celery + supervisor + nginx
- 掌握C++（基本语法，以及C++11常用特性）、java（基本语法）、javascript、PHP

数据结构与算法：掌握基础数据结构与算法

大数据：了解Hadoop、Hive

web开发：熟悉Python Flask框架以及相关技术栈

数据库：熟悉MySQL/MariaDB等关系型数据库、掌握MongoDB

版本管理：熟悉Git工具，GitHub爱好者

NLP及机器学习：本科阶段一直跟导师学习分词、文本分类、机器阅读理解等技术，偏向工程化，主要协助团队实现模型以及上线服务

大学英语：CET4

实习经历

Webot(智言科技)/python实习生

2018年7月 - 至今，深圳市，研究部

负责知识图谱相关服务的构建，相关模型的编写和上线。

主要运用Tensorflow来实现模型以及使用Python来构建API接口，主要技术栈
docker + python + flask + gunicorn + supervisor + celery + nginx

WeLab(我来贷)/算法实习生

2017年7月 - 2017年10月，深圳市，智能金融客服小组

参与了客服数据的清洗、以及文本分类任务

1. 数据清洗

主要对客服数据进行归类，通过一些常用脚本代码获取小批量数据，然后训练了一个分类器再跑全部数据，相当于用机器学习的方法辅助我完成繁琐的归类任务，最后进行更细致的清洗

2. 文本分类

分别对比实验了几个文本分类模型 svm、fasttext、cnn、rnn 的效果

蒙城汇新媒体/开发工程师

2016年7月 - 2016年9月，太原市

微信软文广告统计系统

负责了整个系统的构建（包括前后端，数据库，服务器），技术栈是：ThinkPHP + Bootstrap + jQuery/html/css，服务器环境：LNMP。

系统实现了微信文章的爬取，客户管理，定期发送报表功能。支持导入导出客户信息Excel表，使用Crontab实现定期发送报表

微信投票活动插件改进

改进了基于discuz论坛系统的投票插件禾金投票，增加了IP限投、防刷票、投票记录的加权统计等功能

1. IP限投：每个IP每天只能投一次票，主要通过获取用户真实IP实现
2. 防刷票：通过增设题目回答来限制机器刷票
3. 投票记录加权：由于所在公司主要服务加拿大华人，所以加拿大地区的投票权重较高

项目

科研训练

2017年3月 - 2018年4月，太原市

小语智能聊天机器人

山西大学本科一等科研项目，基于深度学习技术构建的智能聊天机器人，在项目中主要负责：

1. 开发NLP工具包、研究和改进文本分类模型

基于Python开发的NLP工具包xmnlp，以TextCNN为baseline延伸了一系列改进模型，包括：highway+multi-conv, 空洞卷积残差网络, 自注意卷积残差网络

2. 构建了对话模板检索系统

使用Python实现了一个简单的信息检索系统，主要用于对话的检索。基本方案是：倒排表+bm25，倒排表先对用户查询进行文档的快速检索，bm25计算空间向量相似度找到最匹配的文档

3. 百科知识图谱的构建

使用Python scrapy爬虫框架爬取百度百科，将数据存入MongoDB，为了加快检索速度使用了ElasticSearch作为检索系统，d3.js做可视化

4. 微信接口以及小组官网的开发

采用了Python Flask + Bootstrap V4 及其相关技术栈构建

开源项目

1. [xmnlp](#)：轻量级中文自然语言处理工具, 提供中文分词, 词性标注, 文本纠错, 文本摘要, 文本转拼音, 情感分析等功能 (star: 80+, fork: 20+)
2. [nnclf](#)：文本分类器，PyTorch实现，集成了CNN，RNN，CNN+RNN等分类算法
3. [artf](#)：一个轻量级tensorflow库，主要封装了Highway，Multihead-Attention，以及一些低版本的Tensorflow中不支持的函数
4. [LL1-Parser](#)：用C++实现的LL(1)文法编译器，支持文法的自定义

在校情况

社团经历

1. 2015-2016 山西大学校报记者团/视觉部干事/图片的后期处理
2. 2016-2017 CCF-SXU学生分会/副主席/负责中国计算机学会的宣传，认证考试的动员工作

荣誉及奖项

1. [MRC2018](#)(Cips-Baidu DuReader机器阅读理解大赛) Top 10% (11/105)
2. SMP2017中文人机对话测评任务一第八名
3. 2017 山西省"互联网+"创新创业大赛三等奖
4. CNCC2016 优秀志愿者 (CNCC: 中国计算机大会)
5. 2016 山西大学三好学生

已修课程

1. 编程语言：C、C++、JAVA、汇编语言程序设计

2. 算法基础：数据结构
3. 系统：操作系统原理、编译原理、数据库、计算机体系、计算机组成原理、计算机网络
4. 其他：信息检索，数据挖掘导论