

Xxxx Xxxx

+(86) 132XXXXX975 ◊ xxxxxxxx@gmail.com
北京市海淀区颐和园路 5 号北京大学, 邮编 100871

教育

- 北京大学** 2016.9 - 至今
硕士研究生, 网络所, GPA:xxxx(100)
- 北京大学** 2012.9 - 2016.6
本科生, 计算机系
总 GPA: xxx/4.00 专业课 GPA: xxxx/4.00
- 爱丁堡大学** 2015.7 - 2015.8
信息学院交流

荣誉 / 奖项

- NOI 一金牌** 2011.7
- 北京大学游戏 AI 编程对抗赛第一名** 2015.6
- SIGMOD2017 编程竞赛第二名** 2017.3
- 五四青年奖学金** 2014

工作经历

- 谷歌中国** 2018.6 - 2018.9
实习生, *Web Development Service*
对网站内容进行分类, 判断是否适用于 *markup* 标签, 并给出 *markup* 推荐配置
- 编写分布式数据处理 Pipeline
 - 用 TensorFlow 训练模型, 并根据结果对模型进行调整和改进
- 微软互联网工程院** 2017.9 - 2018.2
实习生, 小冰组
主要负责尝试各种 *NLP* 特征, 以及实现一些工程性的 *pipeline*, 编程语言主要是 *C#*.
- 在 *C#* 中手写 CNN 模型预测。一共写了两个版本, 内存优化版本比原有版本快 40 倍; 调用 Intel-MKL *C++* 库的版本比原有版本快 2 倍。
 - 在自动生成评论的模型中, 发现了一种效果很好的特征。
 - 实现了从数据库到 *ElasticSearch* 实时对接的 Pipeline, 具有良好的容错性和 Log 机制。
- 今日头条** 2017.6 - 2017.8
实习生, *AI-Lab* 组
主要负责维护服务器端的视频消重服务, 应用了 *Thrift-Framework* 框架, 编程语言 *C++*.
- 尝试了不同的视频消重特征, 主要有深度特征、orb 特征以及 k-means 聚类方法。
 - 编写了一个 Web 界面, 方便人工标数据。
 - 负责一些清洗数据的工作。
- 旷视科技 (Face++)** 2016.3 - 2016.9
实习生, 文本识别研究组
负责维护使用模型识别文字的 *Pipeline*, 以及编写 *Demo*. *Pipeline* 以及 *Demo* 的服务器引擎主要由 *C++* 编写。
- 用 *C++* 实现了 *Opencv* 中的 *minAreaRect* 函数, 主要算法是凸包以及旋转卡壳。由于考虑到代码要在移动端运行, 因此为了轻便, 不能直接集成 *Opencv*。
 - 用最小费用流实现了文字成行算法, 效果很好。
 - 负责调整 Pipeline, 为不同模型编写 Web 端的 *Demo*。

项目 / 研究经历

自动调参系统

2017.12 - 2018.1

用 Java 实现了一个基于贝叶斯优化的自动调参系统，总共约 2000 行。

Github: <https://github.com/xxxxxxxxx>

- 模型: Gaussian Process.
- Maximizer: Random Sampling.
- Acquisition Function: EI, LCB, GP-UCB

SIGMOD 2017 编程竞赛

2017.1 - 2017.3

版面排名: 第 1 名. 线下最终测试第 2 名

给定短语的字典, 和很多长文档, 然后要求动态地对字典进行插入、删除操作。目标是在每篇文档中, 查找出现过哪些字典中的短语。总数据量 10G。提交代码, 根据代码运行时间排名

- 使用 Hash 表和 Trie 建索引。(尝试过 AC-自动机, 效果不好)
- 实现了一套复杂的并行机制, 使得插入、删除、查询操作可以做到完全并行。
- 编程语言 C++, 总共约 3000 行。

阿里大数据竞赛

2016.10 - 2016.12

给定 2016.1-2016.6 的线下优惠券使用情况, 要求预测 2016.7 线下优惠券的使用情况, 初赛排名 (1/1500). 复赛排名 (19/1500)

三人组队, 负责模型和特征选择, 以及数据清洗。

- 复赛时学习使用阿里数加平台。主要工作分为两部分: 用 Sql 处理数据, 以及用 Java 编写 MapReduce 处理数据;
- 尝试过 LR、随机森林、Xgboost 等机器学习模型, 以及模型融合;
- 尝试加入时序特征, 使 AUC 提高了 1% 左右;

3D 联机对战手游

2016.7 - 2016.8

独立开发

一个 MOBA 类的安卓手游, 每局游戏支持 4 人在线对战, 提供了玩家积分机制, 以及根据战斗力自动匹配的机制。

- 游戏逻辑部分由 Unity3D 引擎开发, 编程语言是 C#;
- 服务器端选择了阿里云 Linux 服务器, 登录系统和积分系统由 Django 框架开发;
- 总代码量约 7000 行;

Java 代码和注释的自动匹配

2015.7 - 2015.8

暑期在爱丁堡大学交换了 2 个月。

在 Charles Sutton 的指导下, 使用深度学习匹配 Java 代码块和注释块。最终目标是 Java 代码注释自动补全。

- 数据集是从 Github 上选的 Java 工程;
- 编写 Java 语法分析器, 提取 Java 代码块和与之对应的注释块。
- 生成训练数据: 尝试各种规则, 生成匹配好的 < 代码, 注释 > 对。

技能

- 编程语言: 熟练使用 C/C++, C#, java, python 和 Go 语言, 了解 shell, matlab 和一些函数式语言, 比如 scala 和 scheme。
- 处理大数据: 熟悉 Hadoop、Spark, 以及 Parameter Server。
- 机器学习: 熟练使用 numpy, matplotlib, sklearn 和 xgboost, 了解 pandas 和 tensorflow。

其他

- 数学水平: 数学分析、高等代数和离散数学等均 A 或 A+
- 英语水平: CET-6.
- GitHub: <https://github.com/xxxxxx>