

# 邵国栋

个人主页: [ssd227.github.io](https://ssd227.github.io)

邮箱: [guodong\\_shao@foxmail.com](mailto:guodong_shao@foxmail.com)

电话: (+86)17600109680

## 简介

互联网算法工程师, 工作方向为推荐广告、拍卖机制策略。曾参与百度凤巢搜索广告-多目标拍卖框架的研发工作。研究方向为推荐系统、广告拍卖机制、自然语言处理(NLP)。技术兴趣广泛, 详见:[[主页链接](#)]。近期在研究生成模型、大模型(LLM)系统和应用相关的工作。

期望找一份算力充足、模型相关的工作。大模型、搜广推或其他有趣的项目。工作地点: 长三角地区优先考虑。

## 教育经历

- 中国人民大学-信息学院-数据仓库与商务智能实验室(DWBI) 海淀, 北京  
学术硕士 - 计算机软件与理论; GPA: 3.59 Sep. 2016 - June. 2019
- 合肥工业大学-数学学院-信息与计算科学 合肥, 安徽  
理学学士; GPA: 3.73 (排名: 1/72) Aug. 2012 - July. 2016

## 工作经历

- 百度-商务搜索部(凤巢)-商务机制策略组 海淀, 北京  
研发工程师 July 2019 - May 2021
  - auction微服务(拍卖框架)重构开发: 将ranker中的多物品拍卖框架剥离成微服务, 便于其他团队调用拍卖pipeline。具体拍卖策略采用Strategy Pattern, 业务代码与拍卖主流程分离, WISE和PC统一拍卖流程框架, 便于策略代码的维护和部署。历时一年打平推全
  - 日常业务策略开发: ranker(PC、WISE)多物品广告策略部分的日常开发上线工作
  - 数据流分析: 线上策略数据的监控、调研和分析
- 中国保险信息技术管理有限责任公司(CIITC), 实习 石景山, 北京  
NLP、IR算法工程师 Sep 2017 - Apr 2018
  - 目标文本抽取: 对全国车辆保险出险日志的分析挖掘, 使用HMM、CRF、RNN等seq2seq模型对出险位置文本的提取。基于spark对20年文本数据批量化处理
  - 文本匹配、检索系统: 利用车险文本抽取结果和全国地图POI数据库, 搭建车险查询系统。基于文本匹配技术(倒排、局部敏感hash、Lucene库等), 关联出险文本和地图信息, 提供高效的双向查询。
- Infosys, 实习 Bengaluru, India  
数据库开发工程师 June 2017 - Aug 2017
  - 分布式列存储DB开发: 基于scala-akka的通讯机制, 对底层文件做分发同步管理
- 国家互联网信息安全中心, 实习 海淀, 北京  
大数据、数据挖掘工程师 Sep 2016 - June 2017
  - 海量数据分析与处理: 基于spark/hadoop对日志数据进行模式发掘、预警分析

## 个人项目经历

- sgd RecSys: 推荐系统全流程框架算法实现、论文研读
- DeepLearning System: CMU深度学习系统needle实现 (天奇、Zico Kolter)
- Distribution Sysetem: 分布式系统实践 (cos418、mit6.824、golang patterns)
- sgd RL: 深度强化学习在棋类和Gym上的应用(PG、Q-learning、AC、AGZ、搜索树)
- sgd NLP: 自然语言处理算法(统计学习算法、深度学习模型) from Scratch
- sgd CV & Gen: 计算机视觉&生成模型from Scratch
- sgd GNN: 图编码、图卷积、cs224w、PyG
- From Nand to Tetris: P1: 逻辑电路到cpu实现 P2: 汇编器、Virtual Machine、编译器

- **sgd CG:** 图形学项目汇总
- **sgd Algorithm:** 算法库实现、leetcode

(Project链接)

## 编程能力

---

编程语言: Python, Go, C/C++, Cuda, Java, Scala, SQL, JS

技术栈:

- **熟悉:** 广告拍卖机制, 推荐系统, 自然语言处理, 数据挖掘, 深度学习系统实现
- **了解:** CV, 强化学习, 生成模型, LLM, 概率图模型, 图卷积, 分布式系统, 高性能计算, 数据库实现, 图形学, 编译原理, 大数据处理 (Hadoop, Spark)

(学习路线链接)