

邵国栋

个人主页: ssd227.github.io

邮箱: guodong_shao@foxmail.com

电话: (+86)17600109680

教育经历

- 中国人民大学-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 Aug. 2016 - Dec. 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)-机制策略组 海淀, 北京
研发工程师 July 2019 - June 2021
 - **Tensorflow:** TensorFlow is an open source software library for numerical computation using data flow graphs; primarily used for training deep learning models.
 - **Apache Beam:** Apache Beam is a unified model for defining both batch and streaming data-parallel processing pipelines, as well as a set of language-specific SDKs for constructing pipelines and runners.
- 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 July 2019 - June 2021
 - **Tensorflow:** TensorFlow is an open source software library for numerical computation using data flow graphs; primarily used for training deep learning models.
 - **Apache Beam:** Apache Beam is a unified model for defining both batch and streaming data-parallel processing pipelines, as well as a set of language-specific SDKs for constructing pipelines and runners.
- Infosys, 实习 Bengaluru, India
分布式数据库开发工程师 July 2019 - June 2021
 - **Tensorflow:** TensorFlow is an open source software library for numerical computation using data flow graphs; primarily used for training deep learning models.
 - **Apache Beam:** Apache Beam is a unified model for defining both batch and streaming data-parallel processing pipelines, as well as a set of language-specific SDKs for constructing pipelines and runners.
- 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 July 2019 - June 2021
 - **Tensorflow:** TensorFlow is an open source software library for numerical computation using data flow graphs; primarily used for training deep learning models.
 - **Apache Beam:** Apache Beam is a unified model for defining both batch and streaming data-parallel processing pipelines, as well as a set of language-specific SDKs for constructing pipelines and runners.
- Coursera Mountain View, CA
Senior Software Engineer Jan 2014 - Oct 2016
 - **Notifications:** Service for sending email, push and in-app notifications. Involved in features such as delivery time optimization, tracking, queuing and A/B testing. Built an internal app to run batch campaigns for marketing etc.
 - **Nostos:** Bulk data processing and injection service from Hadoop to Cassandra and provides a thin REST layer on top for serving offline computed data online.
 - **Workflows:** Dataduct an open source workflow framework to create and manage data pipelines leveraging reusable patterns to expedite developer productivity.
 - **Data Collection:** Designed the internal survey and crowd sourcing platform which allowed for creating various tasks for crowd sourcing or embedding surveys across the Coursera platform.
 - **Dev Environment:** Analytics environment based on docker and AWS, standardized the python and R dependencies. Wrote the core libraries that are shared by all data scientists.

- **Data Warehousing:** Setup, schema design and management of Amazon Redshift. Built an internal app for access to the data using a web interface. Dataduct integration for daily ETL injection into Redshift.
- **Recommendations:** Core service for all recommendation systems at Coursera, currently used on the homepage and throughout the content discovery process. Worked on both offline training and online serving.
- **Content Discovery:** Improved content discovery by building a new onboarding experience on coursera. Using this to personalize the search and browse experience. Also worked on ranking and indexing improvements.
- **Course Dashboards:** Instructor dashboards and learner surveying tools, which helped instructors run their class better by providing data on Assignments and Learner Activity.

• Lucena Research

Atlanta, GA

Data Scientist

Summer 2012 and 2013

- **Portfolio Management:** Created models for portfolio hedging, portfolio optimization and price forecasting. Also creating a strategy backtesting engine used for simulating and backtesting strategies.
- **QuantDesk:** Python backend for a web application used by hedge fund managers for portfolio management.

项目经历

-
- **QuantSoftware Toolkit:** Open source python library for financial data analysis and machine learning for finance.
 - **Github Visualization:** Data Visualization of Git Log data using D3 to analyze project trends over time.
 - **Recommendation System:** Music and Movie recommender systems using collaborative filtering on public datasets.

编程能力

-
- **编程语言:** Python, Go, C/C++, Java, Scala, Cuda, SQL, JS(按熟悉程度排序)

技术栈:

熟悉: 广告拍卖机制, 推荐算法, 自然语言处理, 数据挖掘, 深度学习系统实现, 分布式系统,

入门: 强化学习, 图卷积, 高性能计算, 数据库实现, 图形学, 编译原理, 大数据处理(Hadoop, Spark), 前端(React, nodejs), 爬虫, ,

更多详见: 《个人学习路线》