SaiNageswar Satchidanand

Senior Software Developer

A passionate developer with experience in Deep learning, Big data & Full-Stack. Well experienced in System Design, building ML-Pipelines, gRpc APIs, Devops (Infra-As-Code), Kubernetes, Queues/Topics, Sql/NoSql Databases, Operational Excellence etc. In Deep Learning worked on variety of problems like Causality Analysis, text, Images, Source Code etc. and engineered distributed system for Feature Engineering and Model Inference at scale.

sainageswar@alumni.iitm.ac.in

+917022374569

Hyderabad, India

linkedin.com/in/sainageswar

github.com/SaiNageswarS

WORK EXPERIENCE

Senior Software Engineer - ML
Microsoft

05/2022 - Present

Page Insights, Core AI & Search

Achievements/Tasks

- Developed an offline pipeline for generating LLM-based titles, captions, and web answers for Bing's top search queries. Engineered a vector index using HNSW to cache this content and serve it on the Search Results Page via nearest neighbor searches. Additionally, experimented with various embedding queries to optimize performance.
- Built online LLM captions generation without impacting overall page rendering time. Built a high scale system that triggers LLM preemptively before ranking and save inferences to a distributed cache. As ranking is finished, generated captions would be picked up from cache. Built comprehensive observability to record metrics for each step to optimize system end-to-end latency.
- Enhanced Bing Search Results by using LLMs by chaining prompts for intent generation, alternative queries, and re-ranking. Further, this utilized Chain of Thought Prompting, where the LLM was guided step-by-step to provide results with reasoning. Additionally, developed a streaming engine that sends LLM output tokens to the UI, masking the reasoning.

Senior Software Engineer
Microsoft

01/2021 - 05/2022

Substrate Platform, Office 365

Achievements/Tasks

- Worked on building Kubernetes for Kubernetes.
- As Kubernetes mgmt plane handles resources like pods, deployments, services in multiple nodes, our system handled multiple Kubernetes clusters across multi-location, multi-tenant and multiple types (GPU clusters, Memory Optimized Clusters etc.)
- Following principles of Kubernetes, worked on building distributed system which handled reconciling actual state of multiple clusters with desired state of multiple clusters. Further it handled resources like Keyvault, common containers in pods for logging, monitoring, Api-to-Api auth etc.
- Our system had spec definition framework similar to Kubernetes, however, typesafe and general programming language based instead of yaml/json.
- Worked on designing a composite Inventory to persist topology of system in mutliple storage implementations with single access pattern.

WORK EXPERIENCE

SDE-II
Amazon

06/2017 - 12/2020

Consumer Behavior Analytics

Achievements/Tasks

- Worked on applying Causality Analysis to measure effectiveness of Ads in purchase and High Value UserActions like Prime SignUp, Video/Music streaming etc.
- Causality Analysis involved identifying Control features (intrinsic User features) and Treatment features (Ads viewed/clicked), then determine incremental effect of treatment variables on outcome - purchase/user action.
- Trained DNN models to learn functions that map (Control & Treatment) variables to User Actions - Choosing optimizers, activation functions for multi-class problems etc.
- Further, worked on feature engineering tasks such as determining correlation between different user actions. For e.g. a Prime Video user is likely to signup for Prime Music. Based on that identified past user actions that can be considered as Control Variables.
- Worked on applying graph based representations to capture similarity between user actions based on Jaccard index of UserId-Action vectors. Further, applied spectral analysis on these graphs etc.
- Engineered pipelines to process terabytes of data in distributed Spark clusters. The pipelines apply feature transformations and perform model inference on large scale.
- Implemented regular re-training of model, regularization of weight changes between models to keep explanation of treatment variables consistent etc.
- Also, experiemented applying Sequential Models to capture effect of sequence of Ads/Campaigns.

SDE II
Codenation/Trilogy Innovations

05/2016 - 06/2017

Achievements/Tasks

- Worked on Mining source code Repos.
- Developed Embedding of AST in source code to find duplicate code or similar code.
- Worked with Clone Digger, Eclipse JDT to develop an Auto-Refactoring solution.
- Also, developed an org wide dashboard for executives to monitor developer productivity.
- Developer productivity was measured in terms of commits, fixes, designs submitted and code reviews. The dashboard further used statistical analysis to compute Normal Distribution and fit developers in different buckets.
- Used Docker based orchestration for deploying all services.

SDE
Gozoomo

07/2015 - 05/2016

Achievements/Tasks

- Developed a trusted peer-to-peer used car marketplace.
- Built scalable micro-services, auto-generating boilerplate typesafe client and service side code using protobufs.
- Built reliable Api-to-Api authentication and monitoring.
- Built operational excellence involving alarms, log analytics etc.
- Developed custom chat framework using Firebase and Angular.

SDE
Codenation/Trilogy Innovations

07/2014 - 06/2015

Achievements/Tasks

- Worked on Analytics Enterprise Application based on Microsoft SQL Server which performs different Data Warehousing tasks.
- Upgraded legacy system by upgrading frameworks, refactoring code, adding unit tests, mocking etc.
- Added CI/CD pipelines with integration tests.

EDUCATION

M.Tech in Computer Science Engineering
Indian Institute of Technology Madras (IIT Madras)

07/2012 - 07/2014

Thesis

- Researched on Semi-Supervised Multi-View Multi-Relation Collective Inference using Hypergraphs. The model can capture features to output variable relationship and relationship between instances of training data for Semi-Supervised inference. The work got published in IJCAI.

B.Tech in Computer Science Engineering
National Institute of Science and Technology (NIST)

07/2006 - 07/2010

Project

- Developed a Scientific Engineering Mathematics Tool.
- Parses free form mathematical equations and plots graphs between the variables.
- Built solver for first order and second order differential equations.

PUBLICATIONS

IJCAI'15

Extended Discriminative Random Walk: A Hypergraph Approach To Multi-View Multi-Relational Transductive Learning

Author(s)

Sai Nageswar Satchidanand, Harini Ananthapadmanaban, Balaraman Ravindran

2015

IJCAI 2015: 3791-3797

https://dl.acm.org/doi/abs/10.5555/2832747.2832778

OPEN SOURCE PROJECTS

go-api-boot (https://github.com/SaiNageswarS/go-api-boot)

- gRpc API Framework to build a Microservice.
- Supports grpc-web out of box, odm for mongo. Further, adds Auth JWT, logging middleware by default.
- Provides support for cloud (aws/azure) resources.
- Provides Zero Config SSL Support.

GeoSpatialAnalysis (https://github.com/SaiNageswarS/GeoSpatialAnalysis)

- Uses Temporal Workflows and gdal for analysis of GeoTif remote sensing images.

agent-boot (https://github.com/SaiNageswarS/agent-boot)

- A simple agent framework for intent classification and personal response generation using Chain-Of-Thought and RAG.

GraphMind (https://github.com/SaiNageswarS/GraphMind)

- A LLM-powered tool that constructs semantic graphs from Golang/gRPC microservices to automate multi-repo code changes. Built in solo hackathon - CRED RabbitHole. Rated in top 3 ideas out of 50 submissions of the hackathon.