Soumyajit De

Last Updated on 19th April 2025

career.soumyajit.de@gmail.com | 0091-70-222-51513 | GitHub: github.com/lambday | LinkedIn: linkedin.com/in/desoumyajit

Education

IIT BOMBAY
MTech in Computer Science (ML)
2014 | Mumbai, IN

WB UTECH
BTech in Computer Science
2011 | Kalyani, IN

Technical Experience

DOMAINS
CTR-prediction, Language Models

LANGUAGES
C++, Python, Java, SQL

LIBRARIES & TOOLS
PyTorch, ONNX, Huggingface, Keras, Pandas, SkLearn, Matplotlib, SciPy, NumPy, Jupyter, Docker, Kubernetes, Azure Data Factory, Azure Data Lake Storage, Azure Data Lake Analytics, Distributed FS (Cosmos), Map-Reduce, Kafka, BLAS, GDB, Valgrind, Perf, Git, RESTful APIs, OAuth, Conda, Pip, LaTeX.

Publications

[1] Danica J. Sutherland, Hsiao-Yu Tung, Heiko Strathmann, Soumyajit De, Aaditya Ramdas, Alexander J. Smola, and Arthur Gretton. Generative models and model criticism via optimized maximum mean discrepancy. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017. Link.

Honours

10th Board Exam: State Rank: 1st
Recipient, Chief Ministers Gold Medal.

Industry Experience

MICROSOFT | Senior Data & Applied Scientist, Search Advertising R&R
Dec 2018 – Present | Bangalore, IN

CTR PREDICTION
- Introduced ranking of all ad decorations in top 5 INTL markets using low-latency CTR-prediction models using statistical features (+1.5-2% ΔCTR).
- Unified feature stores, trained a single model to serve across multiple clusters, extending ranking support to 100+ INTL markets across the globe.
- Improved model utilising semantic signals from query, creatives and decorations.
- Addressed signal sparsity in low-volume markets by fine-tuning a pretrained encoder and adopted distillation with pseudo labels. Improved overall ΔAUC by 4%, helping with the cold-start problem. Met latency demands with lightweight student model and cached embeddings (+0.3-2.0% ΔCTR across markets).

PERSONALISATION
- Improved personalisation in decoration ranking by introducing dense-match features utilising user-profile signals (+0.2% ΔCTR on personalisable slice).
- Generated text features from 2 sources using in-context learning. Leveraged similar signals developed by another team. Examined user-interest clusters and explored approaches to capture recency and diversity in user embeddings.
- Customised and fine-tuned an encoder to output lower dimensional embeddings, meeting capacity budgets while maintaining quality (+0.21% ↓ +0.16% ΔAUC).
- Led addition of platform support and integration across the globe (AMER + INTL), collaborating with Engineering and partner teams. Driving adoption in other related areas including creative ranking, decoration relevance, decoration retrieval.

OFFLINE FILTRATION
- Addressed Key-Value store capacity limitation by introducing an estimated click-yield based filtration strategy for algo-generated decorations. (+0.1-0.3% ΔCTR).
- Utilised historical queries for the ad, used sampling to address the scale of scoring 10B (approx) items daily. Worked with partner team on training a global model.

AD DECORATION/CREATIVE GENERATION
- Addressed lower candidate density in smaller markets and overall quality with zero-shot decoration generation offline in INTL markets. Utilised an instruction tuned encoder decoder model. (+1.5-2.0x ΔCoverage, +0.22% ΔRevenue).
- Worked on improving attractiveness of creatives with query-pivoted asset generation online. Explored frozen M/LLMs as distillation targets for zero-shot generation. Proposed dataset curation strategies for supervised fine-tuning of an SLM, addressing different demands across categories, volume and language (Ongoing).

ORACLE | Senior Software Engineer, Cloud Infrastructure
Jul 2014 – Apr 2016, Sep 2016 – Dec 2018 | Bangalore, IN

- Designed and implemented a majority of the Marketplace REST API.
- Employed batch-processing and application-layer caching to reduce the response times of multi-page GET-calls from ∼2 mins to ∼10 secs.

Research Experience

UNIVERSITY COLLEGE LONDON | Research Assistant, Gatsby Unit
May 2016 – Jul 2016 | London, UK

- Devised a cache-friendly algorithm for a class of statistical tests involving MMD estimator that showed ∼300x speed-up over naïve implementation.
- Proposed and implemented a multi-threaded variant that outperformed competing algorithms, built with state-of-the-art solvers, by an order of magnitude [1].

Open Source Experience

SHOGUN ML LIBRARY | Core Contributor | 94,221 LOC changes
2013, 2014, 2016 | Google Summer of Code

2016 Co-mentored in designing Shogun's Linear Algebra library.
2014 Designed and developed a framework for kernel-based hypothesis tests. Added a family of feature selection algorithms on this framework.
2013 Implemented an estimator for log-det of large, sparse matrices arising in the log-likelihood computation of high-dimensional Gaussians in real-world datasets.