Sambuddha Roy

+1 (8xx) x0x-xx0x • /in/sambuddharoy
Bellevue 98004, WA USA • https://sambroy.github.io/
sambroy • Permanent Resident (Green Card holder), USA.

Summary and Goals

I am primarily interested in challenging roles in the space of data science & machine learning, with a leaning towards deep learning and natural language processing.
My primary focus is on deeper understanding of techniques/approaches, teasing out meta-patterns, while grounding such approaches by diving deep into data to curate actionable insights. I derive thrills from unique perspectives that a data driven approach can provide one, as also nifty mathematical arguments/theory underpinning such approaches. I maintain an active interest in mathematical and coding puzzles, as well as exploring the art of exposition; I firmly believe that the best way to understand a topic is to break it up into appropriate chunks and write about it. In this direction, I present some expository material on my blog.

Experience

Microsoft — Bellevue, WA, USA

Principal Data & Applied Scientist — *May '2018 – present*

I am a Principal Applied Scientist in Microsoft, in Bellevue, Washington. Among recent achievements, I shipped the *first generative* models for a high QPS Outlook offering under severe latency restrictions (of ∼90 ms). More recently, I have been working on productionizing super-huge language models.

Redmart — Singapore

Principal Data Scientist/Consultant — *2017 – 2018*

Redmart is the leading e-commerce company in Singapore. My tasks and responsibilities at Redmart included:

- *Recommendations:* I worked on improvements to the existing recommendation engines onsite, with an aim to increase add-to-cart actions.
- *Mentoring:* Hiring and mentoring data scientists to help build a data science team at Redmart.
- Surveying and demarcating data science opportunities pan company in order to improve end-to-end service to customer. The main verticals were forecasting, routing, search and recommendations, marketing.

LinkedIn — India

Staff Data Scientist — *Oct '15 – Dec '16*

At LinkedIn, my work primarily involved *unstructured data* such as text and images. Some highlights:

- *Near Duplicate Detection*: Detecting *near-similarity* between different documents, using different hashing techniques, for eg. Locality Sensitive Hashing (LSH). This work resulted in improvements of 60% in key business metrics.
- *Matching between shorter and longer texts*: This arose in the context of assigning *quality scores* to job postings on LinkedIn. We developed semisupervised techniques, based on word embeddings, achieving precision measures of $\geqslant 0.78$.
- *People Management*: I managed a team of 5 data scientists (including 2 interns) for part of my stay, and mentored the data scientists in the broader team.

Amazon — India

Sr. Machine Learning Scientist — *Mar '14 – Sept '15*

I worked in the Advertising Optimization/Advertising Technology group at Amazon, focusing on fraud detection in online advertising. My work consisted of the following broad themes:

- *Unsupervised Learning/clustering at scale:* The objective was to detect publisher fraud in an advertising network. We translated the problem to the world of graph clustering, thereby achieving significant improvements in detecting publisher fraud.
- *Bid optimization:* The task was to optimize acceptance of bids in the presence of seller-specified thresholds/reserve prices. We applied techniques from *online algorithms* to achieve $5 - 10\%$ improvements in the key metrics. This work also resulted in a publication in the *Amazon Machine Learning Conference*.
- *Pricing and Monetization:* This work involved designing combinatorial auctions, mechanisms in the context of mobile advertising.

IBM Research — India

Research Staff Member — *Aug '06 – Mar '14*

My work at IBM consisted of the following main highlights:

- *Online Optimization:* Worked on a few problems in the space of online convex optimization and applications of the same to the smart energy sector.
- *Densest Subgraphs:* This was work with the telecommunications group at IBM. The objective was to design *fast* algorithms for finding the densest subgraph in a graph subject to various natural constraints. This work resulted in improvement in business metrics as well as several publications.
- *Scheduling:* This was work involving the *staffing of contact centers*, and we explored various primal dual techniques to devise efficient and effective algorithms for such problems. This work resulted both in business value as well as publications in top tier conferences.

Delhi University — Delhi, India
Adjunct Faculty — *2011 – 2014*
Taught undergraduate courses on *Algorithms*, as well as advanced topics such as *Approximation Algorithms*.

Rutgers University — New Jersey, USA
Computer Science Instructor — *2003*
Taught an undergraduate course on *Discrete Mathematics and Probability*, as also served as a guest lecturer on a graduate course on *Complexity Theory*.

Education

Rutgers University — New Jersey, USA
Ph.D. in Computer Science — *1999 – 2006*

University of Chicago — Illinois, USA
started work towards Ph.D. in Computer Science — *1998 – 1999*

IIT Kanpur, India — Uttar Pradesh, India
Bachelor degree (B.Tech.) in Computer Science — *1994 – 1998*

Awards & Honors

Program Committee Member PC member for the conferences ACM CODS COMAD 2021, ACM CODS COMAD 2020, ACM CODS COMAD 2019, ACM CODS COMAD 2018, ACM Compute 2017, ACM Compute 2016.

Invited Speaker Gave invited talks at Veritas, Pune, India, 2016, CODS 2016, School of Approximability, Microsoft Research, 2011.

Outstanding Technical Achievement Award (OTAA) Awarded at IBM Research, 2010 for work on the project OptiManage, involving staffing of contact centers.

McCormick Fellowship, University of Chicago Awarded a McCormick Fellowship at the University of Chicago, for the year 1998–1999.

Two time Silver Medallist, International Mathematics Olympiad (IMO) Awarded silver medals in IMO 1993 (Turkey) and IMO 1994 (Hong Kong).

Top Ranks in Undergraduate Entrance Examinations Secured 4^{th} rank pan India in the prestigious IIT entrance examination, 1994. Secured 1^{st} rank in the state of West Bengal in the WBJEE exam, 1994.

Skills

Technical expertise: Machine Learning, Online Learning, Deep Learning, Natural Language Processing, Optimization, Primal Dual methods, Approximation Algorithms.

Languages/Frameworks: Python (Tensorflow, Keras, Pandas, Numpy/scipy, scikit-learn), Pig, LaTeX, R, PostGRES, (*limited working proficiency* in) Scala/Spark.

Representative Publications

Published papers in various top-tier conferences/journals, such as PODS, PODC, ICALP, IPCO, etc.

Decision trees for entity identification: approximation algorithms and hardness results. PODS, 2007.
Coauthors: Venkat Chakaravarthy, Vinayaka Pandit, Pranjal Awasthi, Mukesh K. Mohania

Density Functions subject to a Co-Matroid Constraint. FSTTCS, 2012.
Coauthors: Venkat Chakaravarthy, Natwar Modani, Sivaramakrishnan R. Natarajan, Yogish Sabharwal

Contact Center Scheduling with Strict Resource Requirements. IPCO, 2011.
Coauthors: Aman Dhesi, Pranav Gupta, Amit Kumar, Gyana R. Parija

Parity Problems in Planar Graphs. CCC, 2006.
Coauthors: Mark Braverman, Raghav Kulkarni.

Distributed Algorithms for Scheduling on Line and Tree Networks PODC, 2012.
Coauthors: Venkat Chakaravarthy, Yogish Sabharwal