Sudipto Mukherjee — Curriculum Vitae

Contact Details

Building 99, 14820 NE 36th St,
Redmond, WA 98052.

Email : sudipto.ece.ju@gmail.com
Github: sudiptodip15, Tel: (206)747-4808

Work Experience

- **Senior Applied Scientist, Microsoft.**
  Microsoft Customer & Partner Solutions (MCAPS) — Nov 2023 - Present
  Research (Project Science Engine Graduatation to MCAPS) — Dec 2021 - Oct 2023
  Microsoft Search, Assistance and Intelligence (MSAI). — Sept 2020 - Dec 2021

Education

- **Doctor of Philosophy** (Electrical and Computer Engineering)
  University of Washington, Seattle, WA.
  Cumulative GPA : 3.92 out of 4.0 — Sept 2015 - Aug 2020
- **Master of Science** (Electrical Engineering)
  University of Washington, Seattle, WA. — Sept 2015 - March 2018
- **Bachelor of Electronics and Telecommunication Engineering**
  Jadavpur University, Kolkata, India.
  Cumulative GPA : 9.67 out of 10.0 — Aug 2011 - June 2015

Research Internship

- **Data and Applied Scientist Intern, Microsoft**
  Communications Intelligence Group, MSAI,
  (*Collaboration with Microsoft Research, Redmond*). — June - Sept. 2019.
- **Data and Applied Scientist Intern, Microsoft**
  Substrate and Query Intelligence (AI + R). — June - Sept. 2018.
- **Indo-German Exchange Scholarship (Daad-Wise)**
  Computer Vision and Pattern Recognition Group, Institute of Computer Science,
  Westfälische Wilhelms-Universität Münster, Germany. — May - August 2014.

Technical Skills

- **Programming Languages** : Python (Proficient), C++, Java (Prior exposure)
- **Machine Learning Frameworks** : PyTorch, Semantic Kernel, Langchain.
- **Data and MLOps** : Azure Machine Learning, Databricks, SQL, PromptFlow.

Selected Research Projects

- **Meeting Summarization from Teams transcripts**
  - As part of a research and incubation effort inside MSAI, built a two stage extractive and abstractive summarization solution to summarize the key discussions for recorded meetings. Large Language Model (LLM) was used to create synthetic dialogs as well as summaries for distilled model fine-tuning. Feature is currently planned for release to Teams Premium customers.
- **Smart To-Do : Automatic Generation of To-Do List from Emails**
  - *This work was featured in online publications and AI news websites such as TechZine, Beebom and The Next Web (TNW).*
  - (*Microsoft Internship, Summer 2019*) Created an entire machine learning pipeline - (a) Defining problem structure and annotation guidelines, (b) Data generation through Crowd-sourcing, (c) Design of extractive and abstractive summarization algorithms for To-Do item generation from emails. The generated To-Do lists were remarkably similar to those written by humans.
- **Clustering in Generative Adversarial Networks**
  - This project found traditional Gaussian and Uniform priors to be unsuitable for clustering in GANs. We designed a new architecture, ClusterGAN, that achieves low-dimension embeddings with cluster structure and preserves interpolation properties in latent space.
- **Estimation of Conditional Mutual Information**
  - (*Collaboration with IBM Research*) In this project, novel estimators for mutual information and conditional mutual information were developed using generative models and classfiers. A test for conditional independence was also built from these estimators.

- **Project Science Engine: Search and Recommendations**
  - As part of Project Science Engine, trained and deployed models for:
    (a) Multimodal search to retrieve reaction recipes for a new query reaction. It involved jointly embedding reactions and known reaction recipes in a vector space using multimodal transformer based architecture. (Patent application WO2024005977A1)
    (b) Ranking papers based on predicted impact score. Trained ranking model based on text embeddings, journal and author reputation features and created automatic pipeline in Azure to continuously rank newly published papers.
- **Content-Based Email Response Prediction**
  - (*Microsoft Internship, Summer 2018*) Scalable user representations based on the content of historically received emails was used for email reply prediction on publicly available Avocado Email Collection. This approach solved the acute "cold-start" problem and was able to accommodate the inherent temporal nature of the problem, where users and email content constantly evolve with time.

Service to Community

- **Review**: Served as reviewer of top-tier journals and conferences in the field of Machine Learning and AI including Nature Communications, IEEE TIP, ISIT, IJCAI, IEEE TSMC, SIGGRAPH and JSAIT.
- **Teaching Assistant** : Data Science (Spring '17), Representation Learning (Spring '18), Information Theory (Winter '19), Probability Models and Inference (Spring '20).

Selected Publications

- *"Smart To-Do: Automatic Generation of To-Do Items from Emails"*- Sudipto Mukherjee, Subhabrata Mukherjee, Marcello Hasegawa, Ahmed Hassan Awadallah, Ryen White. *58th annual meeting of the Association for Computational Linguistics (ACL), 2020.* (17.5 % Acceptance)
- *"CCMI : Classifier based Conditional Mutual Information Estimation"*- Sudipto Mukherjee, Himanshu Asnani, Sreeram Kannan. *Conference on Uncertainty in Artificial Intelligence (UAI-19).* (26 % Acceptance)
- *"ClusterGAN : Latent Space Clustering in Generative Adversarial Networks"*- Sudipto Mukherjee, Himanshu Asnani, Eugene Lin, Sreeram Kannan. *33rd AAAI Conference on Artificial Intelligence (AAAI-19).* (16.2 % Acceptance, Oral Presentation)
- *"Resolving multicopy duplications de novo using polyploid phasing "*- Mark J. Chaisson*, Sudipto Mukherjee*, Sreeram Kannan, Evan E. Eichler. 21^{st} *International Conference on Research on Computational Molecular Biology, RECOMB 2017.* (21.6 % Acceptance)
- *"A deep adversarial variational autoencoder model for dimensionality reduction in single-cell RNA sequencing analysis"*- Eugene Lin, Sudipto Mukherjee, Sreeram Kannan. *BMC Bioinformatics*, 2020.
- *"Robust Community Detection for resolving segmental duplications : The last frontier to assembly "*- Sudipto Mukherjee, Sreeram Kannan. *NeurIPS Workshop on Machine Learning in Computational Biology (MLCB)*, 2017.
- *"Variants of k-regular nearest neighbor graph and their construction"*- Klaus Broelemann, Xiaoyi Jiang, Sudipto Mukherjee, Ananda S.Chowdhury. *Information Processing Letters, Elsevier*, October 2017.
- *"A Lagrangian Approach to Modeling and Analysis of a Crowd Dynamics"*- Sudipto Mukherjee, Debdipta Goswami, Sarthak Chatterjee. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, June 2015.

Major Academic Achievements and Accolades

- Recipient of Prof. Jnansaran Chatterjee Memorial *Gold Medal* in Jadavpur University Convocation for highest aggregate score in Tele-Communication subjects.
- Recipient of the *Supriya Basu Scholarship* for being the *Best Student from Faculty of Engineering - 3^{rd} Year* (2014).
- Recipient of *Meera Rani Mitra Memorial Award (Gold Plated Silver Medal)* for securing the highest marks in the Department in 3^{rd} year Engineering (2014).
- Recipient of *Mamraj Agarwal Rashtriya Purashkar* in August, 2011 based on Standard 12 result from His Excellency Governor of West Bengal, India.