Soumya Saha

thisissoumyasaha@gmail.com
413-570-5644
soumyaumass
soumyaumass

Education

- **University of Massachusetts, Amherst** — Amherst, MA
  Master of Science in Computer Science with concentration in Data Science; GPA: 4.0/4.0 — Sep 2018 - May 2020
  Courses: Intro to NLP, Neural Networks, Machine Learning, Deep Learning for NLP, Reinforcement Learning, Algorithms for Data Science
- **Indian Institute of Technology Kharagpur** — Kharagpur, India
  Bachelor of Technology in Electronics and Electrical Communication Engineering; GPA: 8.23/10.0 — Jul 2012 - May 2016

Skills Summary

- **Languages**: (Proficient in) Java, C++, Python (Familiar with) Typescript, C, SQL
- **AWS Technologies**: API Gateway, Lambda, Step Functions, DynamoDB, SageMaker, ECS, ECR, S3, CloudFormation
- **Tools**: Git, Agile Methodology, Vim, Docker, PyCharm, Perforce, Jira, LaTeX
- **Machine Learning**: Logistic Regression, SVM, Random Forest, Gradient Boosting, Clustering, LSTM, Transformer
- **Deep Learning and NLP**: RNNs, Language Model, Machine Translation, Text Summarization, Text Recommendenation

Industry Experience

- **Amazon** — Seattle, WA
  Applied Scientist II — Oct 2024 - Present
  - Working on improving Math, Instruction Following capabilities of Amazon's LLMs during post-training stage.

  Software Development Engineer II (Transitioning to Applied Scientist) — April 2022 - Sep 2024
  - Working on a new Alexa feature to provide customers with opportunities to learn languages and evaluate their learnings.

  Software Development Engineer — Jul 2020 - March 2022
  - Integrated Alexa's How To Say (HTS) Translation feature with Help Me Get Started feature and migrated HTS GUI to APL. Both projects contributed heavily to the YoY growth of HTS making it the 2nd highest growing Alexa feature in 2021.
  - Developed and facilitated the release of Alexa's Live Translation feature and expanded into multiple language pairs.

  Software Development Engineer Intern — Jun 2019 - Aug 2019
  - As a member of Alexa AI Natural Language Understanding (NLU) team implemented a self-serve Experimental Model Building Service by dynamically interacting with different AWS technologies using Python boto3 AWS client.
  - Evaluated the functionality of the Service by creating and executing Intent Classification and Named Entity Recognition models written in MXNet based Alexa DeepNLU framework in a workflow on AWS ECS Cluster.
- **Lexalytics Inc. & Umass Amherst** — Amherst, MA
  Graduate Student Researcher — Jan 2019 - April 2019
  - Worked on an NLP task of generating abstractive summaries of longer form of documents (scientific research papers).
  - Using a dataset of scientific articles from arXiv, we trained a seq2seq Bi-LSTM encoder-decoder model with sectional and sentence attentions in PyTorch using sentence embeddings from BERT as inputs to the encoder.
  - Summaries generated from the model are generally syntactically correct and have a Rouge-1 score of 21.45.
- **Samsung R&D Institute Delhi** — Noida, India
  Software Development Engineer — July 2016 - July 2018
  - As a C++ developer, led a team of 4 to develop a Native Client (NaCl) based IPTV Reference Application with Javascript based frontend for playback of Live and VOD Channels in Samsung Smart TV running on Tizen OS.
  - Successfully integrated RTP, RTSP, HLS Streaming Protocols and Sample AES DRM in the above solution.
  - Contributed to the maintenance of GStreamer based Player in Netflix application for Samsung Tizen UHD TV Models.

Publication

- **Text Embellishment using Attention based LSTM and Transformer Network (pdf)**
  Presented in CC-NLG, INLG 2019 Workshop — Oct 2018 - Dec 2018
  - Implemented a neural text embellishment model using attention-based LSTM and state-of-the-art Transformer encoder-decoder network. The model was trained with WikiLarge dataset and pre-trained Glove Embeddings.
  - The LSTM model achieved 91 and Transformer achieved 66 BLEU scores with better FKGL Readability scores.

Academic Projects

- **Graph Representation Learning for cold-start Scientific Paper Recommendations** — Aug 2019 - Dec 2019
  - Explored, analyzed and tested current cold-start algorithms for scientific research paper recommendations to users.
  - Used SciBERT model and Node2vec algorithm to generate user and paper embeddings to improve recommendations.
- **German to English Machine Translation with Linguistic Features** — March 2019 - May 2019
  - Extracted German linguistic features such as POS tags, Lemma using Parzu and achieved BLEU score of 29.1 on IWSLT dataset, using the word embeddings and the linguistic features as input to 6-layer Transformer in PyTorch.
  - Achieved 31.3 BLEU score by training the previous Transformer model with Byte Pair Encoded (BPE) input.