Rohan Kumar

rohankum@cs.cmu.edu
rohankumar.github.io

EDUCATION

Carnegie Mellon University   Pittsburgh, PA
MS in NLP and ML (Intelligent Information Systems, Advisor: Ruslan Salakhuditnov)   *Aug 2021 – Dec 2022*
Coursework: Intro to Machine Learning (PhD-level), ML for Text & Graph Mining, Advanced NLP, Convex
Optimization, Advanced Deep Learning, Computational Ethics for NLP, Multimodal ML, Probability & Statistics

VIT University   Vellore, India
B. Tech. Computer Science and Engineering (ABET Accredited)   *Apr 2012 – May 2016*

EXPERIENCE

Rakuten Research, RIT   San Mateo, CA
Research Intern (Host: Dr. Young-joo Chung)   *May 2022 - Aug 2022*

- Proposed a new task of Personalized Keyphrase Extraction for Explaining Product Recommendations.

Flipkart (a Walmart company)   Bangalore, India
Data/Applied Scientist   *April 2017 - July 2021*

Search Query Rewriting and Expansion

- Improved retrieval performance by reducing vocabulary mismatch between user searches and product catalog.
- Developed a neural query rewriting model applying co-attention on Siamese Manhattan LSTMs.
- Deployed a BERT-based query-phrase expansion model trained on 20M+ queries.
- Improved user metrics: clickthrough rate (1.37%), add-to-cart ratio (6.74%) and null-search ratio (-15.84%).
- Reduced inference latency for online deployment by 79% over BERT using knowledge distillation and quantization.

Machine Learned Ranking for Core Search and Query Auto-Completion

- Developed clickthrough(ctr) and conversion(cvr) prediction models for item ranking in Core Search and Pairwise
  Learning to Rank model for Suggestion ordering in Query Auto-Completion.
- Created Apache Hive ETL pipelines and feature transformers to process 100M+ datapoints.
- Used linear models that meet latency constraints.
- Significant reduction of click-depth and improvement in clicks and purchases.

Carnegie Mellon University   Pittsburgh, PA
Visiting Scholar, Computer Science Department (Advisor: Christos Faloutsos)   *Sept 2016 - Mar 2017*

- Worked on a logs based approach to measure user satisfaction and detect anomalous behavior in search.

Flipkart (a Walmart company)   Bangalore, India
Data Science Intern (Host: Dr. Mohit Kumar)   *Jan 2016 - June 2016*

- Quantification and categorization of unfulfilled user demand of items by exploratory analysis of rare and low CTR
  search queries to identify gaps in product selection.

PUBLICATIONS

Rohan Kumar, Surender Kumar, and Samir Shah. *"PSimBERT: Query Phrase Expansion for Low Recall Vocabulary Gap Queries."* **SIGIR 2022 eCom**

Subhadeep Maji[*], Rohan Kumar[*], Manish Bansal, Kalyani Roy, and Pawan Goyal. *"Logic Constrained Pointer Networks for Interpretable Textual Similarity."* **IJCAI 2020**

Subhadeep Maji, Rohan Kumar[*], Manish Bansal[*], Mohit Kumar, Kalyani Roy, and Pawan Goyal. *"Addressing Vocabulary Gap in E-commerce Search."* **SIGIR 2019 Short**

Rohan Kumar, Mohit Kumar, Neil Shah, and Christos Faloutsos. *"Did We Get It Right? Predicting Query Performance in E-commerce Search."* **SIGIR 2018 eCom**

Tejaswin Priyam, Rohan Kumar, and Siddharth Gupta. *"Tweeting Traffic: Analyzing Twitter for generating real-time city traffic insights and predictions."* **IKDD CoDS 2015**

PROJECTS

Knowledge Graph based reasoning for Question Answering (Ongoing)
Exploring combination of Large Language Models and Graph Neural Networks for Question Answering.

Course Project on Podcast Summarization
Experimented with Transformers with pointer generator network, Sparse Attention Models and sentence-masking pretraining (Pegasus) for long document summarization of conversational text. | *Report* | *Code*.

Course Project on Framing in News Media
Identified correlations between economic indicators and New York Times opinion articles categorized using domain-adapted RoBERTa model. | *Report*.

INTERESTS AND ACTIVITIES

- **Research Interests**: Natural Language Processing, Machine Learning, Information Retrieval

- **Skills**: Python, SciPy stack (numpy, pandas, matplotlib), PyTorch, Spark, Bash, SQL/Hive, Git and AWS,GCP.

- **Selected Participant:** Machine Learning Summer School(MLSS), Taipei 2021, Advanced Language Processing Winter School, Univ. Grenoble & Naver Labs Europe 2021, Deep Learning Reinforcement Learning Summer School by Mila & CIFAR 2020

- Invited talk on Query Rewriting at SlashN, Flipkart's annual technical conference

- First Prize, Data Challenge, Conference on Data Science by ACM India SIGKDD 2015

- Student Achiever's Award, VIT University.

- Reviewer for SIGIR eCom'19