Rohan Kumar

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

EDUCATION

Carnegie Mellon University   Pittsburgh, PA
MS in NLP and ML (Intelligent Information Systems, Advisor: Yiming Yang)   *Aug 2021 – Dec 2022*
Coursework: Intro to Machine Learning (PhD-level), ML for Text & Graph Mining, Advanced NLP

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

SKILLS AND INTERESTS

Skills: Python (SciPy stack), PyTorch, PySpark, Bash, SQL/Hive, Git and Hive
Research Interests: Natural Language Processing, Machine Learning, Information Retrieval

EXPERIENCE

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

Search Query Rewriting and Expansion
- Developed models for addressing vocabulary gap between user query and product catalog.
- Deployed a neural network rewriting model incorporating co-attention in Siamese Manhattan LSTMs.
- Implemented a BERT-based phrase-level query 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%).
- Published in SIGIR 2019 as a short paper and filed as a patent.

Popularity scores for machine learned ranking
- Clickthrough rate (CTR) and Conversion (CVR) prediction models used in search ranking.
- Developed Apache Hive ETL pipelines and feature transformers creating datasets with upto 100M observations.
- Trained linear models for high-volume product categories.

Suggestion ranking for query auto-completion
- Pairwise ranking with linear model optimizing for latency constraints.
- Reduction of click-depth and improvement in downstream metrics.

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   *Jan 2016 - June 2016*
- Quantification and categorization of unfulfilled user demand of items by means of exploratory data analysis on rare and low CTR search queries to identify gaps in product selection.

PUBLICATIONS

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

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

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

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

SELECTED ACTIVITIES

- Reviewer for SIGIR eCom'19
- Invited talk on Query Rewriting at SlashN, Flipkart's annual technical conference
- Student Achiever's Award, VIT University.