SUMANTA KASHYAPI

sumantakashyapi@gmail.com | +1 603-617-6681
www.linkedin.com/in/sumantakashyapi
https://www.cs.unh.edu/~sk1105/

Passionate researcher with a focus on Deep Learning and NLP, exploring how AI can be harnessed for the good of mankind.

PUBLICATIONS

- **Kashyapi, S.**, & Dietz, L. (2022, June). Query-specific subtopic clustering. In Proceedings of the 22nd ACM/IEEE **JCDL** (Joint Conference on Digital Libraries) (pp. 1-9). ***Best student paper nominee.***
- Dietz, L., Chatterjee, S., Lennox, C., **Kashyapi, S.**, Oza, P., & Gamari, B. (2022, July). Wikimarks: Harvesting Relevance Benchmarks from Wikipedia. In Proceedings of the 45th International ACM **SIGIR** Conference on Research and Development in Information Retrieval (pp. 3003-3012).
- **Kashyapi, S.**, & Dietz, L. (2022, December). Topic-Mono-BERT: A Joint Retrieval-Clustering System for Retrieving Overview Passages. In Proceedings of the 14th Annual Meeting of the **FIRE** (Forum for Information Retrieval Evaluation) (pp. 54-59).
- **Kashyapi, S.**, & Dietz, L. (2021). Learn The Big Picture: Representation Learning for Clustering (Representation Learning for NLP Workshop at ACL-IJCNLP 2021).
- Lennox, C., **Kashyapi, S.**, & Dietz, L. (2023). Retrieve-Cluster-Summarize: An Alternative to End-to-End Training for Query-specific Article Generation. arXiv preprint arXiv:2310.12361.
- **Kashyapi, S.,** Chatterjee, S., Ramsdell, J., & Dietz, L. (2018). TREMA-UNH at TREC 2018: Complex Answer Retrieval and News Track. in TREC (Text REtrieval Conference).

RESEARCH PROJECTS

Joint Clustering-Retrieval System for Overview Passage Retrieval – Python, PyTorch *May 2021 – February 2022*

- Developed Topic-Mono-BERT, a neural retrieval model that is jointly supervised by two complementary tasks: retrieval and clustering. While the clustering task learns the subtopic embedding space of the document collection, the retrieval task optimizes the query-passage relevance. We show that this combination is particularly beneficial for overview-style passage retrieval that requires knowledge about the subtopics pertaining to the query achieving about 16% improvement in MAP.
- https://github.com/nihilistsumo/ORCA

Clustering Optimization as Blackbox (COB) – Python, PyTorch, CUDA *August 2020 – March 2021*

- Developed COB, a scalable training strategy for supervised clustering, that is at least 100 times faster than traditional approach while achieving better or comparable accuracy. Unlike traditional approaches, it directly optimizes for a discrete clustering metric. A BERT-based LLM embedding model is trained using COB which outperforms another BERT-based embedding model employing Triplet loss and other unsupervised baselines on two clustering benchmarks.
- https://github.com/nihilistsumo/Blackbox_clustering

Context-Aware Trimaese Similarity metric (CATS) – Python, PyTorch *May 2019 – April 2020*

- Conducted research to show that while clustering text-passages relevant for a particular query, it is beneficial to incorporate the query-context information into the clustering algorithm. Based on this research, CATS is designed to calculate context-aware pairwise similarity score that improves the clustering accuracy by about 12%.
- https://github.com/nihilistsumo/CATS

Effective Prediction of Interesting Data Points for MMS – R, Python *January 2018 – May 2018*

- Developed a model to forecast data points (with upto 92% accuracy) pertaining to strong interaction between Sun and Earth's magnetic fields from low-resolution satellite data, useful for scientists involved in Magnetospheric Multiscale Mission (MMS), a NASA funded project supervised by UNH Space Science department.

PATENTS

- **Kashyapi, Sumanta.** 2023. Machine learning/ deep learning engines used to determine path root cause of failures. U.S. Patent DC-131740.01, filed April 27, 2023. Patent pending.
- **Kashyapi, Sumanta.** 2023. Transformer-based automatic labeler for misaligned anomalous event with time series data. U.S. Patent DC-132501.01, filed June 05, 2023. Patent pending.
- **Kashyapi, Sumanta.** 2023. Time series anomaly detection with rare event failure prediction. U.S. Patent DC-134123.01, filed Sep 13, 2023. Patent pending.

RELATED WORK EXPERIENCE

Senior Data Scientist – Python, PyTorch, Kedro, mlflow *July 2022 – Present*

Dell Technologies – Hopkinton, MA

- Developed predictive models for rare catastrophic events with around 75% accuracy from input telemetry time-series data of storage devices.
- Implemented strategies to handle extreme class imbalance (10000:1) of telemetry time-series.
- Been the lead inventor in multiple patents related to root cause analysis of failures from input telemetry time-series data.
- Involved in the orchestration of the full spectrum of Mlops – including model development, training, evaluation and serving.

- Participated in a hackathon to develop a deployment pipeline for models with extreme space and bandwidth constraints.

AI/ML Intern – Python, PyTorch, FIO *May 2021 – August 2021*
Dell Technologies – Hopkinton, MA

- Developed predictive models to forecast storage usage/ IO request patterns for different customers.
- Led a team of interns to execute the IO patterns generated by the prediction model on test devices and study the responses.
- Improved the process of replicating customer data in terms of compressibility and reproducibility.
- Participated in a hackathon to develop an ML application to reduce zoom-fatigue by facilitating personalized break schedules.

Machine Learning Intern – Python, PyTorch, Github *May 2020 – August 2020*
MMS Analytics – Portsmouth, NH

- Automated the process of evaluating health provider groups, saving hours of manual labor.
- Developed a neural similarity metric suitable for clustering health providers data with 98% accuracy.

Teaching Assistant – Java, Algorithms *August 2017 – May 2022*
University of New Hampshire – Durham, NH

- Assisted in teaching two UNH courses (Intro to Java, Advanced Java, Intro to algorithms) by mentoring diverse groups of about 25 students.

Systems Engineer – Java, J2EE, SQL *June 2012 – June 2014*
Tata Consultancy Services – India

- Worked as the primary support engineer of client-side business modules; solved critical front-end and back-end issues impacting user experience and business workflow.
- Analyzed, designed and implemented enhancement requests from the users; independently as well as a part of dev teams.

TECHNICAL SKILLS

Programming Languages:	Java, Python, R, C, SQL
Library, Packages, API, Tools:	Keras, Pytorch, Tensorflow, Kedro, Mlflow, Numpy, Pandas, Lucene, git
Concepts:	Information Retrieval, Deep Learning, Machine Learning, Statistical Learning, Natural Language Processing, Large Language Models (LLM), Algorithms

EDUCATION

University of New Hampshire – Durham, NH *May 2022*
PhD in Computer Science, focusing on Information Retrieval
Thesis: Query-Specific Subtopic Clustering in Response to Broad Queries
Supervisor: Prof. Laura Dietz
In the early stages of information seeking process when the query is not yet well formulated, the candidate set of relevant documents may span a wide range of topics. In such scenarios, an effective information retrieval engine has to consolidate the large set of relevant documents in a meaningful way. This thesis explores the formulation, optimization and novel methodologies related to subtopic clustering of documents specific to vague queries and investigates its relation to the overall retrieval quality.

National Institute of Technology Hamirpur – India *June 2016*
Master of Technology in Computer Science

Kalyani Government Engineering College– India *May 2012*
Bachelor of Technology in Information Technology