Pooja H. Oza

📞 +1 603-767-5153   ✉ oza.pooja.h@gmail.com
in pooja-oza   ⌂ poojahoza

Summary

Passionate and accomplished researcher with a strong background in natural language processing, search systems and deep learning; seeking to leverage my expertise to tackle complex challenges and drive scientific discovery.

Education

Doctor of Philosophy   *University of New Hampshire, NH*
Computer Science   *Expected Aug 2024*
Focus: Natural Language Processing, Large Language Models, Graph Neural Networks, & Information Retrieval

MS Program   *International Institute of Information Technology, India*
Advanced Software Technologies   *2007 – 2009*

Work Experience

Research Assistant (NLP/ Deep Learning)   *Aug 2018 – Present*
University of New Hampshire   *NH, USA*

- **Designed** and trained textual **representation learning** model that combines unstructured text data with symbolic knowledge generated by **fine-tuned Large Language Models** (LLMs) leading to improve natural language processing and information retrieval tasks.
- **Designed** and **trained** retrieval-augmented generation **(RAG)** model that utilize in-context learning and zero-shot, CoT **prompts** to retrieve explanations and entities for navigational search system.
- Successfully developed predictive model that **improves over the attention** mechanism by modeling relevance in graph encoder for structured symbolic knowledge representation learning to improve performance of question answering and information retrieval systems.
- Developed a supervised ranking model that generate knowledge representations through transfer learning, achieving more than **25%** improvement in information retrieval systems.

Machine Learning Intern   *Jun 2022 – Aug 2022*
DeepSee.ai   *Utah, USA*

- Participated in Knowledge automation workflow development by modeling similarity to detect discrepancy in financial documents, leading to **increased** user efficiency.
- Designed and trained a **transformer-based** text classification model to attain an accuracy of **85%**.
- Designed and implemented a data processing pipeline, including data de-duplication, to automate the feature extraction and data cleaning process.

Senior Software Engineer   *Aug 2016 – Jun 2018*
iQuanti Inc.   *Bengaluru, India*

- Developed a near **real-time data processing** framework based on distributed micro-service architecture that processed approximately 10,000 processes in parallel for a Search Engine Optimization product.
- Built services to calculate relevancy of the documents using statistical models and numerical calculations and to expand the query using relevance feedback, resulting in **increased** retrieval performance.

Technical skills

Programming Languages	Python, JAVA, R, C++
Library Tools	PyTorch, Pytorch-Geometric, HuggingFace, OpenAI, Transformers, PEFT, Accelerate, Numpy, Pandas, Scikit-Learn, Spacy, Lucene, Git, AWS, MLOps
Databases and Indexing	FAISS, MongoDB, SQL, LINUX

Research & Publications

Graph Relevance Networks
Jan 2021 - May 2023

- Significantly **improved** over the graph attention mechanism by proposing a relevance mechanism that enhances representation learning for Natural Language Processing and Information Retrieval systems.
- **Trained** and **Fine-tuned** foundational models to generate knowledge representations resulting in performance improvement over information retrieval systems.
- **Publications**:

    - "Neural Entity Context Models" **Oza, Pooja**, Chaterjee, Shubham and Dietz, Laura., IJCKG'23.

    - "Entity Embeddings for Entity Ranking: A Replicability Study." **Oza, Pooja**, and Dietz, Laura., ECIR'23

    - "Identify Relevant Entities Through Text Understanding." **Oza, Pooja**., CIKM'22

    - "Which entities are relevant for the story?." **Oza, Pooja**, and Dietz, Laura., Text2Story@ECIR'21

Fine-grained Representations Learning
Jan 2023 - Sep 2024

- Substantially **improved** textual representation learning by modeling generated symbolic knowledge with the unstructured textual data for question answering and information retrieval tasks.
- **Developed** retrieval-augmented generation model to generate relevant explanations through symbolic knowledge and evaluate the LLM-generated text.
- **Poster**:

    - "Addressing Exploratory Information Needs with Relevant Entities" **Oza, Pooja**, Choi, Jaekeol and Dietz, Laura., Task Focused IR in the Era of Generative AI. 2023.