Yelugam Pranay Kumar

(+413) 2754654 | pyelugam@umass.edu | linkedin.com/pranayyelugam | github.com/pranayyelugam

Education

University of Massachusetts Amherst, MA
Master of Science in Computer Science (GPA: 4/4) *Feb 2021 – Present*
Relevant course work: Advance Natural Language Processing, Advance Machine Learning, Algorithms for Data Science, Statistics, Applied Information Retrieval, Distributed Systems, and Intelligence Visual Computing

Indian Institute of Information Technology Allahabad, India
Bachelor of Technology in Computer Science *Aug 2014 – July 2018*
Relevant course work: Data Structures and Algorithms, Operating Systems, Data Mining, Machine Learning, Artificial Intelligence

Experience

Research Intern — Advisors: Prof. Andrew McCallum & Xiang Lorraine Li Amherst, MA
Information Extraction and Synthesis Laboratory [link] *May 2021 - Present*

- **Commonsense Frame Completion:** Working on commonsense knowledge reasoning.
  - Currently working on developing solutions to integrate the knowledge from Generative Language Models and Knowledge Bases to create more robust and structured systems for commonsense question answering.
  - Developed a dataset for commonsense question answering using AMR representation of commonsense context sentences.
  - Developed a probabilistic model for evaluating the answer distributions from the commonsense knowledge questions.

Graduate Research Student — Advisors: Prof. Andrew McCallum & Neha Kennard Amherst, MA
Information Extraction and Synthesis Laboratory [link] *Feb 2021 - May 2021*

- **Discourse Structure:** Worked on problems in discourse structure at the document level.
  - Developed and annotated a large dataset of scientific peer review text to highlight discourse structure.
  - Developed classification and span selection models to automatically detect the discourse structure using this dataset.

Software Engineer, Samsung Delhi, India
Unified Metadata Team *July 2018 - Dec 2020*

- Implemented a Convolutional Neural Networks and Longest Common Subsequence based solution to infer content similarity from different content providers.
- Implemented *Simhash* for clustering duplicate content from the content description.
- Built the TV Plus parser from scratch for Samsung TVs with data ingestion occurring every 60 seconds.
- Designed and implemented an ETL pipeline that ingests real time TV programs and schedules from major streaming applications like Apple TV, Amazon Prime etc in different formats to the database.
- Re-launched Search Data Exporter for Samsung TV search with 6% reduced latency and developed new features to improve the search data quality.

Publications and Preprints

Kennard Neha Nayak, Tim O'Gorman, Akshay Sharma, Chhandak Bagchi, Matthew Clinton, **Pranay Kumar Yelugam**, Rajarshi Das, Hamed Zamani, and Andrew McCallum. "A Dataset for Discourse Structure in Peer Review Discussions." preprint arXiv:2110.08520 (2021).

Pranay Kumar Yelugam, Bharadwaju, K. Anudeep, A. Vamshi Krishna, Bakshi Rohit Prasad, Sonali Agarwal "Real time mining of ego networks for exploring social associations" CICT, 2017

Projects

PixeltoVoxel — *Python, PyTorch, CNNs* Spring 2021

- Improved upon the existing Pix2Vox for converting the single/multiple images of the object to 3D represention of the objects using a encoder, decoder architecture. The model is trained on the ShapeNet dataset. The resultant model performed better when evaluated on Intersection Over Union metric.

Exploring Circles in Ego Networks — *Python, Flask, PageRank Algorithm and Clustering Algorithms* Fall 2016

- Proposed and developed clustering based approach for mining of ego networks to explore ego's social associations and rank them using PageRank.

Co-Authorship Analysis — *Python, Networkx, Link Prediction Algorithms* Spring 2017

- Implemented a Link Prediction approach to create a recommender system that helps in finding potential collaborators for an author.

Languages and Technologies

- **Python**: PyTorch, sckit-learn, pandas, numpy, NLTK, matplotlib
- **Others**:C/C++, Java, Scala, JavaScript, PostgreSQL, ReactJS, Redux, Akka, Git, AWS (EC2, S3), Google Colab