

ANKITA RAJARAM NAIK

(+1) 413-404-580 ♦ arnaik@umass.edu

ankitanaik.github.io
linkedin.com/in/naik-ankita

EDUCATION

University of Massachusetts Amherst

Master of Science in Computer Science (4.0/ 4.0)

Amherst, MA

Sep 2019 - May 2022

Courses - Natural Language Processing, Reinforcement Learning, CNN for Visual Recognition, Probabilistic Graphical Models, Advanced Machine Learning, Theory and Practice of Software Engineering, Advanced Algorithms, Database Design and Management

Indian Institute of Technology Madras

Dual degree (BS & MS) in Materials Engineering (Minor in Operations Research)

Chennai, India

Jul 2012 - Jul 2017

Courses - Machine Learning, Data Structures & Algorithms, Data Mining & Warehousing, MOOCs - Deep Learning (Andrew NG)

PUBLICATIONS

Rajarshi D., Ameya G., Ankita N., Elliot T., ... McCallum, A. (2022). **Knowledge Base Question Answering by Case-based Reasoning over Subgraphs**. International Conference on Machine Learning. ICML, 2022.

Michael G., Gaetano R., Md Faisal C., Ankita N., Pengshan C., Alfio G. **Re2G: Retrieve, Rerank, Generate** Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics, NAACL 2022.

Ankita N.*, Apurva S.*, Kartik M.* **"Shallow-UWnet: Compressed Model for Underwater Image Enhancement. (Student abstract)"** Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 18. 2021.

Ankita N.*, Nikhil M.*, Surya M.*, Ho Y.*, Meet V., Khaled Y., Katelyn N., Gireesha R. **"Leveraging Knowledge Distillation for Efficient on-device Deployment of Deep Learning Models in Medical Imaging."** *Conference on Machine Intelligence in Medical Imaging*. SIIM 2020

RESEARCH PROJECTS

Case-based Reasoning (CBR) for Interpretable Bio-medical Graph Reasoning**

Jan 2022 - Apr 2022

Research Project : Information Extraction and Synthesis Laboratory

Advisor : Dr. Andrew McCallum

- Participated in the OpenBioLink 2021 - a link prediction challenge on large-scale, heterogeneous bio-medical data.
- Implemented probabilistic case-based reasoning approach which predicts answers for query entity by gathering reasoning paths from similar entities in the KB. The model was the top scoring explainable model and came fourth overall.

Knowledge Base Question Answering by Case-based Reasoning over Subgraphs**

Apr 2021 - Dec 2021

Research Project : Information Extraction and Synthesis Laboratory

Advisor : Dr. Andrew McCallum

- Developed a semi-parametric model which leverages structural similarity between local neighborhoods of subgraphs
- Created query specific subgraphs and used contrastive loss objective for learning parameters of Graph Neural Network
- Published findings in ICML, 2022 for improving state-of-art performance on MetaQA and WebQSP dataset

Reasoning on Knowledge Graphs using box embeddings

Jan 2020 - May 2020

Semester Project : Information Extraction and Synthesis Laboratory

Advisor : Dr. Andrew McCallum

- Represented entities (nodes) of graphs as boxes in place of traditional vector embeddings for knowledge graph completion
- Achieved 5% increase in answering complex logical queries using such graph embeddings on incomplete knowledge graphs

Underwater Image Enhancement using Deep Neural Networks**

Sep 2019 - Dec 2019

Course Project : Neural Networks - A Modern Introduction

Advisor : Dr. Erik Miller

- Conceptualized a shallow residual network (5 layers) using improved perpetual loss function for Image enhancement task
- Increased MSE by 2% & reduced computation time by 50% compared to state-of-art GANs containing 14 layers

PROFESSIONAL EXPERIENCE

IBM Research AI, Research Intern**

Knowledge Induction Group, Knowledge integration and representation for Data Lakes applications Yorktown Heights, NY

Neural retriever-reader architecture for Table Augmentation

May 2022 - Aug 2022

- Developing neural retrieval (Dense Passage Retrieval) and span selection based model for cell filling on WikiTables Corpus
- Conceptualizing self-supervised architecture for Dense Passage Retrieval training for various table augmentation tasks

Neural retriever-generator architecture for Fact Verification (Re2G**)

May 2021 - Aug 2021

- Developed end-to-end neural retrieval, reranker and BART-based sequence-to-sequence generator for Fact Verification
- Published findings in NAACL, 2022 for improving state-of-art performance by 5% on facebook's KILT dataset - FEVER

Walmart Inc., Machine Learning Intern Jul 2020 - Aug 2020
Global People Analytics Team, US-based multinational retail corporation operating hypermarket chain Bentonville, USA

- Redesigned loss function and optimizer of Learning Fair Representation, a fairness algorithm used for hiring process
- Reduced convergence time from 8hrs to 1hr by using 1st order optimizer with 10% improvement in fairness metric
- 1 of 2 interns selected to present internship work to the Leadership team of Walmart Technology and Data Organization

GE Healthcare, Graduate Student Researcher** Jan 2020 - Apr 2020
Medical Imaging Division, US-based conglomerate developing Health technology for medical imaging Amherst, USA

- Compressed DenseNet model using Knowledge Distillation for carrying out disease detection on portable X-Ray machines
- Reduced computational cost while preserving accuracy (70 times less memory, 300 times less operations, 5 times less data)
- Published finding in Conference on Machine Intelligence in Medical Imaging SIIM 2020

ZS Associate, Data Scientist Jul 2017 - Jul 2019
Advanced Data Science Group, US-based Healthcare consultancy specializing in Marketing Solutions Pune, India

Disease prediction using low-dimensional medical embedding learnt on Electronic Health Records(EHR)

- Developed generalized patient & medical event embeddings to diagnose cervical cancer using patient claims data
- Achieved 80% prediction accuracy using LSTM, 5% higher than traditional feature creation based models like XGBoost
- Awarded **\$50,000 seed funding** during Annual Hackathon by leadership to scale framework for other disease

Clinical Trial Optimization using Patient claims database

- Modeled drug's market share at launch using XGBoost on clinical trial endpoints (ClinicalTrials.gov) & patients claims
- Extracted trial endpoints from research article by combining results from StanfordNERtagger with established Ontology
- 1 of 10** (among 100+ projects) awarded at client's **Business & Analytics Innovation summit** for revamping traditional market research based launch share prediction method using machine learning on clinical trial and claims data

TECHNICAL SKILLS

Computer Languages	Python, R, C++, C, HTML, SQL
Python Libraries	PyTorch, TensorFlow, HuggingFace, NLTK, scikit-learn, pandas, numpy, matplotlib, PySpark
Software & Tools	MATLAB, Simple Linux Utility for Resource Management (Slurm), L ^A T _E X, AWS (EC2, S3), GitHub
Operating Systems	Windows, Linux
Biomedical Datasets	MIMIC - III, PubMed, ClinicalTrials.gov, Electronic Health Records (EHR) DrugBank UMLS

TEACHING EXPERIENCE

- Teaching Assistant, CS 589 Introduction to Machine Learning, UMass - Amherst** Jan 2022 - May 2022
- Created assignments, examinations and tutorials for 250+ undergraduate and graduate students.
- Teaching Assistant, CS 682 Neural Networks, UMass - Amherst** Sep 2021 - Dec 2021
- Conducted PyTorch and deep learning model debugging session, assignment & exam evaluations for 350 graduate students.
- Graduate Grader, CS 119 Introduction to Python Programming, UMass - Amherst** Jan 2020 - Dec 2020
- Assisted in conducting TA sessions, assignment and exam evaluations papers for 200 undergraduate students.
- R-programming Training Sessions, ZS Associates, Pune, India** Jun 2017 - Aug 2017
- Conducted comprehensive R sessions and industrial programming standards lectures for 100+ freshers
- Teaching Assistant, Mechanical and Metallurgy Laboratory, IIT-M** Jul 2016 - Nov 2016
- Assisted in conducting, supervising experiments and evaluating exam papers for 50 undergraduate students.

EXTRA-CURRICULARS

- Reviewer, NewInML Workshop @ ICML 2022** May 2022 - Jun 2022
- Reviewed a research publication submitted to NewInML Workshop at ICML 2022
- Student Scholar, Grace Hopper and Tapia Conference** Sep 2020 - Aug 2021
- Obtained 100% scholarship for attending Grace Hopper and Tapia conference
- Organizing Committee, Voices of Data Science @ UMass - Amherst** Feb 2020 - Feb 2021
- Reviewed a research publication submitted to NewInML Workshop at ICML 2
- Co-founder Institute Analytics Club, IIT Madras** Apr 2014 - Apr 2015
- Facilitated 6 industrial projects and Summer School for 40 students. Conducted professor talks and seminars for 200+
- Community Service** Jul 2015 - Jul 2017
- Setup multiple libraries in Orphanages in Chennai, India as a part of the Pledge-a-book Campaign
 - Mentored students as a volunteer at Deepam, an Non-Profit Organization serving education to the underprivileged