Nidhi Piyush Vakil

Boston, MA, USA
(Green Card holder; no visa sponsership needed.)

Contact Information
(+1) 714-726-9425
nidhivakil1@gmail.com || Personal Page || Google Scholar || LinkedIn

Research Interest
Natural Language Processing, Large Language Models (LLM), Health Informatics, Machine Learning, Graph Neural Network

Highlights

- Final year Phd student with 4+ years in AI, ML, NLP, and Graph domains.
- Developing ML algorithms to integrate structured (graphs) and unstructured (NLP) knowledge.
- Papers accepted at ACL, EMNLP, NAACL, INTERSPEECH.
- Hands-on experience in healthcare and general data applications.
- Fine-tuned LLMs, extended framework for custom tasks, and leveraged models in zero-shot settings. Built models for classification, regression, and generation tasks.
- Experience working with text and audio data.
- Hands-on experience with MIMIC-III/IV and SNOMED ontologies and extracting features for medical tasks, such as Mild Cognitive Impairment (MCI)
- Hands-on experience using Pytorch, HugginFace, Pytorch geometric, Ollama, LangChain, Docker, GitHub, Python.

Publications

- Nidhi Vakil, and Hadi Amiri. "Controlled Transformation of Text-Attributed Graphs" (**EMNLP 2024**).
- Hadi Amiri, Nidhi Vakil, Mohamed Elgaar, Jiali Cheng, Mitra Mohtarami, Adrian Wong, Mehrnaz Sadrolashrafi, Leo Anthony G Celi. "Analysis of Race, Sex, and Language Proficiency Disparities in Documented Medical Decisions" (**MedRxiv 2024**).
- Jiali Cheng, Mohamed Elgaar, Nidhi Vakil, Hadi Amiri "CogniVoice: Multimodal and Multilingual Fusion Networks for Mild Cognitive Impairment Assessment from Spontaneous Speech" (**INTERSPEECH 2024**)
- Mohamed Elgaar, Jiali Cheng, Nidhi Vakil, Hadi Amiri, Leo Anthony Celi. "MedDec: A Dataset for Extracting Medical Decisions from Clinical Narratives" **(ACL 2024)**
- Nidhi Vakil, and Hadi Amiri. "Complexity-Guided Curriculum Learning for Text Graphs" (**EMNLP 2023**).
- Nidhi Vakil, and Hadi Amiri. "Curriculum Learning for Graph Neural Networks: A Multiview Competence-based Approach" (**ACL 2023**).
- Nidhi Vakil, and Hadi Amiri. "Generic and Trend-aware Curriculum Learning for Relation Extraction in Graph Neural Networks" (**NAACL 2022**).

Education

Ph.D. in Computer Science 2020 - 2025 (Expected)
Advisor: Dr. Hadi Amiri,
Computational Language Understanding (CLU) Lab
University of Massachusetts, Lowell, United States
GPA: 3.95/4.0

Master of Science 2014 - 2016
Major in Electrical Engineering
California State University, Fullerton, California, United States
GPA: 3.57/4 (*First Class with Distinction*)

Bachelor of Engineering 2009 - 2013
Major in Electronic and Communication Technology
Gujarat Technological University, Gujarat, India
GPA: 7.49/10 (*First Class with Distinction*)

Technical Skills

- Programming Language: Python
- Machine Learning Packages: Numpy, SciPy, Pandas, NTLK toolkit, HuggingFace, spaCy, LangChain, Ollama
- Machine learning algorithms: GNNs, CNN, Transformer based models-BERT, RoBERTa, Phi-3, Flan-T5, T5, and others
- Deep Learning Framework: PyTorch, PyTorch-Geometric
- Database Language and IDE: SQL, Jupyter Notebook, PyCharm

Projects

Detecting Alzheimers and related dementias through Social Determinants of Health

- Trained a regression model and a multi-class classification model to obtain the MCI score based on the social determinants of the datapoints.
- Fine-tuned small and large language models using Hugging Face.
- Implemented zero-shot multi-class classification.
- Extended the Hugging Face framework to build a custom regression model.
- Evaluated small and large language models: BERT, RoBERTa, T5, and Flan-T5. algorithms and analytic approaches for early prediction of Alzheimer's disease and related dementias based on social determinants of health,

MedDec: A Dataset for Extracting Medical Decisions from Discharge Summaries

- Conducted an exhaustive analysis of medical disparities across protected variables (e.g., gender, race) using the MIMIC-III and MIMIC-IV datasets.
- Performed a detailed analysis of different decision types (e.g., drug, evaluation, legal) on professionally annotated discharge summaries.
- Contributed to the creation of a benchmark medical decision dataset .
- Published paper in ACL 2024 and MedArXiv 2024.

Controlled Transformation of Text-Attributed Graphs

- Designed and implemented a controlled graph transformation approach that modifies graphs based on user requirement to provide control over fine-grained topological structures.
- Demonstrated the utility of the algorithm in low-resource data settings without modifying existing model.
- Applied the method to low-resource molecular datasets such as MUTAG and MOLBACE, and achieved performance gain of 10.5 and 4.6 accuracy points respectively.
- Published paper in EMNLP 2024.

CogniVoice: Multimodal and Multilingual Fusion Networks for Mild Cognitive Impairment Assessment from Spontaneous Speech

- Converted audio clips into textual data to predict for classification task.
- Extracted DisVOICE features from audio clips.
- Classified language characteristics of the generated text for MCI diagnosis.
- Published paper at INTERSPEECH 2024.

Curriculum Learning for GNNs: A Multiview Competence-based Approach

- Developed a new effective curriculum learning approach for better data ingestion during training. New Approach leverages multiple views of difficulty for a given samples and model's competence.
- Ability to incorporate a fine-grained spectrum of difficulty criteria in the training paradigms.
- Evaluated the data ingestion curricula on node classification and edge prediction tasks from general domain and medical domain.

- Provided insights about evaluating multi-view difficulty of the samples.
- Achieved at least an improvement of 5 points in F1 score and Accuracy
- Published paper at ACL 2023

Complexity-Guided Curriculum Learning for Text Graphs

- Designed a novel data scheduler which uses "spaced repetition" and employs graph and text complexity for training Graph Neural Networks (GNNs) on text-graph data.
- Gained performance boost of 5.1 absolute points compared to state-of-the-art baselines while using 39.2% less data
- Provided insights into the learning dynamics of GNNs, i.e., which complexity features are learned by GNNs during training.
- Learned curricula is transferable across GNN models and datasets.
- Published paper at EMNLP 2023.

Generic and Trend-aware Curriculum Learning for Relation Extraction in Graph Neural Networks

- Implemented supervised graph algorithm to effectively integrate textual and structural information for relation extraction in text graphs.
- Conducted extensive experiments on real world datasets in both general and specific domains.
- Designed generic and trend-aware curricula, "Trend-SL" that incorporates sample-level loss trajectories (trends) to better discriminate easier from harder samples and schedule them for effective learning.
- Achieved an average improvement of 8.6 points in F1 score
- Published paper at NAACL 2022.

Large Scale Graph Multiclass Classification

- Developed an algorithm for node classification using an English Wikipedia hyperlink network
- Used Wikipedia articles as nodes and hyperlinks between the articles as edges
- Used textual description of the articles
- Trained using Glove and BERT embeddings
- Trained GraphSAGE on 13K class labels Used 20K training instances, 10K validation instances, 60K testing instances. Obtained Accuracy of 32.21%

Certifications

- Oxford Machine Learning Summer School (**OxML** 2021)
- Machine Learning, Stanford University and Coursera (2020)

Academic Service

- Reviewer for COLING 2024, ARR 2024, EMNLP 2023, NAACL 2025
- Served as a on-site volunteer for ACL 2023

References

Prof. Hadi Amiri **E-mail**: hadi@cs.uml.edu
Assistant Professor
Department of Computer Science at University of Massachusetts, Lowell
Department of Biomedical Informatics at Harvard University
Massachusetts, USA.

Last update on December 29, 2024