Laiba Mehnaz

+1 929 584 4743| lm4428@nyu.edu | Website | linkedin | github

Education

New York University, Tandon School of Engineering May 2023
Master of Science in Computer Science
Relevant coursework: Design and Analysis of Algorithms, Machine Learning, Operating Systems

Delhi Technological University, New Delhi, India May 2020
Bachelor of Technology in Software Engineering
Relevant coursework: Data Structures in Java, Database Management Systems, Object Oriented Programming, Software Testing, Compiler Design

Technical Skills

Languages: Python, Java, C/C++, R, SQL
Frameworks: PyTorch, TensorFlow, Keras, SpaCy, Sci-Kit, NLTK, Gensim, D3.js, Git

Projects

Unsupervised Domain Adaptation for Sentiment Analysis using BERT [**poster**] [**code**]
- Built a system using adversarial neural networks(DANs) and increased the accuracy of large-scale language models on new data by 12.5% without using the new data(unsupervised), for the task of sentiment analysis.
- Awarded the first place for this project in the The 17th Term Project Showcase, at National University of Singapore. [**certificate**]

Work Experience

Applied Scientist Intern May 2022 – Aug 2022
Amazon *Seattle, USA*

Multilingual domain adaptation for zero-shot cross-lingual transfer
- Built deep learning architectures that can adapt multilingual language models to specific domains, and show high performance in zero-shot settings, i.e., show competitive performance on new languages that are not seen during training.

Research Assistant June 2020 – June 2021
MIDAS lab, Indraprastha Institute of Information Technology, Delhi *New Delhi, India*

Conversation summarization and translation
- Led a team to curate and release a large scale conversation summarization dataset for conversational AI systems.
- Built and implemented several state-of-the-art large neural language models to summarize as well as translate conversations written in code-mixed Hindi-English to English.
- Conducted an exhaustive analysis to show why large neural language models do not generalize to different languages and fail in a similar manner.
- Built a Python framework to compute automated code-mixed metrics for Hindi-English code-mixed data.

Domain Robustness of pretrained language models
- Conducted a thorough study to understand how well do the large neural language models generalize to new unseen data using probing and domain divergence.
- Performed exhaustive experiments to show that these models encode linguistic information about new unseen data in similar layers in the models, as they do for training data.

Selected Publications

"GupShup: An Annotated Corpus for Abstractive Summarization of Open-Domain Code-Switched Conversations."
EMNLP-2021, Dominican Republic.
Laiba Mehnaz, Debanjan Mahata, Rakesh Gosangi, Uma Sushmitha Gunturi, Riya Jain, Gauri Gupta, Amardeep Kumar, Isabelle Lee, Anish Acharya, Rajiv Ratn Shah. [**paper**] [**code**]

[**Full list of publications**]