Anvesh Rao Vijjini

linkedin.com/in/anvesh-rao/
github.com/nvshrao
nvshrao.github.io
✆ +1(984) 322-8184
✉ anveshrao1@gmail.com

Summary

PhD Student with enthusiasm for research in natural language generation, robustness in NLP, interpretability, and learning in resources poor settings. I research building NLP models efficient in parameters and robust to variations in data. I consider interpreting models to be a crucial step in achieving this. Only when we find out what ML models learn; can we guide them to learn better. I have also contributed to improving resource-scarce languages' representation by curating resources in manners that require minimal human intervention. I have applied my research interests across NLP applications such as Sentiment Analysis, Reading Comprehension, Question Generation, WER Estimation and Story Generation. I am presently working in narrative summarization, something I am currently excited about.

Education

2021 - present   **PhD in Computer Science**,
University of North Carolina at Chapel Hill.
Advisor: Dr. Snigdha Chaturvedi

2013 - 2021   **MS in Computational Linguistics**,
International Institute of Information Technology, Hyderabad.
Advisor: Dr. Radhika Mamidi

2013 - 2017   **B.Tech in Computer Science**,
International Institute of Information Technology, Hyderabad.
Awarded Dean's List for being in the top 10% of the class in academics for Monsoon 2015 and Spring 2016 semesters.

Experience

May 2022 - July 2022   **Research internship at Adobe**.
Worked on transcript segmentation at Project Blink, Adobe's video sharing platform.

- Transcript segmentation can help break down long videos into multiple coherent subsections based on change of topic or content style in transcripts.
- Proposed a pretraining strategy that improves segmentation by training the model on synthetic samples that closely resemble the downstream task. The model is trained in a curated order, with samples arranged in increasing similarity. This curriculum-based approach overcomes the challenges of transcript segmentation, such as the noisy and sporadic nature of the data.
- Published our work "Curricular Next Conversation Prediction Pretraining for Transcript Segmentation" at EACL 2023 (Findings)

August 2018 - August 2021	**Lead Engineer (Research) at Samsung R&D Institute India-Bangalore**. Worked at Bixby Intelligence, Samsung's voice assistant.

- ○ Automated WER Prediction. Automated WER estimation of Bixby's ASR based on Deep Learning models.
- ○ Automated Conversation Quality Estimation. For performance evaluation of Bixby conversations, we implemented an automated feedback prediction system based on state-of-the-art text classification algorithms.
- ○ Data Augmentation to eschew manual annotations. Conversation Quality Estimation relies heavily on laborious tagging. Worked on retaining performance on feedback prediction with much lesser labeled data by utilizing ideas from Domain Adaptation and Semi-Supervised learning.
- ○ Authored research papers based on Text classification and Domain Adaptation at EMNLP 2019, CICLing 2019, NLDB 2019, NLDB 2020, ICPR 2020 and EACL 2021.

Spring 2017	**Teaching Assistant**, *Natural Language Application*. The course covers popular Machine Translation (statistical) methods at the time such as IBM Models, Phrase Based Model and Neural Machine Translation.
Monsoon 2016 and 2017	**Teaching Assistant**, *Natural Language Processing*. An overview of various areas in NLP such as Tokenization, Parts Of Speech tagging, Natural Language Generation and Parsing.

Volunteer

- ○ Reviewer for **ACL 2020, EMNLP 2020, NLP+CSS @ EMNLP 2020, EACL 2021, ACL 2021**.
- ○ Sessions Chair for **NLDB 2020**

Conference Publications

January 2023	**Curricular Next Conversation Prediction Pretraining for Transcript Segmentation**, ***Anvesh Rao Vijjini***, *Hanieh Deilamsalehy, Franck Dernoncourt and Snigdha Chaturvedi*. EACL 2023 (Findings).
October 2022	**Towards Inter-character Relationship-driven Story Generation**, ***Anvesh Rao Vijjini***, *Faeze Brahman and Snigdha Chaturvedi*. EMNLP 2022.
April 2021	**WER-BERT: Automatic WER Estimation with BERT in a Balanced Ordinal Classification Paradigm**, ***Anvesh Rao Vijjini⋆***, *Akshay Krishna Sheshadri⋆ and Sukhdeep Kharbanda*. EACL 2021.
January 2021	**Sequential Domain Adaptation through Elastic Weight Consolidation for Sentiment Analysis**, *Avinash Madasu and **Anvesh Rao Vijjini***. ICPR 2020
June 2020	**A SentiWordNet Strategy for Curriculum Learning in Sentiment Analysis**, ***Anvesh Rao Vijjini⋆***, *Kaveri Anuranjana⋆ and Radhika Mamidi*. NLDB 2020

June 2020	**A Position Aware Decay Weighted Network For Aspect Based Sentiment Analysis**, *Avinash Madasu and **Anvesh Rao Vijjini**.* NLDB 2020
November 2019	**Sequential Learning of Convolutional Features for Effective Text Classification**, *Avinash Madasu and **Anvesh Rao Vijjini**.* EMNLP-IJCNLP 2019
June 2019	**Gated Convolutional Neural Networks for Domain Adaptation**, *Avinash Madasu and **Anvesh Rao Vijjini**.* NLDB 2019
April 2019	**HindiRC: A Dataset for Reading Comprehension in Hindi**, ***Anvesh Rao Vijjini**⋆, Kaveri Anuranjana⋆ and Radhika Mamidi.* CICLing 2019
April 2019	**Effectiveness of Self Normalizing Neural Networks for Text Classification** , *Avinash Madasu and **Anvesh Rao Vijjini**.* CICLing 2019
August 2018	**Twitter corpus of Resource-Scarce Languages for Sentiment Analysis and Multilingual Emoji Prediction**, *Nurendra Choudhary, Rajat Singh, **Anvesh Rao Vijjini** and Manish Shrivastava.* COLING 2018

Workshop Publications

April 2021	**Analyzing Curriculum Learning for Sentiment Analysis along Task Difficulty, Pacing and Visualization Axes**, ***Anvesh Rao Vijjini**⋆, Kaveri Anuranjana⋆ and Radhika Mamidi.* Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA) at EACL 2021.
August 2019	**Hindi Question Generation Using Dependency Structures**, ***Anvesh Rao Vijjini**⋆, Kaveri Anuranjana⋆ and Radhika Mamidi.* Workshop on Humanizing AI (HAI) at IJCAI 2019.
August 2018	**Towards Enhancing Lexical Resource and Using Sense-annotations of OntoSenseNet for Sentiment Analysis**, *Sreekavitha Parupalli, **Anvesh Rao Vijjini** and Radhika Mamidi.* Workshop on Semantic Deep Learning (SemDeep-3) at COLING 2018.
July 2018	**BCSAT: A Benchmark Corpus for Sentiment Analysis in Telugu Using Word-level Annotations**, *Sreekavitha Parupalli, **Anvesh Rao Vijjini** and Radhika Mamidi.* Student Research Workshop (SRW) at ACL 2018.
July 2018	**Towards Automation of Sense-type Identification of Verbs in OntoSenseNet**, *Sreekavitha Parupalli, **Anvesh Rao Vijjini** and Radhika Mamidi.* Workshop on Natural Language Processing for Social Media (SocialNLP) at ACL 2018.