Khandokar Md. Nayem

nayem.github.io

+1 (812) 369-6358

knayem@iu.edu

TECHNICAL EXPERTISE

• Sequence-to-sequence modeling • Speech Enhancement, Translation & Recognition • Machine Learning • Deep Learning • Natural Language Processing (NLP) • Classification & Regression • Knowledge Distillation • Machine Learning Models (Bayesian Network, HMM, GMM, Clustering, Decision Tree, Ensemble Methods) • Deep Neural Network (DNN) • Recurrent Neural Network (RNN) • Long-Short Term Memory (LSTM) • Convolutional Neural Network (CNN) • Transformers • Deep Learning Libraries (PyTorch, TensorFlow, Keras, Fairseq) • NLP Libraries (NLTK, Scikit-Learn) • Human Survey Platform (Qualtrics, Amazon MTurk) • Computer Vision & Graphics Libraries (CImg, OpenCV, OpenGL) • Python • Matlab • C\C++ • Java • HTML, CSS & Javascript • SQL • R • Shell Script

EMPLOYMENT

Amazon Services LLC, Cambridge, MA, *Applied Scientist Intern, Alexa AI* — Fall 2022

- Researched the development of a real-time, end-to-end compressed multi-lingual speech translation system. Investigated the use of large transformer-based models and applied knowledge distillation approach to transfer their performance to smaller models with 50% and 75% fewer parameters. (Will submit, at INTERSPEECH 2023)

Microsoft Corporation, Redmond, WA, *Audio & Acoustics Research Intern* — Summer 2022

- Focused on analyzing and improving the performance of speech enhancement algorithms to generate high-fidelity (Hi-Fi) speech by removing distortions and extending speech bandwidth. Applied causal LSTM models with various augmentation to recover codec and clipping distortions, and performed deep noise suppression. (Will submit, at INTERSPEECH 2023)

BOSE Corporation, Framingham, MA, *Machine Learning/Neural Signal Processing Intern* — Summer 2020

- Researched on enhancing speech in remote microphone applications by removing self-speech in order to provide better quality sound with low latency to hearing aids and voice-assistive wearable devices. Utilized an LSTM-based architecture with speaker-dependent d-vector for speaker identification, to ensure real-time operation.

Indiana University, Bloomington, IN, *Research Assistant, ASPIRE research lab* — Fall 2016 - Present

- Developed an attention-based monaural speech enhancement model with the objective of maximizing human perceptual rating of enhanced speech. This was accomplished by incorporating embedding vectors from a human Mean-Opinion Score (MOS) prediction model and jointly training the model utilizing real-world noisy speech data. (INTERSPEECH-2021)
- Proposed & implemented a quantized speech prediction model that classifies speech spectra into a corresponding quantized class, and applies a language-style model to generate more realistic speech. Acceptable quantization level was determined by listener study conducted on Amazon MTurk, designed using Qualtrics. (ICASSP-2021, poster, slides, video)
- Designed a recurrent layer, named Intra-Spectral Recurrent (ISR) layer to capture spectral dependencies within the magnitude and phase responses of noisy speech using Markovian recurrent connections. This was successfully integrated into a LSTM-based single-channel speech enhancement model.(ICASSP-2020, slides, video)
- Formulated a new type of recurrent output layer that enforces spectral-level dependencies within each spectral time frame, by modeling the Markovian assumption along the frequency axis in both uni-directional and bi-directional ways. This was tested in a magnitude speech enhancement model. (MLSP-2019, poster)
- Engineered a deep architecture, named Recurrent Stacked Generative Adversarial Network (RSGAN) to generate video clips based on a precondition, such as a sentence description, action classes, or fMRI signals. (IU-VISION-2017, poster)

REVE Systems, Dhaka, *Jr. Software Engineer, Team Media Gateway* — January 2015

- Programmed media gateway controller to facilitate both calls and faxes between the telephone network and VoIP network or another telephone network using the Megaco 1.0 protocol. Additionally, designed a front-end panel using the .JSP framework for easy use by VoIP administrators and customers.

EDUCATION

Ph.D. in Computer Science, — April 2023 (Anticipated)
Indiana University, Bloomington, IN — Advisor: Prof. Donald S. Willamson

M.Sc. in Computer Science, — December 2019
Indiana University, Bloomington, IN

B.Sc. in Computer Science & Engineering (CSE), — July 2014
Bangladesh University of Engineering & Technology (BUET), Dhaka, Bangladesh