Syeda Nahida Akter

Mobile: +880-1821874052 | Email: 1505027.sna@ugrad.cse.buet.ac.bd | Website: snat1505027.github.io | Google Scholar | Github

Research Interest

- Natural Language Processing
- Data Mining
- Bioinformatics
- Distributed Deep Learning

Education

Bachelor of Science in Computer Science and Engineering February 2016 – Present
Bangladesh University of Engineering and Technology *Dhaka, Bangladesh*

Thesis: Geo-Distributed Machine Learning, Natural Language processing
Supervisor: Dr. Muhammad Abdullah Adnan, Associate Professor, Department of CSE, BUET
Current CGPA: 3.72 out of 4.00 (7/8 terms, Expected Completion: February, 2021)
Major CGPA: 3.83 out of 4.00 (7/8 terms)
Final Two Year CGPA: 3.88 out of 4.00

Publications

WeightGrad: Geo-Distributed Data Analysis Using Quantization for Faster Convergence and Better Accuracy
Authors: *Syeda Nahida Akter, Dr. Muhammad Abdullah Adnan*

Proc. of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2020), San Diego, CA, USA, August 23-27, 2020. [PDF]

Research Experience

WeightGrad: Geo-Distributed Data Analysis Using Quantization for Faster Convergence and Better Accuracy
Supervisor: Dr. Muhammad Abdullah Adnan (BUET)
Keywords: Distributed Deep Learning, Gradient Quantization
Status: Accepted for publication in **ACM SIGKDD'20** as a full-track research paper
Resources: [PDF], [Promo Video], [Code]

- We provide loss-aware weight-quantized network with quantized gradients to alleviate the limitations of state-of-the-art quantization and propose a synchronous two-level structure to ensure global convergence within a fixed time interval.
- We observe that gradient clipping with momentum correction along with layer-wise ternarizing with a smaller learning rate gives the best convergence rate.
- We attain $5.36\times$ speedup over baseline and $1.4 - 2.26\times$ speedup over the four state-of-the-art distributed ML systems without hurting the accuracy like traditional distributed systems. In fact, we obtain 1.06% gain in top-1 accuracy.

Contextualized Entity-Aware Question Answering June 2020 – Present
Supervisor: Dr. Muhammad Abdullah Adnan (BUET)
Keywords: NLP, Entity-Representation, ALBERT, XLM-Roberta, Contextual QA
Status: Preparing manuscript to be submitted to ACL'21
Resources: [Abstract]

- We develop a model for contextual entity-aware Question Answering (QA) using pretrained XLM-Roberta, ALBERT.
- We propose a new fine-tuning structure: embedding layer with a topic modeling and multi-stage attention layer with feedback-loop to increase accuracy for context-based question answering system.
- We address multiple persons/organizations/events in the context (create ambiguity while generating answer) issue by applying an attention layer followed by an entity extraction layer to ensure query-based tagging which will narrow down the search space and eliminate ambiguity.
- We are emphasizing on cost-effective well-structured fine-tuning steps rather than modifying the pretrained model to address challenges of incorporating entity-based commonsense into the language models.

Responding to the Stigma and Misinformation Related to the COVID-19 Disease Outbreak April 2020 – Present
Supervisor: Dr. Syed Ishtiaque Ahmed (University of Toronto)
Keywords: NLP, Data mining, Topic Modeling, Ensemble Learning, Text Classification
Status: Submitted initial findings to CPHA'20
Resources: [Abstract]

- We use Natural Language Processing (NLP) and Social Network Analysis (SNA) to study collected anonymized Twitter data. I am working on the NLP section.

- We use state-of-the-art NLP techniques to analyze the following from social media posts: keyword gathering, frequency analysis, information extraction, automatic categorization and clustering, automatic summarization, sentiment analysis and finding associations within the data.
- We are working on different topic modeling architectures (LDA, GSDMM, BTM, lda2vec, BERT, Twitter-LDA) for short texts like tweets, and trying to retrieve the true meaning from tweets using ensemble learning.
- We are fine-tuning models for sentiment analysis and text classification (DistilBert, ABSA) and developing novel architecture to detect stigma in tweet texts.

Phylogenetic Tree Estimation Using Deep Neural Network June 2020 – Present

Supervisor: Dr. Md. Shamsuzzoha Bayzid (BUET)

Keywords: Deep Learning in Bioinformatics, Phylogenetic Tree Estimation

Status: In progress (expected completion: December 2020)

- We design an end-to-end deep neural network to predict a species tree from a set of (embedded) gene trees. The model is designed to work well for a wide range of taxa.
- Our novelty is ensuring unique vector representation of the triplets generated from each gene tree by incorporating an embedding layer. This way, different gene trees can be easily differentiated during the entire training process.
- We introduce subsequent self-attention layers to so that the model can capture similarities and discrepancies among the embedded triplets and thus can provide a better summarization.

Image Analysis based Dehydration Detection August 2020 – Present

Supervisor: Dr. Tanzima Hashem (BUET)

Keywords: Computer Vision

Status: In progress (expected completion: July 2021)

- We are working on identifying the image features of skin and body parts (e.g., eye,tongue and lips) that are responsible for dehydration, and plan to build a classifier to predict dehydration level (mild, moderate, acute).
- This work is a collaboration with a medical team from International Centre for Diarrhoeal Disease Research, Bangladesh [icddr,b]

Notable Projects

ML Based Credit Risk Prediction System for Commercial Banks | *XGBoost, Django, MongoDB, JS* 2019 – 2020

Supervisor: Dr. Muhammad Abdullah Adnan (BUET); Dr. Md. Shamsuzzoha Bayzid (BUET)

Resources: [Repository]

A ML based banking software system (a collaboration with Bank Asia) which aims at predicting trustworthiness of a client who has applied for a loan by predicting credit risk. It also predicts the allowable loan amount that can be granted without causing any risk. I worked on the machine learning application that included data cleaning, feature extraction, experimenting on different classifiers and regression models (random forest, decision tree, XGBoost). Our classifier, built using customized boosting and bagging method, achieved 82% test accuracy (SOTA accuracy is **78%** for the same dataset).

ChatBot with Emotion | *FACE API, Telegram API, Google Speech API* January 2019

Achievement: Champion in the BUET CSE Fest Hackathon, Cloud Computing Category, 2019.

Resources: [Repository]

A chatbot which processes texts using Google's Natural Language API and interacts accordingly, analyzes user's facial emotion using Google's FACE API and suggests user specific genre of songs or videos depending on the mood, takes voice command using Speech-to-Text and replies back to user using Text-to-Speech APIs of Google.

Tori: Mental Health Care App using Machine Learning | *Android API, Python, NLTK* June 2018

Achievement: Champion in the national hackathon, Hack_A_Day, 2018.

Resources: [Repository]

Mental Health Care app using machine learning which stores users activities on the mobile and interacts with users through a chatbot which collects textual responses from users and using NLP, tries to predict users' mood and conducts further interaction accordingly (e.g. sending songs, videos, sending suggestions to contact with close friends, informing users about time spent in mobiles each day etc.)

Pooh: Website with Java | *WebView, WebEngine, Image4j, PDFRenderer* December 2016

Supervisor: Mohammed Kaysar Abdullah (BUET)

Resources: [Repository]

Pooh is a web browser built using JAVA, browses web pages like other web browsers (e.g. Gooogle Chrome, Mozila Firefox etc). It covers the basic features of a web browser such as navigation bar (home, history, bookmarks), settings (zoom, fullscreen, new tab/window), a PDF reader, etc.

Awards and Honors

- **KDD Student Registration Award**, 2020
- **University Merit Scholarship**, 2019
- **Dean's List Scholarship**, Bangladesh University of Engineering and Technology (BUET)
- **Champion on BUET CSE Fest Inter-University Hackathon** (Two consecutive years 2018, 2019)
- **Board Merit Scholarship: Government of Bangladesh** (Year: 2007, 2010, 2013, 2015)

Technical Skills

Languages: Python, Java, C/C++, C#, SQL (Postgres), JavaScript, HTML/CSS, Shell, Matlab, Intel 8086 Assembly Language
Frameworks: PySpark, MXNet, BERT, XLNet, RoBERTa, ALBERT, OpenGL, Django, Arduino
Libraries: Pytorch, Tensorflow, Keras, OpenCV, NLTK, Theano, Pandas, NumPy, Matplotlib, Scikit-Learn
Operating System: Linux, Windows
Developer Tools: Latex, Git, Aamazon Web Service, Google Cloud Platform

Reference

- Dr. Muhammad Abdullah Adnan
  Associate Professor
  Department of Computer Science and Engineering
  Bangladesh University of Engineering and Technology (BUET)
  Email: abdullah.adnan@gmail.com, adnan@cse.buet.ac.bd
  Mobile No: +880 1552 336926

- Dr. Syed Ishtiaque Ahmed
  Assistant Professor
  Department of Computer Science
  University of Toronto, Canada
  Email: ishtiaque@cs.toronto.edu
  Mobile No: +1 647 220 3482