SAI BHARATH CHANDRA GUTHA

Email: guthas@informatik.uni-freiburg.de
Mobile: +91-7989509808

Educational Qualification

Degree	Institution	Grade	Year of Graduation
Master of Science, Computer Science (enrolled summer sem. 2021)	University of Freiburg, Germany	-	-
Bachelor of Technology (Computer science & Engineering)	Indian Institute of Technology Kharagpur, West-Bengal, India	8.32 / 10	2018
12th / Intermediate	Narayana Sri Chaitanya Jr college, Hyderabad, Telangana, India	966 / 1000	2014
10th	SPR school of Excellence, Nalgonda, Telangana, India	9.7 / 10	2012

Work Experience

MAY – JULY 2017	Summer Internship at Samsung Research Institute Bangalore (SRIB), India
JULY 2018 – AUG 2020	Senior Software Engineer (Speech Recognition and NLP Team), SRIB, India
OCT 2020 – FEB 2021	Intern at VIGIL lab, IIT Hyderabad

Courses Taken (Math, AI related)

Machine Learning, Theoretical Machine Learning (Statistical Learning Theory), Deep Learning, Natural Language Processing, Artificial Intelligence, Social Computing, Single/Multivariable calculus, Matrix Algebra, Probability & statistics, Signals & Networks

Projects

[1] Mining of Mobile error log data

Developed a machine learning approach and a scalable implementation to classify/group similar events of mobile device error logs based on similar patterns for internal device diagnostics team at Samsung.

[2] YouTube video classification (NLP Course Project)

We tried to classify YouTube videos into Sleeping Beauties / Non Sleeping Beauties by leveraging on the video object information and several language features based on the title and comments of a video. A video that attains peak popularity (views) at least a year after its upload are considered as Sleeping Beauties.

[3] Bachelor's Thesis Project (Paper publication in IEEE BigData Conference 2018)

In Twitter, tweets are shared with followers of the seed user who in turn may share it with their own set of followers. Long chain of such retweets form cascades. We estimate the structure of a cascade tree by developing a model (Cascon) that leverages on temporal pattern of retweets obtained from time series of cascades and underlying follower network to predict cascade structure that closely resembles the gold standard cascades. we provide a closed-form analytical approximation of the degree distribution of the cascade trees and derive the structural metrics from this closed-form solution and validate it through numerical simulations demonstrating its effectiveness to resemble the real cascade structure whenever the retweet count is high.

[4] Rescoring Hybrid Language Model for LVCSR

Developed a novel architecture for Hybrid word-character language model to counter the out-of-vocabulary (OOV) issues in open domain large vocabulary continuous speech recognition (LVCSR), and also used it as spell checker to reduce ASR output spelling errors and word mis-recognitions.

[5] Improved Feed Forward Attention for Sequence Classification (Paper published in SPCOM 2020)

For Sequence Classification, we developed an improved attention mechanism in Bidirectional RNNs that involves computing attention weights separately for forward and backward layers and taking weighted combination of hidden states to generate output logits at every time step of the input sequence, followed by temporal averaging to compute final probability over K classes for which the input sequence belongs to. Our model outperformed the traditional Feed Forward Attention method in RNN for diverse classification tasks such as Speech Emotion, Accent Detection and Whispered speech detection.

[6] Whispered to Natural Speech Conversion & Vice-versa (on arXiv, Submitted to IEEE TASLP Journal, 2020)

Developed a Transformer based architecture to map short chunk of input speech feature vectors to the corresponding output speech feature vectors. Investigated the use of auxiliary decoder to predict phoneme label per frame and included this loss in the overall loss during training phase. Experiments were conducted to convert whispered speech to Natural speech and vice-versa. The Generated audio quality is evaluated by using a pretrained end-to-end neural ASR system to transcript the audios. Improvements were reported in Word error rate for the generated natural speech audios compared to the whispered speech audios.

[7] Real-time Video anomaly detection solution for edge devices (Project at VIGIL Lab, IIT Hyderabad)

Developed a real-time, compute-efficient solution for video motion anomaly detection and localization for edge devices. A completely unsupervised and iterative approach of pixel-smoothing + hierarchical clustering method was developed to predict an anomaly score for each voxel and thereby using it for frame-wise anomaly detection and pixel-wise anomaly localization. Our method out-performed supervised non Neural-network based models on some standard datasets with major advantages of our approach being completely unsupervised and compute efficient making it ideal for edge computing devices.

Technical Skills & Expertise

Programming Languages: C, C++, Java, Python, Shell Scripting

Libraries: Tensorflow, PyTorch, Horovod, Scikit-learn, OpenCV, NLTK, Numpy, Librosa, Scipy etc..

Tools & Software: git, AWS, Android Studio

Research Interests and Experience

Working experience in Computer Vision (CV), Speech recognition (ASR), Natural Language Processing (NLP) and so dealing with handling different types of data (images, audio-signals, text). Current research interests include Data representation & visualization, understanding deep neural networks.

Profile Links

semantic_scholar_profile linkedin_profile google_scholar_profile

Publications

[1] Bhowmick, Ayan Kumar, Sai Bharath Chandra Gutha, Y. Singh and B. Mitra. “Constructing Influence Trees from Temporal Sequence of Retweets: An Analytical Approach.” *2018 IEEE International Conference on Big Data (Big Data)* (2018): 624-633.

[2] Gutha, Sai Bharath Chandra, M. B. Shaik, Tejas Udayakumar and Ajit Ashok Saunshikhar. “Improved Feed Forward Attention Mechanism in Bidirectional Recurrent Neural Networks for Robust Sequence Classification.” *2020 International Conference on Signal Processing and Communications (SPCOM)* (2020): 1-5.

[3] Niranjan, Abhishek, Mukesh Sharma, Sai Bharath Chandra Gutha and M. Shaik. “WHALETRANS: E2E WHisper to nAturaL spEech conversion using modified TRANSformer network.“ *ArXiv* abs/2004.09347 (2020): n. pag.

References

HR, Samsung Research & Development Institute, Bangalore (SRIB), India (*Tel: 91-80-6126-1000, 080-4681-3000*)

Dr. Rer. Nat. Mahaboob Ali Basha Shaik, Senior Technical Manager, ASR Team, Voice Intelligence, SRIB, India (*Email: m.shaik@samsung.com, mobile: +91 9742313071, +91 7995050328*)

Dr. Bivas Mitra, Professor, Dept. of Computer science & Engineering, IIT Kharagpur, West Bengal, India (*Email: bivas@cse.iitkgp.ac.in, bivasmitra@gmail.com, Telephone: +91-3222-282358*)

Dr. C. Krishna Mohan, Professor, Dept. of Computer science & Engineering, IIT Hyderabad, Telangana, India (*Phone: (040) 2301 6021; Fax: (040) 2301 6032, Email: ckm@cse.iith.ac.in*)