Sangeet Sagar

Personal Website | Github

Email: sangeetsagar2020@gmail.com
Mobile: (+49) 1525 7657433
Address: München, Germany

EDUCATION

- **Universität des Saarlandes** — Saarbrücken, Germany
  Master of Science in Language Science and Technology; GPA: 1.8 (ECTS, lower is better) — *Oct. 2020 – April 2023*
  Thesis: Noise Robust Speech Recognition for Search and Rescue Domain [Report] [Code] [Video]

- **The LNM Institute of Information Technology** — Jaipur, India
  Bachelor of Technology in Electronics and Communication Engineering; GPA: 7.13/10.0 — *Aug. 2015 – June. 2019*
  Thesis: Analysis of Emotion Recognition using Speech Features

PROGRAMMING SKILLS

- **Languages**: Python, C++, Unix, MATLAB, R
- **Toolkits/Libraries**: PyTorch, SpeechBrain, K2/Icefall (next-gen-Kaldi), Scikit-learn, Numpy, Huggingface, Matplotlib
- **Misc.**: GIT, Advanced Linux user, AWS, Cluster Computing (HPC, SLURM), Docker.

EXPERIENCE

- **EML Speech Technology GmbH** — Munich, Germany
  Research Engineer for Automatic Speech Recognition — *Sept 2023*
  - Spearheading the development of end-to-end models to facilitate real-time Automatic Speech Recognition (ASR) during live conferences within a commercial setting.
  - Developed a C++ runtime for a streaming faster Conformer-Transducer (NeMo) and integrated its CPU-based decoder with our in-house end-to-end ASR decoder
  - Contributed to training a target speaker extraction system for isolating speech in overlapping scenarios, collaborating with team members to integrate with ASR models, resulting in an 18% improvement in WER on the reconstructed signal.

- **Airbus Defence and Space GmbH** — Munich, Germany
  Speech-to-Text Internship — *(5 months) May 2023 – Sept 2023*
  - Utilized SOTA models such as the Wav2Vec2 and Whisper ASR models to enhance communication between pilots and air traffic control (ATC) by developing state-of-the-art speech-to-text systems for aerospace domain data.

- **German Research Center for Artificial Intelligence (DFKI) GmbH** — Saarbrücken, Germany
  Research Assistant | HiWi — *(1.6 years) June 2021 – Feb 2023*
  - Designed and developed a noise-robust automatic speech recognition system (STT) (German language) as a component of MS thesis, enabling functionality under hostile noisy conditions such as search and rescue operations.
  - Trained open-source attention-based BiRNN punctuation restoration system+TruCasing for the German language. The system outperformed the baseline model- Vosk model by over 14% in recall metric.

- **Institute of Formal and Applied Linguistics, Charles University** — Prague, Czechia
  University research assistant — *(1 year) Oct. 2019 – Dec 2020*
  - Principle tester of the live speech-language translation (SLT) system at live test sessions for the ELITR project.
  - Training and testing (with in-domain/out-of-domain data) Czech punctuator system for live ASR.
  - Perform domain adaptation for live ASR by fine-tuning LM by using speaker-specific data.
  - Set up pipelines for the ELITR test set and to get it ready for evaluation.

- **Faculty of Information Technology, Brno University of Technology** — Brno, Czechia
  University research assistant — *(8 months) Feb. 2019 – Sept. 2019*
  - Developed a system for cross-lingual topic identification in low resource scenarios for Kinyarwanda, Zulu, Hindi and Reuters corpus.
  - Achieved a remarkable weighted average precision score of 0.52 (on 10 different topics) on Kinyarwanda language by learning a simple linear transformation from this language to English language space.
  - Primary task: text feature extraction, training classifier, generating document embeddings, and learning transformation from unknown language to English space.

Projects

- **English and Chinese poetry generation** *April 2022*
  English poetry generation and comparison (using PPL as evaluation metric) using LSTM, encoder-decoder based models, and transformer model (GPT). We conclude that fine-tuning GPT-2 model generated the highest quality poems i.e. with the least PPL score. Also trained a topic-prediction model to study how well a machine-generated poem is interpreted by a system trained on human-written poems. [Report] [Code]

- **Span extraction based slot-filling using attention and RNNs** *April 2022*
  Performed slot-filling (and intent recognition) using RNN on a multi-head attention mechanism. The main idea was to model the slot-filling task as a span extraction and to utilize available information about slot type for which value is to be provided. F1 score of 0.83 was achieved compared to 0.96 of the baseline model. [Report] [Code]

- **Out-of-vocabulary(OOV) word estimation using subword representation** *Aug. 2021*
  Achieved a better OOV rate and perplexity score than the baseline for three levels of granularity (char level, small, large vocab) with appropriate hyperparameter tuning. This was done by training RNN based language model to artificially generate corpus and compute OOV rate on varying sizes of the generated corpora. [Report] [Code]

- **[Bachelors Thesis] Analysis of Emotion Recognition using Speech Features** *2018*
  7% improvisation over baseline on the classification of speech signals based on human emotions like angry, disgust, fear, happy, etc. Implemented on SAVEE and Emo-DB datasets using classifiers like GMM, CNN, MLPNN, we propose the use of feature extraction algorithms like S-Transform and image spectrogram of the speech signal to perform emotion recognition. [Report] [Code]

Foreign Languages

- **Fluent**: English (C1)
- **Beginner**: German (A1)

Publications

[1] Sangeet Sagar et al. "RescueSpeech: A German Corpus for Speech Recognition in Search and Rescue Domain". In: ASRU 2023. arXiv: 2306.04054 [eess.AS].

[2] Sangeet Sagar, Abhinav Bhatt, and Abhijith Srinivas Bidaralli. *Defending Against Stealthy Backdoor Attacks*. 2022. arXiv: 2205.14246 [cs.CR].

[3] Ondřej Bojar et al. "ELITR Multilingual Live Subtitling: Demo and Strategy". In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. Online: Association for Computational Linguistics, Apr. 2021, pp. 271–277. DOI: 10.18653/v1/2021.eacl-demos.32. URL: https://aclanthology.org/2021.eacl-demos.32.

[4] Dario Franceschini et al. "Removing European Language Barriers with Innovative Machine Translation Technology". English. In: *Proceedings of the 1st International Workshop on Language Technology Platforms*. Marseille, France: European Language Resources Association, May 2020, pp. 44–49. ISBN: 979-10-95546-64-1. URL: https://aclanthology.org/2020.iwltp-1.7.

[5] Dominik Macháček et al. *ELITR Non-Native Speech Translation at IWSLT 2020*. 2020. DOI: 10.48550/ARXIV.2006.03331. URL: https://arxiv.org/abs/2006.03331.

[6] Peter Polák et al. "CUNI Neural ASR with Phoneme-Level Intermediate Step for~Non-Native~SLT at IWSLT 2020". In: *Proceedings of the 17th International Conference on Spoken Language Translation*. Online: Association for Computational Linguistics, July 2020, pp. 191–199. DOI: 10.18653/v1/2020.iwslt-1.24. URL: https://aclanthology.org/2020.iwslt-1.24.

[7] Hemanta Kumar Palo and Sangeet Sagar. "Characterization and Classification of Speech Emotion with Spectrograms". In: *2018 IEEE 8th International Advance Computing Conference (IACC)*. 2018, pp. 309–313. DOI: 10.1109/IADCC.2018.8692126.

[8] Hemanta Kumar Palo and Sangeet Sagar. "Comparison of Neural Network Models for Speech Emotion Recognition". In: *2018 2nd International Conference on Data Science and Business Analytics (ICDSBA)*. 2018, pp. 127–131. DOI: 10.1109/ICDSBA.2018.00030.