Status of Nepali Speech Processing

Introduction:

Speech recognition, also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, is a capability that enables a program to process human speech into a written format. Nepali speech recognition involves the conversion of Nepali speech to its correct Nepali transcriptions and can be used for interaction with the devices and instructing them to perform specific tasks. Compared to other languages like English, there hasn't been much research and development in Nepali speech and language systems. At present, the Nepali language comes as a low-resource language due to the lack of effort and contribution in collection of data and other resources for Nepali language.

Conclusion:

Automatic Speech Recognition is an exciting area of research in the application of ML these days. With the development of new technologies and growing data, different ML and DL models like RNN and LSTM have come as frameworks for developing ASR applications. These frameworks mostly provide a way to develop models whose parameters can be learned by providing a sufficient amount of labeled data. Pre-processing of collected Nepali speech is necessary for sampling and background noise removal. Then, the data can be supplied to train the deep learning model and the Automatic Speech Recognition (ASR) system based on deep learning can be used to translate spoken the Nepali language to its textual representation. The validation of the transcriptions will be done with the available Nepali Corpus by calculating Character Error Rate and Word Error Rate. This work summarizes all the efforts that have been done in the context of Nepali Speech processing and challenges ahead.

Keywords:

Artificial Intelligence, Unsupervised Learning, Low-Resource Languages, Audio

Basanta Joshi
Tribhivan University, Nepal

Biography:

Basanta Joshi received a Doctor of Engineering from Osaka Sangyo University, Japan in 2013. He did both Bachelor of Electronic and Communication Engineering and Masters of Science in Information and Communication Engineering from the Institute of Engineering (IOE), Tribhuvan University (TU), Nepal. Currently, he is working as Assistant Professor at the Department of Electronics and Computer Engineering, Pulchowk Campus and Deputy Director at Center for Applied Research and Development, IOE, TU .He is also associated with IOE as a Member of Laboratory for ICT Research and Development. Formerly, he used to work coordinator of Master's in Information and Communication Engineering, IOE, Senior Software Engineer in D2Hawkeye and as a Research Consultant at LogPoint. He is also actively involved in valuable researches in the field of Machine Learning and its application in Big Data, especially Images and Speech. He has been actively publishing national & international research papers. He is member of NEC, NEA, IEEE, ISCA speech & AEHIN website:

basanta@ioe.edu.np

Basanta Joshi
Tribhivan University, Nepal