Abstract

This paper presents a Neural Network based Nepali Speech Recognition model. RNN (Recurrent Neural Networks) is used for processing sequential audio data. CTC (Connectionist Temporal Classification) [1] technique is applied allowing RNN to train over audio data. CTC is a probabilistic approach of maximizing the occurrence probability of the desired labels from RNN output. After processing through RNN and CTC layers, Nepali text is obtained as output. This paper also defines a character set of 67 Nepali characters required for transcription of Nepali speech to text.

References


**Index Terms**

Computer Science

Signal Processing

**Keywords**
Artificial Intelligence, Machine Learning, Automatic Speech Recognition, Recurrent Neural Network, Connectionist Temporal Classification, Softmax, Hidden Markov Model, Nepali Speech Recognition, Long-Short Term Memory (LSTM), Backpropagation, Character Error Rate