

Accepted Manuscript

Machine learning based sample extraction for automatic speech recognition using dialectal Assamese speech

Swapna Agarwalla, Kandarpa Kumar Sarma

PII: S0893-6080(15)00279-8

DOI: <http://dx.doi.org/10.1016/j.neunet.2015.12.010>

Reference: NN 3573

To appear in: *Neural Networks*



Please cite this article as: Agarwalla, S., & Sarma, K. K. Machine learning based sample extraction for automatic speech recognition using dialectal Assamese speech. *Neural Networks* (2015), <http://dx.doi.org/10.1016/j.neunet.2015.12.010>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Machine Learning Based Sample Extraction for Automatic Speech Recognition using Dialectal Assamese Speech

Swapna Agarwalla^a, Kandarpa Kumar Sarma^b

^aDepartment of Electronics and Communication Engineering, Gauhati University, Guwahati, 781014, Assam, India

^bDepartment of Electronics and Communication Engineering, Gauhati University, Guwahati, 781014, Assam, India

Abstract

Automatic Speaker Recognition (ASR) and related issues are continuously evolving as inseparable elements of Human Computer Interaction (HCI). With assimilation of emerging concepts like big data and Internet of Things (IoT) as extended elements of HCI, ASR techniques are found to be passing through a paradigm shift. Oflate, learning based techniques have started to receive greater attention from research communities related to ASR owing to the fact that former possess natural ability to mimic biological behavior and that way aids ASR modeling and processing. The current learning based ASR techniques are found to be evolving further with incorporation of big data, IoT like concepts. Here, in this paper, we report certain approaches based on machine learning (ML) used for extraction of relevant samples from big data space and apply them for ASR using certain soft computing techniques for Assamese speech with dialectal variations. A class of ML techniques comprising of the basic Artificial Neural Network (ANN) in feedforward (FF) and Deep Neural Network (DNN) forms using raw speech, extracted features and frequency domain forms are considered. The Multi Layer Perceptron (MLP) is configured with inputs in several forms to learn class information obtained using clustering and manual labeling. DNNs are also used to extract specific sentence types. Initially, from a large storage, relevant samples are selected and assimilated. Next, a few conventional methods are used for feature extraction of a few selected types. The features comprise of both spectral and prosodic types. These are applied to Recurrent Neural Network (RNN) and Fully Focussed Time Delay Neural Network (FFTDNN) structures to evaluate their performance in recognizing mood, dialect, speaker and gender variations in dialectal Assamese speech. The system is tested under several background noise conditions by considering the recognition rates (obtained using confusion matrices and manually) and computation time. It is found that the proposed ML based sentence extraction techniques and the composite feature set used with RNN as classifier outperforms all other approaches. By using ANN in FF form as feature extractor, the performance of the system is evaluated and a comparison is made. Experimental results show that the application of big data samples have enhanced the learning of the ASR system. Further, the ANN based sample and feature extraction techniques are found to be efficient enough to enable application of ML techniques in big data aspects as part of ASR systems.

© 2015 Published by Elsevier Ltd.

Keywords: Automatic Speech Recognition (ASR), big data, Artificial Neural Network (ANN), Multi Layer Perceptron (MLP), Deep Neural Network (DNN), Recurrent Neural Network (RNN), Fully Focussed Time Delay Neural Network (FFTDNN)

1. Introduction

Speech based applications are one of the fastest means of communication as far as the field of Human Computer Interaction (HCI) is concerned. Present levels of interaction are oriented more towards natural aspects of speech

Email addresses: swapwiz16@gmail.com (Swapna Agarwalla), kandarpaks@gmail.com (Kandarpa Kumar Sarma)

Download English Version:

<https://daneshyari.com/en/article/6863220>

Download Persian Version:

<https://daneshyari.com/article/6863220>

[Daneshyari.com](https://daneshyari.com)