ABSTRACT

Pattern recognition is an important research area offering valuable solutions to the real world pattern identification problems. It is used to recognize and portray precisely the patterns hidden in data through statistical and machine learning approaches. Speech recognition is an application domain where in the pattern recognition techniques are being explored to find appropriate solutions. Speech recognition concentrates on developing systems that are capable enough to recognize or translate the speech. Many challenges like speaker variability, channel variability, channel charactersitics, noise, language variability, etc. are posed in the research concerned with speech recognition. This research titled "Graph Based Segmentation and Deep Learning for Phoneme Pattern Classification in Tamil Continuous Speech" focuses on modelling phoneme classifiers and building efficient models to recognize the phonetic units of Tamil continuous speech.

Continuous Speech Recognition (CSR) systems involve a large vocabulary which aids in intensive training and requires a huge dataset. These systems are capable of handling non-linear input data from different speakers. The greater complexity of the CSR systems requires an efficient generative modelling technique to solve the problem. With the advent and evolution of machine learning and pattern recognition, it has become possible to tackle complex tasks like speech recognition efficiently with affordability and tolerable time constraints. A wider range of problems are open in this area. This research takes the problem of constructing best acoustic models for the representing the phonetic units of the Tamil Language and thus support in improving the human machine interaction.

The main objective of this research is to develop a framework that uses deep learning models to efficiently recognize the phonemes in Tamil continuous speech. The overall objective is subdivided and listed out as a line of investigation objectives.

- To build a speech corpus for Tamil that includes the speech signals, its transcripts and phoneme database to help further research on Tamil phoneme pattern recognition.
- To develop an efficient and dynamic algorithm to automatically segment the Tamil continuous speech into phonetic segments.
- To design and develop a methodology to build an acoustic model for Tamil phoneme pattern classification using Deep Belief Networks performing with good classification accuracy.

- To design a framework that enables to build optimized acoustic models using powerful computational methods like PSO to optimize the acoustic model parameters for better phoneme pattern classification task.
- To enhance the efficiency of the acoustic model by addressing the data imbalance problem through dynamic loss function.

The effort to meet the objective of the research is proceeded by formulating the phoneme pattern classification problem in Tamil continuous speech as a pattern recognition problem. The methodology to build an acoustic model for phoneme classification in continuous speech involves filtering, feature extraction, identifying the phonetic segments in the speech, building phoneme datasets and acoustic model building. The model is built using supervised learning algorithm for training DBN and its variants which are designed to improve the performance of the models. The performances of the models are evaluated using standard performance metrics like precision, recall, F-measure, recognition accuracy, root mean square error, phoneme error rate, etc.

Kazhangiyam, a Tamil speech corpus is created due to lack of free benchmark Tamil speech corpus. The speech from 39 speakers in the age group from 18 to 40 years is collected. A set of forty five sentences are given to the speakers. Utterances of all the sentences by the speakers have been recorded using mobile phones and laptops in a controlled environment. The sentences spoken by the speakers are hand segmented and the phoneme database has been built.

A graphcut based segmentation algorithm using a statistical approach is proposed to automatically segment the phonetic units in the Tamil continuous speech. The input continuous speech is filtered and DWT features are extracted for each time unit. A multigraph is built with the feature vectors as nodes of the graph and the distance between the vectors representing the edges of the graph. The weight matrix representing the edges of the graph is then transformed to an eigenvalue problem. Appropriate eigenvectors are chosen to aid the graph cut process. The dataset Discrete Wavelet Transform Feature Set (DWTFS) is constructed from the segmented phonemes accounting to 6,67,260 instances contributing to 39 phoneme classes.

A pilot study is conducted by building two acoustic models, one using Artificial Neural Network (ANN) and other using Adaptive Neuro-Fuzzy Inference System (ANFIS). ANN acoustic model is built using backpropagation training. ANFIS acoustic model is built using backpropagation training to build the knowledge base of the phoneme classifier. The

dimensionality of the feature space is reduced with LDA. The performances of the ANN and ANFIS phoneme classifiers are studied.

A framework to build Deep Belief Network (DBN) based acoustic model is proposed. Two types of DBNs namely Bernoulli-Bernoulli DBN (BBDBN) and Gaussian-Bernoulli DBN (GBDBN) are explored to build phoneme pattern recognition model. The model building process involves contrastive divergence pretraining and backpropagation training. Experiments are conducted to study the performance of DBNs of various depths have been analyzed.

The contrastive divergence based pre-training of DBNs is replaced by Particle Swarm Optimization (PSO) based pre-training to reduce the time complexity and to drive the model with optimized set of parameters. The model is termed as PSO pretrained DBN. The optimal DBN is thus evolved through the generations of PSO. In addition to basic PSO, the use of its variants Second Generation PSO and New Method PSO in pretraining the DBNs are also explored. Once the optimal DBN is attained through PSO, the DBN is fine tuned using gradient based backpropagation training. The PSO-DBN acoustic models have been built and their performances are reported.

Temperature controlled PSO (TPSO), an improved PSO to handle the problem of stagnation of particles through generation observed in PSO and to help faster convergence of particles towards the optimized solution is proposed. In the proposed TPSO algorithm for parameter optimization, a new term, temperature is introduced in evaluating the velocity of the particles in the population during each generation. The PSO based pre-training undergone in the previous approaches is replaced with TPSO based pre-training. The TPSO-DBN acoustic model is trained and tested for its performance.

Finally, a loss function Weighted Mean Square Error (WMSE) is proposed to handle the influence of imbalanced dataset in training a model. The models built using an imbalanced dataset suffer from being dominated by the instances of majority classes. The proposed loss function is used while pre-training and training the DBN to solve the problem. The efficiency of various DBNs built with WMSE as loss function is experimented and analyzed. The contributions of this research are believed to add few solutions that help in the area of Tamil speech pattern recognition to enhance the usability and user friendliness.