Abstract

Speech information retrieval refers to the task of retrieving the relevant speech utterances to a user query, from a large collection of spoken archives. The main aim of this thesis is to present a novel approach to perform speech information retrieval based on the principle of pattern discovery and clustering of speech utterances. The key idea behind this proposed approach mainly focuses on finding the structure of repeating patterns from the raw speech utterances in an unsupervised manner. In this approach, we utilize the image processing, depth first search and machine learning techniques to find the matching acoustic patterns (keywords or phrases) between the speech utterances. We aggregate all matching acoustic patterns from the entire speech corpus, and segregate related speech utterances into broader domain-specific classes (clusters) using clustering techniques. In each cluster, speech utterances are indexed using unique keywords. This process of clustering the speech corpus helps in performing the retrieval task more efficiently by matching the user query with the unique keywords present in each domain.

A web-based user interface is developed with which a user is allowed to upload a spoken query. Once the user uploads the query, the matching process between the spoken query and the unique keywords is carried out by the unsupervised pattern discovery technique. The list of speech utterances associated with the matched unique keyword is retrieved and displayed to the user. In this thesis, novel methods for unsupervised pattern discovery are devised using image processing, depth first search and machine learning techniques. The proposed pattern discovery techniques are evaluated on Hindi and Bengali speech corpora. The obtained results by the proposed method are better in comparison to the state-of-the-art methods. Also, the performance of the developed speech retrieval technique is observed to be quite decent and accurate. This methodology can be generalized and extended to various real-world scenarios.

The major contributions of the thesis are summarized as follows :

1. A robust unsupervised pattern discovery method is proposed at frame level, based on the image processing techniques to determine the matched pairs of speech documents.

2. An unsupervised pattern discovery technique is proposed at phoneme/sound unit level using 3-neighbor depth-first search (3-NDFS) traversal method.

3. A convolutional neural network (CNN) based unsupervised classification model is proposed to enhance the accuracy of image processing method for unsupervised pattern discovery.

4. A method is devised to generate the indexed keywords and their associated speech utterances in each cluster.

5. A web-based user interface is developed to perform the speech information retrieval task.

Keywords: *Unsupervised Pattern Discovery, Clustering of Speech Utterances, Gaussian Mixture Model, Convolutional Neural Network, Dirichlet Process Gaussian Mixture Model, 3-Neighbor Depth First Search, Speech Information Retrieval, Extraction of Keywords, Indexed Keywords, Speech Indexing System*