International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 4, April 2015
Available at http://internationaljournalofresearch.org

Study of Prosodic Features in Languages Speech Signals

Niraj Kr. Singh
Research Scholar (M.Tech) Vivekananda Global University, Jaipur Jaipur, India
email.nirajsingh@gmail.com

Prof. Anoop Singh Poonia
Professor , E & C Engineering Vivekananda Global University, Jaipur Jaipur, India
anoopsingh_25@rediffmail.com

INTRODUCTION

An LID system can serve as a front end for multi-lingual translation software. An LID system can be trained once and then run on multiple machines simultaneously in order to correctly identify a particular language from a set of languages. Therefore, it is beneficial to go for automatic LID systems. Our model is made to classify Indian languages (Hindi, Assamese, Bengali and English) with the purpose to convert the speech waveform into a set of features or rather feature vectors for further analysis. Prosody is the part of speech where rhythm, stress, and intonation are reflected. In language identification tasks, these characteristics are assumed to be language dependent, and thus the language can be identified from them. In our project/model, an automatic language recognition system that extracts prosody information from speech has been designed.

CLASSIFICATION

The problem of language identification belongs to a much broader topic in scientific engineering called as pattern recognition. The goal of pattern recognition is to classify objects of interest into of a number of categories or classes. The objects of interest are generally called patterns and in this case are sequences of acoustic feature vectors that are extracted from an input speech using the techniques described in feature extraction section. The classes here refer to individual languages. Since the classification procedure in our case is applied on extracted features, it can be also referred to as feature matching. Classification is a method of representing the patterns of the different classes with some common classifiers. The classifiers so chosen should be such that they represent all the patterns of the different classes.

NEED FOR CLASSIFICATION IN LID

A normal human speech contains much redundant information such as noise and silence periods. The main purpose of the feature extraction process is to extract the most relevant information from the speech waveform and discard as much of the redundant information as possible. The relevant information is a set of feature vectors which contains prosodic features. To correctly identify a language, we need a large dataset of feature vectors. But such large datasets offer many computational problems while testing. So we represent the datasets by certain classifiers such as mean, variance and mixture weights. All the feature vectors of a particular language can be represented by a set of common classifiers. The classifiers should to be able to represent the wide range of feature vectors. Also the classifiers should model the underlying hidden features of a language. The set of classifiers is also referred to as a language model. A language model λ contains the classifiers such as mean, variances & weights [1]. The purpose of the back end of any system for automatic language identification (LID) is to train some form of model *λL* for each of the L languages to be recognized by the system.

THE BACK-END OF LID SYSTEM

The purpose of the back end of the LID system for automatic language identification is to train some models λL for each of the languages to be recognized by the system. Here, we have trained a single GMM for all the languages called the Universal Background Model (UBM) and then adapted a separate GMM for each language from the UBM. The UBM trained is considered to represent the characteristics of all the languages under consideration accurately. GMM stands for Gaussian Mixture Model. It is a parametric probability density function (pdf) represented as a weighted sum of Gaussian component densities. The complete GMM is parameterized by mean vectors of the cepstral information, their covariance matrices and the mixture weights of all the component densities.

The UBM is created using a portion of the training data from all the languages. The advantage of using UBM is that the quantity of training data required can be reduced. GMM for all the languages is adapted from the UBM using a probabilistic adaption procedure.UBM also reduces the model training time significantly as the models are being adapted. During testing, speech in unknown language is applied at the front-end where it is converted into a set of feature vectors. Then it is compared with each of the language models λL which is modelled using GMM-UBM.

In the identification phase, the same kind of speech feature is extracted from the unknown speech utterance in a particular language. The feature vector set thus obtained is then compared to the model set λL (L=1, 2, 3…) where L is the number of possible languages that the system is capable of identifying. The system must then determine which of the L languages is most likely related to the feature vector extracted from the speech in unknown language.

TYPES OF CLASSIFICATION

There are different types of classification used for classifying the input feature vectors. Some of the common approaches are as aforesaid below:

Gaussian Mixture Models (GMM): A Gaussian Mixture Model (GMM) is a parametric representation of a probability density function, based on a weighted sum of multi-variate Gaussian distributions. A Gaussian distribution can be completely described by its mean and variance. A GMM with K component densities (or mixtures) can be parameterised K mixture weights, K mean vectors and K covariance matrices [1-4].

GMM-UBM: The training phase of operation of this system occurs in two distinct stages. First a set of feature vectors taken from a number of different languages (typically data from all languages to be tested will be used) are used to train a single GMM. This GMM is referred to as the Universal Background Model (UBM) and is considered to represent the characteristics of all different languages. From the UBM, a GMM is then adapted for each of the languages in the system (using only data from that language) using Bayesian adaptation (maximum a-posteriori or MAP adaptation) [1, 5, 6].

Support Vector Machines (SVM) : Support Vector Machines (SVM) uses a linear kernel in a super-vector space for rapid computation of language distance scores. SVM classifiers are designed through an optimization process, which is discriminative in nature. In SVM classifier design, the kernel plays a central role [1, 7, 8].

Hidden Markov Models (HMM): A hidden Markov model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network. In a hidden Markov model, the state is not directly visible, but the output which is dependent on the state is visible [9, 26].

GMM-UBM Classification: Gaussian Mixture Model (GMM) is a generative model widely used in speaker verification. It represents the state-of-the-art in this field. This model was introduced and apphed for the first time in speaker verification in (Reynolds et Rose, 1995)(Reynolds et al., 2000). It is a semi-parametric probabilistic method that offers the advantage of adequately representing speech signal variability. Frequently, speaker verification systems based on GMMs are combined with other systems based on other types of models to improve their performance.

GMM-UBM classification has become one of the dominant techniques for acoustic based language identification [1]. The UBM is a large GMM trained to represent the language independent distribution of features. Specifically, we want to select speech that is reflective of the expected alternative speech to be encountered during recognition. This applies to both the type and the quality of speech, as well as the composition of the speakers speaking in a particular language. The GMM-UBM model is implemented in this project as the back-end of the LID system [1].

Expectation – Maximization (EM) Algorithm : The Expectation Maximization (EM) algorithm (Dempster et al., 1997) is used to learn the GMM parameters A = ***{wi,*** *p^*, S^) based on maximizing of the expected log-likelihood of the training data. In most speaker verification systems, we do not have enough data to train the speaker GMM using the EM algorithm. To overcome these difficulties, a speaker venfication system uses a GMM Universal Background Model (UBM), under the assumption that this model will adequately descnbe the underlying characteristics of a large speaker population. Generally, the UBM is trained on a large set of speakers, the identities of whom are different from the target speaker. The speaker GMM model is then derived from the UBM by Maximum *A Posteriori* (MAP) adaptation using the target speaker data.

CONCLUSION AND FUTURE SCOPE

Conclusion: Voice based biometric systems may prove to be the only feasible approach for remote access control. This novel approach is based on continuous approximations of the prosodic contours contained in a pseudo-syllabic segment of speech. Each of these contours is fitted to a Legendre polynomial, whose coefficients are modeled by a Gaussian mixture model. Prosodic information models the speaker’s speaking style. It is related to the pitch (vibration of the vocal cords), sound duration and the energy used to produce speech sounds. Using these prosodic features we design a automatic character recognition system.

Future scope of the project: This project has tremendous future scope. The field of automatic language identification is relatively new and it is progressing at a fast pace. Many new feature extraction and classification techniques have been developed (refer Appendix) which will increase the identification rate significantly. The LID system developed in this project can also be implemented with other feature extraction and classification techniques and a comparative study can be performed between them. Also other Indian languages such as Nepali, Tamil, Malayalam etc. can be included. More number of speakers in each language can be included in the existing database and it can be checked whether the system performs better for a larger database.

REFERENCE

[1] C. Pradeep, "Text dependent speaker recognition using MFCC and LBG VQ," National Institute of Technology, Rourkela, 2007.

[2] J. P. Campbell Jr., "Speaker recognition: a tutorial," in *Proc. IEEE*, vol. 85, issue 9,Sept. 1997.

[3] H. Seddik, A. Rahmouni and M. Samadhi, "Text independent speaker recognition using the mel frequency cepstral coefficients and a neural network classifier," in *Proc. First Int. Symp. Control, Communications and Signal Processing*, 2004, pp. 631 – 634.

[4] E. Ambikairajah, H. Li, L.Wang, B. Yin and V. Sethu, "Language identification: a tutorial," *IEEE Circuits and Systems Magazine*, Second Quarter, pp. 82 – 108,May 2011.

[5] D. Reynolds, "Gaussian mixture models," MIT Lincoln Laboratory, Massachusetts,2002.

[6] D. A. Reynolds and R. C. Rose, "Robust text independent speaker identification using Gaussian mixture models," *IEEE Trans. Speech Audio Processing*, vol.3, no.1, pp. 72 – 83, Jan. 1995.

[7] D. A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models," *Speech Commun*. , vol. 17, pp. 91 – 108, Mar. 1995.

[8] D. A. Reynolds, T. F. Quatieri and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 3, pp. 19 – 41,2000.

[9] Y. Xu, J. Yang and J. Chen, "Methods to improve Gaussian mixture model for language identification," in *Proc. Int. Conf. Measuring Technology and Mechatronics Automation*, 2010, pp. 656 – 659.

[10] A. Ziaei, S. M. Ahadi, H. Yeganeh and S. M. Mirrezaie, "Spoken language identification using a new sequence kernel-based SVM back-end classifier," *IEEE Digital Signal Processing Journal*, pp. 324 – 329, 2008.

[11] K. Markov and S. Nakamura, "Language identification with dynamic hidden Markov network," in *Proc. ICASSP*, 2008, pp. 4233 – 4236.

[12] Dustor and P. Szwarc, "Spoken language identification based on GMM models," in *Proc. Int. Conf. Signals and Electronic Systems (ICSES)*, Sept. 2010, pp. 105 – 108.