Université de Lorraine - IDMC
Masters SC & TAL – M1
2019-2020

Fiche de projet tutoré / Project form

Speaker Adaptation Techniques for Automatic Speech Recognition

Encadrement / Supervisors

1. équipe, laboratoire / team, lab :
   MULTISPEECH
2. encadrant·e principal·e (nom, email) / main supervisor (name, email)
   Tugtekin Turan (tugtekin.turan@inria.fr)
3. autres encadrant·es / other supervisors

Description / Description

In statistical speech recognition, there are usually mismatches between the conditions under which the model was trained and those of the input. Mismatches may occur because of differences between speakers, environmental noise, or differences in channels [1]. They should be compensated to obtain sufficient recognition performance. Therefore, acoustic model adaptation is the process of modifying the parameters of an acoustic model used for speech recognition to fit the actual acoustic characteristics by using a limited amount of utterances from the target users.

Speech recognition techniques using hidden Markov models (HMMs) have become significantly popular since the late 1980s. In particular, these algorithms often employ continuous density HMMs using triphones as recognition units and a Gaussian mixture distribution as the output distribution. In other words, Gaussian mixture models (GMM) represent the relationship between HMM states and the acoustic input. Over the last few years, advances in both machine learning algorithms and computer hardware have led to more efficient methods for training deep neural networks (DNNs). It has presented in many papers that DNNs can outperform GMMs at acoustic modeling for speech recognition on a variety of datasets including large datasets with large vocabularies [2].

For GMM models, speaker adaptation has proven to be effective in mitigating the effects of this mismatch. In general, it modifies speaker-independent (SI) models towards particular testing speakers or transforms the features of testing speakers towards the SI models [3]. Although displaying superior generalization ability than GMMs, DNN models still suffer from the mismatch between acoustic models and testing speakers. As is the case with GMMs, DNN models experience a degradation of accuracy when ported from training speakers to unseen testing speakers [4].

The use of utterances from many speakers for training enables these models to represent not only phonetic features but also speaker features. Although this ability has made SI systems practical, the systems still do not perform as well as speaker-dependent systems in

which the parameters are estimated from a sufficient amount of utterances from one target user [5]. This means that speaker adaptation techniques are very important for any recognition system. In this project, we mainly deal with speaker adaptation to optimize the performance by transforming SI models towards particular speakers or modifying the target features to match a pre-trained SI model based on a relatively small amount of adaptation data from the target speakers. We will analyze different techniques and make a comparison using state-of-the-art adaptation systems.

Informations diverses : matériel nécessaire, contexte de réalisation /
Various information: material, context of realization

The related corpus for adaptation is available in the MULTISPEECH team of the INRIA research center. We will be mainly using Librispeech which is large-scale corpus based on English audio books (http://www.openslr.org/12) and Verbmobil which contains multilingual dialogues (https://www.phonetik.uni-muenchen.de/Bas/BasVM2eng.html). Experiments will be performed over both Kaldi speech recognition toolkit (https://kaldi-asr.org) and PyTorch machine learning library (https://pytorch.org).

During this project, the students will,

- learn the basics of speech recognition systems
- analyze the spectral content which carries information
- understand the components of a speech signal
- be familiar with bash scripting and python programming
- implement machine learning solutions for speaker adaptation
- run several experiments and compare their results

Livrables et échéancier / Deliverable and schedule

December, 2019 : Cover the relevant literature and review basic methodology
January, 2020 : Learn how to use main tools and datasets
March, 2020 : Start programming experiments
May, 2020 : Finalize the project with presenting a report

Bibliographie /References

[1] Yamagishi, Junichi et al. "Analysis of Speaker Adaptation Algorithms for HMM-Based Speech Synthesis" IEEE Transactions on Audio, Speech, and Language Processing (2009).

[2] Hinton, Geoffrey et al. "Deep Neural Networks for Acoustic Modeling in Speech Recognition" IEEE Signal Processing Magazine 29 (2012).

[3] Leggetter, Christopher et al. "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models" Computer Speech & Language (1995).

[4] Gupta, Vishwa et al. "I-Vector-Based Speaker Adaptation of Deep Neural Networks for French Broadcast Audio Transcription." IEEE International Conference on Acoustics, Speech and Signal Processing (2014).

[5] Liu, Chaojun et al. "Investigations on Speaker Adaptation of LSTM RNN Models for Speech Recognition" IEEE International Conference on Acoustics, Speech and Signal Processing (2016).