Summary report for project

Exploiting Language Information for Situational Awareness (ELISA)

For year 2017

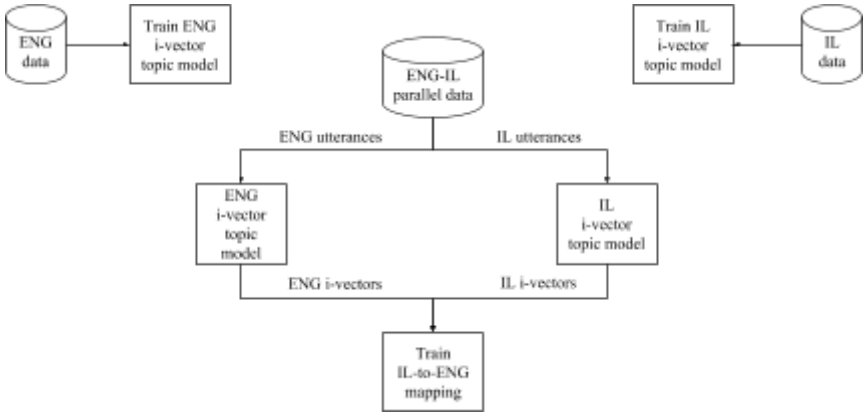
Submitted to Information Sciences Institute, University of South California, USA

By Brno University of Technology

Lead author: Dr. Ondřej Glembek

Cross lingual topic id in latent space

Initial experiments were conducted on a multi-lingual (EN, FR, DE, IT) corpus of political texts comprising of 7 topics and approximately 4400 samples per language. At this moment, there are no results yet. Further analysis is required to analyze model suitability for learning the transformation in the latent space.



Classification using uncertainties in the learned representations

We propose a model that learns the document / sentence embeddings along with the uncertainties. These learned uncertainties can be used in training a classifier. This would allow us to predict the confidence scores for target (topic) labels more reliably. Initial results on Fisher conversational telephone speech showed minor and consistent improvements (in classification accuracy) over baseline systems and major improvements in terms of cross-entropy.

Model evaluation without supervised test data

Problem Statement

The speech data provided during evaluation just consists of development (dev) and evaluation (eval) speech data without any text labels. Thus evaluating the pre-built acoustic model was difficult and had to rely on native informant to transcribe a small set of dev data.

Experimental Focus

To mitigate this issue, we build a simple autoencoder model with unsupervised training data and then pass the unsupervised test data to evaluate the mean square errors [1]. This metric helps to evaluate the acoustic data or model built using it and how well it correlates with the test data provided for evaluation. A simple extension of this approach is to use de-noising and variational auto-encoder as they are robust to new data conditions.

Progress / Plan

Initial experiments with de-noising autoencoders are done with standard TIMIT dataset. Further experiments will be done using Lorelei data from previous evaluations.

Adapting acoustic model to focus on evaluation data

Problem Statement

The acoustic model training done for previous evaluations used speech features after performing several normalization methods such as multilingual bottleneck feature extraction, speed/volume perturbations, noise perturbations and speaker normalization (VTLN). However the model performs poorly when it sees data from extremely new conditions.

Experimental Focus

To adapt the acoustic model parameters with test data before decoding. This is done in an unsupervised manner by modeling the un-certainities using Bayesian dropout technique and introduce it into the loss function [2].

Progress / Plan

The experimental setup is yet to be completed.

References

1. Sri Harish Mallidi, Tetsuji Ogawa, Hynek Hermansky, Uncertainty estimation of DNN classifiers, ASRU 2015
2. Yarin Gal, Zoubin Ghahramani, Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning