1 Introduction

• Multilingual pre-training has huge importance on low-resource data: how does it perform on various amounts of data?

• Incorporating context into BLSTM features.

• System complexity: multilingual pretraining should allow to train more complex architectures.

2 Data


3 GMM system

• Used to produce phoneme alignments for NN training.

• GMM features are based on multilingual Region Dependent Transform trained on 24 Babel languages (Year 1-4) (Released on http://speech.fit.vutbr.cz/software).

4 BLSTM systems

• Standard hybrid DNN-HMM acoustic models

• Feature extraction:

• Single: 24 log-mel-filter-bank + different pitch features FBMK_F0.

Contextual FBMK_F0 feature trajectories spanning 11 frames with Hamming window and Discrete cosine transform - 11FBANK_F0.

4.1 Multilingual architecture

• Trained on Y1-Y3 = 17 languages or Y1-Y4 = 24 languages.

• block-softmax output layer with context-independent phoneme state targets.

• Porting of multilingual models into target language:

1. The final multilingual layer is replaced with randomly initialized target-language initial layer.

2. Only this new layer is trained with a standard learning rate.

3. The whole NN is fine-tuned with low learning-rate.

4.2 Analysis of feature extraction

• Context information is advantageous for multilingual systems.

4.3 Training epochs

• Final multilingual NN should be taken around the first halving of learning rate (20th epoch for 17L, 19th for 24L NN).

• Well trained multilingual NN is suitable only if target language is part of multilingual training set.

4.4 Training data analysis

• More diverse data leads to better results.

5 Conclusion

• Analysis of improvement from multilingual approaches for large scale of training data - significant gain even for 270h of training data.

• Final multilingual NN should be taken around the first halving of learning rate (20th epoch for 17L, 19th for 24L NN).

• Well trained multilingual NN is suitable only if target language is part of multilingual training set.

• Additional gain with adding more parameters into systems.

Acknowledgment

This work was supported by Technology Agency of the Czech Republic project No. TA0411311 “MINT”, European Union’s Horizon 2020 project No. 645523 BISON and Czech Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU) project “IT4Innovations excellence in science” - LGI1602.