

Multilingual Speech Processing: Presentation of Progress

Training and Decoding

Yao Jiacheng, Tobias Pöppke | February 4, 2015

COGNITIVE SYSTEMS LAB

- Create new recognizer with 1 Gaussian
- Do k-means flatstart via `storeTokenSequenceForInit`
- Train the recognizer via `storeTokenSequenceForTrain`

Initialization of next recognizer

- Increase number of gaussians by 1
- Create new recognizer with new number of gaussians
- Initialize new recognizer with old recognizer
- Get paths via `forcedSequenceAlignment` from old recognizer
- Use `storePathForInit` on new recognizer

Training of next recognizer

- Get paths via `forcedSequenceAlignment` from old recognizer
- Use `storePathForTrain` on new recognizer
- Repeat training five times

- Initialize clustering with context of 2
- Using a maximum of 2000 leaves, minimum of 1000 samples and 5 Gaussians per model
- Use paths from the recognizer for `storePathForClustering`
- Run the clustering with `Recognizer.cluster`
- Train the clustered model with `storeTokenSequenceForTrain`
- Problem: segmentation fault when starting to train the model

- Model with 5 Gaussians
- Trained with 5 training iterations
- Average WER on development set: 65.24%
- Average WER on our recorded data set: 91.38%

- Initialize GaussianContainerSet, GaussMixturesSet, AtomManager, MixtureTree and TopologyInfo
- Using stored files from recognizer
- Register attribute handlers for FILLER, SILENCE and TOKEN_SCORE in dictionary
- Load n-gram language model via CacheTokenSequenceModel

- Language model and language model lookahead weight is 22.5
- Maximum language model lookahead tree depth is 12
- Create `SearchGraphHandler` and initialize it with needed values
- Create a `BioKIT Decoder`
- Use decoder to decode samples

- Extract the search results from the decoder
- Use `dictionary.getBaseForm` to get cleaned hypothesis
- Use `align.tokenErrorRateInsDelSubCount` on cleaned hypothesis and reference
- Compute $WER = \frac{Ins+Del+Sub}{Words\ in\ Reference}$
- To compute average WER update global insertions, deletions, substitutions and words