Assignment 1 — continuous speech recognition

Hiroshi Shimodaira and Steve Renals

Automatic Speech Recognition— ASR Assignment 1
5 February 2015

Outline of Assignment 1

- Task: Experiments of continuous speech recognition with HTK, especially find the optimal configuration of parameters.
- Corpus: WSJCAM0 Cambridge Read News
- Software tools to run HTK are provided (shell scripts)
- Experiments
  1. Using single Gaussian component monophone models, investigate the effect of three parameters: pruning level; language model scaling factor; and word insertion penalty.
  2. Training and recognition with multiple mixture component monophone models
  3. Training and recognition with decision-tree-based tied-state triphone models
- For details, see the assignment sheet.
- Drop-in lab sessions: see the course web page.
- Q&A page

Submission

- Deadline: Wednesday, 25th February 2015 at 16:00.
- Submission consists of
  1. Report (4-7pp) and coversheet in PDF or Word format via the DICE electronic submission system – "submit" command.
  2. The best ASR system that achieved the highest recognition accuracy. See the assignment sheet for details. Do not use the submit command for this.

Computing environment

- Informatics DICE
- HTK (Hidden Markov Model toolkit) version 4.3.1
- Work directory ('*WorkDir*' here after) that you should use for the assignment:

`/afs/inf.ed.ac.uk/group/teaching/asr/Work/`*YourLoginName*

where *YourLoginName* denotes your DICE login name.

- If you cannot login to any DICE workstations, please contact the Informatics computing support.
- If you find your *WorkDir*non-existent or non-accessible, contact me.
- If you wish to run experiments in your own laptop, speak to me.

Computing environment (*cont.*)

Initialisation of your environment (for the first time only)

```
% cd WorkDir
% /afs/inf.ed.ac.uk/group/teaching/asr/bin/init-asr.sh
```

You will find the following directories under *WorkDir*.

dir names	*contents*
corpus/	*original corpus (wsjcam0) and parameterised data (WSJCAM0)*
file_lists/	*lists of files for training and recognition*
labels/	*phone and word labels*
model_lists/	*list of HMM model names*
dictionaries/	*word lists and pronunciation dictionaries*
language_models/	*language models*
scripts/	*Shell(Bash) script files for training, recognition, and summarising*
python/	*Python version of the above*
configs/	*HTK configuration files*
edfiles/	*HTK edit command files*
logs/	*log files*
recognised/	*recognition output*
manuals/	*copy of the HTK book, reference papers*

Data (WSJCAM0)

- You don't need to parameterise it, as it's already converted to MFCCs, which can be found under `corpus/WSJCAM0`.

  Do not copy the data either, use the central copy (the scripts are set up to do this).
- There are about 8000 training utterances (from 90 speakers), 700 development utterances (from 20 speakers), and 20000 test ones (from 20 speakers).
- The corpus is divided into sub sets for experiments as follows:

training set	*si_tr*
development sets	*si_dt5a*, *si_dt5b*, *si_dt20a*, *si_dt20b*
evaluation sets	*si_et5a*, *si_et5b*, *si_et20a*, *si_et20b*

  "5": 5k, and "20":20k word vocabulary
- In this coursework, we will use si_dt5a and its subset si_dt5a-div3 to evaluation the recognition performance of the system. NB: The best performance should be reported for si_dt5a (not si_dt5a-div3).

HTK

- HTK consists of a set of commandline tools, which can be run in a terminal window.

HCopy	feature extraction
HCompV	estimation of global mean and gloval variance
HInit	parameter estimation based on Viterbi alignment
HRest	forward-backward parameter estimation
HERest	embedded training
HVite	Viterbi decoding
HResult	calculation of recognition accuracy

- Manuals
  See *WorkDir*/manuals

Typical training/test procedures and scripts

	Descriptions	Provided script files (HTK commands)
Step 0:	Preparation of speech data dictionary language model	
Step 1:	Model training	
(a)	model initialisation	`./scripts/prep_monophones` (HCompV)
(b)	initial training	`./scripts/init_monophones` (HInit,HRest)
(c)		`./scripts/merge_monophones`
(d)	embedded training	`./scripts/train_monophones_wo_sp` (HERest)
(e)	silence model refinement	`./scripts/mk_sp_model`
(f)	label realignment	`./scripts/align_mlf` (HVite)
(g)	embedded training	`./scripts/train_monophones` (HERest)
Step 2:	Recognition	`./scripts/recognise_with_monophones` (HVite)
Step 3:	Result analysis	`./scripts/show_results` (HResult)

Command tools (shell scripts)

- Most of the command tools (scripts) for HTK can be found in *WorkDir*/`scripts/`.
- They should be run in *WorkDir* as your current directory.
- They are write-protected, i.e. you cannot edit.
- To modify a script, you need to copy the original one to a new one of your own, so that you can edit the new one.
  E.g.

```
% cd scripts
% cp -p train_monophones my_train_monophones
% cd ..
```

- Useful examples for writing B-shell scripts can be found in *WorkDir*/ShellScriptExamples.

Report

- Report template files should be used. They can be found in the ASR course web page.
- Assessment of coursework:
  - quality of experiments/investigation
  - quality of report and discussions
- "Scientific reports" are expected. For higher marks, give good discussions based on both theories and experiments, and also try the optional items.
- Your coursework should be done solely by yourself – you cannot use reports/documents, programs, files, or experimental results of someone else (except the programs / scripts provided in the course). However, you may have discussions with your colleagues. For plagiarism / misconduct, please see:
  `http://www.inf.ed.ac.uk/teaching/plagiarism.html`

Submission of the best ASR system

- This part can be minimal as long as you use *WorkDir* and you keep all the relevant files (i.e. all the models you trained, and other files needed to run the system) there.
  If it is not the case, please contact the course lecturer for advise.
- Having all the relevant files stored in *WorkDir*, go to "scripts" directory under *WorkDir*, and create a recognition script whose file name is "`run_the_best_system`". It could be just a copy of your recognition script that gave the best result on the evaluation set "si_dt5a" (not "si_dt5a-div3").
- Make sure that HMM model directory and HVite options are set properly in the script so that it gives the same result (accuracy) as the one for the best system described in your report.