Automatic Speech Recognition 2018-19: Assignment

Hiroshi Shimodaira and Steve Renals (Ver. 1.0)

1 Outline

In this assignment, you will carry out various experiments of continuous word recognition on the TIMIT speech data set and own recordings using the Kaldi automatic speech recognition toolkit. The purposes of the assignment is to learn basic techniques for continuous speech recognition, and familiarise yourself with Kaldi's commands and shell scripts so that you can write scripts of your own to run experiments.

You should submit a report online, and make your ASR systems and audio recordings available in you work directory ("*WorkDir*" hereafter) allocated to you in the course so that the marker can check your work (e.g. code and models) if necessary. Marks will be given to the information provided in the report, and not to the systems developed. Your systems will be considered as the evidence of experiments, and therefore tasks without corresponding systems will not be marked.

Regarding the recordings of your own voice, you will be provided with a list of 20 utterances to record. To carry out optional tasks, you will be able to access recordings of other students who have agreed to share their data with others for the coursework. To that end, you are invited to upload your audio recordings to a specified directory by Wednesday, 27th February.

Working in pairs

This assignment is intended to be done in pairs: by working with another student, you can discuss ideas and work things out together. Ideally, try to find a partner with a different skill set or degree programme to your own, although this may not be possible in all cases.

You may discuss any aspects of the assignment with your partner and divide up the tasks however you wish; but we encourage you to collaborate on each part rather than doing a strict division of tasks, as this will enable better learning for both of you.

You may also discuss high-level concepts and general programming questions with others in the class; however you may NOT share code, designs and results of experiments, or coursework reports directly with other groups.

For plagiarism/misconduct, please see:
`http://web.inf.ed.ac.uk/infweb/admin/policies/academic-misconduct`

Note that you are required to take reasonable measures to protect your assessed work from unauthorised access. For example, if you put any such work on a public repository then you must set access permissions appropriately (permitting access only to yourself, or your group in the case of group practicals like this one).

Once you have identified your partner, please indicate the information in a file 'partner' in your *WorkDir*, which can be done by running the following command in your *WorkDir*.

```
echo Partner's_UUN > partner
```

(NB: Replace *Partner's_UUN* above with the actual UUN of your partner.)

2 Coursework submission

The submission deadline is Wednesday, 20th March 2019 at 16:00. Your coursework submission is complete only if (i) you submit your report and (ii) make all your ASR systems available in *WorkDir*. For late coursework submission, see the following document:
`http://web.inf.ed.ac.uk/infweb/student-services/ito/admin/coursework-projects/late-coursework-extension-requests`

Submission of report

Only one student in your pair needs to submit online. Make sure both student's UUNs (e.g. s1234567) are clearly shown at the top the report. Please do NOT include your names, only UUNs.

- You report should be either a PDF (.pdf) or MS Word (.doc(x)) document, with a double-column format.
- Submit your report file with the "submit" command on a DICE workstation. The following shows an example of submitting a file, "report.pdf".

  ```
  submit asr cw1 report.pdf
  ```

- You should receive an email of acknowledgement from the system as soon as your submission has been received successfully. Keep the message as an evidence of your coursework submission.

Submission of ASR systems and own recordings

The ASR systems (including scripts and models) you used for the assignment should be found in *WorkDir* of the student who submit the report. The audio recordings of your own [1] should be found in your *WorkDir*/`MyAudio`. After the submission deadline, your *WorkDir* will be locked so that no further changes will be possible.

All the scripts you created/modified should be found in *WorkDir*/`my-local`. You should create the directory by yourself. In each task shown in 3.2 below, "[Scripts]" specifies the file name(s) of script you should put in the directory. For example, for Task 1.1, you should put `exp_task1_1.sh` – a script you used to explore different numbers of Gaussian mixture components, which should include all necessary steps of training and evaluation. You can call other (your) scripts from the script to avoid clutter. Task 1.1 also requires you to put `run_task1_1_best.sh` – a script to run a recognition evaluation (decoding, scoring, and displaying WER) with the best model you found. It should not include training steps.

Due to a limited disk space for *WorkDir*, please keep only the models that gave the best performance in the corresponding task, and delete other models and irrelevant files as soon as you finished the task.

3 Assignment specifications

3.1 Data sets

In addition to the TIMIT speech corpus we used in the labs, you will use a small data set of your own voice, ASR19_OWN. Optionally, you will be able to use ASR19_ALL - a collection of ASR19_OWN of the students who consent to sharing their recordings with other students for the coursework.

If you work in a pair, ASR19_OWN represents two sets of own recordings, one is of your own, and the other is of your partner. Please use ASR19_OWN-1 and ASR19_OWN-2 to distinguish them, clarifying which one represents whose.

For details about the recording and consent forms, please follow the instructions:
/afs/inf.ed.ac.uk/group/teaching/asr/doc/asr2019_cw_recording_instructions.pdf.

[1] In case you are unable to record your own voice or you do not wish to do so, please contact Hiroshi Shimodaira as soon as possible.

3.2 Tasks

You will carry out continuous *word* recognition experiments using the Kaldi automatic speech recognition toolkit provided in the course. There are three Tasks - Task 1 and Task 2 concern HMM-based systems only, whereas Task 3 allows you to consider not only HMM-based systems, but also HMM-DNN hybrid systems. Recognition performance should be measured with word error rate (WER) on a test set. To get higher marks, you will need to consider not only WER, but also other measures such as log likelihoods on training/test sets and run time, and give decent discussions.

It should be noted that evaluation experiments in this assignment are not formal, because we use a test set rather than a validation set to seek optimal configurations of parameters.

In the following, square brackets, [ ], are used to denote allocated marks or data sets.

Task 1 Monophone models with HMMs [40 marks]

1.1 Investigate how the number of Gaussian mixture components influences WER, and find the optimal number that gives the lowest WER. You should present a graph which summarises the result of your experiment. [TIMIT] [15 marks]

[Scripts] `exp_task1_1.sh, run_task1_1_best.sh`

1.2 Investigate how the dynamic features (i.e. delta and delta delta features) of MFCCs and CMN/CVN influence WER. Note that the sample scripts used in the labs employ dynamic features as default. [TIMIT] [15 marks]

[Scripts] `exp_task1_2.sh`

1.3 Using the best system you found through Tasks 1.1 and 1.2, carry out a speech recognition experiment on your own recordings. [TIMIT(training), ASR19_OWN(test)] [10 marks]

[Scripts] `run_task1_3.sh`

Task 2 Tied-state triphone models with HMMs [30 marks]

2.1 Investigate how the number of clusters and the number of Gaussian mixture components influence WER, and seek the optimal configuration of parameters and models that gives the lowest WER. It is acceptable that you just seek a local optimum rather than the global one. [TIMIT] [20 marks]

[Scripts] `exp_task2_1.sh, run_task2_1_best.sh`

2.2 Using the best configuration you found in Task 2.1, carry out a speech recognition experiment on your own recordings. [TIMIT(training), ASR19_OWN(test)] [10 marks]

[Scripts] `run_task2_2.sh`

Task 3 Advanced tasks [30 marks]

Further to the tasks described above, define tasks by yourself and carry out investigation. The following are examples. Marks shown below are for reference only - actual marks will vary depending on the contents and quality of experiments and discussions. Additional marks will be given when you go beyond HMM-based systems and the TIMIT data set.

- Develop gender dependent acoustic models and carry out recognition experiments. [10 marks]
- Investigate feature transformation and speaker adaptive training to improve WER. [15 marks]

- Investigate speaker adaptation. [15 marks]
- Investigate the above topic(s) with HMM-DNN hybrid systems [+10 marks]

  [Scripts] `e.g.  exp_task3_speaker_adaptation.sh`

(NB: Script file names for Task 3 can be arbitrary, please indicate them in the report)

Tips on experiments

- You cannot expect a good WER even for a sophisticated system due to the small size of TIMIT training data and mismatch of language model. It is not surprising that you get a WER of around 40% with an HMM-DNN hybrid system.
- You will find there is a large number of possible combinations or a large range of parameters you might need to explore. It will not be feasible to try all the possible combinations / ranges with a fine resolution. Try a coarse resolution first, which would give you some idea as to what/how you should try next.
- Please keep only the models that gave the best performance in the corresponding task, and delete other models and irrelevant files as soon as you finished the task. (See below to see why this is important)
- Due to a limited disk space available, please keep the total disk usage of *WorkDir* less than 4GB, or the total usage of your pair less than 8GB. You can check your disk usage by running the following command in your *WorkDir*.

  ```
  du -hs .
  ```

  Note that in case of having the disk space run out, you will be no longer able to write files – current/further results of experiments will be lost.

3.3 Report

Keep the length of your report between 4 and 7 pages (with a double-column format) including figures, tables, and references. Results of experiments should be well summarised using figures or tables. You should not only show the results, but also explain your experiments and results, and give discussions. Marking will be done based on the contents and quality of experiments, presentation, and discussions, which will count approximately 50%, 20% and 30% of the allocated marks, respectively.

Please read the following instructions and write a “scientific report”. For higher marks, you need to give good discussions based on both theories and experiments.

- Results of experiments should be efficiently summarised using figures or tables (but not both if possible) - avoid using a separate graph/table for each experiment.
- Figures/tables should be numbered and captioned.
- Conditions of experiments and methods that are different from those in original scripts should be shown clearly and concisely - consider using a table for example. Providing sufficient information is essential in scientific reports so that other people could redo the same experiments.
- Show results even if there was no improvement. It is important to discuss/analyse why there was no improvement.