Report

Note: This document is divied into 2 parts: the report itself containning informations about the experimentation and also the a log part containg the HTK command used in order to train the model.

The main object objectif of the project is to build a digit recognation system based on HTK toolkit. We will use 2 datasets to train a model (`train`) and to evaluate his performance (`dev`).

Project structure

Before jumping into the explaination of the experiment, it's important to understand the code-structure of the project:

```
.
├── Configs                     # Configurations of HTK commands (.config)
├── Data
|   ├── Recorded                # Initial datasets
|   |   ├── dev
|   |   ├── test
|   |   └── train
|   └── Lab                     # Feature extracted for each dataset
|       ├── dev
|       ├── test
|       └── train
├── Dictionary
|   └── DictionarySources       # Inital data to build phones lists and
dictionary
|       ├── dict
|       ├── grammar
|       ├── grammar.wordnet
|       └── words.list
├── Labels                      # Transriptions folder (.mlf file)
├── Mappings                    # Mapping folder (.mapping) for
├── Labels                      # Transriptions folder (.mlf file)
├── Models                      # Generated models
|   ├── hmm0
|   ├── hmm13
|   ├── ...
|   └── prototype               # Model template
├ provision                     # Virtual machine provisionning script
├ ASSIGNMENT
├ README.md
├ REPORT.md
├ start-htk.sh                  # Script to execute trainning and testing
└ Vagrantfile
```

Experiement approach

Each step of the experiment are described in the Log section of this document. For more details conserning the overall pipeline refer to the `start-htk.sh` script in the root dicrectory of the project.

- In order to build the digit recognation system, we need first to **prepare the data** for both training and testing. We need first to create a grammar for the system, wich will represent the chain of a spoken digits beginning and ending with a silence `sil`. We will use also the Voxforge dictionary, this dictionnary will allow to create a list of monophones representing each digit. We will have also to extract the features of each recording from

the `train` and `dev` datasets, using the `HCompV` command.

- In a second time, we will train a **monophone** Hidden Markov Model. The model will be re-estimate after introducing short pauses into the transcription in order to increase it's accuracy. We will also use the `HVite` command to align the phones with the training data.
- Finally, we will convert monophones into **triphones** to introduce context awareness into our existing model. By doing this, each monophones will be analysed in the context of the previous and the next one.

Once the model is trained, it's possible to evaluate his preformance against the `dev` dataset.

Results

Model	Model identifier	Correctness
Monphones Aligned	*hmm7*	59.30%
Triphones	*hmm12*	62.88%
Triphones Tied	*hmm15*	63.64%

However it should be noticed that when converting the monophone model to the triphone one, `HLEd` command show 2 warings:

```
WARNING [-2631]  ApplyTie: Macro T_sp has nothing to tie of type t in HHEd
WARNING [-2631]  ApplyTie: Macro T_sil has nothing to tie of type t in HHEd
```

This could indicate and issue this the configuration of the model concerning silence and short pauses. We can notice that the transcription generated by the final model (`hmm15`) on the `dev` dataset don't produce short pauses.

If we delete all the `sp` from the `dev.ref.mlf`, we obtain a better correctness.

Model	Model identifier	Correctness
Triphones Tied on `dev.nosp.mlf`	*hmm15*	90.06%

Log

Step-by-step explanation of the experimentation. For more details `start-htk.sh` contains all

the HTK commands described in this document and also the other bash commands.

Data preparation

Create the dictionary and the list of phones composing each words.

```
HDMan -A -T 1 -m -w Dictionary/Src/word.list -n Dictionary/phones-with-sp.list -
g ./Configs/global.ded -l dlog Dictionary/phones.dict
```

Convert the grammar file into a wordnet.

```
HParse -A -T 1 Dictionary/Src/grammar Dictionary/Src/grammar.wordnet
```

Transcribe the training `train.nosp.mlf` into a phones transcript with and without `sp`.

```
HLEd -A -d Dictionary/phones.dict -i Labels/train.phones.mlf -l ./Data/Lab/train
Configs/HLEd.config Labels/train.nosp.mlf
HLEd -A -d Dictionary/phones.dict -i Labels/train.phones-with-sp.mlf -l
./Data/Lab/train Configs/HLEd-with-sp.config Labels/train.nosp.mlf
```

Extract features from audio files for the `train` dataset.

```
HCopy -A -D -C Configs/HCopy.config -S Mappings/HCopy_train.mapping
HTK Configuration Parameters[20]
  Module/Tool     Parameter                  Value
  #HWAVE          BYTEORDER                 NONVAX
                  NATURALREADORDER             FALSE
                  LOPASS                        64
                  HIFREQ                      4000
                  LOFREQ                        64
                  SOURCERATE           1250.000000
                  NUMCEPS                       12
                  CEPLIFTER                     22
                  NUMCHANS                      23
                  PREEMCOEF               0.970000
                  ZMEANSOURCE                FALSE
                  ENORMALISE                 FALSE
                  USEHAMMING                  TRUE
                  WINDOWSIZE          250000.000000
                  SAVEWITHCRC                FALSE
                  SAVECOMPRESSED             FALSE
                  TARGETRATE          100000.000000
  #HNET           TRACE                         1
                  SOURCEFORMAT              NOHEAD
                  TARGETKIND              MFCC_E_D_A
```

Extract features from audio files for the `dev` dataset.

```
HCopy -A -D -C Configs/HCopy.config -S Mappings/HCopy_dev.mapping
HTK Configuration Parameters[20]
  Module/Tool     Parameter                 Value
  #HWAVE          BYTEORDER                NONVAX
                  NATURALREADORDER            FALSE
                  LOPASS                         64
                  HIFREQ                       4000
                  LOFREQ                         64
                  SOURCERATE            1250.000000
                  NUMCEPS                        12
                  CEPLIFTER                      22
                  NUMCHANS                       23
                  PREEMCOEF                0.970000
                  ZMEANSOURCE                FALSE
                  ENORMALISE                 FALSE
                  USEHAMMING                  TRUE
                  WINDOWSIZE          250000.000000
                  SAVEWITHCRC                FALSE
                  SAVECOMPRESSED             FALSE
                  TARGETRATE          100000.000000
  #HNET           TRACE                          1
                  SOURCEFORMAT              NOHEAD
                  TARGETKIND             MFCC_E_D_A
```

Model training

Note: In the training part we will use the same parameters for `HCompV`, `HERest`, `HVite`

```
HTK Configuration Parameters[5]
  Module/Tool     Parameter                 Value
                  NATURALREADORDER            FALSE
# HNET            TRACE                          2
                  ACCWINDOW                      2
                  DELTAWINDOW                    3
                  TARGETKIND             MFCC_E_D_A
```

1. Monohpones

Train the first model in `hmm0` using the `prototype` as **initial model**.

```
HCompV -A -D -C Configs/HCompV.config -f 0.01 -m -S Mappings/HCompV.mapping -M
Models/hmm0 Models/prototype
```

Reestimate the model the model using `HERest` command 3 times.

```
HERest -A -D -C Configs/HERest.config -I Labels/train.phones.mlf -t 250.0 150.0
30000.0 -S Mappings/HERest.mapping -H Models/hmm0/macros -H Models/hmm0/hmmdefs
-M Models/hmm1 Dictionary/phones.list
HERest -A -D -C Configs/HERest.config -I Labels/train.phones.mlf -t 250.0 150.0
30000.0 -S Mappings/HERest.mapping -H Models/hmm1/macros -H Models/hmm1/hmmdefs
-M Models/hmm2 Dictionary/phones.list
HERest -A -D -C Configs/HERest.config -I Labels/train.phones.mlf -t 250.0 150.0
30000.0 -S Mappings/HERest.mapping -H Models/hmm2/macros -H Models/hmm2/hmmdefs
-M Models/hmm3 Dictionary/phones.list
```

Fix the silence by adding into the existing model the short pause (`sp`).

```
HHEd -A -H Models/hmm4/macros -H Models/hmm4/hmmdefs -M Models/hmm5
Configs/HHEd.config Dictionary/phones-with-sp.list
```

Reestimate the model twice using the transcription containing short pauses.

```
HERest -A -D -C Configs/HERest.config -I Labels/train.phones-with-sp.mlf -t
250.0 150.0 30000.0 -S Mappings/HERest.mapping -H Models/hmm5/macros -H
Models/hmm5/hmmdefs -M Models/hmm6 Dictionary/phones-with-sp.list
HERest -A -D -C Configs/HERest.config -I Labels/train.phones-with-sp.mlf -t
250.0 150.0 30000.0 -S Mappings/HERest.mapping -H Models/hmm6/macros -H
Models/hmm6/hmmdefs -M Models/hmm7 Dictionary/phones-with-sp.list
```

Re-align the model and create a new transcription.

```
HVite -A -D -l ./Data/Lab/train -b sil -o SWT -C Configs/HVite.config -H
Models/hmm7/macros -H Models/hmm7/hmmdefs -i Labels/aligned.mlf -m -t 250.0
150.0 1000.0 -y lab -a -I Labels/train.nosp.mlf -S Mappings/HVite_align.mapping
Dictionary/phones.dict Dictionary/phones-with-sp.list
```

Retraining the model using the generated aligned transcription.

```
HERest -A -D -C Configs/HERest.config -I Labels/aligned.mlf -t 250.0 150.0
30000.0 -S Mappings/HERest.mapping -H Models/hmm7/macros -H Models/hmm7/hmmdefs
-M Models/hmm8 Dictionary/phones-with-sp.list
HERest -A -D -C Configs/HERest.config -I Labels/aligned.mlf -t 250.0 150.0
30000.0 -S Mappings/HERest.mapping -H Models/hmm8/macros -H Models/hmm8/hmmdefs
-M Models/hmm9 Dictionary/phones-with-sp.list
```

2. Triphones

Since so far, each phones are analyzed independently, the existing model (`hmm9`) is not enough accurate. In order to make it more robust in this part we will consider a triphone model. We will consider each phone in his *context*, the one before and the one after.

Conversion of the existing phone model to a triphone one.

```
HLEd -A -D -n Dictionary/triphones.list -l ./Data/Lab/train -i
Labels/triphones.mlf Configs/HLEd-triphone.config Labels/aligned.mlf
HHEd -A -H Models/hmm9/macros -H Models/hmm9/hmmdefs -M Models/hmm10 mktri.hed
Dictionary/phones-with-sp.list
 WARNING [-2631]  ApplyTie: Macro T_sp has nothing to tie of type t in HHEd
 WARNING [-2631]  ApplyTie: Macro T_sil has nothing to tie of type t in HHEd
```

Reestimate the model using the triphone model.

```
HERest -A -D -C Configs/HERest.config -I Labels/triphones.mlf -t 250.0 150.0
30000.0 -S Mappings/HERest.mapping -H Models/hmm10/macros -H
Models/hmm10/hmmdefs -M Models/hmm11 Dictionary/triphones.list
HERest -A -D -C Configs/HERest.config -I Labels/triphones.mlf -t 250.0 150.0
30000.0 -s stats -S Mappings/HERest.mapping -H Models/hmm11/macros -H
Models/hmm11/hmmdefs -M Models/hmm12 Dictionary/triphones.list
```

Making the triphone tied.

```
HDMan -A -D -b sp -n Dictionary/fulllist.list -g Config/global.ded -l flog
Dictionary/tri.dict Dictionary/Src/dict_fixed
HHEd -A -H Models/hmm12/macros -H Models/hmm12/hmmdefs -M Models/hmm13
Configs/tree.hed Dictionary/triphones.list
```

Reestimate 2 more times the model.

```
HERest -A -D -C Configs/HERest.config -I Labels/triphones.mlf -t 250.0 150.0
30000.0 -s stats -S Mappings/HERest.mapping -H Models/hmm13/macros -H
Models/hmm13/hmmdefs -M Models/hmm14 Dictionary/tiedlist.list
HERest -A -D -C Configs/HERest.config -I Labels/triphones.mlf -t 250.0 150.0
30000.0 -s stats -S Mappings/HERest.mapping -H Models/hmm14/macros -H
Models/hmm14/hmmdefs -M Models/hmm15 Dictionary/tiedlist.list
```

Tada!!! The model has been trained!

Testing

Regenerate the transcript of the `dev` dataset using the trained model.

```
HVite -A -D -H Models/hmm15/macros -H Models/hmm15/hmmdefs -S
Mappings/HVite.mapping -i Labels/aligned_15.mlf -w
Dictionary/Src/grammar.wordnet -p 0.0 -s 5.0 Dictionary/phones.dict
Dictionary/tiedlist.list
```

Run the tests

```
HResults -I Labels/dev.ref.mlf Dictionary/tiedlist.list Labels/aligned_15.mlf
```