List of modifications made in HTS (for ver. 2.1)

Heiga Zen

May 23, 2008

1 Model definition

In HTS, the HTK HMM definition (please see HTKBook Chapter 7) has been modified to support MSD, stream-level tying, and adaptation of multi-stream HMMs. This section gives its brief description.

First, <MSDInfo> is added to global options of the HTK HMM definition language The arguments to the <MSDInfo> option are the number of streams (default 1) and then for each stream, 0 (non-MSD stream) or 1 (MSD stream) of that stream. The full set of global options in HTS is given below.

```
globalOpts  =  option { option }
option      =  <HmmSetId> string |
               <StreamInfo> short { short } |
               <MSDInfo> short { short } |
               <VecSize> short |
               <ProjSize> short |
               <InputXform> inputXform |
               <ParentXform> ~a macro |
               covkind |
               durkind |
               parmkind
```

Second, the number of mixture specification is modified to support stream-level tying structure as follows:

```
HTK
 <State> 2
  <NumMixes> 1 2
  <SWeights> 2 0.9 1.1
  <Stream> 1

     <Mixture> 1 1.0
      <Mean> 4
        0.3 0.2 0.1 0.0
      <Variance> 4
        0.5 0.4 0.3 0.2
    <Stream> 2

     <Mixture> 1 0.4
      <Mean> 2
        1.0 2.0
      <Variance> 2
        4.0 8.0
     <Mixture> 2 0.6
      <Mean> 2
        2.0 9.0
      <Variance> 2
        3.0 6.0
```

```
HTS
 <State> 2

  <SWeights> 2 0.9 1.1
  <Stream> 1
     <NumMixes> 1
     <Mixture> 1 1.0
      <Mean> 4
        0.3 0.2 0.1 0.0
      <Variance> 4
        0.5 0.4 0.3 0.2
    <Stream> 2
     <NumMixes> 2
     <Mixture> 1 0.4
      <Mean> 2
        1.0 2.0
      <Variance> 2
        4.0 8.0
     <Mixture> 2 0.6
      <Mean> 2
        2.0 9.0
    <Variance> 2
        3.0 6.0
```

As you can see, <NumMixes> is moved from state-level to stream-level. This modification enables us to include the number of mixture component in the stream-level macro. Based on this implementation, stream-level macro was added. The various distinct points in the hierarchy of HMM parameters which can be tied in HTS is as follows:

∼s	shared state distribution
∼p	shared stream
∼m	shared Gaussian mixture component
∼u	shared mean vector
∼v	shared diagonal variance vector
∼i	shared inverse full covariance matrix
∼c	shared Cholesky U matrix
∼x	shared arbitrary transform matrix
∼t	shared transition matrix
∼d	shared duration parameters
∼w	shared stream weight vector

Note that the ∼p macro is used by the HMM editor HHEd for building tied mixture systems in the original HTK macro definition.

The resultant state definition of in the modified HTK HMM definition language is as follows:

```
state       =  <State> short stateinfo
stateinfo   =  ∼s macro |
               [ weights ] stream { stream } [ duration ]
macro       =  string
weights     =  ∼w macro | <SWeights> short vector
vector      =  float { float }
stream      =  [ <Stream> short ] streaminfo
streaminfo  =  ∼p macro | [mixes] (mixture { mixture } | tmixpdf | discpdf)
mixes       =  <NumMixes> short {short}
tmixpdf     =  <TMix> macro weightList
weightList  =  repShort { repShort }
repShort    =  short [ * char ]
discpdf     =  <DProb> weightList
mixture     =  [ <Mixture> short float ] mixpdf
mixpdf      =  ∼m macro | mean cov [ <GConst> float ]
mean        =  ∼u macro | <Mean> short vector
cov         =  var | inv | xform
var         =  ∼v macro | <Variance> short vector
inv         =  ∼i macro |
               (<InvCovar> | <LLTCovar>) short tmatrix
xform       =  ∼x macro | <Xform> short short matrix
matrix      =  float {float}
tmatrix     =  matrix
```

Third, to support multi-stream HMM adaptation, the HTK HMM definition language for baseclasses is modified. A baseclass is defined as

```
baseClass   =  ∼b macro baseopts classes
baseopts    =  <MMFIdMask> string <Parameters> baseKind [<StreamInfo>] <NumClasses> int
StreamInfo  =  short { short } |
baseKind    =  MIXBASE | MEANBASE | COVBASE
classes     =  <Class> int itemlist { classes }
```

where <StreamInfo> is optionally added to specify the stream structure.

2 Added configuration variables

A number of configuration variables have been added to HTK to control new functions implemented in HTS. Their names, default values, and brief descriptions are as follows:

Module	Name	Default	Description
HADAPT	SAVEFULLC	F	Save transformed model set in full covariance form
	USESMAP	F	Use structural MAP criterion
	SMAPSIGMA	1.0	Prior parameter for SMAP criterion
	BANDWIDTH		Bandwidth of transformation matrices
	DURUSEBIAS	F	Specify a bias with linear transforms
	DURSPLITTHRESH	1000.0	Minimum occupancy to generate a transform for state duration model set
	DURTRANSKIND	MLLRMEAN	Transformation kind
	DURBLOCKSIZE	full	Block structure of transform for state duration model set
	DURBANDWIDTH		Bandwidth of transformation matrices for state duration model set
	DURBASECLASS	global	Macroname of baseclass for state duration model set
	DURREGTREE		Macroname of regression tree for state duration model set
	DURADAPTKIND	BASE	Use regression tree or base classes to adapt state duration model set
HFB	MAXSTDDEVCOEF	10	Maximum duration to be evaluated
	MINDUR	5	Minimum duration to be evaluated
HMAP	APPLYVFLOOR	T	Apply variance floor to model set
HGEN	MAXEMITER	20	Maximum # of EM iterations
	EMEPSILON	1.0E-4	Convergence factor for EM iteration
	RNDPARMEAN	0.0	Mean of Gaussian noise for random generation
	RNDPARVAR	1.0	Variance of Gaussian noise for random generation
	USEGV	F	Use speech parameter generation algorithm considering GV
	CDGV	F	Use context-dependent GV model set

Module	Name	Default	Description
	LOGGV	F	Use logarithmic GV instead of linear GV
	MAXGVITER	F	Max iterations in the speech parameter generation considering GV
	GVEPSILON	1.0E-4	Convergence factor for GV iteration
	MINEUCNORM	1.0E-2	Minimum Euclid norm of a gradient vector
	STEPINIT	1.0	Initial step size
	STEPDEC	0.5	Step size deceleration factor
	STEPINC	1.2	Step size acceleration factor
	HMMWEIGHT	1.0	Weight for HMM output prob
	GVWEIGHT	1.0	Weight for GV output prob
	OPTKIND	NEWTON	Optimization method
	RNDFLAGS		Random generation flag
	GVMODELMMF		GV MMF file
	GVHMMLIST		GV model list
	GVMODELDIR		Dir containing GV models
	GVMODELEXT		Ext to be used with above Dir
	GVOFFMODEL		Model names to be excluded from GV calculation
HMODEL	IGNOREVALUE	-1.0E+10	Ignore value to indicate zero-dimensional space in multi-space probability distribution
HCOMPV	NSHOWELEM	12	# of vector elements to be shows
	VFLOORSCALE	0.0	variance flooring scale
	VFLOORSCALESTR		variance flooring scale vector for streams
HEREST	APPLYVFLOOR	T	Apply variance floor to model set
	DURMINVAR	0.0	Minimum variance floor for state duration model set
	DURVARFLOORPERCENTILE	0	Maximum number of Gaussian components (as the percentage of the total Gaussian components in the system) to undergo variance floor for state duration model set
	APPLYDURVARFLOOR	T	Apply variance floor to state duration model set
	DURMAPTAU	0.0	MAP tau for state duration model set
	ALIGNDURMMF		State duration MMF file for alignment (2-model reest)
	ALIGNDURLIST		State duration model list for alignment (2-model reest)

Module	Name	Default	Description
	ALIGNDURDIR		Dir containing state duration models for alignment (2-model reest)
	ALIGNDUREXT		Ext to be used with above Dir (2-model reest)
	ALIGNDURXFORMEXT		Input transform ext for state duration model set to be used with 2-model reest
	ALIGNDURXFORMDIR		Input transform dir for state duration model set to be used with 2-model reest
	DURINXFORMMASK		Input transform mask for state duration model set (default output transform mask)
	DURPAXFORMMASK		Parent transform mask for state duration model set (default output parent mask)
HHEd	USEPATTERN	F	Use pattern instead of base phone for tree-based clustering
	SINGLETREE	F	Construct single tree for each state position
	APPLYMDL	F	Use the MDL criterion for tree-based clustering
	IGNORESTRW	F	Ignore stream weight in tree-based clustering
	REDUCEMEM	F	Use reduced memory implementation of tree-based clustering
	MINVAR	1.0E-6	Minimum variance floor for model set
	MDLFACTOR	1.0	Factor to control the model complexity term in the MDL criterion
	MINLEAFOCC	0.0	Minimum occupancy count in each leaf node
	MINMIXOCC	0.0	Minimum occupancy count in each mixture component
	SHRINKOCCTHRESH		Minimum occupancy count in decision trees shrinking
HMGENS	SAVEBINARY	F	Save generated parameters in binary
	OUTPDF	F	Output pdf sequences
	PARMGENTYPE	0	Type of parameter generation algorithm
	MODELALIGN	F	Use model-level alignments given from label files to determine model-level durations

Module	Name	Default	Description
	STATEALIGN	F	Use state-level alignments given from label files to determine state-level durations
	USEALIGN	F	Use model-level alignments to prune EM-based parameter generation algorithm
	USEHMMFB	F	Do not use state duration models in the EM-based parameter generation algorithm
	INXFORMMASK		Input transform mask
	PAXFORMMASK		Parent transform mask
	PDFSTRSIZE		# of PdfStreams
	PDFSTRORDER		Size of static feature in each PdfStream
	PDFSTREXT		Ext to be used for generated parameters from each PdfStream
	WINEXT		Ext to be used for window coefficients file
	WINDIR		Dir containing window coefficient files
	WINFN		Name of window coefficient files

Other configuration variables in HTK can also be used with HTS. Please refer to HTKBook Chapter 18 for others.

3 Added command-line options

Various new command-line options have also been added to HTK tools. They are listed as follows:

HInit

```
Option                                                    Default
 -g      Ignore outlier vector in MSD                      on
```

HRest

```
Option                                                    Default
 -g s    output duration model to file s                   none
 -o fn   Store new hmm def in fn (name only)               outDir/srcfn
```

HERest

```
Option                                                    Default
 -b      use an input linear transform for dur models      off
 -f s    extension for new duration model files            as src
 -g s    output duration model to file s                   none
 -n s    dir to find duration model definitions            current
 -q s    save all xforms for duration to TMF file s        TMF
 -u tmvwapd update t)rans m)eans v)ars w)ghts              tmvw
                 a)daptation xform p)rior used
                 s)semi-tied xform
                 d) switch to duration model update flag
```

```
  -y s    extension for duration model files               none
  -N mmf  load duration macro file mmf
  -R dir  dir to write duration macro files                current
  -W s [s] set dir for duration parent xform to s          off
          and optional extension
  -Y s [s] set dir for duration input xform to s           none
          and optional extension
  -Z s [s] set dir for duration output xform to s          none
```

HHEd

```
 Option                                                    Default
  -a f    factor to control the second term in the MDL      1.0
  -i      ignore stream weight                              off
  -m      apply MDL principle for clustering                off
  -p      use pattern instead of base phone                 off
  -r      reduce memory usage on clustering                 off
  -s      construct single tree                             off
  -v f    Set minimum variance to f                         1.0E-6
```

HMGenS

```
 Option                                                    Default
  -a      Use an input linear transform for HMMs            off
  -b      Use an input linear transform for dur models      off
  -c n    type of parameter generation algorithm            0
           0: both mix and state sequences are given
           1: state sequence is given,
              but mix sequence is hidden
           2: both state and mix sequences are hidden
  -d s    dir to find hmm definitions                       current
  -e      use model alignment from label for pruning        off
  -f f    frame shift in 100 ns                             50000
  -g f    Mixture pruning threshold                         10.0
  -h s [s] set speaker name pattern to s,                   *.%%%
          optionally set parent patterns
  -m      use model alignment for duration                  off
  -n s    dir to find duration model definitions            current
  -p      output pdf sequences                              off
  -r f    speaking rate factor (f<1: fast  f>1: slow)       1.0
  -s      use state alignment for duration                  off
  -t f [i l] set pruning to f [inc limit]                   inf
  -v f    threshold for switching spaces for MSD            0.5
  -x s    extension for hmm files                           none
  -y s    extension for duration model files                none
  -E s [s] set dir for parent xform to s                    off
          and optional extension
  -G fmt  Set source label format to fmt                    as config
  -H mmf  Load HMM macro file mmf
  -I mlf  Load master label file mlf
  -J s [s] set dir for input xform to s                     none
          and optional extension
```

```
-L dir  Set input label (or net) dir                        current
-M dir  Dir to write HMM macro files                        current
-N mmf  Load duration macro file mmf
-S f    Set script file to f                                none
-T N    Set trace flags to N                                0
-V      Print version information                           off
-W s [s] set dir for duration parent xform to s             off
        and optional extension
-X ext  Set input label (or net) file ext                   lab
-Y s [s] set dir for duration input xform to s              none
        and optional extension
```

Please also refer to HTKBook Chapter 17 for other command-line options.

4 Added commands and modifications in HHEd

Some HHEd commands have been added in HTS. They are as follows:

```
AX filename         - Set the Adapt XForm to filename
CM directory        - Convert models to pdf for speech synthesizer
CT directory        - Convert trees/questions for speech synthesizer
DM type macroname   - Delete macro from model-set
DR id               - Convert decision trees to a regression tree
DV                  - Convert full covariance to diagonal variances
IX filename         - Set the Input Xform to filename
PX filename         - Set the Parent Xform to filename
// comment          - Comment line (ignored)
```

In the HHEd command `MU`, HTS additionally supports additive and multiplicative mixture incrementation. For example,

```
MU  6 {*.state[3].mix}
MU +6 {*.state[3].mix}
MU *6 {*.state[3].mix}
```

if the the mixture components per state is 2, the first command increases the numbers of mixtures in state 3 of all phonemes of `aa` to 6, the second one increases them to 8, and the last one increased them to 12.

In many HHEd commands, we are required to specify item lists to specify a set of items to be processed. In HTS, item list specification has been modified to specify stream-level items.

itemList = "{" itemSet { "," itemSet } "}"
itemSet = hmmName . ["transP" | "state" state]
hmmName = ident | identList
identList = "(" ident { "," ident } ")"
ident = < char | metachar >
metachar = "?" | "*"
state = index ["." stateComp]
index = "[" intRange { "," intRange } "]"
intRange = integer ["-" integer]
stateComp = "dur" | "weights" | stream
stream = [" stream" index] [".mix" mix]
mix = index ["." ("mean" | "cov")]

For example,

```
TI str1 {*.state[2].stream[1]}
```

denotes tying streams in state 2 of all phonemes.