How to generate I-Vectors with ALIZE

anthony.larcher@i2r.a-star.edu.sg

13 avril 2012

Table des matières

1 Total Variability paradigm

Initially introduced for speaker recognition, i-vectors (Dehak *et al.*, 2011) have become very popular in the field of speech processing and recent publications show that they are also reliable for text-dependent speaker verification (Larcher *et al.*, 2012) language recognition (Martinez *et al.*, 2011) and speaker diarization (Franco-Pedroso *et al.*, 2010). I-vectors convey the speaker characteristic among other information such as transmission channel, acoustic environment or phonetic content of the speech segment.

Detailed descriptions of the Total Variability paradigm could be found in (Dehak *et al.*, 2011; Martinez *et al.*, 2011; Kanagasundaram *et al.*, 2011). The i-vector extraction could be seen as a probabilistic compression process that reduces the dimensionality of speech-session super-vectors according to a linear-Gaussian model. The speaker- and channel-dependent super-vector $\mathcal{M}_{(s,h)}$ of concatenated Gaussian Mixture Model (GMM) means is projected in a low dimensionality space, named Total Variability space, as follows

$$\mathcal{M}_{(s,h)} = \mathbf{m} + \mathbf{T}\mathbf{w}_{(s,h)} \tag{1}$$

where $\mathbf{m}$ is the mean super-vector of a gender-dependent Universal Background Model (UBM), $\mathbf{T}$ is called Total Variability matrix and $\mathbf{w}_{(s,h)}$ is the resulting i-vector.

2 ComputeJFAStat

ComputeJFAStat computes Baum-Welch statistics for a list of files given a Gaussian Mixture Model.

2.1 Inputs-Outputs

Universal Background Model Baum-Welch statistics are computed over this Universal background Model.

Index file the list of files to process is given in an index file (NDX). NDX file should be ascii file with one file per line.

Null Order Statistics per Session will not be used for TotalVariability estimation.

Null Order Statistics per Speaker this matrix contains zero order Baum-Welch statistics for each file. Dimensions of this matrix are number of files $\times$ number of distributions of the UBM.

First Order Statistics per Session will not be used for TotalVariability estimation.

First Order Statistics per Speaker this matrix contains zero order Baum-Welch statistics for each file. Dimensions of this matrix are number of files $\times$ dimension of super vector (which is equal to the number of distributions of the UBM multiplied by the dimension of acoustic features).

2.2 ComputeJFAStat options

Option	Value	function
featureServerMemAlloc	integer	Specify the memory to allocate for the ALIZE buffer
featureServerMode	FEATURE_WRITABLE	Specify the read/write mode of the featureServer (read-only by default)
featureServerBufferSize	ALL_FEATURES	Algorithm Buffer
minLLK	integer	Specify minimum likelihood values
maxLLK	integer	Specify maximum likelihood values
labelFilesPath	path	directory to load label files, could be relative or absolute
featureFilesPath	path	directory to load feature files, could be relative or absolute
matrixFilesPath	path	directory to load and save matrices model, could be relative or absolute
mixtureFilesPath	path	directory to load GMM model, could be relative or absolute
loadMixtureFileFormat	RAW | XML	format of the GMM to load
loadMatrixFormat	DB | DT	format of the matrix to load, could be binary (DB) or text (DT)
saveMatrixFormat	DB | DT	format of the matrix to save, could be binary (DB) or text (DT)
loadFeatureFileFormat	SPRO4 | HTK | RAW	ALIZE can read different format of features including HTK, SPRO and RAW binary files
loadMixtureFileExtension	string	extension of GMM files to load
loadMatrixFilesExtension	string	extension of Matrices files to load
saveMatrixFilesExtension	string	extension of Matrices files to save
loadFeatureFileExtension	string	extension of feature files to load
ndxFilename	string	name of the index file (NDX) containing the list of files to process
inputWorldFilename	string	name of the Universal Background Model to be used for statistics computation.
nullOrderStatSpeaker	string	name of the file containing null order Baum-Welch statistics per speaker (sum of all sessions for one speaker)
nullOrderStatSession	string	name of the file containing null order Baum-Welch statistics per session
firstOrderStatSpeaker	string	name of the file containing first order Baum-Welch statistics per speaker (sum of all sessions for one speaker)
firstOrderStatSession	string	name of the file containing first order Baum-Welch statistics per session
addDefaultLabel	boolean	add a default label in case no label is found for a feature file
defaultLabel	string	if no label file is found, all features from this file will be given the defaultlabel
labelSelectedFrames	string	feature label to use for statistics computation
frameLength	float	period of the frame extraction given in seconds

3 Total Variability

Total Variaiblity matrix can be computed by using EigenVoice program. However, some of the local variables used for JFA estimation are not used for Total Variability estimation and strongly increase the memory requirement. Total Variability is then another version of the EigenVoice program optimized to process a large number of sessions. However, optimization is still balancing between computational time and memory usage and TotalVariability may still require a large amount of memory.

3.1 Mandatory inputs/outputs

Universal Background Model the UBM on which statistics have been computed. The super-vector of the co-variance coefficients from the UBM will be used by TotalVariability.

TotalVariability matrix output of this executable, TotalVariability matrix contains the main eigenvectors of the Total Variability space.

3.2 Optional inputs/outputs

Zero and first order Baum-Welch statistic matrices could be computed directly by the TotalVariability program. Those inputs are then not mandatory. However, considering that the quantity of files to process is important, we recommend to split the list of files to process and to parallelize the computation by using scripting and to concatenate eventually the matrices by using tools provided in ALIZE/LIA_RAL.

Null Order Statistics this matrix contains zero order Baum-Welch statistics for each file. Dimensions of this matrix are number of files $\times$ number of distributions of the UBM.

First Order Statistics this matrix contains zero order Baum-Welch statistics for each file. Dimensions of this matrix are number of files $\times$ dimension of super vector (which is equal to the number of distributions of the UBM multiplied by the dimension of acoustic features).

3.3 TotalVariability options

Option	Value	function
numThread	integer	number of thread to run in parallel. Only available if LIA_RAL is compiled with pThread library (see ALIZE FAQ for more information)
minLLK	integer	Specify minimum likelihood values
maxLLK	integer	Specify maximum likelihood values
labelFilesPath	path	directory to load label files, could be relative or absolute
featureFilesPath	path	directory to load feature files, could be relative or absolute
matrixFilesPath	path	directory to load and save matrices model, could be relative or absolute
mixtureFilesPath	path	directory to load GMM model, could be relative or absolute
loadMixtureFileFormat	RAW | XML	format of the GMM to load
loadMatrixFormat	DB | DT	format of the matrix to load, could be binary (DB) or text (DT)
saveMatrixFormat	DB | DT	format of the matrix to save, could be binary (DB) or text (DT)
loadFeatureFileFormat	SPRO4 | HTK | RAW	ALIZE can read different format of features including HTK, SPRO and raw binary files
loadMixtureFileExtension	string	extension of GMM files to load
saveMatrixFilesExtension	string	extension of GMM files to save
loadFeatureFileExtension	string	extension of feature files to load
addDefaultLabel	boolean	add a default label in case no label is found for a feature file
defaultLabel	string	if no label file is found, all features from this file will be given the defaultlabel
labelSelectedFrames	string	feature label to use for statistics computation
frameLength	float	period of the frame extraction given in seconds
ndxFilename	string	name of the index file (NDX) containing the list of files to process
inputWorldFilename	string	name of the Universal Background Model to be used for statistics computation.
loadAccs	boolean	if true, then load existing statistic matrices, if false the statistics are first computed
nullOrderStatSpeaker	string	name of the file containing null order Baum-Welch statistics per file (in TotalVariability framework, speaker refer to one only file)
firstOrderStatSpeaker	string	name of the file containing first order Baum-Welch statistics per file (in TotalVariability framework, speaker refer to one only file)
nbIt	integer	number of EM iterations to perform
totalVariabilityNumber	integer	rank of the TotalVariability matrix
loadInitTotalVariabilityMatrix	boolean	if true, load an existing matrix for initialization
randomInitLaw	normal | uniform	distribution used to randomly initialize the matrix
initTotalVariabilityMatrix	string	name of the initialization matrix to load
saveInitT	boolean	save the initialization matrix
totalVariabilityMatrix	string	name of the file to store the final TotalVariability matrix
saveAllTVMatrices	boolean	if true, save the TotalVariability matrix after each iteration
minDivergence	boolean	use minimum divergence criteria to update T matrix
meanEstimate	string	name of the new mean estimate after minimum divergence
checkLLK	boolean	if true compute the likelihood after each iteration (Likelihood is computed by creating a GMM for each session and has not been optimized yet)
computeLLK	integer	number of sessions to use to compute the likelihood after each iteration

4 TrainTarget

Extract i-Vectors for a list of files.

4.1 Input-Output

Baum-Welch statistics have to be computed first by using ComputeJFAStats. I-Vectors can be extracted by using the same configuration used for JFA TrainTarget. However, TrainTarget run in mode "*iVector*" is faster than in mode "*JFA*" due to the Multi-Thread implementation.

Null Order Statistics per Speaker this matrix contains cumulated zero order Baum-Welch statistics for each file. Dimensions of this matrix are number of files $\times$ number of distributions of the UBM.

First Order Statistics per Speaker this matrix contains cumulated zero order Baum-Welch statistics for each file. Dimensions of this matrix are number of files $\times$ dimension of super vector (which is equal to the number of distributions of the UBM multiplied by the dimension of acoustic features).

4.2 TrainTarget options for i-Vectors extraction

Option	Value	function
numThread	integer	number of thread to run in parallel. Only available if LIA_RAL is compiled with pThread library (see ALIZE FAQ for more information)
featureServerMemAlloc	integer	Specify the memory to allocate for the ALIZE buffer
featureServerMode	FEATURE_WRITABLE	Specify the read/write mode of the featureServer (read-only by default)
featureServerBufferSize	ALL_FEATURES	Algorithm Buffer
minLLK	integer	Specify minimum likelihood values
maxLLK	integer	Specify maximum likelihood values
MAPAlgo	string	not used in *iVector* mode but need to be fill
matrixFilesPath	path	directory to load and save matrices model, could be relative or absolute
mixtureFilesPath	path	directory to load GMM model, could be relative or absolute
vectorFilesPath	path	directory to store i-Vector files, could be relative or absolute
loadMixtureFileFormat	RAW | XML	format of the GMM to load
loadMatrixFormat	DB | DT	format of the matrix to load, could be binary (DB) or text (DT)
saveMatrixFormat	DB | DT	format of the matrix to save, could be binary (DB) or text (DT)
loadMixtureFileExtension	string	extension of GMM files to load
loadMatrixFilesExtension	string	extension of Matrices files to load
saveMatrixFilesExtension	string	extension of Matrices files to save
vectorFilesExtension	string	extension of i-vector files
channelCompensation	iVector	must be iVector to set this mode
ndxFilename	string	name of the index file (NDX) containing the list of files to process
inputWorldFilename	string	name of the Universal Background Model to be used for statistics computation.
minDivergence	boolean	use minimum divergence reestimated mean to compute iVectors
meanEstimate	string	name of the estimate of mean after minimum divergence
nullOrderStatSpeaker	string	name of the file containing null order Baum-Welch statistics per file
firstOrderStatSpeaker	string	name of the file containing first order Baum-Welch statistics per file
totalVariabilityMatrix	string	name of the TotalVariability matrix file to load
totalVariabilityNumber	integer	rank of the TotalVariability matrix
labelSelectedFrames	string	feature label to use for statistics computation
frameLength	float	period of the frame extraction given in seconds

5 First run

Example of TotalVariability matrix estimation and i-Vector extraction is provided in *IV.sh*.

Références

DEHAK, N., KENNY, P., DEHAK, R., DUMOUCHEL, P. et OUELLET, P. (2011). Front-End Factor Analysis for Speaker Verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4):788–798.

FRANCO-PEDROSO, J., LOPEZ-MORENO, I., TOLEDANO, D. T. et GONZALEZ-RODRIGUEZ, J. (2010). ATVS-UAM System Description for the Audio Segmentation and Speaker Diarization Albayzin 2010 Evaluation. *In FALA "VI Jornadas en Tecnología del Habla" and II Iberian SLTech Workshop*, pages 415–418.

KANAGASUNDARAM, A., VOGT, R., DEAN, D., SRIDHARAN, S. et MASON, M. (2011). i-vector Based Speaker Recognition on Short Utterances. *In International Conference on Speech Communication and Technology*.

LARCHER, A., BOUSQUET, P.-M., LEE, K. A., MATROUF, D., LI, H. et BONASTRE, J.-F. (2012). I-vectors in the context of phonetically-constrained short utterances for speaker verification. *In IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*.

MARTINEZ, D., PLCHOT, O., BURGET, L., GLEMBEK, O. et MATEJKA, P. (2011). Language Recognition in iVectors Space. *In Interspeech*.