Unsupervised Keyword Spotting using Bounded Generalized Gaussian Mixture Model with ICA

Muhammad Azam[1] Nizar Bouguila[2]

Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada

Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada

December 16, 2015

Outline

Unsupervised Keyword Spotting

- Automatic Speech Recognition (ASR)
- Problems associated with ASR
- Unsupervised Methods for ASR
- Keyword Spotting and ASR
- Dynamic Time Warping and Mixture Model

Unsupervised Keyword Spotting

- Automatic Speech Recognition (ASR)
- Problems associated with ASR
- Unsupervised Methods for ASR
- Keyword Spotting and ASR
- Dynamic Time Warping and Mixture Model

Unsupervised Keyword Spotting

- Automatic Speech Recognition (ASR)
- Problems associated with ASR
- Unsupervised Methods for ASR
- Keyword Spotting and ASR
- Dynamic Time Warping and Mixture Model

Unsupervised Keyword Spotting

- Automatic Speech Recognition (ASR)
- Problems associated with ASR
- Unsupervised Methods for ASR
- Keyword Spotting and ASR
- Dynamic Time Warping and Mixture Model

Unsupervised Keyword Spotting

- Automatic Speech Recognition (ASR)
- Problems associated with ASR
- Unsupervised Methods for ASR
- Keyword Spotting and ASR
- Dynamic Time Warping and Mixture Model

Development in Mixture Models

- Problems in Mixture Models: Senitivity to the outliers in GMM
- Development in Mixture model: SMM, GGMM
  - SMM: Improved robustness of algorithm and more flexible approach
  - GGMM: Extensible to fit different shapes using shape parameters, i.e. can generalized into GMM and LMM
- Bounded Support Mixture Models: BGMM, BGGMM

Development in Mixture Models

- Problems in Mixture Models: Senitivity to the outliers in GMM
- Development in Mixture model: SMM, GGMM
  - SMM: Improved robustness of algorithm and more flexible approach
  - GGMM: Extensible to fit different shapes using shape parameters, i.e. can generalized into GMM and LMM
- Bounded Support Mixture Models: BGMM, BGGMM

Development in Mixture Models

- Problems in Mixture Models: Senitivity to the outliers in GMM
- Development in Mixture model: SMM, GGMM
  - SMM: Improved robustness of algorithm and more flexible approach
  - GGMM: Extensible to fit different shapes using shape parameters, i.e. can generalized into GMM and LMM
- Bounded Support Mixture Models: BGMM, BGGMM

Development in Mixture Models

- Problems in Mixture Models: Senitivity to the outliers in GMM
- Development in Mixture model: SMM, GGMM
  - SMM: Improved robustness of algorithm and more flexible approach
  - GGMM: Extensible to fit different shapes using shape parameters, i.e. can generalized into GMM and LMM
- Bounded Support Mixture Models: BGMM, BGGMM

Development in Mixture Models

- Problems in Mixture Models: Senitivity to the outliers in GMM
- Development in Mixture model: SMM, GGMM
  - SMM: Improved robustness of algorithm and more flexible approach
  - GGMM: Extensible to fit different shapes using shape parameters, i.e. can generalized into GMM and LMM
- Bounded Support Mixture Models: BGMM, BGGMM

ICA Mixture Model

- Data comes from a mixture model and it can be categorized into mutually exclusive classes
- Each class of data is modeled as ICA
    - Computation of Data log-likelihood
    - Computation of Posterior
    - Adaptation of Parameters using standard ICA Model
- Applications: Blind source separation, classification, image Segmentation and speech and image processing etc.

ICA Mixture Model

- Data comes from a mixture model and it can be categorized into mutually exclusive classes
- Each class of data is modeled as ICA
    - Computation of Data log-likelihood
    - Computation of Posterior
    - Adaptation of Parameters using standard ICA Model
- Applications: Blind source separation, classification, image Segmentation and speech and image processing etc.

ICA Mixture Model

- Data comes from a mixture model and it can be categorized into mutually exclusive classes
- Each class of data is modeled as ICA
  - Computation of Data log-likelihood
  - Computation of Posterior
  - Adaptation of Parameters using standard ICA Model
- Applications: Blind source separation, classification, image Segmentation and speech and image processing etc.

ICA Mixture Model

- Data comes from a mixture model and it can be categorized into mutually exclusive classes
- Each class of data is modeled as ICA
  - Computation of Data log-likelihood
  - Computation of Posterior
  - Adaptation of Parameters using standard ICA Model
- Applications: Blind source separation, classification, image Segmentation and speech and image processing etc.

ICA Mixture Model

- Data comes from a mixture model and it can be categorized into mutually exclusive classes
- Each class of data is modeled as ICA
  - Computation of Data log-likelihood
  - Computation of Posterior
  - Adaptation of Parameters using standard ICA Model
- Applications: Blind source separation, classification, image Segmentation and speech and image processing etc.

ICA Mixture Model

- Data comes from a mixture model and it can be categorized into mutually exclusive classes
- Each class of data is modeled as ICA
  - Computation of Data log-likelihood
  - Computation of Posterior
  - Adaptation of Parameters using standard ICA Model
- Applications: Blind source separation, classification, image Segmentation and speech and image processing etc.

Multivariate Bounded Generalized Gaussian Mixture Model

- The data $\mathcal{X} = (\vec{X}_1, ..., \vec{X}_N)$, with a mixture of K distributions can be modeled as :

$$p(\mathcal{X}|\Theta) = \prod_{i=1}^{N}\sum_{j=1}^{K} p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j)p_j \tag{1}$$

Where Θ represents the parameters of mixture model having K classes as $\Theta = (\xi_1, \xi_2, \xi_3, \xi_4)$, with $\xi_1 = (\vec{\mu}_1, ..., \vec{\mu}_K)$, $\xi_2 = (\vec{\sigma}_1, ..., \vec{\sigma}_K)$, $\xi_3 = (\vec{\lambda}_1, ..., \vec{\lambda}_K)$ and $\xi_4 = (p_1, ..., p_K)$.

- The term $p(\vec{X}|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j)$ represents the bounded generalized Gaussian distribution (BGGD).

$$p(\vec{X}|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) = \frac{f(\vec{X}|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j)\mathrm{H}(\vec{X}|\Omega_j)}{\int_{\partial_{\Omega_j}} f(\mathrm{X}|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j)dX} \tag{2}$$

Multivariate Bounded Generalized Gaussian Mixture Model

- The data $\mathcal{X} = (\vec{X}_1, ..., \vec{X}_N)$, with a mixture of K distributions can be modeled as :

$$p(\mathcal{X}|\Theta) = \prod_{i=1}^{N} \sum_{j=1}^{K} p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) p_j \tag{1}$$

Where Θ represents the parameters of mixture model having K classes as $\Theta = (\xi_1, \xi_2, \xi_3, \xi_4)$, with $\xi_1 = (\vec{\mu}_1, ..., \vec{\mu}_K)$, $\xi_2 = (\vec{\sigma}_1, ..., \vec{\sigma}_K)$, $\xi_3 = (\vec{\lambda}_1, ..., \vec{\lambda}_K)$ and $\xi_4 = (p_1, ..., p_K)$.

- The term $p(\vec{X}|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j)$ represents the bounded generalized Gaussian distribution (BGGD).

$$p(\vec{X}|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) = \frac{f(\vec{X}|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) \mathrm{H}(\vec{X}|\Omega_j)}{\int_{\partial\Omega_j} f(\mathrm{X}|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) d\mathrm{X}} \tag{2}$$

Multivariate Bounded Generalized Gaussian Mixture Model

- Where

$$\mathrm{H}(\vec{X}|\Omega_j) = \begin{cases} 1 & \text{if} \quad \vec{X} \in \partial_{\Omega_j} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

and
$$f(\vec{X}|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) = \prod_{d=1}^{D} B(\lambda_{jd}) \exp\left(-A(\lambda_{jd}) \left|\frac{X_d - \mu_{jd}}{\sigma_{jd}}\right|^{\lambda_{jd}}\right) \tag{4}$$

with

$$B(\lambda_{jd}) = \frac{\lambda_{jd}\sqrt{\Gamma(3/\lambda_{jd})}}{2\sigma_{jd}\Gamma(1/\lambda_{jd})\sqrt{\Gamma(1/\lambda_{jd})}} \text{ and } A(\lambda_{jd}) = \left[\frac{\Gamma(3/\lambda_{jd})}{\Gamma(1/\lambda_{jd})}\right]^{\lambda_{jd}/2} \tag{5}$$

where $\Gamma(.)$ is the gamma function and $\vec{\mu} = (\mu_1, ..., \mu_D)$, $\vec{\sigma} = (\sigma_1, ..., \sigma_D)$, and $\vec{\lambda} = (\lambda_1, ..., \lambda_D)$ are the mean, standard deviation and shape parameters respectively.

Bounded Generalized Gaussian Mixture Model with ICA

- For the ICA mixture model, each D-dimensional data vector $\vec{X}_i = (X_{i1}, ..., X_{iD})$ can be represented as:

$$\vec{X}_i = A_j \vec{s}_{j,i} + \vec{b}_j \tag{6}$$

where A_j is $L \times D$ basis matrix, $\vec{s}_{j,i}$ is D-dimensional source vector and $\vec{b}_j$ is an L-dimensional bias vector for a particular mixture j. For the simplicity, number of linear combinations (L) is considered to be equal to the number of sources (D) for each observation of the dataset.

Bounded Generalized Gaussian Mixture Model with ICA

- The complete data likelihood is given below.

$$p(\mathcal{X}, Z|\Theta) = \prod_{i=1}^{N} \sum_{j=1}^{K} \left(p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) p_j \right)^{Z_{ij}} \tag{7}$$

- The log-likelihood function is given below.

$$L(\mathcal{X}, Z|\Theta) = \sum_{i=1}^{N} \sum_{j=1}^{K} Z_{ij} \log \left(p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) p_j \right) \tag{8}$$

where Z_{ij} is the posterior probability and can be written as:

$$Z_{ij} = p(j|\vec{X}_i) = \frac{p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) p_j}{\sum_{j=1}^{K} p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) p_j} \tag{9}$$

and $Z = \{\vec{Z}_1, ..., \vec{Z}_N\}$.

Bounded Generalized Gaussian Mixture Model with ICA

- The complete data likelihood is given below.

$$p(\mathcal{X}, Z|\Theta) = \prod_{i=1}^{N} \sum_{j=1}^{K} \left(p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) p_j \right)^{Z_{ij}} \quad (7)$$

- The log-likelihood function is given below.

$$L(\mathcal{X}, Z|\Theta) = \sum_{i=1}^{N} \sum_{j=1}^{K} Z_{ij} \log \left(p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) p_j \right) \quad (8)$$

where Z_{ij} is the posterior probability and can be written as:

$$Z_{ij} = p(j|\vec{X}_i) = \frac{p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) p_j}{\sum_{j=1}^{K} p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) p_j} \quad (9)$$

and $Z = \{\vec{Z}_1, ..., \vec{Z}_N\}$.

Parameter Estimation

- **Maximization of the log-likelihood**
  Mean, standard deviation and prior probability

- **Standard ICA Model**
  Basis function, bias vector and shape parameter

Parameter Estimation

- **Maximization of the log-likelihood**
  Mean, standard deviation and prior probability

- **Standard ICA Model**
  Basis function, bias vector and shape parameter

Parameter Estimation: Maximization of the log-likelihood

- Prior Probability

$$\hat{p}_j = \frac{1}{N}\sum_{i=1}^{N} p(j|\vec{X}_i) \tag{10}$$

- Mean

$$\hat{\mu}_{jd} = \frac{\sum_{i=1}^{N} p(j|\vec{X}_i)\left(\left|X_{id}-\mu_{jd}\right|^{\lambda_{jd}-2}X_{id}+T_{jd}\right)}{\sum_{i=1}^{N} p(j|\vec{X}_i)\left|X_{id}-\mu_{jd}\right|^{\lambda_{jd}-2}} \tag{11}$$

where

$$T_{jd} = \frac{\sum_{m=1}^{M} \mathrm{sign}(\mu_{jd}-s_{j_{md}})\left|\mu_{jd}-s_{j_{md}}\right|^{\lambda_{jd}-1}\mathrm{H}(s_{j_{md}}|\Omega_j)}{\sum_{m=1}^{M} \mathrm{H}(s_{j_{md}}|\Omega_j)} \tag{12}$$

Parameter Estimation: Maximization of the log-likelihood

- Prior Probability

$$\hat{p}_j = \frac{1}{N}\sum_{i=1}^{N} p(j|\vec{X}_i) \tag{10}$$

- Mean

$$\hat{\mu}_{jd} = \frac{\sum_{i=1}^{N} p(j|\vec{X}_i)\left(\left|X_{id}-\mu_{jd}\right|^{\lambda_{jd}-2} X_{id} + T_{jd}\right)}{\sum_{i=1}^{N} p(j|\vec{X}_i)\left|X_{id}-\mu_{jd}\right|^{\lambda_{jd}-2}} \tag{11}$$

where

$$T_{jd} = \frac{\sum_{m=1}^{M} \mathrm{sign}(\mu_{jd}-s_{jmd})\left|\mu_{jd}-s_{jmd}\right|^{\lambda_{jd}-1} \mathrm{H}(s_{jmd}|\Omega_j)}{\sum_{m=1}^{M} \mathrm{H}(s_{jmd}|\Omega_j)} \tag{12}$$

Parameter Estimation: Maximization of the log-likelihood

- Standard Deviation

$$\hat{\sigma}_{jd} = \left(\frac{\lambda_{jd} A(\lambda_{jd}) \sum_{i=1}^{N} p(j|\vec{X}_i) \left| X_{id} - \mu_{jd} \right|^{\lambda_{jd}}}{\sum_{i=1}^{N} p(j|\vec{X}_i)(1 + Q_{jd})} \right)^{1/\lambda_{jd}} \tag{13}$$

where

$$Q_{jd} = \frac{\sum_{m=1}^{M} (-1 + \lambda_{jd} A(\lambda_{jd}) \left| s_{mjd} - \mu_{jd} \right|^{\lambda_{jd}} (\sigma_{jd})^{-\lambda_{jd}}) \mathrm{H}(s_{mjd}|\Omega_j)}{\sum_{m=1}^{M} \mathrm{H}(s_{mjd}|\Omega_j)} \tag{14}$$

with $i = 1, ..., N$, $j = 1, ..., K$, $d = 1, ..., D$ and $m = 1, ..., M$.

Parameter Estimation: Standard ICA Model

- The gradient of log-likelihood for the parameters of each class is given below.

$$\nabla_{\theta_j} L(\mathcal{X}, Z|\Theta) = \sum_{i=1}^{N} Z_{ij} \left(\nabla_{\theta_j} \log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) + \nabla_{\theta_j} \log p_j \right) \tag{15}$$

The ∇_{θ_j} represent gradient with respect to basis function, bias vector and shape parameter.

- The standard ICA model is used for the log-likelihood as follows:

$$\log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) = \log \frac{p(\vec{s}_{j,i})}{|\det A_j|} \tag{16}$$

Parameter Estimation: Standard ICA Model

- The gradient of log-likelihood for the parameters of each class is given below.

$$\nabla_{\theta_j} L(\mathcal{X}, Z|\Theta) = \sum_{i=1}^{N} Z_{ij} \left(\nabla_{\theta_j} \log p(\vec{X}_i | \vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) + \nabla_{\theta_j} \log p_j \right) \tag{15}$$

The ∇_{θ_j} represent gradient with respect to basis function, bias vector and shape parameter.

- The standard ICA model is used for the log-likelihood as follows:

$$\log p(\vec{X}_i | \vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) = \log \frac{p(\vec{s}_{j,i})}{|\det A_j|} \tag{16}$$

Bounded Generalized Gaussian Mixture Model with ICA

- The gradient ascent is used to estimate the shape parameter by maximizing the log-likelihood and represented as:

$$\nabla_{\vec{\lambda}_j} \log[p(\mathcal{X}|\Theta)] = \sum_{i=1}^{N} p(j|\vec{X}_i)\nabla_{\vec{\lambda}_j} \log p(\vec{X}_i|\vec{\mu}_j,\vec{\sigma}_j,\vec{\lambda}_j) \quad (17)$$

- The gradient ascent is used for the adaptation, with the gradient of the component density with respect to shape parameter vector $\vec{\lambda}_j$ for each component of the mixture model.

$$\Delta\vec{\lambda}_j \propto p(j|\vec{X}_i)\frac{\partial}{\partial\vec{\lambda}_j} \log p(\vec{X}_i|\vec{\mu}_j,\vec{\sigma}_j,\vec{\lambda}_j) \quad (18)$$

$$\frac{\partial}{\partial\vec{\lambda}_j} \log p(\vec{X}_i|\vec{\mu}_j,\vec{\sigma}_j,\vec{\lambda}_j) = \vec{\lambda}_j[\mathrm{I} - 2\tanh(\vec{s}_{j,i})\vec{s}_{j,i}^{T}] \quad (19)$$

- By combining Eqs. (18) and (19), we get:

$$\Delta\vec{\lambda}_j \propto p(j|\vec{X}_i)\vec{\lambda}_j[\mathrm{I} - 2\tanh(\vec{s}_{j,i})\vec{s}_{j,i}^{T}] \quad (20)$$

- The gradient ascent is used to estimate the shape parameter by maximizing the log-likelihood and represented as:

$$\nabla_{\vec{\lambda}_j} \log[p(\mathcal{X}|\Theta)] = \sum_{i=1}^{N} p(j|\vec{X}_i) \nabla_{\vec{\lambda}_j} \log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) \quad (17)$$

- The gradient ascent is used for the adaptation, with the gradient of the component density with respect to shape parameter vector $\vec{\lambda}_j$ for each component of the mixture model.

$$\Delta\vec{\lambda}_j \propto p(j|\vec{X}_i) \frac{\partial}{\partial \vec{\lambda}_j} \log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) \quad (18)$$

$$\frac{\partial}{\partial \vec{\lambda}_j} \log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) = \vec{\lambda}_j [\mathrm{I} - 2 \tanh(\vec{s}_{j,i}) \vec{s}_{j,i}^{T}] \quad (19)$$

- By combining Eqs. (18) and (19), we get:

$$\Delta\vec{\lambda}_j \propto p(j|\vec{X}_i) \vec{\lambda}_j [\mathrm{I} - 2 \tanh(\vec{s}_{j,i}) \vec{s}_{j,i}^{T}] \quad (20)$$

- The gradient ascent is used to estimate the shape parameter by maximizing the log-likelihood and represented as:

$$\nabla_{\vec{\lambda}_j} \log[p(\mathcal{X}|\Theta)] = \sum_{i=1}^{N} p(j|\vec{X}_i) \nabla_{\vec{\lambda}_j} \log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) \tag{17}$$

- The gradient ascent is used for the adaptation, with the gradient of the component density with respect to shape parameter vector $\vec{\lambda}_j$ for each component of the mixture model.

$$\Delta\vec{\lambda}_j \propto p(j|\vec{X}_i) \frac{\partial}{\partial \vec{\lambda}_j} \log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) \tag{18}$$

$$\frac{\partial}{\partial \vec{\lambda}_j} \log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) = \vec{\lambda}_j [\mathrm{I} - 2\tanh(\vec{s}_{j,i})\vec{s}_{j,i}^{\,T}] \tag{19}$$

- By combining Eqs. (18) and (19), we get:

$$\Delta\vec{\lambda}_j \propto p(j|\vec{X}_i) \vec{\lambda}_j [\mathrm{I} - 2\tanh(\vec{s}_{j,i})\vec{s}_{j,i}^{\,T}] \tag{20}$$

- An estimate of the shape parameter using gradient ascent is as follows:

$$\hat{\lambda}_j = \vec{\lambda}_j + \alpha(p(j|\vec{X}_i)\vec{\lambda}_j[\mathrm{I} - 2\tanh(\vec{s}_{j,i})\vec{s}_{j,i}^{T}]) \tag{21}$$

where the source is represented as: $\vec{s}_{j,i} = A_j^{-1}(\vec{X}_i - \vec{b}_j)$.

- The adaptation of basis functions for each class is performed by maximizing the log-likelihood with respect to basis functions A_j for each component of mixture model.

$$\nabla_{A_j} \log[p(\mathcal{X}|\Theta)] = \sum_{i=1}^{N} p(j|\vec{X}_i)\nabla_{A_j} \log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) \tag{22}$$

- The adaptation performed by the gradient ascent with respect to the basis functions is given as:

$$\Delta A_j \propto p(j|\vec{X}_i)\frac{\partial}{\partial A_j} \log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) \tag{23}$$

Bounded Generalized Gaussian Mixture Model with ICA

- An estimate of the shape parameter using gradient ascent is as follows:

$$\hat{\lambda}_j = \vec{\lambda}_j + \alpha\left(p(j|\vec{X}_i)\vec{\lambda}_j[\mathrm{I} - 2\tanh(\vec{s}_{j,i})\vec{s}_{j,i}^{\,T}]\right) \quad (21)$$

where the source is represented as: $\vec{s}_{j,i} = A_j^{-1}(\vec{X}_i - \vec{b}_j)$.

- The adaptation of basis functions for each class is performed by maximizing the log-likelihood with respect to basis functions A_j for each component of mixture model.

$$\nabla_{A_j}\log[p(\mathcal{X}|\Theta)] = \sum_{i=1}^{N} p(j|\vec{X}_i)\nabla_{A_j}\log p(\vec{X}_i|\vec{\mu}_j,\vec{\sigma}_j,\vec{\lambda}_j) \quad (22)$$

- The adaptation performed by the gradient ascent with respect to the basis functions is given as:

$$\Delta A_j \propto p(j|\vec{X}_i)\frac{\partial}{\partial A_j}\log p(\vec{X}_i|\vec{\mu}_j,\vec{\sigma}_j,\vec{\lambda}_j) \quad (23)$$

Bounded Generalized Gaussian Mixture Model with ICA

- An estimate of the shape parameter using gradient ascent is as follows:

$$\hat{\lambda}_j = \vec{\lambda}_j + \alpha(p(j|\vec{X}_i)\vec{\lambda}_j[\mathrm{I} - 2\tanh(\vec{s}_{j,i})\vec{s}_{j,i}^{\,T}]) \tag{21}$$

where the source is represented as: $\vec{s}_{j,i} = A_j^{-1}(\vec{X}_i - \vec{b}_j)$.

- The adaptation of basis functions for each class is performed by maximizing the log-likelihood with respect to basis functions A_j for each component of mixture model.

$$\nabla_{A_j} \log[p(\mathcal{X}|\Theta)] = \sum_{i=1}^{N} p(j|\vec{X}_i)\nabla_{A_j} \log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) \tag{22}$$

- The adaptation performed by the gradient ascent with respect to the basis functions is given as:

$$\Delta A_j \propto p(j|\vec{X}_i)\frac{\partial}{\partial A_j} \log p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j) \tag{23}$$

Bounded Generalized Gaussian Mixture Model with ICA

- By using the standard ICA model for log-likelihood, we get:

$$\Delta A_j \propto p(j|\vec{X}_i)A_j[\mathrm{I} - 2\tanh(\vec{s}_{j,i})\vec{s}^{\,T}_{j,i}] \tag{24}$$

- The bias vector $\vec{b}_j$ is estimated using an approximate method as below:

$$\vec{b}_j = \frac{\sum_{i=1}^{N} \vec{X}_i p(j|\vec{X}_i)}{\sum_{i=1}^{N} p(j|\vec{X}_i)} \tag{25}$$

- In the adaptation of the shape parameter, basis functions and bias vector, the gradient of the component of the mixture model is weighted by $p(j|\vec{X}_i)$.

Bounded Generalized Gaussian Mixture Model with ICA

- By using the standard ICA model for log-likelihood, we get:

$$\Delta A_j \propto p(j|\vec{X}_i) A_j [\mathrm{I} - 2\tanh(\vec{s}_{j,i})\vec{s}_{j,i}^{T}] \quad (24)$$

- The bias vector $\vec{b}_j$ is estimated using an approximate method as below:

$$\vec{b}_j = \frac{\sum_{i=1}^{N} \vec{X}_i p(j|\vec{X}_i)}{\sum_{i=1}^{N} p(j|\vec{X}_i)} \quad (25)$$

- In the adaptation of the shape parameter, basis functions and bias vector, the gradient of the component of the mixture model is weighted by $p(j|\vec{X}_i)$.

Bounded Generalized Gaussian Mixture Model with ICA

- By using the standard ICA model for log-likelihood, we get:

$$\Delta A_j \propto p(j|\vec{X}_i)A_j[\mathrm{I} - 2\tanh(\vec{s}_{j,i})\vec{s}_{j,i}^{\,T}] \tag{24}$$

- The bias vector $\vec{b}_j$ is estimated using an approximate method as below:

$$\vec{b}_j = \frac{\sum_{i=1}^{N} \vec{X}_i p(j|\vec{X}_i)}{\sum_{i=1}^{N} p(j|\vec{X}_i)} \tag{25}$$

- In the adaptation of the shape parameter, basis functions and bias vector, the gradient of the component of the mixture model is weighted by $p(j|\vec{X}_i)$.

Model Learning with BGGMM using ICA

```
Input:Dataset $\mathcal{X} = \{\vec{X}_1, \ldots, \vec{X}_N\}$, $t_{min}$.
Output: $\xi_1$, $\xi_2$,$\xi_3$, and $\xi_4$.
{Initialization}: K-Means Algorithm. Set $\xi_3 = 2$.
while relative change in log-likelihood $\geq t_{min}$ do
  {[E Step]}:
    for all $1 \leq j \leq K$ do
      Compute $p(\vec{X}_i|\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j)$ for $i = 1, \ldots, N$.
      Compute $p(j|\vec{X}_i)$ for $i = 1, \ldots, N$.
    end for
  {[M step]}:
    for all $1 \leq j \leq K$ do
      start ICA Algorithm
        Update the basis functions $A_j$.
        Update the bias vector $\vec{b}_j$.
        Update shape parameter $\vec{\lambda}_j$.
      end ICA
      Update the mixing parameter $p_j$.
      Update the mean $\vec{\mu}_j$.
      Update standard deviation $\vec{\sigma}_j$.
    end for
end while
```

Keyword Spotting framework

- Feature Extraction: MFCC
- Training ICA Mixture Model
- Gaussian Posteriorgrams for keyword example and test data
- Segmental Dynamic Time Warping
- Keyword detection: Score Fusion

Keyword Spotting framework

- Feature Extraction: MFCC
- Training ICA Mixture Model
- Gaussian Posteriorgrams for keyword example and test data
- Segmental Dynamic Time Warping
- Keyword detection: Score Fusion

Keyword Spotting framework

- Feature Extraction: MFCC
- Training ICA Mixture Model
- Gaussian Posteriorgrams for keyword example and test data
- Segmental Dynamic Time Warping
- Keyword detection: Score Fusion

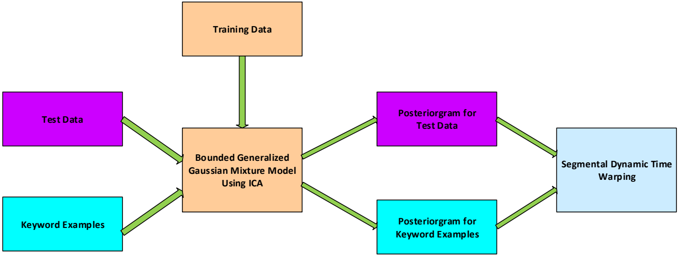
Keyword Spotting framework

- Feature Extraction: MFCC
- Training ICA Mixture Model
- Gaussian Posteriorgrams for keyword example and test data
- Segmental Dynamic Time Warping
- Keyword detection: Score Fusion

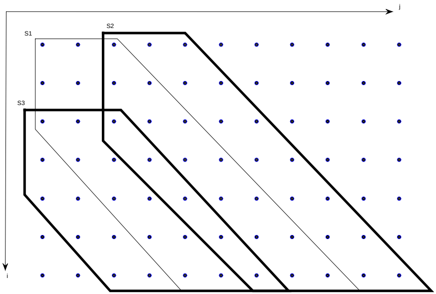
Keyword Spotting framework

- Feature Extraction: MFCC
- Training ICA Mixture Model
- Gaussian Posteriorgrams for keyword example and test data
- Segmental Dynamic Time Warping
- Keyword detection: Score Fusion

Keyword Spotting framework



Segmental Dynamic Time Warping



Experimental Results

- TIMIT Speech Corpus, 6300 speech utterances (Training:4620, Test:1680)
- Same Parameters for keyword spotting from the literature
- Evaluation Matrices, average precision (P@10 and P@N), Equal Error Rate (EER) and the ranking of several keywords based on EER
- Comparison of detection results: ICA mixture model with GMM

Experimental Results

- TIMIT Speech Corpus, 6300 speech utterances (Training:4620, Test:1680)
- Same Parameters for keyword spotting from the literature
- Evaluation Matrices, average precision (P@10 and P@N), Equal Error Rate (EER) and the ranking of several keywords based on EER
- Comparison of detection results: ICA mixture model with GMM

Experimental Results

- TIMIT Speech Corpus, 6300 speech utterances (Training:4620, Test:1680)
- Same Parameters for keyword spotting from the literature
- Evaluation Matrices, average precision (P@10 and P@N), Equal Error Rate (EER) and the ranking of several keywords based on EER
- Comparison of detection results: ICA mixture model with GMM

Experimental Results

- TIMIT Speech Corpus, 6300 speech utterances (Training:4620, Test:1680)
- Same Parameters for keyword spotting from the literature
- Evaluation Matrices, average precision (P@10 and P@N), Equal Error Rate (EER) and the ranking of several keywords based on EER
- Comparison of detection results: ICA mixture model with GMM

Experimental Results

- TIMIT 10 Keyword List

age(3:10)	warm(10:8)	year(11:177)	money(19:17)
artists(7:7)	problem(22:9)	children(18:15)	
surface(3:7)	development(9:8)	organizations(7:7)	

- Evaluation matrix using ICA Mixture Model

# of Examples	P@10	P@N	EER
1	28.37%	26.43%	29.20%
5	57.75%	51.39%	13.79%
10	64.87%	58.27%	12.35%

- Evaluation matrix using Gaussian Mixture Model

# of Examples	P@10	P@N	EER
1	27.0%	17.3%	27.0%
5	61.3%	33.0%	16.8%
10	68.3%	39.3%	15.8%

Experimental Results

- TIMIT 10 Keyword List

age(3:10)	warm(10:8)	year(11:177)	money(19:17)
artists(7:7)	problem(22:9)	children(18:15)	
surface(3:7)	development(9:8)	organizations(7:7)	

- Evaluation matrix using ICA Mixture Model

# of Examples	P@10	P@N	EER
1	28.37%	26.43%	29.20%
5	57.75%	51.39%	13.79%
10	64.87%	58.27%	12.35%

- Evaluation matrix using Gaussian Mixture Model

# of Examples	P@10	P@N	EER
1	27.0%	17.3%	27.0%
5	61.3%	33.0%	16.8%
10	68.3%	39.3%	15.8%

Experimental Results

- TIMIT 10 Keyword List

age(3:10)	warm(10:8)	year(11:177)	money(19:17)
artists(7:7)	problem(22:9)	children(18:15)	
surface(3:7)	development(9:8)	organizations(7:7)	

- Evaluation matrix using ICA Mixture Model

# of Examples	P@10	P@N	EER
1	28.37%	26.43%	29.20%
5	57.75%	51.39%	13.79%
10	64.87%	58.27%	12.35%

- Evaluation matrix using Gaussian Mixture Model

# of Examples	P@10	P@N	EER
1	27.0%	17.3%	27.0%
5	61.3%	33.0%	16.8%
10	68.3%	39.3%	15.8%

Experimental Results

- Ranking of Keywords by EER
- 5 examples of each keyword
- The words with more syllables have better performance

organizations(6.1%)	development(6.7%)	childern(11.3%)
problem(12.6%)	artists(13.5%)	money(15.8%)
warm(21.4%)	year(22.9%)	

Experimental Results

- Ranking of Keywords by EER
- 5 examples of each keyword
- The words with more syllables have better performance

organizations(6.1%)	development(6.7%)	childern(11.3%)
problem(12.6%)	artists(13.5%)	money(15.8%)
warm(21.4%)	year(22.9%)	

Experimental Results

- Ranking of Keywords by EER
- 5 examples of each keyword
- The words with more syllables have better performance

organizations(6.1%)	development(6.7%)	childern(11.3%)
problem(12.6%)	artists(13.5%)	money(15.8%)
warm(21.4%)	year(22.9%)	

Conclusion

- Multivariate Bounded Generalized Gaussian Mixture Model
- Bounded Generalized Gaussian Mixture Model with ICA
- Application of ICA Mixture Model in unsupervised keyword spotting
- TIMIT Speech Corpus for evaluation of keyword spotting framework
- Keyword detection based on average precision (P@10 and P@N), EER and the ranking of several keywords for EER

Conclusion

- Multivariate Bounded Generalized Gaussian Mixture Model
- Bounded Generalized Gaussian Mixture Model with ICA
- Application of ICA Mixture Model in unsupervised keyword spotting
- TIMIT Speech Corpus for evaluation of keyword spotting framework
- Keyword detection based on average precision (P@10 and P@N), EER and the ranking of several keywords for EER

Conclusion

- Multivariate Bounded Generalized Gaussian Mixture Model
- Bounded Generalized Gaussian Mixture Model with ICA
- Application of ICA Mixture Model in unsupervised keyword spotting
- TIMIT Speech Corpus for evaluation of keyword spotting framework
- Keyword detection based on average precision (P@10 and P@N), EER and the ranking of several keywords for EER

Conclusion

- Multivariate Bounded Generalized Gaussian Mixture Model
- Bounded Generalized Gaussian Mixture Model with ICA
- Application of ICA Mixture Model in unsupervised keyword spotting
- TIMIT Speech Corpus for evaluation of keyword spotting framework
- Keyword detection based on average precision (P@10 and P@N), EER and the ranking of several keywords for EER

Conclusion

- Multivariate Bounded Generalized Gaussian Mixture Model
- Bounded Generalized Gaussian Mixture Model with ICA
- Application of ICA Mixture Model in unsupervised keyword spotting
- TIMIT Speech Corpus for evaluation of keyword spotting framework
- Keyword detection based on average precision (P@10 and P@N), EER and the ranking of several keywords for EER

Future Work

- More simulations with large vocabulary database
- Use of ICA Mixture Model in different applications to validate the algorithm
- Variational Bayesian learning or minimum description length to automatically optimize the parameters

Future Work

- More simulations with large vocabulary database
- Use of ICA Mixture Model in different applications to validate the algorithm
- Variational Bayesian learning or minimum description length to automatically optimize the parameters

Future Work

- More simulations with large vocabulary database
- Use of ICA Mixture Model in different applications to validate the algorithm
- Variational Bayesian learning or minimum description length to automatically optimize the parameters

References

Outline
Introduction
Proposed Algorithm
Experiment and Results
Conclusion and Future Work

- Y. Zhang, Unsupervised speech processing with applications to query by-example spoken term detection, Ph.D. thesis, MIT. Department of Electrical Engineering and Computer Science., 2013.
- T.J. Hazen, W. Shen, and C. White, Query-By-Example Spoken Term Detection using Phonetic Posteriorgram Templates, in ASRU 2009. IEEE Workshop on, Nov 2009, pp. 421426.
- T. W. Lee and M. S. Lewicki, The generalized Gaussian mixture model using ICA, 2000.
- T. M. Nguyen, Q.M. Jonathan Wu, and H. Zhang, Bounded generalized Gaussian mixture model, Pattern Recognition, vol. 47, no. 9, pp. 3132 3142, 2014.
- J. Lindblom and J. Samuelsson, Bounded support Gaussian mixture modeling of speech spectra, Speech and Audio Processing, IEEE Transactions on, vol. 11, no. 1, pp. 8899, Jan 2003.

References

Outline

Introduction

Proposed Algorithm

Experiment and Results

Conclusion and Future Work

- Y. Zhang, Unsupervised speech processing with applications to query by-example spoken term detection, Ph.D. thesis, MIT. Department of Electrical Engineering and Computer Science., 2013.
- T.J. Hazen, W. Shen, and C. White, Query-By-Example Spoken Term Detection using Phonetic Posteriorgram Templates, in ASRU 2009. IEEE Workshop on, Nov 2009, pp. 421426.
- T. W. Lee and M. S. Lewicki, The generalized Gaussian mixture model using ICA, 2000.
- T. M. Nguyen, Q.M. Jonathan Wu, and H. Zhang, Bounded generalized Gaussian mixture model, Pattern Recognition, vol. 47, no. 9, pp. 3132 3142, 2014.
- J. Lindblom and J. Samuelsson, Bounded support Gaussian mixture modeling of speech spectra, Speech and Audio Processing, IEEE Transactions on, vol. 11, no. 1, pp. 8899, Jan 2003.

References

- Y. Zhang, Unsupervised speech processing with applications to query by-example spoken term detection, Ph.D. thesis, MIT. Department of Electrical Engineering and Computer Science., 2013.
- T.J. Hazen, W. Shen, and C. White, Query-By-Example Spoken Term Detection using Phonetic Posteriorgram Templates, in ASRU 2009. IEEE Workshop on, Nov 2009, pp. 421426.
- T. W. Lee and M. S. Lewicki, The generalized Gaussian mixture model using ICA, 2000.
- T. M. Nguyen, Q.M. Jonathan Wu, and H. Zhang, Bounded generalized Gaussian mixture model, Pattern Recognition, vol. 47, no. 9, pp. 3132 3142, 2014.
- J. Lindblom and J. Samuelsson, Bounded support Gaussian mixture modeling of speech spectra, Speech and Audio Processing, IEEE Transactions on, vol. 11, no. 1, pp. 8899, Jan 2003.

References

- Y. Zhang, Unsupervised speech processing with applications to query by-example spoken term detection, Ph.D. thesis, MIT. Department of Electrical Engineering and Computer Science., 2013.
- T.J. Hazen, W. Shen, and C. White, Query-By-Example Spoken Term Detection using Phonetic Posteriorgram Templates, in ASRU 2009. IEEE Workshop on, Nov 2009, pp. 421426.
- T. W. Lee and M. S. Lewicki, The generalized Gaussian mixture model using ICA, 2000.
- T. M. Nguyen, Q.M. Jonathan Wu, and H. Zhang, Bounded generalized Gaussian mixture model, Pattern Recognition, vol. 47, no. 9, pp. 3132 3142, 2014.
- J. Lindblom and J. Samuelsson, Bounded support Gaussian mixture modeling of speech spectra, Speech and Audio Processing, IEEE Transactions on, vol. 11, no. 1, pp. 8899, Jan 2003.

References

- Y. Zhang, Unsupervised speech processing with applications to query by-example spoken term detection, Ph.D. thesis, MIT. Department of Electrical Engineering and Computer Science., 2013.
- T.J. Hazen, W. Shen, and C. White, Query-By-Example Spoken Term Detection using Phonetic Posteriorgram Templates, in ASRU 2009. IEEE Workshop on, Nov 2009, pp. 421426.
- T. W. Lee and M. S. Lewicki, The generalized Gaussian mixture model using ICA, 2000.
- T. M. Nguyen, Q.M. Jonathan Wu, and H. Zhang, Bounded generalized Gaussian mixture model, Pattern Recognition, vol. 47, no. 9, pp. 3132 3142, 2014.
- J. Lindblom and J. Samuelsson, Bounded support Gaussian mixture modeling of speech spectra, Speech and Audio Processing, IEEE Transactions on, vol. 11, no. 1, pp. 8899, Jan 2003.

Thank You!