UNDERWATER ACOUSTIC SIGNAL CLUSTERING WITH PROBABILISTIC LINEAR DISCRIMINANT ANALYSIS AND DIRICHLET PROCESS

Wang Qiang, Zeng Xiangyang, Wang Lu
Key Laboratory of Acoustics and Sensing, Northwestern Polytechnical University
School of Marine Science and Technology, Northwestern Polytechnical University,
China, Xi'An, 710072
E-mail: wqiang0212@mail.nwpu.edu.cn

In underwater acoustic signal clustering system, traditional frame-by-frame hand-craft features failed to catch the invariant discriminative features because of the changeable environmental noise. In this paper, a probabilistic linear discriminant analysis(PLDA) based feature re-extraction model is introduced to extract shared features of multiple frame signal by making use of posterior expectation. PLDA based feature re-extraction method can not only unify the feature dimensions of signal with different last time and extract shared features of multiple frame by minimizing the covariance within the class. Then a Dirichlet process based infinite Gaussian mixture model(DP-GMM) is introduced to deal with the variable number of classes of targets. Aggregation degree and adjust rand index(ARI) are defined to compare the performance of unsupervised clustering methods in this paper. The clustering results of measured data indicate that proposed method can achieve higher aggregation degree and ARI by comparing the shared features with MFCCs.

Keywords: Underwater acoustic signal clustering, PLDA, DP-GMM, Invariant features.

1. Introduction

Automatic underwater signal classification is very interesting and important issue in sonar and autonomous Underwater Vehicles(AUVs) systems. The classification system involves two key parts: feature extraction and pattern recognition. Frame-by-frame hand-craft features, like MFCCs[1], are widely accepted in this application. Since the underwater acoustic environment is very complex, the extracted features of received signal are mutability when the environment acoustic noises are variant with time. Traditional features are failed to make full use of the relevance between the adjacent frame which are generally assumed to be generated by the same target. The essences of noisy signal are difficult to extract by only one frame signal. Though several pattern recognition algorithms can find out discriminative statistic characters of data, it often failed to catch the redundancy but discriminating information especially when the data set is in a small size with high dimensional feature space. The pre-extraction of shared features is effective to enhance the classification system.

Many methods, such as LDA and PCA et.al., are proposed to transform the original features into low dimensional space via subspace projection. Probabilistic version of those methods are also present in several literatures[2, 3]. Here we introduce the probabilistic linear discriminant analysis(PLDA) to extract the shared features of multiple frame signal. PLDA assumes that the density of each class follows a Gaussian distribution with shared covariance structure and aims at reducing the within-class variance and enlarging the between-class variance. Different from the typically use of

LDA as a subspace method, we apply the posterior expectation of hidden factor as the shared features with the given multiple frames. This method can not only transforms the signal with unequal last length into unified dimension but also makes full use of the characters that adjacent segment of signal are always come from the same target. Since the invariant features are of great importance for clustering and classification, by extracting the shared features from multiple frame signal, PLDA make the clustering system more robust.

Furthermore, since many classification algorithms are not well designed for unsupervised learning, such as SVM (support vector machine) and artificial neural networks. While most of underwater signals are collected without label information in reality, unsupervised clustering methods raise more and more concerns in this field. However the number of classes need to be predefined in traditional clustering method, such as k-means and Gaussian mixture model(GMM). The number of target is a variable in real applications, thus Dirichlet process mixture model(DPMM)[4] is applied to model the number of category given by the data set. Dirichlet process is also known as one of the nonparametric Bayesian models, which can automatically find the number of mixtures and estimate the parameters of distribution for each class. The complexity of model can varies with different data set which makes it more flexible in different clustering applications.

This paper is consist of the following parts. In section II we first introduce the clustering procedure with PLDA and DP-GMM. Then the algorithms of PLDA and DP-GMM are given to extracted shared features and cluster the targets into several components, respectively. Finally in the experiments, the aggregation degree and ARI (adjust rand index) are defined to compare the performance of proposed and the original system.

2. Clustering of underwater acoustic signal

The schematic diagram for underwater acoustic signal clustering method in this paper is shown in Fig.1. MFCCs are extract as features for the measured acoustic signals. Then PLDA is trained with data set that contains MFCCs of all frames. In clustering stage, after the MFCCs of multiple segment signals are extracted, the posterior expectation of hidden factor is calculated and re-extracted as new features by PLDA. DP-GMM is finally applied to cluster the new features. The algorithms of PLDA and DP-GMM are shown below in details.

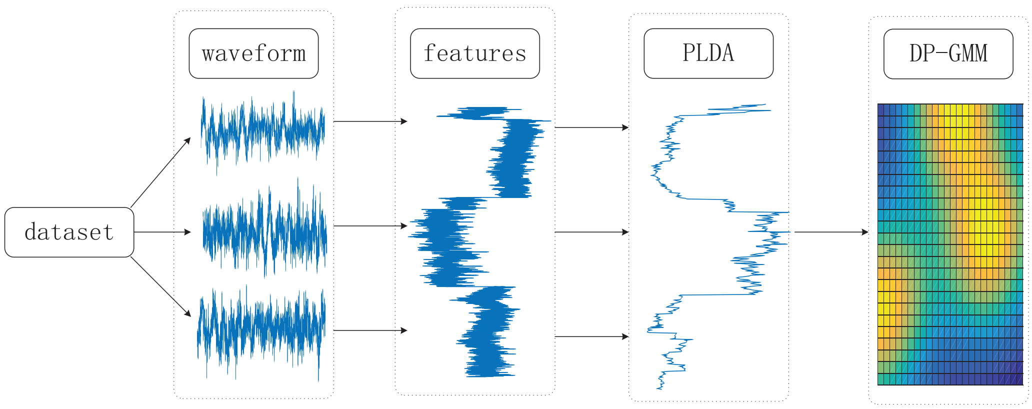

Figure 1: Schematic diagram for underwater acoustic signal clustering method.

3. Probabilistic LDA

Traditional frame-by-frame features ignore the relevance of different frames. Those statistic characters are generally modeled by pattern recognition method. When the data set is in a small size, pattern recognition method are not efficient enough to recover the discriminative features especially when the unsupervised clustering method is adopted. While it has been proved to be valid by many articles[5, 6, 7] that proper transformation of original features can improve the performance of classification system. Prince and Elder[8] proposed a probabilistic LDA(PLDA) approach to enlarge the discriminative information in facial. As mentioned above, PLDA is also introduced in our procedure to mapped the original acoustic features into a low dimensional space. LDA is widely used to enlarge the between class covariance and reduce the within class covariance. Here we exploit this quality to reduce the covariance of the features extracted from adjacent frames which are assumed to be generated by the same target.

In Gaussian PLDA, assumed that the useful information are contains in low-dimentional spaces, the observed dataset $\left\{\mathbf{x}_{ij}|\mathbf{x}_{ij} \in \mathscr{R}^d\right\}, i = 1, 2, \cdots, C, j = 1, 2, \cdots, H_i$ is draw from a probabilistic model of the form:

$$\mathbf{x}_{ij} = \mu + \mathbf{V}\mathbf{z_i} + \mathbf{n} \tag{1}$$

In this model, μ is the global mean of the data, $\mathbf{z}_i$ for $i = 1, 2, \cdots, C$, are the factor of each class that follow a standard Gaussian distribution with k dimension. n is the noise which also follow a Gaussian distribution with zero means, thus $n \sim \mathscr{N}(0, \mathbf{\Sigma})$, $\mathbf{\Sigma}$ is a full $d \times d$ covariance matrix. $\mathbf{V}$ is a low rank matrix, where $k < d$.

Here we note the parameter set as $\theta = \left[\mathbf{V}, \mathbf{\Sigma}, \mu\right]$, the parameters are usually estimated by EM algorithm (see Alg.1). When the parameters and observed dataset $\mathscr{X} = \left\{\mathbf{x}_j|\mathbf{x}_j \in \mathscr{R}^d\right\}$ is given, the expectation of factor can be expressed as:

$$E(\mathbf{z}|\mathscr{X};\theta) = (I + \sum\nolimits_j \mathbf{V}^T\mathbf{\Sigma}\mathbf{V})^{-1}\mathbf{V}^T \sum\nolimits_j \mathbf{\Sigma}^{-1}(\mathbf{x}_j - \mu) \tag{2}$$

Since the data set $\mathscr{X}$ belongs to one class, $E(\mathbf{z}_i|\mathscr{X})$ represents the shared features of given dataset. In underwater target classification, within a given time frame, the input signal can always assumed to be one class, the covariance of features are mainly caused by the time variant noises. Otherwise, the posterior expectation $E(\mathbf{z}_i|\mathscr{X})$ can also unify the dimensions of signals with different last time.

Algorithm 1 EM for PLDA

```
1:  Input: observe set X = {x_nj, n = 1,2,···,C, j = 1,2,···,H_n},MaxIter.
2:  Initialize: Σ = eye(d,d), randomly select k samples from X to form the V.
3:  Output: Σ, V, μ.
4:  μ = (Σ_{i=1}^{C} Σ_{j=1}^{H_i} x_ij) / (Σ_{i=1}^{C} Σ_{j=1}^{H_i} 1).
5:  for i = 1 to MaxIter do
6:      for n = 1 to C do
7:          ⟨z_n|X⟩ =
            (I + H_n V^T Σ^{-1} V)^{-1} V^T Σ^{-1} Σ_{j=1}^{H_n} (x_nj − μ).
8:          ⟨z_n z_n^T|X⟩ =
            (I + H_n V^T Σ^{-1} V)^{-1} + ⟨z_n|X⟩⟨z_n|X⟩^T.
9:      end for
10:     V = [Σ_nj (x_nj − μ)⟨z_n|X⟩] [(Σ_nj ⟨z_n z_n^T|X⟩)]^{-1}.
11:     Σ = 1/(Σ_{n=1}^{C} H_n) Σ_{n=1}^{C} Σ_{j=1}^{H_n} [(x_nj − μ)(x_nj − μ)^T − V⟨z_n|X⟩(x_nj − μ)^T].
12: end for
```

3.1 DP-GMM

It is straightforward to applied a mixture of Gaussian distributions to model the whole data set by inheriting the distribution assumption of PLDA. In this paper, we sample the label assignment of the given signal set from a Dirichlet process. More specifically, we applied an Dirichlet process based infinite Gaussian mixture model (DP-GMM) to solve clustering problem. Dirichlet process is a distribution over distributions, which can handle the number of labels in a generative and self-adapting ways. As the data set becomes larger, the model tends to be more complex. The Dirichlet process based model is suitable choice to deal with the clustering problem of underwater target.

Given the data set $X = \{\mathscr{X}_1, \mathscr{X}_2, \cdots, \mathscr{X}_K\}$,where K is the total number of components. The features of each class are assumed to follow a multivariate Gaussian distribution, then a Gaussian-inverse-Wishart distribution is applied for mean vector and covariance matrix as fully conjugate distribution:

$$\mu, \mathbf{\Sigma}|\mu_0, \mathbf{\Sigma}_0, \alpha_0, \beta_0 \sim \mathscr{N}(\mu|\mu_0, \frac{1}{\alpha_0}\mathbf{\Sigma}) IW(\mathbf{\Sigma}|\mathbf{\Sigma}_0, \beta_0) \tag{3}$$

where μ_0 and $\mathbf{\Sigma}_0$ can be calculated by the maximum likelihood estimation. α_0 and β_0 are hype-parameters which control the similarity between the sampled distribution and the prior distribution. The graphic model of DP-GMM is shown in Fig.2.

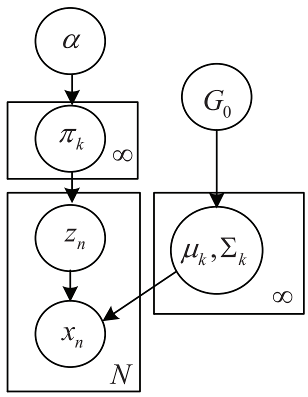

Figure 2: Graphical model for DP-GMM.

Considering following model for DP-GMM:

$$G_0 \sim DP(\alpha_0, G) \tag{4}$$

$$\mu_k, \mathbf{\Sigma}_k \sim G_0 \tag{5}$$

$$\pi|\alpha \sim Dir(\pi|\alpha) \tag{6}$$

$$\mathbf{z}|\pi \sim Mult(\mathbf{z}|\pi) \tag{7}$$

$$\mathbf{x}_n|z_n = k, \mu_k, \mathbf{\Sigma}_k \sim N(\mathbf{x}_n|\mu_k, \mathbf{\Sigma}_k) \tag{8}$$

The marginal distribution of data set $\mathscr{X}_k$ of k^{th} component is given by integrating out the $\mu_k, \mathbf{\Sigma}_k$:

$$p(\mathscr{X}_k) = \int_{\mu_k} \int_{\mathbf{\Sigma}_k} p(\mathscr{X}_k|\mu_k, \mathbf{\Sigma}_k) p(\mu_k, \mathbf{\Sigma}_k)\, d\mu_k\, d\mathbf{\Sigma}_k \tag{9}$$

For a new sample $\mathbf{x}_i$, the component assignment z is sampled from:

$$\begin{aligned} p(z_i = k|\mathbf{z}, \mathbf{x}_i, \mathscr{X}) &\propto p(z_i = k|\mathbf{z}) p(\mathbf{x}_i|z_i = k, \mathbf{z}, \mathscr{X}) \\ &\propto p(z_i = k|\mathbf{z}) p(x_i|\mathscr{X}_k) \end{aligned} \tag{10}$$

where, $p(\mathbf{x}_i|\mathscr{X}_k) = p(\mathbf{x}_i, \mathscr{X}_k)/p(\mathscr{X}_k)$ can calculated by Eq.9, and $p(z_i = k|\mathbf{z})$ are draw from a Chinese restaurant process (CRP):

$$p(z_i = k|\mathbf{z}) = \begin{cases} \frac{N_k}{N-1+\alpha}, if\ N_k > 0 \\ \frac{\alpha}{N-1+\alpha}, if\ N_k = 0 \end{cases} \tag{11}$$

N_k denote the number of samples belongs to k^{th} class and hyper-parameters are ignored in above formulas for the simplicity. DP-GMM is trained by a collapsed Gibbs sampler (see Alg.2).

Algorithm 2 Collapsed Gibbs sampler for DP-GMM

Input: observe set $X = \{\mathbf{x}_n, n = 1, 2, \cdots, N\}$,MaxIter
Initialize: choose an initial label assignment z
Output: z
for $i = 1$ to $MaxIter$ **do**
  for $n = 1$ to N **do**
    remove the label assignment z_n
    for $k = 1$ to K **do**
      calculate $p(z_n = k|\mathbf{z})$ by Eq.11).
      calculate $p(x_n|\mathscr{X}_k) = p(x_n, \mathscr{X}_k)/p(\mathscr{X}_k)$.
      calculate $p(z_n = k|\mathbf{z}, \mathbf{x}_n, \mathscr{X})$ by Eq.10.
    end for
    for a new class, calculate $p(z_n = k|\mathbf{z})$ by Eq.11.
    calculate $p(\mathbf{x}_n)$ by Eq.9.
    calculate $p(z_n = k|\mathbf{z}, \mathbf{x}_n, \mathscr{X}) \propto p(z_n = k|\mathbf{z})p(\mathbf{x}_n)$.
    calculate $p(z_n = k|\mathbf{z}, \mathbf{x}_n, \mathscr{X})$ by normalizing the above formula.
    sample k from the $p(z_n = k|\mathbf{z}, \mathbf{x}_n, \mathscr{X})$, set $z_n = k$.
    remove the empty component and adjust K.
    update $\mu_k, \mathbf{\Sigma}_k$ for each component.
  end for
end for

4. Experiment

The measured data contains 3 different ships with fixed sample frequency 8 kHz. Data of each ship are separated into 15 pieces within different time intervals with label from 1 to 15. Those signals are further segmented into frames and each frame contains 256 data point (almost 30 ms). For each frame, MFCCs are firstly extract with 24 Mel filters with frequencies range from 20 Hz to 4 000 Hz. The first 13 dimension MFCCs are reserved for further analysis. For the DP-GMM, the samples in the database are firstly separated into 20 components randomly. The MaxIter in the Alg.2 is set to be 50. The number of components will be learned from the sampling process of DP-GMM.

To compare the performance of different methods we defined following two parameters, the aggregation degree and ARI. Firstly P_{ij} is defined as follow:

$$P_{ij} = \frac{N_{ij}}{N_i} \tag{12}$$

N_{ij} denotes the number of samples which belong to i^{th} class and divided into j^{th} component of GMM. $j = 1, 2, \cdots, K$, N_i is number of samples that belongs to i^{th} class. N is the total number. Set $m_i = \arg\max(P_{i1}, P_{i2}, \cdots, P_{iM})$. For each class with the same label, aggregation degree is given by following formula:

$$S_i = P_{im_i} \tag{13}$$

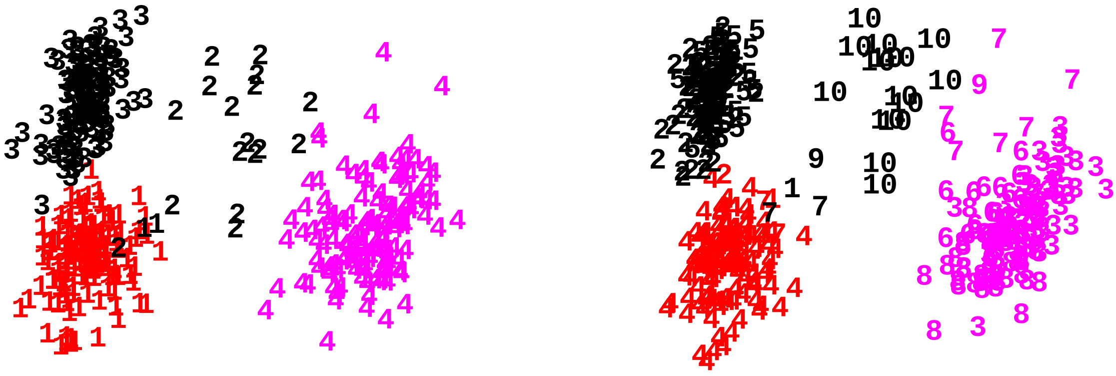

Figure 3: Clustering result of proposed method(left) and MFCCs(right).

And ARI is defined by following equation:

$$ARI = \frac{\sum_{ij}\binom{N_{ij}}{2} - \frac{2ab}{N(N-1)}}{\frac{1}{2}(a+b) - \frac{2ab}{N(N-1)}} \tag{14}$$

where $a = \sum_j \binom{\sum_i N_{ij}}{2}$, $b = \sum_i \binom{\sum_j N_{ij}}{2}$ and $\binom{N}{K}$ denotes the number of combinations of N things taken K. A higher ARI value indicates a better clustering performance.

In this paper MFCCs of adjacent 18 frames (0.54s) are used to estimated posterior expectation of hidden factor by Eq.2. Then those expectations are applied to training DP-GMM. The clustering results are shown in Fig.3. Three target are drawn with different colors, the labeled number 1 to 4 denote the different components of GMM. We can see that our method works well for this data set except 3^{th} class in black are segmented into two components of the GMM. Compared with proposed method, the clustering results based on MFCCs tend to have a larger number of mixtures than that based on the proposed method. This means the aggregation of features re-extracted by PLDA is much higher than the MFCCs.

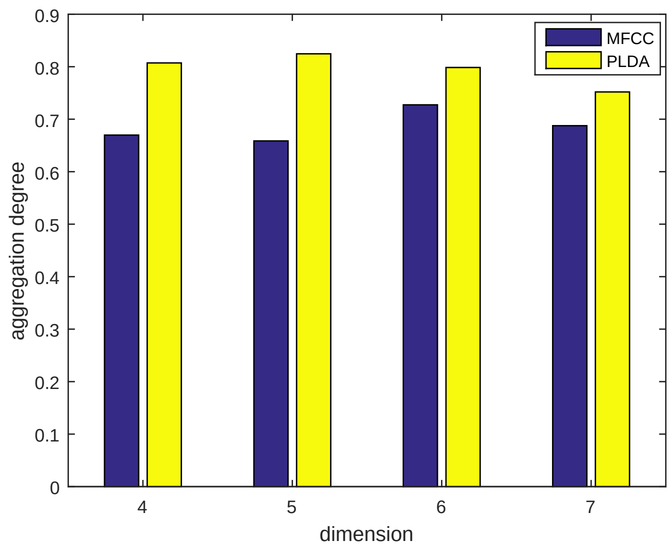

Figure 4: Aggregation degree with different dimensions of features.

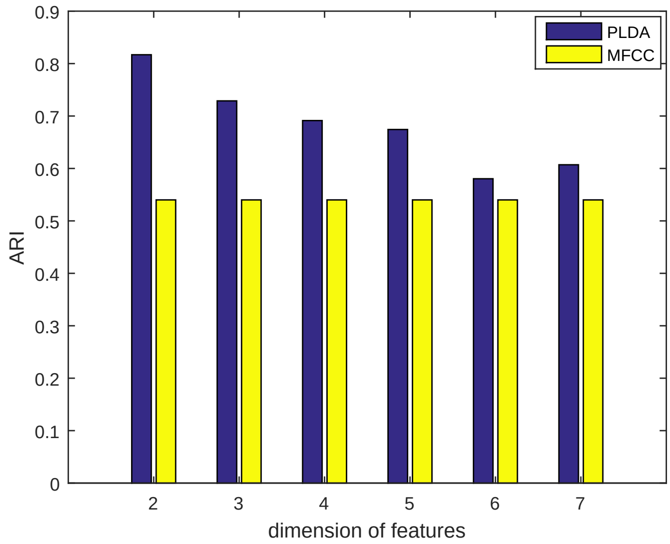

Figure 5: ARI of different features.

In the experiment, because of the randomness of Gibbs sampling in DP-GMM, we repeat clustering process 50 times to generate more comprehensive results. The average of aggregation and ARI are shown in Fig.4 and Fig.5 respectively.

Fig.4 shows that the the aggregation degree of features re-extracted by PLDA are higher than original frame-by-frame MFCCs as was expected. Meanwhile the ARIs of re-extracted features are also much higher than MFCCs in the low dimensions. The number of clustering components of proposed features is closer to the practice than that of MFCCs. Those indicators show that the proposed features outperform the MFCCs in the clustering task. That means the extraction of invariant features can efficiently improve the performance of the clustering system.

5. Conclusion

In this paper we proposed a PLDA based feature re-extraction method to acquire shared features of signals from multiple frames. Due to the complex underwater acoustic environment, the clustering performance of traditional features can not meet the expectation. By making a reasonable assumptions that the signals of adjacent frames are the same target, the invariant feature of those frames extracted by PLDA achieved higher aggregation degree and ARI in the new feature space. As a clustering method, DP-GMM is also introduced to model the number of targets which is crucial important in the real application. The clustering results of proposed method outperform the original method as expect. This work advances the real application of unsupervised clustering system in underwater target classification.

REFERENCES

1. Lim, T., Bae, K., Hwang, C. and Lee, H. Classification of underwater transient signals using MFCC feature vector, *Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on*, pp. 1–4, IEEE, (2007).

2. Li, N. and Mak, M.-W. SNR-invariant PLDA modeling in nonparametric subspace for robust speaker verification, *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, **23** (10), 1648–1659, (2015).

3. Tipping, M. E. and Bishop, C. M. Probabilistic principal component analysis, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61** (3), 611–622, (1999).

4. Neal, R. M. Markov chain sampling methods for Dirichlet process mixture models, *Journal of Computational and Graphical Statistics*, **9** (2), 249–265, (2000).

5. Sahoolizadeh, A. H., Heidari, B. Z. and Dehghani, C. H. A new face recognition method using PCA, LDA and neural network, *International Journal of Computer Science and Engineering*, **2** (4), 218–223, (2008).

6. Pang, Y., Zhang, L., Li, M., Liu, Z. and Ma, W., (2004), A novel Gabor-LDA based face recognition method. *Advances in Multimedia Information Processing-PCM 2004*, pp. 352–358, Springer.

7. Anderson, D. N. and Taylor, S. R. Application of regularized discrimination analysis to regional seismic event identification, *Bulletin of the Seismological Society of America*, **92** (6), 2391–2399, (2002).

8. Prince, S. J. and Elder, J. H. Probabilistic linear discriminant analysis for inferences about identity, *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8, IEEE, (2007).