SSPD 2023
Conference

Underwater Passive Target Classification based on β Variational Autoencoder and MFCC

Adarsh Sunilkumar Shamju Joseph K Manoj Kumar K
adarsh.npol@gov.in shamju.npol@gov.in manojkk.npol@gov.in
Naval Physical and Oceanographic Laboratory (DRDO),Kochi, India

Introduction

- Classification of underwater passive target refers to processing of the radiated noise from the target and identifying the type of the target.
- Open set in nature and it is a challenging task due to the intrinsic complexity of the radiated noise from the target.
- Conventional classification architectures with spectral processing often fail.
- Supervised learning methods like deep learning, offers higher success rate but they require enormous amount of data for training and their performance in open set classification is again a challenge.

Contribution

- Based on Beta Variational Autoencoder (β − VAE) model with Mel Frequency Cepstral Coefficients (MFCCs) features.
- MFCC effectively utilises the non-linear auditory effect of the human ear with different frequencies.
- β – VAE, being one of the generative models, is capable of generalizing with less amount of data.
- Followed Open set architecture.
- Unsupervised learning.

Proposed Method

Mel Frequency Cepstral Coefficients(MFCC)

- Human auditory system has a non-linear characteristics, and it is more sensitive to low frequencies.
- MFCC is based on Mel frequency, which can well characterise this non-linearity[1].

The process involved in obtaining MFCC feature vector from the sound signal is given below,

a) In the pre-process step, framing and windowing is applied on the signal.
b) Apply the Short Time Fourier Transform and perform the power spectrum calculation.
c) Map the linear power spectrum into non-linear one on mel scale with the application of triangular filter banks.
d) Apply the log of these spectrum values to obtain the log filter bank energies.
e) Take the discrete cosine transform of this log filter bank energies.
f) MFCCs are the amplitude of the resultant spectrum.

β−VAE

VAEs are deep generative networks which have both encoder and decoder networks similar to auto encoders. They learn to map their input X to latent representation z, by learning the probabilistic distribution Q(z/X). VAE assumes that input X and z follow isotropic Gaussian distribution[2].

The encoder and decoder network of VAE is given by $Q_\varphi(z|X)$ and $P_\theta(X|z)$ respectively. φ and θ are neural network parameters.

The objective function of VAE is given below in (1),

$$L_{VAE}(\varphi, \theta; X) = L_{RC} + L_{KL} \quad (1)$$

where L_{RC} is the reconstruction loss and L_{KL} is the Kullback-Leibler (KL) divergence loss.

β−VAE introduces the use of Lagrange multiplier β on the KL divergence term in the original VAE formulation.

The objective function of the β−VAE is denoted as shown in (2) ,

$$L_{VAE} = L_{RC} + \beta * L_{KL} \quad (2)$$

Algorithm

- Let S be the sample audio signal to be classified, it is split into multiple frames.
- Sf represents MFCC feature vector of each frame of S. N represents total number of frames.
- T denotes the threshold, which is the minimum number of frames out of N, in which the proposed method demands that the frequency of any one of the models belongs to a particular class, to declare that the sample belongs to that particular class.
- Let $clfr_t$ is a classifier created for some class t, its 99th percentile error is et^{99} and clfr is the set of classifiers created for each class of targets.

Algorithm 1: Classifier Algorithm
Input: Audio of the target.
Output: Class of the target.

1: Create $clfr_t$, et^{99}, t=1,2,...,M, M is the total number of trained classes.
2: **For** each Sf in S do
3: {minError,class}={}
4: **For** each $clfr_t$ in clfr **do**
5: e = getReconstructionError($clfr_t$,Sf)
6: **if** (e <= et^{99}) then
7: update({minError,class})
8: **end if**
9: **end for**
10: **end for**
11: {class,frequency} = getMostFrequentClass()
12: **if** (frequency < T) then
13: return “ ”
14: **end if**
15: return class

Results

- compared with One Class Support Vector Machine (SVM) and Isolation Forest on two data sets.
- first one consists of three underwater targets, one ship and two submarine targets, they are named as class A, B and C respectively.
- Second data set consist of five ship targets, namely class D, E, F, G and H.
- All data have been collected during various expeditions conducted in the Indian ocean with passive sonar systems.
- First data set is collected from sonar fitted on ship 1 and second set is collected from sonar fitted in ship 2.
- Data from underwater targets belong to classes other than the ones used for training, called unknown classes is also included in the test phase. U denotes unknown class data.
- Those data which are declared as unclassified by the proposed method is denoted by UC.
- Each audio data in data set 1 & 2 consists of 512 millisecond duration.

Method	Feature Used	Accuracy	Precision	Recall
Proposed Method	Delta MFCC	94.11%	93.61%	94.67%
One Class SVM	Delta MFCC	71.92%	84.8%	73.63%
Isolation Forest	Delta MFCC	68.45%	84.07%	73.68%

Table 1. Result of data set 1.

Method	Feature Used	Accuracy	Precision	Recall
Proposed Method	Delta MFCC	96.14%	95.91%	96.13%
One Class SVM	Delta MFCC	74.06%	84.74%	75.78%
Isolation Forest	Normal MFCC	80.11%	86.14%	81.41%

Table 2. Result of Data Set 2.

Research Impact in Defence and Conclusion

- Compared with one class SVM and isolation forest models and it performs better than both of the methods.
- Method attained maximum accuracy, when the feature used is delta MFCC, with β value 4 and SELU activation function.
- Suitable in anti-submarine warfare scenarios, where it is desirable to minimize dependency on human operator.
- Scalable and it can perform online learning without retraining the entire models.
- Less misclassification error and it is effective in identifying unknown classes as well.

References

[1] Yuze Tong, Xin Zhang, Yizhou Ge, “Classification and Recognition of Underwater Target Based on MFCC Feature Extraction”, 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC).

[2] Decebal Constantin Mocanu and Elena Mocanu. “One-Short Learning using Mixture of Variational Autoencoders: a Generalization Learning approach”. 17th International Conference on Autonomous Agents and Multiagent systems (AAMAS 2018).