AUTOMATIC PATTERN ANALYSIS OF BIOACOUSTIC SIGNALS: EXPLORING SHALLOW AND DEEP LEARNING FRAMEWORKS

A THESIS

submitted by

ANSHUL THAKUR

for the award of the degree

of

DOCTOR OF PHILOSOPHY

SCHOOL OF COMPUTING AND ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY MANDI

To deal with hyper-planes in a 14-dimensional space, visualize a 3-D space and say "fourteen" to yourself very loudly. Everyone does it.

Geoffrey E. Hinton, *A geometrical view of perceptrons.*

Declaration

I hereby declare that the entire work embodied in this thesis is the result of investigations carried out by me in the **School of Computing and Electrical Engineering, Indian Institute of Technology Mandi**, under the supervision of **Dr. Padmanabhan Rajan**, and that it has not been submitted elsewhere for any degree or diploma. In keeping with the general practice, due acknowledgments have been made wherever the work described is based on finding of other investigators.

Mandi, 175005

Anshul Thakur

Date:

THESIS CERTIFICATE

This is to certify that the thesis titled **AUTOMATIC PATTERN ANALYSIS OF BIOACOUSTIC SIGNALS: EXPLORING SHALLOW AND DEEP LEARNING FRAMEWORKS**, submitted by **Anshul Thakur**, to the Indian Institute of Technology, Mandi, for the award of the degree of **Doctor of Philosophy**, is a bonafide record of the research work done by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other institute or university for the award of any degree or diploma.

Mandi, 175005

Date:

Dr. Padmanabhan Rajan

(Ph.D. Supervisor)

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Padmanabhan Rajan, for his motivation, guidance and incredible support. Without his guidance and constant feedback, this PhD would not have been achievable. I would also like to extend my sincere gratitude towards my doctoral committee members: Dr. Dileep A. D., Dr. Satyajit Thakor and Dr. Viswanath Balakrishnan. They have inspired me through their own research, and their willingness to answer my questions has helped me greatly over the years.

I am indebted to all my collaborators and fellow scholars at MANAS Lab, IIT Mandi for creating a harmonious learning environment. In particular, I am thankful to Vinayak Abrol, Pulkit Sharma, Arjun Pankajakshan, Arshdeep Singh and Daksh Thapar for their frequent help and motivation.

I gratefully acknowledge the funding received towards my PhD from the HTRA fellowship, provided by Ministry of Human Resource Development, India.

Last but not the least, I would like to thank my parents for supporting me in all of my "adventures".

Anshul Thakur

Abstract

Sounds produced by living organisms are called bioacoustic signals. These bioacoustic signals can be analysed to track organisms like birds, amphibians and mammals in their natural habitats. This thesis presents various machine learning frameworks to automatically analyse the bioacoustic signals. One of the challenges in developing machine learning frameworks for bioacoustic pattern analysis is scarcity of labelled training data. As a result, there is a requirement of machine learning frameworks that can overcome this problem, and work effectively under the low-training data conditions. This thesis mainly addresses the development of such data-efficient frameworks. It also deals with the development of standard data-intensive machine learning methods for bioacoustic applications where a sufficient amount of labelled training data is available.

This thesis explores the contrastive paradigms of shallow and deep learning to introduce frameworks for bioacoustic pattern analysis, in particular, bioacoustic activity detection, segmentation and classification. In shallow learning based frameworks, the concepts of dynamic kernels, semi-supervision and matrix factorization are utilised. These frameworks are demonstrated to have low training data requirements and hence, are suitable for many bioacoustic applications. On the contrary, for bioacoustic applications where enough labelled training data is readily available, deep learning frameworks are proposed to emphasize the performance. Apart from the standard deep learning methods, this thesis also explores meta-learning, in particular, deep metric learning to train large neural networks effectively in data-scarce scenarios.

In this thesis, a computationally efficient variant of probabilistic sequence kernel (PSK) is proposed for the task of bioacoustic activity detection. Unlike the existing formulation of PSK, the proposed PSK does not require background modelling and utilises only a Gaussian mixture model (GMM) for bioacoustic activity class. Moreover, only a few most relevant components of this GMM are utilised for the kernel formulation, making the

whole setup computationally efficient. Apart from this, an all-convolutional neural network (all-conv net) is also proposed for activity detection. This neural network consists of only convolutional layers, and utilises *learned pooling* or strided convolutions to down-sample the feature maps. In contrast to max-pooling, the learned pooling helps in capturing the inter-feature map correlations, leading to a better representation.

Next, this thesis proposes a semi-supervised framework and a weakly supervised neural network for the task of bioacoustic signal segmentation. The proposed semi-supervised framework requires only a few strongly labelled training examples, and utilises the correlation between training examples and the test audio recordings to discriminate between the target bioacoustic events and the background. On the other hand, multi-instance learning is incorporated in the all-conv net to provide weakly supervised segmentation.

Next, this thesis explores the utilisation of archetypal analysis (AA), a matrix factorization method, to model the bioacoustic data using its convex hull or extremal elements. Building on AA, a deep matrix factorization framework, referred to as *deep archetypal analysis* (DAA) is proposed. DAA improves the modelling capabilities of AA as it can model both extremal as well as average behaviour of the data. Both AA and DAA are employed in simplex projection based dictionary learning framework and in dynamic kernel formulations for developing bioacoustic classification frameworks. In comparison to other acoustic modelling methods, AA/DAA requires a lesser amount of data to effectively model the variations present in a class, making them appropriate for bioacoustic classification.

Finally, this thesis explores deep metric learning (DML) to propose a data-efficient bioacoustic classification framework that utilises the triplet loss function with dynamically increasing margin. This dynamically varying margin allows the framework to re-use the training data without introducing redundancy in the training process.

The experimental evaluation on publicly available and licensed datasets demonstrates that the proposed frameworks provide either better or comparable performance than state-of-the-art bioacoustic methods.

TABLE OF CONTENTS

Acknowledgments i

Abstract ii

List of Tables x

List of Figures xii

List of Abbreviations xvii

Chapter 1: Introduction 2

1.1 Motivation 2

1.2 Machine learning for bioacoustic pattern analysis 3

1.3 Challenges 6

1.4 Objectives and scope of the thesis 7

1.5 Original contributions of the thesis 8

1.6 Publications 10

1.7 Outline 11

Chapter 2: Literature Overview 15

2.1 Existing approaches for bioacoustic activity detection 15

2.1.1 Shallow learning frameworks 16

2.1.2 Deep learning frameworks 17
2.1.3 Domain adaptation used in existing studies 19
2.2 Existing approaches for bioacoustic signal segmentation 21
2.2.1 Unsupervised segmentation methods 22
2.2.2 Supervised segmentation methods 23
2.2.3 Weakly supervised segmentation 25
2.3 Existing approaches for bioacoustic classification 27
2.3.1 Traditional machine learning frameworks 28
2.3.2 Deep learning frameworks 31
2.4 Possible research directions 33

Chapter 3: Bioacoustic Activity Detection 35

3.1 Problem formulation 35
3.2 Computationally efficient PSK for bioacoustic activity detection 36
3.2.1 Dynamic kernels 36
3.2.2 Proposed formulation of Probabilistic sequence kernel (PSK) 37
3.2.3 Proposed bioacoustic activity detector 39
3.3 All-convolutional neural network 43
3.3.1 Feature representation 43
3.3.2 All-conv architecture 43
3.3.3 Domain adaptation using generative adversarial networks (GAN) . . 49
3.4 Experimental design 51
3.4.1 Datasets used 52

3.4.2 Experiments 53
3.4.3 Comparative methods 55
3.4.4 Parameter setting 56
3.5 Results and Discussion 58
3.5.1 Performance comparison 58
3.5.2 Short-term analysis in bioacoustic applications 63
3.5.3 Computational efficiency of the proposed PSK based framework . . 64
3.6 Conclusion 65

Chapter 4: Bioacoustic signal segmentation 68

4.1 Problem formulation 68
4.2 Major challenges in bioacoustic signal segmentation 69
4.3 Semi-supervised bioacoustic segmentation 70
4.3.1 Directional Embedding 71
4.3.2 Proposed semi-supervised framework 81
4.4 Weakly supervised bioacoustic segmentation 85
4.4.1 Refining segmentation predictions 89
4.5 Experimental design 90
4.5.1 Datasets used 90
4.5.2 Experiments 91
4.5.3 Comparative Methods 93
4.5.4 Parameter settings used 94
4.6 Results and Discussion 96

4.6.1 Performance comparison 96

4.6.2 Second Experiment: Generic nature of the proposed semi-supervised framework 98

4.6.3 Third Experiment: Performance evaluation of MM-CNN 98

4.6.4 Effect of w and Z on segmentation performance of the proposed semi-supervised framework 100

4.7 Conclusion 101

Chapter 5: Bioacoustic signal classification 103

5.1 Archetypal Analysis 103

5.2 Modelling extremal as well as average behaviour: Local Archetypal Analysis and Deep Archetypal Analysis 105

5.2.1 Local Archetypal Dictionaries 106

5.2.2 Deep Archetypal Analysis 107

5.3 Convex representation for bioacoustic classification 110

5.3.1 Compressed super-frames 110

5.3.2 Training 111

5.3.3 Testing: Classifying an input vocalization 114

5.3.4 Decreasing inter-dictionary correlation 114

5.4 Deep archetypal analysis based intermediate matching kernel 115

5.4.1 Intermediate matching kernel (IMK) 116

5.4.2 Proposed AA/DAA-IMK 118

5.5 Experimental Design 121

5.5.1 Experiments and datasets 121

5.5.2 Comparative methods 123

5.5.3 Train-test distribution 124
5.5.4 Parameter Setting 125
5.6 Results and Discussion 127
5.6.1 Performance comparison 127
5.6.2 Size of pruned dictionaries vs classification performance 130
5.6.3 Effect of context window size (W) 130
5.6.4 Effect of depth on classification performance in DAA-IMK 131
5.7 Conclusion 133

Chapter 6: Deep metric learning for bioacoustic classification 136

6.1 Why deep metric learning (DML)? 137
6.2 Proposed DML framework for bioacoustic classification 138
6.2.1 Feature Extraction 139
6.2.2 Neural Network Designs 140
6.2.3 Multiscale CNN Training: Dynamic Triplet Loss 145
6.2.4 Classification 147
6.3 Proposed DML framework for open-set classification 148
6.4 Experimental Setup 149
6.4.1 Datasets Used 149
6.4.2 Data pre-processing and train-test distribution 151
6.4.3 Comparative studies and performance metric 153
6.4.4 Parameter Setting 155
6.5 Results and discussion 156

6.5.1 Classification Performance 156
6.5.2 Performance of open-set classification module 158
6.5.3 Generalization of the proposed DML framework 159
6.5.4 Dynamic vs. fixed margin triplet loss 160
6.5.5 Ablation Study 161
6.6 Conclusion 164

Chapter 7: Conclusion and future work 166
7.1 Overcoming the major challenges 166
7.2 Trade-off between shallow and deep learning frameworks 167
7.3 Issues not addressed 168
7.4 Directions for further work 169
7.5 Introspection 170

References 186

Appendix A: List of datasets used 188

Appendix B: Links to code 189