A Classification Scheme Based on Directed Acyclic Graphs for Acoustic Farm Monitoring

Stavros Ntalampiras
University of Milan
Milan, Italy
stavros.ntalampiras@unimi.it

***Abstract*—Intelligent farming as part of the green revolution is advancing the world of agriculture in such a way that farms become evolving, with the scope being the optimization of animal production in an eco-friendly way. In this direction, we propose exploiting the acoustic modality for farm monitoring. Such information could be used in a stand-alone or complimentary mode to monitor constantly animal population and behavior. To this end, we designed a scheme classifying the vocalizations produced by farm animals. More precisely, we propose a directed acyclic graph, where each node carries out a binary classification task using hidden Markov models. The topological ordering follows a criterion derived from the Kullback-Leibler divergence. During the experimental phase, we employed a publicly available dataset including vocalizations of seven animals typically encountered in farms, where we report promising recognition rates outperforming state of the art classifiers.**

I. Introduction

The area of Computational Bioacoustic Scene Analysis has received increasing attention by the scientific community in the last decades [1], [2], [3], [4]. Such interest is motivated by the potential benefits that can be acquired towards addressing major environmental challenges including invasive species, infectious diseases, climate and land-use change, etc. Availability of accurate information regarding range, population size and trends is crucial for quantifying the conservation status of the species of interest. Such information can be obtained via classical observer-based survey techniques; however these are becoming inadequate since they are a) expensive, b) subject to weather conditions, c) cover a limited amount of time and space, etc. To this end, autonomous recording units (ARUs) are extensively employed by biologists [5], [6]. An ARU which could be useful for the specific application is available at https://www.wildlifeacoustics.com/products/song-meter-sm4. This is also motivated by the cost of the involved acoustic sensors which is constantly decreasing due to the advancements in the field of electronics.

One of the first approaches employed for classifying animal vocalizations is described in [7]. The authors extracted Linear predictive coding coefficients, cepstral coefficients based on the Mel and Bark scale, along with time-domain features describing the peaks and silence parts of the waveform. The classifier was a Support Vector Machine, while three kernels were considered, i.e. polynomial, radial basis function, and sigmoid. These were compared with nearest neighbor and linear vector quantization schemes. The specific dataset included sounds of four animal classes, i.e. birds, cats, cows, and dogs. The literature further includes several approaches which concentrate on specific species, classification of Australian anurans [8], interpretation of chicken embryo sounds [9], classification of insects [10], etc. However, a systematic approach addressing the specific case of farm monitoring, is not present in the literature. This work intended to cover exactly this gap.

Indeed, the acoustic modality could provide complementary information to monitor the health as well as population of animals. For example it could be used in combination with solutions such as [11], [12], [13] which record physiological parameters of the animals, such as rumination, body temperature, and heart rate with surrounding temperature and humidity. The valuable information that can be obtained via the acoustic modality could assist an overall assessment of the current status of the animals as well as the farm in general. More precisely, acoustic farm environment monitoring could assist in the following applications:

- tracking of similar breed animals and parturitions,
- identification of specific animal(s) for several reasons (vaccination, medication, diseases, diet, etc.),
- animal health monitoring,
- population monitoring,
- detect animals missing from the farm, and
- intruder detection and identification.

Of course, this is a non-exhaustive list of the potential applications, while the overall aim is to optimize animal production.

This work aims at constructing a comprehensive classification scheme, the operation of which does not follow the black-box logic, i.e. where one is able to 'open' the classifier, and by inspecting the misclassifications, obtain clear insights on how the performance can be boosted. At the same time, the proposed system is designed keeping in mind that it may have to operate under non-stationary conditions [14], where distributions followed by the known classes may evolve over time (e.g. due to noise, reverberation effects, etc.), new classes may appear (e.g. new species), etc. Such obstacles require a scheme able to incorporate changes during its operation, and address the evolving phenomena by appropriately altering its structure.

Keeping these in mind, we employed a well-known feature set in combination with a classification scheme adopting a directed acyclic graph structure. There, the topological ordering problem is addressed by means of an approach based on the Kullback-Leibler divergence measured among the different sound classes. During the experiments, we used part of the dataset called Environmental Sound Classification-10 described in [15] which includes the animals typically

ISSN 2305-7254

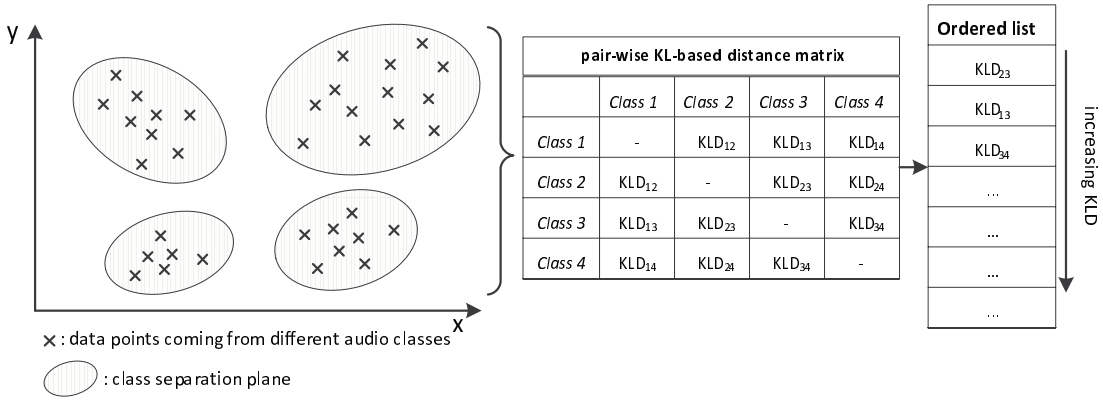

Fig. 1. The determination of the topological ordering (for simplicity, only four classes are considered)

encountered in a farm environment, i.e. *dog*, *rooster*, *pig*, *cow*, *cat*, *hen*, and *sheep*. There, a preliminary classification analysis on the entire dataset provided a recognition rate of 72.7%, while a more recent effort [16] based on convolutional neural networks achieved approximately 80%. Moreover, we performed a comparison with other classification schemes (echo state network, class-specific and universal hidden Markov models, support vector machines, and random forest) and feature sets (MPEG-7 audio standard and perceptual wavelet packets), a process which demonstrated the superiority of the proposed approach.

The rest of this article is organized as follows: section II formulates the problem, while section III details the proposed sound classification framework including the formalization of the Directed Acyclic Graph (DAG) HMM and its topological ordering. Section IV provides information on the dataset we employed, the contrasted approaches, and presents and analyzes the experimental results. Finally, section V concludes this work.

II. Problem Formulation

We suppose a single-channel audio datastream, y_t the duration of which is unknown. y may be emitted by various sources belonging to the set $\mathcal{C} = \{C_1, \ldots, C_m\}$, where m is the number of a-priori known sources. It is further assumed that each source is stationary, i.e. it follows a consistent, yet unknown probability density function $\mathcal{P}$. Moreover, we assume that each recording includes data of one animal, leaving the case of overlapping sounds as future research.

We assume that an initial training sequence $TS = y_t, t \in [1, T_0]$ is available characterized by stationary conditions and containing supervised pairs (y_t, C_i), where t is the time instant and $i \in [1, m]$. The overall aim of the sound recognition system (possibly operating after a separation framework, e.g. [17]) is to correctly label the incoming novel audio stream.

III. The proposed sound classification framework

The proposed framework relies on the Directed Acyclic Graph logic [18], i.e. the classification scheme is a graph denoted as $\mathcal{G} = \{N, L\}$, where $N = \{n_1, \ldots, n_m\}$ represents the nodes and $L = \{l_1, \ldots, l_k\}$ the links associating the nodes. Each node in N is responsible for a binary classification task conducted via a set of hidden Markov models (HMM) which fit well the specifications of audio pattern recognition tasks, thus the DAG-HMM notation.

The motivation behind creating such a graph-based classification system is that in this way, one is able to limit the problem space and design classification algorithms for two mutually-exclusive classes rather than having to deal with the entirety of the different classes at the same time. Essentially, the proposed methodology breaks down any C_m-class classification problem into a series of 2-class classification problems.

DAGs can be seen as a generalization of the class of Decision Trees, while the redundancies and repetitions that may occur in different branches of the tree can be observed more efficiently since different decision paths might be merged. In addition, DAGs are able to collect and conduct a series of tasks in an ordered manner, subject to constraints that certain tasks must be performed earlier than others. The sequential execution of tasks is particularly important and directly related to the efficacy with which the overall task is addressed [19].

The DAG-HMM architecture used in this paper includes $m(m-1)/2$ nodes, each one associated with a two-class classification problem. The connections between the different nodes in $\mathcal{G}$ have only one orientation without any kind of loop(s). As an result each node of a so-called *rooted* DAG has either 0 or 2 leaving arcs.

The following subsections provide a detailed analysis of the way the DAG-HMM is constructed and subsequently operates. The principal issue associated with the design of every DAG is the *topological ordering*, i.e. ordering the nodes in a way that the starting endpoints of every edge occur earlier than the corresponding ending endpoints. In the following, we describe how such a topological ordering is discovered based on the Kullback-Leibler divergence.

A. Determining the topological ordering of the DAG-HMM

Naturally, one would expect that the performance of the DAG-HMM depends on the order in which the different classification tasks are conducted. This was also evident from early experiments. This observation motivated the construction of the DAG-HMM so that ”simple” tasks are executed earlier in the graph. In other words, these are placed in the top nodes of the DAG-HMM, in a way that classes responsible for a

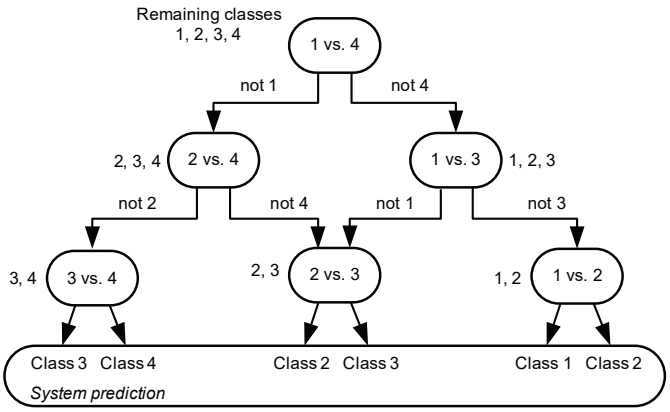

Fig. 2. An example of a DAG-HMM addressing a problem with four classes

high amount of misclassifications are discarded early in the graph operation. In order to get an early indication of the degree of difficulty of a classification task, we employed the metric representing the distance of the involved classes in the probabilistic space, i.e. the Kullback-Leibler Divergence (KLD) between per-class GMMs in the feature space. The basic motivation is to place early in the DAG-HMM tasks concerning the classification of classes with large KLD, as they could be completed with high accuracy. The scheme determining the topological ordering is illustrated in Fig. 1.

The KLD, denoted hereon as D, between two J-dimensional probability distributions A and B is defined as [20]:

$$D(A||B) = \int_{R^J} p(X|A) log \frac{p(X|A)}{p(X|B)} dx \tag{1}$$

KLD provides an indication of how distant two models are in the probabilistic space. It is important to note that KLD as given in Eq. 1 comprises an asymmetric quantity. The symmetrical form can be inferred by simply adding the integrals in both directions, i.e.

$$D_s(A||B) = D(A||B) + D(B||A) \tag{2}$$

In the special case where both A and B are Gaussian mixture models KLD can be defined as follows:

$$KLD(A||B) = \int A(x) log \frac{B(x)}{A(x)} dx \tag{3}$$

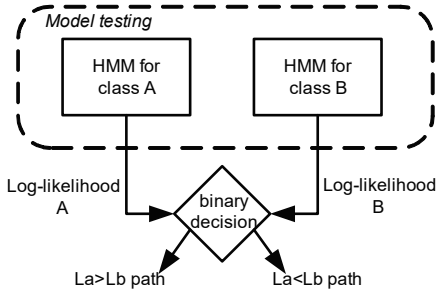

Fig. 3. The binary operation carried out in each node of the DAG-HMM

Unfortunately, there is not a closed-form solution for Eq. 3, thus we employed the empirical mean as follows

$$KLD(A||B) \approx \frac{1}{n} \sum_{i=1}^{n} log \frac{B(x_i)}{A(x_i)} \tag{4}$$

given that the number of Monte Carlo draws is sufficiently large. During our experiments we set n = 2000.

It should be noted the KLD between HMMs was not used since computing distances between HMMs of unequal lengths, which might be common in this work as HMMs representing different classes might have different number of states, can be significantly more computationally demanding without a corresponding gain in modeling accuracy [21], [22].

After computing the KLD for the different pairs of classes, i.e. reach the second stage depicted in Fig. 1, the KLD distances are sorted in a decreasing manner. This way the topological ordering of the DAG-HMM is revealed placing the classification tasks of low difficulty on its top. Each node removes a class from the candidate list until there is only one class left, which comprises the DAG-HMM prediction. The elements of the distance matrix could be seen as early performance indicators of the task carried out by the corresponding node. The proposed topological ordering places tasks likely to produce misclassifications at the bottom of the graph. This process outputs a *unique* solution fo the topological sorting problem, as it is usually met in the graph theory literature [23].

B. The DAG-HMM Operation

The operation of the proposed DAG-HMM scheme is the following: after extracting the features of the unknown audio signal, the first/root node is activated. More precisely, the feature sequence is fed to the HMMs, which produce two log-likelihoods showing the degree of resemblance between the training data of each HMM and the unknown one. These are compared and the graph flow continues on the larger log-likelihood path. It should be stressed out that the HMMs are optimized (in terms of number of states and Gaussian components) so that they address the task of each node optimally. That said, it is possible that a specific class is represented by HMMs with different parameters when it comes to different nodes of the DAG-HMM.

An example of a DAG-HMM addressing a problem with four classes is illustrated in Fig. 2. The remaining classes for testing are mentioned beside each node. Digging inside each node, Fig. 3 shows the HMM-based sound classifier responsible for activating the path of the maximum log-likelihood.

The operation of the DAG-HMM may be parallelized with that of investigating a list of classes, where each level eliminates one class from the list. More in detail, in the beginning the list includes all the potential audio classes. At each node the feature sequence is matched against the respective HMMs and the model with the lowest log-likelihood is erased from the list, while the DAG-HMM proceeds to the part of the topology without the discarded class. This process terminates when only one class remains in the list, which comprises the system's prediction. Hence, in case the problem deals with m different classes, the DAG's decision will be made after the evaluation of $m - 1$ nodes.

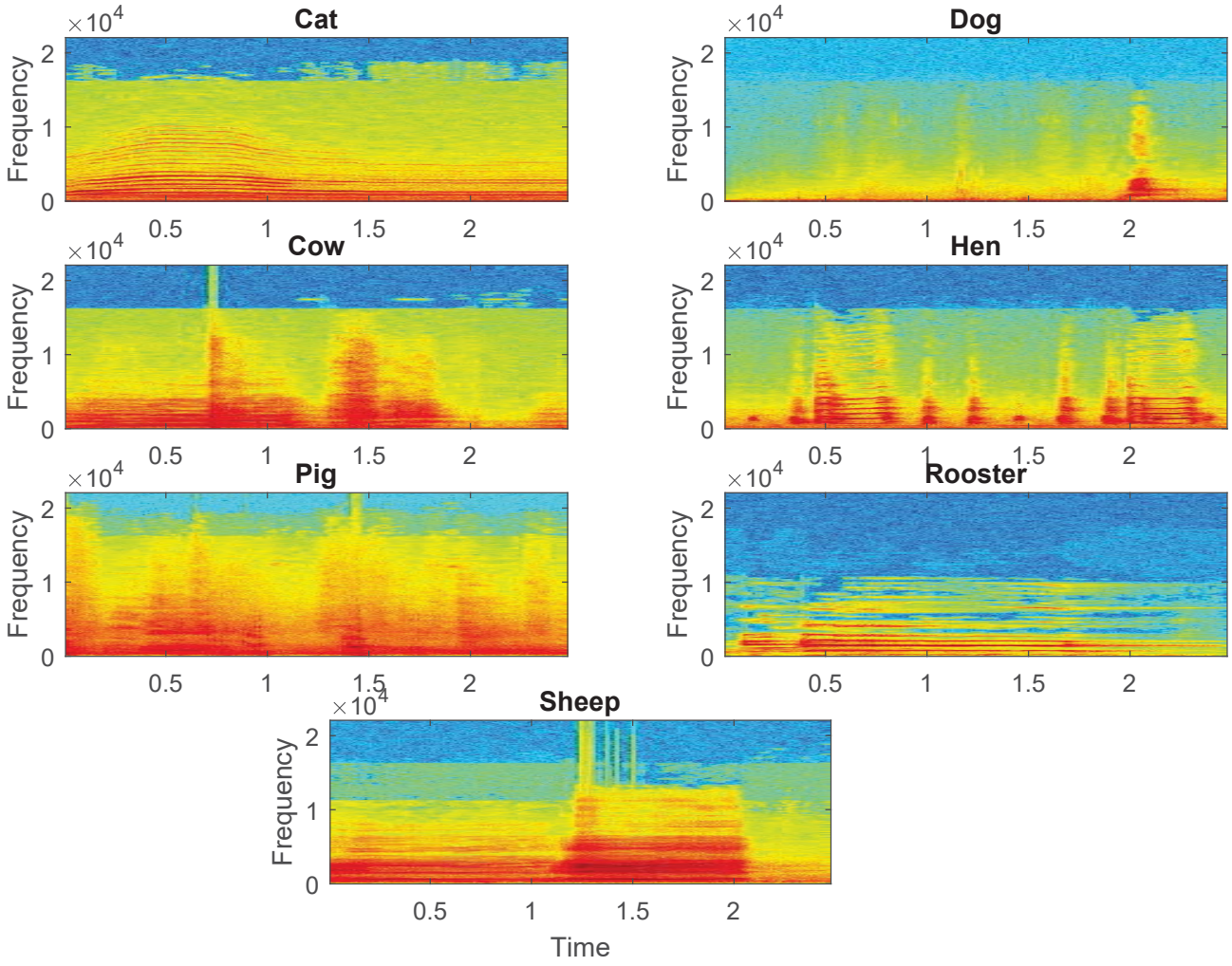

Fig. 4. Characteristic spectrograms extracted out of the available sound classes

C. The feature set

This feature set is composed of the first thirteen Mel frequency cepstral coefficients including the 0-th coefficient which reflects upon the energy of each frame. For MFCC's derivation we compute the power of the short time Fourier transform with respect to every frame and pass them through a triangular Mel scale filterbank. Subsequently, the log operator is applied and the energy compaction properties of discrete cosine transform are exploited in order to decorrelate and represent the majority of each energy band with just a few coefficients. Lastly, a thirteen-dimension vector is formed by the most important thirteen coefficients. Three derivatives of the initial vector are appended resulting to 52 dimensions. The processing stage was based on the openSMILE feature extraction tool [24].

IV. Experimental evaluation

In this section, we analyze the: *a)* dataset used to acoustically simulate a farm environment, *b)* parametrization of both DAG-HMM and feature extraction module, *c)* contrasted approaches, and *d)* we present and comment the achieved results.

A. Dataset

We collected data associated with the following typical farm animals: *dog*, *rooster*, *pig*, *cow*, *cat*, *hen*, and *sheep*. These are taken from the Environmental Sound Classification-10 described in [15], and they are sampled at 44.1 KHz. Each class includes 40 recordings, each one with a duration of 5 seconds. The vocalizations come from multiple animals. Fig. 4 demonstrates characteristic spectrograms extracted out the available sound classes.

B. System parametrization

Following the MPEG-7 standard recommendation, the low-level feature extraction window is 30 ms with 10 ms overlap, so that the system is robust against possible misalignments. The sampled data are hamming windowed to smooth potential

TABLE I. THE CONFUSION MATRIX (IN %) WITH RESPECT TO THE DAG-HMM. THE AVERAGE CLASSIFICATION RATE IS 93.1%.

Presented \ Responded	*Dog*	*Rooster*	*Pig*	*Cow*	*Cat*	*Hen*	*Sheep*
Dog	**99.4**	-	-	-	-	-	0.6
Rooster	-	**99.7**	-	-	-	0.3	-
Pig	14.4	-	**85.6**		-	-	-
Cow	-	-	-	**99.8**	0.2	-	-
Cat	13.9	-	-	-	**86.1**	-	-
Hen	-	0.7	-	-	-	**99.1**	-
Sheep	-	-	-	-	-	17.7	**82.3**

discontinuities while the FFT size is 512. Standard normalization techniques, i.e. mean removal and variance scaling, were applied.

The HMMs of each node are optimized in terms of number of states and nodes following the Expectation-Maximization and Baum Welch algorithms [25]. As the considered sound events are characterized by a distinct time evolution, we employed HMMs with left-right topology, i.e. only left to right states transitions are permitted. Moreover, the distribution of each state is approximated by a Gaussian mixture model of diagonal covariance, which may be equally effective to a full one at a much lower computational cost [26].

The maximum number of k-means iterations for cluster initialization was set to 50 while the Baum-Welch algorithm used to estimate the transition matrix was bounded to 25 iterations with a threshold of 0.001 between subsequent iterations. The number of explored states ranges from 3 to 7 while the number of Gaussian components used to build the GMM belongs to the $\{2, 4, 8, 16, 32, 64, 128, 256, \text{and } 512\}$ set. The final parameters were selected based on the maximum recognition rate criterion. The machine learning package Torch (freely available at http://torch.ch/) was used to construct and evaluate GMMs and HMMs.

C. Contrasted approaches

The proposed approach was contrasted to the following ones: class-specific HMM [27], universal background modeling (UBM) HMM with a KLD based data selection scheme [28], support vector machine (SVM) with radial basis function kernel [29], random forest (RF) [30], and echo state network (ESN) [31], [32]. The parameters of these classification schemes were optimized on TS.

The ESN implementation is based on the Echo State Network toolbox (freely available at https://sourceforge.net/projects/esnbox/) and the SVM on the libsvm library [33].

TABLE II. THE RECOGNITION RATES ACHIEVED BY THE PROPOSED AND CONTRASTED APPROACHES. T HE APPROACH PROVIDING THE HIGHEST RATE IS EMBOLDENED

Classifier	Average recognition rate (%)
DAG-HMM	**93.1**
Class-specific HMMs	77.1
Universal HMM	68.6
SVM	52.3
ESN	60
RF	54.3

D. Experimental results

Table II includes the results achieved by the proposed DAG-HMM as well as the contrasted approaches. The data division protocol is the ten-fold cross validation one. Identically selected folds were used during the training and testing processes of all approaches, enabling a reliable comparison.

A first observation is on the difficulty of the task which is relatively high since many classifications schemes fail to provide a satisfactory recognition rate. Then, as we can see, the proposed DAG-HMM outperforms the rest of the approaches. The second one is based on class specific HMMs, while the ESN achieved the third best recognition rate. The UBM logic provides lower rate than the class-specific one showing the high degree of diversity characterizing the common feature space. Same conclusions can be derived for the SVM, which cannot find reliable boundaries between the classes and the RF, the rules of which do not classify the feature space in a reliable manner. We conclude that limiting the problem space using a DAG-HMM is particularly beneficial in the specific application scenario providing encouraging recognition rates.

The confusion matrix achieved by the DAG-HMM is tabulated in Table I. We observe that the class recognized with the highest accuracy is *cow* (99.8%), while the one presenting the worst rate is the *sheep* one (82.3%). The misclassifications' source is the great variability among sound samples of the same class as it is assessed by a human listener. Moreover, several sound clips are acoustically similar even though they belong to different categories. This is particularly evident in the cases of *sheep-hen*, *cat-dog*, and *pig-dog* pairs. We conclude that the DAG-HMM classification approach provides promising performance; even though the associated computational cost of the training phase is rather high, it is to be conducted only once and offline. At the same time, the testing phase includes simple log-likelihood comparisons and estimations carried out using the Viterbi algorithm, which is computationally inexpensive as it is based on recursive dynamic programming.

V. CONCLUSIONS

This paper presented a novel classification scheme addressing the scientific area of acoustic farm monitoring. We outlined a classification scheme based on a DAG composed of HMMs trained on an MFCC feature set. The superiority of the proposed scheme over state of the art classifiers was proven on a publicly available dataset encompassing vocalizations of seven farm animals. In the future, we wish to enhance the present framework so that it is able to operate in a concept drift environment, i.e. being able to evolve itself during operation and

recognize altered (noisy or reverberant) versions of the existing classes, increase the dictionary of the animal vocalizations, etc. To this direction we plan to explore transfer learning technology [34], [35]. Finally, we wish to enhance the feature extraction part following the unsupervised learning direction, as reported e.g. in [36], [37].

REFERENCES

[1] D. Stowell, *Computational Bioacoustic Scene Analysis*. Cham: Springer International Publishing, 2018, pp. 303–333. [Online]. Available: https://doi.org/10.1007/978-3-319-63450-0_11

[2] D. Blumstein, D. Mennill, P. Clemins, L. Girod, K. Yao, G. Patricelli, J. Deppe, A. Krakauer, C. Clark, K. Cortopassi, S. Hanser, B. Mccowan, A. Ali, and A. Kirschel, "Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus," *Journal of Applied Ecology*, vol. 48, no. 3, pp. 758–767, 6 2011.

[3] M. W. Towsey, A. M. Truskinger, and P. Roe, "The navigation and visualisation of environmental audio using zooming spectrograms," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, Nov 2015, pp. 788–797.

[4] X. Dong, M. Towsey, J. Zhang, and P. Roe, "Compact features for birdcall retrieval from environmental acoustic recordings," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, Nov 2015, pp. 762–767.

[5] T. Grill and J. Schlter, "Two convolutional neural networks for bird detection in audio signals," in *2017 25th European Signal Processing Conference (EUSIPCO)*, Aug 2017, pp. 1764–1768.

[6] S. Ntalampiras, "Bird species identification via transfer learning from music genres," *Ecological Informatics*, vol. 44, pp. 76 – 81, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574954117302467

[7] D. Mitrovic, M. Zeppelzauer, and C. Breiteneder, "Discrimination and retrieval of animal sounds," in *2006 12th International Multi-Media Modelling Conference*, 2006, pp. 5 pp.–.

[8] N. C. Han, S. V. Muniandy, and J. Dayou, "Acoustic classification of australian anurans based on hybrid spectral-entropy approach," *Applied Acoustics*, vol. 72, no. 9, pp. 639 – 645, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003682X11000314

[9] V. Exadaktylos, M. Silva, and D. Berckmans, "Automatic identification and interpretation of animal sounds, application to livestock production optimisation," in *Soundscape Semiotics - Localization and Categorization*, H. Glotin, Ed. Rijeka: InTech, 2014, ch. 04. [Online]. Available: http://dx.doi.org/10.5772/56040

[10] J. J. Noda, C. M. Travieso, D. Snchez-Rodrguez, M. K. Dutta, and A. Singh, "Using bioacoustic signals and support vector machine for automatic classification of insects," in *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, Feb 2016, pp. 656–659.

[11] A. Kumar and G. P. Hancke, "A zigbee-based animal health monitoring system," *IEEE Sensors Journal*, vol. 15, no. 1, pp. 610–617, Jan 2015.

[12] S. K. Nagpal and P. Manojkumar, "Hardware implementation of intruder recognition in a farm through wireless sensor network," in *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, Feb 2016, pp. 1–5.

[13] V. M. Anu, M. I. Deepika, and L. M. Gladance, "Animal identification and data management using rfid technology," in *International Confernce on Innovation Information in Computing Technologies*, Feb 2015, pp. 1–6.

[14] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, "Learning in non-stationary environments: A survey," *IEEE Computational Intelligence Magazine*, vol. 10, no. 4, pp. 12–25, Nov 2015.

[15] K. J. Piczak, "Esc: Dataset for environmental sound classification," in *Proceedings of the 23rd ACM International Conference on Multimedia*, ser. MM '15. New York, NY, USA: ACM, 2015, pp. 1015–1018. [Online]. Available: http://doi.acm.org/10.1145/2733373.2806390

[16] ——, "Environmental sound classification with convolutional neural networks," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, Sept 2015, pp. 1–6.

[17] P. Smaragdis, M. Shashanka, and B. Raj, "A sparse non-parametric approach for single channel separation of known sounds," in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, Eds. Curran Associates, Inc., 2009, pp. 1705–1713.

[18] S. Ntalampiras, "Directed acyclic graphs for content based sound, musical genre, and speech emotion classification," *Journal of New Music Research*, vol. 43, no. 2, pp. 173–182, 2014. [Online]. Available: https://doi.org/10.1080/09298215.2013.859709

[19] T. J. VanderWeele and J. M. Robins, "Signed directed acyclic graphs for causal inference," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 1, pp. 111–127, 2010. [Online]. Available: http://dx.doi.org/10.1111/j.1467-9868.2009.00728.x

[20] P. Taylor, "The target cost formulation in unit selection speech synthesis," in *INTERSPEECH 2006 - ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA, September 17-21, 2006*, 2006. [Online]. Available: http://www.isca-speech.org/archive/interspeech_2006/i06_1455.html

[21] Y. Zhao, C. Zhang, F. K. Soong, M. Chu, and X. Xiao, "Measuring attribute dissimilarity with hmm kl-divergence for speech synthesis," in *In*, 2007, pp. 6–2007.

[22] P. Liu, F. K. Soong, and J. L. Zhou, "Divergence-based similarity measure for spoken document retrieval," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 4, April 2007, pp. IV–89–IV–92.

[23] S. A. Cook, "A taxonomy of problems with fast parallel algorithms," *Information and Control*, vol. 64, no. 1, pp. 2 – 22, 1985, international Conference on Foundations of Computation Theory. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0019995885800413

[24] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM International Conference on Multimedia*, ser. MM '13. New York, NY, USA: ACM, 2013, pp. 835–838. [Online]. Available: http://doi.acm.org/10.1145/2502081.2502224

[25] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb 1989.

[26] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72–83, Jan 1995.

[27] H.-G. Kim and T. Sikora, "Comparison of mpeg-7 audio spectrum projection features and mfcc applied to speaker recognition, sound classification and audio segmentation," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, May 2004, pp. V–925–8 vol.5.

[28] S. Ntalampiras, "A novel holistic modeling approach for generalized sound recognition," *IEEE Signal Processing Letters*, vol. 20, no. 2, pp. 185–188, Feb 2013.

[29] L. Chen, S. Gunduz, and M. T. Ozsu, "Mixed type audio classification with support vector machine," in *2006 IEEE International Conference on Multimedia and Expo*, July 2006, pp. 781–784.

[30] M. M. Al-Maathidi and F. F. Li, "Audio content feature selection and classification a random forests and decision tree approach," in *2015 IEEE International Conference on Progress in Informatics and Computing (PIC)*, Dec 2015, pp. 108–112.

[31] S. Scardapane and A. Uncini, "Semi-supervised echo state networks for audio classification," *Cognitive Computation*, vol. 9, no. 1, pp. 125–135, Feb 2017. [Online]. Available: https://doi.org/10.1007/s12559-016-9439-z

[32] S. Ntalampiras, "Moving vehicle classification using wireless acoustic sensor networks," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 2, pp. 129–138, April 2018.

[33] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[34] D. Banerjee, K. Islam, G. Mei, L. Xiao, G. Zhang, R. Xu, S. Ji, and J. Li, "A deep transfer learning approach for improved post-traumatic stress disorder diagnosis," in *2017 IEEE International Conference on Data Mining (ICDM)*, Nov 2017, pp. 11–20.

[35] S. Ntalampiras, "A transfer learning framework for predicting the emotional content of generalized sound events," *The Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. 1694–1701, 2017. [Online]. Available: http://dx.doi.org/10.1121/1.4977749

[36] H. Lee, P. Pham, Y. Largman, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, Eds. Curran Associates, Inc., 2009, pp. 1096–1104.

[37] S. Chaudhuri and B. Raj, "Unsupervised structure discovery for semantic analysis of audio," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1178–1186. [Online]. Available: http://papers.nips.cc/paper/4661-unsupervised-structure-discovery-for-semantic-analysis-of-audio.pdf