The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Source recognition in acoustic sensor arrays using self-organizing hidden Markov models

Edgar E. Vallejo[1], Julio G. Arriaga[1], and Charles E. Taylor[2]

[1] ITESM Campus Estado de México, Atizapán de Zaragoza, Estado de México, 52926, México
[2] UCLA, Los Angeles, CA, 90095-1606, USA
(Tel: 1-310-825-6850, Fax: 1-310-206-3987)
(taylor@biology.ucla.edu)

Abstract: In this paper, we present a computational model for source recognition in acoustic sensor arrays. The proposed model uses a hybrid method of neural networks to create symbols then hidden Markov models that collectively learn to discriminate sequences of symbols from a collection of recordings from bird species with minimum human intervention. Preliminary simulation results indicate that this model is capable of producing acceptable levels of classification performance.

Keywords: Acoustic sensor arrays, hidden Markov models, bird species classification

1 INTRODUCTION

Over the last few years, we have been engaged in a research program that aims at understanding the capabilities and limitations of sensor arrays in habitat monitoring applications [1]. Particularly, we have developed acoustic sensor arrays for monitoring the diversity and behavior of bird species at several locations in Mexico and in the US. More recently, our research program has evolved towards the application of this technology for understanding the structure and function of bird song.

As part of this study, we have focused on developing robust filters that are able to discriminate bird species, and in some cases, bird individuals with reasonably high accuracy. Specifically, we have explored the use of unsupervised learning methods such as self-organizing maps [11][13], and supervised learning methods such as support vector machines [12], among others. We have had success with the use of hidden Markov models (HMMs) for this problem [9].

We believe that given the appropriate conditions, an array of sensors should be able to self-organize so as to behave as a single ensemble [4][8]. In this idealization, sensors can identify and collectively learn to classify events of interest from an arbitrary acoustic recording stream, with minimal human intervention. This ability is at the core of what is required for achieving adaptive behavior and communication in sensor arrays.

Towards this end, we developed a computational model for source recognition in acoustic sensor arrays. The proposed model uses hidden Markov models that self-organize to discriminate a collection of input sequences with minimal human intervention. We chose HMMs for its ability to discriminate input sequences of arbitrary length, which have proven difficult for alternative classification methods.

Particularly, we introduce a collective learning algorithm that enable a collection of HMMs to learn to classify four different species of antbirds. In the collective learning procedure, a sensor detects an input sequence from the input stream and uses its HMM to determine its corresponding output. The sensor output is then combined with those of its neighbor sensors to produce a consensus output. The sensor then computes the maximum likelihood estimators from these consensus sequences to update its parameters, and the learning process continues.

The overall objective of this self-organizing learning process is to enable HMMs to accurately classify the input sequences and to converge to consistent collective classifications along the array. We used recordings obtained from our field studies in a tropical rainforest in Chiapas, Mexico. From these recordings, we constructed a collection of training sets consisting of tokenized sequences of bird songs.

2 METHODS

2.1 Self-organizing HMMs

In the proposed model, the classification of bird species is produced by hidden Markov models (HMMs). HMMs are extensions of Markov chains in which an observation is a probabilistic function of the state [7].

Formally, a *hidden Markov model* is a 4-tuple (Q, Σ, A, E), where

1. Q is a finite set of states,
2. Σ is an alphabet of symbols,
3. $A = (a_{kl})$ is a $|Q| \times |Q|$ matrix of state transition probabilities,
4. $E = (e_k(b))$ is a $|Q| \times |\Sigma|$ matrix of emission probabilities

Let $M = (Q, \Sigma, A, E)$ be a hidden Markov model. A path $\pi = \pi_1 \ldots \pi_n$ is a sequence of states. The probability that a sequence x of symbols from the alphabet Σ was generated by the path π is

$$P(x|\pi) = \prod_{t=1}^{n} P(x_i|\pi_i)P(\pi_i|\pi_{i+1}) = \prod_{i=1}^{n} e_{\pi_i}(x_i) \cdot a_{\pi_i,\pi_{i+1}}$$

The decoding problem is formulated as to find the optimal path $\pi^* = \arg\max_\pi P(x|\pi)$ for x such that $P(x|\pi)$ is maximized.

The solution of the decoding problem is provided by the Viterbi algorithm [7]. The idea is that the optimal path for the $(i+1)$ prefix $x_1 \ldots x_{i+1}$ of x uses a path for an i-prefix of x that is optimal among the paths ending in an unknown state $\pi_i = k \in Q$.

Define $s_k(i)$ as the probability of the most probable path for the prefix $x_1 \ldots x_i$ that ends with state k ($k \in Q$ and $1 \leq i \leq n$). Then

$$s_l(i+1) = e_l(x_{i+1}) \cdot \max_{k \in Q}\{s_k \cdot a_{kl}\}$$

Let $M = (Q, \Sigma, A, E)$ be a hidden Markov model and $x = x_1 \cdots x_n$ be a string over the alphabet Σ. The HMM M produces the output o in response to input x, $M(x) = o$, if a sequence of states $\pi = \pi_0 \ldots \pi_n$ exists in Q with the following conditions:

1. π is the optimal path π^* given the sequence x and the hidden Markov model M, and

2. $\pi_i^* = o_i$ ($1 \leq i \leq n$) is the state of the optimal path π^* given the sequence x and the hidden Markov model M.

2.2 Learning algorithm

For this study, we devised a collective learning algorithm for HMMs. The model comprises a collection of randomly generated HMMs. For each input in the training set, an HMM uses the consensus optimal path of the HMMs in its neighborhood as its target output. The consensus optimal path is obtained from the neighbor HMMs by voting. The maximum likelihood estimators are then computed from the set of target outputs as follows. We count the number of times each particular transition or emission is used in the set of training inputs. Let these be A_{kl} and $E_k(b)$. Then the maximum likelihood estimators for a_{kl} and $e_k(b)$ are given by:

$$a_{kl} = \frac{A_{kl}}{\sum_{l'} A_{kl'}}$$

and

$$e_k(b) = \frac{E_k(b)}{\sum_{b'} E_k(b')}$$

The learning algorithm is described in Table 1.

Table 1. Learning algorithm.

1. Create an initial random set P of HMMs
2. `Do until` number simulation steps N is met
   (a) `For each` HMM $M \in P$ `do`
      i. Select a subset $S \subseteq P$ of HMMs at random
      ii. Compute the output sequence $M'(x) = o$ for each $M' \in S$ and $x \in X$ using the Viterbi algorithm
      iii. Compute the maximum likelihood estimators a_{kl} and $e_k(b)$ from the consensus outputs of S.
      iv. Update the parameters A and E of M to be the maximum likelihood estimators.

   `End for`

`End do`

Table 2. Training set.

label	species	samples
BAS	Barred antshrike	12
DAB	Dusky antbird	12
GAS	Great antshrike	12
MAT	Mexican antthrush	12

2.3 Data set

The samples used in the experiments presented here were provided to us by Martin L. Cody. The dataset consists of songs from four different antbird species that are abundant at the Montes Azules Biosphere Reserve in Chiapas, Mexico. They are listed on Table 2.

The preparation of data proceeded as follows. The songs were segmented using the Raven bird song analysis program [3]. Syllables were identified by small discontinuities in the corresponding spectrogram. Using this procedure, we obtained a collection a syllable samples as listed in Table 3. For each sample, we obtained a series of temporal and spectral measurements using Raven. These parameters were extracted from the sound signal using the short-time Fourier transform (STFT) and selected according to previous studies on bird species classification [3] [6]. These measurements are described in Table 4.

A normalization process was applied to this data because the selected measurements span different orders of magnitude. Such differences might result from how close the subject is to the recorder, for example. Using the mean and the standard deviation of each measurement, we obtained a collection of feature vectors described as z-scores.

The collection of feature vectors describing syllables were classified using a simple competitive learning neural network [5]. Once the syllables have been categorized we proceeded to represent the original songs as strings of symbols using the label from each syllable category. Table 5 shows the string representation of a subset of the songs obtained us-

Table 3. Syllable samples

label	samples
BAS	216
DAB	129
GAS	339
MAT	117

Table 4. Selected features of syllables

parameter	description
Low frequency	The lower frequency bound of the syllable
High frequency	The upper frequency bound of the syllable
Delta time	The duration of the syllable
Max amplitude	The upper amplitude bound of the syllable
Max power	The upper power bound of the syllable

ing a two-unit competitive learning network, with learning constant $\eta = 0.1$ and epochs = 1000.

Similarly, table 6 shows the string representation of a subset of the songs obtained using a four-unit competitive learning network, with learning constant $\eta = 0.1$ and epochs = 1000.

3 EXPERIMENT AND RESULTS

Multiple simulations were conducted using different combinations of parameter values presented in table 7. The following were the major results:

1. HMMs produced a reasonably good categorization performance ($\sim 82\%$) of the training sets.

2. Acceptable numbers of training steps (~ 100) were required for the collection of HMMs to converge to consistent classifications.

3. The description of bird songs using fewer distinct syllables and HMMs with fewer states produced better classification performance.

Table 5. Strings representation of the training set using 2 syllables

label	string
BAS_1	BBBBBBBBBAAAABAAAAAAABA
BAS_2	BBBBAAAAAAAAAAAAAAAAAAAA
BAS_3	BBBBABBBBAAAAABAAAAABABBA
DAB_1	BAAAAAAAAABBBBB
DAB_2	BAAAAAABBBBBBB
DAB_3	BBAAAAAAAAABBBB
GAS_1	BBBBBBBAAABBABBBABAAAAAABAAABBABBAAAAAB
GAS_2	BBAAAAAABABBABBAAAAAAAAAAAAAAAAAAAAB
GAS_3	BBBBBBBBBABBAAAAAAAAAAAAAAAAAAAAAAAAB
MAT_1	BBBBBBBBBBBBB
MAT_2	AAAAAAAAAABBB
MAT_3	AAAAAAAAAAAAA

Table 6. Strings representation of the training set using 4 syllables

label	string
BAS_1	DDDAAAAAAABAAAAAAAAAABAC
BAS_2	DDDAAAAAAAAABBBABBBBBBBAAB
BAS_3	DDDAAAAAAAAAABAAAAAAAAAAAC
DAB_1	DBBBBBBBBBBDDDDD
DAB_2	DBBBBBBBDDDDDDD
DAB_3	DDBBBBBBBBBDDDD
GAS_1	DDDDCCCDCCCCCCCCCCCCBCCACBAACCAAAABACCD
GAS_2	DDACCCCCCBCCCCBCCCCCACAACABBABBBBAAD
GAS_3	DDDCDCCCCCCBCCCCBBBCCBCCAABAABABBBBBCD
MAT_1	DDDDDDDDDDDDD
MAT_2	DBBBBBBBDDDBBB
MAT_3	BBBBBBBBBBBBB

Table 7. Parameters for simulations

parameter	value
Simulation steps	100–200
$\|P\|$	128–512
$\|S\|$	8–32
$\|Q\|$	8–16
$\|\Sigma\|$	2-16

To calculate the classification performance we obtain the consensus output sequence from all of the samples in the same category. Then we compute the pairwise similarity of each output sequence produced for each sample in the category with respect to the consensus output sequence of the category.

Table 8 shows the accuracy of classification obtained for the two-syllable experiment using HMMs with 8 states and 100 simulation steps.

Similarly, Table 9 shows the accuracy in classification obtained for the four-syllable experiment using HMMs with 16 states and 200 simulation steps.

4 CONCLUSION

Despite its preliminary character, the results shown here seem to indicate that acceptable categorization of bird species can emerge using self-organizing HMMs. The results also show that the accuracy in classification depends on the number of syllables describing the bird songs. This sug-

Table 8. Simulation results

species	classification
BAS	82%
DAB	86%
GAS	81%
MAT	77%

Table 9. Simulation results

species	classification
BAS	79%
DAB	82%
GAS	78%
MAT	75%

gests the existence of a number that is optimum for accurate species classification. Moreover, categorization requires minimum human intervention in contrast with supervised learning methods. In general, experimental results showed that there is an advantage for simulations in which hidden Markov models posses fewer states.

We think the proposed method could be extended in several ways. For instance, the most relevant in practice would be to use this model for event detection and to identify the species that are sharing the acoustic space, in both the temporal and spectral spaces. This capabilities would enable the technology of sensor arrays to explore important questions in ecology regarding both the inter- and intra- species interaction of birds.

It should be noted that the proposed model has only been tested in a simple simulated setting. We will test the proposed model in real settings in the near future. We believe that self-organizing HMMs hold much promise to contribute for the development of fully automated sensor arrays for habitat monitoring applications.

Acknowledgements

We thank Martin L. Cody for kindly providing to us the recordings of birds vocalizations used in our experiments. This work was supported by the US National Science Foundation, Award No. IIS-1125423. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agency.

REFERENCES

[1] Blumstein, DT, Mennill, DJ, Clemins, et al (2011) Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*, Volume 48, Issue 3, pages 758767

[2] Catchpole, C. K., Slater, P. L. B.: *Bird song biological themes and variations.* Cambridge University Press, 1995.

[3] Charif, R. A., Clark, C. W., Fistrup, K. M.: *Raven 1.2 user's manual.* Cornell Laboratory of Ornithology, Ithaca, NY, 2004.

[4] Collier, T. C., Taylor C.E.: Self-Organization in Sensor Networks. *Journal of Parallel and Distributed Computing* 64:7 (2004) pp.866–873.

[5] Hertz, J., Krogh A., Palmer, R. G. (1991) *Introduction to the theory of neural computation*. Addison Wesley, 1991.

[6] Nelson, D. A.: The importance of invariant and distinctive features in species recognition of bird song. *Condor*: Vol. 91, No. 1, pp.120-130,1989.

[7] Rabiner, LR (1989), A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77: 257–286.

[8] Taylor, C. E.: From cognition in animals to cognition in superorganisms. In M. Bekoff, C. Allen and G. Gurghardt, (eds.), *The Cognitive Animal. Empirical and Theoretical Perspectives on Animal Cognition* The MIT Press, 2002.

[9] Trifa, V., Kirschel, A., Vallejo, EE and CE Taylor (2008). Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models. *J. Acoust. Soc. Am*. Volume 123, Issue 4, pp. 2424-2431.

[10] Vallejo EE and CE Taylor (2004), A simple model for the evolution of a lexicon. In M. Sugisaka, H. Tanaka (eds), *Proceedings of the Ninth International Symposium on Artificial Life and Robotics, AROB9th*.

[11] Vallejo EE, Cody, ML and CE Taylor (2007), Bird species recognition using hierarchical self-organizing maps. In (eds), *Proceedings of the Australian Artificial Life Conference, ACAL 2007*. Springer-Verlag.

[12] Vallejo EE and CE Taylor (2009), Sensor arrays for acoustic monitoring of bird behavior and diversity. Preliminary results on source identification using support vector machines. *Artificial Life and Robotics*. Springer-Verlag.

[13] Vallejo EE and CE Taylor (2010), A self-supervised classifier ensemble for source recognition in acoustic sensor arrays. In (eds), *Proceedings of the Twelfth International Conference on the Simulation of Living Systems, Artificial Life 12*. The MIT Press.