Semantic Inference of Bird Songs Using Dynamic Bayesian Networks

Keisuke Daimon
Department of Complex Systems Science,
Graduate School of Information Science
Nagoya University
Furo-cho Chikusa-ku, Nagoya 464-8601, Japan
+81-52-789-4235, k.daimon@alife.cs.is.nagoya-u.ac.jp
http://www.alife.cs.is.nagoya-u.ac.jp/%7ek.daimon

Richard W. Hedley and **Charles E. Taylor**
Department of Ecology and Evolutionary Biology,
University of California, Los Angeles
USA

Abstract

This paper presents preliminary research aiming to understand the semantics of bird songs. We investigate relationship between context information and bird song output using dynamic Bayesian networks, and show the potential of our approach for semantics inference.

Introduction

Knowledge representation and natural language processing are core interests to the field of artificial intelligence (AI). While most research has been directed toward machines and humans, the principles and methods developed for AI might be extended to other species as well. Birds frequently behave in a manner that is intelligent and convey information in their vocalizations that is meaningful to others (Suzuki 2016). Here we report on efforts to use dynamic Bayesian networks (DBNs) to describe semantics of songs among Cassin's Vireos (*Vireo cassinii*) in their natural habitat.

Cassin's Vireo is a songbird found throughout the montane regions of western North America. Males of this species deliver songs that are typically comprised of strings of 0.5-second phrases separated by about 2 seconds of silent intervals. Each phrase can be unambiguously classified as a phrase type using spectrographic characteristics (Tan et al. 2013), Each male delivers about 50 phrase types and shares about 50% of its repertoire with other birds (Hedley 2016). Phrase types are organized into clusters of several phrase types that consistently occur together in sequences. We hypothesize that these clusters may convey meaningful information, such as about nesting behavior or territorial defense.

DBNs are directed graphs that are useful for identifying causal relationships between observed events over timeslices (Murphy and Mian 1999). Nodes in DBNs represent random variables of specific events and edges describe causal relationships from nodes to nodes at next time step. Consecutive data points in bird song data are highly dependent on previous ones, hence the suitability of DBNs as models of bird songs. If semantic content is present in the songs, there should be causal relationships between behavioral contexts and song output, which can be detected and tested through DBN analysis. The behavioral contexts investigated here consist of four different pieces of information. Though these behavioral contexts represent only a small subset of the possible behavioral contexts that could be examined, this analysis will provide a foundation for future attempts to understand the interplay between vocalizations and behavior that may make use of more diverse behavioral categories. To our knowledge, this is the first attempt at using DBNs to infer semantic content in animal vocalizations.

Data and Methods

The data of two distinct birds were analyzed. For convenience, we named the two birds individual 1 and 2. Individual 1 and 2 have 7 and 11 tracks, 2172 and 3905 phrases, 51 and 53 phrase types (25 shared phrase types) and were recorded in 2014 and 2015 respectively. The songs of the birds were recorded in the field (details are described by Hedley (2016)), and behavioral observations were dictated into the microphone. In the lab, spectrograms of each recording were manually annotated, with phrases classified to phrase type and behavioral context information noted at each point in the recording. We look into four behavioral contexts: *Distance (Far)* (whether or not the focal bird is far from the nest), *Background* (were other birds singing?), *Female* (was their mate on the nest?) and *Time (Late)* (was the track recorded late in the day?). *Phrase Type Clusters* (was the cluster sung?) and *Interval (Long)* (was the interval between phrases long?) are used as song output information.

We eliminated data points with missing values before the analyses for the sake of simplicity. Our first step is bringing the number of vocal units from about 50 phrase types to seven clusters to reduce the dimensionality and sparsity of the data based on community detection (modularity-based non-hierarchical clustering). In the second step, to find out relationships between behavioral contexts and song output, we build DBNs that show the best fit to our data according to Bayesian-Dirichlet equivalence (BDe) using a Python library, BNFinder2 (Dojer et al. 2013).

Results

At first, we conducted community detection and found seven clusters for each individual. Those clusters are called 1a-1g (individual 1) and 2a-2g (individual 2) for convenience.

We found the best DBNs through testing all DBN structures with only one regulation: time node cannot be a child

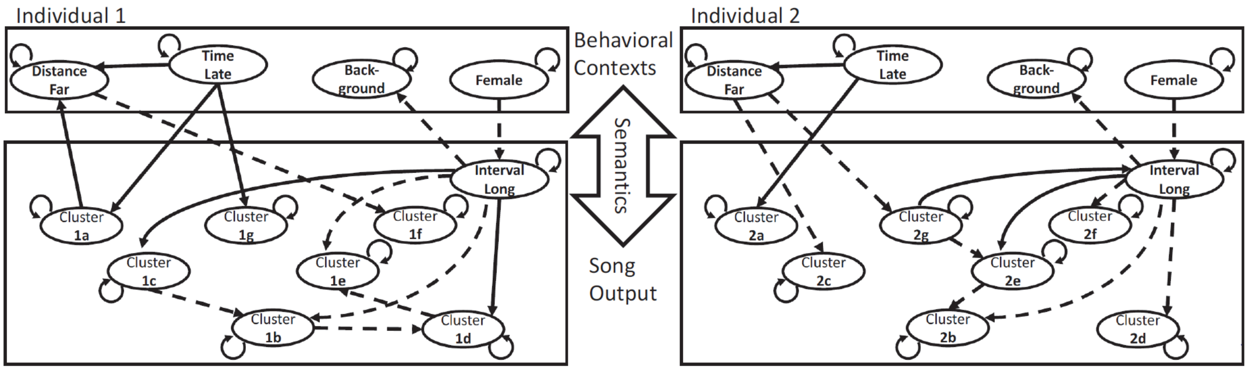

Figure 1: BDe chooses the best DBN structures for each bird (left: individual 1, right: individual 2). All nodes are binary (yes or no). Each directed edge describes how a parent node affects a child node: a solid arrow represents a positive effect, and a dashed arrow represents a negative effect. Positive effect means child nodes are more likely to be the same values as their parent nodes, with the opposite being true for a negative effect.

of the other nodes (Figure 1). In both DBNs distance and time affect one or two phrase type clusters, but only distance of individual 1 is positively affected by one cluster (1a).

All nodes have positive effects to themselves, meaning previous timeslices affect current timeslices. Thinking about arrows other than self regulatory ones, most nodes only have one or two arrows, while interval node has the largest number of arrows (six and seven) in individual 1 and 2, respectively, and two of the arrows for each are connected with behavioral context nodes, background and female.

Because birds do not share the same phrase type clusters, comparing arrows from/to cluster nodes of one bird with those of the other is not appropriate. It should, however, be mentioned that three arrows which are not related to cluster nodes (time to distance, female to interval and interval to background) show up in both of the DBNs.

Discussion

Our approach identified relationships between context and song output in this species, revealing a considerable amount of context-sensitivity in the singing behavior, and demonstrating the utility of DBNs for the representation of semantic content. Though the phrase type clusters of both individuals appear to be influenced by context, it is worth noting that these relationships are not consistent across birds — different birds deliver different phrase types in different matters. This is, at first, counter-intuitive, since transmission of information would be greatly facilitated if all individuals utilized common vocalizations to convey a mutually-agreed-upon set of meanings. However, males in this system directly compete for mates, food, and territories. Under competitive scenarios, evolutionary theory does not predict the emergence of honest species-wide signaling since cheaters introduce instability into the system, whereas interests within a male-female pair bond are more closely aligned: both individuals partake in incubation, nestling provisioning, and predator defense. We propose that this communication system might facilitate the exchange of information within each mated pair at the exclusion of eavesdropping rivals by the usage of highly local vocabularies.

On the other hand, the fact that relationships among non-cluster nodes are the same for both individuals leads to an interesting question. Males have different phrase repertoires, but do they have the same behavioral/semantic characteristics? This is yet to be answered by further research on different individuals. We believe our method (i.e. a combination of community detection and DBN analysis) is suitable for representing semantic content in each individual and general semantic signaling system shared by all individuals, in this and other species.

Acknowledgments

We thank Martin Cody and Joe Suzuki for their valuable support. This work was supported by NSF grant 1125423.

References

Dojer, N.; Bednarz, P.; Podsiadło, A.; and Wilczyński, B. 2013. Bnfinder2: Faster bayesian network learning and bayesian classification. *Bioinformatics* btt323.

Hedley, R. W. 2016. Composition and sequential organization of song repertoires in cassin's vireo (*Vireo cassinii*). *Journal of Ornithology* 157(1):13–22.

Murphy, K., and Mian, S. 1999. Modelling gene expression data using dynamic bayesian networks. Technical Report 104, Computer Science Division, University of California, Berkeley, CA.

Suzuki, T. N. 2016. Semantic communication in birds: evidence from field research over the past two decades. *Ecological Research*.

Tan, L. N.; Alwan, A.; Kossan, G.; Cody, M. L.; and Taylor, C. E. 2013. A sparse representation-based classifier for in-set bird phrase verification and classification with limited training data. In *38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 763–767. IEEE.