Moving from shallow to deep ecoinformatics

Higher-order co-occurrence and path analysis applied to multiple aspects of green-rumped parrotlets

A PhD dissertation proposed by Oliver Muellerklein (@Thru-Echoes / it me!)

I describe two approaches to the *30 years of green-rumped parrotlet* data:

1. High-Order Supervised Machine Learning

Path analysis offers an analysis of **lifetime reproductive success (LRS)** through both direct and indirect (*latent information*) co-occurrence paths. In other words, there is a wide range of ecological / environmental, social, and nest-based effects that can influence an individual bird's LRS.

LRS as a Function of Xs

For the sake of this example, let's label LRS as **Y**. With building a model that attempts to explain the relationships between all variables described above, **X_1, X_2, ... , X_n**, we are attempting to classify or regress unto **Y** (*please forgive me if I just said that backwards*).

I.e. optimize Y based on a range of unknown Xs.

Direct and Indirect Effects

The following may be true:

- casual relationship between X_1 and X_2
- casual relationship between X_1 and Y
- no direct relationship between X_2 and Y
- possible latent variable L_1 affecting X_1
- no direct relationship between L_1 and X_2

The indirect effect of **X_2** on **Y** can be the product of the coefficients on the path from **X_1** to **Y** (*on a single path*). With multiple paths, we could sum all of the products on every possible path.

How does **L_1** influence the paths and relationships?

Model of Variables Influencing LRS

model_of_lrs.png

Use vegetation variables in *resources & environment*

Use climate variables to distinguish between temporal changes in rainfall and spatial differences

Could variables from nest behaviors be used as a new *oval (i.e. set of variable types)* or directly impacts *maternal effects*?

hidden_markov_chain.png

X: states

Y: possible observations

a: state transition probabilities

b: output probabilities

Very Silly Example!

hidden_markov_chain_example.png

Start: start of an individual's breeding lifetime

Rainy: influence of the environment here?

Sunny: other variables that influence from the environment?

Walk: individual does not breed

Shop: individual is breeding away...

Clean: individual does not make it :'(

Model Approaches and Time

Time-dependent models, such as *30 years of parrotlet data* or a *time-series of images such as videos of nests...*, offer a complex problem and interesting approaches to it.

Path Analysis: special case of Structural Equation Modeling (SEM), aka analysis of covariance structures or latent variable models.

Path Analysis is non-recursive in finding correlations between latent (*i.e. hidden*) variables and the paths of the model. Path coefficients are standardized regression coefficients.

Non-parametric Bayesian Inference: non-stationary hidden Markov models that are time-dependent and produce state transition probabilities driven by analyzing high-dimensional data. In other words, these models offer a time-dependent "decision tree" of when individual parrotlets go from *(example)* healthy as a state (*i.e. high LRS*) to moderate (*i.e. average LRS*) to most likely dead as a state. All of these state changes are time-dependent and are probabilistic.

High-Order Structures and Relationships

The underlying assumption of many models (parametric) is the instances (*i.e. observations / data points*) are **independent and identically distributed (IID)**. This constraint results in many models, *including machine learning algorithms*, being stuck analyzing direct relationships and correlations between variables, while missing latent relations and high-order structure in the learned models. I believe applying high-order supervised machine learning approaches, *that do not assume IID and are non-parametric*, will potentially offer novel insight into the relationship between all of these variables and LRS.

Deeper Learning Approaches

Hidden Markov models have been extremely successful in their application to **reinforcement learning** and **temporal pattern recognition**. These successes include gesture recognition, musical score following, and bioinformatics (**vaguely...please do not write such vague words**)

2. Image Analysis and a Behavioral Classifier

Below are steps that can be applied to images (such as videos as time-series images) to perform object recognition, segmentation, finding disscriminative features (*i.e. variables*), and covarianve across images.

Library of image processing, object recognition, and feature detection were created and can be applied on video recordings of green-rumped parrotlets in nest boxes.

Below are example steps for processing and analying nest recordings:

1. Counting Objects

Segment the image, and then calculate the area and eccentricity of each coin. As input, use the image skimage.data.coins.

There are various ways to do this. One would be:

```
1. Equalize the input image (see skimage.exposure)
2. Threshold the image (skimage.filters.otsu)
3. Remove objects touching the boundary (skimage.segmentation.clear_border)
4. Apply morphological closing (skimage.morphology.closing)
5. Remove small objects (skimage.measure.regionprops). Visualize the results
 if you want with skimage.color.label2rgb.
6. Calculate the area and eccentricity of each coin, and display the origina
l image with this information on it (matplotlib.pyplot.text or matplotlib.py
plot.annotate)
```

2. Segment and Compare Regions

3. Area and Eccentricity

4. Panorama Stitching

5. Feature Detection

Example of Object Recognition from Camera Trap Project

camera_trap_processing_example.png

The figure above is taken from a manuscript I am a co-author on that is currently being reviewed in Science. We used deep learning to perform object recognition automatically across over 100k images and identified 20 different species. Most importantly, we decomposed the learning that the deep networks did to find what discriminative variables it learned to distinguish species by. We then could relate those variables back to ecologically meaningful information. This would allow us to create ecological variables of importance for future model building based on what correlations and co-variance structure the deep learning model found.

Mapping Stereotyped Behaviors of Chicks in the Nest

Based on 8 years of nest box recordings covering over 70+ nests and 350+ individual chicks across weeks (*i.e. weeks of recording while in the nest per chick*).

Stereotypy: can we decompose chick behavior (*especially chick-to-chick interactions*) in the nest as **discrete, reproducible behaviors**?

This may be a precursor for future successful genetic and neuroscience approaches to the green-rumped parrotlets.

The workflow for image processing of the nest recordings is described in the next section. Here are some methods that were used:

wavelet: a spectrogram that acts as a spatio-temporal representation of postural time series data

t-SNE: t-distributed stochastic neighbour embedding (*dimensional reduction into 2d time-dependent energy plots*)

An Ensemble Workflow

1. *Pre-process and standardize images of recordings*

2. *Align chicks in same directions (may be multiple clusters of different directions)*

3. *Enforce translational and rotational invariance*

4. *Decompose postures into time-series postural energy plots*

5. *Create spectrograms of postural decomposition via wavelet analysis*

6. *Spectrograms act as spectral feature vectors*

7. *Embed spectral feature vectors into 2-dimensional space via t-SNE*

8. *Estimate probability disctribution over 2-d space to extract peaks*

9. *Pauses near peaks correspond to discrete behavioral states*

3. To the Birds

LRS Questions Addressed by Path Analysis (or analogous approach)

1.a. Quantify the importance of ecological / environmental factors on LRS

1.b. Quantify (and identify) LRS as a function of timing of breeding

Post-Behavioral Classifier (image processing)

2.a. Analyze the percentage of stereotyped behaviors of chicks in nest as a function of age and hatch sequence

2.b. Does percentage of stereotyped behaviors decrease as a time-based decay function (as they age etc)?

Example of Classifying Behaviors

image_processing_example.png

Some Extra Ideas

1. Relationship of vocal babbling, hatch sequence, and brood size
2. Relationship between social intelligence (proxied as paths on social network model), contact calls, and allopreening
3. Relationship between hatch sequence, brood size, and onset of contact calls or babbling

4. Other Projects

T-SEC: Threshold-Smoothing Ensemble Clustering - a clustering algorithm I made that uses entropy-weighted subspace clustering aimed at small sample size, high dimensional data. Written in R and Python but hopefully in a full functional language *(i.e. Clojure and Clojurescript)* soon.

T-SEC is currently used in two published scientific papers.

PEARL: Parasite Extinction Assessment and Red-Listing - a web analytic tool for assessing the risks of extinction across *(currently)* 450+ species of parasites based on **18 climate change scenarios.** Future development would involve creating analytics directly in the web app that would produce dynamic species distribution models based on user parameter tweaks and uploaded species and environmental data. Suggested red-listings for all 450+ species would be self-updating.

http://pearl.berkeley.edu/ (http://pearl.berkeley.edu/)

First version of PEARL: https://github.com/Thru-Echoes/PEARL1.0 (https://github.com/Thru-Echoes/PEARL1.0)

Example of PEARL app

pearl_example.png

```
In [ ]:
```