Machine Learning Engineer Nanodegree

Capstone Proposal

Sam Hiatt
Aug 5, 2019

Background

Many social animals communicate using vocalizations that can give clues to their species as well as to their intent. Studies of animal vocalizations have been hindered by the cost required for manual expert analysis of audio signals. Machine learning offers tools than can automate classification of these audible signals, opening up countless opportunities for sound-aware computer applications and could help accelerate studies of these animals. For example, imagine a computer trained to recognize the call of a specific species of bird. This model could be used to trigger a camera recording, or automatically tag a live audio stream containing avian calls with the species of the bird that made it, producing a timeseries record of the presence of this species.

Sonograms (spectrograms based on sound frequencies) are commonly used for visual representation of audio information and have long been used for interpreting recordings of animal vocalizations. Bird Song Research: The Past 100 Years describes how a device called the Sona-Graph™, developed by Kay Electric in 1948, began to be used by ornithologists in the early 1950's and greatly accelerated avian bioacoustical research.

By applying machine learning techniques for image classification on these sonograms, automated classification of audio clips is possible. The project DeepSqueak at the University of Washington in Seattle takes this approach for classifying recordings of ultrasonic vocalizations of rodents. Their publication in Nature, DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations, uses this classifier to prove correlations between specific behaviors and types of vocalizations.

Problem Statement

This project will build two classifiers to experiment with machine learning for labeling audio recordings of avian vocalizations. One classifier will predict the most prevalent / most likely species of bird in the clip from a specific subset of species commonly heard in the United Kingdom. The other classifier will assume the species is a Zebra Finch and will predict call type.

Datasets and Inputs

Xeno Canto is an online community and database of crowd-sourced recordings of birds from around the world, indexed by species and labeled by call type. The British Birdsong Dataset is a specific subset of the Xeno Canto database, originally compiled by Dan Stowell, and contains 264 recordings with Creative Commons licensing from 88 species commonly heard in the United Kingdom, and it includes a balanced number of samples per class. It is available as a Kaggle dataset here.

The study Individual recognition of opposite sex vocalizations in the zebra finch by researchers at the Max Planck Institute for Ornithology in Germany showed that zebra finches can recognize their mates' vocalizations with a purely audible stimulus. While this study focused on individual recognition, it also produced a high-quality, publicly available dataset of individual Zebra finch calls labeled by six types of vocalizations, called "stack", "kackle", "tet", "hat", "distance", and "male song".

Solution Statement

Drawing on insights from the project DeepSqueak, this effort will attempt to improve upon existing machine learning models for avian vocalization classification by applying a CNN architecture trained on sonograms produced from the input audio files. It is expected that the temporal invariance introduced through convolution and max pooling will achieve improved results compared to models that do not preserve the temporal relationship between input features. Two classifiers will be trained, one for detecting bird species will be trained using the British Birdsong dataset, and the other for detecting call type will be trained using the Zebra Finch dataset.

Benchmark Model

The Kaggle kernel by Edoardo Ferrante uses the same British Birdsong dataset and creates sonograms for each audio sample. It includes a benchmark model showing a Naive Bayes species classifier achieving 86% test accuracy. Another Kaggle kernel by Edoardo Ferrante uses the same sonograms from the previous kernel and implements classifiers with improved accuracy, 97.7% by using a Multi-layer Perceptron model, and 98.3% by using a random forest model.

No existing call type classifiers based on the Zebra Finch dataset were found, so a new benchmark will be needed. A Naive Bayes classifier will be trained on sonograms and used as a benchmark model for the call type classifier.

Evaluation Metrics

sklearn.metrics.accuracy_score will be used to evaluate the overall accuracy of each classification model. It is simply defined as the number of correctly labeled samples divided by the total number of samples. During model selection and hyperparameter tuning 5-fold cross-validation will be used. Final test accuracy will be evaluated against a subset kept aside from the beginning to be used as a test set.

Project Design

First, analyze the chosen datasets looking at the distributions of represented classes and the length of each audio file. Create sonograms and analyze statistics.

Next, review the approach taken for chopping up the audio files and creating the training and test datasets in Edoardo Ferrante's Kaggle kernel. Determine if the implemented data partitioning strategy was appropriate. Could the classifier be picking up on environmental artifacts present in an individual recording, presenting data leakage? If appropriate, readdress partitioning strategy.

Attempt to replicate the results of Edoardo Ferrante's benchmark model then train and evaluate the following classifiers and compare to Edoardo Ferrante's results:

- Naive Bayes Classifier, use this as a benchmark.
- Logistic Regression
- Random Forest Classifier
- MLP Classifier

Implement this same workflow on the Zebra Finch dataset to create a classifier that labels clips by call type.

Now try to improve on these results by implementing a CNN-based classifier. Try several different configurations of CNNs. Try networks that only apply convolution and max pooling along the temporal dimension. Produce a confusion matrix to see which classes are most difficult to distinguish.

Evaluate the best-performing models against the benchmarks and summarize the results of each classifier.

References

- Baker, Myron C. (2001). Bird Song Research: The Past 100 years (PDF). Bird Behavior. 14: 3–50.
- Kevin R. Coffey, Russell G. Marx & John F. Neumaier (2019) DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations Neuropsychopharmacology vol 44, pp 859–868.
- D’Amelio PB, Klumb M, Adreani MN, Gahr ML, ter Maat A (2017) Individual recognition of opposite sex vocalizations in the zebra finch. Scientific Reports 7(1): 5579.
- D'Amelio PB, Klumb M, Adreani MN, Gahr M, ter Maat A (2017) Data from: Individual recognition of opposite sex vocalizations in the zebra finch. Dryad Digital Repository.
- Xeno Canto, founded by Bob Planqué and Willem-Pier Vellinga.
- Edoardo Ferrante. Extract features with Librosa, predict with NB (2019) Kaggle kernel. Version 10.
- Edoardo Ferrante. Bird Visualisation and Classification (2019) Kaggle kernel. Version 2.