Data Mining Applied to Acoustic Bird Species Recognition

Erika Vilches, Ivan A. Escobar, Edgar E. Vallejo
Tecnológico de Monterrey, Campus Estado de México
Atizapán de Zaragoza, 52926, México
A00461595@itesm.mx, A00447647@itesm.mx, vallejo@itesm.mx

Charles E. Taylor
University of California Los Angeles
Los Angeles, CA 90025, USA
taylor@biology.ucla.edu

Abstract

In this work we explore the application of data mining techniques to the problem of acoustic recognition of bird species. Most bird song analysis tools produce a large amount of spectral and temporal attributes from the acoustic signal. The identification of distinctive features has become critical in resource constrained applications such as habitat monitoring by sensor networks. Reducing computational requirements makes affordable to run a classifier on devices with power consumption constraints, such as nodes in a sensor network. Experimental results demonstrate that considerable dimensionality reduction can be achieved without significant loss in classification efficiency.

1. Introduction

A significant amount of our knowledge on bird diversity and behavior is the result of field observations made by expert ornithologists. Bird species identification and the study of their interactions rely crucially on the visual and acoustic abilities of these experts. On the other hand, the identification of individual birds often requires the usage of visual aids, such as color banding.

As a result, the understanding of interactions among individual birds with complex societies has remained elusive. We also have very little information about the interactions amongst species and the influence that environmental facts as rain, earthquakes, predator invasions, etc, have on the behavior of a particular group of birds. Other facts also interfere with the reliability and accuracy of the information, such as human errors and the animal's behavioral changes induced through their interactions with humans.

There are works where other approaches have been explored, such as canonical discriminant analysis [5] which demonstrates that invariant features don't actually provide the most important recognition cues, contradicting some common assumptions in literature.

We propose automatic bird species and individual recognition through acoustic data in conjunction with the existing technology of sensor networks. Automation reduces the limitations of traditional recognition facilitating the ornithologist's work and improving the quality of their research. This allows the experts to focus just on high level bird behavior interpretation, working either directly from their labs or even through the Internet.

For this work, the goal of sensor networks [4, 8, 9] is to introduce a certain number of small sensors or motes in a natural environment in order to acquire data from their surroundings, without human intervention. The large amount of data collected this way demands the use of sophisticated computational tools for their processing. The work reported in this paper is part of collaboration between UCLA and ITESM in the ongoing project "Sensor Arrays for Acoustic Monitoring of Bird Behavior and Diversity" [9] whose specific goal is to monitor different bird species from the ecological reserves in California, USA and Chiapas, Mexico.

2. Methods

2.1. The Bird Songs

Bird Songs for this study were obtained from the Cornell Lab of Ornithology, Macaulay Library [3]. Songs from three species were provided: great antshrike, *Taraba major* (49 song files); dusky antbird, *Cercomacra tyrannina* (79 song files); and barred antshrike, *Thamnophilus doliatus* (76 song files). Each song file has from a few seconds to several minutes of bird calls, with either one, two or more birds singing simultaneously.

The reason to choose these species is because they are abundant in Montes Azules, Chiapas, an ecological reserve where the sensor network will be deployed in the near future.

2.2. Feature Extraction

The study of bird species can be improved, thus we need to use software tools to extract the different features of the signal which will be used later on to analyze and interpret the sound using computers.

Once we obtained the songs in *.wav* format, we loaded them into Sound Ruler [7]. With this software, we are able

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

to see the oscillogram and spectrogram of the signal and within the oscillogram we are able to locate each call from the recording and each pulse within a call. “Calls tend to be shorter, simpler and produced by both sexes throughout the year. Unlike songs, calls are less spontaneous and usually occur in particular contexts.” [2]. It’s important to mention that these songs were only preprocessed through low and high pass filters to facilitate an accurate call and pulse recognition. These filters are species dependant as we can see in Table 1.

	Taraba Major	Cercomacra Tyrannina	Thamnophilus Doliatus
Low-pass filter	3597 Hz	4200 Hz	3597 Hz
High-pass filter	517 Hz	920 Hz	686 Hz

Table 1. Low-pass and High-pass filters per species

Spectrograms are used in bird biology to identify phonetic sounds and analyze the bird songs. They are the result of calculating the frequency spectrum of windowed frames of a compound signal, a three-dimensional plot of the energy of the frequency content of a signal as it changes over time.

The pulse-by-pulse analysis results were saved as comma delimited files. These files contain the 71 attributes of each pulse from the processed samples, representing the bird’s song data. The resulting datasets’ size is as follows: Taraba Major – 21,360 pulse samples, Cercomacra Tyrannina – 5373 pulse samples, and Thamnophilus Doliatus – 911 pulse samples.

2.3. Crossvalidation

Once we obtained the comma-delimited file with the 71 attributes representing a bird’s song, we are ready to begin the data mining. First we must define the crossvalidation scheme to be used that will ensure the proper evaluation, accuracy and reliability of the data mining algorithms. The database constructed from the comma-delimited file will be divided to form training and testing samples. The training part of the database will consist of approximately 70% of the total samples and the testing part will have the remaining 30%.

2.4. Data Mining

When working with bird songs, we have unfortunately to deal with information that is represented as raw data. This information may contain valuable records that may be hidden from the naked eye. We have to apply different computational tools in order to extract the information we require from the raw data. The approach we took was to apply different data mining techniques in order to obtain the most relevant information from the raw data. “Data mining is the extraction of implicit, previously unknown, and potentially useful information from the data.”[11] Once the important data is extracted, we can use only the significant information to feed our classifying algorithms in the sensor nodes in order to recognize different bird species based on their song and call production.

During the development of this project, several data mining algorithms were studied and some were considered and applied to the data obtained from the song samples. The algorithms selected were the decision tree based ID3 and J4.8, the probabilistic classifier Naïve-Bayes and vector quantization. Decision tree based algorithms were chosen to reduce the dimensionality of the problem, to eliminate data set redundancy and for classification. Naïve-Bayes was chosen in order to verify the classification results obtained from the decision trees and because of its affinity with non-redundant, independent data sets, such as the one produced after the reduction with decision tree algorithms’ execution. Vector quantization was chosen in order to convert our original numeric data set into nominal data, a requirement to run ID3 algorithm. By using this algorithm combination, we will be able to compare the full data set classification with the reduced data set classification in order to improve it while reducing the processing power required for its use in sensor networks. The classification improvement on the reduced data set is caused by the attribute dependency elimination by means of the decision tree algorithm.

2.4.1. Vector Quantization. This algorithm was implemented because of the ID3’s lack of numeric support. Quantization [6] is a process in which numeric to nominal data conversion is possible. The algorithm takes an original numeric vector and returns a quantized equivalent numeric vector which can be easily represented by nominal values.

The quantization process calculates two intermediate vectors, partition and codebook. The partition vector is ordered and contains the minimum and maximum values from the original vector plus intermediate values calculated from adding the increase factor to the minimum value of the vector up to the maximum value from the vector. Increase factor is calculated as follows:

$$Increase = \frac{\max(vector) - \min(vector)}{2^{bits-1} - 1}$$

The codebook vector is also ordered and includes values from zero to $2^{bits-1} - 1$ in increments of one. The partition’s vector size is one element lesser than codebook’s vector size. Finally we take each value from the original vector and check in which partition’s vector interval it falls and map it with the corresponding codebook vector’s value for that position. The easiest way to pass these quantized values to nominal values is to set a character equivalent for each codebook value so that you can map them directly. An example of this would be to have the next codebook for a 3 bit quantization: [0, 1, 2, 3, 4, 5, 6, 7] and map it directly with the following vector: [‘0’, ‘1’, ‘2’, ‘3’ ‘4’, ‘5’, ‘6’, ‘7’]. As we can see, the “labels” contained in the last vector are equivalent to the values in the codebook vector. In this way, we obtain nominal representations from numeric values for any set of quantities making it possible to run the ID3 algorithm with them.

In Figure 1 we present a plot comparison from a full original signal with values from 0 to 5000 approximately versus a quantized signal with values from 0 to 6. We can clearly appreciate how the relationship among the attribute values is preserved in the quantized set, even though we can identify some information loss.

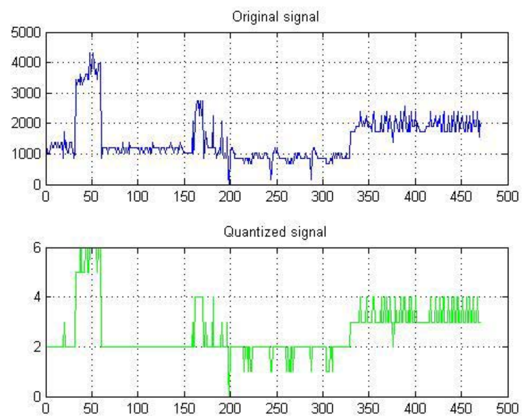

Figure 1. Original vs. quantized signal

2.4.2. ID3. Once we converted the entire species data sets into quantized data, we proceeded to process the information with a decision tree algorithm. "Decision tree algorithms use full binary trees that represent the comparisons between elements that are performed by a particular sorting algorithm operating on an input of a given size." [11] The ID3 algorithm was used to generate the decision tree with Weka [11] software. ID3 is a decision tree algorithm that takes all unused attributes and counts their entropy concerning the test samples to be used. We define entropy as:

$$Entropy(p1, p2,...pn) = \sum_{i=1}^{n} - pi \log_2 pi$$

where $pi = \frac{p_{inst}}{p} g$ and information gain is:

$$Gain(P, xP) = Entropy(P) - Entropy(x \mid P)$$

The algorithm calculates the class's and attribute's entropy and performs a system gain. Then it compares the sample entropies and chooses the one with the maximum information gain or smallest entropy to be the next center node. When the tree is completed, the resulting nodes will be the most significant attributes used to classify the different instances or bird species (the leaves of the tree).

Once we obtained the corresponding decision tree, we only preserved in our data set the attributes that were used in the nodes of the tree (an attribute can be repeated in many nodes). This reduced data set will be used to attempt a reliable classification with the Naïve-Bayes algorithm.

2.4.3. J4.8. This algorithm is an extension of the ID3 algorithm, which solves some deficiencies that the original ID3 algorithm had. Some of the improvements are that J48 avoids over-fitting, uses a reduced-error pruning focus that is based on the consideration that each node of the tree is a prune candidate reducing this way the error, rule post-pruning to find high precision hypothesis and numeric attribute handling. The two main advantages that made us select this algorithm are the computational cost savings and the numeric attribute handling. Weka was used to test this algorithm with our original data sets. The extracted attributes in the reduced data sets were also used to attempt a reliable classification with the Naïve-Bayes algorithm.

2.4.4. Naïve-Bayes. We decided to introduce the Naïve-Bayes algorithm usage as a final classifier because of the main disadvantages that decision tree algorithms have. One if them is that they are unstable. Slight variations in the training data can result in different attribute selections at each choice point within the tree. The effect can be significant since attribute choices affect all descendent sub-trees. Another important disadvantage with decision trees is that trees created from numeric data sets can be quite complex since attribute splits for numeric data are binary.

Naïve-Bayes was executed in Weka, for the original, post-ID3 and post-J4.8 datasets. It is a statistical method based on Bayes rule that naively assumes independence. The Bayes rule says that if you have an hypothesis H and an evidence E then $\Pr[H \mid E] = \frac{\Pr[E \mid H]\Pr[H]}{\Pr[E]}$. Numeric values are handled by this algorithm assuming they have a "normal" or Gaussian probability distribution $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. The mean and standard deviation are calculated for each class and each numeric attribute.

"We know that it is only valid to multiply probabilities when the events are independent. The assumption that attributes are independent in real life certainly is a simplistic one." [10]. In this work, we attempt to eliminate redundancy or dependency in data by means of decision trees (ID3 and J4.8). We use only its extracted attributes to construct the data set that will be fed into Naïve-Bayes trying to assert that we are working only with independent attributes and thus assuring that the learning process is being skewed as less as possible by redundancy and that the maximum efficiency is being obtained.

3. Results

In Figure 2 we can see that the most accurate algorithm is J4.8 (without Naïve-Bayes) obtaining a 98.39% of accuracy. The original attribute number was 71 which this algorithm reduced to 47. We can also appreciate that regarding Naïve-Bayes, the reduced data sets produce a slightly better performance, up to 4.5% improvement.

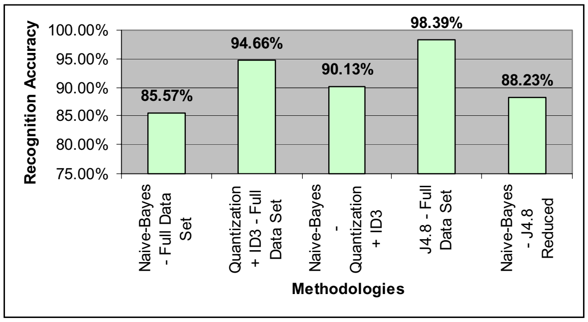

Figure 2. Accuracy percentages graph

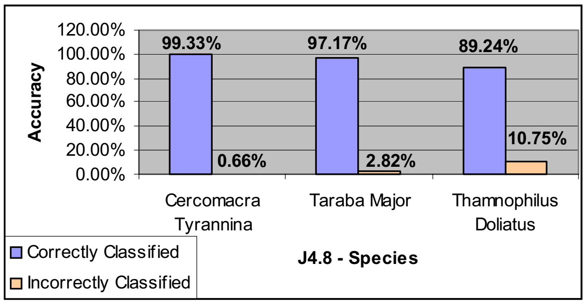

Figure 3. J4.8 accuracy percentages graph

Besides the reliable accuracy preservation, the required processing power is also directly affected from the attribute reduction, since the number of calculations needed to classify in a smaller data set is lower and so is the power consumption.

In the J4.8 tree, the most informative attribute was pulse dominant frequency, the root of the tree. In the next level we find the width of the dominant frequency peak at half of its height divided by the frequency of the peak. One more level down, we find the maximum of dominant frequency in the pulse, the total number of pulses in the call and the dominant frequency at final 50% peak amplitude. These five attributes which J4.8 identified as the most informative ones, contrast with the song duration, number of phrases and number of notes identified by Nelson [4] and the speed, duration, frequency range, and center frequency identified by Bard [1]. The reasons of these discrepancies are probably the use of songs from different bird species and different algorithms for attribute selection, such as canonical discriminant analysis.

4. Conclusions and Future Work

The increase of performance obtained through the combination of decision tree algorithms and Naïve Bayes is due to the elimination of redundant information performed by these algorithms, although an 88.23% or 90.13% are still not enough for reliable classification.

Since this work achieved good species recognition results, we plan to test these same algorithms using recordings from identified individuals expecting to see these excellent results when classifying different individuals from the same species. We also plan to test the efficiency of J4.8 algorithm against the efficiency of Hidden Markov Models [10] for this problem, since they have been tested to be the best human speech recognition algorithms and it is possible that this excellent performance also extends to bird songs.

Finally we plan to take this work into the field and test our algorithms using adapted beamforming microphones with sensor networks and performing live monitoring and classification, expecting to see if our results hold, considering ambient noise and tropical weather interference.

5. Acknowledgement

This research was supported by the National Science Foundation under Award Number 0410438.

6. References

[1] Bard, S. *et. al.;* "Vocal Distinctiveness and Response to Conspecific Playback in the Spotted Antbird, a Neotropical Suboscine". The Cooper Ornithological Society, The Condor 104:387-394, 2002.

[2] Cathpole, C. K., Slater, P.J.B.: *Birdsong Biological Themes and Variations*. Cambridge University Press. 2003.

[3] Cornell Lab of Ornithology – Macaulay Library, http://birds.cornell.edu/MacaulayLibrary/

[4] Mainwaring, A. *et al*.; "Wireless Sensor Networks for Habitat Monitoring". *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications,* September 2002

[5] Nelson, D. "The importance of Invariant and Distinctive Features in Species Recognition of Bird Song", The Cooper Ornithological Society, The Condor 91:120-130, 1989.

[6] Rusell, S.; Norvig, P.; *Artificial Intelligence: A Modern Approach*. Second Edition, Prentice Hall, Englewood Cliffs, New Yersey, 2003. 1132 p.

[7] Sound Ruler, version 0.941, http://soundruler.sourceforge.net/, by Marcos Gridi.

[8] Szewczyk, R. *et al*. "Habitat Monitoring with Sensor Networks", Communications of the ACM, Vol. 47, No. 6, p. 34-40, June 2004.

[9] Taylor, C. *et al*.; "Sensor Arrays for Acoustic Monitoring of Bird Behavior and Diversity", project proposal submitted to National Science Foundation., http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0410438

[10] Wilde, M.; Menon, V.; "Bird call recognition using Hidden Markov Models". *Technical report*. EECS Department Tulane University, 2003.

[11] Witten, I.; Frank, E.; *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. 1st edition, San Francisco, California: Morgan Kaufmann, October 1999. 416 p.