

Using Song to Identify *Cassin's Vireo* Individuals. A Comparative Study of Pattern Recognition Algorithms

Julio G. Arriaga¹, Hector Sanchez¹, Richard Hedley²,
Edgar E. Vallejo¹, and Charles E. Taylor²

¹ Computer Science Dept., Tecnológico de Monterrey, Campus Estado de México
Atizapán de Zaragoza, Estado de México, 52926, México

² Dept. of Ecology and Evolutionary Biology, University of California, Los Angeles
Los Angeles, CA, 90095-1606, USA

Abstract. In this paper, we present a comparative study on the application of pattern recognition algorithms to the identification of bird individuals from their song. A collection of experiments on the supervised classification of *Cassin's Vireo* individuals were conducted to identify the algorithm that produced the highest classification accuracy. Preliminary results indicated that Multinomial Naive Bayes produced excellent classification of bird individuals.

Keywords: Sensor networks, acoustic monitoring of birds, pattern recognition.

1 Introduction

Recent advances in sensor networks technologies hold the potential to transform research in ecology. In effect, the use of sensor networks in natural habitats for monitoring animal behavior and diversity is rapidly coming to be a vital tool in ecology studies [14].

The effective use of sensor networks in ecology relies crucially on the ability to identify relevant events from sensor data. Particularly, the accurate identification of individual animals from sensor data is a necessary condition for analyzing their behavior. Moreover, the identification task must be conducted efficiently given the tight energy constraints imposed by battery-operated sensor networks.

Our research is currently focused on the acoustic monitoring of different species of birds in several areas of the US and Mexico where they are abundant [2]. Our long term goal is to understand the structure and function of bird song. Particularly, the research described here aims at exploring the capabilities and limitations of different pattern recognition algorithms for the identification of *Cassin's Vireo* individuals from their song.

Toward that goal, we have designed a collection of datasets from song recordings representing different properties of the songs, then we applied different pattern recognition algorithms to identify the algorithm that produced the highest

classification accuracy. In the experiments reported here, Multinomial Naive Bayes produced the highest accuracy at the classification of bird individuals among the collection of algorithms used in this study.

2 Methods

The Cassin's Vireo (*Vireo cassinii*) is a small North American songbird which inhabits the area ranging from British Columbia through the western costal states of the U.S. During winter, the bird migrates, traveling from the Sonoran Desert to the south of Mexico. The CaVi is an excellent and persistent singer, with a repertoire of many different phrases strung together while singing.

2.1 Recordings

All recordings were performed on private land five kilometers north from the town of Volcano in Amador County, California between April 25 and June 8, 2013 by Richard Hedley. Recordings were performed opportunistically, beginning the recording when the researcher identified a bird singing and stopping when either the bird stopped for a significant amount of time or changed position becoming inaudible.

Recordings were subsequently segmented and tagged into phrase types -distinct bursts of song of less than a second long, identifiable by its stereotyped delivery- through visual inspection of their spectrograms using the Praat software [3]. Each phrase type was assigned a unique two letter code (aa, ab, ac, etc.) and added to a key file used for further phrase identification. Figure 1 shows spectrogram representations of a small subset of the phrase types. A total of 110 distinct phrase types were identified among the 7 different individuals in our study.

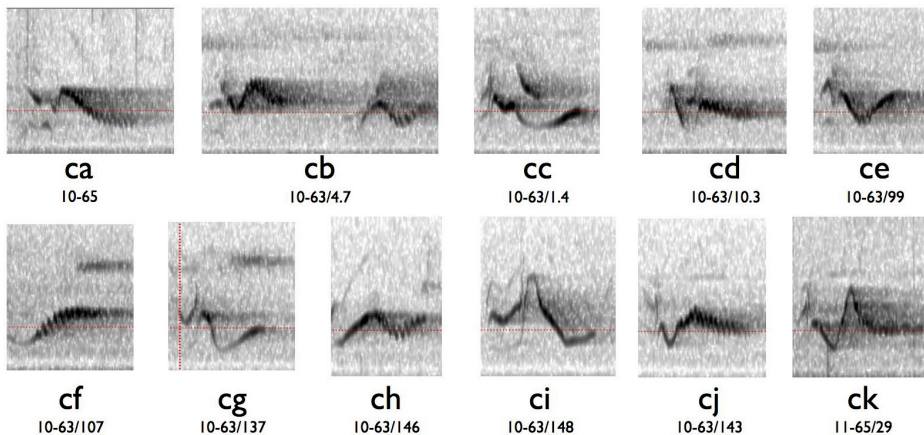


Fig. 1. Sample spectrogram

3 Individual Classification

All available samples were divided into two data sets: one certain and one uncertain. The certain data set contains samples with a high level of individual identity certainty as defined by the recordist: the recordist was able to verify the identity of the bird singing while performing the recording; the uncertain data set contains the remaining samples, those with a low level of identification certainty, determined by the recordist through careful consideration of territories and dates. The certain data set consists of 46 samples, while the uncertain one consists of only 9. For the purpose of this paper, the uncertain data set was not used.

Our approach for individual classification consisted of encoding each sample into a vector using different criteria, labeling it as the appropriate individual, and trying different supervised classification algorithms with 10-fold cross validation to measure their accuracy. Table 1 shows the classification algorithms tested. All algorithms were tested using the implementations found in the WEKA software package [10] with default parameters.

Table 1. Classification algorithms tested

Classifier	Description
Naive Bayes	Probabilistic classifier with strong independence assumptions based on Bayes' theorem. [11]
Multinomial Naive Bayes	Naive Bayes classifier using a multinomial model. [13]
SVM	Non-probabilistic linear or non-linear classifier. LIBSVM implementation. [6]
Multilayer Perceptron	Feedforward artificial neural network classifier learning through backpropagation. [16]
IBK	K-nearest neighbors classifier. [1]
K*	Instance-based classifier using an entropy-based distance function. [7]
Classification Via Regression	Classes are binarized and one regression model is built for each. [8]
PART	Generates a PART decision list by separate and conquer, building a partial C4.5 decision tree in each iteration and choosing the best leaf to be made into a rule. [9]
J48	Generates a pruned or unpruned C4.5 decision tree. [15]
LMT	Classifies by building classification trees with logistic regression functions at the leaves. [12]
Random Forest	Constructs a multitude of decision trees (in our case 10). [4]

Songs were encoded into vectors using different attributes, from acoustical properties of the song to more abstract representations of its structure.

3.1 Classification through Song Acoustic Features

Audio samples were split into smaller files consisting of exactly one phrase per file using the Praat software [3]. Using the Marsyas package [17], 124 audio features were extracted from each phrase and arranged into single vectors.¹

$$\langle AF_0, AF_1, \dots, AF_N, I \rangle. \quad (1)$$

Where AF_i represents an audio feature, and I stands for the individual identification.

Classification through acoustic features was conducted at two different levels: by song and by phrase type.

- Song level: All available vectors were arranged into a single data set.
- Phrase level: Vectors were divided according to their phrase type. The phrase *ch* was chosen for experimentation being the most used phrase type shared by all individuals; this way the largest amount of data possible was available for training and it was possible to assess its suitability for classifying all the individuals.

Results. Experimentation showed that applying a Principal Components filter [10] to the data set, reducing the vector's dimensionality by choosing enough eigenvectors to account for 95% of the variance in the original data, significantly improved the results. Both song and phrase approaches achieved poor results overall managing less than 70% accuracy, with the best results obtained by SVM and LMT for song and phrase respectively (see Figure 2). The main limitation of this approach is its high sensibility to noise in the recordings, as well as the great number of variables (recording equipment, terrain, proximity, weather conditions, etc.) resulting from field recordings.

3.2 Classification through Repertoire

Each sample song was encoded into a single vector.

$$\langle P_0, P_1, \dots, P_N, I \rangle. \quad (2)$$

Where each P_i represents a different phrase type, and I stands for the individual identification.

Repertoire analysis was performed at two different levels: phrase usage and phrase frequency. For phrase usage, each P_n holds a binary value of 0 if phrase Ph_n is not used in the song and 1 otherwise. For phrase frequency, each P_n is given by Equation 3.

$$P_i = \frac{|Ph_i|}{|song|}. \quad (3)$$

Where $|Ph_i|$ is the number of occurrences of phrase Ph_i in the song, and $|song|$ is the length (number of phrases) of the song.

¹ For a complete list of the features extracted, check the Marsyas documentation (<http://marsyas.info/documentation>)

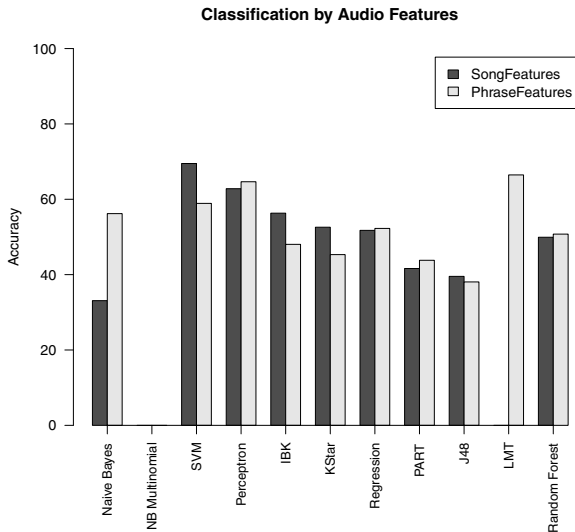


Fig. 2. Classification through song acoustic features. Comparison of results.

Results. Of the methods tested, the combination of the Multinomial Naive Bayes classifier with the binary repertoire vector produced the highest accuracy with 100% correctly classified instances. For the frequency repertoire vector, the multilayer perceptron proved the best match with a 95.65% accuracy (see Figure 3). Previous research suggests CaVis are capable of learning new phrase types from other individuals [5], changing their repertoires over time as a result of this learning process. Repertoire analysis seems sufficient to identify different individuals, however close interactions among neighboring individuals could potentially lead to near identical repertoires diminishing the accuracy of this method.

3.3 Classification through n -Gram Repertoire

Each song sample was encoded into a single vector.

$$\langle NG_0, NG_1, \dots, NG_N, I \rangle. \quad (4)$$

Where I stands for the individual identification, and each NG_i represents a different n -gram. An n -gram is defined as a combination of n phrases occurring consecutively in a song.

N -Gram analysis was performed at two different levels: n -gram usage and n -gram frequency. For n -gram usage, each NG_n holds a binary value of 0 if n -gram Ph_n is not used in the song and 1 otherwise. In the case of n -gram frequency, each NG_i is given by Equation 5.

$$NG_i = \frac{|NG_i|}{|n - grams|}. \quad (5)$$

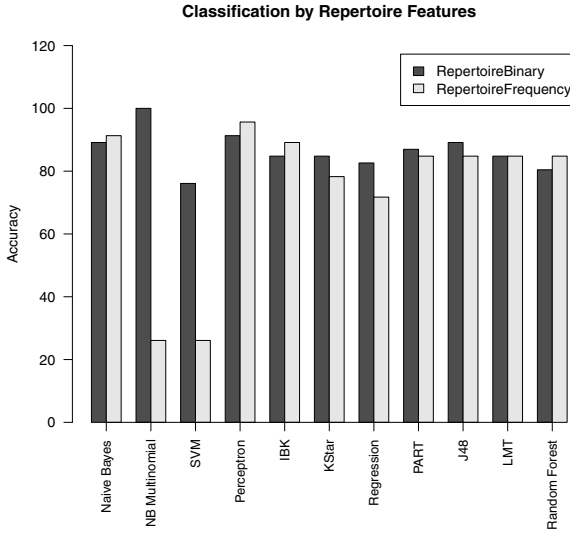


Fig. 3. Classification through repertoire. Comparison of results.

Where $|NG_i|$ is the number of occurrences of different n -gram NG_i in the song, and $|n - grams|$ is the number of n -grams in the song.

Results. N -gram analysis provides the benefit of retaining information about an individual’s repertoire, while also encoding basic extra information about the song’s structure. However, the amount of attributes required to represent it can grow exponentially when compared to just repertoire. For our experiments, we only considered bigrams (combinations of two phrases). As expected, the Multinomial Naive Bayes classifier achieved 100% accuracy (see Figure 4), however the dimensionality of the vector (1347 attributes) proved extremely large for certain algorithms like the multilayer perceptron.

We consider this approach not very useful considering the relatively little amount of extra information gained versus its huge impact on the representation’s size.

3.4 Classification through Song Structure

Directed graphs were created from each sample song, with nodes representing different phrases and edges signifying adjacency between phrases in the song. Figure 5 shows a graph representation of a sample song. Songs were then encoded into vectors.

$$\langle GP_0, GP_1, \dots, GP_N, I \rangle. \quad (6)$$

Where I stands for the individual identification, and each GP_i represents a graph derived property for each phrase type:

- Degree. The total number of edges to and from a node.
- Degree Centrality. The degree of a node, normalized by dividing by the maximum possible degree of the graph.
- Eccentricity. The maximum distance from one node to all other nodes.
- Clique number. The size of the largest maximal clique containing the node.

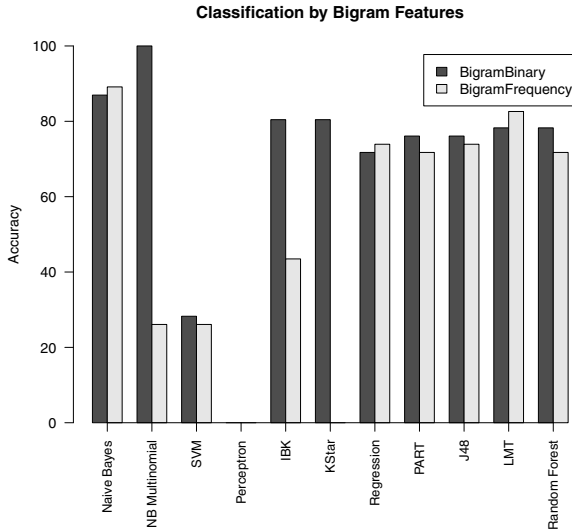


Fig. 4. Classification through bigram repertoire. Comparison of results.

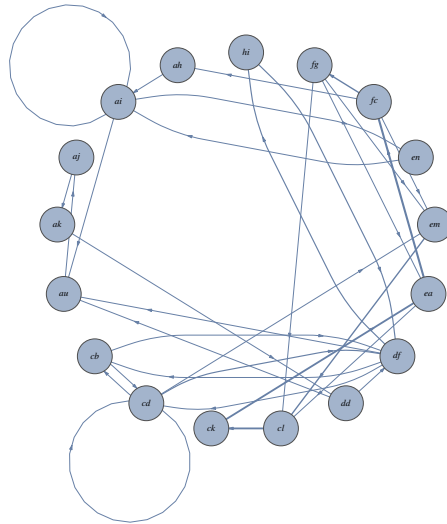


Fig. 5. Graph representation of sample song. Nodes denote distinct phrase types, edges denote adjacency between phrases in the sample song.

Results. Vectors derived from graph analysis have the benefit of containing repertoire information plus extra information about the song’s structure without exploding in the size needed to encode them. Once again, the Multinomial Naive Bayes classifier achieved the best results, with 100% accuracy, for all our approaches except for graph centrality, for which the multilayer Perceptron proved to be the most effective (see Figure 6).

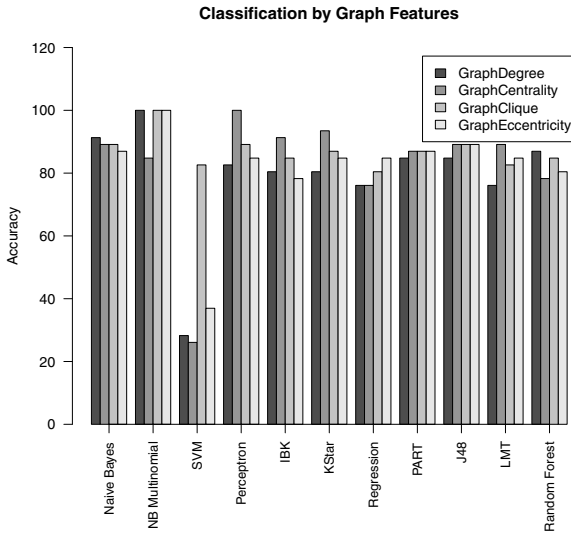


Fig. 6. Classification through song structure. Comparison of results.

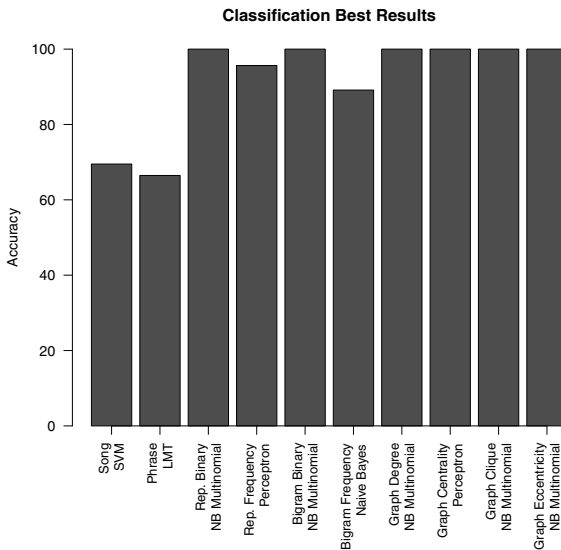


Fig. 7. Comparison of results between best methods for each approach

4 Discussion

In this study, we conducted a series of experiments on the classification of bird individuals from their song. We conducted a comparative study on the classification performance of different pattern recognition algorithms using a collection of dataset representing different properties of the bird song. See Figure 7 for a comparative view of the best results achieved per approach.

Preliminary experimental results indicated that Multinomial Naive Bayes outperformed the other pattern recognition algorithms employed in this study, including more sophisticated and highly regarded algorithms, such as SVMs.

Most notably, being a simple algorithm the generalization capabilities of Multinomial Naive Bayes proved to be highly accurate in contrast to the results produced by the other, more complex, methods. These impressive capabilities have been replicated by this method in other domains such as document classification, among others [18]. All in all, Multinomial Naive Bayes seems to be an excellent candidate algorithm for implementing real time recognition of bird individuals in sensor networks.

Similarly, the dataset encoding attributes extracted from the graph representation of the song allowed the highest accuracy in classification, suggesting that the organization of phrases in songs provides informative attributes that are useful to increase the accuracy of classification of *Cassin's Vireo* individuals.

Further experiments are required to assess the validity of the results reported here, including experiments on larger datasets and different species of birds. We expect that those experiments would contribute to increase the usefulness of sensor networks in ecology studies, generally.

Acknowledgments. This work was supported by the US National Science Foundation under Award Number 1125423 and by Consejo Nacional de Ciencia y Tecnología under Award Number I010/214/2012.

References

1. Aha, D.W., Kibler, D., Albert, M.K.: Instance-based learning algorithms. *Machine Learning* 6(1), 37–66 (1991)
2. Arriaga, J.G., Kossan, G., Cody, M.L., Vallejo, E.E., Taylor, C.E.: Acoustic sensor arrays for understanding bird communication. Identifying Cassins Vireos using SVMs and HMMs. In: 12th European Conference on Artificial Life (ECAL 2013), pp. 287–288. The MIT Press, Cambridge (2013)
3. Boersma, P., Weenink, D.: Praat: Doing Phonetics by Computer (Computer Program). Version 5.3.44., <http://www.praat.org/> (retrieved April 7, 2013)
4. Breiman, L.: Random Forests. *Machine Learning* 45(1), 5–32 (2001)
5. Catchpole, C.K., Slater, P.J.B.: Bird Song. Biological Themes and Variations, 2nd edn. Cambridge University Press, New York (2008)
6. Chang, C., Lin, C.: LIBSVM - A Library for Support Vector Machines, Computer Program (2001), <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

7. Cleary, J.G., Trigg, L.E.: K*: An Instance-based Learner Using an Entropic Distance Measure. In: 12th International Conference on Machine Learning, pp. 108–114 (1995)
8. Frank, E., Wang, Y., Inglis, S., Holmes, G., Witten, I.H.: Using model trees for classification. *Machine Learning* 32(1), 63–76 (1998)
9. Frank, E., Witten, I.H.: Generating Accurate Rule Sets Without Global Optimization. In: Fifteenth International Conference on Machine Learning, pp. 144–151 (1998)
10. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. *SIGKDD Explorations* 11(1) (2009)
11. John, G.H., Langley, P.: Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, pp. 338–345 (1995)
12. Landwehr, N., Hall, M., Frank, E.: Logistic Model Trees. *Machine Learning* 95(1&2), 161–205 (2005)
13. McCallum, A., Nigam, K.: A Comparison of Event Models for Naive Bayes Text Classification. In: AAAI 1998 Workshop on Learning for Text Categorization (1998)
14. Porter, J.H., Nagy, E., Kratz, T.K., et al.: New Eyes on the World: Advanced Sensors for Ecology. *BioScience* 59(5), 385–397 (2009)
15. Quinlan, R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo (1993)
16. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning Internal Representations by Error Propagation. In: Rumelhart, D.E., McClelland, J.L. and the PDP research group (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, MIT Press (1986)
17. Tzanetakis, G.: MARSYAS: A Framework for Audio Analysis (Computer Program). Version 0.2., <http://marsyas.info/download> (retrieved January 8, 2013)
18. Witten, I.H., Eibe, F., Hall, M.A.: *Data Mining. Practical Machine Learning Tools and Techniques*, 3rd edn. Morgan Kaufmann, San Francisco (2011)