How to recognize animal species based on sound – a case study on bumblebees, birds, and frogs

Anton Gradišek, Gašper Slapničar, Jure Šorn, Boštjan Kaluža, Mitja Luštrek, Matjaž Gams
Jožef Stefan Institute
Jamova 39, SI-1000 Ljubljana, Slovenia
anton.gradisek@ijs.si

He Hui
Shanghai University
Shangda Road 98, Baoshan District
200444 Shanghai
China

Tomi Trilar
Prirodoslovni muzej Slovenije
Prešernova 20, SI-1001 Ljubljana, Slovenia
ttrilar@pms-lj.si

Janez Grad
janez.grad@siol.com

ABSTRACT
We present a machine learning-based approach to recognize different types of animal species based on the sound they produce. We focus on bumblebee classification - the algorithm was first developed to recognize bumblebees (roughly 15 most common species found in Slovenia) according to their species and type (queen or worker). Later, it was tested on a set of birds (different species of cuckoos) and frogs of Slovenia. We discuss the sound sample preprocessing, machine learning algorithm, results of algorithm testing, and possible further improvements. A web-based service was developed where users can upload their recordings and further contribute to the learning dataset.

General Terms
Algorithms

Keywords
Sound recognition, machine learning, animal sounds, MFCC

1. INTRODUCTION
Bumblebees (genus *Bombus* from the bee family Apidae) play a key role in the ecosystem as important pollinators. Their different body structure gives them certain advantages over other bees. For example, they can be active in a wider range of weather (bees won't leave the hive when the outside temperature is below 10 °C while a bumblebee is active even below 5 °C). Certain plant species rely on bumblebees as pollinators exclusively, including some cultural plants. For example, bees won't pollinate tomatoes but bumblebees will, which makes them important in the economic sense as well. Selling bumblebee colonies to greenhouses has become a lucrative business in the last decade [1].

There are over 250 species of bumblebee species known worldwide. The biggest diversity is found in Asia while bumblebees are also distributed in Europe [2], North Africa and in the Americas. The highest diversity is found in mountain ranges in temperate climate zones, so perhaps it is not surprising that there have been 35 bumblebee species recorded in Slovenia. Some of these species are either rare or were recorded several decades ago, therefore it is more realistic to say that one can encounter around 20 different species. Bumblebees are social insects; their colonies consist of queens, workers, and males. These types are called castes.

Experts can identify species and caste based on body features, such as the hair colour pattern and body size. For non-experts, some applications have been developed to help with classification, such as *Bumblebees of Britain & Ireland* [3], which provides photos and descriptions of the common species of the British Isles, and *Ključ za določanje pogostih vrst čmrljev* (Key for determination of common Bumblebees), which also provides drawings, photos and descriptions of the common species of Slovenia [4]. Here, we attempt to classify the species and castes automatically, using a computer algorithm. Image recognition is perhaps not the most practical approach due to complications arising from photo quality, light condition, bumblebee orientation, background, etc. Recognition based on the buzzing sound is more promising. In past, there have been attempts to use machine learning-based algorithms to classify different types of insects [5] and also different bird species [6],[7].

In our approach we used Mel-Frequency Cepstrum Coefficients (MFCC) as a feature vector alongside hundreds of others audio features, similar to what was done in the studies mentioned above. Data was preprocessed using Adobe Audition software. Features were extracted using openAUDIO feature extraction tool [8]. Classification algorithms were created using WEKA open source machine learning software. The approach was tested on three groups of animals: bumblebees, with the largest number of samples (11 species, with queens and workers both represented in most cases, 20 classes in total), Slovenian frogs (13 species), and different species of cuckoos (7 species). The recordings of bumblebee were obtained in the field, frog sounds were obtained from the CD Frogs and toads of Slovenia [9] produced by Slovenian Wildlife Sound Archive [10], and the sounds of the cuckoos were obtained from the Chinese database 鸟类网.

In order to make the sound recognition application available to broader audiences, we have developed a web-based service where users can, apart from using only the species classification feature, upload their recordings to be later used in the learning set for further improvement of the classification. The application is now available at animal-sounds.ijs.si It runs in Slovenian, English, and Chinese.

2. PREPROCESSING
First, original sound recordings were manually cut to fragments a couple of seconds long and the sections with no bumblebee sound were excluded. Figure 1 shows a typical (unprocessed) sound file in time domain (*B. hypnorum*, worker) while Figure 2 shows the Fourier transform (absolute value) of data in Figure 1. As seen from the Fourier transform, the relevant frequency window for bumblebee sound is roughly between 100 and 1500 Hz, what is out of this window, can be considered noise. We can clearly see the main frequency at around 200 Hz and the higher harmonics at

multiples of this value. The recordings of bumblebees were typically of good quality and there was no need to additionally filter out background noises since the buzzing sound was by far the most prominent part of the recording.

For frog sounds, the situation was somewhat different. The recordings often contained other sounds, such as other animal sounds (other birds, frogs, insects, etc.) or sounds from sources such as running water etc. Here, background noise was removed by selecting a part of the recording that contains only noise and using standard noise cancellation software tools.

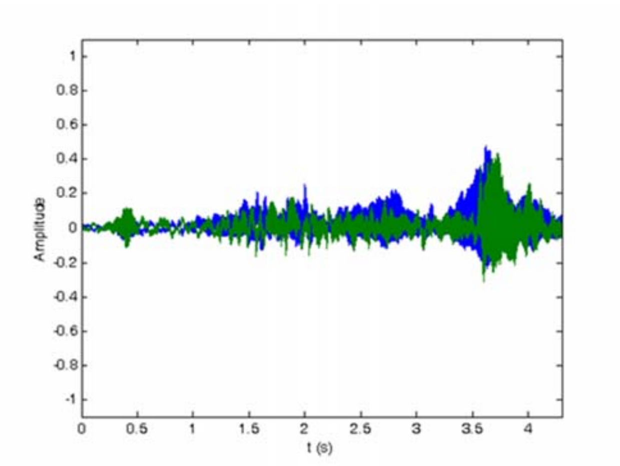

Figure 1. Time-domain representation of a typical sound recording, *B. hypnorum*, worker. Blue and green lines represent the two components of the signal that was recorded in stereo technique.

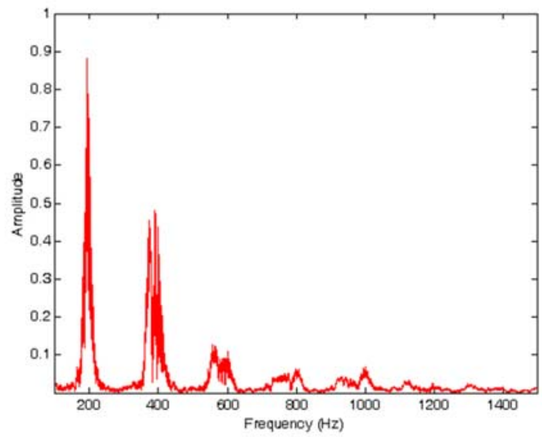

Figure 2. Fourier transform of time-domain data from Figure 1.

3. MACHINE LEARNING AS A SERVICE

Machine learning application was designed following the Machine Learning as a Service (MLaaS) paradigm. This ensures that the data processing, classification model creation, and interaction with client are available within a single cloud service. This animal classification service comprises of three main parts, as shown in Figure 3:

1. audio feature extraction,
2. creation of classification models,
3. user recording processing and serving of results.

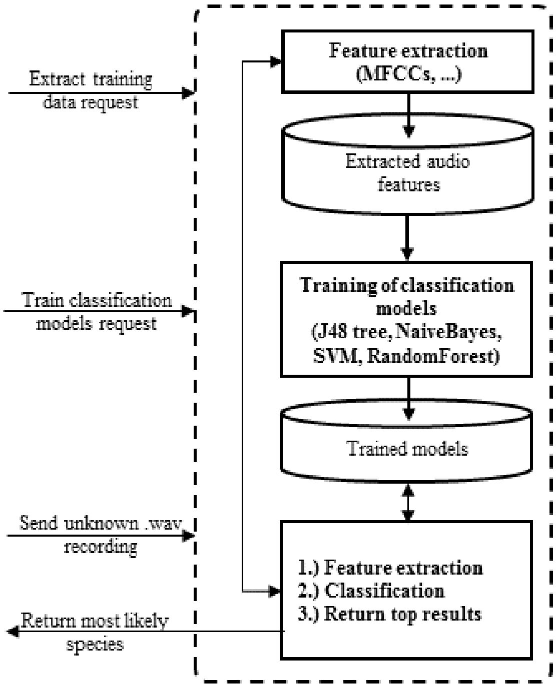

Figure 3. Architecture of animal classification machine learning service.

Audio feature extraction part is responsible for obtaining relevant data, which is then used to create classification models. As input it takes audio files in .wav format. It then computes numerical values representing a large number of different properties of audio signal. Most important among these are Mel-Frequency Cepstrum Coefficients. Following the extraction, the system chooses the best 100 among all extracted features using information gain as a feature quality measurement. These best extracted features are then saved into a database.

The training part takes the extracted features data and uses it to build classification models. It builds the models based on the following algorithms: decision tree, naïve Bayes, Support Vector Machine and Random Forest. All four models are always built, for each animal group. This allows for comparison of classification accuracy based on which we can choose the best performing algorithm. This is described further in section 4.

Third part of the system allows for client interaction. A query with a .wav recording is taken as input and then the system proceeds to extract the same features as were extracted for the training data. It then forwards these features as input into the chosen best classification model, which returns the most likely species.

4. EVALUATION

First we evaluated the results using WEKA built-in kFold cross-validation on all recordings. In this case the results are over optimistic, since parts of the same long recording can appear in both the learning and testing set. This issue was resolved by using evaluation with separate testing (80% of the data) and learning set (20% of the data) where recording slices from each set never belong to the same mutual recording.

In three cases Random Forest algorithm has shown the best classification accuracy while in the case of frogs, Support Vector Machine was slightly superior.

Test results are best presented by means of confusion matrices, which are, in case of bumblebees, too large to be presented in the paper. Evaluation of bumblebee classification shows that the quality of recognition of particular species depends on several factors. Recognition is best in cases where there were several recordings available whereas a small number of clips can result in overfitting the data and the results should therefore be treated with caution. For the classes with at least 20 instances in the classification works best for *B. pascorum*, workers (85%), *B. hypnorum*, worker (100%), *B. sylvarum*, worker (96%), while the classification of *B. humilis*, worker is only 18% accurate. In the attempt to improve the recognition accuracy, we have then decided that the output of the program are three most probable results (as opposed to only the most probable one), together with the pictures of the corresponding species. This additionally helps the user to decide which species was observed.

The results on a set of cuckoos were, on the other hand, surprisingly accurate, with the algorithm correctly classifying 73 out of 74 test instances. The confusion matrix is presented in Table 1:

a	b	c	d	e	f	g	h	<-	-	classified as
15	0	0	0	0	0	0	0	a	=	black-cuckoo
0	6	0	0	0	0	0	0	b	=	himalayan-cuckoo
0	0	12	0	0	0	0	0	c	=	indian-cuckoo
0	0	0	11	0	0	0	0	d	=	lesser-cuckoo
0	0	0	0	12	0	0	0	e	=	madagascan-cuckoo
0	0	0	1	0	10	0	0	f	=	red-chested-cuckoo
0	0	0	0	0	0	7	0	g	=	sunda-cuckoo
0	0	0	0	0	0	0	0	h	=	Unknown

Table 1. Confusion matrix for recognition of seven different species of cuckoos, using the SMO classifier.

The reason for this very high accuracy could be the fact that all the recordings were of extremely high quality (meaning that there was no background noise and the voice was clear) and the songs of different species also differ to the level where an amateur can recognize them only by listening (which is certainly not the case with the bumblebees). The question what would happen if recordings of worse quality were introduced remains open.

In the case of frogs, the recordings were first manually preprocessed with the noise removing software. Original recordings included other animal sounds and sounds of non-animal origin. Furthermore, several species of frogs have more than one type of call and all different calls for each species were grouped into a single class. Nevertheless, the overall classification accuracy was still reasonably high, with 148 out of 179 instances correctly classified (83%).

5. IMPLEMENTATION

Play Framework (Java) was chosen to develop a cloud-based REST service, which offers three endpoints, one for each animal group. WEKA open source machine learning library was used alongside Play Framework to implement the mentioned classification algorithms.

We wanted to offer a unified web application, which would allow users to upload their audio recordings and get the names and images of the most likely species for this recording. Extra functionality is a database in which registered users can save their recordings. Since only good quality recordings are desired in the database we added the feature that only an administrator or a bumblebee/frog/bird expert can confirm these user recordings as suitable, to be permanently added to the database and the learning set.

To do this we developed a Ruby on Rails web application. Web application is easy to use, common to all devices using libraries as Bootstrap and jQuery. The application separates users to ordinary users and administrators, which have different rights to different actions. For authentication of users we take classic session system. The goal of our application was to implement some kind of web portal with audio recordings. Any registered user can add audio recordings of specific animal, which are saved on our server. These audio recordings can be edited by animal experts and be saved to confirmed recordings database. For database we use well known MySQL.

6. DISCUSSION

We have demonstrated that a machine-learning based approach to classify different species of animals by their sounds produces good results. Mel-Frequency Cepstrum Coefficients and other audio features were calculated for each recording and 100 features with the highest information gain were chosen to build classification models. The classification accuracy is excellent in the case of cuckoos, very good for frogs, and variable for bumblebees – some species are classified with high accuracy while some are not. To improve this, three most likely results, together with the corresponding photos, are presented as the output. It is expected that the performance of the classification application will improve when more recordings for each species are available, since some of the classes currently consist of only one or two recordings.

Currently, the preprocessing of the recordings is done manually, the plan is to make this feature automatic as well. In future, we aim to expand the application to include even more groups of animals.

7. REFERENCES

[1] Grad, J., Gogala, A., Kozmus, P., Jenič, A., Bevk, D. (2010): Pomembni in ogroženi opraševalci – Čmrlji v Sloveniji. Čebelarska zveza Slovenije, Lukovica. ISBN 978-961-6516-30-3

[2] P. Rasmont, S. Iserbyt. (2010): Atlas of the European Bees: genus Bombus, Project STEP Status and Trends of European Pollinators. http://www.zoologie.umh.ac.be/hymenoptera/page.asp?ID=169

[3] https://itunes.apple.com/nz/app/bumblebees-britain-ireland/id657076684?mt=8&ign-mpt=uo%3D8

[4] http://www2.pms-lj.si/kljuci/cmrlji/

[5] Diego F. Silva, Vinícius M. A. De Souza, Gustavo E. A. P. A. Batista, Eamonn Keogh, Daniel P. W. Ellis, Applying Machine Learning and Audio Analysis Techniques to Insect Recognition in Intelligent Traps, ICMLA '13 Proceedings of the 2013 12th International Conference on Machine Learning and Applications

[6] Chang-Hsing Lee, Chin-Chuan Han, Ching-Chien Chuang, Automatic Classification of Bird Species From Their Sounds Using Two-Dimensional Cepstral Coefficients, IEEE transactions on audio, speech, and language processing, vol. 16, no. 8, November 2008

[7] Marcelo T. Lopes, Lucas L. Gioppo, Thiago T. Higushi, Celso A. A. Kaestner, Carlos N. Silla Jr., Alessandro L. Koerich, Automatic Bird Species Identification for Large Number of Species, 2011 IEEE International Symposium on Multimedia

[8] Björn Schuller, Florian Eyben, Felix Weninger, Technische Universitaet Muenchen, Germany, openAudio.eu

[9] Tomi Trilar. Slovenske žabe = Frogs and toads of Slovenia. Ljubljana: Prirodoslovni muzej Slovenije: = Slovenian Museum of Natural History, 2003. 1 CD, stereo. ISBN 961-6367-07-2.

[10] http://www.pms-lj.si/si/o-naravi/zivali/oglasanje-zivali/arhiv-zivalskih-zvokov