BirdCLEF 2015 submission: Unsupervised feature learning from audio

Dan Stowell

Centre for Digital Music, Queen Mary University of London
dan.stowell@qmul.ac.uk

Abstract. We describe our results submitted to BirdCLEF 2015 for classifying among 999 tropical bird species. Our test attained a MAP score of over 30% in the official results. This note is not a self-contained paper, since our system was largely the same as used in BirdCLEF 2014 and described in detail elsewhere. The method uses raw audio without segmentation and without using any auxiliary metadata. and successfully classifies among 999 bird categories.

The BirdCLEF 2015 challenge, as part of the LifeCLEF evaluation campaign [1], challenged researchers to build systems which could classify audio files across 999 bird species encountered in South America.

For our participation we submitted a single run from our classifier based on unsupervised feature learning and random forest classifier. This was broadly as used in BirdCLEF 2014, and described in detail in [3]. We refer the reader to that paper for a full system description. We used a single instance of the two-layer unsupervised feature learning process. Figure 1 illustrates the main steps involved in processing.

Differences from the 2014 system included:

- in the downsampling step between the two feature-learning layers, we used L2 pooling rather than max-pooling, which gave a slight improvement;
- we reduced the size of our random forest to 100 due to memory constraints. Our system is fully streaming except for the construction of the random forest; in future it would be interesting to use a streamed implementation such as [2].

Our unsupervised feature learning scales well with increasing data size: linearly, as described in the main paper. However, in our case, due to the compute resources available in the time leading up to the competition deadline we were not able to submit more than one run, nor to apply model averaging.

Our own tests using a two-fold split of the training data confirmed an observation that we made in [3]: adding more layers gives a benefit up to a certain limit, which appears to be related to the size of the available data set. In our tests (Figure 2) the available data appeared insufficient to support a three-layer variant, hence we submitted a two-layer run.

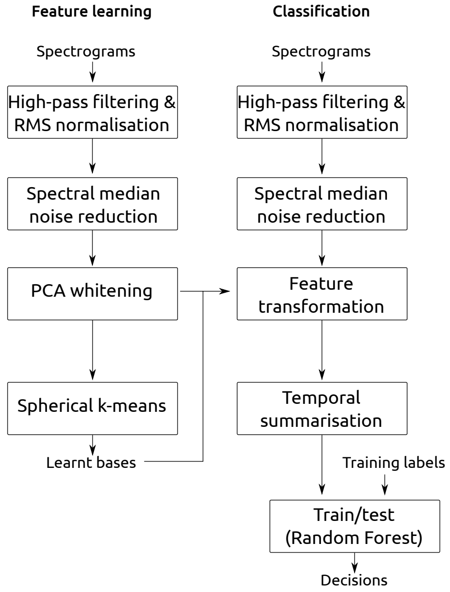
Feature learning
Classification
Spectrograms
High-pass filtering & RMS normalisation
Spectral median noise reduction
PCA whitening
Spherical k-means
Learnt bases
Spectrograms
High-pass filtering & RMS normalisation
Spectral median noise reduction
Feature transformation
Temporal summarisation
Training labels
Train/test (Random Forest)
Decisions

Fig. 1. Summary of the classification workflow, here showing the case where single-layer feature learning is used.

For this 2015 challenge (across 999 bird species with 33,203 audio files) our final MAP score was 30.2% (considering only foreground species), and 26.2% (including background species). These results are a few percentage points lower than the results for the similar systems submitted to the 2014 challenge, as one might expect given that the number of species to identify had been increased from 501 to 999.

Acknowledgments

We would like to thank the people and projects which made available the data used for this research—the Xeno Canto website and its many volunteer contributors—as well as the SABIOD research project for instigating the contest, and the CLEF contest hosts.

This work was supported by EPSRC Early Career Fellowship EP/L020505/1.

References

1. Cappellato, L., Ferro, N., Jones, G., San Juan, E. (eds.): CLEF 2015 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org) (2015), `http://ceur-ws.org/Vol-1391/`

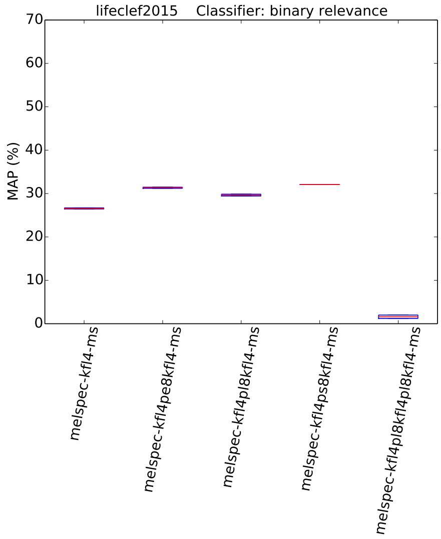

Fig. 2. Evaluation using a two-fold crossvalidation split on the training data. The columns represent a single-layer run, three two-layer runs and one three-layer run.

2. Lakshminarayanan, B., Roy, D.M., Teh, Y.W.: Mondrian forests: Efficient online random forests. arXiv preprint arXiv:1406.2673 (2014)
3. Stowell, D., Plumbley, M.D.: Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. PeerJ 2, e488 (2014)