Study of the robustness of descriptors for musical instruments classification.

Xavier Siebert XAVIER.SIEBERT@UMONS.AC.BE
Mathematics and Operational Research, University of Mons, 20, Place du Parc 7000 Mons, Belgium

Hadrien Mélot HADRIEN.MELOT@UMONS.AC.BE
Algorithms Lab, University of Mons, 20, Place du Parc 7000 Mons, Belgium

Carolien Hulshof C.HULSHOF@MIM.BE
Museum of Musical Instruments, Rue Montagne de la Cour 2, 1000 Brussels, Belgium

Keywords: musical instruments classification, sounds databases, feature selection

Abstract

The current system to classify musical instruments (von Hornbostel & Sachs, 1914; von Hornbostel & Sachs, 1961) has a reducing effect by only considering morphological features. Our research project NeoMI aims at developing a new environment for the organization of musical instruments that takes into account their many aspects. In particular we will focus here on sound-based classification (Herrera-Boyer et al., 2003; Barbedo, 2011; Schedl et al., 2014) by studying two standard databases of instruments, as well as a more challenging databases of instruments of the same family (fiddles). We will present our results in terms of classification and feature selection, and discuss the possibilities and limits of sound-based classification.

This abstract intents to present an extension of a paper accepted at the 6th International Workshop on Folk Music Analysis, Dublin, June 15-17 (http://fma-2016.sciencesconf.org/).

1. Introduction

Our study is based on sound descriptors typically used in the literature (Lartillot et al., 2008), along with several feature selection methods (Fourer et al., 2014) to try to minimize redundancy among these descriptors. It uses two widely used databases, composed of sounds from a variety of instruments, recorded in standard conditions: "MIS", from the University of Iowa[1], and "PHIL", from the London Philharmonia Orchestra[2].

Several classification algorithms (K-nearest neighbors, naive Bayes, Support Vector Machines) were used, along with a complete exploration of all the combinations of the 22 features from MirToolbox (Lartillot et al., 2008) on the MIS and PHIL databases.

We also study a collection of bowed string instruments (15 fiddles families, hereafter called the "MIM" database), consisting of sound recordings of different fiddle types gathered from libraries, personal archives and online sources. The recordings were edited using SoundStudio[3] to get smaller samples of 2 to 4 seconds with minimal environmental noise.

2. Results

We started with the MIS and PHIL databases, for which 30% of the sounds were used as a test set to estimate the percentage of correct classifications, while the other 70% were used as a training set.

The confusion matrices for the PHIL database are shown in Fig. 1, respectively, using one representative classifier (kNN with k=3). Confusion matrices with the MIS database and with other classifiers (kNN with k=1,5; Naive Bayes; SVM) are similar (data not shown).

[1]http://theremin.music.uiowa.edu/MIS.html
[2]http://www.philharmonia.co.uk/explore/make_music
[3]http://felttip.com/ss/

Appearing in *Proceedings of Benelearn 2016*.

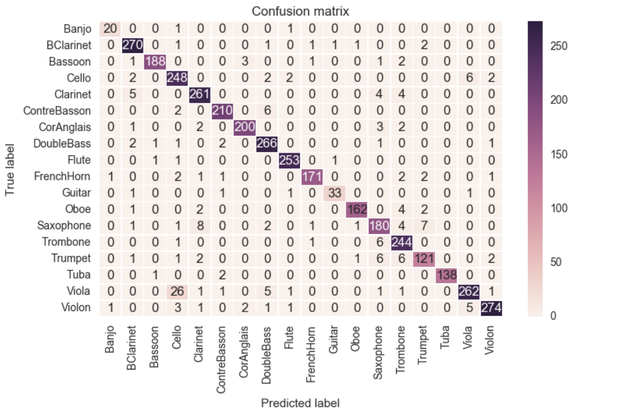

Figure 1. Confusion matrices for kNN classifiers (k=1,3) for the PHIL dataset, using the MIRTOOLBOX descriptors.

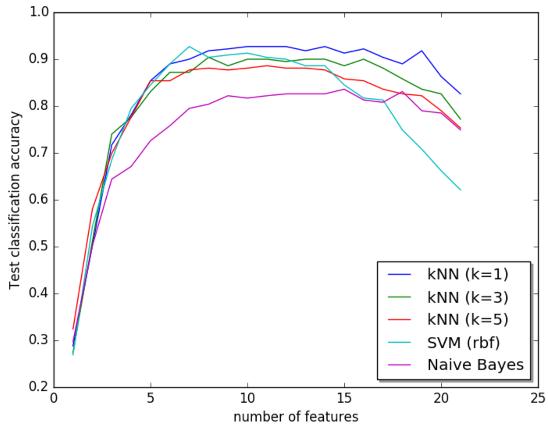

Figure 2. Dependence of the classifier's accuracy on the number of features, for various classifiers.

For the MIS database, the precision is 77%, while the recall is 73%. Some confusion occurs for example among the different types of flutes (altFlute, bassFlute and Flute) or among clarinets. This indicates some difficulty to distinguish between instruments of the same family or whose timbre is similar. For the PHIL database, precision and recall are both around 95%. This reflects the fact that the PHIL database is bigger, but mostly that it contains shorter recordings, each producing a specific note, which simplifies the task of the classifier. Some confusion occurs for example between Cello and Violin, which make sense considering the proximity of these instruments.

To improve these results, we performed feature selection, starting from the observation that not all features were contributing efficiently to the classification. We thus performed a complete combinatorial analysis to find the best combinations among the 22 descriptors initially used, by comparing the best results obtained with several classifiers: k nearest neighbours (kNN) with k values ranging from 1 to 5, naive Bayes and SVM. The results in Fig. 2 show indeed that the classification rate reaches a maximum between 10 to 15 features, before decreasing progressively when increasing the number of features until 22.

Feature selection allows the precision for the MIS database to increase to 86%, and the recall to 84%. However, for the PHIL data-base, the precision and recall remain stable around 94%.

Afterwards, the MIM database was also grouped into classes using either all features or the subset of features identified by feature selection. Because the MIM database is too small to allow 30% of the sounds to be kept aside, we performed an n-fold cross-validation, with a stratified scenario to preserve the percentage of samples for each class and n=9, which corresponds to number of samples in the smallest class.

For the MIM sounds the precision is only 60%, and the recall 51%, with slight improvements upon feature selection. This collection thus poses some difficulties in terms of sound-based classification, notably because of the intrinsic proximity of instruments of the same family, the sparsity and diversity of the recordings.

Together, these results show that a sound-based classification is feasible, but with difficulties arising from instruments of the same family or non-standard recordings. One of the contributions of this type of classification is to help discover possible new links between certain instruments, that were not apparent from classical system to classify musical instruments.

Acknowledgments

This work was supported by a Belspo grant "BRAIN-be" (Belgium Research Action through Interdisciplinary Networks).

References

Barbedo, J. G. A. (2011). Instrument recognition. *Music Data Mining*, 95.

Fourer, D., Rouas, J.-L., Hanna, P., & Robine, M. (2014). Automatic timbre classification of ethnomusicological audio recordings. *International Society* *for Music Information Retrieval Conference (ISMIR 2014)*.

Herrera-Boyer, P., Peeters, G., & Dubnov, S. (2003). Automatic classification of musical instrument sounds. *Journal of New Music Research*, *32*, 3–21.

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A matlab toolbox for music information retrieval. In *Data analysis, machine learning and applications*, 261–268. Springer.

Schedl, M., Gómez, E., & Urbano, J. (2014). Music information retrieval: Recent developments and applications. *Foundations and Trends in Information Retrieval*, *8*, 127–261.

von Hornbostel, E. M., & Sachs, C. (1914). Systematik der musikinstrumente. ein versuch. *Zeitschrift für Ethnologie*, 553–590.

von Hornbostel, E. M., & Sachs, C. (1961). Classification of musical instruments: Translated from the original german by anthony baines and klaus p. wachsmann. *The Galpin Society Journal*, 3–29.