

Title: Automatic tagging in world music collections

Keywords: automatic tagging, music similarity, world music

Abstract:

We are interested in studying music similarity within a corpus of world music recordings. The task of modeling music similarity is subjective. When no annotations are available, music metadata can be used to train models of music similarity. For example, algorithms can be trained to predict music as similar whenever music tracks share tags of genre, mood, or artist. Music similarity algorithms in past research have focused mainly on Western music and the tags associated with it [1], and little has been done on non-Western music. We aim to investigate content-based approaches for automatic tagging in world music for the purpose of studying music similarity in a large corpus later on.

The challenges associated with the analysis of world music result from, amongst others, the degraded audio quality and, with respect to the metadata, the inconsistent taxonomy of world music genres [2]. Large world music archives usually contain material recorded under a variety of recording conditions, e.g., field recordings, and are preserved to different degrees of fidelity. This imposes challenges in the content-based analysis of the audio signal. On the other hand, metadata associated with such collections are usually limited to spatio-temporal information such as the country and year of the recording and cultural background of the performers for example the language and ethnic group or culture. Therefore, the tags describing world music are different from the tags we encounter in Western music and pre-trained models for automatic tagging are not suitable for world music tagging. What is more, world music archives combine the work of many individual collectors and the archive as a whole often suffers from incomplete or inconsistent metadata.

We investigate content-based tagging techniques to overcome the problem of unlabeled data in world music collections. We represent sound recordings by Mel spectrograms and use a multi-layer Convolutional Neural Network (CNN) to learn high-level features. Mel spectrograms and CNNs have been successfully applied in content-based music models [3]. For the purpose of automatic tagging the output of the network is treated as a multi-label classifier. Labels in this case consist of world music tags such as the country, language and culture of a recording. The output of classification indicates which labels (between none or all) are most suited for each recording. We investigate further classification instances with rare outcomes (for example the case where none of the labels were matched) to decide whether a recording denotes a new tag, e.g., a new country, language or culture not included in our training set. The discovery of new tags in this manner can have a significant impact in the musicological analysis of world music collections. Automatic tagging facilitates further computational analysis and enables the exploratory study of music similarity in a large corpus of world music.

References:

- [1] Choi, K., Fazekas, G., & Sandler, M. (2016). Automatic tagging using deep convolutional neural networks. In *International Society for Music Information Retrieval Conference* (pp. 805–811).
- [2] Serra, X. (2011). A Multicultural Approach in Music Information Research. In *International Society for Music Information Retrieval Conference* (pp. 151–156).
- [3] Dieleman, S., & Schrauwen, B. (2014). End-to-end learning for music audio. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6964–6968).