Music Genre Classification
COS424

Vladimir Feinberg, Siddhartha Jayanti

9 February 2016

1 Abstract

blah blah

2 Introduction

A basic problem in supervised machine learning for audio data is to label a song into a genre such as blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock. This problem is difficult because the representation of sound may translate into these genres in a nontrivial fashion. Furthermore, a given song can have multiple genres - two people may disagree on its labelling even if they agree on a large set of other songs. This implies that even in nature the learning process for genre classification leads to inconsistent results. Thus, for a given labelling, it is difficult to expect there to be an arbitrarily accurate classifier. Finally, any given pair of genres differ in amount of variation, with some being easier to distinguish than others.

We will explore a variety of supervised learning methods and evaluate their performance for the GTZAN dataset (TODO cite).

3 Data representation

We work with a small version of the dataset, with 1000 songs. Each song is labelled with one of the aforementioned 10 genres.

The features available in the genres are (TODO, make a latex list and describe dimensionslaity). Don't mention fisher vectors here.

TODO describe need for padding, why justified (can help distinction compared to chopping)

4 Previous Work

The nature of the sound data includes many features which are not predictive of the labels - there is noise in the information-theoretic sense. Secondly, there are variations for songs within genres whose differences are present in the feature vectors for the individual clips, but these types of more subtle musical notes do not interest us.

An often-applied dimensionality reduction technique is PCA; however, its unsupervised nature makes the reduction uninformed with respect to predictive features [3].

This method is used on *timbre-related* musical features in Enrique et al. [1]. They obtain a 4.09% error when they use the MFCC and Fisher-LDA. A later paper by Chang et al. [2] uses both short and long term features of music and the *compressive sampling* technique to keep the dimensionality low. They obtain 92.7% accuracy when they use multiple short and long term features.

References

[1] ALEXANDRE, E., CUADRA-RODRGUEZ, L., GIL-PITA, R., AND ROSA-ZURERA, M. Application of fisher linear discriminant analysis to speech/music classification. In *Audio Engineering Society Convention 120* (May 2006).

[2] CHANG, K. K., SHING ROGER JANG, J., AND ILIOPOULOS, C. S. Iliopoulos: music genre classification via compressive sampling. In *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR* (2010), pp. 387–392.

[3] WELLING, M. Fisher linear discriminant analysis. unpublished note.