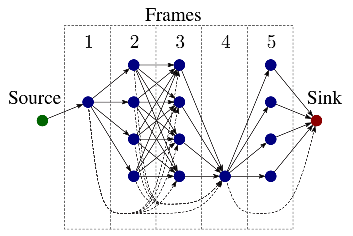

Music Classification using Neural Networks

Amlan Kar, Chaitanya Ahuja

Introduction

Music Classification is a well known problem and has been researched fairly well in literature. From techniques involving traditional **signal processing** approaches involving handcrafted features (FFT, Cepstrum) and modern learning algorithms like **Random Forests(RFs) and Deep Neural Networks(DNNs)** [1], music classification reduces to a problem of good feature extraction. **AdaBoost and Aggregation** of features is used in [2, 3] for the same purpose.

Methodology

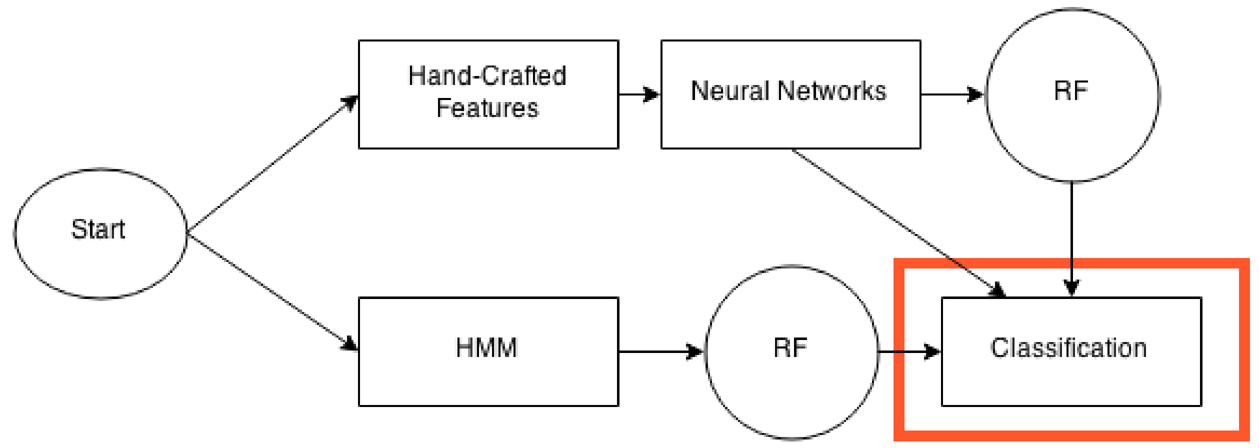

Fig.2: Flowchart of the methods followed

Theory

DNN: - A Deep Neural Network is just another artificial neural network with more than one hidden layers between the input and the output layers.

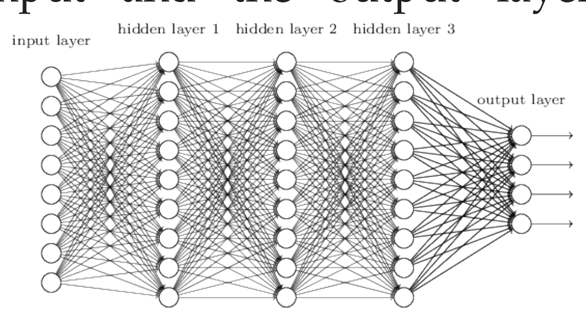

Fig.3: A typical DNN [a]

Dropout: - A way to prevent neural nets from over-fitting. Basically every node in the neural net is given a probability with which it could be present in the net during a training epoch.

HMM: - A Hidden Markov Model is **a special bayes net with properties that make it particularly applicable to temporal data modelling**. Using HMMs to generate features doesn't seem to be a very well-researched topic and we have made an attempt towards it in our project.

Here, we have used the HMM to generate sequences of hidden state numbers for each 30 second splice, which is used further as a feature in the random forest classifier.[4]

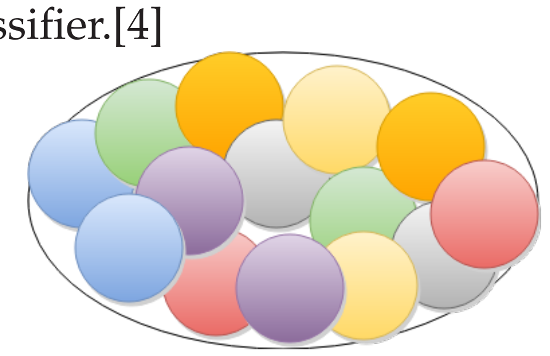

Fig.4:Visualization of a RF Classifier

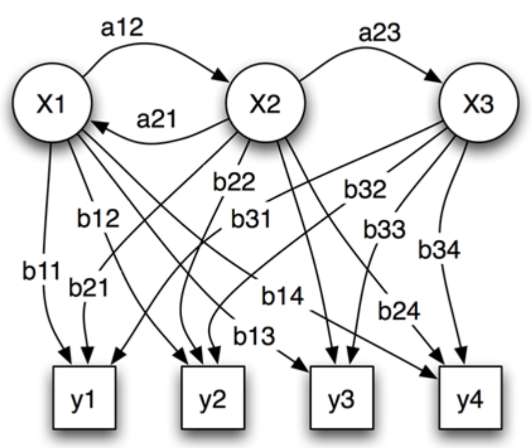

Fig.5: A typical HMM [b]

RF Classifier: - Random forests are an **ensemble learning method** for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. [c]

[a] Image taken from neuralnetworksanddeeplearning.com

[b] http://homepages.inf.ed.ac.uk/group/sli$_{archive/slip0809_c/s0562005/img/HiddenMarkovModel.png}$

[c] from wikipedia.org

Problem Statement

Take a music sample and predict its genre or artist

Fig.1: Black-Box Model of the problem

Results

After training the neural nets on the training data(90%) for 50 and 500 epochs, we tested the results on the testing data (remaining 10%). The HMM model was also trained on the same datasets to maintain comparability.

Method	Accuracy	Benchmark
DNN-50 epochs	0.48	NA
DNN-500 epochs	0.56	NA
DNN-RF-50 epochs	**0.62**	0.718
DNN-RF-500 epochs	**0.63**	0.656
HMM-RF		NA

Table 1: Genre Classification on GZTAN dataset

Method	Accuracy
DNN-50	**0.7573**
DNN-RF-50	**0.8738**
HMM-RF	

Table 2: Artist Classification on Self-Created dataset[a]

[a] All the rights of the music files remain with the respective production houses

Future Work

1. Dimensionality Reduction on feature vectors
2. Using HMM features on Neural Net
3. Expanding artist classification problem to multiple genres
4. Using autoencoders for feature generation
5. Extend the better feature extraction to transcription of music notes

Experiments

Case-1: We trained a neural network of the form 513-50-50-50-10 [a]
Input: FFT of a 512 point window of 30s audio clip.
Output: Probabilities of the classification output.
Activation function: Sigmoid function $\frac{1}{1-\exp^{-(Wx+b)}}$
Case-2: Use the first hidden layer to train a RF classifier to predict classes
Case-3: Use MFCC features to train an HMM model which is then used to train an RF classifier.

Dataset: For Genres we used GZTAN [5] dataset containing 10 genres with 100 30s audio clips**. For Artists we created a similar dataset using music of 10 artists of the genre (blues).**

[a] Code Adopted from: https://github.com/sidsig/ICASSP-MLP-Code

Conclusions

1. The accuracies obtained for a simple DNN is surprisingly well, because it is able to perform at **around 50% accuracie**s on 513 points of data.
2. After **aggregation of different frames** of a given audio and using maximum pooling in an RF classifier, the accuracy boosts up by a significant percentage.
3. HMM modelling have been shown to extract useful features with speech signals [4], they do not seem to work well for music features.

References

[1] Siddharth Sigtia and Simon Dixon. Improved music feature learning with deep neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 6959–6963. IEEE, 2014.

[2] James Bergstra, Norman Casagrande, Dumitru Erhan, Douglas Eck, and Balázs Kégl. Aggregate features and adaboost for music classification. *Machine learning*, 65(2-3):473–484, 2006.

[3] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *Speech and Audio Processing, IEEE transactions on*, 10(5):293–302, 2002.

[4] R Gajšek, F Mihelič, and S Dobrišek. Speaker state recognition using an hmm-based feature extraction method. *Computer Speech & Language*, 27(1):135–150, 2013.

[5] Thierry Bertin-Mahieux, Daniel PW Ellis, Brian Whitman, and Paul Lamere. The million song dataset. In *ISMIR 2011: Proceedings of the 12th International Society for Music Information Retrieval Conference, October 24-28, 2011, Miami, Florida*, pages 591–596. University of Miami, 2011.