LEARNING REPRESENTATIONS FOR MUSIC AUDIO DATA WITH FEED FORWARD NEURAL NETWORKS

Siddharth Sigtia and Simon Dixon

Queen Mary University of London, Centre for Digital Music
{siddharth.sigtia,simon.dixon}@eecs.qmul.ac.uk

ABSTRACT

Finding relevant representations is an important requirement for MIR tasks to be performed successfully. The recent advancements in the field of representation learning provide several ways to automate the process of finding features. Amongst these feature learning methods, Deep Belief Networks have enjoyed substantial success in the fields of vision, natural language processing and speech. However one of the drawbacks of using these systems is that they take very long to train. In this study, we use feed-forward neural networks with rectifier activations to learn features for music audio data.We show that the features learnt with the rectifier networks perform as well as or better than the features learnt by networks with sigmoid activations, while providing a substantial reduction in training time. These learnt features perform better than most of the standard handcrafted features. We show that the use of regularization techniques like dropout during training, reduces the error on classification and auto-tagging tasks on unseen data and eliminates the need for an unsupervised pre-training stage. We also explore the use of Nesterov's Accelerated Gradient method to learn better features, without using second order optimization techniques.