

Research on Music Automatic Annotation Model Based on Deep Learning

Numao Laboratory • HAN Ning



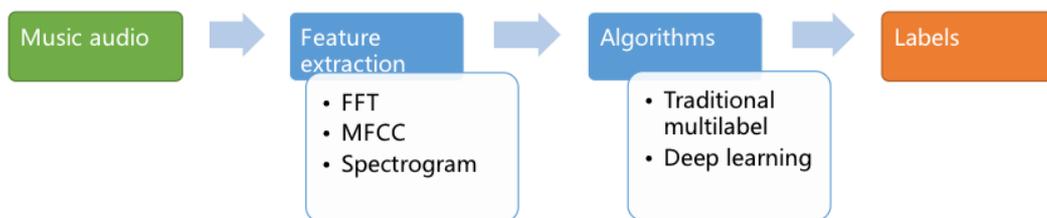
Introduction

Music annotation generally refers to the generation of music labels to describe the music semantics. The rapid development of the digital music market has brought huge resources of digital music. So music labels as structured information organization is increasingly important. It will help the massive music information retrieval, efficient management, and personalized recommendations.



Music automatic annotation process

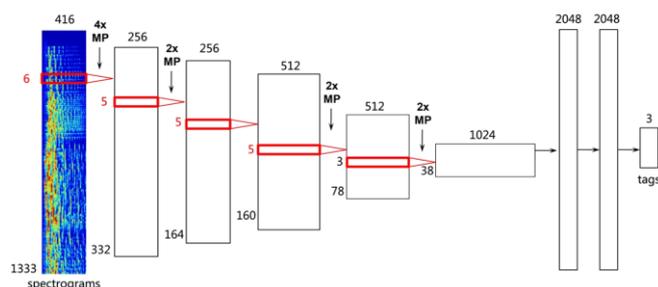
Traditional automatic annotation algorithm is a two-stage process. It still has some problems to be solved. Recently deep learning algorithm in the academic concern has made huge progress in image and voice field, which show that the algorithm has great potential to solve the music automatic annotation task.



Deep learning with CNN

This research implements convolutional neural network (CNN) which represents deep learning algorithms. CNN is a type of feed-forward artificial neural network. It has several distinguishing features, such as 3D volumes of neurons, local connectivity and weight sharing. These features help to get high-level semantic information extracted from low-level semantic information of the music, and effectively reflect the time-related features of music.

When implement CNN to train music automatic annotation model, first cut off music audio to several clips, then generate spectrogram as CNN's input. Through convolution, pooling and full-connected layers, learn probability of different tags. Finally, join probabilities of all clips. The structure of mine CNN is as follow image.



Dataset and experiments

The dataset, consisting of 1040 Chinese popular songs, is labeled manually. The labels set has 23 labels that cover beat, rhythm, mood and so on, which describe the different characteristics of songs.

As comparison, FFT features of audio with traditional multi-label classification algorithm is also implemented. When choose precision and recall as criteria, the results as below show that deep learning performs better in key labels.

