
MLP Coursework 3: Deep Networks Music Classification Based on FMA

s1738623, sXXXXXXXX, sXXXXXXXX

Abstract

The abstract should be 100–200 words long, providing a concise summary of the contents of your report.

1. Introduction

Machine learning is a

Deep learning(DL) is a subset of machine learning and there are three main types of learning: supervised, semi-supervised and unsupervised. In a deep learning model, there are multiple layers each of whose input is from the former layer's output. Many of current deep learning models are based on Artificial neural network (ANN) which includes deep neural networks (DNNs). A deep neural network contains many hidden layers between the input layers and output layers. One of advantages of DNNs is the ability of finding non-linear relationships which are often complex. Besides DNN, there are some other deep learning algorithms, such as Recurrent neural networks (RNNs) and Convolutional deep neural networks (CNNs). RNNs allows data going in any direction and are mainly used as language model. CNNs are mainly used in computer vision domain.

Music information retrieval(MIR), as a branch of information retrieval(IR), is the science of retrieving information from music. There are many developed IR application in the world, for example, the text information retrieval from search engine. Unlike text information retrieval, MIR research is still not much due to many reasons. One of the reason above is MIR often requires a comprehensive background in music, psychology, machine learning etc. Another reason is the lack of numbers of large, complete and available datasets. (?)

In the rest of the report, there are mainly eight sections. Motivation, research questions and objectives will be stated in section 2 to 4 respectively. Section 5 is a brief introduction on FMA datasets. In methodology part (section 6), a concise conclusion of machine learning practical will be given and some novel knowledge used in our experiment will be introduced as well. After experiments part, there is a interim conclusion containing results analysis. At the last, future plan of the project is pointed out.

2. Motivation

Machine learning performs good at classification, and deep learning often has a better performance on finding non-linear relationships. Music is a kind of thing containing a

lot of information which is hard to find the relative relationships inside. Therefore, compared other methods, machine learning, especially deep learning, is a suitable approach to deal with music issues.

Besides deep learning or machine learning, music, which is a natural habit of human, has a remarkable potential market to be explored. For example, a good music classification can bring clear learning schedules for new learner. Also, music recommend systems based on personal preference will be popular because of the increasing cognition of self-value.

Therefore, in this report, our motivation is to find an approach based on deep learning to classify the music well. Besides that, we also want to find some other classification models and make some comparison. All the experiments are based on FMA datasets and classification accuracy is the evaluation of models.

3. Research questions

Unlike the wealth of other information retrieval, lacking large, complete and available datasets for MIR makes MIR research develop slow. Except for FMA dataset, other existing datasets for MIR are either in small scale or not complete. For example, despite containing 2524739 clips, the dataset named AcousticBrainz has no information on artists or audios. Dataset called Unique provides a more complete information, but its capacity(3115 clips) is far away from the FMA(106547 clips). (?)

dataset	clips	artists	year	audio
RWC	465	–	2001	yes
CAL500	500	500	2007	yes
Ballroom	698	–	2004	yes
GTZAN	1000	300	2002	yes
MusiClef	1355	218	2012	yes
Artist20	1413	20	2007	yes
ISMIR2004	1458	–	2004	yes
Homburg	1886	1463	2005	yes
103-Artists	2445	103	2005	yes
Unique	3115	3115	2010	yes
1517-Artists	3180	1517	2008	yes
LMD	3227	–	2007	no
EBallroom	4180	–	2016	no
USPOP	8752	400	2003	no
CAL10k	10271	4597	2010	no
MagnaTagATune	25863	230	2009	yes
Codaich	26420	1941	2006	no
FMA	106574	16341	2017	yes
OMRAS2	152410	6938	2009	no
MSD	1000000	44745	2011	no
AudioSet	2084320	–	2017	no
AcousticBrainz	2524739	–	2017	no

A complete, large dataset containing more information may reveal better relationships between data by fitting them into deep networks. Therefore, our project, which is based on FMA, aims at build MIR classification models in a new, better data environment than before. Also, in order to find a suitable model, we build different models and compare them by their accuracy. Although FMA, which contains both text and audio information both, provides a good audio resource, we still want to explore the text information in this report. The reason for this is mainly about the casual habits that ordinary people often are used to classify the types of music by artists, albums and others belonging to text information. Therefore, we really want to figure out one or several models classifying the music type based on these text characters.

4. Objectives

In this interim report, we mainly illustrate the relationship among the machine learning, deep learning and music information retrieval. After a comprehensive research, we have found some problems leading to several worthy research points. As the priority, we want to build a classification based on DNNs by using the FMA dataset. Besides that, there are two optional goals as well. One is finding more suitable classification models based on the FMA dataset, such as RNNs etc. The other one is trying to build a recommending model based on our classification results.

5. Data

The dataset in our experiment is called Free Music Archive (FMA). It is a dataset opening for free and suitable for evaluating various MIR tasks. FMA dataset contains both texture and audio content inside. For texture content en-

code as csv format, there are some documents recording different information about songs, albums, artists etc. For audio content encoded as mp3, there are four different size of packages, which contains 8000 tracks in 30s (7.2GiB), 25000 tracks in 30s (22GiB), 106574 tracks in 30s (93GiB) and 106574 tracks untrimmed (879GiB) respectively.

In this report, as a mid-term report, our work is mainly about the classification issues. We extract the main text features, such as article, created date, length etc., from the dataset and split it into train set, validation set and test set. Then build different models to fit the data respectively. The evaluation of our task is accuracy of classification.

6. Methodology

7. Experiments

8. Interim conclusions

9. Plan