Music Genre Recognition using Machine Learning

Project Proposal for Machine Learning Course

Ayush K. Rai
Ecole CentraleSupelec
Paris, France
ayush.rai2512@student-cs.fr

Louis De Vitry
Ecole CentraleSupelec
Paris, France
louis.devitry@student-cs.fr

Alami C. Mohamed
Ecole CentraleSupelec
Paris, France
m.alamichehboune@student-cs.fr

MOTIVATION AND PROBLEM DEFINITION

Music Information Retrieval (MIR) is a field that involves retrieval of useful information from music and has many real world applications. Music Genre Recognition is one such application, which over the years has received a lot of attention not only from MIR research community but also from giant tech firms. Nowadays companies like Spotify and SoundCloud recommend music to their customers using personalized music recommender systems and music genre recognition is an important component of building those recommendation engines. Another application of music genre recognition might be to use the predicted genres or sub-genres and combine them with the music meta-data and acoustic features (extracted using signal processing techniques) in order to group similar songs together. Finally music genre recognition can also help a music artist or music composer to understand which music genre is popular among different section of the audience.

We strongly believe that this problem has a great relevance from an educative point of view and can also be further extended into a research project.

Therefore motivated by the above reasons, we are proposing to work on music genre recognition as a multi-label classification problem by applying various machine learning techniques on Free Music Archive (FMA) dataset [2] as a part of our graduate level course in machine learning at Ecole CentraleSupelec.

1 RELATED WORK

Music Genre Recognition is certainly not a novel problem in the field of Music Information Retrieval (MIR) and many studies and research papers have been published on tackling this problem using the whole spectrum of machine learning methods. In the pre-deep learning era, Mel Frequency Cepstral Coefficients (MFCC) based features were a prominent way of addressing this task. [5] trained a Hidden Markov Model (HMM) model using MFCC features. [7] worked on the problem of audio music mood classification by empirically selecting multiple descriptors to extract spectral, temporal, loudness features and further trained a SVM (Support Vector Machine Model) to perform classification.

With the rise of deep learning algorithms many researchers [1] have used Convolutional Neural Networks (CNNs) and Long Short-Term memory (LSTMs) based models over traditional machine learning algorithms to attack the problem of music genre classification. [10] proposed that spectrograms can be considered as images and can be further used to train a convolutional neural network.

2 METHODOLOGY

In this paper, we will explore music genre recognition as a supervised and unsupervised categorization and how they interact with each other. Firstly we will give a description of the datset we are using for this work. After that we explain our approach, which will be three fold.

First, we will discuss and outline our intended supervised machine learning pipeline, which will handle data pre-processing, feature engineering, model training, hyper-parameter optimization, post-processing.

Furthermore, on this pipeline powered by hyper-parameters optimization[4], we will test and benchmark conventional machine learning algorithms, namely random rorests, support sector Machines ensemble learning based models like adaptive boosting, gradient boosting and k-Nearest-Neighbors (with and without Dynamic Time Warping) to the dataset [3].

Finally, we will jointly apply clustering and dimensionality reduction technique to see how unsupervised categorization can further help us discriminate genres. This step will be executed in parallel of making the baseline in order to get sense of the data. For most of our work we plan to use Scikit Learn Library in Python [8].

If we have time, we will explore deep learning in the same settings and also look into other datasets like [3].

2.1 Dataset Description

Other fields of artificial intelligence like computer vision have many established and benchmark datasets like ImageNet and MSCOCO but in MIR there has been always been a dearth of such large scale datasets. In ISMIR 2017 Free Music Archive dataset [2] was published to aid this issue. The data includes:

- Audio track (encoded as mp3) of each of the 106,574 tracks. It is on average 10 millions samples per track.
- Nine audio features (consisting of 518 attributes with statistics such as mean, standard deviation, skew, kurtosis, median, minimum, maximum) for each of the 106,574 tracks.
- Metadata about includes song title, album, artist, genres; play counts, favorites, comments; description, biography, tags.
- The dataset is split into four sizes: small (8 balanced genres), medium (16 unbalanced genres), large (161 unbalanced genres), full (161 unbalanced genres). We may be able to work

on the small one on our computers, but at some point, the experiments and model training will need to be supported by cloud computing (AWS or similar services).

The dataset comes with many baselines models which will be helpful to measure the accuracy of our adopted approach. Also for the purpose of cross validation, we plan to divide our dataset in such a way that 60% of the data will be used for training, 20% for validation and rest of the 20% for testing (generalization) purpose.

2.2 Supervised Machine Learning pipeline

In most Machine Learning settings, transforming raw data to machine learning ready models can be gruesome. To ensure we leave no room for error, we envision to do the following:

2.2.1 ***Pre-processing:*** A thorough Exploratory Data Analysis (distributions of features, multicollinearity matrix, unsupervised learning...) will be done. Additionally, techniques such as scaling (standard and min-max) will be applied and and features transformation (log, sin, ...) will be envisioned.

2.2.2 ***Feature engineering:*** In most papers based on this data set, the features used are statistical and acoustical measures of the song, such as Chroma, Tonnetz, Mel Frequency Spectral Coefficient, Spectral centroid, Spectral bandwidth, Spectral contrast, Spectral roll-off, Root Mean Square energy, Zero-crossing rate with their statistics. Meta-data, such as the year of publications or the artist constitute an essential prior knowledge that must be taken into account. We planned to incorporate the meta-data into the standard dataset. To do so, one-hot encoding and meta-data embedding will be explored.

2.2.3 ***Supervised Methods:*** In a supervised setting a natural framework for music genre recognition is multi-label classification. We aim to scrutinize algorithms such as random forests, support vector machines, ensemble learning based models like boosting and gradient boosting.

2.2.4 ***Hyper-parameter optimization:*** Machine learning based pipeline demand a fine tuning of its hyperparameters, which define the structure of many data-processing (min-max scaling vs. standard scaling) and Machine Learning algorithms (number estimators and depths of Random Forests for instance). In most research, grid search and random search are the go to options. We will however go further by using Bayesian Hyperparameters optimization on the full pipeline[4], from pre-processing to calibration.

2.2.5 ***Post-processing:*** The outputs of classifier models are probabilities. Those probabilities need to be cast into the actual classes (the genres in our case). To solve this, a typical threshold function is applied to the probabilities. However, imbalanced classes and other phenomenons tend to make the probabilities ill-calibrated. We will adjust this problem using Platt's scaling and Isotonic Regression.

2.2.6 ***Evaluation:*** We plan on using F-score and area under the curve of ROC (receiver operator characteristics) plot as the evaluation metrics. Confusion matrix will be computed to better understand the number of false positives and true negatives. We will also investigate feature selection, model selection, K-fold cross-validation techniques to generalize our results.

3 ADDITIONAL METHODS

We also plan to dive into some additional machine learning methods if we have enough time by the end of quarter.

3.0.1 ***Unsupervised Techniques:*** Clustering algorithms discern hidden and latent structures in the data. Our aim is to leverage their power in predicting the genre. We will test a few algorithms (k-Means, Hierarchical Density Based Clustering) and confront their results to the true genres labels. To do so, we envision to make use of:

- Powerful visualization tools such as t-SNE[6]
- Dimensionality reduction techniques designed specifically for time series (such as a transformed version of Principal Component Analysis).
- Compare the output of the clustering algorithms to the true labels with the same metrics discussed in the last paragraph. We will also add the newly obtained observations as features to see if it helps improve the baseline.

3.0.2 ***Deep Learning:*** Inspired from the impact of deep learning in computer vision, we also plan to use very popular deep learning models like VGG [9] in our work on music genre recognition.

REFERENCES

[1] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu. 2014. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing* 22, 10 (2014), 1533–1545.

[2] Michaël Defferrard, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson. 2017. FMA: A Dataset for Music Analysis. In *18th International Society for Music Information Retrieval Conference.* https://arxiv.org/abs/1612.01840

[3] Jort F. Gemmeke, Daniel P. W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio Set: An ontology and human-labeled dataset for audio events. In *Proc. IEEE ICASSP 2017.* New Orleans, LA.

[4] David D. Cox James Bergstra, Dan Yamins. 2013. Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms. *PROC. OF THE 12th PYTHON IN SCIENCE CONF. (SCIPY)* (2013), 2825–2830.

[5] Igor Karpov and Devika Subramanian. 2002. Hidden Markov classification for musical genres. *Course Project* (2002).

[6] Geoffrey Hinton Laurens van der Maaten. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9 (2008), 2579–2605.

[7] C. Laurier and Perfecto Herrera. 2007. Audio music mood classification using support vector machine. In *International Society for Music Information Research Conference (ISMIR).* files/publications/b6c067-ISMIR-MIREX-2007-Laurier-Herrera.pdf

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[9] K. Simonyan and A. Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* abs/1409.1556 (2014).

[10] Lonce Wyse. 2017. Audio spectrogram representations for processing with convolutional neural networks. *arXiv preprint arXiv:1706.09559* (2017).