Multi-class classification of music genres*

Bartosz Karpiński[1] and Olga Wawryk[1]

Wrocław University of Technology, Wrocław, Poland pwr.edu.pl

Abstract. This research shows the summary of work that we have done in relation to the classification of music genres on the music data gathered by us.

Keywords: Classification · Music data

1 Introduction

In recent years due to the popularity of music streaming services the music business has changed dramatically. Users oftentimes want to be able to catalogue or simply categorize and organize their music collection. Music genres are one of the best ways to do so. Additionally, the ability to easily recommend users new music which they would find interesting has become one of the most important problems for the music providing services. Also here, music genres are one of the best ways to classify music, as they look at the music in a much broader perspective than simpler parameters, and provide much more insight about the piece of music in relation to other pieces. Thus the tools that would ease and help automate such work could prove very useful. A multi-class classifier trained on a broad-spectrum data can be very effective to manage such task. The more objective results of one will be able to detect more nuances and as a result be able to describe the data in the more precise and meaningful way. Most of the works in this field are based on just few datasets, that can be found in most of the studies. Our goal is to train and validate the classifiers on the new dataset, completed from scratch from the real-world data, that is sure a good representation of the current trends and in general the state of the music industry.

2 Related works

The music classification topic has been brought to the surface in many research articles, discussing the various approaches and methodologies that can be applied to the problem. The first work [1] did a case study of the performance of different classification approaches, varying from the very basic ones like kNN to more advanced ones like Convolutional Neural Networks. This research gave interesting insight about the solutions, as depending on the length of the sample (as the two discussed were 3 and 30 second samples) the results varied drastically. In the shorter samples kNN was definitely dominant, whereas in case of

* Supported by PWr.

30 second samples its result were rather poor. Another article [2] discussed another approach to analyzing the sound data, by means that are oftentimes more known to the researches, so by analyzing spectrograms. Those are the projections of sounds onto an image, thus providing the ability to use Neural Networks designed for pattern-recognition in images to be used for the music classification tasks. The next work [3] dives deeper into the usage and usefulness of simpler Machine Learning algorithms such as SVM or Logistic Regression to fulfill this task. Similar idea is presented in the other work [4], but here the authors extract the information through the use of Mel Frequency Cepstral Coefficients. The use of representation learning in pair with Convolutional Networks is brought up and tested in the next article [5]. Besides using the spectrograms to perform the classification, the authors also performed experiments on the fusion of both learned features and handcrafted features, to assess the complementarity of those two approaches. The use of Convolutional Recurrent Neural Networks[6] shows, that the usage of CNN and RNN hybrid brings a very strong performance.

3 Experimental Evaluation

3.1 Dataset

The dataset is created from the popular music from years 1960 to 2023. It consists of 6 classes, and the samples were chosen so they belong to one class only in a broader classification. The samples were taken at both 5 second length, and 15 second length from the same songs. The samples were picked from the songs that were at least one minute long, at the 40 second mark and 30 second mark for 5 second and 15 second samples, respectively. That gives a total of 2670 samples, but together with the spectrograms, it theoretically gives 5340 samples. For each file, there are 15 distinct features.

Table 1. Dataset classes and samples

Class	N. of samples
Classical	438
Electronic	443
Hip-Hop	477
Metal	364
Pop	504
Rock	444

3.2 Research questions

The main research questions we want to answer are: - Does the length of the snippet change significantly the results of the classification? - How does the classification performed on the pure music files compare to image classification done on the spectrograms?

3.3 Experiment scenarios

First, we create the tabular data model.

- For each dataset (5s and 15s) we perform:
- We split the data using Repeated Stratified K Fold for 2 splits and 5 repeats
- We create a Sequential model
- For the model layers, we have a Dense layer of 32 neurons with Rectified Linear Unit activation, followed by Batch Normalization, and followed by a Dropout of 30%.
- Then a same Dense layer followed by Batch Normalization and Dropout, this time of 50%.
- Then again same Dense layer, Batch Normalization and Dropout of 70%. - Lastly a Dense layer of 6 neurons with Softmax activation.
- For the model compilation, we use 'adam' for optimizer, Sparse Categorical Crossentropy for loss function, and for metrics we specify accuracy.
- Model is then fit on 10000 epochs, and batches of size 128
- The evaluation is then performed, on batches of size also 128
- Lastly we make the prediction on the test data, and using those results, we compare it against actual labels, calculating: accuracy, f1 score, balanced accuracy score, precision score and recall score

Secondly, for the Spectrogram model:

- For each dataset (5s and 15s) we perform:
- We split the data using Repeated Stratified K Fold for 2 splits and 5 repeats
- We create the generator for the data using flow_from_dataframe, with batch-size of 128 and Categorical class mode
- We create a Sequential model
- First, Conv2D layer of 32 neurons, kernel of size (3, 3) and Relu activation
- Then MaxPooling2D layer with pool size of (2, 2)
- Conv2D with the neurons count changed to 32
- Another MaxPooling2D layer
- Conv2D with 128 neurons
- MaxPooling2D layer
- Flatten layer
- Dense layer of 128 neurons and Relu activation
- Dropout of 50%
- Dense layer of 6 neurons and softmax activation
- Then we compile the model with Categorical Crossentropy loss and adam optimizer
- We fit the model for 200 epochs
- The evaluation is then performed, on batches of size also 128
- Lastly we make the prediction on the test data, and using those results, we compare it against actual labels, calculating: accuracy, f1 score, balanced accuracy score, precision score and recall score

3.4 Experiment evaluation

For the evaluation the comparison of achieved scores: accuracy, f1 score, balanced accuracy score, precision score and recall score are compared. The average, min, max, median for each dataset: so each length and both tabular and spectrogram.

3.5 Experimental Environment

The experiment will be done on a Linux machine based on Arch, namely EndeavourOS. The programming environment is going to be Python, basing on Tensorflow Keras library. For the hardware specification, a 6-core 12-threads modern processor along the CUDA-supporting graphic card will be used.

References

1. Ndou, Ndiatenda, Ritesh Ajoodha, and Ashwini Jadhav. "Music genre classification: A review of deep-learning and traditional machine-learning approaches." 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS). IEEE, 2021.
2. N. M R and S. Mohan B S, "Music Genre Classification using Spectrograms," 2020 International Conference on Power, Instrumentation, Control and Computing (PICC), Thrissur, India, 2020, pp. 1-5, https://doi.org/10.1109/PICC51425.2020.9362364
3. Bahuleyan, Hareesh. "Music genre classification using machine learning techniques." arXiv preprint arXiv:1804.01149 (2018).
4. Ali, Muhammad Asim, and Zain Ahmed Siddiqui. "Automatic music genres classification using machine learning." International Journal of Advanced Computer Science and Applications 8.8 (2017).
5. Costa, Y. M. G., Oliveira, L. S., Silla, C. N. (2017). An evaluation of Convolutional Neural Networks for music classification using spectrograms. Applied Soft Computing, 52, 28–38. https://doi.org/10.1016/j.asoc.2016.12.024
6. Choi, K., Fazekas, G., Sandler, M., Cho, K. (2017). Convolutional recurrent neural networks for music classification. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). https://doi.org/10.1109/icassp.2017.7952585