

Beatsort: Classifying Music Subgenres Using a Neural Network

Ryan Cafarelli^[800397899], Sam Schrader^[800544026], Rohith Kumar Sura^[800691268], and Daniel Tucek^[800621142]

Southern Illinois University Edwardsville, 1 Hairpin Dr, Edwardsville, IL 62026, USA

1 Summary

1.1 Background and Motivation

Genre classification is a problem with substantially more research than subgenre classification when it comes to music. The problem of automatically identifying the subgenre of a given piece of music can be a lot more difficult to get high accuracy. There are several research papers that can get over 90% accuracy [1] when identifying music genres; however, on studies of subgenre classification, such as [2], the accuracy is reduced to 75% with ten or fewer subgenres. The accuracy gets lower the more subgenres that are added.

One of the applications of solving this problem could be to suggest what subgenres musicians should tag to their songs when they upload them. The current manual classification process introduces a lot of subjectivity which can produce inconsistent labels. A deep learning approach could provide useful insights for creating more accurate taxonomies of music. Another application of this algorithm could be to suggest music based on user preference.

One article, [3], described their approach at classifying the subgenres of EDM. The original project used three software packages: [4] an automatic music genre classifier created by the author, [5] an audio-based music information retrieval system, and [6] an audio analysis tool. In the paper, Caparrini uses [5] and [6] to extract a total of 92 statistics from each track then uses various decision tree algorithms to categorize each track into the target subgenre, [4].

Our project will use two distinct neural network models to classify audio as belonging to specific subgenres of electronic music.

1.2 Method

Our team will approach this problem using multiple algorithms and comparing the usefulness of each. The first approach described in [7] uses a recurrent neural network (RNN) with long short-term memory (LSTM). LSTMs were intended to improve the performance of RNNs when dealing with information dependencies across arbitrary lengths of time. In this approach we will preprocess the raw audio files to generate mel-frequency cepstral coefficients (MFCC) then use Keras and Tensorflow to train the model.

The second approach is to use convolutional neural networks (CNN), as discussed in [8]. CNNs are a class of deep learning algorithm commonly applied to analyzing images. To use a visual algorithm on audio data, the audio recordings will be translated into mel spectrograms, a visual representation of the audio data based on the frequency domain.

1.3 Intended Experiments

The original dataset will be divided into the training set (80% of the files from each subgenre) and the testing set (the remaining 20% from each subgenre). The training set will be applied to each approach (LSTM and CNN) using the pre-defined subgenre classification. Once the training of each model has concluded the testing set will be passed into each model to determine which approach provides better results for this dataset. This may be expanded into a training/development/test split if deemed necessary.

Dataset The data for this project will come from two sources: a Kaggle repository that contains a pre-processed dataset of EDM track characteristics, [9], and the sample MP3 files for tracks described by the previous dataset which were scraped from the Beatport website. Each sample track file is between 30 and 120 seconds in length, with the vast majority being 120 seconds. The first dataset was provided as the input dataset for [3].

References

1. Nanni, L., Costa, Y. M., Aguiar, R. L., Silla Jr, C. N., & Brahnam, S. (2018). Ensemble of deep learning, visual and acoustic features for music genre classification. *Journal of New Music Research*, 47(4), 383-397.
2. Diakopoulos, D., Vallis, O., Hochenbaum, J., Murphy, J. W., & Kapur, A. (2009). 21st Century Electronica: MIR Techniques for Classification and Performance. In *ISMIR* (pp. 465-470).
3. Caparrini, A., Arroyo, J., Pérez-Molina, L., & Sánchez-Hernández, J. (2020). Automatic subgenre classification in an electronic dance music taxonomy. *Journal of New Music Research*, 49(3), 269-284.
4. Caparrini, Antonio. (2020). GitHub repository. <https://github.com/Caparrini/pyGenreClf>
5. Music Technology Group, University of Pompeu Fabra. <https://essentia.upf.edu/>
6. Giannakopoulos, Theodoros. (2021). GitHub repository. <https://github.com/tyiannak/pyAudioAnalysis>
7. Tibadiya, Prem. (2020). Medium article. <https://medium.com/@premtibadiya/music-genre-classification-using-rnn-lstm-1c212ba21e06>
8. Roberts, Leland. (2020). Medium article. <https://towardsdatascience.com/musical-genre-classification-with-convolutional-neural-networks-ff04f9601a74>
9. Caparrini, Antonio. (2016). Electronic Music Features - 201611 Beatport-Top100, <https://www.kaggle.com/caparrini/electronic-music-features-201611-beatporttop100>