Automatic Music Genre Classification

Best Banda
University of Witswatersrand)
Johannesburg, South Africa
1227484@students.wits

Ritesh Ajoodha
University of Witswatersrand)
Johannesburg, South Africa
Ritesh.Ajoodha@wits.ac.za

***Abstract*—Classification by music genre has traditionally been performed manually even though humans have been observed to only classify with less than 60% accuracy. This process is not only inefficient but also repetitive, tedious, and expensive. Automatic music genre recognition and classification is tricky as genres overlap and there is discovery and creation of new genres and subgenres every day.Automatic music genre classification has been attempted in the machine learning space using several algorithms which include SVM (Support Vector Machine), KNN (K nearest neighbors) Logistic Regression, Random Forest and many more This paper contributes by investigating the perfomance of a Convoluted Neural Network trained by different datasets i.e. GTZAN dataset and FMA dataset. Does having more entries mean that the model will be surperior? What other factors affect the accuracy of a model**

***Index Terms*—Music genre classification**

I. Introduction

Music can be divided into genres. But what makes genre? Is it specific measurable features hidden in the beat? if yes then with the right combination of musical features, we should be able to accurately classify songs into different genres. This could enable automatic classification of music moving us away from the current manual annotation.

This would allow us to significantly improve convenience and the effectiveness of music suggestion algorithms that are used on media players and audio players like those from spotify, Youtube music, Itunes and Tidal. The benefits of automatic music genre are not only at the entertainment level but also have an impact at the commercial level through capacity creation.

Companies could benefit from automatic annotation as it would save time needed by employees to manually classify individual songs that are on these digital platforms.The aim of this paper is to explore how to approach audio processing and classification of music by genre using content-based features and to highlight the gaps, advances in MIR (Music Information Retrieval) and to further contribute to the pool of knowledge and development of a model that can accurately predict and classify music genre.

II. Related Work

As suggested by [Panagakis et al., 2008] there is an increasing need to automatically classify music genres. This is a result of the rapid rate at which music is becoming digitized and available online and on digital media platforms [Lau and Ajoodha, 2020]. Achieving automatic genre classification would lead to breakthroughs not only in automatic genre classification but also in other analysis problems related to content-based signal analysis in music [Tzanetakis and Cook, 2002]

The preferred models in Machine Learning with respect to music genre classification are classifiers such as Convoluted neural network (CNN), Random-Forest(RF), K-nearest neighbor (KNN), Support Vector Machine (SVM), Naïve Bayes and Multilayer Perceptron(MLP) [Ajoodha et al., 2015] [Lau and Ajoodha, 2020]. There have been cases of applying Sparse representation techniques, but the accuracy is below accepted standards [Sturm and Noorzad, 2012]. Genre Classification has also been attempted using statistical pattern recognition classifiers, and the results were on par with those achieved by people
[Tzanetakis and Cook, 2002].

Most of the implementation and exploration of music genre classification using machine learning models has been achieved through using the GTZAN data set [Sturm, 2013] [Defferrard et al., 2016] which contains about a thousand songs and ten genres [Lau and Ajoodha, 2020] There has been little investigation into its accuracy and integrity. It remains the popular choice for researchers this is despite that there are known inconsistencies that ultimately affect the prediction accuracies of the machine learning models. These faults and inaccuracies have been analyzed and suggested by [Sturm, 2013] Previous studies have also highlighted the importance of extracting relevant, non-redundant content features that can be used to represent and describe components of music accurately, comprehensively, effectively and are compact [Nkambule and Ajoodha, 2020]. These features usually fall into four groups i.e., Tempo-Based, Pitch-Based, Magnitude-Based and Chordal progression features
[Ajoodha et al., 2015]. Extracting designing these features is one of the biggest hurdles that must be overcome to break free from the limitations of current machine learning models.

Music genre classification is a complex problem [Bergstra et al., 2006] and as such there have been a number of suggested features, many approaches and several metrics to evaluate the precision of the models being used. The metrics of evaluating the performance are the confusion matrix, 3 repeated 1o-fold validation, training time and accuracy. Accuracy being the most preferred.

III. Methodology

With the initial hypothesis being that the difference in performance for the convolutional neural network on the different dataset will be similar even though the Free Music Archive(FMA) dataset contains more training examples than the GTZAN dataset, because in both datasets the labelling has been done by humans who can only classify music with an accuracy of 60%.

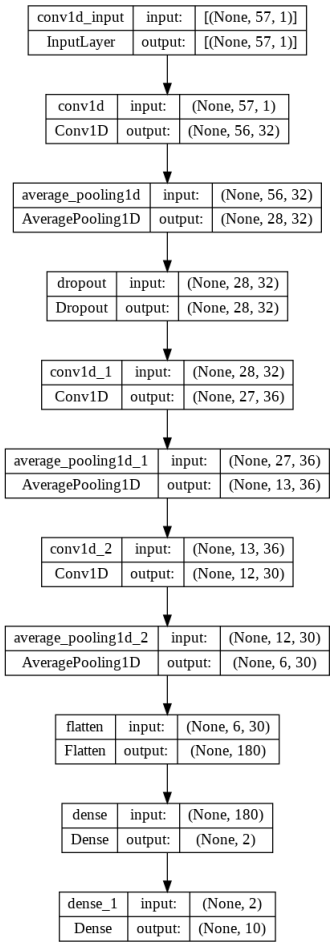

Fig. 1. Architecure of the Neural network

Furthermore there is more to music and genre than just what can be quantified into mean and standard deviationCite A convolutional neural network was created and it was made from the tensorflow library using the keras function. The network was made of a relu activation function, average pooling of 2x2, and a drop out regularization of 0.5, Each of the 3 convolutional layers had filters which were 32,36, and 30 respectively. Convoluted Neural Networks were chosen because they were highlighted by [Lau and Ajoodha, 2020] and [Nkambule and Ajoodha, 2020] to outperform other machine learning algorithms when it comes to automatic music genre classification. After all the layers have run, they are then flattened, and after flattening comes the dense layers. The dense layers are the final layers one with a relu activation function and another one with a sigmoid activation function which contains the probability of all 10 output labels. The CNN model was trained at 1000 epochs and a batch size of 20 and then the model was ran multiple times.

A. Data Sets

The GTZAN dataset appears to be the most widely used dataset in automatic music genre classification in machine learning environments, while the Free Music Archive(FMA) is a dataset that was created to overcome the problem of relatively small datasets when it comes to machine and music genre classification. It contains about 106 574 songs, 16341 artists and 161 genres in its entirety. It has been scaled down to different sizes, that is a small, a medium, a large and a full size dataset. The scope of this research was only limited to the small size FMA dataset, which contains about 8000 songs, in 30 second snippets and 8 genres. Which is similar to the GTZAN dataset which also contains 30 second snippets, but has 10 genres and 1000 songs

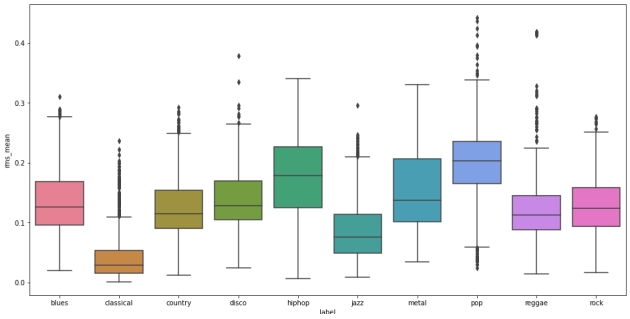

Fig. 2. Genres in the GTZAN dataset

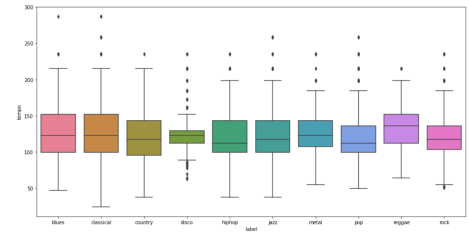

Fig. 3. BPM ranges for ten genres in GTZAN dataset

IV. Discussion and Results

Music is more often than not classified according to genre [Ajoodha et al., 2015]. Genre is however plagued by inconsistencies and ambiguities that arise from how it is defined [Nkambule and Ajoodha, 2020].These distinct definitions of genre support that there is more to genre than aspects/characteristics that can be extracted and quantified [Sturm and Noorzad, 2012].

Other factors that contribute to the definition of a particular genre include fashion, history, dance, culture, politics, marketing and also where and when the music the music originated [Tzanetakis and Cook, 2002]. This prompted [Panagakis et al., 2008] to suggest that MGR (Music Genre Classification) should be approached in a trans-disciplinary manner

Describing music by genre is not without opposition, as a result of differing opinions on what defines a certain music genre, it has been proposed that categorizing music by genre be disregarded [McKay and Fujinaga, 2006] in favor of similarity based algorithms.

Another major issue in classifying music by genre is the lack of a ground truth [Ajoodha et al., 2015] there is often a disparity in opinions with regards to how certain music pieces are to be classified. The existence of a ground truth can contribute positively to the accuracy of a classification but the lack of a ground truth can have a negative impact and thus limit how accurate classifiers are [Nkambule and Ajoodha, 2020] The lack of a large and free to use audio data set bottlenecks the curiosity and advancements of the MIR(Music Information Retrieval) community, as it prevents comparisons and benchmarking which are important aspects in the space of experimental sciences [Defferrard et al., 2016].

The most commonly used data set in audio classification is the GTZAN data set [Lau and Ajoodha, 2020]. Which has been criticized for being too small, not containing metadata and for multiple integrity problems including but not limited to replicas and repetitions [Defferrard et al., 2016].

Despite its flawed character [Sturm, 2013] highlights the important role the GTZAN data set has played in MIR and suggests that the data set should not be done away with but rather when it is employed, there should be a level of consideration for its faults and shortcomings as it has already contributed greatly to the MIR community.

[Defferrard et al., 2016] Argues in favor of giving preference to the FMA over the GTZAN as the FMA is a large data set containing 106 574 songs compared to the 1000 in the GTZAN and hypothesizes that the FMA is the solution that has been missing in MGR (Music Genre Recognition) because of its size, eclectic genres and that it contains full audio files of higher quality. [Sturm, 2013] in contrasts stresses that just because a data set is larger in size and more modern it does not guarantee that it is free from all imperfections that plague the GTZAN data set

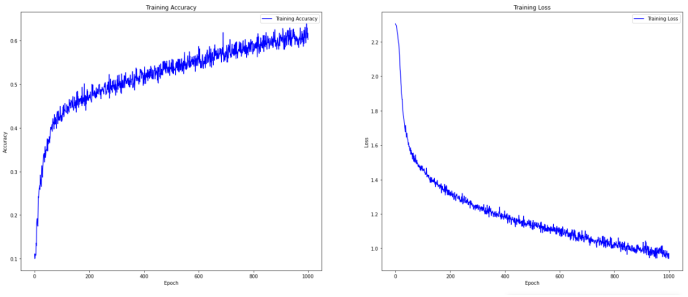

Fig. 4. Accuracy Plot for GTZAN dataset

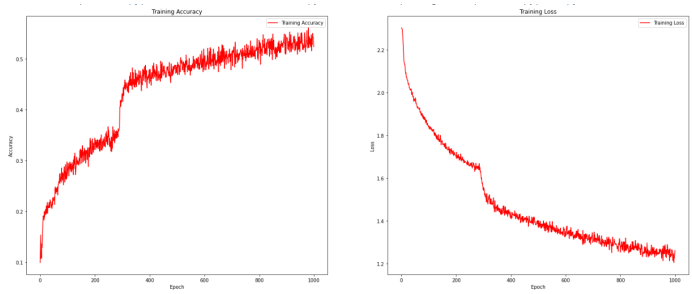

Fig. 5. Accuracy Plot for FMA dataset

A. *Results*

From Figure 4 and 5 we can see the accuracy plots after a 1000 epoch for each of the datasets. They each represent an accuracy of around 60%. The figues also highlight an accuracy loss, which is decreasing, and moves closer to zero

V. Conclusion and Recommendations

In this research the objective was to observe the difference in the performance of a Convoluted Neural Network(CNN) model that is trained on two different music genre classification dataset. According to the observed results the Neural Network performed similarly when trained on the FAM(Free Music Archive) dataset that contains 8000 entries and also on the GTZAN dataset which only has 1000 entries. This seems to suggest that we have not yet accurately identified features that can accurately describe music enough for it to be translated and taught to machines

The limitations that plague both datasets is that there has never been a well-founded ground-truth, in terms of what genre is. Even though both these datasets contain only a small subset of the genres that are in the real world, we still were not able to achieve an accuracy that is out of the ordinary. Which leads the author to believe that there is still some work needed in identifying the right features to use when trying to accurately classify genre and also coming up with the best architecture that can be employed to a CNN model in order to automatically classify music into genres, before

it is applied on a large scale and be part of our day to day life.

Seeing the role that music has played and continues to play in human life, developing fast, efficient, and accurate models to classify music is crucial to both music streaming companies and their customers. The bottlenecking in performance experience during this experiment is not a sign of bounds that cannot be overcome, but rather highlights the need for new perspectives on what makes up music, are there ways to digitize the cultural influences on genre? What makes a genre a genre? What is the ground truth on which definitions of genre are founded? It would be ideal to approach addressing this issues in ways that are not only technical but also ways that touch on different schools of thought i.e. the psychology, culture and social structures behind music genre might be the key to unlocking autonomous music genre classification

REFERENCES

[Ajoodha et al., 2015] Ajoodha, R., Klein, R., and Rosman, B. (2015). Single-labelled music genre classification using content-based features. In *2015 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*, pages 66–71. IEEE.

[Bergstra et al., 2006] Bergstra, J., Casagrande, N., Erhan, D., Eck, D., and Kégl, B. (2006). Aggregate features and adaboost for music classification. *Machine learning*, 65(2):473–484.

[Defferrard et al., 2016] Defferrard, M., Benzi, K., Vandergheynst, P., and Bresson, X. (2016). Fma: A dataset for music analysis. *arXiv preprint arXiv:1612.01840*.

[Lau and Ajoodha, 2020] Lau, D. S. and Ajoodha, R. (2020). Music genre classification: A comparative study between deep learning and traditional machine learning approaches. In *Proceedings of Sixth International Congress on Information and Communication Technology*, pages 239–247. Springer.

[McKay and Fujinaga, 2006] McKay, C. and Fujinaga, I. (2006). Musical genre classification: Is it worth pursuing and how can it be improved? In *ISMIR*, pages 101–106. Citeseer.

[Nkambule and Ajoodha, 2020] Nkambule, T. and Ajoodha, R. (2020). Classification of music by genre using probabilistic models and deep learning models. In *Proceedings of Sixth International Congress on Information and Communication Technology*, pages 185–193. Springer.

[Panagakis et al., 2008] Panagakis, I., Benetos, E., and Kotropoulos, C. (2008). Music genre classification: A multilinear approach. In *ISMIR*, pages 583–588.

[Sturm, 2013] Sturm, B. L. (2013). The gtzan dataset: Its contents, its faults, their effects on evaluation, and its future use. *arXiv preprint arXiv:1306.1461*.

[Sturm and Noorzad, 2012] Sturm, B. L. and Noorzad, P. (2012). On automatic music genre recognition by sparse representation classification using auditory temporal modulations. In *Computer music modeling and retrieval*, pages 379–394.

[Tzanetakis and Cook, 2002] Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302.