DATA-DRIVEN TECHNIQUES FOR MUSIC GENRE RECOGNITION

Sergio Santiago Rentería, Jesus Leopoldo Llano and
Francisco Javier Cantú-Ortiz

Tecnológico de Monterrey, México

ABSTRACT

After the digital revolution, it is not strange to see data science taking interest in music. The sheer amount of available content opens a plethora of possibilities for studying music and its social impact from a data analytic perspective. This paper studies the relationship that exists between, song features and their corresponding genre, to provide data-mining tools for music recommendation and sub-genre identification. For the first task, we compared different classification models, including Random Forests, Fully-connected neural networks and Logistic Regression. For the latter, we carried out cluster analysis and dimensionality reduction for data visualisation. Overall, Random Forest models had better performance in genre classification than Fully-connected networks, but they suffered from overfitting. Moreover, the highest accuracy obtained was too low (64%) to be of use for genre recognition applications. Nevertheless, we think our results show the limitations of hand-crafted features and point towards more sophisticated deep learning techniques.

KEYWORDS

Music Information Retrieval, Data Mining, Automated Music Recommendation, Classification

1. INTRODUCTION

One of the most important skills that humans possess is the capacity of pattern recognition. This skill has allowed the emergence of systematic studies of nature such as physics and drives innovation and development in many other fields ranging from arts to engineering. Out of all the patterns that surround us, music stands out as a universal activity with multiple forms and complex manifestations across cultures. There is increasing evidence that all humans, not just highly trained individuals, share a predisposition for music in the form of musicality, which is defined as a spontaneous developing set of traits based on and constrained by our cognitive abilities and their underlying biology [4].

On the other side, the emergence of digital technologies has dramatically shifted how music is produced, distributed and consumed. It has made possible data-driven studies on the impact of music on human behaviour and the many ways in which we enjoy it. Nowadays, more information about artists and songs is available, new music is capable of reaching more people further and faster than ever before, and recommendations are easily tailored automatically based on user behaviour.

The area of Music Information Retrieval (MIR) arises as a way to tackle the challenges associated with effectively interacting and accessing increasingly large collections of music and all the associated data such as styles, genres, artists, lyrics, and music reviews. Algorithms

David C. Wyld et al. (Eds): ICAITA, CDKP, SAI, NCO, CMC, SOFT, MLT, AdNLP - 2020
pp. 49-63, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.100905

developed for these purposes employ sophisticated digital signal processing and machine learning techniques to extract musical features from audio signals and metadata [7]. Other relevant applications of this field include audio identification, score following, digital musical instruments and audio compression algorithms [7].

When listening to a song most people will not identify all of their features, like the tonality or the metre, but the overarching patterns that appear in songs allow individuals to recognise musical genres. Moreover, there are certain patterns and structures in music that people particularly enjoy, meaning that the most popular (hit) songs are those with patterns that better resonate with the listener. However, we should have in mind that what makes us like a song depends also on extra-musical factors such as the exposure effect and the social context in which music is consumed. Hit song science attempts the herculean task of predicting music success by finding features correlated with music popularity [7].

Understanding music from a data-driven perspective has become vital for the music industry. Techniques for identifying the contrasting and common characteristics in musical genres, extracting the similarities between music that become popular in contrast to music that doesn't could be used to enhance music recommendation and music production. Within these lines, the present article is focused on the application of data mining techniques to music genre recognition and sub-genre identification.

Our main interest is to explore the applications of data science in music recommendation systems and to study how they can improve streaming services such as Spotify and Apple Music. We structure this paper using the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework [10] to provide an implementation of this methodology in the field of Music Information Retrieval.

2. METHODOLOGY

To study common and contrasting characteristics of different music genres, we employ a series of data mining algorithms to extract patterns from the GTZAN genre collection [11]. This is an audio database containing a total of 1000 audio recordings across ten musical genres with 100 instances of 30 seconds each [11]. Below we give an overview of data mining techniques that have been used in this problem.

Different approaches to song recommendation and genre classification have been tried before including traditional algorithms and deep neural networks. Aaron et al. trained a Convolutional Neural Network to predict latent factors from music audio [12]. Gwardys and Grzywczak used transfer learning to leverage features from a model trained on ImageNet dataset in a genre recognition model based on MFCC spectrograms. They obtained an accuracy of 78% in GTZAN database by using SVMs over the CNN representations [3].

More recently, Yang and Zhang used a map of eight musical features as inputs of a CNN, their best model obtained 91 % accuracy on the GTZAN database [15]. Banitalebi-Dehkordi proposed a music genre classification method using Fast Fourier Transform (FFT) to extract short-term features from segments of each spectrogram. They obtained 95.7% Accuracy over GTZAN database [2].

Besides these approaches, there are other techniques and generic data mining algorithms that can be adapted to tackle the problem of music genre recognition. The most common approach is the use of statistical classification methods. An example of this can be seen in a paper by Schindler and Rauber [17] where the authors studied the effects of using a series of features extracted using

the Echonest Analyser [18]. They evaluated this approach in the task of music genre classification with commonly used classifiers in the MIR field, such as KNN, SVM, Random Forest, Naive Bayes and J48 Decision Tree. The authors applied these methods to a set of 4 databases with different subsets of extracted features. They obtained at most 66.9% accuracy in the GTZAN music genre recognition task.

Another example of the application of data mining algorithms adapted and applied to MIR can be seen in a paper by Kotsifakos et al. They carry out genre classification by combining the k-Nearest Neighbours classifier with Subsequence Matching with Bounded Gaps and Tolerances (SMBGT). SMBGT is used to extract similarity features between pairs of songs that characterise them. While k-NN using the extracted features assigns a genre to each song based on the votes of its closest neighbours. While this was applied only to MIDI data, short segments of musical pieces could have been used with the SMBGT for feature extraction [19].

In this work, we do not cover the fundamentals of data mining algorithms, but we recommend the following references in case further reading is required. Goodfellow et al. [20] give an introduction to a broad range of topics in deep learning, covering mathematical and conceptual aspects of different artificial neural network models used in industry and academia. Rokach and Maimon [21] present an in-depth overview of the many techniques based on decision trees used for data mining including an analysis of the specifics of the Random Forest algorithm. Finally, Hastie et al. cover the fundamentals of computational techniques for statistical learning, inference and prediction [22].

2.1. Business Understanding

This step involves understanding business goals, limitations and expectations. The main objective is to cast the business problem as one or more data science problems. Framing it in terms of expected value can allow decomposing the problem into data mining tasks to which well-studied methods exist [9]. In the following paragraphs, we describe howdigital transformation has made critical the use of data-mining techniques in the music industry.

Digital technologies permeate every market operation, the music industry is not an exception. On the contrary, music businesses have quickly adapted to the changes of the digital revolution since the advent of MP3 and portable music players. But nowadays the challenge is greater, given the amount of music being released daily has sky-rocketed. A rough estimate suggests that tracks released daily in streaming platforms doubled from 2018 to 2019, while the weekly average time spent listening to music reported by IFPI is circa 18 hrs [5]. Being able to recognise genres or musical categories is an important functionality for any music business, especially for streaming platforms seeking to engage their users by providing customized musical experiences.

In practice, we would like to know what music our users prefer (even if they do not know it yet) and provide recommendations on that basis. Glenn McDonald, Spotify's leading data scientist mentioned in an interview that the company has used a complex algorithm capable of analysing and categorizing up to 60 million songs on a molecular level, and the micro-classifications now amount to 1,387 sub-genres in total [8]. By understanding the inherent composition of songs within a certain genre and their sub-genres, similarities can be identified in terms of features and then leveraged for genre-based recommendation systems. In this way, music streaming services provide a customised dashboard of new releases and artists in real-time.

The quality of music recommenders and music discovery technologies becomes a powerful differentiator when the music catalogue is similar in size across streaming platforms. Around a third of Spotify's listening time is spent listening to Spotify-curated playlists, while slightly more

than half of that amount goes on playlists personalised to each listener based on their listening history. Nevertheless, while these add-ons might be beneficial, they have raised ethical concerns around the role of streaming services as tastemakers by biasing the visibility of certain artists [6].

Overall, based on the behaviour and profiles of customers, music features and information that is generated during playlist creation, companies can find patterns hidden within historical data and transform music recommendation and playlist curating into machine learning tasks such as classification, clustering, link prediction and association rule mining.

2.2. Data Understanding

Rarely there is an exact match between the data format required by data mining models and available data. Moreover, historical data might have been collected for purposes unrelated to our problem as framed in the business understanding step, or even for no explicit purpose at all. All of these situations have an impact on the reliability of data [9]. In the case of music genre recognition, data is obtained through licensing and partnerships with recording labels and independent artists. Any kind of automatic audio analysis relies on feature extraction to estimate meaningful patterns that are later fed to predictive and pattern recognition models. Audio feature extraction is the process of distilling huge amounts of raw audio into compact and high-level representations about the underlying musical content. Common choices for audio representations reflecting the way humans and many other organisms make sense of their auditory environment are wavelets, filterbanks and Fourier transforms. Nevertheless, more abstract features capture high-level aspects of music recording such as rhythm, harmony and timbral texture [7].

It is important to mention that audio is not the only source of music data. Other sources include metadata, lyrics, review, social tags, user profiles and playlists, MIDI files and music scores [7].

Audio tracks from the GTZAN collection were collected between 2000-2001 from a variety of sources including personal CDs, radio, microphone recordings, to represent a variety of recording conditions. All tracks are in WAV format, have a sampling rate of 22.5 kHz and a bit depth of 16. The 10 genres included in the database are Blues, Classical, Country, Disco, Hip-hop, Jazz, Metal, Pop, Reggae, Rock.

Before the modelling stage, we carried out a series of exploratory analyses. Figure 2 shows the differences between the means of the spectral centroid feature across genres, while Figure 3 does the same for Chroma STFT. Since distributions overlap, we will need more than one feature to tell genres apart. Figure 4 shows the Spearman correlation between four of the features. Interestingly, some of the Mel frequency cepstral coefficients (MFCC) are inversely correlated with zero crossing. We decided not to show the Pearson correlation since the results were similar to Spearman. Finally, we carried out a t-test between genres in terms of the first MFCC. By using this feature, we can reliably tell apart Classical from the rest of the genres. ANOVA tests carried out for all the features of the data shows that the means of each group are significantly different from one another, Figure 6 presents the test carried out in terms of the ninth MFCC per genre. See the features list in the Data Preparation section for more information.

In addition to the aforementioned tests, we carried out a Shapiro-Wilk normality test per feature, where normality was the null hypothesis. All features but MFCC4, MFCC11, MFCC14, MFCC15, MFCC16, MFCC17 and MFCC18 had highly significant p-values. Finally, we computed a two-dimensional non-linear projection of the features using t-distributed stochastic neighbor embedding (TSNE) [14] (See Figure 1). The fact that genres are not perfectly segregated using a non-linear dimensionality reduction technique reflects the complexity of the

problem of genre classification. This gives us an idea of the predictive power of low-level features that ignore the contextual factors of music.

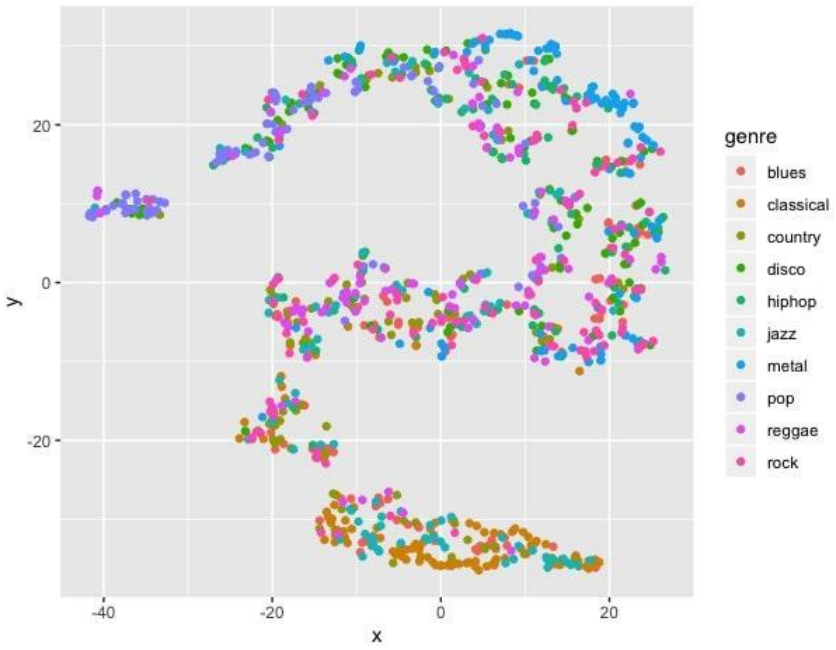

Figure 1. Dimensionality reduction using TSNE

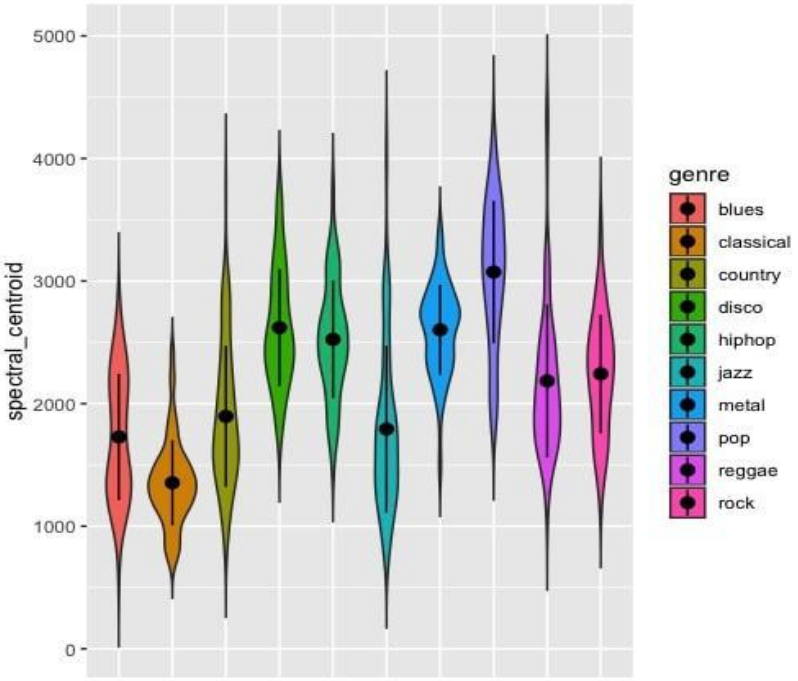

Figure 2. Spectral centroid distribution across genres

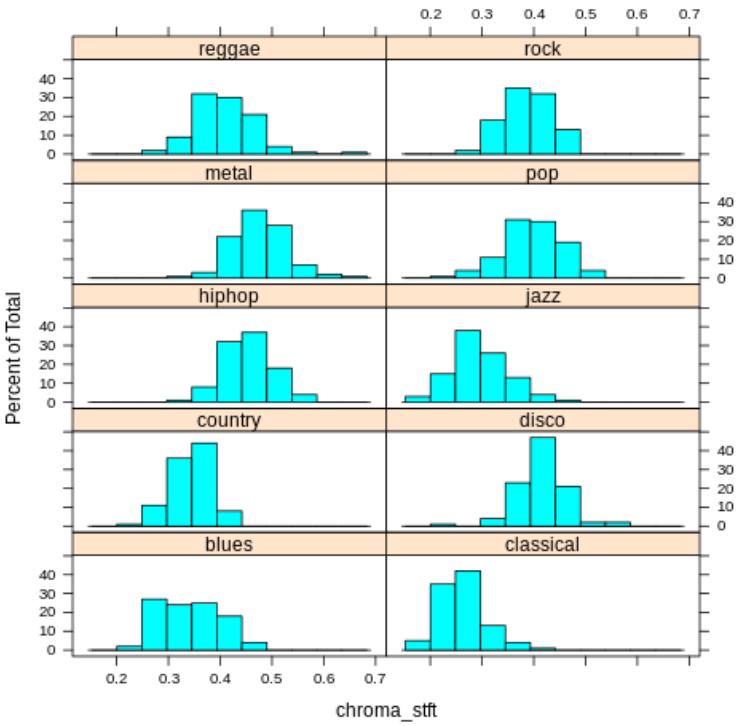

Figure 3. Chroma STFT comparison

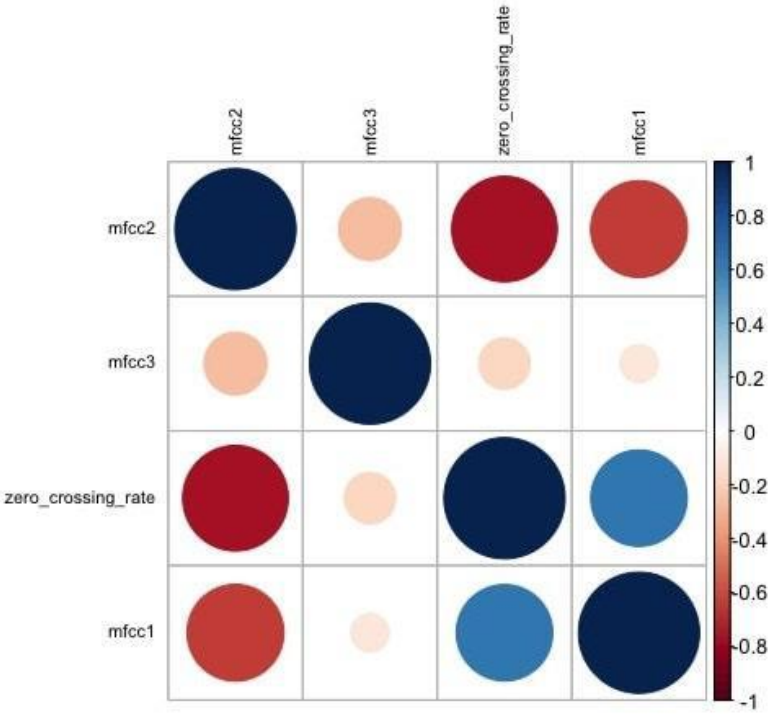

Figure 4. Spearman correlation of 4 features

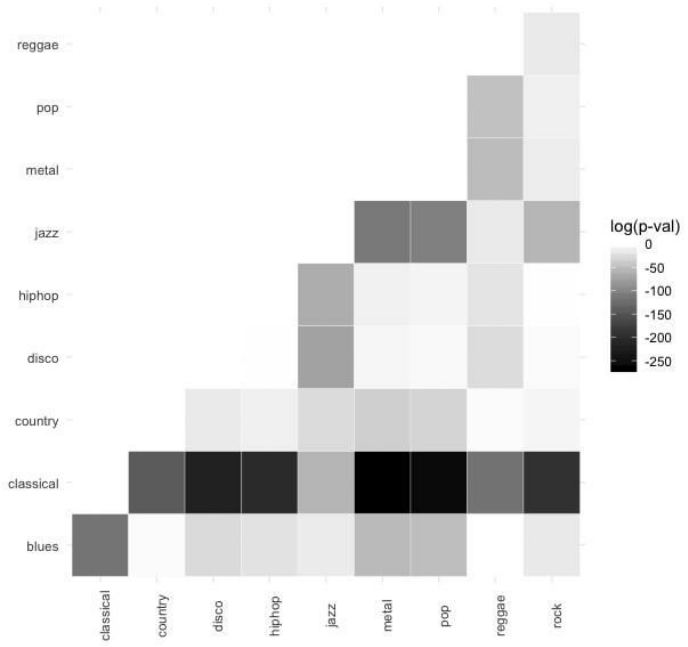

Figure 5. MFCC-1 t-test log(p-value) matrix

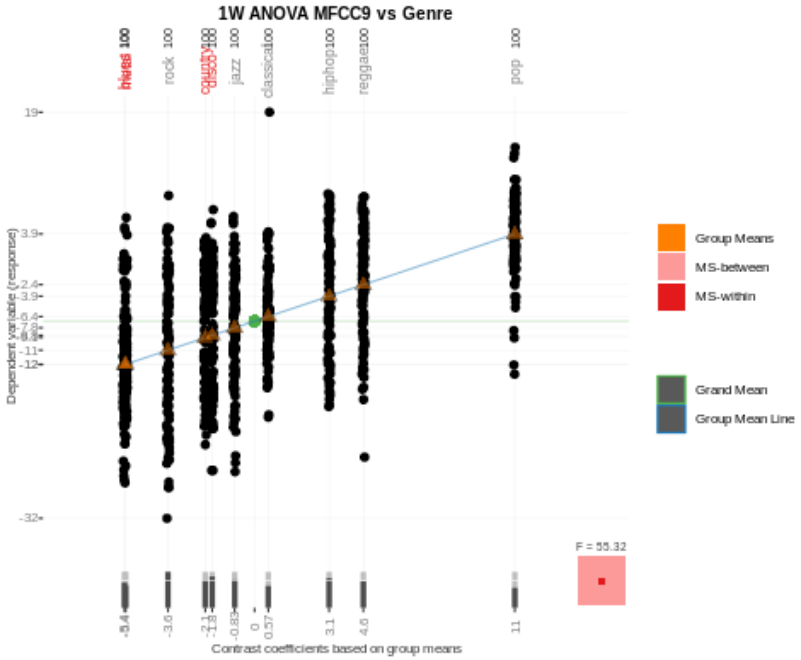

Figure 6. MFCC9 one-way ANOVA test

2.3. Data Preparation

We have to consider the nature of data before preparing it. For instance, some attributes may be categorical, whereas others text-based; All of these require different treatment. Overall, we have to make sure data is clean, free from leakages, and in the right format to carry out model fitting.

The 26 features used in this study were directly extracted from the audio tracks in GTZAN database using Librosa, a python package for music and audio analysis. A list with a short description is shown in the table below:

Table 1. Feature description

Feature	Description
Zero crossing rate	The rate of sign-changes along with a signal, i.e., the rate at which the signal changes from positive to negative or vice versa.
Spectral centroid	The "centre of mass" of the spectrum. Calculated as the weighted mean of the frequencies present in the sound
Spectral rolloff	Represents the frequency below which a specified percentage of the total spectral energy lies.
Root Mean Square Energy (RMSE)	Total magnitude of the signal.
Mel frequency cepstral coefficients (MFCCs)	A set of 20 features or coefficients describing the overall shape of the spectral envelope.
Chroma frequencies	A representation for music audio in which the entire spectrum is projected onto 12 bins accounting for the 12 distinct semitones (or chroma) of the musical octave.
2nd order Spectral Bandwidth	A weighted standard deviation from the spectral centroid

These 26 features were arranged in a table and appended to a .csv along with their corresponding filename and genre label.

Additionally, we used Wavenet as a feature extractor. These features are not human-readable but leverage the power of a convolutional neural network (CNN) trained directly on audio waveforms [13]. Since the shape of the feature map for each song was (125, 16) we also tested different aggregation methods including, flatten, column averages and row averages. These methods lead, respectively, to the following feature vector sizes: 2000, 125 and 16.

2.4. Modelling

We are concerned with the following data-mining tasks. These are related to recommendation operations in music streaming services. This paper covers model evaluation for the first two tasks.

- Classification: carry out class probability estimation for each song in our database to recommend more songs of users preferred genres.
- Clustering: Group songs by similarity to create sub-genres and fine-grained categories within and between genres.
- Link Prediction: attempt to predict connections between genres likeability. For instance, estimating how likely it is that someone who likes rock will like metal.
- Association rule mining: As part of the pattern extraction of songs, this task would help to understand how different types of genres and particular songs are grouped altogether by users and playlists.

For the first phase of the modelling step, we implemented two linear models: (I) a multinomial logistic regression to estimate genre (class) probability from the set of 20 Melfrequency cepstral coefficients (MFFCs), and (II) a generalized linear model to individually predict the respective values of Zero crossing rate, spectral centroid, spectral rolloff and RMSE from the nominal genre variable.

The multinomial logistic regression coefficients shown in the attached file are far from zero. This confirms our intuition that there was at least a linear relationship between predictors (MFCCs) and the probability of being of a certain genre. Differences across MFCCs weights reflect how the coefficients account for local variations in the spectral envelope that are correlated with the acoustic properties of genres.

With the linear models, we seek to test the hypothesis that the genres are meaningfully related to audio features, in particular those that describe musical and acoustic characteristics of the waveform such as the chroma frequencies, the centre of mass of the spectrum, etc. By generating a set of contrast variables using the genres and then fitting a set of linear models to predict each of the variables that do not fit under the umbrella of MFCCs, we corroborated this assumption. Moreover, the t-test showed significant differences across genres in terms of these covariates. To corroborate the generalised linear hypothesis of each model, a set of pairwise contrast variables was created using all the possible combinations of genres. These were tested using the Tukey method, again results show statistically significant differences across genres. Proving, further, the relation between features and genres.

Table 2. The architecture of the Fully Connected Network used for classifying the dataset.

Layer (type)	Output shape	Param #
Flatten	(None, 26)	0
Dense 1	(None, 32)	864
Dense 2	(None, 32)	1056
Dense 3	(None, 32)	1056
Dense 4	(None, 32)	1056
Dense 5	(None, 32)	1056
Dense 6	(None, 32)	1056
Dense 7	(None, 10)	330
Total trainable params	-	6,474

For the second phase, we explored the usage of more complex models for classifying the different music genres: A Fully Connected Network (FCN) described in Table2, and a family of Random Forest models. We compared their performance using thirty randomly drawn 70-30% training-testing splits. As a means of comparison, since the data is non-linearly separable, we fitted two logistic regression models using the same splits. Below we describe the implementation details of the models.

1) Fully-connected neural network (FCN): Implemented in Keras framework. The architecture consists of 6 layers with 32 units each. It was trained with RMSProp optimizer (default parameters) for 100 epochs and using a batch size of 32. See Figure 7. FCN-WF, FCN-WC and FCN-WR refer to the results of FCN using flattened, column-averaged and row-averaged Wavenet features, respectively.
2) Random Forest 1 (RF1): Implemented using sklearn ensemble classifiers. The model consists of 10,000 estimators trained over the dataset using Gini's impurity as splitting criteria.
3) Random Forest 2 (RF2): Preserves the same parameters as the previous model, only changing the splitting criterion from Gini's impurity to Information gain.
4) Random Forest 3 (RF3): Based on RF1 this model incorporated Minimal Cost-Complexity Pruning (CCP) to reduce overfitting. Only estimators with a CCP < 0.15 were selected to be part of the model.
5) Random Forests with Wavenet features: Following the same models as RF1 and RF3. Each pair of RF-WR, RF-WC and RF-WF show the results of using row-averaged, column-averaged and flattened Wavenet features, respectively.
6) Linear Regression model (LIN1): A regularised logistic regression model based on a quasi-Newton method (LBFGS) and Ridge Regression (L2) as implemented in sklearn.
7) Linear Regression model (LIN2): Similar to LIN1 but using an improved version of the Stochastic Average Gradient and Elastic net regularisation.

To test whether accuracy was significantly different across models, we carried out a Wilcoxon signed-rank paired test between the accuracy series. These were generated by evaluating the performance of trained models on 30 Test sets generated with the same random seeds (i.e. paired test). Figure 8 reports the p-values and Table 3 the performance results for each model.

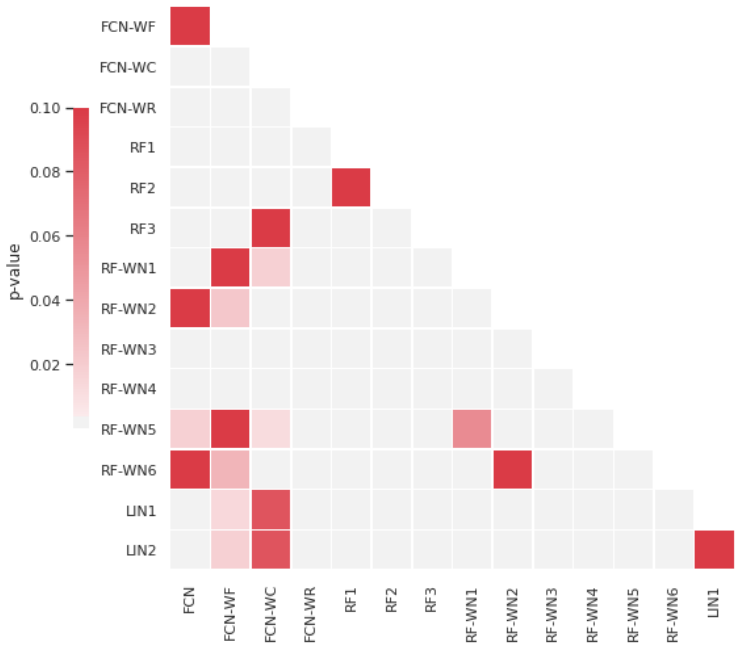

Figure 8. Wilcoxon signed-rank paired test for Test set

2.5. Evaluation

The purpose of the evaluation stage is to assess the data mining results and gain confidence that they are valid and reliable. These aspects include assessing models stability across time, ensuring that models satisfy the original business goals and spotting spurious correlations [9].

Even if we know from our musical experience there are systematic relationships across genres, the residual deviance of multinomial logistic regression is relatively high (2367.332), which means a linear model is a bad fit. Nevertheless, Non-parametric models such as Random Forest performed better than linear models and the fully-connected network (FCN). See Table 1.

By the other side, we believe some of the standard errors for multinomial logistic regression parameters are high, given that the sample size per genre is small and there is multicollinearity in our predictors, which means some MFCCs can be predicted from one another.

Table 3. Results: Mean accuracy for all models.

Model	Training %	Test %	Delta %
FCN	44.07	38.76	5.32
FCN-WF	45.46	40.66	4.81
FCN-WC	64.38	50.43	13.95
FCN-WR	72.82	56.34	16.48
RF1	99.93	64.48	35.45
RF2	99.93	64.19	35.74
RF3	61.18	49.38	11.80
RF-WR1	99.91	42.40	57.51
RF-WR2	48.74	37.49	11.25
RF-WC1	99.91	32.56	67.35
RF-WC2	47.94	24.84	23.10
RF-WF1	99.91	41.66	58.25
RF-WF2	80.15	37.68	42.47
LIN1	48.65	45.03	3.61
LIN2	48.81	44.71	4.10

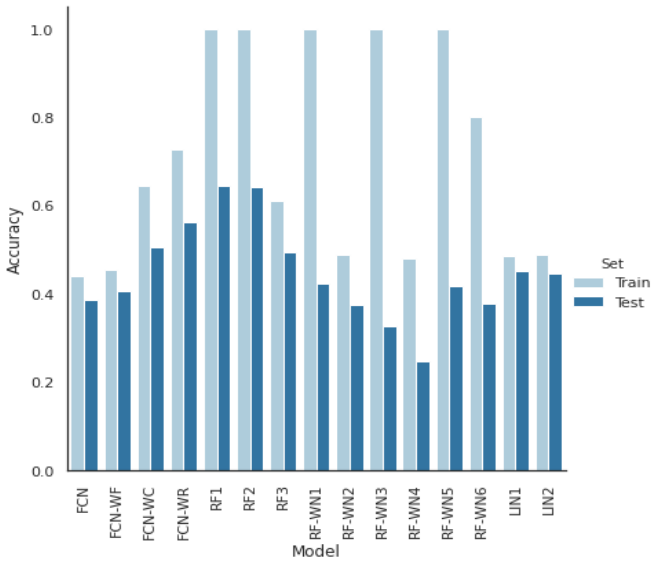

Figure 9. Mean accuracy comparison for Training and Test sets.

3. DISCUSSION

The main goal of this project is to identify characteristics within music genres to gain a better understanding of what patterns characterise them and which are shared amongst classes (i.e. genres). With this knowledge, we expect to be able to identify new songs that belong to a specific genre according to acoustic features and also to find sub-genres within them by using clustering techniques.

We noticed that some MFCCs are correlated among themselves but inversely correlated with other frequency related features such as zero crossings. The differences between genres in terms of MFCC1 were enough to tell apart Classical from the rest of the genres. During the data visualisation phase, we found that Pearson and Spearman's correlations gave similar results. For this reason, we decided to only include the latter.

Overall, the performance of the linear models, especially for class probability estimation, is low. Non-parametric models such as Random Forest performed better than linear models but overfit the training dataset. No significant differences were found between RF1 and RF2, which means for this task there is no noticeable advantage between Gini's impurity and information gain criteria. Despite the fully-connected network leveraging non-linear activation functions (ReLu), its accuracy was below the best Random Forest (RF1) and linear models. Interestingly, as opposed to Random Forests, the FCN accuracy improved using Wavenet features instead of handcrafted ones (i.e. MFCC, zero-crossings, etc.).

We also observed that the aggregation method (flatten, column average, row average) influenced the performance with Wavenet features for both Fully-connected neural networks and Random Forests. This might be explained by the fact using fewer features, as obtained with the column average method, reduces overfitting.

We confirmed that more expressive models than logistic regression, such as random forest, performed better at the classification task, but still are far from the required performance expected for music recommendation tasks (maximum average Test accuracy was 64%). Nevertheless, it was a good exercise to evaluate models with different assumptions: linear, non-parametric and hierarchical non-linear (i.e. artificial neural networks) with handcrafted and automatically extracted features (i.e. Wavenet). In this way, we compared how much improvement is coming from modelling non-linear interactions of covariates and learning structure from data without distributional assumptions.

4. CONCLUSION

Streaming services have disrupted the music industry, from how it is distributed to the way music is produced. Data mining techniques underlying this transformation have allowed companies to extract features with increasing precision and to use them for various purposes such as music recommendation, classification, trend prediction, sub-genre discovery and to design tailored user experiences for streaming apps.

We have demonstrated that there are significant differences in the feature values of each genre, seemingly enough to characterise each of them. The different tested linear models, leaving aside their poor performance, show that there is a strong relationship between the response (genre) and the proposed predictive variables. Furthermore, we found that non-parametric models, such as Random Forest with MFCC features, performed better than linear and Fully-connected neural

network models using Wavenet features. Our best model was a Random Forest (RF1) and achieved around 64% accuracy in the test set using handcrafted features (MFCC).

Despite not having met state-of-the-art results, the main contribution of this work is to compare a wide range of modern data mining techniques to study the dimensions of music, particularly its genre and the differences that exist amongst them. Moreover, to our best knowledge, we are the first to evaluate the effect of Wavenet features in the context of music genre recognition, particularly in the GTZAN database. The stark contrast between our models and state of the art approaches speaks about the relevance of finding good representations of musical data.

4.1. Future Work

Potential avenues for extending this work are: (1) exploring unsupervised clustering models, (2) multimodal models incorporating extra-musical and non-audio data, (3) transfer learning, data augmentation, and (4) extending GTZAN with new labelled instances (i.e. use a larger database). Clustering might find sub-structures within genres and help to determine useful patterns for music recommendation. Multimodal models [1] can leverage extra-musical information while transfer learning might be capable of finding better audio representations than MFCCs. Finally, we believe model training and testing results can be improved by applying other data mining and statistical techniques including but not restricted to LSTM neural networks, Dynamic Time Warping, K-means and Siamese Neural Networks.

REFERENCES

[1] TadasBaltrusaitis, Chaitanya Ahuja, and Louis Philippe Morency. Multimodal Machine Learning: A Survey and Taxonomy, 2019.
[2] Mehdi Banitalebi-Dehkordi and Amin Banitalebi-Dehkordi. Music genre classification using spectral analysis and sparse representation of the signals. Journal of Signal Processing Systems, 2014.
[3] Grzegorz Gwardys and Daniel Grzywczak. Deep image features in music information retrieval. International Journal of Electronics and Telecommunications, 2014.
[4] H. Honing, W.T. Fitch, B. Merker, I. Morley, W. Zuidema, L. Trainor, A. Patel, S.E. Trehub, J. Becker, M. Hoeschele, et al. The Origins of Musicality. The MIT Press. MIT Press, 2018.
[5] IFPI. Ifpi releases music listening 2019, Sep 2019.
[6] Mansoor Iqbal. Spotify usage and revenue statistics (2019), May 2019.
[7] Tao Li, MitsunoriOgihara, and George Tzanetakis. Music Data Mining. CRC Press, Inc., USA, 1st edition, 2011.
[8] Nick Patch. Meet the man classifying every genre of music on spotify - all 1,387 of them, Jan 2016.
[9] Foster Provost and Tom Fawcett. Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking, 2013.
[10] Colin Shearer. The crisp-dm model: The new blueprint for data mining. Journal of Data Warehousing, 5(4), 2000.
[11] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 2002.
[12] Aaron van den Oord, Sander Dieleman, and Benjamin Schrauwen. Deep content-based music recommendation. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 26, pages 2643–2651. Curran Associates, Inc., 2013.
[13] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, NalKalchbrenner, Andrew Senior, and KorayKavukcuoglu. Wavenet: A generative model for raw audio, 2016.
[14] Laurens Van Der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. Journal of Machine Learning Research, 2008.
[15] Hansi Yang and Wei Qiang Zhang. Music genre classification using duplicated convolutional layers in neural networks. In Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2019.

[16] Hansi Yang and Wei Qiang Zhang. Music genre classification using duplicated convolutional layers in neural networks. In Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2019.

[17] Alexander Schindler and Andreas Rauber. Capturing the temporal domain in echonest features for improved classification effectiveness. In Andreas Nürnberger, Sebastian Stober, Birger Larsen, and Marcin Detyniecki, editors, Adaptive Multi-media Retrieval: Semantics, Context, and Adaptation, pages 214–227, Cham, 2014.

[18] Tristan Jehan and David DesRoches. Analyzer documentation (analyzer version 3.08). Website, 2011. Available online at http://developer.echonest.com/docs/ v4/_static/AnalyzeDocumentation.pdf; visited on May 30th 2020.

[19] AlexiosKotsifakos, Evangelos E. Kotsifakos, Panagiotis Papapetrou, and VassilisAthitsos. Genre classification of symbolic music with smbgt. In Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments, 2013.

[20] Ian Goodfellow, YoshuaBengio, and Aaron Courville. 2016. Deep Learning. The MIT Press.

[21] LiorRokach and Oded Maimon. Data mining with decision trees. Theory and applications, volume 69. 01 2008.

[22] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Elements of Statistical Learning. Springer, New York, NY, USA, 2008.

AUTHORS

Santiago Renteria is a computer scientist and audio engineer working at the intersection of artificial intelligence and biology. As part of his masters, he developed a "Shazam" for birdsong based on siamese neural networks, a few-shot machine learning technique capable of recognizing birds' complex melodic sequences. Beyond machine learning one of his main interests is to develop and understand non-human forms of intelligence through artistic experimentation. In his practice, he plays with different media such as virtual reality, immersive audio and biosignal sensors. Currently, he is part of WATS, an interdisciplinary initiative generating projects at the interface of art, technology, and science. Furthermore, his work as a creative developer has been showcased at Laboratorio de Arte Alameda, Centro Cultural Universitario Tlatelolco, Carnaval de Bahidorá and Tecnológico de Monterrey.

Jesus Leopoldo Llano is a computer scientist whose work has mainly focused on the area of evolutionary computing and multi-objective optimisation. During his masters, he designed an evolutionary algorithm for the numerical treatment of equality constrained multi-objective optimization problems. His areas of professional interest focus mainly on the study and development of bio-inspired algorithms and computer systems as ways to solve high complexity problems. Currently, he is part of the machine learning research group of Tecnológico de Monterrey. His work has been showcased at the 50th Research and Innovation Congress of Tecnológico de Monterrey.

Francisco J. Cantu-Ortiz is Professor of Computer Science and Artificial Intelligence at Tecnológico de Monterrey. He is member of the Advisory Board for QS-World University Rankings and associate editor of various journals and conferences. His research interests include data science, AI analytics, science & technology management, and philosophy of science & religion. He has published more than 100 scientific documents and is a certified researcher by the National Council for Science and Technology, Mexico. He holds a PhD in Artificial Intelligence from the University of Edinburgh, UK, an MSc in Computer Science from North Dakota State University, USA, and a BSc in Computer Systems Engineering from Tecnológico de Monterrey (ITESM), México. http://semtech.mty.itesm.mx/fcantu/

© 2020 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.