Training and Optimizing Music Recommendation Algorithms Using Self-Similarity Matrices

Hope Mogale and Michael Esiefarienrhe

May 1, 2022

Training and Optimizing Music Recommendation Algorithms Using Self-Similarity Matrices

Abstract—**Most music recommendation algorithms such as proprietary ones designed for big music streaming platforms such as Spotify, Apple Music, Tidal, Deezer, and etc rely mainly on song ratings data to be able to recommend songs to users. This does not always provide a great music experience for many users who pay for these services. As an example there are over 35 million songs on Spotify alone and according to recent statistical report released by Spotify there are about 4 million songs on Spotify that have never been played not even once. Clearly this is not good news to content creators who trust and pay Spotify as a platform to distribute their music. The root of the problem lies on the music recommendation algorithm adopted and used by many music streaming platforms like Spotify. The algorithm fails to recommend great music to the users which they can enjoy and this results in the platform experiencing "dark music" or music that has never received any play in the platform. We propose a framework which provides a unique strategy that can be adopted by music recommendation algorithms to give users a better music experience. We utilize self-similarity matrices developed in R to visualize patterns of repetition in text extracted song lyrics. The way this works is that the lyrics of a song played by the user are extracted and a visual pattern print of a song is generated and this pattern is then compared with many other patterns of songs existing on the platforms especially ones that have not been played before. A comparison is then performed and if the comparison similarity index is above 70 percent then a song is recommended to the user. A song with high a similiarity index gets first priority. We believe this will ensure great music experiences for the listeners and also benefit content creators since the likelyhood that their music will reach users will be high.**

Keywords -*Self-similarity Matrices, Music Streaming, Music Recommendation Algorithm, And Similiarity Index*

I. Introduction

Music is something that we all love to listen to and enjoy everyday. For every different day many users love listening to different kinds of music genre ranging from House to Jazz, Blues, Hip-Hop, Pop, RnB and etc. Since there exists a different variety of genres it can become really difficult for users to decide which songs to play hence why many users stick to playing the same music over and over again. To overcome this problem music streaming platforms such as Spotify and Apple Music utilize music recommendation algorithms [1] such as one describe by researchers in [2][3][4]. Over the years this has worked well however, recently this has become a problem as the number of songs in music streaming platforms grew. Over 20 000 songs are added on daily basis and most of them don't get to be recommended to users unless they are coming from big artists that many users are already following. Most of the existing music recommendation algorithms are still using old methods such as analyzing the audio signal and using low level information such as song acoustics [5] [6] features to determine whether listeners will love it. Over the years this hasn't worked well and so most listeners rely on verbal recommendations from family and friends and also through what they are exposed to on social media platforms. This should not be the case and music streaming platforms should innovate and come with a innovative solution to make the music recommendation algorithms useful again. We propose a useful solution to overcome this. Our solution utilizes self-similarity matrices [7] [8] developed in R Studio to determine similarity in songs.

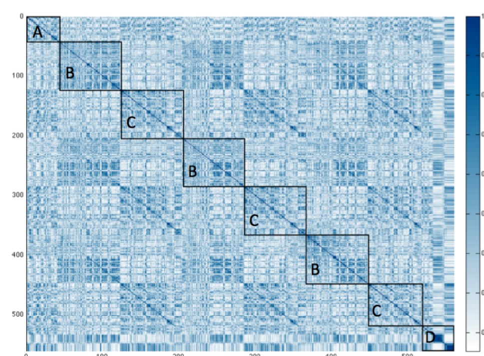

Figure 1. Visual Matrix for the Beatles - Help

As visible in figure 1 visual matrices can be thought of as the DNA print or visual print of song since they are constructed using rate of repetition in songs. Biologists use a similar technique of self similarity matrices to model and visualize DNA sequences. The Visual Matrice method utilizes individual words of the song which are used as the names of the columns and the names of the rows in a matrix. In every point of the song where the row name equals the column name a dot is mapped on the visual matrix and the the diagonal of every similarity matrix is then filled. The duration of the song can be envisioned along a diagonal from top left to bottom right. The patterns that appear to be away from the diagonal represent two different points in time with the same words.

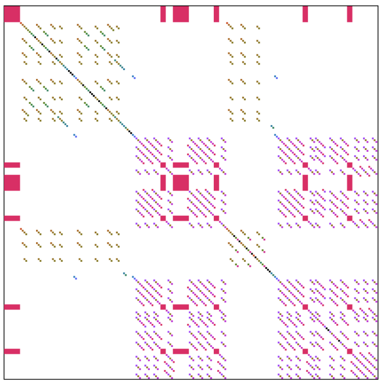

Figure 2. Visual Matrix for Beatles - All you need is love

In figure 2 above, we see a visual matrix of the song All you need. We see that the matrix in contrast to figure 1 Silento Watch me has a different visual matrix. This is both songs are totally different and our algorithm should never recommend the Beatles song to a user after the user has listened to Silento song. We believe that Visual Matrices will be the key to training and optimizing music recommendation algorithms in streaming applications. The rest of the paper is organized as follows the next section will outline related work to our study. Section three will detail self similarity matrices used as tool to optimize song recommendation algorithms. Lastly, Section four will provide conclusion to our work.

II. Related Work

We live in the digital age where everything has moved to be accessed online. Music is no exception hence the study of its digital nature is imperative. There has been substantial amount of studies published on implementing self similarity matrix on digital audio. We enlist the ones we believe are relevant to our study. Foote and Cooper [9] published one the earliest studies on visualizing self similarity techniques and have published a paper that studied the acoustic similarity between any two instants of an audio recording which is displayed in a static 2D representation, which makes structural and rhythmic characteristics visible. Rafii and Pardo alluded in [10] that, while the existing methods are effective in utilizing repetition as a fundamental element in generating and perceiving structure in music. These methods depend on an assumption of periodically repeating patterns hence, they proposed a method that generalizes the repetition based source separation approach to handle cases where repetitions also happen intermittently or without a fixed period, thus allowing the processing of music pieces with fast varying repeating structures and isolated repeating elements. With preliminary evaluation which reported encouraging results, Kaiser et. el conclude in [11] that over the dimensions of the NMF decomposition, structural parts can easily be modeled and based on this observation, they introduce a clustering algorithm that can explain the structure of the whole music piece. To tackle background noise which can at times become a problem, Peeters in [12] introduced the use of higher-order (2nd and 3rd order) similarity matrices in order to reinforce the diagonals corresponding to common repetitions and reduce background noise. We take this technique into consideration and where necessary we will use it in our study. Lastly, O'Dair et. al exclaimed in [13] the impact of recommendation systems upon artists. They further outlined that although there has been important work on algorithms in the context of music streaming, the focus on music streaming remains relatively unusual. With this paper we plan to help bridge the gap that exists and blackbox created by recommendation algorithms on music streaming services.

III. Self Similarity Matrix Method

The reason why we utilize visual matrices is because we believe that people love music because of repetition and this is attested by this study conducted by researchers in [14]. Also according to recent reports that were published by billboard songs with more repetition that is easy for people to understand are likely to be in the top ten. Hence, why this approach is the secret recipe of our approach where by we utilize visual matrices as seen in figure 1, 2 and 3 to determine the repetition rate in songs.

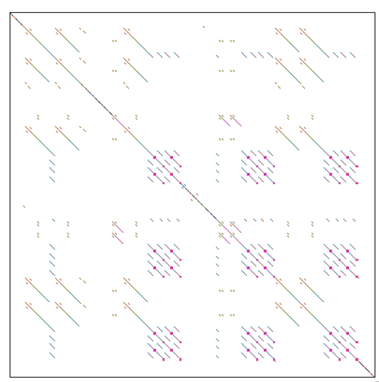

Figure 3. Visual Matrix for Aqua - Barbie

We take this further by generating a word frequency histogram from the song lyrics to extrapolate the rate of repetition. Once we have this information we go over all the new songs that have been added on the platform of the same genre and perform the same action on the new song. Once this is completed we go over and do over 100 different songs that fall in the same genre and category as the currently playing song. Once that is done we compute the similarity index by surveying the repetition rate of each song then we compare it to the original chosen song. We then follow an Axiom which states that the first five songs with the highest repetition rate obtained from the word frequency histogram gets the priority to be the one chosen to be recommended for the user. To counter the dark music problem we prioritize with new music first that has just been added to the platform that has never received any play so that users can be able to discover new music.

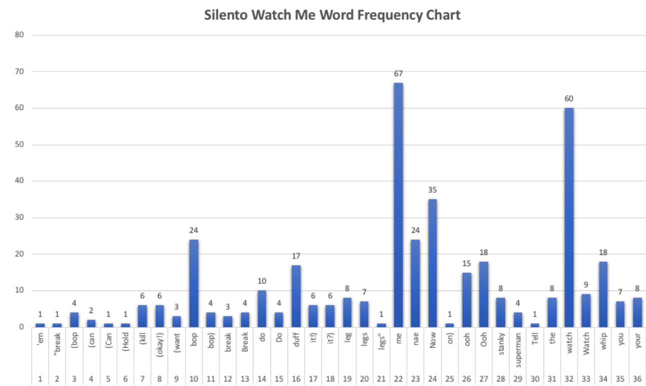

Figure 4. Watch Me Word Frequency Histogram

as seen in figure 4 above we have generated a word frequency histogram chart from the song lyrics of the song "Silento - Watch Me". We see from the histogram that the word "Watch" has the high repetition frequency rate of 60 and the word "me" has the high repetition frequency rate of 67. We also see that the word "hold" and "can" both have low repetition frequency rate of 1. This is all extrapolated from the songs lyrics.

Table1. Watch Me Five High Frequency Word List

Count	Word	Frequency
1	"me"	67
2	"Watch"	60
3	"Now"	35
4	"nae"	24
5	"whip"	18

As can be seen in Table 1 above we can see that the top five of the most popular songs from the Watch Me song by Silento. Hence from this word list we can compute the average from the given frequencies and use that as our similarity index to compare with other songs.

$$SNR(\overline{x}_0) = \frac{S_{t_0}(\overline{x}_0)}{\sum_{t \neq t_0} e^{-\beta |c_0 - c(t)|} S_t(\overline{x}_0)}; \quad (1)$$

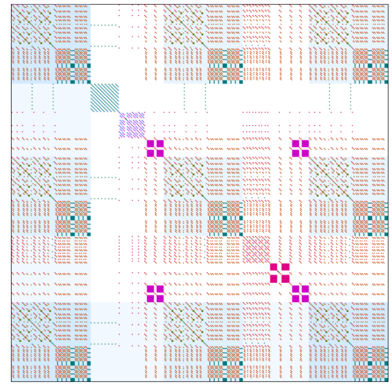

Figure 5. Visual Matrix for Silento - Watch me

Lastly, we compare the generated visual matrix pattern which we also call the song DNA print seen in Figure 5 to assure that there is simillarity between songs recommended the user. To achieve precision we utilize Signal to Noise (SNR) Ratio (1) on the song prints to determine how precise is the match. If all conditions are met then the song gets recommended to the user. We are confident that this unique approach will benefit a lot of listeners of music on gigantic music streaming platforms such as Apple Music, Spotify, Tidal and the likes.

IV. Conclusion

This paper demonstrates the use of self similarity matrix used to optimize song recommendation algorithms in music streaming platforms. Self similarity matrices can be used to to counter the dark music problem often experienced by music streaming platforms. In future we plan to optimize our framework and algorithm for dedicated music players that users use everyday on their personal computers, smartphones and other digital music players such ones found in portable music players and cars [15] so that users stop relying on playlists and the shuffle button.

References

[1] Ò. Celma and P. Lamere, "If you like the beatles you might like...: a tutorial on music recommendation," in *Proceedings of the 16th ACM international conference on Multimedia*. ACM, 2008, pp. 1157–1158.

[2] H.-S. Park, J.-O. Yoo, and S.-B. Cho, "A context-aware music recommendation system using fuzzy bayesian networks with utility theory," in *International conference on Fuzzy systems and knowledge discovery*. Springer, 2006, pp. 970–979.

[3] X. Wang, D. Rosenblum, and Y. Wang, "Context-aware mobile music recommendation for daily activities," in *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 2012, pp. 99–108.

[4] B. B. Sree and S. Demissie, "A context-aware music recommendation," *International Journal of Research*, vol. 5, no. 7, pp. 497–504, 2018.

[5] P. Cano, M. Koppenberger, and N. Wack, "Content-based music audio recommendation," in *Proceedings of the 13th annual ACM international conference on Multimedia*. ACM, 2005, pp. 211–212.

[6] J. Donaldson, "A hybrid social-acoustic recommendation system for popular music," in *Proceedings of the 2007 ACM conference on Recommender systems*. ACM, 2007, pp. 187–190.

[7] J. T. Foote and M. L. Cooper, "Media segmentation using self-similarity decomposition," in *Storage and Retrieval for Media Databases 2003*, vol. 5021. International Society for Optics and Photonics, 2003, pp. 167–176.

[8] J. Foote, "Visualizing music and audio using self-similarity," in *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*. ACM, 1999, pp. 77–80.

[9] J. Foote and M. Cooper, "Visualizing musical structure and rhythm via self-similarity." in *ICMC*, vol. 1. Citeseer, 2001, pp. 423–430.

[10] Z. Rafii and B. Pardo, "Music/voice separation using the similarity matrix." in *ISMIR*, 2012, pp. 583–588.

[11] F. Kaiser and T. Sikora, "Music structure discovery in popular music using non-negative matrix factorization." in *ISMIR*, 2010, pp. 429–434.

[12] G. Peeters, "Sequence representation of music structure using higher-order similarity matrix and maximum-likelihood approach." in *ISMIR*, 2007, pp. 35–40.

[13] M. O'Dair and A. Fry, "Beyond the black box in music streaming: the impact of recommendation systems upon artists," *Popular Communication*, vol. 18, no. 1, pp. 65–77, 2020.

[14] D. J. Hargreaves, "The effects of repetition on liking for music," *Journal of research in Music Education*, vol. 32, no. 1, pp. 35–47, 1984.

[15] L. Baltrunas, M. Kaminskas, B. Ludwig, O. Moling, F. Ricci, A. Aydin, K.-H. Lüke, and R. Schwaiger, "Incarmusic: Context-aware music recommendations in a car," in *International Conference on Electronic Commerce and Web Technologies*. Springer, 2011, pp. 89–100.