SMU Data Science Review

Volume 5 | Number 3 Article 4

2021

Emotion Integrated Music Recommendation System Using Generative Adversarial Networks

Mrinmoy Bhaumik
Southern Methodist University, mrinmoy.bhaumik@gmail.com

Patrica U. Attah
Southern Methodist University, trishattah@gmail.com

Faizan Javed
Southern Methodist University, fjaved@mail.smu.edu

Recommended Citation

Bhaumik, Mrinmoy; Attah, Patrica U.; and Javed, Faizan (2021) "Emotion Integrated Music Recommendation System Using Generative Adversarial Networks," *SMU Data Science Review*: Vol. 5: No. 3, Article 4.
Available at: https://scholar.smu.edu/datasciencereview/vol5/iss3/4

Emotion Integrated Music Recommendation System Using Generative Adversarial Networks

Mrinmoy Bhaumik[1], Patricia U Attah[1],, Dr. Faizan Javed,
[1] Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA
mbhaumik@mail.smu.edu
pattah@mail.smu.edu

Abstract Music can stimulate emotions within us; hence is often called the "language of emotion." This study explores emotion as an additional feature in generating a playlist with a deep learning model to improve the current music recommendation system. This study will sample emotions from certain subjects for each song in a sample of the data. Since the effect of music on emotion is subjective and is different person to person, this study would need a considerable number of subjects to reduce subjectivity. Due to the limited resources, a portion of the data will be labeled with emotion from subjects and the rest of the data will be labeled by using an active learning model. A content-based recommendation system will be built using a GAN (Generative Adversarial Network). This research led to creating two recommendation models, one utilizing emotion while the other did not. The Cosine Similarity and the Euclidean Distance where the two metrics used to judge validity of the models. The results showed that the model that utilized emotion performed better than the model that did not but the difference between the two was not statistically significant. One can conclude that there is promise in using emotion as a feature when recommending music. Further research would have to be done to mitigate certain obstacles as well as utilizing better resources to enhance emotional data extraction.

1. Introduction

Advances in technology have given us an information overload, and web-based music services can employ recommender systems as a powerful tool to harness this information and create a better experience for the user. The Merriam-Webster online dictionary defines emotion as "a conscious mental reaction (such as anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body." This definition underlines the impact of emotion on people. Emotions can determine a person's outlook on life and influence his or her actions towards the people around them. They can either help maintain or deteriorate a person's health, decision-making skills, and productivity. Many benevolent acts as well as crimes occurred because of strong emotions. For example, in criminal law, there is a term called a “crime of passion.” A “crime of passion” is when a person impulsively commits a crime based on a strong emotional response to some sort of provocation. The legal system reduces the sentence if proven that a crime is a “crime of passion.” It is widely recognized how a strong emotional impulse like rage can cause a person to act spontaneously without thinking of the consequences of his or her own actions. An article stated that a study was conducted on college students where students played a game that measured their level of cooperation when listening to music conveying different emotional states. The group that listened to happier music showed a higher level of cooperation than the group that listened to angry or sad music which is another clear demonstration of how emotions can affect actions [20]. Emotions can even affect your health and a direct health issue related to emotion is clinical depression. Clinical depression is expressed as feeling of sadness for extended periods of time and is a serious medical illness [18]. There are even physical symptoms that usually accompany depression are vague aches and pains, including chronic joint pain, limb pain, back pain, gastrointestinal problems, tiredness, sleep disturbances [19]. Due to the strong correlation between music, emotion and mental health, the American Music Therapy Association (AMTA) association was developed in 1998. AMTA uses music to accomplish personalized goals such as managing stress, enhancing memory, improving communications, alleviating pain, promoting wellness as well as physical rehabilitation, and more. AMTA states the treatments are clinical and evidence-based.

Emotions from music also have a significant effect on people in different age ranges from young to old. In an article by Shahram Heshmet, he explains that young people may derive a sense of identity from the music they listen to [16]. In a report by Ryan Hill, he states music has been recognized as one of the most influential elements of society. He investigates the different genres of rap and their influence on people. The article referred to a study from Emory University that discovered that teens between 14-18 who listened to rap music were 2.5 times more likely to get arrested and 1.5 times more likely to partake in illegal activities [17]. Researchers project that 20% of older adults are estimated to have some type of mental disorder such as mood disorder or cognitive impairment. They expressed depression as detrimental to older people due to its direct influence on emotional, cognitive, and physical health, so they decided to research music-based emotion regulation [21]. The study mentioned is critical as music is one of the most subtle and potent ways of altering our emotions. Although this research does not aim to alter the emotional state, emotion will be included as a feature to help the study select more engaging and pleasurable music for the listeners.

There are a variety of recommendation systems currently in use today. Most use a combination of the features of the specific item and user preference to generate recommendations. These recommendation systems can be a collaborative filtering system, content-based system, or a combination of both. Considering the significant effect music has on emotion, this study intends to investigate the use of a content-based music recommendation system by incorporating the emotional aspect of the song with the advantages of using a generative adversarial model.

Recommender Systems (RS) can be classified into three groups Collaborative Filtering (CF), Content-Based (CB) and a Hybrid method which incorporates both prior named methods. CF relies on how other users rate the item and compares the similarities between different users and then recommends items to others based on their rating. There are two types of CF systems: model-based and memory-based.

Memory-based CF uses the user rating historical data to calculate the similarity between users or items. There are two varieties of memory-based CF. There are the item-based and user-based, where item-based looks for users who like comparable items and outputs recommendations based on items those users liked. While user-based predicts the items a user might have interest in based on ratings given on that item by other users with a similar taste. The memory-based approach uses underlying algorithms such as cosine similarity or Pearson correlation. Model-based CF uses machine learning algorithms to predict users' ratings of unrated items. Some methods include matrix factorization, clustering, and deep learning. Content-based systems do not require user ratings and instead make recommendations using the features of the items from the user's history.

Today music streaming services such as Apple Music, Pandora, Amazon Music, and Spotify have their own music recommender systems. These systems are usually unique to each streaming service and are a combination of more than one model. For example, Spotify applies a combination of three different recommendation systems: collaborative filtering, natural language processing and audio file models [15]. Their collaborative models recommend songs by collecting information about what songs other users like. The recommendations are based on a series of features determined for each song, and user history. In previous studies there have been recommender systems built with a variety of deep learning models. Still, there have not been any that utilized emotion as a key feature in combination with the advantages of an adversarial generative model. There has been a study that used a Convolutional Neural Network (CNN) for emotion-based recommendations. However, this differs from this study since GANs has a built-in self-checking mechanism that attempts to detect whether an item is from the sample data i.e., the user's original tracklist. This unique trait of GAN will better help in identifying which music a specific user would prefer.

This study uses the Spotify API, which included over 18,000 songs with a useful set of features: danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, duration, and language. There is also another dataset with some of the emotion labels with the music, which will be combined with the one previously mentioned. To classify the rest of the unlabeled data, this study will have to have a few subjects label the emotion for a small portion of the songs. Since there are limited resources and time, the rest will be predicted by using an active learning model. Ideally, the study would have a substantial number of subjects to label the emotion data for the songs since emotion can be very subjective, especially regarding how different people perceive music.

After having all the emotions labeled, a GAN recommender system will be implemented to generate a playlist for the user. A GAN model is mostly used for image recognition, text generation, and even video editing. There is now a lot of research going into GAN recommender systems. Some studies are showing promising results in building effective GAN recommender systems since there are certain issues with recommender systems that GAN models can seem to address. "Originally proposed for unsupervised learning, the GAN approach can be applied in reinforced learning [3], which gives a high potential for its applications in the generation of recommendations" [7]. This study aims to continue that research and see how it applies to recommending music with the help of emotion.

2. Related Works

2.1 Emotions & Music

Authors have studied the music recommendation systems in context to emotions with different approaches [8]. Previous research has investigated music track selection to change or maintain emotional and psychophysiological states to support mental wellbeing. The study emphasizes the use of music's influence on humans by identifying the current mental state the individual is in by finding emotional and other features in the music in combination with external factors, personal data, and behaviors to alter their mental state. This study collects a wide variety of data such as the current state in general personal data, physical, sensory data, external factors such as social network-based applications, and interactive user feedback as the individual listens to the track. All this data collection makes it difficult to narrow down what factors influenced the individual's emotional state. This angle chooses to alter the emotional state and aims to generate a satisfying experience, but a dynamic state selected end–point. There have also been recent studies that have investigated facial expressions to determine emotional states which can be used to generate a playlist. The results were not expressed in mathematical terms of error or otherwise but were just a comparison of similarly selected features assigned by their feature detection software. But they reported the system successfully determined the mood changes during listening to music [9]. This study assumes that if the facial expression can be properly identified into one of the following categories happy, sad, fear, anger, surprise, or disgust, this would generate an engaging playlist, which is another interesting approach [12]. There has been research on emotion-based music recommendations for films as well. The investigated soundtracks of films and generated emotions from the music features and film video where captions, sound effects, visual features, and speech can also be factored in. This study sought to create certain emotions that could be used for articulating film expression. It used a modified Mixed Media Graph (MMG) to extract the emotion features and a Music Affinity Graph (MAG) to discover the relationship between the music features and the emotions. The result was that the best algorithm used between the two models they compared was 85 percent accurate.

Researchers have also used a music perception model to detect the emotion of audio files in terms of valence and arousal index. This was done by continuously observing songs aired on popular radio stations and creating a radio-induced emotion dataset. Back then, radio music was an effective way of inducing emotions to influence decision-making for marketing or Selling products [1].

There are also examples of studies where emotion is studied in other mediums related to music. There has been research on emotion-based music recommendations for films, in this study they investigated soundtracks of films and generated emotions from the music features and film video where captions, sound effects, visual features and speech can also be factored in. This study seeks to create certain emotions that can be used for articulating film expression [10].

2.2 Analysis of Music

Ashu Abdul et al, [2] proposed an emotion-aware personalized music recommendation system (EPMRS) to extract the correct song based on the mood. This system combines the Deep Convolutional Neural Networks (DCNN) approach and the weighted feature extraction (WFE) approach. The DCNN approach helped to extract the latent features from music data for classification. While in the WFE approach, implicit user ratings for the music are generated to extract the correlation between the user data and the music data. The system recommends the songs to the user based on calculated implicit user rating for the music.

In related research, authors used a large dataset and weighted feature extraction from user-to-song relationships that are determined from user data. A deep convolution neural network is deployed to get the song's latent music features. The user's listening history and audio signals from the specific songs help classify and recommend songs to the user. Overall, the study suggests some positive outcomes from this type of recommender system, but it would also need to assess what the user is listening to now continuously. This research heavily relies on CNN to do this type of classification and it would be interesting to see if this classification can be done more effectively with another type of deep learning model.

The objective of this following study is to find a way to classify emotions through music using SVM. Emotional states considered anger, sadness, and happiness. Subjectivity when it comes to emotion classification with music is a problem, to reduce the issue they build a classification system applied to different subjects. This is also a project which uses subjects to determine emotion. Subjects were asked to assign an emotion after hearing a certain song and then there were a total of 24 emotion cognitive tracks made for each subject. There are two SVMs that are created, one for arousal and one for valence. This can be an effective way to classify music within the main emotional states. One of the ways to determine the strength of their model was to use another music recommendation system approach to see the validity of their method. It seemed to compete well with the other systems, but the only issue was that there needed to be a larger collection music to pull from to get a more accurate model [20].

2.3 GANs

A GAN is a generative model using deep learning techniques such as neural networks. GANs are frameworks consisting of two sub-models known as the generator, and the discriminator. The generator creates similar data to the training data while the discriminator attempts to determine which data is real (from the training data) and which is fake (from the generator). Prosvetov [7] did a study where he created GAN recommender system where the system would recommend airline tickets to customers and compared it to a recommender system that was based on a Deep Neural Network and was able to successfully compete with it.

Most recommender systems suffer from data sparsity which occurs because users only interact with a small portion of the objects when being used with matrix operations, this can be overcome with GAN since the premise of the model is not matrix operation but just reproducing synthetic samples, another problem that recommender systems face is data noise this is negative or misleading samples that are uninformative and cause inaccurate results this particular situation rarely applies for this case since the music selection of an individual should not be wrong or inaccurate however if someone does have a song in their playlist they do not like a GAN will help determine this since the main function is to identify the true distribution of the selection of songs [11].

3. Data

There are two data sets used in this research and they are combined on the track id column. These data sets were combined because one has the feature emotion labeled while the other does not. Along with additional data that will be gathered from subjects this information would be used to generate the emotion classification for the entire data set using active learning. The first data set was originally sourced from Spotify through the Spotify API, downloaded from Kaggle. It consists of 25 features and 18,454 instances of songs. At the same time, the second data set was also sourced from Kaggle and consisted of 19 features and 686 instances. Only the eight most key features were used for the model to reduce noise and for compatibility of the neural network. For the data set without emotion, they were loudness, key, speechiness, acousticness, danceability, liveness, tempo, and energy. While for the dataset with emotion the features used were speechiness, acousticness, danceability, liveness, tempo, energy, valence, and emotion. When identifying emotions in music it is prone to subjectivity since people experience emotions in different ways. The main objective of this paper is to find out if we can generate a more effective playlist by utilizing emotion as a key feature. With that said, this study does not delve into the complexity of manually generating emotion classification of songs from audio features, environmental sensors, user interface activities or other methods. Emotions generated through these methods must be tested and authenticated to prove the accuracy of the predicted emotion. In the study, some of the songs have emotions classified while others do not so we are working with partially labelled data. Each unclassified song needs to be assigned either one or a combination of the two primary emotions. Ideally, for this study, data would be collected from a larger sample group of people, but this is not feasible due to limited resources and time. The emotion sampling capability is limited, and it is not possible to obtain emotions for every song, however a portion of the data will be labeled. An active learning machine learning model labels the rest of the songs and completes the process.

3.1 Data Exploration

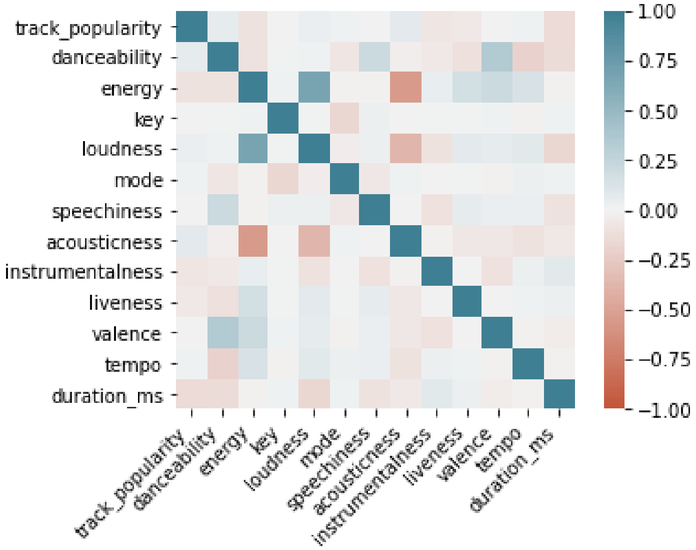

Fig. 1 The above figure is a correlation plot between the numerical features in the Spotify dataset.

Going through the Spotify dataset that was made available through the Spotify API there is a special focus on valence. Valence is the measure of positivity in a track where a high valence means the song is happy, cheerful, or euphoric while a low score would be considered sad, depressed, or angry. From Fig 1. The plot shows there is not much correlation with valence and the other numerical variables. The only two other features that show a correlation around .5 with valence are energy and danceability. So, looking at the energy and danceability specifically compared to valence one can see a positive linear regression with both.

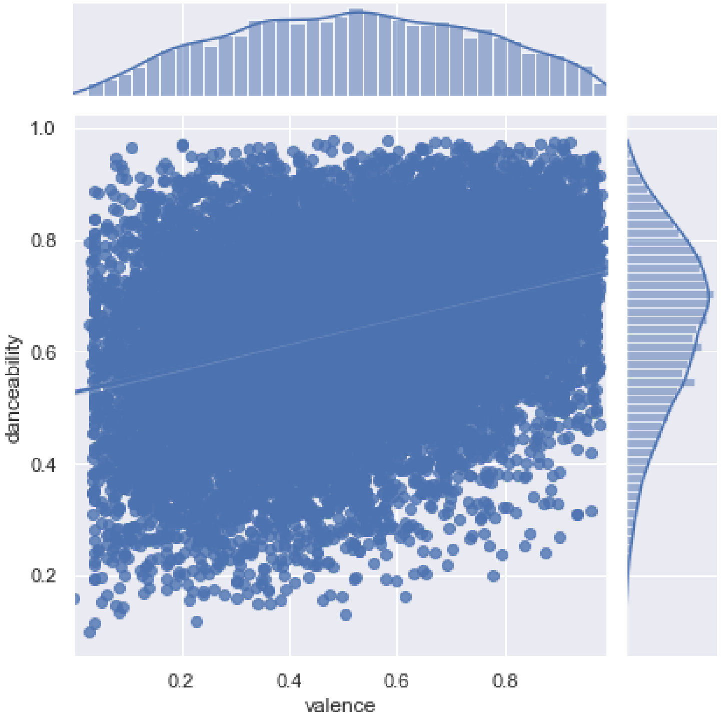

Fig. 2 The above figure is scatterplot of the danceability and valence features of the Spotify data showing that there is a positive correlation.

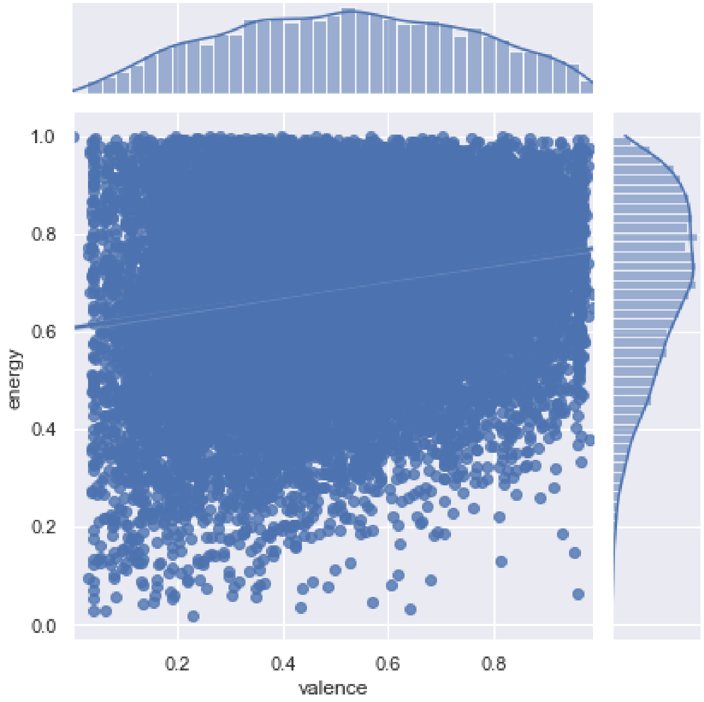

Fig. 3 The above figure is scatterplot of the energy and valence features of the Spotify data showing that there is a positive correlation.

There were 6 major playlist genres which are rock, r&b, pop, edm, latin and rap, these were divided into playlist sub-genres to a total of 24. Genre is a significant feature which was seen in an initial analysis. There seems to be a clear distinction between the playlist genre and various features in the data set including valence, danceability, energy, acousticness and loudness. There is a large variation in the median acoustic values between genres, r&b has the highest median value of acousticness at about 0.18 followed by Latin music at 0.17 while EDM and rock has the lowest values at about 0.12. The acousticness measures how confident that the song is acoustic. The genres with highly danceable music are rap, latin and r&b with median values of about 0.71, 0.7 and 0.68 respectively, while rock had an exceptionally low danceable value of 0.5. The genres have similar valence levels except latin which has a high median value of 0.62, while edm has a low median value of 0.38.

Investigating individual scatter plots a slightly negative correlation between energy and the how acoustic the song was as the energy increased the acoustic level slightly decreased, and slightly positive correlation between the loudness and the energy. Also, it's been shown that speechiness, liveliness and acousticness were strongly right skewed, while loudness was strongly left skewed.

4. Methodology

4.1 Emotion Labeling using Active Learning

The data used for this project is a combination of two data sets. The smaller data set has emotion identified; however, this data set does not have the genre of the song. While the larger data set that will be concatenated to the first does not have emotion classified. In this study there will be several sample emotions that will be obtained by members of the research group identifying and labeling the emotions they experienced by listening to the song. Due to limited time and resources this can only be completed for a certain portion of the songs in dataset. This is a frequent problem within the data science community since data collection can be costly and time consuming.

Typically, when there is a certain amount of unlabeled data, a good sizeable portion would need to be labeled and then a model would have to be trained on that dataset to make some sort of prediction for the rest of the unlabeled data. This method of classifying data is what is known as passive learning. Now there is a new way to classify unlabeled data known as active learning. It is still a topic that is being heavily researched, but it is showing promising results, especially in Natural Language Processing where it must deal with copious amounts of unlabeled data. Active learning gives researchers the opportunity to classify unlabeled data by reducing the amount of labeled data that is primarily needed to train the model. The main idea behind active learning is that if the learning algorithm is given the opportunity to choose the data it wants to learn from it will be able to perform better than traditional passive learning methods with less data [22].

When it comes to active learning, there are three different sampling scenarios that are usually applied. The three scenarios are membership query synthesis, stream-based sampling, and finally pool-based sampling. In this study, the active learning model is being used in pool-based sampling. Pool based sampling is used when there is a small set of labeled data and a large pool of unlabeled data. Instances are drawn from the pool with an informativeness measure. This informativeness measure is derived from something called uncertainty sampling. Uncertainty sampling in an active learning framework is when the active learning model queries the instances which it is least certain how to label [23]. There are three different ways of doing so. The first is called the least confidence strategy which selects the instance that has the least confidence it is most likely label. Then there is the margin sampling strategy which picks an instance and differences probabilities of the first and second most probable labels. Then finally, the sampling method that is being used in this study, the entropy formula, where x is the instance and y_i is the specific label:

$$x_H^* = \underset{x}{\operatorname{argmax}} - \sum_i P_\theta(y_i|x) \log P_\theta(y_i|x)$$

Even though the purpose of the active learning model is to use a small set of labeled data to determine the unlabeled data. The approach in this study was to first use the model only with the labeled data and cross validate the results with the actual label to assess the accuracy of the model. When the model is effective a certain accuracy then the active learning model will be implemented to classify the rest of the unlabeled data.

4.2 Generative Adversarial Networks

Generative adversarial networks (GAN) are a framework consisting of two neural networks, a generative and discriminative network, these two models are trained concurrently. The generative model uses supervised learning to discover patterns and trends in the data that can capture the data distribution to generate new instances of the data, while the discriminative model estimates the probability of the data coming from the actual data set or the generated one. The overall idea is that these two models are competing, the generator is creating synthetic music features that it passes to the discriminator with hopes to maximize the probability of the discriminator not identifying that the data is synthetic. The goal of this competition is to improve the model until the model generated music is for an individual is indistinguishable from the music an individual would have listened to, hence creating an improved recommender system. As related to our project the GAN will produce a set of synthetic songs. These songs are compared to a list of songs and the most similar songs to the synthetic songs are selected as recommendations.

The generative model applied in this system is multilayer layer perceptron (MLP). There are several types of GAN that can generate diverse types of synthetic data. Some types include Conditional GAN (CGAN) which has conditional parameter incorporated in the model, the Deep Convolution (DCGAN) which has a network of convolution layers and is suitable for image generation. In this paper we use a vanilla GAN which consists of two layers of MLPs. MLP uses supervised learning technique called backpropagation to train feed forward neural networks which is what a MLP is. The mathematical equation is optimized using gradient descent, in other words losses generated by the discriminator are sent to the generator and back to itself to improve the next batch of results. The inputs for the model are real data and latent space samples. Latent space samples were random continuous numbers from 0 to 1 that are converted to tensors with the same number of columns as real data. The real samples were normalized before input and there were no features with negative values. When results were output, they were inverse transformed back to be compared with the test data which were not transformed. The formula for the full is seen below

$$\min_{G} \max_{D} V(D,G)=\mathrm{E}_{x\sim p_{data}(x)}\Big[\log D(x)]+\mathrm{E}_{x\sim p_{z}(z)}[\log(1-D(G(z)))\Big]$$

where,
G = generator
D = discriminator
$p_{data(x)}$ = distribution of real data
$p_{(z)}$ = distribution of generated data
x= sample from real data($p_{data(x)}$)
z = sample from generated data ($p_{(z)}$)
D(x) = discriminator network
G(z) = generator network

Training the GAN

For the number of epochs in k steps the noise is passed into the generator while the discriminator is idle in this stage only forward propagation is done. The output from the generator as well as the input is passed to the discriminator predictions as the data labels are also passed into the discriminator. The loss function calculates the error and is passed back to the discriminator, which is updated by ascending its stochastic gradient. The losses for the generator are passed to it and the update occurs by descending its stochastic gradient. Discriminator update algorithm:

$$\nabla_{\theta d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D(x^{(i)}) + \log(1 - D(G(z^{(i)}))) \right]$$

Generator update algorithm:

$$\nabla_{\theta g} \frac{1}{m} \sum_{i=1}^{m} \left[\log(1 - D(G(z^{(i)}))) \right]$$

5. Results

Table 5.1. Accuracy results

	Euclidean Distance		Cosine similarity	
	Emotion	No Emotion	Emotion	No Emotion
Accuracy	0.61	0.47	0.54	0.41

Table 3.1 is a summary of accuracy from all test subjects combined. In both metrics the emotion incorporated model had better accuracy results. The model with emotion was 13% higher in accuracy for euclidean distance and 12% higher for cosine similarity. The accuracy results do not correlate to the results in the statistical tests, since we see a higher accuracy in the results with emotion. This may be because of a variety of reasons such as similarity metric suitability, a wider range of test subjects, more avenues of capturing emotions conveyed by songs and other reasons described in discussion section. In this study all other features were held constant except for the two features that were used to incorporate emotion. Neural networks must have a certain number of inputs that will affect the number of neurons in hidden layers of the network, for this reason when the emotion feature is removed it had to be replaced. Since the features were replaced by other features, the replaced features may have contributed to the results, positively or negatively.

Table 5.2 Statistical test results

	Euclidean Distance	Cosine similarity
Paired t-test p value	0.051	0.080

A Paired t-test was carried out on the recommendations to see if there was any difference in the recommended songs with or without emotion incorporated in the model. The null hypothesis states the mean difference between sample distributions is equal to zero. These statistical tests were performed by collecting the scores from the similarity metrics of both models and comparing them both distributions. From the results in table 3.2 both values are above 0.05 significance so there is insufficient evidence from these results we fail to reject the null hypothesis. This means the recommendations from both models show no difference in performance.

Table 5.3 Recommendation similarity scores per song for Subject

No-Emotion		Emotion	
Subject 1		Subject 1	
Euclidean distance	Cosine similarity	Euclidean distance	Cosine similarity
1.5197	0.999984	2.198920108	0.9999841
1.1323	0.999970	3.459064694	0.99997017
1.8956	0.999968	5.802268461	0.99996841
1.4493	0.999988	3.513973914	0.99998848
2.6448	0.999965	3.669265847	0.99996484
1.9315	0.999974	3.169606848	0.99997355

1.0853	0.999971	1.035656121	0.9999709
2.9141	0.999971	3.280530614	0.99997137
9.8326	0.999970	6.309128958	0.99997004
0.3771	0.999988	2.875532337	0.9999875
Subject 2		Subject 2	
1.5083	0.999975	2.881095847	0.99997501
1.6152	0.999982	3.009589017	0.99998153
2.3953	0.999958	4.227493154	0.99995785
1.1170	0.999965	3.273390279	0.99996479
4.0958	0.999969	3.08622661	0.9999686
1.6357	0.999995	2.589977153	0.99999511
1.4562	0.999967	3.792029003	0.99996691
1.4572	0.999986	3.008263865	0.9999855
3.4155	0.999966	6.032963217	0.99996634
1.1862	0.999960	2.469443298	0.99996045
8.1032	0.999989	27.94812987	0.99998871
Subject 3		Subject 3	
5.2419	0.999972	4.183723891	0.99997232
2.5075	0.999977	3.660045486	0.99997699
1.0085	0.999990	2.239902906	0.99999009
4.1664	0.999989	2.181314233	0.99998899
1.0148	0.999984	2.242708951	0.99998379
11.9504	0.999980	8.234428238	0.99998042
1.8422	0.999985	3.033822014	0.99998487
1.6167	0.999928	3.273022019	0.99992767
1.7640	0.999990	3.499429732	0.99998999
3.4408	0.999980	5.816164715	0.99998022
1.0957	0.999957	3.318144173	0.99995668

In the table above the similarity scores for the recommended songs for both metrics are shown. The cosine similarity averaged at 0.99 accuracy while the euclidean distance averaged at 2.3 for the model with emotion. The model without emotion average scores were 0.99 for cosine similarity and 3.7 for the euclidean distance. These are much higher than the rankings by the test subjects, which implies the similarity metrics were not efficient in finding personalized recommendations of the subjects. Although not covered in this paper further studies can be explored for methods to check for overfitting in neural networks and the effectiveness of similarity metrics in neural networks.

6. Discussion

This study implements a GAN to reproduce features of a song that will closely represent the type of songs the test subject would like. The GAN would be tested with two datasets. One dataset incorporates emotion using two features and the other dataset without emotion. The effectiveness of the recommendations is measured with cosine similarity and euclidean distance, to determine the accuracy and statistical significance of the recommendations. It is imperative to point out that the similarity algorithms play a significant role in this recommender system since the GAN does not recommend songs. The full responsibility of song selection falls on the similarity measures to identify the most appropriate songs. The GAN may generate exceptionally superior results but if the similarity measures cannot properly identify the combination of similar features between songs a less than desirable set of songs will be selected. The recommendations were generated by comparing the features of the songs that were most closely related to the synthetic music produced by the generator, the songs with the closest similarity measures were selected, and compared to the user's playlist. The average cosine similarity score was above 0.98% for all recommendations for both models, with and without emotion. The average euclidean distance is 2.16 which is impressive since it is a small value. Since recommendations and emotions are subjective the subjects ranked the playlists to verify if these metrics produced the best songs.

The accuracy score suggests that songs recommended from the data incorporating emotion had a higher likeability rate than the data without emotion. However, the statistical tests did not have the same conclusion it showed there were no difference between the recommendations with without emotion incorporated. Accuracy scores show that the highest score was 61%. This means that the models did not perform well in recommending personalized songs for the test subjects. One major reason may have been the similarity metrics chosen. Currently there are a variety of similarity metrics being used in machine learning, and they can be divided into two separate groups namely similarity based, and distance based. One metric was chosen from each group, cosine similarity from similarity based and euclidean distance from distance based. These metrics were chosen because they are popularly used for recommender systems with satisfactory results and cover the two categories of similarity metrics. These metrics may not have performed well due to the inherent nature in which they function. Cosine similarity is usually implemented with matrix factorization recommendation systems. Matrix factorization normally suffers from data sparsity which makes data farther apart in length, cosine similarity can overcome this since it uses direction and not length to determine similarity. Neural networks do not suffer from data sparsity and the output of neural networks may not be conducive for this metric however further analysis should be conducted on the internal structure of the output in relation to cosine similarity method of measurements. The euclidean distance measured the direct distance between two points, and it performed slightly better than the cosine similarity in both models. In prior studies it has been noted that high dimensional data has a negative effect on the performance of distance metrics and can be argued to lose meaning. Further studies on the effects of high dimensional data with distance metrics should be conducted to determine the effectiveness of the euclidean distance.

This study primarily depends on two stages, the quality of the synthetic music generated and the ability of the similarity algorithms to properly identify similar songs. The loss per epoch chart gave us an insight into how the GAN was performing. In both the model with and without emotion the epochs did not smooth out at the end on the runs this may imply that a higher number of epochs may produce better results. However, the number of epochs does not guarantee that there will be little or no loss variation at the end and weighing cost versus returns we could run no more than 600 epochs on the data due to memory and computing limitations.

The results, however, do not specify the extent to which the emotions played, there has not been an analysis on the feature importance's in the model. The attempt to magnify the role of emotion was implemented by holding all other features constant with exception of the replacement variable, but there are no calculations to further support this method. Another scope of the results that were not covered is how effective the similarity measured the recommendations. The similarity scores for the recommendations were remarkably high but these did not translate to the rates by the subject, more studies into the methodology of how the comparison should be made would benefit this analysis. An aspect of analysis that investigates how well the emotion features were assigned to the songs was not in the scope of this study.

One of the uncertainties that this paper faced is having decisive answers to whether the emotion incorporated data impacted the results, since the euclidean distance metric showed that emotion feature did have a positive effect on the recommendations from the statistical tests while the cosine similarity metric implies there was no difference in the recommendations using both data with and without emotion. A solution to this would be to create a GAN model that would recreate music by correctly classifying the song to the correct emotion, this would tell reveal how well the model can learn the emotion of a song. If the GAN classifies the emotion well, then it would be understood that when these emotions are in the data as compared to out of the data, if there is no difference in outcome of recommendations it would be give stronger indication that the emotion did not have an impact on the recommendations since it was the GAN properly understands the correct emotion for each song.

There was a limitation with this study. For one, there was an issue with the number of test subjects. Since the perceived emotion through music is subjective from person to person, there was a need for a larger number of test subjects to label the emotions but due to the limited resources of this study this could not be accomplished. The emotion labels were created by a small number of test subjects so when the active learning was applied to the rest of the unlabeled data the emotion labels were based on the experiences of a small portion of people. In the future, to get better and more accurate results, a study would need to have a larger sample size of people and make sure each person labels a certain minimum number of emotions for each song (i.e., 25 songs labeled happy, 25 songs labeled sad, etc.) this way there will be a better distribution of each emotion.

To get a better understanding of the emotional state as well as the listening preferences of the user there should be an understanding of the overall listening history of the individual. Data in other aspects of the user's life would also aid in achieving optimum analysis. Relationship status, current occupation, active lifestyle or otherwise, recent concerts, categories of media consumption, state of health may affect listener's emotional state while ranking songs. These details are important in not only understanding the user's emotional state but also to get a better idea of what type of music the individual prefers. The time-of-day user listens to music and what type of music the user may be open to listening to in the future can also be helpful.

6.1 Ethics

As mentioned before, emotional labeling was taken from a few subjects in this study, but it was done with their consent. If this study were to be done with more test subjects, then it would also have to be done with their consent as well because they would be giving up a little glimpse into their emotional insight which could be considered a breach of their privacy if permission were not asked. Previously when it was stated how the study could be better by developing a better understanding of a person's emotional state with access to more data, that could be an ethical concern. Retrieving data from other aspects of a user's life could be a breach of consumer privacy laws. If this were applied, whichever company wanted this data would have to ask for permission to access the rest of the user's data but even then, this could still be a bit invasive.

7. Conclusion

Overall, the results of the research do not point to any definitive conclusion on whether emotion can play a key role in making a better recommendation system. Although there was no statistical significance between the model with or without emotion, the accuracy was better for both metrics in the model with emotion. Though both p-values were not statistically significant they were near to 0.05 significance level. With additional measures taken, as specified in the discussion and ethic section emotion may be found to be statistically significant. Some of the improvements that could have increased statistical significance are additional subject participants, research into the most effective similarity metric, additional methods of obtaining emotions including lyrics analysis and test subject behavioral analysis during listening process. Further research that can be done to improve results include using better resources to extract emotional data (i.e., machines that can read heartrates, breathing signals, or facial expressions).

This study shows that there is potential for emotion to be a significant feature when recommending music. Future research and even implementation of this subject area can really help users discover music through another avenue outside of features one might see in the Spotify data set. Research would have to be done to find better and more sophisticated ways to label the emotion of the songs but also find the real time emotions of the user as well, with this in conjunction, I can see emotion can be a valuable asset in music recommendation.

References

1. Panwar, S., Roopaei, M.,Rad,P., Choo, K. (2019). Are you emotional or depressed? Learning about your emotional state from your music using machine learning. Journal Of Computing. Vol. 75 Issue 6, p2986-3009. 24p.
2. Ashu A., Chen, J., Hua-Yuan, L., Shun-Hao C. (2018). An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach. **Applied Sciences; Basel** Vol. 8, Issue. 7
3. Oramas, S., Nieto, O., Sordo, M.,Serra,X. (2017). A Deep Multimodal Approach for Cold-start Music Recommendation. ACM Proceedings of the 2nd Workshop on deep learning for recommender systems, 2017-08-27, p.32-37
4. Yepes, Fabio A ; López, Vivian F ; Pérez-Marcos, Javier ; Gil, Ana B ; Villarrubia, Gabriel (2018). Listen to This: Music Recommendation Based on One-Class Support Vector Machine. Cham: Springer International Publishing Hybrid Artificial Intelligent Systems, 2018-06-08, p.467-478
5. Vall, Andreu ; Widmer, Gerhard (2019). Machine Learning Approaches to Hybrid Music Recommender Systems. Cham: Springer International Publishing Machine Learning and Knowledge Discovery in Databases, 2019-01-18, p.639-642
6. Antal, D., Fletcher, A., & Ormosi, P. L. (2021, September 20). Music streaming: Is it A level playing field? Retrieved November 20, 2021, from https://www.competitionpolicyinternational.com/music-streaming-is-it-a-level-playing-field/
7. Prosvetov, A. V. (2019). Gan for recommendation system. *Journal of Physics: Conference Series, 1405*(1), 012005. doi:10.1088/1742-6596/1405/1/012005

8. Xinxi Wang and Ye Wang. 2014. Improving Content-based and Hybrid Music Recommendation using Deep Learning. In Proceedings of the 22nd ACM international conference on Multimedia (MM '14). Association for Computing Machinery, New York, NY, USA, 627–636. DOI:https://doi.org/10.1145/2647868.2654940
9. H. Zhang, H. Yang, T. Huang and G. Zhan, "DBNCF: Personalized Courses Recommendation System Based on DBN in MOOC Environment," 2017 International Symposium on Educational Technology (ISET), 2017, pp. 106-108, Doi: 10.1109/ISET.2017.33
10. Van den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Weinberger (Eds.), Advances in Neural Information Processing Systems 26 (2013) (Vol. 26). Presented at the Neural Information Processing Systems Conference (NIPS 2013), Lake Tahoe, NV, USA: Neural Information Processing Systems Foundation (NIPS).
11. Min Gaoa, Junwei Zhanga, Junliang Yuc , Jundong Lid, Junhao Wena, and Qingyu Xiong (2020). Recommender Systems Based on Generative Adversarial Networks: A Problem-Driven Perspective, Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University), Ministry of Education, Chongqing, 401331, China
12. Shan, M.-K., Kuo, F.-F., Chiang, M.-F., & Lee, S.-Y. (2009). Emotion-based music recommendation by affinity discovery from film music. *Expert Systems with Applications*, *36*(4), 7666–7674. https://doi.org/10.1016/j.eswa.2008.09.042
13. H. Immanuel James[1], J. James Anto Arnold[2], J. Maria Masilla Ruban[3], M. Tamilarasan[4], R. Saranya[5] (2013) Emotion based music recommendation system e-ISSN: 2395-0056, p-ISSN: 2395-0072
14. Mikhail Rumiantcev, Oleksiy Khriyenko, Emotion Based Music Recommendation System, ISSN 2305-7254
15. Heshmat, S., Ph.D. (2019, August 25). Music, emotion, and well-being. Retrieved November 21, 2021, from https://www.psychologytoday.com/us/blog/science-choice/201908/music-emotion-and-well-being
16. Hill, R. (2020, April 23). The influence of rap music in society. Retrieved November 21, 2021, from https://spokeonline.com/2020/04/the-influence-of-rap-music-in-society/
17. Clinical depression. (n.d.). Retrieved November 21, 2021, from https://uhs.berkeley.edu/health-topics/mental-health/clinical-depression
18. Trivedi MH. The link between depression and physical symptoms. *Prim Care Companion J Clin Psychiatry*. 2004;6(Suppl 1):12-16.
19. Monroe, J. (2021, February 22). Spotify top songs list highlights the effects of music on emotion. Retrieved November 21, 2021, from https://www.newportacademy.com/resources/treatment/effects-of-music/
20. *Jang, S. (2017, December 13). MUSIC-BASED EMOTION REGULATION (MBER) INTERVENTION MANUAL FOR PREVENTION OF DEPRESSION IN OLDER PERSONS [Pdf]. Kansas: Graduate Faculty of the University of Kansas.*
21. Stefan, H. (2018, February 9). Guide to active learning in machine learning (ML). Retrieved November 21, 2021, from https://www.datacamp.com/community/tutorials/active-learning
22. Settles, B. (2009, January 9). *Active Learning Literature Survey* [Pdf]. Wisconsin: Computer Sciences Technical Report 1648 University of Wisconsin–Madison.

23. Mosquera, D. G. (2020, May 19). Gans from scratch 1: A deep introduction. with code in pytorch and tensorflow. Retrieved November 21, 2021, from https://medium.com/ai-society/gans-from-scratch-1-a-deep-introduction-with-code-in-pytorch-and-tensorflow-cb03cdcdba0f
24. Lüthe, M. (2021, May 13). Calculate similarity - the most relevant metrics in a nutshell. Retrieved November 21, 2021, from https://towardsdatascience.com/calculate-similarity-the-most-relevant-metrics-in-a-nutshell-9a43564f533e
25. Z_ai. (2021, September 26). The surprising behaviour of distance metrics in high dimensions. Retrieved November 21, 2021, from https://towardsdatascience.com/the-surprising-behaviour-of-distance-metrics-in-high-dimensions-c2cb72779ea6