Alexander Ho
5/10/23
ECEN 5712

# Using HMMs to Create EDM Music

---

Electronic Dance Music (EDM) is a vibrant genre known for its energetic beats, infectious melodies, and immersive electronic soundscapes. With its diverse subgenres and global appeal, EDM has captured the hearts of music enthusiasts worldwide. As an avid music listener, I have been captivated by the dynamic compositions and pulsating rhythms of EDM. Although I always wanted to make music, I lack the musical talent to create such music on my own, I began to explore how modern tools and techniques from my area of expertise could be harnessed to generate new music in this genre. This realization led me to delve into the potential of Hidden Markov Models (HMMs) as a powerful analytical tool for music composition. HMMs can learn intricate statistical patterns, including drum beats, note frequencies, and chord progressions, which are fundamental elements of EDM. By harnessing the analytical capabilities of HMMs, I aim to create a simple EDM song that embodies the essence and excitement of this genre.

The motivation behind this project stems from the belief that HMMs offer a unique opportunity to unlock new possibilities in music generation. This project serves as a testament to the boundless potential of HMMs in the realm of music composition and highlights the untapped opportunities for creativity within the EDM genre. Through this exploration, I seek to showcase how HMMs can be employed as a tool for music generation, enabling individuals without traditional musical abilities to express their creativity and contribute to the rich landscape of EDM. By combining the creative freedom of EDM with the analytical power of HMMs, I aspire to create a unique musical experience that embodies the spirit and allure of EDM.

A Hidden Markov Model (HMM) is a statistical model widely used in various fields, including speech recognition, natural language processing, and pattern recognition. At its core, an HMM is a probabilistic model that represents a system with hidden states and observable states. I believe an HMM can best be explained with an example. Let's consider an example where the hidden states represent the weather conditions as either rainy or sunny, and the observable states represent whether someone who just walks into a room is dry or wet. In this scenario, the transition between weather conditions (rainy or sunny) is modeled as a Markov process, where the current state depends only on the previous state and not the one before ( that would be a Markov chain). However, the weather conditions themselves are not directly observable; instead, we can only observe whether the person entering the room is dry or wet.

By analyzing the sequence of observable states and making assumptions about the underlying hidden states and their transitions, an HMM can estimate the most likely sequence of hidden states given the observed data. This allows us to infer the weather conditions based on the wetness or dryness of individuals entering the room, even if we cannot directly observe the weather outside. HMMs provide a powerful framework for modeling systems with hidden states, which will further help us create new EDM music such as predicting the next beat.

One of the main challenges in this project is the lack of a dedicated dataset specifically comprising EDM music with all the necessary features required for creating new music. The absence of such a dataset poses a significant obstacle, as existing music datasets often contain a diverse collection of genres, making it difficult to isolate and extract the specific features unique to EDM. Moreover, even if an EDM dataset is available, it may not provide all the required features for music generation. To address this challenge, I have taken the initiative to create my own dataset. As a fan of the artist Seven Lions, I have selected a collection of my favorite songs from their many albums. These songs serve as a source of inspiration and a reference for the style I aim to recreate using Hidden Markov Models (HMMs). In creating the dataset, I have considered two complementary approaches: a traditional music analysis approach and a more analytical approach. By taking these two approaches, I will see which is better to create new music.

The objective of the music approach takes traditional sound properties such as pitch, duration, and velocity of each note and uses that as a feature for the HMM. The process begins by converting input MP3 files into WAV format to facilitate further analysis. Next, by extracting a variety of audio features from the WAV files, we can capture important characteristics such as pitch, intensity, and spectral properties. These features are then saved to a CSV file for future reference. Following the feature extraction step, the code preprocesses the audio by detecting note onsets and calculating the pitch, duration, and velocity for each note. This information is stored in a format that represents the musical composition. The note sequences are concatenated into a single array, preparing the data for training the HMM model. The HMM model is then initialized with 30 states and trained on the concatenated note sequences. The number of states was chosen as an arbitrary value to encapsulate many not combinations and patterns. During training, the model learns the statistical patterns and transitions between notes in EDM music. The trained model's transition matrix, covariances, and averages are printed and visualized, providing insights into the learned patterns (the transition matrix can be seen in the appendix). Once the model is trained, a function is defined to generate new sequences of notes. This function utilizes the trained HMM model to sample notes based on transition probabilities and the pitch distribution of each state. The generated note sequences are then used to create a MIDI file, with tempo and note tracks specified. Once, this entire process is done we can listen to the output file and see how well we have done. After, listening to the file, the duration is too long at hundreds of minutes, and only a few notes are played lasting minutes at a time. This process shows that this approach to analysis may not be sufficient enough to encapsulate all the features of EDM to generate new ones and so a more analytical approach may be required.

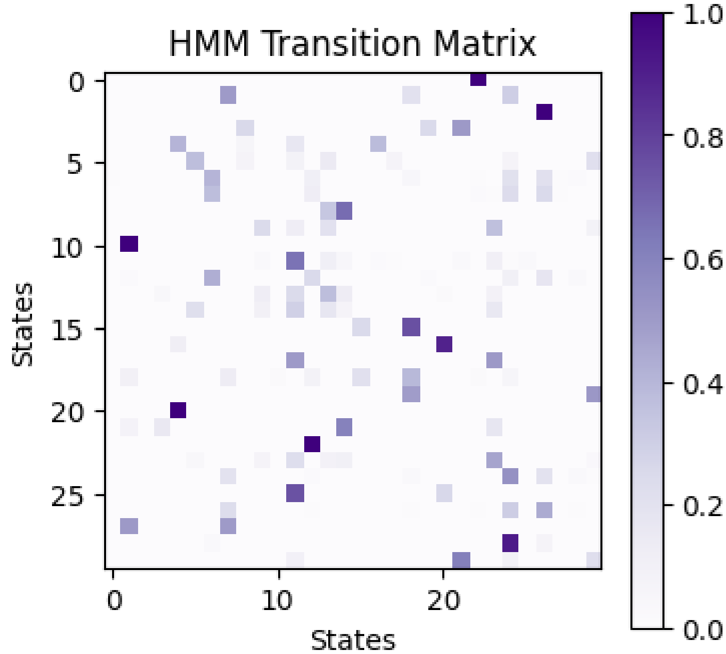
Since tracking simple music patterns did not yield a good result, a more empirical analytical analysis might yield something better. As opposed to the last approach, I will be looking at six different features. The MFCC (Mel-frequency cepstral coefficients) captures timbral information by transforming the short-term power spectrum of a sound signal using the Mel scale. Chroma represents the pitch class of a sound and is useful for analyzing the harmonic content. Spectral flux measures change in spectral content over time and help identify transitions. Spectral bandwidth measures the spread of spectral content, while spectral roll-off indicates the frequency below which a specified percentage of energy is contained. Zero_crossing_rate quantifies the number of times a waveform crosses the zero-axis, useful for analyzing rhythmic structures and transients. We will refer to these features as the MFCC features.

After the features are found, the music files need to be converted to beat frames which show when the beats occur in the music. Beat frames are computed using beat tracking. The MFCC features are clustered using K-means clustering, and the beat frames are labeled according to the dominant rhythmic patterns in each section of the song (intro, lyrics, chorus, base drop, ending). Next, a Hidden Markov Model (HMM) is trained using the labeled beat frames and feature vectors. The transition matrix between states is computed, along with the starting state probability distribution and emission probabilities for each state (the transition matrix for this implementation can be seen in the appendix). As opposed to the prior implementation, the number of states was specified to 5. This was an attempt to reflect the five sections of a song. After, using the trained HMM model, a sequence of states is generated using the Viterbi algorithm. A corresponding sequence of feature vectors and states is sampled from the HMM model. The generated sequence of states is then used to sample beat frames, forming the foundation for new music generation. The generated sequence of states is mapped to audio features using the centroids obtained from K-means clustering. These features are concatenated to create a feature matrix for the entire song. Using inverse MFCC transformation, the audio is synthesized from the feature matrix. Finally, the tempo of the synthesized audio can be adjusted, and the resulting audio is saved as a WAV file. On initial inspection, the output file seems to just be beats played very fast, once it is slowed down to about 99 beats per minute a cohesive song is created. Unfortunately, the output song is identical to one of the input songs. These files can be heard with syntesized_music.wav and new_synthesized_music.wav.

Although I have experimented with various methods to generate new EDM music, unfortunately, the results I obtained were not satisfactory in terms of quality or novelty. However, this process has provided valuable insights and learning experiences. Specifically, I gained a deeper understanding of the functioning of Hidden Markov Models (HMMs) and their relevance to music generation. One significant issue I encountered in the second implementation was the problem of overfitting. Initially, I used a limited dataset consisting of only three songs. However, due to the computational constraints and the first implementation resulting in an excessive number of features (over 300,000), I expanded the dataset to include ten songs. This adjustment aimed to mitigate overfitting and enhance the diversity of musical patterns captured by the model. Despite the challenges faced and the limitations encountered, this journey has allowed me to acquire valuable knowledge and refine my understanding of music generation techniques. It has also highlighted the importance of dataset size and variety when training machine learning models for complex tasks such as EDM music generation. With further exploration and experimentation, I believe there is potential to improve the results and unlock the true potential of algorithmic music composition.

## Appendix

Music analysis transition matrix




Analytical analysis transition matrix