Ahmad Moussa

Research Proposal

Introduction:

Music is multi-dimensional. It has rhythm, pitch, timber and tone, harmony and structure, patterns and cycles. The goal of a piece of music is to tell a story, or much rather, to convey a certain feeling. When we listen to a song, we can very distinctly say “this is a sad song” or “this song makes me happy”. This same argument can be made about other forms of Art, such as literature and cinematography. These other vessels of Art and expressionism present the same aforementioned features, such as structure, patterns and harmony. Ultimately, the goal of each creative work is to convey a certain idea, an emotion (however complex it may be) or even an abstract concept.

Motivation:

For almost a decade now, in the informal setting of my bedroom, I have been attempting to compose little pieces on the guitar, and along the way I have begun studying the compositions of successful musicians, that I enjoy listening to and inspire me. Many contemporary musicians across the genres that I listen to, attempt to conceptualize and translate a certain feeling or concept they are exposed to into a concrete composition and a piece of music. This process is sometimes difficult to explain with words, because music is such an abstract medium of expression.

I have also been exposed to the power of Artificial Intelligence and Machine Learning through my studies at the American University of Beirut. I learned that it is possible to conceptualize abstract human concepts into machine understandable data, such as the sentiment

analysis of human speech, or the creation of neural networks that can be taught to generate new pieces of music. And it has become something that I am very interested in and at the same time intrigued by.

Research question:

The question that I want to pose: Is translation between these modes of expression is achievable?

To pose the question differently, if there exists a book or story that has an emotional content (which speaks about love and friendship for instance), can we generate a piece of music that has the same underlying tone and emotional content as said book?

Through neural networks and deep learning we have already been able to create intelligent systems that can extract the emotional content and of subjective information, such as attitude and opinion, from written pieces of literature, as well as generative systems that can learn the fundamental components of musical pieces and create new compositions from scratch. Therefore I would like to invest time into researching how to combine these two fields, such as to create a system that is capable of translating and bridging between both ends. I have exposed myself and studied music for quite a while now and believe a solid enough understanding of the structure of music that I can attempt this project, a sample of my music can be heard here https://soundcloud.com/ahmad-moussa-16/asymmetric-flavours. I also have grown a big interest in machine learning throughout my bachelor studies, that I would like to utilize to realize this project.

Research methodologies:

For now I would like to draw the focus on emotional content of written works of literature and musical compositions.

In more technical terms, one would have to examine what attributes a certain "feel" to a work of literature. Extensive work has been done on sentimental analysis of text, and many "sentimental" lexicons have been created https://web.stanford.edu/class/cs424p/materials/ling287-handout-09-21-lexicons.pdf that assign a score for different categories of sentiments to each word. Hence the vocabulary an author chooses can draw an outline of the overall sentiment or underlying tone of his work. But some difficulties arise by simply attributing an "emotional score" to a chunk of text, for example we could encounter a structure such as:

I really like eating at this restaurant, but it is very expensive.

This sentence would happen to have a neutral score, due to it's overall bipolarity. Therefore one would have to investigate in a much smarter way to evaluate texts, such as to outline dependencies within it and attribute a different type of score http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/6459/pdf/imm6459.pdf. On the other hand, one would also have to look at the grammar utilized as well as the pace of the story. Are events happening in quick succession, or is the plot stagnant? This could yield important information about the rhythmical component of the composition that is to be created.

Now, how would we create a piece of music from all this data and values that have been gathered? Musical notes can be played simultaneously, or successively. Played simultaneously they form intervals and chords that can be attributed a certain degree of consonance or dissonance.

Our modern ear (western music) has been conditioned to perceive a certain feeling for each interval, for example, a major third interval is perceived to be happy and bright sounding and evoke a sensation of joy, in opposition to a minor third interval, which has a consensus of evoking a feeling of sadness. Another example would be the "diabolus in musica" (the devil in music), which refers to the diminished fifth interval, and is regarded as the most dissonant interval, and is described to sound "hostile and averse". But this is only a brief glance at how certain combinations of sounds can evoke a certain "feel", and composition techniques go much further than that to create more complex and intricate soundscapes, which in turn elicit much more delicate feelings.

The next step would be to feed the emotional data, that we extracted from our text, and feed it into a generative neural network that will create the composition. Input and output would be MIDI (Musical Instrument Digital Interface) files, which are simply said, files that describe the notes of our composition and can be played through a digital synthesizer in any music production software (such as Logic Pro X, Ableton, Cubase). The neural network that we would be using would need to be trained beforehand on a certain set of pieces from a certain genre, which would in turn give a certain "flavor" to our output. A lot of research has been conducted on such networks, and many papers on the topic exist. We would be using generative adversarial networks as described in the following two papers:

http://mogren.one/publications/2016/c-rnn-gan/mogren2016crnngan.pdf

https://arxiv.org/pdf/1406.2661.pdf

Hence the biggest difficulty that this research poses is most likely how the intermediary Data between the extraction and the generation steps would have to be modeled, such as construct

a meaningful and intelligent input model for the generative phase of the process. Another difficulty would be finding the optimal formats for the neural networks that will be used.

Goals and Outlook:

The goal of this research would be to create a scalable system, that can expand beyond the sentimental text analysis and be applied to other mediums. To list a couple of applications that this could be used for:

- Video: the pixel input from each frame of a video could be taken as input and music could be generated that could serve as accompaniment for that video, or that could be used as draft for film composers to embellish.
- Video games: in the same manner, we could take the pixel input from art or frames of a video game and create a preliminary soundtrack for the game.
- Film and theater scripts: the transcription of film scripts can be analyzed in the same manner the text from books can be used as input.
- Creating text from music: this would be using the model in reverse order such as to create a story from a piece of music.