ARTIFICIAL CREATION OF MEANING? THE FASCINATION AND LIMITATIONS OF A.I.-BASED MUSIC COMPOSITION

ALI NIKRANG
KEY RESEARCHER AND ARTIST AT ARS ELECTRONICA FUTURELAB

Introduction

The question of computer created art will certainly concern us in the coming years if not decades. Almost every form of art will be affected. Recent developments in the field of A.I.-based creative applications have shown highly promising results. Nowadays A.I.-based generative systems are able to create realistic images, videos and even text and music in high quality. However, the question that arises here is: Can machines create something that can trigger emotional responses in us, in the same manner as the art does?

In this article, we will discuss the relationship between A.I. and creativity with a focus on music and how A.I.-creativity can be used for music composition. We will discuss both the fascination behind these systems as well as their limitations and issues. We will also ask questions about the artistic value of a completely A.I.-generated piece of music and how this technology can be beneficial for both professional artists and for amateurs. The author of this article is himself an artist and A.I.-researcher at the Ars Electronica Futurelab in Linz, Austria. He has his background in both music and computer science.

The Technology Behind Artificial Creativity

Even if we consider the fundamental issues and limitations of today's A.I.-systems regarding creative and artistic tasks, recent developments have shown such a strong creative potential that can gain an impression of upcoming systems and their possibilities. Even with these limitations, artists have been able to harness the creative potential inherent in these systems in numerous fascinating works of art and projects. The corresponding A.I. technology in the background is called Deep Learning. It is a probabilistic A.I. technology that uses artificial neural networks to learn relevant features of a data domain that are needed to solve a predefined task of interest. The basic intention is simply to build a system that, based on the input data, outputs data that can solve a predefined problem.

However, since this assignment between input and output can be very complex, the neural network must be trained using examples. The model needs to learn independently all characteristics of the input data that are relevant for the solution of the task it has to perform.

Sometimes the task of the model is to generate data similar to the input data. For example, autoregressive models are often used for text and music generation, or the frequently used Generative Adverserial Neural Networks (GAN[1]) for many types of image generation. It should be mentioned that "similar" does not mean "the same". In most cases, the model should generate data that is not present in the training set. I want to give an example: Let's assume a model that is trained with human faces. The model should be able to generate images that can be recognized as faces, but they should not be included in the training set. There has been enormous progress in the field of this type of model in recent years thanks to the great academic interest in applying Deep Learning technology to creative tasks. But why are these types of models interesting for academic research at all?

The research interest in creative models is based on the assumption that the generation of new, similar data is only possible if a basic understanding of the nature of the input data is gained. For example, a model that is capable of generating human faces should have learned the essential characteristics of human faces and what makes a human face a face, but also the variations and differences between human faces (e.g., different genders, ages, complexions, eye colors, etc.). Furthermore, the model should also learn that the eye colors can differ from person to person, but a person has the same eye color in both eyes.

Generally speaking, Deep Learning systems are able to work with large amounts of training data. Often, in order to solve their tasks, they also learn features in the data that human observers may not have been aware of. They can identify relationships and dependencies that human observers might never have paid attention to, and can therefore serve as a new source of inspiration for creative tasks in general. Specially for artists, A.I. can also be seen as a new creative field for experimenting and discovering new possibilities for artistic works.

The Technology Behind A.I.-based Music Generation

In this section we will focus on some technical background aspects of A.I. systems and how these systems are trained and used for creating music. From a technical point of view, several research projects have shown that recent developments in A.I.-based natural language processing (NLP) can be successfully used for music generation.

To use these systems, musical data must first be represented as a stream of "words". There are several possibilities here. Pitches, for example, can be represented as unique "words", and durations and pauses can also be represented as word symbols. A sequence of these special words could represent symbolic musical data (like a score or a MIDI file) without any information lost. The representation of music data as a sequence of words may seem very unintuitive for musicians, but research has shown that A.I. systems make very effective use of these representations.

MuseNet[2] created by Christine McLeavey Payne at OpenAI in 2019 is the best example. It uses a technology similar to some of the most powerful natural language processing models (such as GPT-2[3], GPT-3[4] or Bert[5]) in the background and is trained with a sequential representation of music data similar to that of a language. The technology it uses, is called "transformer"[6]. Neural networks with transformer architecture learn to attend to a particular part of their context that is important for generating the current output. Using attention mechanism seems to be crucial for music generation as notes in the music (like words in the language) are highly dependent on the context in which they occur. Every word in a natural language can only be understood in the context in which it occurs. Likewise, the musical perception of every pitch in music depends on its musical context.

Besides, music consists of many repetitive structures that can be learned by referring to the previous context. These repetitive structures can be close together or even far apart. They include themes and structures that human listeners can remember when they appear later in the piece. Similarly, remembering past contexts is also very important for natural language processing: Consider a story in which, at the beginning, the main actors are introduced and later

1 Ian J. Goodfellow et al., Generative Adversarial Networks, 2014. url: https://arxiv.org/abs/1406.2661

2 Payne, Christine. "MuseNet." OpenAI, 25 Apr. 2019. url: openai.com/blog/musenet

3 Alec Radford et al., Language Models are Unsupervised Multitask Learners, 2019. url: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

4 Tom B. Brown et al., Language Models are Few-Shot Learners, 2020. url: https://arxiv.org/abs/2005.14165

5 Jacob Devlin et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018. url: https://arxiv.org/abs/1810.04805

6 Ashish Vaswani et al., Attention Is All You Need, 2017. url: https://arxiv.org/abs/1706.03762

in the text the actors are referred to by the use of words such as he, she, his, her. An A.I. system should be able to find intended references of all these words, even in cases they are ambiguous.

Models like MuseNet belong to the so-called autoregressive models.

The goal is to train a model which is able to predict the next word for given previous words. Autoregressive models are based on the idea that a model which is able to generate the next word from previous words must have learned a basic understanding of the domain from which the training data is taken. For language, this includes the meaning and relationships of words and the grammar.

Therefore, once a model has learned to "internalize" the data (which technically means learning a nonlinear internal representation of data), it can be used for tasks other than text generation. These tasks can include text summarization (the model can summarize a given text), sentiment analysis (finding the mood of a sentence, e.g. positive or negative), questions answering (finding answers to given questions), and many others. To sum up, text generation can be used as a task to understand the inner workings of a language. Once a representation of data is learned, it can be used for other related tasks.

In the case of music, the model is trained to generate the next note based on the musical context represented as previous notes. The length of the context varies from model to model. Depending on the memory and the type of model, it can consist of several hundred to several thousand elements.

In this section, we have seen that even though music and natural language are two completely different worlds, there are similarities in the challenges that A.I. systems face. We need the musical context to generate the next note in the music, similarly we need the musical context to generate the next notes.

Faszination

Today's A.I.-based composition systems are capable of composing music in a quality that was unimaginable even two years ago. For non-experts it is often almost impossible to distinguish between A.I.-generated music and music composed by humans. We demonstrated some of these abilities at the 2019 Ars Electronica Festival with the "Mahler Unfinished" project. We used the then recently released MuseNet to show how an A.I. system can compose music based on the theme of Gustav Mahler's last symphony (his unfinished symphony). The theme of the symphony is very unusual. It begins monophonic, without accompaniment and in a vague state of tonality. However, MuseNet was able to continue the theme, combine it with new musical ideas and create a piece of music. The result was arranged for orchestra and it was played by a professional symphony orchestra (Bruckner Orchestra Linz). The orchestra also played the original music of Gustav Mahler. It was also our intention to create a discussion basis for A.I. and creativity especially in the field of (classical) music. According to the feedback of the audience, A.I.-composed music sounded natural, emotional, and it was not possible for most listeners to recognize A.I.-composed music by its sound (which is certainly also related to the emotional performance of the symphony orchestra, which offered all its enthusiasm to the music, regardless of the composer). Of course, the Mahler experts were able to recognize the differences between the way the master worked with the theme and how the A.I. system developed the theme.

Markus Poschner conducts the Bruckner Orchestra Linz in Mahler Unfinished Project. The orchestra plays A.I.-composed music based on the theme of Gustav Mahler's last symphony (Ars Electronica Festival 2019). Credit: Jürgen Grünwald

However, this and similar projects will intensify the discussion about the artistic significance of A.I.-generated music. From a purely technological point of view, it is of course fascinating to see how A.I. systems are increasingly able to compose music that sounds natural and emotional and can trigger emotional responses in us as if it had been composed by an experienced human composer. Even from a music theory perspective, some of the results, which contain complex harmonies and rhythmic patterns, are highly fascinating. And yet this question arises: Can we call music produced by an A.I. system art? Can technology create artistic meaning without human intervention?

Despite all these fascinations, the question of the artistic meaning of a completely A.I.-generated piece of music is still unanswered.

In the next section we will discuss some of the fundamental limitations of A.I.-based music creation systems. The section is divided into two parts. In the first part we will discuss the limitations of current A.I. systems for music composition. It covers issues related to the technical nature of the models that could be overcome in the future. In the second part of this section we will discuss limitations that are not necessarily caused by current technology. These are general concerns about art created by A.I. and how A.I. can be useful for human art and culture.

Limitations

A.I.-based music generation systems might be able to generate the next correct note based on some given previous notes. However, just because a generated next note matches the previous notes (based on a statistical point of view) does not mean that the generated output has any creative value or higher meaning.

But how can we judge the quality of music composed by an A.I. system? The "novelty" of the output could be a criterion. Deep Neural Networks for music generation are trained with large amounts of data in order to generate similar data. Models used to generate natural language also have a similar goal. For a language model, however, the novelty of the data is not of the same importance as for music generation: It is not important, for example, that the English translation of a certain sentence into German does not occur in the training set and is novel. This could be a very important issue for music generation: Music can be easily recognized by the human listener if it contains familiar melodies, even if the melodies contain variations, occur in different tonalities and are played at different tempos.Therefore the output of the model should not be recognizable as one of the pieces that were included in the training set. However, modern music generation systems such as MuseNet seem to be able to generate new data that is different from the data set. When generating, these models also use some random strategies that make it very unlikely that the generated music was included in the training set (at least over a longer sequence of notes).

In addition to novelty, the quality of the output data can also be estimated on the basis of its content. It is not possible to objectively assess the artistic quality of a completely A.I.-generated content. However, we can still use music theory to estimate the quality of the output based on some music theoretical aspects. These points may include harmonic complexity, rhythmic complexity along with other music theoretical aspects. However, these music theoretical points cannot tell us anything about the artistic value of the output. Some of the greatest artistic moments in classical music have very simple musical structures.

Besides the novelty and excitement of the content, the "originality" is another important point: Current Deep Learning based computer models learn statistical patterns of training data, which makes it more difficult to learn and recognize what human listeners would perceive as the "originality" or the "uniqueness" of a piece of music.

A.I.-based composition systems tend to learn and create common rather than special structures. Their output contains harmonic, melodic and rhythmic structures that are more common in the data set. In music (and in art in general), however, "originality" is exactly what distinguishes the works of the greatest artists from others. The preference of common words over special words is also a well-known problem in the field of natural language processing. Research[7] has shown that (even beyond the field of art and creativity) humans tend to use more special words than machines. Tonal music contains moments that can be considered highly predictable (such as the end of phrases or the resolution of dissonant chords), but at the same time it attracts our attention by including special, unique and unpredictable moments. It is precisely these moments that give a piece of music its uniqueness.

Originality and uniqueness are not something that can be achieved by simply using structures with lower

7 See for example: Sebastian Gehrmann et al., GLTR: Statistical Detection and Visualization of Generated Text, 2019. url: https://arxiv.org/abs/1906.04043

probabilities. It requires a very deep understanding of both the data and the nature of human art.

Furthermore, special structures can only appear in the right places with the right intensity and frequency. There is no formal definition of "originality". However, from a technical point of view, future models might be able to learn these specialities better than current models.

Conceptual thinking is another issue with the current A.I.-based generation systems. Concepts are attractions that summarize the ideas or even the intention of an artwork. Especially in music there are many conceptual structures. Concepts in artworks can be very complex and even represent the intention behind the artwork. In the following we will only consider the basic types of concepts that can be found in almost every piece of music and are easily recognizable by human listeners. The main theme of a piece of music, for example, is such a concept: it can be seen as an overall musical idea. In classical music a particular theme and its variations can be stretched over several movements. Certain types of repetition are also types of concepts in which a section of a piece of music is repeated alternately. Several aspects of these concepts can be reproduced to some extent by learning the statistical patterns using attention mechanisms or by incorporating other ideas from the A.I. research. However, these aspects are very limited and cannot be compared to the way human composers work. For example, it is not possible to have a main theme that is present throughout the piece and that appears regularly with variations over and over again. The lack of ability to learn these kinds of concepts can be seen as a fundamental limitation of A.I.-based music composition.

Finally, the most important question (which is not limited to today's technology but also to that of the future) is the question of meaning:

Assume in the following that all technical problems are solved and we have the perfect system capable of creating music of the highest possible quality and in any style, simply by pressing the "compose" button. What is the meaning of the music generated? Can technology create meaning without the intention of creating meaning?

We need intention to create meaning, and intention is something that machines can not provide. Even the very first cave paintings in human history had the intention to create and draw meaning. Suddenly it was possible to make a fleeting imagination permanent in a way that could be communicated to others. Even several thousand years after creation we are still able to interpret drawn scenes. Nobody knows what the real intention behind the drawing of the pictures was. Nor can we say what kind of intention is behind a particular work of art. We just see out interpretation in artworks as intentions of the artist. Nonetheless, we can say that there is no creation without intention. Of course, readers of this article could argue that in modern music the intention[8] could even be to "create" a work of art that is not intended. But even in these situations there is an intentional creation of a new concept.

In addition, the selection and publication of a piece of music composed by A.I. is of course also a conscious act of creation. When we use A.I.-based music composition software, we must intentionally "press the button" and save the result. We could also publish the result and share it with others, which is also intentional. This gives the selected piece of music "its" personal human meaning. Of course, from an artistic point of view, the intention could also be to build the system itself, which composes music without any human involvement. But in this case, the creator of the system would be the source of the intention.

Collaborative A.I.

We discussed the fascination of the novel A.I.-based music composition techniques together with their limitations and problems. We showed how A.I.-based music creation systems are able to create the next correct note based on a given context that has previously occurred. We also discussed the question of the artistic meaning of the output which is generated without any intention and only on the basis of statistical patterns of data which are learned during the training of the A.I. model.

Creating "meaning", however, requires the intention to create "meaning", and this is why human collaboration is crucial for such creative systems as an A.I.-based music composition system. Given the rather limited interaction

8 We can find similar ideas in music of John Cage. We can also think about musical pieces where artists make use of randomness and leave the musical decisions to the coincidence or the performers

possibilities of today's creative A.I. systems for music, however, they are used more as simple tools than as counterparts. In general, the aspect of interaction and collaboration between creative A.I. systems and their human users has not received much attention in recent years.

But how can the collaboration between human artists and A.I. systems look like? Collaboration means working together as equals and as counterparts. A trained neural network has its own expectations and assumptions about the music learned during training. Based on these assumptions, it calculates the probability of the next note based on some given previous notes. A human artist has different assumptions and expectations. Our goal as researchers and artists should be to bring these two worlds together by creating a new environment that can be used by both professional and non-professional artists. The potential of current A.I. systems for creative tasks can only be used in cooperation with human beings.

Ricercar: An A.I.-based Music Companion

At the Ars Electronica Futurelab, we are investigating these aspects of collaboration between creative A.I. systems and human artists. It is reasonable that creative tasks require creative ways of collaboration. In order to find an optimal collaboration environment for A.I.-based music composition and human artists, we are developing a new interactive A.I.-based music composition system called "Ricercar" that provides intuitive ways for collaboration between human users and the A.I. system. The word Ricercar refers to a musical form of the Baroque and Renaissance and means "search" in its Italian origin. Composers used this term for pieces in which they experimented with a theme or musical idea and discovered its qualities such as permutation and variation possibilities as well as its harmonic potential.

"Ricercar" follows a similar idea. It aims to create an intuitive interface between human artists and an A.I.-based composition system, where human users and the A.I. system can discover the potential of a musical idea (given to the system by the users or initialized by the system itself) in a collaborative way.

In addition to composition, A.I. models like Ricercar could also provide us with some insight into the inner workings of the music itself by showing us new aspects of musical data that have not yet been considered by human observers. Maybe one day we can even understand more about the nature of music and its power to move us and our emotions so deeply. Remember that the main idea of training a text or music generation model is not necessarily to generate text or music at the end, but to learn a statistical representation of the data. This representation can also be used for tasks other than generation. For instance, to understand the inner workings of Ricercar, we used its inner representation of music in a project called "How Machines See Music".

Instead of composing music, we fed the model with an existing piece of music as input and observed the reaction of the neurons in each layer while the music continued. For each moment in the piece of music, it shows us the neurons that are reacting to that particular moment of music. Finally, a so-called "similarity matrix" shows the similarity between the activated neurons from any given moment to any other moment in the music. In other words, it shows us which parts in the piece are similar (the so-called repetitive structures) as they cause the activation of similar neurons.

The result can be seen as a novel way of visualizing music that shows higher-level structures of a piece of music (e.g. the repetitive structures). It also shows us once again that music is not only pleasant for our ears, but also contains beautiful hidden structures that span the entire piece and can be uncovered by using A.I. techniques.

Johann Sebastian Bach, Goldberg Variations No. 11. Visualization shows the activation maps of two selected layers in Ricercar.

To find the optimal way of collaboration between human artists and machines as eligible counterparts, we use the interface of Ricercar as a research tool to evaluate different ways of interaction and collaboration. The ideal collaboration should take advantage of the benefits of artificial creativity, together with the ability to personalize the result by having the artistic input come from the users. We believe that the aspect of interaction and collaboration will play a very important role in how successfully A.I. can be used for creative tasks. Despite the current limitations of A.I.-based music composition systems, even the current potential (not concerning the future potential of A.I.) can be much more beneficial for artists (both professional and non-professional) to expand their creativity and find new ways to create art by using better ways of interacting and collaborating with them. In the end, music is a very human and intuitive art. Composing music should be accessible to everyone in an intuitive way. A.I. technology could make composing more intuitive and make it something that anyone can do.