Schumann's Music Generation

Student Name: Jiahao Liu

Student ID: 21212299

jiahao.liu@ucdconnect.ie

Table of Contents

Introduction 1
Definition of Computational Creativity 1
Core Idea and Source of Inspiration 2
Dataset and Process 3
Model 5
Evaluation 6
Conclusion 12
Reference 12

Introduction

In the rapidly evolving world of artificial intelligence (AI), one intriguing frontier is the creative domain of music generation. As AI continues to break barriers and redefine possibilities, a compelling question arises: can a machine learn to compose music like a human? And more specifically, can it capture the essence of a renowned composer's style and create novel compositions that reflect their musical identity? This project aims to explore these queries by applying a deep learning model to generate music in the style of the illustrious composer, Robert Schumann.

Schumann is highly admired for his inventive and emotive compositions, which profoundly influenced the Romantic period of Western classical music. His work is characterized by complex harmonies, expressive melodies, and innovative structural designs, all contributing to a musical language distinctly his own. Emulating such intricate and nuanced musical elements poses a significant challenge for AI, and this project will explore the capabilities of the model in this regard.

The model is based on Long Short-Term Memory (LSTM), an architecture well-suited for sequence prediction tasks. The model is trained on a dataset comprising MIDI files of Schumann's piano compositions, transformed into numerical data sequences representing notes and chords. After successful training, the AI uses seed data from the corpus to generate new compositions, which are then converted back into MIDI files.

The performance and output of this music generation model are meticulously evaluated through various approaches. A detailed analysis of music theory aspects, including melody, rhythm, and harmony, scrutinizes the compositional structure of the generated piece. Cosine similarity measures assess the similarity of the AI-produced music to the original Schumann pieces, providing a quantitative metric of novelty. Furthermore, a listening test conducted by Xiaochi Li offers valuable subjective insights into the creative merit and stylistic authenticity of the generated music.

Definition of Computational Creativity

Creativity has been a defining trait of humanity, a beacon of our innate ability to generate novel and valuable outcomes in various domains of life, including the arts, sciences, and culture (Runco & Jaeger, 2012). It is often seen as the engine of cultural evolution, shaping societies, nurturing progress, and cultivating our collective imagination. Musical creativity, which weaves together the threads of emotion, expression, and innovation, stands as a remarkable manifestation of this trait.

The concept of creativity is not simple to define, owing to its multifaceted nature and its intersections with various cognitive, personal, and environmental factors (Kaufman & Sternberg, 2019). Batey (2012) describes creativity as the human ability to produce outputs that are novel and valuable within a given context. It involves both the generation of new ideas and the realization of these ideas into tangible outcomes. This definition not only encompasses the production of original artifacts but also the ability to problem-solve and innovate within specific frameworks and constraints, a feature that is prominently displayed in musical composition.

Building on this definition of creativity, the concept of Computational Creativity extends creativity to the realm of artificial intelligence and machine learning. As defined by Wiggins (2006), Computational Creativity focuses on creating software that can generate outputs that would be considered creative if produced by a human. This includes not only the capability to generate novel ideas or artifacts but also the ability to evaluate the quality of these outputs, to refine them, and to learn from the process. It signifies an attempt to replicate or simulate the human creative process within the confines of an algorithmic system.

By imitating or enhancing human creativity, Computational Creativity broadens our grasp of creativity itself and boosts our creative abilities. This exploration allows us to uncover the methods through which machines foster creativity and generate novel outputs. Moreover, it opens up fresh avenues for artistic and scientific expression and promotes creative synergy between humans and machines.

Core Idea and Source of Inspiration

This project delves into the confluence of human artistic expression and modern computational capacities, with a distinct emphasis on musical creativity in piano composition (McCormack et al., 2019). Harnessing the field of Computational Creativity, the project's objective is to devise an AI system that can conceive original, aesthetically pleasing piano compositions (hope so) inspired by Schumann's works (Jordanous & Keller, 2016).

Schumann's piano music serves as the inspirational backbone of this endeavor due to its monumental influence and unwavering significance in classical music. Born in 1810, Robert Schumann left an indelible impression on the musical world through his varied and emotionally resonant compositions (Sadie & Tyrrell, 2001). His piano works, celebrated for their evocative melodies, innovative structures, and emotional richness, provide a fertile ground for this project.

Schumann's musical oeuvre represents a high point in creative expression during the Romantic era. His piano compositions, distinguished by their harmonic innovations, intricate blend of poetic delicacy and dramatic intensity, and a remarkable range of emotional expression, are considered among the paramount in the repertoire. Utilizing his distinct creative signature as a blueprint, this project aspires to create an AI system capable of generating novel music that is both enjoyable and stylistically compatible with Schumann's works.

This initiative aligns with the wider ambitions of Computational Creativity research: the integration of AI and machine learning into creative practices to discover innovative methods and provide new tools and perspectives for artists. Moreover, it contributes to the ongoing discourse on the role of AI in creative industries, proposing machine contribution as independent creative entities rather than merely supporting human creativity.

Current technological advancements, especially in deep learning, enable the formulation of machine learning models that can understand and replicate complex data patterns (Goodfellow, Bengio, & Courville, 2017). Given music's inherent pattern-based construct, these technologies present promising opportunities for understanding and emulating the unique stylistic elements of Schumann's piano works.

The project entails training a deep learning model on Schumann's piano compositions, thus enabling the system to apprehend the intricate interplay of melody, rhythm, and harmony that characterizes his music (Briot, Hadjeres, & Pachet, 2019). With this training, the model is anticipated to generate new music that retains the unique "Schumannian" character while incorporating original elements. Finally, there is a comprehensive systematic evaluation of the generated music, including feedback from professional listener.

Dataset and Process

The dataset for this project consists of 24 MIDI files of piano pieces composed by Schumann (Rakshit, S. (2019), which were loaded into the Python environment using the Music21 library, a powerful toolkit for computer-aided musicology. The chosen compositions range across various periods of Schumann 's work, allowing the model to capture a comprehensive overview of Schumann's unique and innovative style. Here are some MIDI files of Schumann's work in Figure 1.

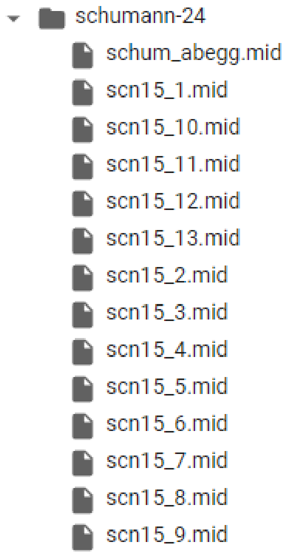

Figure 1: some Schumann's work

Each MIDI file represents a sequence of musical events or messages, including information such as note pitch, velocity (corresponding to volume or intensity), and duration, making it a rich source of musical data (Briot, Hadjeres, & Pachet, 2019).

To extract the musical elements from the MIDI files, two primary elements were considered: notes and chords. Notes represent a single musical tone with a specific pitch and duration, while chords represent a set of notes played simultaneously. The notes with low recurrence in the corpus, specifically those occurring less than 100 times, were considered as outliers and removed from the dataset. This process helped in reducing the complexity of the learning model and in avoiding overfitting to infrequent musical patterns. After cleaning, we can view the scores of the first 100 notes in the corpus in the Figure 2.

Figure 2: The part of corpus's sheet

The pre-processed corpus was then converted into a sequence of numerical data. A unique integer was assigned to each distinct note or chord, which led to the creation of a mapping dictionary. This dictionary became a crucial tool in translating the MIDI data into a format that the neural network could process, and in converting the network's output back into musical data. Then, in Figure 3, we can clearly observe the frequency distribution of notes in this corpus.

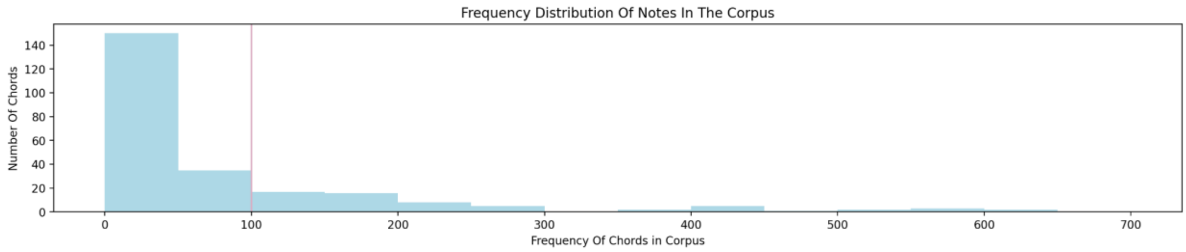

Figure 3: The frequency distribution of corpus' notes

The numerical data was then reshaped into a format suitable for the LSTM (Long Short-Term Memory) model. The LSTM model requires input in the shape of (samples, time steps, features), hence, the corpus was divided into sequences of a fixed length (40 notes/chords), and each sequence was paired with a target note that followed the sequence in the original corpus. The resulting input data was then normalized, as neural networks perform better with small input values.

Finally, to provide the model with a ground truth for learning Schumann 's style, a portion of the data was separated as a seed for the generative process. This seed data, constituting 20% of the dataset, was used to stimulate the initial output of the model and guide the learning process.

Model

The Schumann piano music generator is a deep learning model utilizing Long Short-Term Memory (LSTM)(Kapoor, K. (2021)), an effective Recurrent Neural Network (RNN) architecture for sequence prediction tasks. I adjust the model parameters to be more suitable for this project.

The model takes sequences of notes and chords from Schumann 's compositions as input, with each sequence being a 40-element array. The model consists of several layers, starting with an LSTM layer of 512 units. This layer outputs a value for each time step, capturing music patterns throughout the sequence. A dropout layer follows, randomly zeroing 10% of the input units during training to prevent overfitting and enhance the robustness of generated music. Another LSTM layer of 256 units is next, condensing the learned sequence information into a single output.

Subsequent layers include a dense layer of 256 units, which adjusts the model's output dimensionality to the number of unique notes and chords in the corpus, followed by another 10% dropout layer for further generalization. The final dense layer is size-matched to the corpus's unique elements and uses a softmax activation function to output a probability distribution for the next note or chord.

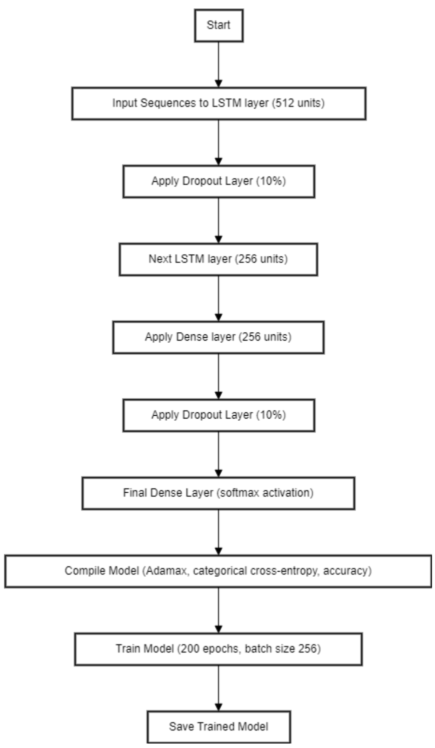

Figure 4: Model Flowchart Part 1

The model is compiled using the Adamax optimizer with a 0.01 learning rate and a categorical cross-entropy loss function. It's trained for 200 epochs with a batch size of 256. The trained model is saved for future use, and the training history is stored for analysis. In Figure 4, we can clearly see the process of the first part.

Post-training, the model generates new sequences from seed data, transformed back into musical notation using a mapping dictionary. The resulting sequences are output as MIDI files, which can be played to hear the model generating music in the style of Schumann. In Figure 5, we can clearly see the process of the second part.

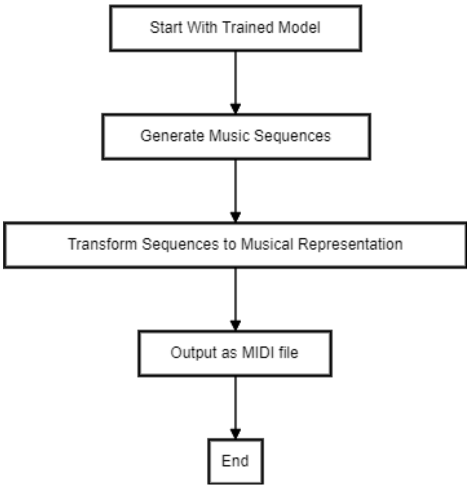

Figure 5: Model Flowchart Part 2

Evaluation

In this section delves into the evaluation phase, where the effectiveness of our musical model is assessed in terms of its capacity for creativity and innovation. After the model has been trained and a new musical piece has been generated, several techniques were employed to examine the musical elements of the generated piece.

We can follow these 3 main parts:

1. Music Theory Analysis: Consider music theory aspects such as melody, rhythm, and harmony (Krumhansl, 1990). Novelty is assessed based on the use of different scales, chord progressions, rhythmic patterns, etc. Conduct a deeper analysis of music, providing a more comprehensive understanding of the music that is generated (Cook, 1987).

2. Cosine similarity: By converting the generated music and training data into sequences of notes, they can be represented as vectors. Then, use cosine similarity to find the similarity between these vectors. A high cosine similarity indicates that the generated music is very similar to a piece in the training data, while a low cosine similarity may indicate that the generated music is more novel. In this way, it can be reflected that the generated music has absorbed the rules in the training set, but it is not exactly the same based on the original tone. The novelty portion of the score is reflected in the value (Briot, Hadjeres, & Pachet, 2019).

3. Listening test: In the subjective listening test, I invited Xiaochi Li, a future musician with 10 years of music experience, a UCD student of this course, to comment on the generated music. His evaluation can provide valuable insights into the novelty of the generated music.

First, let's look at the training effect of the model. we can see an overall trend of decreasing loss and increasing accuracy in the following Figure 6 over the course of the training epochs, which is a good sign of the model learning effectively from the data. Then, we can see the sheet music of the generated piece in Figure 7.

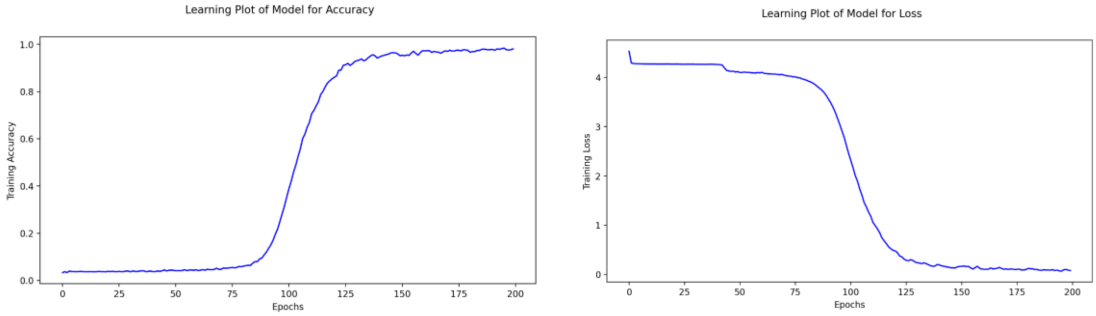

Figure 6: Model’s loss and accuracy

These two Figures indicate a successful model training. Over the course of 200 epochs, the model's performance significantly improved. Starting from an initial loss of 4.5308 at epoch 1, the model steadily decreased this value to a low of 0.0972 by epoch 200. Simultaneously, accuracy increased from a starting point of 0.0333 to a high of 0.9760 by the end of training. These improvements show the model learned and became more accurate over the training period.

Figure 7: New_schumann’s sheet

Music Theory Analysis

The first method implemented was the analysis of the key of the piece. This assessment gives a basic understanding of the tonal center and key signatures used in the melody. Alongside this, chord progressions in the piece were examined, revealing the relationships between chords and how they contribute to the harmonic structure of the music.

The Key of the piece:

```
Key of the piece: g minor
```

Figure 8: The Key of the piece

In Figure 8, we can see that the model generating piece is in G minor, a key often used by Schumann for its emotive qualities.

The Chord Progressions:

```
Progression: Minor Third above G -> Minor Third above F#
Progression: Minor Third above F# -> Minor Third above C
Progression: Minor Third above C -> Minor Sixth above D
Progression: Minor Sixth above D -> Major Sixth above C
Progression: Major Sixth above C -> Minor Third above G
Progression: Minor Third above G -> Bb-major triad
Progression: Bb-major triad -> Minor Third above G
Progression: Minor Third above G -> Minor Third above G
Progression: Minor Third above G -> Minor Sixth above D
Progression: Minor Sixth above D -> Major Sixth above C
```

Figure 9: The Chord Progressions

In Figure 9, we can see that the model generating chord progressions blend traditional and novel elements. Notable transitions include a chromatic step-down and a key change, both reminiscent of Schumann's style. The model uses a borrowed chord and creates a recurring theme, reflecting Schumann's techniques. The final progressions suggest tension and resolution, indicating the model's learned creativity and understanding of Schumann's musical language.

The Melodic Intervals:

```
Interval: Major Third
Interval: Minor Fourteenth
Interval: Augmented Octave
Interval: Augmented Fifth
Interval: Perfect Unison
Interval: Minor Third
Interval: Major Ninth
Interval: Perfect Nineteenth
Interval: Minor Fourteenth
Interval: Perfect Fourth
Interval: Perfect Fourth
Interval: Perfect Eleventh
Interval: Perfect Unison
Interval: Minor Seventh
Interval: Major Second
Interval: Major Ninth
Interval: Major Second
Interval: Perfect Unison
Interval: Minor Tenth
Interval: Major Third
Interval: Minor Second
```

Figure 10: The Melodic Intervals

In Figure 10, we can see that the model generating music demonstrates a complex melodic interval structure, capturing a balance of small and large intervals, similar to Schumann's style. The use of Major/Minor Seconds, Thirds, and Perfect Fourths/Fifths suggest a strong tonal center and conventional harmony, while larger intervals like Major Ninth, Perfect Eleventh, and above indicate dramatic, expressive lines. The presence of Octave and Double

Octave intervals adds variety while preserving tonal coherence. This suggests the AI's proficiency in generating musically compelling and harmonically coherent compositions.

The Note Range:

```
Note range: 710.5746799841472 Hz
```

Figure 11: The Note Range

In Figure 11, we can see that the note range of the model generating tune is about 710.57 Hz. This indicates a broad span of pitches used in the composition, which is characteristic of Schumann's style, known for its expressivity and dramatic contrasts

Note Distribution:

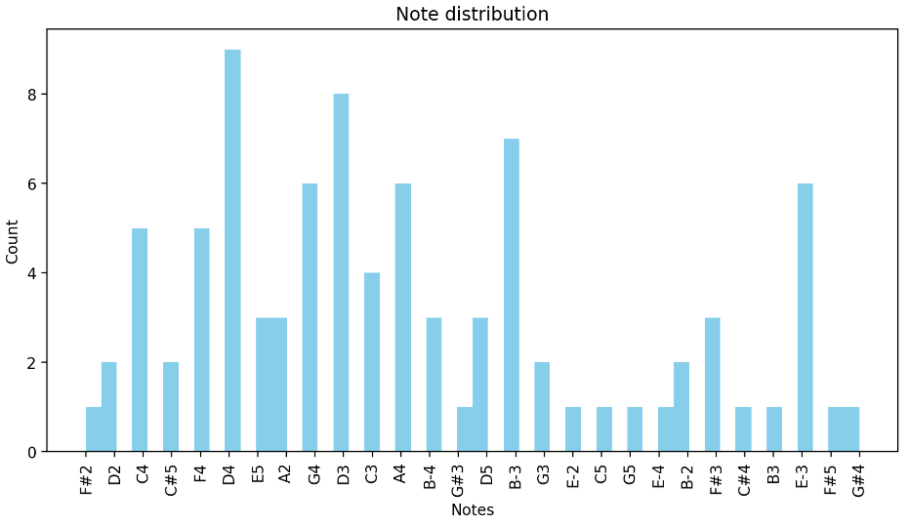

Figure 12: Note Distribution

In Figure 12, we can see that the Counter object indicates that 'D4' is the most frequent note in the composition, occurring 9 times, followed closely by 'D3', and 'B-3'. These notes play a central role in the composition's melodic and harmonic structure.

Pitch Distribution:

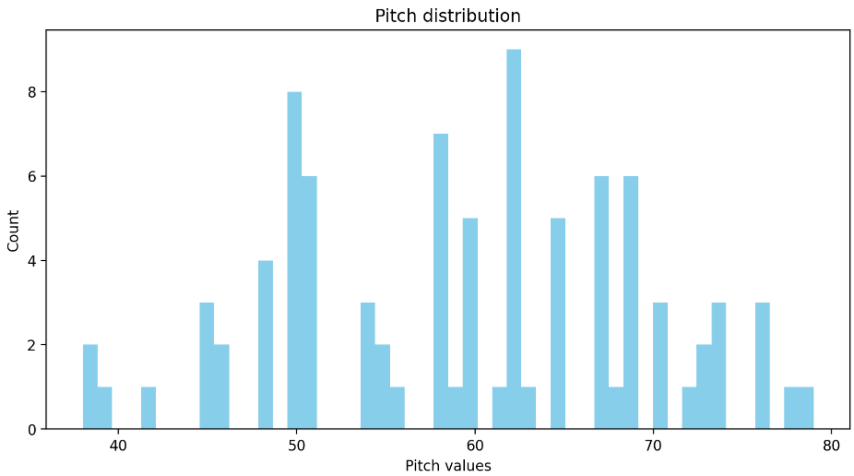

Figure 13: Pitch Distribution

In Figure 13, we can see that the pitch distribution spans from roughly 38 Hz to 79 Hz. This is a broad range, indicating a mix of low, middle, and high pitches, contributing to the piece's dynamic contrast and tonal richness. The most frequent pitches are around 50.3 Hz and 66.7 Hz (approximately 'D3' and 'E4' in the MIDI note range), which ties in well with the note distribution.

Chord Distribution:

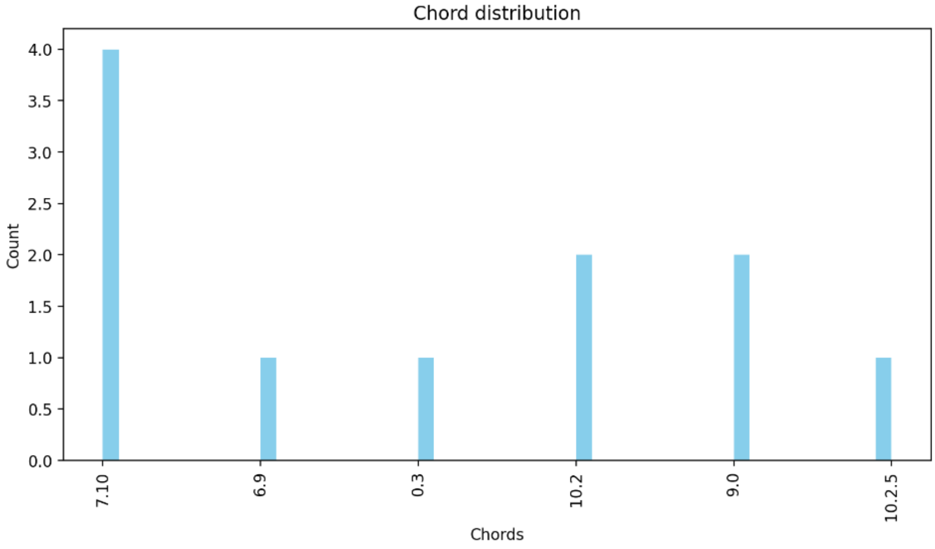

Figure 14: Chord Distribution

In Figure 14, we can see that the model has primarily used the chords '7.10', '10.2', and '9.0'. In music theory, these numbers refer to pitch class sets, which represent a particular group of pitches and the intervals between them. These three chords seem to be the primary harmonic structures used in the model generated piece. The usage of these specific chord combinations may reflect a learned understanding from Schumann's compositions, who is known for his sophisticated and often unconventional harmonic structures.

Unique Notes:

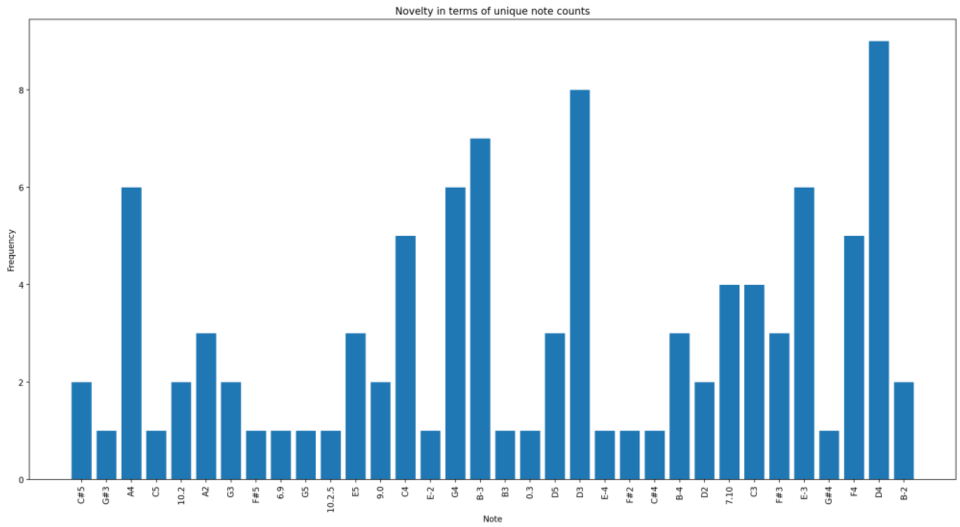

Figure 15: Unique Notes

In Figure 15, we can see that the model generating piece exhibits a wide range of notes (from E-2 to F#5), featuring most frequently D4, D3, and B-3, with some notes used infrequently for emphasis. This expansive use of notes, including accidentals, aligns with Schumann's compositional style. The chord frequencies show a particular harmonic preference, reflecting Schumann's distinctive harmonies. The model shows a nuanced use of notes and chords, which echoes Schumann's varied and expressive style.

Cosine similarity

Similarity analysis was conducted between the model generating piece and the original corpus. This method involves converting the MIDI file to a sequence of notes, converting this to a string, and then computing cosine similarity with the original corpus. The novelty score was then derived by subtracting the maximum similarity from one, thus providing a quantitative measure of the novelty of the model generating music.

	Score
Similarities	0.897
Novelty score	0.103

Table 1: Cosine similarity and Novelty score

In Table 1, we can see that the cosine similarity score of 0.897 suggests that the model generating tune has a high similarity to Schumann's music. This is expected as the model generating was trained on Schumann's compositions, and thus it would learn and replicate the musical structures and elements present in his music.

The novelty score, on the other hand, is 0.103. This suggests that approximately 10.13% of the model generating tune is different from Schumann's music. This difference could be due to new musical ideas, structures, or elements that were not present in the training data. It also indicates that the model can reproduce new and unique musical ideas based on imitating the training data.

Listening test

I have sent the generated music files to Xiaochi Li. After he listened to it, I provided him with some leading questions, and asked him to evaluate the generated piece comprehensively based on these questions.

a) Melody Analysis: How do pitch and rhythm interact in a melody? Are there any recurring themes or phrases? Are there any developments or changes to these patterns?
b) Harmony analysis: What chords are used in the piece? How do these chords work? Are there any common chord progressions? How does the harmony support the melody?
c) Rhythmic analysis: what rhythmic patterns were used? How do these patterns interact with melody and harmony?
d) Form analysis: What is the overall structure of the work? Are there any repetitions or variations?

His answer is as follows: *"Regarding this generated music, my first impression is that there are a lot of dissonant notes, and it seems to sound a bit melody-less. Moreover, each note feels like a beat, and in the overall structure, there is no repetition. Haha, anyway, it has a bit* *of Schumann's shadow. The tune is still very innovative and has a shadow of modern atonal music. It would work extremely well as background music in a haunted house!"*

Conclusion

Through a systematic analysis of the model generating piano piece in the style of Schumann, we find that as a composition, it satisfies the basic requirements of music theory. It absorbs elements of Schumann's musical style, such as the G minor key signature, unconventional chord progressions, a varied note range, and a mixture of small and large melodic intervals. Moreover, it manages to produce music in a style of its own. From my personal listening experience, my comments align with those of Xiaochi Li. The piece is suitable for use as background music in a haunted house. I am grateful to Xiaochi Li for his insightful observations.

However, due to the limitations of the small training set, there still exists a gap between artificial intelligence and human creativity. The music generated this time represents but a small step in the realm of computational creativity in music. At present, if a machine is to generate music with aesthetic appeal, it is worth exploring further the intersections of creativity, music, and artificial intelligence.

Reference

Runco, M.A. & Jaeger, G.J. 2012, "The Standard Definition of Creativity", Creativity research journal, vol. 24, no. 1, pp. 92-96.

Kaufman, J.C. & Sternberg, R.J. 2010, The Cambridge handbook of creativity, Cambridge University Press, Cambridge.

Batey, M. 2012, "The Measurement of Creativity: From Definitional Consensus to the Introduction of a New Heuristic Framework", Creativity research journal, vol. 24, no. 1, pp. 55-65.

Wiggins, G.A. 2006, "A preliminary framework for description, analysis and comparison of creative systems", Knowledge-based systems, vol. 19, no. 7, pp. 449-458.

McCormack, J., Gifford, T., Hutchings, P., Llano Rodriguez, M.T., Yee-King, M. & d'Inverno, M. 2019, "In a Silent Way: Communication Between AI and Improvising Musicians Beyond Sound", ACM, , pp. 1.

Jordanous, A. & Keller, B. 2016, "Modelling Creativity: Identifying Key Components through a Corpus-Based Approach", PloS one, vol. 11, no. 10, pp. e0162959-e0162959.

Sadie, S., Tyrrell, J. & Grove, G. 2001, The new Grove dictionary of music and musicians, 2nd edn, Macmillan Reference, London.

Goodfellow, I., Bengio, Y. & Courville, A. 2017, Deep learning, MIT Press, London;Cambridge, MA;.

Briot, J., Hadjeres, G. & Pachet, F. 2019, Deep Learning Techniques for Music Generation -- A Survey, Cornell University Library, arXiv.org, Ithaca.

Rakshit, S. (2019) Classical Music Midi. Available at: https://www.kaggle.com/datasets/soumikrakshit/classical-music-midi (Accessed: 20 May 2023).
Kapoor, K. (2021) Music Generation LSTM. Available at: https://www.kaggle.com/code/karnikakapoor/music-generation-lstm/notebook (Accessed: 20 May 2023).
Cook, N. 1994, A guide to musical analysis, Oxford University Press, Oxford.
Krumhansl, C.L. & ProQuest (Firm) 1990;2001;2010;, Cognitive foundations of musical pitch, Oxford University Press, New York.