

UNIVERSIDAD DEL VALLE DE GUATEMALA

CC3045 - Informe de Proyecto

Sección 10

Ing. Alberto Suriano



Deep Learning y Sistemas Inteligentes

- Proyecto Final -

Nombre de Integrante(s)

Kenneth Galvez 20079

Esteban Aldana 20591

Jessica Ortiz 20192

GUATEMALA, 13 de Noviembre de 2023

Índice:

Descripción del problema.....	3
Análisis.....	4
Propuesta de solución.....	5
Descripción de la solución.....	5
Herramientas aplicadas.....	6
Resultados (métricas).....	7
Conclusiones:.....	8
Bibliografía:.....	8
Anexos:.....	9

Descripción del problema

La creación de música es una forma de expresión artística compleja que ha evolucionado a lo largo de la historia. Históricamente, la composición musical ha sido una habilidad exclusiva de los seres humanos, pero con los avances en inteligencia artificial y aprendizaje profundo, ahora es posible automatizar la generación de música. En este proyecto, nos enfocamos en el uso de las Redes Generativas Adversarias (GANs) para generar secuencias musicales que sean coherentes y estéticamente agradables.

La música se compone básicamente de notas y acordes organizados en el tiempo. Los archivos MIDI son una representación digital de estas estructuras musicales, donde cada nota y acorde se codifica como un evento en una secuencia temporal. El desafío consiste en entrenar un modelo capaz de comprender la complejidad y coherencia de estas secuencias para generar música nueva y original.

El conjunto de datos utilizado contiene una amplia variedad de estilos y géneros musicales. Cada archivo MIDI en este conjunto proporciona información detallada sobre las notas, acordes e instrumentos utilizados en la composición. Nuestro objetivo es utilizar esta información para entrenar un modelo GAN que pueda aprender los patrones y estructuras musicales, permitiendo así crear nuevas composiciones automáticamente.

La industria musical y la inteligencia artificial son entidades que a diferencia de lo que muchos creen, siempre deberían de ir de la mano tal y como lo demostro la artista canadiense de música pop-electrónica Grimes quien es quizás la artista más famosa y la que más se ha pronunciado a favor de la inteligencia artificial. Lanzó una herramienta de IA llamada Elf.Tech, diseñada para permitir a los usuarios crear musica utilizando su voz (Reiff, 2023).

Análisis

El problema se aborda utilizando archivos MIDI como conjunto de datos de entrada. Los archivos MIDI son archivos digitales que contienen información de secuencias de comandos que crean sonido, en lugar de información de audio. Estos archivos contienen una serie de comandos que describen cómo se debe reproducir una pieza musical, como qué notas tocar, cuándo tocarlas y con qué instrumento. Los archivos MIDI se pueden manipular de muchas maneras, como cambiar la clave, la instrumentación o el tempo de una pieza musical, e incluso editar notas individuales (*Classical Archives: MIDI*, 2023).

Para el modelado de nuestro proyecto se optó por utilizar GAN que es un marco de aprendizaje automático utilizado para la modelización generativa. Está compuesto por dos redes neuronales, un generador y un discriminador, que compiten entre sí en un juego de suma cero para mejorar la precisión de sus predicciones. El modelo generador se entrena para generar nuevos ejemplos, mientras que el modelo discriminador intenta clasificar los ejemplos como reales o falsos. Los dos modelos se entrenan juntos hasta que el modelo discriminador se engaña aproximadamente la mitad del tiempo, lo que significa que el modelo generador está generando ejemplos plausibles (Brownlee, 2019).

La representación de datos fue un paso crucial en el procesamiento de información musical. En este proyecto, se utilizó secuencias de notas y acordes para entrenar el modelo GAN. La elección de representar la música como secuencias de enteros permite al modelo aprender patrones y relaciones entre notas de manera secuencial. La notación de las notas y acordes se ajusta y corrige para garantizar la consistencia y la interpretación adecuada.

También es importante considerar los desafíos específicos de representar música. La música tiene una estructura temporal única, con patrones rítmicos y melódicos que pueden variar significativamente. La interpretación musical puede depender de la elección del instrumento y otros aspectos expresivos. En consecuencia, el modelo debe capturar no solo las relaciones de nota a nota, sino también las complejidades de la estructura musical y la expresión artística.

Es importante destacar que la generación de música con inteligencia artificial plantea cuestiones éticas y artísticas. La música tiene un componente humano único que va más allá de la mera secuencia de notas, y el uso de algoritmos para generar música plantea preguntas sobre la creatividad y la autenticidad artística. Además, el respeto a los derechos de autor y la originalidad deben ser considerados al trabajar con conjuntos de datos musicales existentes.

Propuesta de solución

Un modelo con las siguientes características:

- Un modelo que aborde la corrección de la notación de las notas y acordes presentes en los archivos MIDI, con la finalidad de que la notación se ajuste garantizando uniformidad y coherencia. Y que cree un diccionario que asigna índices a notas y acordes, lo que facilite la representación numérica y la entrada al modelo GAN.
- Un modelo GAN que se componga de un generador y un discriminador. El generador utilizara capas LSTM para aprender patrones temporales en las secuencias de notas, incorporando capas de Dropout para mejorar la generalización y evitar el sobreajuste. El discriminador, evaluara la autenticidad de las secuencias, proporcionando una señal de retroalimentación al generador. Ambos modelos se entrenados de manera adversarial, buscando mejorar continuamente la calidad de las secuencias generadas.

Descripción de la solución

La solución implementa un flujo de trabajo completo que comienza con la carga y procesamiento de archivos MIDI. La notación de notas y acordes se corrige para asegurar una interpretación consistente. Se utiliza la librería “music21” para analizar la estructura de los archivos MIDI y extraer información relevante. Se crea un diccionario que asigna índices a las notas y acordes, facilitando la representación numérica de las secuencias.

El modelo GAN consta de un generador que crea nuevas secuencias de notas y un discriminador que evalúa la autenticidad de las secuencias. Durante el entrenamiento, el generador y el discriminador se actualizan de manera adversarial, buscando mejorar continuamente sus capacidades como se menciono anteriormente.

Para la generación de música, se utiliza el generador entrenado con ruido aleatorio como entrada. Las secuencias generadas se decodifican utilizando el diccionario corregido, y se convierten en un archivo MIDI reproducible. La conversión de las secuencias a archivos WAV permite una evaluación auditiva de las composiciones generadas.

Herramientas aplicadas

1. FluidSynth: Conversión de MIDI a Audio

FluidSynth es un sintetizador de software de código abierto que convierte datos de notas MIDI en una señal de audio utilizando la tecnología SoundFont, sin necesidad de una tarjeta de sonido compatible con SoundFont. FluidSynth puede actuar como un dispositivo MIDI virtual, capaz de recibir datos MIDI de cualquier programa y transformarlo en audio sobre la marcha (*FluidSynth | Software Synthesizer Based on the SoundFont 2 Specifications*, 2023). En este proyecto, se descargó una SoundFont específica que contiene muestras de instrumentos musicales y se utilizó para convertir las secuencias de notas generadas en archivos de audio reproducibles.



2. Music21: Manipulación y Análisis de Datos Musicales

Music21 es una biblioteca de Python que se utiliza para analizar, buscar y transformar música en formas simbólicas (basadas en partituras). Music21 se basa en marcos y tecnologías preexistentes, como Humdrum, MusicXML y MIDI, pero utiliza un marco orientado a objetos que facilita el manejo de datos complejos (*CuthbertLab*, 2023). En este proyecto, se utiliza para cargar y analizar archivos MIDI, así como para corregir la notación de las notas y acordes. Además, facilita la creación de secuencias de notas y acordes a partir de la salida generada por el modelo.



3. Pygame: Reproducción de Archivos WAV en Jupyter Notebook

Pygame es una biblioteca de Python de código abierto y multiplataforma diseñada para escribir videojuegos y aplicaciones multimedia (Fincher, 2019). Pygame se utilizó para reproducir archivos de audio WAV directamente dentro del entorno de Jupyter Notebook.



4. Midi2audio: Conversión de Archivos MIDI a Audio

Midi2audio es una biblioteca de Python que se utiliza para convertir archivos MIDI en archivos de audio, como WAV o MP3. Midi2audio utiliza la biblioteca FluidSynth para sintetizar los datos MIDI en una señal de audio (Bzamecnik, 2022). Se utilizó en combinación con FluidSynth para convertir las secuencias generadas por el modelo en archivos WAV reproducibles.



Resultados

Durante el entrenamiento, se monitorea el progreso mediante las pérdidas del generador y del discriminador. Se utilizan métricas como la binary cross-entropy para evaluar la capacidad del generador para engañar al discriminador y la capacidad del discriminador para distinguir entre secuencias reales y generadas. En la generación final de música, se evalúan subjetivamente los resultados. Aunque el modelo puede aprender patrones y estructuras, la calidad de la música generada puede variar. Es posible que las métricas de pérdida no capturen completamente la percepción subjetiva de la calidad musical.

Evaluación auditiva

El archivo de audio final resultante de la generación del modelo presenta algunos indicios de ritmo y melodía, pero se percibe más como ruido debido al sample de piano utilizado. Esto puede deberse a varias razones:

- Complejidad del modelo: Si bien la arquitectura del modelo GAN puede aprender patrones y estructuras, la generación de música es inherentemente compleja. Puede que el modelo no haya capturado completamente la coherencia melódica y rítmica.
- Limitaciones del conjunto de datos: La calidad de los resultados también depende en gran medida de la calidad y diversidad del conjunto de datos utilizado para el entrenamiento. Puede que el conjunto de datos careciera de variabilidad musical, lo que provocó que el modelo tuviera dificultades para generar secuencias musicales más complejas.
- Hiperparámetros del modelo: Los resultados también están influenciados por la elección de hiperparámetros, como la arquitectura específica del modelo, la tasa de aprendizaje y la duración del entrenamiento.
- Elección del sample: La elección del sample pudo haber tenido un impacto significativo en la percepción de la generación. Algunos samples pueden sonar más realistas y expresivos que otros.

Conclusiones:

- ❖ La generación de música con inteligencia artificial presenta desafíos únicos y complejidades. A pesar de los avances en el uso de GANs, la coherencia melódica y rítmica sigue siendo un área desafiante. La complejidad inherente a la música puede resultar difícil de capturar completamente, y la elección de parámetros y conjuntos de datos resulto ser crucial para mejorar los resultados.
- ❖ La representación adecuada de los datos musicales, en este caso utilizando archivos MIDI y corrigiendo la notación de notas y acordes, es esencial para el éxito del modelo. La utilización de un diccionario para asignar índices a notas y acordes facilita la entrada numérica al modelo GAN, mejorando la calidad de las generaciones.
- ❖ La generación de música con inteligencia artificial plantea cuestiones éticas y artísticas, especialmente en términos de creatividad y autenticidad. La música tiene un componente humano único que va más allá de las secuencias de notas, y es esencial abordar estas consideraciones al trabajar con algoritmos generativos. Siempre se deben respetar los derechos de autor y garantizar la originalidad al utilizar conjuntos de datos musicales existentes.
- ❖ La convergencia entre la industria musical y la inteligencia artificial es evidente en iniciativas como la herramienta de inteligencia artificial lanzada por la artista Grimes. Estas tendencias sugieren un interés creciente en la exploración de nuevas formas de colaboración entre humanos y algoritmos en la creación musical.

Bibliografía:

- Reiff, N. (2023, August 9). Desde Grimes Hasta Drake: ¿Qué Opinan Los Artistas Famosos Sobre la IA en la Música? Decrypt; Decrypt.
<https://decrypt.co/es/resources/desde-grimes-hasta-drake-que-opinan-artistas-famosos-sobre-ia-musica>
- Classical Archives: MIDI. (2023). Classicalarchives.com.
<https://www.classicalarchives.com/midi.html>
- Brownlee, J. (2019, June 16). A Gentle Introduction to Generative Adversarial Networks (GANs) - MachineLearningMastery.com. MachineLearningMastery.com.
<https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/>
- FluidSynth | Software synthesizer based on the SoundFont 2 specifications. (2023). Fluidsynth.org. <https://www.fluidsynth.org/>
- cuthbertLab/music21: music21 is a Toolkit for Computational Musicology. (2023, June 16). GitHub. <https://github.com/cuthbertLab/music21>
- Jon Fincher, Real Python. (2019, September 16). PyGame: A Primer on Game Programming in Python. Realpython.com; Real Python. <https://realpython.com/pygame-a-primer/>
- Bzamecnik/midi2audio: Play and synthesize MIDI to audio - easy to use Python/CLI API to FluidSynth. (2022, September 28). GitHub. <https://github.com/bzamecnik/midi2audio>