

Composición asistida por red neuronal para piano en jazz

Ismael Medina Muñoz

Microsoft, Redmond,
USA

`ismedina@microsoft.com`

Resumen. La inteligencia artificial ha asumido un papel importante en actividades que anteriormente se consideraban exclusivamente humanas. La IA generativa es un área de investigación vibrante, con un creciente interés en campos de aplicación relacionados con las artes. Esta investigación tiene como objetivo demostrar una red neuronal recurrente para crear nuevas secuencias de n -notas a partir de un conjunto inicial de n -notas, y se aplica un enfoque probabilístico para establecer la duración de cada nota en el conjunto resultante. Esto sirve como asistente en la composición de piano para melodías de jazz, produciendo una representación simbólica de la música en forma de partituras. Al aprovechar las técnicas de aprendizaje automático, la IA puede analizar patrones y estructuras en composiciones de jazz existentes, permitiéndole generar secuencias nuevas y armoniosas. Las secuencias generadas pueden ser luego refinadas y ajustadas por compositores humanos, resultando en un proceso colaborativo entre la IA y la creatividad humana. Esta integración de la IA en el proceso creativo abre nuevas posibilidades para la innovación y exploración musical.

Palabras clave: Recurrent neural network, composición musical, asistente.

Neural Network-Assisted Composition for Piano in Jazz

Abstract. Artificial intelligence has taken on an important role in activities that were previously considered exclusively human. Generative AI is a vibrant area of research, with growing interest in application fields related to the arts. This research aims to demonstrate a recurrent neural network for creating new sequences of n -notes from an initial set of n -notes, and a probabilistic approach is applied to establish the duration of each note in the resulting set. This serves as an assistant in piano composition for jazz melodies, producing a symbolic representation of the music in the form of scores. By leveraging machine learning techniques, AI can analyze patterns and structures in existing jazz compositions, allowing it to generate novel and harmonious sequences. The generated sequences can then be further refined and adjusted by human composers,

resulting in a collaborative process between AI and human creativity. This integration of AI into the creative process opens up new possibilities for musical innovation and exploration.

Keywords: Recurrent neural network, musical composition, assistant.

1. Los datos de entrenamiento de la red neuronal

1.1. Partituras como representación simbólica de la música

Un gran número de partituras producidas con el software libre **MuseScore** están disponibles de forma gratuita. Se descargaron partituras de jazz ejecutadas en piano que posteriormente fueron analizadas programáticamente utilizando la librería **MS3**, desarrollada para Python por Johannes Hentschel [2].

Los datos que las partituras proporcionaron reflejan que un ejecutante de piano utiliza ambas manos para ejecutar piezas de jazz. Esto se representa mediante 2 pentagramas, el superior sirve para describir las notas pulsadas en la clave de Sol por la mano derecha, mientras que el inferior sirve para describir las notas pulsadas en la clave de Fa con la mano izquierda. Se demostró que la mano izquierda tiende a tocar teclas con sonidos más graves mientras que la mano derecha sirve para tocar teclas con sonidos más agudos, con tendencia a pulsar un máximo de 5 teclas al mismo tiempo por cada mano. Para aquellas excepciones, donde cada mano pulsó más de 5 teclas se tomaron sólo las primeras 5 notas de forma ascendente de acuerdo con su valor del Musical Instrument Digital Interface (MIDI), donde los valores menores representan sonidos más graves mientras que los valores mayores representan sonidos más agudos.

A las notas pulsadas al mismo tiempo para cada mano se le denominó como “rebanada” de n -notas. Así pues, la mano izquierda registra que en el pentagrama en clave de Fa (staff 2) pulsa 5 notas a lo más al mismo tiempo, mientras que la mano derecha registra que en el pentagrama en clave de Sol (staff 1) pulsa 5 notas a lo más al mismo tiempo (vea la figura 1). Una rebanada de 0-notas en ambas manos representa el silencio en la música.

Las notas pulsadas están descritas por su valor MIDI (vea la figura 2). Estas rebanadas de n -notas de la mano derecha e izquierda pulsadas al mismo tiempo se codificaron para crear el vocabulario, como si de palabras de algún lenguaje se trataran y su representación se da por un identificador numérico.

1.2. Secuencias de rebanadas"de n -notas como palabras de una oración

Con las rebanadas de n -notas ya codificadas con un identificador fue ya posible codificar también secuencias de rebanadas de n -notas como si se trataran de palabras en una oración. Para esta investigación se generaron secuencias de 3 palabras (vea la figura 3).

Cabe aclarar que se removieron rebanadas subsecuentes que representan a la misma palabra para evitar que la red neuronal aprenda a producir secuencias repetitivas de las mismas notas.

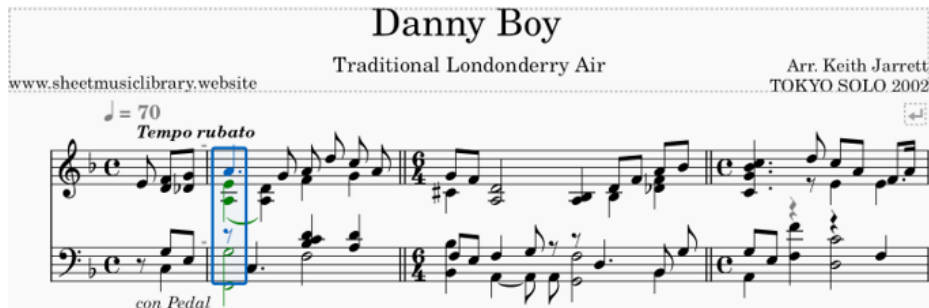


Fig. 1. Rebanada"de 3-notas para la mano derecha y de 2-notas para la mano izquierda.

staff01_note01	staff01_note02	staff01_note03	staff01_note04	staff01_note05	staff02_note01	staff02_note02	staff02_note03	staff02_note04	staff02_note05
0	0	0	0	0	0	0	0	0	0
71	0	0	0	0	0	0	0	0	0
71	0	0	0	0	36	48	0	0	0
0	0	0	0	0	31	43	0	0	0
0	0	0	0	0	36	48	0	0	0

Fig. 2. Rebanada"de n -notas para cada mano (staff 1 o staff 2) y representación de las teclas pulsadas con su valor MIDI

prev_n_notes_id	actual_n_notes_id	next_n_notes_id	next_n_notes_id
0	0	1	0
1	0	2	3
2	2	3	4
3	3	4	3
4	4	3	4
...
78846	179	18226	0
78847	18226	179	311
78848	311	178	18228
78849	178	18228	0
78850	18228	0	0

Fig. 3. Secuencias únicas de 3 rebanadas"de n -notas.

1.3. La duración de las notas y la función de masa de probabilidades

El proceso de construcción del vocabulario también produjo información fuertemente asociada a la nota pulsada en cada pentagrama, la duración de dichas notas. La duración de cada nota se contabilizó para producir distribuciones de probabilidad de las notas de acuerdo con la cantidad de n -notas pulsadas por

Tabla 1. Diferencia de distribución de rebanadas de n -notas.

Métrica	Staff 1	Staff 2
2/2	p-value = 0.000, K-S 2c = 0.4269	p-value = 0.000, K-S 2c = 0.8958
2/4	p-value = 0.000, K-S 2c = 0.4164	p-value = 0.000, K-S 2c = 0.8958
4/4	p-value = 0.000, K-S 2c = 0.2564	p-value = 0.000, K-S 2c = 0.8958
3/4	p-value = 0.000, K-S 2c = 0.3010	p-value = 0.000, K-S 2c = 0.8958
5/4	p-value = 0.000, K-S 2c = 0.4988	p-value = 0.000, K-S 2c = 0.8958

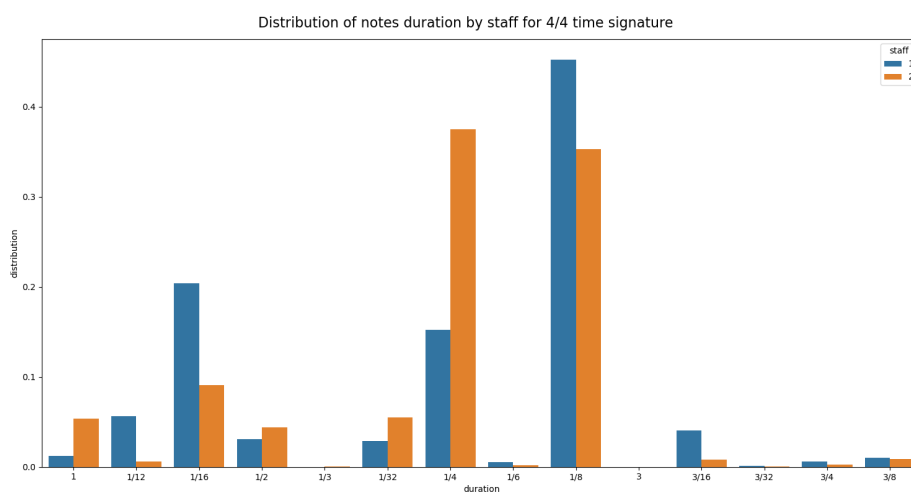


Fig. 4. Distribución de la duración de las notas por cada mano (staff 1 y 2) para una pieza en métrica de 4/4

cada mano, agrupadas por la métrica del pentagrama. Estos datos produjeron funciones de masa de probabilidad de donde se toman valores aleatorios que complementan a las notas producidas por la red neuronal (vea la figura 4).

2. El modelo resultante

Las secuencias de rebanadas de n -notas como palabras de una oración fueron codificadas como tensores de entrada. Este modelado de secuencias y pilas para entrenar una red neuronal recurrente que produce secuencias fue tomado de las ideas de [1], como es similar a secuencias de palabras en el procesamiento de lenguaje natural (vea la figura 5).

Se entrenaron dos redes neuronales, una basada en arquitectura LSTM y otra basada en arquitectura GRU. La red con arquitectura GRU se desempeñó mejor durante el entrenamiento y esta fue utilizada para producir nuevas secuencias de n -notas (vea la figura 6).

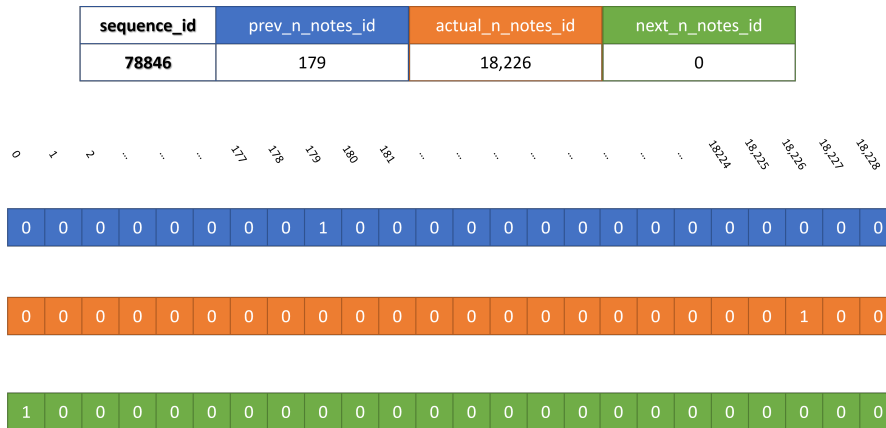


Fig. 5. Descripción del tensor de entrada que presenta un formato de secuencias de palabras

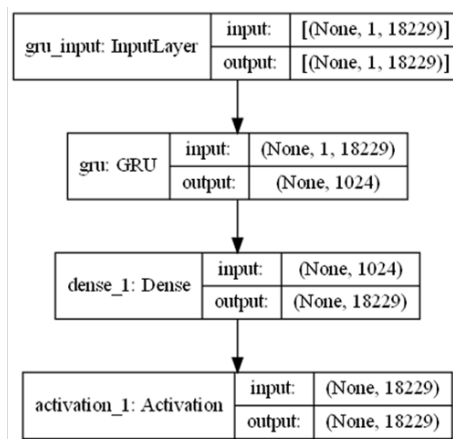


Fig. 6. La arquitectura de la red neuronal entrenada

3. Resultados y conclusiones

Con el modelo elegido, se generaron 100 nuevas secuencias de 64 rebanadas de n -notas con duraciones de nota tomadas de forma aleatoria de las distribuciones de masa de probabilidad construidas en el experimento. Dichas secuencias fueron representadas en cinco métricas de partitura que luego fueron analizadas y comparadas contra el conjunto de partituras usadas para el entrenamiento utilizando la prueba de Kolmogórov-Smirnov de 2 colas (estadístico K-S 2c).

Ismael Medina Muñoz



Fig. 7. Archivo PNG creado utilizando la respuesta JSON de la REST API del modelo

Los resultados muestran que las nuevas piezas son estadísticamente diferentes respecto de las partituras de entrenamiento (vea la tabla 1).

El modelo fue desplegado como una Representational State Transfer Application Programming Interface (REST API) que produce secuencias de n -notas con

sus duraciones en formato Comma-separated values (CSV), Extensible Markup Language (XML), MIDI y Portable Network Graphics (PNG). La representación PNG es útil para los ejecutantes de piano (vea la figura 7).

Esto constituye a la red neuronal definida como un asistente para la composición musical de piezas de jazz que brinda ideas nuevas para la composición.

Referencias

1. Goldberg, Y.: *Neural Network Methods for Natural Language Processing*. pp. 164–166. Springer International Publishing, Cham (2017)
2. Manual, <https://johentsch.github.io/ms3/build/html/manual/index.html>, last accessed 2023/02/17