Deep Learning for Music

Allen Huang e Raymond Wu
Stanford University

Aluna: Leticia Silvério Moreira

Objetivo - Introdução

O objetivo é ser capaz de construir um modelo generativo a partir de uma arquitetura de rede neural profunda para tentar criar música que tem harmonia e melodia e é passável como música composta por humanos. Anterior trabalho em modelagem de música polifônica centrou-se em torno da estimativa de densidade de probabilidade de séries temporais. Hoje, existem muitos trabalhos baseados em Redes Neurais Recorrentes combinadas com Máquinas Boltzmann restritas (RNN-RBM) e as máquinas de Boltzmann restritas Recorrente Temporal (RTRBM) apresentam trabalhos bem sucedidos. A abordagem do artigo, no entanto, é realizar aprendizado e geração de ponta a ponta com profundidade redes neurais sozinhas.

Dados

Um dos principais desafios na formação de modelos para geração de música é escolher os dados corretos para a representação. Para este trabalho foram escolhidos dois tipos principais:

arquivos midi com pré-processamento mínimo e uma representação de “piano roll” de arquivos midi.

Midi data

Os arquivos Midi são estruturados como uma série de faixas, cada contendo uma lista de meta mensagens e mensagens. Foram extraídas as mensagens referentes às notas e suas durações e em seguida, a mensagem foi toda codificada como um único símbolo. Foram achatados os trilhos para que os tokens das faixas separadas de uma peça seriam concatenadas de ponta a ponta.

Corpus	Words	Unique Tokens
Bach Only	1,663,576	35,509
Full Classical	24,654,390	175,467
Truncated Classical	11,413,884	132,437

Muse Piano Roll Data

Os arquivos midi foram representados como uma série de etapas de tempo. Cada etapa de tempo é uma lista de IDs de notas que estão sendo reproduzidas. O conjunto de dados MuseData tinha 524 peças para um total de 245.202 etapas de tempo. Foram codificados cada passo de tempo concatenando os ids da nota juntos para formar um token. Por exemplo, foi codificado um acorde C-Major como "60-64-67". Além disso, a fim de reduzir o número de tokens, foram escolhidas aleatoriamente 3 notas quando a polifonia excedeu 4.

Dataset	Unique Tokens
Muse-All	39,289
Muse-Truncated	21,510

Experimento Bach-Midi

Foi usada uma memória de longo prazo de 2 camadas (LSTM) arquitetura da rede neural recorrente (RNN) no Conjunto de dados "Bach Only". A saída do LSTM é alimentada na camada softmax com uma entropia cruzada correspondente função objetiva. 50 épocas levaram cerca de 4 horas para treinar em uma instância do AWS g2.2xlarge.

Hidden State	128
Token Embedding Size	128
Batch Size	50
Sequence Length	50

Experimento Midi-Clássico

Foi usada a mesma arquitetura do experimento Bach-Midi no conjunto de dados "Truncated Classical" devido às limitações de tempo. 15 épocas levaram 22 horas de para serem processadas. Além disso, devido a limitações na memória do dispositivo no g2.2xlarge da AWS, foram forçados a reduzir o tamanho do lote e a sequência comprimento.

Hidden State	128
Token Embedding Size	128
Batch Size	25
Sequence Length	25

Muse Piano Roll Experiments

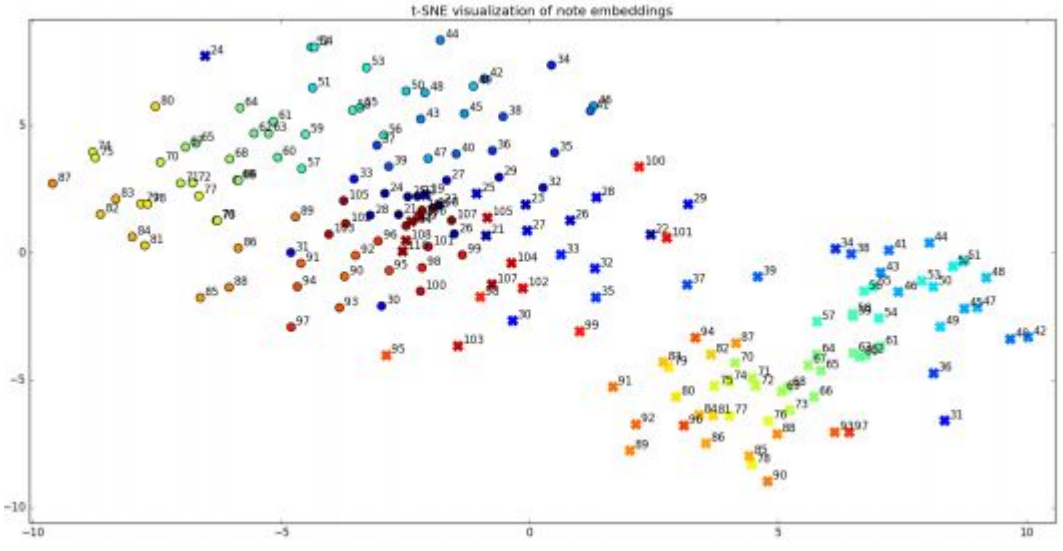
Este mesmo experimento foi realizado com os mesmos parâmetros que o "Experimento Bach-Midi". Nós o rodamos com 800 épocas, que levaram sete horas em uma instância do AWS g2.2xlarge. E também foi executada a mesma configuração no conjunto de dados truncado por 100 épocas, o que levou 7 horas em uma CPU.

Piano-Roll Sigmoid Model

Em vez de codificar uma lista de notas para cada etapa de tempo como um token, cada nota teve seu próprio vetor de incorporação. Para cada etapa de tempo, o vetor de entrada seria uma soma desses vetores. A saída de o LSTM é projetado de volta para o espaço de entrada e alimentado em uma camada sigmóide. A função objetivo é entropia cruzada padrão.

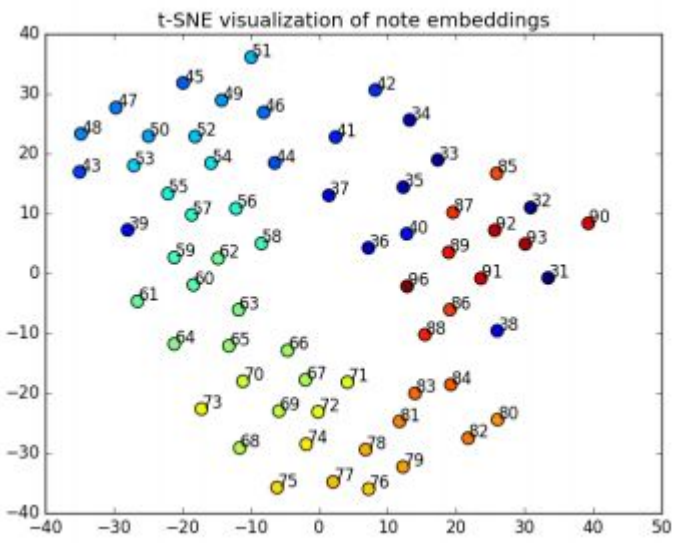
Resultados

Visualização de t-SNE de vetores de incorporação de nota única do experimento Clássico-Midi. Os círculos indicam ON mensagens enquanto os x's representam mensagens OFF. Note que há clusters claros entre as mensagens on e off para o notas de média frequência (as notas tocadas com mais frequência), enquanto as notas azuis e vermelhas correspondem às baixas e altas as notas estão agrupadas em uma nuvem indistinta no centro.



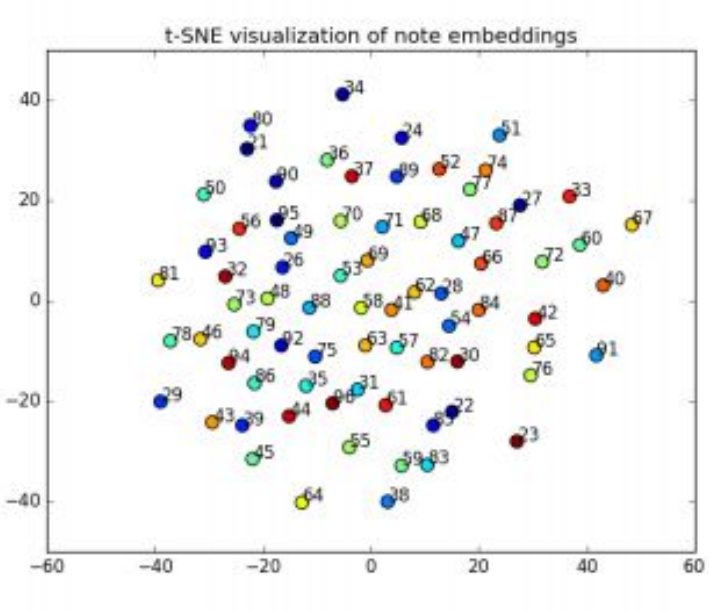
Resultados

Visualização de t-SNE de vetores de incorporação de nota única do experimento Muse-Piano-Roll. Note que os vetores são capa para desambiguar entre as notas baixa, média e alta



Resultados

Visualização de t-SNE de vetores de inclusão de nota única do experimento do Modelo Sigmóide de Piano-Roll. Parece que nosso modelo sigmóide não foi capaz de aprender efetivamente, o que é coincidentemente refletido na música gerada.



Conclusão

A análise inicial dos vetores de incorporação sugere que tiveram sucesso moderado na produção de um modelo genérico viável.