HW3
Symbolic Music Generation

B10901024 謝翔

Google drive link: https://drive.google.com/drive/folders/12ZleSjK9j-BGwrF1nYA6wxP2CHYJ4UbM?usp=share_link

Outline

- Highlight
- Methodology Detail
- Result and analysis
- Takeaway

Highlight: Methodology

- Model
  - GPT2 (12 head, 12 layer)
  - TransformerXL (12 head, 12 layer)/(6 head, 6 layer)
- Dataset
  - POP1K7 (window size = 1024, step size = 512)
  - About 110K data
  - Use REMI (in miditok) to tokenize

Highlight: Result

- Far from applicable. Easy to capture rhythm but hard to capture chord
- With the same amount of head and layer, GPT2 (with only decoder) and TransformerXL (with both encoder and decoder) have similar performance.
- The amount of head and layer plays a crucial role on music quality

Finding

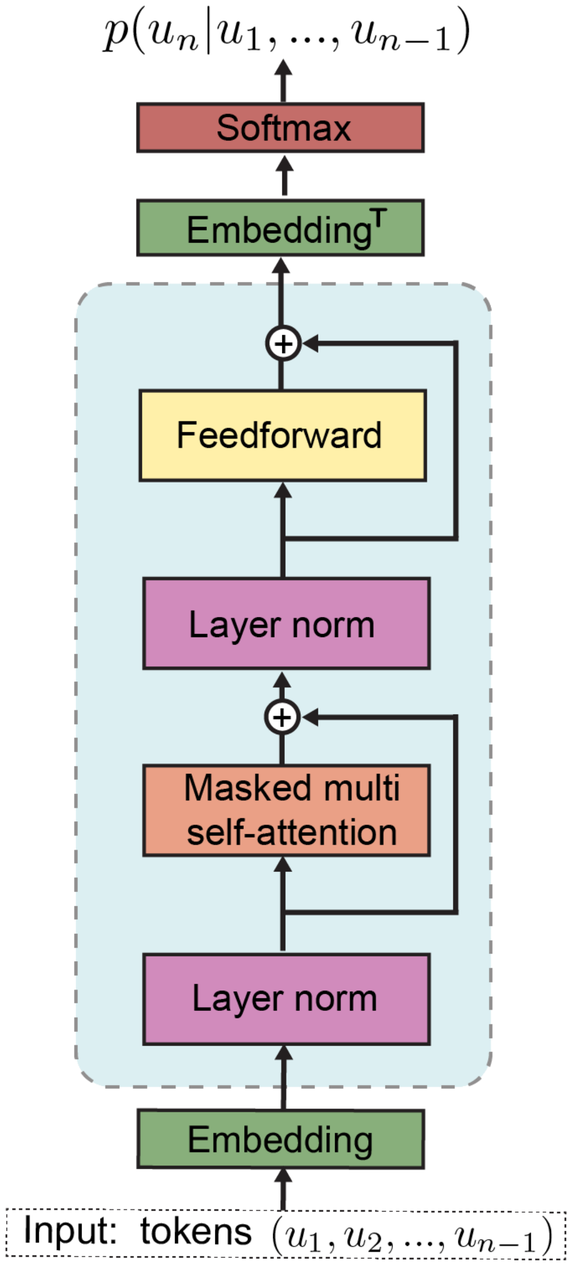
- For transformer based model, only GPT (decoder only) is sufficient
- Finer music model needs more data amount, more head and more layer
- Chord information helps!

Methodology Detail
Model

- Model (and parameter)
  - GPT2 (12 head, 12 layer): 124M
  - TransformerXL (12 head, 12 layer): 188M
  - TransformerXL (6 head, 6 layer): 102M

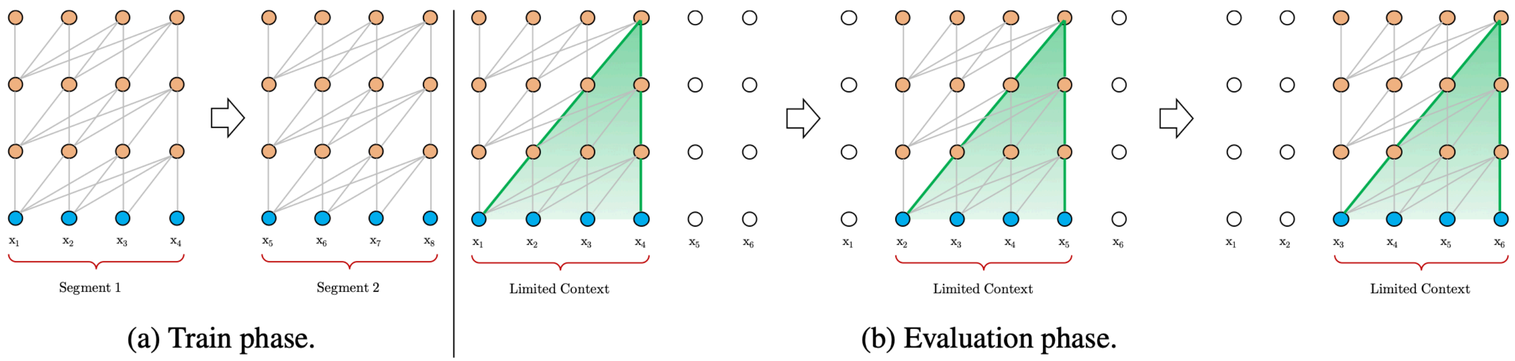
Methodology Detail
GPT2

- Unilateral
- Only utilize decoder, no cross attention
- 124M (With default settings)



Methodology Detail
TransformerXL

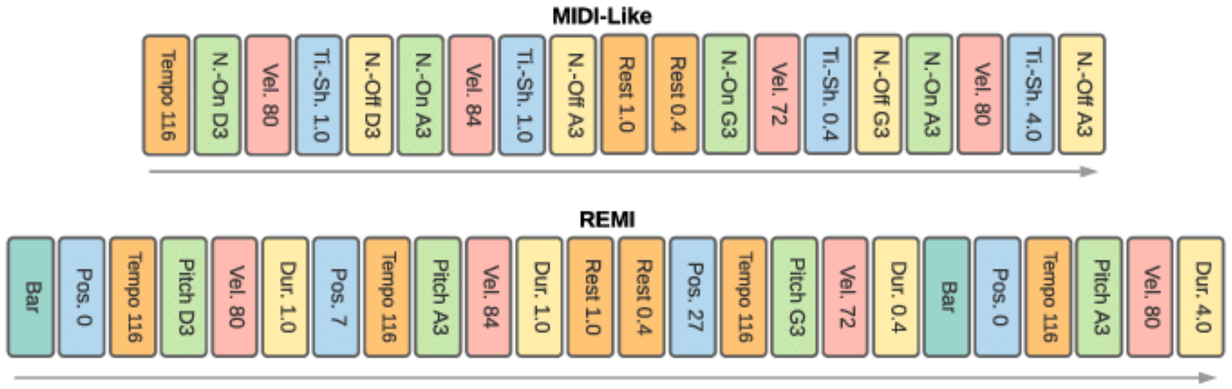
- A bigger transformer (encoder + decoder) to handle long sequence
- Utilize segment memory to for computation efficiency
- 257M (With default settings)



Methodology Detail

Dataset preprocess

- REMI tokenizer in MidiTok
- Token type: special token bar, pitch, velocity, duration, position, pitchDrum (282 tokens)
- Preprocessing: window size = 1024, step size = 512



Methodology Detail

Sampling Policy

- Top K = 50, Top P = 0.9, Temperature = 1
- To make the tokens decodable, The model shall strictly follows pitch-duration-velocity format (no other restriction to bar and position token)
- Window size = 128 (to keep computational efficiency)

Result & Analysis

Model (n_heads/n_layers)	Prompt	H4 score	GS score
GPT2 (12/12)	[“start”, “bar”]	1.168	0.970
TransformerXL (6/6)	[“start”, “bar”]	1.822	0.787
TransformerXL (12/12)	[“start”, “bar”]	0.973	0.925
GPT2 (12/12)	Prompt Songs provided by TA	1.497	2.452

Result & Analysis

- Generally, GPT2 and TransformerXL has the same performance with the same number of head and tokens
- Tends to produce the whole note and wierd chord, occasionally pop rhythm
- Can continue a fine prompt with a little flaw, but the flaw might ruin the rest of the midi
- TransformerXL (6/6) can hardly produce the music. The note transformer produce not as much as the other two model
- GPT2 tends to produce stabler rhythm

Takeaway

- The general LM model as large as GPT2 can’t really handle symbolic music generation. However, more head and more layers might help
- Encoder, though increase computation time, doesn’t help.
- Is it promising that one produce chord token every time after position token?