Homework-3
Symbolic Music Generation
Report

Student：潘品齊 (Pin-Chi Pan)

ID：R12942103

Project Page：
https://github.com/PANpinchi/DeepMIR_HW3_PANpinchi

Novelty highlight

Integration of REMI Representation with Chord Tokens:

- Experimented with **w/ chord** and **w/o chord** configurations to explore the impact of harmonic information on music generation.

Custom Sampling Strategies for Diversity Control:

- Used **top-k sampling** and **temperature scaling** to adjust creativity and randomness in generated outputs.

Comprehensive Metric Evaluation:

- Assessed music quality using H1 and H4 (harmonic diversity) and GS (rhythmic consistency), benchmarking results against real-world MIDI sequences.

Methodology Highlight

Model: GPT-2 Transformer

Representation:

- REMI with and without chord tokens.

Training:

- Loss: Cross-Entropy.
- Input Length: 1024 tokens.
- Sampling Strategies: Top-k = 50 or 100, Temperature = 1.0 or 2.0.

Dataset:

- Pop1K7 Dataset, single-track 4/4 time signature MIDI files.

Result Highlight

Task 1: Objective result comparison

Model	representation	event	loss	top-k	temperature	H1	H4	GS
GPT-2	REMI	w/ chord	3.04	50	1.0	**1.8665**	**3.1021**	**0.7291**
GPT-2	REMI	w/ chord	3.04	50	**2.0**	2.1350	3.2684	0.6705
GPT-2	REMI	w/ chord	3.04	**100**	1.0	**2.0059**	**3.1944**	0.6951
GPT-2	REMI	**w/o chord**	3.02	50	1.0	2.2312	3.2594	**0.7000**
GPT-2	REMI	**w/o chord**	3.02	50	**2.0**	2.1677	3.2822	0.6823
GPT-2	REMI	**w/o chord**	3.02	**100**	1.0	2.1388	3.2521	0.6687
Real data						**1.9355**	**2.5273**	**0.7790**

Red indicates the best performance, and **blue** indicates the second-best.

Result Highlight

Task 1: Generated 20 mid/midi files (32 bars)

Model	representation	event	loss	top-k	temperature	.mid/.midi	.wav	
GPT-2	REMI	w/ chord	3.04	50	1.0			**(Best)**
GPT-2	REMI	w/ chord	3.04	50	**2.0**			
GPT-2	REMI	w/ chord	3.04	**100**	1.0			
GPT-2	REMI	**w/o chord**	3.02	50	1.0			
GPT-2	REMI	**w/o chord**	3.02	50	**2.0**			
GPT-2	REMI	**w/o chord**	3.02	**100**	1.0			

Result Highlight

Task 2: Symbolic music continuation (8 + 24 bars)

Model	representation	event	loss	top-k	temperature	.mid/.midi	.wav
GPT-2	REMI	w/ chord	3.04	50	1.0		
GPT-2	REMI	w/ chord	3.04	50	**2.0**		
GPT-2	REMI	w/ chord	3.04	**100**	1.0		
GPT-2	REMI	**w/o chord**	3.02	50	1.0		
GPT-2	REMI	**w/o chord**	3.02	50	**2.0**		
GPT-2	REMI	**w/o chord**	3.02	**100**	1.0		

Result Highlight

Task 2: Generated the best song among different inference configurations.

Model	representation	event	loss	top-k	temperature	Song 1	Song 2	Song 3
GPT-2	REMI	w/ chord	3.04	50	1.0			
GPT-2	REMI	w/ chord	3.04	50	**2.0**			
GPT-2	REMI	w/ chord	3.04	**100**	1.0			
GPT-2	REMI	**w/o chord**	3.02	50	1.0			
GPT-2	REMI	**w/o chord**	3.02	50	**2.0**			
GPT-2	REMI	**w/o chord**	3.02	**100**	1.0			

Findings Highlight

Main Observations:

- **Temperature and Entropy:** Higher temperature settings (e.g., 2.0) resulted in more diverse melodic content (higher H1 and H4), but this came at the cost of rhythmic consistency (lower GS). This trade-off between creativity and structure is crucial for different musical contexts.
- **Impact of Chord Information:** The inclusion of chord tokens provided harmonically richer sequences. However, the "w/o chord" configuration consistently generated rhythmically tighter outputs, emphasizing the role of harmonic constraints in rhythmic fluidity.

Findings Highlight

Main Observations:

- **Real Data as Benchmark:** While generated sequences displayed promising results, they still fell short of the complexity and balance found in real-world MIDI sequences, highlighting areas for further improvement.

Approach Details

Tokenization & Model:

- **Representation:** REMI tokenization included a rich set of tokens such as pitches, velocities, durations, positions, tempos, and chords. This granularity allowed the model to capture intricate musical details.
- **Model Architecture:** GPT-2 with:
  - 12 transformer layers
  - 12 attention heads
  - Embedding size of 768
- **Objective:** Predict next token in the sequence using Cross-Entropy Loss.

Approach Details

Inference Configurations:

- **Top-k Sampling:** Experimented with k values of 50 and 100 to control the diversity of token generation.
- **Temperature Scaling:** Used temperature values of 1.0 (balanced) and 2.0 (creative) for testing.

Approach Details

Inference Configurations:

- **Metrics Used:**
  - **H1 (Short-term Pitch Diversity):** Evaluates the entropy of pitch class histograms for individual bars.
  - **H4 (Long-term Pitch Diversity):** Extends H1's scope to multiple bars for observing compositional evolution.
  - **GS (Groove Similarity):** Measures the rhythmic consistency across bars to assess structural coherence.

Approach Details

Experiments:

- **Chord Token Impact:** Comparison of "w/ chord" vs. "w/o chord" configurations.
- **Temperature and Sampling:** How inference parameters influence harmonic richness and rhythmic stability.

Result Analysis & Discussion

Impact of Temperature Settings:

- At **temperature = 1.0**, sequences displayed balanced entropy and groove, ideal for structured compositions.
- At **temperature = 2.0**, increased randomness led to higher harmonic entropy (H1 and H4) but compromised rhythmic consistency (GS). Suitable for experimental or avant-garde music styles.

Result Analysis & Discussion

Influence of Chord Tokens:

- **"w/ chord":** Generated sequences exhibited better harmonic transitions but occasionally introduced rhythmic irregularities.
- **"w/o chord":** Focused on rhythmic precision, often resulting in sequences that were rhythmically tighter but harmonically less adventurous.

Result Analysis & Discussion

Generated MIDI Quality:

- **Task 1:** Generated MIDI files (32 bars) closely matched real data in terms of pitch diversity but lagged behind in groove similarity. This highlights the challenge of balancing harmonic complexity and rhythmic consistency.
- **Task 2:** In the 8 + 24 bar continuation task, sequences showed smooth transitions but struggled with maintaining thematic consistency over longer sequences.

Result Analysis & Discussion

Limitations:

- Generated MIDI lacks the nuanced variability of real-world compositions.
- Long-term dependencies, such as thematic development across multiple bars, remain challenging despite using transformers.

End

Student：潘品齊 (Pin-Chi Pan)

ID：R12942103

Project Page：
https://github.com/PANpinchi/DeepMIR_HW3_PANpinchi