



2019年

Published as a conference paper at ICLR 2019

MUSIC TRANSFORMER: GENERATING MUSIC WITH LONG-TERM STRUCTURE

Cheng-Zhi Anna Huang* Ashish Vaswani Jakob Uszkoreit Noam Shazeer
Ian Simon Curtis Hawthorne Andrew M. Dai Matthew D. Hoffman
Monica Dinculescu Douglas Eck

Google Brain

{annahuang, avaswani, usz, noam,
iansimon, fjord, adai, mhoffman, noms, deck}@google.com

ABSTRACT

Music relies heavily on repetition to build structure and meaning. Self-reference occurs on multiple timescales, from motifs to phrases to reusing of entire sections of music, such as in pieces with ABA structure. The Transformer (Vaswani et al., 2017), a sequence model based on self-attention, has achieved compelling results in many generation tasks that require maintaining long-range coherence. This suggests that self-attention might also be well-suited to modeling music. In musical composition and performance, however, relative timing is critically important. Existing approaches for representing relative positional information in the Transformer modulate attention based on pairwise distance (Shaw et al., 2018). This is impractical for long sequences such as musical compositions because their memory complexity for intermediate relative information is quadratic in the sequence length. We propose an algorithm that reduces their intermediate memory requirement to linear in the sequence length. This enables us to demonstrate that a Transformer with our modified relative attention mechanism can generate minute-long compositions (thousands of steps) with compelling structure, generate continuations that coherently elaborate on a given motif, and in a seq2seq setup generate accompaniments conditioned on melodies¹. We evaluate the Transformer with our relative attention mechanism on two datasets, JSB Chorales and Maestro, and obtain state-of-the-art results on the latter.



Music Transformer 主旋律 返回

