Evaluating Algorithms that Learn how to Compose Music from Scratch

New long– and short–term evaluation metrics applied to machine learning models

James Owers

A thesis presented for the degree of

Doctor of Philosophy

Supervised by:

Amos Storkey

Mark Steedman

University of Edinburgh, UK

January 2021

I, James Owers, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Evaluating whether creative content generated by a computer is 'good,' be it music, images, or text, is unsolved and not even well defined. We identify a property of music which is not modelled well, and propose new evaluation metrics for music generation which can be used to distinguish between real and generated data, and thus be useful for automatic quantitative analysis of generation quality.

This is interesting because …TODO… and it has implications for …TODO…

Finally, we make recommendations for how to make progress with respect to music generation and related tasks.

Acknowledgements

Interdum et malesuada fames ac ante ipsum primis in faucibus. Aliquam congue fermentum ante, semper porta nisl consectetur ut. Duis ornare sit amet dui ac faucibus. Phasellus ullamcorper leo vitae arcu ultricies cursus. Duis tristique lacus eget metus bibendum, at dapibus ante malesuada. In dictum nulla nec porta varius. Fusce et elit eget sapien fringilla maximus in sit amet dui.

Mauris eget blandit nisi, faucibus imperdiet odio. Suspendisse blandit dolor sed tellus venenatis, venenatis fringilla turpis pretium. Donec pharetra arcu vitae euismod tincidunt. Morbi ut turpis volutpat, ultrices felis non, finibus justo. Proin convallis accumsan sem ac vulputate. Sed rhoncus ipsum eu urna placerat, sed rhoncus erat facilisis. Praesent vitae vestibulum dui. Proin interdum tellus ac velit varius, sed finibus turpis placerat.

Table of Contents

List of Figures

List of Tables

Abbreviations

API	**A**pplication **P**rogramming **I**nterface
JSON	**J**ava**S**cript **O**bject **N**otation
MDTK	**M**idi **D**egradation **T**ool**k**it
SOTA	**S**tate **o**f **t**he **A**rt

Chapter 1

Introduction

1.1 Historical background

…TODO… Give historical background

1.1.1 First instance of music generation

…TODO… From Section 1.2 (Briot et al. 2019)

The first music generated by computer appeared in 1957. It was a 17 seconds long melody named "The Silver Scale" by its author Newman Guttman and was generated by a software for sound synthesis named Music I, developed by Mathews at Bell Laboratories

1.1.2 Mozart using mechanical aids for idea generation

…TODO… From footnote 7 in Section 1.2 (Briot et al. 2019)

> One of the first documented case of stochastic music, long before computers, is the Musikalisches Wurfelspiel (Dice Music) by Wolfgang Amadeus Mozart. It was designed for using dice to generate music by concatenating randomly selected predefined music segments composed in a given style (Austrian waltz in a given key).

1.1.3 Ada lovelace noting computers could generate music

…TODO… From (Hollings et al. 2018) Ada Lovelace, "Sketch of the Analytical Engine invented by Charles Babbage, Esq., by L. F. Menabrea," Scientific Memoirs, vol. 3, ed. Richard Taylor, 1843, pp. 666-731 (this quote on p 694)

> "Note G" is the culmination of Lovelace's paper, following many pages of detailed explanation of the operation of the Engine and the cards, and of the notation of the tables. The paper shows Lovelace's obsessive attention to mathematical details - it also shows her imagination in thinking about the bigger picture.
>
> Lovelace overseed a fundamental principle of the machine, that

> the operations, defined by the cards, are separate from the data and the results. She observed that the machine might act upon things other than numbers, if those things satisfied mathematical rules.
>
> Supposing that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent.

Lovelace also has the Lovelace Test of Creativity attributed to her - see (Ariza 2009).

1.2 Modern interest and achievements

…TODO…

- Imogen Heap: How AI is helping to push music creativity
- This is the AI Song Contest In the AI Song Contest teams of musicians, artists, scientists and developers take on the challenge of creating a new Eurovision-like hit with the help of artificial intelligence.
- Swooshes, Seaboards, Synths and Spawn
- David Rosen and Scott Miles on the Neuroscience of Music and Creativity
- AI Music Generation Challenge 2020 - (Sturm 2020)

1.3 What are algorithms that learn

…TODO… define/introduce machine learning

1.4 What is composing music

…TODO…

1.5 What does it mean to compose from scratch

…TODO… what is the minimum information we supply as a starting point? What feedback do we give?

1.6 Motivation for this work

…TODO…

- why are we focussing on metrics and not human evaluation
- why do we care about 'from scratch'

1.7 List of contributions in this thesis

…TODO…

- I

- Did
- New
- Stuff

Chapter 2

Literature Review

2.1 Models for composing music

…TODO… 1. state of the art generative models for music - what are the problems people currently try to / can solve, and how do they do it 1. Main point is to motivate need for better evaluation 2. How do people address the issue of modelling long term dependencies - this is what we are going to aim to show is poorly evaluated

2.2 Evaluation methods for creative models

…TODO… how to evaluate generative models with a focus on music - how do people evaluate their success

2.3 Methods for representing music on a computer

…TODO… how to represent music data - (in relation to 'from scratch,' what is the minimal information supplied to the models, and is there evidence of what difference it makes (either by experiment or just by reasoning?)

Chapter 3

New metrics for Evaluating Musical Generations

3.1 The midi degradation toolkit

3.2 A phrase-level metric for short-term structure

3.3 A piece-level metric for long-term structure

Chapter 4

Evaluating State-of-the-Art Music-generating Models

4.1 Comparative analysis using new and existing metrics

- Use phrase and piece level metrics to evaluate state-of-the-art models
- Compare and contrast, outlining the issues identified (e.g. meandering, no high-level structure)

4.2 Strengths and shortcomings of existing models

4.3 Avenues for improvement

Chapter 5

A New Model

Potential ideas:

- An improved generative model for music
  - Training like BERT? http://jalammar.github.io/illustrated-bert/
  - Using mdtk for data augmentation in training (negative examples?), making them more robust
  - Alternative training objectives:
    * crossentropy slow and not musically informed
    * can we use something akin to word error rate (this has been done for text)
- Alternative ways to encode music: encoding chords and phrases in a low-rank continuous space
  - Have done some work on this with convnets and generating continuations
    * low rank was enforced by cross-product ing two vecs

– Could investigate effect of different representations for music on performance

Chapter 6

Conclusion

Chapter 7

References

…TODO… check over using https://www.cl.cam.ac.uk/~ga384/bibfix.html

Ariza, C., 2009. The Interrogator as Critic: The Turing Test and the Evaluation of Generative Music Systems. *Computer Music Journal*, 33(2), pp.48–70. Available at: https://www.jstor.org/stable/40301027 [Accessed January 22, 2021].

Briot, J.-P., Hadjeres, G. & Pachet, F.-D., 2019. Deep Learning Techniques for Music Generation – A Survey. *arXiv:1709.01620 [cs]*. Available at: http://arxiv.org/abs/1709.01620 [Accessed January 22, 2021].

Hollings, C., Martin, U. & Rice, A.C., 2018. *Ada Lovelace: The making of a computer scientist*, Oxford: Bodleian Library.

Sturm, B., 2020. The 2020 Joint Conference on AI Music Creativity. *The 2020 Joint Conference on AI Music Creativity*. Available at: https://boblsturm.github.io/aimusic2020/ [Accessed January 22, 2021].