Evaluating Algorithms that Learn how to Compose Music from Scratch

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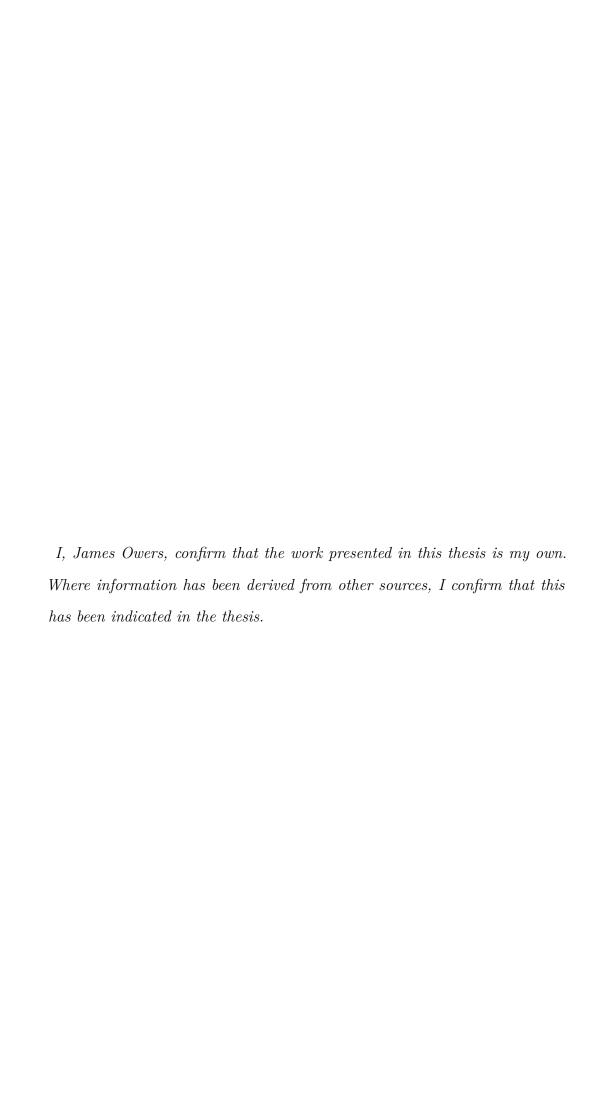
James Owers

A thesis presented for the degree of

Doctor of Philosophy

Supervised by:

Amos Storkey

Mark Steedman

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

5 A New Model

6 Conclusion

7 References

List of Figures

List of Tables

Abbreviations

API   **A**pplication **P**rogramming **I**nterface

JSON   **J**ava**S**cript **O**bject **N**otation

MDTK   Midi **D**egradation **T**oolkit

SOTA   **S**tate **o**f **t**he **A**rt

Chapter 1

Introduction

(Jeong et al. 2019)

(Savage et al. 2015)

1.1 What are algorithms that learn

...TODO... define/introduce machine learning

1.2 What is composing music

...TODO...

1.3 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?

1.4 Motivation for this work

...TODO...

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

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

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

https://www.cl.cam.ac.uk/~ga384/bibfix.html

Jeong, D. et al., 2019. Graph Neural Network for Music Score Data and Modeling Expressive Piano Performance. In K. Chaudhuri & R. Salakhutdinov, eds. *Proceedings of the 36th International Conference on Machine Learning.* Proceedings of Machine Learning Research. PMLR, pp. 3060–3070. Available at: http://proceedings.mlr.press/v97/jeong19a.html.

Savage, P.E. et al., 2015. Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, 112(29), pp.8987–8992. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1414495112 [Accessed January 19, 2021].