Thesis Title

Tapan Vivek Auti (ID)

School of Computer Science

National University of Ireland, Galway

Supervisors

Name of Supervisor(s)

In partial fulfillment of the requirements for the degree of

MSc in Computer Science (PROGRAMME)

August 23, 2019

DECLARATION I, NAME, do hereby declare that this thesis entitled THESIS-TITLE is a bonafide record of research work done by me for the award of MSc in Computer Science (PROGRAMME) from National University of Ireland, Galway. It has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

Signature: _____

Abstract

The earliest work of using Neural Network for music composition, analysis and generation can be found around late 1980's Dolson (1989), What has changed from then to today is the availability of more discrete resources for data gathering and the variations that can be found in the availability and type of data. Musical data before was available in non symbolic form i.e musical notes etc which was hard to train along with the complexity of training with very less data. The availability of resources has paved a new pathway for more advancements in 'deep learning' Deng and Yu (2014) and hence a lot of research has been conducted since then for music generation with Neural Network.

Pons (2015) suggests that all earlier work was done using a single melody.

The majority of the research being conducted for music generation is done using RNN and LSTM Boulanger-Lewandowski et al. (2014), Chen and Miikkulainen (2001), Eck and Lapamle (2008).

The first one to use RNN was MOZER (1994) he provided motivation for Eck and Schmidhuber, who later were the first ones to apply LSTM.Eck and Schmidhuber (2002)

Sturm et al. (2016) - The most significant use of sequential abc files for music generation is done by them using LSTM, they have used two methodologies for it ,one is char rnn which uses a single character

vocabulary and other is folk rnn which used transcription token for generation.

Till date only type of research done on music regeneration was mostly on piano rolls which used midi files for regenartion Yang et al. (2017) , Dong et al. (2017)

Dong and Yang (2018) and all were generally polyphonic music generation models. Till late 2018 Mogren (2016) was the only model that used GAN for music production, it used midi files as data but had drawbacks of not being able to generate music using priming melody or chord sequences , these were removed in Yang et al. (2017) also this used Cnn as their main base model.

After various researches it was proved that CNN is faster than RNN and also easily parallizable van den Oord et al. (2016a) before this maximum model used RNN only.

All the years of research has paved path for various types of music generations, like Tokui (2020) tries to reproduce rhythm patterns of electronic music using GAN.

Like symbolic data for generation, various studies were also conducted for music generation through non symbolic forms such as raw music files and wave forms van den Broek (2021) was one such successfull model that generted music from raw files., they were able to achieve this with low convolutional GAN with a milestone of low computational power.

All the models that used GAN before had one major drawback that due to no or less temporal feature extraction the generated music didn't always sound natural and was unstable to overcome .Guan et al.

(2019) proposed a new model that considered temporal features as well and had a self-attention mechanism to enable GAN and introduced a new method of switchable normalization to stabilize network training.

A lot of research is being made in music generation considering various aspects of music such as music generation based upon the genre or feel of music, one such proposed model is used in Bao et al. (2017) here the authors took a dataset that has emotions tagged along with song so while music generation the model trains on basis of the feel of the music and produces music on what feel note you want like rowdy, romance, quiet, majestic etc.

GANSYNTH in 2019 was the most recent work that successfully synthesised audio by wave files as input using GAN , they used a specific controlled dataset for music generation and were able to achieve great speed which was claimed to be one thousand time faster generation that van den Oord et al. (2016b)

The latest research going on has been music generation from lyrics Yu et al. (2021) it used lstm generator and discriminator for generation. ,where in the dataset the lyrics are mapped to melodies but comparatively such data still is very scarce and hence the result of the mapping is not melodies to lyrics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This section should be short and highly readable, even to non-experts.

The chapters in this template are typical, but not mandatory. Change them, change the titles, change the order, as needed.

The following is a guideline for number of pages per section:Bao et al. (2017)

- Abstract 1

- Introduction 3-4

- Related work 3-6

- Background 2-4

- Data 1-3

- Experimental settings 2-3

- Methodology 6-10

- Results 4-10

- Conclusion 1-3

Figure 1.1: A suitable caption

(total 22-44)

We have given Chapter titles, but you can use `\section` and `subsection` to give more fine-grained structure as needed.

You can refer to other chapters/sections using `\ref`, e.g. "We will describe the proposed new model in detail in Chapter 5."

Notice that we must use paired backquotes and apostrophes for correct quotation marks in Latex: "here is an example". If we use standard quotation marks they are formatted incorrectly: "here is an example".

Simple equations are written using `$...$`, e.g. $e^{i\pi} - 1 = 0$.

Figures should be formatted using `\begin{figure}...\end{figure}`, with captions and labels. Figures should then be referred to from the text, e.g.: Figure 1.1 shows an example of a logo.

Refer to Overleaf documentation and other online resources for basic Latex usage, in particular mathematics, emphasis, and citation.

Chapter 2

Background

Here you can give textbook-level knowledge which you might expect that expert researchers can skip but might be helpful to establish the setting and terminology.

You can also give some general knowledge, eg the number of users of some website or the size of some industry, to help motivate the importance.

Chapter 3

Related Work

The Related Work or Literature Review should stick to strongly relevant papers in high-quality sources. Learn to recognise and avoid spam/predatory/pay-to-publish/vanity journals and conferences. Use newspaper/magazine/blog sources only very rarely and only if truly unavoidable. Cite the originator of an idea, not a random author who used it recently.

If you paste any text from any source, you must quote and cite. If you paste any text and then alter it to avoid quoting and citing, delete it and ask your supervisor for advice on how to avoid plagiarism.

The Related Work section should be synthetic, that is it should identify common themes and issues and connections between papers to form a larger-scale understanding. It should help the reader by giving a taxonomy or categorisation of existing work, i.e. it should not be a bare list of papers. It should demonstrate critical thinking and judgement, not just rephrase what previous authors have claimed.

Do not write paper titles, e.g. do not write *In a paper titled "Community Detection in Graphs"*. Just cite.

For citations, use natbib commands. There are two styles, depending on your

sentence:

- Parenthetical : Community detection in graphs is an interesting problem .

- Textual `\citet{NewmanGirvan2004}`: It was shown by that community detection in graphs is an interesting problem.

Add your citations to references.bib which is in the file-tree in Overleaf.

Chapter 4

Data

You might need a chapter about your data and pre-processing, especially if it is a dataset that has not been previously described.

Chapter 5

Methodology

Here you can describe your proposed new models.

Chapter 6

Experimental Settings

Your goal is to give a complete description of your experiments, sufficient for another researcher to read your document and reproduce your results.

Chapter 7

Results

Results first, using figures and tables, with little commentary and no interpretation.

Then analysis and interpretation.

Chapter 8

Conclusion

Here you must zoom back out to evaluate the thesis. Mention limitations and weaknesses as well as contributions.

References

Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, and Gang Hua. Cvae-gan: Fine-grained image generation through asymmetric training. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2764–2773, 2017. doi: 10.1109/ICCV.2017.299. iv, 1

Nicolas Boulanger-Lewandowski, Gautham J. Mysore, and Matthew Hoffman. Exploiting long-term temporal dependencies in nmf using recurrent neural networks with application to source separation. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6969–6973, 2014. doi: 10.1109/ICASSP.2014.6854951. ii

Chun-Chi J. Chen and Risto Miikkulainen. Creating melodies with evolving recurrent neural networks. In *Proceedings of the INNS-IEEE International Joint Conference on Neural Networks*, pages 2241–2246, Piscataway, NJ, 2001. IEEE. URL `http://nn.cs.utexas.edu/?chen:ijcnn01`. ii

Li Deng and Dong Yu. Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, 7(3–4):197–387, 2014. ISSN 1932-8346. doi: 10.1561/2000000039. URL `http://dx.doi.org/10.1561/2000000039`. ii

M. Dolson. Machine tongues xii : Neural networks. *Computer Music Journal*, 13:28–40, 1989. ii

Hao-Wen Dong and Yi-Hsuan Yang. Convolutional generative adversarial networks with binary neurons for polyphonic music generation, 2018. iii

Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang. Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment, 2017. iii

Douglas Eck and Jasmin Lapamle. Learning musical structure directly from sequences of music, 2008. ii

Douglas Eck and Jürgen Schmidhuber. Learning the long-term structure of the blues. In José R. Dorronsoro, editor, *Artificial Neural Networks — ICANN 2002*, pages 284–289, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg. ISBN 978-3-540-46084-8. ii

Faqian Guan, Chunyan Yu, and Suqiong Yang. A gan model with self-attention mechanism to generate multi-instruments symbolic music. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6, 2019. doi: 10.1109/IJCNN.2019.8852291. iii

Olof Mogren. C-rnn-gan: Continuous recurrent neural networks with adversarial training, 2016. iii

MICHAEL C. MOZER. Neural network music composition by prediction: Exploring the benefits of psychoacoustic constraints and multi-scale processing. *Connection Science*, 6(2-3):247–280, 1994. doi: 10.1080/09540099408915726. URL https://doi.org/10.1080/09540099408915726. ii

Jordi Pons. Deep learning for music information research. Technical report, Universitat Pompeu Fabra, 2015. ii

Bob Sturm, João Santos, Oded Ben-Tal, and Iryna Korshunova. Music transcription modelling and composition using deep learning. *arXiv*, 04 2016. ii

Nao Tokui. Can gan originate new electronic dance music genres? – generating novel rhythm patterns using gan with genre ambiguity loss, 2020. iii

Korneel van den Broek. Mp3net: coherent, minute-long music generation from raw audio with a simple convolutional gan, 2021. iii

Aaron van den Oord, Nal Kalchbrenner, Oriol Vinyals, Lasse Espeholt, Alex Graves, and Koray Kavukcuoglu. Conditional image generation with pixelcnn decoders, 2016a. iii

Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alexander Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. In *Arxiv*, 2016b. URL https://arxiv.org/abs/1609.03499. iv

Li-Chia Yang, Szu-Yu Chou, and Yi-Hsuan Yang. Midinet: A convolutional generative adversarial network for symbolic-domain music generation, 2017. iii

Yi Yu, Abhishek Srivastava, and Simon Canales. Conditional lstm-gan for melody generation from lyrics. *arXiv*, 17(1), 2021. ISSN 1551-6857. doi: 10.1145/ 3424116. URL https://doi.org/10.1145/3424116. iv