

View Reviews

Paper ID

187

Paper Title

MusPy: A Toolkit for Symbolic Music Generation

Track Name

ISMIR2020

Reviewer #1

Questions

2. The title and abstract reflect the content of the paper.

Strongly Agree

3. The paper discusses, cites and compares with all relevant related work.

Strongly Agree

4. The writing and language are clear and structured in a logical manner.

Strongly Agree

5. The references are well formatted.

Strongly Agree

6. The topic of the paper is relevant to the ISMIR community.

Strongly Agree

7. The content is scientifically correct.

Strongly Agree

8. The paper provides novel methods, findings or results.

Agree

9. The paper provides all the necessary details or material to reproduce the results described in the paper.

Strongly Agree

10. The paper provides reusable insights (i.e. the capacity to gain an accurate and deep understanding). Such insights may go beyond the scope of the paper, domain or application, in order to build up consistent knowledge across the MIR community.

Agree

11. Please explain your assessment of reusable insights in the paper.

This paper present a tool for symbolic music generation (MusPy). Moreover, the dataset analysis capability of MusPy can also be exploited for musicology studies. Figures 5 to 10 may illustrate what I mean. Distribution of musical keys for different dataset, or cross-dataset generalizability results can give a precious insight also for musicology studies.

15. The paper will have a large influence/impact on the future of the ISMIR community.

Agree

16. Overall evaluation

Strong Accept

17. Main review and comments for the authors

In this paper, a tool for symbolic music generation is presented.

This tool permits the easy integration of ML tools for dataset analysis like PyTorch or TensorFlow. In the same way it integrates tools for audio rendering, graphical representation, format conversion that are made possible by third party software like music21, pretty_midi and Pypianoroll.

The authors demonstrate that MusPy is 1) a tool for developing music generation systems and 2) statistical analysis of symbolic music datasets.

Reviewer #2

Questions

2. The title and abstract reflect the content of the paper.

Strongly Agree

3. The paper discusses, cites and compares with all relevant related work.

Disagree

4. The writing and language are clear and structured in a logical manner.

Strongly Agree

5. The references are well formatted.

Agree

6. The topic of the paper is relevant to the ISMIR community.

Agree

7. The content is scientifically correct.

Disagree

8. The paper provides novel methods, findings or results.

Disagree

9. The paper provides all the necessary details or material to reproduce the results described in the paper.

Disagree

10. The paper provides reusable insights (i.e. the capacity to gain an accurate and deep understanding). Such insights may go beyond the scope of the paper, domain or application, in order to build up consistent knowledge across the MIR community.

Strongly Disagree

11. Please explain your assessment of reusable insights in the paper.

The author does not present any idea about the way someone can reuse the library, even though it is open source, as far as I can see, so we cannot go further into the experience with the resources provided.

15. The paper will have a large influence/impact on the future of the ISMIR community.

Strongly Agree

16. Overall evaluation

Weak Accept

17. Main review and comments for the authors

The paper presents a great library and I wish you can continue working on this for a long time in order to help the community. The related works should be better discussed, as we can find many libraries in Python for dealing with music and they are not presented or even compared with yours. The paper has a great focus on the datasets and I am pretty sure that this could be the main topic in the title and abstract, but the authors didn't notice it, maybe. The MusPy solution seems to be a great tool, but requires more comparison with other solutions in order to get more attention.

Reviewer #3

Questions

2. The title and abstract reflect the content of the paper.

Strongly Agree

3. The paper discusses, cites and compares with all relevant related work.

Strongly Agree

4. The writing and language are clear and structured in a logical manner.

Agree

5. The references are well formatted.

Strongly Agree

6. The topic of the paper is relevant to the ISMIR community.

Agree

7. The content is scientifically correct.

Agree

8. The paper provides novel methods, findings or results.

Agree

9. The paper provides all the necessary details or material to reproduce the results described in the paper.

Strongly Agree

10. The paper provides reusable insights (i.e. the capacity to gain an accurate and deep understanding). Such insights may go beyond the scope of the paper, domain or application, in order to build up consistent knowledge across the MIR community.

Disagree

11. Please explain your assessment of reusable insights in the paper.

This paper proposes a library for symbolic music data management, but nothing about their system design seems particularly novel or interesting beyond the particular problem it solves.

15. The paper will have a large influence/impact on the future of the ISMIR community.

Disagree

16. Overall evaluation

Weak Accept

17. Main review and comments for the authors

This paper describes the development of a data management library for symbolic music data, then explores some potential applications of the tool to compare symbolic music datasets. While I don't see much value in this paper from a theoretical perspective, the tool they developed addresses an important development bottleneck in deep learning-based symbolic music generation and would be of great value to the researchers working on this problem in the ISMIR community.

The comparative analysis of different common symbolic music datasets is also interesting, and I wish the authors had spent more time analyzing the results there. For instance, why does the hymnal dataset generalize worse than the other multipitch datasets? Would a different preprocessing scheme cause the Bach dataset to generalize better? It seems like there are several potential insights here which may inform the development of future symbolic music datasets.

It would also be interesting to study the design of such a library from a UX perspective. A user study with a few students interested in symbolic music generation could shed light on what design features are important, or what additional uses (besides training a neural network) potential users might try to do with your library. While deep learning approaches are popular now, a successful library design should anticipate potential alternative use cases which may have different needs.

Reviewer #4

Questions

2. The title and abstract reflect the content of the paper.

Strongly Agree

3. The paper discusses, cites and compares with all relevant related work.

Strongly Agree

4. The writing and language are clear and structured in a logical manner.

Strongly Agree

5. The references are well formatted.

Strongly Agree

6. The topic of the paper is relevant to the ISMIR community.

Strongly Agree

7. The content is scientifically correct.

Strongly Agree

8. The paper provides novel methods, findings or results.

Strongly Agree

9. The paper provides all the necessary details or material to reproduce the results described in the paper.

Strongly Agree

10. The paper provides reusable insights (i.e. the capacity to gain an accurate and deep understanding). Such insights may go beyond the scope of the paper, domain or application, in order to build up consistent knowledge across the MIR community.

Strongly Agree

11. Please explain your assessment of reusable insights in the paper.

The manner the software is exposed and present can be understood as an reusable insight, in my opinion.

15. The paper will have a large influence/impact on the future of the ISMIR community.

Agree

16. Overall evaluation

Strong Accept

17. Main review and comments for the authors

Overall:

* The paper is very well-written, the ideas are well developed, and the figures are beautiful and easily readable.

For me it was a pure pleasure to read this paper!

* The subject is VERY important within the MIR community, and the contribution of the software is amazing!

* The documentation and codes available are very welcome and easy to follow. However I was not able to run and test it by myself, because I'm having some problems with Python on my new machine, unfortunately...

* I think that a more detailed explanation of the experiment reported in the paper with the neural nets would be good. I am aware of the limited space, but the documentation of the software could contain a better description of it, instead of needing to read the associated .py file. Another examples of usage of the package would be good also, and I hope they will be available when the software is officially published.

It was quite hard to find points to improve the paper, but here I comment about some stuff:

Introduction:

* Support the claim in lines 37-40 about datasets, representations and metrics using a reference;

* Include a summary of the paper at the end of the section.

Related work:

* At the end of the first paragraph (lines 102~105), it is mentioned that MusPy deals with more than one Machine Learning framework, but additionally to TensorFlow, only PyTorch is mentioned. However, the text, up to this point, indicates that more frameworks are also compatible. It would be interesting to clarify this point here.

Dataset Analysis:

* At line 244 you want to mention "Figure 7".

Experiments and results:

* The explanation of the large log-perplexity for JSBach Chorale Dataset in lines 287~293 is not clear.

* In Figure 10, both plots could follow the same color scheme: note that the top plot associates white with 160 and the bottom associates white with 40. It is very quick to note this and understand both plots, but for comparison reasons between both, sharing the same scale could be better.

* It is mentioned in line 340 that the information about the stratified and unified datasets are present in Figures 8 and 10, but it is only in Figure 10!

View Meta-Reviews

Paper ID

187

Paper Title

MusPy: A Toolkit for Symbolic Music Generation

Track Name

ISMIR2020

META-REVIEWER #1

META-REVIEW QUESTIONS

2. The title and abstract reflect the content of the paper.

Disagree

3. The paper discusses, cites and compares with all relevant related work.

Disagree

4. The writing and language are clear and structured in a logical manner.

Strongly Agree

5. The references are well formatted.

Strongly Agree

6. The topic of the paper is relevant to the ISMIR community.

Strongly Agree

7. The content is scientifically correct.

Agree

8. The paper provides novel methods, findings or results.

Disagree

9. The paper provides all the necessary details or material to reproduce the results described in the paper.

Strongly Agree

10. The paper provides reusable insights (i.e. the capacity to gain an accurate and deep understanding). Such insights may go beyond the scope of the paper, domain or application, in order to build up consistent knowledge across the MIR community.

Disagree

11. Please explain your assessment of reusable insights in the paper.

Although creating a toolkit to parse and work with a number of datasets is a very laudable endeavor, the paper does not provide insights to deeply understand the intrinsic characteristics of datasets.

15. The paper will have a large influence/impact on the future of the ISMIR community.

Disagree

16. Overall evaluation.

Weak Accept

17. (Initial) Main review and comments for the authors

General comments

Although creating a toolkit to manage a number of datasets and analyze the qualities of music generation models created with them is a very laudable endeavor, I think there is a lack of serious thoughts about the different research needs involved in the task they want to address. I think the authors should do research and determine what are the "essential" requirements of a general-purpose library for generative modeling of music, and then decide what are the best data representations, design principles, and tools they need to provide in order to satisfy these needs. Some of these characteristics are briefly mentioned, but are scattered through the different sections. I

think the paper will benefit of stating this in a very clear and straightforward manner. The authors can use Section 2 for this, which currently is quite short.

In the paper, the authors spend large amounts of space in figures and tables that are not quite informative for the reader, especially if they do not perform any kind of analysis of the data obtained. On the contrary, there is a lack of description of important characteristics of the library. For example, they mention “the lack of consistent implementation of commonly used metrics” for the objective evaluation of music generation systems. The reader would appreciate if the authors provide a comprehensive list of the metrics and their specific implementation.

Detailed comments

Title

- Since the model creation and training are performed by an external machine learning library. I think the title is misleading.

Abstract

- The abstract is a bit long. You may get rid of the first section and start with “In this paper, we present Muspy, an ...”

Introduction

Section 1

Section 2

- line 110: “Muspy focuses more on providing tools that are essentials in developing a music generation system” Such as what specific tools? Management and I/O are quite generic. What are the specific needs you are dealing with?

- line 155: “Muspy provides a general-purpose framework for generative modeling of music“ You should define what are the characteristics that are needed to create a general-purpose framework for the generative modeling of music.

- line 120: “This is consistent with our findings ... between various datasets” These sentences should go in Section 5, when you talk about your experiments.

Section 3

- line 129: “It contains three main components——dataset management ...“ It would be great if you show in Figure 2 where these three main component are located. As of now, you are now only showing “Data management”

- line 133: “Design principles” You should state what are the “essential” requirements of a general-purpose library for generative modeling of music, and then decide what are the best data representations, design principles, and tools you need to provide in order to satisfy these needs. You have briefly mentioned this in the previous sections, but I think you should be very clear and straightforward. You can use Section 2 for this, which currently is quite short.

- line 154: “we follow MIDI's standard for playback-related data such as note velocities and program information ... for song and source metadata.” You should elaborate on some of these concepts and your decision. For example, what do you mean by “program information”? I guess it is not MIDI's program change message. Why MusicXML's song and source metadata are relevant for your project?

- line 161: “Table 1 summarizes the key differences ... and the proposed MusPy format.” You should really try to put your figures and tables in page you are using them.

- line 183: “... using the event-based representation.” Is this correct? Perhaps I am missing something, but the

attribute of the last line in your code example states: `representation = "pianoroll"``

- line 195: "... several objective metrics proposed in the literature ..., including empty beat rate, in scale rate and polyphony rate." Could you be more specific and provide more examples of the metrics you are integrating? If there are too many, you can talk about the different categories. If space is a concern, you can get rid of Figure 6, 5, or 4.

- line 209: "The modularized and flexible design make it easy to extend support for new datasets in the future." What are the principles, or steps, other researchers have to follow in case they want to contribute a library?

Section 4

- line 244: "Finally, Figure 6" -> "Finally, Figure 7"

- line 247 and Figure 7: "... due to the fact that C major is often set as the default key in music notation programs and MIDI editors/sequencers." Do you have any explanation about why there are mostly Major keys for all datasets? I think minor keys are underrepresented. I tend to think that this is not a characteristic of the datasets. If this holds true, it is quite bizarre.

Section 5

- line 263: "... which leads to in total 357 possible events, " Since your range goes from 0 to 387, does not this imply 356 events? (i.e., 388-32)

- line 277: "We can see that all models have similar tendencies ... That is, the test perplexity can serve as an indicator for the complexity of a dataset." I think this is a stretch. The word "complexity" is complicated. Perhaps, "diversity" would be a better word to use? In any case, you have not provide any metrics for the complexity or diversity of the datasets, just their size measured in "songs" and "hours".

- line 319: "... which is possible because its samples are downsampled ... making its note duration distribution distinct to that of the other datasets." For this reason, do not you think the J.S.Bach Chorale Dataset shall be left out of the MusPy datasets?

- line 337: "... and then randomly pick one sample from that dataset." So that did you end with a balanced number of samples per dataset? How many?

Conclusions

- line 352: "Our findings provide a guide for ..." Could you please elaborate on the guides you are actually providing?

- line 354: "... we showed that combining heterogeneous datasets could help improve generalizability." Were not you expecting this characteristic?

- line 358: "This would allow a music generation system to provide a general tempo and dynamic directions without specifying actual bpm's and velocities." First, since you are using "qpms", you should stick to that. More importantly, do you have any thought on how you plan to do this mapping? If so, why you did not include them in this research?

18. Meta-review and final comments for authors. (To be completed after discussion phase.)

The creation of a toolkit to manage a number of datasets and analyze the qualities of music generation models created with them is a laudable endeavour made by the authors of this contribution. The reviewers consider the

ISMIR community will benefit from having access to this tool and, therefore, would accept it.

However, the paper present a number of issues that have to be addressed for the final publication in the conference proceedings. In general, the reviewers agree in the lack of serious thoughts about the different research needs involved in the task the authors address in their research and the lack of depth in the comparative analysis. The authors should investigate and determine what are the “essential” requirements of a general-purpose library for the generative modelling of music, and then decide what are the best data representations, design principles, and tools they need to provide in order to satisfy these needs. Also, there is a lack of description of important characteristics of the library. For example, they mention “the lack of consistent implementation of commonly used metrics” for the objective evaluation of music generation systems. However, the authors themselves do not provide a thorough documentation and these metrics. Readers of the paper would appreciate if the authors provide a comprehensive list of the metrics and their specific implementation.

The library and tools presented in this paper seem to be useful and may contribute to the community. However, the authors must address all comments, edits, and suggestions by the reviewers for the inclusion of this paper in the conference proceedings.
