SuPP & MaPP: Adaptable Structure-Based Representations for MIR Tasks

Claire Savard [1] Erin H. Bugbee [2] Melissa R. McGuirl [2] Katherine M. Kinnaird [3]

[1]University of Colorado-Boulder [2]Brown University [3]Smith College

Abstract

Accurate and flexible representations of music data are paramount to addressing MIR tasks, yet many of the existing approaches are difficult to interpret or rigid in nature. This work introduces two new song representations for structure-based retrieval methods: **Surface Pattern Preservation (SuPP)**, a continuous song representation, and **Matrix Pattern Preservation (MaPP)**, SuPP's discrete counterpart. These representations come equipped with several user-defined parameters so that they are adaptable for a range of MIR tasks. Experimental results show MaPP as successful in addressing the cover song task on a set of Mazurka scores, with a mean precision of 0.965 and recall of 0.776. SuPP and MaPP also show promise in other MIR applications, such as novel-segment detection and genre classification, the latter of which demonstrates their suitability as inputs for machine learning problems.

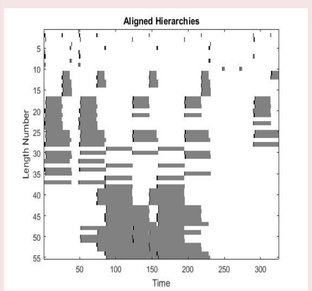
Background

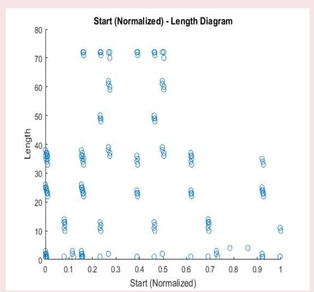
I. Aligned Hierarchies (AHs) [1]:

- Encodes all possible hierarchical structure decompositions on a common time axis
- Pros: great visual representation of all repetitions occurring in a song
- Cons: built for cover song task, computationally inefficient, rigid for comparisons

II. Start(Normalized)-Length (S_NL) Diagrams [2]:

- Motivated by TDA, collection of points whose coordinates represent structure in AHs by start time and length of structural representations in a song
- Pros: not as rigid in comparison and more computationally efficient than AHs
- Cons: built for cover song task, not machine learning friendly, complicated computationally

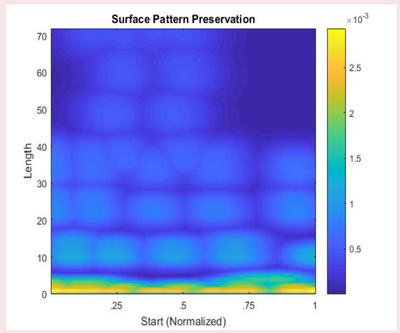




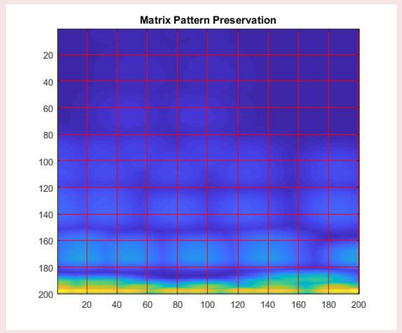
Goal: Create a new representation (influenced by [3]) that is more computationally efficient, that can be extended to more MIR tasks, and is machine learning friendly.

Methods

I. Surface Pattern Preservation (SuPP): 2D Gaussians are applied to each point on the S_NL diagram with user-defined standard deviations in both the start and length dimensions. A user-defined surface weight function is also applied to allow for emphasis of different parts of a piece. The flexibility in SuPP parameter selection allows this representation to be applied to various MIR tasks. SuPP balances repetitive structure information with user-preference on how prominent each structure should be. As a surface, however, SuPP is computationally complex.



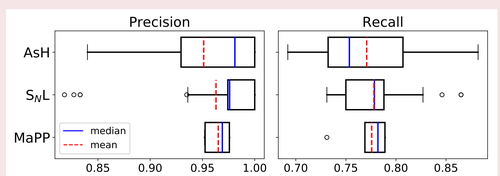
II. Matrix Pattern Preservation (MaPP): To simplify computations on SuPP, we create an alternate representation which is a pixelation of SuPP. The volume under the surface within each pixel is calculated and stored in a discrete matrix, the MaPP. As a matrix, computations on MaPP are simple, and they can also be extended to machine learning tasks. For example, in the cover song task you can measure dissimilarity between any two MaPPs by computing a distances between each pair. The SuPP to MaPP conversion condenses information but greatly simplifies computation.



Experimental Results

Cover Song Task:

- 104 Chopin Mazurka scores in dataset, each has expanded and non-expanded pair
- SuPPs have constant standard deviations in time and length, linearly decreasing surface weight in length
- Measure MaPP dissimilarities using Frobenius distance
- Mutual k-nearest neighbors ($k = 1$) to pair songs

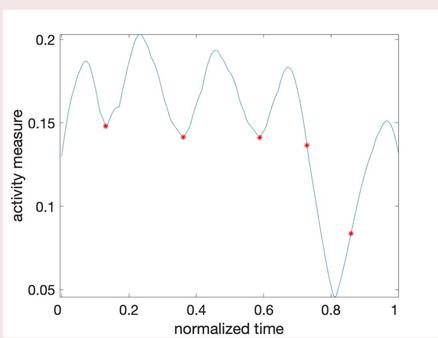


Genre Classification

- 104 Chopin Mazurkas and 676 Jazz lead sheets
- Same SuPP as in cover song task
- MaPP matrix elements are the features input to machine learning classifiers
- Logistic regression, a Gaussian kernel SVM, polynomial kernel SVM each classify with $> 94\%$

Novel-Segment Detection:

- Same SuPP as in cover song task
- Sum columns in MaPP to create 1D-vector to measure structure activity at given time
- Local minima = regions between repeated sections
- Minima outliers = middle of novelty sections, 2nd derivative used for boundary



SuPP Parameter Selection

In creating the SuPP, there are 3 parameters that can be altered, allowing users to emphasize different parts of a musical piece. These parameters may be altered for specific MIR tasks as well.

$\boldsymbol{\sigma_s}$: standard deviation along start axis
ex: constant, each time-slice in song is equally important

$\boldsymbol{\sigma_\ell}$: standard deviation along length axis
ex: constant, each structure length is equally important

$\boldsymbol{F(\bar{s}, \ell)}$: surface weight applied on top of aggregation of 2D gaussians
ex: linearly decreasing, larger repetitive structures not as reliable (harder to find computationally)

Conclusion

We present two new music representations, SuPP and MaPP, which extend from AHs [1] and S_NL diagrams [2] with TDA inspiration [3] and which are adaptable to many MIR tasks.

- Pros: more flexible and computationally friendly than predecessors, MaPP can be applied to machine learning tasks
- Cons: manual selection of SuPP parameters
- Cover song task: comparable results on average, less variance
- Genre classification: accuracy above 94% in preliminary results
- Novel-segment detection: method shows promise, further testing required

Future work will focus on more in-depth studies of applying MaPP to genre classification and novel-segment detection, and transitioning from working with score data to audio data.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. DMS-1439786 while the authors were in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the Summer@ICERM 2017 and Collaborate@ICERM programs. The second author was supported by The Karen T. Romer Undergraduate Teaching and Research Awards. The third author was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1644760. The last author is the Clare Boothe Luce Assistant Professor of Computer Science and Statistical & Data Sciences at Smith College and as such, is supported by Henry Luce Foundation's Clare Boothe Luce Program.

References

[1] K. M. Kinnaird, "Aligned hierarchies: A multi-scale structure-based representation for music-based data streams," ISMIR, 2016.
[2] M. R. McGuirl, K. M. Kinnaird, C. Savard, and E. H. Bugbee, "SE and SNL diagrams: Flexible data structures for MIR," ISMIR, 2018.
[3] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier. Persistence images: A stable vector representation of persistent homology. Journal of Machine Learning Research, 18(8):1-35, 2017.