7100 Project Proposal Fall 2020

Tianxue(Tess) Hu

October 4, 2020

1 About

Automatic chord recognition in symbolic music is a hallmark research topic in the field of music information retrieval (MIR). In the past decade, this problem has been tackled by using a variety of deep learning methods, despite the achievement or the glass-ceiling of adopting these models, further improvements of chord recognition accuracy appears to be improving music representations, using a different pre-processing architecture, or refining results with post-processing steps.

Based on previous research (in the next section), this project will focus on building a deep learning model for automatic chord symbol recognition in symbolic polyphony music datasets.

2 References

Chord symbol analysis focuses on recording chords names, specifying the triad quality, seventh, inversion (bass note), and any modifications (such as added and altered notes). Past related research (automatic harmony analysis) are either about chord symbol or roman numeral identification, depending on whether to consider local/global keys or not. These studies mainly serve research purposes in the field of computational musicology, music information retrieval, and machine learning, some of them [12][3] were applied to chord recognition applications or plug-ins for general audiences.

Early efforts were made by building rule-based, algorithmic, or hierarchical models [12][3][4]. With the booming of machine learning in recent years, automatic harmony analysis has also been considerably developed by adopting such methods [11][7] [2][1]. These research proposed different models on different datasets/data representations in order to break the glass-ceiling of identification accuracy. In the latest research in the field, Micchi et al. collected all existing Roman numeral analysis datasets, examined the best practices in the encoding of pitch, time, and harmony for machine learning tasks, and tested several neural network architectures [8]. They proposed an input representation of polyphony music with:

- **Timing: 'frame-based' encoding,** where each input vector denotes an individual time frame, padding sustained notes.

- **Pitch: two dimensions**
  - Pitch spelling: pitch class representations (12 per octave, and no difference between the enharmonic equivalent pairs like G and Ab), or maintaining the full pitch spelling (with 21 possibilities per octave for single sharps/flats and 35 for double).
  - Registral information: whether or not to include octave info.

While their output representations are rumerals using quantized label (one class per slice).

However, there are several setbacks in their representations that might causes the poor performance of identifying Roman numerals (Figure 1). In the input representation, note padding in quantized slices doesn't indicate real onsets (new attacks vs. held from the previous slice) of notes, duration of notes, metric positions (strong beat vs. weak beats), and voice leading information, whereas these information are essential in the real-word harmony analysis. Moreover, the output representation in Roman numerals may result in too many categories for a model to predict (under-fitting?).

	Key	Degree	Quality	Inversion	RN
ConvGRU + PSb + global (all data)	**82.9**	**68.3**	**76.6**	**72.0**	**42.8**
ConvGRU + PSb + global	80.6	66.5	76.3	68.1	39.1
Chen and Su (2019)	78.4	65.1	74.6	62.1	
Chen and Su (2018)	66.7	51.8	60.6	59.1	25.7
Local model after Temperley (1999)	67.0				

Figure 1: Result table from Micchi et al.'s paper.

3 Project Plan

In this project, we will propose several representation improvements based on Micchi et al's research. For the input representations, we could encode timing in onset slices [5] instead of quantized slices to comprise metric information, add other features such as durations, approach/depart by step to include voice leading characteristics. For the output representations, we choose to use chord symbols without inversions in order to reduce categories and substitute chord inversion identifications through a post-processing step by identifying bass note in the score.

Besides improvements on input/output representations, another feature of our project will be adding a pre-proccesing step that eliminates/weights non-chord tones in the score. Since we have already built a non-chord tone recognition model for melody last semester, we decide to use multi-parts symbolic music datasets as they are easier to split into parts of melodies (for feeding into the non-chord tone model) comparing to other music that constantly changing number of voices. Here are several available open-source datasets with both 4-parts symbolic music and chord labels in the format of Roman numeral and chord letters:

- Annotated Beethoven Corpus (ABC) [10]: all Beethoven string quartets (16 string quartets, 70 movements) in MuseScore format, annotated with Roman numerals. **Micchi et al. used in their research.**
- "Sun" Quartets [9]: the Op. 20, six string quartets from Joseph Haydn in **kern format, annotated with Roman numerals.
- The Rameau dataset [6]: 371 Bach chorales in Lilypond format, and 156 of them are annotated with chord letters.
- Roman Text [13]: contains 24 preludes from the first book of Bach's Well Tempered Clavier, and 48 romantic songs from France and Germany, annotated with Roman numerals. **Micchi et al. used in their research. We only consider 24 preludes from Bach's Well Tempered Clavier in our project.**

In the string quartets datasets, we need to take further considerations because sometimes an instrument plays multiple notes at the same time. Researchers also refer Bach Chorales as homorhythmic music where all parts share a very similar rhythm[5]. The nature may also simplify the onset slices procedure as well. This pre-processing step of eliminating NCT could be skipped if we want to explore the performance only with improved representations, we will compare both scenarios (with NCT and without NCT) in our research.

On the other hand, Micchi et al. mainly focuses on training a CovGRU model on 4 datasets and applied other models only on specific datasets to compare results. If our approach has significant improvements, we will further explore the performances using other deep learning methods that have been applied in automatic harmony recognition task, such as RNN-LSTM [2] and Transformer[1].

See workflow (Figure 2) on page 5!
See tentative weekly schedule (Figure 3) on page 6!

Main stages of project

Semester 1

- Getting familiar with Micchi et al.'s method
- Data collecting, pre-processing for music and chords (convert Roman numeral to chord symbol if necessary. Start with the Haydn Bach Chorales dataset.), and data augmentation via transpose.
- Feed score into non-chord tone recognition model. Assemble new score.
- Convert score to input representations. Convert chord symbols into output representations.
- Train and test ConvGRU model.

References

[1] T.-P. Chen and L. Su. Harmony transformer: Incorporating chord segmentation into harmony recognition. *neural networks*, 12:15, 2019.

[2] T.-P. Chen, L. Su, et al. Functional harmony recognition of symbolic music data with multi-task recurrent neural networks. In *ISMIR*, pages 90–97, 2018.

[3] W. B. De Haas, M. Rohrmeier, R. C. Veltkamp, and F. Wiering. Modeling harmonic similarity using a generative grammar of tonal harmony. In *Proceedings of the Tenth International Conference on Music Information Retrieval (ISMIR)*, 2009.

[4] P. R. Illescas, D. Rizo, and J. M. I. Quereda. Harmonic, melodic, and functional automatic analysis. In *ICMC*. Citeseer, 2007.

[5] Y. Ju, S. Howes, C. McKay, N. Condit-Schultz, J. Calvo-Zaragoza, and I. Fujinaga. An interactive workflow for generating chord labels for homorhythmic music in symbolic formats. In *ISMIR*, pages 862–869, 2019.

[6] P. Kröger, A. Passos, M. Sampaio, and G. De Cidra. Rameau: A system for automatic harmonic analysis. In *ICMC*, 2008.

[7] K. Masada and R. C. Bunescu. Chord recognition in symbolic music using semi-markov conditional random fields. In *ISMIR*, pages 272–278, 2017.

[8] G. Micchi, M. Gotham, and M. Giraud. Not all roads lead to rome: Pitch representation and model architecture for automatic harmonic analysis. *Transactions of the International Society for Music Information Retrieval*, 3(1), 2020.

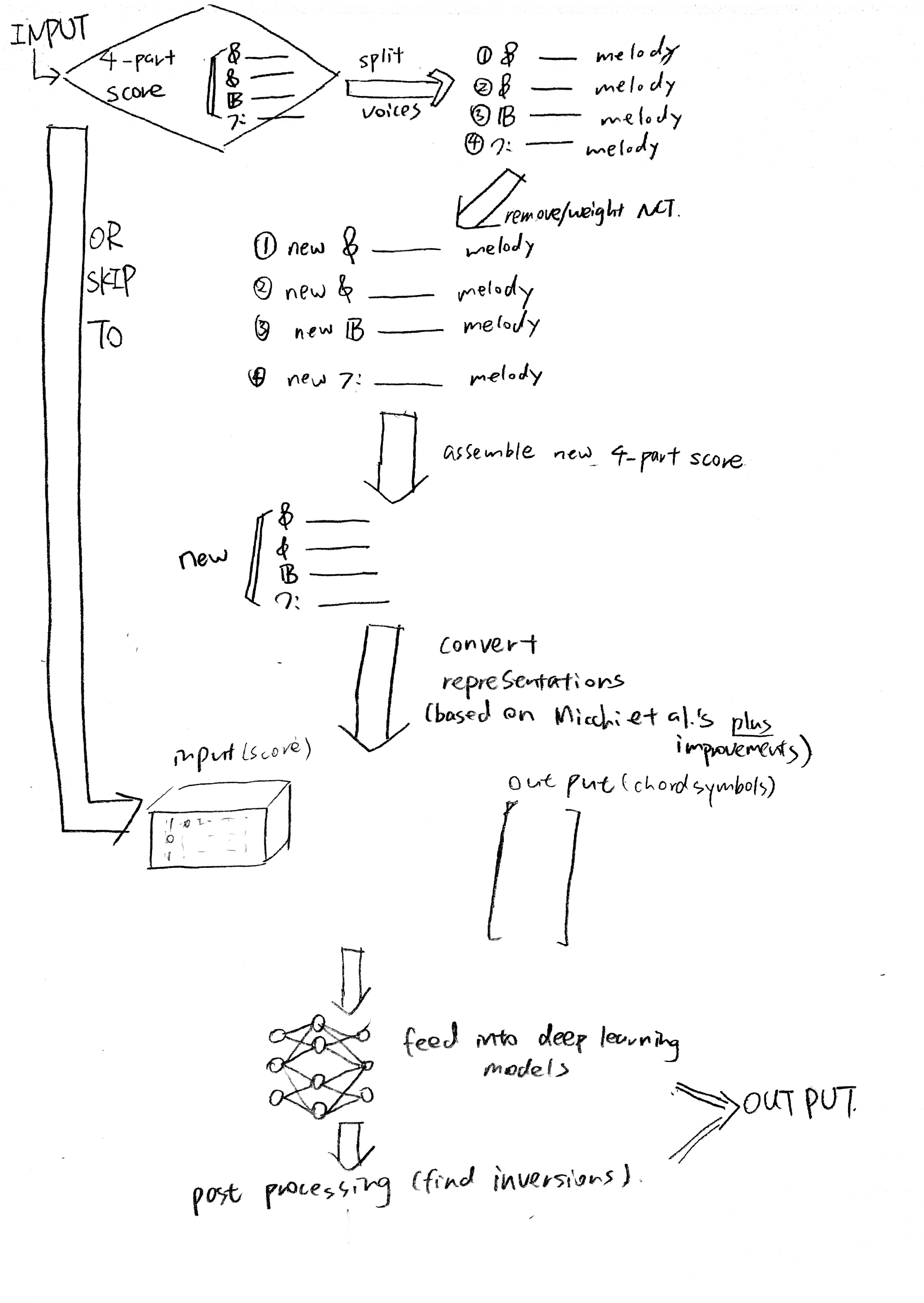

Figure 2: Workflow example of a 4-part score.

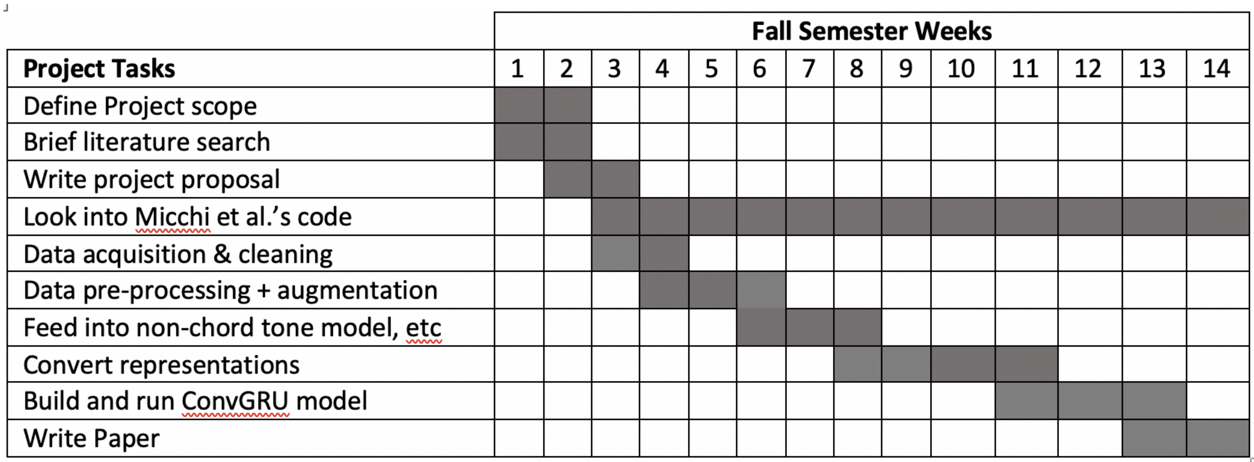

Figure 3: Weekly Schedule 1st Semester.

[9] N. Nápoles López. Automatic harmonic analysis of classical string quartets from symbolic score. Master's thesis, 2017.

[10] M. Neuwirth, D. Harasim, F. C. Moss, and M. Rohrmeier. The annotated beethoven corpus (abc): A dataset of harmonic analyses of all beethoven string quartets. *Frontiers in Digital Humanities*, 5:16, 2018.

[11] D. P. Radicioni and R. Esposito. Breve: an hmperceptron-based chord recognition system. In *Advances in Music Information Retrieval*, pages 143–164. Springer, 2010.

[12] D. Temperley and D. Sleator. Modeling meter and harmony: A preference-rule approach. *Computer Music Journal*, 23(1):10–27, 1999.

[13] D. Tymoczko, M. Gotham, M. S. Cuthbert, and C. Ariza. The romantext format: A flexible and standard method for representing roman numeral analyses. In *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR*, 2019.