PRELIMINARY REPORT ON GATHERING A LARGER ANNOTATED DATASET FOR PATTERN DISCOVERY TASKS

Klara Žnideršič, Iris Yuping Ren, Anja Volk, Matevž Pesek

University of Ljubljana, Faculty of computer and information science
Utrecht University

1. INTRODUCTION AND BACKGROUND

The topic of pattern discovery has been tackled in musicology, music theory and, in recent decades, in music information retrieval (MIR) through a variety of different approaches. The basis for determining a pattern seems to lay in its repetitiveness [1], as it aids the listener in perceiving the musical information more clearly by compressing the repeated segments [2].

In an attempt to formalise the definition and evaluation of pattern discovery, the MIREX task is named *Discovery of repeated themes & sections*. The task aims to find repetitions representing one of the more significant aspects of a music piece [3]. Perhaps the most valuable contribution of the MIREX tasks is the collection of standardised datasets for evaluation. Although the discovery of repeated themes & sections is almost a decade old, the JKU-PDD dataset, which the task provided for the development of new approaches, still remains a benchmark to this day (e.g. [4]), regardless of its limitations in size (5 songs) and genre (classical music). In addition, the evaluation metrics used in the MIREX task treat annotations as a single "ground truth" entity, regardless of multiple human annotators.

We present a new collection of annotations for pattern discovery, which includes 22 music pieces, annotated by five annotators. The annotations were collected using the PAF tool [5]. The dataset includes the monophonic symbolic music pieces from existing datasets (JKU-PDD [6], Anomic [7]), as well as different genres (folk songs and their variations, jazz solos). The motivation behind the collection is to 1) explore the pattern discovery past the classical musical genre (therefore adding new genres), 2) to analyze the inter-annotator agreement (including five independent annotators), and 3) to collect more detailed annotations, which include subpatterns and musical transformations (prime forms, inversions, retrograde, retrograde inversions). In this paper, we report on the preliminary statistical observations of the collected data.

2. DATA COLLECTION

In theory, the term pattern indicates a repetitive sequence, which represents an atomic conceptual part of a music piece. However, individual perception of pattern instances differs between the listeners—the underlying reason for these differences or disagreements may lay in the listeners' background, experience or familiarity [8]. Annotations should therefore be treated as reference labels, which is at least in its majority correct from an objective standpoint. Additionally, the annotators' metadata could reveal additional insights and explain the inter-annotator agreement based on the annotators' background [9].

2.1. Challenges

One of the biggest challenges for gathering a (larger) annotated dataset is the number of annotators and the amount of time needed to provide annotations. To address the problem of annotators, we employed five musicology experts, all of whom had a bachelor's degree in musicology and were enrolled on a musicology master's programme. The number of experts was considered sufficient for inter-annotator analysis, and was implicitly capped by a cost total cost of expert work. Due to a large number of songs, the experts' time was financially compensated.

Second, the medium for gathering annotations could represent a significant overhead, if the materials are gathered in a physical form. A few digital tools for pattern annotations were presented [5, 10], which can significantly reduce the time needed to digitise and minimise the digitisation (transcription of annotations into digital form) errors. We used the PAF interface [5], which provides a digital interface for annotating patterns, automatic annotation of (strict) pattern occurrences, and pattern labels. The latter was used for describing individual subvariants of the occurrence, as described later.

2.2. Music pieces and annotators

The dataset that was used in the annotator tool consists of the subsets, music pieces and compositions of various composers and genres. Classical composers were represented by the songs from JKU-PDD (Gibbons, Bach, Mozart, Beethoven, Chopin) and Anomic datasets (Bach, Haydn, Mozart, Beethoven), five pieces are taken from jazz repertoire (Bob Berg, Charlie Parker, Lester Young, John Coltrane), and six from Dutch Song Database. The songs were annotated by five annotators, two females and three males, aged 20 to 25.

3. PRELIMINARY ANALYSIS

A preliminary analysis of the collected 4026 patterns shows that the average pattern range is 9.93 halftones, the average pattern length, expressed in quaver notes, is 8.18, and the most common last note of an annotated pattern is a quaver.

klara.znidersic@gmail.com

Based on labels provided by the annotators, the average occurrence of a pattern in each individual song per user is 2.10. The highest number of occurrences of a specific pattern annotated in a song by one annotator was 36, and the highest number of other annotators were 19, 19, 6, and 16.

3.1. Inter-annotator agreement

The computing of the inter-annotator agreement is based on matching annotations' starting and ending timestamps. This agreement measure was previously used and accurately described in [9]. After each annotator is compared to every other, the precision, recall and F1 scores are used to create matrices, (Figure1), which then serve as a summary of the agreed patterns between all pairs of annotators. The agreement between annotators is also represented by timelines of patterns made for each song and annotator, where blue lines represent the patterns and the intensity of color shows additional subpatterns of the overlaps between separate patterns (Figure2).

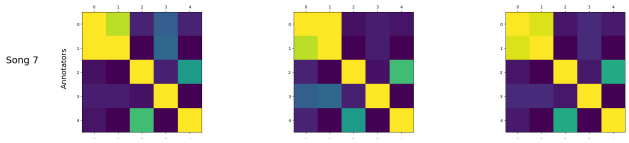

Fig. 1. Precision, recall and F1 Matrices of the inter-annotator agreement for one example.

Results acquired by described methodology show the strongest agreement between annotators in cases of folk songs and the weakest agreement when it comes to jazz solo excerpts. Those results will be supported by further feature analysis, which will provide a more detailed insight into individual patterns.

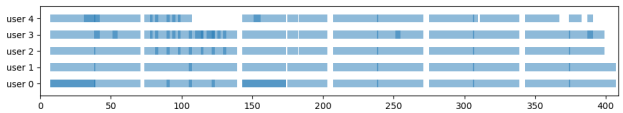

Fig. 2. Example of pattern agreement between annotators.

4. FUTURE WORK

The presented collection of annotations offers new possibilities for diverse analyses and (re-)evaluation of pattern discovery approaches. The preliminary inter-annotator analysis revealed visible differences between the genres. This not only highlights the potential for further analyses of individual patterns, but also calls for a comparison of these patterns within and across different genres. By instructing the annotators to also pay attention to the subpatterns that might occur within the larger patterns, this part of the presented dataset provides an opportunity to explore additional insights into the underlying structure of musical compositions, as well as aid in the classification and understanding of different genres. In addition, the dataset could be used to develop new pattern discovery algorithms and re-evaluate the existing approaches, which have so far reported results mostly on the JKU-PDD dataset. Finally, the currently used metrics will be further expanded to take into consideration multiple annotators separately, their backgrounds and music experience.

5. REFERENCES

[1] David Meredith, Kjell Lemström, and Geraint A Wiggins, "Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music," 2003.

[2] David Meredith, "Cosiatec and siateccompress: Pattern discovery by geometric compression," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2013.

[3] Tom Collins, "2015:discovery of repeated themes & sections - mirex wiki," 2016.

[4] Otso Björklund, "Siatec-c: Computationally efficient repeated pattern discovery in polyphonic music," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2022, International Society for Music Information Retrieval Conference, ISMIR 2022 ; Conference date: 04-12-2022 Through 08-12-2022.

[5] Matevž Pesek, Darian Tomašević Tomaševič, Iris Yuping Ren, and Matija Marolt, "Ismir late-breaking/demo [unrefereed] an opensource web-based pattern annotation framework-paf," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2019.

[6] Tom Collins, "Jku patterns development database," 2013.

[7] Iris Yuping Ren, Hendrik Vincent Koops, Anja Volk, and Wouter Swierstra, "In search of the consensus among musical pattern discovery algorithms," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2017, pp. 671–678.

[8] Matevž Pesek, Špela Medvešek, Anja Podlesek, Marko Tkalčič, and Matija Marolt, "A comparison of human and computational melody prediction through familiarity and expertise," *Frontiers in Psychology*, vol. 11, 2020.

[9] Darian Tomašević, Stephan Wells, Iris Yuping Ren, Anja Volk, and Matevž Pesek, "Exploring annotations for musical pattern discovery gathered with digital annotation tools," *https://doi.org/10.1080/17459737.2021.1943026*, vol. 15, pp. 194–207, 5 2021.

[10] Frank Zalkow, Angel Villar Corrales, T J Tsai, Vlora Arifi-Müller, and Meinard Müller, "Ismir late-breaking/demo [unrefereed] tools for semi-automatic bounding box annotation of musical measures in sheet music," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2019.