SYMCHMMERGE: AN EXTENSION TO THE COMPOSITIONAL HIERARCHICAL MODEL FOR PATTERN DISCOVERY IN SYMBOLIC MUSIC REPRESENTATIONS

Matevž Pesek
University of Ljubljana
Faculty of computer
and information science
matevz.pesek@fri.uni-lj.si

Aleš Leonardis
Centre for Computational
Neuroscience and Cognitive Robotics
School of Computer Science
University of Birmingham
ales.leonardis@fri.uni-lj.si

Matija Marolt
University of Ljubljana
Faculty of computer
and information science
matija.marolt@fri.uni-lj.si

ABSTRACT

This submission is an extension to the SymCHM approach, submitted to Mirex 2015 Discovery of Repeated Themes & Sections. While retaining the compositional approach, we further improved the results by refining the output the SymCHM, using a technique of pattern merging and output refinement.

1. INTRODUCTION

As an alternative to the existing deep learning architectures, a compositional hierarchical model (CHM) was introduced to the MIR field by Pesek et al. [3], based on the model in computer vision, developed by Leonardis and Fidler [2]. Its main difference between the model and other deep architectures is in its transparent structure, thus allowing representation and interpretation of the signal's information extracted on different levels.

Based on the Mirex 2015 submission which presented a novel application of SymCHM for pattern discovery in symbolic music representations [4], we developed the SymCHMMerge. While retaining the structure and the applied methods presented in [3], the SymCHM was adjusted for the two-dimensional symbolic input. The SymCHMMerge provides a pattern merging approach, further refining the output of the SymCHM model and thus achieving better results.

2. COMPOSITIONAL HIERARCHICAL MODEL

The compositional hierarchical model provides a hierarchical representation of the audio signal, from the signal components on the lowest level, up to individual musical events on the highest levels. The model is built on the belief of the signal's ability of hierarchical decomposition into atomic blocks, denoted as *parts*. According to their complexity, these parts can be structured across several layers from less to the more complex. Parts on higher layers are expressed as compositions of parts on lower layers — similarly as a chord is composed of several pitches, or a pitch represents a composition of several harmonics. A part can therefore describe individual frequencies in a signal, their combinations, as well as pitches, chords and temporal patterns, such as chord progressions.

2.1 Input

The input layer $\mathcal{L}_0$ of the model is a symbolic representation of the music signal, consisting of a set of pitches, each defined by an onset and an offset. It contains a set of atomic parts (pitches), which are activated (is present in the signal) at any MIDI location and any given time. Similar to the original model, where any time-frequency representation can be used for the input layer, any two-dimensional representation — in this case pitch-time representation — can be used as an input to the adjusted version.

The SymCHM is used in the same two-stage manner as the CHM. During the first, the build stage, the model is developed layer-by-layer. By composing atomic $\mathcal{L}_0$ parts, the model first produces compositions of two pitches ($\mathcal{L}_1$). To retain the compositions which cover the most information in the input layer, a statistical approach is employed. Based on the compositions' occurrence, the learning process retains the compositions which are more frequently activated.

3. CHM FOR PATTERN DISCOVERY

Due to the statistical nature of the model's learning behaviour, more frequently activated parts are retained on each layer. The activations can be observed as locations of part's occurrences, thus, the amount of part's activations indicates the significance of the part's structure (i.e. a repeated pattern) in the signal. A part can thus be observed as a medium for aggregation of re-occurring patterns. For this task, we build a new model for each musical piece provided in the input. After the learning process, the model's

Table 1. Comparison of SymCHM and SymCHMMerge for discovery of repeated patterns on the JKU PDD dataset.

	P_{est}	R_{est}	F_{1est}	$P_{occ(c=.75)}$	$R_{occ(c=.75)}$	$F_{1occ(c=.75)}$
SymCHM	67.92	45.36	51.01	93.90	82.72	86.85
SymCHMMerge	67.96	50.67	56.97	88.61	75.66	80.02
	$P_{occ(c=.5)}$	$R_{occ(c=.5)}$	$F_{1occ(c=.5)}$	P	R	F_1
SymCHM	78.53	72.99	75.41	25.00	13.89	17.18
SymCHMMerge	83.23	68.86	73.88	35.83	20.56	25.63

activations are produced through the output of the model where every part is observed as a pattern and each part activation belonging to that part as a pattern occurrence.

4. SYMCHMMERGE

For a thorough analysis, we focused on the results produced on the JKU PDD dataset. The result of analysis offered an improvement in the model's output by removing the redundant parts in the output and identifying the missing patterns. The analysis the output of the pattern-picking approach shows it can still produce a several redundant patterns. We therefore implemented a part matching procedure. By matching the parts' activations the matching procedure outputs the percentage of co-occurrences of compared parts. For the comparison of co-occurring activations, we first identify occurrences of both parts, which commonly cover at least one event in the input representation. A similarity coefficient is calculated across all pattern instances. If there is a sufficient amount of co-occurrence of pattern instances, the patterns are merged.

5. RESULTS

The results of the evaluation on the JKU PDD and the Mirex 2015 Discovery of Repeated Themes & Sections Results task evaluation on the JKU PTD database are shown in Table 1. For the evaluation and comparison of the output to the ground truth, we used the script, provided by Tom Collins [1]. The results display the commonly used metrics for the task.

Compared to the MIREX 2015 SymCHM, the improved SymCHM achieves better results (6.3% increase) for the F_{1est} establishment measure which evaluates to to what extent an algorithm can discover one occurrence, up to time shift and transposition [1]. The occurrence measures $F_{1occ(c=0.75)}$ and $F_{1occ(c=0.5)}$, that evaluate the algorithm's ability to find all occurrences of the established patterns, have dropped for almost 5 percent on the JKU PDD dataset. The TLF_1 measure has increased for 1.8% where as the first five target proportion and first five precision increased significantly for 18.8% and 4% respectively.

6. REFERENCES

[1] Tom Collins. 2015:Discovery of Repeated Themes & Sections - MIREX Wiki, 2015.

[2] Aleš Leonardis and Sanja Fidler. Towards scalable representations of object categories: Learning a hierarchy of parts. *Computer Vision and Pattern Recognition, IEEE*, pages 1–8, 2007.

[3] Matevž Pesek, Aleš Leonardis, and Matija Marolt. A compositional hierarchical model for music information retrieval. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 131–136, Taipei, 2014.

[4] Matevž Pesek, Urša Medvešek, Aleš Leonardis, and Matija Marolt. SymCHM: a compositional hierarchical model for pattern discovery in symbolic music representations. In *11th Annual Music Information Retrieval eXchange (MIREX'15)*, pages 1–3, Malaga, 2015.