A TWO-LAYER APPROACH FOR MULTI-TRACK SEGMENTATION OF SYMBOLIC MUSIC

Brigitte Rafael and Stefan Oertl
Re-Compose GmbH
Vienna, Austria

Music segmentation is a key issue in music information retrieval (MIR). Structural information about a composition achieved by music segmentation can improve several tasks related to MIR such as searching and browsing large music collections, visualizing musical structure, lyric alignment, and music summarization. Various approaches have been introduced to MIR, many of them recently concentrating on digital audio. The authors of this paper present a two-layer approach for the segmentation of symbolic music. The first step uses exact and approximate string matching methods to detect the best candidate segmentations for each track of a multi-track composition independently. The second step combines all single track results and determines the best segmentation for each track in respect to the global structure of the composition.

The most common approach in recent research projects aims at detecting structure boundaries with the aid of a novelty score. Methods using that score are limited to compositions following certain rules and principles as they require the existence of domain knowledge (extraopus). However, the authors' method is not based on this kind of a priori knowledge but focuses on the information provided within the piece itself (intraopus). In consequence, it can be applied to a broader musical spectrum. Music is analyzed based on its self-similarity and repetitions are used to detect and cluster segments.

Since the information in MIDI files is separated into single tracks, each track can be analyzed isolated from the remaining composition and its optimal solution can be found independently from the other tracks. Musically relevant data of each track (e.g., pitch, note duration, rhythmic position, velocity, etc.) is transformed into a string sequence. The algorithm then applies several string matching techniques to find repetitions within the sequence. It combines exact and approximate string matching to benefit from the advantages of both methods. For exact string matching, the authors use a parameterized version of the correlative matrix (a concept similar to the similarity matrix). The approximate string matching algorithm borrows the concept of dynamic programming from the field of bioinformatics. Dynamic programming targets at the alignment of two music sequences, thus allowing insertion, deletion, and replacement of notes. The exact and approximate string matching algorithms both result in one or multiple lists of pattern groups that form nonoverlapping segmentations of the composition. The algorithm evaluates each list and calculates a total score for each potential segmentation to find the optimal local segmentation for the track. The evaluation function considers various features of segments and segment groups. It is fine-tuned by parameters that can be set manually by music experts. The list of pattern groups with the best score represents the best local segmentation for the current track. All lists exceeding a threshold score based on this highest score are saved for future processing.

The result of the first step of the algorithm is a list of multiple candidate segmentations for each track. The segmentation with the highest score is the optimal local segmentation for the respective track. From a global view, however, also candidates having a lower score might turn out as the best segmentation if they fit into the global structure of the composition better than the best single track segmentation. The second part of the algorithm calculates a geometric score for each candidate segmentation of each track by comparing it to all candidate segmentations of all other tracks. The local score of each segmentation is weighed by its global score from geometric comparisons, resulting in a best global segmentation for each track.

The two-layer approach presented in this paper already achieves good results for a large pool of test compositions. Considering both exact and approximate string matching techniques, segmentations for tracks with various characteristics can be found. For tracks allowing multiple segmentations with similar scores, the algorithm finds the best solution in respect to the geometry of the composition.

This paper presents a two-layer approach to achieve a segmentation for each track that is influenced by the global geometry of the composition. A third layer can be introduced to merge all segmentations of the single tracks to get one global segmentation for the whole composition. Implementing such a third layer will be part of future developments.