Title	和音、機能、調性の相互依存性を考慮した教師なし認識
Author(s)	上原, 由衣
Citation	
Issue Date	2022-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/18139
Rights	
Description	Supervisor: 東条 敏, 先端科学技術研究科, 博士

Abstract

Music has provided indispensable pleasures to humans and is one of the large markets in entertainment. Regularities that govern music have been studied for philosophical interests in ancient times and practical uses in the present age. As a result, music theories have converged into a certain degree of common sense: harmony theory. The harmony theory has been employed in music education and recent artificial intelligence. Despite its popularity, composers and even listeners have not fully been satisfied with it, and thus musical works have not been restricted by the theory.

The motivation of this study is to conduct a harmony analysis that reflects the characteristics of diverse musical expressions. Harmony analysis can be generalized as the following four processes. Firstly, define an appropriate set of chord labels. Then, segments scores and assign chord labels. Finally, analyze the labeled chord sequence by the chord functions. These processes seem simple but not trivial in practice because of their interdependency. It is especially significant for polyphonic music in contrast to music where melody and harmony can be easily distinguished (e.g., homophonic music). In addition, the notion of tonality would influence all these processes.

Previous efforts of unsupervised statistical learning for harmony have independently simulated chord labeling, chord function identification, or key detection. This study argues that simulations for these three attributes should be performed in a unified manner, considering the mutual dependency between them. In addition, chords and keys may not be easily annotated when we analyze a broader style of musical pieces. Therefore, we propose a model that does not require pre-annotations for not-targeted attributes and analyzes chord, function, and tonality in unified unsupervised learning.

To this end, this study attempts to combine a probabilistic generative model and neural networks. As the generative model, we select the hidden semi-Markov model (HSMM), an extension of the hidden Markov model (HMM). The HMM has been employed in previous works and showed promising results that well simulated known chord functions. However, considering that this study aims to automatically segment scores and classify chords instead of relying on pre-annotated chord symbols, we employ an HSMM that explicitly models duration probabilities for hidden states that are expected to represent latent chord categories. Furthermore, a more difficult problem arises in considering interdependencies between chord functions and tonality; chord functions, which is a notion that represents chord transition properties, are changed by local modulations, as H. Riemann pointed out.

In other words, a single transition matrix in conventional H(S)MM is no longer sufficient to analyze chord progressions when considering local modulations. Therefore, this study employs the idea of the neural hidden Markov model, which can adjust the hidden state transition probability by the contexts, and extends it to the semi-Markov model. The neural networks can utilize additional contexts, such as preceding chord sequences, pitch-classes, and beat information, for calculating categorical distributions that comprise the HSMM. Experiments show the added contexts considerably improve the model's generalization performance in terms of perplexity.

According to the aforementioned H. Riemann's view, tonality can be recognized by analyzing chord transition properties. We further introduce a teacher-student architecture to classify tonalities. While the teacher model equips the elaborated neural network for calculating the transition probability, the student model simplifies it to learnable matrices like a conventional HSMM. We prepare multiple student transition matrices and expect them to represent prototypes of tonalities. The student model classifies a predicted (labeled) chord sequence into a tonality by comparing the count of chord transitions with the transition probability matrices. Experiments show that the three-students model is the most consistent with the human analysis in terms of the F1-score; obtained three students can be interpreted as major, minor, and dorian modes, respectively. The transition matrices of the model reflect the difference between tonalities, consistently with known functions of tonic, dominant, and subdominant.

Thus, the neural HSMM and the extension of teacher-student architecture enable an unsupervised machine to recognize chords, chord functions, and tonality. We consult J. S. Bach's four-part chorales as the corpus of this study and give qualitative analysis in comparison with the conventional harmony theory. The consistency between the self-emergent chord functions and the known harmony theory suggests the potential of the proposed model to apply to a wider variety of music styles.

Keywords: Unsupervised Learning; Hidden Semi-Markov Model; Neural Network; Music Harmony Analysis; Chord Segmentation; Chord Function Recognition; Tonality Recognition.