氏名	上原　由衣		
学位の種類	博士（情報科学）		
学位記番号	博情第 483 号		
学位授与年月日	令和 4 年 9 月 22 日		
論文題目	Unsupervised Recognition of Chords, Functions, and Tonality		
論文審査委員	主査	東条 敏	北陸先端科学技術大学院大学　教授
		上原 隆平	同　教授
		鵜木 祐史	同　教授
		北原 鉄朗	日本大学文理学部　教授
		持橋 大地	統計数理研究所数理・推論研究系　准教授

論文の内容の要旨

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.

論文審査の結果の要旨

人工知能が人間と同様の創造性を有するかは，情報科学のみならず未来の人間社会に大きな影響を与える問いである．この試金石として，機械学習を用いた技術によって，楽譜に書かれた音楽において計算機が人間同様に和声や調，和声の機能を認識できるかは，そもそも聴覚や情動を持たない計算機に人間の模倣可能性を問うことの第一歩として重要な意味を持つ．これまで音楽情報処理は検索や推薦など実用システムの開発が主であり，楽曲の構造理解においても，調を与えての和声機能認識，和音列を与えての調認識など，前提知識を必要とした認識問題が主であった．本研究で取り組んだ課題は，和音・和声機能・調をいずれも前提とすることなく，相互に自律的に(unsupervised)発見する試みである．実際，和音列と調認識の関係理解は人間にとっても難しい問題である．Brahms のオルガンコラール op.127, no.4 の冒頭部分や，Faure の最晩年の弦楽四重奏曲の第一楽章は音楽学者にとっても一意の調解釈は困難である．本研究では，これら困難な問題を念頭に置きながら，その取り掛かりとして J. S. Bach の四声体コラールをターゲットとした．計算機実験においては，まず，一つの和音を発出するのは一つの隠れ状態であるとして，これら隠れ状態の連鎖が和音の連鎖を生み出すマルコフ過程を仮定する．しかし，一和音の長さは和音によって異なるため，時系列モデルにおいて隠れ状態の時間持続を組み入れられるセミマルコフモデルとした．さらに，和音列が特定の機能を持つかどうかは調というコンテクストによる．このコンテクストを与えるために，セミマルコフモデルに LSTM (long short term memory) を付加したニューラル隠れセミマルコフモデルを用いた．隠れ状態の数の妥当性においては perplexity を評価メトリックとした．和音は純粋に同時点で鳴っているピッチクラスの集合とし，和音名を周到に避けた．この設定で和音群は 8 個のクラスターを形成し，それぞれが隠れ状態をなし，かつその隠れ状態間の遷移関係のうちいくつか顕著に現れる遷移が和声機能であるとした．さらに，この遷移関係を観察することで調認識がなされ，調間のクラスターを Teacher-Student モデルによって形成した．以上により，和音・和声機能・調を相互依存的に自律的に発見するという試みは，音楽ジャンルの制限がある中では実現され，人工知能のフィージビリティへの知見に寄与するところが大きい．よって博士（情報科学）の学位論文として十分価値あるものと認めた．