

深層学習を用いた音楽の和音進行の推定

陳 偉博^{1,a)} 小坂 直敏^{1,b)}

概要：近年、SNS と文化産業の発展に伴い、自作の音楽を発表する場が増加している。しかし、音楽の制作、すなわち作曲や変奏などの工程は、専門知識を持っていないければその実現は困難で、自動作曲技術の進展とその利用が望まれる。筆者らは、自動作曲技術中の自動変奏工程を支援することを目指して、与えられた譜面からその曲の和声進行を自動的に抽出する問題について検討した。まず、入力した譜面を自然言語の構文解析のように分節し、その分節結果から指定した形の入力データに変換した。次に、入力データを DNN と LSTM を利用した二段階構造の和音推定システムへ入力して、和音を推論する。ベートーヴェンのソナタ全 32 曲の譜面と和声のデータベースを入力としてネットワークの学習を行った。本システムのセグメンテーション結果と和音推定結果について、音楽的に判断された正解結果と比べ、それぞれ正答率 94.4% と 88.5% で推定できた。

キーワード：GTTM、クロマベクトル、DNN、LSTM

1. はじめに

楽曲の作成、変奏は音楽家の専門知識を要するために、一般人には困難である。しかし、自動作曲技術が利用できると、楽曲作成、SNS、ゲームなどに応用でき、一般人も作曲を十分に楽しむことができる。

作曲には、メロディ生成、和声付け、伴奏付け、リズム楽器の付与、変奏、編曲など、数多くの工程がある。われわれは、クラシック音楽の変奏曲のように、一つのテーマに、希望する曲の情動を与えて、さまざまな変奏を自動的に作ることを大きな目標にしている。変奏には、メロディのアイデンティティを保ってリズムや調を変形させる歌謡的変奏と、和声進行を保って、メロディは変更するものとある。

変奏は、楽器の割り当てなどを配慮しつつメロディの同一性は保つ編曲とは異なる。また、変奏曲で最初に提示されるテーマも、楽曲としての魅力を出すため、和声進行そのものではなく、メロディとリズムも備わっている。本研究では、和声進行から自動変奏を生成することを目標とするが、その際、テーマさえも和声進行が陽に与えられていないため、まず、譜面からの和声進行を推定することを研究の目的とする。

以下では、GTTM 理論値に基づくセグメンテーションを行い、深層学習法として LSTM を用いて和声進行を推定した結果について述べる。

2. 提案手法

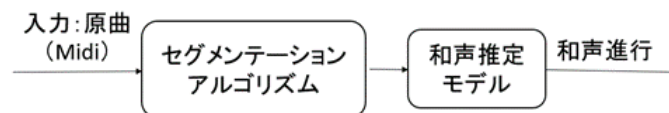
2.1 和音と和声進行

和音は、高さが異なる複数のピッチの楽音が複合したものである。古典的な西洋音楽の音楽理論では、3 和音を基本として考えることが多く、ポピュラー音楽では 7 和音を基本として考えることが多い。

和声進行は和音の時系列であり、音楽の表層構造の背後にある構造である。言い換えると、楽曲は複数の和音の時系列の上に存在して進行する。和音の役割として、メロディを装飾し、音楽表現を豊かにする。

2.2 提案手法の概要

図 1 に本研究で提案する和音推定システムを示す。原曲の MIDI データを入力すると、まず GTTM ルールを実装したセグメンテーションアルゴリズムでセグメントのデータを作る。次にセグメントデータを擬似クロマベクトルに変換し、DNN と LSTM を用いた二段階の和音推定モデルへ入力して、和音を取得する。取得した和音をまとめると、入力曲の和声進行を取得できる。以下、このシステムの各部について、詳細に述べる。



1 東京電機大学
Tokyo Denki University
a) 19fmi21@ms.dendai.ac.jp
b) osaka@mail.dendai.ac.jp

図 1 和音推定システム

2.3 セグメンテーション

Grouping Analysis (3)

GPR3 (change) Consider a sequence of four notes, n2-n3 may be heard as a group boundary, if

- a. register
- b. dynamics
- c. articulation
- d. Length changes.



図 2 GTTM のルールの例([1] より引用)

GTTM は、Lerdahl と Jackendoff により提案された。図 2 に示すように音符の情報から作ったルールに従って、階層的な構造を構築していく認知的音楽理論である[2]。

GTTM による音楽の構造解析は、音楽の認知に必要とされる以下 4 つのサブ理論から成る。それらはグルーピング構造、拍節構造、タイムスパン簡約、延長的簡約である。本研究は、グルーピング構造(Grouping Structure)を中心にして検討する。これにより、音楽的なまとまりを階層的に解析する。

グルーピング構造の解析では、ルールに従って、自然言語の構文解析のように楽曲の動機やリズム変化から曲を分節し、その分節の結果はグループと言われる。更に、階層的な分節結果も取得できる。

しかし、ルールが数多く存在するために、ルールの中に矛盾、または競合がある場合はよくある。この問題を解決するために、浜中らはパラメータを使って、各規則の重要度を決めて、境界になれるかどうかを分析し、矛盾を解決する手法を提案した[3]。図 3 はその手法の適用例を示す。ルール 3a とルール 2b とが競合する場合、パラメータを用いた計算結果によって 3a の方が重要と判断され、ルール 3a のところに境界を指定する。



図 3 ルールの競合を解決する例

本研究は、浜中らの提案手法を参照し、グルーピング構造のルールを用いて、セグメンテーションアルゴリズムを Python で実装した[3]。[3]と異なるのは、本研究は和音を推

論するため、階層的なグルーピング結果は不要である点である。

2.4 データ処理

データは訓練データと、推論用データの 2 つの部分がある。訓練データは BPS-FH (Beethoven Piano Sonata With Function Harmony) データセット[4]から作り、推論データは 2.3 のセグメンテーションアルゴリズムの出力データから作る。

BPS-FH データセットには、ベートーヴェンの全 32 曲のピアノソナタの譜面と和声進行がある。学習時には、大量のデータが必要である。BPS-FH にある約 7500 個のセグメントのデータ数は不十分と考えられ、そこでデータの数を拡張することが必要である。

データの拡張手法について、一般的に使われる手法は 32 曲全てを転調することである。これにより、1 オクターブ内の 12 個の半音に転調し 1 曲から 12 曲が得られる。最終的に、384 曲を転調による増分を含んで学習データとして使う。セグメントの数は 136128 である。

2.5 和音推定モデル

提案する 2 段階構造の和音推定モデルの構造を図 4 に示す。12 次元の擬似クロマベクトルデータを取得すると、まず第 1 段階の DNN モデルへ入力し、推論する和音を取得する。その次に、第 1 段階で取得した推定したいセグメントの前 3 セグメントの和音を擬似クロマベクトル構造に変換し、推定したいセグメントの擬似クロマベクトルと一緒に 4 連続セグメントの入力データを作る。その入力データを第 2 段階の LSTM モデルへ入力し、第 2 段階の和音を推論する。最後に、二つの推論和音を選択アルゴリズムに入力すると、より精度の高い最終結果を取得する。

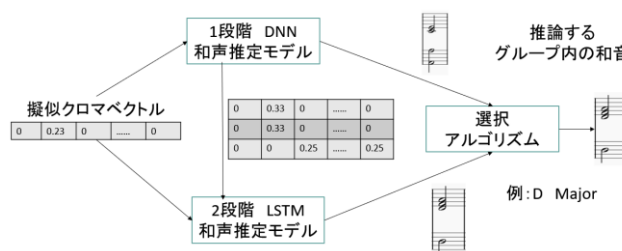


図 4 和音推定モデル構造

以下、その和音推定システムをさらに詳細に述べる

2.5.1 セグメントデータと擬似クロマベクトル

ある和音のクロマベクトルは、信号の周波数領域で構成要素となる周波数の周波数の振幅強度を特徴量としたベクトルである[5]。ここでは、この考えを MIDI データに適用し、ある和音の構成要素となるピッチのヒストグラムを特徴量とし、以下これを擬似クロマベクトルと呼ぶ。

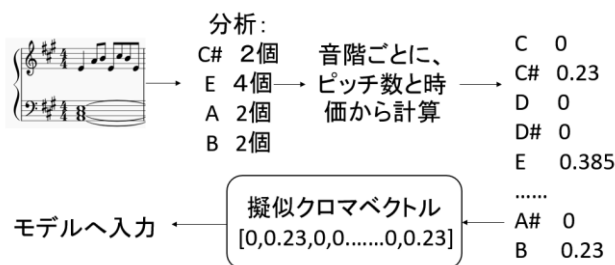


図 5 擬似クロマベクトル処理の例

図 5 に示す通り、セグメント化された MIDI データを読み込んだ後、12 音階ごとにピッチの頻度を計算し、各音符の音価（長さ）を重みとして 12 次元の擬似クロマベクトルを算出する。

2.5.2 DNN モデル

DNN モデル構造を図 6 に示す。入力層が 12 次元で、第 4 層の全接続層が 300 次元、出力が 84 次元である。今回処理できる和音は、M, m, M7, m7, Dominant7, dim, dim7 の 7 種類を定義している。これが 12 音あるため、合わせて 84 種類とした。また、転回形まで考慮すると、先和音は 3、3、4、4、4、3、4 種で合計 25 種類あり、これが 12 個のピッチに対して存在するため、全和音数は 300 となる。

出力を 84 次元の one-hot ベクトルとしたため、そのベクトルから和音を番号に変換し、簡単に推論和音を取得できる。

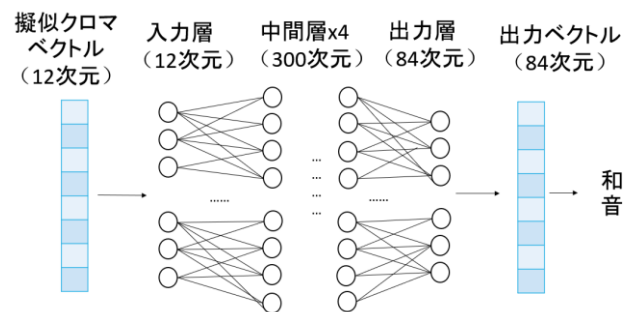


図 6 DNN モデル

2.5.3 LSTM モデルの入力データ

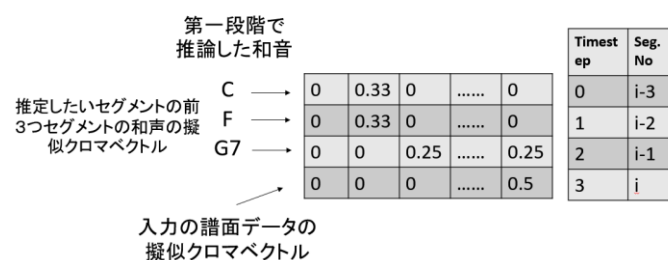


図 7 入力データの構造

入力データの作成方法を図 7 に示す。入力データは 4 連続和音列を用いる。前の 3 和音列は、推論したい該当セグメントの前 3 セグメントの和音から変換した 12 次元ベクトルである。その和音は第 1 段階の DNN モデルから取得する。和音が入力されると、擬似クロマベクトル構造と同じ 12 次元のベクトルを作る。例えば、C major の和音は C,E,G の 3 つで、12 次元のベクトルに、C,E,G のところは $1/3=0.333$ に設定し、他は 0 に設定する。最終的に取得するのは $[0.333, 0, 0, 0, 0.333, 0, \dots]$ である。

図 7 に示す通り、G7 等の 7 の和音の場合、和音の構成要素数が 4 つあるために、構成要素に該当する次元は $1/4=0.25$ に設定する。

最後に、3 つの和音ベクトルと推定したいセグメントの擬似クロマベクトルから、4 つの和音系列の入力データに変換する。

2.5.4 LSTM モデル

図 8 に本提案で用いる LSTM モデルを示す。和音の 4 系列を 4 層の LSTM モデルに入力する。LSTM モデルは、推定したいセグメントの譜面情報（擬似クロマベクトル）と共に、次の和音を推論する。これにより、第 1 段階推論の誤りを修正できる。例として、前 3 つのセグメントの和音が C,F,G7 の場合、4 つ目のセグメントについて、第 1 段階の推論は D（誤り）が、第 2 段階は、C、F、G7 の和音系列と 4 つ目セグメントの譜面から、正解の C 和音を取得でき、選択アルゴリズムで推論結果を正解として修正可能である。

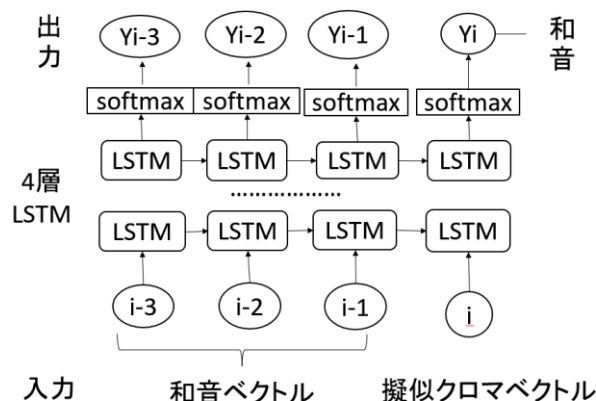


図 8 LSTM モデル

2.5.5 選択アルゴリズム

2 つの段階の推論和音を入力の前 3 セグメントの和音と一度比較する。第 1 段階の結果と第 2 段階の結果と違う場合、和音は譜面に多く含まれる方を正解と判断する。ただし、両方の比率が同じ場合、2.5.3 の説明の通り、第 2 段階の和音を選択する。

3. 実験結果と考察

3.1 セグメンテーション

実験対象として、ベートーヴェンの 24 バリエーションの No.15 変奏曲の MIDI ファイルをセグメンテーションアルゴリズムへ入力した。結果の一例を図 9 と図 10 に示す。図 9 は音楽常識的に判断されたセグメンテーション結果で、図 10 はアルゴリズムで自動的に取得したセグメンテーション結果である。



正解

図 9 正解セグメンテーション



実験結果

図 10 セグメンテーション実験結果

図 10 の例は推定結果が正解と全く同じであるが、全曲のセグメンテーションを分析すると、最終の正答率は 94.4%であった。

3.2 モデルの精度

訓練データを DNN モデルと LSTM モデルへ入力し、学習した。このときのモデルの精度を表 1 に示す。表の通り、第 2 段階は、第 1 段階の結果を多少修正したため、精度が第 1 段階より高いといえる。

表 1 2 つ段階のモデル精度

a)DNN 推定モデル

学習正答率	推論正答率	Dim7 の誤り排除
92.5%	72.5%	76.18%

b)LSTM 推定モデル

学習正答率	推論正答率	Dim7 の誤り排除
87.5%	78.5%	80.17%

dim7 和音は構成音が特殊のため、データベースの正答ラベルとは異なる例がよくある。例えば、Cdim7 和音と E $\flat$ dim7 和音は、構成音が全く同じ E $\flat$ 、G $\flat$ 、A、C である。このため、擬似クロマベクトルを入力すると、正解の和音の推定精度は低くなる。

3.3 和音推定

3.1 の結果から、セグメントの擬似クロマベクトルを作成する。そのデータを和音推定システムへ入力した結果の例を図 11 に示す。赤い文字は誤りと判断された結果である。全曲の和声推論の正答率は 88.5%であった。



図 11 和声推論の結果

3.4 考察

今回の結果から見ると、簡単な曲、即ち dim 和音が少ない曲の場合、提案した和音推定システムは精度が高く和声進行を抽出できた。一方モデルの精度を見ると、dim 和音のデータが数多く存在する場合、その正答率はやや低い。擬似クロマベクトルは、非和音が多くある場合、和音の構成要素以外の存在比率が増し、推定精度に限界があるため、と考えられる。

3.5 書き換え規則と変奏生成

本来、最終的にやりたいことは、和声進行を取得した上での変奏である。本検討では、和声進行の推定が最も大きな課題であり、変奏生成については、今後検討していく。ここでは、和音の一つに書き換え規則を適用することにより、変奏生成を行った事例を紹介する。

推論和音の最終結果を入力とし、個別の和音に対する書き換えルールを作成する。これを全和声に適用し、変奏曲の譜面を生成する。図 12 は、入力と声声を任意に作成し、これを書き換え規則に照らして変奏生成をした譜例である。



図 12 変奏曲の生成

次に、実際に 3 の提案手法により和音推定した結果に基

づいた変奏生成事例を示す。図 13 に示す書き換え規則を用いて、変奏曲を作成した。譜面を図 14 に示す。これらは、誤りの和音のところ違和感はあるが、全体では正常な曲と判断される。

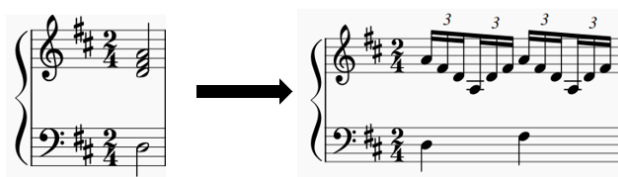
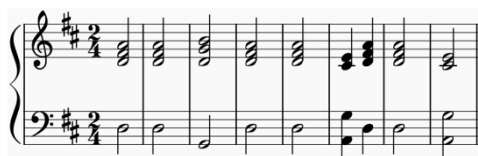


図 13 実験で使用する書き換え規則

和声進行: D D G D D A7 D D A7



書き換え規則



図 14 生成した変奏曲デモの譜面

4. おわりに

本研究は、GTTM 理論と深層学習技術を用いて、入力 of 音楽譜面から自動的に曲をセグメンテーションして、和声進行を推定するシステムを提案した。譜面情報から擬似クロマベクトルに変換した後、DNN と LSTM2 つの段階がある和音推定モデルに入力し、結果を取得し、最後に、選択アルゴリズムを用いて推論の精度を一層向上させた。簡単な曲の和声処理する場合、94.4%のセグメンテーション正答率と 88.5%の和音推論正答率を得た。しかし、dim7 等、複雑な和音が多くある曲に対しては改善する必要がある。

謝辞

本研究に日頃熱心な議論をしていただいた研究室のメンバーに心から感謝します。

参考文献

- [1] Keiji Hirata, Satoshi Tojo, Masatoshi Hamanaka: Techniques for Implementing the Generative Theory of Tonal Music, ISMIR 2007 (7th International Conference on Music Information Retrieval) Tutorial, September, 2007.
- [2] F. Lerdahl, R. Jackendoff, "A Generative Theory of Tonal Music", Cambridge, Mass. MIT Press, 1983.
- [3] 浜中雅俊, 平田圭二, 東条敏, "GTTM グルーピング構造分析の実装: ルールを制御するパラメータの導入", 情処研報. MUS, 56, 1-8, 2004.
- [4] Tsung-Ping Chen and Li Su, "Functional Harmony Recognition with Multi-task Recurrent Neural Networks," ISMIR, September 2018.
- [5] C. Harte and M. Sandler, "Automatic Chord Identification Using a Quantized Chromagram," Proc. of Audio Eng. Soc. May 2005.
- [6] 松岡, 渡部, 「音楽のコード認識技術とその応用」, NTT DOCOMO テクニカル・ジャーナル, 2017, Vol.25, No.2