

本著作物の著作権は情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。利用にあたっては、著作権法ならびに[情報処理学会倫理綱領](#)に従ってください。

The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). The material has been made available on the website by the author(s) under the agreement with the IPSJ. Please be complied with Copyright Law of Japan and [Code of Ethics of the IPSJ](#) if any users wish to reproduce, distribute, or modify the whole or any part of the material.

Attention-based LSTMを用いた クラシック楽曲の時代識別

森山 治紀^{1,a)} 平賀 譲^{1,b)}

概要: クラシック音楽では、バロック・古典派・ロマン派といった時代区分があるように、楽曲が書かれた時代によって書法に違いがみられる。従来の音楽理論がそれらの違いについて定性的な説明を行う一方で、音楽情報処理では、時代や作曲家ごとの音楽的特徴を定量的に取り扱う研究が行われている。本研究では、時代ごとの音楽的特徴を定量的に表現する手法として、和音系列から楽曲の時代区分を識別するニューラルネットワークモデルを提案する。本研究で提案する手法は、次の2つのプロセスからなる。(1) word2vecによる単音の分散表現の獲得。(2) Attention機構を導入したLong Short-Term Memory (LSTM)による識別モデルの構築。(2)の識別モデルの入力として、(1)で得られた分散表現を用いる。(1)については、獲得したベクトルをt-Distributed Stochastic Neighbor Embedding (t-SNE)を用いて二次元に圧縮すると、同じピッチクラスの音がクラスターを形成し、調の構成音と非構成音が分離し、さらに構成音は3度間隔で環状に並ぶことが確認された。(2)については、クラシック音楽の作曲家13人の楽曲2,396曲を対象に、バロック・古典派・ロマン派の3クラス分類を行なった結果、モデルの識別精度は約76%であった。以上より、本研究で提案する手法は、時代による音楽的特徴の違いを定量的に扱う方法として有用であると考えられる。

キーワード: 時代識別, 様式分析, 分散表現, word2vec, Attention機構, LSTM

Attention-based LSTM for Period Identification of Classical Music

HARUKI MORIYAMA^{1,a)} YUZURU HIRAGA^{1,b)}

Abstract: Classical music is divided into such eras as Baroque, Classical, and Romantic periods, each of which has its characteristic style. While traditional music theory qualitatively analyzes such peculiarities, the quantitative difference among periods or composers are studied in music information retrieval. Aiming for quantitative analysis, this paper proposes a neural network that classifies a series of chords according to when it was inferred to be written. The proposed model is comprised of two stages: generation of distributed representations of chord tones using word2vec and chord sequence classification with Attention-based Long Short-Term Memory (LSTM). The distributed representations are employed as the input to the LSTM classifier. Some musicologically reasonable phenomena were confirmed in the obtained word2vec vector space visualized in 2-D with t-Distributed Stochastic Neighbor Embedding (t-SNE). For example, notes belonging to the same pitch class were gathered into a cluster, scale notes and non-scale notes were separated, and the scale notes formed a cyclic form of diatonic third intervals. The LSTM classifier achieved 76% accuracy in 3-class (Baroque/Classical/Romantic) classification where 2,396 works of 13 classical music composers were used. The results showed the capability of the model to capture the stylistic difference among periods quantitatively.

Keywords: Period identification, Style analysis, Distributed representation, word2vec, Attention, LSTM

1. はじめに

クラシック音楽には、バロック・古典派・ロマン派といった時代区分が存在する。時代背景の変化とともに、作曲家

¹ 筑波大学大学院 図書館情報メディア研究科
Graduate School of Library, Information and Media Studies,
University of Tsukuba

^{a)} s1821637@s.tsukuba.ac.jp

^{b)} hiraga@slis.tsukuba.ac.jp

が新しい響きや表現を求めて既存の様式を発展させていく過程で、各時代を特徴付ける音楽様式が形成されてきた。

今日、長調・短調の区別や曲全体を統一する拍子が音楽の基本様式として認識されている。それらの概念が確立されたのはバロック時代のことである。音楽におけるバロックは17世紀初頭から18世紀中頃を指し、社会的には絶対主義の時代であった。この時代の音楽は、王侯貴族の権威を誇示する宮廷の音楽として作られたものであり、作曲技法の点では厳格対位法から自由対位法への転換および通奏低音の使用が特徴的である [1], [2]。

18世紀中頃から19世紀初頭にかけて、市民階級の台頭および啓蒙主義の浸透を背景に、宮廷のための音楽は市民のための音楽へと変容していった。この時代の音楽が古典派である。古典派音楽を特徴付ける作曲技法の変化として、対位法や通奏低音の廃止、ポリフォニーからホモフォニーへの転換がある [1], [2]。

19世紀には、聴衆のさらなる拡大とロマン主義の繁栄を背景に、作曲家は芸術家として独創性を追求するようになった。このロマン派時代の音楽は、半音階を駆使した和声や感情的な旋律による多彩な表現が特徴的である [1], [2]。

そして、20世紀以降、調性・和声・拍子といったそれまでの原則を否定する音楽様式が現れ、現代音楽へと続いている。 [1], [2]

従来の音楽理論では、各時代の音楽的特徴の違いについて和声学や楽式論に基づく分析が行われる [2], [3], [4], [5], [6], [7]。それらのアプローチによる分析は、定性的で形式的記述のない知識を含んでいるため、計算機上での実現が困難である。一方、音楽情報処理の観点からは、音楽に対して情報学の知見を活用することにより、時代や作曲家ごとの音楽的特徴を定量的に分析する枠組みを提示できる可能性がある。一つの方法として、音楽的に類似するものが特徴空間上で近い位置に配置されるような特徴ベクトルを用いて、音楽を表現することができれば、音楽を定量的に取り扱う理論の構築に貢献できると考えられる。

そこで、本研究では、クラシック音楽を対象に、時代ごとの音楽的特徴を捉えた定量的表現手法の確立を目指して、和音系列から楽曲の時代区分を識別するニューラルネットワークモデルを提案する。モデルを構築するにあたっては、自然言語と音楽の類似性に着目し、自然言語処理の分野で得られた知見を取り入れることで、よりの確に音楽の時系列的性格を捉えることを目指す。

2. 関連研究

2.1 作曲家識別

本研究に深く関連する研究として、楽譜情報に基づく作曲家識別がある。Pollastriらは、メロディを入力とし、Hidden Markov Model (HMM) を用いて作曲家を識別する研究を行った [8]。旋律音程と音長比を特徴量として、

Mozart/Beethoven/Dvořák/Stravinsky と The Beatles の5クラス分類を行い、42%の識別精度を得ている。

長谷川らはラルーらの様式分析手法 [9] の定量化を試みる研究を行った [10]。様々な時代・地域のクラシック音楽を対象に、独自に定義した14種の特徴量を用いて26作曲家の正準判別分析を行い、56%の精度を得ている。判別分析の結果に対して階層的クラスタ分析を行い、共通する時代や文化をもつと考えられる作曲家が特徴空間上で近い位置に配置されることを示した。

2.2 自然言語と音楽の類似性

自然言語と音楽の間にはいくつかの共通点がある。順序をもつ離散的記号列として表現される点や階層的構造をもつ点などがその例である。このような類似性に基づいて自然言語と音楽を対応付け、自然言語の枠組みを音楽に適用することで音楽の構造を理解しようとする研究がこれまでに行われてきた [11]。音楽分析に自然言語処理の技術を用いる研究も行われてきたが、これまでのところ、従来の音楽理論に対して新しい枠組を提示できるような成果が得られているとはいえない [12]。

2.3 自然言語処理

近年、自然言語処理においてニューラルネットワークモデルが大きな成果をあげている [13]。こうした研究の多くは、単語の分散表現 [14] を基に句や文の分散表現を生成し、それらを用いて種々のタスクを解いている [15]。単語の分散表現獲得からタスク処理に至るまでの各プロセスは、すべてニューラルネットワークにより実現される。単語の意味を的確に捉えた分散表現を獲得する手法として様々なものが提案されているが、それらの手法の多くは分布仮説 [16] という考えに基づいている。分布仮説とは、「ある単語の意味はその周辺に出現する単語が何であるかによって定まる」という考えである。word2vec [17] は、分布仮説に基づいて単語の分散表現を獲得するための代表的手法である。word2vec では、文章中のある単語を入力とし、その周辺に出現する単語を推論するニューラルネットワークモデルを学習することにより、単語の分散表現を得る。word2vec により得られる単語の分散表現は加法構成性を備えており、 $\vec{\text{king}} - \vec{\text{man}} + \vec{\text{woman}} \approx \vec{\text{queen}}$ [18] に代表されるように、単語の意味的関係をベクトル演算を用いて表現できることが示されている。

自然言語のように順序が意味をもつ時系列データを扱うためのモデルとして、Recurrent Neural Network (RNN) [19] が用いられている。通常の RNN では、入力系列が長くなると学習時に勾配消失を起こしやすくなり、系列内の遠く離れた要素の間に存在する依存関係の学習が困難となる場合がある [20]。これは誤差逆伝播法により損失の勾配を計算する際、系列の長さの分だけ重みを乗じる回数が増加

し、勾配の大きさが指数関数的に減少することが原因である。RNNのネットワークに新たなユニットを追加することで勾配消失の問題に対処したモデルが Long Short-Term Memory (LSTM)[21] である。LSTM では、遠距離要素間の依存関係を学習できることが示されており、自然言語処理の多種多様なタスクで使用されている。

現在の自然言語処理では、様々なタスクにおいて Attention 機構 [22] をベースとしたモデルが最も優れたパフォーマンスをあげている。Attention 機構とは、入力系列の各要素に対して重み付けをする手法である。このとき、入力系列のうち、タスクを解くうえで重要な情報をもつ要素に対して大きな重みを与えられるよう、ネットワークの学習が行われる。

2.4 和音の分散表現

自然言語と音楽の構造的類似性に着目し、word2vec を音楽の構成要素に適用する研究には先例が存在する。[23], [24], [25], [26], [27], [28], [29] は和音の分散表現を、[30], [31] はメロディの分散表現を獲得している^{*1}。前者の研究においては、楽曲中のある和音を入力として、その周辺に出現する和音を推論するニューラルネットワークモデルを学習することで、和音の分散表現を得ている。こうして得られた分散表現を二次元に圧縮した際の相対的位置関係を観察することで、和音間の音楽的関係が反映されていることを示唆する結果が得られている [24], [25]。しかし、音楽は自然言語ほど「意味」が明瞭でなく、曖昧であるため、得られた分散表現を評価するのが難しい。自然言語処理では、単語の分散表現を種々のタスクに活用して得られた結果から、分散表現の有用性を評価する。しかし、音楽情報処理では、タスクの処理に応用することで分散表現の有用性評価を試みる研究は、あまり行われていない。

本研究では、まず、word2vec を用いて単音の分散表現を獲得する。先行研究では、和音系列を学習することで、和音の分散表現を得ているのに対し、本研究では、より根本的な音楽の構成要素に着目し、和音からその和音を構成するそれぞれの音の分散表現を学習する。さらに、和音から得られた単音の分散表現を用いて和音の分散表現を生成し、和音系列の時代区分の識別タスクに応用する。これにより、各時代の音楽的特徴を定量的に捉える手段を模索するとともに、得られた分散表現の有用性の検証を行う。

3. 提案手法

3.1 単音の分散表現

本研究では、分布仮説における「単語」を「単音」に置き換えて、「ある音の音楽的性格は同時に鳴る音が何であるかによって定まる」と仮定する。この仮定に基づき、

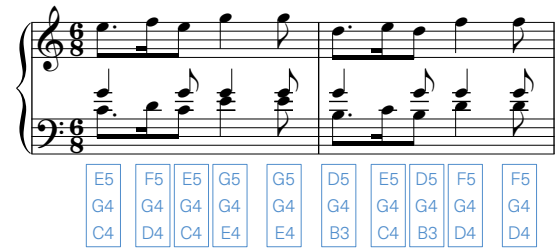


図 1 楽曲データから和音系列への変換

Fig. 1 Conversion from MIDI to chord sequence.

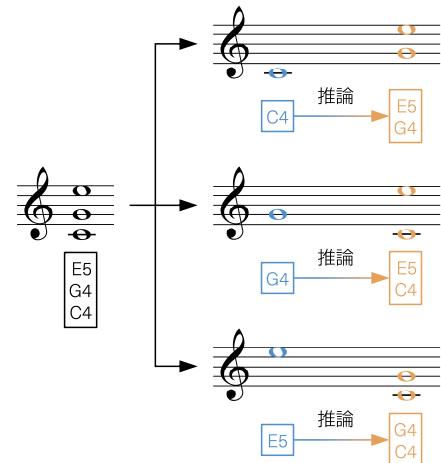


図 2 和音構成音を推論する skip-gram モデル

Fig. 2 Skip-gram model for chord tones.

word2vec を用いることで、各音のもつ音楽的性格を的確に捉えたベクトル表現の獲得を目指す。分散表現の学習のための楽曲データとして、クラシック音楽を対象に幅広い時代・地域の作曲家を含む 8,900 曲分の MIDI ファイルを用いる [32]。すべての楽曲について、長調の曲はハ長調、短調の曲はイ短調に移調することで、調性の違いによる影響を排除する。なお、移調の際、一般的なピアノの音域である A0 から C8 までの範囲から外れる音が生じる場合、その音はオクターブシフトすることで範囲内に収める。学習にあたっては、同時に鳴る音の音高の集合を和音と定義したうえで、楽曲全体を和音系列に変換する (図 1)。このとき、和音交替のタイミングは、一つ以上の和音構成音が変化した時点とする。word2vec における skip-gram モデルにない、ある和音の構成音の一つを入力とし、その和音の他の構成音を推論するニューラルネットワークモデルを構築する (図 2)。このモデルに和音の構成音を学習させることで、各音について 12 次元の分散表現を獲得する。

自然言語処理では、単語の分散表現は数百次元とするのが一般的であるが、これをそのまま適用するのは適当ではないと考えられる。本研究で取り扱う単音はピアノ音域内の 88 音であり、自然言語の語彙数と比べて数が少ないためである。頑健性のあるベクトル表現を得るためには、低次元で的確に特徴を捉える必要がある。したがって、本研究では自然言語処理で一般的な値より低い次元数を用いる。

^{*1} ただし、[28] は分散表現の獲得に word2vec ではなく、Pennington らの分散表現獲得手法である GloVe を使用している。

3.2 識別モデル

本研究では、Attention 機構を導入した双方向 LSTM を用いて、和音系列から楽曲の時代区分を識別するモデルを構築する。モデルの構成を図 3 に示す。

紙面の都合上、図 3 における入力とは 3 つの和音からなる系列となっているが、実際には、楽曲中に出現する連続した 64 個の和音を含んだ系列をモデルの入力として用いる。楽曲データから窓幅 64、ステップ幅 32 で和音系列を切り出していく。最後に幅 64 未満の系列が余った場合はそれを除いて入力ユニット集合とする（図 4）。入力ユニットを構成する各和音を構成音の音高の集合として考え、和音構成音に 3.1 節で獲得する単音の分散表現を割り当てることで、各和音はベクトルの集合に変換され、入力ユニットはベクトル集合の系列に変換される。モデル内部では、まず、Attention 機構により入力ユニットの各和音について構成音の重み付けを行い、加重和を各和音のベクトル表現とする。次に、得られた和音ベクトルの系列を双方向 LSTM へ入力し、Attention 機構を用いて、LSTM 層の各時点における隠れ状態に重み付けを行う。最後に、得られた特徴ベクトルに対して線形変換を行い、softmax 関数による正規化を経て、入力ユニットがバロック・古典派・ロマン派の

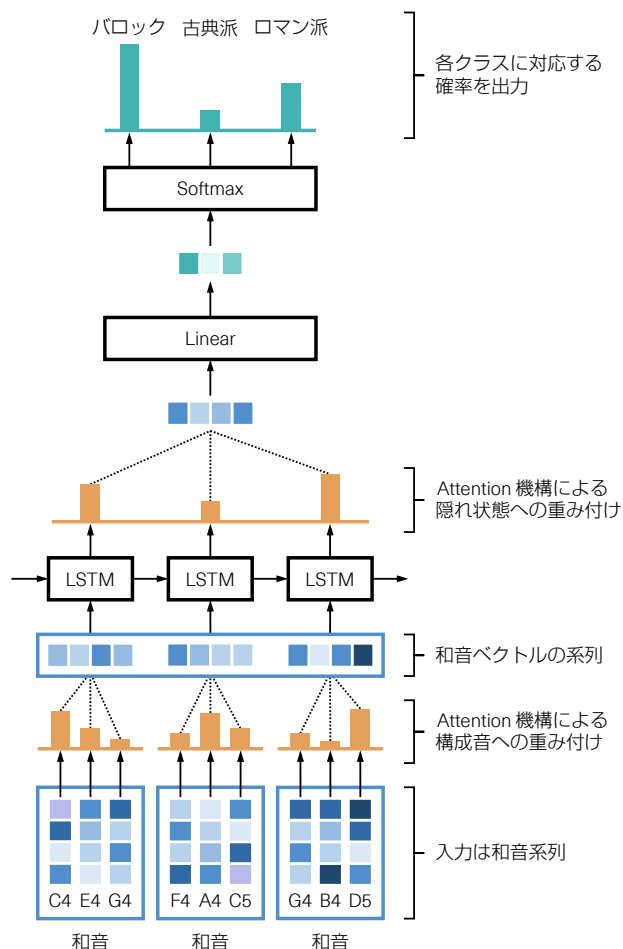


図 3 和音系列の時代区分識別モデル

Fig. 3 Period classifier of chord sequence.

表 1 評価用データ

Table 1 Test data.

時代区分	作曲家	曲数	入力ユニット数
バロック 4 作曲家・1,028 曲	A. L. Vivaldi	20	23,997
	G. P. Telemann	49	
	G. F. Händel	238	
	J. S. Bach	721	
古典派 2 作曲家・683 曲	F. J. Haydn	462	25,063
	W. A. Mozart	221	
ロマン派 7 作曲家・685 曲	M. Mendelssohn	46	24,988
	F. F. Chopin	107	
	R. Schumann	67	
	J. Brahms	127	
	A. L. Dvořák	94	
	P. I. Tchaikovsky	175	
	G. U. Fauré	69	
合計		2,396	74,048

クラスに属する確率をそれぞれ出力する。

モデルの評価は 5 分割交差検証により行う。3.1 節で利用したデータセット [32] からクラシック音楽の作曲家 13 人の楽曲 2,396 曲を選び（表 1）、各時代区分について曲数が均等になるように楽曲集合を 5 分割して訓練データ・テストデータとする。なお、楽曲ごとに和音系列長に差があるため、5 分割した楽曲集合のそれぞれにおいて、含まれる入力ユニット数は異なる。

なお、本研究の提案手法は、単音のベクトル表現を基に和音のベクトル表現を生成している点で、先行研究 [23], [24], [25], [26], [27], [28], [29] と異なっている。和音のベクトル表現を直接学習する場合、データセット中の出現回数が低い和音や一度も現れない和音については、適切なベクトル表現を得るのが困難、あるいは不可能という問題が生じる [33]。単音ベクトルの獲得を経ることで、本手法ではこの問題を解消している。

4. 結果と考察

4.1 単音の分散表現

3.1 節で獲得した単音のベクトルを t-Distributed Stochastic Neighbor Embedding (t-SNE)[34] を用いて二次元に圧縮し、可視化した（図 5）。注目すべき点は 3 つある。

第一に、同じピッチクラスの音がクラスタを形成している点である。オクターブ違いの音は本質的には同種の音として捉えられる。オクターブについての同値関係を特に与えていないにも関わらず、それらの音が近接する位置に配置されていることから、モデルはオクターブについての類似関係を的確に捉えることができていると考えられる。

第二に、調の構成音と非構成音、すなわちハ長調・イ短調における白鍵音と黒鍵音が分離して現れている点である。構成音と非構成音の間には音楽的な意味、つまりは用法に大きな違いがあると考えられる。モデルに区別を与え

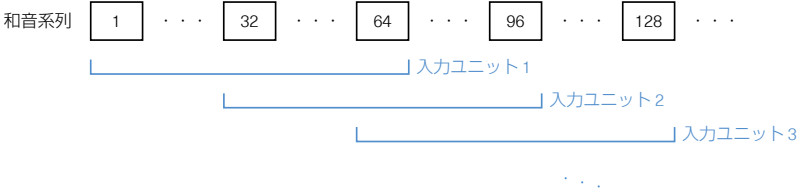


図 4 和音系列から入力ユニットへの変換
Fig. 4 Conversion from chord sequence to input unit.

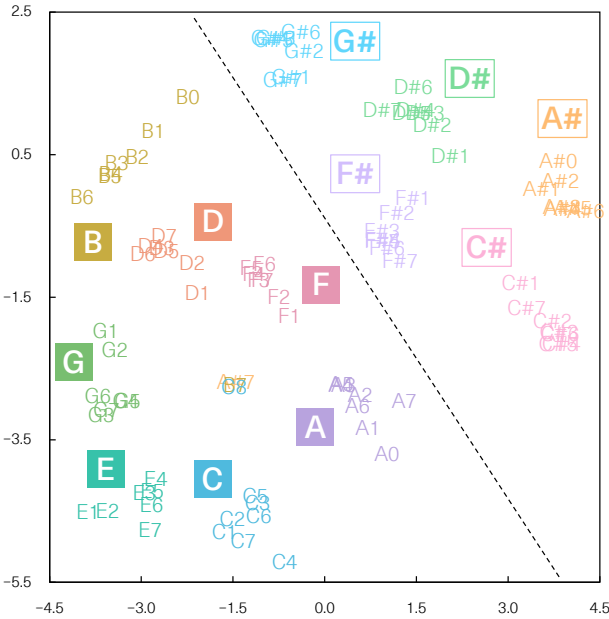


図 5 単音分散表現の t-SNE による二次元可視化
Fig. 5 Visualization of distributed representations for notes using t-SNE.

ていないのにも関わらず、それらが特徴空間上で分離することは、音楽の意味を捉える力がモデルに備わっていることを示す結果であると考えられる。

第三に、調の構成音が C-E-G-B-D-F-A-C のように全音階的な 3 度間隔で環状に並んでいる点である。長調の I の和音なら C-E-G, VI の和音なら A-C-E といったように、調の任意の固有三和音は、この循環列の連続する 3 音から構成されている。このように 3 度音程は和音構成において基礎となる音程関係であり、特徴空間上の隣接関係として 3 度音程が現れることは、興味深い結果である。モデルに特別な情報を与えていないことは、他の 2 点と同様である。

本研究で獲得した音の分散表現は、少なくとも上記の点において、各音の間に存在する音楽的な意味の関係を捉えることができていると考えられる。

4.2 識別モデル

3.2 節の時代識別タスクにおいて、各クラスの入力ユニットが実際にどのクラスと識別されたかを表す混同行列を図 6 に示す。図中の数値は、5 分割交差検証の各回において、識別された入力ユニット数の平均値である。また、こ

識別 \ クラス	バロック	古典派	ロマン派	合計
正解				
バロック	4,080	448	271	4,799
古典派	853	3,709	450	5,013
ロマン派	644	877	3,476	4,998

図 6 時代識別の混同行列
Fig. 6 Confusion matrix of period classification.

表 2 時代識別モデルの性能評価
Table 2 Performance evaluation of period classifier.

クラス	精度	適合率	再現率	F 値
バロック	—	73.0%	85.0%	78.4%
古典派	—	74.1%	74.0%	73.7%
ロマン派	—	82.7%	69.6%	75.5%
平均	76.0%	76.6%	76.2%	75.9%

の識別結果から算出される性能評価値を表 2 に示す。
モデルによる識別精度は約 76% であったが、クラスごとの結果に着目すると、時代が下るほど再現率が低下し、適合率は上昇していることがわかる。これは、後続する時代は先行する時代の様式を踏襲・応用しているために後続時代と先行時代の区別が難しくなり、後の時代ほど誤認識するケースが増えているためではないかと考えられる。

5. おわりに

本研究では、クラシック音楽における各時代の音楽的特徴を捉えた定量的表現手法の確立を目指して、和音系列から楽曲の時代区分を識別するニューラルネットワークモデルを構築した。はじめに、word2vec を用いて単音の分散表現を獲得した。得られた分散表現を t-SNE により二次元に圧縮すると、同じピッチクラスの音がクラスタを形成し、調の構成音と非構成音が分離し、さらに構成音は 3 度間隔で環状に並ぶことが確認された。このことから、得られた分散表現は、各音の間に存在する音楽的な関係を捉えられていると考えられる。次に、Attention 機構を導入した LSTM を用いて、和音系列の時代識別モデルの構築を行っ

た。モデルの入力として、楽曲中に出現する連続した 64 個の和音からなる系列を用い、クラシック音楽の作曲家 13 人の楽曲 2,396 曲を対象にバロック・古典派・ロマン派の 3 クラス分類を行ったところ、モデルの識別精度は約 76% であった。識別で良好な結果が得られたことから、提案手法は各時代の音楽の特徴を定量的に捉える手法として有用であると考えられる。

今後は、本研究で用いた手法を応用し、作曲家識別に取り組む予定である。さらに、より応用的なタスクへの活用として、音楽的類似度に基づく楽曲検索や作曲家ごとの特徴を捉えた楽曲生成などの実現を目指していきたいと考えている。

参考文献

- [1] 岡田暁生：西洋音楽史——「クラシック」の黄昏，中公新書（2005）。
- [2] 田村和紀夫：アナリーゼで解き明かす——新 名曲が語る音楽史——グレゴリオ聖歌からポピュラー音楽まで，音楽之友社（2008）。
- [3] 長谷川良夫：対位法，音楽之友社（1955）。
- [4] 石桁真礼生：新版 楽式論，音楽之友社（1966）。
- [5] 熊田為宏：演奏のための楽曲分析法，音楽之友社（1974）。
- [6] 島岡謙，野田暉行，尾高惇忠，川井學，佐藤眞，永富正之，南弘明，浦田健次郎，野平一郎：総合和声——実技・分析・原理，音楽之友社（1998）。
- [7] 林達也：新しい和声——理論と聴感覚の統合，アルテスパブリッシング（2015）。
- [8] Pollastri, E. and Simoncelli, G.: Classification of Melodies by Composer with Hidden Markov Models, *Proc. First International Conference on WEB Delivering of Music (WEDELMUSIC'01)*, Nesi, P., pp.88-95, IEEE Computer Society (2001).
- [9] ヤン・ラルー，大宮眞琴：スタイル・アナリシス——総合的様式分析——方法と範例，音楽之友社（1988）。
- [10] 長谷川隆，西本卓也，小野順貴，嵯峨山茂樹：楽譜情報からの作曲家らしさ認識のための音楽特徴量の提案，情報処理学会論文誌，Vol.53, No.3, pp.1204-1215（2012）。
- [11] Lerdahl, F. and Jackendoff, R.: *A Generative Theory of Tonal Music*, MIT Press (1996).
- [12] 平田圭二，東条敏：バーンスタインの「答えのない質問」再考——計算論的音楽の理論の枠組みについて，人工知能学会全国大会論文集，Vol.JSAI2014, No.1K4-OS-07a-1, pp.1-4（2014）。
- [13] Goldberg, Y.: A Primer on Neural Network Models for Natural Language Processing, *JAIR*, Vol.57, pp.345-420 (2016).
- [14] Rumelhart, D.E., McClelland, J.L. and PDP Research Group (Eds.): *Parallel distributed processing—explorations in the microstructure of cognition—Vol.1*, Hinton, G.E., McClelland, J.L. and Rumelhart, D.E.: *Distributed Representations*, pp.77-109, MIT Press (1986).
- [15] 岡崎直観：特集「ニューラルネットワーク研究のフロンティア」——言語情報における分散表現学習のフロンティア，人工知能，Vol.31, No.2, pp.189-201（2016）。
- [16] Harris, Z.: Distributional Structure, *WORD*, Vol.10, No.2-3, pp.146-462 (1954).
- [17] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J.: Distributed Representation of Words and Phrases, *Proc. 26th International Conference on Neural Information Processing Systems (NIPS'13)*, Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z. and Weinberger, K.Q., Vol.2, pp.3111-3119, Curran Associates Inc. (2013).
- [18] Mikolov, T., Yih, W.-T. and Zweig, G.: Linguistic Regularities in Continuous Space Word Representations, *Proc. 2013 North American Chapter of the Association for Computational Linguistics—Human Language Technologies 2013 (NAACL-HLT 2013)*, Vanderwende, L., Daumé III, H. and Kirchhoff, K., pp.746-751, ACL (2013).
- [19] Elman, J.L.: Finding structure in time, *Cognitive Science*, Vol.14, No.2, pp.179-211 (1990).
- [20] Bengio, Y., Simard, P. and Frasconi, P.: Learning Long-Term Dependencies with Gradient Descent is Difficult, *IEEE Transactions on Neural Network*, Vol.5, No.2, pp.157-166 (1994).
- [21] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, Vol.9, No.8, pp.1735-1780 (1997).
- [22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I.: Attention Is All You Need, *Proc. 31st Conference on Neural Information Processing Systems (NIPS'17)*, Guyon, I., Luxburg, U. von, Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. and Garnett, R., pp.5998-6008, Curran Associates Inc. (2017).
- [23] Madjiheurem, S., Qu, L. and Walder, C.: Chord2Vec—Learning Musical Chord Embeddings, *Proc. Constructive Machine Learning 2016 (CML2016)*, pp.1-5 (2016).
- [24] Huang, C.-H.A., Duvenaud, D. and Gajos, K.Z.: ChordRipple—Recommending Chords to Help Novice Composers Go Beyond the Ordinary, *Proc. 21st International Conference on Intelligent User Interfaces (IUI2016)*, Nichols, J., O'Donovan, J., Conati, C. and Zancanaro, M., pp.241-250, ACM (2016).
- [25] Herremans, D. and Chuan, C.-H.: Modeling Musical Context Using Word2vec, *Proc. First International Workshop on Deep Learning and Music (DLM2017)*, Herremans, D. and Chuan, C.-H., pp.11-18 (2017).
- [26] Nikrang, A., Sears, D.R.W. and Widmer, G.: Automatic Estimation of Harmonic Tension by Distributed Representation of Chords, *Proc. 13th International Symposium on Computer Music Multidisciplinary Research (CMMR 2017)*, Kronland-Martinet, S., Ystad, S. and Aramaki, M., pp. 21-32, PRISM (2018).
- [27] 塚本康太，饗庭絵里子，南泰浩：単語埋め込みを利用した和音進行分析，情報処理学会研究報告，Vol.2019-MUS-122, No.9, pp.1-5（2019）。
- [28] 石田颯人，木村昌臣：度数表記と Chord2Vec を利用した楽曲類似度指標の提案，情報処理学会研究報告，Vol.2019-MUS-123, No.21, pp.1-5（2019）。
- [29] Chuan, C.-H., Agres, K. and Herremans, D.: From Context to Concept—Exploring Semantic Relationships in Music with Word2Vec, *CoRR*, Vol.abs/1811.12408, pp.1-21 (online), available from <<https://dblp.org/rec/bib/journals/corr/abs-1811-12408>>.
- [30] Hirai, T. and Sawada, S.: Melody2Vec—Distributed Representations of Melodic Phrases based on Melody Segmentation, *JIP*, Vol.27, pp.278-286 (2019).
- [31] Arronte Álvarez, A. and Gómez-Martin, F.: Distributed Vector Representations of Folksong Motifs, *Proc. 7th International Conference, Mathematics and Computation in Music (MCM 2019)*, Montiel-Hernández, M., Gómez-Martín F. and Agustín-Aquino, O.A., pp.325-332, Springer (2019).
- [32] Kunst der Fuge (online), available from <<http://www.kunstderfuge.com/>>.
- [33] Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T.: Enriching Word Vectors with Subword Information, *TACL*, Vol.5, pp.135-146 (2017).
- [34] Maaten, L. van der and Hinton, G.: Visualizing Data using t-SNE, *JMLR*, Vol.9, pp.2579-2605 (2008).