

時系列解析手法を用いた楽曲間類似度

澤井 賢一[†], 平田 祥人[†], 富岡 亮太^{††}, 合原 一幸^{†,‡}

[†] 東京大学 ^{††} 東京工業大学 [‡] ERATO 合原複雑数理モデルプロジェクト

本稿では、時系列解析の手法を用いて楽曲間の関係を調べる手法について述べる。ここで対象とする音楽情報は、楽譜から得られる記号情報とする。本研究では音楽を、作曲者というシステムから生成された時系列データとみなし、クロスリカレンスプロットという時系列解析手法を音楽情報に適用した。この手法を適用するにあたり、各時刻におけるデータ値の間に距離関数が必要であるが、本研究では二通りの距離関数を考えた。そして、一つ目の距離関数を用いることで、楽曲の雰囲気の種類度が測れる可能性を示した。また二つ目の距離関数によるクロスリカレンスプロットは、 n -グラムモデルを用いた楽曲の2クラス判別に関連があることを示した。

Music Similarity Measures using a Time Series Analysis Technique

Ken'ichi SAWAI[†] Yoshito HIRATA[†] Ryota TOMIOKA^{††} Kazuyuki AIHARA^{†,‡}

[†] The University of Tokyo

^{††} Tokyo Institute of Technology

[‡] ERATO Aihara Complexity Modelling Project, JST

This paper describes two applications of a time series analysis method for music information. In our study, we treat symbolic music information obtained from musical scores, and regard music as time series data. The time series analysis technique we used is the cross recurrence plot. It needs a distance function between two time points, and thus we propose two distance functions. With one of the functions, we suggested that we can quantify the atmospheres of musical pieces. We also show that the cross recurrence plot with the other distance function is related to the discrimination problem using the n -gram model.

1 はじめに

近年の科学技術の発達により、個人が所有、アクセスできる情報の量は急速に増えている。またそれに伴い、文字情報だけでなく、音声や映像情報も多く流通するようになった。そのため音楽の類似度に関する研究は、音楽情報検索システムとの関わりから多くの研究が行われている。

音楽における類似度の研究は、音響情報と記号情報のどちらを扱うか、旋律と楽曲全体のどちらを対象にするかの二点から分けることができる。旋律を対象とした類似度は、“鼻歌検索”と呼ばれる検索システムによく用いられ、たとえば文献 [3] などの音響情報を用いたものや、文献 [1] などの記号情報によるものがある。楽曲全体の音響情報を用いた例としては、部分から求まる特徴量の全楽曲中での変化を見ることで二つの楽曲が似ているかどう

かを判断する研究 [8] などがある。本研究では、楽譜から得られる楽曲の記号情報を用いて、楽曲全体に関する類似度を考える。類似する研究として、電子ファイルとして表現されたあらゆるものを対象として類似度を測る手法 [4] がある。これは情報圧縮の理論より考案されたもので、MIDI 形式のファイルに適用することで楽曲間の類似度も測ることができる。しかし、この手法は作曲家分類に関して良い結果を示しているが、結果に対して音楽的視点から考察を与えることが難しい。本研究は、音楽の知識をもとに類似度を考えることで、音楽情報検索への応用以外にも楽曲間の関係を分析する道具としても用いることができる。

また本研究では、提案手法と n -グラムモデルの関連についても述べる。楽曲に対する n -グラムモデルの適用は、たとえば文献 [5] のような、索引付

けによる検索の研究が多い。本研究では、音楽に対する n -グラムモデルの、これまでとは異なる応用についても論じる。

2 リカレンスプロット

リカレンスプロットは、本来複雑な時系列データを視覚的に解析するために提案された手法である [6]。本研究では、リカレンスプロットから派生したクロスリカレンスプロットという手法を扱う。ここで時系列 $\mathbf{u}_i (i = 1, 2, \dots, M)$ とは、時間の経過に従って観測されたデータ列で、各時刻 i でベクトル量 \mathbf{u}_i が観測されたことを表す。クロスリカレンスプロットは、二つの時系列 $\mathbf{u}_i, \mathbf{v}_j (i = 1, 2, \dots, M, j = 1, 2, \dots, N)$ の関係を調べる手法で、二つの点の間に距離を定義し、時刻 i, j におけるそれぞれの点の値がある閾値より近いときに、平面上に点 (i, j) をプロットする。具体的には次のように定義される：

$$\mathbf{R}(i, j) = \begin{cases} 1 & (d(\mathbf{u}_i, \mathbf{v}_j) < \epsilon), \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

ただし ϵ は閾値、関数 $d(\cdot, \cdot)$ はデータ間の距離で、 $\mathbf{R}(i, j) = 1$ なる (i, j) をプロットする。

RR (recurrence rate)

ある二つの時系列データに対するクロスリカレンスプロットには、二つのデータの注目しているところが似ているとき、似ていないときに比べてより多くの点がプロットされる。そのため、プロットされる点の数に対する実際にプロットされた点の数の割合

$$RR = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \mathbf{R}(i, j) \quad (2)$$

は、二つの時系列データの類似度を測る指標とされている。

3 リカレンスプロットによる楽曲間類似度

本研究では、 i 番目から $i + n - 1$ 番目までの n 個の音符の列を、時刻 i におけるデータとみなす。そして、 n 個の音符列間に距離を定めることで、リカレンスプロットを楽譜に適用する。

本節で述べる一つ目の手法では、音高とリズムの情報をもとにした編集距離を音符列間に定義し、楽

曲間にクロスリカレンスプロットを適用する。編集距離は、音符の挿入、削除、置換のそれぞれに対してコストを定め、この三つの動作を用いて一方の音符列を他方に変形するのにかかる最小のコストとして定義される (図 1)。ここでは音符の挿入、削除にかかるコストは 1 とし、置換については、音の開始時刻の差 Δt と音高の差 Δp にまた、それぞれパラメータ c_t, c_p をかけて足し合わせたものとした。以下の適用結果では、 $c_t = 1/2(n-1), c_p = 0.5$ としている。最小となるコストは、動的計画法により求めることができる。

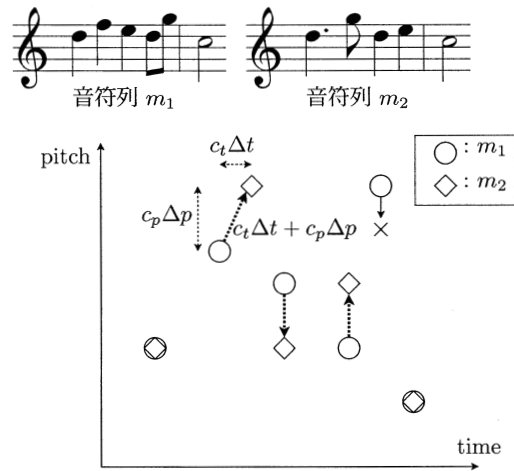


図 1 編集距離を測る際の音符列の変形。音符列 m_1 の 2, 3, 4 番目の音符を置換し、5 番目を削除して、音符列 m_2 に変形している。

ここでは、E. Köehler: Easy Exercise 1, 5 および 6 の 3 曲 (以下 S_1, S_2 および S_3) について、二つずつ取り出した 3 組の楽曲に対してクロスリカレンスプロットを適用した。近さを決める閾値 (式 (1) 参照) は、 $\epsilon = 4$ としている。得られたクロスリカレンスプロットのうち $n = 4$ の場合を図 2-4 に示す。また、音符列の長さ n に対する前節で述べた特徴量 RR の値を図 5 に示す。これらの楽曲は、 S_1, S_2 は似ているが、 S_3 は他の 2 曲とは雰囲気や異なると思われる。図 5 を見ると S_1, S_2 の組に対する RR の値が他の組より大きくなっており、この手法により楽曲の雰囲気の類似度を測れる可能性があると言える。

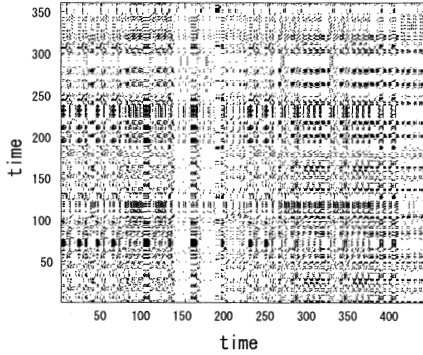


図2 楽曲に対するクロスリカレンスプロット (横軸: S_1 , 縦軸: S_2).

4 n -グラムモデルによる作曲者判別とリカレンスプロット

本節では、二人の作曲者のいずれのものであるかの分かっている楽譜がいくつか与えられているときに、 n -グラムモデルを用いて未知の楽曲の作曲者が二人のうちのどちらであるかを判別する手法について述べる。また、クロスリカレンスプロットを用いることで、この判別問題の計算量を減らせることを示す。

4.1 n -グラムモデルによる判別問題の定式化とカーネルトリック

n -グラムモデルは、もともと自然言語処理の分野で広く用いられてきた手法である。 n -グラムとは、文章に含まれる長さ n の部分文字列のことで、各 n -グラムの頻度を数え、その分布から文章の特徴を抽出するのが n -グラムモデルである。本研究では、音符 1 個を言語の場合の 1 文字と見て、音楽における n -グラムを考える。ただし、ここで扱う音高は $\{C, C\#, \dots, B\}$ の 12 種類とし、異なる音域にある同じ音名のものは同一視する。そして、 n -グラムの頻度分布をその楽曲の特徴ベクトルと考え (図 6), 以下のような判別問題を考える。

ベクトル \mathbf{x} に対する線形判別関数

$$f(\mathbf{x}) = \beta^T \mathbf{x}, \quad (3)$$

とは、値の正負によって \mathbf{x} が二つのクラスのどちらに属するかを判別するものである。今、関数を

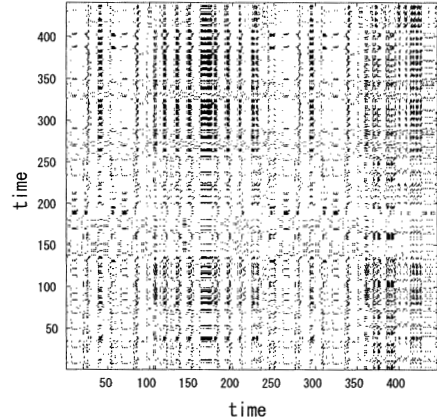


図3 楽曲に対するクロスリカレンスプロット (横軸: S_2 , 縦軸: S_3).

特徴づける定ベクトル β を、訓練用の楽譜から得られる特徴ベクトルから最小二乗法で求めることを考える。訓練用楽譜を s_1, s_2, \dots, s_N , N_s を訓練用楽譜の数とし、各訓練用楽譜の特徴ベクトルを $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_s}$, とする。つまり、長さ n の音高パターンの集合を $P = \{p_1, p_2, \dots, p_q\}$ ($q = 12^n$) とし、楽曲 s_i 中にパターン p が出現する回数を $n_i(p)$ と書くと、

$$\mathbf{x}_i = \begin{pmatrix} n_i(p_1) \\ n_i(p_2) \\ \vdots \\ n_i(p_q) \end{pmatrix} / \sum_{p \in P} n_i(p) \quad (4)$$

となる。特徴ベクトル $\mathbf{x}_1, \dots, \mathbf{x}_{N_s}$ を横に並べた行列を X とおく。また、第 i 成分が \mathbf{x}_i の属するクラスを表した教師信号ベクトルを $\mathbf{y} \in \{-1, 1\}^{N_s}$ とする。すなわち、訓練用楽譜がそれぞれ二人の作曲者 A, B のいずれかによるものであるとして、楽譜 s_i が作曲者 A の作ったものである場合は \mathbf{y} の第 i 成分を -1 , 作曲者 B の作ったものである場合は 1 のようにする。

このもとの

$$\min_{\beta} \|\mathbf{y} - X^T \beta\|^2 + \lambda \|\beta\|^2, \quad (5)$$

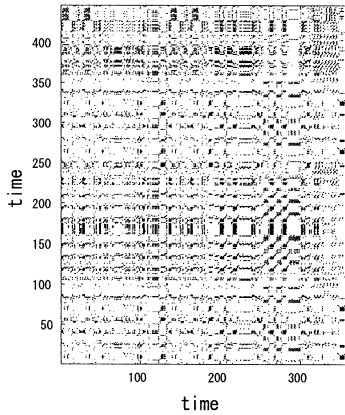


図4 楽曲に対するクロスリカレンスプロット (横軸: S_3 , 縦軸: S_1).

を解くことで, β が求まる. 式 (5) の第 1 項は, 特徴ベクトルを判別関数 (式 (3)) に代入した値と教師信号との差の二乗和である. 第 2 項は, 判別関数が訓練用楽譜に対する当てはまりを強め過ぎないようにし, 未知の楽譜に対する判別結果を改善するためのものである. 変数 λ は当てはまり具合を抑制する強さを表すパラメータである. この問題の解は解析的に,

$$\beta = (XX^T + \lambda I)^{-1} X y, \quad (6)$$

と求めることができる.

式 (5) の最小化問題は, 変数変換

$$\beta = X \alpha, \quad (7)$$

を行っても等価であることが知られている. 式 (7) を式 (3), (5) に代入すると, 判別関数と, 判別関数を求めるための最小化問題はそれぞれ

$$f(x) = \alpha^T X^T x, \quad (8)$$

$$\min_{\alpha} \|y - K \alpha\|^2 + \lambda \alpha^T K \alpha, \quad (9)$$

となる. ただし行列 $K := X^T X$ は, (i, j) 成分が特徴ベクトル x_i, x_j の内積を表す. また, 式 (8) の判別関数に現れるベクトル $X^T x$ は, 判別したい楽譜の特徴ベクトル x と楽譜 s_i の特徴ベクトル x_i と

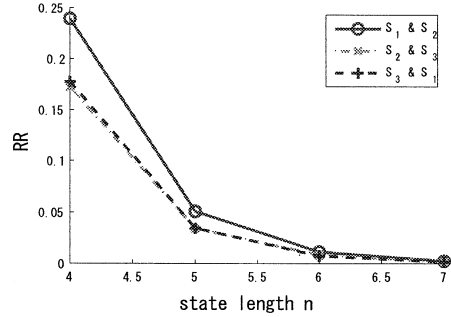


図5. 音符列の長さ n に対する特徴量 RR .

の内積を第 i 成分に持つ. 式 (9) の最小化問題を解くと,

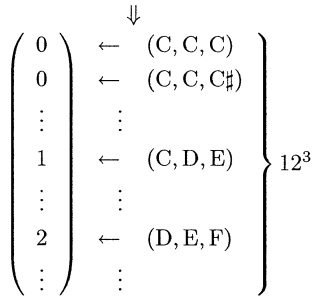
$$\alpha = (K + \lambda I)^{-1} y, \quad (10)$$

となる.

以上より, 訓練用楽譜 s_1, s_2, \dots, s_{N_s} をもとに, 特徴ベクトルが x である未知の楽譜の作曲者を判別するには, 式 (5) によりベクトル β を求め, 式 (3) を計算して値の符号を見れば良い. しかし, すべての特徴ベクトル同士の内積が分かっている場合は, 式 (9) を解いてベクトル α を求め, 判別関数に式 (8) を用いても良い. すなわち, 式 (7) 変数変換を行うことで, 各特徴ベクトルが陽に分かっていたとしても, お互いの内積が別に計算できればこの判別問題を考えることができる. 今考えている特徴ベクトルは大きさが 12^n で, 行列 X を含む演算には大きな空間計算量が必要となる. しかし, 式 (7) の変数変換により, 計算すべき問題の規模を小さくすることができる. このように, 高次元の特徴空間を介さずに内積の値を用いることで問題の規模を縮小化することを, 一般にカーネルトリックと呼ぶ. また, 二つの特徴ベクトルの内積を与える関数をカーネル関数という. 次節では, 楽曲間のクロスリカレンスプロットによりカーネル関数が計算できることを示す.

4.2 クロスリカレンスプロットによるカーネルトリック

ここで考える楽譜間のクロスリカレンスプロットでは, 二つの音符列間の距離を, 音高の列が完全に一致したら 0, 一カ所でも音高が異なれば ∞ と



特徴ベクトル
(ただし、和が 1 になるよう正規化する)

図 6. 特徴ベクトルの作り方 ($n = 3$ の場合).

する。この場合、式 (1) の閾値 ϵ は、 $\epsilon > 0$ であればどの値でも結果は変わらないので特に定めなくて良い。本節では、このようなクロスリカレンスプロットにより、楽譜の特徴ベクトル間の内積が求まることを示す。

まず、二つの楽曲 s_i, s_j に対する特徴ベクトル $\mathbf{x}_i, \mathbf{x}_j$ の内積は、

$$\mathbf{x}_i^T \mathbf{x}_j = \sum_{p \in P} n_i(p)n_j(p) / \left(\sum_{p \in P} n_i(p) \sum_{p \in P} n_j(p) \right), \quad (11)$$

と表わされる。

また、この二つの楽曲に対するクロスリカレンスプロットにおいて、パターン p に一致したことによる点だけを取り出すと、二辺が $n_i(p), n_j(p)$ 個の点からなる長方形の格子となる。そして、その長方形をなす点の数は $n_i(p)n_j(p)$ である。図 7 はこの長方形を表して、各軸上の矢印はパターン p の n -グラムが現れた時刻を表す。たとえばパターン p が楽譜 s_i で 4 回、楽譜 s_j で 3 回出現したとすると、これらの楽譜に対するクロスリカレンスプロットには、 $4 \times 3 = 12$ 個の点がプロットされることが分かる。すべてのパターンに対して長方形格子を考えれば、このクロスリカレンスプロットの RR

は

$$RR = \sum_{p \in P} n_i(p)n_j(p) / \left(\sum_{p \in P} n_i(p) \sum_{p \in P} n_j(p) \right) \quad (12)$$

と表わされ、この値は特徴ベクトルの内積 (式 (11)) に等しい。すなわち、それぞれの楽譜に対する特徴ベクトルを求めることなく互いの内積を計算できるため、前節での考察より、クロスリカレンスプロットによって作曲者判別問題を考えることができる。

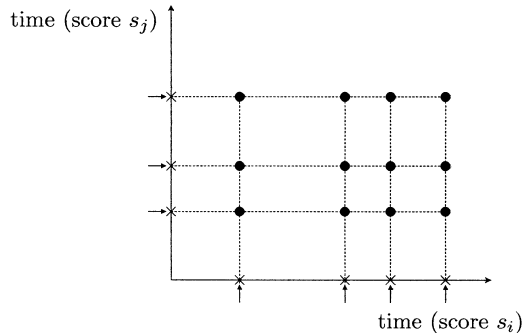


図 7 楽譜 s_i, s_j に対するクロスリカレンスプロットの一部。

4.3 作曲者判別の結果

この判別問題を、次の 3 人の作曲家による楽曲に対して考える：

- G. Gariboldi: “*Etudes Mignonnes Op. 131*” (20 曲),
- E. Köehler: Easy Exercise 1–15 and Medium Difficult Exercise 1–12 from “*35 Exercises for Flute Op. 33*” (27 曲),
- J. Andersen: “*Eighteen Studies Op.41*” (18 曲).

これらはすべてフルートの練習曲で、以下ではそれぞれの作曲者を C_1, C_2, C_3 と表記する。本研究で提案した手法は、楽曲が単音で構成されていることを仮定しているため、手法の適用は主に管楽器の楽曲に限定される。また、すべての楽曲は主音を C に移調したものを考える。

作曲者の判別は 2 人の作曲者の組ごとに次の手順で行った。まず、すべての楽曲の組についてクロ

スリカレンスプロットを求め、 RR を算出する。次に、2人の作曲家の楽曲の中で1曲を実験用、それ以外を訓練用とし、式(10)によって訓練用楽譜から判別関数を求める。そして、式(8)により、実験用楽譜が正しく判別できるかを見る。上記の判別を、すべての楽曲を一度ずつ実験用として楽曲の数だけ行い、正答率を算出する。このような正答率の算定方法をクロスバリデーションという。ただし、手法にはパラメータ λ が含まれており、 λ の適当な値を決めるために、訓練用楽譜の中でもクロスバリデーションを行っている。

部分列の長さ n と、上記により求めた正答率との関係を図8に示す。判別自体はリカレンスプロットを用いなくてもできるが、リカレンスプロットを用いることで計算量が減り、より大きな n に対する問題も容易に扱うことができる。この結果より、作曲家 C_1 は n -グラムモデルにより特徴が表せると言える。しかし、作曲家 C_2, C_3 は、この手法では区別することができない。

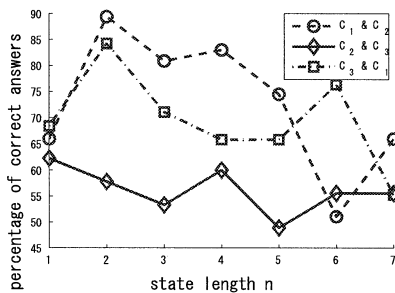


図8. 3人の作曲家判別の結果。

5 まとめと今後の課題

本研究では、時系列解析手法の一つであるリカレンスプロットを用いて、楽曲全体の類似度を測る手法を提案した。また、 n -グラムの頻度分布をもとに作曲家判別を行い、判別問題を解く際にリカレンスプロットが有用であることを示した。

今回提案した手法は単音の楽曲にしか適用できず、手法の有用性の検討や音楽情報検索などの応用が現時点では難しい。和音の n -グラムの頻度分布は非常に疎になるため、本手法をそのまま適用すると誤差が大きくなると考えられるからである。そこで、音高の組合せに対する同値関係 [2] を適切に

定めることで、特徴空間を縮約することや、グラフの判別において疎な特徴ベクトルを効率的に扱う手法の応用 [7] など検討している。

また、1-グラムに対応するリカレンスプロットにおいて、たとえば対角方向と平行な長さ3の点列は、3-グラムの一致があったことを意味する。つまり、1-グラムのリカレンスプロットには $\forall n \geq 2$ に対する n -グラムの情報が含まれる。すべての n に対する情報を包括的に扱う手法を提案することも、今後の課題の一つである。

参考文献

- [1] Will Archer Arentz, Magnus Lie Hetland, and Bjørn Olstad. Retrieving musical information based on rhythm and pitch correlations. *Journal of New Music Research*, Vol. 34, No. 2, pp. 151–159, 2005.
- [2] Clifton Callender, Ian Quinn, and Dmitri Tymoczko. Generalized voice-leading spaces. *Science*, Vol. 320, pp. 346–348, 2008.
- [3] Lujia Chen and Bao-Gang Hu. An implementation of web based query by humming system. In *2007 International Conference on Multimedia & Expo*, pp. 1467–1470, 2007.
- [4] Rudi Cilibrasi, Paul Vitányi, and Ronald de Wolf. Algorithmic clustering of music based on string compression. *Computer Music Journal*, Vol. 28, No. 4, pp. 19–67, 2004.
- [5] Shyamala Doraisamy and Stefan Rüger. Robust polyphonic music retrieval with n -grams. *Journal of Intelligent Information Systems*, Vol. 21, No. 1, pp. 53–70, 2003.
- [6] Jean-Pierre Eckmann, Sylvie Oliffson Kamphorst, and David Ruelle. Recurrence plots of dynamical systems. *Europhysics Letters*, Vol. 4, No. 9, pp. 973–977, 1987.
- [7] Koji Tsuda. Entire regularization paths for graph data. In *the 24th International Conference on Machine Learning*, pp. 919–926, 2007.
- [8] 大野和久, 鈴木優, 川越恭二. 楽曲全体における特徴量の傾向に基づいた類似検索手法. In *DEWS 2008*.