

ピアノ運指データを用いた統計学習手法による運指推定と 演奏難易度の定式化

中村 栄太^{1,a)} 齋藤 康之² 吉井 和佳¹

概要：ピアノ運指の自動推定は、音楽演奏過程を情報学的に理解するために重要であり、演奏支援や演奏学習支援技術へ応用可能である。運指の良さを定義する自然な方法は演奏の制約やコストのモデルを構成することであるが、一般的にこれらのモデルでは適切なパラメータの値を見つけるのは難しい。本稿では、統計モデルに基づくデータ駆動型のアプローチを考え、与えられた運指の自然さを確率に基づいて記述する方法について調べる。具体的には、2種類のHMM（隠れマルコフモデル）とその高次の拡張を構成する。比較手法として、DNN（深層ニューラルネットワーク）に基づく方法も調べる。新しく公開したピアノ運指のデータセットを用いて、これらの手法の学習と評価を行い、制約に基づく代表的な手法との比較評価も行う。評価に関しては、運指の個人的差異を考慮して、複数の正解運指データがある場合に使える新たな評価指標を考案した。評価の結果、高次HMMに基づく手法がその他の手法よりも推定精度が高いことが明らかになった。運指モデルに基づく演奏難易度の定式化およびピアノ用編曲への応用についても議論する。

1. はじめに

音楽演奏は高度な技能を要する動作であり、それを実現するには関連する動作の計画と調整が欠かせない [1, 2]。鍵盤楽器と弦楽器では、適切な運指の習得が基礎的な演奏技能と見なされている [3–8]。計算論的な観点では、これらの楽器のための適切な運指を見つける問題は、しばしば複雑な組み合わせ最適化問題と見なせる。自動運指推定は音楽情報処理のトピックの一つであり [9–26]、演奏過程を情報学的に理解する目的で研究されているほか、演奏支援や学習システム [27–30] や編曲技術 [31, 32] に応用されている。本稿では、統計モデリングに基づくピアノ運指の推定手法について論じる。

運指の良さは人間工学的、認知的、音楽解釈的な制約に基づき決まると考えられており [10]、これまで人間工学的な制約や演奏コストに基づく運指の良さのモデルが考えられてきた [10, 12, 19, 25]。しかし、一般的にこれらの制約の重みやコスト関数のパラメータの適切な値を求めることは、理論的にも実験的にも難しい。一方、統計モデルに基づく方法は、運指の制約やコストを文脈に依存した運指の確率を通してモデルに取り込むものである [17, 24]。この

方法は、パラメータはデータからの統計学習により最適化できるという利点を持つ。もう一つの方法として、DNN（深層ニューラルネットワーク）を用いて入力データ（音符列）と出力データ（運指）の間の非線形写像を学習するものが考えられる [33]。DNNもデータ駆動方式で学習することができ、様々なパターン認識の問題に適用され成功している。今のところ、我々の知る限り、DNNを用いたピアノ運指推定手法の文献はまだない。

統計学習、深層学習、およびその他のデータ駆動型手法をピアノ運指に応用するためには、楽譜に対して運指が与えられたデータが必要である。同様のデータは運指推定手法を系統的に評価する際にも必要になる。しかし、これまでこのようなデータで公開されているものはなく、運指推定手法の大規模な評価も報告されていない。

本稿では、HMM（隠れマルコフモデル）およびDNNに基づく運指推定手法を調べる。HMM [34] は、潜在変数（ここでは運指）の系列依存性と、潜在変数と観測データ（ここでは音符列）との確率的対応を記述する代表的な統計モデルであり、ピアノ運指推定にも応用されてきた [17, 24]。本稿では、これまで調べられてきたHMMの高次への拡張と、多声（ポリフォニック）音楽における制約を取り込む改良を提案する。また、これまでのモデルのように音符単位ではなく、和音単位で潜在状態が定義される別の構造を持つHMM（和音HMMと呼ぶ）も提案する。このモデルは、コストに基づくモデル [19] の確率的定式化と見なせ

¹ 京都大学
Kyoto 606-8501, Japan

² 木更津工業高等専門学校
Chiba 292-0041, Japan

^{a)} enakamura@sap.ist.i.kyoto-u.ac.jp

る。DNN に基づく手法としては、音声認識や機械翻訳などの他の領域で一般的である、フィードフォワード型ネットワークおよび LSTM (長短期記憶) ネットワークを調べる。これらの手法および代表的な制約に基づく手法 [25] の比較評価を行う。さらに、統計的運指モデルに基づく定量的演奏難易度尺度の定式化とその評価を行う。

本研究の主な結果は次の通りである。

- 運指推定の新規手法の考案。高次 HMM は現在の最高性能であることが明らかになった。和音 HMM と DNN に基づく手法は、それぞれの長所を持ち、今後有用になる可能性がある。
- 複数の正解運指データと推定された運指を比較するための新規の評価尺度の考案
- 系統的評価による現在の最高性能、HMM の異なる設計の効果、学習データ量の影響、典型的な推定誤り、および現在の限界の調査結果
- 運指モデルに基づく演奏難易度を用いた演奏誤り予測性能の評価

2 章では、ピアノ運指推定の代表的な既存手法について解説する。3 章では、統計モデルおよび DNN に基づくピアノ運指推定手法について説明する。4 章では、運指推定性能の評価結果について論じる。5 章では、演奏難易度尺度の定式化と評価、そしてピアノ用編曲手法への応用について論じる。6 章に結論を記す。

2. ピアノ運指推定の既存手法

Parncutt ら [10] は、ピアノ運指推定の計算論的手法の最も初期のものの一つである、制約に基づく方法を提案した。複数の運指の可能性がある少数の短い単旋律楽譜を用いて動作確認が行われており、この手法が最も難易度が低いと判断した運指が、人間の演奏者が選択した運指としばしば一致することが確認された。この手法の改良として、音高の表現の変更と制約の追加を行ったものが提案されている [13]。Lin ら [27] はこの手法のオンライン拡張を行い、演奏支援へ応用した。

最近、Balliauw ら [25] は、Parncutt らの制約に基づく手法 [10] を多声音楽の取り扱いができるようにさらに拡張した。彼らは VNS (variable neighborhood search; 可変近傍探索法) を応用し、定式化した制約に基づく最適な運指を効率的に見つける方法を提案した。論文では、いくつかの例でこのモデルの効果が確認されたが、定量的・系統的な評価は報告されていない。

Hart ら [12] は単旋律の曲に対する最適な運指を見つけるためのコストに基づく方法を提案した。各指のペアに対して定義された音符ペアを弾く局所的コストの和として運指のコストを定式化した。与えられた音高系列に対して最適な運指を見つけるための動的計画法に基づくアルゴリズムが提案された。De Prisco ら [20] は遺伝的アルゴリズム

ムに基づく手法を調べた。この手法で運指コストに対応する評価関数はデータからの学習により得られた。高い精度 (89.2%) が報告されているが、この手法は二つの曲のみで評価されており、長さや推定結果などの情報は明らかにされていない。多声部の曲に適用できるコストに基づく方法が Al Kasimi ら [19] により提案されている。このモデルは、特定の和音を弾く際の手の広がりを表す垂直コストと、和音の連なりを弾く際の手の伸びを表す水平コストからなり、Hart らのモデル [12] の一般化と見なせる。このモデルの詳細は 3.4 節に記す。

統計モデルに基づく初めてのピアノ運指推定の方法は Yonebayashi ら [17] により提案された。彼らは運指動作の統計的な自然さを表す遷移確率と手の形の自然さを表す出力確率からなる HMM を提案した。Nakamura ら [24] は HMM に基づく手法を拡張し、両手の運指に適用できる方法を開発した。この HMM は多声部の楽譜に適用でき、改良された出力確率の構造を持つ (詳細は 3.2 節を参照)。

3. 運指推定のためのモデルと方法

3.2 節では、単純な HMM に基づく既存手法 [17,24] について解説し、3.3 節では高次へ拡張した提案モデルについて説明する。3.4 節と 3.5 節では、それぞれ和音 HMM と DNN に基づく手法について論じる。

3.1 問題の定式化

本研究では、運指推定として以下の具体的な問題を考える。入力は MIDI 信号と同様の系列 $(p_n, t_n, \bar{t}_n)_{n=1}^N$ により表されるピアノ演奏である。ここで、各音符 n に対して、 p_n は音高、 t_n は発音あるいは打鍵時刻、 \bar{t}_n は消音あるいは離鍵時刻を表し、 N は音符数を表す。音符は打鍵時刻の順番に並んでいると仮定する。また、右手パートと左手パートは別々に与えられるものとする (両手の音符が混ざった系列から、各パートへの分離を自動で行う方法については例えば文献 [24] を参照されたい)。出力は各音符 n に対する運指番号のリスト $(f_n)_{n=1}^N$ である。各 f_n は $\{1, \dots, 5\}$ の中の値をとる (標準的な音楽記法と同様に、1 は親指、2 は人差し指、 \dots 、5 は小指を表す)。

3.2 単純な HMM

本節では、先行研究 [17,24] で提案されているピアノ運指のための 1 次の HMM について説明する。

3.2.1 生成モデル

ピアノ運指推定のための HMM はピアノ演奏の生成過程を確率的に記述する。このモデルはピアノ演奏 $\mathbf{p} = (p_n)_{n=1}^N$ と運指 $\mathbf{f} = (f_n)_{n=1}^N$ のあらゆるペアに対する確率 $P(\mathbf{p}, \mathbf{f})$ を推定する手段を与えるものである。まず、運指番号の系列がマルコフモデルにより生成される。

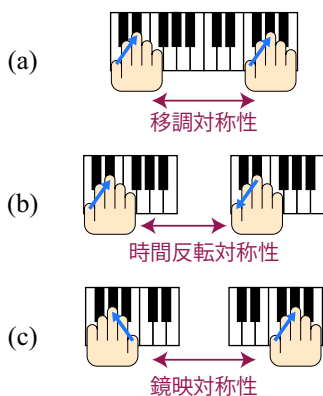


図 1 HMM の出力確率の対称性

$$P(\mathbf{f}) = P(f_1) \prod_{n=2}^N P(f_n | f_{n-1}) \quad (1)$$

ここで、 $P(f_1)$ は初期確率、 $P(f_n | f_{n-1})$ は遷移確率と呼ばれる。遷移確率は運指番号の順番の自然さを統計的に記述するものである。次に、音高列 \mathbf{p} は運指番号列 \mathbf{f} に条件づいて次のように生成される。

$$P(\mathbf{p} | \mathbf{f}) = P(p_1 | f_1) \prod_{n=2}^N P(p_n | p_{n-1}, f_{n-1}, f_n) \quad (2)$$

この式の右辺の確率 $P(p_n | p_{n-1}, f_{n-1}, f_n)$ は出力確率と呼ばれ、音高のペアと運指番号のペアの対応の自然さを統計的に記述するものである。一番目の因子 $P(p_1 | f_1)$ は音高系列を生成するために、形式的に必要である。同時確率 $P(\mathbf{p}, \mathbf{f}) = P(\mathbf{f})P(\mathbf{p} | \mathbf{f})$ は式 (1) と (2) の確率の合成により得られる。

以上の定式化では、モデルは打鍵時刻や離鍵時刻を記述しておらず、音符の時間順序には依存するが時間間隔には依存しないモデルになっている。音符の時間間隔はほとんどの既存研究で考慮されておらず、本研究でも考慮しない。しかし、和音（あるいは、ほぼ同時に演奏される音符）に関する例外を記述するため、一部の時間情報を用いるモデルについて 3.2.3 項で後述する。

3.2.2 出力確率の設計

最も一般の設定では、出力確率 $P(p_n | p_{n-1}, f_{n-1}, f_n)$ は $88^2 \cdot 5^2 (\sim 2 \times 10^5)$ 個のパラメータを持つ。限られたデータ量でこれらすべてのパラメータを正確に推定するのは困難なため、パラメータ数を効果的に減らすためにモデル構造に対する仮定をおく必要がある。出力確率におく仮定の違いによりいくつかの種類 HMM が運指推定に用いられる。

一般的な仮定は、出力確率が相対的な音程に依存し、絶対的な音高には依存しないというものである。この出力確率の「移調対称性」(図 1(a)) は数学的には次で表される。

$$P(p_n | p_{n-1}, f_{n-1}, f_n) = P(p_n + p | p_{n-1} + p, f_{n-1}, f_n) \quad (3)$$

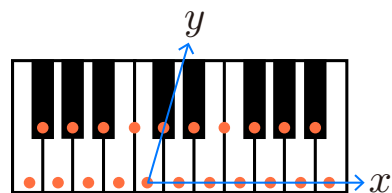


図 2 二次元格子によるピアノ鍵盤の表現

ここで p は任意の音高である。これにより確率分布は次の形を持つ。

$$P(p_n | p_{n-1}, f_{n-1}, f_n) = F(p_n - p_{n-1}; f_{n-1}, f_n) \quad (4)$$

この仮定は本稿では常に適用されるものとする。これらの式では、音高に関する足し算と引き算を定義する必要があるが、これについては後述する。

音高の表現法にはいくつかの可能性があるが、本研究では次の二つを考える。

- 整数音高表現
- 格子音高表現

一番目の表現では、音高を半音単位の整数で表す。これは数学的に単純であり、足し算と引き算の演算は率直に定義される。この表現法の問題は、運指に重要である白鍵と黒鍵からなるピアノの鍵盤の幾何的な特徴を反映できない点である [13]。この問題を解決するため、二番目の表現では、ピアノの鍵盤は二次元格子によって表され、音高はこの格子点上の点で表される (図 2)。この表現法では、足し算と引き算は二つの次元で独立に適用されるものとして定義される。例えば、 $p_n - p_{n-1}$ の y 方向の値は 0, 1, または -1 をとる。

もう一つの基本的な仮定は、大きな音高の遷移(跳躍)を等しく扱うというものである。オクターブを超えるような大きな跳躍はまれであるため、個々の場合について確率を正確に推定するのは現実的ではない。この問題に対処するため、整数音高表現では上限値 δp_{\max} を定め、 $p_n - p_{n-1} > \delta p_{\max}$ の場合は等しく扱う ($p_n - p_{n-1} < -\delta p_{\max}$ の場合も同様)。格子音高表現では、 x 方向に対して同様の上限値を適用することができる。

先行研究ではさらに二つの仮定が提案されている [24]。一つ目は「時間反転対称性」(図 1(b)) で、次のように表される。

$$F(p_n - p_{n-1}; f_{n-1}, f_n) = F(p_{n-1} - p_n; f_n, f_{n-1}) \quad (5)$$

この式は、順方向の運指の動作と逆方向の動作がコスト的には同じであることを意味している。もう一つの仮定は、両手のモデルのパラメータ間の関係に関するもので、それが x 方向の反転(鏡映)により関係しているというものである。この「鏡映対称性」(図 1(c)) は次のように表される。

$$F_L(p_n - p_{n-1}; f_{n-1}, f_n) = F_R(-[p_n - p_{n-1}]_x; f_{n-1}, f_n) \quad (6)$$

ここで F_L と F_R は左手と右手の出力確率であり、 $-[\cdot]_x$ は x 方向に関するマイナス演算を表す。整数音高表現では標準的なマイナス演算であり、格子音高表現では x 方向の要素の符号を逆転するものである。

3.2.3 和音の記述

ここまで、入力のある音符列 $(p_n, t_n, \bar{t}_n)_{n=1}^N$ は打鍵時刻の順に並んでいると仮定したが、時間情報はモデルには組み込まれていない。この結果、上述のモデルは、同時あるいはほぼ同時の打鍵時刻を持つ複数音符の集まりとして表される和音を含む一般の多声部の曲にもそのまま適用できる。しかし、時間情報を用いなければ、和音の運指に特化した制約（例えば和音内での指くぐりがまれであることなど）を設けることができない。

和音に特化した運指の制約を考慮したモデルの改良を行うため、和音内の音符遷移と和音をまたがる音符遷移の場合に分けて二つのセットの遷移確率と出力確率を用いることが考えられる。しかし、使える学習データ量が限られていることから、この方法による改良の効果は限定的であることが予備的実験で分かった。よって、ここではもっとも重要と考えられる制約である、和音内での指くぐりの禁止を組み入れた最小のモデルを考える。これは和音内での音符遷移に関して音高遷移の方向と指の遷移の方向が逆である時に出力確率を 0 とすることで実現できる。二つの音符が一つの和音に属するか否かはそれらの打鍵時刻の差を見ることで検出できる。具体的には、 $\Delta = 30$ ms として $|t_n - t_{n-1}| \leq \Delta$ の条件のときに $(n-1)$ 番目と n 番目の音符は高確率で同じ和音に属することが知られている [35]。改良モデルでは、この条件が満たされるときに上記の制約を適用するものとする。

3.2.4 運指推定アルゴリズム

統計的な枠組みでは、運指推定は確率的最適化問題として定式化される。与えられた音符列 \mathbf{p} に対して、確率 $P(\mathbf{f}|\mathbf{p}) \propto P(\mathbf{f}, \mathbf{p})$ を最大化する最も確からしい運指 $\hat{\mathbf{f}}$ として最適な運指を推定する。本研究で扱う HMM では、標準的なビタビアルゴリズム [34] を用いることでこの最適な運指を効率的に見つけられる。

3.3 高次の HMM

3.2 節で述べた単純な HMM には、接続する音符間に関する最も局所的な運指の制約だけしか記述できないという限界がある。この限界は一般に次のような遷移確率と出力確率により記述される高次の HMM により克服できる。

$$P(f_n | f_{n-m}, \dots, f_{n-1}) \quad (7)$$

$$P(p_n | p_{n-m}, \dots, p_{n-1}, f_{n-m}, \dots, f_n) \quad (8)$$

ここで、 m は HMM の次数を表す。

m 次の HMM の一般形の出力確率のパラメータ数は $88^{m+1} \cdot 5^{m+1}$ であり、 m の増加とともに急激に増加する。少数のパラメータにより効率的に出力確率を近似する方法として、次のように対因子の集合により分解する方法がある。

$$P(p_n | p_{n-m}, \dots, p_{n-1}, f_{n-m}, \dots, f_n) \propto \prod_{l=1}^m P(p_n | p_{n-l}, f_{n-l}, f_n)^{\alpha_l} \quad (9)$$

他分野において、このような対因子への分解が時系列モデルに対して有効であることが示されている [36, 37]。上式では、 l 番目の因子に対する重みを表す係数 α_l を導入した。右辺の各因子は移調対称性を持つと仮定し、式 (4) のように表されるものとする。3.2.2 項で述べた出力確率に対するさらなる仮定も同様に高次 HMM に対して適用できる。

遷移確率のパラメータ数も同様に m に関して指数関数的に増加するが、その程度は出力確率のパラメータ数に比べれば小さい。データがスパースになる可能性を避けるため、線形補間に基づく遷移確率のスムージング法を導入する。

$$P(f_n | f_{n-m}, \dots, f_{n-1}) = (1 - \lambda_1 - \dots - \lambda_{m-1}) P_{\text{ML}}(f_n | f_{n-m}, \dots, f_{n-1}) + \sum_{l=1}^{m-1} \lambda_l P_{\text{ML}}(f_n | f_{n-l}, \dots, f_{n-1}) \quad (10)$$

ここで、 $\lambda_l \in [0, 1]$ は補間係数を表し、 $P_{\text{ML}}(f_n | f_{n-l}, \dots, f_{n-1})$ はスムージング前の l の遷移確率を表す。

運指情報付きのピアノ演奏の学習データが与えられた時、遷移確率と式 (9) の出力確率の各対因子 $P(p_n | p_{n-l}, f_{n-l}, f_n)$ は容易に推定できる。一方、係数 α_l と λ_l は、運指推定精度に関しての最適化により決定できる（詳細は 4 章を参照）。運指推定アルゴリズムはビタビアルゴリズムを単純に拡張することで導出できる。

3.4 和音 HMM に基づく手法

ここでは、文献 [19] で提案されたモデルの確率的定式化である、「和音 HMM」と呼ぶもう一つの HMM を構築する。3.2 節と 3.3 節で定式化したモデルのように音符ごとに状態遷移を考える代わりに、このモデルでは状態を和音ごとに定義する。ここで、和音とはほぼ同時に打鍵される音符の集合として定義し、具体的には 3.2.3 項で述べたように $\Delta = 30$ ms を閾値として用いて音符列を和音列に分割する。以下に詳しく説明するように、このモデル構造のおかげで、和音単位での系列依存性を取り込めること（それにより音符単位では長距離の依存性を扱えること）および

音長の影響を取り込めることが和音 HMM の利点である。

このモデルの状態は各和音に対する全ての運指の組み合わせにより構成する。和音内では指くぐりはないこと、および各々の音高が別の指で弾かれることを仮定すると、 K 個の音高からなる和音に対する可能な運指の数は $\binom{5}{K}$ である。このモデルの特徴は、離鍵時刻も考慮して、音符が弾かれている間は後続の和音の構成音に含まれるようになっていることである。替え指の可能性を無視すると、音符が弾き続けられる時は同じ指が使われるように状態遷移が制約される。

元のモデル [19] では垂直コストと水平コストが考えられている。ここでは、和音とその先行和音は Q (構成音は K 個とする) と Q' (構成音は K' 個とする)、それらの構成音高は Q_k ($k \in \{1, \dots, K\}$) と Q'_k ($k \in \{1, \dots, K'\}$)、対応する運指番号は G_k と G'_k で表すことにする。垂直コスト V はある和音を弾く際の指の広がりを表し、次のように表される。

$$V(Q, G) = \sum_{\substack{k, k'=1 \\ k \neq k'}}^K v(Q_{k'}, Q_k; G_{k'}, G_k) \quad (11)$$

一方、水平コスト H は和音の連なりを弾く際の手の伸びを表し、次のように表される。

$$H(Q', G'; Q, G) = \sum_{k'=1}^{K'} \sum_{k=1}^K h(Q'_{k'}, Q_k; G'_{k'}, G_k) \quad (12)$$

ここで、 $v(p', p; f', f)$ と $h(p', p; f', f)$ は音高対 (p', p) と運指番号対 (f', f) を弾く際のコストであり、前者は和音内の垂直音程に対して適用され、後者は和音をまたぐ水平音程に対して適用される。状態遷移のコストは垂直コストと水平コストの和として定義される。

確率的定式化においては、垂直コストと水平コストは音符対に対する遷移確率と出力確率の積として表される。これは高次 HMM の出力確率 (式 (9)) の場合と同様の定式化である。遷移確率と出力確率は次のように記述される。

$$P(G|G') \propto \left[\prod_{k'=1}^{K'} \prod_{k=1}^K P(G_k|G'_{k'})^{\beta_1} \right] \left[\prod_{\substack{k, k'=1 \\ k \neq k'}}^K P(G_k|G_{k'})^{\beta_2} \right] \quad (13)$$

$$P(Q|Q', G', G) \propto \left[\prod_{k'=1}^{K'} \prod_{k=1}^K P(Q_k|Q'_{k'}, G'_{k'}, G_k)^{\gamma_1} \right] \cdot \left[\prod_{\substack{k, k'=1 \\ k \neq k'}}^K P(Q_k|Q_{k'}, G_{k'}, G_k)^{\gamma_2} \right] \quad (14)$$

式 (14) の確率の要素に対して、パラメータ数を減らすために移調対称性を課し、格子音高表現を用いることにする。

表 1 PIG データセットの内容

サブセット	作曲家	曲 (小節; 音符)	異なる運指の曲 (小節; 音符)
Bach	1	10 (218; 3,657)	40 (872; 14,628)
Mozart	1	10 (185; 2,546)	60 (1,110; 15,276)
Chopin	1	10 (244; 4,022)	50 (1,220; 20,110)
Miscellaneous	24	120 (2,533; 38,501)	159 (3,355; 50,030)
全体	24	150 (3,180; 48,726)	309 (6,557; 100,044)

前述の HMM の出力確率と同様に、時間反転対称性と鏡映対称性を課すこともできる。

学習段階では確率の各要素は最尤推定法により推定でき、推論段階ではビタビアルゴリズムを適用できる。式 (13) と (14) の中の積は全ての構成音高の対に適用されるため、構成音が多い和音に対して重みを与えすぎる可能性がある。このバイアスの可能性を低減するため、ビタビ更新において K 個の音高からなる和音の寄与を重み付けする因子 $K^{-\zeta}$ ($\zeta \geq 0$) を導入する。係数 $\beta_1, \beta_2, \lambda_1, \lambda_2$, および ζ は運指推定精度を目的関数とする最適化により決定できる。3.3 節で述べた前述の高次 HMM と同様に、高次の和音 HMM を構成できる。

3.5 DNN に基づく手法

近年の深層学習の広範囲の適用を考慮し、比較評価の参照として二つの単純な DNN に基づくピアノ運指推定手法を考える。一つはフィードフォワード (FF) 型ネットワークであり、もう一つは LSTM (長短期記憶) ネットワークである。どちらのネットワークも、入力として整数音高列を入力とし、対応する運指番号列を出力とする。また、左手と右手それぞれで学習と評価を行った。予備実験により決定したネットワークの構造は、それぞれ三層の隠れ層からなる FF 型ネットワークと 32 ユニットの LSTM 三層と一つの softmax 層からなるネットワークである。窓幅は対象とする記号の周り ± 5 個とした。

4. 運指推定精度の評価

4.1 データ

運指推定手法の学習および評価には、新しく収集・公開した運指情報付きピアノ楽曲データである PIG データセット (PIano fingerinG Dataset) を用いた。PIG データセットは西洋クラシック作曲家のピアノ曲から成り、ピアノ演奏者により付けられた運指が含まれている。現在、150 曲の楽曲が収められており、各楽曲には一つあるいは複数の運指が与えられている。なるべく多様なスタイルの音楽を含めるため、少量の楽曲の全曲に対して運指を付けるのではなく、多くの曲の一部に対して運指のラベル付けをしてある。運指付けされた部分の典型的な長さは、楽譜 1 ページ程度で、小節数は約 20、音符数は約 300 である。

データ収集に当たっては、有名な作曲家の曲で演奏される頻度が高いものを含め、音楽スタイルや作曲年代に関してなるべく広い範囲をカバーするように楽曲を選択した。一部の楽曲には、個人性を調べるために最低4人の異なるピアノ演奏者による運指が与えられている。これらの楽曲には J. S. Bach と W. A. Mozart と F. Chopin の各10曲が含まれており、それぞれ「Bach セット」「Mozart セット」「Chopin セット」と呼ぶことにする。これらに含まれないその他の楽曲は、作曲家24人^{*1}による120曲であり、「Miscellaneous セット」と呼ぶことにする。Miscellaneous セットにも、Bach と Mozart, Chopin の楽曲は含まれるが、それらは上記の作曲家別サブセットに含まれるものとは異なる曲である。各セットのデータ量は表1にまとめられている。データおよび全ての楽曲のリストは Web ページ^{*2}からダウンロード可能である。

PIG データセットの運指は音楽大学ピアノ学科の卒業生あるいは20年以上のピアノ演奏経験を持つピアノ熟練者により与えられたものである。演奏者は自身が演奏可能な楽曲を選択して、その楽曲を実際に弾くときに用いる運指を提供するように指示を与えられている。PIG データセットに関するより詳しい記述およびこのデータを用いた運指の統計解析の結果については文献 [26] を参照されたい。

4.2 設定

3章で述べたピアノ運指推定手法の性能を系統的に比較するため、4.1節で説明したデータを用いて学習と評価を行った。学習には Miscellaneous セットを用いて、評価には三つの作曲家別セットを用いた。評価の際は、楽譜の上段と下段をそれぞれ右手パートと左手パートとして扱った。

3.2節と3.3節で説明したHMMを3次のモデルまで、3.4節の1次のHMM、および3.5節の二つのDNN (FF と LSTM) を実装した。HMMの重み係数 α_l と和音HMMの係数 $\beta_1, \beta_2, \gamma_1, \gamma_2$, および ζ はベイズ最適化法 [38] を用いて最適化した。パラメータ α_l は遷移確率に対する出力確率の重要度を制御する効果を持つため、1次のHMMに対しても対応する係数 α_1 を導入し、同様に最適化する。 δp_{\max} は15とした。DNNの学習には、Adam最適化法 [39] を初期設定で用い、バッチサイズは3にした。

2章で解説した先行研究で提案された手法のうち、VNS法 [25] のみがソースコードを公開している。この方法は [10] の手法の改良であり、制約に基づく代表的な手法と見なせる。提供されているプログラムは MusicXML ファ

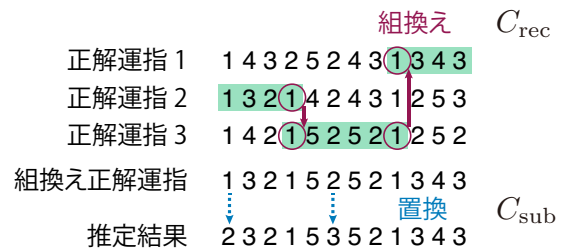


図3 組換え一致率に関する編集コスト

イルを入力としているため、評価に用いた曲に対して MusicXML ファイルを生成して、得られた出力を他の手法の出力結果と同じ形式へ変換した。この手法のパラメータ値は初期設定のものを用いて評価を行った。

4.3 評価尺度

適切なピアノ運指は唯一ではないことを考慮すれば、複数の正解データに対して使える評価尺度が必要である [26]。一つの完全な評価尺度を見つけることは難しいため、いくつかの可能な尺度をここで提案する。一つだけの正解運指が与えられる場合には、最も単純な尺度は、正しく運指が推定された音符の割合として定義される「一致率」である。複数の正解運指が与えられる場合には、それぞれの正解に対して一致率を計算して、その平均値である「一般的一致率」 M_{gen} を用いることができる。これは、推定結果が全ての正解に比べてどれだけ似ているかを表す指標である。また、推定結果に最も近い正解運指に注目して、「最高一致率」 M_{high} を定義することもできる。一方で、各音符に対する正しい推定の最も緩い判定基準は、推定された運指番号が正解の運指のどれか一つにでも一致するというものである。この基準を用いて、「柔軟一致率」 M_{soft} を定義できる。

文献 [26] で論じられている通り、運指には局所的にはいくつかの候補があり得るが、同時にそれらの間の運指遷移の一部を除外する強い系列依存性があることが分かっている。この性質を鑑みて、複数の正解運指データを組換えて作られる参照系列と推定結果を比較する評価尺度を考える (図3)。この尺度は編集距離に基づき定式化できる。「組換え正解運指」と呼ぶこの参照系列を構成するにあたり、正解運指の間の組換えは運指番号が一致する音符の箇所のみで許し、各組換えに対して編集コスト「組換えコスト」 C_{rec} を割り当てる。また、推定結果と組換え正解運指との間の運指番号の不一致に各々対して、編集コスト「置換コスト」 $C_{\text{sub}} = 1$ を割り当てる。複数の正解を推定結果に編集するのに必要な合計コスト E_{rec} は推定誤りの数として解釈できる。推定結果の誤りを最小にする組換え正解運指を考え、「組換え一致率」 $M_{\text{rec}} = (N - E_{\text{rec}})/N$ を定義する (N は音符の総数)。以下では、 $C_{\text{rec}} = 1$, 即ち一回の組換えは一回の置換と同じだけペナルティーを科す場合を考え

^{*1} 具体的には、Albeniz, Bach, Bartók, Beethoven, Brahms, Chopin, Debussy, Dvořák, Fauré, Grieg, Joplin, Liszt, Mendelssohn, Mozart, Mussorgky, Rachmaninoff, Ravel, Saint-Saëns, Satie, Scarlatti, Schubert, Schumann, Scriabin, Tchaikovsky.

^{*2} <http://beam.kisarazu.ac.jp/research/PianoFingeringDataset-j/>

表 2 ピアノ運指推定手法の比較評価. 一致率は評価データの 30 曲に関する平均値を表している. * 「人間」と記された参考値は, 評価データの各演奏者の運指に対して, それ以外の全ての演奏者の運指データ全てと比較して計算した一致率を示している.

手法	M_{gen}	M_{high}	M_{soft}	M_{rec}
1 次 HMM	61.7	68.3	82.8	74.0
2 次 HMM	64.3	70.8	85.3	77.6
3 次 HMM	64.5	71.0	85.5	77.8
和音 HMM	61.2	67.7	81.7	73.8
DNN (FF)	61.5	66.2	82.5	69.5
DNN (LSTM)	61.3	66.1	82.8	69.5
人間 *	71.4	79.1	90.8	84.3

表 3 表 2 の結果を得るのに用いられたパラメータ

モデル	パラメータ
1 次 HMM	$\alpha_1 = 0.964$
2 次 HMM	$\alpha_1 = 0.556, \alpha_2 = 0.407, \lambda_1 = 0.474$
3 次 HMM	$\alpha_1 = 0.448, \alpha_2 = 0.292, \alpha_3 = 0.194,$ $\lambda_1 = 0.470, \lambda_2 = 0.504$
和音 HMM	$\beta_1 = 0.94, \beta_2 = 4.70, \gamma_1 = 7.53,$ $\gamma_2 = 5.29, \zeta = 0.10$

る. 組換一致率の計算アルゴリズムの詳細については, 付録 A.1 で説明する.

4.4 精度比較

まず全ての手法を比較した結果について論じる (表 2). HMM と和音 HMM の出力確率に対しては, 移調対称性のみ課した場合の結果を示している. これらの結果を得るのに用いた, モデルの係数の値は表 3 に示している. HMM に対しては, 和音に対する制約 (3.2.3 項) を適用した結果である. VNS 手法に対する結果はここには示していないが, これはこの手法が評価データの一部の曲に対して運指を出力できなかったからである. この手法との比較については後述する.

この結果より, 2 次と 3 次の HMM が他の手法を有意に上回っていることが分かる. 特に, 1 次の HMM と 2 次の HMM の一致率の間の明確な差異は, 長距離依存性を取り入れることが一致率の向上に有効であることを示している. 一方, 3 次の HMM は 2 次の HMM より全ての一致率において上回ったが, その差は小さいことが分かる. これは, 少なくとも現在の学習データ量では, 高次の HMM による改善はこれらの次数で飽和している可能性を示している. これらの現在の最良の手法の一致度と人間の演奏者の間の一致度との間には依然としてギャップがある. この点については, 後により詳しく考察する.

文献 [19] のモデルの確率的定式化である和音 HMM は 1 次の HMM と比べて同等の一致率であったが, 後者の方が常に上回っている. 和音 HMM は和音単位の遷移およ

表 4 VNS セットに対する比較評価の結果. 表 2 の説明文を参照.

手法	M_{gen}	M_{high}	M_{soft}	M_{rec}
1 次 HMM	60.5	67.9	82.4	73.7
2 次 HMM	63.1	70.6	85.1	77.6
DNN (LSTM)	59.8	64.6	82.6	67.9
VNS 法 [25]	56.7	62.4	78.2	67.5
人間 *	70.2	78.9	91.0	84.5

び音長の影響も記述していることから, この結果はやや予想外であったが, その理由として二つの可能性が考えられる. 一つ目は, モデルの構造, 特に式 (13) と (14) の確率の分解が不正確である可能性である. もう一つは, 入力のパiano演奏データが人工的に生成されたものであり, 物理的に許されない音長や音符の時間的重なりを含むことが原因で, 継続音符に対する制約が望ましくない形で働いている可能性である. この結果から, 和音 HMM に対しては高次のモデルの評価を行わないことにした.

二つの DNN 手法 (FF と LSTM) はどちらも同様の一致率であり, その一般的に一致率は 1 次の HMM に比べ若干低かった. 組換一致率に関しては, 1 次の HMM よりも大きく低いことは注目すべきである. これは学習した DNN は, HMM では遷移確率に取り込まれている, 運指の系列的整合性を捉えられていないことを示している. この結果は, piano運指推定の問題が DNN に基づく手法により単純に解決できるものではないことを示している. ただし, 一般的に DNN を用いる手法の性能は学習データの量に大きく依存していること, そして 3.2.3 項と 3.2.2 項に示されている運指の対称性に基づくデータ拡張や時間特徴量の使用などの改良の可能性が残っていることに注意が必要である.

VNS 法 [25] は評価データの 30 曲のうち 14 曲のみ結果を出力することができた (これはおそらく MusicXML ファイルのパラーサーの問題と考えられる). VNS 法とその他の手法を比較するために, これらの 14 曲のデータ (「VNS セット」と呼ぶ) に対して一致率を計算した結果を表 4 に示す. ここでは, HMM と DNN に基づく手法のうち代表的なものの結果のみ示している. この結果より VNS 法の一致率は 1 次の HMM よりも有意に低いことが分かる. これは HMM に基づく統計学習手法の有効性を示している. LSTM ネットワークの結果との比較では, VNS 法の方が一致率が低いが, 組換一致率に関しては同程度であった.

4.5 HMM の構造の影響

HMM の出力確率の構造による影響を調べるための比較評価も行った. 表 5 の結果は $\alpha_1 = 1$ で移調対称性のみ課した場合のものである (他の対称性を課した場合の影響については 4.6 節で論じる). まず, 和音の制約に関しては, 常に一致率の改善が見られる. また, 整数音高表現よりも格子音高表現の方が良いことが分かるが, 一致率の改善は

表 5 異なる出力確率の構造を持つ 1 次 HMM に対する一致率

音高表現	和音の制約	M_{gen}	M_{high}	M_{soft}	M_{rec}
格子	✓	61.6	68.3	82.8	74.0
整数	✓	61.4	67.9	82.6	74.0
格子		60.0	66.7	81.2	72.7
整数		59.5	66.2	80.6	72.2

比較的小さい。これらの改良はモデルをより正確にするべく導入されたものであり、結果は期待通りであった。

4.6 学習データ量による影響

学習データが増えたことによる効果を調べるため、また HMM の出力確率に追加で対称性（時間反転と鏡映）を課すことによる影響を調べるため、学習データ量を変えた時の影響を調べる。図 4 に学習データ量を変化させた時の 1 次と 2 次の HMM の一般的一致率を示す。全ての場合で格子音高表現を用い、表 3 にあるハイパーパラメータ α_l と λ_l の値を用いた。一番右の点は学習データを全て使った場合の結果であり、表 2 の値に対応する。他の点は学習データを減らしてモデルを学習した場合の結果である。具体的には、学習データ全部より少ない曲数を選び、この曲数で 100 の異なるデータセットをランダムに選択し、100 セットの学習データそれぞれでモデルを学習し、評価データに対する一致率の平均を計算した。

1 次と 2 次の HMM の結果で同様の傾向が見られた。学習データを全て使った時は、追加で対称性を課さない基本モデルの一致率が一番高く、対称性を両方追加したもの的一致率が一番低かった。時間反転対称性を課すと一致率がわずかに低下し、鏡映対称性を課すと一致率がより大きく低下した。これは十分大きな学習データ量がある時は、対称性を課さないモデルの方が記述能力が高く、予測性能も高いことを示している。また、時間反転対称性に対する非対称度合いよりも鏡映対称性の非対称度合いの方が大きいことも示唆している。それぞれのモデルの一致率の順位は学習データの量により変化した。学習データ量が小さい極限では、基本モデルの一致率が一番低く、対称性を追加することで一致率が向上していることがわかる。これは対称性を追加することで有効なモデルパラメータの数が減り、学習データが少ない時の一致率の低下の主な原因であるオーバーフィットに対する頑健性が大きくなることによるものである。この結果は、考えた二つの対称性が良い近似になっていることも示している。

興味深いことに、一致率は次の簡単な関数でよくフィットできる。

$$A(N) = a - b/\sqrt{N} \quad (15)$$

ここで N は学習データ内の音符数であり、 a と b は定数である。付録 A.2 で示すように、この関数はモデルパラメータが漸近値に近いことを仮定すると統計理論により導出で

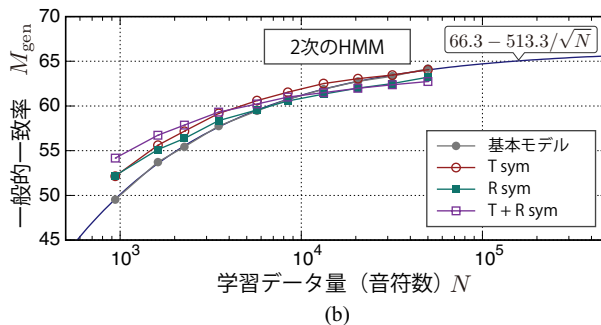
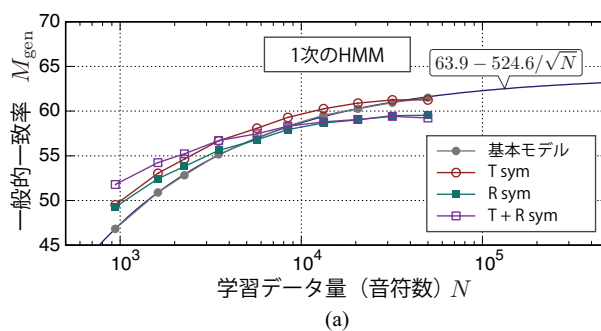


図 4 学習データ量を変化させたときの一貫率の変化。基本モデルは追加の対称性を課さないモデル、「T sym」は時間反転対称性を課したモデル、「R sym」は鏡映対称性を課したモデル、「T + R sym」は両方の対称性を課したモデルを示している。

きる。重要な帰結として、係数 a が学習データ量の無限極限での一致率の予測値を与える点である（図 4 に具体的な値がある）。この結果はこのモデルの可能性と限界の両方を示している。即ち、学習データをさらに増やすことで、一致率がさらに約 2 ポイント向上する見込みがある一方で、無限に学習データを増やしたとしてもこれらのモデルの一致率は、人間の演奏者同士での一致度（ $\sim 71\%$ ）には達しないと考えられる。ただし、一致率の値は評価データに依存すること、そしてこれらの予測値は学習データをさらに増やしてモデルパラメータの漸近値が変われば変更される可能性があることに注意されたい。

4.7 結果の例と誤り解析

HMM に基づく手法の能力と限界を示し、また高次のモデルの効果を明らかにするために、図 5 に示す結果例について議論する。図 5(a) の楽譜の右手パートでは、二つのモデルによる推定結果の最初の半分は一致しており、E5 の音の後から分岐している。どちらの運指も大きな問題はなく演奏に使えるが、2 次の HMM のものの方が正解により近い。実際、楽譜の中程の G5-E5-A5 に対する運指としては 4-2-5 の方がより自然で、2 次の HMM では考慮される G5 から A5 に移る指の動きがこの判断に重要である。同様に左手の運指では、1 次の HMM による推定結果は連続する二つの音符に着目すれば許容される指の動きを含んでいるが、全体的に見ると 2 次の HMM の結果の方がはるかに自然であり、正解の運指にもより近い。2 次の指の動きを考慮す

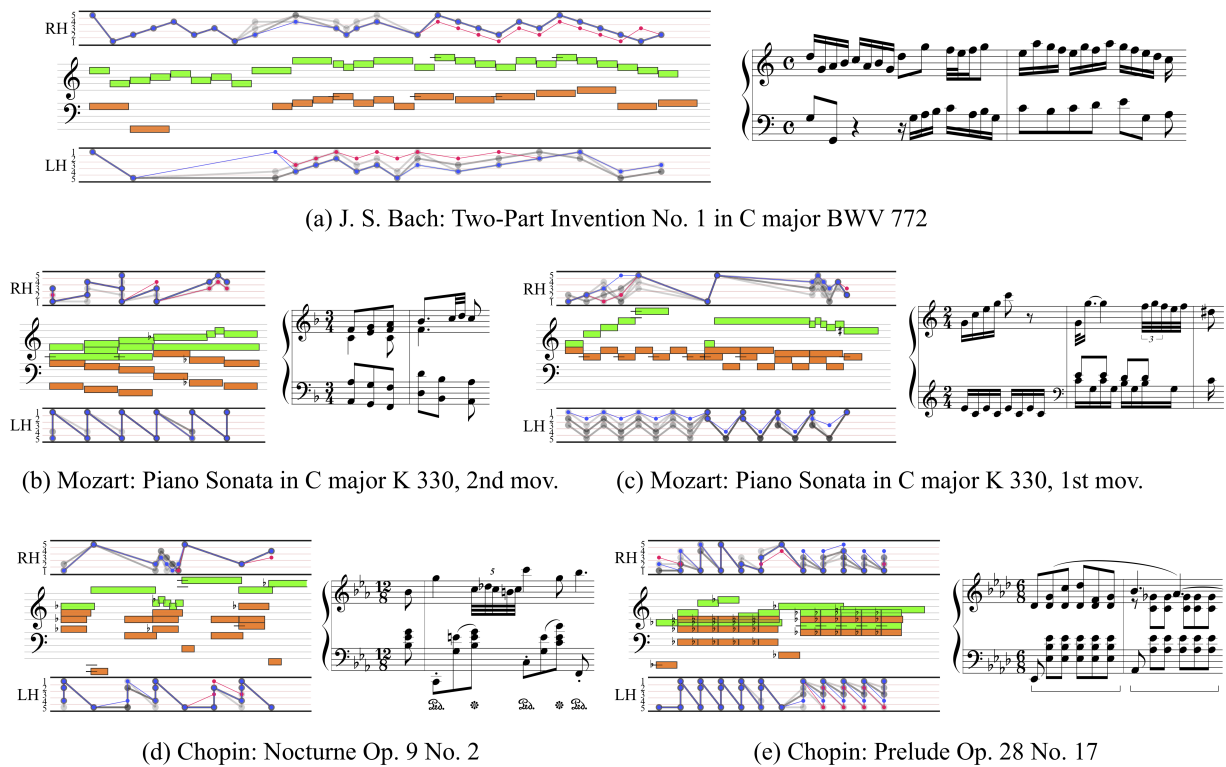


図 5 HMM によるピアノ運指推定の結果例。ピアノロール表現、楽譜、正解の運指、および推定された運指を各楽曲ごとに示す。ピアノロール表現では、緑とオレンジの長方形はそれぞれ右手と左手の音符を表している。複数の演奏者による正解運指は半透明のグレーの円と辺で示されている（色が濃いところは一致度が高いことを示す）。1 次と 2 次の HMM による推定結果は同様に赤と青で示されており、両方の結果が同じ箇所は 2 次の HMM の結果のみが示されている。

る重要性は、多声部の例である図 5(b) の結果からも理解できる。2 次の HMM による右手パートの運指は正解と一致しており、メロディーライン (F4-G4-A4-B♭4-C5-D5-C5) に対して好ましいものになっている。このメロディーラインは音符レベルでは高次の遷移を含んでいる。表 2 に示した比較的高い一致率からも示唆される通り、HMM により推定された運指は大部分において許容できるものであり、系列の最適化に対するモデルの非自明な能力を示している。これらのモデルによる運指の推定結果は、他の例も含めて付随する Web ページ^{*3}から入手でき、その内容は可視化ツール^{*4}により見ることができる。

以下では、HMM に基づく手法による推定結果の典型的な誤りに注目する^{*5}。図 5(a) の左手パートでは、三番目の音 G3 に対する運指誤り 1 はフレーズ境界により生じている。フレーズ境界をまたぐ指の動きはフレーズ内での動きに比べより制約を受けにくい。二番目から三番目の音符への遷移 G2-G3 に対する運指 5-1 は 5-5 よりも一般的には

自然であるが、人間の演奏者の選択が後者であるのは、この遷移が休符で遮られた“dead interval”と見なされ、後続の遷移に対しては G3 を指 5 で弾く方がより適切であるという理由による。同様のフレーズ境界による影響は、図 5(d) の左手の運指での推定誤りにも見られる。バス音の後に和音が二つ続くこの三つ組み和音進行では、1 次と 2 次の HMM は両方とも二番目の和音で誤りを起こした。もしバス音と二番目の和音がスラーでつながっていたら推定された運指は適しているとも言えるが、この箇所では与えられたアーティキュレーションは二番目と三番目の和音の間につながりにより重要性を与えていて、その結果異なる運指になっている。

図 5(c) では、HMM により推定された左手の運指が人間の演奏者によって用いられる運指よりも小さい指の範囲を使っており、同様のケースがしばしば見られた。この楽譜の前半では同様に、両方の手において、通常 1-3 や 2-4 など弾かれる長三度音程や短三度音程に 1-2 や 2-3 の運指が割り当てられている。通常 1-5 で弾かれる五度の音程にも他の指が割り当てられることが時々あった。この楽譜の後半のバス音に対する運指番号 4 の誤った割り当てと同様に、通常は 1 または 5 の指で弾かれる上向きまたは下向きの頂点音に他の指が割り当てられることもあった。

*3 <http://statpianofingering.github.io/demo-ja.html>

*4 <https://fingeringdata.github.io/FingeringVisualizer-ja.html>

*5 ここでは形式的に「誤り」とは、推定した運指番号が正解データのどれとも異なる場合として定義する。前述の通り、正しい運指はただ一つではないため、このように定義した誤りが不可能な運指とは必ずしも言えない。

図 5(e) の右手パートと左手パートの後半では、HMM により推定された運指が人間の演奏者のものよりも広い指の範囲を使うケースが観察される。この例のように両手がオーバーラップする楽譜では、通常よりも狭い範囲の指がしばしば使われるが、これは両手を独立に記述している HMM では捉えられない現象である。さらに、右手パートの Ab4 に対する HMM による指 5 の割り当ては、メロディーラインの Bb4-Ab4 の滑らかな遷移を困難にするため適していない。これらの音符は和音により分断されているため、適切な運指を見つけるには長距離の依存性を考慮に入れる必要がある。人間の演奏者による運指ではこのような非常に長距離の依存性が時に観察されるが、それは推定誤りにつながる。また両手パートの誤った割り当てによる推定誤りもあった。

上に述べたタイプの誤りは音符系列の高レベルの文脈に関係しており、計算論的に容易に記述できないため、定量的な誤り解析はできないが、これらは著者により見つかった典型的な推定誤りのタイプである。また、推定誤りの一部は運指の強い系列依存性により他の誤りにより引き起こされることもあり、この状況はさらに複雑である。一方で、その出現頻度に関わりなく、これらの典型的な誤りは現在のモデルの限界を示しており、モデルを改良する方法について示唆を与えるものである。

4.8 議論

上の結果はピアノ運指推定の統計学習アプローチの可能性を明確に示すものであると同時に、このアプローチの限界と拡張の可能性を明らかにした。ここではこれらを今後の課題としてまとめる。

【時間特徴量と高レベルの文脈】 打鍵時刻間の間隔、音長、あるいは現在の音符の打鍵時刻と一つ前の音符の離鍵時刻の間隔などの時間特徴量は、系列依存性の規則性をより正確に捉えるために重要である。例えば、時間的に重複している音符の間に生じる制約や運指の選択の自由度の変化などである。時間特徴量は運指に影響を与える、意図されたフレーズ構造を推定するためにも重要である。フレーズ、声部、頂点音、あるいは反復構造を捉えるための高レベルの文脈も最適な運指を見つけるために重要な役割を果たしている。和音 HMM と DNN はこれらの問題の解決に重要となる可能性がある。前者は離鍵時刻の情報や和音レベルの長距離依存性を取り込み、後者は効率的な計算により多くの特徴量を扱うのに適しているからである。

【両手の相互依存性】 両手の運指の相互依存性は、両手が重なり合う場合や並行的な動作をするときなどに重要である [6, 40]。文献 [24] で扱われた両手パートの推定もピアノ運指を完全に理解するためには必要である。

【個人性とモデル適応】 本稿ではピアノ運指の一般的な性質に注目したが、運指を理解するためには個人性も重要

である [26]。指の長さや演奏の習得レベルなどの個人的差異は、運指の決定に重要な要因である [23, 41]。なるべく少ないデータで運指の個人的スタイルに適用できるようにするためには、これらの要因を変数で表し、一般的な性質を定数として表すモデルが理想的であろう。

【コスト的観点 対 統計的観点】 今回の評価結果では、コストに基づく手法や制約に基づく手法に比べて、統計学習アプローチの長所が示された。また、運指決定の多くの側面が統計学習と確率的最適化の原理で説明できることもわかった。一方で、その限界についても明らかになった。1 章で論じたように、運指のメカニズムを理解するには、最終的には運指のコストの定量化が必要だと考えられる。図 5(c) の運指にある小さな指の範囲の選択は、高い確率の運指と、コストが低い運指とが一般には異なることを例示している。運指のコストと確率は背後にある決定過程の原因と結果の関係にあり、前者は後者から単純に推定することはできない。しかし、運指の決定過程において統計学習それ自体が重要である可能性も気に留めるべきである。例えば、現代の運指の実践（例えば [6]）と古典的な実践（例えば [3]）とは大きく異なる。これらの問題についてのさらなる研究は今後の課題とする。

5. 演奏難易度尺度とピアノ用編曲への応用

5.1 演奏難易度の定式化

統計モデルに基づくピアノ運指モデルの応用として、定量的演奏難易度尺度の定式化とピアノ用編曲への応用について説明する [32]。ピアノ用編曲やピアノ採譜の定式化において、適切なピアノ楽譜を出力するために、演奏可能なピアノ楽譜を計算論的な言葉で特徴付ける必要がある。従来、演奏可能なピアノ楽譜の特徴として、「片手で同時に発音できる音符数は 5 まで」や「片手で弾ける最大の音程は 14 半音」などの条件が考えられてきた [42, 43]。しかし、これらの条件は演奏可能な楽譜の必要条件とも十分条件とも見せせない。現実にはアルペジオなどで演奏される 6 音符以上からなる和音や 1 オクターブを大きく超える音程などを含む楽譜が存在すること、および和音を含まないメロディーでも速いテンポでは演奏不可能になることもあるからである。演奏可能である条件は、演奏者の技術レベルに依存し、テンポに対しても連続的に変化するため、全ての状況で妥当である完全な記述は難しいであろう。この問題の解決案として、定量的な演奏難易度を定式化して、それを各々の技術レベルやテンポなどにおける演奏可能な楽譜の尺度として用いるアプローチが考えられる [29, 44]。

演奏難易度の定量的尺度は演奏に必要なコストに基づき定義できると考えられる。統計的観点では、運指モデルの確率値の対数の負値である確率コストが演奏コストの自然な候補である。テンポの依存性を取り入れるため、確率コストの時間率として演奏難易度 $D(t)$ を次のように定義

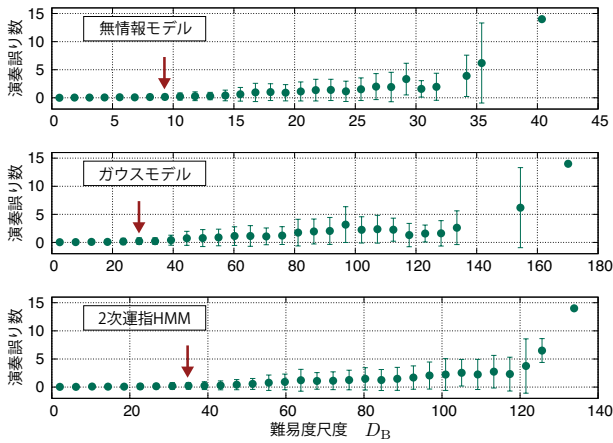


図 6 演奏難易度尺度 D_B の値と演奏誤りの数の関係. 点と縦線は平均値と標準偏差を示している. 矢印は演奏誤りのオンセットを表している (本文参照).

する.

$$D(t) = -\ln P(\mathbf{p}(t)) / \Delta t \quad (16)$$

ここで, Δt は時間幅, $\mathbf{p}(t)$ は時間 $[t - \Delta t/2, t + \Delta t/2]$ 中にある音高系列, および $P(\mathbf{p}(t))$ は音高系列の生成確率を表す. 運指モデルにおいては, 次のように音高と運指の同時確率を用いて難易度を定義できる.

$$D(t) = -\ln P(\mathbf{p}(t), \mathbf{f}(t)) / \Delta t \quad (17)$$

ここで $\mathbf{f}(t)$ は音高系列 $\mathbf{p}(t)$ に対する運指番号の系列である. 運指が未知のときは, 推定された運指 $\hat{\mathbf{f}}(t)$ を代入して計算できる. 打鍵時刻 t_n を持つ各音符 n に対して, $D(n) = D(t_n)$ と記す.

運指モデルに基づく演奏難易度は両手それぞれに対して定義でき, $D_L(t)$ や $D_R(t)$ と記す. さらにこれらの和として両手の総合的な難易度 $D_B(t) = D_L(t) + D_R(t)$ を定義できる. ピアノ曲では, 片手パートは両方とも難易度が高くない場合でも, 両手では難易度が高い曲があるため, $D_L(t)$ と $D_R(t)$ に合わせて合計難易度 $D_B(t)$ も必要になると考えられる.

先行研究 [29, 44] では, 演奏速度, 音符密度, 音高のエントロピー, 手の移動率, 手の広がり度合い, および運指の複雑さなどの特徴量がピアノ楽譜の難易度を推定するために考えられている. これらの特徴量は, 明示的ではないが, 上記の難易度尺度に取り込まれている. 式 (16) の生成確率として音高の一様分布を用いれば, 音符密度と演奏速度のみを考慮した演奏難易度を定義できる (これを「無情報モデル」による難易度と呼ぶ). また, 音高の生成確率として音高遷移確率が音程に関するガウス分布に比例するモデルを考えることで, これらに加えて音高エントロピー, 手の移動率, および手の広がり度合いを取り込んだ難易度を定義できる (これを「ガウスモデル」による難易度と呼ぶ). 式 (17) のように運指モデルを用いることで, さらに運指の複雑度を取り込んだ難易度が定義できる.

表 6 演奏誤り予測の精度. 閾値は F 値に対して最適化した表中の値を用いた.

モデル	閾値 ($D_R^{\text{th}}, D_L^{\text{th}}, D_B^{\text{th}}$)	\mathcal{F}	\mathcal{P}	\mathcal{R}
無情報モデル	(9, 10, 14)	52.4	43.0	67.1
ガウスモデル	(30, 30, 42)	54.2	46.3	65.2
1 次運指 HMM	(37, 37, 50)	54.3	48.7	61.3
2 次運指 HMM	(39, 34, 49)	54.6	47.4	64.5

5.2 演奏誤り予測精度に基づく演奏難易度の評価

前節に述べた尺度が実際の演奏難易度をどれだけ反映しているかを定量的に調べるため, 演奏誤りの頻度との関係を調べる. ここでは, 文献 [45] で用いられたクラシックピアノ曲 60 曲の演奏 MIDI ファイル 180 曲からなるデータを用いる. これは, 各曲に対して国際的ピアノコンクールにおいて録音された 3 人の異なる演奏者の演奏が収められたデータである. このデータでは, 演奏中の音符は対応する楽譜の音符と対応づけがとられており, 音高誤り, 余分な音符, および不足音符の演奏誤りラベルが与えられている. 時間に関する演奏誤りに関してはラベル付けされておらず, ここでも時間に関する誤りは扱わないものとする. この演奏データに含まれる誤りは, ピアノ上級者が十分時間をかけて練習した曲においてなお生じた誤りであると言える.

演奏内の各音符の発音時刻に対して演奏難易度の値を計算し, 同時にこの発音時刻の周り時間幅 Δt 中にある演奏誤りの数を計算する. 以下では Δt は 1 秒とする. 運指モデルとしては, 1 次と 2 次の HMM に基づくモデルを用いる. またガウスモデルでは, 標準偏差の値は 6.38 半音とした [24].

図 6 に 3 つのモデルを用いて計算した演奏難易度尺度 D_B の値と演奏誤りの数の関係を示す. それぞれのモデルにおいて演奏誤りのオンセットに対応する難易度の値があることがわかる. 難易度がこの値よりも小さい時は誤りの数はほぼゼロであり, この値よりも大きい時は誤りの数が徐々に増加している. これは, 演奏難易度尺度が演奏誤りの指標として用いられることを示している.

比較評価として, 3 つのモデルそれぞれに対して, 難易度に対して閾値処理により演奏誤りを予測して, その予測精度を測定した. 片手と両手の難易度に対して 3 つの閾値 $D_R^{\text{th}}, D_L^{\text{th}}, D_B^{\text{th}}$ を考え, いずれかの閾値よりも対応する難易度の値が高いときに誤りが起こると予測した. 実際に演奏誤りがあるかどうかを調べることで, 適合率 \mathcal{P} , 再現率 \mathcal{R} , および F 値 \mathcal{F} を計算した. 閾値は F 値に関して最適化した. 表 6 に結果を示す. この結果より, モデルの精密さを上げるほど予測性能が向上していることが確かめられる. 同時に, ガウスモデルと運指モデルを用いた場合の差は大きくないこともわかる.

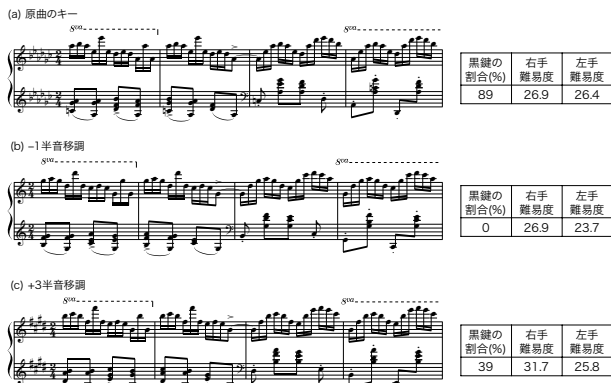


図 7 ショパンの黒鍵のエチュード (Op. 10-5) の一部を移調した際の黒鍵の割合および演奏難易度の変化

5.3 移調に対する演奏難易度の変化の解析

次に、ある楽曲を移調したときに、運指モデルに基づく演奏難易度がどのように変化するかを調べる。楽曲を移調すると、含まれる黒鍵の割合が変わり、ピアノの鍵盤上での音符の配置が変わるため、演奏に必要な運指の動作とその難易度も変わると考えられる。これは単に音高配置の半音単位での相対関係のみを考慮するガウスモデルでは捉えられない側面である。

いくつかの楽曲で、原曲を移調した時の演奏難易度と黒鍵の割合を調べたところ、多くの場合は黒鍵の割合が最小になるキー、つまり一定の調の場合はハ長調やイ短調などに移調した場合に、難易度も最小になることが分かった。これは一般的に黒鍵をより多く含む曲の方が演奏難易度が高いという傾向に合致する結果である。

一方で、異なる移調をした時の黒鍵の割合の大小関係と難易度の大小関係が異なる場合もしばしば観察された。例として、ショパンの黒鍵のエチュード (Op. 10-5) の一部を移調した際の演奏難易度の変化を図 7 に示す。図には黒鍵の割合の変化も示してある。この曲は原曲のキーではほとんどの音符が黒鍵上にあるが (図 7(a)), 1 半音低く移調すると全ての音符が白鍵上に移る (図 7(b))。右手の難易度はこの移調に対して変化していないが、これは全ての音が黒鍵上にある場合と白鍵上にある場合で、鍵盤上の相対位置関係が変わっておらず、運指に変化がないためである。一方で、図 7(c) にある 3 半音高く移調した結果では、黒鍵の割合は原曲よりも少なくなっているが、右手の難易度は高くなっている。実際に、この移調では右手パートで白鍵と黒鍵が両方現れており、演奏の際に必要な手指の動きがより大きくなっていることから、この移調での演奏は前述の 2 つの場合よりも難易度が高いと考えられる。

もう一つの例として、メシアンの楽曲の一部に対する結果を図 8 に示す。この曲では右手パートに速い動きと広い音程を持ち、全音音階から外れた音形が現れる。原曲を 3 度低く移調した時に黒鍵の割合が最小になる一方で、右手の難易度は原曲よりも高くなった。難易度が最小になった

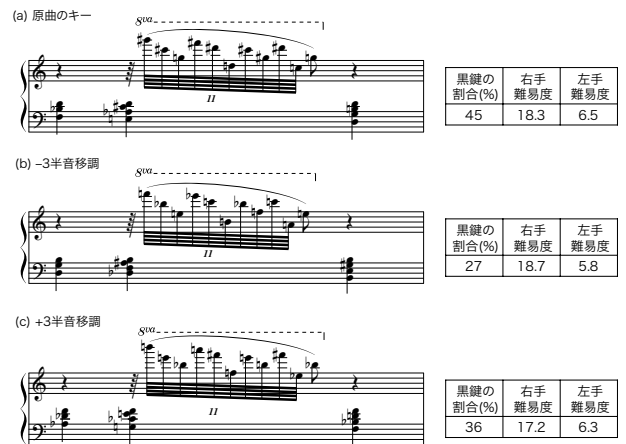


図 8 メシアンの子供イエスに注ぐ 20 の眼差しの第 11 曲の一部を移調した際の黒鍵の割合および演奏難易度の変化

のは、3 度高く移調した時であった。著者が実際に演奏をした際の主観評価では、これらの 3 つの場合の演奏難易度の順番は、運指モデルから計算されたものと同じ順序と一致した。以上の結果は、黒鍵の割合を見るなどの単純な論理では難易度が評価しにくい楽曲に対しても、運指モデルに基づく演奏難易度が実際の主観的演奏難易度の指標として使える可能性を示唆している。

複数の作曲家のピアノ楽曲に対して、移調に対する難易度および黒鍵の割合の変化を計算した結果を図 9 に示す。運指モデルを学習したデータにはほぼ曲が含まれていない図中の作曲家 5 名の曲を分析対象とした。楽曲は Web 上で入手できた MIDI ファイルから無作為に選択肢した。用いた楽曲数は、クーペラン 36 曲、ヘンデル 70 曲、ブゾーニ 33 曲、スクリャービン 42 曲、カプースチン 37 曲である。図中に示した生誕年の通り、最初の 2 人はバロック時代の作曲家であり、主にチェンバロのための鍵盤楽曲を多く作曲した。残りの 3 人はいずれも 19 世紀後半以降の近代・現代の作曲家であり、現代のものに近いピアノのための鍵盤楽曲を多く作曲した。図 9 では各々の作曲家に対して、全ての曲を移調した場合の難易度 (右手と左手の合計値) と黒鍵の割合の平均値を示している。

図 9 の結果では、バロック時代の 2 人の作曲家では、移調度数 0 の場合に黒鍵の割合が最小になっている。難易度は移調度数が 0 付近で小さくはなっているものの、移調度数 0 で最小にはなっていないことが分かる。一方で、近現代の作曲家の内、ブゾーニとカプースチンでは、難易度が移調度数 0 の時に最小になっていることが分かる。黒鍵の割合は、ブゾーニは移調度数 0 の時に最小であるが、カプースチンでは移調度数 2 の時に最小であり、移調度数 0 から 3 でほぼ同じ値であった。スクリャービンは、他の作曲家とは対照的に、移調度数が 0 から大きく離れたところで難易度も黒鍵の割合も小さくなった。

移調に対する黒鍵の割合の変化は、楽曲データにおける調の分布を表していると考えられる。よって、バロックの

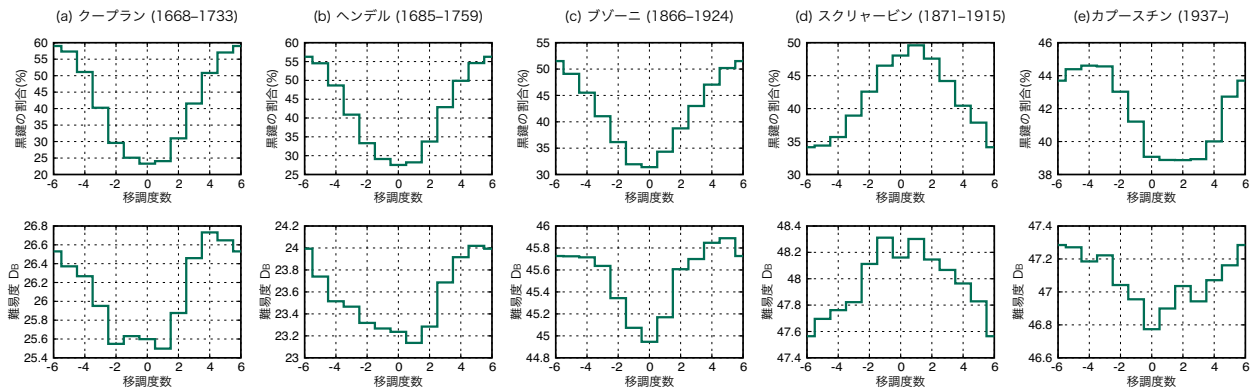


図9 複数の作曲家のピアノ楽曲を移調した時の難易度および黒鍵の割合の変化。移調度数は移調でのシャープあるいはフラットの数の変化を表しており、例えば +1 は完全 5 度上 (7 半音上) への移調を表し、-2 は長 2 度下 (2 半音下) への移調を表す。

作曲家とブゾーニでは、変化記号が少ない調がより多く使われていると解釈できる。カプースチンではシャープが増える方向に移調した場合、黒鍵の割合がほとんど変わらないが、これはこの作曲家の楽曲がジャズのスタイルを取り入れており、そこに現れるブルーノートの影響が原因として考えられる。難易度に関しては、バロックの作曲家では移調度 0 で難易度が最小でない要因として、これらの作曲家の時代では現在と異なる運指が用いられていたことが考えられる [3]。一方で、ブゾーニとカプースチンの楽曲では移調度 0 で難易度が最小となっていることは、これらの作曲家の楽曲が演奏難易度に関して最適化されている、即ちピアノ演奏に適した音高配置を持っていることを示唆している。スクリャービンの場合は、意識的に演奏難易度が高い調を用いることで、独特の音楽表現をしている可能性が考えられる。詳しい音楽学的分析は今後の課題であるが、これらの結果は、提案した演奏難易度尺度によって、演奏動作が楽譜上での音形に与える影響などの音楽学的研究にも応用できる可能性を示していると言える。

5.4 運指モデルに基づくピアノ用編曲手法

運指モデルおよび上記の演奏難易度尺度は合奏曲をピアノ用に編曲するなどのピアノ用編曲手法に応用できる。ここでは文献 [32] の結果について簡単に紹介する。合奏曲のピアノ用編曲では、編曲が演奏可能であることと原曲になるべく忠実であることが基本的な要請と考えられる。一般に原曲の音符をなるべく残せば忠実度が上がるが、音符の数が多すぎて演奏が困難なことが多い。また、演奏者の技術レベルによって、忠実度よりも難易度の低さを優先した編曲が欲しいこともあるであろう。そこで一般には演奏難易度と原曲への忠実度の間にはトレードオフ関係があると考えられる。

文献 [32] では、このトレードオフを統計機械翻訳と同様の方法により計算論的に定式化している。運指モデルに基づくピアノ楽譜の生成モデルにより自然なピアノ編曲を記

述する。一方、編曲の際の音符編集のモデルにより原曲への忠実度を確率的に記述する。これらの二つのモデルの統合モデルにより、確率的最適化問題としてピアノ用編曲アルゴリズムが導出できる。また、編曲結果の難易度をコントロールするために、演奏難易度尺度に対する制約を考え、入力としてユーザーが与える演奏難易度の値を超えないような編曲結果を求める制約付き最適化アルゴリズムを導出できる。

文献 [32] で得られたピアノ編曲の例は Web ページ*6 から見る事ができる。プロの編曲者による自動編曲の評価結果において、客観的な難易度尺度の制約により主観的な難易度尺度の制御ができていないこと、難易度と原曲への忠実度のトレードオフ関係が編曲結果に反映されていること、および運指モデルを用いることでより自然で演奏不可能な音符の少ない編曲が自動生成できることが確認された。文献 [32] では少量のデータで学習された 1 次の運指 HMM を用いているが、高次の HMM を用いることや PIG データセットを用いて学習したモデルを用いることで、編曲の品質のさらなる向上が見込まれる。

6. 結論

本稿では、新たに公開したピアノ運指のデータセットと統計学習と推論に基づく運指推定手法について論じた。統計モデルに基づく手法は、データから効率的にパラメータが推定できる長所を持ち、十分な量の学習データがあれば、1 次の HMM ですら代表的な制約に基づく手法 [25] よりも推定精度が高いことがわかった。高次の HMM により長距離の系列依存性を取り込むことにより精度はさらに改善し、現在の最高性能を達成した。基礎的なネットワーク構造を持つ DNN に基づく手法も比較し、HMM よりも性能が低いことを明らかにした。本研究で達成した最高の推定精度は人間の演奏者間の運指の一致度よりもまだ低く、さらなるモデルの改良の方向性について議論した。また、運

*6 <https://pianoarrangement.github.io/demo.html>

指モデルに基づく定量的演奏難易度尺度の定式化およびその応用可能性についても論じた。

提案したピアノ運指の生成モデルは応用を広げる上でも長所が多い。さらにモデルや手法を改良するには、より多くの学習データがあることが望ましい。データ量を大きく増やすには、動画処理 [46,47], 光学的楽譜認識 [48], およびセンサ技術 [49] などを用いるデータ収集の自動化手法が必要であろう。そうした自動手法に対して、ノイズを含むデータを補完して精度を向上するため、本運指モデルを事前モデルとして用いることも考えられる。また、本モデルは自動ピアノ採譜やピアノ音楽生成 [32,50] のための事前モデルとして、演奏難易度の制御や出力楽譜の自然さを高めるために用いることもできる。

謝辞 有用な議論をして頂いた嵯峨山茂樹氏と渡邊健二氏、運指データの収集にご協力頂いた古屋晋一氏を含む多くの皆様、VNS法のコードを提供して頂いたDorien Herremans氏、DNN手法の実装にご協力頂いた上野聖氏に感謝の意を表す。本研究は、科研費16H01744, 16H02917, 16K00501, 16J05486, 17H00749, 19K20340, JST ACCEL No. JPMJAC1602, および柏森情報科学振興財団からの支援を受けた。

付 録

A.1 組換え一致率の計算アルゴリズム

この付録では、4.3節で導入した組換え一致率の計算アルゴリズムについて説明する。(運指番号などの)推定された系列を $(f_n^{\text{est}})_{n=1}^N$ で表し(N は音符数)、複数の正解系列を $(f_n^{(g)})_{n=1}^N$, $g \in \{1, \dots, N_g\}$ で表す(N_g は正解系列の数)。組換え編集コスト E_{rec} はこの複数の正解から推定系列を再構成するためのコストの合計として定式化する。組換え正解系列 $(f_n^{\text{rec}})_{n=1}^N$ は、 $f_n^{\text{rec}} = f_n^{(z_n)}$ として表現できる。ここで $z_n \in \{1, \dots, N_g\}$ は n 番目の音符において参照する正解系列を示している。

組換えコストは系列 z_n に対する遷移コストとして次のように表される。

$$C(z_{n-1}, z_n) = \begin{cases} 0, & z_n = z_{n-1} \\ C_{\text{rec}}, & z_n \neq z_{n-1} \text{ and } f_n^{(z_n)} = f_n^{(z_{n-1})} \\ C'_{\text{rec}}, & \text{otherwise} \end{cases} \quad (\text{A.1})$$

ここで、 C_{rec} は組換えられる正解系列の運指番号が一致している箇所における組換えコストであり、それ以外の箇所における組換えコストを C'_{rec} として導入した。 $C'_{\text{rec}} = \infty$ とすれば4.3節で定義した組換えコストが復元される。置換コストは組換え正解系列の f_n^{rec} と推定系列の f_n^{est} 関数として次のように表される。

$$C(f_n^{\text{rec}}, f_n^{\text{est}}) = \begin{cases} 0, & f_n^{\text{rec}} = f_n^{\text{est}} \\ C_{\text{sub}}, & f_n^{\text{rec}} \neq f_n^{\text{est}} \end{cases} \quad (\text{A.2})$$

この時、組換え編集コスト E_{rec} は次のように表現される。

$$E_{\text{rec}}(z_{1:N}) = \sum_{n=2}^N C(z_{n-1}, z_n) + \sum_{n=1}^N C(f_n^{(z_n)}, f_n^{\text{est}}) \quad (\text{A.3})$$

組換え一致率を計算するには、 $E_{\text{rec}}(z_{1:N})$ を最小化する $z_{1:N}$ で表される組換え正解系列を求める必要がある。これはビタビアルゴリズムと同様の標準的な動的計画法を用いて行える。最適な組換え $\hat{z}_{1:N}$ を求めた後、組換え一致率は $M_{\text{rec}} = (N - E_{\text{rec}}(\hat{z}_{1:N}))/N$ として計算できる。

A.2 一致率の漸近形

この付録では一致率の漸近形である式(15)を導出する。パラメータ θ を持つ統計モデルがあるとす。本文で考えたHMMの場合は、 θ は遷移確率と出力確率の集合である。まず、サンプル数 N が大きい極限で次の漸近公式が成り立つことを示す。

$$\theta \sim \theta_0 + \frac{f(\theta_0)}{\sqrt{N}} + \mathcal{O}\left(\frac{1}{N}\right) \quad (\text{A.4})$$

ここで、 θ_0 は θ の漸近値であり f は θ と同じ次元を持つ何らかの関数である。これは不偏推定量に対する次のCramér-Rao不等式[51]の帰結である。

$$\langle (\theta - \theta_0)^2 \rangle \geq I(\theta_0)^{-1}/N \quad (\text{A.5})$$

ここで左辺は θ の分散であり、 θ の標準誤差の二乗として解釈される。また、 $I(\theta_0)$ は単サンプルのフィッシャー情報量である。本稿で扱っている離散分布を含む広いクラスの分布に対する最尤推定量に対しては、漸近的に上記の下限に到達することが知られている[52]。 $N \rightarrow \infty$ の時に $\theta \rightarrow \theta_0$ であることに注意すると、(A.5)の等式は式(A.4)と等価であることがわかる。

統計量 θ の滑らかな関数 $A(\theta)$ を考える。式(A.4)を $A(\theta)$ に代入してテイラー展開の1次を見ると、次が得られる。

$$A(\theta) \sim A(\theta_0) + \frac{\partial A(\theta_0)}{\partial \theta_0} \cdot \frac{f(\theta_0)}{\sqrt{N}} + \mathcal{O}\left(\frac{1}{N}\right) \quad (\text{A.6})$$

N を固定して θ に関する期待値を求めると、これは式(15)と N が大きい極限で等価である。

上の導出からわかる通り、式(15)はモデルパラメータ θ の関数が漸近値 θ_0 の近傍で滑らかである場合に成り立つ。また、式(15)の定数 a と b は漸近値 θ_0 の関数であることも注意が必要である。即ち、これらの定数の値は θ の漸近値によって変わる。

参考文献

- [1] C. Palmer. Music performance. *Annual Review of Psychology*, 48(1):115–138, 1997.
- [2] S. Furuya, M. Flanders, and J. F. Soechting. Hand kinematics of piano playing. *Journal of Neurophysiology*, 106(6):2849–2864, 2011.
- [3] F. Couperin (transl. by M. Halford). *The Art of Playing the Harpsichord*. Alfred Pub. Co. (Original work in French published 1717), 1974.
- [4] C. P. E. Bach (transl. by W. J. Mitchell). *Essay on the True Art of Playing Keyboard Instruments*. W. W. Norton and Company, NY, USA (Original work in German published 1753), 1949.
- [5] D. G. Türk. *Klavierschule, oder Anweisung zum Klavierspielen für Lehrer und Lehrende, mit kritischen Anmerkungen*. Bärenreiter, Kassel (Original work published 1789), 1962.
- [6] J. Musafia. *The Art of Fingering in Piano Playing*. MCA Music, 1971.
- [7] L. Mozart (transl. by E. Knocker; 2nd ed.). *A Treatise on the Fundamental Principles of Violin Playing*. Oxford University Press, Oxford (Original work in German published 1756), 1985.
- [8] F. Sor (ed. by F. M. Harrison). *Method for the Guitar*. Robert Cocks and Co., London (Original work in French published 1830), 1896.
- [9] 野口賢治, 野池賢二, 乾伸雄, 野瀬隆, 小谷善行, 西村恕彦. n グラムの手法を用いたピアノ運指の推論. 情処学会全国大会講演論文集, 2-101, 1996.
- [10] R. Parncutt, J. A. Sloboda, E. F. Clarke, M. Raekallio, and P. Desain. An ergonomic model of keyboard fingering for melodic fragments. *Music Perception*, 14(4):341–382, 1997.
- [11] 関口博之, 英保茂. 計算機によるピアノ演奏動作の生成と表示. 情処学会論文誌, 40(6), pp. 167–177, 1999.
- [12] M. Hart, R. Bosch, and E. Tsai. Finding optimal piano fingerings. *The UMAP Journal*, 21(2):167–177, 2000.
- [13] J. P. Jacobs. Refinements to the ergonomic model for keyboard fingering of Parncutt, Sloboda, Clarke, Raekallio, and Desain. *Music Perception*, 18(4):505–511, 2001.
- [14] 林田教裕, 水谷哲也. 楽曲構造に基づくピアノ運指ルールの論理表現. 情処学会全国大会講演論文集, Vol. 2, pp. 203–204, 2003.
- [15] M. Miura, I. Hirota, N. Hama, and M. Yanagida. Constructing a system for finger-position determination and tablature generation for playing melodies on guitars. *Systems and Computers in Japan*, 35(6):10–19, 2004.
- [16] D. R. Tuohy and W. D. Potter. A genetic algorithm for the automatic generation of playable guitar tablature. In *Proc. ICMC*, pages 499–502, 2005.
- [17] Y. Yonebayashi, H. Kameoka, and S. Sagayama. Automatic decision of piano fingering based on hidden Markov models. In *Proc. IJCAI*, pages 2915–2921, 2007.
- [18] 米林裕一郎, 亀岡弘和, 嵯峨山茂樹. 手の自然な動きを考慮した隠れ変数付き隠れマルコフモデルに基づくピアノ運指決定. 情報処理学会音楽情報科学研究報告, Vol. 2007-MUS-71, No. 29, pp. 1–6, 2007.
- [19] A. Al Kasimi, E. Nichols, and C. Raphael. A simple algorithm for automatic generation of polyphonic piano fingerings. In *Proc. ISMIR*, pages 355–356, 2007.
- [20] R. De Prisco, G. Zaccagnino, and R. Zaccagnino. A differential evolution algorithm assisted by ANFIS for music fingering. In *Swarm and Evolutionary Computation*, pages 48–56. Springer, 2012.
- [21] 春原雅彦, 松原正樹, 前田俊太郎, 齋藤博昭. 左手と和音に対応したピアノ運指の自動推定システム. 情処学会全国大会講演論文集, 2-147, 2010.
- [22] 若松万紗子, 松井知己. 和音に対するピアノ運指決定法. 京都大学数理解析研究所数理解析研究所講究録, Vol. 1773, pp. 87–95, 2012.
- [23] W. Nagata, S. Sako, and T. Kitamura. Violin fingering estimation according to skill level based on hidden Markov model. In *Proc. ICMC*, pages 1233–1238, 2014.
- [24] E. Nakamura, N. Ono, and S. Sagayama. Merged-output HMM for piano fingering of both hands. In *Proc. ISMIR*, pages 531–536, 2014.
- [25] M. Balliauw, D. Herremans, D. P. Cuervo, and K. Sörensen. A variable neighborhood search algorithm to generate piano fingerings for polyphonic sheet music. *International Transactions in Operational Research*, 24(3):509–535, 2017.
- [26] 中村栄太, 齋藤康之, 吉井和佳. ピアノ運指データを用いた運指の個人性の解析. 第 123 回情報処理学会音楽情報科学研究報告, Vol. 2019-MUS-123, No. 42, pp. 1–6, 2019.
- [27] C.-C. Lin and D. S.-M. Liu. An intelligent virtual piano tutor. In *Proc. ACM International Conference on Virtual Reality Continuum and Its Applications*, pages 353–356, 2006.
- [28] Y. Takegawa, T. Terada, and S. Nishio. Design and implementation of a real-time fingering detection system for piano performance. In *Proc. ICMC*, pages 67–74, 2006.
- [29] V. Sébastien, H. Ralambondrainy, O. Sébastien, and N. Conruyt. Score analyzer: Automatically determining scores difficulty level for instrumental e-learning. In *Proc. ISMIR*, pages 571–576, 2012.
- [30] Y. Saito, E. Nakamura, R. Sato, S. Agata, Y. Igarashi, and S. Sagayama. Conversion from standard MIDI files to vertical line notation scores and automatic decision of piano fingering for beginners. In *Proc. International Conference on Technologies for Music Notation and Representation*, pages 200–211, 2016.
- [31] G. Hori, H. Kameoka, and S. Sagayama. Input-output HMM applied to automatic arrangement for guitars. *J. Info. Processing Soc. Japan*, 21(2):264–271, 2013.
- [32] E. Nakamura and K. Yoshii. Statistical piano reduction controlling performance difficulty. *APSIPA Transactions on Signal and Information Processing*, 7(e13):1–12, 2018.
- [33] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press Cambridge, 2016.
- [34] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77(2):257–286, 1989.
- [35] E. Nakamura, N. Ono, S. Sagayama, and K. Watanabe. A stochastic temporal model of polyphonic MIDI performance with ornaments. *J. New Music Res.*, 44(4):287–304, 2015.
- [36] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440:1007–10012, 2006.
- [37] J. Sakellariou, F. Tria, V. Loreto, and F. Pacht. Maximum entropy models capture melodic styles. *Scientific Reports*, 7(9172):1–9, 2017.
- [38] B. Shahriari, K. Swersky, and Z. Wang. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE*, 104(1):148–175, 2016.
- [39] D. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proc. ICLR*, pages 1–15, 2015.

- [40] R. Parncutt, J. A. Sloboda, and E. F. Clarke. Interdependence of right and left hands in sight-read, written, and rehearsed fingerings of parallel melodic piano music. *Australian Journal of Psychology*, 51(3):204–210, 1999.
- [41] J. A. Sloboda, E. F. Clarke, R. Parncutt, and M. Raekallio. Determinants of finger choice in piano sight-reading. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1):185–203, 1998.
- [42] Onuma, S.; Hamanaka, M.: Piano arrangement system based on composers’ arrangement processes, in *International Computer Music Conference*, New York, 2010, 191–194.
- [43] Huang, J.-L.; Chiu, S.-C.; Shan, M.-K.: Towards an automatic music arrangement framework using score reduction. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 8(1) (2012), 8:1–8:23.
- [44] Chiu, S.-C.; Chen, M.-S.: A study on difficulty level recognition of piano sheet music, in *IEEE International Symposium on Multimedia*, Irvine, California, 2012, 17–23.
- [45] Nakamura, E.; Yoshii, K.; Katayose, H.: Performance error detection and post-processing for fast and accurate symbolic music alignment, in *International Conference on Music Information Retrieval*, Suzhou, 2017, 347–353.
- [46] A. Oka and M. Hashimoto. Marker-less piano fingering recognition using sequential depth images. In *Proc. Korea-Japan Joint Workshop on Frontiers of Computer Vision*, pages 1–4, 2013.
- [47] J. MacRitchie and N. J. Bailey. Efficient tracking of pianists’ finger movements. *J. New Music Res.*, 42(1):79–95, 2013.
- [48] L. Chen and C. Raphael. Human-directed optical music recognition. *Electronic Imaging*, 2016(17):1–9, 2016.
- [49] D. A. Randolph and B. Di Eugenio. Dactylize: Automatically collecting piano fingering data from performance. In *Proc. Late-Breaking Demo Session ISMIR*, pages 1–3, 2016.
- [50] E. Nakamura, E. Benetos, K. Yoshii, and S. Dixon. Towards complete polyphonic music transcription: Integrating multi-pitch detection and rhythm quantization. In *Proc. ICASSP*, pages 101–105, 2018.
- [51] T. M. Cover and J. A. Thomas. *Elements of Information Theory (2nd ed.)*. John Wiley & Sons, 2006.
- [52] L. E. Lehmann. *Elements of Large-Sample Theory*. Springer, 1998.