

## 弦楽器練習支援のための音符からの運指推定 Fingering Proposal from Music Note for Practicing String Instrument

新村 祐加<sup>†</sup> 堰澤 映<sup>†</sup> 中島 克人<sup>†</sup>  
Yuka Shimmura Akira Sekizawa Katsuto Nakajima

### 1. はじめに

弦楽器はボウイングと呼ばれる右手で弓を動かす技術や弦をはじいて音を鳴らすピチカート等様々な技術や奏法が必要である。特に非熟練者が苦勞するのは左手で弦を押さえ音を鳴らす操作である。弦楽器はギターのようなフレットがないため、正しいポジション（ある音に対する指板の位置）を当てるにはある程度時間がかかる。また、ピアノのように音と鍵盤が一意に決まっておらず、1つの音に対し複数のポジションがあり、曲を演奏する際、どのように指を動かして弾くかという運指によってこのポジションが決まる。運指は前後の音の繋がりや表現の仕方によって決まるが、楽譜にはそれを示す指番号は書かれていない。つまり奏者は、自分の経験を基に運指を決定しなければならないが、非熟練者にとって楽譜から運指を考えることは大変困難なことである。

そこで我々は、機械学習を用いた運指推定システムの構築を目指している。ユーザがシステムに楽譜を入力すると、音符の上に自動的に指番号が付与されるというものである。これにより非熟練者は運指を検討する時間の短縮が見込め、曲の練習により多くの時間を割くことが可能となる。

本稿では、時系列推定で最も一般的な Recurrent Neural Network (RNN) を用いた運指推定システムを提案する。RNNの具体的なモデルとして Seq2Seq を用いて運指推定の精度を検証した。

### 2. 関連研究

これまで、様々な楽器に対して運指推定を行う研究がなされてきた。長田ら[1]は、Hidden Markov Model (HMM) の内部状態を手の形と位置に見立て、出力を音符列としてモデル化することでバイオリンの楽譜の運指推定を行った。習熟度に応じ最適な運指を推定させるため、音符長や休符長、運指の変化の起こりやすさや表現の優先度を考慮したモデルを構築した。評価実験では、教本との一致率やバイオリン経験者による主観評価によって習熟度に応じた運指が推定されていることを確認した。Raboanary[2]らは動的計画法を用いてギターの最適運指を決定するアルゴリズムを2種類提案した。1種類目は物理的に容易な最適運指に基づいたもの、2種類目は演奏者の嗜好に基づいたものである。最適指番号はフレットの位置や音符と指番号の組同士の間隔、手の位置、などを難易度によって重みづけしたものをを用いて求める。演奏者の嗜好に基づく第2の方法では、この重みづけを演奏者自身に変えてもらうことにより達成される。実験の結果、第2の方法の方が第1の方法より好ましい運指が生成されたが、演奏時の精度は第1の方法の方がより向上していることが分かった。

これらの研究では規則性を手動で与えているが、弦楽器の運指を決定する規則性は様々ある点や考慮すべき楽譜

<sup>†</sup> 東京電機大学 Tokyo Denki University

情報が多くあるため、全てに対応する様に地道に重みを与えていくのは困難であると思われる。

### 3. 提案手法

#### 3.1 RNN のネットワーク構造

RNNの具体的なモデルとして、系列から系列への変換で代表的な Seq2Seq[3]を選択した。Seq2Seqは Long short-term memory (LSTM) を用いたエンコーダ・デコーダ型の RNN である。エンコーダは入力系列を高次元の潜在ベクトルへと変換し、デコーダは潜在ベクトルから目的の系列を生成する。LSTMは隠れ状態の数と層の深さという2つの主要なハイパーパラメータを持つ。隠れ状態の数とは LSTM のメモリユニットが持てる状態の数を表す。層の深さとはエンコーダとデコーダそれぞれの層の深さを表す。

#### 3.2 教師データの形式

本稿では弦楽器として具体的にヴィオラを想定し、教師データ (X, Y) の形式を次のように設計した。音符情報を X, 出力である指番号を Y とし、X は音名・変化記号・オクターブ・音符長の四つ組の情報であり、Y は押弦する左手の人差し指 (1番) から小指 (4番) に対応する指番号である。開放弦は 0 番とする。音名は C~B の 7 種類である。オクターブは五線譜の第 3 線からはじまる C~B を 4 とし、オクターブ上を 5, 下を 3 などとしている。変化記号とは音名に対してシャープ、フラット、ダブルシャープ、ダブルフラットが付与されているか否かを表す値であり、それぞれを 1, -1, 2, -2 を割り当てる。変化記号が無い場合は 0 を割り当てる。音符長は 32 分音符を 1 とした場合の相対的な音の長さを整数で表現する。Seq2Seq の出力値は指番号の整数値である。

教師データの X を Seq2Seq に与える際には、 $N+1+M$  個の系列で与える。N は運指推定の対象となる音符の前 N 個分の音符情報を表し、M は対象音符の後ろ M 個分の音符情報を表す。すなわち N, M の大きさが Seq2Seq が考慮する文脈長を表す。

我々が目指す運指推定システムを図 1 に示す。また、データ形式例を表 1 に示す。

### 4. 実験

#### 4.1 データセット

使用するデータセットは初心者用弦楽器教本[4], [5]から抜粋及びクラシック曲[6]から抜粋したフレーズの 5 曲から自作したデータである。今回は休符及び拍子は考慮しない。実験に用いた楽譜は全音符から 32 分音符までの 6 種類で構成されており、合計の音符数は 663 個である。このうち訓練データを 80% にあたる 530 個、テストデータを残りの 20% にあたる 133 個とした。

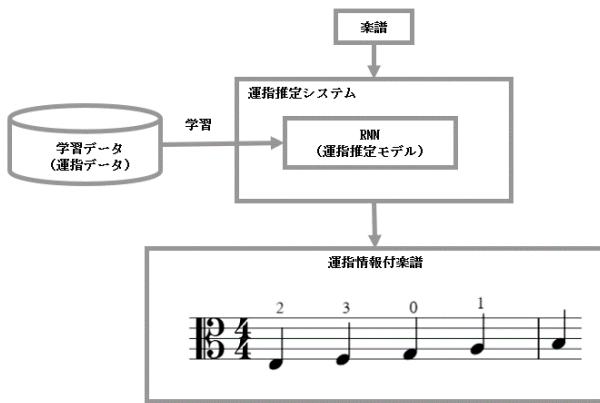


図 1 運指推定システム

表 1 データ形式例

音名	オクターブ	変化記号	音符長	指番号
B	3	-1	4	2
E	4	-1	8	1
G	3	1	8	1
F	3	2	8	0
F	3	1	4	4
E	3	0	8	2
D	4	0	8	0
A	3	0	12	1
B	3	0	4	2

## 4.2 学習

全てのモデルはバッチサイズ 32 で 250 エポックまで学習した。オプティマイザには RMSProp を使い、学習率の初期値には 0.001 を用いた。実験では 1 回ごとに訓練データとテストデータを混ぜてシャッフルし、5 回分の推定精度の平均値を最終的な推定精度として用いた。

## 5. 結果及び考察

実験結果を表 2 に示す。表 2 において、運指推定の対象の音符の前後の合計文脈長が 2 である場合の推定精度を第 3, 4 列に示し、合計文脈長が 4 である場合の推定精度を第 5, 6 列に示している。文脈長の 2-2 という表記は、運指推定の対象音符の前に 2 つの音符、後ろに 2 つの音符を与えた際の推定結果を示す。また層の深さの 5-5 という表記は、エンコーダが 5 層でデコーダが 5 層という意味を表す。

隠れ状態の数については、30 の方が 15 よりも概ね 5 ポイント以上推定精度が高い。層の深さについては、推定精度との相関は見られなかった。隠れ状態の数と文脈長  $N$ ,  $M$  が同じ場合の推定精度を比較すると、層の深さが深い方が推定精度が高くなる場合もあれば、その逆の場合もあった。このことから、運指推定においては層の深さ、すなわちネットワークの表現力よりも、LSTM のユニットの隠れ状態の数、すなわちユニット間で受け渡す文脈情報の量が推定精度に大きな影響を与えたと考えられる。文脈長  $N$ ,  $M$  については、 $N \approx M$  となる場合に推定精度が高くなる傾向が見られた。人手で運指を検討する際も音符の前後の運指を考慮するため、これは直感に沿う結果である。しかし

表 2 実験結果

隠れ状態の数	層の深さ (エンコーダ-デコーダ)	文脈長 (前-後)	精度	文脈長 (前-後)	精度
30	5-5	2-0	52.9	4-0	54.4
15	5-5	2-0	48.0	4-0	46.5
30	3-3	2-0	54.9	4-0	54.4
15	3-3	2-0	53.8	4-0	44.8
30	1-1	2-0	53.8	4-0	48.6
15	1-1	2-0	54.0	4-0	48.0
30	5-5	1-1	<b>57.4</b>	2-2	<b>57.3</b>
15	5-5	1-1	49.2	2-2	46.0
30	3-3	1-1	<b>57.4</b>	2-2	53.2
15	3-3	1-1	52.8	2-2	49.0
30	1-1	1-1	52.8	2-2	53.1
15	1-1	1-1	52.2	2-2	54.6
30	5-5	0-2	50.7	0-4	52.6
15	5-5	0-2	49.3	0-4	46.3
30	3-3	0-2	55.9	0-4	51.7
15	3-3	0-2	52.5	0-4	46.0
30	1-1	0-2	51.9	0-4	54.4
15	1-1	0-2	48.6	0-4	49.8

合計文脈長を 2 から 4 に増やしても精度が上がらないことも確認された。

今回の実験での推定精度は、運指をランダムに推定した場合の期待値である 20% より 20 ポイント以上は高いが、最高推定精度は 57.4% に留まっており、満足な値とは言えず、今後の精度向上が課題である。

## 6. まとめ

我々は、RNN の一種であり系列から系列への変換で代表的な Seq2Seq を用いて弦楽器の運指推定システムを構築した。そして Seq2Seq を構成する LSTM の隠れ状態の数、ネットワークの層の深さ、文脈長を変えて推定精度を検証した。その結果、隠れ状態の数と文脈長が大きな影響を与えることがわかった。しかし最高精度は 60% にも届かず、実用的な運指推定システムの実現にはさらなる検討が必要である。

### 参考文献

- [1] 長田若奈, 他, “隠れマルコフモデルによる習熟度に対応したバイオリン運指推定”, 情処学研報, Vol.2014-MUS-102, No.11, pp.1-6 (2008).
- [2] T.H.Raboanary, et al., “Finding optimal bass guitar fingerings”, in Proc. IEEE AFRICON, pp.65-71 (2017).
- [3] I.Sutskever, et al., “Sequence to sequence learning with neural networks”, In Advances in Neural Information Processing Systems (NIPS), pp.3104-3112 (2014).
- [4] S.Suzuki, “Suzuki Viola School Viola Parts Vol.3”, Alfred Publishing (2009).
- [5] 磯良男, “初心者のためのやさしいヴィオラ入門”, ドレミ楽譜出版社 (2011).
- [6] V.Kalinnikov, “Symphony No.1 in G minor I. Allegro moderato”.