

## LSTM を用いた四声体和声の生成

## Four-part Harmonization Using LSTM

山田竜郎 \*1  
Tatsuro Yamada北原鉄朗 \*2  
Tetsuro Kitahara有江浩明 \*1  
Hiroaki Arie尾形哲也 \*1  
Tetsuya Ogata

\*1 早稲田大学 理工学術院

Faculty of Science and Engineering, Waseda University

\*2 日本大学 文理学部

College of Humanities and Sciences, Nihon University

We build LSTM-RNN models that learn to generate four-part harmonies in accordance with the melody of soprano voices. We conduct learning experiments in which models different in architectures are trained with a hymn dataset. By evaluating their generation performances by using some objective criteria, we compare their characteristics with each other and with an existing Bayesian network model. It is revealed that the bi-directional LSTM model scores higher performance than uni-directional models.

## 1. はじめに

計算機を用いた和声の生成は、自動作曲における一課題として広く取り組まれている。[Ebcioglu 90] はエキスパートシステムを構築したが、和声は共時構造（同時に鳴る音の協和性）と通時構造（各声部の旋律のスムーズ性やコード進行）の双方を持つため、全てのルールを満たす和声を生成するのは困難な場合がある。ルールベース推論に代わる手法としては、遺伝アルゴリズムによる探索 [Phon-amnuaisuk 99] やベイジアンネットワーク (BN) を用いた方法 [Suzuki 14] が提案されている。

近年、自然言語処理分野では、時系列処理の方法として Long short-term memory (LSTM)-RNN [Hochreiter 97] が注目されているが、これを用いた音楽生成の事例も報告されている。たとえば、DeepBach [Hadjeres 16] は、バッハの賛美歌を用いて LSTM を学習することで、ギブスサンプリングによりバッハ風の曲を生成することができる。

しかし深層学習手法を用いた既存研究では、モデルの構成や規模の決定に関する指針が十分報告されていない。本研究は和声生成課題のうち、特にソプラノへの伴奏づけに焦点を当て、複数の LSTM モデルを構築し学習実験を行う。モデル構成を変えた時に、生成の性能や特性がいかに変化するか調査し、従来研究の BN モデルとも比較しながら報告する。

## 2. LSTM による四声体和声生成手法

## 2.1 ベクトル表現への変換

四声体和声からなる楽曲の MIDI ファイルを 148 次元ベクトルの時系列に変換する。ここで四声体和声はソプラノ、アルト、テノール、バスからなる和声であり、各パート高々1つの音高のみが演奏されるものとする。4分音符を1拍とし、変換された系列の個々のベクトルは、拍単位の演奏を表すとする。

いま、 $n$  拍目のソプラノパートの旋律を

$$s_n = \left( s_b^{(n)}, s_r^{(n)}, s_L^{(n)}, s_{L+1}^{(n)}, \dots, s_{L+34}^{(n)} \right) \quad (s_i^{(n)} \in \{0, 1\})$$

と表す。ここで、 $s_b^{(n)}$  は楽曲の開始および終了時にそれぞれ1回ずつのみ1を取る変数、 $s_r^{(n)}$  はソプラノパートが休符（無音）のときに1を取る変数、それ以外是对応するノートナンバーの音が演奏される時に1を取る変数である（ $L$  は対象

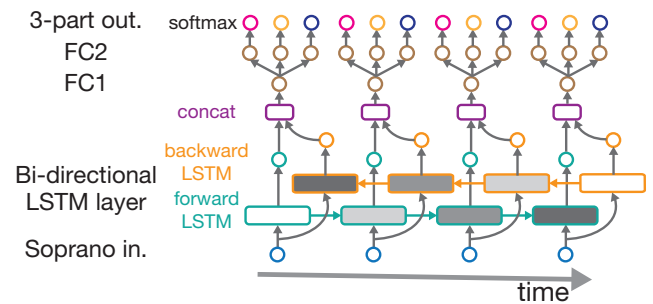


図 1: LSTM を用いた和声生成モデル。

とする最低音ノートナンバー)。ソプラノパートが複数の音を同時に演奏することはないと仮定しているため、このベクトルは 1-hot ベクトルとなる。同じ時刻のアルト、テノール、バスパートの旋律も同様に 37 次元ベクトルとして表し ( $a_n, t_n, b_n$  とする)、これらを連結したベクトル  $h_n = (s_n, a_n, t_n, b_n)$  を、 $n$  拍目の四声体和声を表すベクトルとする。

上述のように、4分音符単位でベクトル化されるため、与えられる楽曲には4分音符よりも細かい音符はないものと仮定する。また、同じ音高で4分音符が2つ連続する旋律と2分音符が1つのみの旋律は区別されないものとする。

## 2.2 LSTM モデルの学習

2.1 節の方法でベクトル化された四声体和声を用いて図1に示すモデルを学習する。図1のモデルでは、順方向の時間的依存関係と同時に逆方向の時間的依存関係をも考慮した双方向のモデルになっている。このようなモデルにしたのは、和声は一般に直前の和音に依存して決まるだけでなく、ある和音に解決するためにその直前にある和音を置くなど、直後の和音にも依存して決まるからである。双方向のモデルの有効性を検証するため、(1) 順方向のみ、(2) 逆方向のみでも実験を行う（双方向のモデルは (3) とする）。

LSTM 層に  $[s_0, s_1, \dots]$  が入力されると、LSTM 層の出力が全結合層に (FC1, FC2) に入力される。活性化関数には tanh 関数を用いる。この出力に対してソフトマックス関数を適用し、 $[(\hat{a}_0, \hat{t}_0, \hat{b}_0), (\hat{a}_1, \hat{t}_1, \hat{b}_1), \dots]$  を得る。この各ベクトルは、各パート、各拍における音高の確率分布を表す。この値と正解  $[(a_n, t_n, b_n)]$  との交差エントロピーが最小になるようにモデ

ルの学習を行う．なお，LSTM 層，FC1 層のノード数はそれぞれ 148 とした．FC2，ソフトマックス層のノード数は，出力データの次元と同じとなるため  $37 \times 3=111$  である．

**2.3 ソプラノパートが与えられた際他パートの生成**  
ソプラノパートの旋律のみが与えられ，アルト，テノール，バスパートを LSTM で生成する場合は，まず，ソプラノパートの旋律を 2.1 節と同様の方法でベクトル化し， $[s_0, s_1, \dots]$  を得る．その後，2.2 節で述べたモデルを用いて  $[(a_0, t_0, b_0), (a_1, t_1, b_1), \dots]$  を推定する．最後に，各拍において  $a_n, t_n, b_n$  の最大値をそれぞれ求め，それに対応するノートナンバー（または休符）を出力する．

### 3. 実験 1

#### 3.1 実験条件

実験は，結果を直接 [Suzuki 14] と比較できるようにするため，[Suzuki 14] と同じデータおよび評価基準で行った．具体的には，[日本基督教団讃美歌委員会 83] に収録されている四声体和声の賛美歌から抽出した 224 曲を学習データとし ([Suzuki 14] は学習データに 254 曲を用いているが，そのうち 30 曲は 4 分音符よりも細かい音を含んでいるため除外した)，[池内 67] から選んだ 32 曲を評価データとした．モデルには，

- (1) 順方向のみの LSTM
- (2) 逆方向のみの LSTM
- (3) 双方向の LSTM

の 3 種類を用い，[Suzuki 14] に掲載されている

- (4) ベイジアンネットワーク (BN)

の結果と比較する．比較基準は，[Suzuki 14] に合わせ

- Crit. 1 不協和音率
- Crit. 2 ノンダイアトニック (ND) 音含有率
- Crit. 3 最後のコードが C である割合
- Crit. 4 3 声以上が同じ音名である割合
- Crit. 5 バスが連続して完全 4 度より大きく跳躍している箇所の割合
- Crit. 6 曲中で各声部が使用する平均音高数を用いる．Crit. 1, 2, 4, 5 は値が小さい方が，Crit. 3, 6 は値が大きい方が良い結果である．LSTM の最適化手法は Adam を採用し，各モデルとも学習率を 0.001 とし，10,000 回学習を行った．

#### 3.2 実験結果

結果を表 1 に示す．BN モデルに比べてどの LSTM モデルも Crit. 1 (不協和音率) が高かった．一方で Crit. 4 (3 声以上の重複率) や Crit. 5 (バスの連続跳躍率) は LSTM モデルの方が少なかった．Crit. 6 (各声部の平均出現音高数) は BN モデルよりも少なかった．これらの原因として，生成結果に同じ音の連続が多いことが挙げられる．図 2 は双方向 LSTM モデルによる生成例であるが，ソプラノが動いているにもかかわらず，アルト，テノール，バスが同じ音を生成し続けている．これにより，不協和音の割合が高くなり，バスの跳躍箇所や平均使用音高数が少なくなったと考えられる．

この同音連続生成という傾向は，モデルと学習データの双方の特性の組み合わせに起因するものであると思われる．今回用いた学習データは，2 分音符以上の長さで旋律が遷移する箇所が多く，25 小節以上からなるのに対し，評価データはほとんどが 4 分音符ごとに音が変わる 4 小節の楽曲である．LSTM モデルは全時間（ただし順方向モデルの場合前側のみ，逆方向モデルの場合後側のみ）の情報を考慮して現在時刻の出力を決



図 2: 双方向モデルによる生成例．

めるよう学習するため，この学習データと評価データ間のタイムスケールの差異に適応できなかった可能性がある．一方で BN モデルは一つ前と後の和音のみから現在の和音が決まるため，この差異にはほとんど影響を受けなかったと考えられる．

LSTM の 3 つのモデルについて比べると，順方向モデルよりも逆方向，双方向モデルは Crit. 3 (最後のコードが C である割合) が高かった．これは，最終音を生成するときに，順方向モデルはそれが最終音であるという情報を持っていないのに対し，逆方向および双方向モデルは持っている（曲終端記号の情報が逆方向 LSTM によって得られる）ことによると考えられる．

### 4. 実験 2

#### 4.1 実験条件

実験 1 では，評価データにおけるソプラノパートの旋律が学習データに比べて頻繁に音高が遷移するため，そういった旋律に適したアルト，テノール，バスの旋律が生成されなかった可能性が示唆された．そこで，評価データと学習データとで旋律の特徴が共通している場合を確認するため，学習データと評価データを同じ楽曲集から抜粋した場合の実験も行った．具体的には，[日本基督教団讃美歌委員会 83] から抜粋した 124 曲を学習データとし，同じ楽曲集から抜粋した別の 100 曲を評価データとした．それ以外の実験条件は実験 1 と同様である．

#### 4.2 実験結果

実験結果を表 2 に示す．表中において「正解」は [日本基督教団讃美歌委員会 83] のデータセットにおけるアルト，テノール，バスに関する結果である．和声付けの特性上データセット中の和声が必ずしも唯一の正解ではないが，ここでは便宜上「正解」と称する．Crit. 1 (不協和音率) をみると「正解」よりは高いものの，いずれも実験 1 よりも低い値となっている．これは，この実験における評価データのソプラノパートは，実験 1 のものに比べて 2 分音符や全音符など長い音符が多く，実験 1 のように，ソプラノが頻繁に動くにもかかわらずアルト，テノール，バスで同じ音が長く続くために不協和音が生じるという事態がなかったからと考えられる．一方，Crit. 2 (ND 音含有率) は実験 1 よりも高い値となっている．しかし，これらの多くは C7 コードにおける B $\flat$  音や D7 コードにおける F $\sharp$  音など音楽的に適切なものであった．実際，正解における Crit. 2 はさらに高い値となっており，このデータセットには一定のノンダイアトニックコードを含み，これを適切に学習したことによる結果であるとも解釈できる．Crit. 3 (最後のコードが C である割合) は，実験 1 と同様，順方向モデルに比べて逆方向および双方向モデルでは十分に高い値となった．

次に，データセット中の和声（便宜上の「正解」との一致

表 1: 実験 1 の結果 . (1)~(3) の LSTM モデルは 10 回異なる初期値から学習を行い, その平均値を記した . (4) は [Suzuki 14] の結果である . ただし, (4) についても集計は我々がやり直したため, [Suzuki 14] に記載されている値とは微妙に異なる .

	Crit. 1 不協和音	Crit. 2 ND 含有	Crit. 3 C 終了	Crit. 4 3 声重複	Crit. 5 Bass 跳躍	Crit. 6-1 Alt 音高	Crit. 6-2 Tenor 音高	Crit. 6-3 Bass 音高
(1) 順方向	17.5 %	0.3 %	76.9 %	5.6 %	11.0%	3.4	4.1	4.3
(2) 逆方向	17.9 %	1.0 %	100.0 %	5.9 %	6.3 %	4.1	4.7	3.4
(3) 双方向	18.4 %	0.6 %	98.4 %	7.5 %	5.6%	3.5	4.2	3.3
(4) BN	8.8 %	1.5 %	100.0 %	24.0 %	24.8 %	5.2	4.9	4.8

表 2: 実験 2 の結果 . 「(4) 正解」は [日本基督教団讃美歌委員会 83] において実際に記載されている和声を表す .

	Crit. 1 不協和音	Crit. 2 ND 含有	Crit. 3 C 終了	Crit. 4 3 声重複	Crit. 5 Bass 跳躍	Crit. 6-1 Alt 音高	Crit. 6-2 Tenor 音高	Crit. 6-3 Bass 音高
(1) 順方向	11.5 %	3.0 %	79.4 %	5.7 %	8.4 %	7.1	7.4	8.8
(2) 逆方向	11.5 %	3.3 %	93.9 %	6.3 %	7.9 %	7.0	7.6	8.7
(3) 双方向	10.5 %	2.9 %	93.9 %	6.0 %	7.6 %	6.8	7.3	8.6
(4) 正解	8.6 %	4.2 %	91.0 %	5.0 %	5.2 %	6.8	6.9	8.7

表 3: 実験 2 における「正解」との一致率

	Alto	Tenor	Bass
(1) 順方向	45.2 %	44.2 %	36.6 %
(2) 逆方向	47.5 %	47.8 %	41.5 %
(3) 双方向	50.7 %	50.4 %	43.9 %

率 (表 3) について考察する . 表より, (1) 順方向のみのモデル, (2) 逆方向のみのモデル, (3) 双方向のモデルの順に「正解」との一致率が高くなっていることが分かる . このことから, 双方向の依存関係をモデル化することが有効であることが示唆される . ただし, 前述の通り, ここでの「正解」は唯一の正解ではなく, 「正解」と一致しなかった和声が音楽的に必ずしも適切ではなかったとは限らないことに注意すべきである .

## 5. まとめ

本研究では複数の LSTM モデルを構築し, それらの和声生成結果を, 従来の BN モデル [Suzuki 14] と比較しながら調査した . その結果, 今回用いたデータおよび評価指標では, LSTM は [Suzuki 14] の BN モデルよりも不協和音が多く出力されること, LSTM のモデル化では双方向の依存関係を考慮することが有効であることが分かった . 主観評価をも含めたさらなる評価やモデルの獲得した内部表現の解析が今後の課題である .

## 謝辞

本研究は, 日本学術振興会 科学研究費補助事業 (16H05878, 16K16180, 16H01744, 26280089, 26240025, 16KT0136) の助成を受けた .

## 参考文献

[Ebcioğlu 90] Ebcioğlu, K.: An Expert System for Harmonizing Chorales in the Style of J. S. Bach, *Journal of Logic Programming*, Vol. 8, No. 1-2, pp. 145-185 (1990).

[Hadjeres 16] Hadjeres, G. and Pachet, F.: DeepBach: a Steerable Model for Bach Chorales Generation, *arXiv: 1612.01010 [cs.AI]*, pp. 1-20 (2016).

[Hochreiter 97] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, Vol. 9, No. 8, pp. 1735-1780 (1997).

[Phon-amnuaisuk 99] Phon-amnuaisuk, S. and Wiggins, G. A.: The Four-Part Harmonisation Problem: A comparison between Genetic Algorithms and a Rule-Based System, *Proceedings of the AISB '99 Symposium on Musical Creativity*, pp. 28-34 (1999).

[Suzuki 14] Suzuki, S. and Kitahara, T.: Four-part Harmonization Using Bayesian Networks: Pros and Cons of Introducing Chord Nodes, *Journal of New Music Research*, Vol. 43, No. 3, pp. 331-353 (2014).

[池内 67] 池内, 島岡.: 和声-理論と実習 別巻 課題の実践, 音楽之友社 (1967).

[日本基督教団讃美歌委員会 83] 日本基督教団讃美歌委員会 (編): 三編合本 (讃美歌・讃美歌第二編・ともいうたおう), 日本キリスト教団出版局 (1983).