2T-01

条件付き自己符号化器を用いた旋律の協和度を制御できる旋律生成方式の検討

城田 晃希[†1,a)]　竹川 佳成[†2,b)]　平田 圭二[†2,c)]

1. はじめに

本稿は，旋律生成とそれに用いられる手法の内，モーフィングに着目する．モーフィングとは異なる複数の楽曲の，中間的な出力を連続して生成する手法である．モーフィングを行うことにより，データ入力者はある程度旋律的概形をイメージでき，想定する出力の期待をすることが出来る [1]．深層学習によるモーフィングを行う手法として変分自己符号化器 (Variational Autoencoder) があり，VAE を用いることにより，学習させたモデルの特徴量を操作することにより様々な音楽的特徴を反映させた生成が可能であることが示されている [2][3]．Adam らは小節間の音楽的構造を考慮するすることにより，最大 32 小節のモーフィングを行わせた [2]．村田らは，旋律とコードのモーフィングを同時に行わせ，異なるコードを持つ旋律間のコードのモーフィングをはかった [3]．ただし，それらは音楽的なドメイン知識の学習が明示的でないため，入力者の音楽的意図が反映されるとは限らない．補助ソフトウェアとして生成モデルを用いる場合，この問題点は作曲者が目的とする旋律の生成と乖離を生む．具体的には、作曲者は和声音と非和声音の登場を暗黙的にコントロールしながら作曲を行うが、従来のモデルではそのフローが反映されていないといえる。

音楽的なドメイン知識の一つとして，旋律とそのコードとの関係性が存在する．その関係性は聴取者にとって，旋律の感じ方に作用する．本稿では条件付き変分自己符号化器 (Conditional Variational Autoencoder) を用いて，モーフィング出力を行う手法についての検証を行う [4]．具体的には，CVAE に用いる条件付け入力に，旋律から評価されるコードに対しての評価値を与える．以後，その評価値のことを協和性と呼称する．これにより，モデルの旋律とコードの関係性の制御をはかる．データ入力者はその関係性を制御することにより，旋律生成に自身の音楽的意図の一部反映を行うことができる．提案手法では学習時には CVAE に対し協和性と旋律を対応付け，旋律の復元を行わせるよう学習させ，生成時にデータ入力者が協和性を操作し出力のコントロールを行う．

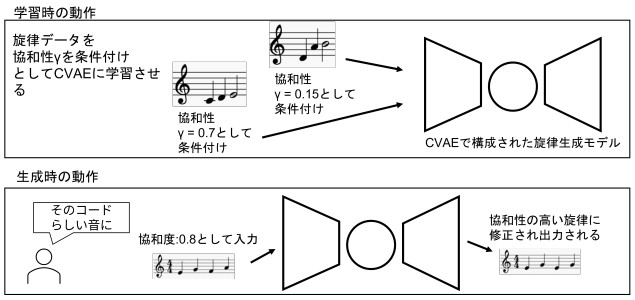


図 1: 提案手法全体のフロー

2. 協和性の評価手法

モデルの旋律とコードの関係性の評価として，そのコードトーンに着目した手法をとる．評価を行うアルゴリズムのフィルターはコードにより決定され，その際のコードトーンに対して重みが高くなるように設定する (図 2)．協和性は γ とし，$(-1 \leq \gamma \leq 1)$ で定義される．それにより，協和性の操作による生成結果に旋律がコードに則している度合などの，直観的な音楽的ドメイン知識の反映をはかる．

3. 提案する CVAE モデル

提案モデルではエンコーダの層を時系列データの前方

†1 現在，公立はこだて未来大学大学院
Presently with Graduate School of Future University Hakodate
†2 現在，公立はこだて未来大学
Presently with Future University Hakodate
a) g2120019@fun.ac.jp
b) yoshi@fun.ac.jp
c) hirata@fun.ac.jp

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.

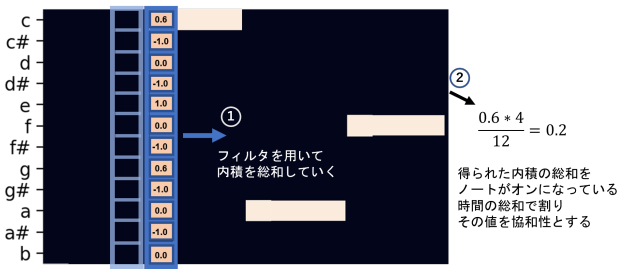


図 2: C メジャーの際の協和性の計算方法

向，後方向の両方向を考慮するため Bidirectional LSTM (BLSTM) にし，デコーダを一つ前あるいは，t 次元前の出力を考慮するため再帰的構造を持つ LSTM にした (図 3)．また，実装は Roberts らの実装を参考にした [2]．

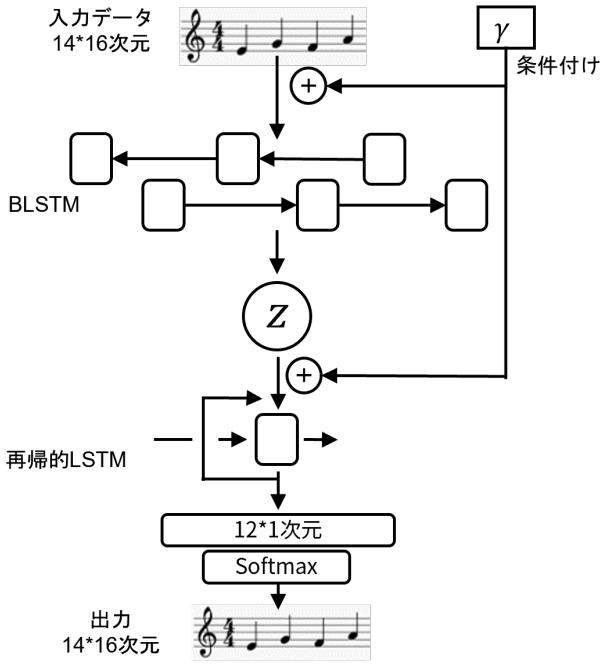
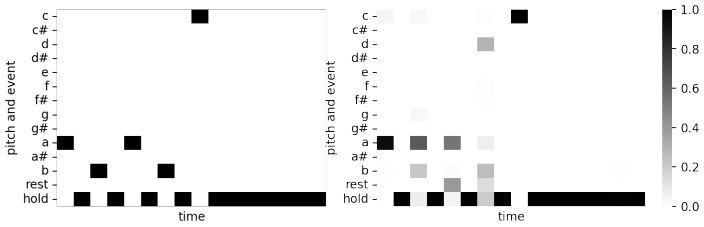


図 3: モデルの概要

4. 生成実験

データベースは Western music lead sheets を実験に用いた [5]．簡単化のため三連符を含む旋律は除外し，16 分音符でクォンタイズを行った．1 オクターブ分の音高情報と休符情報，音の持続を表すホールドのフラグの 14 次元，時系列方向の 16 次元で 14*16 次元のワンホットベクトルにデータの前処理を行った．トレーニングデータの総数は 33609 個となった．LSTM の隠れ状態を 256 次元，潜在空間を 128 次元とした．最適化手法は Adam (lr = 0.0005) としその他パラメータは Kingma らの手法に則る [6]．誤差関数にはクロスエントロピーを用いた。バッチサイズは 64 とし，5000 epoch の学習を行った．

結果としてテストデータを入力として，条件付けである協和度を変化させた際の出力を例示する．入力データのコードは F_M である．協和度を 0.3 に設定を行うと，F_M のコードトーンである a に出力が遷移するような観察が得られた (図 4a)(図 4b)．



(a) F_M の入力データ例　**(b)** 協和度が 0.3 の出力

図 4: 入出力の例

5. おわりに

モデルの動作として，協和度に合わせた出力が例として確認をすることができた．但し全体の出力の妥当性の評価，聴取実験を交えた調査は必要である．また学習データの数から，各協和度においてのサンプル数の不足が見受けられた．今後の方針として学習データの拡張及び増加を行う必要がある．

参考文献

[1] Briot, J. P., Hadjeres, G., Pachet, F. D. (2017). Deep learning techniques for music generation–a survey. arXiv preprint arXiv:1709.01620.

[2] Roberts, A., Engel, J., Raffel, C., Hawthorne, C., Eck, D. (2018, July). A hierarchical latent vector model for learning long-term structure in music. In International conference on machine learning (pp. 4364-4373). PMLR.

[3] 村田叡, 坂東宜昭, 糸山克寿, 吉井和佳. (2018). VAE を用いたメロディのモーフィング. 第 80 回全国大会講演論文集, 2018(1), 175-176.

[4] Kingma, D. P., Mohamed, S., Rezende, D. J., Welling, M. (2014). Semi-supervised learning with deep generative models. In Advances in neural information processing systems (pp. 3581-3589).

[5] Lim, H., Rhyu, S., & Lee, K. (2017). Chord generation from symbolic melody using BLSTM networks. arXiv preprint arXiv:1712.01011.

[6] Kingma, D. P., Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.