The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

深層学習によるユーザ評価モデルを導入した遺伝的プログラミングによる音楽自動生成手法の提案

Automatic Music Composition System Based on Genetic Programming and Surrogate Model with Deep Learning

山本 周典 *1 Hironori Yamamoto　長谷川 拓 *2 Taku Hasegawa　森 直樹 *2 Naoki Mori　松本 啓之亮 *2 Keinosuke Matsumoto

*1大阪府立大学 Osaka Prefecture University　*2大阪府立大学工学研究科 Graduate School of Engineering, Osaka Prefecture University

Automatic music composition is one of the most difficult and attractive challenges in the artificial intelligence (AI) field. In order to tackle this challenge, an approach using interactive evolutionary computation (IEC) is drawing attention because IEC takes human emotions into consideration. The major problem in using IEC is that the number of evaluations from one user is limited due to user fatigue. To tackle this problem, a surrogate model is often introduced into IEC. An approach based on deep learning (DL) is also common in this field because of many quantitative futures. However, the approach hardly considers human emotions.

In this study, we proposed the automatic music composition system based on IEC and a surrogate model called evaluation model. The model is constructed with a DL model, thus our system can compose music reflected human emotions quantitatively. The experiments are carried out to show the effectiveness of the proposed method.

1. はじめに

近年，計算機による音楽の自動生成に関する研究が積極的になされており，新しい文化や産業の創造に繋がるとして多方面から強く期待されている．しかしながら，作品の評価は個人の嗜好や感性に大きく依存するため，これを定量的に評価することは非常に困難とされている．このような問題に対して，人間の評価系そのものを評価関数として最適化システムに導入した対話型進化型計算 (Interactive Evolutionary Computation: IEC)[1] が注目をされている．

IEC は定量化が難しい人の感性を評価できるという利点がある一方で，ユーザ負荷の観点から解を評価できる回数に大きな制約があり，特に音楽の自動生成の分野においては各々の解を逐次的に評価する必要があることから，評価回数の制約が顕在化しやすいという問題がある．また，深層学習 [2] を用いた音楽の自動生成の手法も提案されている．特に，Variational Autoencoder (VAE)[3] をはじめとする生成モデルを活用した方法は，入力データが潜在的に内包している意味を反映し写像した潜在空間を構築可能な点からも幅広い応用が期待されている．しかしながら，深層学習を用いた手法は主に事前に学習した分布の再現になるため，生成物に対してユーザの感性を直接的に反映させることは困難であるという課題も存在する．

本研究では以上の点を背景として，適応度景観を学習し近似的に評価関数を推定する surrogate model[4] を応用した IEC による対話型音楽自動生成システムを提案する．本システムでは音の高さおよび長さの概念を反映した木構造によって楽曲を表現した．また，Genetic Programming (GP) の拡張手法である Genetic Programming with Multi-Layered Population Structure (MLPS-GP)[5] を音楽の探索アルゴリズムとして用いた．更に，既存の楽曲データから音楽的特徴を抽出，学習し，個体に対する近似評価をする評価モデルを surrogate model として導入した．評価モデルの学習に当たっては VAE を用い，楽曲学習時に生成される潜在空間に着目しつつ IEC と組み合わせることで，定量的かつユーザの嗜好を反映した音楽の自動生成を提案する．

連絡先: 山本 周典，大阪府立大学，h.yamamoto@ss.cs.osakafu-u.ac.jp

2. 音楽データの分散表現の獲得

本研究では Variational Recurrent Autoencoder (VRAE)[6] を使用することで音楽の潜在変数を獲得し，これを音楽の分散表現として扱う．VRAE は Variational Autoencoder (VAE)[3] と呼ばれる生成モデルの拡張手法で，エンコード部およびデコード部で Recurrent Neural Network を使用しているため音楽のような時系列データに適していると考えられる．VAE および VRAE の最適化項目としては，潜在変数の確率分布の形状に関わる項（latent loss）および潜在変数を事前分布として元データの復元に関わる項（reconstruction loss）がある．

3. 音楽の自動生成システムの概要

本章では進化型計算で音楽を扱うための個体表現の定義，surrogate model の利用方法，提案システムの流れおよび入力曲の定義について述べる．

3.1 曲における個体表現

本研究では Genetic Programming (GP) を適用するために曲を木構造によって表現する．この木構造によって曲の音高および音価が表現可能である．図 1 に，GP における曲の個体表現と楽譜の対応関係を示す．終端ノードを音高とし，非終端ノードを子ノードへの音高の倍率および子ノードへの分岐数とした．i を子ノードの音価への倍率，j を子ノードへの分岐数として，非終端ノードを S_j^i で表す．i，j に関しては，$i \in \{1.0, 1.5, 2.0\}$，$j \in \{1, 2, 3, 4\}$ と定めた．また本研究では，根ノードの音価を 16 分音符で固定する．なお，$S_j^{1.5}$ に関しては付点音価を表現することを目的としているため，ある終端ノードの音価を計算するに際して，その親ノードが $S_j^{1.5}$ の場合においてのみ音価を 1.5 倍とし，それ以外の $S_j^{1.5}$ に関しては音価を 1 倍とするように設定した．木構造表現から楽譜に変換する際は，終端ノードに関して深さ優先探索の先行順で辿ることで変換がなされる．

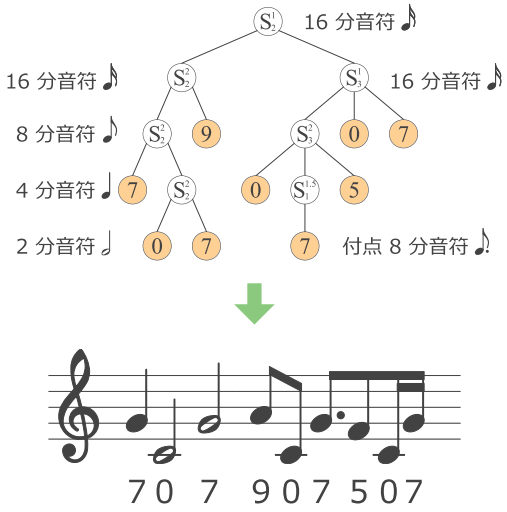

図 1: 木構造と楽譜の対応関係の例

3.2 surrogate model の導入

IEC を利用する際はユーザ負荷を考慮し，できるだけ探索に必要な解の評価回数を減らす必要がある．そこで本研究では，ユーザの代わりに適応度景観を学習し近似的に解を評価する surrogate model を導入した．これによってユーザに直接的な負荷をかけることなく十分な探索を実現する事ができる．さらに，ユーザの評価は，直接的に解の優劣を示す適応度のみではなく，主にモデルを作成し更新するために利用されるため，ユーザ負荷を小さく抑える一方で精度の高い近似評価を実現している．モデルと実際の評価関数をどのように利用し使い分けるかを示す evolution control に関しては，一定世代ごとにユーザの評価をモデルに反映する generation-based evolution control[7][8] として捉えることができる．一般的な generation-based evolution control では一定世代ごとに個体群すべての個体を評価するが，本手法においてはモデルにおける “適応度” が高い個体，いわゆるエリート個体をユーザに提示し，評価させるという枠組みになっている．以下 surrogate model として構築したモデルを評価モデルと呼ぶ．

3.3 提案システムの流れ

本システムはユーザの嗜好に合った曲をユーザの評価に基づいて進化的に獲得することを目的としたシステムである．なお，評価モデルが楽曲の特徴量を学習するのに使用した曲を入力曲，進化型計算による探索中の個体を探索曲，探索終了後のエリート個体を進化曲と呼ぶ．

1. 入力曲に基づいて評価モデルが構築される．
2. 評価モデルによって探索曲を近似評価し，この評価に従い MLPS-GP が進化曲を生成する．
3. ユーザが進化曲を実評価する．
4. 評価モデルがユーザの実評価に基づいてユーザの嗜好を推定し，さらに IEC により評価モデルを更新し，初期個体群に進化曲を反映する．
5. 2. に戻る．あるいは，ユーザが満足する曲が生成できた場合は終了する．

3.4 入力曲の定義

本研究では，入力曲を一般入力曲および選好入力曲の 2 種類の入力曲に大別し，評価モデルに学習させる．一般入力曲とは人が普段から耳にし聴き心地の良いと感じるような普遍的な曲のことで，これらを大量に学習することによって汎用的な楽譜情報の獲得が期待される．一方，選好入力曲とは，進化曲の特徴付けのために一般入力曲の中から抽出された曲のことで，ユーザの嗜好を獲得することが期待される．本システムでは，一般入力曲の学習によって獲得した普遍的な音楽の特徴およびユーザに選択させた選好入力曲の特徴を組み合わせることで，ユーザの嗜好に合った進化曲の獲得を目的としている．

4. 評価モデル

本章では，提案システムにおいて重要な評価モデルによる曲の定量的評価手法について述べる．

4.1 音楽データの規格化

VRAE を評価モデルとして用いるために，VRAE で学習する音楽データを規格化した．この規格化の 1 つ目の理由に MLPS-GP による効率的な探索の実現が挙げられる．MLPS-GP では局所探索の際にランダムなノード選択を含むため，すべての音楽データが含む音について探索をするには莫大な実行時間を要する．特に，本システムではユーザの実評価を定期的に必要とするため，近似評価に必要な実行時間はボトルネックとなる．よって，学習データを規格化した上で MLPS-GP の探索範囲を限定することで，短時間でより良い解の探索が期待できる．

また，2 つ目の理由として VRAE の汎用的な学習の実現が挙げられる．生の音楽データの場合，音高毎に存在する音楽データに偏りがあることが予想され，特に音楽データが少ない音高の範囲では十分な学習が成されない可能性がある．そこで，規格化により出現する音の種類を限定することで，より汎用的で時系列的な情報に基づいた学習をすることが可能になると考えられる．

規格化の方法は以下のとおりである．

- 音高に対して，階名を保ったまま 3 オクターヴ内に収まるように変換．
- 音価に対して，曲全体の値の比を保ったまま 16 分音符～全音符に収まるものに変換．
- 変換の過程で 1. および 2. で示した範囲に入らなかった音については “unk” タグを付与．

規格化を経て “unk” タグの割合は全体の 9.398e-5 程度になった．このことから，ほとんどの音楽データは全体を通して音が 3 オクターヴかつ 16 分音符～全音符内に存在しており，上記の規格化を通してもほとんど元のデータ構造を保持できていることがわかる．また，MLPS-GP の音の探索範囲を限定することについても合理的であると考えられる．

4.2 個体評価値

次に評価モデルにおける個体評価値を定める．まず，VRAE に対して一般入力曲を学習させる．次に，選好入力曲をユーザが設定し，これらを VRAE のエンコード部によって潜在空間に写像することで，潜在変数 $\boldsymbol{z}_t$ を得る．また，探索曲 x についても同様に潜在ベクトル $\boldsymbol{z}_x$ を算出し，得られた潜在ベクトル $\boldsymbol{z}_x$ をデコードすることで $\hat{x}$ を得る．以上の変数を用いて，

表 1: MLPS-GP のパラメータ

適応度評価回数 N_{e}	20000
初期化における減衰率 r_{dump}	0.8
初期の深さ制限 D_{init}	3

探索曲 x に対する個体評価値 $f(x)$ を次のように定義した．

$$f(x) \quad = \quad (\frac{1}{T}\sum_{t}|\boldsymbol{z}_t - \boldsymbol{z}_x| + \alpha|g(x,\hat{x}) - \beta|)^{-1} \qquad (1)$$

$g(x,\hat{x})$ は x の reconstruction loss に相当し，本研究では関数 g を softmax cross entropy とした．また，α および β はパラメータ，T は選好入力曲の個数である．

（1）式において $\frac{1}{T}\sum_t|\boldsymbol{z}_t - \boldsymbol{z}_x|$ はユーザが設定した選好入力曲の集合に対して潜在空間において探索曲 x のユークリッド距離が近くなることを目的とする項である．これにより，ユーザの嗜好を進化曲に取り入れることが期待される．また，(1) 式の $|g(x,\hat{x}) - \beta|$ は探索曲 x の reconstruction loss がパラメータ β に近くなることを目的とする項である．この β は学習済みの VRAE に対してテストデータを入力した際に得られる reconstruction loss によって決定される値である．このようにパラメータ β を定める理由として VRAE によって形成される潜在空間には曲の特徴量をうまく写像している部分とそうでない部分があり，単純に選好入力曲とのユークリッド距離を近くするだけでは楽曲としての特徴を十分に反映した進化曲が得られないと予想されるためである．よって，テストデータの reconstruction loss に近づくような個体を探索することで，VRAE が上手く写像できる空間の中での探索が可能となり，楽曲の一般的な特徴を反映した進化曲を得ることが期待できる．このようにテストデータの reconstruction loss に基づいて，教師無しで入力と学習時のデータセットとの差異を測る研究 [9][10] も複数報告されており，本研究でも有用であると考えられる．

4.3 適応度

ユーザの嗜好を反映する上で，生成する曲の長さは重要な要素である．故に，本システムでは，ユーザが必要とする曲の曲長を t_{d}，進化曲の曲長を t_x とし，MLPS-GP における適応度 $F(x)$ は (1) 式 を用いて，次式のように与えられる．

$$F(x) = \frac{t_{\mathrm{d}}}{t_{\mathrm{d}} + |t_{\mathrm{d}} - t_x|} f(x) \qquad (2)$$

5. 実験

ここでは音楽的特徴に基づく評価モデルによる探索曲の進化について解析をする．具体的には生成された進化曲の個体評価値および個体評価値を構成する各類似度について示すことで，評価モデルにおける個体評価値の妥当性，および進化曲が適切に進化しているかどうかを調べる．

5.1 実験条件

選好入力曲を “SCHOENE AUGEN SCHOENE STRAHLEN” の 1 曲に設定し，（2）式の適応度に基づき音楽を自動生成した．

表 1 に本実験で用いた MLPS-GP のパラメータを示し，表 2 に今回学習した VRAE のパラメータを示す．latent loss についてはその係数を 0 から 0.2 の範囲において指数関数的に上昇させた．その際の更新率は 0.9999 に設定した．また，VRAE で用いる最適化アルゴリズム Adam のパラメータ alpha については更新率を 0.999 として 1e-3 から 1e-6 の範囲で指数関数的に減少させた．また，表 3 に個体評価値（1）式のパラメータを示す．β は学習後の VAE に対してテストデータを入力した際の reconstruction loss の平均値とした．以上の設定に基づいて乱数シードを変えた進化曲およびランダム曲を 140 曲用意した．ただし，ランダム曲とは 3.1 節で示した曲の個体表現で表現可能な個体をランダムで生成した曲のことで，進化のベースラインとして用いた．

表 2: VRAE のパラメータ

batch size	512
epoch	5000
embed units	256
hidden state	512
latent units	128
optimizer	Adam
alpha (Adam)	[1e-6, 1e-3]
beta1 (Adam)	0.05
beta2 (Adam)	0.001
dropout	0.8
loss fuction	softmax cross entropy

表 3: 評価モデルパラメータ

個体評価値 α	0，1
個体評価値 β	3.075
曲長 t_{d}	30 秒

5.2 実験結果

表 4 にランダム曲および進化曲間で Welch の t 検定にかけた際の p 値を示す．また，図 3 に得られた進化曲の一例を示す．

以下，表 4 の結果より考察する．まず，進化曲（$\alpha = 1$）において，ランダム曲と比較するとどちらの項についても $p < 0.01$ より有意水準 1% において有意差が見られ，進化が個体評価値にもとづいて成功していることがわかる．次に，進化曲（$\alpha = 0$）においても同様に $p < 0.01$ より有意水準 1% において有意差が見られた．進化曲（$\alpha = 0$）の場合，(1) 式における $\alpha = 0$ であるため reconstruction loss の項の最適化は本来行われないはずである．しかしながら，ベースラインであるランダム曲とはこの項について有意差が生じている．この理由として，VRAE によって形成されている潜在空間が適切であるため，選好入力曲との距離を小さくすることによってランダムな音列から曲らしい音列へと近づいていっていると考えられる．一方で，進化曲（$\alpha = 1$）および進化曲（$\alpha = 0$) において，すべての項目について $p < 0.01$ より有意水準 1% で有意差が見られた．この結果より，個体評価値 1 のパラメータ α の選定が非常に重要であることが分かる．

また，図 2 に，ある進化の過程における選好入力曲，テストデータ，および最大個体評価値を更新した個体（以下優良個体）を潜在空間上に写像し t-SNE で可視化したものを示す．図 2 において選好入力曲は水色のクロス，テストデータは灰色のサークル，優良個体は色の付いたサークルによって表現されており，優良個体については青から赤にかけて個体評価値が高くなっていくように色が付けられている．この結果から，MLPS-GP によって潜在空間上で選好入力曲の周辺を中心に探索がなされていることがわかる．一方で，図 2 に示した結果より総評価個体数に対する優良個体数が非常に少ないことも示しており， MLPS-GP による探索が不十分であることも

表 4: ランダム曲および進化曲間で検定をした時の p 値

	選好入力曲との距離 $\frac{1}{T}\sum_t \lvert \boldsymbol{z}_t - \boldsymbol{z}_x \rvert$	reconstruction loss $g(x, \hat{x})$
ランダム曲 - 進化曲（$\alpha = 1$）		
p 値	1.569e-212	7.720e-127
ランダム曲 - 進化曲（$\alpha = 0$）		
p 値	3.409e-179	2.496e-23
進化曲（$\alpha = 1$）- 進化曲（$\alpha = 0$）		
p 値	5.654e-60	4.471e-88

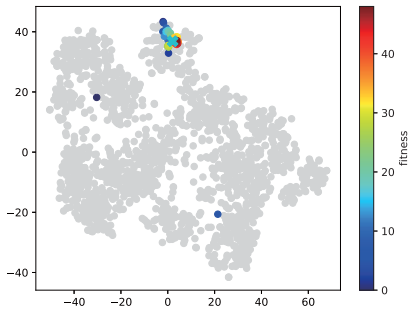

図 2: 優良個体の潜在空間上での変遷

示している．このことから，より効率的な探索手法として深層学習に基づく手法の探索アルゴリズムおよび探索オペレータの構築が必要であると考えられる．また，これらの探索性能には VAE の潜在空間の精度も重要な要因となっているため，VAE 側のより適切な構造およびパラメータの調整も重要な課題である．

5.3 アンケート実験

本システムを用いて得られた進化曲を用意し，アンケートを用いてその評価をした．被験者は，大学生および大学院生計 14 名 (男性 : 12 名，女性 : 2 名) で，選好入力曲およびそれぞれの選好入力曲から進化した進化曲を 3 曲ずつ用意し，被験者に選好入力曲に対応していると思われる進化曲を選択させた．この結果，すべての被験者が適切な選好入力曲および進化曲の組み合わせを選択した．以上の結果より本システムが選好入力曲の情報を反映しつつ進化曲を生成していることがわかった．

6. まとめと今後の課題

本論文では対話型進化型計算および深層学習を用いた音楽の自動生成のシステムを提案した．まず，システム概要を示し，次に評価モデルの構築方法および楽曲の近似評価について示した．実験では優良個体および進化曲に基づき進化の解析をした．また，アンケート実験によって本システムの有用性を示した．今後の課題としては，深層学習に基づく手法の探索アルゴリズムおよび探索オペレータの構築，コードの自動付与が挙げられる．

7. 謝辞

なお，本研究は一部，日本学術振興会科学研究補助金基盤研究 (C) (課題番号 26330282) および特別研究員奨励費 (課題番号 16J10941) の補助を得て行われたものである．

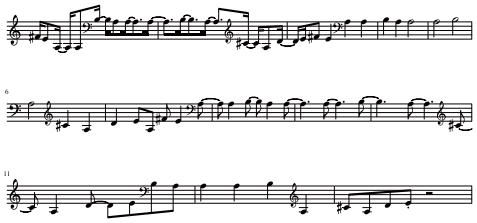

図 3: 本システムによって得られた進化曲

参考文献

[1] Hideyuki Takagi. Interactive evolutionary computation: Cooperation of computational intelligence and human kansei. In *Proceedings of the 5th International Conference on Soft Computing and Information/Intelligent Systems*, pp. 41–50, 1998.

[2] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, Vol. 61, pp. 85–117, 2015.

[3] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[4] Yaochu Jin. Surrogate-assisted evolutionary computation: Recent advances and future challenges. *Swarm and Evolutionary Computation*, Vol. 1, No. 2, pp. 61–70, 2011.

[5] Taku Hasegawa, Naoki Mori, and Keinosuke Matsumoto. Genetic programming with multi-layered population structure. *Proceedings of the 2017 Annual Conference on Genetic and Evolutionary Computation*, 2017.

[6] Otto Fabius and Joost R van Amersfoort. Variational recurrent auto-encoders. *arXiv preprint arXiv:1412.6581*, 2014.

[7] Yaochu Jin. A comprehensive survey of fitness approximation in evolutionary computation. *Soft computing*, Vol. 9, No. 1, pp. 3–12, 2005.

[8] JE Dennis and Virginia Torczon. Managing approximation models in optimization. *Multidisciplinary design optimization: State-of-the-art*, pp. 330–347, 1997.

[9] Erik Marchi, Fabio Vesperini, Florian Eyben, Stefano Squartini, and Björn Schuller. A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional lstm neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pp. 1996–2000. IEEE, 2015.

[10] Loïc Bontemps, James McDermott, Nhien-An Le-Khac, et al. Collective anomaly detection based on long short-term memory recurrent neural networks. In *International Conference on Future Data and Security Engineering*, pp. 141–152. Springer, 2016.