

対話型進化型計算を用いた音楽自動生成システムの提案

Automatic Music Composition System with Interactive Evolutionary Computation

土江 海輝 *¹ 長谷川 拓 *² 上野 未貴 *² 森 直樹 *² 松本 啓之亮 *²
 Kaiki Tsuchie Taku Hasegawa Miki Ueno Naoki Mori Keinosuke Matsumoto

*¹大阪府立大学 工学部 知能情報工学科

Osaka Prefecture University Department of Computer Science and Intelligent Systems

*²大阪府立大学 工学研究科 電気・情報系 知能情報工学分野

Osaka Prefecture University Electrical Engineering and Information Science

Recently, automatic music composition is one of the most difficult and attractive issues in Kansei engineering. However, it is difficult for computer to compose automatic music because computer cannot evaluate the quality of music by numerically. To introduce the human evaluation results into computers, IEC (Interactive Evolutionary Computation) has been proposed and reported lots of applications for design fields. However, in most applications it is necessary for users to evaluate individuals many times. To prevent users from evaluating excessively, surrogate model is often introduced into IEC which expresses evaluation function artificially by learning fitness landscape of human evaluation models. In this paper, we proposed a novel interface of IEC for automatic music composition with surrogate model. The computational experiments are carried out in order 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多くなされており、その中でも進化型計算 (Evolutionary Computation: EC) を用いたメロディの自動生成が注目されている。しかしながら、ユーザの嗜好を反映した音楽を得るための評価はユーザの感性に大きく依存しているため、これを定量的に数値化し実装することは、非常に困難とされている。そこで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) という手法が提案されており、音楽の自動生成に関しても IEC がさまざまなシステムに用いられている [1]。

IEC を用いる場合は、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存在するという問題を考慮する必要がある。特に、音楽の自動生成の分野においては、絵や文字のように並列に解を提示して評価することが困難であり、各々の解を逐次的に評価せねばならず、評価回数の制約が顕在化しやすい。そのため、評価回数に制限がある状況で探索することが求められ、その結果として、本来進化型計算が持つ探索能力を十分に発揮できないことが多い。

上記の問題を解決するために、適応度景観を学習し評価関数を擬似的に表現する surrogate model [2] を導入した進化型計算が数多く提案されている。surrogate model を用いた進化型計算は有用性が数多く報告されている一方で、問題に対してどのようなモデルと枠組みを用いるかによって大きく探索性能が変わってしまうことが指摘されている [3]。そのため surrogate model を用いた進化型計算を利用する際には、対象とする問題に応じてよりよいモデルを選び、枠組みを作ることが重要となる。

以上のことを踏まえ、本研究では上記の surrogate model を応用した IEC による音楽自動生成システムを提案する。

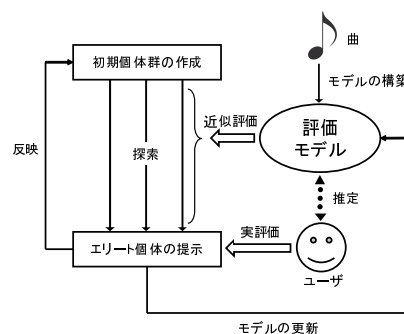


図 1: 提案手法の構成

2. 提案手法

本章では、はじめに提案システムの概要を述べ、進化型計算による数値表現について示した後、surrogate model によるユーザ推定における手法について述べ、最後に対話システムとユーザの嗜好推定について説明する。

2.1 提案システムの概要

提案システムは、ユーザがシステムと対話することで好ましい楽曲を進化的に得ることを目的としている。また、音楽的知識がまったくないようなユーザも対象にしており、より幅広いユーザが使用可能となるシステムの実装も目標としている。今回は、まずシステムの構築の第一段階としてメロディ (主旋律) のみの自動生成を考える。図 1 に、提案手法の構成を示す。また図 2 に、本研究で実装した提案システムのインタフェースを示す。本手法の大きな特徴としては、適応度景観を学習し評価関数を擬似的に表現する surrogate model を導入し、ユーザの嗜好に従って対話的にモデルを更新していき、そのモデルに曲の評価をさせるという点である。また、任意の入力曲に対応しており、少ない楽曲情報でもメロディの進化を可能としている点で、ユーザにとって利用しやすいシステムとなっている。

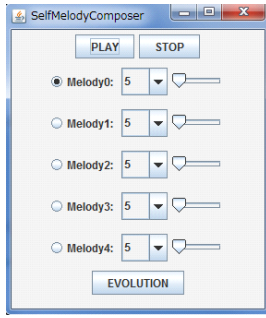


図 2: システムの外観

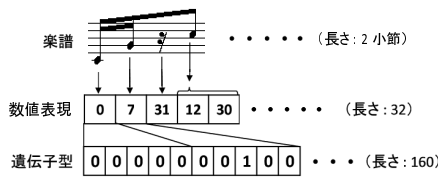


図 3: 楽譜と数値表現と遺伝子型の対応関係の例

2.2 進化型計算による数値表現

本システム中で用いられる進化型計算の個体は 2 小節分のメロディを表している。数値表現としては 1 つのメロディを 16 分の 1 の音価ずつに等分し、等分された各要素を 1 つの数字で表わした。具体的には 0 ~ 29 をある高さのドから始まる 2 オクターブ (24 音) + その上の 6 音とし、30 を音の伸ばし、31 を休符とした。以下、この音の伸ばしを duration と呼ぶ。さらに遺伝子型としてはこれらの数値列を 5 bit のグレイコードを用いてコード化している。図 3 に、楽譜と数値表現と遺伝子型の対応関係の例を示す。

2.3 surrogate model によるユーザ推定

対話型進化型計算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。そこで本手法では、ユーザの評価を模したモデルを作り、このモデルによる近似的な評価値に基づいて解を進化させることにより、ユーザに直接的な負荷をかけずに十分な探索を実現する。また、ユーザの評価は解の優劣を表す適応度として直接使用されるわけではなく、主にモデルを作成し、更新するために利用される。その利用法に関しては??で詳細に述べる。

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は、surrogate model を用いた進化型計算の分野においては evolution control として考えられており、基本的にモデルを用いて進化し、一定世代ごとにユーザの評価をモデルに反映するこの枠組みは、generation-based evolution control [2] としてとらえることができる。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。さらに複数の個体群を並列化して進化させ、多様な個体のユーザへの提示を実現している点においても、一般的な generation-based の手法とは異なる。

本研究で surrogate model として構築したモデルを、以下では評価モデルと呼ぶ。

2.3.1 評価モデルの構築

評価モデルを構築するにあたってある特定の曲、あるいはある特定のメロディフレーズの特徴量を抽出することでモデル

を構築していく。本研究では、入力曲は既存の曲を数曲用い、それらの中から任意に選択させるという方法をとった。選択された曲が複数個であれば、それらの音楽的特徴が混在したメロディの生成が期待される。

楽曲の特徴量は多数存在するため、本研究では音の繋がりに着目して評価した。以下では、音の時系列的な繋がりを音列と呼ぶ。音を 1 オクターブの 12 種類の音名と定義し、入力曲のある音名から次のある音名への跳躍を 2 音および 3 音ずつ評価モデルに記憶させていく。音名間の跳躍には、絶対音として上への跳躍と下への跳躍が存在する。例えば、音の跳躍を 1 オクターブ以内に限定した場合、MIDI におけるノートナンバーで考えると「ド」→「ソ」なら「60」→「67」と「60」→「55」の 2 種類が存在する。本手法では、音の跳躍を 1 オクターブ以内に限定する。つまり現在の音は 1 オクターブで 12、休符を 1 つの音と捉えて 1 の合計で 13 通りの音、次の音は、上下 1 オクターブずつで 24、同音で 1、休符を 1 つの音と捉えて 1、合計で 26 通りの音となる。

特徴を抽出する曲 m_i に含まれる 2 音の音列それぞれの出現回数を $13 \times 26 = 338$ 次元の特徴ベクトル $t_{m_i}^2 = (t_{m_i,1}^2, t_{m_i,2}^2, \dots, t_{m_i,j}^2, \dots, t_{m_i,338}^2)$ 、3 音の音列それぞれの出現回数を $13 \times 26^2 = 8788$ 次元の特徴ベクトル $t_{m_i}^3 = (t_{m_i,1}^3, t_{m_i,2}^3, \dots, t_{m_i,k}^3, \dots, t_{m_i,8788}^3)$ で表す。特徴を抽出するすべての曲の数を n とすると、これらを用いて特徴ベクトル集合 S^2 、 S^3 は次のように表される。

$$S^2 = \{t_{m_0}^2, t_{m_1}^2, \dots, t_{m_i}^2, \dots, t_{m_{n-1}}^2\}$$

$$S^3 = \{t_{m_0}^3, t_{m_1}^3, \dots, t_{m_i}^3, \dots, t_{m_{n-1}}^3\}$$

また、評価モデルによる個体評価のために入力曲全体の長さに対する duration の割合 d を保存しておく。

さらに評価モデルは以上の S^2 、 S^3 、 d に加え、特徴を抽出した曲それぞれに対する重みベクトル $w = \{w_{m_0}, w_{m_1}, \dots, w_{m_i}, \dots, w_{m_{n-1}}\}$ を持つ。この曲の重みベクトルの決定方法については 2.3.3 にて述べる。

2.3.2 評価モデルによる個体評価

個体を評価モデルによって評価する際、音列、および duration によって点数付けがなされる。

音列による評価得点は特徴ベクトル集合から単純に出現回数を点数として与える評価方法では、同一のフレーズが頻繁に繰り返されるような曲を入力曲とした場合、ある特定の音列のみの評価得点が著しく大きくなってしまい、同じ音列が過剰に進化曲に反映されてしまう可能性が高くなる。これを防ぐために、既に出現している同一の音列が再度出現するごとに、与える評価得点を割引いて点数を与える。さらに 2 音の音列に比べ 3 音の音列の出現確率は低いため、3 音の音列に対しては s 倍の点数を与える。曲 x の音列による評価値 $f_s(x)$ は可調整パラメータ α と x の特徴ベクトル t_x^2 、 t_x^3 を用いて次のように与えられる。

$$f_s(x) = \sum_{i=0}^{n-1} w_i f_s^i(x) \quad (1)$$

$$f_s^i(x) = \sum_{j=0}^{338} \sum_{u=0}^{t_{m_i,j}^2} \alpha^{u-1} t_{xj}^2 + \sum_{k=0}^{8788} \sum_{v=0}^{t_{m_i,k}^3} \alpha^{v-1} s t_{xk}^3 \quad (2)$$

音列による評価においては、duration に関して一切評価得点は与えられない。その場合、エリート個体はすべて 16 分音

符および 16 分休符で構成されたメロディとなってしまう。これを防ぐために、曲 x の長さに対する duration の割合 d_x 、可調整パラメータ β を用いて、duration による評価値 $\sigma(x)$ を式 (3) で定義する。

$$\sigma(x) = 1 - \beta |d - d_x| \quad (3)$$

duration による評価値 $\sigma(x)$ は、 $0 \leq \sigma(x) \leq 1$ となり、 $\sigma(x)$ は音列による評価値に対する割引率として機能している。

以上、音列による評価値 $f_s(x)$ 、duration による評価値 $\sigma(x)$ を用いて、曲 x の評価モデルによる評価値 $f(x)$ は次のように与えられる。

$$f(x) = \sigma(x)f_s(x) \quad (4)$$

2.3.3 対話システムとユーザの嗜好推定

対話システムにおいて、評価モデルがユーザの嗜好を推定するためには、ユーザによるフィードバックが必要となる。本システムではユーザによって決定される 2 種類の評価値を定義する。

1 つはエリート個体に対する 5 段階の個体評価点 s_i ($i = 0, \dots, N_m$) である。 i は曲番号、 N_m は最大曲番号である。特徴を抽出した曲それぞれに対する重みベクトル w を本システムではユーザに決定させる。 s_i は、その w を決定付ける要素となっている。システム上での 5 段階評価は 1 ~ 5 で、評価値としては $0 \leq s_i \leq s_{\max}$ とした。個体評価値 s_i を用いて、 w の要素 w_i を式 (5) と定義する。また、ユーザによってエリート個体を評価し、その評価に基づいて新たに評価モデルを更新するまでの一連の過程を本論文ではステップと呼ぶ。

$$w_i = \frac{s_i}{s_{\max}(N_m + 1)} \quad (5)$$

$s_i = 0$ の場合、そのエリート個体の特徴量はまったく評価モデルに反映されない。

もう一方の評価値は、初期個体反映率 p_i である。これは、GA で初期個体群を生成する際に、エリート個体を初期個体として反映させる割合である。エリート個体を初期個体群に混入させることで、前のステップのエリート個体の遺伝子構造が残りがつくなり、次のステップでのエリート個体の劣化を防ぐことが期待される。本システムでは、 p_i はスライダによって制御される。スライダの値を a とし、可変長パラメータ γ を用いて、初期個体反映率 p_i を以下の式 (6) で定義する。

$$p_i = 10^{1-(\frac{1}{a})^\gamma} \quad (0 < a \leq 1) \quad (6)$$

$p_i = 0$ の場合、初期個体群はすべてランダムに生成され、 $p_i = 1$ の場合、初期個体群はすべてエリート個体となる。

これらの更新をステップを重ねるごとに適用することで、徐々に各エリート個体の音楽的特徴が評価モデルに反映されていく。

2.4 提案手法のアルゴリズム

以下に提案手法のアルゴリズムを示す。

1. 個体数 N_p 、最大世代数 N_g 、最大曲番号 N_m 、最大ステップ数 N_s を与える。
2. 曲番号 i に対応した任意の入力曲 m_i をエリート個体 e_i として与え、ステップ $T = 0$ とする。
3. ステップ T においてユーザにすべてのエリート個体を聴取させ、それぞれに個体評価点 s_i 、初期個体反映率 p_i を決定させ、曲番号 $i = 0$ とする。

表 1: GA のパラメータ

個体数	300
世代	500
遺伝子長	160
交叉	一点交叉
交叉率	1.0
突然変異率	$\frac{1}{2L}$ (L : 遺伝子長)
選択	トーナメント選択
トーナメントサイズ	2

4. 曲番号 i における e_i の特徴ベクトル集合と s_i に従って評価モデルを構築する。
5. 世代 $t = 0$ とし、初期世代 $P_i(0)$ において初期個体反映率 p_i で e_i を反映し、 $1 - p_i$ でランダムに生成する。
6. 世代 t の個体群 $P_i(t)$ 中の全個体からランダムに 2 個体ずつ選択し、これらを親として交叉を適用し 2 個体の子を生成する。これを繰り返すことで、 N_p 個の個体を生成する。この子個体に突然変異を施し、この N_p 個体の個体群を $P'_i(t)$ とする。
7. 評価モデルを用いて全個体の適応度を評価し、新たな e_i が見つければ e_i を更新する。この適応度に基づくトーナメント選択により、 $P'_i(t)$ の中から次世代に残す個体を $N_p - 1$ 個体選択し、これを $P''_i(t)$ とする。 $P''_i(t)$ に e_i を加えた個体群を $P_i(t+1)$ とする。
8. $t = t + 1$ とし、 $t < N_g$ なら 6. に戻る。 $t = N_g$ なら $i = i + 1$ とする。 $i < N_m$ なら 4. に戻る。 $i = N_m$ なら $T = T + 1$ とし、 $T < N_s$ なら、3. に戻る。 $T = N_s$ なら終了する。

3. 実験

実験 1 として、アンケートによる評価モデルの妥当性の評価を、実験 2 として、ユーザの提案システムの使用によるシステムの評価をした。以下に実験 1 および実験 2 の概要を示す。実験で設定した各可調整パラメータの値を、 $s = 25$ 、 $\alpha = 0.5$ 、 $\beta = 8.0$ 、 $\gamma = 1.3089$ 、 $s_{\max} = 4$ とした。 γ に関しては、スライダの値 $a = 0.5$ のとき $p_i = 0.033$ (10 個体) となるように設定した。また、 s_i を整数とした。

3.1 実験 1

本システムを用いて入力曲を 1 ステップだけ進化させたメロディをユーザに聴取させ、それぞれの進化曲がどの入力曲から進化したと感じ取れたかをアンケート調査した。ユーザは大学生 11 人を対象とし、入力曲、進化曲の順にスピーカーを用いて全員同時に聴取させた。入力曲として、「大きな古時計」、「荒城の月」、「お正月」、「螢の光」および「故郷」の日本人に馴染みのある 5 曲を使用した。これらの曲の調は、それぞれ「大きな古時計」が C メジャースケール、「荒城の月」が C マイナースケール、「お正月」、「螢の光」、「故郷」が F メジャースケールとなっている。回答する曲名に関しては重複を許し、入力曲、進化曲ともに聴き直しを許した。また曲名の回答と併せて、進化曲のメロディとしての良さを 5 段階評価 (1 ~ 5 点) として点数付けさせた。

表 1 に、実験 1 で設定した GA のパラメータを示す。表 1 における「世代」とは、1 ステップで個体群を進化させる回数のことを指している。また表 2 に、実験 1 のアンケート結果を示す。

一致率とは、各ユーザの回答した曲名と入力曲の曲名との一致数を、ユーザの人数で割ったものである。表 2 より、ユーザ

表 2: 実験 1 の結果

曲名	一致率	評価平均
大きな古時計	54.5%	2.6
荒城の月	72.7%	3.3
お正月	81.8%	3.2
蛍の光	45.5%	2.8
故郷	72.7%	3.0



図 4: 「荒城の月」から生成されたメロディ例

の回答の平均一致率が 65.5% となり、ベースライン (20%) を有意に超えていることがわかる。図 4 に、「荒城の月」から生成された進化曲を示す。図 4 より、進化曲は入力曲と同一の調になっていることがわかる。一致率が高くなった原因として、両曲で調が一致していることが大きく影響していると考えられる。

3.2 実験 2

提案システムの評価として、ユーザの嗜好に合ったメロディを作成することを目的としてユーザにシステムを使用させ、その後感想と改善点を自由形式で回答させた。使用する入力曲としては実験 1 と同様の 5 曲を使用した。なお、各曲の調を統一するために「大きな古時計」を F メジャースケール、「荒城の月」をその平行調である D マイナースケールに移調したものを使用した。また、各ステップで生成される進化曲は 5 曲とした。ユーザによる評価は、「5 曲とも良くなった」あるいは「良いと思える曲がもうできそうにない」と思った時点で終了させた。実験 2 における GA の条件は表 1 に従うが、世代数のみを 300 に設定した。

実験 2 の結果として、各ユーザの最終ステップまで要したステップ数の平均は 5.7 ステップで、要した時間としては平均約 15 分であった。表 3 に、各ユーザの初期ステップおよび最終ステップにおける進化曲の評価値の平均を示し、図 5 に実験 2 でユーザによって生成されたメロディの例を 1 つ示す。表 3 より、初期ステップで生成されたメロディよりも最終ステップで生成されたメロディの方が高い評価値が得られていることがわかる。また、実験後のアンケートでもすべてのユーザが最終ステップで得られたメロディが最も良かったと回答していたことから、提案システムによってメロディがユーザの嗜好に合わせてある程度進化していったといえる。

しかしながら、明確に自分がどのようなメロディを生成したいかという意思を持ってシステムを使用すればある程度望んだメロディが生成されたという感想があった一方で、自分がどのようなメロディの生成を目指しているのか途中でわからなくなるという感想も多くあった。その改善案として、進化前と進化後でメロディが良くなっているのかどうか不明確になり評価が困難になるということから、進化前のメロディとの比較、あるいは進化前のメロディへの復帰を可能とする実装が考えられる。また、1 つの個体に対して部分的に良いメロディがある場合その個体の評価が困難になるということから、曲中のフレーズ位置を選んで次のステップの進化曲に反映したいという意見も得られた。操作が煩雑になりユーザの評価における負荷が大きくなるという問題が考えられるが、容易かつ明確に部分的評価ができるようなインターフェイスの実装が課題となる。

表 3: 実験 2 の結果

評価値	5	4	3	2	1
初期ステップ	10%	20%	25%	25%	20%
最終ステップ	55%	20%	10%	15%	0%

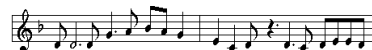


図 5: 実験 2 で生成されたメロディ例

4. まとめと今後の課題

本研究では、surrogate model を用いた進化型計算から着想を得た対話型進化型計算を導入した音楽自動生成システムを提案した。そして、ユーザによるシステムの評価実験をし、評価モデルの妥当性、および提案システムの有効性を示した。

今後の課題としては、様々な GA の適用、評価モデルのさらなる改良、個体表現の変更、より良いパラメータの発見、およびエリート個体の評価の方法などが挙げられる。また、このシステムの応用として、今回の実験 2 で同時に生成したメロディをうまく組合せて、1 つのより長い曲の生成も今後取り組んでいきたいと考えている。また、本研究では、入力曲は既存の曲を数曲用い、それらの中から任意に選択させるというようにしたが、今後の研究では、入力曲自体もユーザに即興的に作曲させ、そのフレーズを基に対話的に進化させていくというシステムの開発も考えている。この場合、ユーザによって入力されたメロディが短いフレーズであれば、そこから得られる音楽的特徴量が少ないため、音楽理論をシステムに導入することでメロディの特徴量を補完するというような手法も考えている。

謝辞

本研究は、日本学術振興会科学研究補助金基盤研究 (C) (課題番号 26330282) の補助を得て行われたものである。

参考文献

- [1] 安藤大地, Palle Dahlstedt, Mats G. Nordahl, 伊庭齊志: 対話型 GP を用いたクラシック音楽のための作曲支援システム, 芸術科学会論文誌 Vol. 4 No. 2 pp. 77-87 (2005)
- [2] Y. Jin. A comprehensive survey of fitness approximation in evolutionary computation. *Soft Comput.*, Vol. 9, No. 1, pp. 3-12, January 2005.
- [3] Y. Jin, M. Olhofer, and B. Sendhoff. On evolutionary optimization with approximate fitness functions. In: *Proc. of the Genetic and Evolutionary Computation Conference GECCO*, pp. 786-793. Morgan Kaufmann, 2000.