

# クラスタリングと主成分分析を用いた 対話型遺伝的アルゴリズムによる感性モデルの推定

Presumption of user's Kansei model by interactive Genetic Algorithm  
using clustering and Principal Component Analysis

田中美里\*<sup>1</sup> 廣安知之\*<sup>2</sup> 三木光範\*<sup>3</sup> 吉見真聡\*<sup>3</sup> 佐々木康成\*<sup>3</sup> 横内久猛\*<sup>2</sup>  
Misato Tanaka Tomoyuki Hiroyasu Mitsunori Miki Masato Yoshimi Yasunari Sasaki Hisatake Yokouchi

\*<sup>1</sup>同志社大学大学院工学研究科 Graduate School of Engineering, Doshisha University  
\*<sup>2</sup>同志社大学生命医科学部 Faculty of Life and Medical Sciences, Doshisha University

\*<sup>3</sup>同志社大学理工学部 Faculty of Science and Engineering, Doshisha University

iGA (interactive Genetic Algorithm) is the optimization method for Kansei, and we study the application of iGA to product recommendation. In product recommendation, a user has many preferences even in the same category. In this paper, the proposed method using clustering and principal component analysis to detect the multiple preferences and search for a preference was discussed. The recommended products by the proposed method give users higher satisfaction than that of conventional method. To find the effectiveness of the proposed method, the following two experiment were performed. In the first experiment, subjects' multiple preferences were confirmed and multiple peak landscapes were obtained. The second experiment examined that the proposed method searched effectively for subjective landscapes which have multiple peak.

## 1. はじめに

対話型遺伝的アルゴリズム (interactive Genetic Algorithm: iGA)[Takagi 01] は遺伝的アルゴリズムの評価関数を、人間の主観的な評価に置き換えることで、印象や嗜好といった人間の曖昧な感性に基づいた最適化を行う手法である。主に服飾デザイン [Kim 00] や補聴器フィッティング [大崎 98] など、人間の感性を必要とする設計問題への応用が研究されている。本研究では、iGA のショッピングサイトにおける商品推薦への応用を検討している。Web 上のショッピングサイトでは、一度に呈示できる商品数に限りがあり、購入される可能性の高い商品を効率的に呈示する必要がある。iGA を用いた推薦により、様々な商品の閲覧を繰り返すような、購買意欲はあるが購入対象が曖昧なユーザの嗜好にそった商品を呈示できると考えられる。

iGA を商品推薦に応用する際に問題となるのが、ユーザの嗜好の多峰性である。商品に対するユーザの好みは一通りではなく、同じカテゴリに属する商品であっても、複数の好みがあると考えられる。例えば T シャツでは、色について白だけでなく、青も好きといった例である。従って、商品推薦の際には、呈示する商品を一種類に集中させるよりは、ユーザの多数の好み (多峰の嗜好) を反映した最適化が必要となると考えられる。しかし、従来の iGA は唯一の最適解を求めるよう設計されており、多峰の嗜好における峰毎の探索に対応していない。

本稿では、このユーザの多峰性の嗜好を推定するアルゴリズムとして、クラスタリングによる多峰の推定と主成分分析による峰毎の探索を行う手法について検討した。まず、予備実験において嗜好の多峰性の有無について検証した。12 人の実験参加者の半数以上で嗜好に多峰性が見られた。また、この実験結果を基に、提案手法によって iGA によって峰の推定と探索が行われることを確認した。



図 1: iGA システム利用の流れ

## 2. 対話型遺伝的アルゴリズム

### 2.1 概要

対話型遺伝的アルゴリズムは、ユーザの主観的な評価と探索を繰り返すことで、解候補全体をユーザの嗜好や印象に沿って進化させる対話型最適化手法の一つである。図 1 にその流れを示す。ユーザは呈示された複数の解候補に対して評価値をつける。システムはその評価値を基に、次の解候補を生成するための親個体を選択し、交叉によって子個体を生成する。以上を 1 世代の操作とし、これを複数世代繰り返すことで、解候補全体がユーザの評価の高い個体へと更新されていく。

これらの解候補の一つ一つは、設計変数の数値列で表現されている。これらの設計変数によって構築された設計変数空間の各座標に対し、ユーザの評価値が描く図 2 のような超曲面を感性のランドスケープ [Tanaka 11] と呼ぶものとする。

### 2.2 多峰性のランドスケープにおける課題

感性のランドスケープは、特にユーザの好みの指向性が多様である商品推薦などの問題において、大きな一つの単峰ではなく、ほぼ等しい高さとなる複数の峰を有した多峰性のランドスケープになることが想定される。

しかし、従来の iGA の多くは、一つの最適解、または最適な領域を求めることを想定して設計されている。例えば、図 3 に示すように、ユーザの評価が同程度の峰が離れている場合、その間の評価値の低い領域が探索領域に含まれる。適切に峰

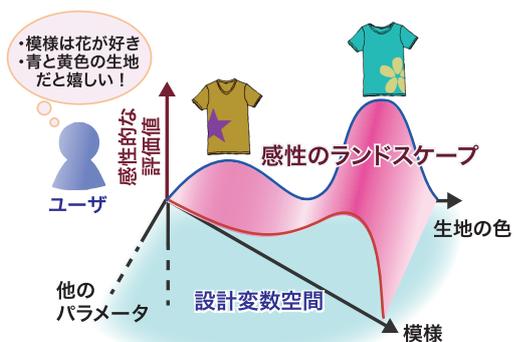


図 2: 感性的ランドスケープ

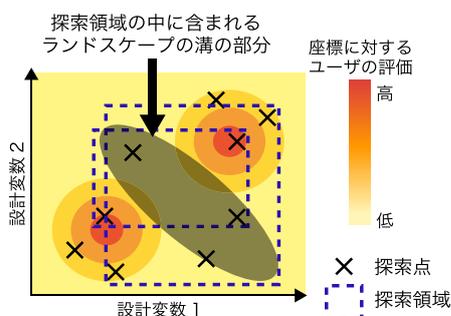


図 3: 多峰性のランドスケープに対する探索の注意点

を抽出して探索領域を絞り込むことが、探索能力を向上させ、ユーザの呈示内容に対する満足度を向上させると考えられる。

また、多峰を並行して探索することで、全体が一つのパターンに収束することを避けることができる。探索の進行による収束であっても、ユーザにとって選択の幅がなくなることは満足度に悪影響を与える場合がある。従って探索においては、ある程度の多様性を維持することが望ましい。

加えて、商品推薦を対象問題とする場合、商品の購入というゴールによって、ある峰に対する評価が大幅に低減すること考えられる。単峰性の峰として探索されていた場合、それまでの探索の履歴が無効となる。しかし、多峰性のランドスケープとしてユーザの嗜好のモデルが推定できている場合、一つの峰に対する評価値が下がっても、検出されたその他の峰から継続して効率的に商品の探索が行えると考えられる。

### 2.3 提案手法

本稿では、ユーザの多峰性のランドスケープを推定し、探索する手法としてクラスタリングと主成分分析を用いた交叉手法について検討した。

まず、多峰を検出するためにクラスタリングを用いる。解候補群から選ばれた親個体に対して、設計変数空間上においてクラスタリングを行い、座標の近い親個体の組合せを抽出し、それを峰とする。次にその峰を探索するために、峰内の分布から確率モデルを構築し、子個体を生成する。クラスタリングで抽出された親個体の組合せを主成分分析し、主成分方向の空間に回転させることで無相関化する。その空間上での親個体の分布から正規分布確率を生成し、子個体の生成を行う。

これらの手法により、多峰の検出と各峰の探索を行い、ユーザに対して、多様性の維持された精確な探索を呈示することができると考えられる。



図 4: 対象問題の呈示個体例

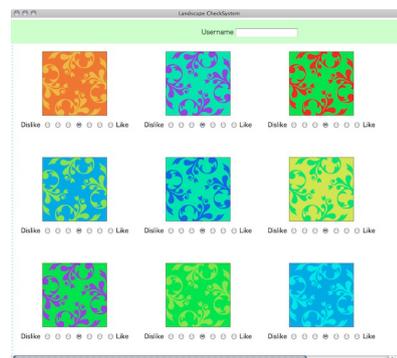


図 5: 評価インターフェース

## 3. 実験

本稿では、提案手法の有効性を検証するために2つの実験を行った。まず、1つ目の実験では、ユーザの嗜好に多峰性が存在することを確認した。そして、この実験の結果を踏まえ、2つめの実験において提案手法について嗜好の多峰性が検出、探索できることを検証した。

### 3.1 嗜好の多峰性の検証

#### 3.1.1 実験概要

ユーザの嗜好に多峰性が存在することを確認するため、複数の対象問題について、実験参加者の嗜好のランドスケープを近似的に抽出した。実験参加者は20代の男女12名である。

実験では、対象問題の各設計変数をそれぞれ10等分する格子を作り、その格子点の座標から生成した100個の個体それぞれに対する好きの度合を、実験参加者に評価させた。その格子点の評価値を基に設計変数空間全体のランドスケープを推定し、その形状からランドスケープの多峰性を評価した。

#### 実験システム

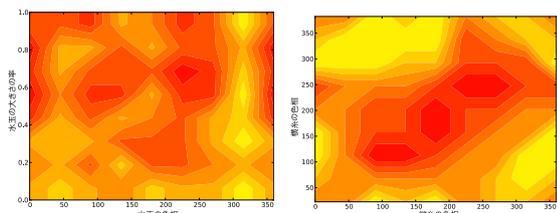
実験で用いた対象問題は表1に示す3種類の模様である。設計変数の数はいずれも2である。色はHSB表色系を用い、彩度と明度は固定し、色相のみを設計変数に用いた。水玉の大きさの率は隣り合う水玉同士の距離に対する、水玉の半径の大きさの比率である。それぞれの評価時の個体表現の一例を図4に示した。

図5が実験で用いた評価インターフェースである。評価インターフェースには100個の個体が呈示されており、各個体の下にその好ましさの度合を7段階で評価するためのラジオボタンを設置した。

#### 実験手順

表 1: 対象問題の設計変数

対象問題名	設計変数 1	設計変数 2
水玉模様	水玉の色相 (0.0~360.0)	水玉の大きさの率 (0.0~1.0)
唐草模様	唐草の色相 (0.0~360.0)	地の色相 (0.0~360.0)
チェック柄	縦糸の色相 (0.0~360.0)	横糸の色相 (0.0~360.0)



(a)多峰性のランドスケープ (対象:水玉模様) (b)単峰性のランドスケープ (対象:チェック柄)

図 6: 感性のランドスケープの一例

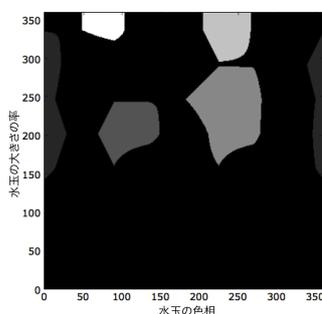


図 7: 峰ラベリング画像 (図 6(a) をラベリング処理した結果)

実験参加者には「自宅をリフォームする際に、新しく購入する布製品の家具 (カーテン, ベッドカバー, ソファーカーバー) の模様として評価する」という教示を与え, 各対象問題の評価インタフェースを操作させた. なお, この対象問題の実験順序についてはカウンタバランスをとった.

### 3.1.2 実験結果

実験により得られたランドスケープの一例を図 6 に示す. 実験参加者が評価していない格子点の間の領域については, 格子点の評価値を基に双一次補完によって評価値を推定した.

ランドスケープから峰を検出するために, まず各ランドスケープについて格子点の評価値ヒストグラムをとり, 上位 25% を分ける値を算出し, 水準面とした. この水準面よりも評価値の高い領域を, 嗜好の峰とした. このランドスケープと水準面の交叉する領域に対してラベリング処理を行った画像が, 図 7 である. これらのラベル 1 つを, 峰の 1 つとして捉え, ユーザの嗜好の峰の数を算出して多峰であるかを検証した.

全対象問題, 全実験参加者を合わせた峰の数の度数分布を図 8 に示す. 全 36 回の試行の中で, 22 回の試行において 2 つ以上の峰が見られた. この結果から, 嗜好のランドスケープに多峰性が存在するケースは多く, それに対応した探索手法が必要であると言える.

## 3.2 嗜好の多峰性の検出と探索の検証

### 3.2.1 実験概要

本節では, 提案手法が嗜好の多峰性を検出した上で, 峰を探索できることを検証するために, 以下の 2 つのシステムを実験参加者に操作させ, その結果を比較した.

- 提案システム: 提案手法の交叉を実装した iGA システム
- 従来システム: 従来の交叉手法としてブレンド交叉 [Eshelman 93] を実装した iGA システム

各システムにおける探索履歴に対し, 予備実験で得られた嗜好のランドスケープを重ねて検証することで, 提案システムにおいて実験参加者が多峰を探索できていることを検証した.

実験参加者は先の実験 (以降, 予備実験と呼ぶ) と同じ 20 代の男女 12 名である.

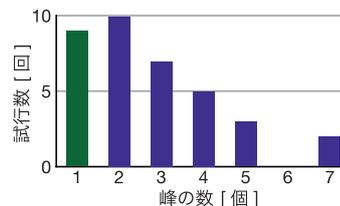


図 8: 峰の数毎の試行数



図 9: 提案手法の実験インタフェース

## § 実験システム

実験で用いた対象問題は予備実験と同じ 3 種類である.

実験インタフェースを図 9 に示す. 実験参加者は一度に呈示される 25 個の個体の画像について, 好きなものをクリックして評価する. クリックすると画像周囲の枠が赤に変化する. 全ての個体を評価した後, インタフェース下部の Next Page ボタンを押すと, システムによってページが切り替わり, 25 個の個体が新たな個体に更新される. この一度に評価する単位をページとする.

提案システムと従来システムに共通するパラメータを表 2 に示す. 両システムは交叉手法のみ異なる手続きを実装しており, その他は全て共通である.

提案システムの交叉は, 親個体群のクラスタリングによる分割と, 分割されたクラスタ内での分布に基づく子個体生成の 2 段階に分かれる. クラスタリングには k-means 法を用いた. k-means 法ではクラスタ数を予め与えておく必要があるが, 実験参加者のランドスケープが不明な状況から探索を開始するため, 最初から正しいクラスタ数を与えることはできない. そのため, クラスタ数を自動的に決定する指標としてシルエット統計量 [Rousseeuw 87] を用いた. これはクラスタ内の個体の分散が小さく, 異なるクラスタとの分散が大きい程, 値が大きくなる. k-means に与えるクラスタ数を 2 個から 9 個までの間で変化させ, シルエット統計量が最大となるクラスタ数をその世代のクラスタ数として採用した. クラスタ内の分布に基づく子個体の生成では, 主成分分析によって分散の大きい方向を求め, その方向を軸とした空間に射影することで分布を無相関化する. その空間上での分布から, 次元毎に正規分布を求め, その分布確率に基づいて子個体の生成を行う. なお, そのまま

表 2: 実験システムパラメータ

パラメータ	値
母集団サイズ	25
世代数	12
選択個体数	20
交叉率	1.0
突然変異率	0.2
突然変異手法	一様突然変異

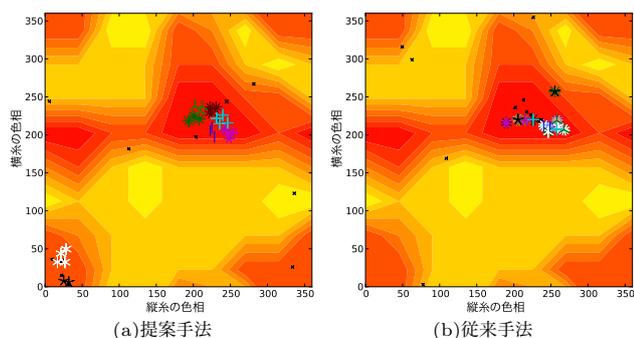


図 10: 多峰性のランドスケープの探索結果 (12 世代目)

の確率分布で子個体を生成すると、分布が中央に偏りすぎるため、標準偏差を 1.4 倍に拡大して、生成範囲を広くしている。

従来システムでは、親個体群から 2 個体をランダムに復元抽出して決定し、ブレンド交叉を行う。α 値は 0.2 とした。

選択で選出する親個体は、実験参加者がインタフェースで選択した個体のみとする。従って初期のページでは選択された個体数が少ないため、設定した数の選択個体数がアーカイブされるまで、遺伝的操作は行わず、ランダムに生成した個体を呈示するページが続く。本実験では、選択個体数の下限を 20 に設定し、最初に 20 に達して遺伝的操作を行うまでは 1 世代目とした。2 世代目以降は、その世代の母集団に対して実験参加者が選択した個体と、実験参加者が選択した個体が 20 個となるまで過去のページを遡って、親個体を選出する。

評価は 12 世代までとし、11 世代目の評価が終了すると Next Page ボタンが無効化し、End ボタンが有効化され、12 世代目の評価が終了すると同時に試行が終了する。

#### § 実験手順

実験参加者には「自宅をリフォームする際に、新しく購入する布製品の家具 (カーテン, ベッドカバー, ソファカバー) の模様として好きなものとして選ぶ」という教示を与えた。実験システムの操作順序はカウンタバランスをとり、また各実験システム中における対象問題の実験順序についてもカウンタバランスをとった。

#### 3.2.2 実験結果

実験参加者の探索履歴の一例を図 10 と図 11 に示す。図 10 は多峰性のランドスケープが得られた実験参加者の、同じ対象問題、同じ世代における探索履歴である。図 11 は単峰性のランドスケープが得られた実験参加者の探索履歴である。提案手法では、多峰性のランドスケープに沿った探索が可能であった。一方でランドスケープ全体が大きな単峰となる場合、親個体群が細かく分割されて探索範囲が小さく収束してしまい、十分な探索が行えない傾向があることも確認された。これは峰を検出するクラスタリングの精確さ、主成分分析の探索範囲の広さの 2 点が問題として考えられる。特に前者では自動決定されるクラスタの数がクラスタリングの精度を大きく左右しており、現在使用しているクラスタ数の指標であるシルエット統計量について今後検討を行っていく必要がある。

探索履歴において多峰を探索できていることを確認するために、予備実験で得られた峰ラベリング画像を用いた。各世代において、生成された子個体の座標と峰ラベリング画像を重

表 3: 多峰性のランドスケープにおける峰探索率

手法	1-6 世代平均	4-9 世代平均	7-12 世代平均
提案手法	0.5733	0.5105	0.4918
従来手法	0.5183	0.4278	0.4062

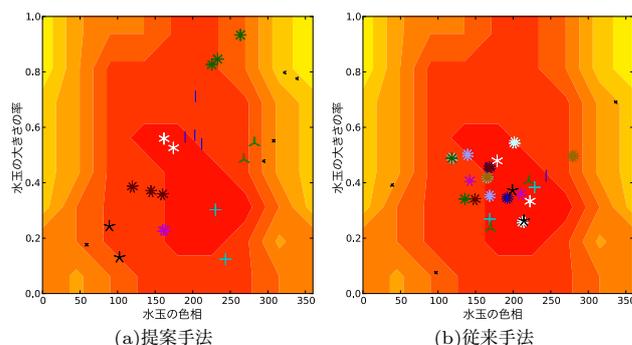


図 11: 単峰性のランドスケープの探索結果 (12 世代目)

ね、子個体を含む峰の数をカウントし、そのランドスケープに存在する峰の何割を探索できているか (峰探索率) を算出した。表 3 は、その大まかな傾向を示した表である。峰探索率を全被験者、全対象問題で、探索の前半 (1-6 世代)、中盤 (4-9 世代)、後半 (7-12 世代) で平均した。なお、ランドスケープにおいて峰の数が 1 で、単峰性と判断された試行は省いているため、値は 6 世代、各 11 試行の平均となっている。平均値はいずれの段階でも提案手法の方が勝っており、提案手法の方がより多くの峰を探索できていることが確認された。

#### 4. まとめ

本稿では iGA の商品推薦への応用を目指し、ユーザの多峰性の嗜好に対応した探索を行う交叉手法について検討した。提案手法ではクラスタリングによって多峰を検出し、主成分分析によって各峰を探索する。実験では、まずユーザの感性のランドスケープを 3 つの対象問題において近似的に抽出し、嗜好の多峰性が存在することを確認した。さらに、提案手法を実装したシステムと、従来の交叉手法としてブレンド交叉を実装したシステムを用いて実験を行い、両者の探索履歴を比較した。実験結果より、提案手法は峰に沿った探索を行い、また従来手法に比べて多くの峰を探索できていた。しかし、広い単峰性のランドスケープに対して複数の峰を検出する傾向があり、峰を検出するクラスタリング部分の改良が必要とされる。また、次元数を挙げた際のアルゴリズムの挙動について検討する必要がある。

#### 参考文献

- [Eshelman 93] Eshelman, L. J. and Schaffer, J. D.: Real-Coded Genetic Algorithms and Interval-Schemata, *Foundations of Genetic Algorithms*, Vol. 2, pp. 187–202 (1993)
- [Kim 00] Kim, H. S. and Cho, S. B.: Application of interactive genetic algorithm to fashion design, *Engineering Applications of Artificial Intelligence*, Vol. 13, No. 6, pp. 635 – 644 (2000)
- [Rousseeuw 87] Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, Vol. 20, pp. 53–65 (1987)
- [Takagi 01] Takagi, H.: Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation, *Proceedings of the IEEE*, Vol. 89, No. 9, pp. 1275–1296 (2001), Invited Paper
- [Tanaka 11] Tanaka, M., Hiroyasu, T., Miki, M., Sasaki, Y., Yoshimi, M., and Yokouchi, H.: Extraction and usage of Kansei meta-data in interactive Genetic Algorithm, in *Proceedings of WCSMO-9*, p. 505.1 (2011)
- [大崎 98] 大崎 美穂, 高木 英行: デジタル補聴器フィッティングへの対話型 EC の応用, 第 14 回ファジーシステムシンポジウム, pp. 193–194 (1998)