未評価標本を含むアンケート集計結果への評価推定と順位付けの一手法

五味恵理華† 斉藤 優理† 伊藤 貴之† 萩田真理子† 高塚 正浩††

† お茶の水女子大学大学院人間文化創成科学研究科 〒 112–8610 東京都文京区大塚 2-1-1
†† Faculty of Engineering & IT，The University of Sydney School of Information Technologies，J12，University of Sydney，NSW 2006 Australia
E-mail: †{erika53,yuri,itot}@itolab.is.ocha.ac.jp, hagita@is.ocha.ac.jp ††masa.takatsuka@sydney.edu.au

あらまし 我々は，印象評価収集を目的とした効率的なアンケート収集システム，およびアンケート集計結果の可視化システムの開発に取り組んでいる．アンケート収集システムでは評価対象となる多数の画像を提示し，ユーザにその評価を入力してもらうことで，印象評価の集計結果につながる情報を収集する．ユーザの評価が高い（または低い）画像に対する回答が重要な回答であるという前提に基づき，対話型遺伝的アルゴリズムを用いて評価が高い（または低い）と推測される画像を積極的にユーザに提示する．可視化システムではユーザ評価が高い（または低い）画像を優先的に提示するために，評価対象となる全ての評価画像を順位付けし，その上位（または下位）の画像を優先的に提示する．しかし我々の開発するアンケート収集システムでは，参加者の誰にも提示されない画像が存在する場合がある．そこで本報告では，アンケート実施時に参加者に全く提示されていない画像の評価を推定する手法，およびアンケート集計結果から全ての画像を順位付けする手法を提案する．

キーワード 対話型遺伝的アルゴリズム，印象評価，順位付け，評価推定，可視化

1. はじめに

一般消費者の評価を収集した集合知としての情報が，近年ではインターネットに広く公開されている．例えばレストラン等を紹介するウェブサイトでは各店舗の評価が数値化されて表示されていることが多い．また自動車などの商品カタログのウェブサイトでも，同様に各商品の評価が数値化されて表示されていることが多い．これらの情報において，一般消費者は全ての店舗や商品に対して評価を入力しているわけではない．むしろ，全ての店舗や商品に対して評価を求めるのは非現実である．逆に言えば，一般消費者の各々はごく一部の店舗や商品に対してのみ評価を入力していたとしても，大人数による評価の集合によって有益な情報の構築が可能であることを意味する．本研究はこの点に着目し，個々のユーザからの少ない入力情報の集合により有益な評価収集を実現できるシステムの構築を目指す．

一般消費者の評価を収集する際のアンケート実施において，その質問数はアンケート被験者のモチベーションを左右する要因に成り得る．特に，印象評価や嗜好調査では様々なパターンの質問を受けなくてはならない傾向にあり，アンケート被験者への大きな負担になると考えられる．また，一被験者の評価，嗜好には傾向があると考えられる．印象評価や嗜好調査でのアンケートで提示される質問は，多次元の要素を組み合わせたものであることが多い [1] [2]．例えばアンケートで T シャツの嗜好選択を行うのであれば，提示される T シャツは色，形，デザインといった複数の要素を組み合せによって生成されている [3]．用意される色，形，デザインの種類が増えるごとに生成される T シャツの種類は増える．この例のように用意される質問数は質問を構成する要素数に依存し，膨大なデータ数を扱うことになる．しかし，要素とその組み合わせに着目することで被験者の傾向が得られ，質問数を操作できるのではないかと考えた．

我々は印象評価収集を目的とした効率的なアンケート収集システム [4] を提案している．このシステムでは，対話型遺伝的アルゴリズム (interactive Genetic Algorithm: iGA) を用いた最適化により，ユーザの評価が高い（または低い）と予想される画像を積極的に提示する．印象評価では多くの場合において，ユーザの評価が高い（または低い）画像が何であるか重要な情報であり，評価が中程度であるという情報はさほど重要ではない．ユーザの評価が高い（または低い）という回答を多く集めることで，少ない回答数で効率よく有益な情報を得ることができると考えられる．

本報告ではアンケート集計結果の可視化を目的として，未評価画像に対する評価推定手法，および全ての評価画像を順位付けする手法を提案する．我々が提案しているアンケート収集システムでは，ユーザに提示される画像はそれぞれ異なる．そのため，評価数の単純加算では各画像への評価数のばらつきが考慮されない．よって本研究では，相対的な評価件数に応じて各画像の得点を算出し，画像を順位付けをする手法を開発している．また本研究では，ユーザに一度も提示されていない画像を対象として，自己組織化マップ（Self-Organizing Map:SOM）を用いて評価を推定する手法をあわせて開発している．これらの手法を用いて評価対象となる全ての画像を順位付けし，上位または下位となる画像を優先的に提示することで，アンケート集計結果を効率よく提示できると考えられる．

本報告では女性の装い印象評価を題材として，提案手法の実装および実験結果を示す．近年では化粧や髪型のシミュレーションサイトやアプリが普及し，さまざまな装いの画像を容易に生成できる．しかし，どのような化粧や髪型が自分に似合っているか，あるいは世間の評価が高いかを把握するのが苦手な

人がいれば，どのような装いを選ぶかに関する自己決定が苦手な人もいる．そこで，装いに関する世間の印象評価を提示することで，自己の装い決定の促進につながるのではないかと考えた．我々の実装では，顔の輪郭やパーツ，化粧，髪型などの組み合わせを「装い」と称し，その組み合わせによって合成される 1500 枚以上の画像を順次ユーザに提示する．そして，その各々が似合っているかを「Good」「Soso」「Bad」の 3 段階で回答させる．最後に，多数のユーザによるアンケート回答結果から評価画像を順位付けし可視化する．

2. 関 連 研 究

近年クラウドソーシングを用いた解析は集合知を得るための研究として注目を集めている．クラウドソーシングを用いた解析は，ウェブデザイン [5]，二次元画像 [6] や三次元モデリング [7] など多くの題材で主観評価を得る手法として役立っている．小山ら [8] は多くの被験者を集め，パラメタ空間の解析を行っている．この研究では，デザインの「良し悪し」の空間分布を推定する際に，探索対象物に数値パラメータを用いている．我々の手法では探索対象物に数値パラメータを与えない．

iGA は画像検索 [9]，音楽推薦 [10] や服装のコーディネイト提案 [11] など多くのアプリケーションで使用されている．これらの研究と本手法の違いは，嗜好個体と非嗜好個体を同時に探索している点，またアンケート収集システムに特化して開発されている点である．

3. iGA を適用したアンケート収集システム

本章では，我々が既に提案しているアンケート収集システムを紹介する．

3.1 アンケート収集においての知見

一例として多数の画像をユーザに印象評価させることを考える．印象評価における多くの場面にて，アンケート収集結果のうち重要な情報は，どの画像への評価が「高い」のか「低い」のかの両極であると考えられる．そして，少ない質問数でこの両者の回答を集中して得られれば，被験者の負担も少なく，効率的に知見を得られると考えられる．そこで我々は，評価が高い（または低い）と予想される画像を積極的に提示するために，iGA を適用したアンケート収集システム [4] を提案している．このシステムを用いることで，個々のユーザが限られた時間の中で少数の画像に対して評価を入力したとしても，多数のユーザによる回答結果の集合によって有用なアンケート集計結果を得ることが期待できる．

iGA とは，遺伝的アルゴリズム (Genetic Algorithm: GA) の遺伝的操作をベースとし，人間の嗜好，印象といった感性を適合度関数として求める解を導き出す手法である．素材空間を散策するための遺伝子選択を繰り返すため，非線形探索空間の探索に優れている．親個体から子個体に要素が受け渡される iGA を用いることで，被験者それぞれの嗜好を引き継ぐ画像が積極的に提示されると予想した．

しかし一般的な iGA は適応度の高い解を求める手法であり，適応度の高い解と低い解を同時に求めるものではない．そこで本システムでは島モデル (islandmodel) [12] の考え方を取り入れる．島モデルは全体の個体群を複数の島に分割し，各島で独立して交叉を行うアルゴリズムのことである．本手法では，島に分割しその中で交叉を行うという点に注目し，評価による島の分割を行う．個体の母集団を分割し，島ごとに独立した遺伝的操作を適用することで,「評価が高い個体群の島」「評価が低い個体群の島」を形成し，2 種類の解を同時に求める．

3.2 アンケート収集システムのアルゴリズム

我々が提案しているアンケート収集システムのアルゴリズムを以下に示す．また，そのフローチャートを図 1 に示す．

【Step1】初期集団の生成　初期集団として，母集団の中からランダムに一定枚数を選出する．

【Step2】提示　ユーザに対して個体を提示する．

【Step3】評価　提示された画像に対し，被験者の主観に基づき評価をし，適応度を与える．我々の実装では「Good」「Soso」「Bad」の 3 個のボタンを表示し，いずれかを押すことで評価を入力させる．

【Step4】選択・移住　選択では，ユーザが「Good」 「Bad」と評価した個体を親個体 とし，遺伝的操作を行う． 「Soso」と評価されたものは，ここでは適応度の低いものとみなし，親個体として選択されにくくする．続 いて親個体を， 「Good」と評価された島， 「Bad」と評価された島，の 2ヶ所に分割する.「Good」と評価された島から「Bad」と評価される個体が出た場合，その個体を「Good」と評価された島から「Bad」と評価された島へ移住させる．逆に,「Bad」と評価された島から「Good」と評価される個体が出た場合も同様に，島の移住をさせる．各島の保有する個体数は一世代の半数を基準とするが，評価の数に応じ可変とする．

【Step5】交叉　それぞれの島ごとに独立して，親個体 2 個体から特徴を受け継いだ子個体 2 個体を生成する．これにより，親個体と同じ評価を受ける子個体が生成されやすくなる．

【Step6】突然変異　個体の多様性を維持するため，突然変異率を設定し各個体に適用する．

【Step7】終了判定　設定した世代数を満たした時，終了とする．

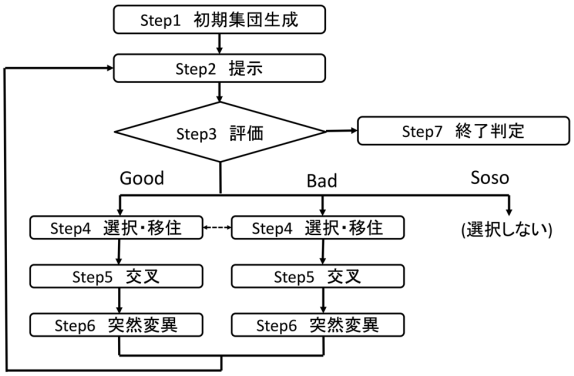

図 1　アンケート収集システムのアルゴリズム

4. 評価画像の順位付け

我々はアンケート集計結果から評価が高い（または低い）と

された評価画像を優先的に提示する可視化システムを開発している．そのためにはアンケート集計結果から相対的に各画像の得点を算出し，順位付けをする必要がある．本報告では，前章で紹介したアンケート収集システムによる回答結果をもとにして評価画像を順位付けする手法，および全ての画像を順位付けするために未評価画像の評価を推定する手法を提案する．本章では両手法の設計方針と処理手順を説明する．

4.1 未評価画像への評価推定手法

我々の開発するアンケート収集システムでは，参加者の誰にも提示されない画像が存在する場合がある．一方で，印象評価に用いた全ての画像を順位付けするためには，全ての画像に対して統一された基準で得点がつけられる必要がある．そこでユーザに全く提示されていない画像の評価を推定するための一手法として，本報告では自己組織化マップ（Self-Organizing Map: SOM）[13] を適用した手法を提案する．

SOM とは教師なしのニューラルネットワークアルゴリズムで，多次元ベクタを有する標本群を低次元へ非線形写像するデータ解析手法である．SOM に適用されるニューラルネットワークは入力層と出力層の 2 層により構成されており，入力層は入力ベクタの次元数と同じ数のノードを持ち，出力層では標本群が格子状に配置される．また多次元ベクタを低次元に非線形写像することで，特徴の似た標本がマップ上の近い位置に配置される．

SOM を用いた評価推定手順を以下に示す．

（1） 評価件数の設定: 画像の属性を表現する遺伝子に等価な 9 変数のベクタと，その画像の「Good」「Soso」「Bad」の各々の評価件数とで，各画像について合計 12 変数の多次元ベクタを構成する．

（2） SOM マップの作成: 「Good」「Soso」「Bad」の評価件数を有する全ての画像を SOM で学習させ，マップを生成する．

（3） 未評価画像類似データの探索 生成したマップを用いて，未評価画像の評価件数を推定する．総評価件数ゼロの画像の遺伝子型の各々について，SOM マップ上のニューロンの中で最も近いものを特定する．その際，未評価画像の特徴量から 9 変数の遺伝子型を生成し，マップ上の各ニューロンが持つ多次元ベクタの上位 9 変数と比較する．

4.2 順位付けのための得点算出手法

評価件数推定が終わったら続いて，全ての画像を順位付けする．順位付けに際して，アンケート集計結果から得られる各画像の評価件数を単純加算するのではなく，評価件数の相対差に応じた重みづけを適用する．それぞれの評価件数をそのまま考慮した場合，各画像間の「Good」「Soso」「Bad」評価件数の相対的な差は考慮されない．そこで，各画像間の「Good」「Soso」「Bad」評価件数の相対的な差を考慮し，各画像の得点を算出し，その得点から各画像の順位を決定する．

本手法では，画像 i の得点 $Score_i$ を以下の式で定義する．

$$\mathrm{Score}_i = \sum_{i \neq j} \frac{pg_i(ps_j + pb_j) + ps_i(pb_j)}{(pg_i + ps_i + pb_i)(pg_j + ps_j + pb_j)} - \frac{pg_j(ps_i + pb_i) + ps_j(pb_i)}{(pg_i + ps_i + pb_i)(pg_j + ps_j + pb_j)} \quad (1)$$

式 (1) は，画像 i 以外の各画像 j との評価結果の比較によって画像 i の得点を求めることに相当する．ここで pg_i は画像 i の「Good」の評価数，ps_i は画像 i の「Soso」の評価数，そして pb_i は画像 i の「bad」の評価数を示す．同様に pg_j は画像 j の「Good」の評価数，ps_j は画像 j の「Soso」の評価数，そして pb_j は画像 j の「bad」の評価数を示す．

画像 i と画像 j の評価件数の比較により，それぞれの画像の勝ち数を算出し得点を割り出す．画像 i と画像 j を比較する際，画像 i の「Good」は画像 j の「Soso」「Bad」よりも価値があるとし，画像 i の「Soso」は画像 j の「Bad」よりも価値があるとする．

$$pg_i(ps_j + pb_j) + ps_i(pb_j) \quad (2)$$

式（2）により，i の j と比較した勝ち数が算出される．逆に画像 j の「Good」は画像 i の「Soso」「Bad」よりも価値があるとし，画像の j「Soso」は画像 i の「Bad」よりも価値があるとする．

$$pg_j(ps_i + pb_i) + ps_j(pb_i) \quad (3)$$

式（3）により，j の i と比較した勝ち数が算出される．i の j に対する勝ち数は式（2）から式（3）を引くことによって算出される．この値を比較をした回数で割ることで，正規化された得点を求める．

以上の処理を画像 i 以外の各画像 j に対して適用して得点を累算することで，画像 i の得点を求める．得点が高い画像を評価が高い画像であるとし，この得点に従って各画像に順位を与える．

5. 装い画像の生成

本報告では女性の「装い」画像を題材とした我々の実装とアンケート集計結果を示す．以降，女性の顔，化粧，髪型の組み合わせを「装い」と表記する．我々の実装では，以下の手順により合成した多数の装いを，順次ユーザに提示している．画像生成のために我々はまず，20 代女性 18 人の顔画像を撮影し．これらのすべての組み合わせに対してモーフィングを適用した．その中から，以下の顔パーツの特徴を満たす 16 種類の中間顔画像を選出した．

- 顔の長さ：「長い」「短い」
- 顎周りの形状：「細い」「丸い」
- 目の印象：「二重」「一重」
- 鼻の印象：「丸い」「細い」

続いて我々は中間顔画像に化粧を施し，髪型を合成した．化粧と髪型は以下のいずれかのものとする．

- 化粧系統：「フレッシュ」「キュート」「クール」「エレガント」
- 髪の長さ：「ロング」「ミドル」「ショート」

- 前髪：「有り」「無し」
- 髪の形状：「直毛」「パーマ」
- 髪色：「茶」「黒」

以上の結果，1536 枚の装いを生成した．各画像は顔パーツ・化粧・髪型の 9 次元の遺伝子型を保有する．生成された装い画像の例を図 2 に示す．

図 2　装い画像例

6.　アンケート収集

我々は上述の装い画像群を題材として，20 代女性 30 人にアンケートを実施した．アンケート収集のために我々が開発したユーザインタフェースの画面を図 3 に示す．本実装では「Good」「Soso」「Bad」の 3 個のボタンを表示し，いずれかを押すことで評価を入力できる．iGA の条件は以下のように設定した．

- 装い画像の総数: 1536 枚
- 一世代の数: 12 枚
- 交叉率: 1.0
- 突然変異率:
  (if $n_{Soso} < 4$)：0.05
  (if $n_{soso} \geqq 4$)：$0.05(n_{Soso} - 2)$
  n_{Soso} は前世代での「Soso」の評価数
- 反復処理: 20 世代

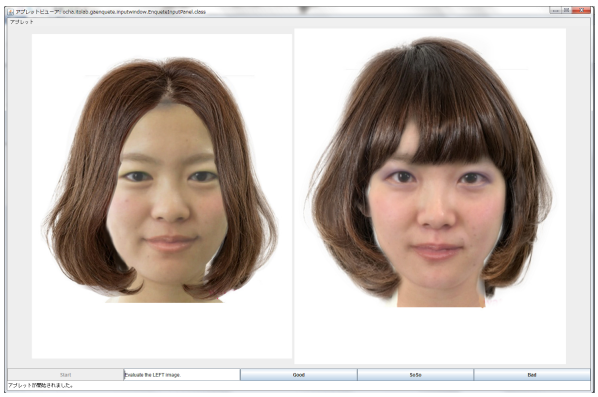

図 3　アンケート実装画面

重複を除き被験者に提示された装いの平均枚数は 169 枚であった．これは総装い数の 11%に当たり，それだけのバリエーションが被験者に提示されたことになる．いずれの被験者にも提示されていない装いは 54 枚であった．被験者 30 人のうち 7 人が「Good」と評価した装いが 1 枚，6 人が「Good」と評価した装いが 4 枚あったが，この 5 枚のうち 3 枚が同じ髪型の装いであった（図 4）．また，6 人に「Bad」と評価された装いは 5 枚みられたが共通した傾向は見られなかった．

各被験者における世代での評価傾向では，20 世代のうち「Soso」の評価がされなかった世代が見られたのは 30 人中 14 人であった．世代の推移例を図 5 に示す．縦軸が評価数，横軸が世代数を表す．また，「Good」の評価数が最多であった世代で「Bad」の評価数が最少であったケースは 30 人中 18 人，「Good」の評価数が最少であった世代で「Bad」の評価数が最多であったケースは 30 人中 14 人にみられた．加えて，20 世代を通して減少傾向が見られたのは 30 人中 8 人であった．これより，進化計算によって「Soso」の評価が少なくなるパターンがみられた．本アンケート収集において，過去に選出された装い画像も重複し提示させたところ，同じ装いでも提示順により評価が変わるパターンが見られた．このように評価の変化が見られる要因として，被験者の評価基準のゆらぎに加えて，2 枚の画像を同時に提示して比較させるインタフェースが評価の変化を誘発させている可能性も考えられる．一方で，再度提示しても再度同じ評価を受けている装い画像も多数見られた．以上により，装い間の「Good」「Soso」「Bad」の評価数の相対的な差を考慮し，アンケート集計結果を評価するべきだと考えた．

図 4　似合っていると評価された装い

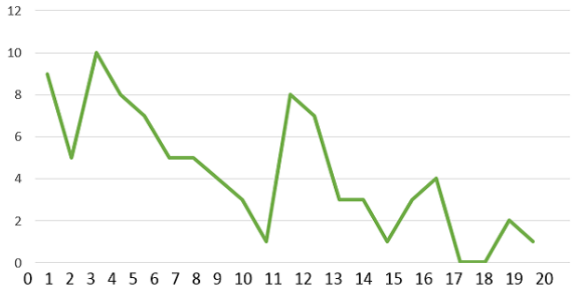

図 5　「どちらともいえない」評価の現象推移例

7.　順位をつけた可視化

前章のアンケート集計結果において，いずれの被験者にも提示されていない装い画像は 54 枚であった．これらの画像に 4.1 節の手法を適用して評価を推定した．評価推定の結果例を図 6

に示す．

	Good	Soso	Bad
000000010	0.241066	0.459615	0.299319
000001010	0.326369	0.515736	0.157895
000002001	0.222055	0.459777	0.318168
000031010	0.413297	0.432402	0.154301
000100000	0.156778	0.384602	0.45862
⋮			
111112110	0.315132	0.516023	0.168845
111120110	0.250324	0.493989	0.255687

図 6　SOM による未評価画像への得点付加結果例

続いて，4.2 節で提案した手法を適用して，装い画像を順位付けした．ここでユーザが装いの評価結果を参照するとき，顔パーツ・化粧・髪型の要素の組み合わせに注目することが多いと考えられる．よって，本報告ではユーザが要素を指定し，その要素を持つ画像間の順位を表示する，という操作手順を想定した可視化ユーザインタフェースを実装した．その画面を図 7 に示す．

このユーザインタフェース上で，評価画像を構成する要素をユーザが指定することで，その要素を持つ評価画像間の順位を閲覧できる．このとき本手法は指定要素を持つ評価画像間の順位を計算し，評価が高い画像から順に画面の左上部に提示する．そして評価が高い画像には青枠を，評価が低い画像には赤枠をあわせて表示する．また，画像の下には各画像評価の割合を表す帯グラフを表示する．「Good」が青，「Soso」が緑，「Bad」が赤で示される．

図 7　アンケート集計結果実装画面

7.1　ユーザテストの実施

装いのアンケート結果を搭載した可視化 UI を用い，ユーザテストを実施した．本ユーザテストでは，SOM による評価推定結果を含めた装いのアンケート集計結果と順位付けの妥当性の検証，またアンケート結果を提示する目的として装いの順位付けは利用者の意思決定につながるかを調査した．本実装で選択できる要素は「長い顔/短い顔」「細い顎/丸い顎」「二重/一重」「丸い鼻/細い鼻」の各顔パーツと「フレッシュ/キュート/クール/エレガント」の化粧傾向となっている．被験者はそれぞれの顔パーツ，化粧傾向から，いずれかを選択，もしくはいずれも選択をしないで結果を見ることができる．本可視化 UI の実行環境は JRE (Java Runtime Environment) 1.8 となっている．20 代女性 12 人を被験者とし，以下の項目について回答してもらった．

1.SOM による評価推定画像の選択: 選択要素を指定したとき，可視化 UI に 24 枚の画像が評価順に提示される．この中に，SOM による評価推定がされた画像が 3 枚組みこまれていると被験者に伝え，いずれの画像であると思うか回答してもらう．また，空欄がある場合は「わからない」「見当たらない」と判断されたものとみなす．指定要素は「長い顔・細い顎・二重・丸い鼻・フレッシュ」と「短い顔・丸い顎・一重・細い鼻・キュート」の二つとし，いずれに対しても回答をしてもらう．

2. 順位の妥当性の評価: 被験者には自由に要素を指定してもらい，結果を見てもらう．それらの順位は妥当であると思ったかを 5 段階で評価をしてもらう．

3. 妥当な組み合わせの調査: 被験者に結果を自由に見てもらい，順位において妥当であると思った組み合わせを提示してもらう．

4. 妥当でない組み合わせの調査: 被験者に結果を自由に見てもらい，順位において妥当でないと思った組み合わせを提示してもらう．

5. 要素の選択方法: 見るときの要素はどのように選んだかを答えてもらう．選択肢は以下とし，検索の仕方が推移した場合は、その順番で続けて記載してもらう．

1. 直観　2. 自分の顔パーツに近い項目から　3. 気になる顔パーツから　4. 好きな化粧傾向から　5. 要素を確認せず完全ランダム　6. その他

6. 意思決定への有用性の調査: 順位の提示は意思決定に役立てると思ったか、5 段階で評価をしてもらう．

7.2　ユーザテストの結果と考察

ユーザテストの各項目に対して，それぞれ結果と考察を述べる．

まず，1.SOM による評価推定画像の選択 に関しての質問の結果と考察を述べる．

本質問において，SOM によって推定された画像がどれであるかわからなかった被験者が多かった場合，その順位は自然であることを意味する．選択要素を「長い顔・細い顎・二重・丸い鼻・フレッシュ」とした場合，可視化 UI に提示される 24 枚の装い画像のうち 3 枚が SOM による評価推定を行い，順位付けに組み込まれたものである．被験者には，評価推定が行われたと思われる画像を 3 つ指定してもらうが，わからなかった場合は画像の指定をしない状態での提出を求めた．その結果，3 枚を指定した被験者は 12 人中 3 人，2 枚を指定した被験者は 12 人中 6 人，1 枚を指定した被験者は 2 人，1 枚もわからなかった被験者は 1 人であった．平均して，1 人当たり 1.9 枚を SOM による評価画像であると判断したことになる．しかし被験者 12 人のうち指定した画像が SOM による評価であった人はわずか 1 人であり，正解した画像の枚数も 1 枚のみであった．正解した被験者は画像指定の際に，「評価では悪いとなっているがそこまで悪くない気がする」とコメントしており画像を見ての主観による選択をしていることがわかる．

同様に，選択要素を「短い顔・丸い顎・一重・細い鼻・キュー

ト」とした場合も，24 枚中 3 枚が SOM による評価推定がされた画像である．SOM による評価であるとして，3 枚を指定した被験者は 12 人中 2 人，2 枚を指定した被験者は 12 人中 4 人，1 枚を指定した被験者は 4 人，1 枚もわからなかった被験者は 2 人であった．平均して，1 人当たり 1.5 枚を SOM による評価画像であると判断したことになる．しかし上記結果と同様に被験者 12 人のうち指定した画像が SOM による評価であった人はわずか 1 人であり，正解した画像の枚数も 1 枚のみであった．正解した被験者は画像指定の際に，特にコメントをしていないためどのように選択したかは不明である．

また，被験者によるコメントとして「全員が同じ評価をしていることに違和感があった」など画像の評価件数の割合だけで，SOM であると判断している人は 12 人中 4 人であり，いずれも正しい画像を指名できていなかった．よって，SOM による評価推定画像は自然な評価件数割合がつけられており，評価件数だけでは SOM による評価推定であると判断できないことがわかる．

以上の結果より，SOM による評価推定を行った画像はおおむね自然に順位に組み込まれていると判断できる．

続いて，

2. 順位の妥当性の評価

3. 妥当な組み合わせの調査

4. 妥当でない組み合わせの調査

の 3 項目に関しての質問の結果と考察を述べる．

被験者に対し可視化 UI を用いて，自由に結果を検索してもらい，提示された装いの順位が全体的に妥当であると感じたか調査を行った．各被験者の検索回数を調査したところ，最高検索回数は 33 回，最低検索回数は 5 回であった．被験者は平均して 16.8 回検索をしていた．

提示された装いの順位の全体的な妥当性を 5 段階で評価してもらった．提示された装いの順位を妥当であると答えた被験者は 12 人中 4 人，どちらかといえば妥当であると答えた被験者は 12 人中 7 人，どちらとも言えないと答えた被験者は 12 人中 1 人であった．被験者からのコメントとして「自身の主観評価とあっている」という意見が多くみられたが，中には「上位，下位の大まかな順位はあっている」「要素を絞り込みすぎないときの高評価，低評価は納得できる」という意見もみられた．後者のコメントはいずれもアンケート収集に参加している被験者からのものであった．アンケート収集に参加した被験者は，自身の細かい評価が各装いに対してつけられていると考えられる．これより，提示された装いの順位は全体的におおむね妥当であったといえる．

また妥当な組み合わせの調査では，12 人中 4 人が「短い顔・丸い顎・二重・細い鼻」の要素を含む装いに対して妥当であったと回答していた．そのうち 2 人が上記要素に加え「キュート」の化粧を選択したときの装い順位を妥当であるとしていた．「短い顔・丸い顎・二重・細い鼻・キュート」を要素と指定したとき，高評価 10 個のうち 8 個の装いが「前髪有り」，低評価 14 個のうち 10 個が「前髪無し」であった（図 8）．これより妥当と評価される装いの傾向を見ることができた．また「無難な髪形が

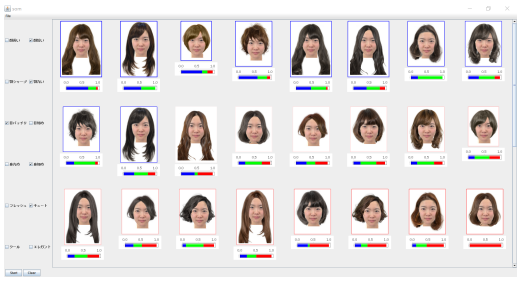

図 8　選択要素を「短い顔・丸い顎・二重・細い鼻・キュート」としたときの結果

高評価として提示されていた」というコメントがあったことから，高評価と低評価に分けて結果を見た時の妥当性が回答に反映されていると考えられる．逆に妥当でない組み合わせの調査では，12 人中 8 人が回答をし，被験者全体の傾向はなかった．

最後に，

5. 要素の選択方法

6. 意思決定への有用性の調査

の 2 項目に関しての質問の結果と考察を述べる．

被験者に対し，要素の選択をどのように行っているか調査した．結果を図 9 に示す．図 9 の横軸は人数を表す．調査の結果 12 人中 7 人が最初は「自分の顔パーツに近い項目から」選ぶと回答した．また，「好きな化粧傾向から」選ぶと回答した人はいなかった．これより自身の顔に合わせて，装い要素をみたいという考えを持つ被験者が多いと推測できる．

順位の提示は意思決定に役立てると思ったかの 5 段階評価では，12 人中 5 人が役立つ，12 人中 7 人がどちらかといえば役立つと回答した．これより，順位の提示は意思決定におおむね役立つと考えられる．被験者からのコメントとして「髪型の一覧を見れて参考になった」という感想があったことからも，特定の装いの善し悪しだけでなく，装いの評価全体の順位を提示したことによって被験者に新しい意思決定の選択肢を与えることができたと考えられる．また「第三者の意見からこれはやめておこうという意思決定につながった」という意見もあり，装いの印象評価に関しては低評価のものを適用しないという意思決定につながっていた．これより，高評価なものだけでなく低評価なものも提示することに有用性があると考えられる．

以上の結果より，本ユーザテストの目的はおおむね達成できたと考えられる．被験者からのコメントとして，「選択要素は文字ではなく画像で提示されたほうが直感的にわかる」「髪型要素も選択できるようにしてほしい」など実装面の改善点が多くあがった．アンケート提示方法によって意思決定の方法も変化する可能性があるため，今後の課題としたい．

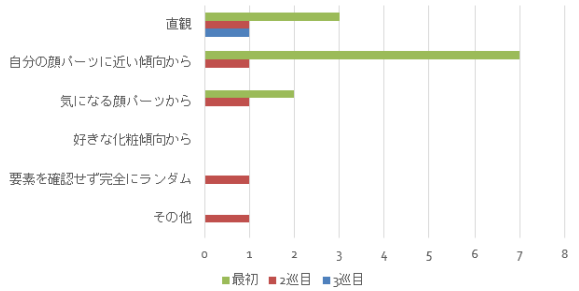

図 9　要素の選択方法の調査

8. まとめと課題

本研究では，iGA を用いたアンケート収集手法および収集結果の順位付け手法，また全ての画像を順位付けするために未評価画像の評価を推定する手法を提案した．

今後の課題として，進化計算の改善があげられる．iGA のパラメータ調整を検討し，より効率的なアンケート収集を目指す．最後に，女性の装い以外の題材へのアンケート手法の適用を進めたい．

謝　　辞

本研究の一部は，日本学術振興会科学研究費補助金の助成に関するものです．

文　　献

[1] 槙, 乾, 中村, 街路景観の評価構造の安定性, 日本建築学会計画系論文集, 458, 27-33, 1994.
[2] 桐谷, 牛窪, 高野, 化粧品配色の印象評価と表現媒体の関係, 感性工学研究論文集, 5(1), 27-32, 2004.
[3] 伊藤, 廣安, 三木, 横内, 対話型遺伝的アルゴリズムにおける嗜好の多峰性に対応可能な個体生成方法, 人工知能学会論文誌, 24(1), 127-135, 2009.
[4] E. Gomi, Y. Saito, T. Itoh, Visualization of Crowd-Powered Impression Evaluation Results, 18th International Conference on Information Visualisation (IV2015), 89-94, 2015.
[5] K. Reinecke, K. Z. Gajos, Quantifying Visual Preferences around the World, ACM SIGCHI Conference on Human Factors in Computing Systems, 2014.
[6] Y. Gingold, A. Shamir, D. Cohen-Or, Micro Perceptual Human Computation for Visual Tasks. ACM Transactions on Graphics, 31(5), 119:1-119:12, 2012.
[7] A. Secord, J. Lu, A. Finkelstein, M. Singh, A. Nealen, Perceptual Models of Viewpoint Preference, ACM Transactions on Graphics, 30(5), 109:1-109:12, 2011.
[8] Y. Koyama, D. Sakamoto, T. Igarashi, Crowd-powered Parameter Analysis for Visual Design Exploration, ACM Symposium on User Interface Software and Technology, 65-74, 2014.
[9] H. Takagi, S. Cho, T. Noda, Evaluation of an IGA-based Image Retrieval System Using Wavelet Coefficients, IEEE International Conference on Fuzzy Systems, 1775-1780, 1999.
[10] Y. Saito, T. Itoh, MusiCube: A Visual Music Recommendation System featuring Interactive Evolutionary Computing, Visual Information Communication and Interaction Symposium, 2011.
[11] S. Cho, Towards Creative Evolutionary Systems with Interactive Genetic Algorithm, Applied Intelligence, 16(2), 129-138, 2002.
[12] D. Whitley, S. Rana, R. B. Heckendorn, The Island Model Genetic Algorithm : On Separability, Population Size and Convergence, Journal of Computing and Information Technology, 7(1), 33-47, 1999.
[13] 宗本, 自己組織化マップによる評価パターンの可視化分析を用いた空間嗜好の類型化の研究：企画展の展示計画を対象として, 日本建築学会計画系論文集 (609), 231-237, 2006.