

進化計算を用いた「合コン」問題の解法と評価

黒岩 将^{1,a)} 安本 慶一¹ 村田 佳洋² 伊藤 実¹

受付日 2012年2月2日, 再受付日 2012年3月22日/2012年5月21日,
採録日 2012年7月27日

概要: 合コン（お見合いパーティ）では、できるだけ多くのカップルを成立させたいという要求が発生する。本論文では、合コン結果から、カップルが成立しやすい男女の属性情報の組（好相性と呼ぶ）を、進化計算を用いて求めることで、理想的な合コンメンバ（合コン参加者名簿）を決定するシステムを提案する。提案システムでは、男女の属性情報の組を進化計算の解集団（初期個体群）としてシステムに持たせ、合コンでのカップル成否を解の評価値（適応度）としてフィードバックしながら、好相性を表現する準最適解集団の獲得を目指す。提案システムの実現には、複数の好相性の同時探索、様々な参加者による多数の合コンの実施が必要である。これらの課題を解決するため、進化計算の新しい選択法、過去の合コン結果の新たな解評価への再利用法を考案した。提案システムを評価するため、カップルになった男女の属性情報を解としてそのまま利用する比較手法を用意し、計算機シミュレーションにより比較を行った。結果、提案手法が、比較手法に比べて、半分の合コン実施回数で、約2倍のカップル成立数を達成できることを確認した。

キーワード: 進化計算, 合コン問題, 多峰性関数, マルチ・ニッチイズ・クラウディング

A Method for “Gokon” Problem Using Evolutionary Computation and Its Computational Evaluation

SHO KUROIWA^{1,a)} KEIICHI YASUMOTO¹ YOSHIHIRO MURATA² MINORU ITO¹

Received: February 2, 2012, Revised: March 22, 2012/May 21, 2012,
Accepted: July 27, 2012

Abstract: There is a demand to maximize the number of successful couples in match-making party called Gokon. In this paper, we propose a method to find good affinity patterns between man and woman from resulting matches of Gokon by encoding their attribute information and using evolutionary computation scheme. We also propose a system to assign the best members to each Gokon based on the method. The purpose of the proposed system is to derive good affinity patterns. For this purpose, a specified number of solutions as chromosome of evolutionary computation (EC) are initially prepared in the system. By feeding back the results of Gokon to the solutions as fitness value of EC, semi-optimal solutions are derived. To realize the proposed system, we need simultaneous search of multiple different good affinity patterns and efficient evaluation of solutions through many Gokons with various attribute members. To these challenges, we devise new methods for efficient selection operation and reuse of the past matches to evaluate new solutions. To evaluate the system, we prepared a comparative method which uses attribute information between a man and a woman who made a match as a solution. Through computer simulation, we confirmed that the proposed system achieves twice as many successful couples as the comparative method with about half of evaluation times.

Keywords: evolutionary computation, matchmaking party, multi-modal, multi-niches crowding

1. はじめに

2005年以降、わが国の人口増加率は0%以下で推移してきた[1]。そのため、政府は、2003年より少子化社会対策基本法を整備、内閣府に少子化対策室を設置するなど、その対策に乗り出している。内閣府によるアンケート調査では、少子化の主たる要因は、結婚適齢層の結婚・出産に対

¹ 奈良先端科学技術大学院大学情報科学研究科
Graduate School of Information Science, Nara Institute of
Science and Technology, Ikoma, Nara 630-0101, Japan

² 広島市立大学大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima City
University, Hiroshima 731-3194, Japan

a) sho-k@is.naist.jp

する価値観の変化であり、未婚の主たる理由は、適切な相手に巡り合わないからであることが分かっている [2]. そのため、適切な相手と巡り合える場を提供するサービスを、民間企業や地方自治体が提供してきた [3]. そうしたセッティングサービスでは、カップル成立数を高めることが課題となっている. しかし、相性の良い男女を同じ合コンに参加させるためには、どのような男女が好相性なのか (予測問題)、また、既知となった好相性をどのように用いて、いくつの合コンをセッティングすればよいのか (組合せ最適化問題)、という 2 つの問題が存在する. そこで、本論文では、この 2 つの問題をあわせて解くシステムを提案する.

前者の予測問題の解決には、進化計算 (EC, Evolutionary Computation) [4] を用いる. 男女の属性情報の組合せを EC の解 (染色体あるいは個体) としてシステムに持たせ、カップル成否を解の評価値 (適応度) としてフィードバックしながら、好相性の準最適解集団を求める. 後者の組合せ最適化問題には、好相性を表す解ごとに、それに似ている属性値を持つ男女を合コンの定員数ずつ集めるという単純な方法を採用し、段階的に「合コン問題」が解けるような構成にする.

相性空間 (男女の属性の組がとりうる値の範囲) は多峰性であり、多様なカップルを成立させるためには、EC によって、この多くの峰を求める必要がある. しかし、従来の EC 研究は、1 つの最適解を求めるものが大半である. 我々の以前行った研究でも、1 つの最適解を求める手法しか考えていない [5]. 提案システムでは、相性空間のような多峰性関数の複数極大値算出手法として、類似の解が増えないように工夫した EC の選択法を導入する (4.2.1 項, 提案選択法). 対象問題の最適解を求めるためには、様々な参加者による多数の合コンを実施する必要がある. そこで、解に対して合コンを行わずに、過去の合コン履歴を活用して任意の解の評価を行う方法を導入する (4.4.2 項, ストック利用評価法).

提案システムを評価するため、カップルが成立した男女の属性を好相性としてそのまま活用する比較手法を用意して、計算機シミュレーションにより比較を行った. 結果、提案システムが、比較手法に比べて、半分の合コン実施回数で、約 2 倍のカップル成立数を達成できることを確認した.

2. 関連研究

男女のマッチングに関する研究として、古くから結婚安定問題が知られている [6]. これは、複数の男女それぞれが、異性に対し、自分の好みで順位付けを行い、その順位リストの「安定性」を満たすマッチング (結婚) を求めるものである. 2004 年度から施行された医師臨床研修制度下の医師の研修先病院を決定するシステムには、同問題を解く手法が実装されている. こうしたグループ分割に、EC を応用したアルゴリズムも数多く提案されており、なかで

も MOCK [7] は、未知のグループ数を自動的に決定する能力を有し、様々なデータセットにおいて高い分割性能を示すことで知られている. しかし、いずれも好みや希望が既知であるものとしてマッチングや分割を行う研究であり、未知である相性 (好み) を予測することはできない.

EC は、対象問題の評価関数をブラックボックスのまま扱うことができるため、幅広い分野に応用されている [8]. 身近な例としては、新幹線のぞみ 700 系統の車両先頭部の形状設計への応用がある [9]. この研究では、形状を EC の解にエンコードし、高速走行時に発生する騒音と空気抵抗を評価値にして最適化を行っている. また、人間の好みや感性などのモデル化が困難なものをそのまま EC の評価系として組み込んだ対話型進化計算と呼ばれる分野の研究も広く行われている [10]. たとえば、楽譜情報を解に、メロディの聞こえ良さを評価値にして、作曲者の好みの曲を生成する研究などがある [11]. しかし、これらの応用例の多くは、1 つの大局的最適解やパレート最適解の探索問題を扱っており、本研究で扱う局所最適解も含めた複数解の探索問題にそのまま適用することはできない. また、EC を多峰性関数に適用する研究もいくつか存在するが、NSGA-II [12] を筆頭にほとんどがパレート最適解を求める多目的最適化を対象としており、問題の種類が異なる.

3. 合コン問題

合コン問題とは、大規模な利用者集団から、好相性の男女混合グループを複数作る問題である.

入力: 問題の入力は男性利用者集団 $B = \{b_1, b_2, \dots\}$ と女性利用者集団 $G = \{g_1, g_2, \dots\}$ と定義する. 男性 $b_i \in B$ と女性 $g_j \in G$ は、それぞれ k 個と l 個の属性情報 $(b_{i1}, b_{i2}, \dots, b_{ik})$, $(g_{j1}, g_{j2}, \dots, g_{jl})$ を持つ. 属性情報は、たとえば、身長や体重、経歴や性格に関する情報である.

出力: B と G をそれぞれ L 人ずつの部分集団に重複を許して分ける、男女同数の部分集団の組 (合コン) に割り当てる. 得られる合コンの数を h とし、 i 組目の合コンを $[B_i, G_i]$ と表す. 問題の出力は、以下の目的関数を最適化する合コン集団 $[B_1, G_1], \dots, [B_h, G_h]$ と未知の評価関数 F である.

目的関数: $[B_i, G_i]$ のカップル成立数を返す評価関数を F とし、 h 回の合コンにおけるカップル成立数の総和を最大化する目的関数を以下のように与える.

$$\text{maximize} \sum_{i \in \{1, \dots, h\}} F(B_i, G_i) \quad (1)$$

4. 提案システム

図 1 に示すように、EC を用いて、合コン問題の評価関数 F を推定し、合コン割当て名簿を出力 (= 参加者を決定) するシステムを提案する. 提案システムは、男女の相性の

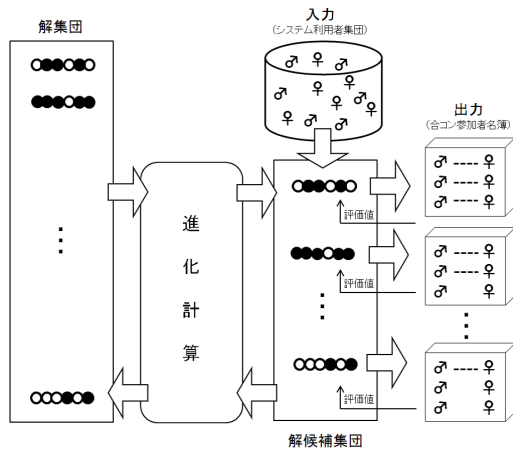


図 1 提案システム概要
Fig. 1 Outline.

表 1 解集団に対する処理の流れ
Table 1 Algorithm.

1: 解集団を初期化
2: 指定した世代数, 以下を適用
3: 進化計算オペレータ
4: 各解候補に以下を適用
5: 合コン割当てオペレータ
6: 合コン評価オペレータ
7: 解候補集団を用いて解集団を更新

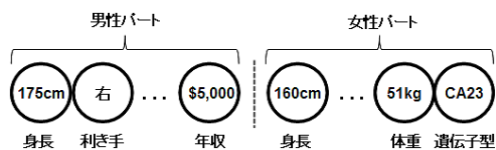


図 2 解の例
Fig. 2 Example of solution representation.

良さを表現する解集団と、その解集団に対し EC を実行する進化計算オペレータ群 (選択, 交叉, 突然変異), 生成された解候補集団に対し合コン割当て名簿を生成する合コン割当てオペレータ, 合コン結果を評価値として解候補に与える合コン評価オペレータから構成される。解集団に対して、進化計算, 合コン割当て, 合コン評価の各オペレータ (それぞれ 4.2 節, 4.3 節, 4.4 節で後述) を適用し, 評価値付きの解候補集団を生成する。この解候補集団を用いて解集団の更新を行う。ここで, 初期解集団を 1 世代, $t-1$ 回目の更新後の世代を t 世代と呼ぶ。世代を経ながら, 最適解集団, すなわち, F の推定を目指す。解集団に対する操作を表 1 に示す。

4.1 解

図 2 に示すように, 解は, 男性と女性の, 身長や収入, 学歴といった属性情報の組合せから構成される。前半分を男性の属性, 後半分を女性の属性に対応させる。これらの情報はアンケートによりシステム利用者から取得す

る。属性には, 男女の相性に関係するといわれるヒト白血球抗原 (HLA: Human Leukocyte Antigen) [13] などの先天的属性 (しかし, HLA は評価関数が既知のため推定の必要がない。したがって今回は取り扱わない) と, TCI (Temperament and Character Inventory) [14] などの性格気質 (後天的) テストの結果を加える。

こうした属性は, 実際に, いくつかの結婚情報サービスで使用されているものである。しかし, これらは業者ごとに異なり, また, あまり学術的研究も進んでいないため, どの情報がどの程度, 相性に影響するかは未知である。身長や収入, 学歴なども, よく調べれば, 平均より高い/低い, 多い/少ない, 高い/低いという分類で十分かもしれない。この場合はそれぞれ 1 ビットで表現できる。また, TCI などは, アンケートの項目数は非常に多いが, 結果は 12 パターン (4 ビット) に分類される。性格気質を 5 パターン (3 ビット) に分類できるとする研究も存在する [15]。既存の結婚紹介サービスでは, 上記に加え, 血液型 (2 ビット), 長男長女かどうか (1 ビット), 家が近いか (1 ビット) などの質問を用意しているが, 以上にあげたすべてを足しても 11 ビットであり, 相性に影響する要因を正確に突き止めれば, 必要な情報量は 20 ビット程度に収めることができると予想している。

本研究では, こうした男女の属性情報によって, 相性の良さが決まるという仮定をおいている。各初期解の各要素は乱数を用いて生成する。

4.2 進化計算オペレータ

本オペレータは, 図 1 に示すように, 解集団に対して選択, 交叉, 突然変異という 3 つの EC の操作を加える。

合コン問題は複数の最適解 (好相性) を持つ問題であり, それら複数の最適解を求めるためには解集団の多様性を維持する必要がある。そこで, 多様性を保つ工夫を施した選択法を提案する (4.2.1 項に後述)。交叉は一点交叉とし, 突然変異とともに, 一般的な EC の定石に従う。

4.2.1 提案選択法

提案選択法では, t 世代解集団 S^t から $t+1$ 世代解集団 S^{t+1} を生成する際, 類似の解が増えないように工夫する。また探索性能を向上させるために, 評価値の高い解を優先的に $t+1$ 世代に残すエリート保存戦略の要素を取り入れる。以下, その選択戦略について詳述する。

図 3 に示すように, まず, S^t から m 個 ($m \leq |S^t|$) 重複なしでランダムに選び出した集団 S^{*t} を生成する。 t 世代解集団 S^t からルーレット戦略で n 個 ($n \leq |S^t|$) 選び出し ($|D|=n$), 一点交叉させ, 突然変異を加え, 合コン参加者を割り当て, 評価を行う。次に, 全解候補 $d \in D$ と距離 (両者の要素の一致率) が最短の $s^* \in S^{*t}$ を探し, d と s^* の評価値を比較する。評価値が大きい方を $t+1$ 世代解集団 S^{t+1} に代入する。以下に具体的なステップを示す。

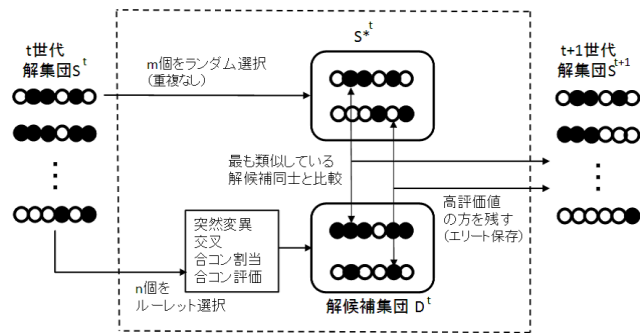


図3 提案選択法のイメージ
Fig. 3 Scheme of proposed selection method.

このステップは図3に対応している。
 ステップ1. S^{t+1} に空の集団を代入
 ステップ2. S^t からランダムかつ重複なしに m 個の要素を選んで S^* を生成
 ステップ3. S^t に対しルーレット選択, 交叉, 突然変異を実行し, n 個の要素からなる D を生成
 ステップ4. すべての $d \in D$ に対して以下を反復
 ステップ4.a. d と最も距離が近い $s^* \in S^*$ を探索
 ステップ4.b. d と s^* の評価値を比較し, d の評価値の方が大きければ S^{t+1} に d を追加. そうでなければ, S^{t+1} に s^* を追加
 ステップ5. t が規定回数に到達していれば終了. そうでなければステップ1へ
 m と n は, $|S^t|$ の8~10割程度の大きさとする. 提案選択法では, 上記のステップ4.bにおいてエリート保存戦略の要素を取り入れ, 高評価値の方のみを残すものとした. この部分は, 差分進化 [16] の目標ベクトルと比較参照ベクトルを比べて良い方を残すという戦略にも通じる.

4.3 合コン割当てオペレータ

合コン割当てオペレータは, 解候補に対し, システム利用者の割当てを行う (合コン割当て名簿の出力). 解候補が好相性を表現しているかどうかを評価するため, 解候補と同じ属性を持つ男女のカップルの成否を確かめる必要がある. しかし, 一般に, 解候補とまったく同じ属性を持つ男女は少ないと予想されるので, できる限り「似ている」属性を持つ男女を割り当てる. 具体的には, 本オペレータは次のステップ1-3を実行する.

- ステップ1: 解候補の前半分 (男性の属性に対応) と利用者の全男性の属性間の距離を計算 (女性についても同様の操作を行う. 解候補の後半分を利用)
- ステップ2: 最短距離の属性を持つ男性1名をその合コン参加者名簿に追加. 利用者集団からその男性を削除 (女性についても同様の操作を実行)
- ステップ3: 合コンの定員が埋まるまで, ステップ1と2を反復 (女性についても同様)

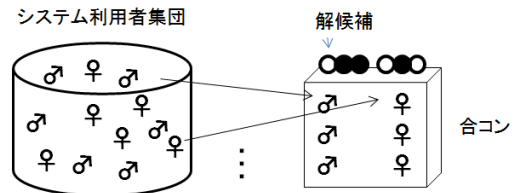


図4 合コン割当てオペレータの動作
Fig. 4 Operation of assignment.

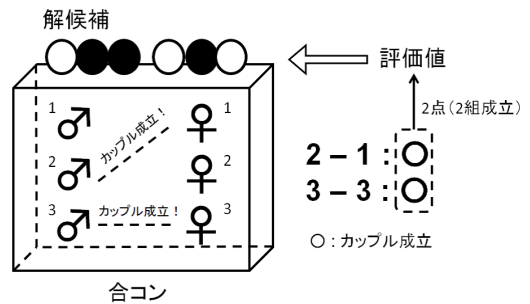


図5 合コン評価オペレータの動作
Fig. 5 Operation of evaluation.

以上の過程により, メンバが確定し合コンが実施可能となる (図4). この割当ての際, ほぼ確実に発生する合コン参加者の属性値とテスト解候補間の距離には注意が必要である. その理由は次節で後述する.

4.4 合コン評価オペレータ

本オペレータは, 合コンを実施し, 解候補に評価値を与えるオペレータである. 評価には, 図5に示すように, その合コンでのカップル成立数を用いる. しかし, その際, 解候補と割り当てられた利用者の属性値間の距離を用いた補正を行う. なぜなら, 本研究では利用者の属性値で相性が決まるという仮定をおり, 解候補が, カップルが成立した男女の属性値と距離が離れていれば, その結果 (カップル成立) が解候補の良さとは異なる (離れる) と考えられるからである. そこで, 解候補と利用者属性値間の距離の指標として, 割当て一致率 (次項で説明) を導入する. また, この割当て一致率を活用し, 合コンを行わずに, 履歴データを使って解候補に評価値を与えるストック利用評価法 (4.4.2項に後述) を導入する.

4.4.1 割当て一致率を用いた評価法

解候補をベクトル $\mathbf{s} = (s_1, \dots, s_{k+l})$ で表す. 男女の組 (b, g) を, b の属性情報 (b_1, b_2, \dots, b_k) と g の属性情報 (g_1, g_2, \dots, g_l) を組み合わせたベクトル $\mathbf{x} = (b_1, b_2, \dots, b_k, g_1, g_2, \dots, g_l) = (x_1, x_2, \dots, x_{k+l})$ で表す. このとき, \mathbf{s} と \mathbf{x} の割当て一致率 $C(\mathbf{s}, \mathbf{x})$ を次のように与える.

$$C(\mathbf{s}, \mathbf{x}) = \frac{\sum_{i=1}^{k+l} \text{match}(s_i, x_i)}{k+l} \quad (2)$$

ここで,

$$\text{match}(s_i, x_i) = \begin{cases} 1 & \text{when } (s_i = x_i) \\ 0 & \text{(otherwise)} \end{cases} \quad (3)$$

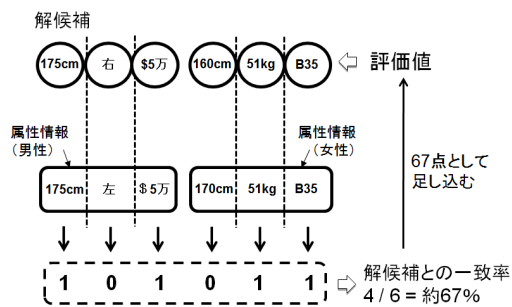
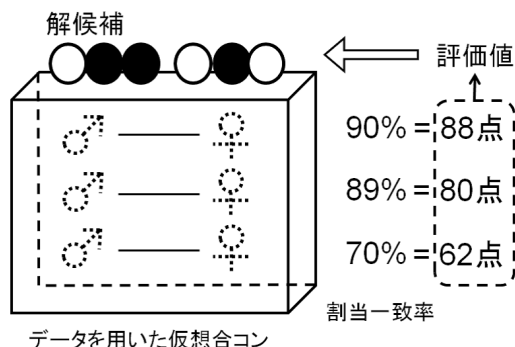


図 6 割当て一致率

Fig. 6 Calculation of matchrate.



データをを用いた仮想合コン

図 8 ストックを利用した評価値の与え方

Fig. 8 Evaluation using stock.

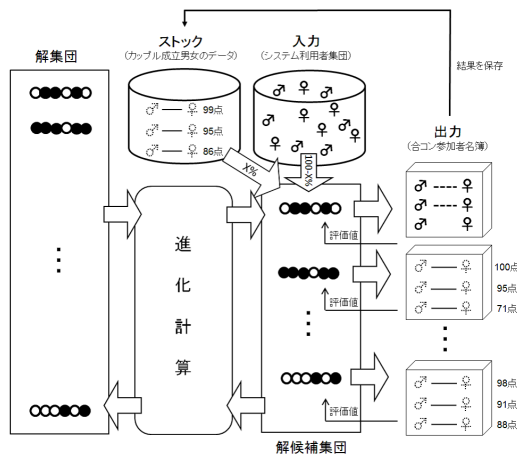


図 7 ストックの導入

Fig. 7 Outline using stock.

図 6 に示すように、その合コンで成立した全カップルの属性情報ベクトルとテスト解候補との割当て一致率の総和を、解候補の評価値として与える。カップル成立数が 0 の場合は評価値を 0 とする。

4.4.2 ストック利用評価法

解候補を評価する際、実際の合コン参加者による評価をなるべく減らしたい。なぜなら、理想的な男女の属性値の組 (属性ベクトルと呼ぶ) が求まるまでは最適な合コンへの割当てが行われなため、参加者はカップル成立数の面で不利になるからである (すなわち、参加者が人柱になる)。利用者を用いずに、解候補の評価を行うことができれば、利用者の時間的、経済的負担、不利益享受を減らすことができる。そこで、利用者の評価履歴データを利用する方法を導入する。

図 7 に示すように、この方法では、まず、カップルが成立した男女の属性ベクトルを保存する。この集団をストックと呼ぶ。図 7 にストックを付け足したシステムの全体像を示す。これ以降、提案システムとは、図 1 ではなく、図 7 を指すものとする。

解候補を評価する際、まず、合コン割当てオペレータの全適用対象の $X\%$ をストックから取り出して割り当てる (システム利用者集団からの割当ては $100\% - X\%$)。 X は、ストックが十分蓄積された段階で 50~100% の間に設定する。

解候補に距離が近い属性ベクトルを持つストック内の男女 (ストック男女と呼ぶ) を、距離に近い順に規定人数割り当てする。これより、図 8 に示すように、解候補とストック男女 (の属性ベクトル) の割当て一致率を計算し、その総和を解候補の評価値として与える。ストックから解候補 s に距離が近い順に L 個取り出したストック男女の集団を O 、ストック男女を $o \in O$ 、 o の s に対する割当て一致率 $C(s, o)$ として、 s の評価値 E_s の計算式を式 (4) に示す。

$$E_s = \sum_{o \in O} C(s, o) \tag{4}$$

5. 実験

提案手法が、最適解集団にどれだけ近い解集団を求めることができるか (最適解到達率, 5.2 節に後述)、また、得られた解集団を何組のカップルを成立させることができるかを、以下の実験 1 と実験 2 の場合において計算機シミュレーションを用いて調べた。実験 2 では、貪欲手法 (5.3 節) との比較を行った。また、実験 3 では、利用者情報量と問題サイズに対する提案手法のスケラビリティを調べた。

実験 1: 解集団の多様性を保つ提案選択法 (4.2.1 項) の有無

実験 2: ストック利用法 (4.4.2 項) の有無

実験 3: 入力情報量 (4.1 節) と問題 (5.1 節) サイズの変更ベンチマークテストとして N 個のピークを持つ多峰性関数である $NMax$ 問題 (次節で説明) を導入した。

5.1 $NMax$ 問題 (ベンチマークテスト)

4.4.1 項で定義した任意の男女の属性ベクトルを表すベクトル \mathbf{x} と好相性の 1 つを表す最適解 \mathbf{p} との一致率を返す関数 $lovematch(\mathbf{x}, \mathbf{p})$ を以下のように与える。

$$lovematch(\mathbf{x}, \mathbf{p}) = \frac{\sum_{i=1}^{|\mathbf{x}|} g(i)}{|\mathbf{x}|} \tag{5}$$

$$g(i) = \begin{cases} 1 & (x_i = p_i) \\ 0 & (\text{otherwise}) \end{cases} \tag{6}$$

ここで i は、ベクトルの i 番目の要素を表す. $NMax$ 問題では N 個の最適解 p^1, \dots, p^N が存在すると定義する. $NMax$ 問題を, 任意の男女の相性の良さを返す関数として, 以下のように与える.

$$f_{NMax}(\mathbf{x}) = \max_{1 \leq j \leq N} (\text{lovematch}(\mathbf{x}, \mathbf{p}^j)) \quad (7)$$

j は $NMax$ 問題の N 個のピークのインデックスである. ビット列 \mathbf{x} と, 各ピークのビット列の間で一致率が計算され, その最大値が f_{NMax} の出力値となる. $NMax$ 問題は, 進化計算のベンチマークとしてよく知られる $OneMax$ 問題 [17] を一般化したもので, $N = 1, \mathbf{p}^1 = \{1, 1, \dots, 1\}$ のとき, $OneMax$ 問題と同値となる.

本実験では, カップル成否を返す評価関数として $NMax$ 問題を与えたい. しかし, $NMax$ 問題の出力は, 好相性との一致度であり, 確率的な要素が含まれていない. 実際の合コンでは, 相性が良くてもカップルにならない場合や, 多少悪くてもなる場合がある. そうした点を, 本実験では評価関数に組み込みたい. そこで, 本実験では, カップル成立有無の評価関数をカップル成立確率 P として, f_{NMax} を用い, 式 (8) として与えた.

$$P = \begin{cases} 1 - \frac{1-f_{NMax}}{1-\beta} & (\beta \leq f_{NMax}) \\ 0 & (\text{otherwise}) \end{cases} \quad (8)$$

β はカップル成立の閾値であり, f_{NMax} の出力が, β 未満のときはカップル成立確率 P は 0 になり, β 以上のときは P が線形に増加するよう設定した.

5.2 最適解到達率

得られた解集団を評価するため, 解が, 最適解にどれくらい到達できているかを示す指標として最適解到達率を定義する. 解 \mathbf{s} の最適解到達率 R は, 最適解集団 \mathbf{p} (図 11 に示す) と, 式 (3) の $match$ 関数を用いて, 以下の式で与える.

$$R = \max_{1 \leq i \leq n} (\text{match}(\mathbf{s}, \mathbf{p}^i)) \quad (9)$$

獲得した解集団がどれくらい相性空間を推定できているかは, この解集団の解の最適解到達率の平均値によって確認することができる.

5.3 比較手法

提案手法と比較するため, 次のようなグリーディな手法を考える. これは, 図 9 に示すように, 合コンのたびに, その合コンでカップルが成立した男女の属性ベクトルの中で, その解候補に最も類似する属性ベクトルを, 解候補に上書きする方法である. 良い組合せが見つかった場合, それがそのまま解として次世代に受け継がれる単純な探索手法である.

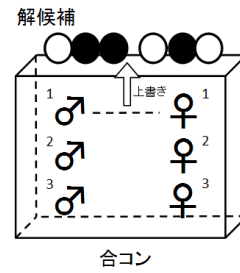


図 9 貪欲手法

Fig. 9 Greedy method.



1010111110110101010101010111

図 10 入力データの簡単化

Fig. 10 Simplification of input data.

表 2 設定値

Table 2 Experimental setup.

合コン	
利用者集団サイズ	6,000 名 ($ B = 300, G = 300$)
利用者属性長 ($ b_{ik} , g_{jl} $)	10 ビット
利用者調整理想度*	0~100% (一様乱数)
合コンサイズ	30 名 ($ B_i = 15, G_i = 15$)
ストックサイズ	500 レコード
カップル成立閾値	0.5
最低参加回数*	5 回
進化計算	
解集団サイズ	30 個
解長	20 ビット
選択個数 m	10 個
交叉率	0.95
突然変異率	0.2
スケーリング	1 : 1 線形

5.4 実験設定

本実験では, まず, 入力データのビット列への簡単化を行った. 概要を図 10 に示す. 入力であるシステム利用者の属性情報をビット列で与える. 既存の結婚情報サービスの利用者情報はすべてビット列に符号化可能である. 本実験では, 相性の判定に必要な情報がビット列化できることを前提に, アルゴリズムがうまく動作するかを調べる. ただし, 情報の種類によっては符号化によってビット数が増大するものもあり, これらの短縮には, 相性に影響する要因の特定と 4.1 節で述べたような工夫が必要である. 実験パラメータの設定値を表 2 に示す.

本実験では, 図 11 の 5 つの好相性に近い男女が, 式 (8) に従いカップルが成立するという設定で実験を行った. そのため, 利用者の属性値を一様乱数で生成すると, カップル成立数が極端に減り (付録 A.1 参照), 探索が進まない.

```
Peak1 00000000000000000000
Peak2 00000000011111111111
Peak3 11111000001111100000
Peak4 11111111110000000000
Peak5 11111111111111111111
```

図 11 $NMax$ 問題 ($N = 5$)

Fig. 11 $NMax$ problem ($N = 5$).

したがって、本実験では、図 11 の N 個の各ピークとの類似度（利用者調整理想度）がある程度高くなるよう利用者の属性情報を生成することにした。 $NMax$ 問題のピークを利用者の属性値にコピーし、ランダムに 1~100 の間から数字 θ を選び、コピーした利用者の属性値 $\theta\%$ をビット反転させることで生成した。これを各ピークに対して順番に、利用者集団が埋まるまで繰り返した。

また、カップルが成立した利用者を逐次、新しい利用者に入れ換える設定とした。ただし、その利用者が保持する参加回数が最低参加回数を超えた場合のみとした。新しい利用者も利用者調整理想度のルールに従って生成する。参加回数は、全利用者がそれぞれ保持し、初期値は 0、合コン参加によりカップルが成立した場合に 1 ずつ加算される。参加回数が最低参加回数を超えると、利用者は利用者集団から排除される。実際のお見合いパーティでも、カップリングが成功（カップリングゲームで両思いが成立）しても、もっと良い相手を探して、何回もパーティに参加し続ける人が多い（最初から何回か参加することを前提としている人も多い）。

合コン問題を、EC 世代ごとに繰り返し解くことを、繰返し合コン問題と定義する。繰返し合コン問題は、世代ごとのカップル成立の累積値を最大化することが目的関数となる。 F の正確な推定（高い最適解達成率に到達）は、合コン問題はもちろん繰返し合コン問題の解決に貢献する。実験 3 では、利用者属性長は 4.1 節での考察に基づき現在の 2 倍の長さ（解長も倍となる）までを調べた。同じく実験 3 の N に関しては 12 個程度までとした。 N の値は、4.1 節で述べた性格パターン数の 2 乗（男女の組合せのため掛け算）で 25 程度を考えてもよいが、組合せには、好相性ばかりではなく、相性が良くない組合せも存在すると考えられる。したがって、25 の約半分の値とした。

5.5 結果

図 12 と図 13 に、実験 1 の結果として、提案選択法を用いた場合と用いなかった場合の、1,000 EC 世代後に得られた解集団の平均最適解到達率 \bar{R} を示す。初期集団をランダムに変更し 30 回実験を行ったところ、 \bar{R} の値はそれぞれ変わらず、標準偏差値は 1% であった。これにより提案選択法が \bar{R} を約 27% 高めることに貢献していることが分かる。提案選択法が導入されていない図 13 の獲得解集団の例では、5 つのピークのうち 1 つが発見できておらず、また \bar{R} も低い。

```
11101111111011111111 peak.5 90%
1111111111000001010 peak.4 90%
11111111110000000000 peak.4 100%
111111111111111101111 peak.5 95%
0000000000001000010 peak.1 90%
000000000100000001 peak.1 90%
1011111111110101111 peak.5 85%
000000000010000000 peak.1 95%
0010000000000000000 peak.1 95%
111110111000010000 peak.4 90%
1111110110010010100 peak.4 80%
0000000001111111111 peak.2 100%
1000010000000000000 peak.1 90%
1111000001111010000 peak.3 95%
1111111111111110101 peak.5 90%
000000000100000000 peak.1 95%
1101011111000010000 peak.4 85%
1101111010000010000 peak.4 85%
0000010100000000000 peak.1 90%
0010001000100000000 peak.1 85%
```

図 12 提案選択法を導入した場合 ($\bar{R} = 91\%$)

Fig. 12 Using proposed selection ($\bar{R} = 91\%$).

```
10011011110011101010 peak.4 60%
11011110010000111000 peak.4 70%
11101010010001100110 peak.3 60%
11011001100011111110 peak.5 65%
1101101010011110010 peak.3 60%
11111001111001111011 peak.5 75%
11011110111011101010 peak.5 70%
00011010010010111010 peak.1 55%
1101101011000101010 peak.4 65%
11001010110101011011 peak.5 60%
11001010110101001011 peak.4 55%
11111001011001100001 peak.3 75%
10101000111001010110 peak.1 50%
1110101001101111010 peak.3 65%
111000001000110010 peak.1 60%
11010010111011100110 peak.3 60%
1100101001101100011 peak.3 65%
11111010110001111110 peak.5 70%
11110001100001110010 peak.3 60%
11101001011011111011 peak.5 70%
```

図 13 提案選択法を導入しなかった場合 ($\bar{R} = 64\%$)

Fig. 13 Not using proposed selection ($\bar{R} = 64\%$).

実験 2 の結果を表 3 に示す。 \bar{R} およびカップル成立数は、合コン回数 10,000 回（1,000 世代）後に得られた解集団における値と、その解集団を用いて合コンを行った場合で得られた値であり、初期集団をランダムに変更し行った 30 回の試行の平均値と標準偏差値で示している。しかし、表内 3 行目はストックを 50% 利用した場合で、実際の合コン実施回数は約半分の 5,250 回後の値となっている。表内 4 行目も 50% ストックを利用した場合であるが、実際の合コン回数が 1,000 回に達するまで（1,950 世代）実験を行った結果である。カップル成立数は、それぞれ得られた解集団に対し合コン割当てオペレータを適用して得られた値である。これらの結果より、提案システムが、比較手法に比べて、約 22% 以上高い最適解到達率を求めることができ、約 2 倍のカップル成立数が多い合コンの実施に、約半分の合コン回数で至ることが確認できた。

実験 3 の結果を図 14, 図 15 に示す。いずれも縦軸に \bar{R} の平均、横軸はそれぞれ利用者属性長と $NMax$ 問題の N である。実験 2 と同じく初期集団をランダムに変更した 30 回の試行の平均値をプロットした。標準偏差値は実験 3 の結果のすべての値において 1 であった。これは図中には記載していない。これより、提案手法が、現在の倍のスケールの入力情報量に対し、 \bar{R} を約 10% しか劣化させない

表 3 提案手法と貪欲手法の比較

Table 3 Comparison of proposed and greedy method.

	最適解到達率 \bar{R}	カップル成立数	実合コン回数
最適解※	100 ± 0%	4,494 ± 6 組	0 回
提案手法 (ストック利用率 0%)	91 ± 1%	3,440 ± 112 組	10,000 回
提案手法 (ストック利用率約 50%)	90 ± 1%	3,358 ± 132 組	5,250 回
提案手法 (ストック利用率約 50%)	91 ± 1%	3,457 ± 125 組	10,000 回
貪欲手法	64 ± 2%	1,456 ± 106 組	10,000 回

※ 最適解集団 (図 11) を解集団に与えた場合

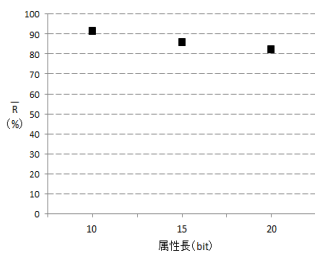


図 14 利用者属性長に対するスケラビリティ ($N = 5$ のとき)

Fig. 14 Length of attribute scalability (when $N = 5$).

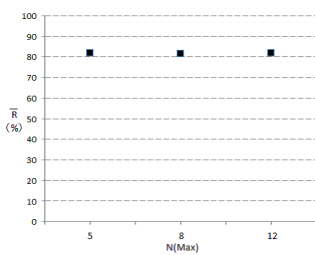


図 15 N に対するスケラビリティ (属性長 = 20 のとき)

Fig. 15 N scalability (when length of attribute = 20).

ことが分かった。また、 N の値が 2 倍以上になっても、ほぼ \bar{R} が変わらないことも分かった。ただし、図 15 においては $N = 8$ の場合、2 回 (30 回中) の試行で 1 個 (8 個中) のピークが未発見、 $N = 12$ の場合、14 回 (30 回中) の試行で 1 個 (12 個中) のピークが未発見であった。

5.6 考察

提案手法が、最適解集団を与えた場合のカップル成立に及ばなかった理由としては、カップル成否の評価関数を確率で与えているため、解候補の評価値が変動し、たまたま良い値をつけた解候補が解集団に残ることが考えられる。とくに、カップル成立閾値が小さい場合、この問題が顕著になるはずである (偶然、カップルになる数が増加するから)。

また、利用者調整理想度が低く、最適解から遠い属性を持つ利用者ばかりの場合も、カップルが成立しにくいことに加え、成立しても偶然の確率が高いため、最適解到達率は下がる。カップル成立確率をどのように与えるかによっても変わる点に注意が必要である。

ストックの利用方法の改善について、今後、検討が必要

である。今回の実験では、ストック利用の有無で結果に差は出ていないが、実世界の問題において、ストックに保存されていないような新しい好相性の解候補が出てきた場合、正しい評価値を与えることができない (類似の属性ベクトルを割り当てることができないため) 可能性がある。したがって、ストックに保存する属性ベクトル群の多様性を保つ工夫が求められる。また、このストックに、過去蓄積されてきたデータを初期データとして入れることができれば、収束までの時間も短縮できる。提案手法を、より大きなスケールで有効に作用させるためには、より豊富で大きなシステム利用者集団と、より大きな解集団サイズが必要と考えられる。

6. おわりに

本論文では、実際に実施されている合コンセッティングサービスで、EC によって複数の未知の相性を予測し、カップル成立数を増やすシステムを提案した。計算機実験を用いて、最適解に平均 86% 一致する解集団を求めることができた。また、一般的に考える単純な貪欲手法に比べて、半分の実合コン実施数で、約 2 倍のカップルを誕生させることが可能なことを示した。

今後の主要な課題は、実用までのギャップの解消である。今回の結果を得るには 5,000 から 10,000 回の実合コンの実施が必要であるが、この回数は、既存の実合コンセッティングサービスにおいては現実的な回数である。たとえばツヴァイ社 [18] は年間約 2,000 回、日本サブライズ社 [19] は年間約 1,200 回の実合コンを実施しており、年間約 15,000 回の実合コン回数を報告している業者 [20] も存在する。これら数社のデータを統合することができれば、数カ月から数年というオーダーでの提案システムの適用が現実的となってくる。

「合コン」問題のような非線形問題にはニューラルネットワーク (NN, Neural Network) の適用が考えられるが、NN の素子数やネットワーク構造の決定は難しい問題であり、また、学習後の NN から好相性を理解することが困難である場合が多い [21]。そのため今回は採用を見合わせた。他方、進化計算であれば、解集合に対してクラスタリング手法を適用したり、統計的解析をかけたりすることで、好相性を抽出できる可能性が高い。筆者は NN のネットワーク構造を進化計算によって最適化する NEAT [22]

分野での研究を行ってきており [23], 今後はこれを「合コン」問題に適用したいと考えている.

その他の検討課題としては, 解集団の多様性を保つ手法として有名なシェアリング [24] や資源共有法 [25] などと提案選択法の性能比較が考えられる. さらに, 属性情報に欠損がある (利用者がアンケートの一部に未回答) 場合の対策や, 評価関数 F が変化した場合の適応性の強化などもあげられる.

参考文献

[1] 国立社会保障・人口問題研究所: 人口統計資料集 (2011). 入手先 (<http://www.ipss.go.jp/syoushika/>).

[2] 内閣府: 少子化社会白書 (2004). 入手先 (<http://www8.cao.go.jp/shoushi/whitepaper/w-2004/html-g/indexg.html>).

[3] 内閣府少子化対策室: 子ども・子育て白書 (2011). 入手先 (<http://www8.cao.go.jp/shoushi/whitepaper/index-w.html>).

[4] Foster, J.A.: Evolutionary Computation, *Nature Rev. Genet.*, Vol.2, pp.428-436 (2001).

[5] Kuroiwa, S., Murata, Y., Kitani, T., Yasumoto, K. and Ito, M.: A Method for Assigning Men and Women with Good Affinity to Matchmaking Parties through Interactive Evolutionary Computation, *Proc. Simulated Evolution and Learning*, pp.645-655 (2008).

[6] Gale, D. and Shapley, L.S.: College Admissions and the Stability of Marriage, *American Mathematical Monthly*, Vol.69, pp.9-15 (1962).

[7] Julia, H. and Joshua K.: Evolutionary Multiobjective Clustering, *Trans. IEEE Evolutionary Computation*, Vol.11, No.1, pp.1081-1091 (2004).

[8] 石淵久生, 半田久志: 進化計算の内外の研究動向, 進化計算学会論文誌, Vol.1, No.1, pp.15-22 (2010).

[9] 伊藤 一: N700 系新幹線車両の概要—遺伝的アルゴリズムを用いた先頭車両形状設計, 計測自動制御学会北陸支部講演会 (2009).

[10] Takagi, H.: Interactive Evolutionary Computation: Fusion of the Capacities of EC Optimization and Human Evaluation, *Proc. IEEE*, Vol.89, No.9, pp.1275-1296 (2001).

[11] Tanji, M. and Iba, H.: ConBreO: A music performance rendering system using hybrid approach of EC and automated evolution, *Proc. GECCO*, pp.1275-1282 (2010).

[12] Deb, K., Agrawal, S., Pratap, S. and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II, *Trans. IEEE Evolutionary Computation*, Vol.6, No.2, pp.182-197 (2002).

[13] Jacob, S., McClintock, M.K., Zelano, B. and Ober, C.: Paternally inherited HLA alleles are associated with women's choice of male odor, *Nature Genetics*, Vol.30, pp.175-179 (2002).

[14] Cloninger, C.R., Svrakic, D.M. and Przybeck, T.R.: A Psychobiological Model of Temperament and Character, *Archives of General Psychiatry*, Vol.50, pp.975-990 (1993).

[15] 宗像恒次, 田中京子, 小林由美: SAT 気質コーチングによる人間関係のコラボレーション, ヘルスカウンセリング学会年報, Vol.13, pp.1-11 (2007).

[16] Storn, R and Price, K.: Differential Evolution — A Simple and Efficient Heuristic for global Optimization over Continuous Spaces, *J. Global Optimization*, Vol.11,

pp.341-359 (1997).

[17] Schaffer, J.D. and Eshelman, L.J.: On Crossover as an Evolutionarily Viable Strategy, *Proc. Int'l Conf. Genetic Algorithms*, pp.61-68 (1991).

[18] 株式会社ツヴァイ. <http://www.zwei.com/>

[19] 株式会社日本サブライズ社. <http://www.surprise-japan.co.jp/>

[20] 株式会社エクシオジャパン. <http://www.exeojapan.com/>

[21] 石川眞澄: ソフトコンピューティングと情報統合, システム制御情報学会誌, Vol.43, No.4, pp.174-179 (1999).

[22] Stanley, K. and Miikkulainen, R.: Evolving Neural Networks Through Augmenting Topologies, *Evolutionary Computation*, Vol.10, pp.99-127 (2002).

[23] 黒岩 将, 吉田 久: コンピュータシミュレーションによる NGL 獲得構造 NN の学習能力に関する比較, 電子情報通信学会総合大会 (2006).

[24] Horn, J., Nafpliotis, N. and Goldberg, D.E.: A niched Pareto genetic algorithm for multiobjective optimization, *Proc. Int'l Conf. Evolutionary Computation*, pp.82-87 (1994).

[25] Juill, H. and Pollack, J.B.: Coevolutionary Learning: A Case Study, *Proc. Int'l Conf. Machine Learning*, pp.251-259 (1998).

付 録

A.1 乱数で利用者を生成する際, 解候補長が増えると好相性にマッチする利用者が極端に減る理由

一様乱数を用いて, n ビットの任意のベクトル \mathbf{p} に $\alpha\%$ 一致する \mathbf{x} を生成できる確率 $P(n, \alpha)$ は次式で与えられる.

$$P(n, \alpha) = {}_n C_{\alpha n} \left(\frac{1}{2}\right)^{\alpha n} \left(1 - \frac{1}{2}\right)^{[n(1-\alpha)]} \quad (A.1)$$

$P(5, 0.8) = 0.156$, $P(15, 0.8) = 0.014$, $P(25, 0.8) = 1.58 \times 10^{-3}$ と指数的に減少する.



黒岩 将

2004 年近畿大学生物理工学部生物工学科卒業. 2008 年奈良先端科学技術大学院大学情報科学研究科修士課程修了. 同年同博士課程進学, 進化計算の応用に関する研究に従事. 同年起業し, 結婚紹介業に取り組みが, 2012 年退社, 譲渡. 著書に『出稼げば大富豪』(KK ロングセラーズ) シリーズ, 『大富豪のオキテ』(日本文芸社) 等.



安本 慶一 (正会員)

1991年大阪大学基礎工学部情報工学科卒業。1995年同大学大学院博士後期課程退学後、滋賀大学経済学部助手。2002年奈良先端科学技術大学院大学情報科学研究科助教授、2011年より同研究科教授。博士(工学)。モバイルコンピューティング、ユビキタスコンピューティングに関する研究に従事。電子情報通信学会、ACM、IEEE各会員。



村田 佳洋 (正会員)

2003年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。同年奈良先端科学技術大学院大学情報科学研究科助手。現在、広島市立大学情報科学研究科准教授。遺伝的アルゴリズム、エージェント技術等の研究に従事。



伊藤 実 (正会員)

1977年大阪大学基礎工学部卒業、1979年同大学大学院基礎工学研究科博士前期課程修了。1979年大阪大学基礎工学部助手、1986年同講師、1989年同助教授を経て、1993年より現在、奈良先端科学技術大学院大学情報科学研究科教授。工学博士。データベース理論、効率的なアルゴリズム開発等の研究に従事。ACM、IEEE、電子情報通信学会各会員。