

修士論文

対話型遺伝的アルゴリズムのための商品間  
の推薦関係を用いた設計変数空間の生成

同志社大学大学院 工学研究科 知識工学専攻  
博士前期課程 2008年度 762番

田中 美里

指導教授 三木 光範 教授

2010年1月23日

## Abstract

Recently, online shopping sites have been attracting more users than conventional physical stores. Online shopping sites can provide vendors with increased sales opportunities and present tremendous choices for consumers. Product recommendation is a technique that can present products that may be bought by users. In this technique, information based on a user's preference is utilized. Many of the shopping sites, such as Amazon<sup>a</sup>, adopt product recommendation schemes, such as Collaborative Filtering<sup>1,2)</sup> and support vector machine (SVM),<sup>3)</sup> to lead users to products purchased by other users with similar preferences and encourage their purchase.

iGA (interactive Genetic Algorithm),<sup>4)</sup> which is an interactive evolutionary computational method, is expected to be useful for recommending products according to users' preferences through communication between systems and users. The iGA is an algorithm derived from GAs (Genetic Algorithms).<sup>5)</sup> The evaluation operation in GAs is replaced with the user's subjective preference. Ideally, iGA in a recommendation system will obtain the users' preference from their browsing history on the online shopping site, search for products matching these characteristics, and improve the personal showcase.

When iGA is applied to shopping site system, phenotype is a product and genotype is a design variable and the mapping between genotype and phenotype should be prepared. However, since there are a wide variety of products on shopping sites that are updated rapidly, extracting mapping relation from existing products into design variables by hand is unrealistic. To address this problem, a method to generate design variables from products automatically is proposed.

To generate the design variables automatically for iGA, the relevance of the products based on collective preference accumulated on the web is applied. The proposed method obtains the adjacency matrix from the recommendation relations among the products on web, computes the genes by Principal Components Analysis and optimizes the product recommendation by iGA. Through the experiments, it was described that the distributions of books generated by the proposed method were influenced by their authors. The subjective experiments of finding the favorite books for users were performed and it was confirmed that subjects searched solutions by their unique preference. It was concluded that the proposed method is very useful for recommendation system using iGA and the produced design variable space is appropriate for iGA search.

---

<sup>a</sup><http://amazon.com/>

# 目次

<b>1</b>	<b>序論</b>	<b>1</b>
<b>2</b>	<b>対話型遺伝的アルゴリズム</b>	<b>1</b>
2.1	対話型進化計算法	2
2.2	遺伝的アルゴリズムの概要	2
2.3	対話型遺伝的アルゴリズム	3
<b>3</b>	<b>商品間の推薦関係に基づく対話型遺伝的アルゴリズムのための設計変数空間の導出</b>	<b>6</b>
3.1	対話型遺伝的アルゴリズムの商品推薦への適用と課題	6
3.2	対話型遺伝的アルゴリズムの商品推薦への適用における課題	7
3.3	商品間の推薦関係に基づく対話型遺伝的アルゴリズムのための設計変数空間の生成	9
<b>4</b>	<b>商品間の推薦関係に基づく設計変数空間の特徴の検討</b>	<b>11</b>
4.1	設計変数空間の生成パラメータ	12
4.2	商品の分布と属性値の関係性に関する検討	13
4.3	次元数の検討	16
<b>5</b>	<b>商品間の推薦関係に基づく設計変数空間上における対話型遺伝的アルゴリズムの検討</b>	<b>17</b>
5.1	実験目的	17
5.2	実験システム	17
5.3	実験計画	20
5.4	結果と考察	21
<b>6</b>	<b>結論</b>	<b>24</b>

# 1 序論

近年、Web上のオンラインショッピングサイトに対する需要が増加している。オンラインショッピングサイトは現実の店舗よりも多くの商品を提示できるため、売り手にはより多くの商売の機会を、客に対しては幅広い選択肢を提供することができる。Amazonなど多くのショッピングサイトでは、嗜好の似た他のユーザの購入した商品へとユーザを誘導し、購入を勧めるために、協調フィルタリング<sup>1,2)</sup>やサポートベクターマシン<sup>3)</sup>、コンテンツフィルタリングを用いた商品推薦を採用している。

本研究では、この商品推薦アルゴリズムとして、対話型遺伝的アルゴリズム<sup>4)</sup>(iGA)を適用する事を考える。iGAは人間とコンピュータとの相互作用、および人間の主観的評価に基づいて最適化を行う対話型進化計算法(Interactive Evolutionary Computing: IEC)の一つであり、GA<sup>5)</sup>の適合度関数に変わり、ユーザが個体の評価を行う。そのため、人の感性という複雑な構造を解析する方法として、定量的な評価が困難な楽曲やデザインなどの生成に多く適用されている<sup>6,7)</sup>。

推薦アルゴリズムとしては、iGAはコンテンツフィルタリングに分類される。従来のコンテンツフィルタリングでは、商品の特徴値とユーザのプロファイルの適合度に基づいて推薦を行う。iGAはユーザの主観的な情報をインタラクティブに取得、解析することで、ユーザの固定的なプロファイルだけでなく、ユーザの時々々の感性を反映した推薦が行えるものと考えられる。

商品推薦をiGAの対象問題とするにあたり、商品推薦を最適化問題の枠組みにモデリングする必要がある。具体的には、ショッピングサイトで扱われる大量の商品の特徴を、各商品の設計変数として数値化する必要がある。これには大変な入力コストがかかる。また、多様な商品を表現するのにふさわしい設計変数の特定、各設計変数の近傍の設計が必要となるなど、ベンダ側の負担が大きい。

そこで本研究では、Web上から商品同士の推薦関係を取得し、そこから設計変数空間を生成する手法を提案する。各商品の設計変数を自動的に算出することで、ベンダ側のiGAシステム運用のコストを低減し、感性的な商品の探索を可能とすることを目的とする。

本稿では書籍を対象問題とし、Amazonから推薦関係を含む多数の商品の情報を取得し、主成分分析を用いて複数の設計変数空間を生成した。また、生成した設計変数空間を用いたiGAシステムを構築し、被験者実験を行うことで生成した設計変数空間上において各被験者の嗜好がどのように表れるかについて検証する。

以下、第2章ではiGAの概要について説明した。第3章ではiGAの商品推薦への適用について述べ、その要件と課題について述べた上で、提案手法であるWeb上の商品の推薦関係から商品の設計変数を生成する手法について簡単に述べた。第4章では提案手法中で用いる主成分分析について説明する。第5章では提案手法の詳細な手順と、生成された設計変数空間の特徴について検討を行った。第6章では、生成した設計変数空間を利用したiGAシステムについて説明し、被験者実験の手順と結果について述べた。そして、第7章において本論文の結論を述べる。

## 2 対話型遺伝的アルゴリズム

本章では、対話型進化計算法、遺伝的アルゴリズム、および対話型進化計算法の1つである対話型遺伝的アルゴリズムについて述べる。

## 2.1 対話型進化計算法

知能科学，特にソフトコンピューティングと呼ばれる分野では，1980年代にニューラルネットワーク (Neural Network: NN)，ファジィシステム，および進化計算 (Evolutionary Computation: EC) などの技術が成長し，それぞれの技術の融合，協調モデルが提案され，1990年代に実用化へと広がっていった。

ECは最適化手法の1つである。最適化とは，ある制約条件の下で目的とする関数の最小値（最大値）とその設計変数を求めることであり，遺伝的アルゴリズム (Genetic Algorithm: GA)，遺伝的プログラミング (Genetic Programming: GP)，進化戦略 (Evolutionary Strategies: ES)，および進化的プログラミング (Evolutionary Programming: EP) などがある。しかし，最適化問題にはデザイン性など定量的に最小化（最大化）できる目的関数を持たない問題も存在する。定量的に最小化（最大化）できない対象を最適化する手法として対話型進化計算法 (Interactive Evolutionary Computing: IEC) がある<sup>4,6,7)</sup>。IECでは，ECの評価部分に人間の主観的評価を用いることで，感性をシステムに組み込むことができるとされている。IECシステムの概念図を Fig. 2.1 に示す。本研究ではIECの最適化系としてGAを利用した対話型遺伝的アルゴリズム (Interactive Genetic Algorithm: iGA) を用いる。GAについては2.2節，iGAについては2.3節で述べる。

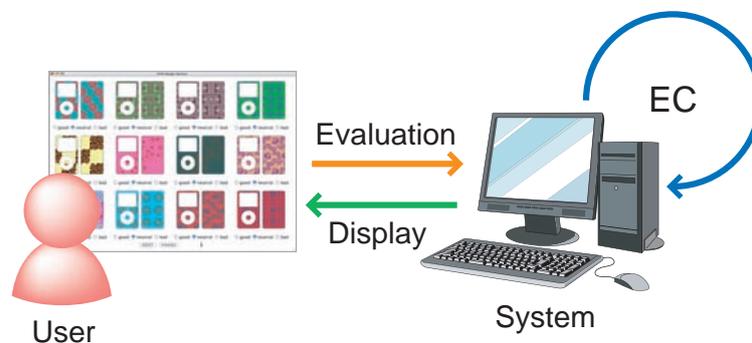


Fig. 2.1 IEC system

## 2.2 遺伝的アルゴリズムの概要

GAは生物の進化の過程を工学的に模倣した確率的な最適化手法であり，対象とする問題を限定せず，モデル化によって多様な問題に適応可能なメタヒューリスティクスである<sup>8)</sup>。

GAではある世代 (Generation) を形成している個体 (Individual) の集合を母集団 (Population) と呼ぶ。また，GAでは自然界の進化過程と同様に環境への適合度 (Fitness) の高い個体が高い確率で選択 (Selection) される。そして，その個体に対して交叉 (Crossover)，および突然変異 (Mutation) が確率的に発生することにより次世代の母集団が形成され，最後に得られた母集団の中で最も適合度の高い個体を最適解とする。

個体は染色体 (Chromosome) によって特徴付けられており，染色体は複数の遺伝子 (Gene) で構成されている。通常GAでは1染色体で1個体を表現する。GAで扱う情報は，表現型 (Phenotype) と遺伝子型 (Genotype) の2つがあり，表現型は個体の形質や特性，遺伝子型は染色体の構造を表す。表

現型によって表される個体の特徴は設計変数とも呼ばれ、また、表現型によって表された個体の分布する空間を設計変数空間という。表現型から遺伝子型へ写像することをコード化 (Coding)、遺伝子型から表現型へ逆写像することをデコード化 (Decoding) という。

遺伝子型が2進数のビット {0,1} や実数値によって表され、前者をビットストリング GA と、後者を実数値 GA と呼ぶ。表現型によって適合度が決まり、適合度の高い個体ほど子孫を残しやすく、低い個体ほど死滅しやすいようになっている。このことにより、次世代の各個体の適合度が前世代より良いことが期待される。

## 2.3 対話型遺伝的アルゴリズム

### 2.3.1 対話型遺伝的アルゴリズムの概要

iGA は、GA における遺伝的操作をベースとして、提示された個体を人間の主観に基づいて評価し、インタラクティブに最適化を行うアルゴリズムである<sup>6)</sup>。GA における評価系に人間の評価を用いて解の探索を行うため、人の感性という複雑な構造を解析する方法として、定量的な評価が困難な楽曲やデザインなどの生成に多く適用されている<sup>6,7)</sup>。

### 2.3.2 対話型遺伝的アルゴリズムの基本動作

iGA では、GA と同様に選択、交叉、および突然変異を行うが、個体の評価を人間が行うため、解の提示、および評価が異なる。また、終了条件は人間の判断により終了する。iGA における基本動作のフローチャートを Fig. 2.2 に示す。また、以下にアルゴリズムの概要を示す。

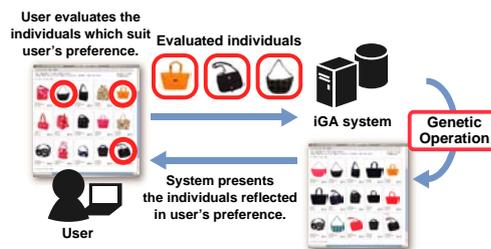


Fig. 2.2 Flowchart of IGA

#### (1) 初期化 (Initialization)

あらかじめ設定された数だけの個体を生成する。生成される個体の数を母集団サイズ (Population size)、あるいは個体数と呼ぶ。個体は通常、何らかの初期情報がない限り一様乱数を用いて生成する。

#### (2) 提示 (Display)

インタフェースを通して提示された個体を、ユーザが評価する。

#### (3) 評価 (Evaluation)

提示された個体に対して、ユーザが主観に基づいて評価し、適合度を与える。iGA でいう適合度とは、提示された個体が、ユーザの主観的評価基準にどの程度沿っているかを数値化したも

のである。評価方法はインタフェースに依存し、全ての個体に対し 100 点満点で点数をつけるインタフェース、5 段階で評価を行うインタフェース、あるいは良い・悪いといった 2 段階の評価を行うインタフェースなどがある。100 点満点などの細かい評価を行うよりも、5 段階評価など粗い評価の方がユーザにとって容易であることが報告されている<sup>9)</sup>。

#### (4) 選択 (Selection)

生物の適者生存を模倣したものである。適合度に依存した一定の規則に従い、次世代に残す個体を選択する。選択により、適合度の低い幾つかの個体は淘汰され、その個体数だけ適合度の高い個体が増殖する。代表的な選択の方法として、トーナメント選択 (Tournament selection)、およびルーレット選択 (Roulette selection) などがある。トーナメント選択は、集団の中からあらかじめ定められた数だけの個体をランダムに選出し、その中で最も適合度の高い個体を選択するという操作を、設定された個体数が選ばれるまで重複を許しながら繰り返す方法である。ルーレット選択については適合度の高い個体ほど選択される確率を上げる方法であり、適合度の低い個体も選択される可能性を残す。この他に、適合度の高い個体を無条件に次世代に残すエリート保存戦略 (Elitism) も選択の 1 つとして考えられる。

#### (5) 交叉 (Crossover)

生物の有性生殖を模倣したものである。交叉の目的は、親個体の優れた部分形質を子個体に継承することである。交叉では、親個体間で染色体情報を交換し、新しい子個体を生成する。適合度の高い個体同士が交叉すると、それぞれの個体の優れた部分解が結合し、より適合度の高い個体が生成されるものと期待される。母集団のうち何割の個体が交叉するかは、交叉率 (Crossover rate) によって決定される。

#### (6) 突然変異 (Mutation)

選択、および交叉のみでは、初期母集団内の遺伝子に依存するような限られた範囲の子個体しか生成されない。そのため、一般的にあまり望ましくない解に収束することが多い。したがって、突然変異によって選択、および交叉だけでは生成できない子個体を生成し、個体群の多様性を維持することが必要である。突然変異では、染色体上のある遺伝子座の遺伝子を、他の対立遺伝子に置き換える。これは自然界における DNA 複写の際のコピーミスに当たる。各遺伝子座に対して何割の確率で突然変異が起きるかは、突然変異率 (Mutation rate) によって決定される。

#### (7) 終了判定 (Terminal criterion)

終了判定の条件としては、世代数、またはユーザの判断が用いられる。前者は一定世代数繰り返した後に終了する方法であり、後者であればユーザの求めるものが得られたのであればユーザは操作を終了し、得られていないのであれば再び評価、選択、交叉、および突然変異といった遺伝的操作を繰り返す。

### 2.3.3 対話型遺伝的アルゴリズムにおける個体表現

iGA では最適化の対象となる個体を，設計変数によって表現する．Fig. 2.3 に T シャツのデザインを最適化する場合の，T シャツの表現を示す．

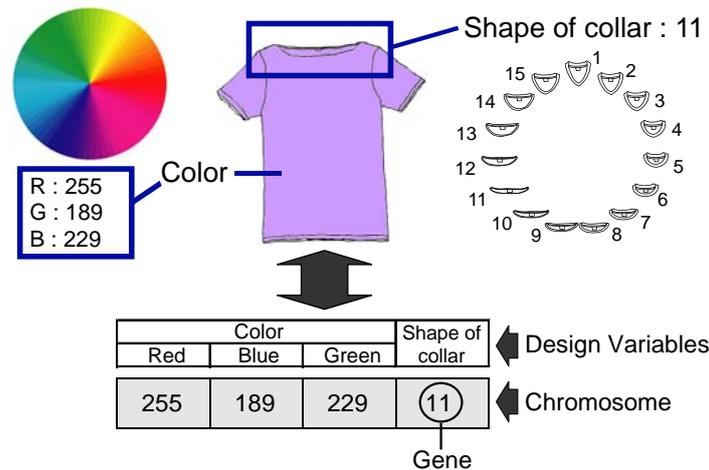


Fig. 2.3 The design variables in T-shirt iGA system

Fig. 2.3 では，T シャツは色，襟の形の属性によって表現される．この他にも模様，素材，装飾などが T シャツの属性として考えられ，これらの設計変数から構成される染色体によって，1 枚の T シャツのデザインが決定される．iGA における最適化は，遺伝子型の数値列を対象に行われる．各個体に対する評価値は，遺伝子型をデコードし，表現型の T シャツとして表現したものをユーザに提示することで獲得する．

### 2.3.4 対話型遺伝的アルゴリズムの特徴

iGA では人間が評価値を与えるため，心理空間上の好みと評価値が揺らぐ可能性がある．しかし，人間の評価の揺らぎについてはいくつかの実験がなされており，測定した主観的評価値の揺らぎに基づいたシミュレーションで解の収束性を調べた結果，ほとんど影響がないと報告されている<sup>9,10</sup>．これは，評価が人間に依存しているため，探索解の収束先も粗いものとなっている場合が多いからである．

また，iGA システムは対象とする問題に対して経験のないユーザの方がより有効に機能することが報告されている<sup>11-14</sup>．青木らが行った iGA を用いた 3 次元 CG ライティングの実験では，ある程度 CG 経験のあるユーザには有意な iGA のデザイン支援効果は見られなかったが，CG 経験の浅い，またはないユーザの場合は統計的に有意な iGA のデザイン支援効果がみられた．

### 2.3.5 対話型遺伝的アルゴリズムの課題

対話型遺伝的アルゴリズムの一般的な課題としては，以下が挙げられる

- 個体評価によるユーザの負担

iGA では，最適化系に用いる適合度を，ユーザが提示される各個体に対して評価を行うことで獲得する．この操作を複数世代に渡って繰り返すため，個体の評価がユーザに疲労や負担を与

えてしまう。また、ユーザが疲労することによって、正しく個体の評価ができなくなることは解の収束の悪化につながる。

そのため、ユーザに負担を与えない評価インタフェースの開発や、評価世代数を減らすために収束を高速化するアルゴリズムの開発、ユーザの過去の入力から未知の個体への評価値を予測するアルゴリズムの開発などが必要とされている。

- モデル化のコスト

iGA の対象問題には楽曲やデザイン、CG ライティングなどがある。これらの問題を iGA によって最適化するには、3.2 項において述べたように、それぞれを計算機で解けるような枠組みにモデル化する必要がある。モデル化は各対象問題の性質を調査した上で、GA の最適化に沿うよう適切に行わなければならない。また、対象毎に全く異なるモデルとなる場合が多いため、非常にコストがかかる。

本研究ではとくに 2 の課題について着目した。3.2 節では、商品推薦アルゴリズムを対象問題とした場合の、モデル化のコストに対する問題提起と、その解決方法の提案について述べる。

### 3 商品間の推薦関係に基づく対話型遺伝的アルゴリズムのための設計変数空間の導出

本章では iGA の商品推薦への適用と課題、そしてその課題を解決する手法の提案について述べる。iGA によるユーザへの適切な商品の推薦が可能であることは、先行研究からも明らかとされており<sup>6)</sup>、本研究では特に Web 上のショッピングサイトにおける商品推薦への適用を検討している。

#### 3.1 対話型遺伝的アルゴリズムの商品推薦への適用と課題

本節では従来の商品推薦手法について述べた上で、iGA による商品推薦の特徴、そしてその課題について述べる。

##### 3.1.1 商品推薦アルゴリズム

オンラインショッピングサイトで用いられる商品推薦のアルゴリズムには、以下のような手法がある。

(1) コンテンツフィルタリング<sup>15)</sup>

コンテンツの特徴とユーザのプロファイルをマッチングさせて、適合度の高い商品を推薦する手法である。各ユーザにフィットした商品が推薦される確率が高いが、事前にコンテンツの特徴に関する解析と、ユーザのプロファイルの収集作業が必要になる。

(2) ルールベースの推薦

ある製品を買った人、ある行動をした人には、この製品やサービスを勧めるというルールに基づく推薦手法である。事前に各商品にルール付けする作業が必要であるため、ショッピングサイト側のコストがかかる。

### (3) 協調フィルタリング<sup>1-3)</sup>

ユーザの過去の行動履歴を基に、類似した行動をとる他のユーザの行動履歴から推薦する商品を選出する手法である。例えば、ユーザ A と B が同じ商品を購入した場合、ユーザ B が購入した別の商品が、ユーザ A に対して推薦される。協調フィルタリングは商品の属性については考慮しないため、コンテンツフィルタリングのように、事前に商品の特徴値を事前に設定する必要はない。しかし、ユーザの行動履歴が大量に集積されないと有意な推薦結果が得られない。Amazon 等の大型ショッピングサイトなどでは、この協調フィルタリングを用いた商品推薦アルゴリズムが利用されている<sup>16)</sup>。

iGA を商品推薦に適用する場合はコンテンツフィルタリングの一種である。iGA が収集するユーザのプロファイルは、商品に対するユーザの評価で表された、ユーザの嗜好の情報である。ユーザの嗜好はショッピングサイトにアクセスする毎に変化する可能性があり、iGA を用いることにより、そのときどきのユーザの嗜好や感情に沿って、推薦内容を変化させることができると考えられる。

#### 3.1.2 対話型遺伝的アルゴリズムの商品推薦への適用

Fig. 3.1 に iGA による商品推薦の流れを示す。ユーザは Web インタフェースを通して、提示された個体、すなわち商品に対して評価を行う。iGA システムはその評価値に基づいて選択、交叉や突然変異といった遺伝的操作を行い、進化させた商品群を再びユーザに提示する。これらの操作をインタラクティブに行うことで、対象となる商品提示の最適化が可能となると考えられる。

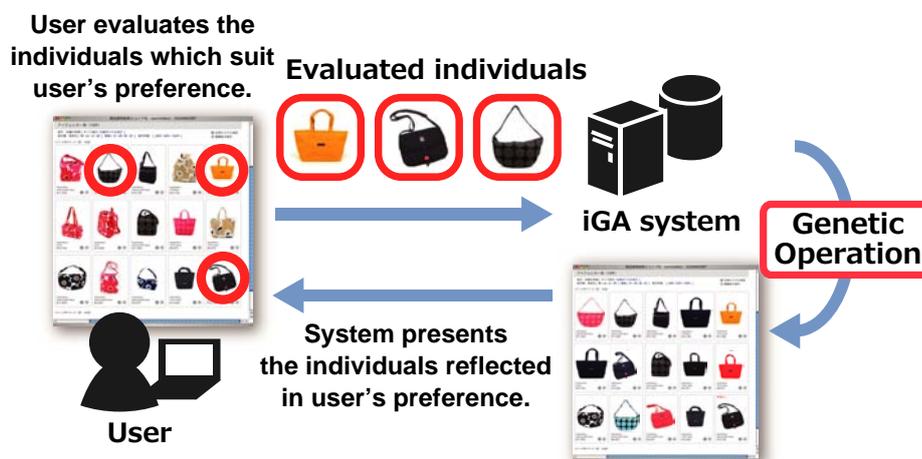


Fig. 3.1 iGA による商品提示の最適化

## 3.2 対話型遺伝的アルゴリズムの商品推薦への適用における課題

項において述べたように、対話型遺伝的アルゴリズムでは最適化の対象となる個体を遺伝子型によって表現する必要がある。従来の iGA では、対象問題の設計変数をシステム設計者が任意に決定し、個体のデザインを計算機上でシミュレートしていた。しかし、iGA を商品推薦に適用する場合、設計変数から個体をシミュレーションするのではなく、既存の商品から、その商品を表現するための設計変数を抽出する必要がある。この場合、従来の iGA と異なり以下のような問題が生じる。

### (1) 設計変数の定義

#### (a) 有効な設計変数の決定

色、大きさ、形、模様、価格、人気など、各種パラメータの中から商品を的確に表現するパラメータを、設計変数として特定する。

#### (b) 設計変数の近傍の設計

色や大きさ、価格といった要素は数値的な表現が容易である。色であれば、RGB 表色系、HSV 表色系などのモデルによって表現できる。しかし、形、模様などの要素は、人間の主観を反映した近傍を決めることが難しい。例えば、模様であれば、模様 A と模様 B が人間の主観において似ているのであれば、設計変数空間上において近く、模様 C と模様 D が似ていないのであれば、設計変数空間上において遠いように近傍が決定される。従来方法では、専門家による判断、アンケートによる統計的な尺度構成などにより、近傍が定義されてきた。

#### (c) 対象問題に依存するモデリング

有効な設計変数や設計変数の近傍は、対象問題によって異なる。例えば、カバンと椅子が全く同じ特徴で表現できるかと言えば、それは難しい。

### (2) 設計変数値の入力のコスト

ベンダ側は、定義された各設計変数に対し、実際の商品の値を測定し、遺伝子として入力する作業が必要となる。商品の特徴的な色を抽出する手法は既に存在するが<sup>17)</sup>、模様や装飾について解析することは難しい。従って、設計変数値を適切に抽出するには人間の作業が必要となる。

特に 2 の問題については、何百何千という商品を取り扱う場合、致命的な問題となり得る。これらの問題を解決するために、自動的に設計変数を定義し、各商品の遺伝子を算出する手法が求められる。

iGA のための設計変数の自動抽出手法としては、Cho らのウェーブレット変換を用いた絵画の遺伝子の自動生成などがある<sup>18,19)</sup>。しかし、これは絵画の視覚的な特徴のみを反映したものであり、商品の大きさや価格といった視覚特性以外の情報を反映しない。

## 3.3 商品間の推薦関係に基づく対話型遺伝的アルゴリズムのための設計変数空間の生成

### 3.3.1 商品間の推薦関係に基づく対話型遺伝的アルゴリズムのための設計変数空間の生成

2.3.5 節で述べた問題を解決するために、本研究では Web 上に蓄積された多数のユーザの集合的な嗜好情報から、対話型遺伝的アルゴリズムに用いる設計変数を自動的に生成する手法を提案する。

本論文では、集合的な嗜好として Amazon Web Services が公開している商品の推薦関係を扱う。推薦関係を用いたのは、商品同士の関連度の取得が容易なためである。提案手法では、各商品が他のどの商品を推薦しているかをその商品の特徴として捉え、遺伝子の原型とする。さらにこれに次元を圧縮する手法である主成分分析<sup>20)</sup>を適用し、その結果を各商品の遺伝子とした。

Amazon Web Services の詳細については 3.3.2 項、主成分分析の詳細については 3.3.3 項において詳述する。

### 3.3.2 Amazon Web Services

本論文では、商品の推薦関係の情報を、AWS (Amazon Web Services)<sup>1</sup>から取得するものとする。AWS とは、Amazon.com, Inc. が提供する商品データベースへのアクセスや技術プラットフォームの利用を可能とするサービスの総称である。特に Amazon Associates Web Services は、API を用いることで Amazon や Amazon の販売プラットフォームを利用するショッピングサイトが扱う商品の情報を取得できる。書籍を例に、取得できる情報の種類を以下に示す。

- 表題
- 著者
- 出版社
- ページ数
- 価格
- ユーザによるレビュー
- 書籍が推薦する他の書籍のリスト (10 冊まで)

本論文では”書籍が推薦する他の書籍のリスト”を用いて設計変数空間の構築を行っている。なお、AWS の仕様により、1つの商品に対して推薦関係が取得できるのは 10 商品までとなっている。AWS の商品推薦アルゴリズムには、3.1.1 にて述べた協調フィルタリングが用いられている<sup>16)</sup>。よって、今回用いる商品間の推薦関係は、Amazon 上におけるユーザの購買行動に基づく関係となる。

### 3.3.3 主成分分析

#### § 主成分分析の概要

主成分分析とは、データに含まれる変数間の相関関係や特徴を容易に把握するために、変数の統合的指標を統計的に設定する手法である。例えば、Fig. 3.2 に示すように「体重」「身長」という情報を表す変数  $x_1$ ,  $x_2$  があつた場合、両者に相関関係があることが予想される。このとき、両者の相関関係から、データのもつ特徴、または傾向が最もよく反映されている新たなベクトル  $z_1$  を算出し、この  $z_1$  によってデータを表現することができる。このように 2次元のデータを 1次元で表現するなど、主成分分析を用いることでデータの持つ情報の損失を最小限に押さえつつ、データを低次元化することができる。

---

<sup>1</sup><http://aws.amazon.com/>

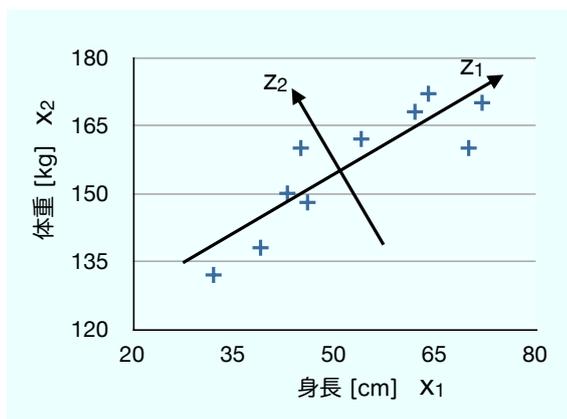


Fig. 3.2 体重，身長を変数とする主成分分析

Fig. 3.2 に示す  $z_1$ ,  $z_2$  のように算出されたデータの特徴を反映するベクトルを主成分と呼ぶ。主成分は，データの共分散行列から求められた固有ベクトルがそれに値する。この固有ベクトルに対する固有値が最大である主成分が，データの傾向を最もよく反映する第 1 主成分であり，以降固有値の大きさに従って，第 2 主成分，第 3 主成分，…，第  $m$  主成分 ( $m \leq$  変数の数) と求められる。この求められた主成分を用いて元のデータを表現したものを，主成分得点という。Fig. 3.2 では，分布の軸として  $z_1$ ,  $z_2$  を用いたときの各データの分布を示している。

各主成分が元のデータに含まれる特徴をどの程度反映しているのかを示す指標を，寄与率という。主成分は固有ベクトルであることを前節にて述べたが，この固有ベクトルに対応する固有値の総和をとり，各固有値の割合を求めたものが各主成分の寄与率となる。また，第一主成分から第  $m$  主成分までの寄与率の総和を累積寄与率と呼ぶ。累積寄与率が 60% を越える主成分数を利用するのが適切と言われているが，分野，用途によってこの閾値は異なり，主成分分析の目的や適用対象に応じて採用する主成分数を適切に決定する必要がある。

また，主成分に対して，各変数が及ぼす影響の度合いを定量化したものを因子負荷量という。因子負荷量は各主成分と各変数の相関係数として定義される。

### § 主成分分析による設計変数の次元圧縮

提案手法では，対象となるデータを  $\{0, 1\}$  の分類尺度のデータとして扱い，主成分分析を行った。主成分分析によって算出された累積寄与率に基づいて適当な主成分数を求め，それらの主成分から得られた主成分得点を各個体の遺伝子として扱った。

提案手法において，主成分分析を用いる理由は以下の通りである。

- 疎行列

推薦情報に基づいて作成される隣接行列は，0 の多い疎行列となる。これは 3.3.2 項で述べたように，AWS によって提供される推薦関係の情報は，1 つの商品に対し，10 商品までであるためである。この隣接行列の各行を遺伝子として用いる場合，iGA の交叉において，子個体が親の形質を適切に受け継ぐことができず，有効な探索が行えない。

- 遺伝子長の商品数に対する依存

商品の推薦関係の有向グラフは、商品数×商品数の隣接行列として表される。その各行をそのまま設計変数として用いる場合、一商品の遺伝子長は、全体の商品数に依存する。これは運用上望ましくない。また、iGAの母集団数に対し、探索領域が非常に広くなりう、GAによる探索能力に影響を与える。

### 3.3.4 商品間の推薦関係に基づく設計変数空間の生成手順

商品の推薦関係に基づく設計変数の生成は、以下の手順で行う。

- (1) 商品の推薦関係を Web 上のサイトから取得する。
- (2) 商品 (Item) の推薦関係を有効グラフとして捉え、隣接行列を生成する。

$$\begin{array}{c}
 \text{Ind}_1 \quad \text{Ind}_2 \quad \text{Ind}_3 \quad \dots \quad \text{Ind}_N \\
 \begin{array}{c} \text{Ind}_1 \\ \text{Ind}_2 \\ \text{Ind}_3 \\ \vdots \\ \text{Ind}_N \end{array} \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 1 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \dots & 1 \end{pmatrix}
 \end{array}$$

$\text{Ind}_1$  が  $\text{Ind}_2$  を推薦するとき、 $[\text{Ind}_1, \text{Ind}_2]$  の値は 1 となり、推薦関係が無い場合は 0 となる。また、行列の主対角線上は全て 1 となる。行列の各行がそれぞれの商品個体の遺伝子型 (genotype) の原型を示し、各列は商品の設計変数を意味するものとする。

- (3) 主成分分析により、商品個体の設計変数の次元数を削減する。
  - (a) 隣接行列から共分散行列を算出し、その固有値、固有ベクトルを求める。
  - (b) 設計変数の削減後の次元数を決定し、固有値の降順に次元数分だけ固有ベクトルを抽出し、回転行列を生成する。
  - (c) 回転行列に元の対象行列を乗じ、各個体の主成分得点を得る。これを各商品の遺伝子型とする。

## 4 商品間の推薦関係に基づく設計変数空間の特徴の検討

提案手法によって生成された設計変数空間を用いた iGA を行うにあたり、設計変数空間の特徴について検討を行った。

分布の特徴、生成された軸の特徴、因子負荷量による操作の 3 項目において考察を行った。

### 4.1 設計変数空間の生成パラメータ

本節では検討対象として Amazon で取り扱われている商品のうち、書籍を用いる。現在 Amazon で扱われている商品数は 300 万を超え<sup>1</sup>、その全ての推薦関係を取得することは難しい。よって、まず

<sup>1</sup>2010 年 1 月 19 日時点

は検討のため、小規模な設計変数空間を3パターン生成した。この設計変数空間は5章に述べたiGAによる被験者実験にも使用し、実験結果の考察が容易になるように設計変数空間の次元数を5として生成した。

Table 4.1 は各設計変数空間のパラメータである。

Table 4.1 設計変数空間の生成パラメータ

設計変数空間名	カテゴリ	初期値商品	データ取得期間	商品数	推薦の距離
mystery	ミステリー・サスペンス・ハードボイルド	ブラックペアン 1988(下)	2009/12/31 - 2009/1/1	649	4
science	コンピュータ・サイエンス	プログラマのための論理パズル 難題を突破する論理思考トレーニング	2010/1/3 - 2010/1/4	703	4
comic	少年コミック	鋼の錬金術師 24 (ガンガンコミックス) (コミック)	2010/1/1-2010/1/2	309	3

Table 4.1 中の”カテゴリ”は Amazon における書籍の分類カテゴリーを示す。それぞれの”カテゴリ”において、”データ取得期間”における売上げが最も良かった書籍を”初期値商品”としている。まず、この”初期値商品”が推薦している商品のリストを AWS から取得し、さらにそのリスト中の一商品が推薦する商品のリストを、そのリストに追加した。この手順を繰り返して、商品の推薦関係を獲得した。なお、このとき、予備実験<sup>21)</sup>より DVD について省くようにした。書籍以外の商品が設計変数空間に含まれる場合、人間の感性を反映した設計変数空間を生成できないと考えられる。これは直接の推薦関係がある商品同士であっても、ユーザは書籍と書籍以外の商品に対して、全く異なる印象を抱くためである。

”推薦の距離”は、推薦関係を取得する商品が、初期値商品から幾つのエッジでつながったものであるかを示している。各設計変数空間は、指定された”推薦の距離”にある商品の全てを取得し、それが各設計変数空間を構成する”商品数”となる。

”推薦の距離”は累積寄与率によって決定した。前述したように本節では5次元の設計変数空間で検討を行う。この5次元の設計変数空間はすなわち、第1主成分から第5主成分によって構成された回転行列によって算出された主成分得点である(3.3.4項の手順3を参照)。すなわち、設計変数の第m次元が第m主成分に基づく分布を示すことになる。この設計変数における各次元の分布が基の隣接行列をどの程度反映しているかの指標が累積寄与率である。累積寄与率としては0.6や0.7が適当とされているが、今回は後述するiGAによる実験を考慮して、5次元で0.1を越えればよいと仮定した。またこの仮定に基づいて”推薦の距離”を決定した。

Fig. 4.1 から Fig. 4.3 に生成した3つの設計変数空間の第1次元、第2次元における商品の分布を示す。黒丸がそれぞれ1冊の書籍を示す。それぞれの次元に対し、偏った分布をしている様子が観察された。

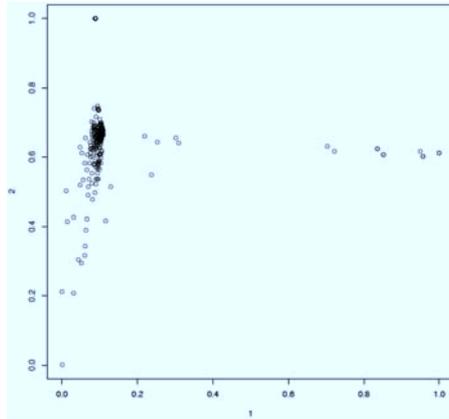


Fig. 4.1 設計変数空間 comic における商品の分布 (第 1-2 次元)

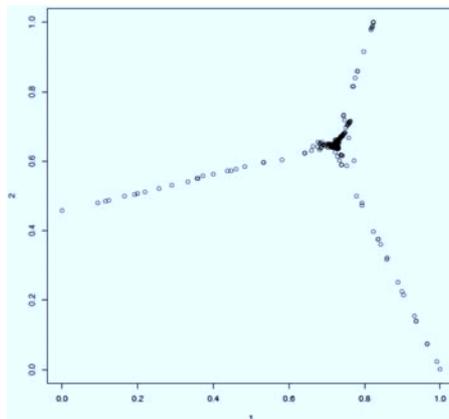


Fig. 4.2 設計変数空間 mystery における商品の分布 (第 1-2 次元)

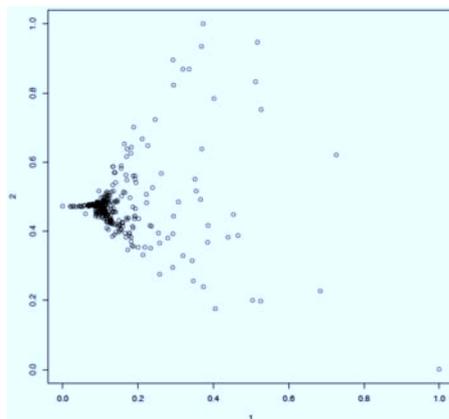


Fig. 4.3 設計変数空間 science における商品の分布 (第 1-2 次元)

## 4.2 商品の分布と属性値の関係性に関する検討

### 4.2.1 商品の分布における著者の偏り

設計変数空間 mystery, comic における分布では, 著者による影響が確認された. Fig. 4.4(a) は設計変数空間 mystery の第 1 次元を横軸に, 第 2 次元を縦軸としたグラフである. Fig. 4.4(b) は, 第 3

次元と第5次元を軸としたグラフである。

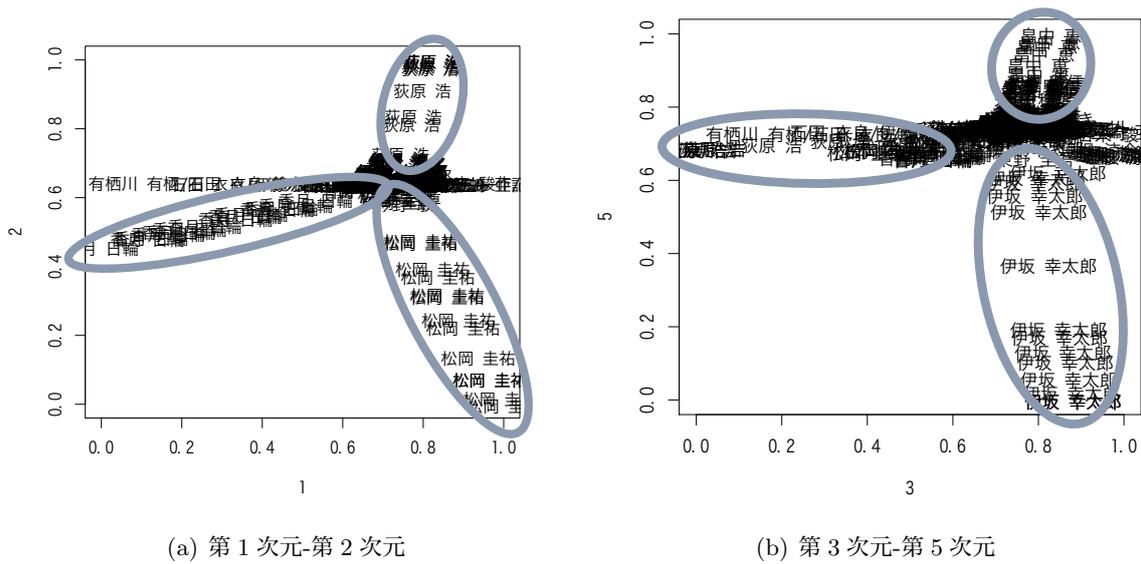


Fig. 4.4 mystery における著者の分布

各軸に対し、特定の著者が狭い幅の中に分布している。同様の傾向は、comic においても確認できた。Fig. 4.5 にその例を示す。

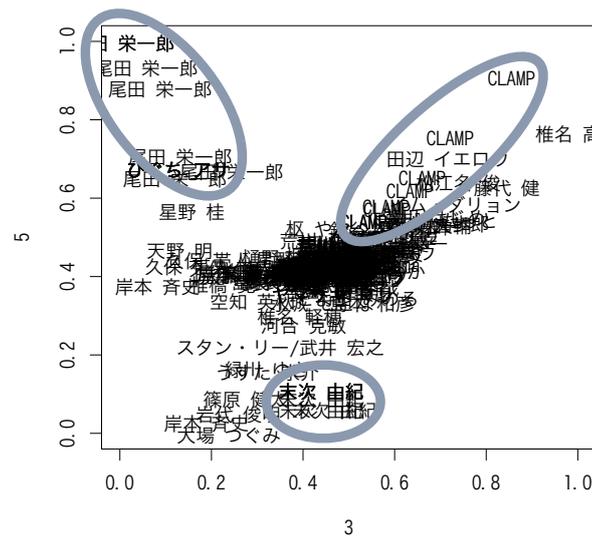


Fig. 4.5 comic における出版社の分布 (第3次元-第5次元)

この原因は Amazon のユーザが同じ著者の書籍を買い揃えようとするためと考えられる。Fig. 4.2 に、各設計変数空間における全ての書籍の数と、重複を排除した著者の数、著者を同じくする書籍の数を示す。mystery における同著者の作品が占める割合は 0.886 である。これは computer science のおよそ 3 倍である。

同じ著者の書籍は互いに推薦し合うような密な推薦関係を持つ傾向がある。3.3.4 項の手順 2 に述

Table 4.2 The numbers of books written by a same author

Category	Total Books	Distinct Authors	Books Written by a Same Author
mystery	649	205	575
comic	309	135	216
computer science	703	498	213

べたように、隣接行列では行の商品が列の商品を推薦している場合 1 を得るので、密な推薦関係を持つ場合、要因としての書籍に与えられる相関は強いものとなる。そのため、主成分分析の結果として Fig. 4.4(a) に見られるようなはっきりとした分布を示す。

#### 4.2.2 商品の分布における出版社の偏り

mystery, comic では分布に強い影響を与えた著者であるが、science においては分布の偏りが見られなかった。これは Table 4.2 に示すように science における同じ著者の作品数が少なかったことが原因であると考えられる。

設計変数空間 computer science における書籍の分布では、出版社による影響が見られた。science の第 1 次元、第 4 次元の分布図を Fig. 4.6 に示す。横軸に対して、複数の出版社が狭い範囲に分布している。

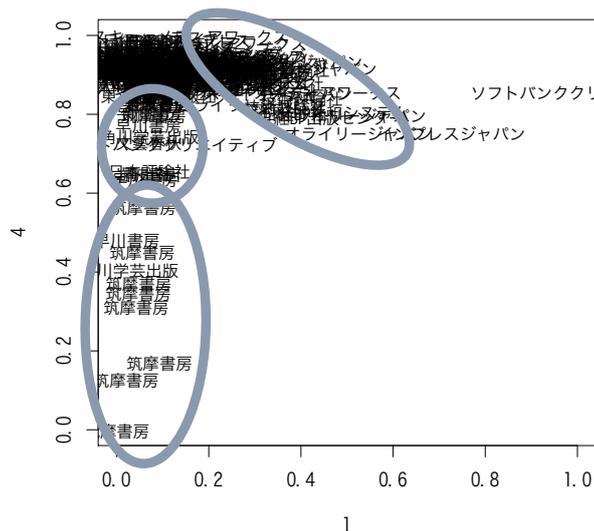


Fig. 4.6 science における出版社の分布 (第 1 次元-第 2 次元)

この分布の原因として computer science に属する書籍が専門的なものであることが挙げられる。そのため、ユーザは著者の個性ではなく出版社の傾向を重視すると考えられる。なお、computer science における同じ出版社から出版された書籍の全体に占める割合は、0.829 であった。しかし、computer science ではこの他の次元の組み合わせにおいて、出版社の分布の偏りは視認できなかった。この設計変数空間の特徴を確認するためには、出版社だけでなく、タイトルやカスタマレビューに含まれる

専門用語などの解析が必要と考えられる。

### 4.3 次元数の検討

mystery の著者の分布から，設計変数の適切な次元数に関する考察を行う。Fig. 4.7(a)，Fig. 4.7(b) は Fig. 4.4(a) の商品の密集部分をそれぞれ 12.5 倍，100 倍に拡大したグラフである。

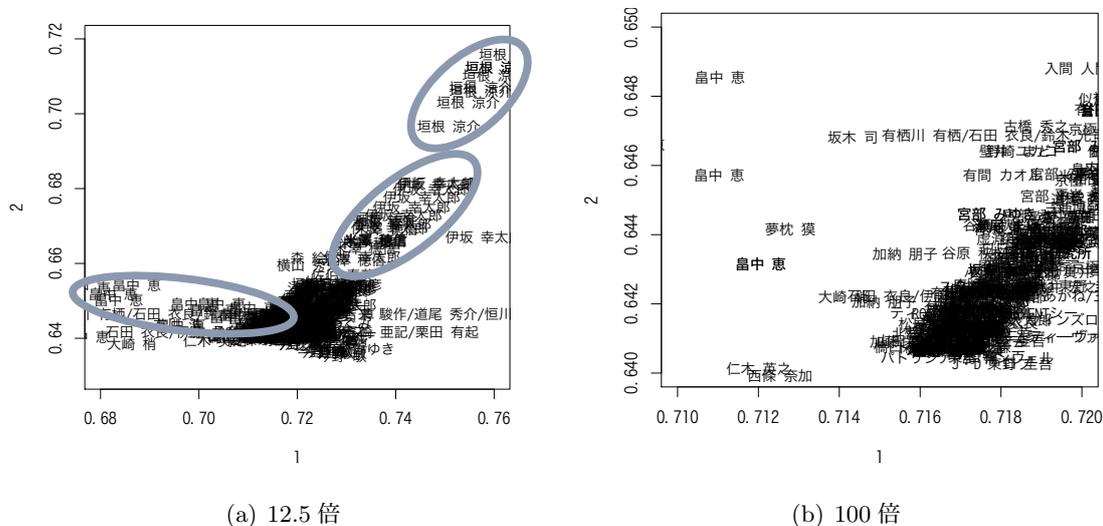


Fig. 4.7 mystery における著者の分布 (第 1 次元-第 2 次元)

Fig. 4.7(a) ではまだ著者の分布の偏りが視認できるが，Fig. 4.7(b) では確認できなかった。で視認できなかった著者の一部は，Fig. 4.4(b) など他の次元にて分布が確認された。しかし，その他の多く商品が，今回検討対象である 1 から 5 次元において著者の分布の偏りを示さなかった。

この原因は，1 つの主成分において有効な著者の数が限られているためと考えられる Table 4.3 に各主成分において，主因子法により求めた因子負荷量が 0.0646 を越える著者のリストを示す。因子負荷量は各要因が，主成分に対してどの程度影響を与えているかを示す指標である。0.0646 という数値は，危険率を 0.10 とした以下の式 4.1 から求めた。

$$r(0.10) = \frac{\sqrt{n-1}}{1.645} \quad (4.1)$$

各主成分に対して高い因子負荷量を持つ著者は，3 人までである。また，全ての主成分に対して高い因子負荷量をもつ著者は，存在しない。この結果から，より多くの著者について分布の偏りを示すためには，より多くの主成分を用いる必要があると考えられる。目安としては，累積寄与率が 0.6 を超える主成分数などが考えられる。

しかし，設計変数の次元数が増えると，本論文で用いる最適化アルゴリズムである iGA の解探索性能が低下する。iGA による探索性能を考慮した次元の削減の必要と，元の隣接行列における相関を適切に反映するための次元の非圧縮のトレードオフが設計変数に生じており，適切な次元数の決定手法が必要である。

Table 4.3 因子負荷量の高い著者

Dimension	Authors
1	Keisuke Matsuoka, Hinowa Kogetsu, Megumi Hatanaka
2	Keisuke Matsuoka, Hiroshi Ogiwara
3	Hiroshi Ogiwara, Ryosuke Kakine
4	Ryosuke Kakine
5	Kotaro Isaka, Megumi Hatanaka, Yu Shibata (the coauthor of Megumi Hatanaka), Ira Ishida/Takuji Ichikawa/Eiichi Nakata/Wataru Nakamura/Takayoshi Honda (the coauthors of Kotaro Isaka)

## 5 商品間の推薦関係に基づく設計変数空間上における対話型遺伝的アルゴリズムの検討

### 5.1 実験目的

提案手法を用いて生成した設計変数空間において、ユーザがどのような探索を行うのかを調査するため、iGA システムによる被験者実験を行った。

### 5.2 実験システム

#### 5.2.1 実験システムの概要

探索履歴の検討のため、提案手法によって生成した設計変数空間を利用する iGA システムを実装した。構築するシステムは、4.1 項で生成した設計変数空間を用いるため、書籍を対象とした iGA システムである。

#### 5.2.2 実験システムのインタフェース

実験システムのインタフェースを Fig. 5.1(a) に示す。インタフェースにより提示された 16 個の商品の中から、被験者は自身の嗜好に合った商品をクリックして選ぶ。被験者に対して提供される商品の情報は、Table 5.1 の通りである。商品に関する詳細な情報は、被験者が書籍の表紙をマウスオーバーすると、ポップアップされる詳細画面 (Fig. 5.1(b)) で確認できる。詳細画面は、マウスが離れると消去される。

一つの提示画面において被験者が選択できる商品の最大数は、提示個体数の半数である 8 とする。被験者は 1 個以上の個体を選択して画面上部右の”次へ”ボタンを押す。すると、システムがユーザの評価に基づいて GA による遺伝的操作を行い、インタフェース上の商品を更新する。

#### 5.2.3 実験システムの最適化系

本システムの最適化系は iGA である。以下、本システムにおける遺伝的操作の詳細について述べる。



(a) インタフェース



(b) 詳細画面

Fig. 5.1 実験システム

Table 5.1 被験者に提示する商品情報

提示ウィンドウ	提示する商品の情報
インタフェース (Fig. 5.1(a))	表紙
詳細画面 (Fig. 5.1(b))	表紙, 表題, 著者, 価格, 出版年, 出版社, カスタマーレビュー

### (1) 選択

- 複製選択

被験者の評価に基づき、交叉、突然変異を行う親個体を、母集団サイズの半数選択する被験者がクリックした個体（以降、選好個体とする）が半数に満たない場合、非選好個体の中から、選好個体に対する設計変数空間上のユークリッド距離が近いものから順に、親個体として選択する<sup>22)</sup>。

- 生存選択

エリート戦略を用いる。エリート戦略とは、適合度の高い個体を無条件に次世代に残す選択手法であり、前の世代における選好個体が、次世代においても表示されるようになる。本システムにおけるエリート戦略であるが、選好個体をその直後の世代では表示せず、2世代後に提示する。これは事前に行った予備実験により、前の世代で提示された個体とその直後の世代で提示されると、被験者が不快感、違和感を覚える場合が多かったため、直後の世代では表示せず、一世代間を置くようにした。

### (2) 交叉

交叉手法として、ブレンド交叉 (BLX- $\alpha$ )<sup>23)</sup> を用いた。BLX- $\alpha$  は、親個体の各次元での区間  $d_i$

を次元軸方向の両側に  $\alpha d_i$  だけ拡張し、その区間内で一様ランダムに子個体を生成する。すなわち、親個体の周辺に次元軸に平行な辺をとり、その各辺によってできる超直方体の内側の領域が子個体の生成領域となる。BLX- $\alpha$ では親個体が探索空間で離れて存在している場合には子個体も広い範囲に生成され、親個体が互いに近くに存在している場合には親個体の近傍に生成される特徴がある。

### (3) 突然変異

突然変異手法として、一様突然変異を用いた。一様突然変異は、変数の定義域の範囲で、一様乱数を用いて値を生成する手法である。

### (4) 引き戻し処理

交叉、突然変異によって生成された遺伝子が、既存の商品の遺伝子と一致するとは限らない。よって、生成された遺伝子を、設計変数空間上のユークリッド距離が最も近い既存の商品の遺伝子に置き換える。このとき前の世代で提示された商品との重複は避け、距離が近いものから順に選択する。

エリート戦略と引き戻し処理において、前の世代で提示した個体をその直後に提示しないように操作している。単純な図形や色の最適化を行うような、より直感的な iGA<sup>22)</sup> では、被験者は前の世代に表示された個体をあまり記憶しない。しかし書籍など、文字情報や追隨する情報が多く含まれる対象問題では、被験者の各個体に対する認知レベルが高く、提示された個体を記憶する。そのため、予備実験では同じものが世代間で連続して提示された場合、被験者は飽きる、また、同じ個体を選択することに違和感を覚えるなどの意見が得られた。

以下に iGA の主なパラメータを示した。

Table 5.2 iGA パラメータ

パラメータ	値
個体数	16
設計変数	5次元
世代数	4
交叉率	1
突然変異率	0.2
$\alpha$	0.2

## 5.3 実験計画

設計変数空間として、4.1 項で生成した mystery, science, comic を用いる。1つの設計変数空間を実装した iGA システムに対し、4世代評価を行う操作を1試行とする。被験者は各設計変数空間に対して iGA を行い、3試行操作する。設計変数空間の順序は、被験者間でカウンタバランスをとった。

Fig. 5.2(a), Fig. 5.2(b), Fig. 5.2(c) に各試行の初期画面の一例を示した。また, Fig. 5.3 に実験風景を示す。



(a) 設計変数空間 comic の初期画面



(b) 設計変数空間 mystery の初期画面



(c) 設計変数空間 science の初期画面

Fig. 5.2 初期画面の例



Fig. 5.3 実験風景

被験者数は 22 歳から 29 歳の男女 6 名で, 男性 5 名, 女性 1 名である。実験は以下の手順で行った。

(1) 被験者に教示を与える

被験者に書籍を選択するモチベーションを与えるため、被験者に対し、「あなたは明日休みをとることになりました。そしてその一日を使って本を読みたいと考えています。そのために、オンラインブックストアで本を探すことにしました。」という内容の教示を与えた。

(2) システムの操作

iGA を 1 試行，被験者に操作させる。

(3) アンケート

以下の項目について，被験者に 5 段階で回答させる。

**項目 1** 最終画面で提示された商品に対し満足しているか。

**項目 2** 最終画面で提示された商品に対し，多様性があるか。

**項目 3** 商品の選び方について，特定の興味のある分野を選んだか，それとも様々な商品を選んだか。

(4) 終了判定

全ての設計変数空間について試行した場合，終了。終了しなければ，2 に戻る。

アンケートでは，満足度，多様性について問う。iGA は，人とシステムがインタラクティブにやり取りする中で，人の嗜好に沿う解を探索するアルゴリズムである。従って，被験者の満足度を向上させることが重要になる。また，iGA は人が評価系であるため世代数が極めて少なく，収束圧が高いアルゴリズムを用いることが多い。そのため，多様性が失われやすく，被験者の満足度を低減させてしまう可能性があるため，多様性についてもアンケートで問う。

## 5.4 結果と考察

本節では，被験者間に置ける探索結果の違いと被験者内における推薦アルゴリズムの比較、被験者へのアンケート結果の 3 つの項目について考察を行う。

### 5.4.1 被験者間の探索結果の違い

実験の結果，被験者毎に異なる探索の傾向が見られた。Fig. 5.4(a), Fig. 5.4(b) は，ミステリの探索空間における被験者 A と被験者 B の探索結果である。黒三角は第 1 世代の提示個体，白丸は最終世代の提示個体を示す。

被験者 A, B の初期個体は同一であるが最終世代における提示個体は異なるものとなった。この探索結果の違いより，生成した設計変数空間において被験者固有の嗜好による探索ができたと考えられる。また，被験者 A の最終個体は縦軸に対して，被験者 B の嗜好は横軸に対して狭い範囲に分布しており，各軸に依存した嗜好が確認された。

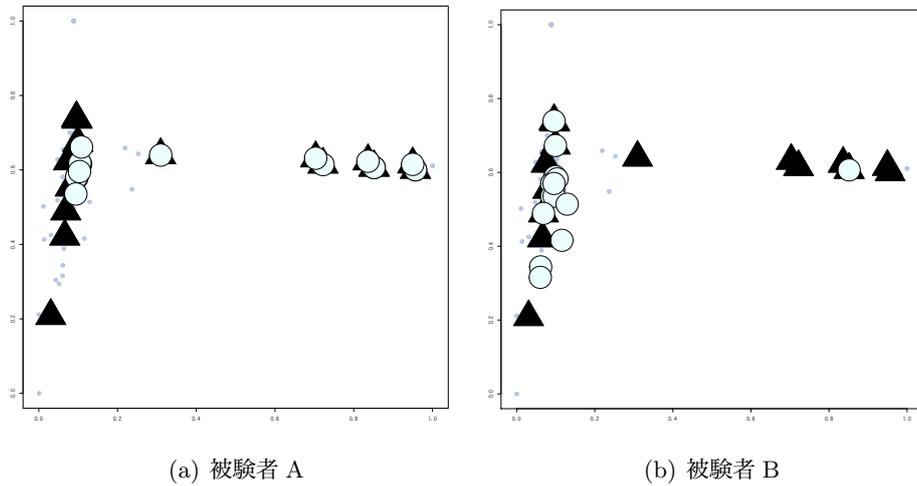


Fig. 5.4 最終世代の分布の比較（ミステリ）

#### 5.4.2 被験者内における推薦アルゴリズムの比較

iGAによる推薦結果と、元のデータであるAmazonの推薦結果の比較を行う。Fig. 5.5は、被験者Bが設計変数空間”少年コミック”の第3世代で選択した個体から、iGAによって選出された個体である。Amazonによって推薦された個体も示した。被験者が第4世代で選択した個体を丸で示している。



Fig. 5.5 被験者Bの推薦結果の比較

iGAにより推薦された個体と、Amazonにおける推薦個体は一部重複した。被験者が第4世代で選択した個体にも、重複個体が含まれる。しかし、iGAによってのみ推薦された個体（以降、非Amazon個体）も被験者による評価を受けており、元の推薦関係とは異なるiGAの推薦が、ユーザの評価に適う可能性を示すことができた。

Table 5.3に、被験者毎に、被験者が評価した個体の中に非Amazon個体が含まれる比率を示した。総平均は0.390であり、Amazonとの重複個体だけでなく、iGAによってのみ推薦された個体であっても評価が高いことが分かる。しかし、非Amazon個体の評価は被験者による偏りも見られ、比率がもっとも低い被験者Aと高い被験者Fを除いて平均をとると0.355と下がる。比率の最も高い被験者Fのミステリにおける探索履歴を確認すると、iGAによって提示された商品とAmazonの推薦個体との重複自体が少なかった。この被験者Fについては、Amazonの推薦による個体の提示経路とは全く異なる探索をiGAで行っていたものと考えられる。

Table 5.3 被験者による評価個体に占める非 Amazon 個体の比率

Subject	A	B	C	D	E	F
comic	0.389	0.250	0.304	0.556	0.625	0.714
mystery	0.118	0.400	0.385	0.087	0.308	1.000
science	0.417	0.412	0.375	0.538	0.417	0.600
average	0.308	0.354	0.355	0.394	0.450	0.771

比率の最も低い被験者 A のミステリにおける探索履歴を確認すると、2 世代目、4 世代目の提示個体の半数以上が、Amazon による推薦個体と一致していた。

被験者によっては Amazon の推薦データと重複した推薦に対して満足度が高く、今後は Amazon によってのみ推薦される個体も被験者に提示し、評価の比率について検討を行っていく必要がある。

今後は Amazon によってのみ推薦される個体も被験者に提示し、評価の比率について検討を行っていく必要がある。

### 5.4.3 アンケート結果の検討

3 つの設計変数空間に対する被験者 6 名のアンケートの項目 1 に対する回答を Fig. 5.6(a) に示す。Fig. 5.6(b) には項目 2 の結果を示した。

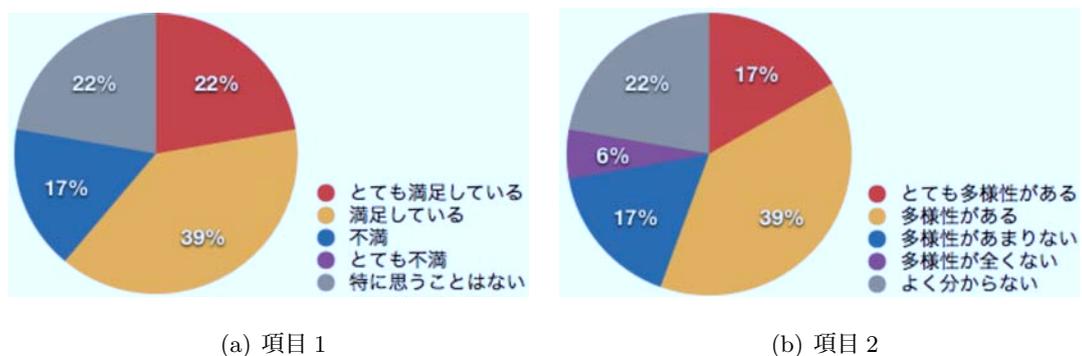


Fig. 5.6 アンケート結果

アンケートの結果では、最終世代で提示された個体が満足できるものだったという回答が、半分を超えた。多様性があるという回答も半分以上だった。

しかし、項目 2 の結果では「よく分からない」という回答が全体の 22% を占めた。そのように回答した被験者にインタビューしたところ、カテゴリー内の書籍について把握しておらず、多様性がよく分からないというコメントが得られた。このことから、被験者が慣れているカテゴリーを用いる必要があることが分かった。

「不満」と答えたユーザに対するインタビューでは、回答の理由として以下の項目を挙げている。

- (1) そのカテゴリーをよく知らないで、自分の中に評価基準がない。
- (2) 同じシリーズの書籍が多く出てきたので、もっと種類を増やして欲しい。
- (3) 読んだことがない本をカスタマーレビューだけで選ぶのは難しい。

1 に対しては、前述した問題と同様である。

2 については、comic の設計変数空間について得られたコメントであった。comic の設計変数空間は、Table 4.1 に示されるように商品の数が他の設計変数空間に比べておよそ半分と少ない。探索の領域が狭かったことが原因の一つであると考えられる。実験に用いる設計変数空間を生成する際には、被験者の探索回数を考慮した上で商品数をそろえる必要があることが分かった。

3 は 1 と同じく、商品に対する被験者の情報量の不足の問題である。今回ユーザに提示する情報は AWS から取得できる情報に限っていた。しかし、提示された商品に対して被験者が適切に評価できるようにするため、あらすじを事前に取得しておく必要があると考えられる。

## 6 結論

本研究では、iGA を商品推薦アルゴリズムとして用いるにあたり、探索対象である商品の設計変数を Web 上の商品の推薦関係から生成する手法を提案した。提案手法では、商品間の推薦関係を有向グラフとし、その隣接行列を主成分分析することにより、各商品の染色体を算出した。また、その染色体の分布を調査することで、導出された設計変数空間の各次元の持つ意味について考察した。

生成された設計変数空間の特徴を検証するために、Amazon の書籍の推薦データを用いて、「ミステリ」、「少年コミック」、「コンピュータサイエンス」の 3 種の設計変数空間を生成した。「ミステリ」、「少年コミック」の設計変数空間においては著者の値が軸に対して狭い範囲に分布する現象が見られた。これは著者や出版社を主眼においた iGA を行うユーザが効率的な探索を行える可能性を示している。しかし、一方で多くの著者による分布を設計変数空間に反映するには、設計変数の高次元化が必要となる。これは iGA により有効な探索を行うための設計変数の低次元化には反する特徴だった。一方、「コンピュータサイエンス」のカテゴリにおいては、さらに下層のカテゴリが与える影響が部分的に確認できたが、分布に支配的な影響をもつ特定の属性は見つからなかった。専門性をもつカテゴリにおいては、書籍のもつ著者や出版社といったプロフィール的な属性値だけでなく、書籍内容に関わる情報を用いた解析の必要性が示唆される。

また、生成された設計変数空間においてユーザがどのような探索を行うのか検証するため、生成した設計変数空間を用いた書籍の iGA システムを構築し、被験者実験を行った。実験結果より、被験者間では探索傾向の違いが確認された。これにより、ユーザ固有の嗜好に基づく探索が行われていることが考えられる。また、設計変数空間を作る元データとの推薦結果の違いを比較し、iGA による推薦結果が被験者の評価を受けることを確認した。

今後の課題としては、設計変数空間の特性についてさらに調査を行う。そして、適切な設計変数の次元数を決定するための手法と、その次元における適切な分布の生成手法について検討する。また、扱う商品数を増やしてさらに大規模な設計変数空間での特性についても調査を行う必要がある。

## 謝辞

本研究を遂行するにあたり、多大なるご指導そしてご協力を頂きました。同志社大学生命医科学部の廣安知之教授に心より感謝いたします。本研究を遂行するにあたり、様々なご指摘、ご助言を頂きました。同志社大学理工学部の三木光範教授、吉見真聡助教に心より感謝いたします。

これまで三木先生、廣安先生、そして吉見先生のご指導の下、研究活動に尽力して今に至りましたことを、この場を借りて重ねて厚く御礼申し上げます。本当にありがとうございました。

本論文の執筆にあたり、多大なご協力を頂きました。佐々木康成氏、知的システムデザイン研究室の小林祐介氏、および藤田宗佑氏に心より感謝いたします。また、同じ研究グループに所属し、本研究を実施するにあたりご協力を頂きました。知的システムデザイン研究室の米田有佑氏に心より感謝いたします。

そして、何事に対しても意欲的に取り組み、常に励まして頂きました知的システムデザイン研究室の各位に深く感謝いたします。お忙しい中、たくさんの方が快く実験にご協力くださいました。本当にありがとうございました。

最後に、本研究を実施するにあたり、理解と献身的な協力を頂きました両親、そして私を支えてくれた全ての方々に心より感謝いたします。ありがとうございました。

## 参考文献

- 1) Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl. Grouplens: applying collaborative filtering to usenet news. *Communications of the ACM*, Vol. 40, No. 3, pp. 77–87, March 1997.
- 2) Badrul Sarwar, George Karypis, Joseph A. Konstan, and John Reidl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pp. 285–295. ACM New York, USA, 2001.
- 3) Thorsten Joachimes. Text categorization with suport vector machines: Learning with many relevant features. In *Proceedings of the 10th European Conference on Machine Learning*, pp. 137–142. Springer-Verlag London, UK, 1998.
- 4) Hideyuki Takagi. 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.
- 5) David E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional, 1989.
- 6) 高木英行, 畝見達夫, 寺野隆雄. インタラクティブ進化計算. 遺伝的アルゴリズム 4, 第 11 章, pp. 325–361. 産業図書, 2000.
- 7) 高木英行, 畝見達夫, 寺野隆雄. 対話型進化計算法の研究動向. 人工知能学会誌, Vol. 13, No. 5, pp. 692–703, Sep. 1998.
- 8) D.E.Goldberg. *Genetic Algorithms in Search Optimization and Machine Learnig*. Addison-Wesley, 1989.
- 9) M. Osaki, H. Takagi, and K. Oya. An input method using discrete fitness values for interractive GA. *J. of Intelligent and Fuzzy Systems*, Vol. 6, No. 1, pp. 131–145, 1998.
- 10) 高木英行, 大宅喜美子, 大崎美穂. 対話型遺伝的アルゴリズムのインターフェース改善手法の提案と評価. 第 12 回ファジィシステムシンポジウム, 東京, pp. 513–516, 1996.
- 11) K. Aoki, H. Takagi, and N. Fujimura. Interactive ga-based design support system for lighting design in computer graphics. *Int'l Conf. on foft Computing, Iizuka, Fukuoka, Japan*, pp. 533–536, 1996.
- 12) 青木研, 高木英行. 対話型 GA による 3 次元 CG ライティング支援. 電子情報通信学会論文誌, Vol. J81-D-2, No. 7, pp. 1601–1608, 1998.
- 13) 青木研, 高木英行. 3 次元 CG におけるライティングデザイン支援. 第 13 回ファジィシステムシンポジウム, 富山, pp. 311–314, 1997.

- 14) K. Aoki and H Takagi. 3-d cg lighting with an interactive ga. *1st Int'l Conf. on Conventional and Knowledge-based Intelligent Electronic Systems*, pp. 296–301, 1997.
- 15) Ken Lang. Newsweeder: Learning to filter netnews. In *in Proceedings of the 12th International Machine Learning Conference (ML95, 1995)*.
- 16) G. Linden, B. Smith, and J. York. Amazon.com recommendations: item-to-item collaborative filtering. *Internet Computing, IEEE*, Vol. 7, No. 1, pp. 76–80, Jan/Feb 2003.
- 17) 義介山川, 研二北. Ec サイトにおける画像の類似性測定 : 『sudachi』の開発と商品推薦エンジンへの応用 (特別セッション 好みの計量). 日本行動計量学会大会発表論文抄録集, Vol. 36, pp. 5–6, 20080902.
- 18) Hideyuki Takagi, Sung-Bae Cho, and Toshihiko Noda. Evaluation of an iga-based image retrieval system using wavelet coefficients. In *IEEE Int'l Conf. on Fuzzy Systems (FUZZ-IEEE'99)*, pp. 1775–1780, Seoul, Korea, 1999.
- 19) Sung-Bae Cho. Towards creative evolutionary systems with interactive genetic algorithm. *Applied Intelligence*, Vol. 16, No. 2, pp. 129–138, 2002.
- 20) I. T. Jolliffe. *Principal component analysis*. Springer, 1986.
- 21) Misato Tanaka, Tomoyuki Hiroyasu, Mitsunori Miki, and Hisatake Yokouchi. Extraction of design variables using collaborative filtering for interactive genetic algorithms. *2009 IEEE International Conference on Fuzzy Systems Proceedings*, 8 2009.
- 22) Tomoyuki Hiroyasu, Nozomi Yamakawa, Fuyuko Ito, Mitsunori Miki, and Yasunari Sasaki. Discussion of the evaluation method and the generation method of the individuals for interactive genetic algorithm. Research note, IPSJ SIG Technical Report 2008-MPS-68, 2008.
- 23) Eshleman, L.J., Schaffer, J.D. *Real-Coded Genetic Algorithms and Interval-Schemata 2*, 第 2 巻, pp. 187–202. 1993.

## 付録：発表論文リスト

1. **田中 美里**, 廣安 知之, 三木 光範  
条件に基づく認可を実現するシングルサインオンシステム  
情報処理学会 第 70 回全国大会 発表 (2008.3.14)
2. **田中 美里**, 伊藤 冬子, 廣安 知之, 三木 光範  
対話型遺伝的アルゴリズムにおける確率モデル構築による子個体生成の検討  
人工知能学会 第 22 回全国大会 発表 (2008.6.13)

3. Tomoyuki Hiroyasu, **Misato Tanaka**, Mitsunori MIki, Hisatake Yokouchi, Fuyuko Ito  
Discussion of a Crossover Method using a Probabilistic Model for interactive Genetic Algorithm  
Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on advanced Intelligent Systems (SCIS&ISIS 2008) (2008.9.19)
4. **田中 美里**, 廣安 知之, 三木 光範, 横内 久猛  
協調フィルタリングを用いた対話型遺伝的アルゴリズムのための設計変数の抽出  
情報処理学会 第72回数理モデル化と問題解決・第15回バイオ情報学合同研究会 発表 (2008.12.18)
5. **田中 美里**, 廣安 知之, 三木 光範, 横内 久猛  
商品推薦のための対話型遺伝的アルゴリズムの設計変数の導出  
人工知能学会 第10回 AI 若手の集い (2009.6.4)
6. Tomoyuki Hiroyasu, **Misato Tanaka**, Mitsunori MIki, Hisatake Yokouchi  
Extraction of Design Variables using Collaborative Filtering for interactive Genetic Algorithms  
IEEE Computational Intelligence Society, Korean Institute of Intelligent Systems (FUZZ-IEEE2009)  
(2009.8.23)