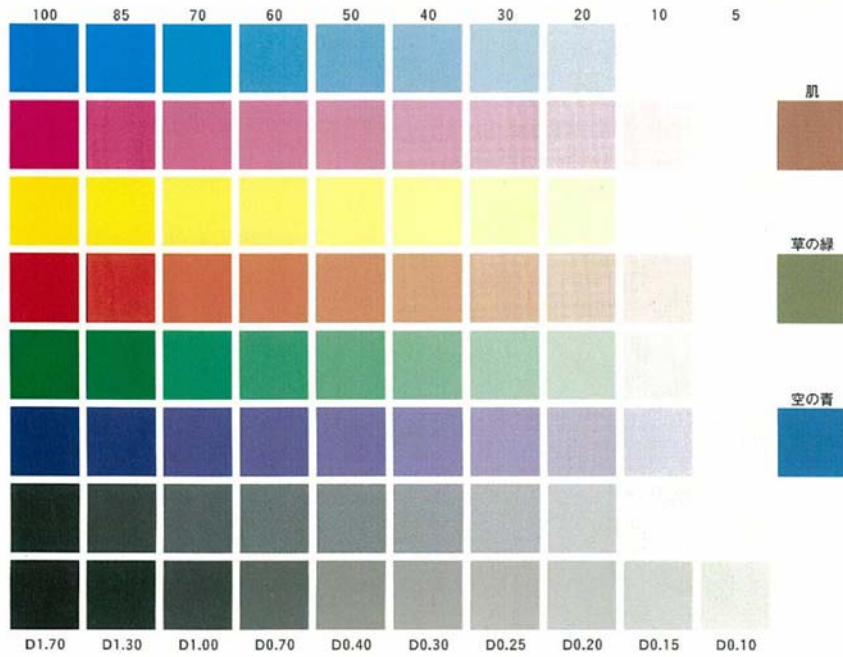


関西大学審査学位論文

対話型進化計算におけるユーザビリティを考慮した
評価インターフェースに関する研究



2013年 3月 31日

竹之内 宏

要旨

〈概要〉

本論文は、単独及び複数のユーザの感性を考慮した官能評価インタフェースの開発が期待されている対話型進化計算手法における、ユーザビリティの向上を目的とした評価インタフェースを提案し、まとめたものである。ユーザビリティの向上を目的とした評価インタフェースとして、本論文では、1個人のユーザの感性に基づいて解候補を評価する(1)トーナメント式評価手法及び対話型タブーサーチ、複数のユーザの感性を反映させることを目的とした(2)複数参加型トーナメント方式を提案している。

トーナメント式評価手法を用いた対話型遺伝的アルゴリズムは、ユーザが解候補をトーナメント形式で評価する手法である。本手法においては、実ユーザの代わりにビット列で作成された評価エージェントを用いた数値シミュレーション及び実ユーザを対象とした評価実験により、有効性を検証した。シミュレーション結果より、トーナメント式評価手法は、対話型進化計算手法の評価インタフェースとして一般に用いられている10段階評価手法と同等の進化性能を有することが確認されている。さらに、評価実験の結果より、トーナメント式評価手法は10段階評価手法と比較して、解評価が有意に行いやすく、ユーザが満足のいく解候補を生成できることが確認されている。

対話型タブーサーチは、対話型進化計算手法の進化計算アルゴリズムにタブーサーチ法を用いた手法である。本手法においては、評価エージェントを用いた数値シミュレーション及び実ユーザを対象とした評価実験により、有効性を検証した。シミュレーション結果より、対話型タブーサーチは、10段階評価手法を上回る進化性能を有することが確認されている。さらに、評価実験の結果より、対話型タブーサーチは10段階評価手法と比較して、解評価が有意に行いやすく、ユーザが満足のいく解候補を生成できることが確認されている。

複数参加型トーナメント方式を用いた対話型遺伝的アルゴリズムは、多くのユーザの投票により解候補の優劣を決定する手法である。本手法においては、まず、評価エージェントを用いた数値シミュレーションにより、有効性を検証した。シミュレーション結果より、複数参加型トーナメント方式においては、全評価エージェントの約60%が投票に参加していれば、全評価エージェントが投票に参加した場合と同等の進化性能を

得られることが確認されている。次に、複数参加型トーナメント方式において想定される問題を解決するため、2点嗜好法及び対戦結果推定を適用した数値シミュレーションを行った。シミュレーション結果より、2点嗜好法の適用はトーナメント対戦の効率化に有効であり、対戦結果推定の適用はユーザの解評価に対するモチベーション維持に有効であることが確認されている。さらに、複数参加型トーナメント方式を用いたWebサイトデザイン支援システムを構築し、実ユーザを対象とした評価実験を行った。評価実験の結果より、複数参加型トーナメント方式は全解候補を同時に提示する投票システムと比較して、解評価が有意に行いやすく、多くのユーザが満足のいく解候補を生成できることが確認されている。

〈各章の要旨〉

本論文は、8章で構成される。

第1章は序論であり、研究の背景、目的及び第2章以降の概要について述べている。

第2章では、ユーザが単独で解候補を評価する場合を想定した評価インタフェースについて述べている。本章では、まず、対話型進化計算手法の基本アルゴリズム及び技術的特徴と課題をまとめ、先行研究を概括している。次に、本論文で有効性を検証する、1個人のユーザの感性を用いることを想定したトーナメント式評価手法及び対話型タブーサーチについて述べている。本章では、実ユーザの代わりにビット列で作成された評価エージェントを用いた数値シミュレーションにより、トーナメント式評価手法及び対話型タブーサーチの有効性を検証した。シミュレーション結果より、トーナメント式評価手法は、対話型進化計算手法の評価インタフェースとして一般に用いられている10段階評価手法と同等の進化性能を有することが確認されている。また、対話型タブーサーチは、10段階評価手法を上回る進化性能を有することが確認されている。

第3章では、実ユーザによる対話型進化計算手法の評価インタフェースのユーザビリティ評価について述べている。本章では、トーナメント式評価手法の有効性を検証するため、音楽・動画・静止画を評価対象とした対話型進化計算システムを構築し、トーナメント式評価手法と10段階評価手法の性能比較を行った。評価実験の結果より、トーナメント式評価手法は、10段階評価手法と比較して、解評価が有意に行いやすく、ユーザが満足のいく解候補を生成できることが確認されている。次に、対話型タブーサーチの有効性を検証するため、ランニングシューズデザインシステムを構築し、対話型タブーサーチと10段階評価手法及びトーナメント式評価手法の性能比較を行った。評価実験の結果より、対話型タブーサーチは、10段階評価手法と比較して、解評価が有意に行いやすく、ユーザが満足のいく解候補を生成できることが確認されている。

第4章では、複数参加型トーナメント方式を用いた対話型遺伝的アルゴリズムについて述べている。本章では、まず、多人数参加型の対話型進化計算システムに関する先行研究について概括している。次に、複数参加型トーナメント方式のシステム概要について述べている。本章では、評価エージェントを用いた数値シミュレーションにより、複数参加型トーナメント方式の有効性を検証した。シミュレーション結果より、複数参加型トーナメント方式においては、全評価エージェントの約60%が投票に参加していれば、全評価エージェントが投票に参加した場合と同等の進化性能を得られることが確認されている。

第5章では、2点嗜好法を適用した複数参加型トーナメント方式について述べている。複数参加型トーナメント方式においては、実際のWebシステムを想定した場合、解評価に必要な投票数を獲得できないという問題が想定される。そこで、統計的手法である2点嗜好法をトーナメント対戦の勝敗決定に適用し、トーナメント対戦の効率化を図った。評価エージェントを用いた数値シミュレーションの結果より、解評価に必要な投票数が約80%減少され、2点嗜好法の適用はトーナメント対戦の効率化に有効であることが確認されている。

第6章では、対戦結果推定を適用した複数参加型トーナメント方式について述べている。トーナメント式評価手法においては、勝ち上がった解候補はユーザに複数回評価されなければならない。そのため、ユーザの解評価に対するモチベーションが低下してしまうことが想定される。このような問題を解決するため、対戦結果推定を適用することを試みた。対戦結果推定は、過去の対戦の投票結果や評価値の履歴を用いて、対戦結果が推定可能な解候補同士の対戦においてはユーザに評価を委ねず、システム側で対戦結果を決定する手法である。評価エージェントを用いた数値シミュレーションの結果より、対戦結果推定において決定された勝敗結果のうち、約70%が正しく推定され、対戦結果推定の適用は、ユーザの解評価に対するモチベーション維持に有効であることが確認されている。

第7章では、複数参加型トーナメント方式における実ユーザを対象とした評価実験について述べている。評価実験においては、複数参加型トーナメント方式を用いたWebサイトデザイン支援システムを用いて、有効性を検証した。評価実験の結果より、複数参加型トーナメント方式は、全解候補を同時に提示する投票システムと比較して、解評価が有意に行いやすく、多くのユーザが満足のいく解候補を生成できることが確認されている。

第8章は結論であり、得られた成果を要約し、今後の課題について述べている。

目次

第1章 序論	1
第2章 ユーザの評価負担軽減を目的としたIECインタフェース	9
2.1 緒言	9
2.2 対話型進化計算	10
2.3 トーナメント式評価手法	16
2.4 対話型タブーサーチ	20
2.5 数値シミュレーションによる性能評価	24
2.6 トーナメント式評価手法におけるシミュレーション結果	30
2.7 対話型タブーサーチにおけるシミュレーション結果	49
2.8 結言	53
第3章 実ユーザによるIECインタフェースのユーザビリティ評価	55
3.1 緒言	55
3.2 評価実験	56
3.3 実験システム	59
3.4 実験システムの評価インタフェース	63
3.5 トーナメント式評価手法における実験結果	74
3.6 対話型タブーサーチにおける実験結果	92
3.7 結言	101
第4章 複数参加型トーナメント方式	103
4.1 緒言	103
4.2 多人数参加型の対話型進化計算	104
4.3 複数参加型トーナメント方式	106
4.4 数値シミュレーションによる性能評価	109
4.5 シミュレーション結果	115
4.6 結言	146

第5章 2点嗜好法の適用	147
5.1 緒言	147
5.2 2点嗜好法を適用した複数参加型トーナメント方式	148
5.3 数値シミュレーションによる性能評価	151
5.4 シミュレーション結果	153
5.5 結言	175
第6章 対戦結果推定の適用	177
6.1 緒言	177
6.2 対戦結果推定を適用した複数参加型トーナメント方式	178
6.3 数値シミュレーションによる性能評価	183
6.4 シミュレーション結果	185
6.5 結言	200
第7章 複数参加型トーナメント方式を用いた Web サイトデザイン支援システム	201
7.1 緒言	201
7.2 Web サイトデザイン支援システム	202
7.3 評価実験	210
7.4 実験結果	213
7.5 結言	217
第8章 結論	219
付録	222
A 評価値の一致率に関する実験	223
B 評価エージェントの集団性の表現方法	229
謝辞	230
参考文献	232
研究業績	239

第1章 序論

近年、インターネット環境の普及や通信技術の発達により、Web サイトで商品を購入する機会が増えている。ユーザが商品を購入する際、高度経済成長期にみられた大量消費時代では商品の機能が重要項目であったのに対して、近年では個人の感性に合った商品のデザインが重要になっている。また、商品は音楽や動画、静止画など様々な形で表現されるようになってきている。このような個人ニーズの多様化に伴い、製品開発者が多様化した個人ニーズを分析し、把握することが困難になっている。このため、製造分野では人々の様々な潜在的ニーズを開拓し、商品を開発する技術が求められている。

人々の様々な潜在的ニーズの把握のためには、多くのユーザに対して製品の仕様やデザインに関するアンケートを行い、アンケート結果を統計的に解析するなどの手法がこれまで一般的であった。しかし、アンケートでは、ユーザに製品の仕様やデザインに関する多くの設問の回答が要求されるため、ユーザの負担が大きかったといった問題がある。さらに、アンケート解析では、人々の潜在的ニーズを把握するのに時間がかかり、その間にユーザの嗜好や流行が変化してしまう可能性がある。

そこで、本論文では、人とコンピュータがコミュニケーションをとりながら、人の感性に合うものを生み出していくとする対話型進化計算 (Interactive Evolutionary Computation: IEC) に着目する。IEC は、進化計算 (Evolutionary Computation: EC) において解候補を評価するために用いる目的関数を、人の直感的評価に置き換えた手法である [1]。このため、IEC は定量的に比較することが困難な商品のデザインの感性評価やユーザ個人の主観評価を必要とする製品のカスタマイズなど、人の好みや主観に依存する問題に有効とされている [2]。IEC を用いたシステムの研究では、ユーザの感性に合った画像を検索するシステム [3] や衣服のデザイン支援システム [4, 5]、配色支援システム [6]、電化製品のサイン音や音楽、ロボットの動作を生成するシステム [7]-[10]、補聴器フィッティングシステム [11]、インテリアレイアウト支援システム [12] など、様々なものをデザインするシステムが提案されている。しかし、これまでの IEC システムにおける評価インタフェースには、複数の解候補を同時にユーザに提示し、それぞれについて5段階や10段階で評価を求めるものが多い。この場合、ユーザは一度に提示

される多数の解候補を見比べて評価値を与えなければならない。このため、提示された解候補が類似している場合には、評価が困難になるといった問題がある。また、IECの応用システムを想定した場合の解候補は、大別して音楽や動画などの時系列データ及び静止画などの非時系列データとして表現される。特に、時系列データが評価対象の場合は、解候補は必然的に時系列に提示されることになるため、静止画とは異なり、ユーザは複数の解候補を順々に確認し、評価することになる。したがって、ユーザに大きな負担を強いることになり、実用の面での問題が多い。ユーザの感性を反映した製品開発などにおけるIECシステムの実用化には、ユーザの解評価負担軽減が重要な課題であると考えられる。

IECシステムにおいて、ユーザの評価負担を軽減する方法はいくつか存在するが、その中でも評価方法を単純にすることが効果的であると考えられている [2]。このような手法として、ユーザが提示された解候補の中から好みの解候補だけを選択する評価インタフェースが考えられる。しかし、単純に全解候補の中から好みの解候補のみを選択する場合でも、ユーザは解候補を順々に確認し、好みの解候補を選択しなければならず、ユーザの解評価負担が大きくなる。このため、好みの解候補を選択する評価インタフェースにおける提示解候補数は、IECの評価対象によって変わってくると考えられる。

IECの評価対象が音楽や動画などの時系列データの場合は、一度に提示する解候補を可能な限り少なくし、好みのものをユーザに選択してもらう手法が有効であると考えられる。これは、ユーザが時系列データを順々に確認し評価する際の負担を軽減するためである。このような手法の中に、一度に提示する解候補を2個に限定し、好みのものをユーザに選択してもらう一対比較法がある [13]。しかし、単純に全解候補の中から2つの解候補を取り出し、ユーザに評価を求めた場合、解候補数が n のとき、ユーザは $(n(n-1)/2)$ 回の評価を行わなければならない。例えば、解候補数が16個の場合、ユーザは毎世代120回もの一対比較評価を行うことになる。

そこで、本論文では、解候補をトーナメント形式で評価する対話型遺伝的アルゴリズム (Interactive Genetic Algorithm: IGA) に着目する。トーナメント式評価手法においては、解候補数が n のとき、ユーザの評価回数は $(n-1)$ 回で済む。例えば、解候補数が16個の場合、ユーザは毎世代15回の一対比較評価を行うことになる。このため、トーナメント式評価手法は一対比較法と比較して、ユーザの評価回数は充分少なく、解評価負担を軽減できると考えられる。トーナメント式評価手法を用いたIECは、これまでにアイデアの提案や手法自体の問題点が指摘されているが、解候補の進化性能やユーザの解評価負担に関する明確な有効性は検証されていない [14]-[16]。そこで、

本論文では、実ユーザの代わりにビット列で生成された評価エージェントを用いた数値シミュレーション及び実ユーザを対象とした評価実験により、次の3手法の性能評価を行う [17]。第1の手法は、IECの評価インタフェースとしてこれまで広く利用されている全解候補に10段階評価値を与える手法 (Normal IGA: NIGA) である。第2の手法は、二者択一方式による通常のトーナメント式評価手法 (Tournament evaluation with two levels: T2) である。第3の手法は、トーナメントの勝敗について大差・僅差の段階評価を設けるトーナメント式評価手法 (Tournament evaluation with four levels: T4) である。本論文ではまず、数値シミュレーションによりNIGA, T2及びT4の進化性能を定量的に評価する。これは、T2, T4ではNIGAよりユーザの解候補評価作業が単純になるため、解候補に正確な評価値が与えられず進化性能が低下することが懸念されるためである。次に、実ユーザを対象とした評価実験により、数値シミュレーションでは検証が不充分であった、実ユーザにおける解評価のしやすさや評価に要した時間及び解候補の進化に要した世代交代数を測定し、トーナメント式評価手法の有効性を検証する。

好みの解候補のみを選択するという評価インタフェースにおいては、IECの評価対象が静止画のような非時系列データの場合は、一度に提示する解候補は通常のIECシステムにおける提示解候補数である10~20個であっても、ユーザの解評価負担は軽減されることが考えられる。しかし、提示された解候補の中から好みの解候補を選択する評価インタフェースをIGAに適用した場合、解候補の正確な評価値を獲得することが困難になり、解候補集団が十分に進化しないという問題が想定される。そこで、IECのECアルゴリズムにタブーサーチ (Tabu Search: TS) を用いた対話型タブーサーチ (Interactive Tabu Search: ITS) を提案する [18]。TSは、1989年にF. Gloverによって考案された局所探索法である [19]。ITSにおいては、ユーザは提示された解候補の中から最も好みの解候補を1つ選択する。そして、ユーザが選択した解候補を基に近傍解候補を生成し、次世代の解候補とする。IECの解探索においてTSを用いた手法は、これまでに提案されている [20, 21]。しかし、これらの研究は、TSにおける解評価そのものをユーザに求める手法ではない。本論文におけるITSは、IECユーザの解評価負担の軽減を目的として、好みのものを1つ選択するのみという評価インタフェースをIECにおいて実現するものである。本論文では、数値シミュレーション及び実ユーザを対象とした評価実験により、ITSの有効性を検証する。ITSでは、TSのアルゴリズム上、解候補が局所的に収束するため十分な進化性能を得られないことが懸念される。そのため、まず数値シミュレーションにより、ITS及びNIGAの進化性能を定量的に評価する。次に、実ユーザを対象とした評価実験により、数値シミュレーションでは検証が不充分

であった，実ユーザにおける解評価のしやすさや評価に要した時間及び解候補の進化に要した世代交代数を測定し，ITSの有効性を検証する。

さらに，本論文では，Web上から多くのユーザの感性を投票により獲得し，IGAの解評価に用いることを想定したシステムを提案する．IECにおけるユーザの解評価負担軽減を支援する評価インタフェースは，ユーザ個人の感性を反映したものでだけでなく，多くのユーザの感性を反映したものを生成する際に有効であると考えられる．なぜなら，多くのユーザの感性評価を獲得する際には，投票のような単純な評価作業が求められるためである．これは，Yahoo!やFacebook，Twitterなど多くのWebサイトで，Webサイトを閲覧したユーザが掲載された記事に対して，共感の有無に関する投票ができるシステムが普及しており，多くの投票が獲得されていることから明らかである．本論文では，このように多くのユーザからの投票を利用することを想定したIGAシステムとして，複数参加型トーナメント方式(Tournament evaluation by multiple people's Kansci: TBMP)を用いたIGAを提案する [22]．TBMPは，トーナメント形式で多くのユーザの投票による勝ち抜き戦を行い，解候補の優劣を決定する手法である．TBMPにおけるユーザの解評価作業は，提示された2つの解候補の中で，好みの解候補に投票するのみとなり，ユーザの解評価負担は軽減できると考えられる．また，一定の期間内にユーザの投票を受け入れるため，全てのユーザが同時に投票に参加する必要がなく，ある程度の投票数が得られれば，多くのユーザの感性を反映した解候補の生成できるようになると考えられる．

なお，本論文では，通常の複数参加型トーナメント方式をTBMP1とする．TBMP1において，まず，数値シミュレーションにより解候補の進化性能に関する有効性を検証する．これは，T2，T4と同様に，TBMPではユーザの解候補評価作業が単純になるため，解候補に正確な評価値が与えられず解候補が十分に進化しないことが懸念されるためである．また，TBMP1においては，(1)想定した投票数を獲得できず，トーナメント対戦を進行できないこと，(2)勝ち上がった解候補は複数回評価されなければならず，ユーザの解評価に対するモチベーションが低下してしまうことが想定される．これらの問題は，ユーザの解評価負担を発生させるおそれがある．そこで本論文では，TBMP1の改良手法であるTBMP2及びTBMP3を提案する．

TBMP2は，2点嗜好法を適用した複数参加型トーナメント方式である [23]．(1)の問題を解決するためには，投票数の効率的な利用，すなわちトーナメント対戦の効率化が求められる．トーナメント対戦の効率化のためには，一定の投票数を獲得不可能であってもトーナメント対戦が停滞せず，少数の投票において解候補の優劣を決定できることが求められる．そこで，TBMP1においてトーナメント対戦の効率化を実現

するため，各対戦の勝敗決定に統計的手法である2点嗜好法の適用を試みる．TBMP2では，数値シミュレーションにおいて解候補の進化性能に関する有効性を検証する．

TBMP3は，対戦結果推定を適用した複数参加型トーナメント方式である [24]．(2)の問題を解決するためには，遺伝子パターンが類似した解候補同士の対戦や勝ち上がった解候補の複数回にわたる評価をユーザに強いることを避け，解候補集団の初期収束性の向上を図る必要がある．そこで，TBMP1において対戦結果推定の適用を試みる．対戦結果推定は，過去の対戦の投票結果や評価値の履歴を用いて，対戦結果が推定可能な解候補同士の対戦の評価はユーザに委ねず，システム側で対戦結果を決定する手法である．TBMP3でも同様に，数値シミュレーションにおいて解候補の進化性能に関する有効性を検証する．

また，本論文では，TBMP1を用いたWebサイトデザイン支援システムを構築し，実ユーザを対象とした評価実験を行う．本実験では，解評価のしやすさや評価時間などを測定し，TBMP1の実ユーザに対する有効性を検証する [25]．本システムにおいては，多くのユーザがWeb上で投票を行い，解候補を評価する．

本論文では，T2，T4，ITS及びTBMPを提案し，IECインタフェースのユーザビリティ，すなわち解評価のしやすさの向上の面からユーザの解評価負担を軽減することを目的とする．従来のIECシステムでは，解候補が5段階や10段階で評価されるため，ユーザの解評価負担が大きく実用化が困難であった．しかし，T2，T4，ITS及びTBMPの評価インタフェースによりユーザビリティの向上が実現され，ユーザの解評価負担が軽減されると考えられる．このためには，従来のIECシステムより解評価がしやすいことに加え，ユーザが満足 of いくものを得るのに要する時間や世代交代数が少ないこと，すなわち解候補の初期収束性が高いことが求められる．そのため，本論文では解評価のしやすさ及び解候補の進化性能の面から，T2，T4，ITS及びTBMPの有効性を検証する．

本論文は，8章で構成される．

第2章では，ユーザが単独で解候補を評価する場合を想定した評価インタフェースについて述べる．本章では，まず，IECの基本アルゴリズム及び技術的特徴と課題をまとめ，先行研究を概括する．次に，1個人のユーザの感性を用いることを想定したトーナメント式評価手法及びITSについて述べる [17, 18]．本章では，実ユーザの代わりにビット列で作成された評価エージェントを用いた数値シミュレーションを行い，トーナメント式評価手法及びITSの解候補の進化性能に関する有効性を検証する．トーナメント式評価手法の数値シミュレーションにおいては，NIGA，T2，T4，ITSの性能比較を行う．ITSの数値シミュレーションにおいては，ITS，NIGAの性能比較を行う．

第3章では、実ユーザによるIECインタフェースのユーザビリティ評価について述べる。本章では、トーナメント式評価手法及びITSにおいて、実ユーザにおける解評価のしやすさや評価に要した時間及び解候補の進化に要した世代交代数を測定し、トーナメント式評価手法及びITSの実ユーザに対する有効性を検証する [17, 18]。トーナメント式評価手法の評価実験では、音楽・動画・静止画を評価対象としたIECシステムを構築し、NIGA, T2, T4の性能比較を行う。ITSの評価実験では、静止画を評価対象としたIECシステムを用いて、NIGA, T2, T4, ITSの性能比較を行う。

第4章では、多くのユーザの投票により解候補を評価する複数参加型トーナメント方式であるTBMP1について述べる [22]。本章では、まず、多人数参加型の対話型進化計算システムに関する先行研究について概括する。次に、複数参加型トーナメント方式のシステム概要について述べる。本章では、評価エージェントを用いた数値シミュレーションを行い、TBMP1の解候補の進化性能に関する有効性を検証する。本シミュレーションにおいては、評価エージェントの投票参加割合に関する検討、通常の投票方式 (General Voting Method: GVM) を比較手法とした性能比較を行う。

第5章では、2点嗜好法を適用した複数参加型トーナメント方式であるTBMP2について述べる [23]。TBMP1においては、実際のWebシステムを想定した場合、解評価に必要な投票数を獲得できないといった問題が想定される。そこで、統計的手法である2点嗜好法をトーナメント対戦の勝敗決定に適用し、トーナメント対戦の効率化を図る。本章では、数値シミュレーションにおいて、2点嗜好法により対戦解候補の獲得投票数に有意差があると判定された対戦数、判定誤り率、解評価に必要な投票数の検討及びTBMP1を比較手法とした性能比較を行い、TBMP2のトーナメント対戦の効率化に対する有効性を検証する。

第6章では、対戦結果推定を適用した複数参加型トーナメント方式であるTBMP3について述べる [24]。トーナメント式評価手法においては、勝ち上がった解候補をユーザが複数回評価しなければならないという問題がある。また、IGAにおいては、世代交代を経るに連れて、遺伝子パターンが類似した解候補が増加する。そのため、ユーザの解評価に対するモチベーションが低下してしまうことが想定される。そこで、対戦結果推定の適用により、ユーザの解評価に対するモチベーションの維持を試みる。本章では、数値シミュレーションにおいて、対戦結果推定における推定率及び推定誤り率に関する検討、TBMP1を比較手法とした性能比較を行う。

第7章では、TBMP1を用いたWebサイトデザイン支援システムについて述べる [25]。本章では、まず、Webサイトデザイン支援システムの概要、Webサイトデザインの構成、評価インタフェースについて述べる。そして、本システムを用いた実ユーザを対

象とした評価実験について述べる。本実験においては、トーナメント対戦の進行に要した時間や各ユーザの投票数、生成された解候補に対する満足度及び各システムにおける解評価のしやすさなどを調査し、TBMP1の実ユーザに対する有効性を検証する。

第8章は結論であり、得られた成果を要約し、今後の課題について述べる。

第2章 ユーザの評価負担軽減を目的としたIECインタフェース

2.1 緒言

本章では、ユーザが単独で解候補を評価する場合を想定した評価インタフェースについて述べる。

まず、IECの基本アルゴリズム及び技術的特徴と課題をまとめ、先行研究を概括する。次に、1個人のユーザの感性を用いることを想定したトーナメント式評価手法であるT2, T4, ITSについて述べる。トーナメント式評価手法においては、ユーザが解候補をトーナメント形式で評価する [17]。ITSは、IECのECアルゴリズムにTSを用いた手法である [18]。ITSにおけるユーザの解評価は、提示された解候補の中から好みの解候補を1つ選択するのみとなる。

本章では、実ユーザの代わりにビット列で作成された評価エージェントを用いた数値シミュレーションにより、トーナメント式評価手法及びITSの解候補の進化性能に関する有効性を検証する。本シミュレーションにおける評価エージェントは、実ユーザの評価特性を模倣した多峰性感性空間を有する。トーナメント式評価手法の数値シミュレーションでは、NIGA, T2, T4の性能評価を行う。ITSの数値シミュレーションでは、ITS及びNIGAの性能評価を行う。性能評価においては、評価エージェントの解評価に対して評価値の揺らぎを発生させた場合における、各手法の進化性能を調査し、有効性を検証する。

2.2 対話型進化計算

本節では、IECの基本アルゴリズム及び技術的特徴と課題をまとめ、先行研究を概括する。

2.2.1 対話型進化計算の概要

IECは、ECにおいて解候補を評価するために用いる目的関数を、人の直感的評価に置き換えた手法である [1]。このため、定量的に評価することが比較的困難な製品のデザインなどの感性評価や、ユーザ個人の主観評価が必要となる製品のカスタマイズなど、人の好みや主観が考慮されるべき問題に対して有効とされている [2]。

IECを用いたシステムの研究では、ユーザの感性に合った画像を検索するシステム [3] や衣服のデザイン支援システム [4, 5]、配色支援システム [6]、電化製品のサイン音や音楽、ロボットの動作を生成するシステム [7]-[10]、補聴器フィッティングシステム [11]、インテリアレイアウト支援システム [12] など、様々なものをデザインするシステムが提案されている。また、芸術分野への応用も行われている。代表的な例として、Dawkins が提案した Biomorph [26] や Unemi が提案している SBArt [27] がある。

IECの代表的な手法として、IGAがある。遺伝的アルゴリズム (Genetic Algorithm: GA) は、1975年に J.H. Holland らによって提案された近似解を探索するメタヒューリスティックアルゴリズムである [28]。IGAは、GA処理における解候補評価をユーザが自らの感性に基づいて行い、解候補を最適化する手法である。図 2.1 に IGA の流れを示す。まず、初期遺伝子集団を生成する。次に、解候補をユーザに提示する。ユーザは、自身の感性に基づいて解候補を評価する。ユーザの評価が終了すると、各解候補の評価値を基に、選択 (Selection)・交叉 (Crossover)・突然変異処理 (Mutation) が行われ、新たな解候補を生成し、再びユーザに提示する。このような処理を繰り返し、ユーザが満足 of いく解候補を生成する。

IGAは、人とGAが心理空間と特徴パラメータ空間の写像関係に基づいて協調的に解候補を探索し、最適化を行うシステムである。ユーザは、心理空間上のターゲットとシステムにより提示された解候補との距離に応じて解候補を評価し、GAはユーザの示した心理尺度距離を評価値として特徴パラメータ空間を探索する。IGAにおいては、ユーザがIECシステムにより提示された音声や動画、画像を評価する際の肉体的心理的負担が大きな問題となっている。

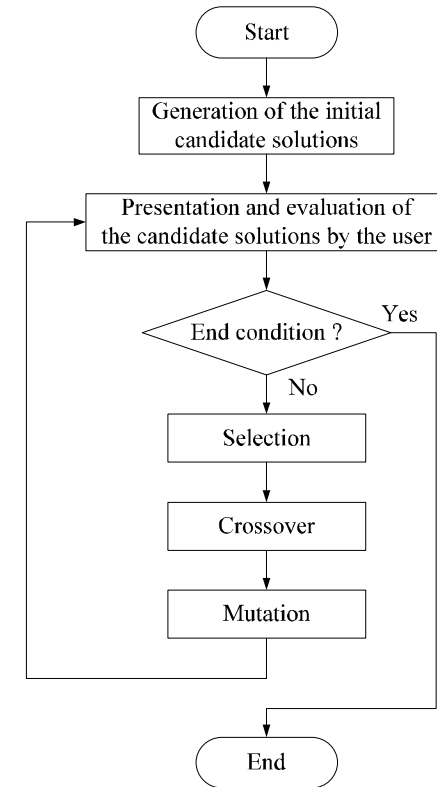


図 2.1: IGA の流れ

2.2.2 対話型進化計算の技術的課題と解決策

IEC の抱える最大の問題は、解候補を評価するユーザの疲労をどのように軽減するかということである。この問題を解決するため、IEC においてはユーザに提示する解候補数と世代交代数を通常の EC 探索に比べて、非常に少なく設定しなければならない。通常の EC 探索の場合、解候補数は数 100 個、世代交代数は数 100 世代といった探索にも耐えられる。しかし、IEC の場合、10~20 個、10~20 世代が広く用いられている。このため、IEC においては、解候補の良し悪しが最初に生成された解候補へ大きく依存する。すなわち、最初に良質な解候補が生成されるかどうかの運の要素が大きくなる。また、交叉する解候補の組み合わせ数の減少により、解候補の多様性の維持が困難になる。このようなことが進化の妨げとなり、解の収束を遅らせることになる。さらに、評価対象が音声や動画の場合は、解候補は必然的に時系列に提示される。この場合、ユーザは記憶にある過去の解候補との良し悪しを比較しなければならない。そのため、音声や動画におけるユーザの評価負担は、一目で相対的に比較できる静止画像に比べ、はるかに大きいと考えられる。

このような課題を解決し、IEC におけるユーザの解評価負担を軽減するため、次のような研究が行われている。

(1) 入力インタフェースの改善

ユーザが解候補を評価し、主観的評価値を入力して IEC システムにフィードバックする過程において、解評価のしやすさはユーザの心理的負担に大きく影響を与える。高木らは、各解候補に 5 段階や 7 段階の評価値を与える手法について、実ユーザによる評価実験を行っている [29]。5 段階や 7 段階の評価値は、100 段階評価などと比較して量子化ノイズを有するため、解候補の収束が悪化してしまうおそれがある。しかし、IGA において実用的な世代交代数である 10~20 世代では、量子化ノイズによる進化性能低下に対する影響は小さいと報告されている。

また、ユーザの解評価負担を軽減するためには、解候補の評価方法を単純にすることが効果的であると考えられている [2]。このような手法を実現するため、差分進化 (Differential Evolution: DE) や粒子群最適化 (Particle Swarm Optimization: PSO) を IEC の EC アルゴリズムに応用する手法が提案されている。

DE を IEC に応用した対話型差分進化 (Interactive Differential Evolution: IDE) は、ユーザが一対比較により解候補を評価する IEC 手法である [30, 31]。シミュレーションにおける検証より、IDE は IGA よりも高い進化性能を有することが確認されている。

PSO を IEC に応用した対話型粒子群最適化 (Interactive Particle Swarm Optimization: IPSO) は、IDE と同様にユーザが一対比較により解候補を評価する IEC 手法である [32]。IPSO はアルコール発酵パラメータの最適化問題に応用され、解候補の進化性能に関する有効性が確認されている。

(2) 評価値予測に基づく解候補提示インタフェース

IEC において生成された解候補をランダムに提示するのではなく、提示する順序に工夫を凝らせば、ユーザが評価しやすくなると考えられる。このような手法として、ユーザの評価特性をシステム側で予測し、ユーザの評価が高いと予測される解候補を生成し提示する評価値予測がある。評価値予測は、近接する解候補同士で大まかな比較ができ、ユーザの評価負担軽減に有効であると考えられる。

長尾らは、解候補間のユークリッド距離により、ユーザの評価を予測する目的関数を作成し、200 個体に対して GA 演算を行い、上位 10 個体をユーザに提示する手法を提案している [33]。佐藤らは、ユーザの評価特性をニューラルネットワーク (Neural Network: NN) で学習する手法を提案し、解候補の進化性能に関する有効性を検証している [34]。大崎らは、解候補間のユークリッド距離により、解候補の評価値を予測する手法及びその改善手法を提案している [35, 36]。渡辺らは、解候補同士の評価値の相関係数より算出した評価値推論精度に基づいて、ユーザが評価する解候補数を変化させる手法を提案している [37]。この手法では、ユーザの実評価回数を減少できることが確認されている。岩下らは、ユーザの評価特性を自動推定するモデルを構築し、一定数の解候補をユーザに代わって評価し、解候補の総数を保ったままユーザが評価する解候補数を削減する手法を提案している [38]。この手法では、ベンチマーク関数を用いたシミュレーションにより、解候補の進化性能に関する有効性が確認されている。

(3) 通常 EC と IEC の統合

IEC においては、1 世代あたりにユーザに提示できる解候補数は、通常の EC に比べて大きく減らされてしまう。そのため、少ない提示解候補数において、ユーザの好みの領域に収束できる技術が求められている。提示できる解候補数が少なくなる原因には、ユーザの疲労の他にも評価インタフェース上の空間的な制限やユーザが記憶可能な解候補数の制限がある。このような問題を解決するため、通常の EC と IEC を統合し、通常の EC 並みの解候補数を確保する取り組みがある。例えば、感性検索エージェントの最適化における水面下個体と相互評価を用いた IEC 手法 [39] や 2 次元コード装飾において通常 EC と IEC の統合を実装したシステム [40] な

どがあり、解候補の進化性能に関する有効性が確認されている。

(4) オンライン知識組み込み

一般にIECにおける解評価は、IECシステムより提示された解候補にユーザが評価値を与えるのみである。しかし、ユーザが解評価の途中で気づいた知見を、IECの解候補生成において積極的に反映することにより、良好な進化性能が期待できる。このような手法を、オンライン知識組み込みという。IECにおける解候補の表現型と遺伝子型に対応関係がある場合、良い部分と悪い部分があるという情報をIECシステムに与え、良い部分の表現型に反映される遺伝子を固定し、以後の探索を行うことで探索空間の次元数を減らすことができる。Sugaharaらの提案した浴衣デザインシステムでは、ユーザが任意の世代において、浴衣の帯の色を突然変異によりカスタマイズできる機能を組み込んでいる [41]。

(5) 解候補の多様性維持

一般にユーザの嗜好は、多峰性を有すると考えられる。これは、例えばユーザが衣服のデザインに対して、「クールな色合いが好き」「カジュアルな色合いも好き」のように複数の好みを示す場面から想像できる。このようなユーザの多峰性を有する感性に合致した解候補を提示するIEC手法は、いくつか提案されている [42]-[44]。

Nishinoらは、免疫アルゴリズム (Immune Algorithm: IA) を用いたIEC手法を提案している [42]。この手法では、IAの適用により、解候補の多様性が確保されることが確認されている。

伊藤らは、クラスタリング手法を用いたユーザの嗜好の多峰性に対応した解候補生成手法を提案している [43]。この手法では、クラスタリング手法を用いることにより、ユーザの嗜好に合った解候補を提示できることが確認されている。

雨宮らは、事前にユーザの嗜好をサポートベクターマシン (Support Vector Machine: SVM) により学習し、解探索を行うIEC手法を提案している [44]。この手法では、SVMの適用により、ユーザの嗜好に合った解候補を提示できることが確認されている。

これらの他にも、IECにおける解探索過程を可視化し、ユーザの解探索に役立てようとする手法が提案されている [45, 46]。

山代らは、看護師スケジューリング問題 (Nurse Scheduling Problem: NSP) において島モデル型多目的GAを適用し、ファジィクラスタリングを用いてIECの解探索過

程を可視化する手法を提案している [45]。この手法では、解探索過程の可視化がユーザの解探索負担軽減に有効であることが示唆されている。

Hayashidaらは、自己組織化マップ (Self-Organizing Map: SOM) を用いて、IECにおける解探索過程を可視化する手法を提案している [46]。この手法では、SOMによる解探索過程の可視化により、解候補の収束性が向上することが確認されている。

しかし、(2)~(5)における評価インタフェースは、解候補を5段階や10段階で評価するものである。そのため、IECにおける世代交代数や解候補の評価回数を削減できても、解評価において評価値を与える作業そのものは(2)~(5)と同様である。したがって、解評価作業そのものにおけるユーザの評価負担軽減は実現されていない。また、(1)のIDEやIPSOの対比較評価インタフェースにおいても、解候補の進化性能に関する有効性は示されているが、音楽や動画などを評価対象とした場合の検証は行われていない。そこで本論文では、IECの解評価作業においてユーザが好みの解候補を選択するという評価インタフェースとして、T2, T4, ITS及びTBMPを提案する。

各手法について、2.3節でトーナメント式評価手法に関する先行研究及びT2, T4の概要、2.4節でITSに関する先行研究及び概要について述べる。さらに、第4章で多人参加型のIECに関する先行研究及びTBMPの概要について述べる。

2.3 トーナメント式評価手法

本節では、トーナメント式評価手法について述べる。まず、トーナメント式評価手法及び一対比較法をIECの解評価に用いた先行研究について概括する。そして、本論文で解候補の進化性能に関する有効性を検証するT2及びT4について述べる。

2.3.1 関連研究

IECシステムの評価インタフェースには、複数の解候補を同時にユーザに提示し、それぞれについて評価を求めるものが多い。これに対し、1度に提示する解候補を2個に限定し、好みのものをユーザに選択してもらう一対比較法がある[13]。このようにユーザの解評価を単純にすることは、ユーザの解評価負担軽減に効果的であると考えられている[2]。

しかし、単純に全解候補の中から2つの解候補を取り出しユーザに評価を求めた場合、解候補数が n のとき、ユーザは $(n(n-1)/2)$ 回の評価を行わなければならない。例えば、解候補数が16個の場合、ユーザは毎世代120回もの評価を行うことになる。そのため、一対比較法の場合は、ユーザの評価回数が膨大になり、解評価負担が大きくなってしまう。Johansonの研究では、解候補の優劣を決定するマージソートのための一対比較をユーザが行う手法が提案されている[47]。しかし、膨大な回数の一対比較をユーザに強いるという問題がある。

そこで、評価回数を削減するため、解候補をトーナメント形式で評価し、相対的な評価値を獲得するトーナメント式評価手法が提案されている[14]-[16]。トーナメント式評価手法においては、解候補数が n のとき、ユーザの評価回数は毎世代 $(n-1)$ 回で済む。例えば、解候補数が16個の場合、ユーザは毎世代15回の評価を行うことになる。このため、トーナメント式評価手法は一対比較法と比較して、ユーザの評価回数は充分少なく、解評価負担を軽減できると考えられる。

しかし、トーナメント式評価手法を用いた場合、トーナメントの組み合わせや各対戦の勝敗により、全解候補に必ずしも適正な評価値が与えられないおそれがある。そのため、解候補が十分に進化しないという問題も想定される。Limらの研究では、これらの問題について言及しているが、IECの進化性能にどの程度の悪影響を及ぼしているかについては検証されていない[14]。

Angelineらの研究では、通常のトーナメントの二者択一方式を改良し、トーナメントの勝敗について大差・僅差などの段階を設ける評価手法が提案されている[15]。この手法においては、対戦結果に基づいて解候補間の相対的な評価値を獲得することが可

能となり、通常のトーナメントに比べて解候補に適正な評価値を与えることができると考えられる。しかし、アイデアの提案は行われているが、解候補の進化性能やユーザの解評価負担軽減に関する有効性は検証されていない。

Watanabeらの研究では、補聴器フィッティングシステムを用いて、トーナメント式評価手法の有効性を検証している[16]。評価実験の結果より、二者択一方式のトーナメント式評価手法を実装したIECは、5段階評価手法を用いたIECよりもユーザの評価負担が軽減されることが報告されている。しかし、段階的な評価を行うトーナメント式評価手法の有効性は検証されていない。

このように、トーナメント式評価手法はこれまでに提案されているが、IECの評価対象の違いや段階的な評価を行うトーナメント式評価手法に関する明確な有効性は検証されていない。本論文ではこれらの知見を踏まえ、IEC応用システムの主な評価対象である音楽・動画・静止画を用いた場合のトーナメント式評価手法の有効性を、ユーザビリティ性及び解候補の進化性能の面から検証する。

2.3.2 T2

本項では、通常のトーナメント式評価手法であるT2について述べる[17]。トーナメント式評価手法においては、ユーザは一度の評価で2つの解候補の優劣判定を行うのみとなり、5段階や10段階評価手法と比べ、ユーザの評価負担を軽減できると考えられる。

図2.2にトーナメント式評価手法による解評価方法を示す。まず、IGAにおいて生成された解候補を、ランダムにトーナメント表に配置する。ランダムな順序で並べられた16個の解候補を $p_1 \sim p_{16}$ とする。次に、図2.2のトーナメント表の対戦組み合わせに基づいて、ユーザは対戦する解候補の相対評価を行い優劣を判定する。まず、 p_1 と p_2 、次に p_3 と p_4 を一対比較により評価して優劣を決定する。図2.2では、 p_2 及び p_4 が勝利し、第2回戦に進んでいる。 p_{15} と p_{16} の対戦まで評価が終了すれば、引き続き第2回戦である p_2 と p_4 の対戦の評価を一対比較により行う。決勝戦まで評価が終了すれば、勝ち上がり対戦数に応じて各解候補に評価値を与える。図2.2では、まず、優勝した解候補である p_4 に10点、準優勝した解候補である p_{15} に8点が与えられる。以降、準決勝で敗退した解候補である p_7, p_{10} に6点、第2回戦で敗退した解候補である p_2, p_5, p_{12}, p_{13} に4点、初戦で敗退した残りの8個の解候補に2点が与えられる。また、本章の数値シミュレーション及び第3章の評価実験におけるT2は、これらの要領で解候補に評価値を与えている。

トーナメント式評価手法では、集団中の最も優秀な解候補に対して確実に高い評価

値が与えられるため、集団中のエリート解候補が高確率で次世代に子孫を残すことが可能である。また、16個の解候補それぞれを10段階で評価する場合、ユーザが評価値を与える回数は1世代あたり16回であるが、トーナメント式評価手法では評価回数は15回で済む。ただし、トーナメント式評価手法では、1回の評価について2つの解候補を比較することになる。そのため、評価を行うために解候補を確認する回数は、評価回数の2倍の30回となる。しかし、16個の解候補それぞれを10段階で評価する場合でも、各解候補を1回だけ確認して評価値を与え、そのまま次世代に進むことは少ない。一般には、何回か解候補の確認を行い、評価値を修正するなどの作業が行われることが多い。さらに、音声や動画などが評価対象の場合、ユーザは16パターンもの解候補を順番に確認し、評価を与えるのは非常に困難であると考えられる。このような場合において、T2は2パターンの解候補を比較して優劣を判定するだけでよい場合、ユーザの評価負担は軽減できると考えられる。

一方、T2には、トーナメント対戦の組み合わせが各解候補に与えられる評価値に大きく影響するという問題がある。例えば、トーナメントの早い段階で優秀な解候補と対戦し、敗退した解候補は、本来は解候補集団中の上位の評価が得られる解候補であったとしても早期に敗退しているため、最終的に低い評価値が与えられてしまう。また、評価の低い解候補ばかりがトーナメントの片方のグループに集中した場合は、本質的に劣っている解候補が決勝戦まで進出してしまい、最終的に高い評価値が与えられてしまう。このような評価の矛盾は、集団全体の進化を多少なりとも妨げてしまうと考えられる。この問題を解決するため、各対戦の段階評価を取り入れたトーナメント式評価手法が提案されている [15]。

2.3.3 T4

T2における評価誤差の問題は、ユーザによる解評価が一对比較による優劣判定のみであることから生じている。すなわち、対戦した解候補の相対評価の大差と僅差を区別せず、勝敗のみを評価値に反映していることが問題であると考えられる。例えば、ある解候補が初戦で敗退したが、そのときに対戦した相手が最終的に優勝したとする。この場合、初戦での対戦結果が僅差であった場合は、敗退した解候補にもある程度高い評価値を与えるべきである。このような手法として、解候補の一对比較におけるユーザの相対評価の大差・僅差の情報を取得し、解候補の評価に反映させる T4がある [17]。

T4においては、まず、T2と同様に、生成された解候補をトーナメント表に配置する。次に、ユーザは提示される2つの解候補 p_i 及び p_j について優劣の評価を行うが、このときに、「大差で p_i の勝利」、「僅差で p_j の勝利」という要領で、大差・僅差の情

報を付加して優劣を決定する。トーナメントの全対戦の勝敗が確定した後、大差・僅差の情報に基づいて各解候補に評価値を与える。

具体的な評価値の与え方は、次のとおりである。まず、T2と同様に、優勝した解候補に10点を与える。次に、決勝戦で敗退した解候補には、大差負けの場合は $(10 - a)$ 点、僅差負けの場合は $(10 - b)$ 点を与える。ここで、 a 及び b を評価減算値とする。ただし、 a 及び b は、 $(10 > a > b)$ を満たす正の整数とする。各対戦において敗退した解候補には、勝者の評価値を q 点とすると、大差負けの場合は $(q - a)$ 点、僅差負けの場合は $(q - b)$ 点を与える。

図 2.2 の対戦結果においては、優勝した解候補である p_4 と対戦して敗退した解候補である p_{15}, p_7, p_2, p_3 には、各対戦におけるユーザの評価が「大差で p_4 の勝ち」の場合は $(10 - a)$ 点、「僅差で p_4 の勝ち」の場合は $(10 - b)$ 点を与える。例えば、第2回戦における p_2 と p_4 の対戦において「僅差で p_4 の勝ち」であった場合、 p_2 の評価値は $(10 - b)$ 点となる。このとき、初戦で p_1 が p_2 に大差で敗れていた場合、 p_1 の評価値は $((10 - b) - a)$ 点となる。

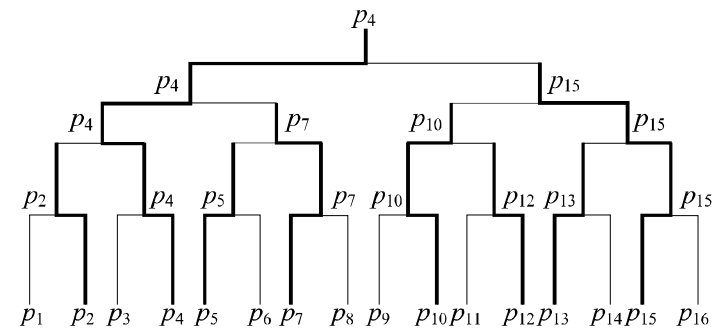


図 2.2: トーナメント式評価手法における解評価方法

2.4 対話型タブーサーチ

本節では、ITSについて述べる。まず、ITSと同様に好みのものを選択することで解候補を評価するIECの先行研究について概括する。そして、TS及びITSの概要について述べる。

2.4.1 関連研究

ITSのように、提示された全解候補の中から好みの解候補のみを選択するという評価インタフェースは、IGAなどにおいても実現可能である。このような手法は、これまでいくつか提案されている [48]-[50]。

石野らは、模擬育種法と呼ばれる評価手法を提案している [48]。模擬育種法においては、ユーザは提示された解候補の中から好みのものだけを選択する。すなわち、選ぶ(評価値1)、選ばない(評価値0)の2段階で解候補に評価値を与える。このため、ユーザは単純で直感的に解候補を評価することができ、解評価負担は小さいと考えられる。しかし、1または0かの極めて粗い評価であるため、評価値1を得た解候補だけが交叉に選択される解候補の大多数を占めてしまい、局所解に陥りやすく解候補が十分に進化しないという欠点がある。

Oliverらは、ユーザが提示された解候補の中から好みの解候補を複数個選択する評価インタフェースを提案している [49]。この手法では、ユーザが選択した解候補数によって、GAにおける次世代の解候補の生成方法を変化させている。しかし、この手法の解候補の進化性能に関する明確な有効性は検証されていない。

鷺見らは、Oliverらと同様の評価インタフェースを提案している [50]。この手法では、Oliverらの手法と次世代の解候補の生成方法が異なる。この手法は、評価実験において、鷺見らの従来手法よりも探索世代数を減少できることが確認されている。

また、IECの解探索において、TSを用いた手法がいくつか提案されている [20, 21]。Klauらは、対話型最適化に関する問題として巡回セールスマン問題(Traveling Salesman Problem: TSP)やスケジューリング問題を対象とし、Human-guided TSを提案している [20]。宗本らは、対話型多目的最適化において、室配置における組み合わせ最適化問題を解く際に、TSを用いている [21]。しかし、これらの手法は、TSにおける解評価そのものをユーザに求める手法ではない。本論文におけるITSは、IECユーザの解評価負担軽減を目的として、好みのものを1つ選択するのみという評価インタフェースをIECにおいて実現するものである。

本論文ではこれらの知見を踏まえ、ユーザビリティ性及び解候補の進化性能の面から、ITSの有効性を検証する。

2.4.2 TS

TSは、1989年にF. Groverによって考案された局所探索法である [19]。

図2.3にTSにおける解探索を示す。まず、ランダムに生成された複数の解候補の中から最も評価の高い解候補 p_1 を選択する。次に、 p_1 を基にして、複数の近傍解候補を生成する。そして、複数の近傍解候補の中から最も評価の高い解候補 p_2 を選択する。TSでは、このような処理による近傍探索を繰り返し、最適解を探索する。

TSにおいては、現在の解候補が局所最適解であっても他の解候補へ移動する。そのため、局所最適解へ陥りにくくなっている。しかし、現在の解候補が局所最適解のとき、現在の解候補から他の解候補に移動した後に同様の処理を行うことによって、再び局所最適解に戻ってしまうサイクリングが発生するおそれがある。そのため、TSでは、タブーリスト(Tabu list)と呼ばれる禁止則に、探索済みの解候補の情報を書き込む処理が行われる。TSにおいて近傍解候補を生成する際には、タブーリストに含まれる解候補を生成することが禁止されている。しかし、タブーリストの要素を長期間保持すると、生成できる近傍解候補がいづれなくなってしまう。そのため、タブーリストが保持できる禁止則数には制限があり、待ち行列の要領で更新される。本論文では、TSをIECのECアルゴリズムとして用いるITSを提案し、ユーザの解評価負担を軽減することを目的とする。

2.4.3 ITS

ITSは、IECのECアルゴリズムにTSを用いた手法である [18]。

図2.4にITSの流れを示す。まず、ランダムに初期解候補を生成する。次に、全解候補をユーザに提示し、ユーザは最も好みの解候補を1つ選択する。ユーザが選択した解候補のビット列をタブーリストに追加する。そして、ユーザが選択した解候補を基にして、近傍解候補を生成する。近傍解候補が生成されれば、再びユーザに解候補の評価を求める。

TSは、近傍解候補の中で最も評価の高い解候補に、現在の解候補を移動する。そのため、数回前の解評価において最適解となった解候補よりも現在の解候補が改悪されている可能性がある。したがって、ITSでは、ユーザは解探索を終了した後に、タブーリスト中の解候補から最も好みの解候補を選択する。

ITSでは、各世代の評価は、最も好みの解候補を1つ選択するのみとなる。そのため、解候補それぞれに5段階や10段階評価値を与える評価インタフェースより、ユーザの解評価負担を軽減できると考えられる。しかし、ITSはユーザによって選択された解候補の近傍探索を行うため、解候補集団が局所的に収束する。また、ユーザは、提示解候補に好みのものが存在しない場合においても、それらの中から1つの解候補を強制的に選択しなければならず、次世代に好みでない解候補が多数提示される場合がある。このため、ITSにおいては、提示解候補中に好みの解候補が存在しない場合、ユーザの好みの解候補を探索できるまでに時間がかかってしまうという問題が想定される。

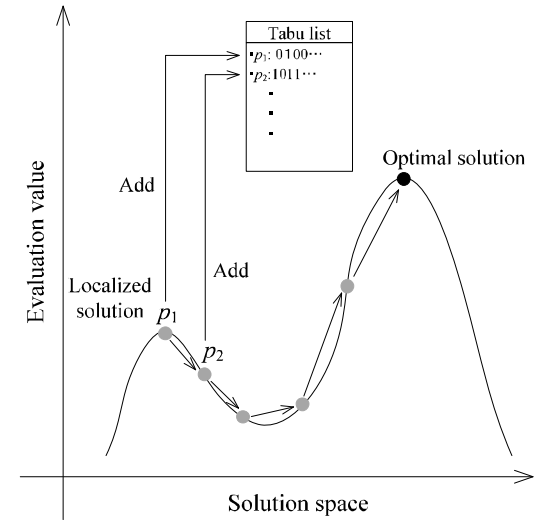


図 2.3: TS の解探索

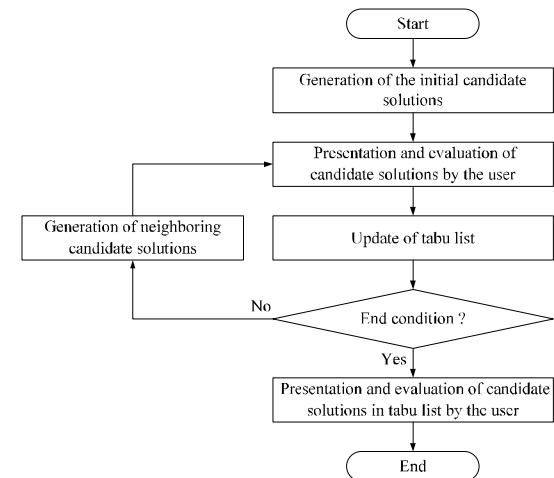


図 2.4: ITS の流れ

2.5 数値シミュレーションによる性能評価

本節では、T2, T4, ITS の解候補の進化性能に関する有効性を検証するために行う数値シミュレーションの概要及び実ユーザの評価特性を表現した評価エージェントの作成方法について述べる。

2.5.1 シミュレーション概要

本章では、実ユーザの代わりにビット列で作成された評価エージェントを用いた数値シミュレーションにおいて、T2, T4, ITS の有効性を検証する。評価エージェントとは、コンピュータ上でユーザに見立てられた擬似ユーザのことをいう。本シミュレーションでは、2.5.2, 2.5.3 項で述べるように、評価エージェントには多峰性感性空間及び評価値の揺らぎを表現させ、実ユーザの評価特性に近い解候補評価を実現する。

本シミュレーションにおいては、T2, T4, ITS の解候補の進化性能が、NIGA と同等であることを示す必要がある。これは、ユーザの解評価負担を軽減する上で非常に重要である。

トーナメント式評価手法においては、解候補の評価値はトーナメントの勝敗により決定される。したがって、トーナメント対戦の組み合わせにより各解候補の勝ち上がりが増減し、本来の解候補の優劣とは異なる評価値が与えられる可能性がある。なぜなら、解候補集団中の最も優秀な解候補とトーナメントの初戦で対戦する解候補は、その解候補本来の評価値に関わらず初戦敗退となるため、結果的に低い評価値が与えられてしまうからである。

ITS においては、TS のアルゴリズムの特性上、局所探索により最適解を探索する。そのため、局所解に陥った際には、そこから抜け出し再び最適解を探索しなければならない。多くの解評価を繰り返さなければならぬおそれがある。

そのため、T2, T4, ITS は、NIGA よりユーザの解評価作業は単純になるが、解候補の進化性能が低下することが懸念される。そこで、数値シミュレーションにおいて、T2, T4, ITS のアルゴリズム本来の進化性能を定量的に比較する。

本シミュレーションにおいては、トーナメント式評価手法のシミュレーションでは、表 2.1 のパラメータを用いる。ITS のシミュレーションでは、表 2.2 のパラメータを用いる。表 2.1, 2.2 の突然変異率 (Mutation rate) 及び表 2.1 の評価減算値 (Evaluation subtraction value) a, b は、予め行ったシミュレーションより、各手法が最良のパフォーマンスを発揮できるように決定した。突然変異処理は、エリート解候補以外の解候補における各遺伝子座を突然変異率で反転する処理とした。また、GA の選択処理におい

ては、全解候補の評価値を 0~100 の範囲で正規化した値を用いている。表 2.1 の θ は、T4 における大差・僅差判定における閾値を示している。これは、T4 において、2 つの解候補の評価値の差が 0.1 より大きい場合は大差判定、0.1 以下の場合は僅差判定となることを示している。表 2.2 より、ITS における近傍解候補生成時の近傍範囲 (Neighboring range) は 1bit とした。これは、ITS における近傍解候補は、現在の最適解を 1bit 反転することで生成されることを示している。なお、確率的な影響を排除するため、シミュレーション試行回数は 100 回とした。

表 2.1: NIGA, T2, T4 のパラメータ

	NIGA	T2	T4
Gene row	Bit string		
Candidate solutions	16		
Selection	Roulette selection + Elite preservation		
Crossover	Uniform crossover		
Mutation rate	3%	4%	4%
Evaluation subtraction value	—	—	$a = 4, b = 2$
Threshold	—	—	$\theta = 0.1$

表 2.2: ITS, NIGA のパラメータ

	ITS	NIGA
Gene row	Bit string	
Candidate solutions	8	
Tabu list	8	—
Neighboring range	1bit	—
Selection	—	Roulette selection + Elite preservation
Crossover	—	Uniform crossover
Mutation rate	—	3%

2.5.2 評価エージェントによる解候補評価

本項では、評価エージェントに対して、実ユーザに近い評価特性を表現させる方法について述べる。本シミュレーションでは、評価エージェントはビット列で作成されているため、評価対象となる解候補のビット列との一致度により解候補を評価できる。この場合、ユーザの解候補評価における評価尺度はビット列の一致度のみとなり、一致度が高い場合は評価が高く、低い場合は評価が低くなることを示す。このような評価を単峰性感性空間における評価とする。単峰性感性空間における評価は、実ユーザの評価に置き換えた場合、嗜好の対象が唯一点に限定される状態を示すことになる。

しかし、実ユーザにおいては、衣服のデザインや配色などに対する嗜好は、唯一点ではないことが多く、様々な嗜好を有していると考えられる。これは、例えばコンセプトが異なる2つの衣服のデザインのうちどちらが好みであるか実ユーザに尋ねた場合、「どちらも好みである」と答えるような場面から想像できる。また、嗜好の強さにおいても、衣服のデザインによって異なる場合がある。これは、例えば実ユーザが衣服や配色のデザインに対して、「好き」、「あまり好きではない」という段階的な評価を行うような場面から想像できる。このように、実ユーザの評価特性は、ある程度の複雑性と非線形性を有する評価尺度であると考えられる。このような評価を多峰性感性空間における評価とする。本シミュレーションでは、実ユーザに近い解候補評価を再現するため、多峰性感性空間を有する評価エージェントを用いる。

次に、評価エージェントの感性空間に対して多峰性を表現させる方法について述べる。評価エージェントの多峰性感性空間における解候補評価は、ランダムに生成された複数のビットパターンに重みを与えることで実現される。まず、1つのビットパターンを生成し、重み $w_0 = 1.0$ を割り当てる。そして、重み w_0 を与えたビットパターンとは異なるビットパターンをランダムにいくつか生成し、0.1 から 0.9 の重みを割り当てる。多峰性感性空間における評価は、重み付けされたビットパターンと評価対象解候補のハミング距離及び予め割り当てられた重みにより行われる。すなわち、多峰性感性空間における解候補の評価値 $fitness$ は、式(2.1)で表現される。

$$fitness = \begin{cases} w_i & \text{if } \min(l_0, l_1, \dots, l_n) = 0 \\ \sum_{i=0}^n w_i \cdot w_i & \text{otherwise} \end{cases} \quad (2.1)$$

式(2.1)において、 w_i は評価エージェントとランダムに生成されたビットパターンに付与した重み、 n はランダムに設定される重みの数、 w_i は重み付けされたビットパターンと評価対象解候補のビットパターンとのハミング距離による重みである。

式(2.1)においては、評価対象解候補のビットパターンが重み付けされたいずれかの

ビットパターンと一致した場合は、一致したビットパターンに設定された重み w_i が評価対象解候補の評価値となる。いずれの重み付けされたビットパターンとも一致しない場合は、評価エージェントとランダムに生成されたビットパターンに付与した重み w_i と重み付けされたビットパターンと評価対象解候補のハミング距離による重み w_l を用いて評価値が算出される。

w_l は、式(2.2)で表現される。

$$w_l = \frac{1}{\sum_{i=0}^n \frac{1}{l_i^k}} \quad (2.2)$$

式(2.2)において、 l_i は重み付けされたビットパターンと評価対象解候補のビットパターンとのハミング距離、 k はハミング距離のスケール乗数である。本論文では予め行ったシミュレーションより、 $n = 4, k = 3$ 、ビットパターンに設定される重みは、1.0, 0.8, 0.1, 0.1, 0.1 とする。したがって、評価エージェントのビットパターンの重みは、 $w_0 = 1.0, w_1 = 0.8, w_2 = 0.1, w_3 = 0.1, w_4 = 0.1$ となる。また、生成されるビットパターンは、シミュレーション試行毎にランダムに生成される。

NIGAにおいては、全解候補に対する評価エージェントの評価値 $fitness$ に10を乗算し、小数第1位を四捨五入した値が解候補それぞれの評価値になる。T2, T4においては、評価エージェントは、対戦する2つの解候補に対する評価値 $fitness$ を求め、評価値が高い方を勝者とする。ITSにおいては、全解候補に対する評価値 $fitness$ を求め、最も評価値の高い解候補が選択される。また、2.6, 2.7節のシミュレーション結果におけるNIGA, T2, T4の絶対評価値は、各世代のエリート解候補に対する評価エージェントの評価値 $fitness$ に10を乗算し、小数第1位を四捨五入した値の試行回数分の平均値である。ITSの絶対評価値は、各世代のタブーリスト中の最良解に対する評価エージェントの評価値 $fitness$ に10を乗算し、小数第1位を四捨五入した値の試行回数分の平均値である。なお、四捨五入した際に評価値が0となった場合は、0.1に切り上げている。

図2.5に可視化した単峰性、多峰性感性空間の一例を示す。本来、感性空間は多次元の空間であり、感性空間内における評価尺度の非線形性が感性の多峰性となる。これを視覚的に確認するため、多次元の感性空間を次のようにして、3次元空間に写像している。まず、評価エージェントのビット数、すなわち感性空間の次元数を20とし、 x 軸を評価エージェントのビット列の前半の10bitsを10進数に変換した値、 y 軸を後半の10bitsを10進数に変換した値とする。そして、 z 軸の値を式(2.1), (2.2)より求めた評価値とする。また、3次元空間における可視化を行っているため、式(2.1), (2.2)

のハミング距離は、ユークリッド距離に置き換えて評価値を算出している。ビットパターンに対する重みは、図2.5(a)においては、ランダムに生成された1つのビットパターンに重み1.0を割り当てている。図2.5(b)においては、評価エージェントのビットパターンに対する重みは、 $w_0 = 1.0, w_1 = 0.8, w_2 = 0.1, w_3 = 0.1, w_4 = 0.1$ を割り当てている。

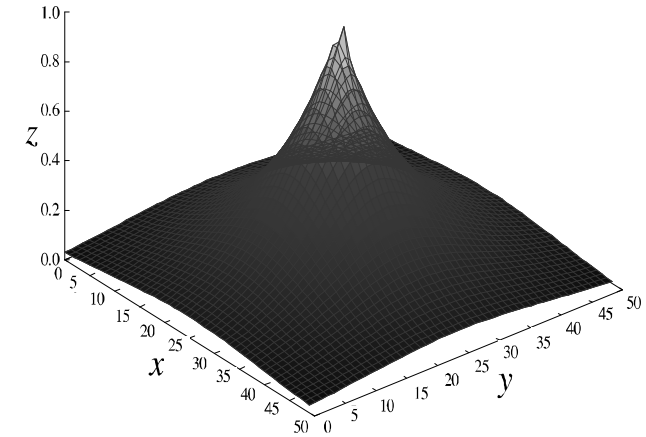
図2.5においては、評価値の低い領域が谷の部分となり、ユーザの嗜好が弱い部分を表している。評価値の高い領域が山の部分となり、ユーザの嗜好が強い部分を表している。このように多峰性感性空間には、ユーザの嗜好が強いとされる領域と弱いとされる領域が複数存在し、実ユーザに近い候補評価を表現できているといえる。

2.5.3 評価エージェントの候補評価における揺らぎの表現

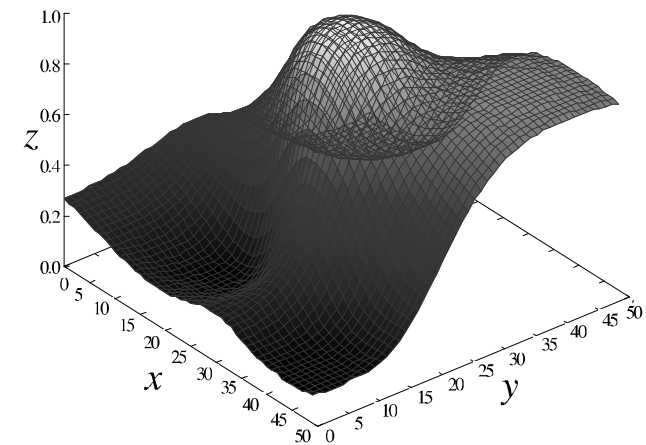
IECにおける評価作業中は、短時間に多くの候補を評価することになるため、ユーザの評価基準が多少なりとも変化してしまう可能性がある。本シミュレーションでは、このような評価特性を表現するため、2.5.2項の多峰性感性空間における候補評価において、評価値の揺らぎを発生させる。

NIGA, T2, T4における評価値の揺らぎは、評価値の揺らぎ量を ± 0.1 とし、評価値の揺らぎ発生確率を様々に変更し、発生させる。評価値の揺らぎは、評価エージェントが候補の評価値を決定する度に、揺らぎ発生確率に基づいて実行される。例えば、トーナメント式評価手法においては、勝ち上がった候補は複数回評価される。この場合、評価値の揺らぎは、勝ち上がった候補の評価値を算出する度に、揺らぎ発生確率に基づいて実行される。なお、揺らぎ量0.1は、実ユーザが10段階評価を行った場合に、6点を与えた候補の評価値が5点または7点になってしまう揺らぎを表現している。

ITSでは、実ユーザにおいて評価の揺らぎが発生した場合でも、提示された候補の中で評価の低い候補が選択されることは少ないと考えられる。そのため、ITSにおける評価値の揺らぎは、評価エージェントが最も評価値の高い候補を選択する際に、評価値の揺らぎ発生確率に基づき、2番目に評価値が高い候補を選択するようにしている。例えば、評価値の揺らぎ発生確率が20%の場合、評価エージェントは、80%の確率で最も評価の高い候補、20%の確率で2番目に評価の高い候補を選択する。



(a) 単峰性感性空間



(b) 多峰性感性空間

図 2.5: 感性空間の可視化の一例

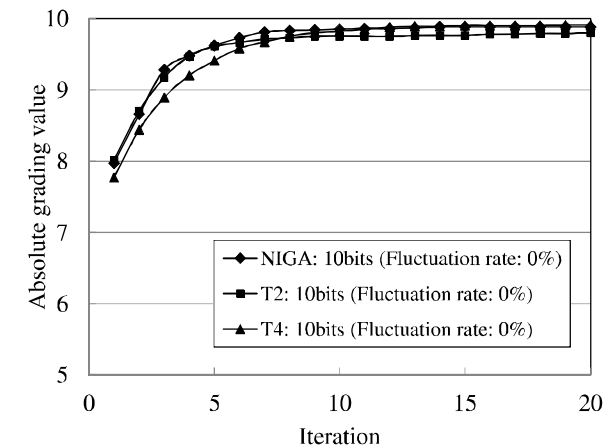
2.6 トーナメント式評価手法におけるシミュレーション結果

本節では、トーナメント式評価手法における数値シミュレーションの結果について考察する。

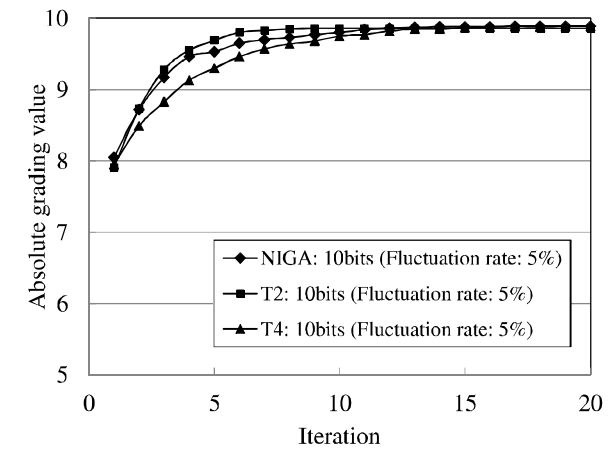
図2.6～2.11に性能比較の結果を示す。評価値の揺らぎを発生させない、すなわち揺らぎ率(Fluctuation rate) 0%の場合、いずれの遺伝子長においてもT2, NIGA, T4の順に進化性能が高いことが確認された。また、T4はT2の改良手法であるが、T2はT4より進化性能が高いことが確認された。これは、GAの選択処理において評価値の正規化を行ったためであると考えられる。評価値の正規化により、T2では第1回戦で敗退した解候補の評価値は0になる。そのため、T2では、第1回戦を勝ち上がった評価の高い解候補同士で交叉が行われ、次世代の解候補が生成される。しかし、T4では、大半の解候補が次世代の解候補生成時に選択される可能性がある。したがって、T2は、T4よりも解の収束性が高くなったと考えられる。

次に、評価値の揺らぎを発生させた場合の結果について考察する。評価値の揺らぎを発生させ、揺らぎ率を増加すると、いずれの手法においても進化性能が低下することが確認された。特に、遺伝子長が長くなると、揺らぎ率の増加に対する進化性能の低下は顕著である。しかし、NIGAは、いずれの遺伝子長においても、揺らぎ率の増加に対してロバストである。NIGAにおける解候補の評価値は、評価値の揺らぎの発生により ± 1 の間で変動する。そのため、NIGAでは、評価値の揺らぎにより、解候補の優劣の順序が入れ替わることは少ないと考えられる。したがって、NIGAでは、評価値の揺らぎによる進化性能への影響が少なかったと考えられる。しかし、T2及びT4においては、評価値の揺らぎにより各対戦の優劣判定が覆ってしまう可能性がある。このため、T2及びT4は、評価値の揺らぎによりNIGAよりも進化性能が低下したと考えられる。

しかし、いずれの手法、遺伝子長、揺らぎ率でも、20世代目において、10段階評価における8点以上が得られている。付録Aより、10段階評価において8点以上が与えられた評価対象のうち、約80%の評価対象は被験者が満足できるものであることが確認された。また、母集団における10段階評価において8点以上が与えられた評価対象に満足できる割合の推定範囲は、76.3～90.3%となった。したがって、いずれの手法、遺伝子長、評価値の揺らぎが発生した場合でも、ユーザが満足のいくものを生成できる可能性があると考えられる。そのため、T2及びT4においては、実ユーザにおける解評価負担や解評価のしやすさに関する有効性を検証する必要がある。

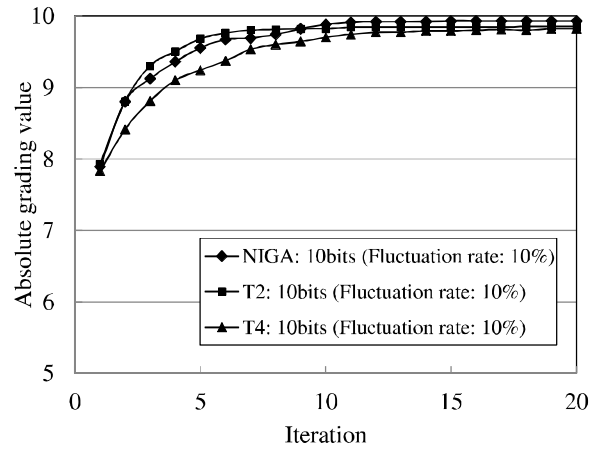


(a) 揺らぎ率: 0%

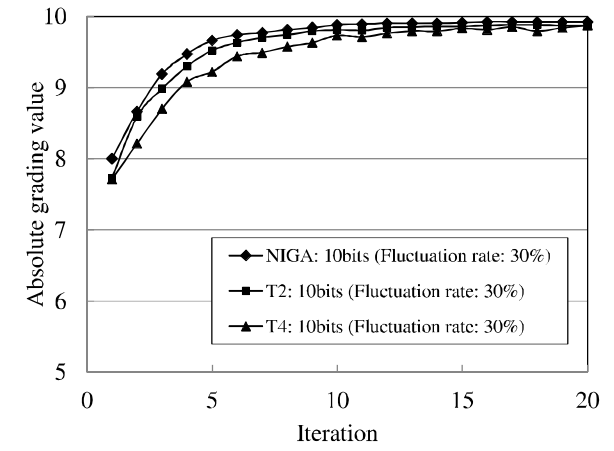


(b) 揺らぎ率: 5%

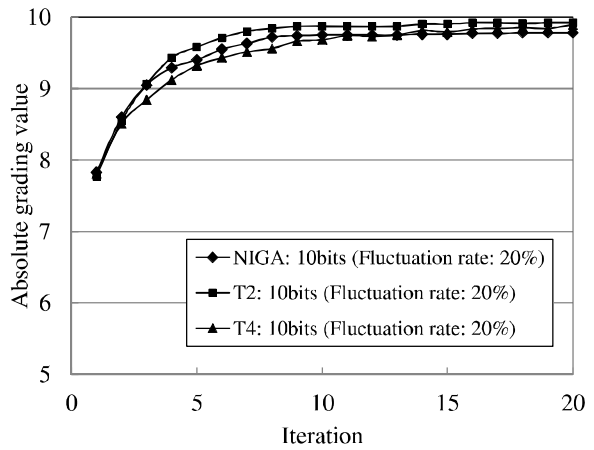
図 2.6: 性能比較の結果 (10bits)



(c) 揺らぎ率 : 10%



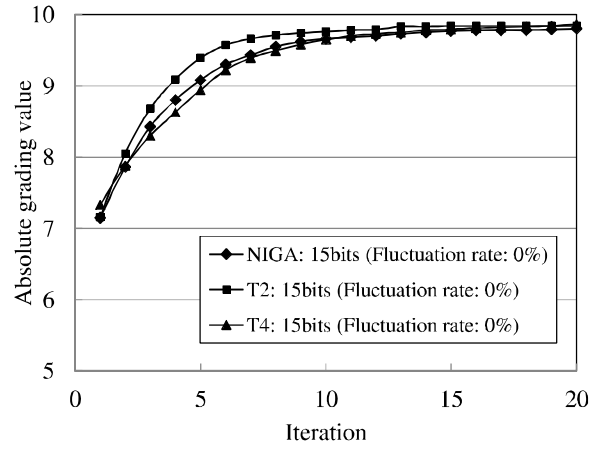
(e) 揺らぎ率 : 30%



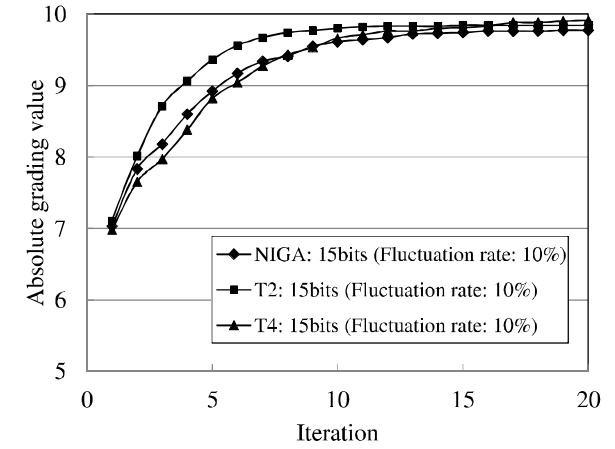
(d) 揺らぎ率 : 20%

図 2.6: 性能比較の結果 (10bits)

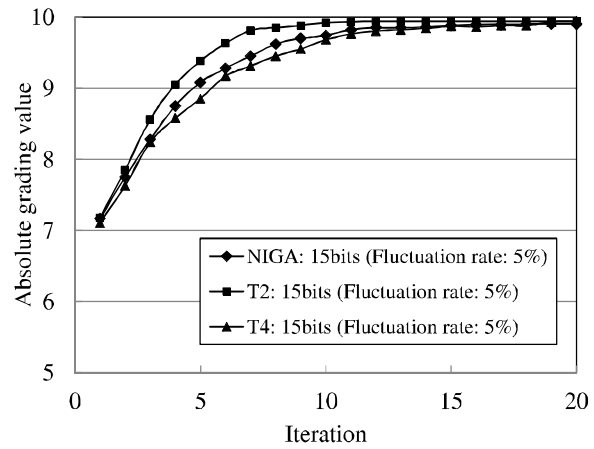
図 2.6: 性能比較の結果 (10bits)



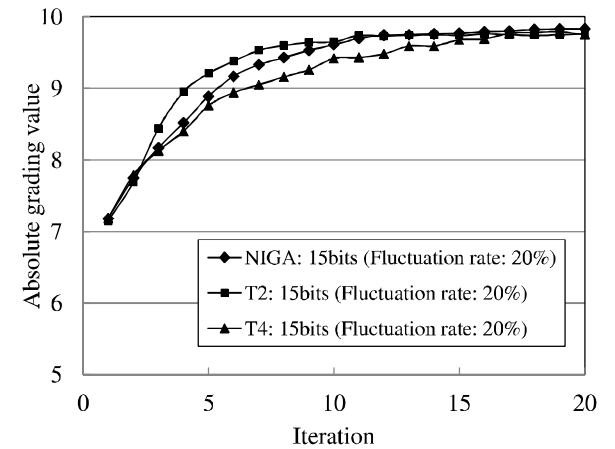
(a) 揺らぎ率 : 0%



(c) 揺らぎ率 : 10%



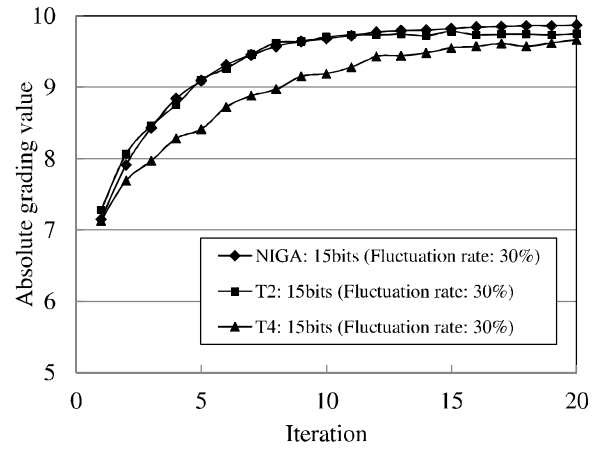
(b) 揺らぎ率 : 5%



(d) 揺らぎ率 : 20%

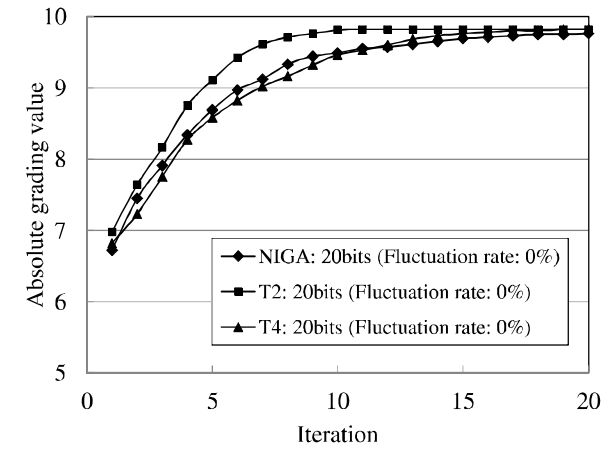
図 2.7: 性能比較の結果 (15bits)

図 2.7: 性能比較の結果 (15bits)

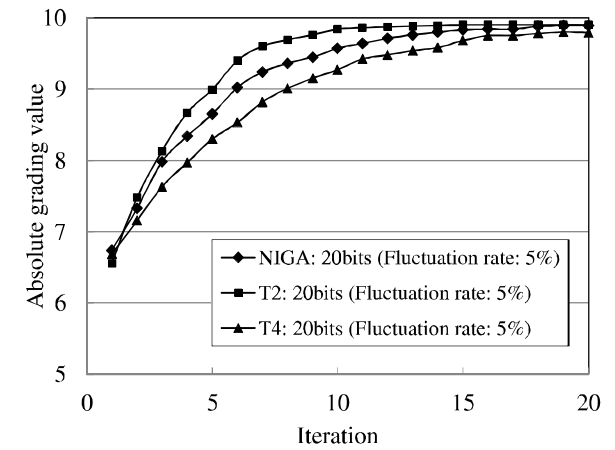


(e) 揺らぎ率 : 30%

図 2.7: 性能比較の結果 (15bits)

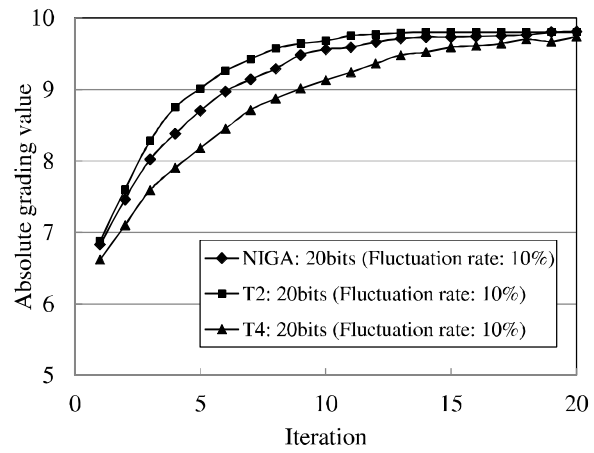


(a) 揺らぎ率 : 0%

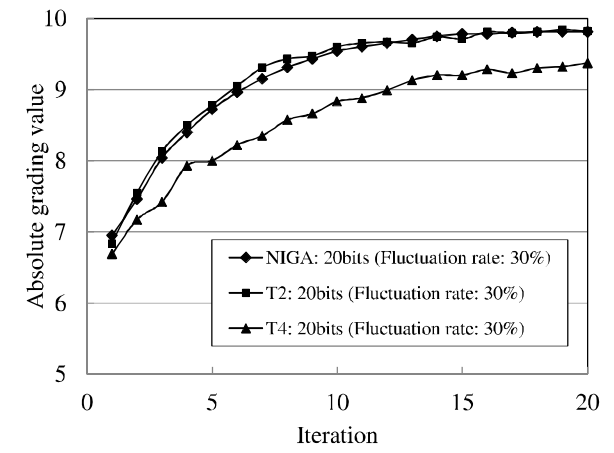


(b) 揺らぎ率 : 5%

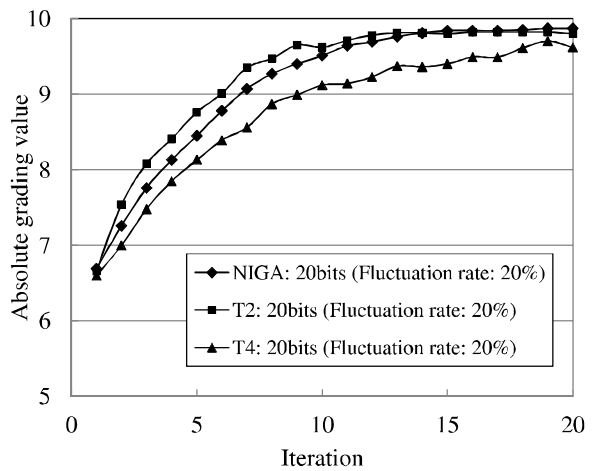
図 2.8: 性能比較の結果 (20bits)



(c) 揺らぎ率 : 10%



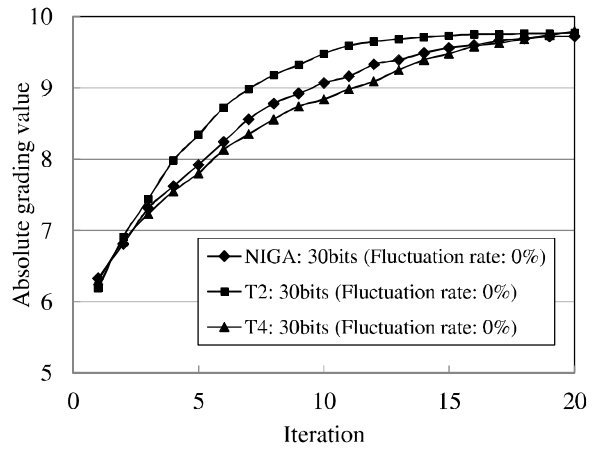
(e) 揺らぎ率 : 30%



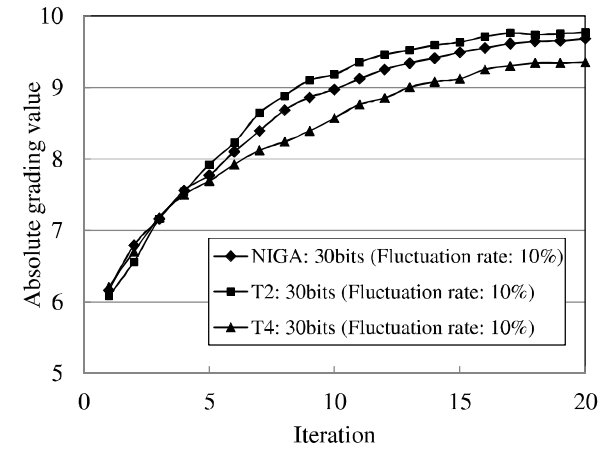
(d) 揺らぎ率 : 20%

図 2.8: 性能比較の結果 (20bits)

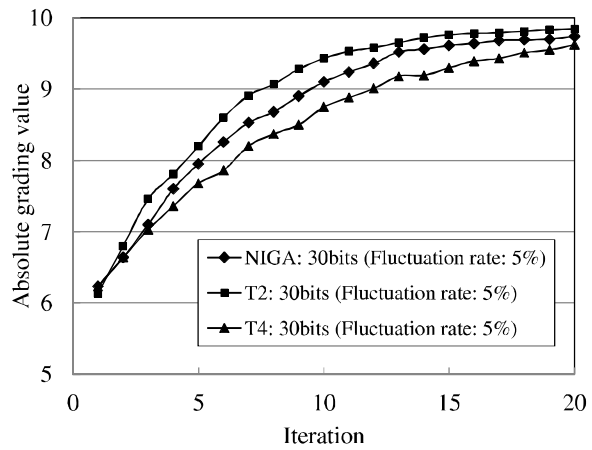
図 2.8: 性能比較の結果 (20bits)



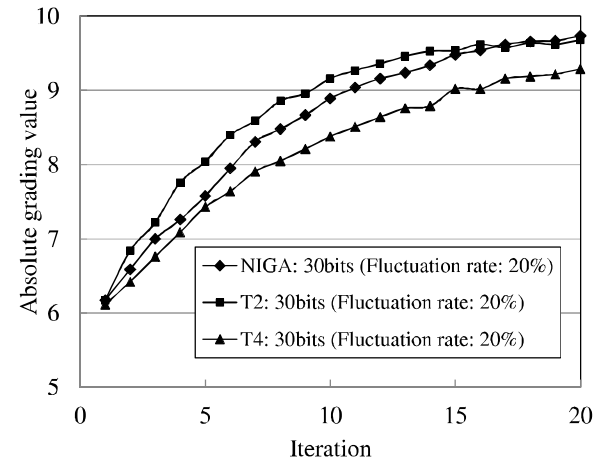
(a) 揺らぎ率 : 0%



(c) 揺らぎ率 : 10%



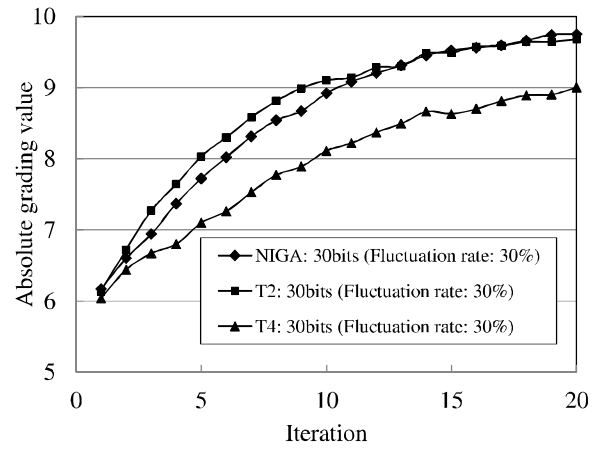
(b) 揺らぎ率 : 5%



(d) 揺らぎ率 : 20%

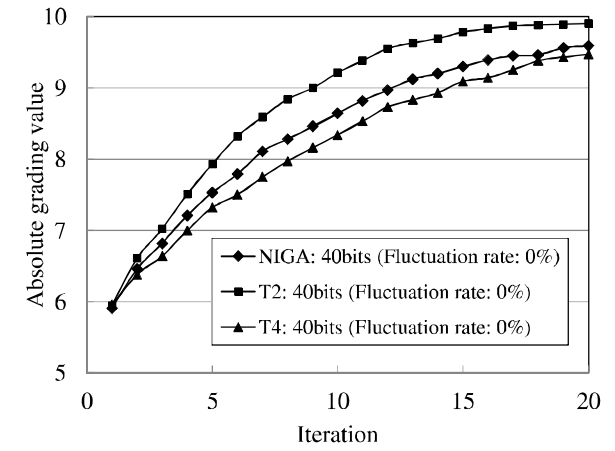
図 2.9: 性能比較の結果 (30bits)

図 2.9: 性能比較の結果 (30bits)

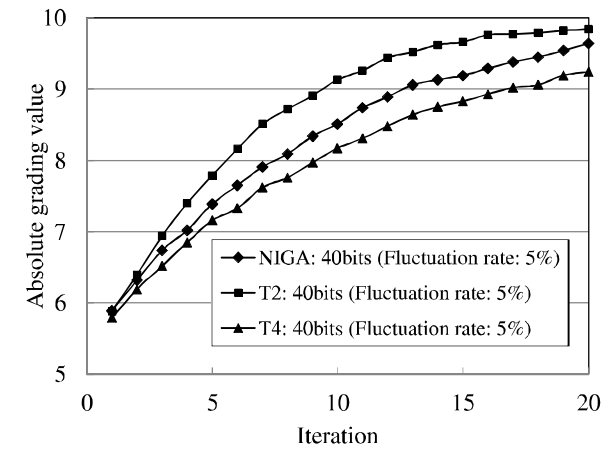


(e) 揺らぎ率 : 30%

図 2.9: 性能比較の結果 (30bits)

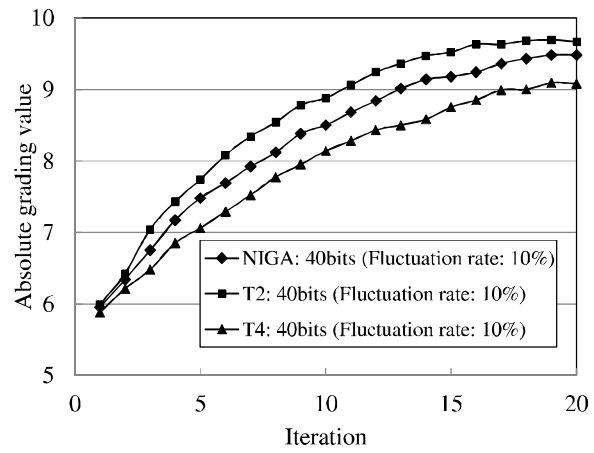


(a) 揺らぎ率 : 0%

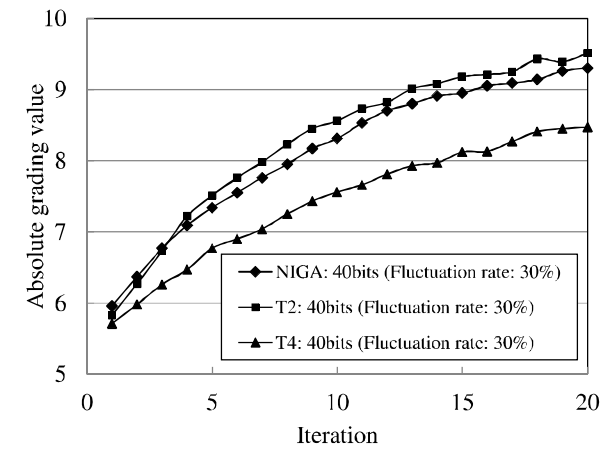


(b) 揺らぎ率 : 5%

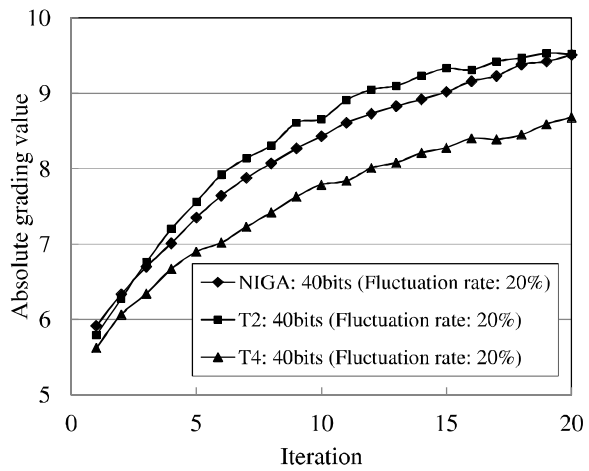
図 2.10: 性能比較の結果 (40bits)



(c) 揺らぎ率 : 10%



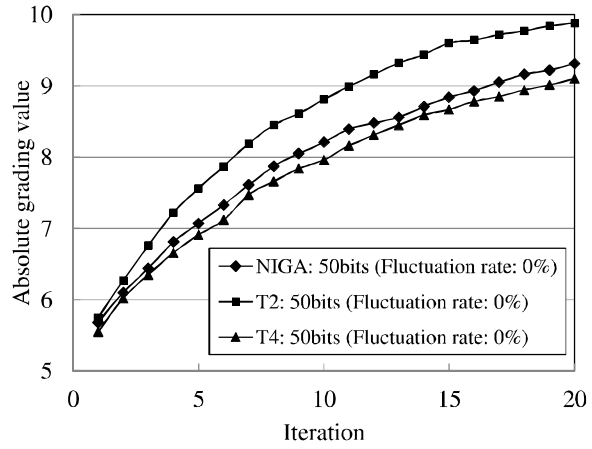
(e) 揺らぎ率 : 30%



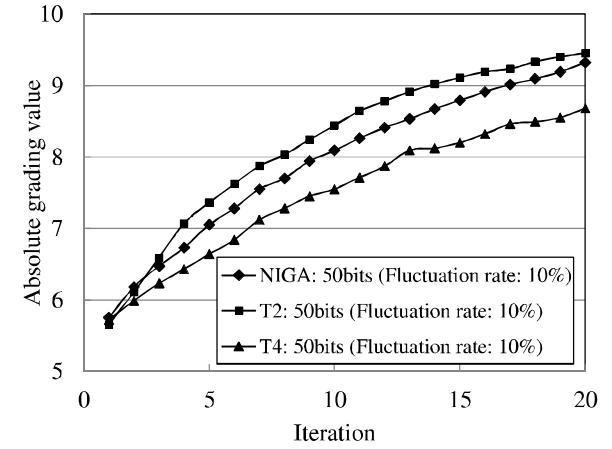
(d) 揺らぎ率 : 20%

図 2.10: 性能比較の結果 (40bits)

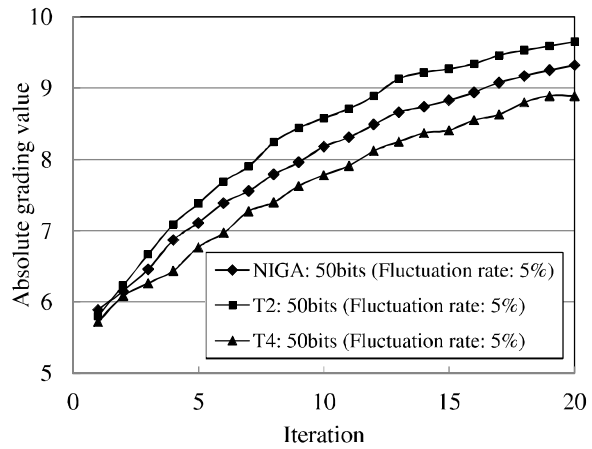
図 2.10: 性能比較の結果 (40bits)



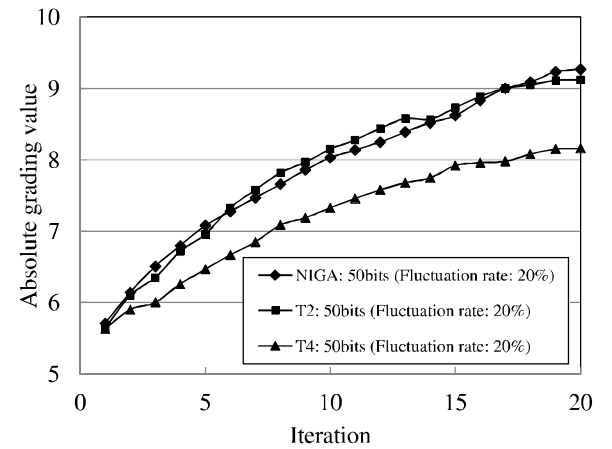
(a) 揺らぎ率: 0%



(c) 揺らぎ率: 10%



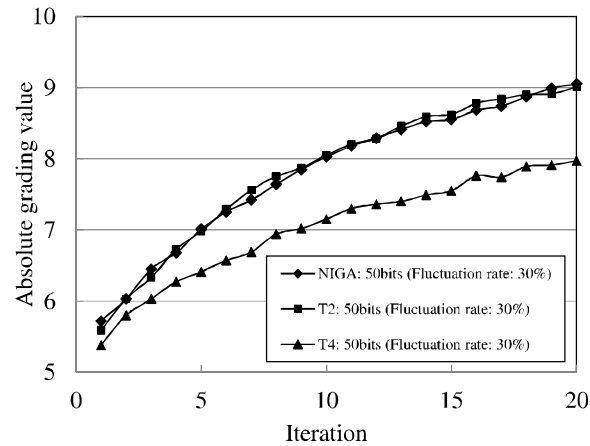
(b) 揺らぎ率: 5%



(d) 揺らぎ率: 20%

図 2.11: 性能比較の結果 (50bits)

図 2.11: 性能比較の結果 (50bits)



(e) 揺らぎ率：30%

図 2.11: 性能比較の結果 (50bits)

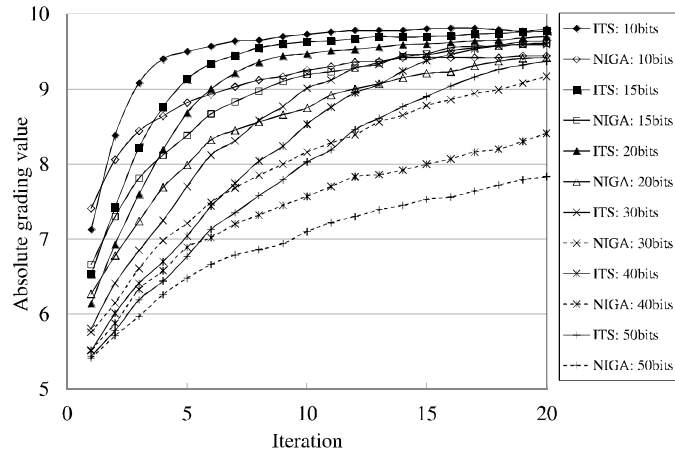
2.7 対話型タブサーチにおけるシミュレーション結果

本節では、ITSにおける数値シミュレーションの結果について考察する。

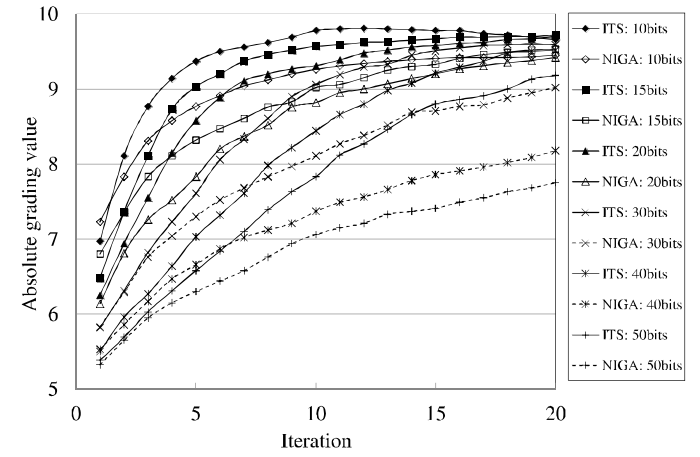
図 2.12 に性能比較の結果を示す。図 2.12(a) より、評価値の揺らぎを発生させない、すなわち揺らぎ率 0% の場合は、遺伝子長が 10~30bits のときは、ITS は NIGA より初期収束性が高いことが確認された。また、遺伝子長が 40, 50bits のときは、ITS 及び NIGA の初期収束性は同等であるが、3 回目以降においては、ITS は NIGA より進化性能が高くなることが確認された。これは、遺伝子長が長くなるに連れて解空間が広くなることが原因であると考えられる。ITS は近傍探索を行うため、1 回目の評価終了後に解候補がある近傍に収束する。遺伝子長が短い場合は解空間が狭いため、近傍探索によって評価の高い解候補へ移動しやすい。しかし、解空間が広くなると、近傍探索によって評価の高い解候補へ移動しにくくなる。

次に、評価値の揺らぎを発生させた場合の結果について考察する。図 2.12(b)~2.12(c) より、評価値の揺らぎを発生させ、揺らぎ率を増加させると、ITS 及び NIGA ともに進化性能が低下することが確認された。ITS においては、初期解候補中から最も好みの解候補を選択した後は、近傍解候補が提示される。ITS における近傍解候補は、近傍範囲 1bit で生成している。そのため、近傍解候補の評価値の差は小さく、評価値の揺らぎにより、2 番目に評価の高い解候補が選択されたとしても、ITS の解候補の進化性能への影響は小さいと考えられる。また、初期解候補がランダムに生成されているため、ITS では揺らぎ率を増加させると、1 回目の絶対評価値が低くなってしまふことが想定される。しかし、ITS においては評価値の揺らぎが発生した場合、提示された解候補の中で 2 番目に評価の高い解候補が選択される。そのため、図 2.12(b)~2.12(c) においても、ITS と NIGA の 1 回目の絶対評価値に大きな差は確認されなかった。

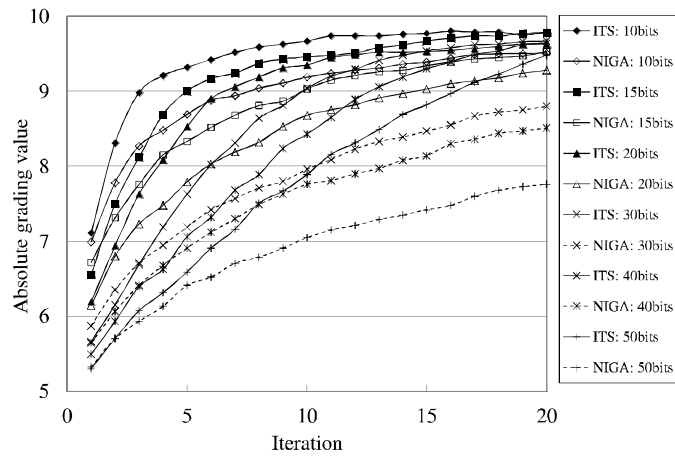
ITS では、いずれの遺伝子長、揺らぎ率でも、20 回目において、10 段階評価における 8 点以上が得られている。付録 A より、母集団における 10 段階評価において 8 点以上が与えられた評価対象に満足できる割合の推定範囲は、76.3~90.3% となった。したがって、ITS においては、いずれの遺伝子長、評価値の揺らぎが発生した場合でも、ユーザが満足のいくものを生成できる可能性があると考えられる。そのため、ITS においては、実ユーザにおける解評価負担や解評価のしやすさに関する有効性を検証する必要がある。



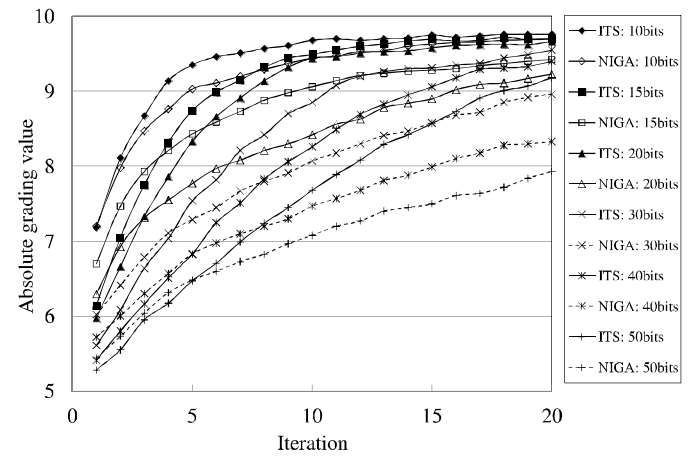
(a) 揺らぎ率 : 0%



(c) 揺らぎ率 : 10%



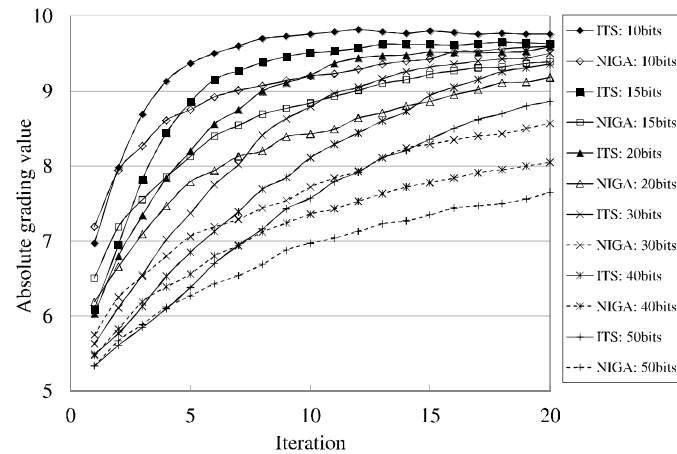
(b) 揺らぎ率 : 5%



(d) 揺らぎ率 : 20%

図 2.12: 性能比較の結果 (ITS)

図 2.12: 性能比較の結果 (ITS)



(e) 揺らぎ率 : 30%

図 2.12: 性能比較の結果 (ITS)

2.8 結言

本章では、まず、IECの基本アルゴリズム及び技術的特徴と課題をまとめ、先行研究を概括した。次に、T2、T4、ITSについて述べ、これらの手法の解候補の進化性能に関する有効性を検証するために数値シミュレーションを行った。数値シミュレーションによる検証結果は以下のとおりとなった。

(1) T2、T4のシミュレーション結果

評価値の揺らぎを発生させない場合は、いずれの遺伝子長においてもT2、NIGA、T4の順に進化性能が高いことが確認された。また、評価値の揺らぎを発生させた場合は、NIGAは揺らぎ率の増加に対してロバストであったが、T2及びT4においては進化性能が低下することが確認された。しかし、付録Aより、母集団における10段階評価において8点以上が与えられた評価対象に満足できる割合の推定範囲は、76.3~90.3%となった。したがって、いずれの手法、遺伝子長、評価値の揺らぎが発生した場合においても、ユーザが満足のいくものを生成できる可能性があると考えられる。

(2) ITSのシミュレーション結果

評価値の揺らぎを発生させない場合は、遺伝子長が10~30bitsのときは、ITSはNIGAより初期収束性が高いことが確認された。また、遺伝子長が40、50bitsのときは、ITS及びNIGAの初期収束性は同等であるが、3回目以降においては、ITSはNIGAより進化性能が高くなることが確認された。評価値の揺らぎを発生させた場合は、ITS及びNIGAともに進化性能が低下することが確認された。また、付録Aより、母集団における10段階評価において8点以上が与えられた評価対象に満足できる割合の推定範囲は、76.3~90.3%となった。したがって、ITSでは、いずれの遺伝子長、評価値の揺らぎが発生した場合においても、ユーザが満足のいくものを生成できる可能性があると考えられる。

本章では、数値シミュレーションにより、T2、T4、ITSの解候補の進化性能に関する有効性が確認された。そのため、T2、T4、ITSについては、実ユーザにおける解評価負担や解評価のしやすさに関する有効性を検証する必要がある。第3章においては、T2、T4、ITSにおける実ユーザを対象とした評価実験について述べる。

第3章 実ユーザによるIECインタフェースのユーザビリティ評価

3.1 緒言

本章では、トーナメント式評価手法及びITSの実ユーザに対する有効性を検証する。第2章の数値シミュレーションより、T2、T4及びITSは解候補の進化性能がNIGAより同等以上であることが示された。そこで、本章では第2章の数値シミュレーションでは検証が不十分であった、実ユーザにおける解評価のしやすさや評価に要した時間及び解候補の進化に要した世代交代数を実ユーザを対象とした評価実験により測定し、トーナメント式評価手法及びITSの有効性を検証する [17, 18].

トーナメント式評価手法の評価実験においては、NIGA、T2、T4を実装したIECシステムを用いる。また、トーナメント式評価手法は、一対比較で解候補を評価するため、音楽や動画などの時系列データが評価対象の場合に有効であると考えられる。そのため、本実験では評価対象に音楽・動画・静止画を用いて、T2、T4の有効性を検証する。

ITSの評価実験においては、NIGA、T2、T4、ITSを実装したランニングシューズデザインシステムを用いる。本実験では、IECのユーザビリティ向上に有効であると考えられる、T2、T4も合わせて実験を行う。

各評価実験においては、生成された解候補に対する満足度と解評価のしやすさ、終了世代数及び評価に要した時間を測定し、実験結果を考察する。

3.2 評価実験

本節では、トーナメント式評価手法及び ITS の評価実験の概要について述べる。

3.2.1 トーナメント式評価手法の実験概要

本実験において性能比較を行う IEC 手法は、NIGA、T2、T4 である。本実験では、IEC の評価対象に音楽・動画・静止画を用いる。トーナメント式評価手法においては、ユーザは2つの解候補を比較し優劣を決定するのみの評価を行うため、NIGA のように複数の解候補を同時に評価する場合よりも、音楽や動画などの時系列データの評価に有効であると考えられる。そのため、本実験では IEC の評価対象として、音楽・動画・静止画を用いた評価実験を行う。

本実験における被験者は、20代の大学生42名である。本実験は、評価対象が音楽・動画・静止画の順に約1週間おきに行った。NIGA、T2、T4の使用順序は、被験者毎にランダムに設定した。被験者は、音楽・動画・静止画を評価対象としたそれぞれの実験において、NIGA、T2、T4によって生成した解候補に対する満足度、各手法における解評価のしやすさ及び実験中に気づいた知見について、アンケートで回答する。また、各 IEC システムの GA パラメータ及び T4 の大差・僅差判定時の評価減算値は、表 2.1 のとおりである。

本実験では、音楽の評価対象にサウンドイコライザシステム、動画の評価対象に万華鏡動画生成システム、静止画の評価対象にランニングシューズデザインシステムを用いた。実験に先立ち、被験者には各システムを次のコンセプトに基づいて使用することを教示した。サウンドイコライザシステムにおいては、「音楽を聴くときの好みに合わせてイコライザを調整する」、万華鏡動画生成システムにおいては、「パソコンのスクリーンセーバーにしたい万華鏡動画を作成する」、ランニングシューズデザインシステムにおいては、「ジョギングをするときに履きたいランニングシューズを作成する」とした。システムの使用終了条件は、被験者が提示された解候補のうち、約80%が満足 of いく解候補になったと判断したときとした。システムの使用を終了する際には、被験者は最終世代において最も好みの解候補を1つ選択する。NIGA では、最終世代に提示された解候補の中から最も好みの解候補を選択する。T2 及び T4 では、最終世代において優勝した解候補となる。実験終了後、被験者が各システムの使用を終了する際に選択した解候補を再度提示し、5段階評価における満足度をアンケートで回答してもらった。

また、本実験では、各システムにおいて同じ評価時間を設ける実験方法が考えられ

る。しかし、本実験では、各システムにおいて、ユーザが満足のいく解候補を生成できるかどうかを検証することに焦点を当てた。また、評価時間に制限を設ける場合、被験者によって評価に要する時間は異なるため、評価時間の設定が困難であり、被験者が満足のいく解候補を得られるかどうか不明である。そのため、本実験においては、被験者が満足のいく解候補を生成できるまで各システムを使用してもらった。

3.2.2 対話型タブーサーチの実験概要

本実験において性能比較を行う IEC 手法は、NIGA、T2、T4、ITS である。本実験では、IEC のユーザービリティの向上に有効であると考えられる、T2、T4 も合わせて実験を行う。ITS では、多くの解候補をユーザに提示するため、音楽や動画などの時系列データが評価対象の場合は、ユーザの解評価負担が発生する可能性がある。そのため、本実験では IEC の評価対象を、ITS における好みの解候補を1つ選択するという解候補評価がユーザの評価負担軽減に有効であると考えられる静止画とした。

本実験における被験者は、20代の大学生21名である。NIGA、T2、T4、ITS の使用順序は、被験者毎にランダムに設定した。被験者は、NIGA、T2、T4、ITS によって生成した解候補に対する満足度、各手法における解評価のしやすさ及び実験中に気づいた知見についてアンケートで回答する。また、NIGA 及び ITS の GA 及び TS パラメータは、表 2.2 のとおりである。T2、T4 のパラメータは、表 2.1 のとおりである。なお、解候補数の違いによる進化性能への影響を考慮するため、いずれの手法においても解候補数は8個に統一した。そのため、T2における解候補の評価値は、優勝解候補に10点、準優勝解候補に8点、準決勝敗退解候補に6点及び第1回戦敗退解候補に4点を与えるものとした。

実験に先立ち、被験者には、「ジョギングをするときに履きたいランニングシューズを作成する」というコンセプトに基づいてシステムを使用することを教示した。システムの使用終了条件は、被験者が提示された解候補のうち、約80%が満足のいく解候補になったと判断したときとした。システムの使用を終了する際には、被験者は、ITS では、タブーリストの中から最も好みの解候補を1つ選択する。NIGA では、最終世代に提示された解候補の中から最も好みの解候補を選択する。T2 及び T4 では、最終世代において優勝した解候補となる。実験終了後、被験者が各システムを終了する際に選択した解候補を再度提示し、5段階評価における満足度をアンケートで回答してもらった。

また、本実験でも、各システムにおいて同じ評価時間を設ける実験方法が考えられる。しかし、3.2.1 項と同様の理由で、評価時間の設定が困難である。さらに、ITS に

おける解探索は局所探索であるが、NIGA, T2, T4においては広域探索となる。そのため、ITSとNIGA, T2, T4では解探索の性質が異なり、評価時間を統一することは平等な条件ではないと考えられる。

3.3 実験システム

本節では、評価実験において音楽・動画・静止画の評価対象として用いる、サウンドイコライザシステム、万華鏡動画生成システム、ランニングシューズデザインシステムの構成及び遺伝子コーディングについて述べる。

3.3.1 サウンドイコライザシステム

本システムにおけるサウンドイコライザは、IIR ピーキングフィルタを用いて実現している [51]。IIR ピーキングフィルタには、フィルタ処理を適用する中心周波数 f_c 、クオリティファクタ Q 値、ゲイン値 g の3つのパラメータがある。本システムでは、中心周波数 f_c を8個 (60, 250, 500, 1,000, 2,000, 4,000, 8,000, 16,000 [Hz]) 設定し、それぞれの中心周波数のゲイン値 g を遺伝子コーディングに用いた。なお、クオリティファクタ Q 値は、 $Q = 1/\sqrt{2}$ とした。

表 3.1 にサウンドイコライザシステムのゲイン値とビットパターンの対応を示す。各中心周波数のゲイン値は、2bits で構成されている。そのため、本システムの遺伝子長は 16bits (2bits \times 8 帯域) となる。評価対象となる音楽の再生時間は 10[s] とした。

3.3.2 万華鏡動画生成システム

本システムにおける万華鏡動画は、色相値変換フィルタ (Hue-value transform filter) を用いて実現している。図 3.1 に色相値変換フィルタを用いた万華鏡動画生成方法を示す。色相値変換フィルタは、元画像の色相値のみを変化させるフィルタであり、色相値空間における4つの基準点 $a_i (0 \leq i \leq 3)$ を結ぶスプライン曲線により構成される。各基準点の初期位置は、 $a_0 = 90, a_1 = 180, a_2 = 270, a_3 = 360$ である。これらの基準点を移動することにより、様々なスプライン曲線が生成される。生成されたスプライン曲

表 3.1: サウンドイコライザシステムのゲイン値 g とビットパターンの対応

Gain g	Bit pattern
0.5	11
0	10
-0.5	01
-1.0	00

線により、元画像の色相値を変換し、万華鏡動画生成に用いる画像を作成する。なお、スプライン曲線を生成する際には、変換前の色相値が0の位置に $\alpha = a_3$ を代入し、色相値が0~90の間に補完している。

本システムでは、これらの基準点の移動量を遺伝子コーディングにより決定する。基準点の移動量は、 $22.5 \times j (0 \leq j \leq 15)$ で決定される。本システムの遺伝子コーディングにおいては、 j が4bits グレイコーディングで表現される。そのため、本システムの遺伝子長は16bits (4bits \times 4基準点)となる。なお、色相値が360を超えた場合は、色相値を360で除算した際の剰余を用いる。

本システムにおける万華鏡動画の再生時間は10[s]とした。万華鏡動画は、図3.1中、右上の画像中の赤と青の2つの三角形が色相値変換された画像上を移動することにより実現される。本処理により、万華鏡において模様が内から外へ広がる様子を実現している。さらに、本動画の万華鏡の模様部分は、 $3^\circ/\text{s}$ で左回転する。本処理により、万華鏡を回転させたときの様子を表現している。

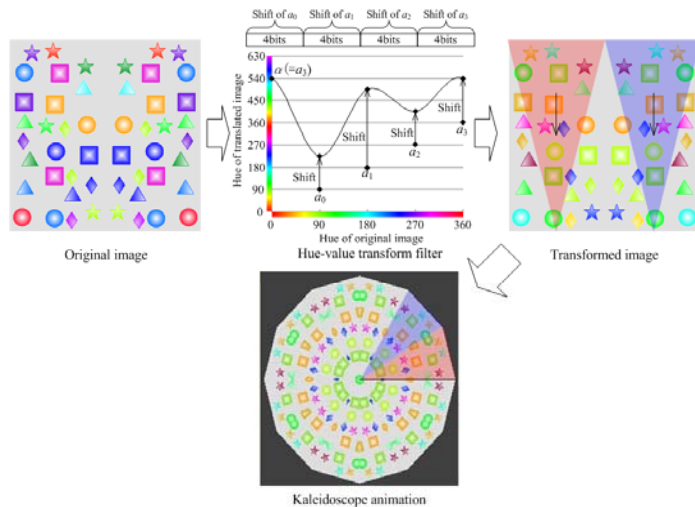


図 3.1: 色相値変換フィルタを用いた万華鏡動画生成

3.3.3 ランニングシューズデザインシステム

本システムにおけるランニングシューズデザインは、5種類のランニングシューズデザインパーツの組み合わせにより構成される。図3.2にランニングシューズデザインの遺伝子コーディングを示す。本システムにおけるランニングシューズのデザインは、靴底 (Sole)・ベース (Base)・靴先 (Toe)・ライン (Line)・靴紐 (Shoelace) の5つのデザインパーツの組み合わせにより、表現されている。各デザインパーツには、8種類のデザインがある。そのため、遺伝子コーディングにおいては、各デザインパーツに3bits が割り当てられる。したがって、本システムの遺伝子長は15bits (3bits \times 5デザインパーツ)となる。

図3.3に本システムで用いたランニングシューズのデザインパーツを示す。本システムにおいては、これらのパーツを組み合わせることで様々なランニングシューズデザインを生成する。各デザインパーツのビットパターン割り当ては、遺伝子型であるビットパターン間のハミング距離と表現型であるデザインパーツの類似性を考慮し、決定した。

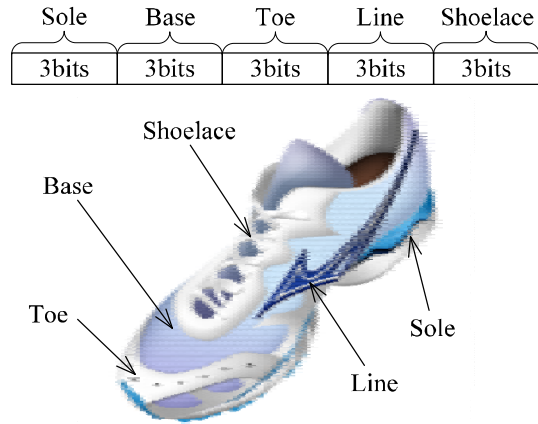


図 3.2: ランニングシューズデザインの遺伝子コーディング

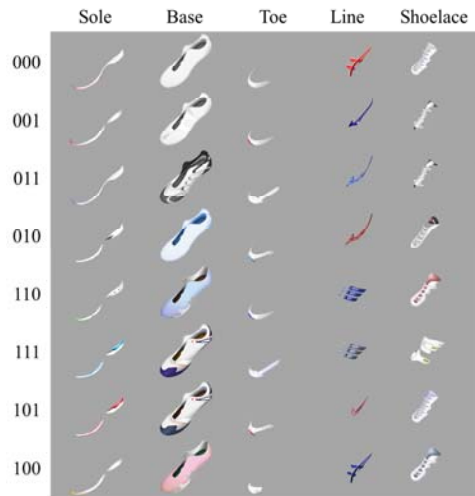


図 3.3: ランニングシューズデザインパーツとビットパターンの対応

3.4 実験システムの評価インタフェース

本節では、サウンドイコライザシステム、万華鏡動画生成システム、ランニングシューズデザインシステムの評価インタフェースについて述べる。

3.4.1 トーナメント式評価手法の評価実験における評価インタフェース

本項では、トーナメント式評価手法の評価実験で用いる各システムの評価インタフェースについて述べる。

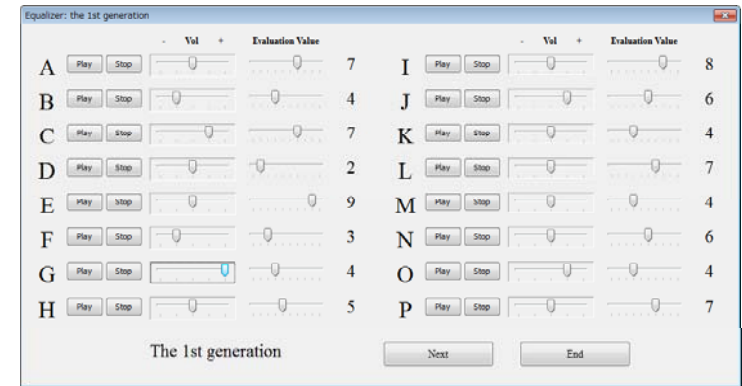
図 3.4～3.6 に NIGA, T2, T4 を用いたサウンドイコライザシステム、万華鏡動画生成システム、ランニングシューズデザインシステムの評価インタフェースを示す。図 3.4(a)～3.6(a) の NIGA においては、ユーザーは評価値入力用スライダで 10 段階の評価値を入力する。1 世代分の全解候補の評価が終了すれば、ユーザーは、“Next” ボタンをクリックし新しい解候補を生成するか、“End” ボタンをクリックし解候補生成を終了するかを決定する。また、図 3.6(a) においては、インタフェース画面領域の制限のため、各ランニングシューズデザインは縮小して表示している。そのため、各ランニングシューズデザインをクリックすると、別ウィンドウに拡大画像が表示されるようにしている。

図 3.4(b)～3.6(b), 図 3.4(c)～3.6(c) の T2, T4 においては、ユーザーはインタフェース画面上部のトーナメント表で対戦経過を確認できる。ユーザーは、対戦解候補下のボタンで各対戦の勝者を決定する。T2 では、ユーザーは 2 つの解候補のうち、好みの方の“Good” ボタンをクリックする。T4 では、ユーザーは 2 つの解候補の優劣を“Very good” ボタン (大差判定) もしくは“Good” ボタン (僅差判定) をクリックすることで決定する。T2, T4 においては、毎世代決勝戦まで終了した時点で、ユーザーは、“Next” ボタンをクリックし新しい解候補を生成するか、“End” ボタンをクリックし解候補生成を終了するかを決定する。

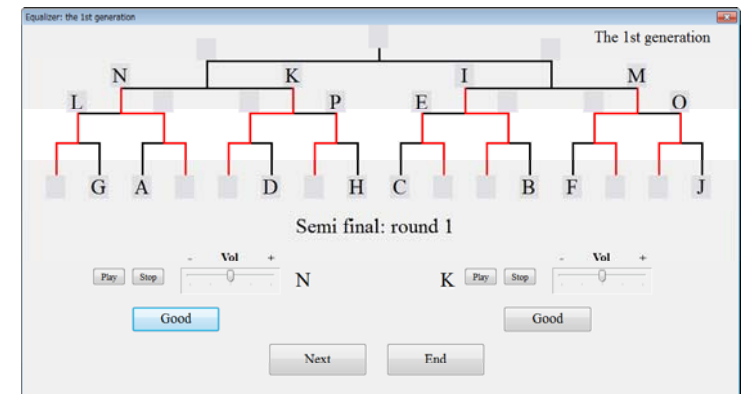
図 3.4 においては、評価対象が音楽であるため、各解候補の“Play”及び“Stop”ボタン、ボリューム調節用スライダが配置されている。図 3.5 においては、評価対象が動画であるため、各解候補の“Play”及び“Stop”ボタンが配置されている。

また、図 3.5 においては、万華鏡動画の再生前に万華鏡の静止画が初期画面として表示される。一般に、パソコンなどで動画を見る場合、動画再生前にどのような動画であるか明確に想像できることは少ないと考えられる。そのため、本実験においても動画を評価対象とした場合、ユーザーが動画の良し悪しを再生前に評価できないことが望ましい。しかし、本システムでは、万華鏡動画の初期画面の印象と万華鏡動画再生時

の印象は異なるが、被験者はある程度解評価に慣れてくると、動画を再生することなく初期画面のみで解候補を評価できてしまう。そのため、動画の評価であっても、静止画と同様な評価が行われてしまう可能性がある。本実験では、静止画として評価されることを回避するため、動画の実験時においては、被験者に「動画再生中の印象によって解候補を評価する」ということを教示した。

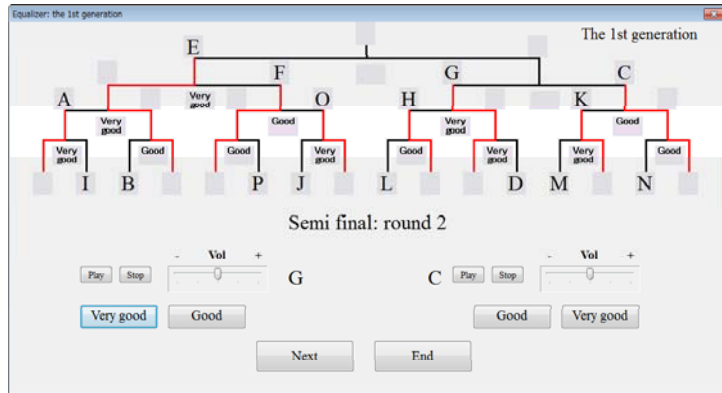


(a) NIGA



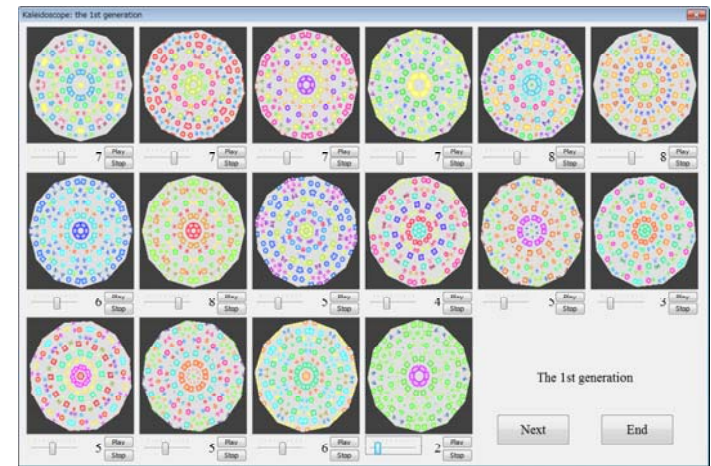
(b) T2

図 3.4: サウンドイコライザシステムの評価インタフェース

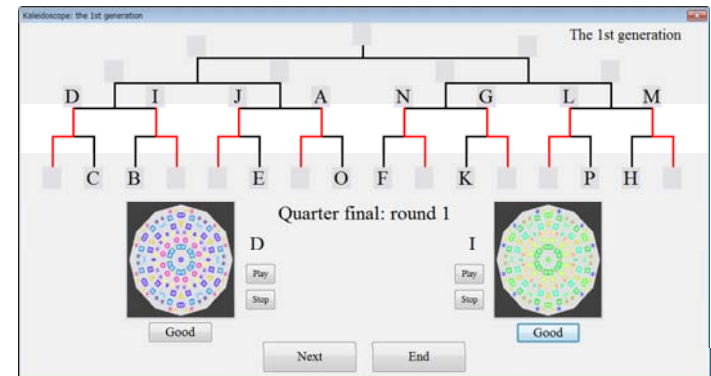


(c) T4

図 3.4: サウンドイコライザシステムの評価インタフェース

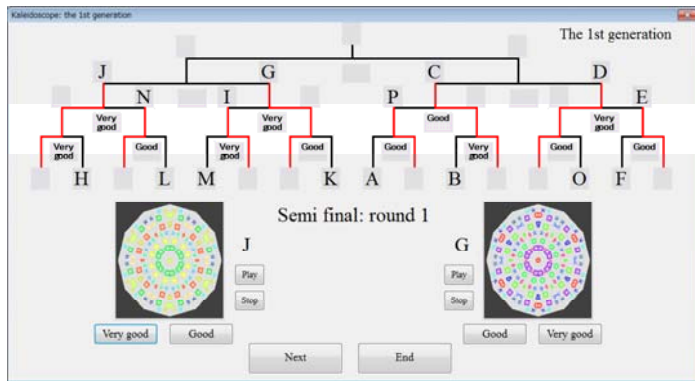


(a) NIGA



(b) T2

図 3.5: 万華鏡動画生成システムの評価インタフェース

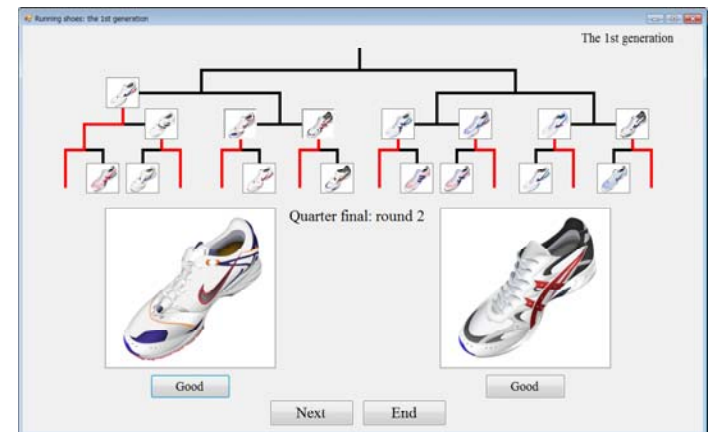


(c) T4

図 3.5: 万華鏡動画生成システムの評価インターフェース



(a) NIGA



(b) T2

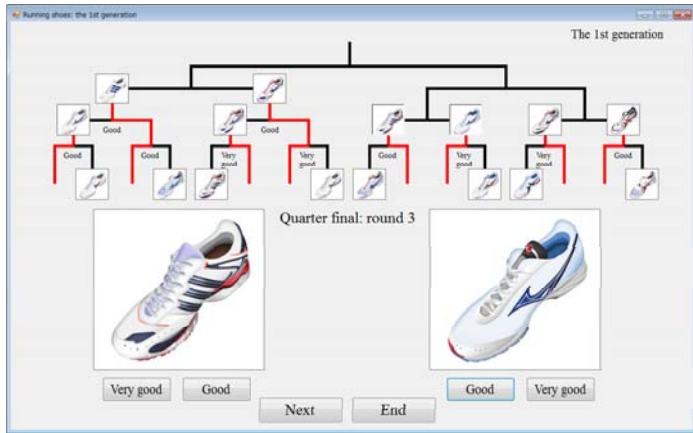
図 3.6: ランニングシューズデザインシステムの評価インターフェース (T2, T4 評価実験)

3.4.2 対話型タブーサーチの評価実験における評価インタフェース

本項では、ITSの評価実験で用いる各システムの評価インタフェースについて述べる。

図 3.7 に NIGA, T2, T4, ITS を用いたランニングシューズデザインシステムの評価インタフェースを示す。図 3.7(a)~3.7(c) の NIGA, T2, T4 の評価インタフェースの操作方法は、3.4.1 項で述べたものと同様である。なお、図 3.7 における各評価インタフェースでは、ランニングシューズデザインの画像の大きさを統一している。そのため、図 3.6(a) の NIGA において実装していた画像拡大の機能は省いている。

図 3.7(d) の ITS においては、ユーザは提示された解候補の中から最も好みの解候補を選択し、その解候補下の“Good”ボタンをクリックする。そして、解候補選択の確認メッセージが表示され、“OK”をクリックすると新しい解候補が生成される。“Cancel”をクリックすると、再度解候補を選択できる。ユーザが解評価を終了する際には、“End”ボタンをクリックする。その後、タブーリスト中の解候補が提示され、その中から最も好みの解候補を1つ選択し、解候補生成を終了する。

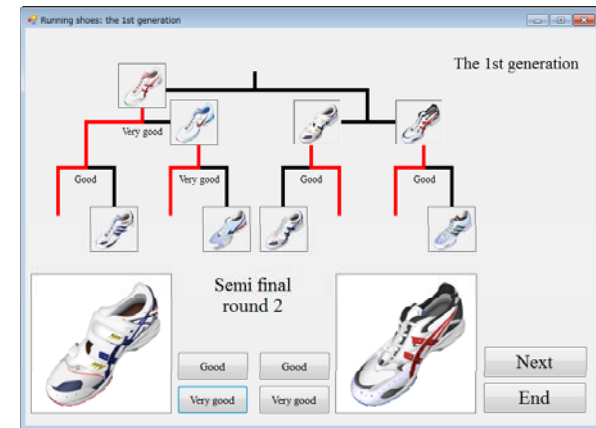


(c) T4

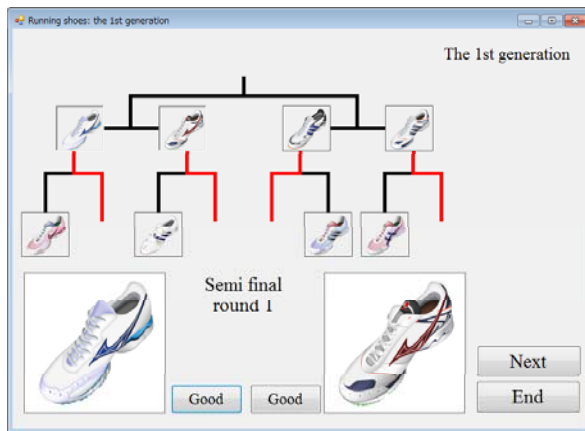
図 3.6: ランニングシューズデザインシステムの評価インタフェース (T2, T4 評価実験)



(a) NIGA



(c) T4



(b) T2



(d) ITS

図 3.7: ランニングシューズデザインシステムの評価インタフェース (ITS 評価実験)

図 3.7: ランニングシューズデザインシステムの評価インタフェース (ITS 評価実験)

3.5 トーナメント式評価手法における実験結果

本節では、トーナメント式評価手法の評価実験の結果について考察する。

3.5.1 生成された解候補に対する満足度

本項では、トーナメント式評価手法の評価実験における、各手法によって生成された解候補に対する満足度 (Satisfaction level) 及び終了世代数 (Final generation) の結果について考察する。

図 3.8～3.10 に各評価対象において生成された解候補に対する満足度の平均値と標準偏差及び人数分布の割合を示す。図 3.8(a)～3.10(a) より、いずれの評価対象、手法でも、生成された解候補に対する満足度は、5段階評価において約4点になっていることが確認された。また、図 3.8(b)～3.10(b) より、いずれの評価対象、手法においても、全被験者の約70%以上が4点以上を付けていることが確認された。そこで、各手法の満足度の統計的有意差を確認するため、フリードマン検定を行った。表 3.2 にフリードマン検定の結果を示す。表 3.2 には、3.5.2 項において考察する解評価のしやすさ及び評価時間に関する検定結果も合わせて示している。満足度においては、いずれの評価対象でも、有意水準5%で統計的有意差は確認されなかった。このため、いずれの手法を用いても被験者が満足いく解候補を生成できることが確認された。

しかし、満足いく解候補が得られるのに要した世代交代数は、手法によって異なる。図 3.11～3.13 に各評価対象における終了世代数の平均値と標準偏差及び人数分布の割合を示す。図 3.11(a)～3.13(a) より、いずれの評価対象においても、NIGA が最も多く世代交代を要していることが確認された。また、図 3.11(b)～3.13(b) より、T2 及び T4 においては、全被験者の約80%が4世代目までに評価を終了していることが確認された。しかし、NIGA では、4世代目までに評価を終了している被験者は、全被験者の約50%に留まっている。表 3.2 のフリードマン検定の結果より、終了世代数においては、評価対象が動画・静止画の場合には、有意水準1%で統計的有意差が確認された。さらに、フリードマン検定において統計的有意差が確認された場合は、シェッフェの多重比較により、どの手法間に統計的有意差があるかを確認した。表 3.3 にシェッフェの多重比較の結果を示す。表 3.3 中の斜体表記は、統計的検定により良好な結果を示した手法である。表 3.3 には、3.5.2 項において考察する解評価のしやすさ及び評価時間に関する検定結果も合わせて示している。表 3.3 より、動画・静止画の場合とも有意水準1%で、T2 及び T4 は NIGA より終了世代数が少ないことが確認された。

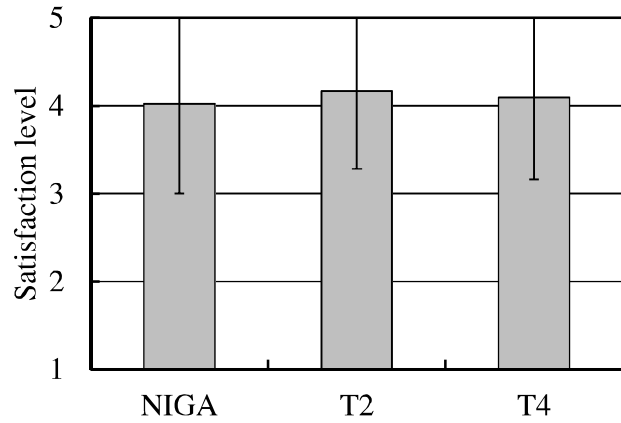
これは、NIGA においては、被験者に全解候補が同時に提示されることに起因して

いると考えられる。被験者が明確な目標デザインを持っていない場合、提示された解候補に対して大差のない評価値を与える場合が多い。そのため、解候補それぞれの評価値の差が生じにくく、解が進化しにくくなると考えられる。T2、T4 においては、アルゴリズムの特性上、初戦で敗退した解候補には低い評価値、勝ち上がった解候補には高い評価値が与えられるため、必然的に評価値の差が生じる。そのため、T2、T4 は NIGA と比較して解収束が早く、世代交代数が少なくなったと考えられる。

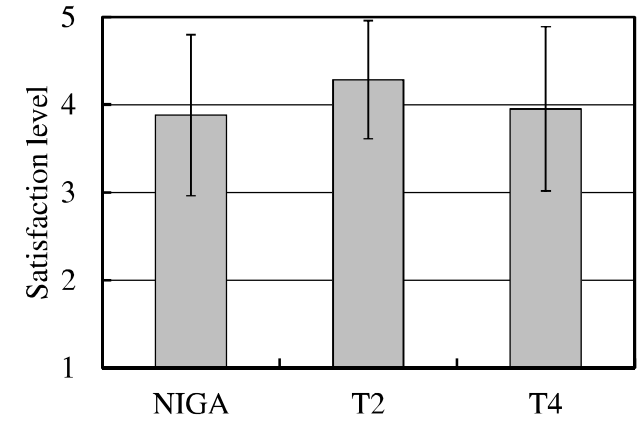
また、次のような被験者のコメントが得られた。

- NIGA では世代交代を重ねる間に目標とするデザインが変化してしまう

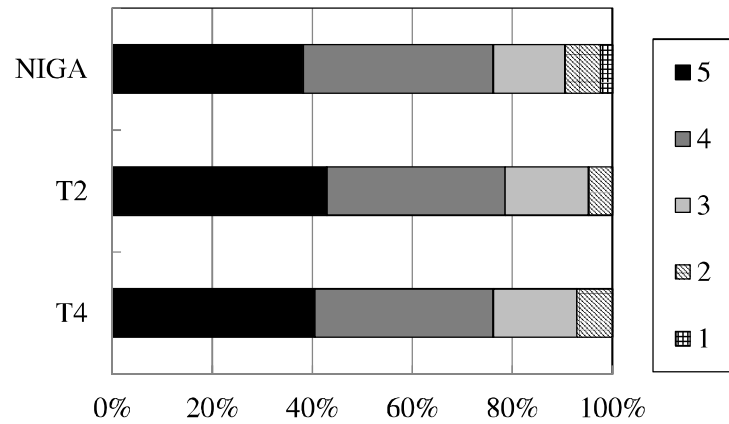
評価対象における目標デザインが明確でない場合、被験者の嗜好は世代交代を重ねる度に变化する可能性がある。これは、被験者は多くの解候補を同時に提示され、解評価に迷いが生じるためである。その結果、NIGA では世代交代が多くなったと考えられる。



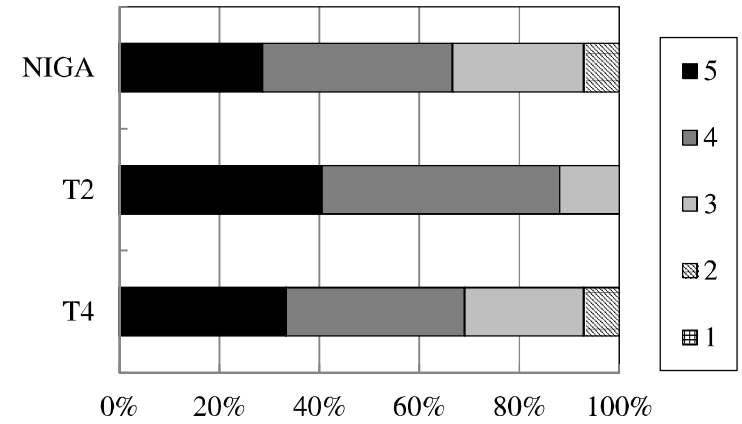
(a) 平均値と標準偏差



(a) 平均値と標準偏差



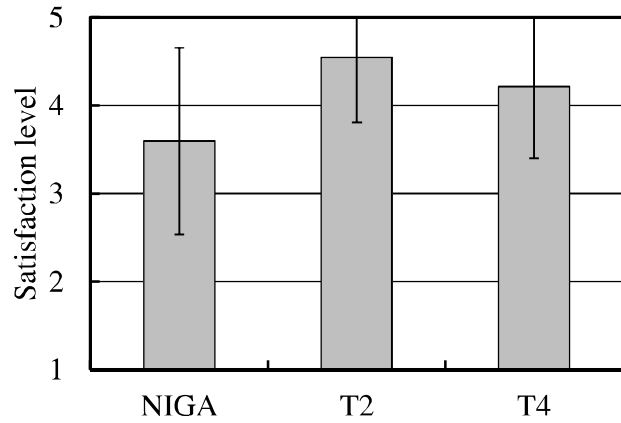
(b) 人数分布の割合



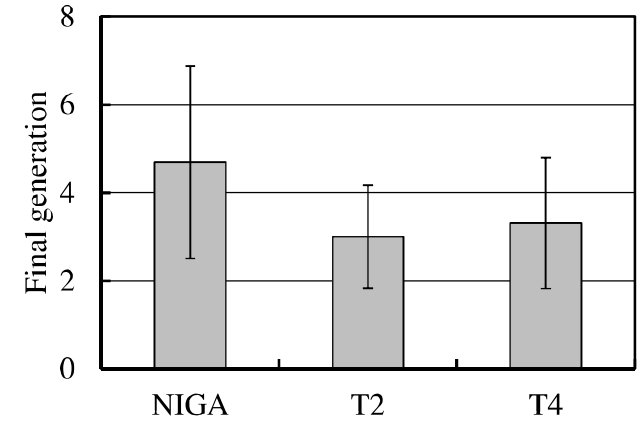
(b) 人数分布の割合

図 3.8: 生成された解候補に対する満足度の結果 (音楽)

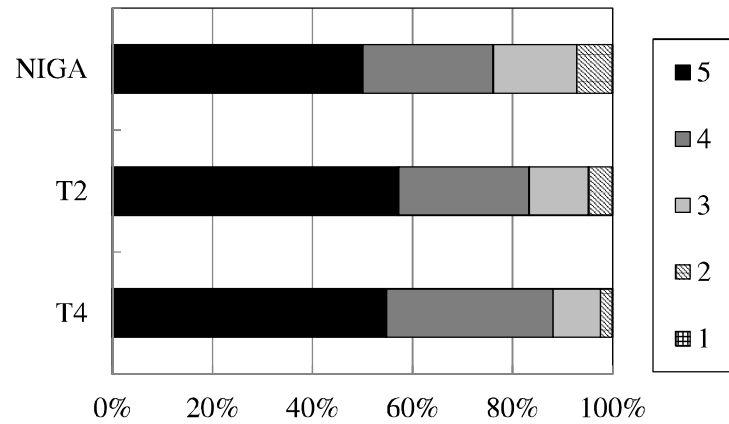
図 3.9: 生成された解候補に対する満足度の結果 (動画)



(a) 平均値と標準偏差

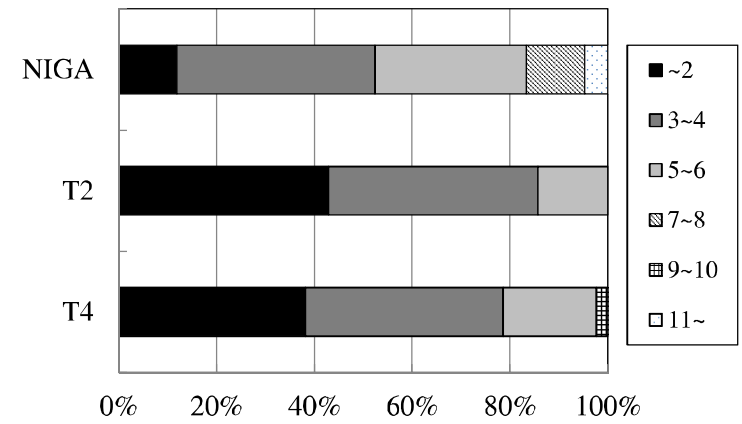


(a) 平均値と標準偏差



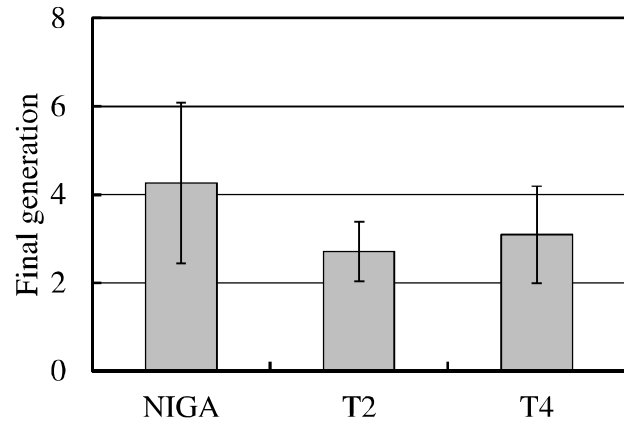
(b) 人数分布の割合

図 3.10: 生成された解候補に対する満足度の結果 (静止画)

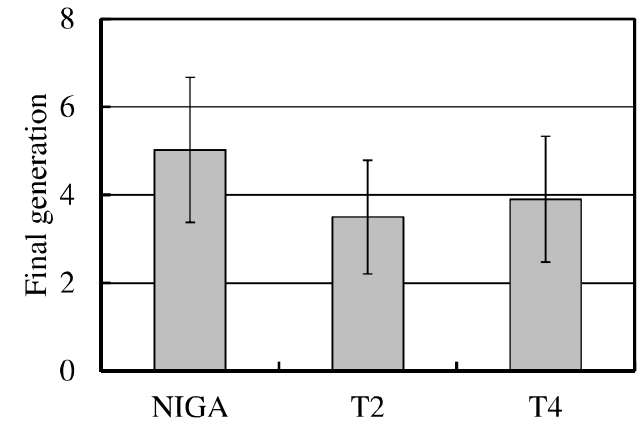


(b) 人数分布の割合

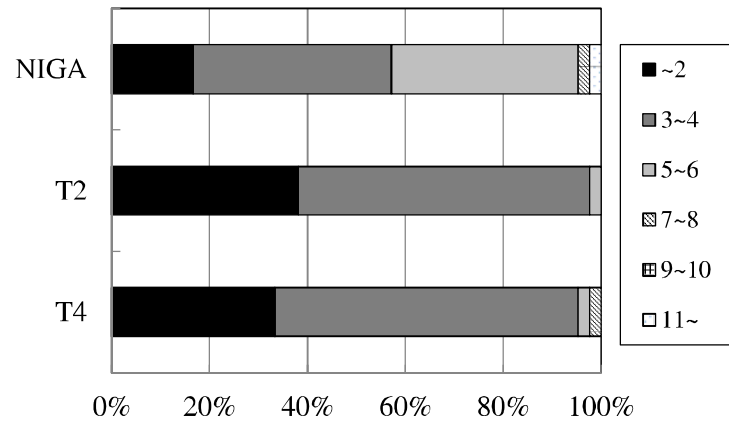
図 3.11: 終了世代数の結果 (音楽)



(a) 平均値と標準偏差

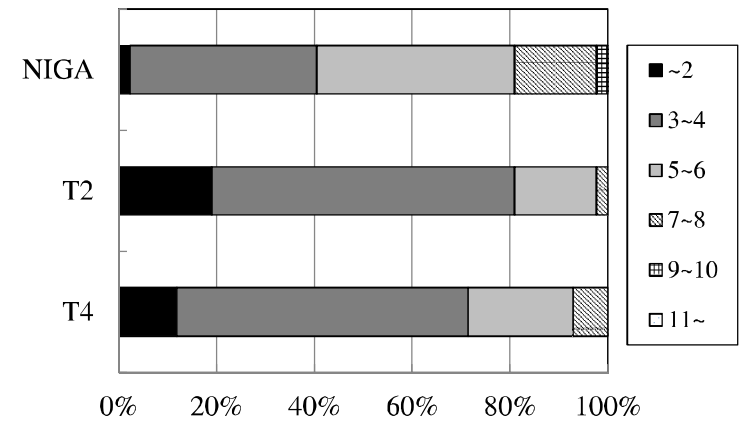


(a) 平均値と標準偏差



(b) 人数分布の割合

図 3.12: 終了世代数の結果 (動画)



(b) 人数分布の割合

図 3.13: 終了世代数の結果 (静止画)

表 3.2: フリードマン検定の結果

	Music		Animation		Image	
	Significance	Statistic ratio	Significance	Statistic ratio	Significance	Statistic ratio
Satisfaction level	n.s.	0.23	n.s.	3.46	n.s.	0.30
Final generation	n.s.	5.65	**	28.58	**	21.54
Usability	**	18.11	**	9.48	**	13.44
Evaluation time	n.s.	5.19	*	7.43	**	13.86

n.s.: Not significant. *: $p < 0.05$ (Critical value: 5.99), **: $p < 0.01$ (Critical value: 9.21)

表 3.3: シェッフェの多重比較の結果

	Music		Animation		Image	
	Significance	Statistic ratio	Significance	Statistic ratio	Significance	Statistic ratio
Satisfaction level	—	—	—	—	—	—
Final generation	—	—	** $T2$ -NIGA	42.48	** $T2$ -NIGA	25.43
	—	—	** $T4$ -NIGA	24.24	** $T4$ -NIGA	15.38
Usability	** $T2$ -NIGA	16.65	** $T2$ -NIGA	12.06	** $T2$ -NIGA	19.18
	** $T4$ -NIGA	15.23	—	—	* $T4$ -NIGA	6.17
Evaluation time	—	—	* $NIGA$ - $T4$	6.86	** $T2$ -NIGA	12.96
	—	—	—	—	* $T2$ - $T4$	6.86

*: $p < 0.05$ (Critical value: 5.99), **: $p < 0.01$ (Critical value: 9.21)

3.5.2 解評価のしやすさと評価時間

本項では、各評価インタフェースにおける解評価のしやすさ (Usability) 及び評価時間 (Evaluation time) の結果について考察する。

図 3.14~3.16 に各評価インタフェースにおける解評価のしやすさの平均値と標準偏差及び人数分布の割合を示す。図 3.14(a)~3.16(a) より、いずれの評価対象においても、T2 及び T4 は、NIGA より解評価のしやすさの平均値が高いことが確認された。また、図 3.14(b)~3.16(b) より、T2 及び T4 では、全被験者の約 70% が 5 段階評価において 4 点以上を付けていることが確認された。しかし、NIGA では、4 点以上を付けた被験者は、音楽の場合は全被験者の約 30%、動画の場合は全被験者の約 40%、静止画の場合は全被験者の約 60% に留まっている。表 3.2 のフリードマン検定の結果より、解評価のしやすさにおいては、いずれの評価対象でも有意水準 1% で統計的有意差が確認された。さらに、表 3.3 のシェッフェの多重比較の結果より、いずれの評価対象においても有意水準 1% で、T2 は NIGA より解評価が行いやすいことが確認された。また、評価対象が音楽・動画の場合は有意水準 1%、静止画の場合は有意水準 5% で、T4 は NIGA より解評価が行いやすいことが確認された。このため、評価対象が音楽・静止画の場合は T2 及び T4、動画の場合は T2 が NIGA より解評価が行いやすいと考えられる。

次に、被験者が評価に要した時間について考察する。IEC ユーザの評価負担軽減には、評価に要する時間も重要である。本実験における評価時間は、各システムにおいて被験者に初期解候補を提示してから、被験者が最終世代で好みの解候補を 1 つ選択し、終了するまでの時間とした。図 3.17~3.19 に各評価インタフェースにおける評価時間の平均値と標準偏差及び人数分布の割合を示す。評価対象が音楽・動画の場合は、解候補は必然的に時系列に提示されるため、静止画の場合より評価時間が長くなっている。図 3.17(a)~3.19(a) より、いずれの評価対象においても、T4 の評価時間が NIGA 及び T2 と比較して長いことが確認された。T4 においては、被験者が各対戦の評価時に大差・僅差の判定に迷うため、評価時間が長くなったと考えられる。また、図 3.17(b), 3.18(b) より、評価対象が音楽・動画の場合は、いずれの手法においても全被験者の約 70% が 900[s] 以内に評価を終えていることが確認された。図 3.19(b) より、静止画の場合は、T2 では、全被験者の約 80% が 240[s] 以内に評価を終えていることが確認された。しかし、NIGA 及び T4 では、240[s] 以内に評価を終えている被験者は、全被験者の約 50% に留まっている。表 3.2 のフリードマン検定の結果より、評価時間においては、評価対象が動画の場合は有意水準 5%、静止画の場合は有意水準 1% で統計的有意差が確認された。さらに、表 3.3 のシェッフェの多重比較の結果より、動画の場合は有意水準 5% で、T4 は NIGA より評価時間が長いことが確認された。静止画の場合は有意水準

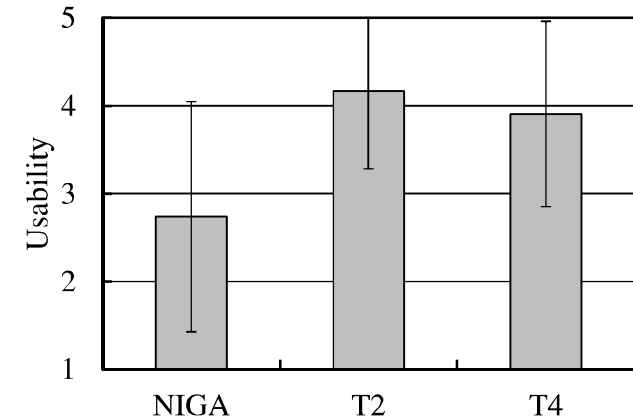
1%で、T2はNIGAより評価時間が短いことが確認された。また、有意水準5%で、T2はT4より評価時間が短いことが確認された。

本実験の結果より、いずれの評価対象、手法を用いても、被験者が満足のいく解候補を生成できることが確認された。

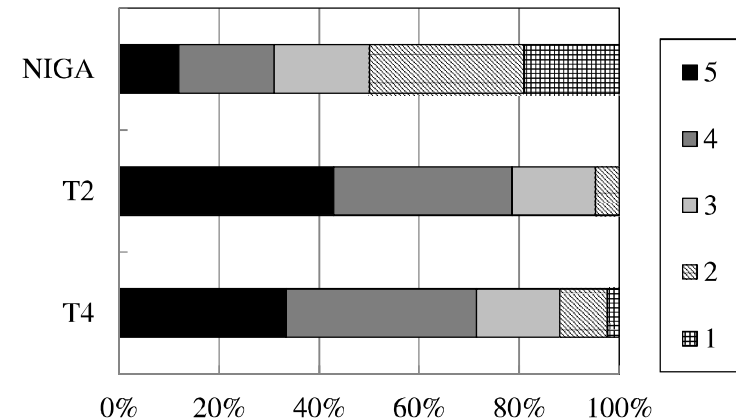
評価対象が音楽の場合においては、T2及びT4は、NIGAより解評価が行いやすいことが確認された。終了世代数、評価時間における統計的有意差は、有意水準5%において確認されなかったが、終了世代数はT2、T4、NIGAの順で少なく、評価時間もT2、T4、NIGAの順で短いことが確認された。

評価対象が動画の場合においては、T2はNIGAより解評価が行いやすいことが確認された。T2及びT4は、NIGAより終了世代数が少ないことが確認された。T2は、NIGA及びT4より評価時間が短い、統計的有意差は確認されなかった。また、T4はNIGAより評価時間が長いことが確認された。このため、評価対象が時系列データ、すなわち音楽・動画である場合は、T2はIECユーザーの解評価負担軽減に有効であると考えられる。

また、評価対象が静止画の場合においては、T2及びT4は、NIGAより解評価が行いやすいことが確認された。T2及びT4は、NIGAより終了世代数が少ないことが確認された。T2は、NIGA及びT4より評価時間が短いことが確認された。T4においては、ユーザーに対して各対戦の評価時に大差・僅差の判定に迷いを与えるという問題がある。しかし、T4は、NIGAより終了世代数が少ないことが確認された。このため、評価対象が静止画である場合は、T2及びT4は、IECユーザーの解評価負担軽減に有効であると考えられる。

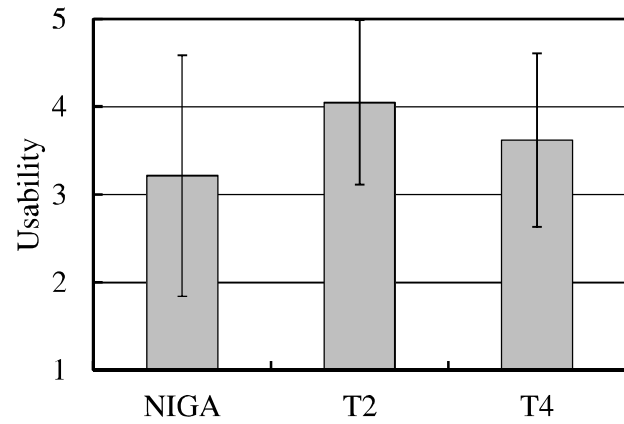


(a) 平均値と標準偏差

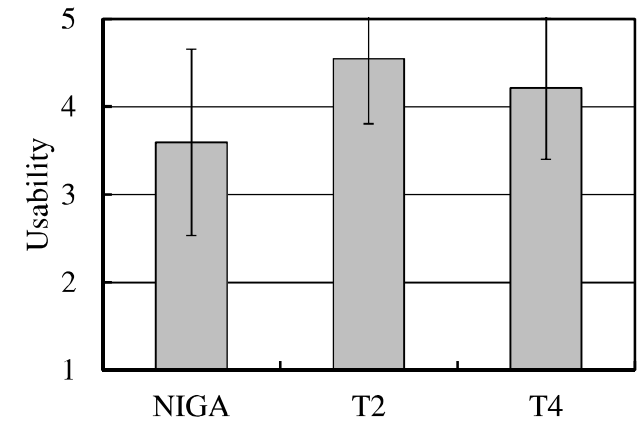


(b) 人数分布の割合

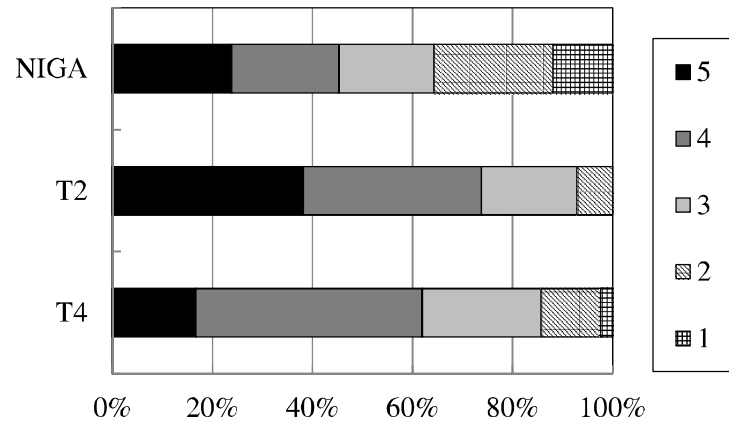
図 3.14: 解評価のしやすさの結果 (音楽)



(a) 平均値と標準偏差

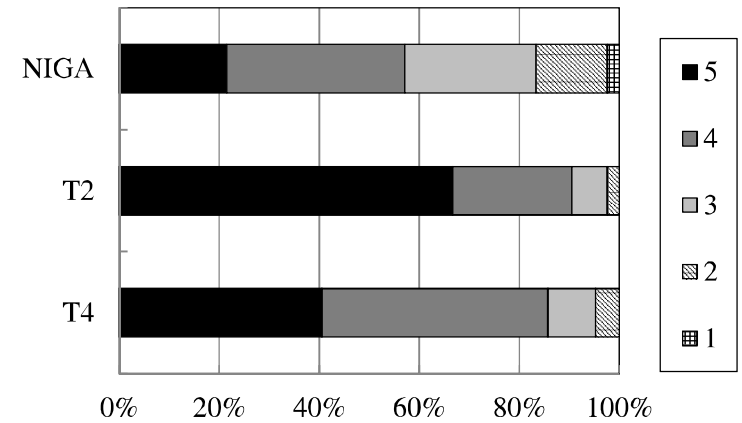


(a) 平均値と標準偏差



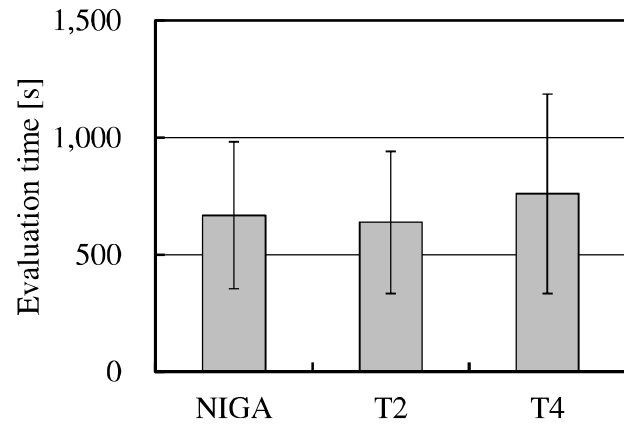
(b) 人数分布の割合

図 3.15: 解評価のしやすさの結果 (動画)

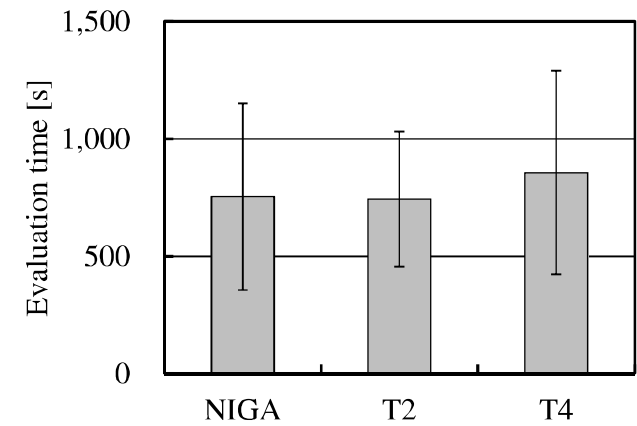


(b) 人数分布の割合

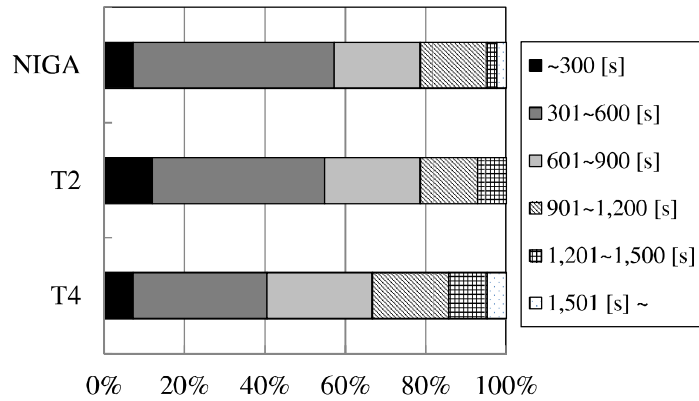
図 3.16: 解評価のしやすさの結果 (静止画)



(a) 平均値と標準偏差

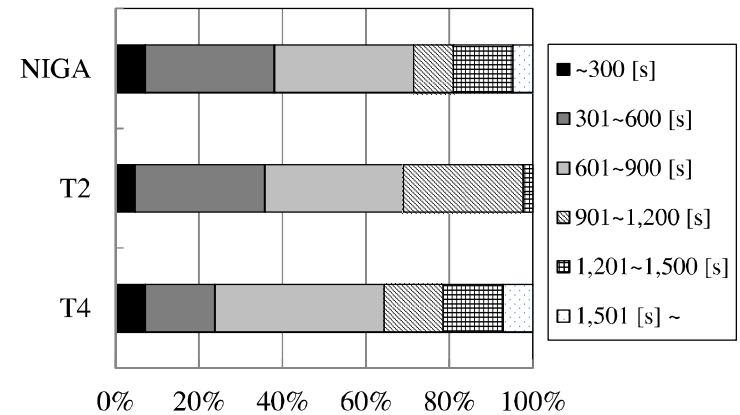


(a) 平均値と標準偏差



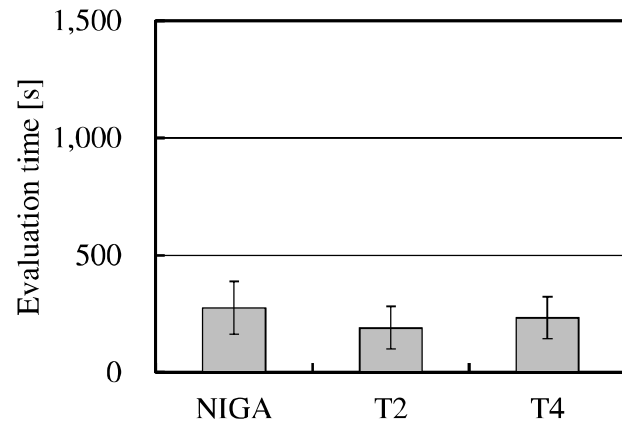
(b) 人数分布の割合

図 3.17: 評価時間の結果 (音楽)

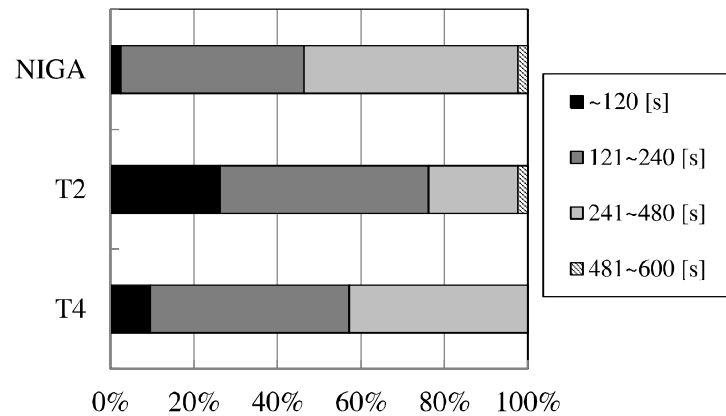


(b) 人数分布の割合

図 3.18: 評価時間の結果 (動画)



(a) 平均値と標準偏差



(b) 人数分布の割合

図 3.19: 評価時間の結果 (静止画)

3.5.3 被験者のコメント

本項では、トーナメント式評価手法に対する被験者のコメントと対処方法についてまとめる。以下に被験者の代表的なコメントを示す。

- (1) 見た目が類似した解候補同士の対戦は評価が困難である
- (2) 初戦で評価の高い解候補同士が対戦してしまう

(1)は、被験者が強制的にどちらかの解候補を勝者または敗者にしなければならないことが原因である。この問題は、「どちらのデザインも良い」、「どちらのデザインも良くない」という判定ができるボタンを追加することで、解決できると考えられる。これらのボタンが押された場合は、システム側でランダムに勝者を決定する必要がある。

(2)は、トーナメント対戦の組み合わせを毎世代ランダムに決定していることが原因である。この問題は、例えば評価値予測などを導入し、トーナメント対戦の組み合わせを決定する際に、評価が高いと予測される解候補と低いと予測される解候補が対戦するように調整することで、解決できると考えられる。

3.6 対話型タブサーチにおける実験結果

本節では、ITS の評価実験の結果について考察する。

3.6.1 生成された解候補に対する満足度

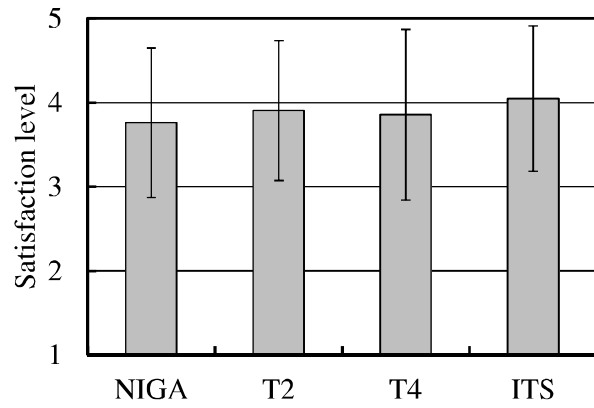
本項では、ITS の評価実験における、各手法によって生成された解候補に対する満足度 (Satisfaction level) 及び終了世代数 (Final generation) の結果について考察する。

図 3.20 に各手法で生成された解候補に対する満足度の平均値と標準偏差及び人数分布の割合を示す。図 3.20(a) より、各手法によって生成された解候補に対する満足度は、5 段階評価において約 4 点になっていることが確認された。また、図 3.20(b) より、いずれの手法においても、全被験者の約 50%以上が 4 点以上を付けていることが確認された。そこで、各手法の満足度の統計的有意差を確認するため、フリードマン検定を行った。表 3.4 にフリードマン検定の結果を示す。表 3.4 には、3.6.2 項において考察する解評価のしやすさ及び評価時間に関する検定結果も合わせて示している。満足度においては、有意水準 5% で統計的有意差は確認されなかった。このため、いずれの手法を用いても被験者が満足のいく解候補を生成できることが確認された。

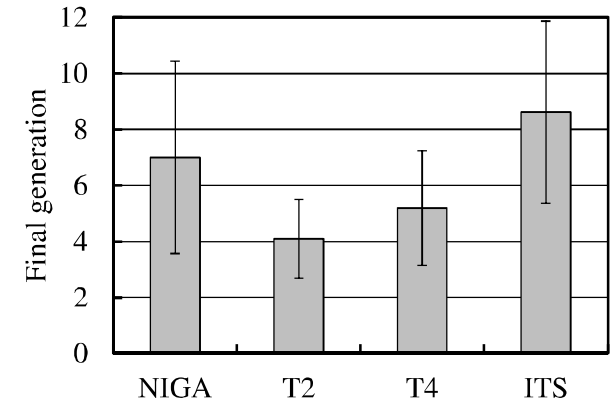
しかし、満足のいく解候補が得られるの要した世代交代数は、手法によって異なる。図 3.21 に各手法の終了世代数の平均値と標準偏差及び人数分布の割合を示す。図 3.21(a) より、ITS が 4 手法の中で最も多く世代交代を要していることが確認された。また、図 3.21(b) より、5 世代目までに評価を終了した被験者の割合は、ITS、T2、T4、NIGA の順で大きく、ITS が最も小さいことが確認された。さらに、表 3.4 のフリードマン検定の結果より、終了世代数においては有意水準 5% で統計的有意差が確認された。フリードマン検定において統計的有意差が確認された場合は、シェッフェの多重比較により、どの手法間に統計的有意差があるかを確認した。表 3.5 にシェッフェの多重比較の結果を示す。表 3.5 中の斜体表記は、統計的検定により良好な結果を示した手法である。表 3.5 には、3.6.2 項において考察する解評価のしやすさ及び評価時間に関する検定結果も合わせて示している。表 3.5 より、有意水準 1% で、T2 は NIGA 及び ITS より、T4 は ITS より終了世代数が少ないことが確認された。

これは、ITS においては、選択された解候補への近傍収束が早いことに起因していると考えられる。ITS では、提示された解候補に好みのものがない場合、次に提示される解候補は好みでない解候補の近傍において生成される。その後、近傍探索を繰り返し、好みの解候補を探索するため、ITS は他の 3 手法より多くの世代交代数を要したと考えられる。

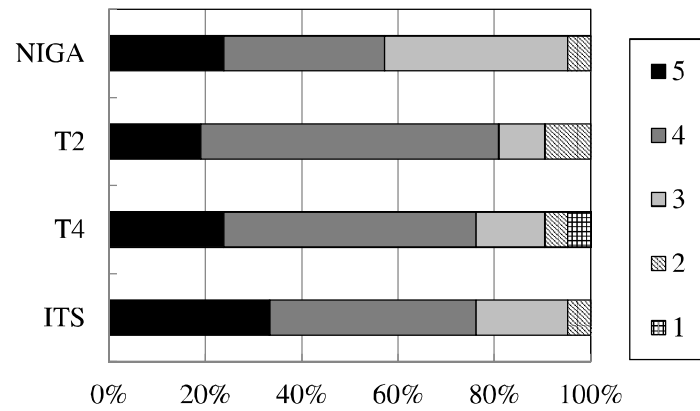
また、2.7 節におけるシミュレーション結果では、ITS の世代交代数は NIGA より少なく、図 3.21 の結果とは異なる。これは、実ユーザを対象とした場合は、ITS の解評価作業が単純なため、被験者は満足のいく解候補を生成できてもさらに良いものを作成したいと考え、その結果多くの世代交代を要したためであると考えられる。



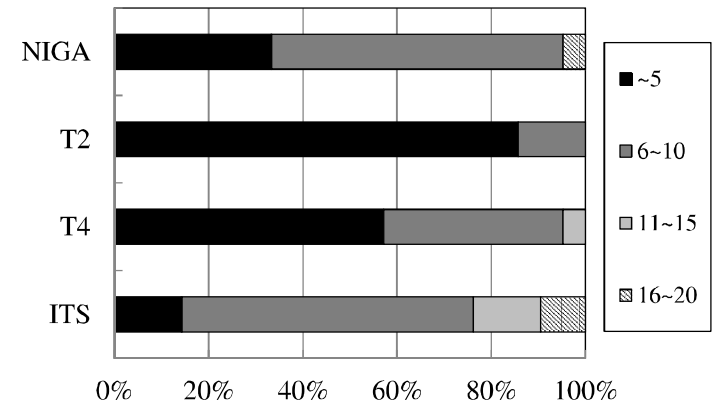
(a) 平均値と標準偏差



(a) 平均値と標準偏差



(b) 人数分布の割合



(b) 人数分布の割合

図 3.20: 生成された解候補に対する満足度の結果 (ITS)

図 3.21: 終了世代数の結果 (ITS)

表 3.4: フリードマン検定の結果 (ITS)

	Significance	Statistic ratio
Satisfaction level	n.s.	1.17
Final generation	**	30.07
Usability	**	16.90
Evaluation time	**	16.77

n.s.: Not significant, * : $p < 0.05$ (Critical value: 7.81)** : $p < 0.01$ (Critical value: 11.34)

表 3.5: シェッフエの多重比較の結果 (ITS)

	Significance	Statistic ratio
Satisfaction level	—	—
Final generation	** $T2$ -NIGA	13.24
	** $T2$ -ITS	31.57
	** $T4$ -ITS	13.24
Usability	** $T2$ -NIGA	11.64
	** ITS -NIGA	13.94
Evaluation time	* $T2$ -T4	9.66
	** ITS -T4	12.86

* : $p < 0.05$ (Critical value: 7.81), ** : $p < 0.01$ (Critical value: 11.34)

3.6.2 解評価のしやすさと評価時間

本項では、ITS の評価実験における、各評価インタフェースにおける解評価のしやすさ (Usability) 及び評価時間 (Evaluation time) の結果について考察する。

図 3.22 に各評価インタフェースにおける解評価のしやすさの平均値と標準偏差及び人数分布の割合を示す。図 3.22(a) より、T2 及び ITS は、NIGA 及び T4 より解評価のしやすさの平均値が高いことが確認された。また、図 3.22(b) より、T2 では全被験者の約 30%、ITS では全被験者の約 50% が 5 段階評価において 5 点を付けていることが確認された。NIGA 及び T4 においては、5 点を付けた被験者は全被験者の約 5% に留まっている。表 3.4 のフリードマン検定の結果より、解評価のしやすさにおいては、有意水準 1% で統計的有意差が確認された。さらに、表 3.5 のシェッフエの多重比較の結果より、有意水準 1% で T2 及び ITS は NIGA より解評価が行いやすいことが確認された。これは、T2 及び ITS の解評価が NIGA と比較して、単純であることが理由であると考えられる。

次に、被験者が評価に要した時間について考察する。IEC ユーザの評価負担軽減には、評価に要する時間も重要である。本実験における評価時間は、各システムにおいて被験者に初期解候補を提示してから、被験者が ITS ではタブーリスト中、NIGA、T2、T4 では最終世代の提示解候補から最も好みの解候補を 1 つ選択し、終了するまでの時間とした。図 3.23 に各評価インタフェースにおける評価に要した時間の平均値と標準偏差及び人数分布の割合を示す。図 3.23(a) より、ITS は、4 手法の中で最も評価時間が短いことが確認された。また、図 3.23(b) より、T2 及び ITS では、全被験者の約 60% が 120[s] 以内に評価を終了していることが確認された。しかし、NIGA 及び T4 では、120[s] 以内に評価を終了している被験者は、全被験者の約 20% に留まっている。表 3.4 のフリードマン検定の結果より、評価時間においては、有意水準 1% で統計的有意差が確認された。さらに、表 3.5 のシェッフエの多重比較の結果より、有意水準 5% で T2 は T4 より評価時間が短く、有意水準 1% で ITS は T4 より評価時間が短いことが確認された。

ITS は、3.6.1 項で述べたように、4 手法の中で最も世代交代数が多い。しかし、ITS の 1 世代あたりの解評価作業は、提示された解候補の中から最も好みの解候補を 1 つ選択するのみである。そのため、ITS における評価時間は、4 手法の中で最も短くなったと考えられる。

また、T2 における解評価のしやすさ及び評価時間の結果は ITS と同等であるが、T2 の終了世代数は ITS より少ない。しかし、T2 においては、毎世代の対比較による解評価は 7 回となる。そのため、T2 の対比較回数、T2 の終了世代数の平均値である

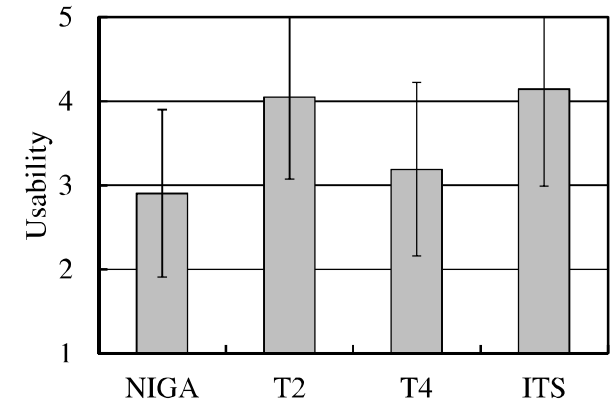
約4世代と掛け合わせると約28回となる。ITSにおいては、一度の評価でユーザが比較する解候補数は8個となるが、解候補の評価回数と世代交代数は同値であり、ユーザの評価回数は約8回で済む。このため、ITSはT2よりユーザの評価回数が少なく、解評価負担軽減に有効であると考えられる。

3.6.3 被験者のコメント

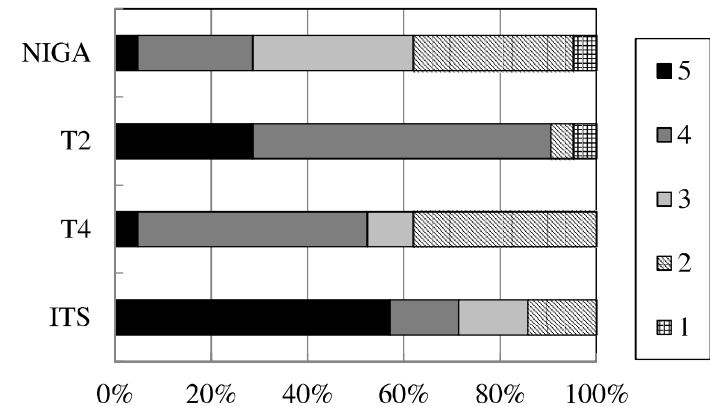
本項では、ITSに対する被験者のコメントと対処方法についてまとめる。以下に被験者の代表的なコメントを示す。

(1) 好みの解候補を得るのに多くの解候補選択が必要である

(1)は、ITSにおける局所探索が原因である。ITSでは、提示された解候補の中に好みの解候補が存在しない場合においても、強制的に好みの解候補を1つ選択しなければならない。この場合、被験者の好みでない近傍解候補が生成されてしまい、好みでない解候補の近傍から抜け出し、被験者の好みの解候補を探索するため、多くの世代交代を要してしまうおそれがある。この問題を解決するためには、ITSにおける近傍解候補生成範囲の調節や並列探索を行うITSの応用が有効であると考えられる。

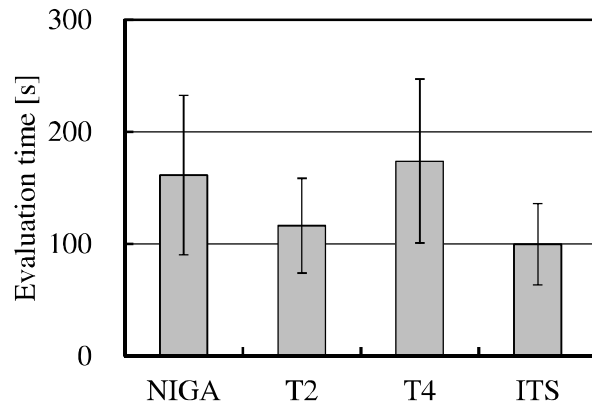


(a) 平均値と標準偏差

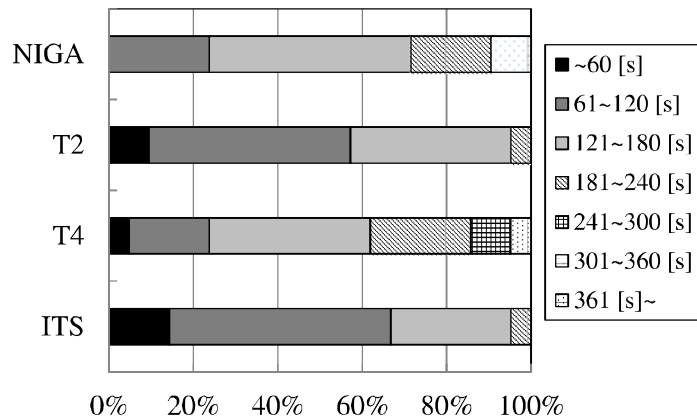


(b) 人数分布の割合

図 3.22: 解評価のしやすさの結果 (ITS)



(a) 平均値と標準偏差



(b) 人数分布の割合

図 3.23: 評価時間の結果 (ITS)

3.7 結言

本章では、トーナメント式評価手法及び ITS の実ユーザに対する有効性を検証するために行った評価実験について述べた。各評価実験の結果は、以下のとおりとなった。

(1) トーナメント式評価手法の評価実験

評価対象が時系列データ、すなわち音楽・動画である場合は、T2 は NIGA よりも解評価が行いやすいことが確認された。また、T2 は T4 より評価時間が短く、終了世代数も少ないことが確認された。このため、評価対象が音楽・動画である場合は、T2 は IEC ユーザの解評価負担軽減に有効であると考えられる。

また、評価対象が静止画の場合においては、T2 及び T4 は、NIGA より解評価が行いやすいことが確認された。さらに、T2 は、NIGA 及び T4 より評価時間が短いことが確認された。T2 及び T4 は、NIGA より終了世代数が少ないことが確認された。このため、評価対象が静止画である場合は、T2 及び T4 は、IEC ユーザの解評価負担軽減に有効であると考えられる。

(2) ITS の評価実験

ITS は NIGA 及び T4 よりも解評価が行いやすいことが確認された。また、ITS における評価時間は、4 手法の中で最も短いことが確認された。このため、ITS においては、世代交代数は 4 手法の中で最も多いが、解評価のしやすさ及び評価時間は 4 手法の中で最も良好であることが確認された。したがって、ITS は IEC ユーザの解評価負担軽減に有効であると考えられる。

本章では、実ユーザを対象とした評価実験により、T2、T4、ITS においてユーザの解評価負担を軽減するという観点より有効性が確認された。第 4 章では、多くのユーザの投票により解候補を評価する TBMP について述べる。

第4章 複数参加型トーナメント方式

4.1 緒言

本章では、多くのユーザの投票によりトーナメントの勝敗を決定する手法である TBMP1 について述べる [22].

まず、多くのユーザの感性を用いた IEC に関する先行研究を概括する。次に、TBMP1 について述べる。TBMP1 は、多くのユーザの感性を投票として獲得し、トーナメント形式で解候補を評価する手法である。ユーザは、提示された 2 つの解候補のうち、好みの方に 1 票を投じる。そのため、ユーザ 1 人 1 人の解評価作業が単純になり、解評価負担を軽減できると考えられる。さらに、全てのユーザが全トーナメント対戦の投票に参加する必要がなくなることも考えられる。

本章では、TBMP1 の解候補の進化性能に関する有効性を数値シミュレーションにより検証する。本シミュレーションでは、実ユーザの代わりにビット列で作成された評価エージェントが解候補を評価する。本シミュレーションにおいては、投票に参加する評価エージェントの割合に関する検討及び通常の投票方式である GVM を比較手法とした性能比較を行う。

4.2 多人数参加型の対話型進化計算

本節では、多くのユーザの感性を用いたIECに関する先行研究について述べる。

複数のユーザの感性を利用したIECシステムに関する研究は、これまでにいくつか行われている [52]-[54]。

田川らの研究では、ネットワーク接続されたIGAシステムを数人のユーザが個別に操作し、建築物の内装をデザインする手法が提案されている [52]。この手法においては、まず、各ユーザに4つの建築物内装デザインを提示し、一対比較法により評価してもらう。そして、各ユーザの一対比較法による評価結果から解候補の評価値を算出する。様々なユーザを被験者とした評価実験より、合意形成にIGAを用いることは有効であるとされている。

三木らの研究では、合意形成を目的とした並列分散IGAが提案されている [53]。並列分散IGAは、複数のユーザでネットワーク接続されたIGAシステムを個別に操作し、生成された解候補をユーザ間で共有するシステムである。各ユーザは、衣服の配色を5段階で評価する。そして、各ユーザが良いと判断した解候補を移住個体として自分以外のユーザの提示解候補に組み込み、ユーザ間で好みの解候補を共有する。評価実験の結果より、並列分散IGAは合意形成に有効であるとされている。

逸見らは、「何もしないよりは、不確かな他人の評価特性を使ってでも、ユーザの疲労軽減や解探索の高速化に役立たせた方が良からう」というコンセプトに基づいたIECシステムを提案している [54]。この手法は、ユーザ本人の評価と予めコンピュータに組み込まれた複数のユーザの評価特性による評価の両方を用いて、解候補を評価するIECシステムである。この手法においては、シミュレーション結果より、複数のユーザの評価特性を用いない場合と比較して、解の収束性が向上することが確認されている。

しかし、これらのシステムにおいては、IECユーザは数人であり、数1,000人程度の不特定多数のユーザが解評価に参加するようなIGAシステムの研究は報告されていない。そこで、Web上で数1,000人程度のユーザの感性を投票として獲得し、IECの解評価に用いることを想定した多人数参加型のシステムを提案する。このような手法により、例えばWebサイト上の投票システムのように、様々なユーザが気軽に投票できるIECシステムを構築できると考えられる。また、一定の期間内にユーザの投票を受け入れるため、全てのユーザが同時に解評価に参加する必要がない。そのため、それぞれのユーザは、個人の都合に合わせて解評価に参加でき、ユーザの評価負担を軽減できると考えられる。

多くのユーザが投票により解候補を評価するIECインタフェースにおいては、IECにより生成された全解候補を一覧表示し、その中で最も好みの解候補にユーザが投票

する方法が考えられる。しかし、全解候補を同時に提示した場合には、ユーザに解候補選択の迷いを与えかねない。そこで、解の評価手法にトーナメント式評価手法を用いる。トーナメント式評価手法においては、ユーザは一対比較により解候補を評価する。トーナメント式評価手法は一対比較法よりも評価回数を削減でき、ユーザの解評価負担を軽減できると考えられる。第2、3章における数値シミュレーション及び評価実験より、トーナメント式評価手法はNIGAより解評価が行いやすいことが示されている。また、TBMPにおいては、Webサイト上で多くのユーザが投票に参加するようなIGAシステムを想定しているため、投票のしやすさという観点よりトーナメント式評価手法は有効であると考えられる。

4.3 複数参加型トーナメント方式

本節では、TBMP1の概要及び解候補の評価方法について述べる。

4.3.1 システム概要

TBMP1は、多くのユーザの投票によりトーナメントの勝敗を決定する手法である[22]。そのため、多くのユーザが満足のいくものや流行の変化に対応した解候補を生成できると考えられる。TBMP1におけるユーザの解評価作業は、提示された2つの解候補のうち、好みの方に1票を投じるのみとなる。そのため、ユーザ1人1人の解評価における負担を軽減できると考えられる。さらに、全てのユーザが全トーナメント対戦の投票に参加する必要がなくなることも考えられる。TBMP1においては、Web上で多くのユーザの投票を獲得し、多くのユーザの感性を反映した様々なものやデザインを自動的に生成することを想定している。

図4.1にTシャツの配色を例としたTBMP1のシステム概要を示す。まず、初期遺伝子集団を生成し、生成された解候補をトーナメント表に配置する。次に、多くのユーザにWeb上で各対戦の2つのデザインを提示し、好みのデザインの方に投票してもらう。TBMP1では、一定の期間内にユーザの投票を受け入れて、期間が過ぎれば次の対戦に移行する。決勝戦まで終了すれば、解候補を評価し、GA処理を行う。そして、新しい解候補を提示し、再びユーザに投票を求める。これらの処理を繰り返して、多くのユーザの感性を反映したデザインを作成する。

トーナメント式評価手法においては、初期解候補の対戦組み合わせにより勝敗結果が左右されてしまう問題がある。このため、優秀な解候補同士がトーナメントの初戦で対戦してしまうと、解候補を正確に評価できないおそれがある。この問題を解決するため、TBMP1においては、各対戦の投票数を基にして解候補に評価値を与える。各対戦における解候補の獲得投票数は、多くのユーザの感性を表現しているといえる。このため、多くのユーザの感性をより反映した評価値を与えることができると考えられる。

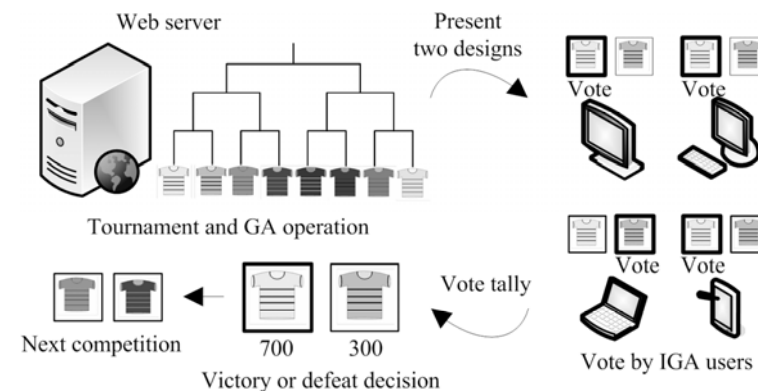


図 4.1: TBMP1 のシステム概要

4.3.2 解候補の評価方法

TBMP1においては、各対戦の獲得投票数に基づいて解候補を評価する。図4.2にユーザが1,000人のときの投票結果の一例を示す。まず、優勝した解候補である P_4 に10点を与える。そして、準優勝以下の解候補には、式(4.1)を用いて評価値を与える。

$$X_{lose} = X_{win} \times \frac{Y_{lose}}{Y_{win}} \quad (4.1)$$

式(4.1)において、 X_{win} は各対戦における勝者の評価値、 X_{lose} は各対戦における敗者の評価値である。また、 Y_{win} は各対戦における勝者の獲得投票数、 Y_{lose} は各対戦における敗者の獲得投票数である。例えば、準優勝した解候補である P_7 の場合は評価値を X_7 として、式(4.2)のようになる。

$$X_7 = 10 \times \frac{300}{700} \cong 4.3 \quad (4.2)$$

同様に、準決勝で P_4 に敗れた P_2 の場合は、評価値を X_2 として、式(4.3)のようになる。

$$X_2 = 10 \times \frac{200}{800} = 2.5 \quad (4.3)$$

なお、評価値が1より小さくなった場合は、1に切り上げるものとした。これらの評価値は、GAにおける選択及び交叉処理で用いる。

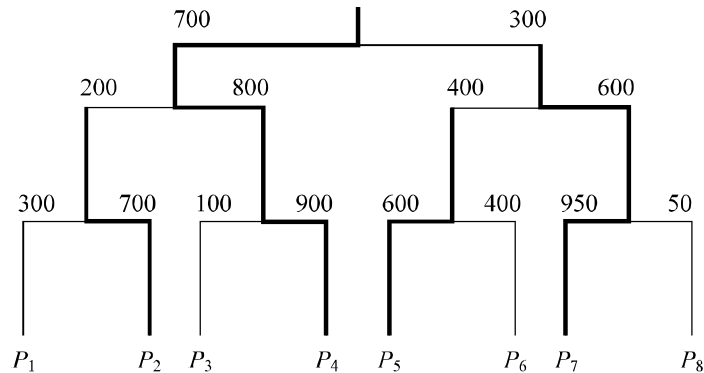


図 4.2: 投票結果の一例

4.4 数値シミュレーションによる性能評価

本節では、TBMP1の解候補の進化性能に関する有効性を検証するために行う数値シミュレーションの概要及び評価エージェントの集団性の表現方法、性能比較における比較手法であるGVMについて述べる。

4.4.1 シミュレーション概要

本章では、実ユーザの代わりにビット列で作成された評価エージェントを用いた数値シミュレーションにおいて、TBMP1の解候補の進化性能に関する有効性を検証する。本シミュレーションでは、解候補を定量的に評価し、TBMP1の解候補の進化性能を検証する。本シミュレーションにおいては、評価エージェントの投票参加割合に関する検討及びGVMを比較手法とした性能比較を行う。GVMは、ユーザが全解候補を一度に比較して、その中で最も好みの解候補に投票する評価インタフェースを想定した手法である。

図4.3にTBMP1のシミュレーションの流れを示す。まず、初期遺伝子集団を生成し、生成された解候補をトーナメント表に配置する。次に、評価エージェントによる投票を行い、トーナメント対戦を進行する。決勝戦まで終了すれば、解候補に評価値を与え、GA処理を行う。そして、新しい解候補を提示し、再び評価エージェントによる投票を行う。

表4.1に本シミュレーションのパラメータを示す。ユーザの解評価負担の軽減のため、IGAでは世代交代数に10~20世代が広く用いられている。しかし、TBMP1では、ユーザは各対戦において1票を投じるのみであり、解評価における肉体的心理的負担を軽減できると考えられる。そのため、本シミュレーションでは100世代目までシミュレーションを行う。表4.1の突然変異率は、予め行ったシミュレーションより、TBMP1が最良のパフォーマンスを発揮できるように決定した。突然変異処理は、エリート解候補以外の解候補における各遺伝子座を突然変異率で反転する処理とした。確率的な影響を排除するため、シミュレーション試行回数は100回とした。なお、4.5節におけるシミュレーション結果は、シミュレーション試行回数分の平均値である。

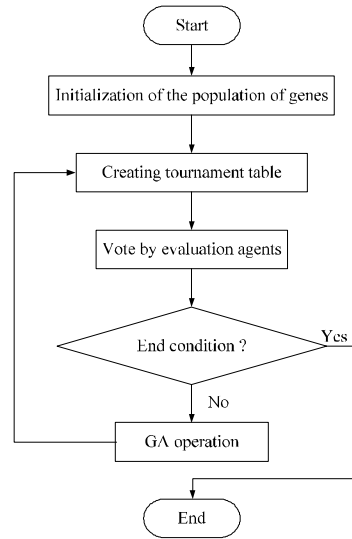


図 4.3: TBMP1 のシミュレーションの流れ

表 4.1: TBMP1 のシミュレーションパラメータ

Gene row	Bit string
Gene length	20, 30, 40, 50bits
Generation	100
Candidate solutions	16
Selection	Roulette selection + Elite preservation
Crossover	Uniform crossover
Mutation rate	5%

4.4.2 評価エージェントの集団性の表現

本項では、本シミュレーションで用いる評価エージェントの集団性の表現方法について述べる。TBMP1では、多くのユーザが解評価に参加するIGAシステムを想定しているため、多くの評価エージェントを作成する。そのため、評価エージェントには好みの傾向、すなわち集団性を表現させている。

図 4.4 に評価エージェントの集団性の概念を示す。Tシャツの配色には多くの種類が存在し、ユーザによって好みは分かれる。しかし、全てのユーザが自分以外のユーザと全く共通性のない感性を持っていることは少なく、図 4.4 のように、ある程度類似した感性を持っている集団が存在すると考えられる。本シミュレーションでは、このような集団性を評価エージェントにおいて表現する。

図 4.5 に本シミュレーションにおける評価エージェントの集団性の表現方法を示す。図 4.5 は、ビット数が 20bits の評価エージェントを 5 グループ作成する場合を示している。まず、5 種類の評価エージェントを他の 4 種類の評価エージェントとのハミング距離が 8bits になるように作成する。そして、これら 5 種類の評価エージェントにおいて、それぞれをハミング距離が 4bits 以下の範囲で反転することにより、評価エージェントの集団を生成する。集団の基となる 5 種類の評価エージェント同士のハミング距離は、ビット数の 40% とした。集団内のハミング距離は、ビット数の 20% とした。すなわち、ビット数が 30bits の場合は、集団間のハミング距離は 12bits、集団内のハミング距離は 6bits となる。本シミュレーションでは、ビット数が 20, 30, 40, 50bits の評価エージェントを用いる。なお、本評価エージェントの集団性は GA を用いて生成しており、詳細は付録 B に記す。

さらに、本シミュレーションでは、2.5.2 項で述べた方法を用いて、各評価エージェントに多峰性感性空間による評価を表現させている。これは、図 4.5 で生成された評価エージェントのビットパターンに重み $w_0 = 1.0$ を与えることにより実現される。 w_1, w_2, w_3, w_4 が割り当てられるビットパターンは、評価エージェント及び試行回数毎にランダムに生成される。評価エージェントは 2.5.2 項と同様の方法により、各対戦において 2 つの対戦候補の評価値を求め、評価値が高い候補に投票する。なお、4.5 節におけるシミュレーション結果の絶対評価値は、各世代のエリート候補に対する全評価エージェントの評価値 $fitness$ に 10 を乗算し、小数第 1 位を四捨五入した値の試行回数分の平均値である。

本シミュレーションにおいては、実ユーザにおける様々な集団のケースに対応するため、評価エージェントの集団性を変化させたシミュレーションを行う。表 4.2 に本シミュレーションで用いる評価エージェントの集団性のパターンを示す。Pattern 1 は、

集団の人数が均等に分かれた状態を表現している。Pattern 2は、集団の人数に偏りを持たせた状態を表現している。Pattern 3は、Pattern 2より集団の人数に偏りを持たせた状態を表現している。

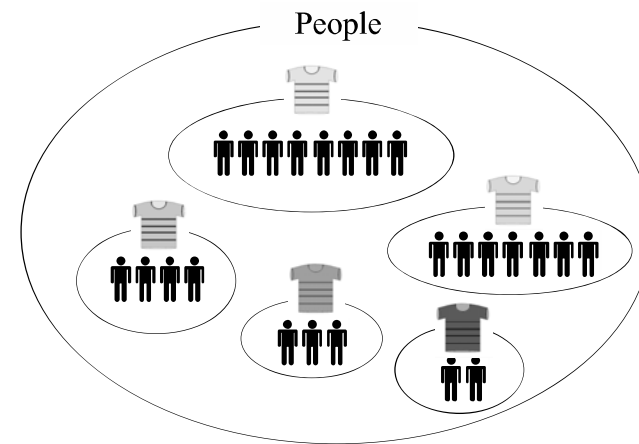


図 4.4: 集団性の概念

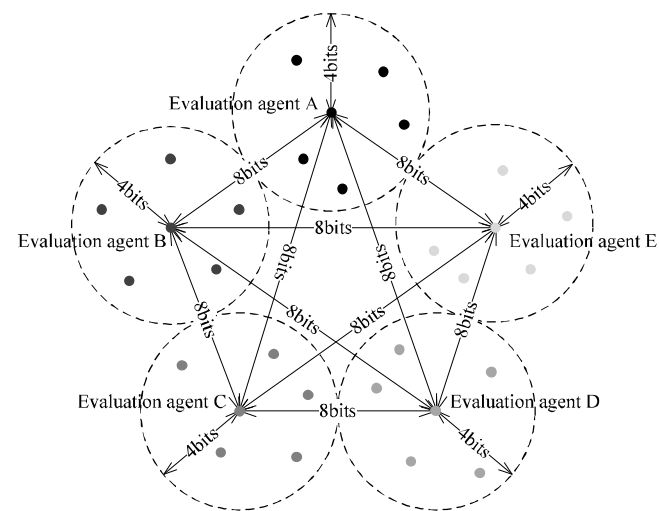


図 4.5: 評価エージェントの集団性の表現

表 4.2: 評価エージェントの集団性のパターン

	Group A	Group B	Group C	Group D	Group E	Total
Pattern 1	200	200	200	200	200	1,000
Pattern 2	500	300	200	0	0	1,000
Pattern 3	800	150	50	0	0	1,000

4.4.3 比較手法

本項では、本シミュレーションにおいて比較手法として用いる GVM について述べる。

GVM では、ユーザが全解候補を一度に比較して、その中で最も好みの解候補に投票する評価インタフェースを想定している。すなわち、評価エージェントは全解候補に対して多峰性感性空間による評価値を求め、その中で最も評価値の高い解候補に投票する。GVM においても TBMP1 と同様に、評価エージェントの投票結果に基づいて、解候補に評価値を与える。まず、最も獲得投票数が多い解候補に 10 点を与える。そして、2 位以下の解候補には式 (4.4) を用いて評価値を与える。

$$X_2 = X_1 \times \frac{Y_2}{Y_1} \quad (4.4)$$

式 (4.4) において、 X_1 は評価値を求める解候補より順位が 1 つ上の解候補の評価値、 X_2 は求める評価値、 Y_1 は X_1 の獲得投票数、 Y_2 は X_2 の獲得投票数である。なお、TBMP1 と同様に評価値が 1 より小さくなった場合は、1 に切り上げるものとした。

また、GVM における GA パラメータは、表 4.1 と同様である。

4.5 シミュレーション結果

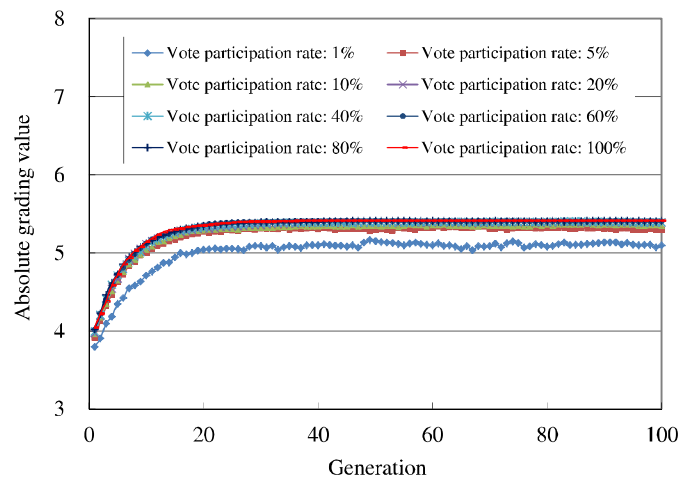
本節では、数値シミュレーションにおける評価エージェントの投票参加割合に関する検討及び GVM を比較手法とした性能比較の結果について考察する。

4.5.1 評価エージェントの投票参加割合に関する検討

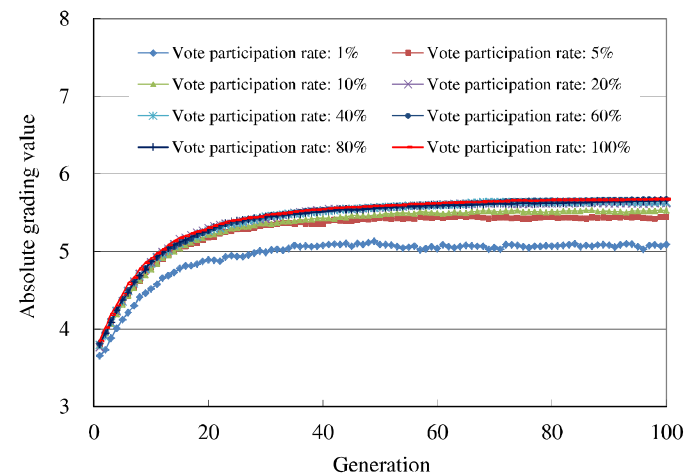
本項では、投票に参加する評価エージェントの割合を変化させた場合の進化性能について考察する。TBMP1 においては、一定の期間内にユーザの投票を受け入れて、期間が過ぎれば次の対戦に移行するシステムを想定している。そのため、全てのユーザが全トーナメント対戦の投票に参加しない場合においても、良好な進化性能が得られることが求められる。

図 4.6~4.8 に評価エージェントの集団性の違いによる投票参加割合 (Vote participation rate) の検討結果を示す。各世代における絶対評価値 (Absolute grading value) は、各世代のエリート解候補に対する全評価エージェントの絶対評価値のシミュレーション試行回数分の平均値である。本シミュレーションでは、投票する評価エージェントは各対戦においてランダムに選択され、同一対戦で同じ評価エージェントが複数回投票することがないようにした。図 4.6~4.8 より、いずれの遺伝子長においても評価エージェントの投票参加割合が低ければ、絶対評価値が低くなることが確認された。しかし、集団性に偏りのある Pattern 3 においては、投票参加割合が低下しても、Pattern 1, 2 と比較して絶対評価値の減少量は小さい。Pattern 3 においては、投票参加割合が低下しても、多数派集団 (Group A) から投票する評価エージェントが選択される確率が高いと考えられる。そのため、評価エージェント集団全体の絶対評価値において、大きな差が確認されなかったと考えられる。

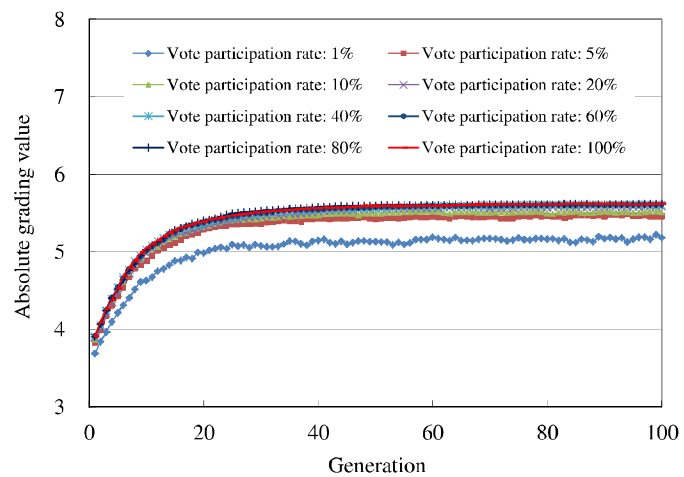
図 4.6~4.8 より、評価エージェントの投票参加割合が 60% 以上であれば、投票参加割合が 100% のときと同等の進化性能が得られることが確認された。したがって、TBMP1 においては、全ての対戦において全ユーザの約 60% の投票を得られれば、多くのユーザが満足のいくものを生成できると考えられる。そのため、それぞれのユーザが時間的に束縛されず、部分的に投票に参加しても多くのユーザが満足のいくものを生成できるシステムが実現可能であると考えられる。



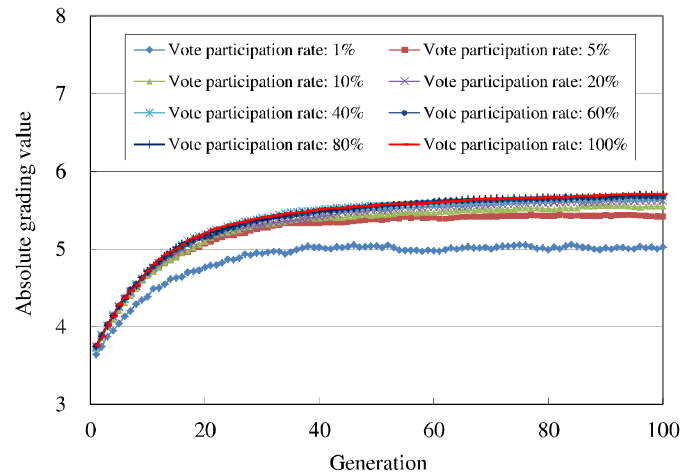
(a) 20bits



(c) 40bits



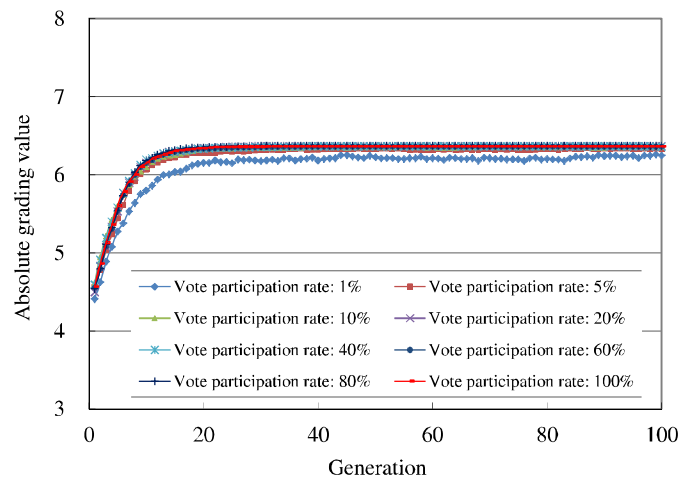
(b) 30bits



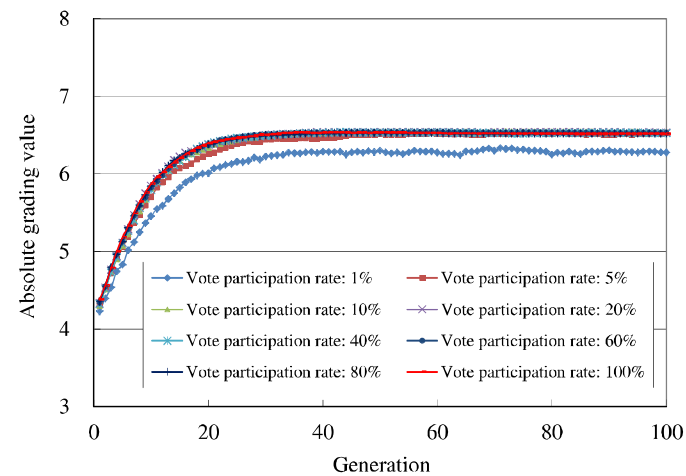
(d) 50bits

図 4.6: 投票参加割合の検討結果 (Pattern 1)

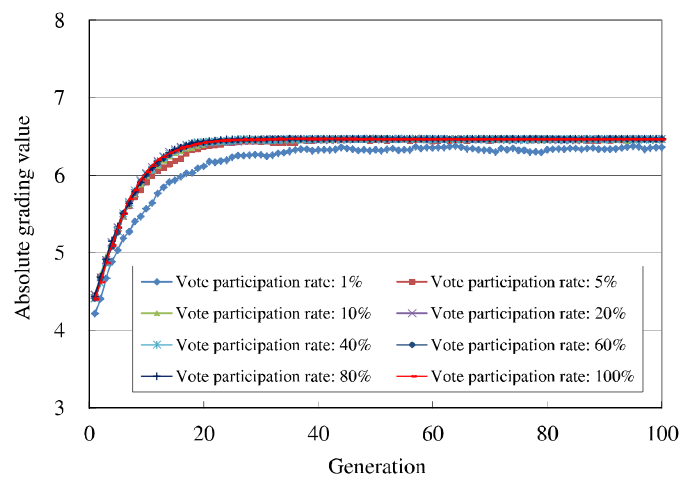
図 4.6: 投票参加割合の検討結果 (Pattern 1)



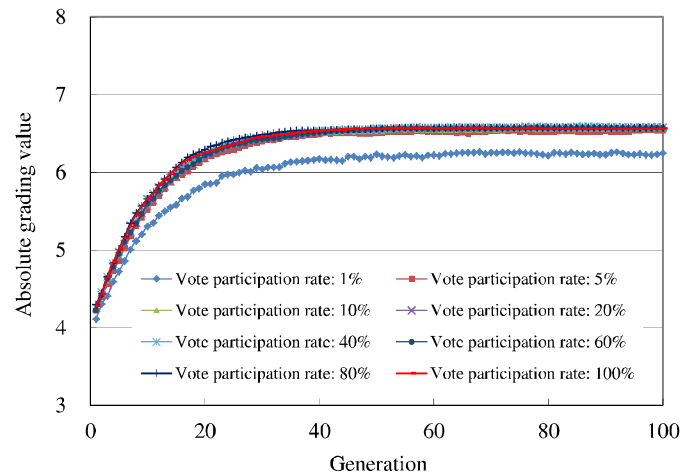
(a) 20bits



(c) 40bits



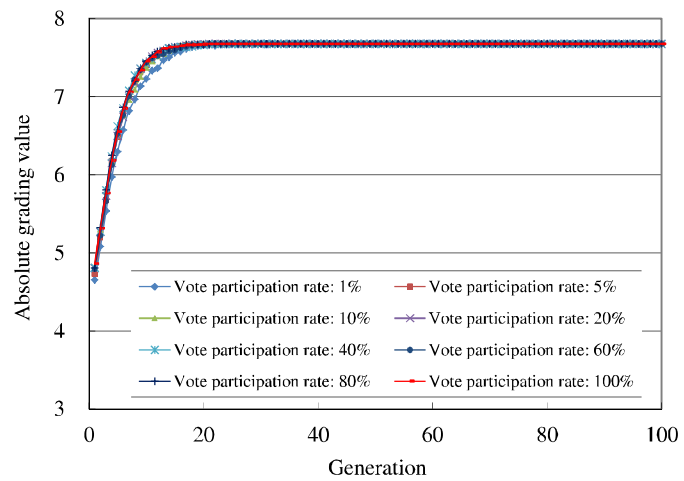
(b) 30bits



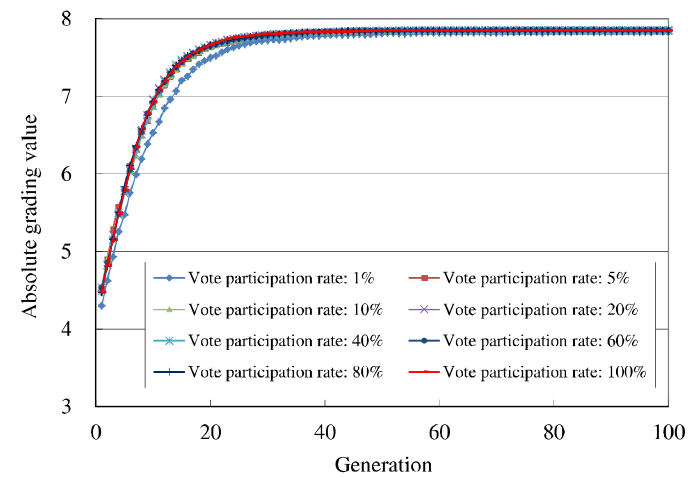
(d) 50bits

図 4.7: 投票参加割合の検討結果 (Pattern 2)

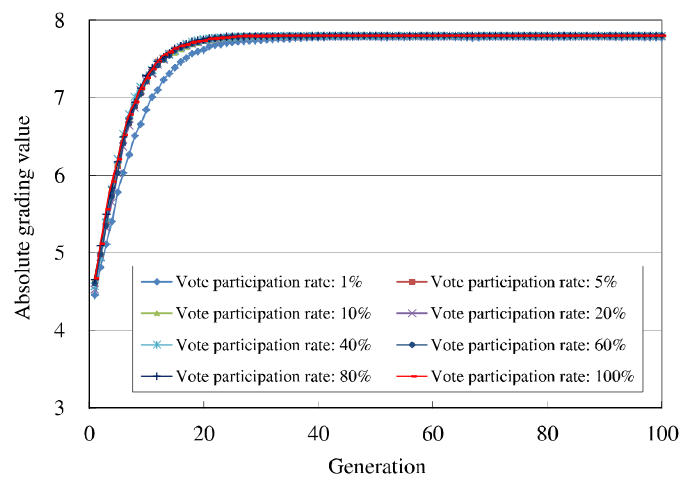
図 4.7: 投票参加割合の検討結果 (Pattern 2)



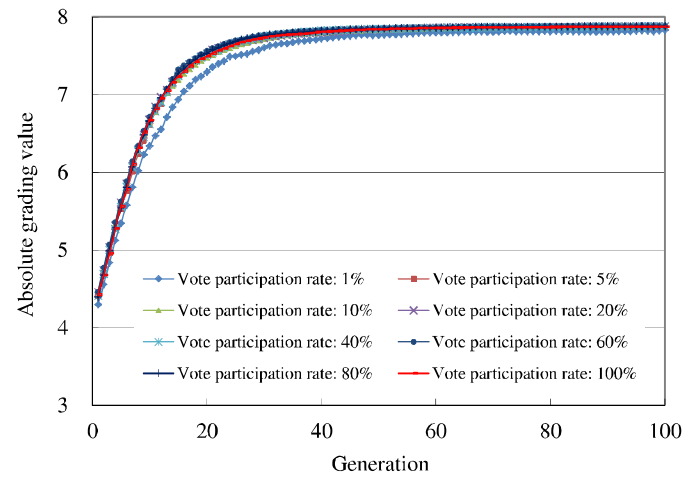
(a) 20bits



(c) 40bits



(b) 30bits



(d) 50bits

図 4.8: 投票参加割合の検討結果 (Pattern 3)

図 4.8: 投票参加割合の検討結果 (Pattern 3)

4.5.2 性能比較

本項では、GVMを比較手法とした性能比較の結果について考察する。本シミュレーションにおいては、評価エージェントの投票参加割合は100%とした。

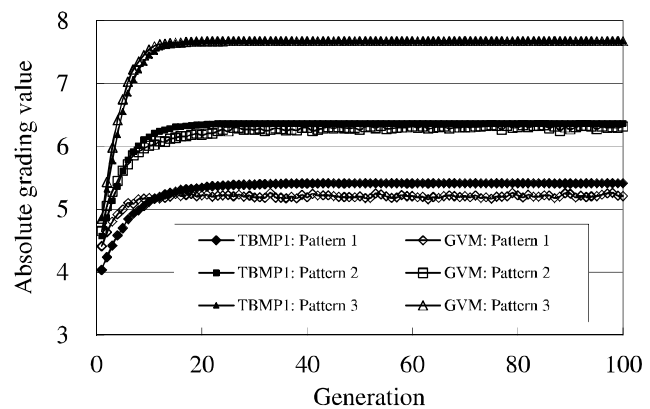
図4.9に性能比較の結果を示す。Pattern 1では、いずれの遺伝子長においても初期世代ではGVMがTBMP1より絶対評価値が高いが、20～30世代目以降はTBMP1の方が絶対評価値が高いことが確認された。Pattern 2では、いずれの遺伝子長においても初期世代ではTBMP1とGVMの絶対評価値は同等であるが、10～20世代目以降はTBMP1の方が絶対評価値が高いことが確認された。Pattern 3では、いずれの遺伝子長においても初期世代よりGVMがTBMP1より絶対評価値が高いが、大きな差は確認されなかった。

そこで、TBMP1及びGVMの絶対評価値に対してt検定を行い、統計的有意差を確認した。表4.3に10, 20, 30世代目におけるt検定の結果を示す。表4.3中の最右列は、統計的検定により進化性能が高いことが示された手法を示す。表4.3(a)より、Pattern 1では、いずれの遺伝子長においても10世代目では、TBMP1及びGVMの絶対評価値は同等であるか有意水準1%または5%でGVMの方が高いことが確認された。しかし、20, 30世代目においては、TBMP1及びGVMの絶対評価値は同等であるか有意水準1%でTBMP1の方が高いことが確認された。表4.3(b)より、Pattern 2では、いずれの遺伝子長においても10, 20, 30世代目では、TBMP1及びGVMの絶対評価値は同等であるか有意水準1%でTBMP1の方が高いことが確認された。表4.3(c)より、Pattern 3では、いずれの遺伝子長においても10, 20, 30世代目では、TBMP1及びGVMの絶対評価値は同等であるか有意水準1%または5%でGVMの方が高いことが確認された。これらの結果より、TBMP1は、Pattern 1, 2において全評価エージェントの折衷案を探索する場合には、GVMより有効であると考えられる。

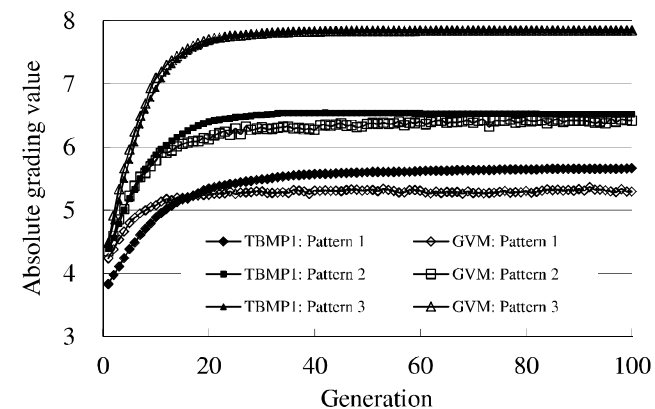
次に、評価エージェントの集団別の絶対評価値について考察する。図4.10～4.12に評価エージェントの集団性の違いによる集団別の絶対評価値の結果を示す。いずれの遺伝子長、集団性のパターンにおいても、TBMP1及びGVMで対応する集団は、同様の絶対評価値の変化を示していることが確認された。このため、各パターンにおける集団間の絶対評価値は、TBMP1及びGVMともに同等であると考えられる。

また、図4.13～4.24に評価エージェントの集団性の違いによる絶対評価値の評価エージェント数の分布を示す。いずれの遺伝子長、集団性のパターンにおいてもTBMP1及びGVMともに、絶対評価値が6以上の評価エージェント数は、世代交代を経る毎に増加していることが確認された。このため、TBMP1及びGVMともに多くの評価エージェントの絶対評価値が高いものを生成できると考えられる。しかし、TBMP1は、

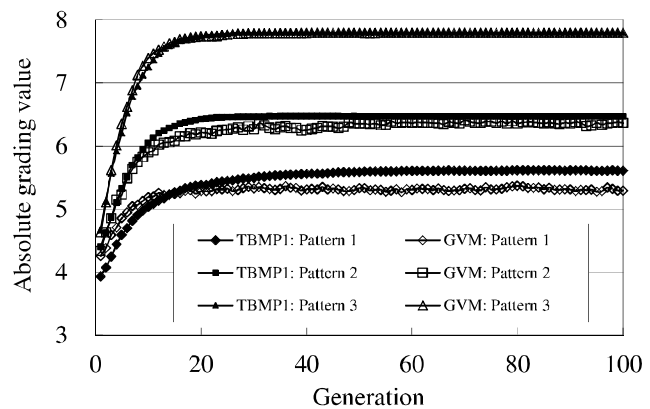
Pattern 1, 2のように各集団の人数に大きな偏りがない状況下で全評価エージェントの折衷案を探索するような場合には、GVMより高い進化性能を示した。また、Pattern 3においては、遺伝子長50bitsの20世代目を除いて、20, 30世代目ではTBMP1とGVMの進化性能は同等であった。したがって、Pattern 1, 2のような状況下において多くのユーザが満足のいくものを生成するという観点では、TBMP1は、GVMより有効であると考えられる。



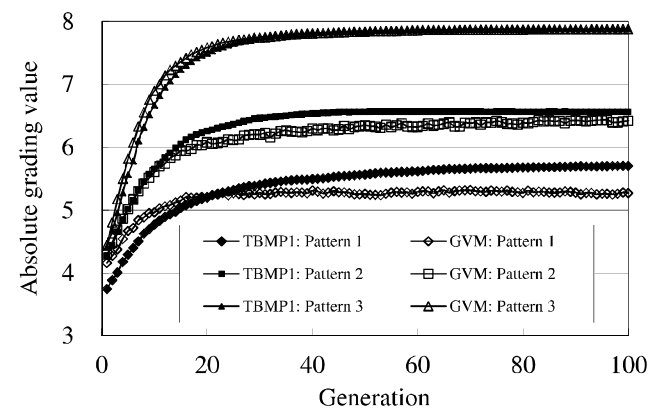
(a) 20bits



(c) 40bits



(b) 30bits



(d) 50bits

図 4.9: 性能比較の結果

図 4.9: 性能比較の結果

表 4.3: t 検定の結果

(a) Pattern 1

Generation	Gene length	TBMP1	GVM	Significant difference
10	20	5.12	5.18	n.s.
20	20	5.35	5.22	** TBMP1
30	20	5.40	5.19	** TBMP1
10	30	5.04	5.19	** GVM
20	30	5.38	5.28	** TBMP1
30	30	5.50	5.33	** TBMP1
10	40	4.89	5.08	** GVM
20	40	5.21	5.24	n.s.
30	40	5.49	5.30	** TBMP1
10	50	4.77	4.96	* GVM
20	50	5.20	5.20	n.s.
30	50	5.40	5.27	** TBMP1

n.s.: Not significant, *: $p < 0.05$, **: $p < 0.01$

(b) Pattern 2

Generation	Gene length	TBMP1	GVM	Significant difference
10	20	6.15	6.01	** TBMP1
20	20	6.35	6.18	** TBMP1
30	20	6.36	6.26	** TBMP1
10	30	6.04	5.90	** TBMP1
20	30	6.42	6.21	** TBMP1
30	30	6.46	6.27	** TBMP1
10	40	5.85	5.79	n.s.
20	40	6.39	6.12	** TBMP1
30	40	6.50	6.31	** TBMP1
10	50	5.66	5.60	n.s.
20	50	6.26	6.08	** TBMP1
30	50	6.46	6.20	** TBMP1

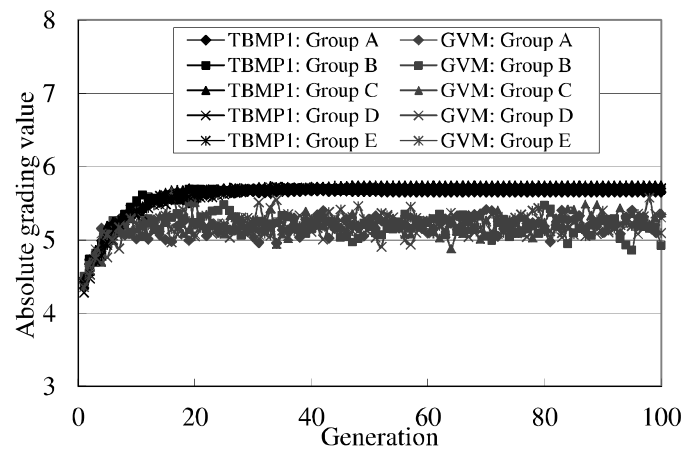
n.s.: Not significant, *: $p < 0.05$, **: $p < 0.01$

表 4.3: t 検定の結果

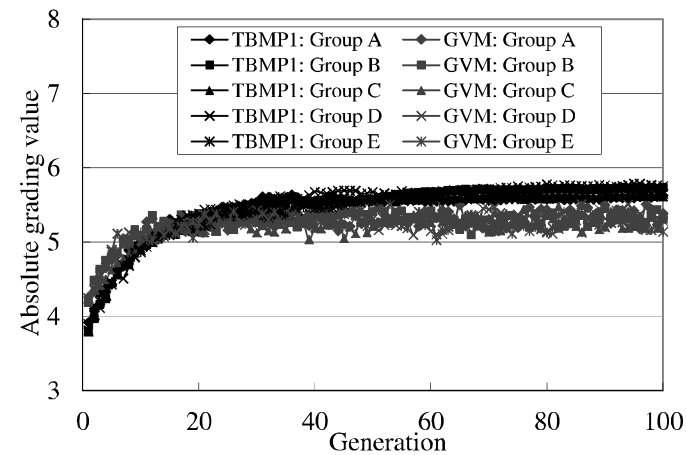
(c) Pattern 3

Generation	Gene length	TBMP1	GVM	Significant difference
10	20	7.46	7.54	* GVM
20	20	7.67	7.67	n.s.
30	20	7.68	7.68	n.s.
10	30	7.26	7.40	** GVM
20	30	7.74	7.74	n.s.
30	30	7.80	7.79	n.s.
10	40	6.93	7.10	** GVM
20	40	7.67	7.71	n.s.
30	40	7.80	7.80	n.s.
10	50	6.67	6.89	** GVM
20	50	7.50	7.59	** GVM
30	50	7.74	7.74	n.s.

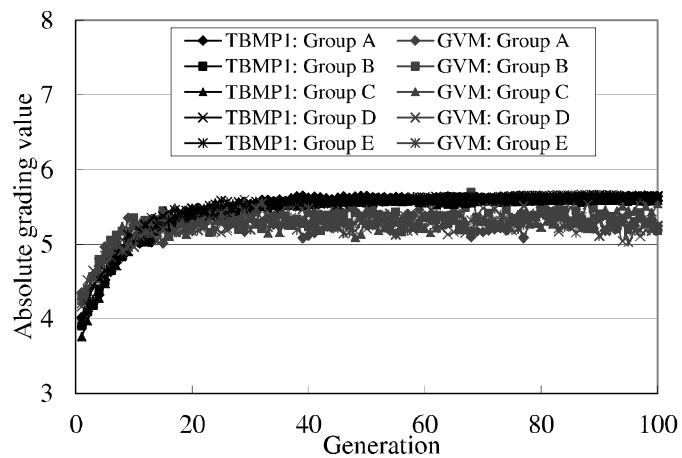
n.s.: Not significant, *: $p < 0.05$, **: $p < 0.01$



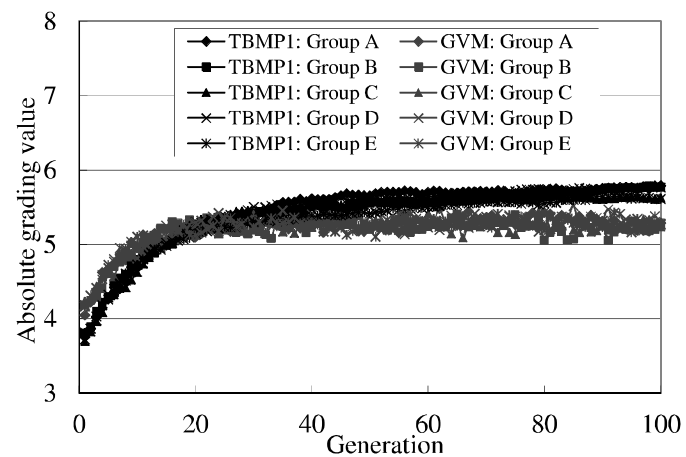
(a) 20bits



(c) 40bits



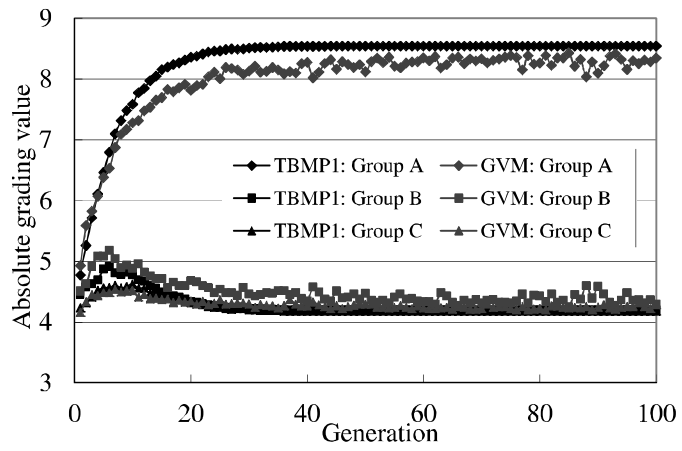
(b) 30bits



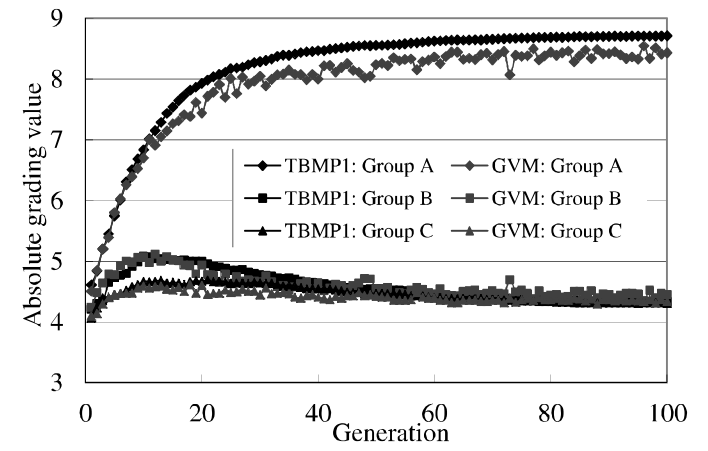
(d) 50bits

図 4.10: 集団別の絶対評価値の結果 (Pattern 1)

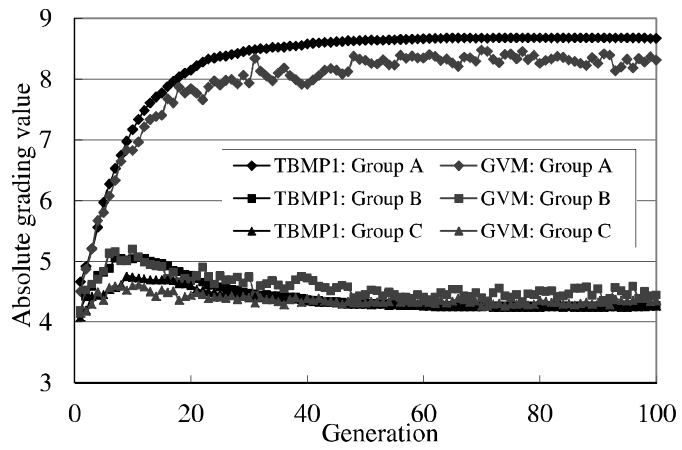
図 4.10: 集団別の絶対評価値の結果 (Pattern 1)



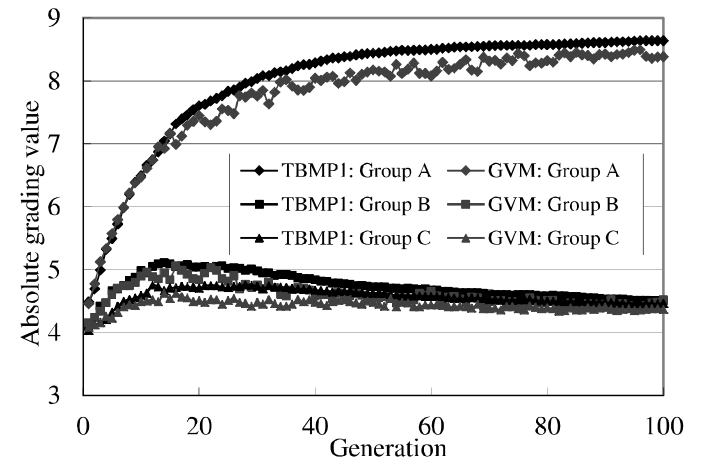
(a) 20bits



(c) 40bits



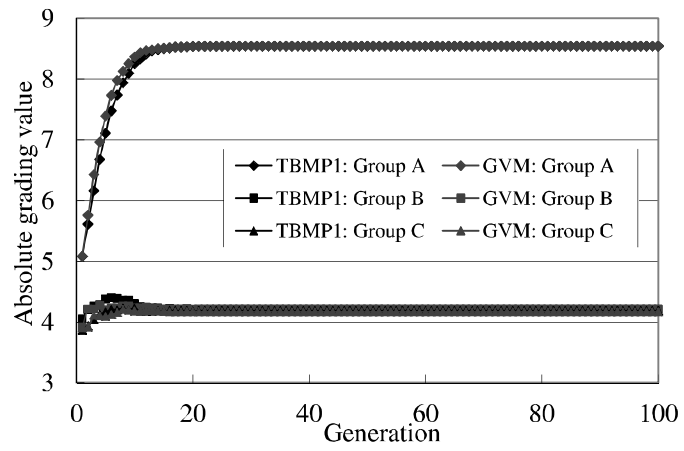
(b) 30bits



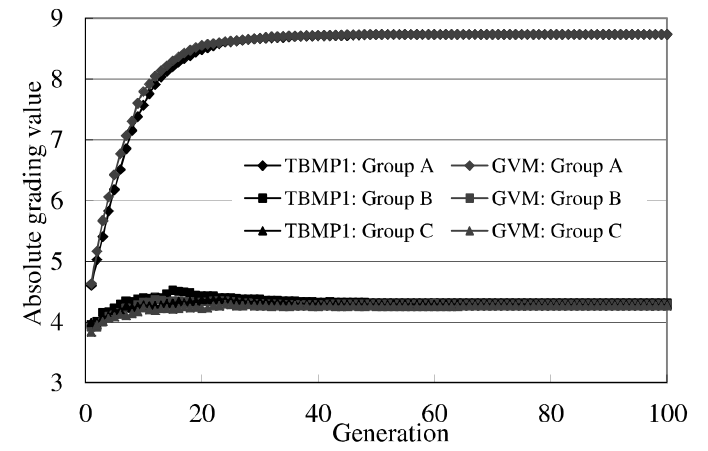
(d) 50bits

図 4.11: 集団別の絶対評価値の結果 (Pattern 2)

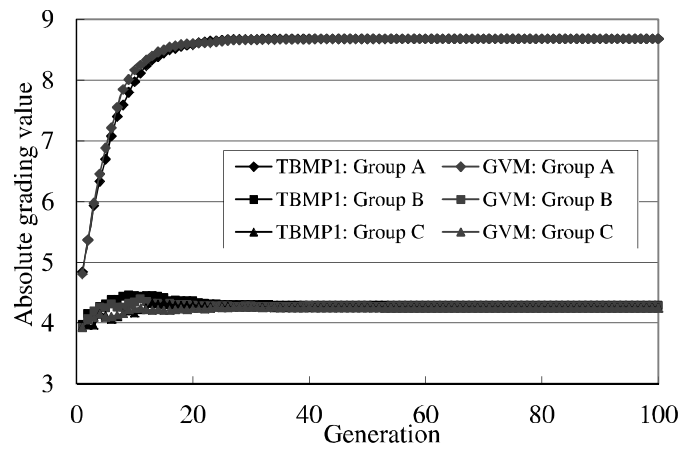
図 4.11: 集団別の絶対評価値の結果 (Pattern 2)



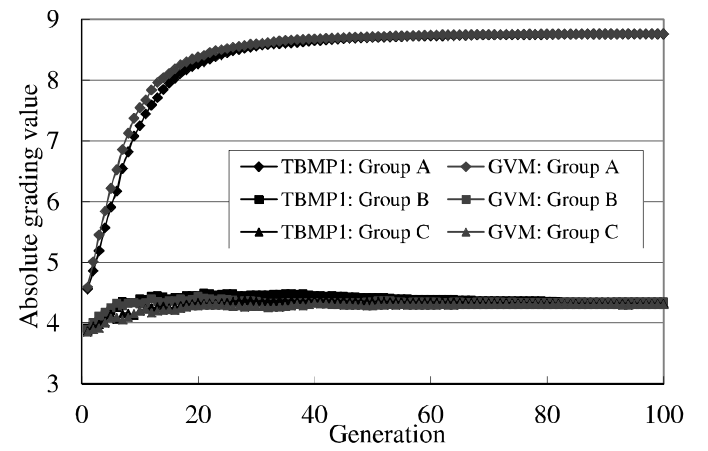
(a) 20bits



(c) 40bits



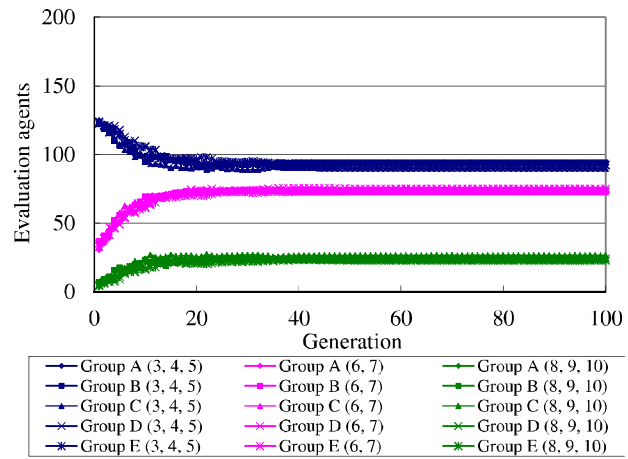
(b) 30bits



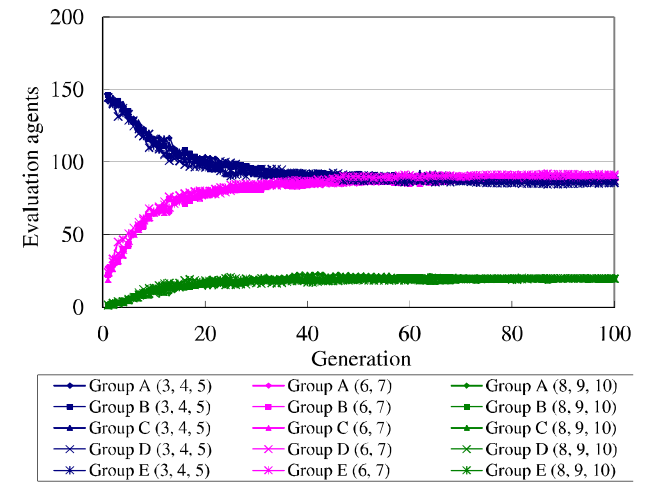
(d) 50bits

図 4.12: 集団別の絶対評価値の結果 (Pattern 3)

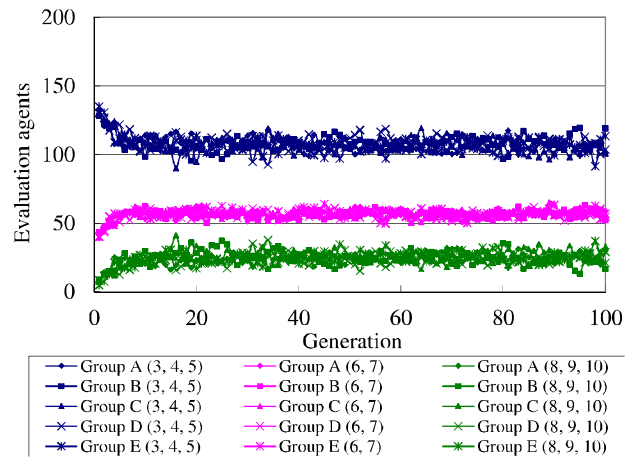
図 4.12: 集団別の絶対評価値の結果 (Pattern 3)



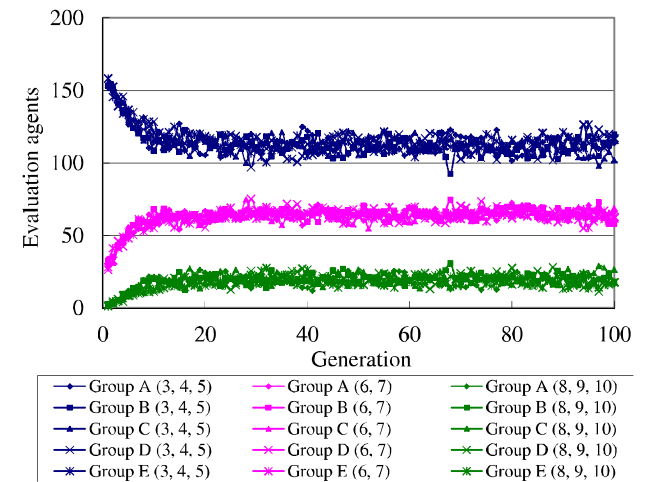
(a) TBMP1



(a) TBMP1



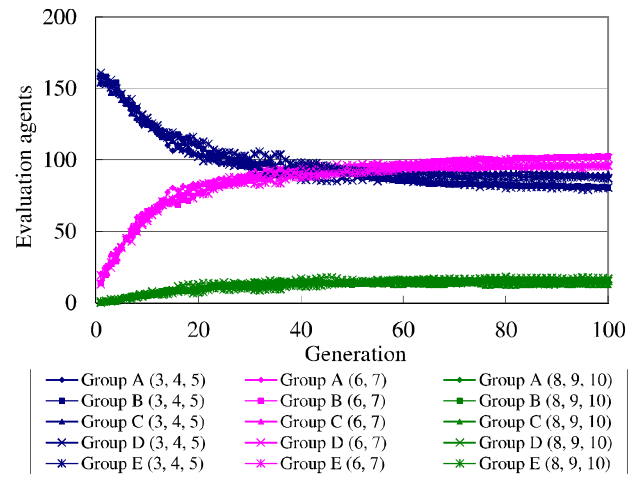
(b) GVM



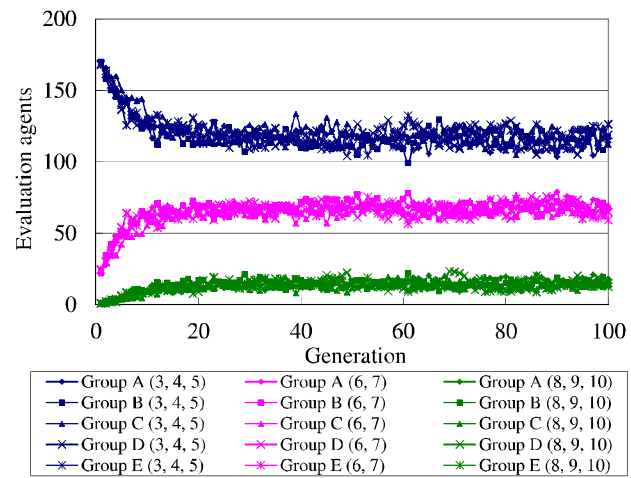
(b) GVM

図 4.13: 絶対評価値の評価エージェント数の分布 (20bits, Pattern 1)

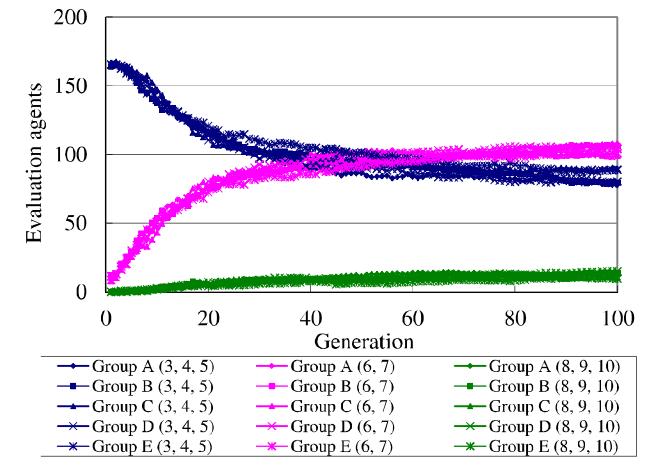
図 4.14: 絶対評価値の評価エージェント数の分布 (30bits, Pattern 1)



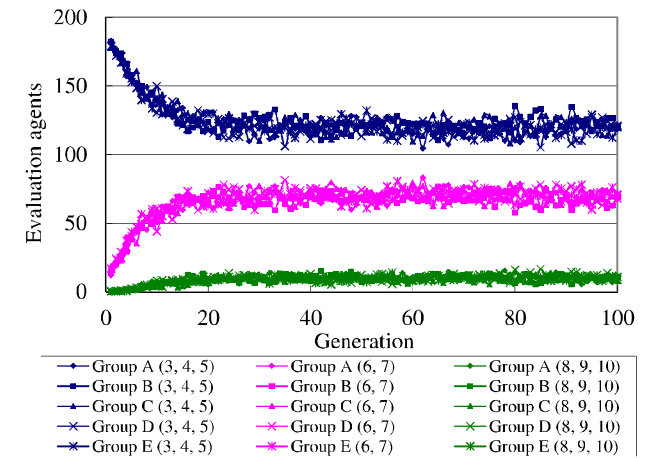
(a) TBMP1



(b) GVM



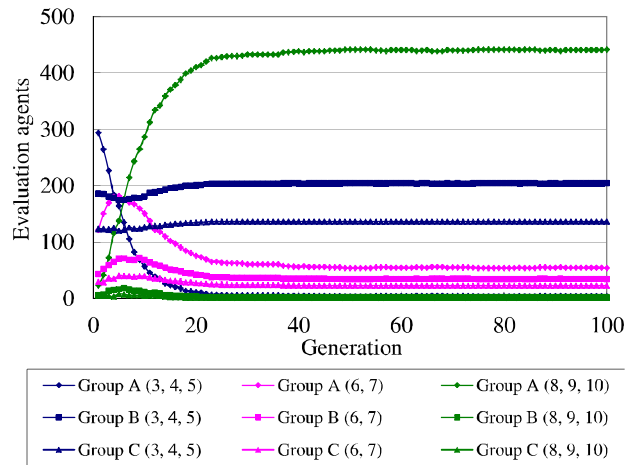
(a) TBMP1



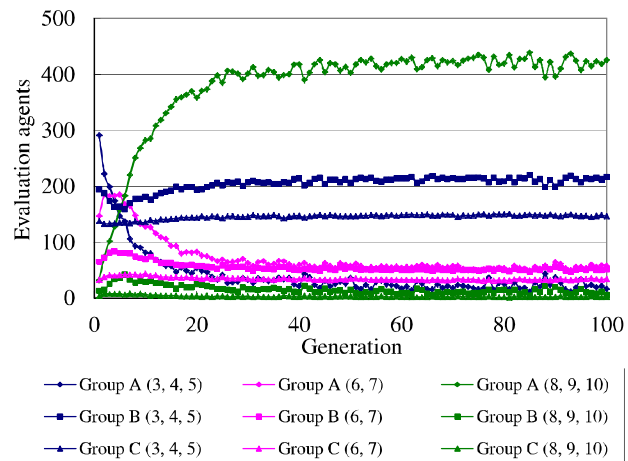
(b) GVM

図 4.15: 絶対評価値の評価エージェント数の分布 (40bits, Pattern 1)

図 4.16: 絶対評価値の評価エージェント数の分布 (50bits, Pattern 1)

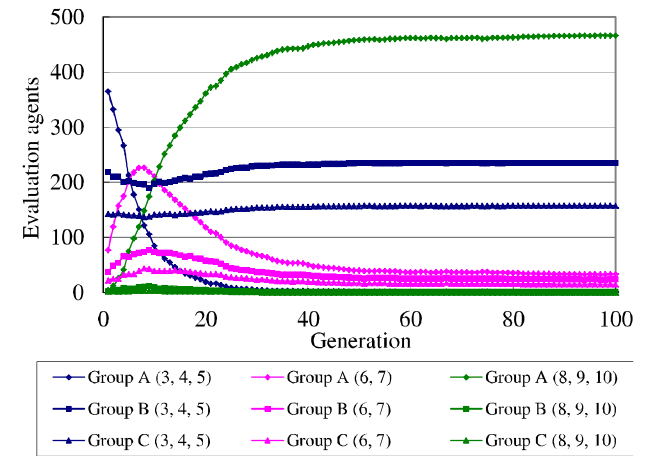


(a) TBMP1

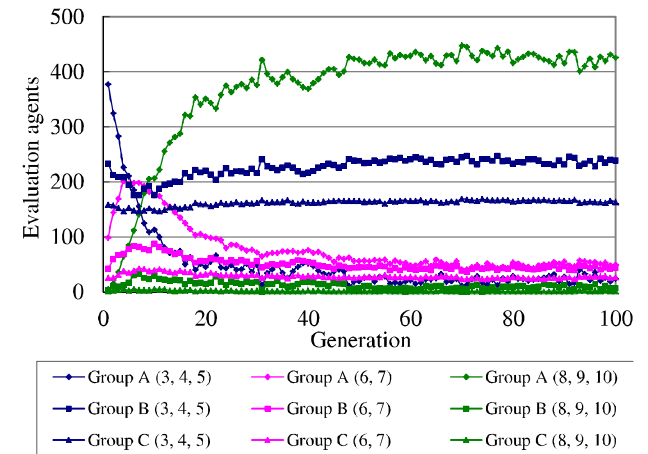


(b) GVM

図 4.17: 絶対評価値の評価エージェント数の分布 (20bits, Pattern 2)

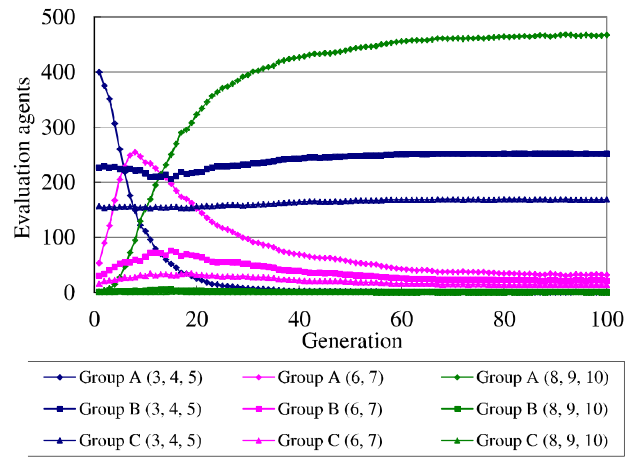


(a) TBMP1

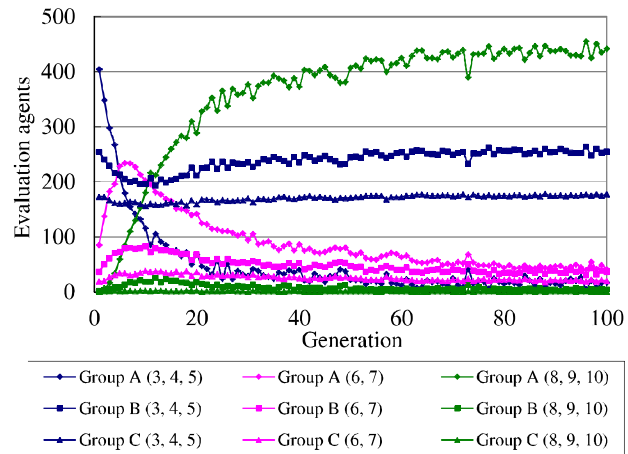


(b) GVM

図 4.18: 絶対評価値の評価エージェント数の分布 (30bits, Pattern 2)

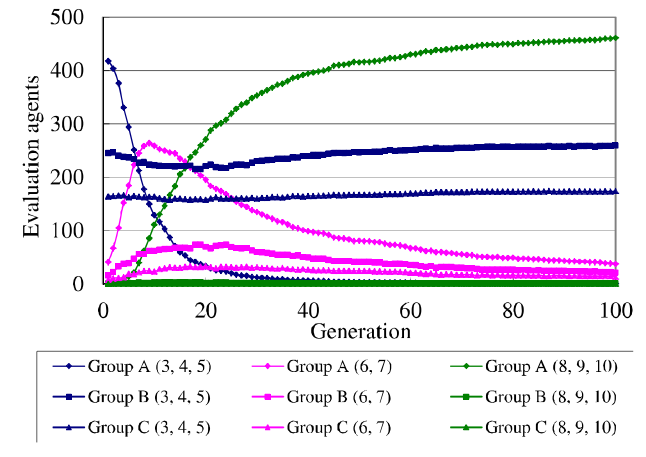


(a) TBMP1

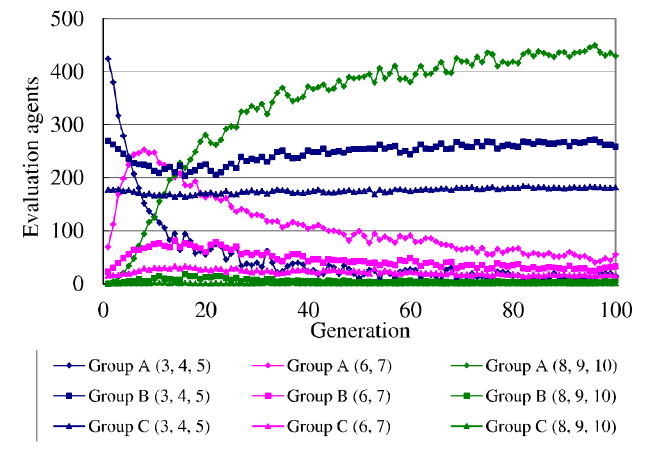


(b) GVM

図 4.19: 絶対評価値の評価エージェント数の分布 (40bits, Pattern 2)

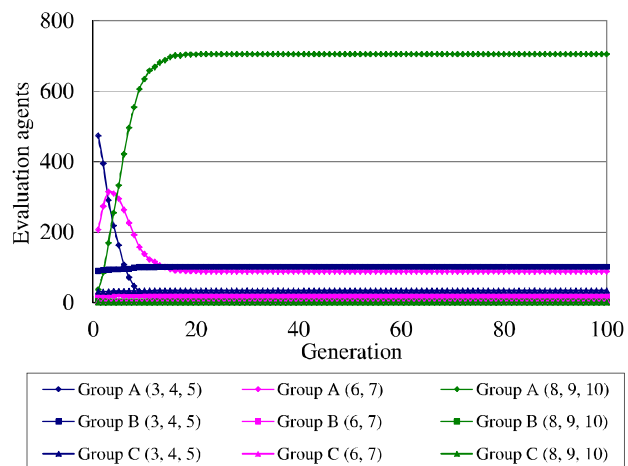


(a) TBMP1

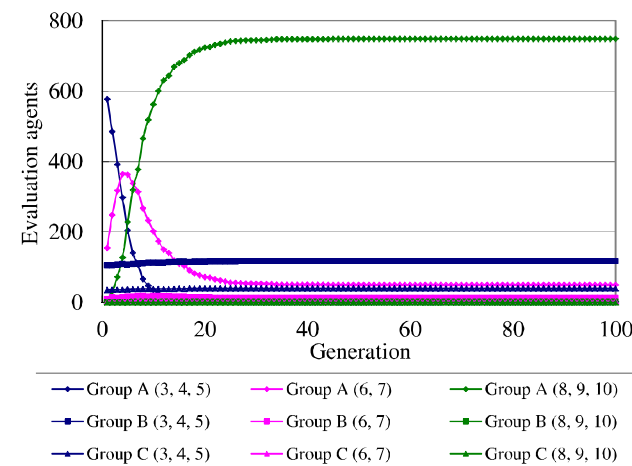


(b) GVM

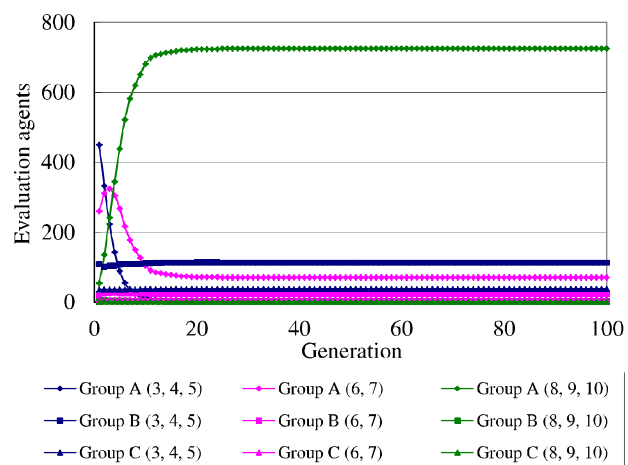
図 4.20: 絶対評価値の評価エージェント数の分布 (50bits, Pattern 2)



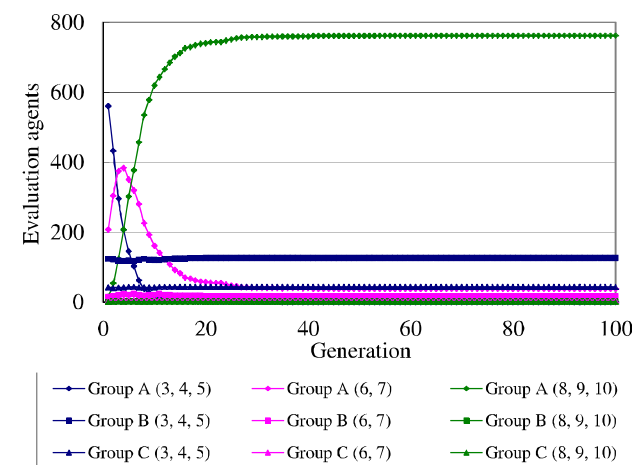
(a) TBMP1



(a) TBMP1



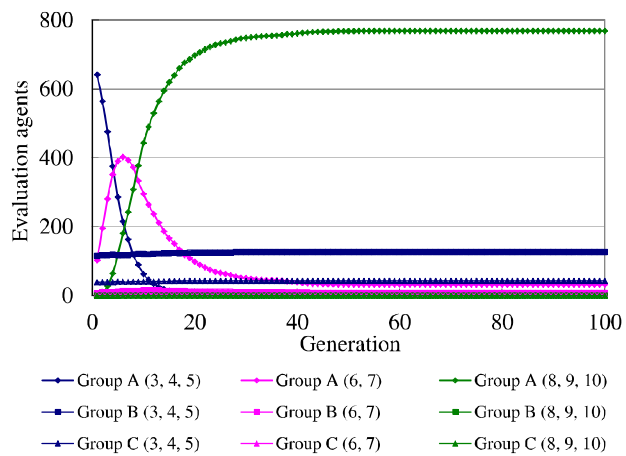
(b) GVM



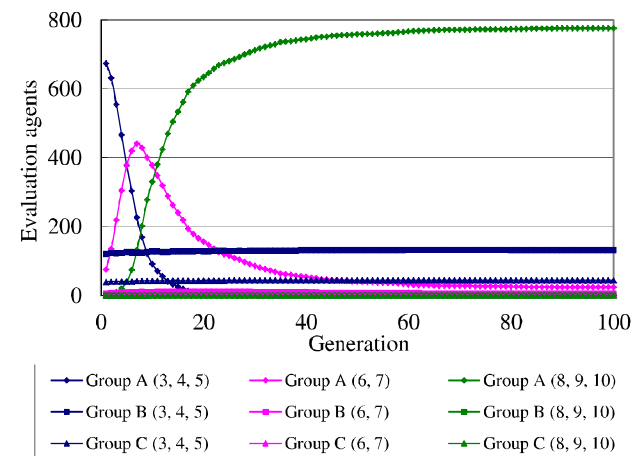
(b) GVM

図 4.21: 絶対評価値の評価エージェント数の分布 (20bits, Pattern 3)

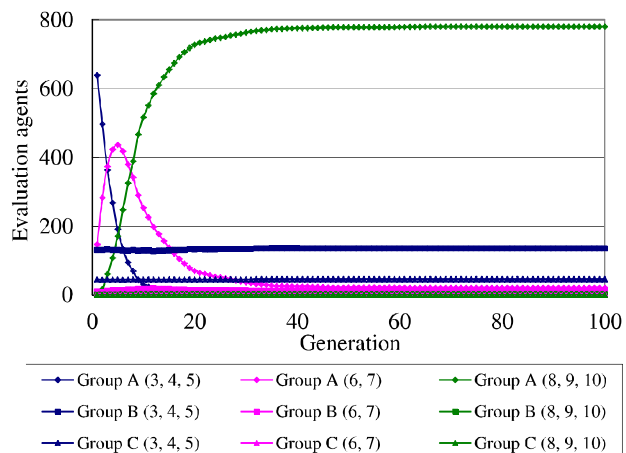
図 4.22: 絶対評価値の評価エージェント数の分布 (30bits, Pattern 3)



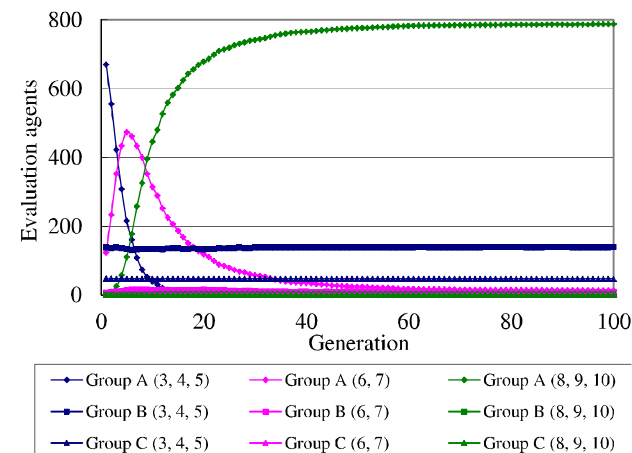
(a) TBMP1



(a) TBMP1



(b) GVM



(b) GVM

図 4.23: 絶対評価値の評価エージェント数の分布 (40bits, Pattern 3)

図 4.24: 絶対評価値の評価エージェント数の分布 (50bits, Pattern 3)

4.6 結言

本章では、まず、多くのユーザの感性を用いたIECに関する先行研究を概括した。次に、TBMP1のシステム概要及び数値シミュレーションによる解候補の進化性能に関する有効性の検証について述べた。数値シミュレーションによる検証結果は以下のとおりとなった。

(1) 評価エージェントの投票参加割合に関する検討

いずれの遺伝子長、集団性のパターンにおいても、評価エージェントの投票参加割合が60%以上であれば、投票参加割合が100%のときと同等の進化性能が得られることが確認された。そのため、TBMP1においては、多くのユーザが時間的に束縛されず、部分的に解評価に参加しても多くのユーザが満足のいくものを生成できるシステムが実現可能であると考えられる。

(2) 性能比較

いずれの遺伝子長においても、Pattern 1, 2のように各集団の人数に大きな偏りが無い状況下で全評価エージェントの折衷案を探索するような場合には、TBMP1はGVMより高い進化性能を示した。したがって、Pattern 1, 2のような状況下において多くのユーザが満足のいくものを生成するという観点では、TBMP1はGVMより有効であると考えられる。

本章では、数値シミュレーションにより、TBMP1の解候補の進化性能に関する有効性が確認された。しかし、TBMP1においては、(1) 想定した投票数を獲得できず、トーナメント対戦を進行できない、(2) 勝ち上がった解候補は複数回評価されなければならない、ユーザの解評価に対するモチベーションが低下してしまうという問題が想定される。本論文では、これらの問題を解決するため、TBMP1の改良手法であるTBMP2及びTBMP3の有効性を数値シミュレーションにおいて検証する。さらに、実ユーザを対象とした評価実験を行い、TBMP1の実ユーザに対する有効性を検証する。

第5章 2点嗜好法の適用

5.1 緒言

本章では、2点嗜好法を適用した複数参加型トーナメント方式であるTBMP2について述べる [23]。

一般に、Web上の投票システムなどで、投票に参加するユーザ数を推測することは困難である。そのため、第4章で述べたTBMP1においては、トーナメント対戦の勝敗判定に必要な投票数を獲得できるまでに要する時間が不明である。また、一定の投票数を獲得できず、トーナメントの勝敗を判定できないなどの問題も想定される。そこで、これらの問題を解決するために、トーナメント対戦の勝敗決定に2点嗜好法を適用する。2点嗜好法は、嗜好など試料間に客観的な順位が存在しない場合に、2種の試料の優劣を統計的に決定する手法である [56]。

本章では、TBMP2の概要及びトーナメント対戦の効率化に関する有効性を検証するために行う数値シミュレーションについて述べる。本シミュレーションでは、2点嗜好法により対戦解候補の獲得投票数に有意差があると判定された対戦数、2点嗜好法の判定誤り率、解評価に必要な投票数の検討及びTBMP1を比較手法とした性能比較を行う。

5.2 2点嗜好法を適用した複数参加型トーナメント方式

本節では、TBMP1の問題点について述べ、2点嗜好法を適用する目的及びシステム概要について述べる。

5.2.1 TBMP1の問題点

一般に、Web上の投票システムなどで、投票に参加するユーザ数を推測することは困難である。そのため、TBMP1においては、トーナメント対戦の勝敗判定に必要な投票数を獲得できるまでに要する時間が不明である。また、一定の投票数を獲得できず、トーナメント対戦の勝敗を判定できないという問題も想定される。このような問題を解決するためには、投票数の効率的な利用、すなわちトーナメント対戦の効率化が求められる。

トーナメント対戦の効率化のためには、(1)投票を獲得できなくとも、トーナメント対戦が停滞しない、(2)一定の投票数を獲得不可能であっても少数の投票において解候補の優劣を決定できる、の2点が求められる。(1)、(2)のような条件を満たすためには、予め定めた投票数を獲得しなければトーナメント対戦を進行させないという処理は行わず、少数の投票であっても統計的手法などにより解候補の優劣を判定できる仕組みが求められる。

そこで、TBMP1においてトーナメント対戦の効率化を実現させるため、2点嗜好法の適用を試みる。2点嗜好法は、嗜好など試料間に客観的な順位が存在しない場合において、一対比較によって2つの試料の優劣を統計的に決定する官能評価手法である[56]。2点嗜好法の適用により、各対戦において投票開始からより早い段階に解候補の優劣を決定することができ、解評価に必要な投票数の減少、すなわちトーナメント対戦の効率化を実現できると考えられる。

5.2.2 2点嗜好法を適用した複数参加型トーナメント方式の概要

本項では、TBMP2の概要について述べる[23]。

図5.1にTBMP2のシステム概要を示す。まず、初期遺伝子集団を生成し、生成された解候補をトーナメント表に配置する。次に、多くのユーザにWeb上で各対戦の2つのデザインを提示し、好みのデザインの方に投票してもらう。決勝戦まで終了すれば、解候補を評価し、GA処理を行う。TBMP2では、2点嗜好法(Paired preference test)を用いて解候補の優劣の差を判断できる投票数が獲得された時点より、提示した2つの解候補の優劣判定を試みる。2点嗜好法の適用により対戦解候補の投票数に統計的有

意差があると判断されれば、対戦解候補の優劣を決定し、次の対戦に移行する。統計的有意差があるとはいえないと判断されれば、さらにユーザの投票を求める。しかし、投票を重ねた後も2点嗜好法により優劣が判定されない場合においては、一定の期間が過ぎれば投票を打ち切り、その時点での投票数により解候補の優劣を決定する。

また、TBMP2においては、対戦によって投票数が異なる可能性がある。これは、統計的に2つ解候補の優劣に差があると判定しているためである。そのため、投票数そのものには意味がなく、獲得投票数を利用して解候補を評価することができない。したがって、TBMP2においては、各解候補が何回戦まで勝ち進んだかという情報のみを利用して、表5.1のように評価値を与える。

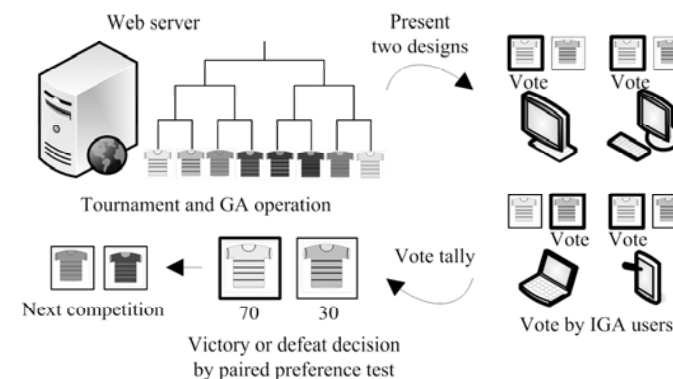


図 5.1: TBMP2 のシステム概要

表 5.1: 評価値の内訳

Winner of the tournament	10
Loser of the final round	8
Loser of the semi-final	6
Loser of the second round	4
Loser of the first round	2

5.2.3 2点嗜好法

2点嗜好法は、嗜好など試料間に客観的な順位が存在しない場合において、2種の試料を組にして、被験者にどちらの試料が質問に該当するか回答してもらい、2種の試料の優劣を統計的に決定する手法である [56]。2点嗜好法においては、 n 回の判定を行ったとき、片方の試料が選ばれる回数 k は $p = \frac{1}{2}$ の二項分布に従うため、 $p = \frac{1}{2}$ の二項検定を行う。2点嗜好法では、帰無仮説 H_0 と対立仮説 H_1 は式 (5.1)、(5.2) で表現される。

$$H_0 : p = \frac{1}{2} \quad (5.1)$$

$$H_1 : p \neq \frac{1}{2} \quad (5.2)$$

帰無仮説の下で、2つの試料のうちいずれかが n 回のうち k 回 ($k \geq \frac{n}{2}$) 以上選ばれる確率は、 $n \leq 100$ のとき、式 (5.3) で表現される。

$$P = \sum_{x=k}^n {}_n C_x \left(\frac{1}{2}\right)^n \times 2 \quad (5.3)$$

実測値から求められる P が有意水準以下であれば、帰無仮説を棄却し対立仮説を採択する。すなわち、2つの試料間に優劣の差があると判断する。2点嗜好法を適用して行われた優劣判定は、その後、多くの投票を重ねていったとしても、有意水準5%の場合は95%の確率で同様の結果となる。

また、 $n > 100$ のときは、式 (5.4) に示す正規分布近似による検定統計量 u_0 を用いる。

$$u_0 = \frac{k - \frac{1}{2}n - 0.5}{\sqrt{n \times \frac{1}{2} \times \frac{1}{2}}} \quad (5.4)$$

式 (5.4) において、分子の0.5は、連続性に対する修正である。 $u_0 \geq u_{\alpha/2}$ ならば、帰無仮説を棄却し、対立仮説を採択する。ただし、 $u_{\alpha/2}$ は、標準正規分布の両側 $100\alpha\%$ [上側 (片側) $100\alpha/2\%$] であり、 $\alpha = 0.05$ のとき、 $u_{0.025} \cong 1.96$ である。2点嗜好法は統計的検定であり、有意なときは2つの試料間に優劣の差があると断定できるのに対して、有意でないときは2つの試料間に優劣の差があるとはいえないという結論になる。したがって、2つの試料間に優劣の差がないという断定はできない。

TBMP2では、有意水準5%で2点嗜好法を適用する。2点嗜好法の適用により、投票開始からより早い段階で解候補の優劣を決定でき、トーナメント対戦の効率化を図ることができると考えられる。

5.3 数値シミュレーションによる性能評価

本節では、TBMP2のトーナメント対戦の効率化に関する有効性を検証するために行う数値シミュレーションの概要について述べる。本シミュレーションでは、2点嗜好法により対戦解候補の獲得投票数に有意差があると判定された対戦数、2点嗜好法の判定誤り率、解評価に必要な投票数の検討及びTBMP1を比較手法とした性能比較を行う。

図5.2にTBMP2のシミュレーションの流れを示す。まず、初期遺伝子集団を生成し、生成された解候補をトーナメント表に配置する。次に、ランダムに選択された評価エージェントが投票を行い、2点嗜好法を適用可能な投票数を獲得できた時点より、2点嗜好法による優劣判定を試みる。2点嗜好法により、対戦解候補の獲得投票数に統計的有意差があると判断されれば、対戦解候補の優劣を判定し、次の対戦に移行する。統計的有意差があるとはいえないと判断されれば、新たにランダムに選択された評価エージェントが投票を行う。そして、評価エージェントが投票を行う度に2点嗜好法による優劣判定を試みる。全評価エージェントが投票に参加しても2点嗜好法により有意差があるとはいえないと判断された場合は、全評価エージェントの投票結果によって対戦解候補の優劣を決定する。決勝戦まで終了すれば、解候補を評価し、GA処理を行う。そして、新しい解候補を提示し、再び評価エージェントによる投票を行う。

本シミュレーションにおけるパラメータは表4.1、評価エージェントの集団性は表4.2と同様である。また、評価エージェントの仕様は、第4章で用いたものと同様である。比較手法として用いるTBMP1の評価エージェントの投票参加割合は100%とした。確率的な影響を排除するため、シミュレーション試行回数は100回とした。なお、5.4節におけるシミュレーション結果は、シミュレーション試行回数分の平均値である。

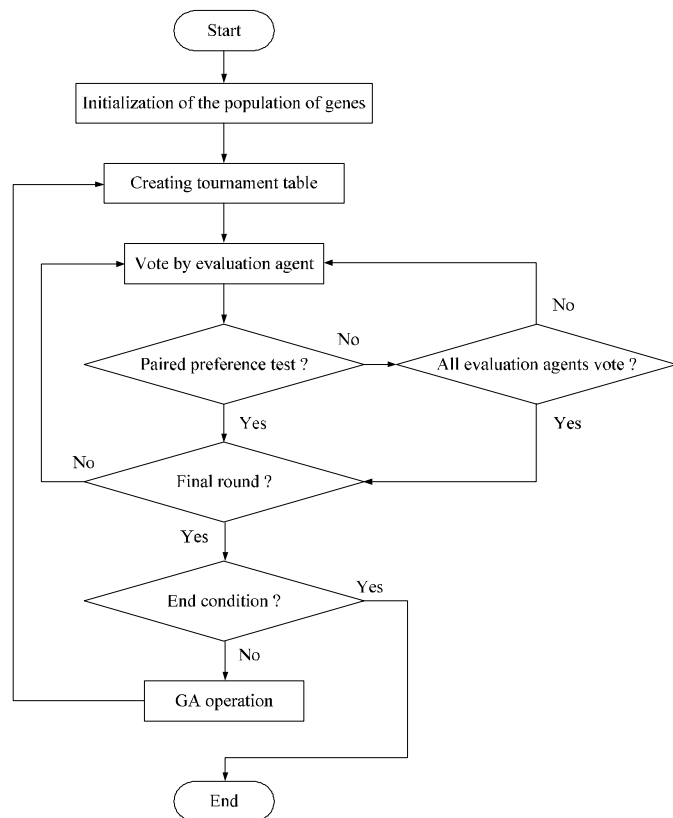


図 5.2: TBMP2 のシミュレーションの流れ

5.4 シミュレーション結果

本節では、数値シミュレーションにおける2点嗜好法により対戦解候補の獲得投票数に有意差があると判定された対戦数、2点嗜好法の判定誤り率、解評価に必要な投票数の検討及びTBMP1を比較手法とした性能比較の結果について考察する。

5.4.1 2点嗜好法による優劣判定対戦数に関する検討

本項では、2点嗜好法により対戦解候補の獲得投票数に有意差があると判定された対戦数の結果について考察する。

図 5.3 に評価エージェントの集団性の違いによる2点嗜好法により対戦解候補の獲得投票数に有意差があると判定された対戦数 (The number of judgments) の検討結果を示す。いずれの遺伝子長においても、1世代あたり約90%の対戦において、2点嗜好法により対戦解候補の獲得投票数に有意差があると判定されていることが確認された。また、集団の人数に偏りを持たせた Pattern 2, 3 では、集団の人数が均等な Pattern 1 より、2点嗜好法により有意差があると判定された対戦数が多いことが確認された。これは、Pattern 2, 3 では、解評価に大きな影響を与える多数派集団 (Group A) の投票行動により、投票の初期段階で2つの解候補の投票数に差が生じやすくなったためであると考えられる。

5.4.2 2点嗜好法の判定誤り率に関する検討

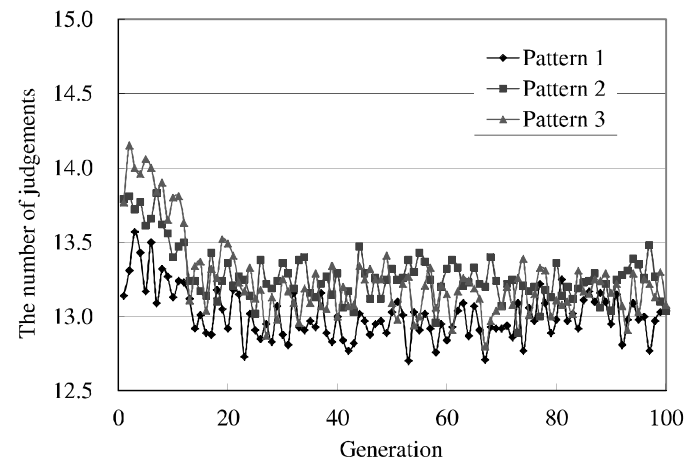
2点嗜好法においては、統計的に解候補の優劣を判定しているため、2点嗜好法による優劣判定と全評価エージェントの投票結果による優劣判定は異なる場合があると考えられる。そのため、2点嗜好法により有意差があると判定された対戦のうち、どの程度の対戦数に誤りが生じているかについて考察する。

図 5.4 に評価エージェントの集団性の違いによる2点嗜好法の判定誤り率 (Judgment error rate) の検討結果を示す。判定誤り率とは、2点嗜好法により有意差があると判定された対戦のうち、2点嗜好法における優劣判定と全評価エージェントが投票を行った場合の優劣判定が異なる割合である。いずれの遺伝子長においても、判定誤り率は約5%となることが確認された。1世代あたりの判定誤り対戦数は、1回に達することはなく、2点嗜好法により優劣が判定された対戦数のうち、約95%が正しいという結果となった。しかし、本シミュレーションにおいては、2点嗜好法を有意水準5%で適用している。そのため、判定誤り率が約5%であるという結果は、妥当であると考えられる。また、集団性の偏りが表現されていない Pattern 1 に関しては、他のパターンと比

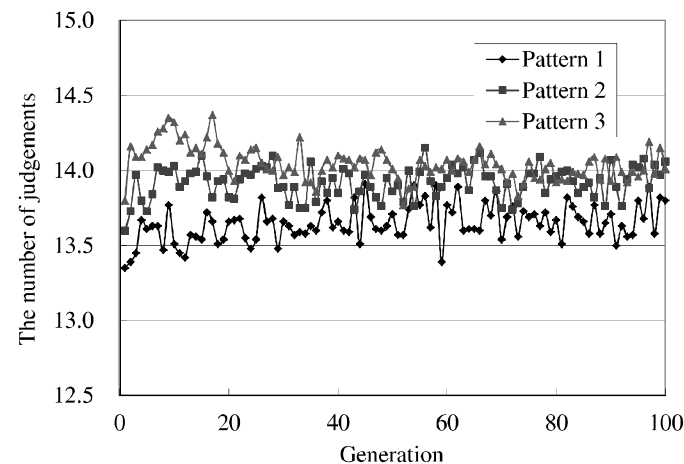
較して判定誤り率が高いことが確認された。Pattern 1においては、Pattern 3のように解候補の優劣判定に大きな影響を与える多数派集団が存在しない。そのため、2点嗜好法による優劣判定の後に投票を重ねた場合、全評価エージェントが投票した場合の判定結果と異なる割合がPattern 2, 3と比較して、高くなったと考えられる。

さらに、本シミュレーションでは、判定誤りを意図的に発生させた場合の検討を行った。図 5.4 より、本シミュレーションにおける2点嗜好法の判定誤り率は約5%であった。しかし、実際のシステムでは、どの程度の判定誤りが発生するかは不明である。そのため、意図的に判定誤りを発生させた場合に、TBMP2においてはどの程度の進化性能が得られるかについて検証した。判定誤りを意図的に発生させる対戦の条件は、2点嗜好法により有意差があると判定され、かつ全評価エージェントの投票による優劣判定結果が一致したときとし、予め定めた確率で判定誤りを実行する。

図 5.5~5.8 に判定誤り率を変化させた場合の検討結果を示す。図 5.5~5.8 の判定誤り率0%とは、TBMP1 において表 5.1 により解候補を評価した場合を示す。これは、TBMP2 において、全トーナメント対戦の結果が正しい場合と同等の結果となる。また、判定誤り操作なし (*Judgment error rate: 5%) とは、2点嗜好法の優劣判定結果を意図的に操作しない場合を示す。すなわち、判定誤り率は約5%となる。各世代における絶対評価値 (Absolute grading value) は、各世代のエリート解候補に対する全評価エージェントの絶対評価値のシミュレーション試行回数分の平均値である。図 5.5~5.8 より、判定誤り率0%と判定誤り操作なしのときの絶対評価値は、同等であることが確認された。また、判定誤り率が10%より高くなると、絶対評価値の減少は顕著であることが確認された。本シミュレーションにおける絶対評価値の減少量と、実ユーザの満足度の間の関連性は示されない。しかし、判定誤り率が増加すると絶対評価値の減少は顕著である。そのため、判定誤り率の増加は、実際のシステムにおいてもユーザの満足度に影響を与えてしまう可能性があると考えられる。

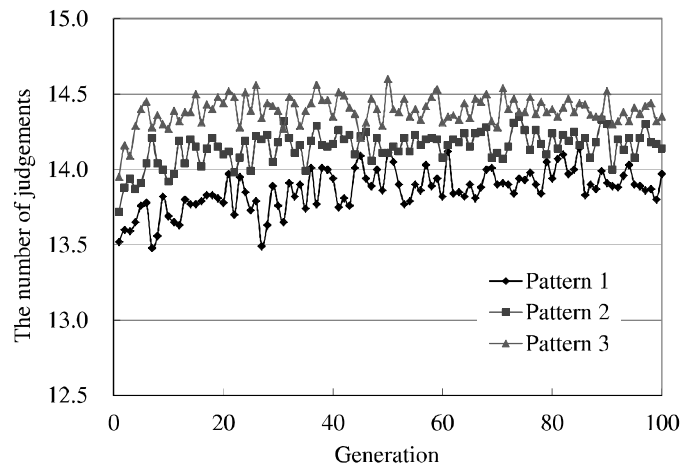


(a) 20bits

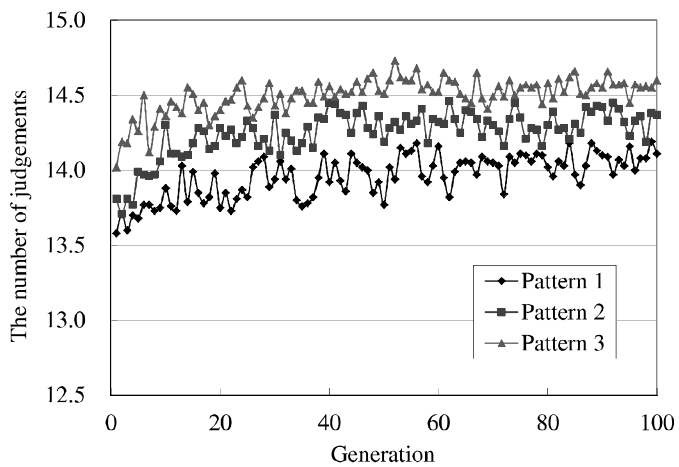


(b) 30bits

図 5.3: 2点嗜好法による優劣判定対戦数の検討結果

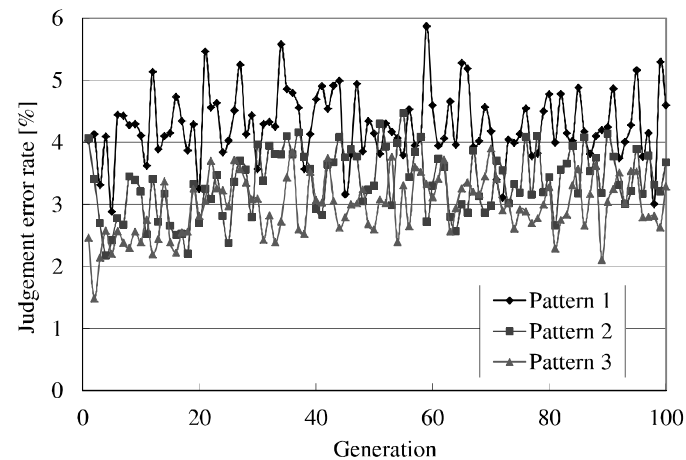


(c) 40bits

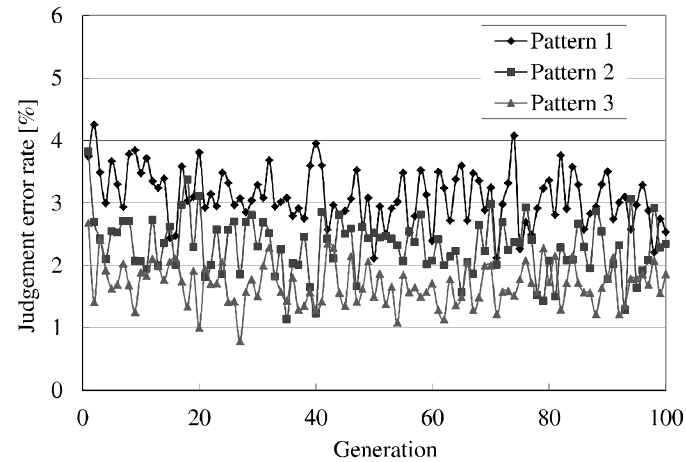


(d) 50bits

図 5.3: 2点嗜好法による優劣判定対戦数の検討結果

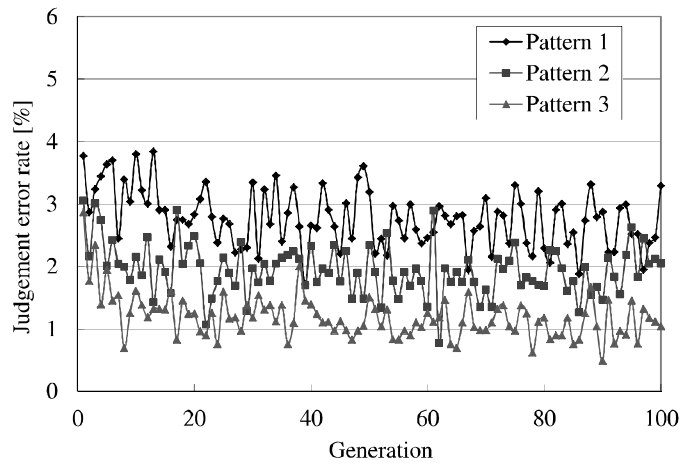


(a) 20bits

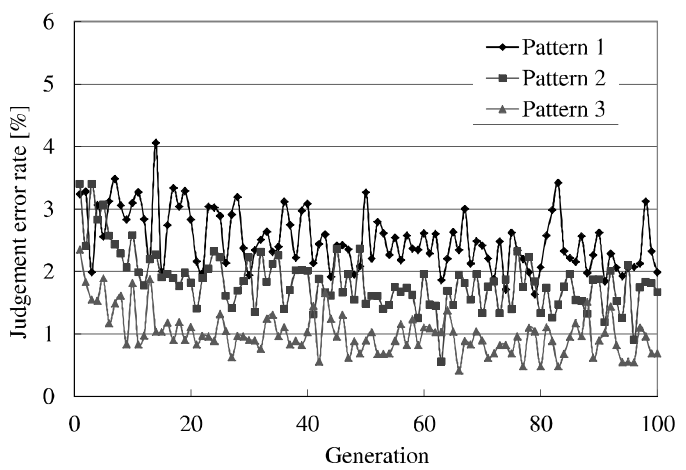


(b) 30bits

図 5.4: 2点嗜好法の判定誤り率の検討結果

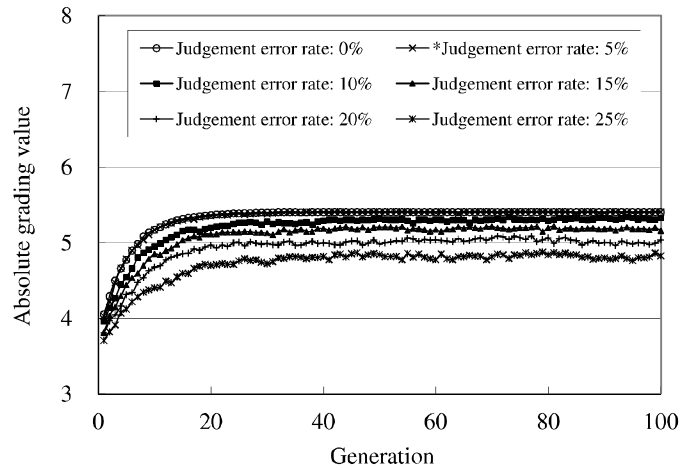


(c) 40bits

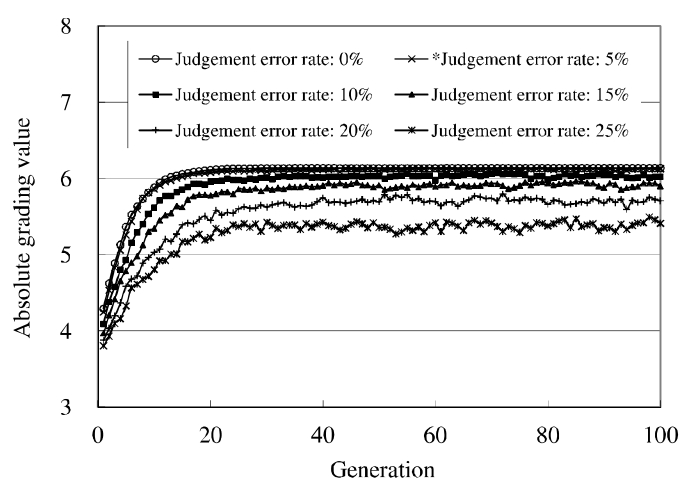


(d) 50bits

図 5.4: 2点嗜好法の判定誤り率の検討結果

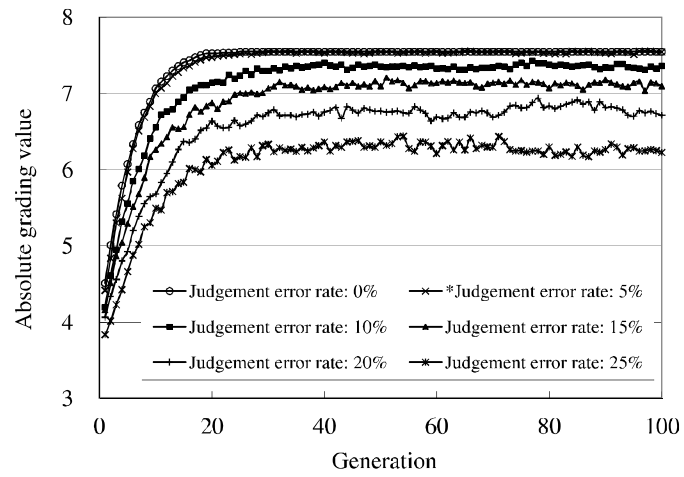


(a) Pattern 1

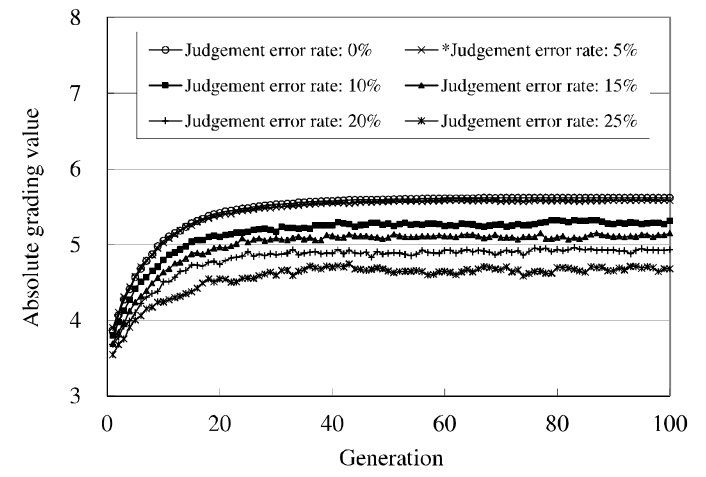


(b) Pattern 2

図 5.5: 判定誤り率を操作した場合の検討結果 (20bits)

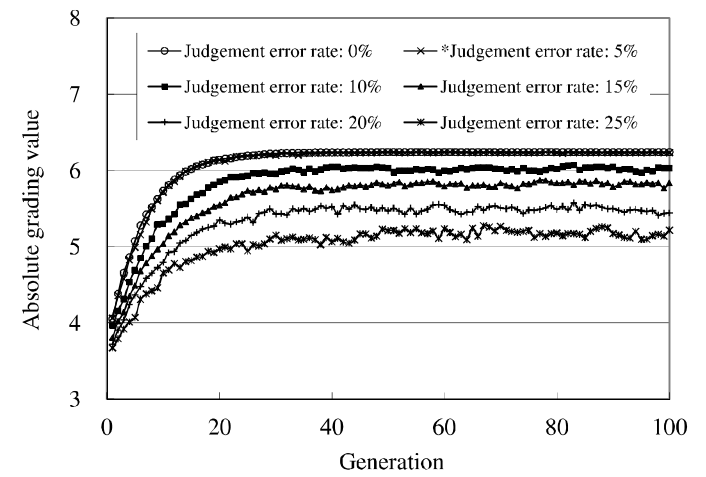


(c) Pattern 3



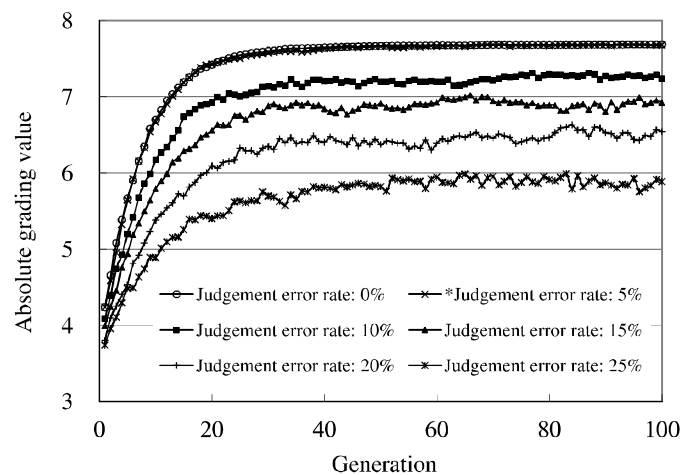
(a) Pattern 1

図 5.5: 判定誤り率を操作した場合の検討結果 (20bits)



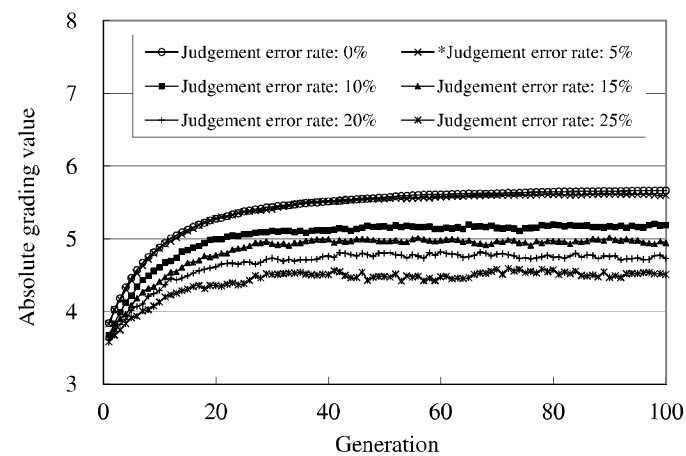
(b) Pattern 2

図 5.6: 判定誤り率を操作した場合の検討結果 (30bits)

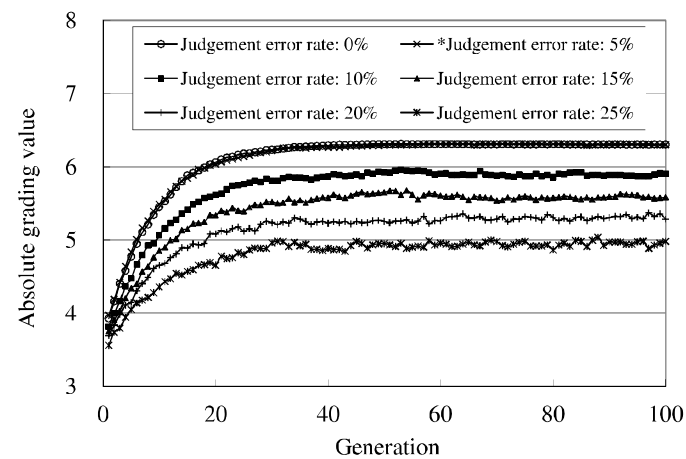


(c) Pattern 3

図 5.6: 判定誤り率を操作した場合の検討結果 (30bits)

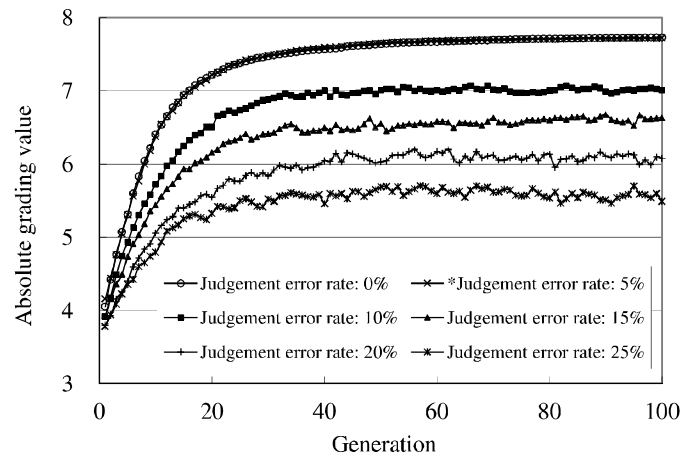


(a) Pattern 1



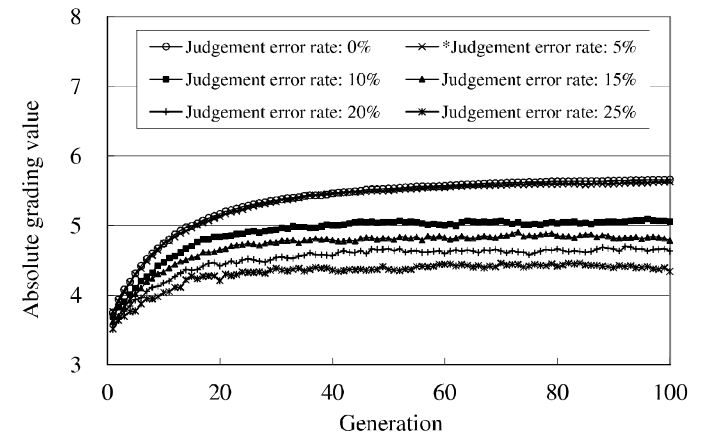
(b) Pattern 2

図 5.7: 判定誤り率を操作した場合の検討結果 (40bits)

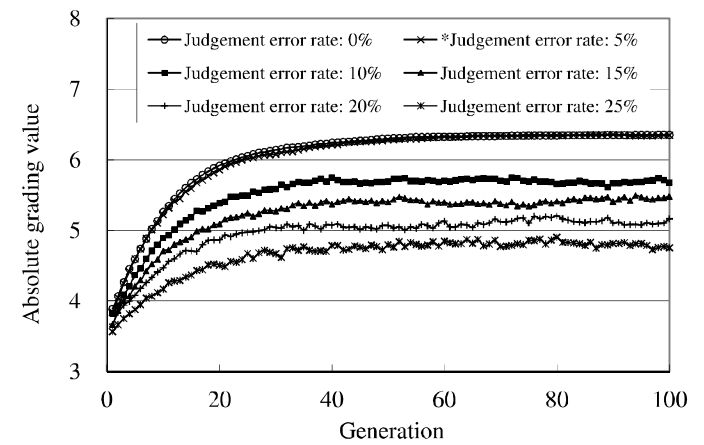


(c) Pattern 3

図 5.7: 判定誤り率を操作した場合の検討結果 (40bits)

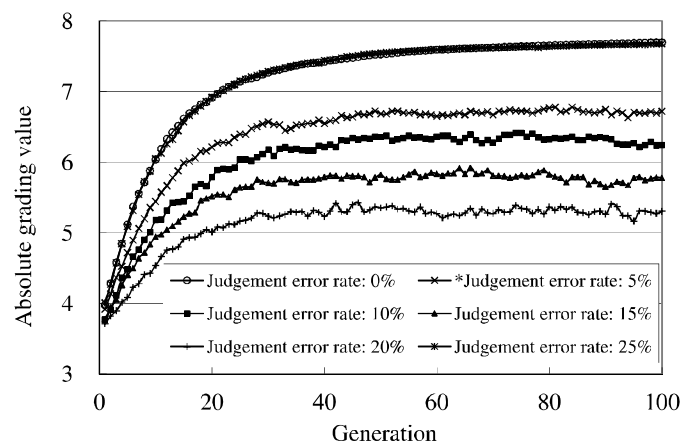


(a) Pattern 1



(b) Pattern 2

図 5.8: 判定誤り率を操作した場合の検討結果 (50bits)



(c) Pattern 3

図 5.8: 判定誤り率を操作した場合の検討結果 (50bits)

5.4.3 解評価に必要な投票数に関する検討

本項では、2点嗜好法適用時の解評価に必要な投票数の結果について考察する。

図 5.9 に評価エージェントの集団性の違いによる解評価に必要な投票数の割合 (Vote rate) の検討結果を示す。図 5.9 では、全評価エージェントが全トーナメント対戦の投票に参加した場合の投票数、すなわち TBMP1 における投票数を 100% とし、TBMP2 において、どの程度の投票数が減少されるかを確認した。いずれの遺伝子長においても、TBMP2 は TBMP1 より、解評価に必要な投票数を約 80% 減少できることが確認された。これは、5.4.1 項のように、1 世代あたりの約 90% の対戦において、2 点嗜好法により対戦解候補に有意差があると判定されているためであると考えられる。したがって、TBMP2 は、解評価に必要な投票数の減少に有効であると考えられる。

5.4.4 性能比較

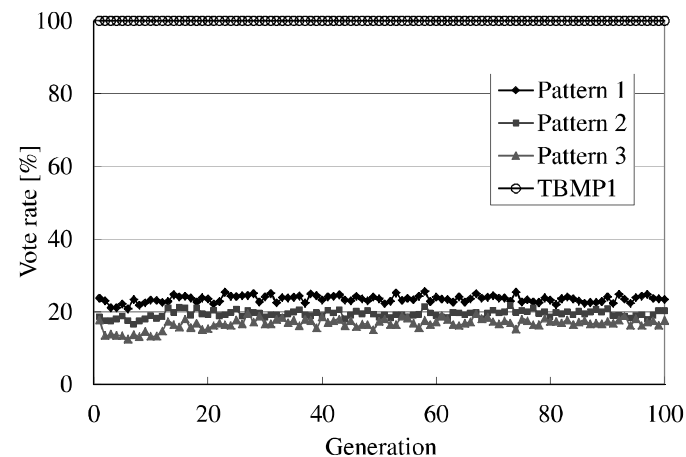
5.4.3 項までの検討において、TBMP2 では、毎世代約 90% の対戦で 2 点嗜好法により対戦解候補に有意差があると判定され、解評価に必要な投票数を約 80% 減少できることが確認された。しかし、多くのユーザの投票を獲得せずに解評価が行われた場合、解候補の進化性能が TBMP1 より劣るとい問題が想定される。そのため、本項では、TBMP2 の解候補の進化性能について考察する。

図 5.10 に評価エージェントの集団性の違いによる性能比較の結果を示す。図 5.10 の横軸は、評価エージェントの投票数 (The number of votes) を示す。いずれの遺伝子長においても、投票のより早い段階で、TBMP2 は TBMP1 と同等の絶対評価値を得られていることが確認された。また、表 5.2 に絶対評価値が収束した時点における投票数の減少率 (Decline rate) を示す。表 5.2 における減少率は、TBMP1 の絶対評価値収束時における投票数を 100% とし、TBMP2 の絶対評価値収束時における投票数を割合で表したものである。いずれの遺伝子長、集団性のパターンにおいても、投票数は 72.5 ~ 87.8% 減少されていることが確認された。

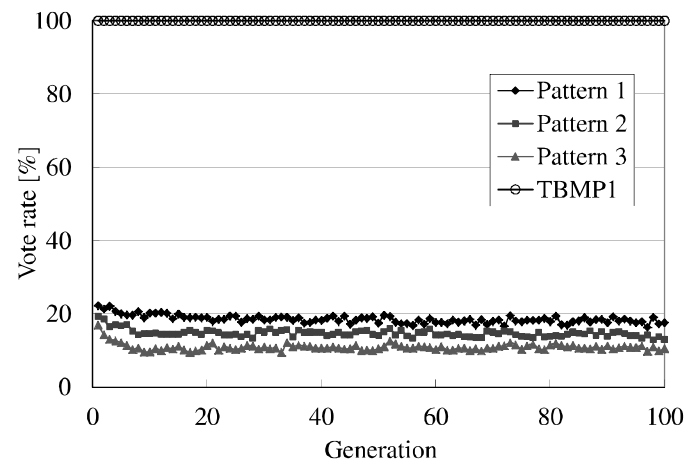
また、遺伝子長が長くなり解空間が複雑になるに連れ、TBMP2 と TBMP1 において絶対評価値に差が確認される集団性のパターンが存在している。そのため、有意水準 5% で t 検定を行い、TBMP2 及び TBMP1 の収束時の絶対評価値の統計的有意差を確認した。その結果、遺伝子長 20bits (Pattern 2)、遺伝子長 40bits (Pattern 1, 3)、遺伝子長 50bits (Pattern 1, 3) において、TBMP2 は TBMP1 より絶対評価値が低いことが確認された。したがって、TBMP2 では、全ての集団性のパターンにおいて TBMP1 と同等の絶対評価値を獲得できない可能性があると考えられる。

一般に、遺伝子長が増加すると解空間が複雑になり、最適解の探索が困難になる。さらに、TBMP2においては、2点嗜好法の適用により統計的に解候補の優劣を決定しているため、判定誤りが生じることが確認された。そのため、TBMP2では、遺伝子長が増加するに連れ、TBMP1と同等の進化性能が得られなくなったと考えられる。

TBMP2では、不特定多数のユーザが介在するWebシステムを想定した場合のトーナメント対戦の効率化を目的としている。シミュレーション結果より、TBMP2ではTBMP1と同等の絶対評価値を獲得できない集団性のパターンが存在している。しかし、TBMP2では、TBMP1より解評価に必要な投票数を約80%減少でき、投票開始からより早い段階で解候補の優劣判定を行える可能性があることが確認された。そのため、TBMP2は、トーナメント対戦の効率化という観点より、有効性は高いと考えられる。

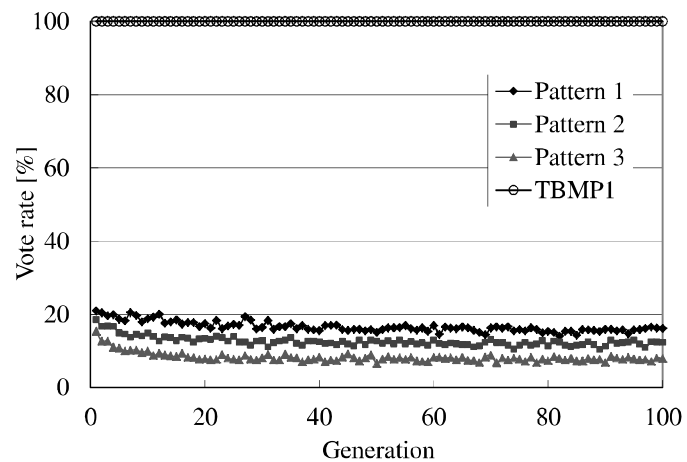


(a) 20bits

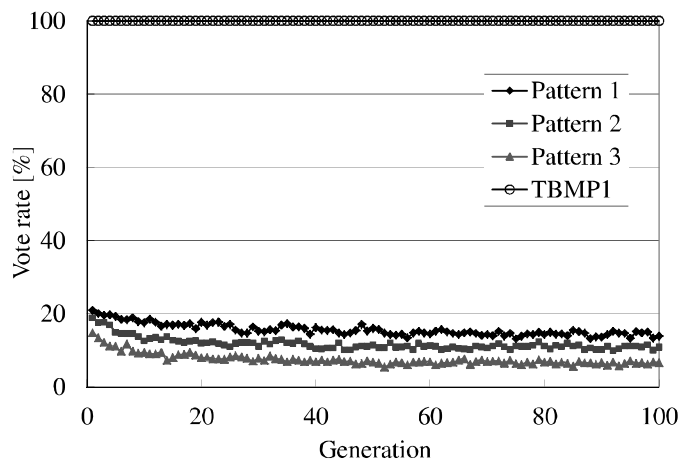


(b) 30bits

図 5.9: 解評価に必要な投票数の検討結果

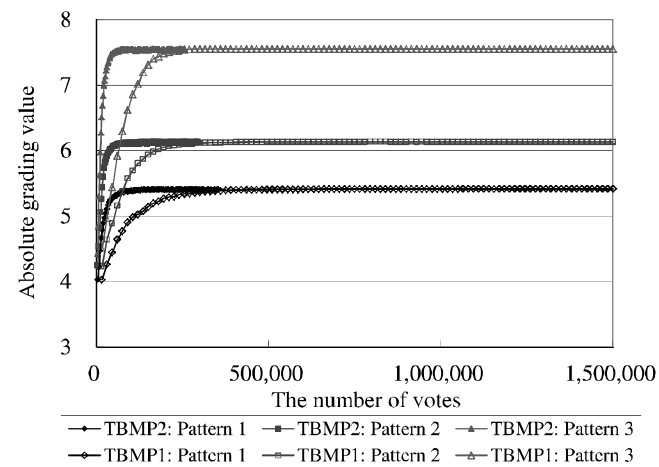


(c) 40bits

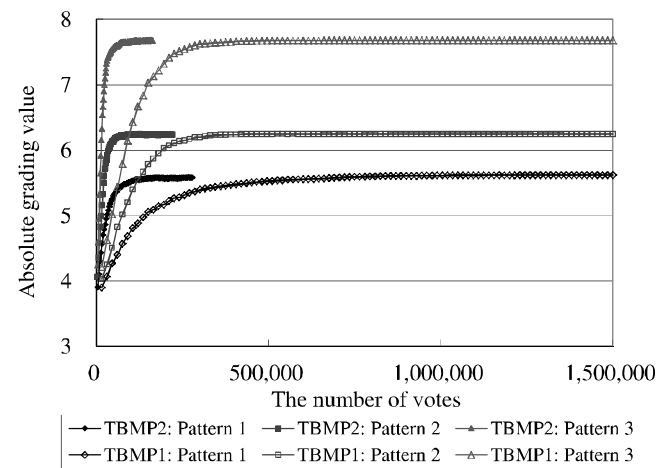


(d) 50bits

図 5.9: 解評価に必要な投票数の検討結果



(a) 20bits



(b) 30bits

図 5.10: 性能比較の結果

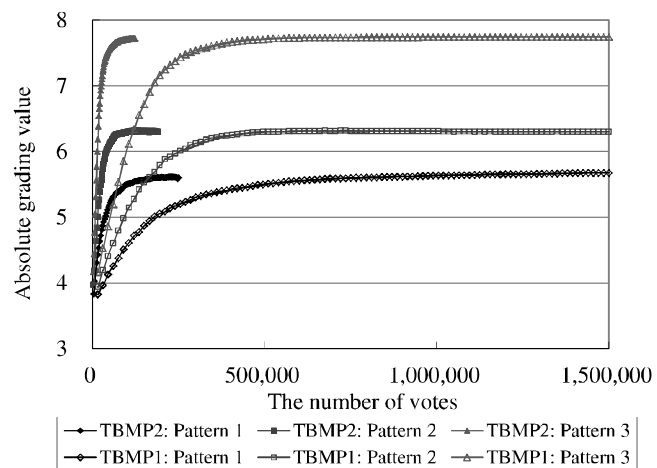
表 5.2: 性能比較の結果

(a) 20bits

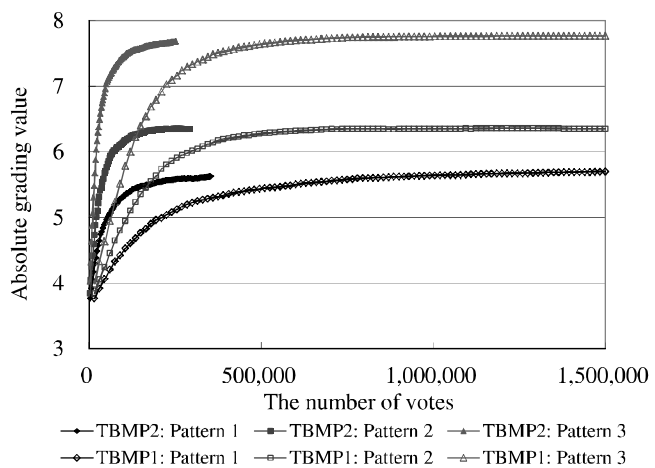
Method	Absolute grading value	The number of votes	Decline rate [%]
TBMP1: Pattern 1	5.42	1,065,000	—
TBMP2: Pattern 1	5.41	130,024	87.8
TBMP1: Pattern 2	6.14	795,000	—
TBMP2: Pattern 2	6.13	127,277	84.0
TBMP1: Pattern 3	7.55	270,000	—
TBMP2: Pattern 3	7.55	70,865	73.8

(b) 30bits

Method	Absolute grading value	The number of votes	Decline rate [%]
TBMP1: Pattern 1	5.62	990,000	—
TBMP2: Pattern 1	5.59	172,865	82.5
TBMP1: Pattern 2	6.25	405,000	—
TBMP2: Pattern 2	6.24	88,660	78.1
TBMP1: Pattern 3	7.68	465,000	—
TBMP2: Pattern 3	7.68	109,269	76.5



(c) 40bits



(d) 50bits

図 5.10: 性能比較の結果

表 5.2: 性能比較の結果

(c) 40bits

Method	Absolute grading value	The number of votes	Decline rate [%]
TBMP1: Pattern 1	5.68	1,440,000	—
TBMP2: Pattern 1	5.62	221,414	84.6
TBMP1: Pattern 2	6.32	675,000	—
TBMP2: Pattern 2	6.32	121,642	82.0
TBMP1: Pattern 3	7.74	705,000	—
TBMP2: Pattern 3	7.72	107,125	84.8

(d) 50bits

Method	Absolute grading value	The number of votes	Decline rate [%]
TBMP1: Pattern 1	5.70	1,395,000	—
TBMP2: Pattern 1	5.63	346,491	75.2
TBMP1: Pattern 2	6.36	810,000	—
TBMP2: Pattern 2	6.35	224,365	72.3
TBMP1: Pattern 3	7.77	1,185,000	—
TBMP2: Pattern 3	7.69	252,031	78.7

5.5 結言

本章では、TBMP2のシステム概要及び数値シミュレーションについて述べた。数値シミュレーションによる検証結果は以下のとおりとなった。

(1) 2点嗜好法による優劣判定対戦数に関する検討

いずれの遺伝子長、集団性のパターンにおいても、1世代あたり約90%の対戦において、2点嗜好法により対戦候補に有意差があると判定されていることが確認された。

(2) 2点嗜好法の判定誤り率に関する検討

いずれの遺伝子長、集団性のパターンにおいても、2点嗜好法により有意差があると判定された対戦数のうち、約95%が正しく判定されていることが確認された。

(3) 解評価に必要な投票数に関する検討

いずれの遺伝子長、集団性のパターンにおいても、TBMP2はTBMP1より、解評価に必要な投票数を約80%減少できることが確認された。

(4) 性能比較

TBMP2においては、TBMP1と同等の絶対評価値を獲得できない集団性のパターンが存在することが確認された。しかし、TBMP2においては、解評価に必要な投票数を約80%減少でき、投票開始からより早い段階で解候補の優劣判定を行える可能性がある。そのため、TBMP2は、トーナメント対戦の効率化という観点より、有効性は高いと考えられる。

本章では、数値シミュレーションにより、トーナメント対戦の効率化という観点においてTBMP2の有効性が確認された。しかし、TBMP1においては、勝ち上がった解候補は複数回評価されなければならない、ユーザの解評価に対するモチベーションが低下してしまうという問題が想定される。第6章では、この問題を解決するため、数値シミュレーションにおけるTBMP3のユーザの解評価に対するモチベーション維持に関する有効性の検証について述べる。

第6章 対戦結果推定の適用

6.1 緒言

本章では、対戦結果推定を適用した複数参加型トーナメント方式である TBMP3 について述べる [24].

TBMP1 においては、ユーザは同一世代に勝ち上がった解候補を複数回評価しなければならない。そのため、ユーザの解評価負担を軽減できると考えられる一対比較による解評価であっても、ユーザの解評価に対するモチベーション低下を引き起こすおそれがある。また、IGA では、解候補集団が収束するに連れて、遺伝子パターンが類似した解候補が増加する。そこで、これらの問題によるユーザの解評価に対するモチベーション低下を防ぐため、対戦結果推定の適用を試みる。対戦結果推定は、過去の対戦の投票結果や評価値の履歴を用いて、対戦結果が推定可能な解候補同士の対戦ではユーザに評価を委ねず、システム側で対戦結果を決定する手法である。

本章では、TBMP3 の概要及びユーザの解評価に対するモチベーション維持に関する有効性を検証するために行う数値シミュレーションについて述べる。本シミュレーションでは、対戦結果推定における推定率、推定誤り率に関する検討及び TBMP1 を比較手法とした性能比較を行う。

6.2 対戦結果推定を適用した複数参加型トーナメント方式

本節では、TBMP1の問題点について述べ、対戦結果推定を適用する目的及びシステム概要について述べる。

6.2.1 TBMP1の問題点

トーナメント式評価手法においては、一度の評価で2つの解候補の良し悪しを判断する一対比較評価を繰り返し、解候補の優劣を決定する。しかし、トーナメント式評価手法では、ユーザは勝ち上がった解候補を複数回評価しなければならない。そのため、ユーザの解評価に対するモチベーション低下を引き起こすおそれがある。また、IGAでは、世代交代を経る毎に解候補集団が収束し、遺伝子パターンが類似した解候補が増加する。TBMP1においては、ユーザは多人数となり、好みの解候補に1票を投じるのみの評価を行うため、評価負担は軽減できると考えられる。しかし、遺伝子パターンが類似した解候補に何度も投票を行わなければならないという問題は解決できない。

そこで、これらの問題を解決するため、TBMP1において対戦結果推定を適用することを試みる。対戦結果推定の適用においては、ユーザの解評価に対するモチベーション維持を目的とする。ユーザの解評価に対するモチベーション維持のためには、(1) 遺伝子パターンが類似した解候補同士の対戦や勝ち上がった解候補の複数回にわたる評価をユーザに強いることを避け、(2) 解候補集団の初期収束性の向上を図る必要がある。

(1)は、対戦結果推定における有効性を検証する上で重要である。(1)については、TBMP3の数値シミュレーションにおいて、対戦結果推定がどの程度の対戦数で実行され、そのうちの程度の勝敗結果が正しく推定されているかを検証する。

(2)は、対戦結果推定の適用により、実現できると考えられる。対戦結果推定が実行されると、ユーザに対戦解候補を提示することなく、より多くのトーナメント対戦を進行できる可能性がある。さらに、対戦結果推定が正しく行われた場合は、ユーザの実評価回数を増加させず、対戦結果推定によりトーナメント対戦を正しく進行でき、解候補集団の初期収束性の向上が可能になると考えられる。(2)については、数値シミュレーションにより、TBMP3及びTBMP1の進化性能を比較し、TBMP3の有効性を検証する。

6.2.2 TBMP3のシステム概要

本項では、TBMP3の概要について述べる [24]。対戦結果推定は、過去の対戦の投票結果や評価値の履歴を用いて、対戦結果が推定可能な解候補同士の対戦においてはユーザに評価を委ねず、システム側で対戦結果を決定する手法である。

図6.1にTBMP3のシステム概要を示す。まず、初期遺伝子集団を生成し、生成された解候補をトーナメント表に配置する。第1世代目においては、履歴情報が獲得されていないため、対戦結果推定は行わず、全トーナメント対戦を多くのユーザの投票により評価してもらう。そして、第2世代目より、対戦結果推定を試みる。対戦解候補の優劣が推定可能であれば、システム側で対戦解候補の優劣を決定する。対戦解候補の優劣が推定不可能であれば、ユーザに評価を委ねる。決勝戦まで終了すれば、解候補を評価し、GA処理を行う。TBMP3における解候補の評価値は、4.3.2項で示した要領で与えられる。

TBMP3においては、対戦履歴及び評価履歴の2つの履歴情報を用いる。対戦履歴は、対戦が終了した2つの解候補の遺伝子パターンと対戦結果を保持する。評価履歴は、解候補の遺伝子パターンと評価値を保持する。対戦履歴は各対戦終了後、評価履歴は各世代終了後に更新される。なお、対戦結果推定が実行された場合は、対戦及び評価履歴により推定された対戦結果が実際の対戦結果となる。このような場合は、解候補の勝敗結果や評価値などは対戦及び評価履歴に追加されない。

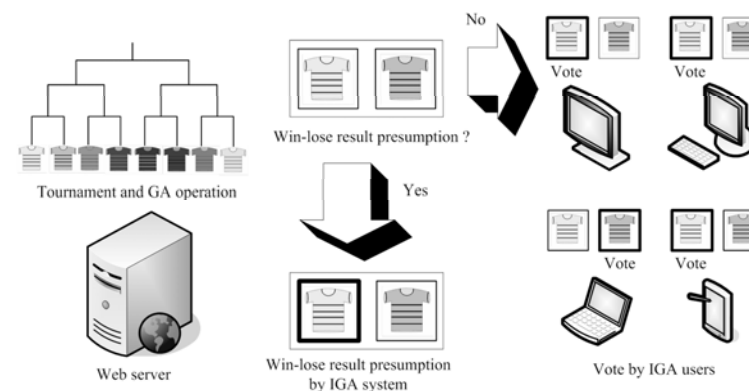


図 6.1: TBMP3のシステム概要

6.2.3 対戦結果推定

対戦結果推定では、まず、2つの対戦解候補を対戦履歴内より検索する。対戦履歴内に2つの対戦解候補が存在すれば、対戦履歴と同様の対戦結果を採用する。対戦履歴内に2つの対戦解候補が存在しなければ、評価履歴を参照し対戦結果推定の適用を試みる。

TBMP3では、評価履歴の参照において2つの対戦結果推定手法、すなわち完全一致手法 (Win-Lose Result Presumption 1: WLRP1) 及び加重平均手法 (Win-Lose Result Presumption 2: WLRP2) を適用した場合の数値シミュレーションを行う。

WLRP1における対戦結果推定は、2つの対戦解候補が評価履歴内に完全一致する遺伝子パターンを有する場合に実行される。WLRP1における解候補の推定評価値は、評価履歴の中で完全一致した遺伝子パターンの評価値をそのまま用いる。

WLRP2における対戦結果推定は、2つの対戦解候補が予め設定した閾値内に評価履歴を有する場合に実行される。本閾値は、遺伝子パターンのハミング距離として定義される。WLRP2における解候補の推定評価値は、評価履歴の中で設定された閾値内に存在する履歴解候補の加重平均による評価値となる。WLRP2における推定評価値 F_a は、式(6.1)で表現される。

$$F_a = \frac{\sum_{i=1}^n f_i \times \frac{1}{l_i}}{\sum_{i=1}^n \frac{1}{l_i}} \quad (6.1)$$

式(6.1)において、 n は閾値内の評価履歴数、 f_i は閾値内の評価履歴解候補の評価値、 l_i は対戦解候補と評価履歴内の解候補のハミング距離である。

例えば、図6.2のように対戦解候補 (Competing candidate solution) と評価履歴解候補 r_1, r_2 が存在した場合、対戦解候補の評価値は、式(6.2)となる。

$$F_a = \frac{4 \times \frac{1}{3} + 6 \times \frac{1}{4}}{\frac{1}{3} + \frac{1}{4}} = \frac{34}{7} \cong 4.9 \quad (6.2)$$

しかし、式(6.1)においては、2つの対戦解候補の遺伝子パターンが異なる場合でも、2つの対戦解候補は同じ推定評価値になってしまう場合がある。例えば、図6.3(a)のように対戦解候補Aと遺伝子パターンが完全一致する評価履歴解候補A'が存在し、かつ対戦解候補Bが閾値内に評価履歴解候補A'しか有しない場合、対戦解候補A、Bは遺伝子パターンが異なるにも関わらず、同じ推定評価値になる。また、図6.3(b)のように対戦解候補A、Bがそれぞれハミング距離が等しい評価履歴解候補A'しか有しない場合も、対戦解候補A、Bは同じ推定評価値になる。このような場合は、2つの対戦解

候補の推定評価値が等しくなり矛盾が生じるため、対戦結果推定を適用せず、ユーザに評価を委ねる。

次に、WLRP1及びWLRP2によって求められた推定評価値より、2つの対戦解候補の獲得投票数を推定する。獲得投票数は、式(4.1)によって求められた式(6.3)、(6.4)で表現される。

$$Y_{win-p} = \frac{X_{win.result}}{X_{win.result} + X_{lose.result}} \times V_{ave} \quad (6.3)$$

$$Y_{lose-p} = \frac{X_{lose.result}}{X_{win.result} + X_{lose.result}} \times V_{ave} \quad (6.4)$$

式(6.3)、(6.4)において、 Y_{win-p} は勝者と推定される解候補の推定獲得投票数、 Y_{lose-p} は敗者と推定される解候補の推定獲得投票数、 $X_{win.result}$ は勝者と推定される解候補の推定評価値、 $X_{lose.result}$ は敗者と推定される解候補の推定評価値、 V_{ave} は対戦結果推定が適用される時点までの各対戦の平均投票数である。実際のWebシステムにおける各対戦の投票数は、全て等しいとは限らない。そのため、 V_{ave} には対戦結果推定が適用される時点までの各対戦の平均投票数を用いる。

例えば、2つの対戦解候補の推定評価値を $F_{a_1} = 5.1, F_{a_2} = 4.2 (F_{a_1} > F_{a_2})$ 、対戦結果推定が適用される時点までの各対戦の平均投票数を $V_{ave} = 1,000$ とすると、 $X_{win.result} = 5.1, X_{lose.result} = 4.2$ となり、 Y_{win-p}, Y_{lose-p} は式(6.5)、(6.6)のように求められる。

$$Y_{win-p} = \frac{5.1}{5.1 + 4.2} \times 1,000 \cong 548 \quad (6.5)$$

$$Y_{lose-p} = \frac{4.2}{5.1 + 4.2} \times 1,000 \cong 451 \quad (6.6)$$

対戦結果推定では、式(6.5)、(6.6)により求められた推定獲得投票数によって、2つの対戦解候補の優劣を決定する。

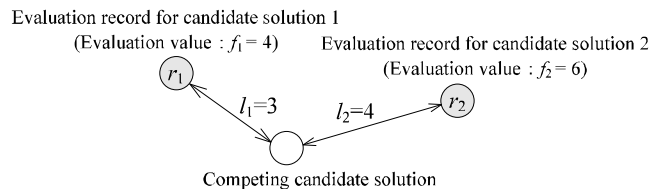


図 6.2: WLRP2 の例

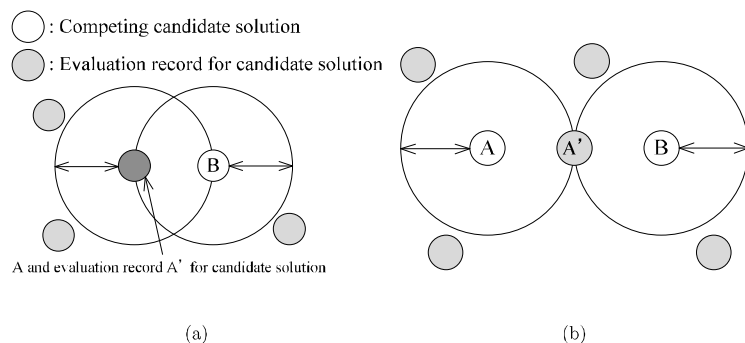


図 6.3: WLRP2 の例外

6.3 数値シミュレーションによる性能評価

本節では、TBMP3のユーザの解評価に対するモチベーション維持に関する有効性を検証するために行う数値シミュレーションの概要について述べる。本シミュレーションでは、対戦結果推定における推定率、推定誤り率に関する検討及びTBMP1を比較手法とした性能比較を行う。

図6.4にTBMP3のシミュレーションの流れを示す。まず、初期遺伝子集団を生成し、生成された解候補をトーナメント表に配置する。次に、評価エージェントによる投票を行い、トーナメント対戦を進行する。第1世代目においては、対戦及び評価履歴が獲得されていないため、対戦結果推定は行われない。第2世代目以降の対戦より、各対戦において対戦結果推定が試みられる。対戦結果推定により2つの対戦解候補の優劣が推定可能であれば、システム側で対戦解候補の優劣を決定する。2つの対戦解候補の優劣が推定不可能であれば、評価エージェントに評価を委ねる。決勝戦まで終了すれば、解候補に評価値を与え、GA処理を行う。そして、新しい解候補を提示し、再び評価エージェントによる投票を行う。

本シミュレーションにおける評価エージェントの仕様、集団性及びシミュレーションパラメータは、第4章で用いたものと同様である。ただし、本シミュレーションでは、GA処理における突然変異率は1%、評価エージェントの投票参加割合は100%とした。WLRP2における評価履歴採用閾値は1bit及び遺伝子長の5%、10%、20%の場合を対象とした。なお、6.4節におけるシミュレーション結果は、シミュレーション試行回数分の平均値である。

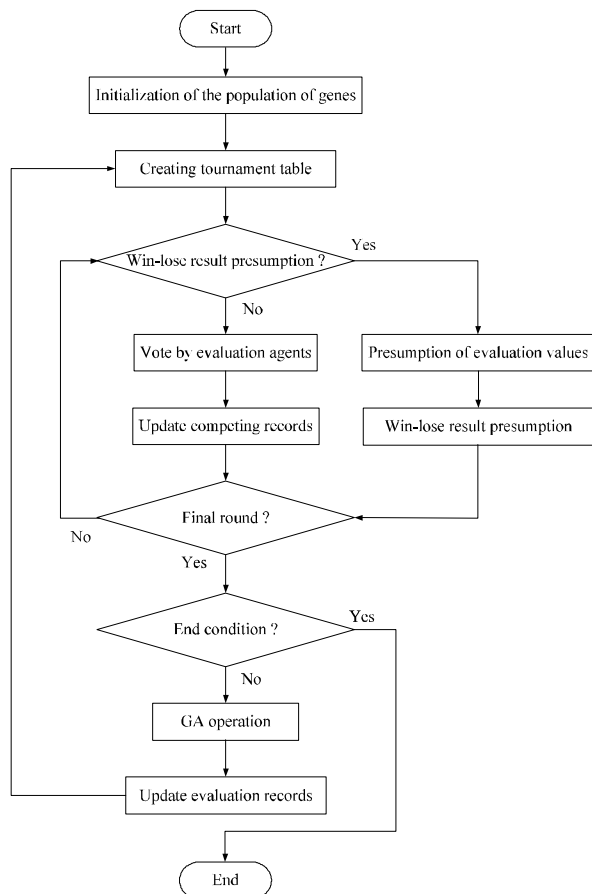


図 6.4: TBMP3 のシミュレーションの流れ

6.4 シミュレーション結果

本節では、数値シミュレーションにおける対戦結果推定の推定率、推定誤り率に関する検討及びTBMP1を比較手法とした性能比較の結果について考察する。

6.4.1 推定率に関する検討

推定率は、1世代あたりにどの程度の対戦数に対して、対戦結果推定による優劣判定が行われたかの割合を示す。

図 6.5 に評価エージェントの集団性の違いによる推定率 (Presumption rate) の検討結果を示す。図 6.5 中の WLRP2 の括弧内のビット数及びパーセンテージの表記は、評価履歴採用閾値を示す。いずれの集団性のパターンにおいても、推定率は世代交代を経る毎に増加することが確認された。また、集団性に偏りを持たせた Pattern 3 は、Pattern 1, 2 と比較して、WLRP1 の推定率が高いことが確認された。これは、Pattern 3 における多数派集団 (Group A) の投票傾向が対戦結果を決定してしまう可能性が高く、対戦結果推定が実行されやすくなったためであると考えられる。さらに、WLRP2 は、対戦結果推定が実行される条件が WLRP1 と比較して緩いため、推定率が高くなることが確認された。

6.4.2 推定誤り率に関する検討

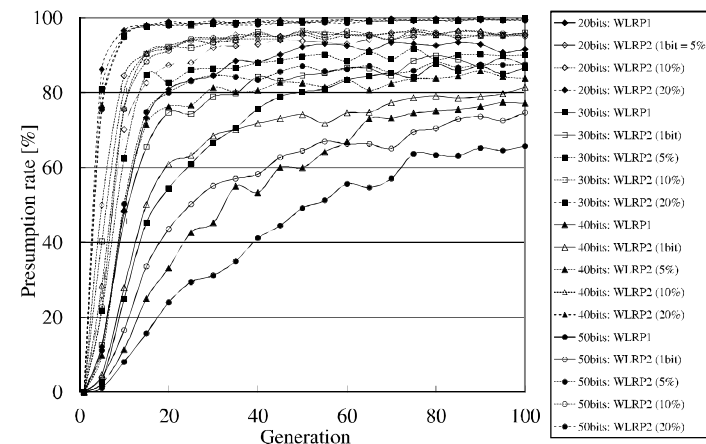
推定誤り率は、対戦結果推定による優劣判定が行われた対戦のうち、対戦結果推定による優劣判定と全評価エージェントが投票を行った場合の優劣判定が異なる割合を示す。

図 6.6 に評価エージェントの集団性の違いによる推定誤り率 (Presumption error rate) の検討結果を示す。いずれの集団性のパターンにおいても、WLRP2 では評価履歴採用閾値が増加すると、推定誤り率が増加することが確認された。しかし、世代交代を経ても推定誤り率は大きく変動せず、WLRP2 (20%) のときに最も高く、約 30% であった。また、Pattern 3 においては、Pattern 1, 2 と比較して、推定誤り率が低いことが確認された。これは、Pattern 3 における多数派集団の投票傾向が対戦結果を決定してしまう可能性が高いため、対戦結果推定における推定誤りが発生しにくくなったためであると考えられる。これらの結果より、推定誤りは、対戦結果推定が実行される条件が緩和されると発生しやすくなると考えられる。

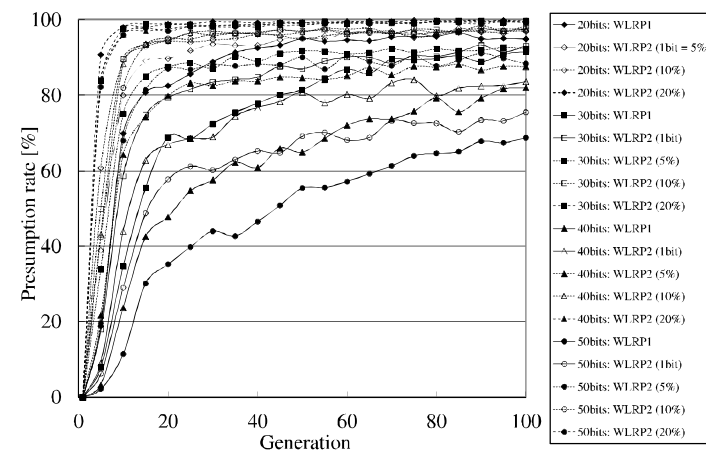
対戦結果推定における推定誤りを決定する要因として、対戦履歴 (Competition record) 及び評価履歴 (Evaluation record) の数が重要である。履歴数が少ない場合は、対戦解

候補の評価値を推定できたとしても、正確な推定評価値を算出できず、推定誤りが発生しやすくなる可能性がある。

図 6.7, 6.8 に対戦及び評価履歴数の推移を示す。対戦及び評価履歴数は、世代交代を経る毎に増加していることが確認された。しかし、WLRP2において評価履歴採用閾値が高くなると、対戦及び評価履歴数の増加率は低くなっている。このような場合は、図 6.6 における推定誤り率は高くなっている。また、WLRP1においては、履歴解候補と対戦解候補が完全に一致しなければ対戦結果推定は実行されないため、履歴数の増加率が高い。これらの結果より、対戦結果推定における評価履歴採用閾値が高い場合は、対戦及び評価履歴数は増加しにくく、推定誤りが発生しやすいことが確認された。

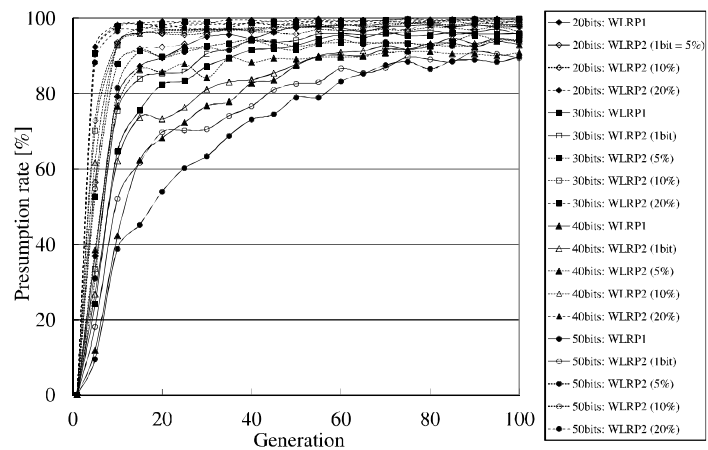


(a) Pattern 1



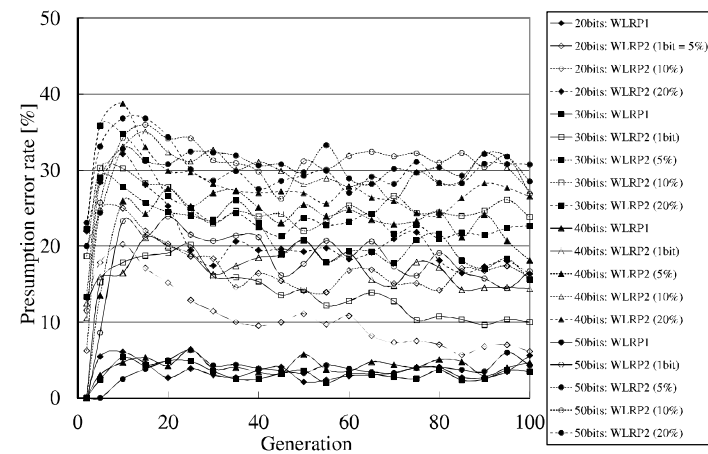
(b) Pattern 2

図 6.5: 推定率の検討結果

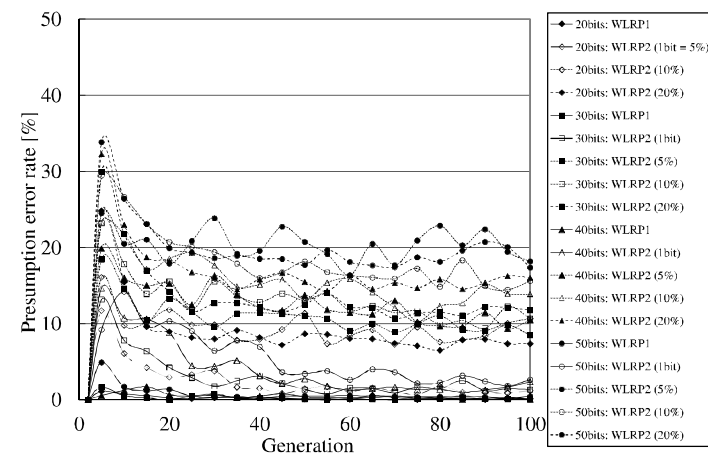


(c) Pattern 3

図 6.5: 推定率の検討結果

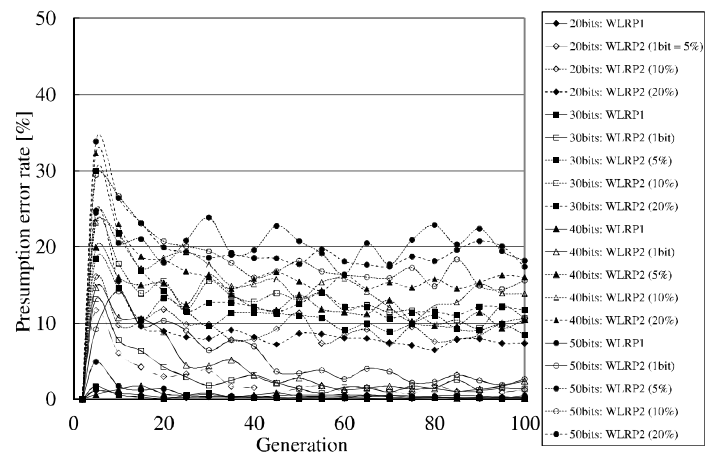


(a) Pattern 1



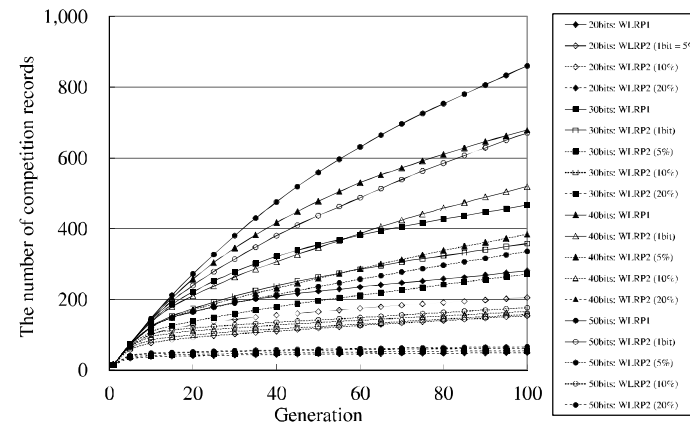
(b) Pattern 2

図 6.6: 推定誤り率の検討結果

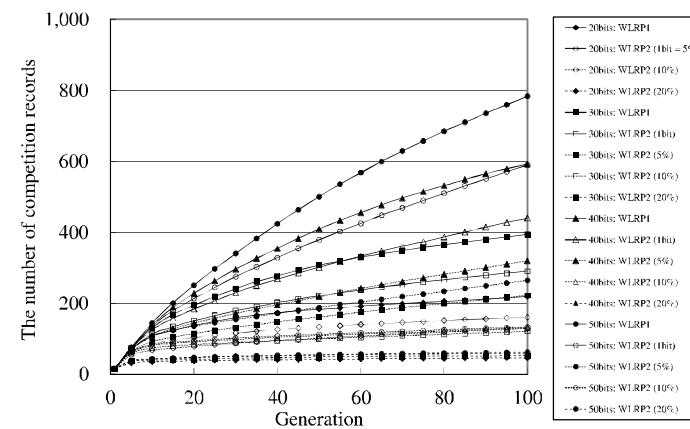


(c) Pattern 3

図 6.6: 推定誤り率の検討結果

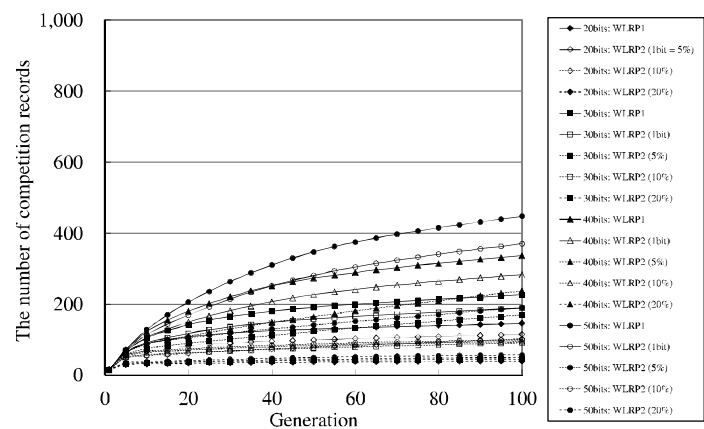


(a) Pattern 1



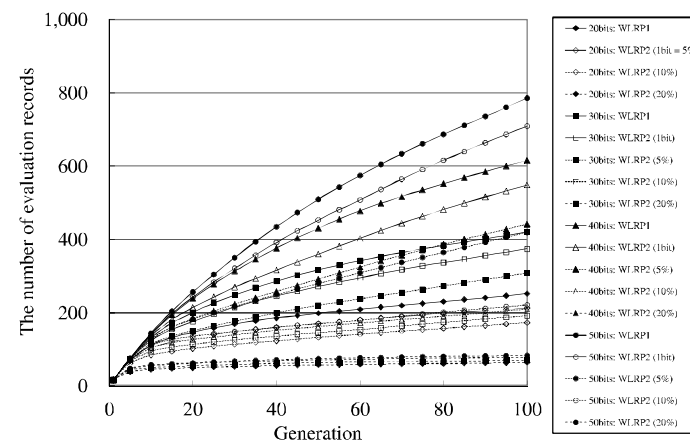
(b) Pattern 2

図 6.7: 対戦履歴数の推移

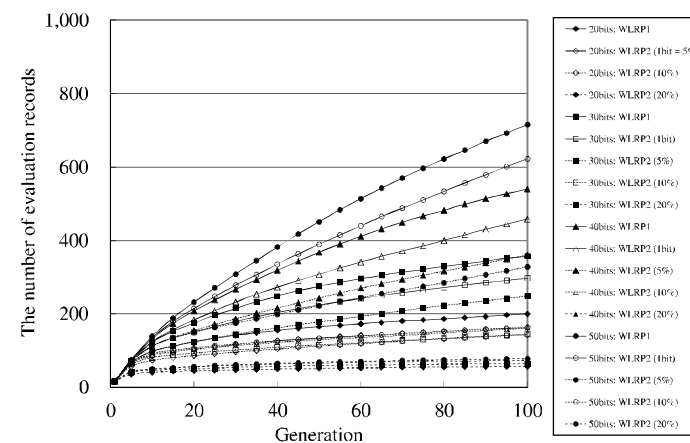


(c) Pattern 3

図 6.7: 対戦履歴数の推移

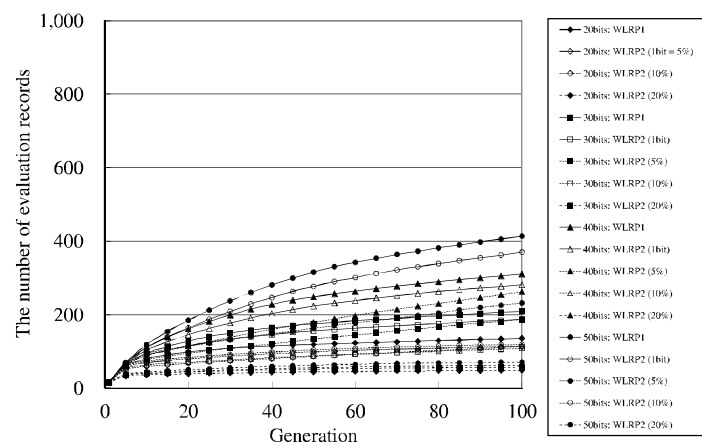


(a) Pattern 1



(b) Pattern 2

図 6.8: 評価履歴数の推移



(c) Pattern 3

図 6.8: 評価履歴数の推移

6.4.3 性能比較

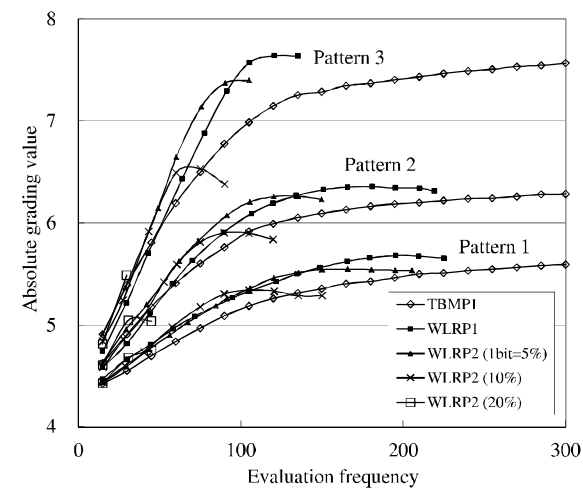
本項では、TBMP3及びTBMP1を比較手法とした性能比較の結果について考察する。

まず、TBMP3及びTBMP1における進化性能について考察する。図6.9に性能比較の結果を示す。図6.9の横軸は評価エージェントの実評価回数 (Evaluation frequency), 縦軸の絶対評価値 (Absolute grading value) は各世代のエリート解候補に対する全評価エージェントの絶対評価値のシミュレーション試行回数分の平均値を示す。いずれの遺伝子長においても、投票開始初期では、TBMP3はTBMP1より絶対評価値が高いことが確認された。TBMP3のうち、WLRP1が最も良好な進化性能を有している。これは、WLRP1は、WLRP2と比較して推定誤り率が低いいため、対戦結果推定により解候補の進化が妨げられなかったためであると考えられる。また、WLRP2 (5%, 10%) は、初期段階において、WLRP1より絶対評価値が高いことが確認された。これは、WLRP2 (5%, 10%) は、WLRP1より推定率が高いためであると考えられる。そのため、WLRP2 (5%, 10%) ではWLRP1よりも対戦結果推定が多く行われる。しかし、WLRP2 (5%, 10%) はWLRP1より新たな対戦、評価履歴を獲得しにくいいため、推定誤りが増加し、最終的に解の進化が停滞したと考えられる。この傾向は、集団性に偏りのあるPattern 2, 3で顕著である。これは、Pattern 2, 3の多数派集団の投票傾向が対戦結果推定に大きく影響したためであると考えられる。

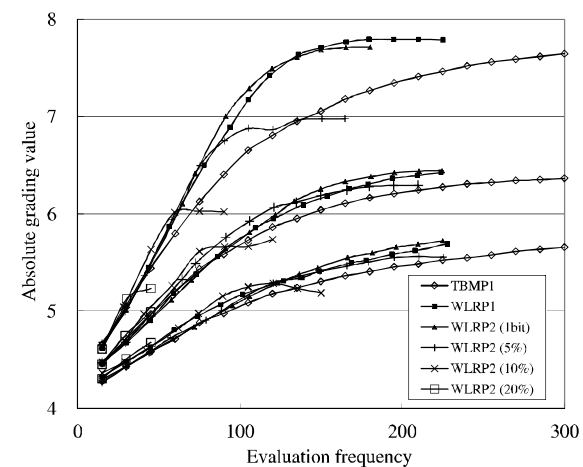
次に、投票初期段階における絶対評価値の統計的有意差について考察する。表6.1にTBMP3とTBMP1の絶対評価値において有意水準5%でt検定を実行した結果を示す。表6.1におけるt検定は、第2世代以降の絶対評価値の結果に対して、評価エージェントの実評価回数が225回までのうち、実評価回数15回毎に行った。表6.1中の統計的有意差の割合 (The rate of significant difference) は、(t検定により有意差があると判定された数/t検定の総実行回数) である。WLRP2の評価履歴採用閾値が高い場合においては、推定率が高く、評価エージェントの実評価回数が少ないため、t検定実行回数は少なくなっている。統計的有意差の割合は、遺伝子長が増加すると減少していることが確認された。これは、遺伝子長の増加により解空間が膨大になるため、TBMP3においては推定誤りにより絶対評価値が低下したためであると考えられる。また、統計的有意差の割合は、評価エージェントの集団性に偏りがあるPattern 2, 3において高いことが確認された。これは、Pattern 2, 3の多数派集団の投票傾向により、推定誤りが発生しにくいためであると考えられる。表6.1より、遺伝子長20bits (Pattern 3, WLRP2: 20%), 遺伝子長30bits (Pattern 2, WLRP2: 20%), 遺伝子長50bits (Pattern 1, WLRP2: 1bit) の場合を除いて、評価エージェントの実評価回数が225回以内においては、TBMP3はTBMP1より絶対評価値が高い場合があることが確認された。

TBMP3は、不特定多数のユーザが介在する Web システムを想定した場合のトーナメント対戦における、ユーザの解評価に対するモチベーション維持を目的としている。本シミュレーションにより、TBMP3は、評価エージェントの実評価回数を増加させることなく初期収束性の向上を実現できることが確認された。また、TBMP3とTBMP1の絶対評価値においてt検定を適用した結果、表6.1のように多くの遺伝子長、集団性の条件において、評価エージェントの実評価回数が225回以内では、TBMP3はTBMP1より絶対評価値が高い場合があることが確認された。これらの結果より、解候補集団の初期収束性の向上という観点では、遺伝子長が20, 30, 40bitsのときはWLRP1とWLRP2 (5%), 遺伝子長50bitsのときは集団性に偏りがある場合にWLRP1とWLRP2 (5%) が有効であると考えられる。

TBMP3においては、対戦結果推定が実行された対戦のうち、約70%が正しく推定されていることが確認された。また、性能比較の結果より、多くの遺伝子長、集団性の条件において、TBMP3はTBMP1より絶対評価値が高いことが確認された。したがって、TBMP3は、ユーザの解評価に対するモチベーションを維持するという観点より、有効性は高いと考えられる。



(a) 20bits

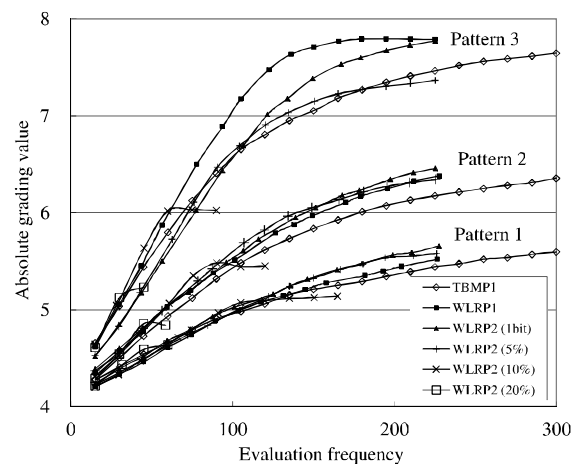


(b) 30bits

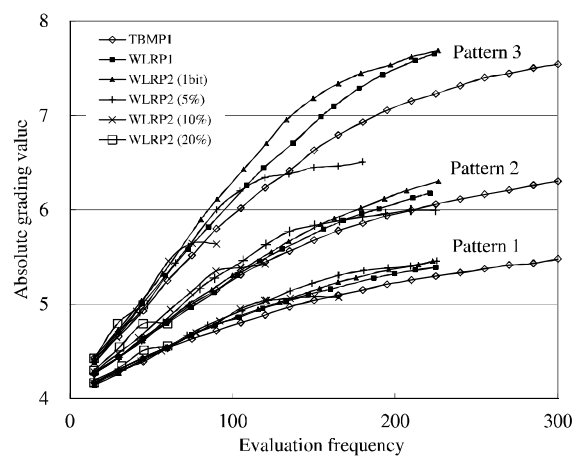
図 6.9: 性能比較の結果

表 6.1: t 検定の結果 (有意水準 : 5%)

Gene length	Presumption method	The rate of significant difference		
		Pattern 1	Pattern 2	Pattern 3
20	WLRP1	14/14	10/14	6/ 8
	WLRP2 (1bit = 5%)	9/13	7/ 9	5/ 7
	WLRP2 (10%)	4/10	4/ 7	2/ 5
	WLRP2 (20%)	1/ 2	1/ 2	0/ 1
30	WLRP1	11/14	8/14	11/14
	WLRP2 (1bit)	9/14	9/14	8/11
	WLRP2 (5%)	8/14	10/13	4/10
	WLRP2 (10%)	5/ 9	3/ 7	3/ 5
	WLRP2 (20%)	2/ 2	0/ 2	1/ 2
40	WLRP1	2/14	8/14	10/14
	WLRP2 (1bit)	8/14	11/14	11/14
	WLRP2 (5%)	8/14	11/14	10/14
	WLRP2 (10%)	4/11	4/ 7	3/ 5
	WLRP2 (20%)	2/ 3	1/ 3	2/ 2
50	WLRP1	1/14	1/14	9/14
	WLRP2 (1bit)	0/14	9/14	11/14
	WLRP2 (5%)	3/14	7/14	6/11
	WLRP2 (10%)	3/11	3/ 7	3/ 5
	WLRP2 (20%)	1/ 3	2/ 3	1/ 2



(c) 40bits



(d) 50bits

図 6.9: 性能比較の結果

6.5 結言

本章では、TBMP3のシステム概要及び数値シミュレーションについて述べた。数値シミュレーションによる検証結果は以下のとおりとなった。

(1) 推定率に関する検討

いずれの遺伝子長、集団性のパターンにおいても、世代交代を経る毎に推定率は上昇することが確認された。また、WLRP2は、WLRP1と比較して対戦結果推定の条件が緩いため、推定率が高いことが確認された。

(2) 推定誤り率に関する検討

いずれの遺伝子長、集団性のパターンにおいても、対戦結果推定が実行された対戦のうち、約70%が正しく推定されていることが確認された。また、WLRP2における評価履歴採用閾値が高い場合は、対戦及び評価履歴数は増加しにくく、推定誤りが発生しやすいことが確認された。

(3) 性能比較

いずれの遺伝子長、集団性のパターンにおいても、投票の初期段階では、TBMP3はTBMP1より絶対評価値が高い場合があることが確認された。これらの結果より、TBMP3により、評価エージェントの実評価回数を増加させることなく、初期収束性の向上を実現できると考えられる。そのため、TBMP3は、ユーザの解評価に対するモチベーションを維持するという観点より、有効性は高いと考えられる。

本章では、数値シミュレーションにより、ユーザの解評価に対するモチベーションを維持するという観点においてTBMP3の有効性が確認された。第7章では、TBMP1における実ユーザを対象とした有効性の検証について述べる。

第7章 複数参加型トーナメント方式を用いたWebサイトデザイン支援システム

7.1 緒言

本章では、TBMP1を用いたWebサイトデザイン支援システム及び本システムを用いた評価実験について述べる [25]。まず、Webサイトデザイン支援システムの概要、Webサイトデザインの構成、評価インターフェースについて述べる。

次に、本システムを用いた実ユーザを対象した評価実験について述べる。本実験では、GVMを用いたWebサイトデザイン支援システムを比較システムとして、TBMP1とGVMの性能比較を行う。本実験においては、トーナメント対戦の進行に要した時間や各ユーザの投票数、生成された解候補に対する満足度及び各システムにおける解評価のしやすさなどを調査し、TBMP1の実ユーザに対する有効性を検証する。

7.2.2 Web サイトデザインの遺伝子コーディング

本システムにおける Web サイトデザインは、レイアウト、背景、文字などのデザイン要素の色、グラデーションなどの特殊効果により表現される。図 7.2 に Web サイトデザインの構成を示す。背景 A, B, C は色情報、文字 A, B, C は色情報及びフォントの種類で表現される。色情報においては、HSV 表色系を用いている。また、背景 A には 4 種類のハッチング効果を使用し、背景 C には 4 種類のグラデーション効果を使用している。さらに、4 種類のレイアウトを用いて、メニューフレーム及び写真画像の位置を決定している。図 7.3 に 4 種類のレイアウトの例を示す。図 7.3(a) はメニューフレームが左、図 7.3(b) は右、図 7.3(c) は上、図 7.3(d) は下の場合のレイアウトを示している。メニューフレームの位置によって、画像素材などの他のデザイン要素の位置も変化する。

表 7.1 に Web サイトデザインの遺伝子コーディングを示す。表 7.1 中の “Place” は、図 7.2 に示された背景及び文字の項目に対応している。レイアウトにおいては、図 7.3 で示した 4 種類のレイアウトを 2bits で表現している。背景及び文字の色情報は色相 (Hue: H) を 4bits、彩度 (Saturation: S) を 3bits、明度 (Value: V) を 2bits で表現している。背景 A におけるハッチング及び背景 C におけるグラデーションは 2bits で表現し、それぞれ 4 種類のデザインを表現している。文字 A, B, C におけるフォントの種類には、Meiryo, DejaVu Serif, Arial Black, Times New Roman, Century, Georgia, Gungsuh, Impact の 8 種類を用いている。色情報における H 及び文字におけるフォントの種類は遺伝子コーディングには、グレイコーディングを用いている。なお、レイアウト、背景のハッチング、グラデーション、文字のフォントの種類はビットパターンの割り当ては、遺伝子型であるビットパターン間のハミング距離と表現型である各デザイン要素の類似性を考慮し、決定している。本 Web サイトデザインにおける遺伝子長は 69bits となり、生成可能な Web サイトデザインの総数は約 5.9×10^{20} 通りとなる。これは、ユーザの感性を反映した解候補を生成するのに十分な数であると考えられる。

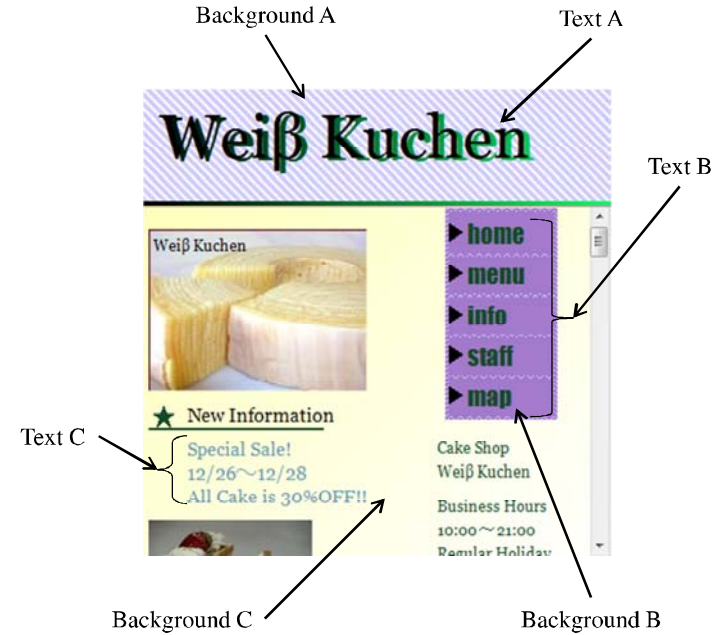


図 7.2: Web サイトデザインの構成

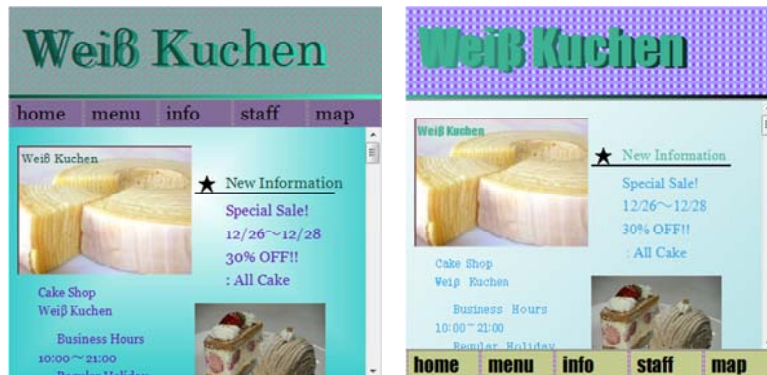
表 7.1: Web サイトデザインの遺伝子コーディング

Place	Gene information	Gene length
All	Layout	2bits
Background A	Color	9bits
	Hatching	2bits
Background B	Color	9bits
	Gradation	2bits
Background C	Color	9bits
	Gradation	2bits
Text A	Color	9bits
	Font type	3bits
Text B	Color	9bits
	Font type	3bits
Text C	Color	9bits
	Font type	3bits
Total		69bits



(a) メニューフレーム : 左

(b) メニューフレーム : 右



(c) メニューフレーム : 上

(d) メニューフレーム : 下

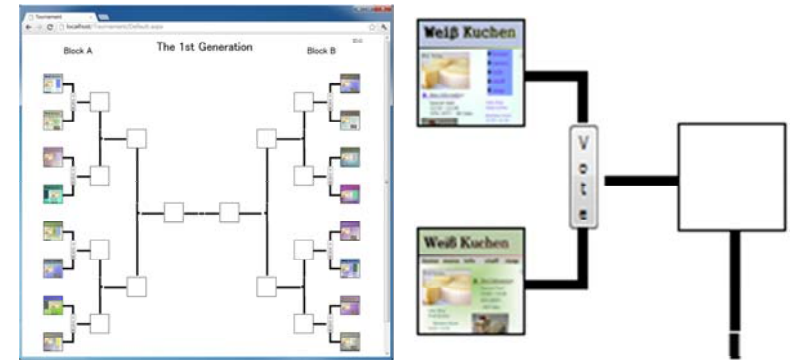
図 7.3: レイアウトの種類

7.2.3 評価インタフェース

本項では、TBMP1を用いた Web サイトデザイン支援システムの評価インタフェースについて述べる。

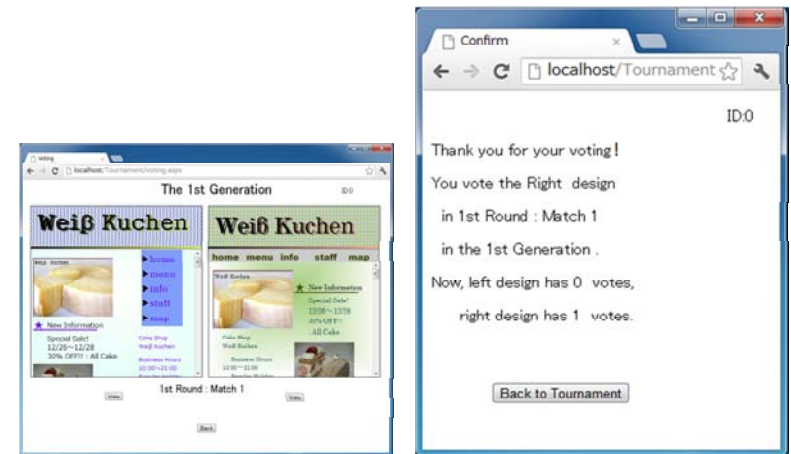
図 7.4 に TBMP1 を用いた Web サイトデザイン支援システムの評価インタフェースを示す。ユーザは、Web サイトデザイン支援システムにアクセスし、予め配布された ID 及びパスワードを使用し、ログインする。ログイン後、図 7.4(a) のように、GA 処理により生成された 16 個の Web サイトデザインが配置されたトーナメント表が表示される。図 7.4(a) において、投票可能な対戦は第 1 回戦の 8 対戦である。ユーザが投票を行う際には、図 7.4(a) のトーナメント表中で投票したい対戦の“Vote”ボタンをクリックする。図 7.4(b) は“Vote”ボタンを拡大したものである。“Vote”ボタンをクリックすると、図 7.4(c) の投票画面へ移行する。図 7.4(c) において、ユーザは各 Web サイトデザイン下にある“Vote”ボタンをクリックし、好みの方の Web サイトデザインに投票する。ユーザによる投票が行われれば、図 7.4(d) の投票確認画面へ移行する。図 7.4(d) では、ユーザが投票した対戦の投票経過が表示される。なお、本システムではユーザが投票を行った後に、投票結果が分かるようにしている。これは、ユーザが投票を行う前に投票結果が分かってしまうと、ユーザの投票行動に影響が生じると考えられるためである。例えば、ユーザが投票を行おうとする対戦において、すでに 2 つの Web サイトデザインの投票数に大きな差が生じていた場合、ユーザが投票数が少ない Web サイトデザインに投票しにくくなる可能性があると考えられるためである。

本システムでは、1 対戦ずつユーザに投票を求めるのではなく、投票可能な対戦の中からユーザが投票したい対戦を選択できるようにしている。そのため、例えば図 7.4(a) 中の Block A の上から 2 つの対戦が終了すれば、Block A の上側の第 2 回戦は投票可能となる。また、本システムでは、ユーザは各対戦において 1 票のみ投票できる。



(a) トーナメント表

(b) トーナメント表 (拡大)



(c) 投票画面

(d) 投票確認画面

図 7.4: TBMP1 を用いた Web サイトデザイン支援システムの評価インタフェース

7.3 評価実験

本節では、Web サイトデザイン支援システムを用いた評価実験の概要及び比較システムについて述べる。

7.3.1 実験概要

本実験においては、Web サイトデザイン支援システムへユーザがアクセスし、投票を行う。被験者は、20代の大学生17名である。また、ユーザは予め設けられた実験期間内であればいつでも投票可能とした。

表 7.2 に本実験の GA パラメータを示す。TBMP1 を用いた Web サイトデザイン支援システムは、各対戦において全被験者の 70% 以上、すなわち 13 名の被験者が投票を行えば、その対戦の投票を打ち切り、勝敗を決定する。これは、4.5.1 項における評価エージェントの投票参加割合の数値シミュレーションによる検証結果を受けてのものである。

また、本実験では、TBMP1 の比較システムとして、GVM を用いた Web サイトデザイン支援システムを用いる。GVM を用いたシステムにおいては、被験者 1 人あたりの持ち票は 15 票である。これは、TBMP1 における 1 世代あたりの対戦数が 15 対戦であり、1 人の被験者が 1 世代あたり最大 15 票を投票できることに起因している。GVM を用いた Web サイトデザイン支援システムでは、全被験者が 15 票を投じた場合、1 世代あたり 225 票獲得できる。しかし、TBMP1 では、各対戦の投票数を全被験者数の 70% で打ち切っているため、GVM においても 1 世代あたりの最大獲得投票数である 225 票の 70%、すなわち 195 票獲得できれば投票を打ち切り、解候補を評価し、GA 処理を行う。なお、比較システムの詳細は 7.3.2 項で述べる。

本実験では、どちらのシステムにおいても各対戦の投票数、世代進行に要した時間を測定した。また、実験終了後、被験者に対してアンケートを行った。アンケートにおいては、最終世代に生成された全解候補及び最終世代のエリート解候補に対する満足度、各システムにおける解評価のしやすさ及び各システムの使用における全体的な満足度を 5 段階評価で回答してもらった。さらに、実験中に気づいた知見についても回答してもらった。

表 7.2: 評価実験における GA パラメータ

Candidate solutions	16
Generation	5
Gene length	69bits
Crossover	Uniform crossover
Selection	Roulette selection + Elite preservation
Mutation rate	5%

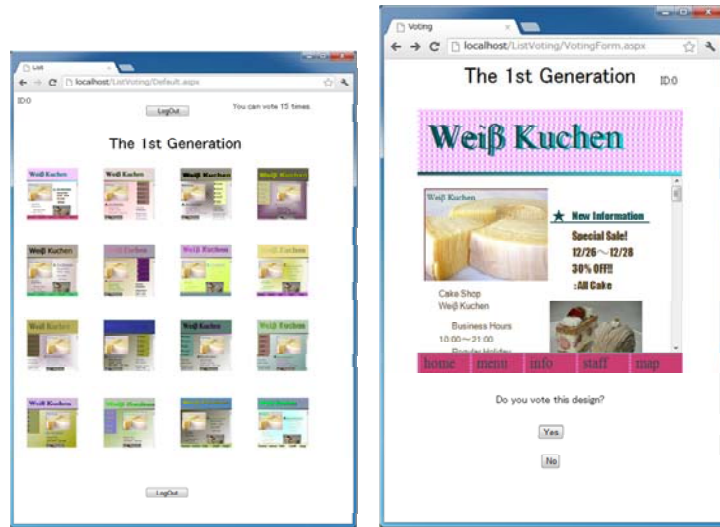
7.3.2 比較システム

本項では、GVM を用いた Web サイトデザイン支援システムの評価インタフェースについて述べる。

図 7.5 に GVM を用いた Web サイトデザイン支援システムの評価インタフェースを示す。TBMP1 と同様にユーザは、Web サイトデザイン支援システムにアクセスし、予め配布された ID 及びパスワードを使用し、ログインする。ログイン後、図 7.5(a) のように、GA 処理により生成された 16 個の Web サイトデザインが表示される。ユーザが投票を行う際には、図 7.5(a) 中の Web サイトデザインから、投票したい Web サイトデザインをクリックする。そして、図 7.5(b) の投票画面へ移行する。図 7.5(b) では、ユーザが選択した Web サイトデザインの拡大画像が表示される。ユーザが “Yes” をクリックした場合は、図 7.5(c) に示す投票確認画面へ移行し、投票処理が完了する。ユーザが “No” をクリックした場合は、再び図 7.5(a) の画面へ戻る。なお、TBMP1 と同様に、本システムにおいてもユーザが投票を行った後に、その時点での投票結果が分かるようにしている。

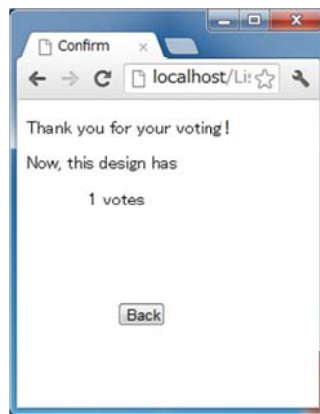
本システムでは、被験者 1 人あたりの持ち票は 15 票である。各被験者は、15 票の持ち票を 1 つの Web サイトデザインに全て投票することも、15 個の Web サイトデザインに 1 票ずつ投票することも可能である。

1 世代あたりの所定の投票数が獲得されれば、4.4.3 項の要領で解候補を評価し、GA 処理を行う。そして、新たな Web サイトデザインを生成し、第 2 世代目の投票を行う。GVM においては、これらの手順を繰り返し、多くのユーザの感性を反映した Web サイトデザインを生成する。



(a) 一覧表示画面

(b) 投票画面



(c) 投票確認画面

図 7.5: GVM を用いた Web サイトデザイン支援システムの評価インターフェース

7.4 実験結果

本節では、Web サイトデザイン支援システムを用いた TBMP1 及び GVM の評価実験の結果について考察する。

図 7.6 に各世代が終了するまでに要した時間の結果を示す。いずれの世代においても、GVM は TBMP1 より所要時間が短いことが確認された。しかし、GVM における 1 世代分の投票は、各ユーザが 1 回アクセスすれば充分であるのに対して、TBMP1 では、各ユーザは第 1 回戦、第 2 回戦、準決勝戦、決勝戦において投票しなければならず、最低でも 4 回アクセスする必要がある。すなわち、TBMP1 においては、1 世代の進行に要する時間が GVM の約 4 倍になると推測される。

しかし、TBMP1 と GVM の 1 世代あたりの進行に要した時間には、4 倍もの差は確認されなかった。さらに、2, 5 世代目においては、TBMP1 と GVM の所要時間は同等である。また、1 世代目開始から 5 世代目終了までに TBMP1 が要した時間は 1,807,946[s] (20[d]22[h]12[mi]26[s]), GVM が要した時間は 955,394[s] (11[d]1[h]23[mi]14[s]) であった。TBMP1 と GVM においては、5 世代目終了までに要した時間に約 2 倍の差があるものの、推測した程の差はないことが確認された。

図 7.7 に 5 世代目終了時点における被験者毎の投票数を示す。図 7.7 より、約半数の被験者が 60 票以上を投票していることが確認された。

図 7.8 に被験者に回答してもらったアンケート結果の平均値と標準偏差を示す。図 7.8 中の“Overall satisfaction”は、生成された Web サイトデザイン、優勝候補補、解評価のしやすさなどを総合した満足度である。優勝候補補に対する満足度は、TBMP1 の方が GVM より高いことが確認された。また、5 世代目の全解候補に対する満足度も、TBMP1 の方が GVM より高いことが確認された。そこで、TBMP1 と GVM のアンケート結果における統計的有意差を確認するため、ウィルコクソンの符号順位和検定を行った。表 7.3 にウィルコクソンの符号順位和検定の結果を示す。生成された解候補に対する満足度において、統計的有意差は確認されなかった。これは、図 7.7 より、TBMP1 及び GVM とともに約半数の被験者から 60 票以上を獲得できているためであると考えられる。そのため、TBMP1 及び GVM とともに満足度は 5 段階評価において 3 以上であり、両システムとも多くのユーザが満足のいく解候補を生成できると考えられる。

次に、図 7.8 中の各システムにおける解評価のしやすさの結果について考察する。解評価のしやすさは、TBMP1 の方が GVM より高いことが確認された。また、表 7.3 より、TBMP1 は GVM よりも有意水準 1% で解評価が行いやすいことが確認された。したがって、TBMP1 は GVM より、解評価のしやすさという点ではユーザの評価負担を

軽減できると考えられる。

また、図 7.8 中の各システムにおける全体的な満足度は、TBMP1の方がGVMより高いことが確認された。しかし、表 7.3 より、全体的な満足度においては統計的有意差は確認されなかった。

これらの結果より、TBMP1は、多くのユーザが満足のいく解候補を生成でき、GVMより解評価が行いやすいという点で有効であると考えられる。

しかし、被験者からは、TBMP1は、トーナメント対戦が進行しにくく、何度も進行状況を確認するのは負担が大きいという意見があった。これは、全被験者のうち70%の被験者の投票を獲得できなければ、トーナメント対戦を進行できないことに起因している。この問題は、例えば2点嗜好法や対戦結果推定の適用により、改善できると考えられる。

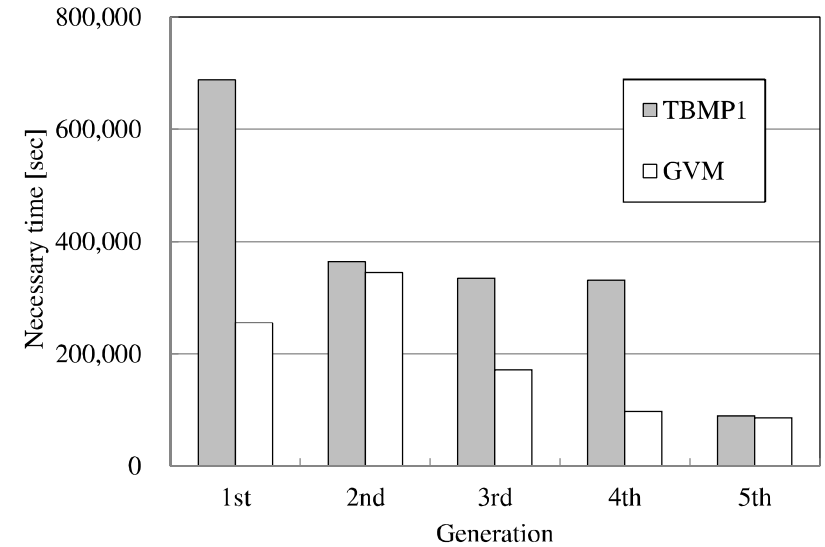


図 7.6: 各世代が終了するまでに要した時間

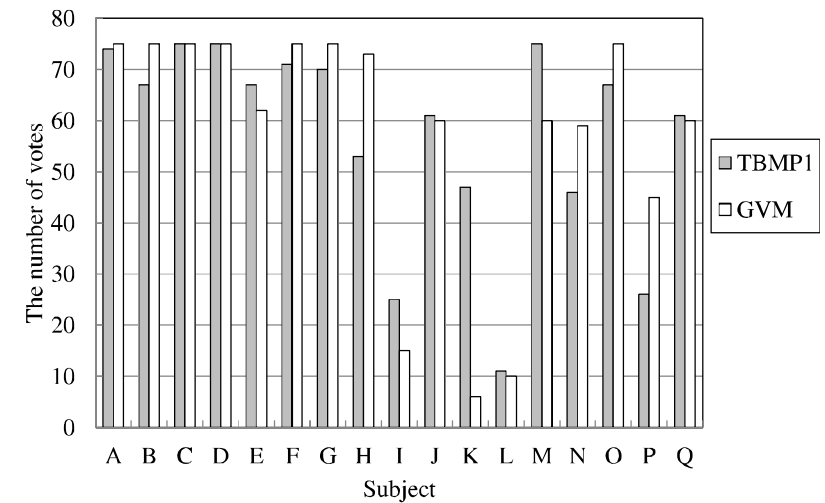


図 7.7: 被験者の投票数の結果

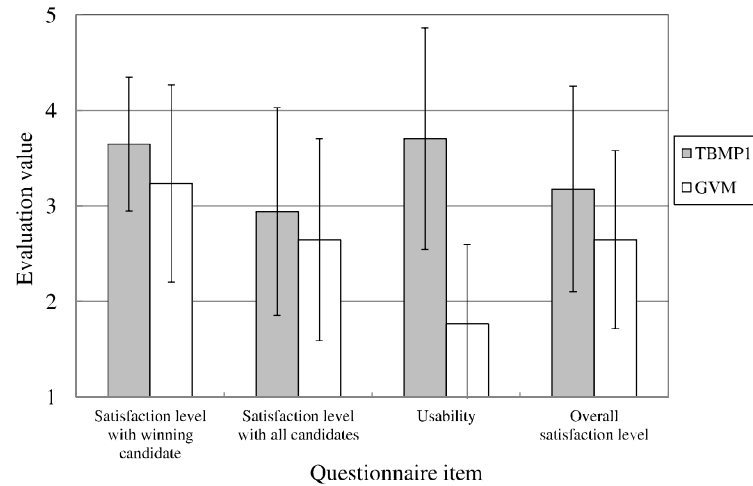


図 7.8: アンケートの結果

表 7.3: ウィルコクソンの符号順位検定の結果

Questionnaire item	P value	Significant difference
Satisfaction level with winning candidate	0.210	n.s.
Satisfaction level with all candidates	0.774	n.s.
Usability	0.004	** TBMP1
Overall satisfaction level	0.092	n.s.

n.s.: Not significant, *: $p < 0.05$ (Critical value: 5.99), **: $p < 0.01$ (Critical value: 9.21)

7.5 結言

本章では、TBMP1を用いた Web サイトデザイン支援システム及び本システムを用いた実ユーザを対象とした評価実験について述べた。本システムは、Web サーバ上で動作し、多くのユーザの投票を受け入れる。そのため、ユーザは各自の携帯端末やパソコンから本システムにアクセスし、投票を行うことができる。また、本システムにより生成される Web サイトデザインの総数は、ユーザの感性を反映するのに充分であると考えられる。

次に、本システムを用いた実ユーザを対象とした評価実験について述べた。本実験においては、GVMを用いた Web サイトデザイン支援システムを比較システムとして、TBMP1とGVMの性能比較を行った。実験結果より、TBMP1及びGVMともに多くのユーザが満足のいく解候補を生成できることが確認された。また、有意水準1%でTBMP1はGVMより解評価が行いやすいことが確認された。したがって、多くのユーザが満足のいくものを生成し、解評価が行いやすいという観点より、TBMP1は、GVMより有効であると考えられる。

第8章 結論

本論文では、単独及び複数のユーザの感性を考慮した官能評価インタフェースの開発が期待されている IEC において、ユーザビリティの向上を目的としたトーナメント式評価手法、ITS、TBMP についてまとめた。これまでの IEC システムでは、ユーザが解候補を 5 段階や 10 段階で評価するインタフェースが主であったため、ユーザの解評価負担が大きく実用化が困難であった。本論文では、IEC インタフェースのユーザビリティ、すなわち解評価のしやすさの向上の面より IEC ユーザの解評価負担軽減を実現するため、T2、T4、ITS、TBMP の評価インタフェースを提案し、評価エージェントを用いた数値シミュレーション及び実ユーザを対象とした評価実験により有効性を検証した。

第2章では、ユーザが単独で解候補を評価する場合を想定した評価インタフェースについて述べた。まず、IEC の基本アルゴリズム及び技術的特徴と課題をまとめ、先行研究を概括した。次に、1 個人のユーザの感性を用いることを想定したトーナメント式評価手法及び ITS について述べ、評価エージェントを用いた数値シミュレーションにより解候補の進化性能に関する有効性を検証した。数値シミュレーションの結果より、T2 及び T4 は、NIGA と同等の進化性能を有することが確認された。また、ITS は、NIGA を上回る進化性能を有することが確認された。T2、T4、ITS では、数値シミュレーションにより 10 段階評価において 8 点以上を獲得する解候補が得られており、ユーザが満足のいくものを生成できる可能性があることが示唆された。

第3章では、実ユーザによる IEC インタフェースのユーザビリティ評価について述べた。本章では、トーナメント式評価手法及び ITS において、実ユーザを対象とした評価実験を行い、実ユーザを対象とした場合の有効性を検証した。トーナメント式評価手法の評価実験の結果より、評価対象が時系列データ、すなわち音楽・動画の場合は、T2 は NIGA より解評価が行いやすく、ユーザの解評価負担軽減に有効であることが確認された。また、評価対象が静止画の場合においては、T2 及び T4 は、NIGA より解評価が行いやすく、ユーザの解評価負担軽減に有効であることが確認された。ITS の評価実験の結果より、ITS は、NIGA より解評価が行いやすく、ユーザの解評価負担軽減に有効であることが確認された。

第4章では、TBMP1について述べた。TBMP1は、多くのユーザの投票により、トーナメントの勝敗を決定する手法である。TBMP1においては、ユーザ1人1人は1票を投じるのみの解評価を行うため、ユーザの解評価負担を軽減できると考えられる。評価エージェントを用いた数値シミュレーションの結果より、全評価エージェントの約60%が投票に参加していれば、全評価エージェントが投票に参加した場合と同等の進化性能が得られることが確認された。

第5章では、TBMP2について述べた。TBMP1においては、実際のWebシステムを想定した場合、一定の投票数を獲得できなければ次の対戦に進めないという問題が想定される。そこで、2点嗜好法を適用することにより、トーナメント対戦の効率化を図った。評価エージェントを用いた数値シミュレーションの結果より、TBMP2はTBMP1と同等の絶対評価値を獲得できない評価エージェントの集団性のパターンが存在することが確認された。しかし、TBMP2においては、TBMP1より解評価に必要な投票数を約80%減少でき、投票開始からより早い段階で解候補の優劣判定を行える可能性があることが確認された。したがって、TBMP2はトーナメント対戦の効率化という観点より、有効性は高いと考えられる。

第6章では、TBMP3について述べた。トーナメント式評価手法においては、勝ち上がった解候補をユーザが複数回評価しなければならない。このため、ユーザの解評価に対するモチベーションが低下してしまうという問題が想定される。そこで、対戦結果推定の適用により、ユーザの解評価に対するモチベーションの維持を試みた。評価エージェントを用いた数値シミュレーションの結果より、対戦結果推定が実行された対戦のうち、約70%が正しく推定されていることが確認された。さらに、投票開始初期の進化性能において、TBMP3はTBMP1より絶対評価値が高いことが確認された。したがって、TBMP3は、ユーザの解評価に対するモチベーションの維持に有効であると考えられる。

第7章では、TBMP1を用いたWebサイトデザイン支援システム及び本システムを用いた実ユーザを対象とした評価実験について述べた。本システムは、Webサーバ上で動作し、多くのユーザの投票を受け入れる。そのため、ユーザはインターネットに接続できる端末より、投票に参加することができる。評価実験の結果より、TBMP1は、多くのユーザが満足のいくWebサイトデザインを生成でき、GVMより解評価が行いやすいことが確認された。したがって、多くのユーザが満足のいくものを生成し、解評価が行いやすいという観点より、TBMP1はGVMより有効であると考えられる。

本論文において、トーナメント式評価手法及びITSは、第2、3章の評価エージェントを用いた数値シミュレーション及び実ユーザを対象とした評価実験によりユーザの解

評価負担軽減に関する有効性が示された。これにより、IEC応用システムの主な評価対象である音楽・動画・静止画それぞれを対象とした場合に、ユーザの解評価負担軽減に有効な評価インタフェースが示された。今後は、トーナメント式評価手法及びITSを用いたIECシステムの実用化について考察する必要がある。TBMPにおいては、第4～7章の評価エージェントを用いた数値シミュレーション及び実ユーザを対象とした評価実験によりユーザの解評価負担軽減に関する有効性が示された。これにより、TBMPはIECにおいて多くのユーザの感性を利用したものを生成することに、解候補の進化性能及び解評価のしやすさの面で有効であることが示された。しかし、IEC応用システムにおける実用化のためには、実ユーザを対象とした評価実験により、さらに評価対象の検討や被験者の投票行動分析などを行い、より明確な有効性を検証する必要がある。

付録

A 評価値の一致率に関する実験

A.1 実験目的・手順

第2章では、NIGA, T2, T4の数値シミュレーションにおいて、10段階評価における8点が得られた。また、ITSの数値シミュレーションにおいても同様の結果となった。そこで、実ユーザーがある評価対象に対して、10段階評価において8点以上を与えた場合、それらの評価対象に満足できるかどうかを検証した。

本実験では、被験者に50枚の風景画像を10段階及び3段階評価により評価してもらい、これら2通りの段階評価の結果がどの程度一致するかを検証した。本実験における被験者は、20代の大学生20名である。具体的な実験手順は、以下のとおりである。

- (1) 被験者に風景画像を1枚ずつ提示し、10段階評価で好みの度合いを回答してもらう。
- (2) 全風景画像を(1)と異なる順序で再度提示し、3段階評価で好みの度合いを回答してもらう。

本実験において、(1)及び(2)の実行順序及び画像の提示順序は、被験者毎にランダムに設定した。また、被験者には、風景画像を直感で評価するように教示した。

図A.1に本実験における画像提示時間を示す。本実験における画像提示時間は、風景画像1枚あたり画像提示3[s]、白紙提示3[s]である。被験者は、風景画像1枚につき、画像及び白紙提示時間を合わせた6[s]以内に評価値を入力する。

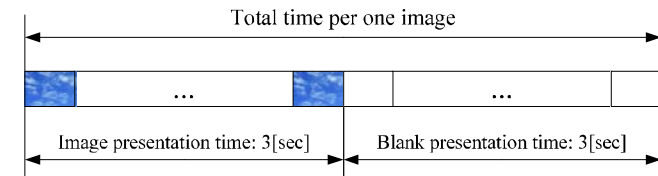
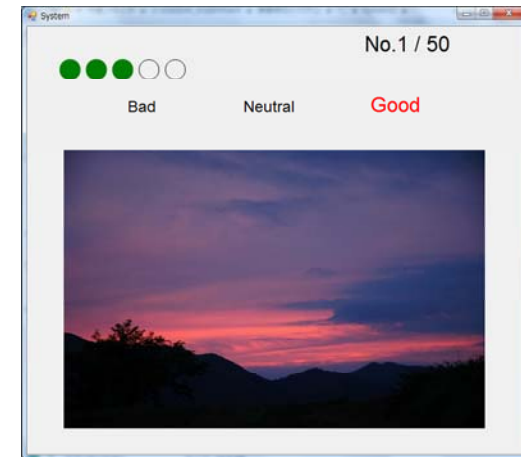


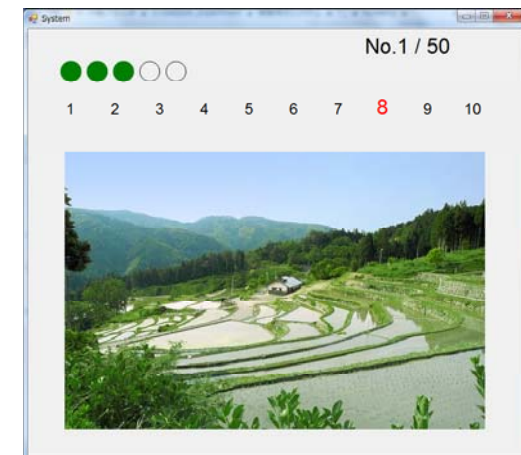
図 A.1: 画像提示時間

図 A.2 に本実験において使用した画像提示アプリケーションのインタフェースを示す。図 A.2 の左上の丸印は、1[s] 毎に左から点灯し、経過時間を示す。画像提示中は緑、白紙提示中は赤の丸印となる。なお、右上の “No.” 表示は、風景画像の提示枚数を示す。本アプリケーションにおいては、被験者はキーボードの数字キー、またはテンキーを使用し、評価値を入力する。

図 A.3 に評価値入力におけるキー割り当てを示す。10 段階評価を行う際は、数字キーである “1”～“0” キー、またはテンキーの “1”～“9” キー及び “+” キーを使用する。3 段階評価を行う際は、数字キー及びテンキーともに “1”～“3” キーを使用する。評価値が入力されれば、図 A.2 のインタフェース上で該当する評価値が強調される。入力した評価値のキャンセルは、“ESC” キーを押すことにより実行される。なお、誤った評価値を入力してしまった場合は、正しい評価値を再入力することで修正できる。また、評価値が入力されなかった風景画像及び評価値がキャンセルされた風景画像については、全風景画像を提示後、被験者に再度評価を求める。



(a) 3 段階評価



(b) 10 段階評価

図 A.2: 画像提示アプリケーションのインタフェース

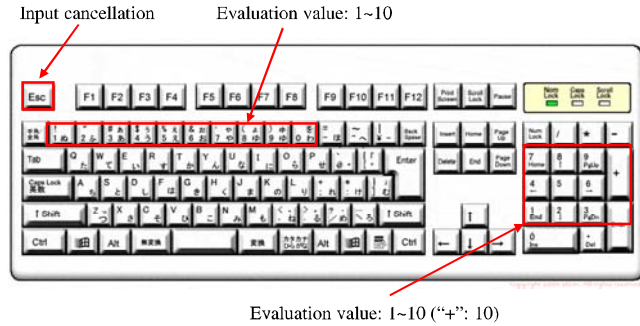
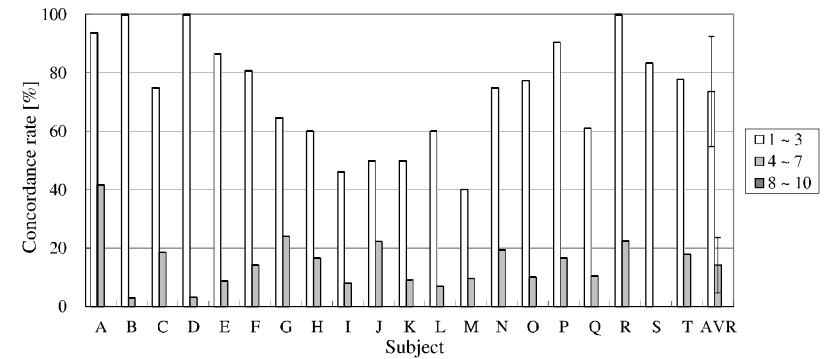


図 A.3: 評価値入力のキー割り当て

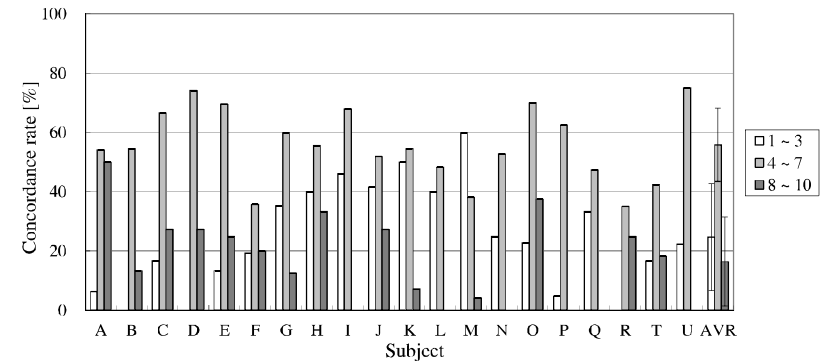
A.2 実験結果

図 A.4 に 10 段階評価値と 3 段階評価値の一致率 (Concordance rate) の結果を示す。図 A.4(a) は、10 段階評価値が 1~3 点、4~7 点、8~10 点と 3 段階評価値が 1 点の風景画像の一致率、図 A.4(b) は、10 段階評価値が 1~3 点、4~7 点、8~10 点と 3 段階評価値が 2 点の風景画像の一致率を示す。図 A.4(a) より、10 段階評価が 1~3 点の風景画像の約 75% に対して、3 段階評価において 1 点が与えられていることが確認された。図 A.4(b) より、10 段階評価が 4~7 点の風景画像の約 60% に対して、3 段階評価において 2 点が与えられていることが確認された。図 A.4(c) より、10 段階評価が 8~10 点の風景画像の約 80% に対して、3 段階評価において 3 点が与えられていることが確認された。

本実験では、図 A.4(c) における結果に着目する。図 A.4(c) より、本実験の被験者集団では、10 段階評価において 8~10 点が与えられた風景画像のうち、約 80% は満足のいくものであると考えられる。そこで、10 段階評価が 8~10 点の風景画像と 3 段階評価が 3 点の風景画像の一致率が、母集団においてはどの程度になるかを推定するため、信頼度 95% で t 推定を行った。その結果、母集団における一致率の推定範囲は、76.3~90.3% となった。したがって、本実験の被験者集団以外のユーザ集団においても、一致率は 76.3~90.3% となると推定される。



(a) 3 段階評価値 : 1 点



(b) 3 段階評価値 : 2 点

図 A.4: 一致率の結果

B 評価エージェントの集団性の表現方法

第4~6章の数値シミュレーションにおける評価エージェントの集団性は、GAを用いて生成される。表B.1に評価エージェントの集団性の生成におけるGAパラメータを示す。遺伝子長は、評価エージェントのビット数の5倍としている。図B.1に遺伝子列の構成を示す。遺伝子列は、各集団の評価エージェントのビット列を、全て足し合わせたものである。遺伝子列を評価する際の適応度 f は、式(B.1)で表現される。

$$f = \frac{100}{\sum_{i=0}^{group} \sum_{j=i+1}^{group} |h_{ij} - 0.4 \times L| + 1} \quad (B.1)$$

式(B.1)において、 h_{ij} は評価エージェント ij 間のハミング距離、 $group$ は集団数、 L は遺伝子長を示す。また、評価エージェントの集団間のハミング距離が評価エージェントのビット数の40%であるため、分子において L を0.4倍している。

これらの条件により、評価エージェントの集団性の生成を10,000回実行したときの成功率、すなわち想定した集団性を有する評価エージェントが生成された割合は、いずれの遺伝子長においても約65%であった。

表 B.1: 評価エージェントの集団性の生成における GA パラメータ

Gene row	Bit string
Gene length	The number of bits for a evaluation agent $\times 5$
Candidates	100
Generation	100
Selection	Roulette selection + Elite preservation
Crossover	Uniform crossover
Mutation rate	1/Gene length

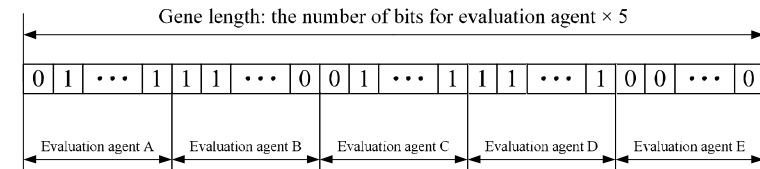
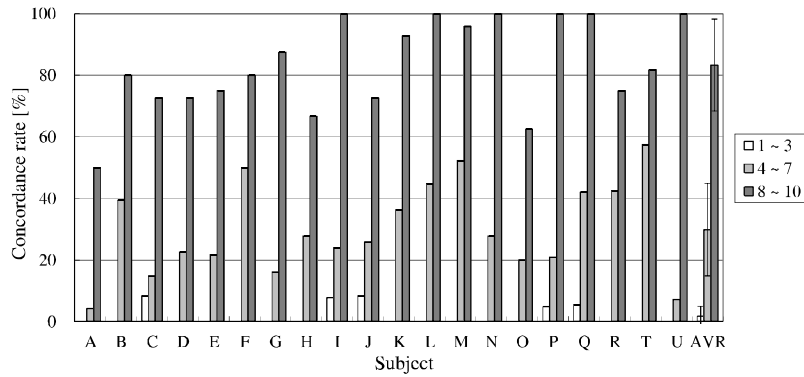


図 B.1: 遺伝子列の構成



(c) 3段階評価値: 3点

図 A.4: 一致率の結果

謝辞

本論文は、著者が関西大学工学部電子工学科 計算機システム工学研究室に配属された当初より、6年間に渡り行ってきた研究成果をまとめたものである。本論文をまとめるにあたり、関西大学大学院理工学研究科 計算機システム工学研究室 村中 徳明 教授には、指導教官として論文の綿密な御校閲並びに貴重な御教示を頂いた。ここに深謝の意を表す。

同研究室 徳丸 正孝 准教授 (現 感性情報システム研究室 准教授) には、研究の端緒より、また著者が同研究科博士課程前期課程在籍時には指導教官として終始変わらぬ御懇篤な御指導、御鞭撻を賜った。同准教授には、著者が同研究科博士課程後期課程進学後も共同研究者として、研究遂行にあたり貴重な御討論、御教示を数多く頂いた。ここに深甚なる感謝の意を表す。

感性情報システム研究室 院生 横山 貴洋氏には、第7章のWeb サイトデザイン支援システムの構築及び評価実験の遂行において、共同研究者として御協力頂いた。ここに感謝の意を表す。

感性情報システム研究室の諸氏には、研究遂行にあたり日頃より有益な御討論、御助言を頂いた。また、第3, 7章, 付録A の評価実験では被験者として御協力頂き、貴重な御指摘を頂いた。ここに感謝の意を表す。

計算機システム工学研究室の諸氏には、第3, 7章, 付録A の評価実験では被験者として御協力頂き、貴重な御指摘を頂いた。ここに感謝の意を表す。

また、関西大学大学院理工学研究科 梶川 嘉延 教授, 棟安 実治 教授には、論文の御校閲並びに貴重な御教示を頂いた。ここに感謝の意を表す。

最後に、著者の院生生活を手厚く支えて頂いた両親、妹に心から感謝の意を表す。

参考文献

- [1] 北野 宏明, 遺伝的アルゴリズム 4, 産業図書, 2000.
- [2] H. 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.
- [3] S.B. Cho and J.Y. Lee, "A Human-Oriented Image Retrieval System Using Interactive Genetic Algorithm," IEEE Transaction on Systems, Man, and Cybernetics part A: Systems and Humans, Vol.32, No.3, pp.452–458, 2002.
- [4] H.S. Kim and S.B. Cho, "Fashion Design Using Interactive Genetic Algorithm with knowledge-based Encoding," Yaochu Jin, Knowledge Incorporation In Evolutionary Computation, pp.411–434, 2005.
- [5] 鬼沢 武久, 緒形 友希, "インタラクシヨン服飾デザイン支援システム," 日本感性工学会論文誌, Vol.9, No.1, pp.25–32, 2009.
- [6] 井上 博行, 袁 丹, 岩谷 香栄, "対話型進化計算による配色支援システム," 知能と情報: 日本知能情報ファジィ学会誌, Vol.21, No.5, pp.757–767, 2009.
- [7] 三木 光範, 織田 博子, 菅原 麻衣子, 和氣 早苗, 森下 千春, 廣安 知之, "対話型遺伝的アルゴリズムを用いたサイン音系列生成システム," ヒューマンインタフェース学会論文誌, Vol.10, No.2, pp.243–251, 2008.
- [8] 前田 陽一郎, 宮下 滋, "対話型遺伝的アルゴリズムを用いたカオティック・インタラクティブ・サウンド生成システム," 知能と情報: 日本知能情報ファジィ学会誌, Vol.21, No.5, pp.768–781, 2009.
- [9] H. Inoue and H. Miyagoshi, "Behavior Evolution of Pet Robots with Human Interaction," Proceedings of the 2nd International Conference on Innovative Computing, Information and Control, pp.23–26, 2007.

- [10] 氏家 良樹, 奈良原 久之, 前野 隆司, 竹村 研治郎, 松岡 由幸, “多空間デザインモデルに基づくロボットの親近感生起動作生成法,” デザイン学研究, Vol.57, No.3, pp.35–42, 2010.
- [11] H. Takagi and M. Ohsaki, “Interactive Evolutionary Computation-Based Hearing Aid Fitting,” IEEE Transaction on Evolutionary Computation, Vol.11, No.3, pp.414–427, 2006.
- [12] 伴場 裕介, 小谷 淳司, 萩原 将文, “評価エージェントを用いた対話型進化計算手法によるインテリアレイアウト支援システム,” 情報処理学会論文誌, Vol.46, No.11, pp.2804–2813, 2005.
- [13] 長町 三生, 商品開発と感性, 海文堂, 2005.
- [14] I.S. Lim and D. Thalmann, “Tournament Selection for Browsing Temporal Signals,” ACM Symposium on Applied Computing, pp.570–573, 2000.
- [15] P.J. Angeline and J.B. Pollack, “Competitive Environments Evolve Better Solutions for Complex Tasks,” Proceedings of the 5th International Conference on Genetic Algorithms, pp.264–270, 1993.
- [16] Y. Watanabe, T. Yoshikawa and T. Furuhashi, “A Study on Application of Fitness Inference Method to PC-IGA,” 2007 IEEE Congress on Evolutionary Computation (CEC 2007), pp.1450–1455, 2007.
- [17] H. Takenouchi, M. Tokumaru and N. Muranaka, “Performance Evaluation of Interactive Evolutionary Computation with Tournament-Style Evaluation,” 2012 IEEE Congress on Evolutionary Computation in 2012 IEEE World Congress on Computational Intelligence (WCCI 2012), pp.193–200, 2012.
- [18] H. Takenouchi, M. Tokumaru and N. Muranaka, “Interactive Evolutionary Computation Using a Tabu Search Algorithm,” IEICE Transactions on Information and Systems, Vol.E96–D, No.3, pp.673–680, 2013.
- [19] F. Glover, “Tabu Search - Part I,” ORSA Journal on Computing, Vol.1, No.3, pp.190–206, 1989.
- [20] G.W. Klau, N. Lesh, J. Marks and M. Mitzenmacher, “Human-Guided Tabu Search,” Proceedings of the 18th National Conference on Artificial Intelligence, pp.41–47, 2002.
- [21] 宗本 晋作, 加藤 直樹, 今村 元一, “直交グラフ描画法を用いた室配置手法: タブ探索法を用いた対話型多目的最適化,” 日本建築学会計画系論文集, No.529, pp.279–286, 2000.
- [22] H. Takenouchi, M. Tokumaru and N. Muranaka, “Tournament Evaluation System Considering Multiple People’s Kansei Evaluation,” Journal of Kansei Engineering International, Vol.9, No.2, pp.43–50, 2010.
- [23] 竹之内 宏, 徳丸 正孝, 村中 徳明, “2点嗜好法を適用した複数参加型トーナメント方式を用いた対話型遺伝的アルゴリズム,” 知能と情報: 日本知能情報フェジ学会誌, Vol.23, No.1, pp.38–53, 2011.
- [24] H. Takenouchi, M. Tokumaru and N. Muranaka, “Tournament Evaluation System Applying Win-Lose Result Presumption Considering Multiple People’s Kansei Evaluation,” Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), Vol.16, No.3, pp.453–461, 2012.
- [25] T. Yokoyama, H. Takenouchi, M. Tokumaru and N. Muranaka, “Website Design System Based on an Interactive Genetic Algorithm Using Tournament Evaluation by Multiple People,” Joint 6th International Conference on Soft Computing and Intelligent Systems and 12th International Symposium on Advanced Intelligent Systems (SCIS&ISIS 2012), pp.2260–2263, 2012.
- [26] R. Dawkins, The Blind Watchmaker, Penguin, 1986.
- [27] T. Unemi, “SBArt4 for an Automatic Evolutionary Art,” 2012 IEEE Congress on Evolutionary Computation in 2012 IEEE World Congress on Computational Intelligence (WCCI 2012), pp.2014–2021, 2012.
- [28] 北野 宏明, 遺伝的アルゴリズム, 産業図書, 1993.
- [29] 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フェース改善方法の提案と評価,” 第12回フェジシステムシンポジウム講演論文集, pp.513–516, 1996.

- [30] H. Takagi and D. Pallez, "Paired Comparisons-based Interactive Differential Evolution," World Congress on Nature and Biologically Inspired Computing (NaBIC'09), pp.475-480, 2009.
- [31] 河野 高英, 堤 和敏, "オフィスビルのファサードを対象としたデザイン発想支援システムの開発 IDES(Interactive Differential Evolution with Score)の提案," 知能と情報: 日本知能情報ファジィ学会誌, Vol.23, No.4, pp.380-390, 2011.
- [32] J. Mádár, J. Abonyi and F. Szeifert, "Interactive Particle Swarm Optimization," Proceedings of the 5th IEEE International Conference on Intelligent Systems Design and Applications (ISDA'05), pp.314-319, 2005.
- [33] 長尾 光悦, 山本 雅人, 鈴木 恵二, 大内 東, "インタラクティブ GA に基づく画像検索システムの評価," 人工知能学会誌, Vol.13, No.5, pp.720-727, 1998.
- [34] 佐藤 義洋, 栢菅 彩, 有田 隆也, "適応度予測に基づく対話型進化計算とその似顔絵生成への応用," 第32回知能システムシンポジウム論文集, pp.199-204, 2005.
- [35] 大崎 美穂, 高木 英行, "対話型 EC 操作者の負担軽減-評価予測値による提示インタフェースの改善-, " 人工知能学会誌, Vol.13, No.5, pp.712-719, 1998.
- [36] 大崎 美穂, "音量の好みをを用いた評価値予測によるインタラクティブ EC フィッティングの改善手法," 知能と情報: 日本知能情報ファジィ学会誌, Vol.24, No.2, pp.680-690, 2012.
- [37] 渡辺 芳信, 吉川 大弘, 古橋 武, 大崎 美穂, "対話型進化計算における実評価数可変型評価値推論法の適用," 知能と情報: 日本知能情報ファジィ学会誌, Vol.20, No.5, pp.810-816, 2008.
- [38] 岩下 志乃, 王 上飛, 高木 英行, "評価尺度変換を用いたユーザ評価特性の学習効率化とその評価," 第21回ファジィシステムシンポジウム講演論文集, pp.610-613, 2005.
- [39] 徳丸 正孝, 村中 徳明, "対話型進化計算による感性検索エージェントの最適化," 日本感性工学会論文誌, Vol.8, No.3, pp.885-892, 2009.
- [40] 小野 智司, 中山 茂, "ユーザシステム協調型進化計算を用いた2次元コード装飾," 情報処理学会研究報告 数値モデル化と問題解決研究報告 (MPS), Vol.2012-MPS-87, No.5, pp.1-6, 2012.
- [41] M. Sugahara, M. Miki and T. Hiroyasu, "Design of Japanese Kimono using Interactive Genetic Algorithm," 2008 IEEE International Conference on Systems, Man and Cybernetics (SMC 2008), pp.185-190, 2008.
- [42] H. Nishino, T. Sueyoshi, T. Kagawa and K. Utsumiya, "An Interactive 3D Graphics Modeler Based on Simulated Human Immune System," Journal of multimedia, Vol.3, No.3, pp.51-60, 2008.
- [43] 伊藤 冬子, 廣安 知之, 三木 光範, 横内 久猛, "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, Vol.24, No.1, pp.127-135, 2009.
- [44] 雨宮 明日香, 三木 光範, 廣安 知之, "ユーザの嗜好に基づく初期個体生成を行う対話型遺伝的アルゴリズム," 同志社大学理工学研究報告, Vol.50, No.1, pp.34-45, 2009.
- [45] 山代 大輔, 吉川 大弘, 古橋 武, "可視化手法を用いた多目的最適化問題における満足解の選択支援," 知能と情報: 日本知能情報ファジィ学会誌, Vol.20, No.6, pp.850-859, 2008.
- [46] N. Hayashida and H. Takagi, "Visualized IEC: Interactive Evolutionary Computation with Multidimensional Data Visualization," 2000 IEEE International Conference on Industrial Electronics, Control, and Instrumentation (IECON 2000), Vol.4, pp.2738-2743, 2000.
- [47] B. Johanson, Automated Fitness Raters for the GP-Music System, University of Birmingham Masters Degree Final Project, 1997.
- [48] 石野 洋子, 寺野 隆雄, "模擬育種法と帰納学習を適用したマーケティング情報分析," 人工知能学会誌, Vol.12, No.1, pp.121-131, 1997.
- [49] A. Oliver, N. Monmarche and G. Venturini, "Interactive Design of Web Sites with a Genetic Algorithm," Proceedings of the International Association for Development of the Information Society (IADIS) International Conference WWW/Internet 2002, pp.355-362, 2002.
- [50] 鷺見 慎一郎, 老沼 樹里, 荒川 薫, 原島 博, "対話型進化計算を用いた顔画像美観システムのユーザビリティ改善とその評価," 電子情報通信学会技術研究報告 スマートインフォメディア研究会, Vol.111, No.457, pp.83-88, 2012.

- [51] 青木 直史, C言語ではじめる音のプログラミング-サウンドエフェクトの信号処理-, オーム社, 2008.
- [52] 田川 和正, 河村 廣, 谷 明勲, 滝澤 重志, “対話型進化計算法を用いた建築物内装デザインの合意形成に関する研究,” 日本建築学会近畿支部研究報告集 計画系, No.44, pp.301-304, 2004.
- [53] 三木 光範, 廣安 知之, 富岡 弘志, “並列分散対話型遺伝的アルゴリズムを用いた合意形成システムの有効性,” 人工知能学会論文誌, Vol.20, No.4, pp.289-296, 2005.
- [54] S. Henmi, S. Iwashita and H. Takagi, “Interactive Evolutionary Computation with Evaluation Characteristics of Multi-IEC Users,” 2006 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2006), pp.3475-3480, 2006.
- [55] J. Horn, N. Nafpliotis and D. E. Goldberg, “A Niche Pareto Genetic Algorithm for Multiobjective Optimization,” 1994 IEEE Conference on Evolutionary Computation in 1994 IEEE World Congress on Computational Intelligence (WCCI 1994), Vol.1, pp.82-87, 1994.
- [56] 天坂 格郎, 長沢 伸也, 官能評価の基礎と自動車における感性のエンジニアリングのために, 日本規格協会, 2000.

研究業績

【学術論文】

- [A-1] H. Takenouchi, M. Tokumaru and N. Muranaka, “Tournament Evaluation System Considering Multiple People’s Kansei Evaluation,” Journal of Kansei Engineering International, Vol.9, No.2, pp.43-50, 2010.
- [A-2] 竹之内 宏, 徳丸 正孝, 村中 徳明, “2点嗜好法を適用した複数参加型トーナメント方式を用いた対話型遺伝的アルゴリズム,” 知能と情報：日本知能情報フェジィ学会誌, Vol.23, No.1, pp.38-53, 2011.
- [A-3] H. Takenouchi, M. Tokumaru and N. Muranaka, “Tournament Evaluation System Applying Win-Lose Result Presumption Considering Multiple People’s Kansei Evaluation,” Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), Vol.16, No.3, pp.453-461, 2012.
- [A-4] H. Takenouchi, M. Tokumaru and N. Muranaka, “Interactive Evolutionary Computation Using a Tabu Search Algorithm,” IEICE Transactions on Information and Systems, Vol.E96-D, No.3, pp.673-680, 2013.

【国際会議】

- [B-1] H. Takenouchi, M. Tokumaru and N. Muranaka, “An interactive genetic algorithm with tournament evaluation of individuals by multiple people,” Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on Advanced Intelligent Systems (SCIS&ISIS 2008), pp.1061-1066, 2008.
- [B-2] H. Takenouchi, M. Tokumaru and N. Muranaka, “Interactive Genetic Algorithm with Tournament Evaluation Applying Win-Loss Result Presumption by Multiple People,” Joint 5th International Conference on Soft Computing and Intelligent Systems and 11th International Symposium on Advanced Intelligent Systems (SCIS&ISIS 2010), pp.687-692, 2010.

- [B-3] H. Takenouchi, T. Hirokata, M. Tokumaru and N. Muranaka, “Running Shoes Design System with Interactive Evolutionary Computation,” Proceedings of the International Conference on Kansai Engineering and Emotion Research 2012 (KEER 2012), pp. 925–932, 2012.
- [B-4] H. Takenouchi, M. Tokumaru and N. Muranaka, “Performance Evaluation of Interactive Evolutionary Computation with Tournament-Style Evaluation,” 2012 IEEE Congress on Evolutionary Computation in 2012 IEEE World Congress on Computational Intelligence (WCCI 2012), pp.193–200, 2012.
- [B-5] T. Yokoyama, H. Takenouchi, M. Tokumaru and N. Muranaka, “Website Design System Based on an Interactive Genetic Algorithm Using Tournament Evaluation by Multiple People,” Joint 6th International Conference on Soft Computing and Intelligent Systems and 12th International Symposium on Advanced Intelligent Systems (SCIS&ISIS 2012), pp.2260–2263, 2012.
- 【学会発表】**
- [C-1] 竹之内 宏, 徳丸 正孝, 村中 徳明, “複数参加型トーナメント式手法を用いた IGA の有効性に関する検討,” 第 13 回電子情報通信学会関西支部学生会研究発表講演会, D5-6, pp.102, 2008.
- [C-2] 竹之内 宏, 徳丸 正孝, 村中 徳明, “複数参加型トーナメント方式を用いた IGA の有効性に関する検討,” 第 10 回日本感性工学会大会予稿集, 21D-03, 2008.
- [C-3] 竹之内 宏, 徳丸 正孝, 村中 徳明, “複数参加型トーナメント方式を用いた IGA の有効性に関する検討,” 第 13 回曖昧な気持ちに挑むワークショップ講演論文集, pp.25–30, 2008.
- [C-4] 竹之内 宏, 徳丸 正孝, 村中 徳明, “2 点嗜好法を適用した複数参加型トーナメント方式を用いた IGA の有効性に関する検討,” 第 25 回ファジィシステムシンポジウム講演論文集, 2F3-01, 2009.
- [C-5] 竹之内 宏, 徳丸 正孝, 村中 徳明, “対戦結果推定を適用した複数参加型トーナメント方式を用いた対話型遺伝的アルゴリズム,” ヒューマンインタフェースシンポジウム 2010 講演論文集, pp.985–992, 2010.
- [C-6] 竹之内 宏, 徳丸 正孝, 村中 徳明, “対戦結果推定を適用した複数参加型トーナメント方式を用いた対話型遺伝的アルゴリズム ～集団性を変化させた検討～,” 平成 22 年電気関係学会関西連合大会講演論文集, pp.653–658, 2010.
- [C-7] 広方 竜也, 竹之内 宏, 徳丸 正孝, 村中 徳明, “対話型進化計算手法を用いたランニングシューズデザイン支援システム,” 第 27 回ファジィシステムシンポジウム講演論文集, P4-08, pp.1049–1054, 2011.
- [C-8] 井上 嵩, 竹之内 宏, 徳丸 正孝, “ポーカー方式を採用した IGA システムの有効性に関する検討,” 第 27 回ファジィシステムシンポジウム講演論文集, TD1-3, pp.607–612, 2011.
- [C-9] 竹之内 宏, 徳丸 正孝, 村中 徳明, “トーナメント式評価手法を用いた対話型進化計算の性能評価,” 2012 年電子情報通信学会総合大会, D-8-2, pp.116, 2012.
- [C-10] 横山 貴洋, 竹之内 宏, 徳丸 正孝, 村中 徳明, “複数参加型トーナメント方式を用いた IGA を組み込んだ Web サイトデザイン支援システム,” 第 14 回日本感性工学会大会予稿集, F3-7, 2012.