

GA によるヒューリスティック探索の最適化

仙石浩明[†] 田口順一[†]

sengoku@sdl.hitachi.co.jp

(株)日立製作所 システム開発研究所[†]

〒215 川崎市麻生区王禅寺 1099

与えられたヒューリスティック探索アルゴリズムに遺伝的アルゴリズム (Genetic Algorithm, GA) を組み込む汎用的手法を提案する。本手法では、探索木の分枝選択に優先度を定め、これを GA で最適化する。従来人手で解かれていたような問題を計算機で解く場合、人手による手順を手続き化して、ヒューリスティック探索アルゴリズムを作ることとは比較的容易である。しかし実問題に適用するには、より詳細にわたる知識の抽出が必須である。本手法に基づいて GA を組み込むことにより、少ない知識で質の高い解が得られる。さらに本手法はヒューリスティックの評価手段を提供する。ヒューリスティックを試行錯誤で変更しつつ本手法で評価することにより、より優れたヒューリスティックを開発する手段として利用することが可能である。

GA based Optimization of Heuristic Search

Hiroaki Sengoku[†] and Jun'ichi Taguchi[†]

Systems Development Laboratory[†],

Hitachi Ltd.

1099 Ohzenji Asao, Kawasaki, Kanagawa 215 JAPAN

We present a generic method adding GA (Genetic Algorithm) to a given heuristic search algorithm. In the method, GA is used to optimize the priority of branches in the search tree. Heuristic search algorithms are easy to develop and have been widely used to solve actual problems, which have been conventionally solved manually, but for practical applications, we have to add more detailed knowledge. The proposed method does not require so much knowledge due to using GA. We also present a measure of heuristic algorithms. The measure helps to improve heuristic algorithms by trial and error.

1 はじめに

遺伝的アルゴリズム (GA) は、生物進化の過程をモデル化した解探索アルゴリズムである。探索に関する知識が不要なため作成が容易、問題の諸条件が変化しても柔軟に適応できる、探索が大域的である、などの長所を持つ。

このような長所がある反面、探索効率の点でヒューリスティック探索などの従来手法より劣る場合が多い。GA では、解の探索に関する知識を明示的には用いないためである。そこで GA にヒューリスティクスを併用することにより探索の効率向上を図り、工学的応用を図る研究がさかんに行われている。

ところがいずれも問題領域に強く依存しており、タイプが異なる問題への応用は困難である。問題固有の知識を多用するあまり、交差方法を始めとする遺伝的操作が複雑化する傾向も見られる。さらに、効率向上に関する議論が各個別問題ごとに閉じており、他の GA 応用分野でそのノウハウを生かすことは難しい。

そこで、我々はヒューリスティック探索に GA を組み込む汎用的手法 [5] を提案した。与えられたヒューリスティック探索を、本手法で定義する標準型に変換できれば、容易に GA を組み込むことができる。

ヒューリスティック探索の部分は問題個別に作成する必要があるが、従来人手で解かれていたような問題に関しては、人手でどのように解いていたかを整理して、ルールあるいは手続きの形で表現することができれば、比較的容易に探索プログラムを開発することができる。そして本手法に基づき GA を組み込むことによって、少ない知識で質の高い解を得ることが可能になる。

ところがヒューリスティック探索の部分がすでに開発済みである場合、これを本手法で定義する標準型へ変換することが難しい場合がある。そこで今回、標準型の要件を緩和した手法を提案する。標準型への変換が容易になるため、より応用範囲が広がると期待される。

また、本提案手法を用いてヒューリスティックを評価する手法を提案する。与えられたヒューリスティックにどの程度改善余地があるか評価でき、より優れたヒューリスティックを開発する際の指針となる。

以下、2 章では、本論文であつかう標準型ヒューリスティック探索アルゴリズムの定義を行い、このアルゴリズムに GA を組み込む手法と本手法を用いてヒューリスティックを評価する手法を 3 章で提案する。4 章で巡回セールスマン問題 (TSP) を例にとり、提案手法を適用する。

2 標準型ヒューリスティック探索

重複を許すアルファベットの並び σ を解とする組み合わせ最適化問題を解く標準型ヒューリスティック探索アルゴリズムを定義する。この探索では、探索範囲を限定するため、最適解が得られるとは限らない。そこでこの探索範囲における最も良い解を「最良解」と呼ぶ。

定義 1 (標準型ヒューリスティック探索) 組合せ最適化問題 [5] $(\Sigma, D, \bar{f}, \underline{f})$ を解く次のアルゴリズムを標準型ヒューリスティック探索と定義する。ただし、 Σ はアルファベットの集合、 D は部分解の集合である。 $\bar{f}(\sigma)$ は、部分解 $\sigma \in D$ の末尾にアルファベットを接続していくことにより生成可能な解のうち、評価値が最小である解の評価値の上界を与え、 $\underline{f}(\sigma)$ は下界を与える。 $\bar{f}(\sigma) = \underline{f}(\sigma)$ を満たす σ を「解」と呼ぶ。

```

procedure main {
   $x \leftarrow \infty$ 
  search( $\Sigma, \lambda$ )
}

procedure search( $S, \sigma$ ) {
  if  $\bar{f}(\sigma) < x$  then {
     $x \leftarrow \bar{f}(\sigma)$ 
    if  $\bar{f}(\sigma) = \underline{f}(\sigma)$  then {
       $\sigma_t = \sigma$ 
      return
    }
  }
   $S \leftarrow \{s \in S \mid \sigma \cdot s \in D$ 
     $\wedge \underline{f}(\sigma \cdot s) < x\}$  ..... (1)
   $\{\forall s' \in T, \forall s \in (S - T), h(\sigma, s') < h(\sigma, s)$ 
    を満たす  $T \subset S$  を選択} ..... (*)
  for all  $s \in T$  do search( $S - \{s\}, \sigma \cdot s$ )
}

```

ただし、 x, σ_t は大域変数、 \cdot は連接演算子である。 σ_t に最良解が返される。

$h(\sigma, s)$ は、ヒューリスティックを用いて求めた $\underline{f}(\sigma \cdot s)$ の近似関数である。つまり部分解 $\sigma \cdot s$ から生成可能な解の中に、良い解が含まれると経験上判断する場合は $h(\sigma, s)$ の値が小さく、良い解が含まれないと判断する場合は $h(\sigma, s)$ の値が大きくなるような関数 $h(\sigma, s)$ を経験的知識に基づいて定める。□

$h(\sigma, s)$ を中間評価関数、その値を中間評価値と呼ぶ。 s を部分解 σ における選択枝と呼び、中間評価関数に基づいて選択した T を「選択枝集合」と呼ぶ。 T の要素数 $w = |T|$ を選択枝数と呼ぶ。

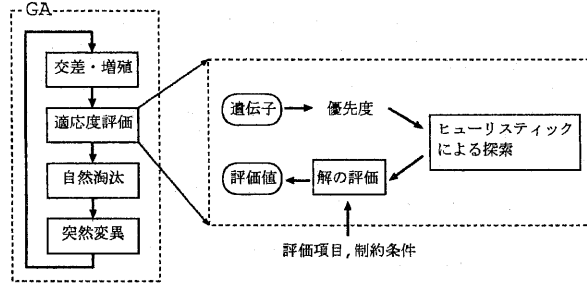


図 1: GA によるヒューリスティック探索の最適化

3 GA による標準型ヒューリスティック探索の最適化

前章で定義した標準型ヒューリスティック探索に GA を組み込み、選択枝集合の求め方を最適化する手法を提案する。最適化された探索で最良解を求めることにより、組合せ最適化問題 (Σ, D, \bar{f}, f) の準最適解を求めることができる。

アルゴリズムの概要を図 1 に示す。各個体の遺伝子から部分解 σ における選択枝 s それぞれの優先度 $g(\sigma, s)$ を求め、標準型ヒューリスティック探索において優先度を参照しながら最良解を生成する。そしてその最良解の評価値を個体の評価値とする。

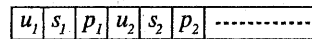
3.1 遺伝子表現法

中間評価値 $h(\sigma, s)$ の算出にあたって、支配的な項目を探索局面と呼び、 $u_{\sigma s}$ で表す。例えば中間評価値が、 σ の末尾のアルファベット $s_{|\sigma|}$ と選択枝 s のみに大きく依存するのであれば、 $u_{\sigma s} = s_{|\sigma|}$ とする。一般に、 $\forall s, h(\sigma, s) \sim h(\sigma', s) \Rightarrow u_{\sigma s} = u_{\sigma' s}$ が成立するように $u_{\sigma s}$ を定めれば良い。

遺伝子は、探索局面 u における選択枝 s の優先度 $p(u, s)$ をコーディングしたものである。

遺伝子列は図 2 に示す、三つ組 $t_i = (u_i, s_i, p_i)$ を要素とする可変長ベクトルである。要素数を $|g|$ で表す。三つ組 (u_i, s_i, p_i) は、探索局面 u_i における選択枝 $s_i \in \Sigma$ の優先度が p_i ($0 < p_i < 1$) であることを示す。

探索局面の間および選択枝の間に適当な全順序関係を導入するものとし、各三つ組は遺伝子列において常に辞書順に並んでいて、かつ探索局面と選択枝の双方が同一である三つ組は存在しないものとする。つまり、遺伝子列 $\{t_i\}$ は次の条件を満たす。



- u_i : 探索局面
- s_i : アルファベット
- p_i : 優先順位

図 2: 遺伝子列

- $\forall i < j, u_i \leq u_j$
- $\forall i < j, u_i = u_j \Rightarrow s_i < s_j$

遺伝子列 g から優先度 $p(u_j, s_k)$ を求めるアルゴリズムを次に示す。

```

for all  $(u_i, s_i, p_i) \in g$  do {
  if  $u_i = u_j \wedge s_i = s_k$  then {
     $p(u_j, s_k) \leftarrow p_i$ 
  }
  return
}
p  $\leftarrow$  rand(0, 1)
g  $\leftarrow$  g  $\cup$   $(u_j, s_k, p)$ 
 $p(u_j, s_k) \leftarrow p$ 

```

ただし、 $'p \leftarrow \text{rand}(0, 1)'$ は $0 < p < 1$ を満たす乱数を発生させる。 $'g \leftarrow g \cup (u_j, s_k, p)'$ は三つ組の集合(遺伝子列)に三つ組 (u_j, s_k, p) を加える。このように、探索局面と選択枝の新しい組合せが現れるたびに $|g|$ は大きくなる。

初期個体の遺伝子列の長さは 0、つまり空集合である。

以上より、部分解 $\sigma \cdot s$ の優先度 $g(\sigma, s)$ は常に求めることができ、 $g(\sigma, s) = p(u_{\sigma s}, s)$ である。

3.2 適応度評価

定義1の標準型ヒューリスティック探索のアルゴリズムにおいて、最後の部分

$$\begin{aligned} & \{ \forall s' \in T, \forall s \in (S - T), h(\sigma, s') < h(\sigma, s) \\ & \quad \text{を満たす } T \subset S \text{ を選択} \} \dots\dots\dots (*) \\ & \text{for all } s \in T \text{ do search}(S - \{s\}, \sigma \cdot s) \\ & \} \end{aligned}$$

を次のように書き換えることにより、優先度 $g(\sigma, s)$ に基づく探索を行う。

$$\begin{aligned} & \{ \forall s' \in T, \forall s \in (S - T), h(\sigma, s') < h(\sigma, s) \\ & \quad \wedge |T| = w \\ & \quad \text{を満たす } T \subset S \text{ を選択} \} \dots\dots\dots (*) \\ & \{ \forall s' \in T', \forall s \in (T - T'), q(\sigma, s') < q(\sigma, s) \\ & \quad \text{を満たす } T' \subset T \text{ を選択} \} \dots\dots\dots (**) \\ & \text{for all } s \in T' \text{ do search}(S - \{s\}, \sigma \cdot s) \\ & \} \end{aligned}$$

ただし、 $q(\sigma, s)$ は優先度と中間評価値の加重和

$$q(\sigma, s) = \beta g(\sigma, s) + (1 - \beta)h(\sigma, s) \dots\dots\dots (2)$$

である。 $q(\sigma, s)$ を選択枝関数と呼ぶ。係数 β ($0 \leq \beta \leq 1$) を GA 係数と呼ぶ。

つまり、中間評価関数 $h(\sigma, s)$ が小さい選択枝から順に w 個選び、そのなかから選択枝関数 $q(\sigma, s)$ が小さい選択枝を選び T' とする。 T' に含まれる選択枝のみが探索対象となる。

すなわち選択枝が中間評価値が上位 w 個に含まれさえすれば、 w 個の中で最悪の中間評価値であったとしても、優先度の値によっては選択対象となり得る。

w を選択枝数と呼ぶ。

以上の探索によって得られる最良解の評価値をその個体の評価値とし、評価値の逆数を適応度とする。

3.3 自然淘汰

個体集合から R 個 (定数) の個体を取り除く。このとき、個体集合の多様性を維持するために [1], 次のように行う。まず適応度が高い順に個体を並べる。適応度が高い方から順に個体の適応度を調べ、直前の個体の適応度との差が ϵ 以下である個体をすべて取り除く。ただし取り除く個体数は R 個までとする。取り除いた個体数を r とする。 $r < R$ の場合は、 $R - r$ 個の個体を適応度が低い方から順に取り除く。

3.4 突然変異

最優良個体を除くすべての個体の遺伝子列のすべての三つ組 $t_i = (u_i, s_i, p_i)$ を、確率 P (定数) で取り除く。

1つ以上の三つ組を取り除いた個体は、評価値および適応度を再計算する必要がある。

再計算の際、取り除いた三つ組 $t_i = (u_i, s_i, p_i)$ の優先度が必要であれば、3.1節で述べたように、優先度 p' がランダムに算出され、三つ組 (u_i, s_i, p') が遺伝子列に追加される。

3.5 交差・増殖

親個体の遺伝子列 $g_1 = t_1 t_2 \dots$, $g_2 = t'_1 t'_2 \dots$ に対して、 $0 \leq r \leq |g_1|$, $0 \leq r' \leq |g_2|$ を満たす乱数 $r, r' \in \mathbb{Z}$ (整数) を生成し、子個体の遺伝子列 $g = t_1 t_2 \dots t_r \cup t'_{r'+1} t'_{r'+2} \dots$ を作る。ただし、探索局面 u と選択枝 s が同一である三つ組 t_i, t'_j がある場合は、 t'_j を子個体の遺伝子列から取り除く。

さて、探索局面が u である三つ組がそれぞれの親個体の遺伝子列中 $t_i, t_{i+1}, \dots, t_j, t'_i, t'_{i+1}, \dots, t'_j$ であるとする。すると、 $r < i$ かつ $j' \leq r'$ ならば、子個体の遺伝子列には探索局面が u である三つ組は含まれない。つまりこの交差法は突然変異と同等の効果を合わせ持つ。

ランダムに R 組の親個体を選んで交差を行い、個体数を R 個増やす。自然淘汰において個体数が R 個減るので、世代間で個体数は一定である。

3.6 ヒューリスティックの評価

GA 係数 β は、共に値を大きくすると GA の比重が大きくなり、値を小さくするとヒューリスティックの比重が大きくなる。GA の比重を大きくしたときにより良い解が得られるのであれば、そのヒューリスティック探索には改善の余地があり、ヒューリスティックの比重を大きくしたときに良い解が得られるのであれば、改善の余地はあまり無い、あるいは十分に優れたヒューリスティック探索であると言える。

一方選択枝数 w は、探索範囲に最適解を含むほど十分大きければ、 w をさらに大きくしても解はそれ以上改善しない。したがって解の質がそれ以上向上しないときの w の値を調べれば、ヒューリスティックがどれだけ最適解の「近く」を探索しているか推定できる。 w の値が大きければ、あまり良いヒューリスティックではなく、特に $w \sim |\Sigma|$ であれば、そのヒューリスティックは不適切であると判断できる。

4 巡回セールスマン問題への適用

TSP は単純化された問題であり本論文が扱うような実問題ではないが、提案手法の効率を他手法と比較評価

表 1: アフリカ 96 都市 TSP

		$\beta = 0.3$				$\beta = 0.7$				
		回	平均	偏差	遺伝子長	回	平均	偏差	遺伝子長	
提	最遠	4	10	55328.63	108.93	372.37	6	55330.23	91.80	376.93
		12	23	55234.67	55.72	707.53	18	55266.47	83.52	733.93
		20	13	55272.60	67.61	855.60	14	55281.80	87.37	881.37
案	最近	4	0	56962.40	450.82	280.97	0	56457.50	392.11	270.50
		12	4	55758.07	416.93	440.93	27	55217.33	28.99	379.93
		20	0	55657.47	296.00	593.43	27	55219.43	41.24	504.17
法	最廉	4	0	57043.53	231.62	463.10	0	56197.67	234.23	270.90
		12	3	55767.40	289.24	741.33	13	55361.37	201.14	558.63
		20	4	55591.90	311.60	918.77	18	55297.67	181.19	781.03
従	局所	19	55238.00	38.52		試行回数: 30, 個体数: 100, 世代数: 500				
来	最遠	0	59222.93	1200.00						
法	最近	0	69827.41	738.59						
	最廉	0	69105.00	1228.88						

するため、例題として用いる。

定義 2 (TSP) n 個の都市の集合 $C = \{1, 2, \dots, n\}$ とそれらの間のコスト $d(i, j)$ ($i, j \in C$) が与えられたとき、すべての都市を重複なく訪れる経路 $O = c_1 c_2 \dots c_n$ ($c_i \in C$) のうち、コスト $\sum_{i=1}^n d(c_i, c_{i+1})$ (ただし $c_{n+1} \equiv c_1$ とする) が最小になるものを求める問題を TSP と定義する。 □

TSP の近似解法としてよく知られる手法に挿入法がある。挿入法のアルゴリズムを次に示す。

```

O ← { 適当に選んだ 1 都市からなる経路 }
S ← { c ∈ C | c ∉ O }
while S ≠ ∅ do {
    { c ∈ S を選択 } ..... (3)
    { c を O へ挿入 } ..... (4)
    S ← S - { c }
}

```

(4) において $O = c_1 c_2 \dots c_m$ ($m < n$) であるとすると、 c を挿入後の O' は次のように求める。まず、 $i = 1, 2, \dots, m$ それぞれについて、 c_i と c_{i+1} の間に c を挿入するときのコストの増分 $\Delta_i = d(c_i, c) + d(c, c_{i+1}) - d(c_i, c_{i+1})$ を求める。 Δ_i が最小となる i を選び、 $O' = c_1 c_2 \dots c_i c c_{i+1} \dots c_m$ を作る。

挿入法では、(3) の c の選び方が重要であり、次の 3 種のヒューリスティックが知られる。

最近挿入法: O と c との距離 $(d(c, t) \ (t \in O))$ の最小値が最小となる c を選ぶ。

最遠挿入法: O と c との距離が最大となる c を選ぶ。

最廉挿入法: コストの増分 Δ が最小となる c を選ぶ。

いずれの方法も探索枝は 1 本のみである。(3) の c の選択は、定義 1 の (*) に対応する。

したがって中間評価関数を、

最近挿入法:

$$h(O, c) = \frac{d(t, c) - \bar{d}}{6S_d} + \frac{1}{2}$$

最遠挿入法:

$$h(O, c) = -\frac{d(t, c) - \bar{d}}{6S_d} + \frac{1}{2}$$

最廉挿入法:

$$h(O, c) = \frac{d(t, c) + d(t', c) - d(t, t') - \bar{d}}{6S_d} + \frac{1}{2}$$

とおけば、2 章で定義した標準型ヒューリスティックと同じ形になる。

ただし、 \bar{d} は全ての二都市間の距離の平均、 S_d は全ての二都市間の距離の標準偏差である。最近挿入法、最遠挿入法においては、都市 $t \in O$ は、 $d(c, t)$ が最小となる都市、最廉挿入法においては、都市 $t, t' \in O$ は、 c を挿入したとき c に隣接する都市とする。

いずれの中間評価関数も、平均が $1/2$ 、標準偏差が $1/6$ になるように正規化している。また、本実験では簡単のため部分解から生成可能な評価値最小の解の上・下界の評価は行わない。

そして、3 章で提案した手法に基づき GA を組み込む。すなわちそれぞれのヒューリスティック探索に対し、3.2 節で述べた書き換えを行う。

(**) で選択する集合 T' の大きさは 1 とする。つまり選択枝関数 $q(\sigma, s)$ の値が最小の選択枝のみを探索する。

中間評価値を求める際、支配的なのは前述した都市 $t \in O$ であるから、探索局面 u として都市 t を用いる。したがって優先度 $g(O, c)$ は、 $p(t, c)$ であり、3.1 節で述べた方法により $p(t, c)$ を遺伝子にコーディングする。

TSPLIB[6] の中から、アフリカの96都市を巡る TSP(Africa-Subproblem of 666-city TSP) をベンチマークとして用い、選択枝数 $w = 4, 12, 20$ それぞれについて、GA 係数 $\beta = 0.3, 0.7$ それぞれの場合について、三つのヒューリスティック探索の最適化を行なった。

比較対象の従来法として、上記3ヒューリスティック(最近, 最遠, 最廉), および GA と局所探索を併用した高速解法 [1, 2] (局所) をとりあげる。なお、後者のプログラムは JAVA 言語で記述されており、WWW (World Wide Web) 上で実行可能 [3] である。

実験結果を表1に示す。3ヒューリスティック以外の手法は確率的求解法であるので、30回実験を行ない、最適解(経路長が55209[km]である解)を発見した回数(「回」の欄)と、解の評価値(経路長[km])の平均と標準偏差、および最終世代における最優良個体の遺伝子列の長さの平均をそれぞれ示した。3ヒューリスティックは、最初に選択する1都市を全ての場合について探索し、得られた解の評価値の平均および標準偏差を示した。

提案法では、個体数100, 淘汰個体数 $R = 30$, 突然変異率 $P = 0\%$ とし、500世代までに得られた最良解の評価値を示した。ただし、最適解を発見した場合はその時点で打ち切る。局所探索併用法では、提案法と同じ個体数・世代数とし、一世代毎に20個体に対し局所探索として2optによる解の逐次改善を試みる。

提案法は、ヒューリスティックのみの場合よりつねに良い解を求めることができた。また β と w のパラメータを適切に設定すれば、局所探索を併用した方法より高い頻度で最適解を発見できた。

β の値を大きくすると、最遠挿入法の場合のみ解の質が悪くなり、最近および最廉の場合は解の質が向上している。特に最近挿入法の場合は、 $\beta = 0.7$ とすることにより9割の確率で最適解を発見している。このことから最遠挿入法のヒューリスティックは改善の余地があまり無く、最近挿入法のヒューリスティックは大いに改善の余地があることが分かる。

w の値を大きくすると、最遠および最近の場合は $w = 12$ から $w = 20$ へ増加させても解の質は向上しないが、最廉挿入法の場合は $w = 20$ の時が一番解の質が良い。このことから最遠および最近に比べ、最廉挿入法のヒューリスティックは、この問題においては適切でな

いことが分かる。

5 おわりに

標準型ヒューリスティック探索を GA で最適化する手法を提案した。この手法を用いることにより、2章で定義した標準型に変換可能なヒューリスティックが知られている問題に、容易に GA を適用することができる。

TSP を例題として、具体的な適用方法を示し、良く知られた TSP の三つのヒューリスティック探索の最適化を行なった。

選択枝数および GA 係数を変化させたときの求解効率を比較することにより、ヒューリスティックの評価を行なった。

提案手法を実問題に適用することにより、枝刈りの調整を GA で自動化でき、開発工数の短縮や、プログラムが単純になることによる保守性の向上、などが期待される。

また、開発したヒューリスティックを提案手法を用いて評価することにより、ヒューリスティックを改善するにあたっての指針を得ることができる。

スケジューリング、レイアウト問題など、知識工学が応用されている分野は数多い。これらの分野に GA が応用できれば、詳細にわたる知識獲得が不要となる。本論文で提案した手法を用いることにより容易に GA の応用分野を拡大することができるものと期待される。

参考文献

- [1] 仙石, 吉原: 遺伝的アルゴリズムによる TSP の高速解法, 第46回情報処理学会全国大会論文集 8D-4 (1993).
- [2] Sengoku, H., Yoshihara, I.: *A Fast TSP Solver Using GA on JAVA*, Proc. of 3rd Int'l Symp. on Artificial Life and Robotics Vol.1 (1998)
- [3] Sengoku, H.: *A fast TSP solver using a genetic algorithm*, <http://www.hitachi.co.jp/Div/sdl/e-naiyo/e-seika22/TSP.html>
- [4] 仙石, 吉原, 今川: GAによるヒューリスティック探索の最適化 —バス仕業ダイヤの作成—, 情報処理学会数理解モデル化と問題解決研究会研究報告 95-MPS-2 (1995).
- [5] 仙石, 吉原: GAによるヒューリスティック探索の最適化, 情報処理学会 論文誌 Vol.37 No.10 (1996).
- [6] Reinelt, G.: *TSPLIB*, <http://www.crcp.rice.edu/softlib/tsplib/>