

# ナップザック問題における 評価関数設定に対する探索空間構造に関する考察

吉澤 大樹 橋本 周司

早稲田大学理工学部

## 1. はじめに

組合せ最適化問題などに対して確率探索法を用いる場合、目的関数にいわゆる騙し関数のような構造が存在したり、問題の制約条件を満たさない解が最適解近傍に存在したりするなどの理由で、最適解の近傍に探索点が集まるのが阻害され、効率的な探索が出来ない場合がある。

このようなとき、元の問題の目的関数を直接最適化するのではなく、別の評価関数を設定してそれを最適化する場合がある。たとえば、GAにおけるスケージング技法はその一例であり、SAにおける温度変化も評価関数の形状を変えているという見方が成り立つ。

本研究においては、ナップザック問題を例に評価関数がどのような統計的な構造を探索空間に与えるかを調べた。

これによって、確率探索法の適切な選択や調整が可能になると考えられる。

## 2. ナップザック問題

ナップザック問題は代表的な組合せ最適化問題の1つである。重量  $w_i$  と価値  $p_i$  をもつ  $n$  個の荷物 ( $i=1, 2, \dots, n$ ) から荷物をいくつか選び、荷物の重量の合計がナップザックの許容重量  $c$  を超えない範囲で合計の価値を最大化する問題である。

$p_j$  : profit of object  $j$ .  $j = 1, 2, \dots, n$ .

$w_j$  : weight of object  $j$ .

$c$  : capacity

$$x_j = \begin{cases} 1 & \text{if object } j \text{ is selected;} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

$$\text{Maximize } \sum_{j=1}^n p_j x_j, \text{ subject to } \sum_{j=1}^n w_j x_j \leq c. \quad (2)$$

## 3. ナップザック問題の目的関数の特徴 と適切とされる評価関数

ナップザック問題は、 $w_i, p_i$  が事前に与えられて

The Effect of Evaluation Functions  
on the Search Space Structure in Knapsack Problems

Hiroki Yoshizawa and Shuji Hashimoto  
Waseda University

いるので、 $w_i, p_i$  の平均や分散は既知である。ここで、 $w_i, p_i$  を 0 から 100 の一様分布から選び、多数問題を生成した場合の問題の構造について我々の従来の研究[5]を基に述べる。

式 1 より、解候補は  $(x_1, x_2, \dots, x_n)$  と表される。解候補  $(0, 0, \dots, 0)$  に対して合計価値 0 が保証されるので、許容重量を超過した場合を 0 以下の適当な数に設定することにより、ナップザック問題の目的関数  $g(x_1, x_2, \dots, x_n)$  は、

$$g(x_1, x_2, \dots, x_n) = \begin{cases} p_j x_j & \text{if } w_j x_j \leq c, \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

と表すことが出来る。ただし、荷物 20 個の問題 1000 題 ( $c = 500$ ) について、各解候補と最適解とのハミング距離とその解候補の目的関数値の相関係数を調べた結果が -0.18 であることから判るように、この関数を直接既存の確率探索法で最適化するのは必ずしも効率が良くない。このように相関が低い原因は、最適解近傍に許容重量を超過する解が多いためである。つまり、最適解に似た解は必ずしも高い評価を得ないので、GA などを用いても探索点が最適解周辺に集まりにくく、効率が悪いのである。

最適解近傍の許容重量超過解に適切な評価値を与えることにより、この問題は解決できる。

解候補  $(x_1, x_2, \dots, x_n)$  に対する評価値  $f(x_1, x_2, \dots, x_n)$  を、

$$f(x_1, x_2, \dots, x_n) = \begin{cases} p_j x_j & \text{if } w_j x_j \leq c, \\ p_j x_j - 2(w_j x_j - c) & \text{otherwise,} \end{cases} \quad (4)$$

として評価関数を設定すると、荷物 20 個の問題 1000 題 ( $c = 500$ ) について調べた各解候補と最適解とのハミング距離とその解候補の評価値の相関係数は、-0.56 となり、効率的な探索が可能になる。

ここで注意すべきは、式 4 の最大値を与える解は必ずしも最適解ではない点ということである。つまり、式 4 の下段で評価された解候補はどんなに高い評価値であっても、許容重量を超えているからナップザック問題の解とは言えないのである。したがって、それまでに得られた最良の解などを記憶する場合、それを更新しないように扱わなければならない。しかしながら、式 4 の評価関数は、最適解近傍に探索点を集めるという意味で式 3 に比べて有効である。

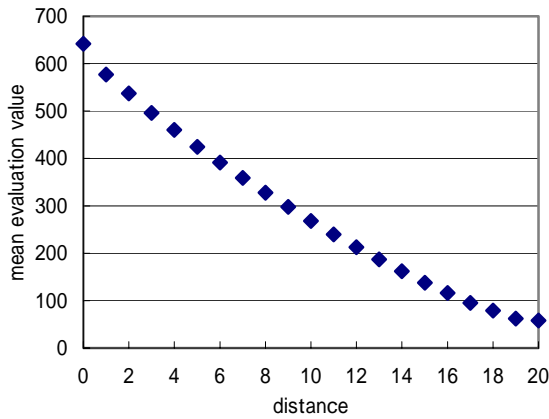


図 1. 最適解からの距離と平均評価値 (実験による)

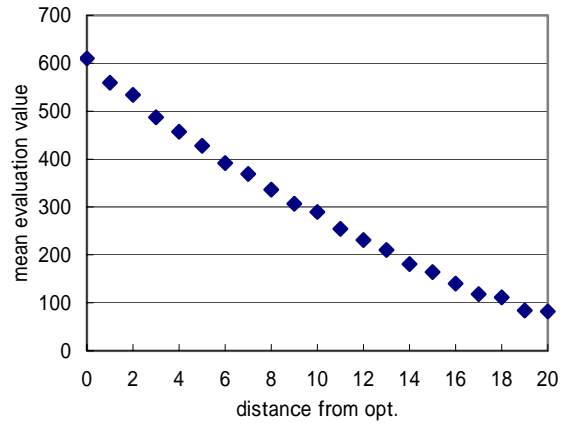


図 2. 最適解からの距離と平均評価値 (式 5 による構成)

図 1 は、荷物 20 個 1000 題の最適解からの距離ごとの平均評価値を表したものである。

#### 4. 評価関数がつくる探索空間構造

ここでは、式 4 を用いた場合、最適解の距離ごとの平均評価値がいくつになるか、モデルを用いて定量的に検討する。

$w_j, p_j$  の各値は、何らかの分布に従う乱数である。ナップザックの許容重量  $c$  と最適解の合計重量  $w_j x_j$  との差、 $c - w_j x_j$  を  $L$  とおく。また、最適解を表すバイナリ列  $(x_1, x_2, \dots, x_n)$  について、その '0' の個数を  $k$  とする。最適解においてナップザックに入らない荷物の平均価値を  $P_1$ 、平均荷重を  $W_1$ 、入る荷物の平均価値を  $P_2$ 、平均荷重と  $W_2$  とする。

このとき、最適解からの距離  $d$  の解を得るために、最適解に対して  $e$  個 '0' を '1' に換え、 $d-e$  個 '1' を '0' に換えると考える。 $e$  を可能な範囲で動かして、最適解からの距離  $d$  の解がもつ平均評価値の最適解からの変化量  $f(d)$  を得る。

$f(d)$  は、以下のように表される。

$$f(d) = \frac{\sum_{e=\text{Max}(0, d-N+k)}^{\text{Min}(d, k)} \left[ \binom{k}{e} C_e \cdot \binom{N-k}{d-e} C_{d-e} \{ e P_1 - (d-e) P_2 - h(d) \} \right]}{n(d)} \quad (5)$$

ただし、

$$h(d) = \begin{cases} 0 & \text{if } d < e - (L - e W_1) / W_2, \\ 2\{e W_1 - (d - e) W_2 - L\} & \text{otherwise,} \end{cases} \quad (6)$$

$$n(d) = \sum_{e=\text{Max}(0, d-N+k)}^{\text{Min}(d, k)} \binom{k}{e} C_e \cdot \binom{N-k}{d-e} C_{d-e} \quad (7)$$

実験条件を図 1 の場合と同じにするために、荷重と価値それぞれの総平均を 50、ナップザックの許容荷重  $c$  を 500 とした場合を具体的に計算した。 $P_2, W_2$  の値を変化させるにしたがって、 $d$  ごとの平均評価値のグラフ形状が変化した。 $P_2$  の値を 61、 $W_2$  を 39 とした場合、期待される  $k$  は 8 と考えられ、そのとき  $L$  は 32 となり、また  $P_1$  は 33.5、 $W_1$  は 66.5 となる。この条件下で式 5 をグラフに表したものが図 2 である。

#### 5. まとめと考察

ナップザック問題の評価関数として式 4 を用いた場合に、少なくとも特定のパラメータ (荷重・価値の分布) 下において既存の確率探索にとって有利な大域的相関構造が形成される。この構造は、モデル化したナップザック問題のパラメータ、すなわち、 $P_1, P_2, W_1, W_2, L, k$  によって理解できることが示された。

今後は、探索空間の局所構造、すなわち探索空間を大域的な山とノイズの和と見た場合のノイズの大きさがどのように決まるのかを調べるつもりである。

#### 参考文献

- [1] Manderick, Weger, Spiessens: The GA and the Structure of Fitness Landscape, 4th ICGA (1991).
- [2] Stadler, Schnabl: The landscape of the TSP, Physics Letters A, Vol.161, pp.337-344 (1992).
- [3] Boese, Kahng, Muddu: A New Adaptive Multi-start Technique for Combinatorial Global Optimization, Operations Research Letters, Vol.16 (1994).
- [4] 山田武士 and Colin, R.R.: フローショップスケジューリング問題の地形解析と遺伝的局所探索による解法, 情報処理学会論文誌, Vol.39, No.7, pp.2112(1998).
- [5] Yoshizawa, Hashimoto: Landscape analyses and global search of knapsack problems, In Proc. IEEE Intl. Conf. on Systems, Man and Cybernetics (2000).