第 1 章

探索とは何か？

1.1　探索は何故難しいのか

次のような問題を考えよう（図 1.1）.

> あなたは明日遠足を控えている．そのために荷物造りをしているとしよう．持って行きたい物はたくさんあって全部を持っていくことはできない．ナップザックが小さいのである．またあまりにも重すぎる荷物は持っていけない．このときどのような物を持っていけばいいのであろうか.

これは**ナップザック問題**（**knapsack problem**）と呼ばれる有名な問題である．日常生活でもこれとよく似た状況に頻繁にお目にかかる．ここではこの問題を例にして探索の難しさについて考えてみよう.

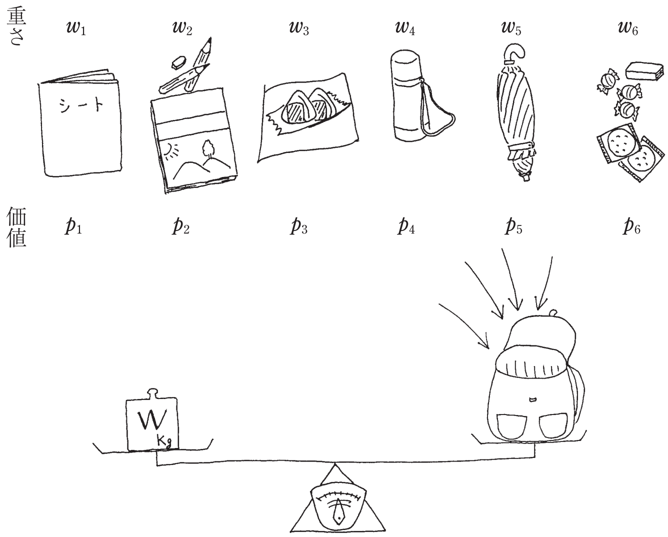

図 1.1　ナップザック問題のイメージ.

ナップザック問題を定式化すると次のようになる.

N 個の物があり各々の重さ w_i と価値 p_i が決まっている.またナップザックには重量制限 W があってこれより重い荷物は詰め込めない.このとき価値の和が最大になるように N 個の物から選んでナップザックに詰め込む方法を求めよ.

つまり,数学的にいうと

$$T \subseteq \{1, 2, \cdots, N\}, \tag{1.1}$$

$$\sum_{i \in T} w_i \leq W \tag{1.2}$$

を満たす $\{1, 2, \cdots, N\}$ の部分集合の中で $\sum_{i \in T} p_i$ を最大にする T を求めよ,ということになる.

この問題は N が小さいときは簡単に解ける.持っていく物の候補が 3 つ,バナナ,リンゴ,ミカンしかないときを考えよう.各々の価値と重さは次の通りとする.

$$w_{\text{バナナ}} = 2, \tag{1.3}$$

$$w_{\text{リンゴ}} = 1, \tag{1.4}$$

$$w_{\text{ミカン}} = 2, \tag{1.5}$$

$$p_{\text{バナナ}} = 2, \tag{1.6}$$

$$p_{\text{リンゴ}} = 2, \tag{1.7}$$

$$p_{\text{ミカン}} = 3. \tag{1.8}$$

重量制限 W は 3 とする.このとき答えは

$$\{\text{リンゴ}, \text{ミカン}\} \tag{1.9}$$

となるが,これを見出すのは難しくない.考えるべき組み合わせ(これを解候補と呼ぶ)は表 1.1 のようになる.ここで o(x) は候補に入っている(いない)ことを示す.重量合計が 3 以下の物しか解候補にはならない.この中で価値の合計が最大のものはリンゴとミカンという組み合わせであり,価値 5 をとることがわかる.

このような解き方は,すべての解候補を生成しているので**全解探索**あるいは**列挙法**(**enumeration method**)と呼ばれている.全解探索では必ず正解が得られるが,実用的には使える手法ではない.その理由は,実際的な問題では解候補が莫大な数になるからである.例えば上のナップザック問題で考えてみると,$N = 3$ 個に対して解候補は $2^3 = 8$ であった.一方,$N = 10$ 個の物があるときを考えると,$2^{10} = 1{,}024 \approx 10^3$ となる.$N = 100$ 個では

表 1.1　ナップザック問題.

バナナ	リンゴ	ミカン	重量合計	価値合計
x	x	x	0	0
x	x	o	2	3
x	o	x	1	2
x	o	o	3	5
o	x	x	2	2
o	x	o	4	5
o	o	x	3	4
o	o	o	5	7

図 1.2　欲張り法.

$2^{100} = 1,024^{10} \approx 10^{30}$ にもなる．どんな高速なコンピュータを使ったとしても 10^{30} 個の解候補を試すのは大変である．例えば 1 秒間に 10 万個の解を試せたとしても全体で $10^{30}/10^5 = 10^{25}$秒 $\approx 3.17 \times 10^{17}$ 年かかる．従ってより効率的な解法が必要となるのである．

効率的な解法の一つとして，**欲張り法**（**greedy algorithm**）が知られている（図 1.2）．まず，物の価値を重さで割った「重さあたりの価値」を求める．この値が大きい順に重さ制限を満たす限りナップザックに詰め込んでいくのである．例えば先の例では，表 1.2 のようになる．従ってまず価値/重さが一番大きいリンゴを詰める．このときナップザックの重さは 1 であり重量制限 3 をオーバーしない．次には価値/重さが 2 番目に大きいミカンを詰める．このときもナップザックの重さは $2 + 1 = 3$ となり重量制限をオーバーしない．しかし次の物（バナナ）を詰めるとオーバーするので，結局 { リンゴ，ミカン } が解として得られるのである．

これは価値/重さが大きい物から欲張って詰め込んでいるため欲張り法と呼ば

表 1.2　重さ当たりの価値.

	価値/重さ
バナナ	1
リンゴ	2
ミカン	1.5

表 1.3　ナップザック問題 (2).

	w	p	p/w
a	4.1	4.1	1
b	3	2.7	0.9
c	2	1.8	0.9

れている．この方法の優れている点は N が増えても使用可能なことである．N が 100 のときを考えてみよう．価値/重さを 100 個の物で計算しなくてはならない．さらに 100 個の価値/重さを順序付ける（ソートする）必要がある．その後はこの比が大きい物から重量制限を満たす限りナップザックに詰め込めばいい．全解探索の場合，N 個の物に対して 2^N の解を試す必要があった．これを計算量が $O(2^N)$ であるという（O はオーダーと読む）．これは計算の手間（比較や加減乗除などの計算数）が高々2^N 程度あることを意味する．O の厳密な定義は 4.1 節で述べる．これに対して欲張り法の計算量は，

$$N\text{ 個の}\frac{\text{価値}}{\text{重さ}}\text{計算} + N\text{ 個のソート} + \text{ナップザックへの詰め込み} \tag{1.10}$$

であり，

$$\begin{aligned} &O(N) + O(N\log N) + O(N) \\ &\quad = O(N + N\log N) = O(N\log N) \end{aligned} \tag{1.11}$$

であることがわかる（ソートの計算量については後述しよう）．本書では特に断らない限り log の底は 2 とする．明らかに計算の手間は (1.11) の方が少ない．$N = 100$ でも，

$$N\log N = 100 \cdot \log_2 100 \approx 667 \tag{1.12}$$

である．全解探索のときの 10^{30} とは文字通り桁が違う．

以上から欲張り法は効率的であると思われる．ではこの方法は万能であろうか．ここで表 1.3 のような問題を考えてみよう．この場合重量制限を $W = 6$ とする．さて欲張り法によると $\frac{p}{w}$ の一番大きい a が加えられる．すると残りの b も c も重量オーバーとなるので加えられない．従って欲張り法で求められる解は

$$\{\mathrm{a}\} \tag{1.13}$$

である．このときの価値の和は 4.1 となる．ところがこの場合の最適解は

$$\{\mathrm{b}, \mathrm{c}\} \tag{1.14}$$

であり，このときの価値の和は $2.7+1.8=4.5$ である．何故こうしたことが起こったのであろうか．それは加える候補を $a \Rightarrow b \Rightarrow c$ の順番に考えたからである．順番を b から先に考えればうまく {b, c} という解が得られるであろう．しかしこうした適切な順番は前もってわからないし，これを考えることこそナップザック問題の探索の本質である．従って欲張り法ではいつもうまくいくとは限らない．

直感的には欲張り法は納得できる戦略である．大体人間は目先の利益（この場合 p/w）が大きい物から考えていく．そしてある程度効率もよく，まあまあの解が得られる．しかしいつも最適であるとは限らないのである．つまり「欲張りは欲張りなりに損をする」ということだ．

先程の例で見たような正しい解

$$\{\mathrm{b}, \mathrm{c}\} \tag{1.15}$$

のことを**最適解**といい

$$\{\mathrm{a}\} \tag{1.16}$$

のようにそれなりの解だがもっと良い解が他にある場合を**局所解**と呼ぶ．局所解は，これをちょっと改善しただけ（ナップザックの例では他の物を加えるなど）では成績が良くならない．従って見たところ（局所的には）最適解であるが真の正解ではない．局所解の意味は次節でより厳密に定義する．

ナップザック問題の状況をもう一度考えてみよう．上の例は「明日遠足に行く」という状況を想定しており，考える時間は十分にあった．また物の重さと価値ははっきりとわかっていた．このような問題を

1. 標準形

と呼ぼう．日常生活では標準形の問題設定に様々なバリエーションがつくことがあろう．

例えば次のようなものが考えられる（図 1.3）．

2. 部分情報の問題
   物の価値や重さの一部がわからないことがある．自分の知らない物を母親が買って来たときを考えてみよう．また友達と遠足に行くのであるが，友達の価値基準（リンゴとミカンのどっちが好きかなど）がわからないとしてみたらどうであろうか．
3. 実時間対応型問題
   「明日の遠足」ではなく「目前に災難が迫ってきて避難しなくてはならない」という状況を考えよう．このときじっくりと長考して探索している暇

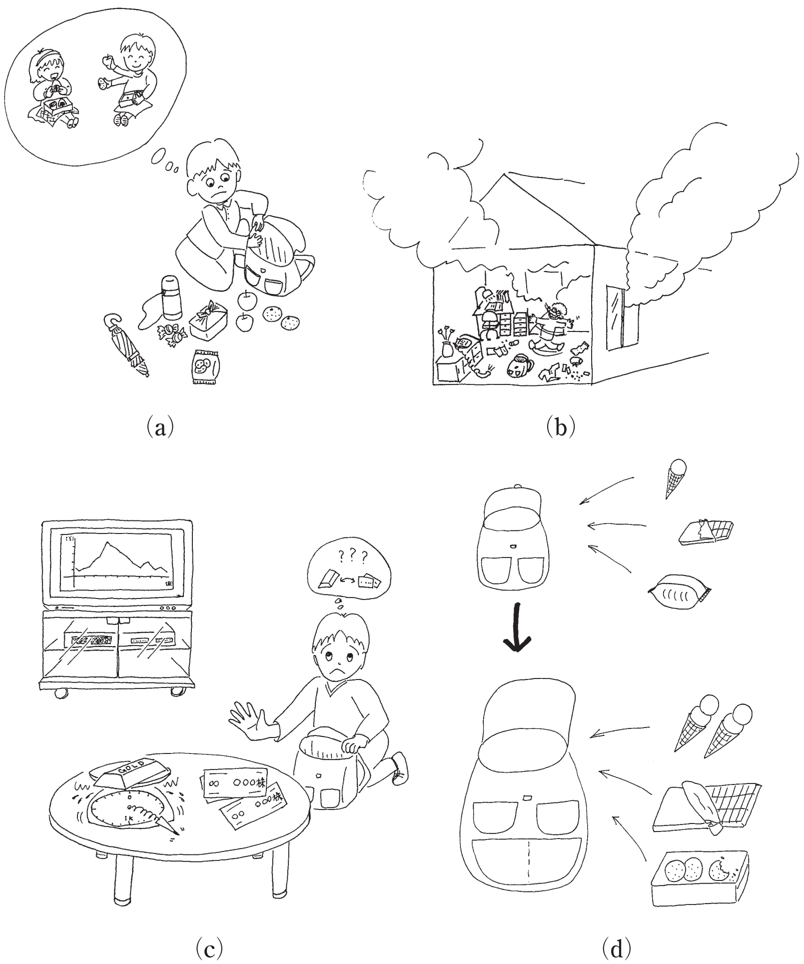

図 1.3　ナップザック問題のさまざまなバリエージション．

はない．とりあえずの解を出して満足しなくては逃げ遅れてしまう．全解探索はもちろんのこと，欲張り法をやっている時間もないかもしれない．限られた少ない時間をできるだけ有効に使うにはどうすればよいか．

4. ノイズのある問題

価値や重さがはっきりわからず誤差や変動があるとしよう．例えば重さを量るはかりがいい加減であるとか，価値が時価で変動する（金や株など）としたらどうなるであろうか．

5. 環境が動的な問題

4 と似ているが価値がある時点で急に変わったり，または重量制限が変わることを考える．例えばあるとき急に新しいナップザックが手に入り重量制限が 2 倍になったとしよう．このとき今まで考えている解（の一部）を使えるであろうか．または再び最初から考え直さなくてはならないのか．

探索の難しさは単に解を求めることのみにあるのではなく，求め方の計算の