

ソート (その4)

山本昌志*

2005 年 11 月 07 日

1 本日の学習内容

- 単純挿入ソート
- 2 分挿入ソート

2 単純挿入ソート

原理については，教科書に分かりやすくかかれていたので，それを使って説明する．説明するまでもなく，教科書を一読すれば，直ちに原理は理解できるであろう．

計算量のオーダーについて，簡単に説明する．左から i 個，整列が完了して， $i + 1$ 個目を整列させることを考える．

- 数列の $i + 1$ 番目が挿入されるべき位置を見つけるための比較の平均回数 (期待値) は， $i/2$ である．
- 位置が見つかった場合，それを挿入するために，配列は平均で $i/2$ の交換が必要である．

このことから，数列の $i + 1$ 番目を正しい位置に挿入するためには，合わせて i 回の計算¹が必要となる．大きさが N の数列全体では，

$$\begin{aligned}\sum_{i=1}^{N-1} i &= \frac{(N-1)(N-2)}{2} \\ &= \frac{1}{2}N^2 - \frac{3}{2}N + 1\end{aligned}\tag{1}$$

となる． N が大きい場合， $N^2/2$ が支配的で N^2 に比例して，計算量が増加する．このような場合，計算量のオーダーを表す記号として $O(N^2)$ と書くのは，以前述べたとおりである．

*独立行政法人 秋田工業高等専門学校 電気情報工学科

¹ここでは，比較と配列の交換を計算と言っている．

2.1 単純挿入ソートの様子

教科書のプログラム (List 1-7) を使った場合の単純挿入ソートの様子を図 1 に示す．ただし，教科書とは異なり，数列 (配列) の大きさは $N = 20$ としている．

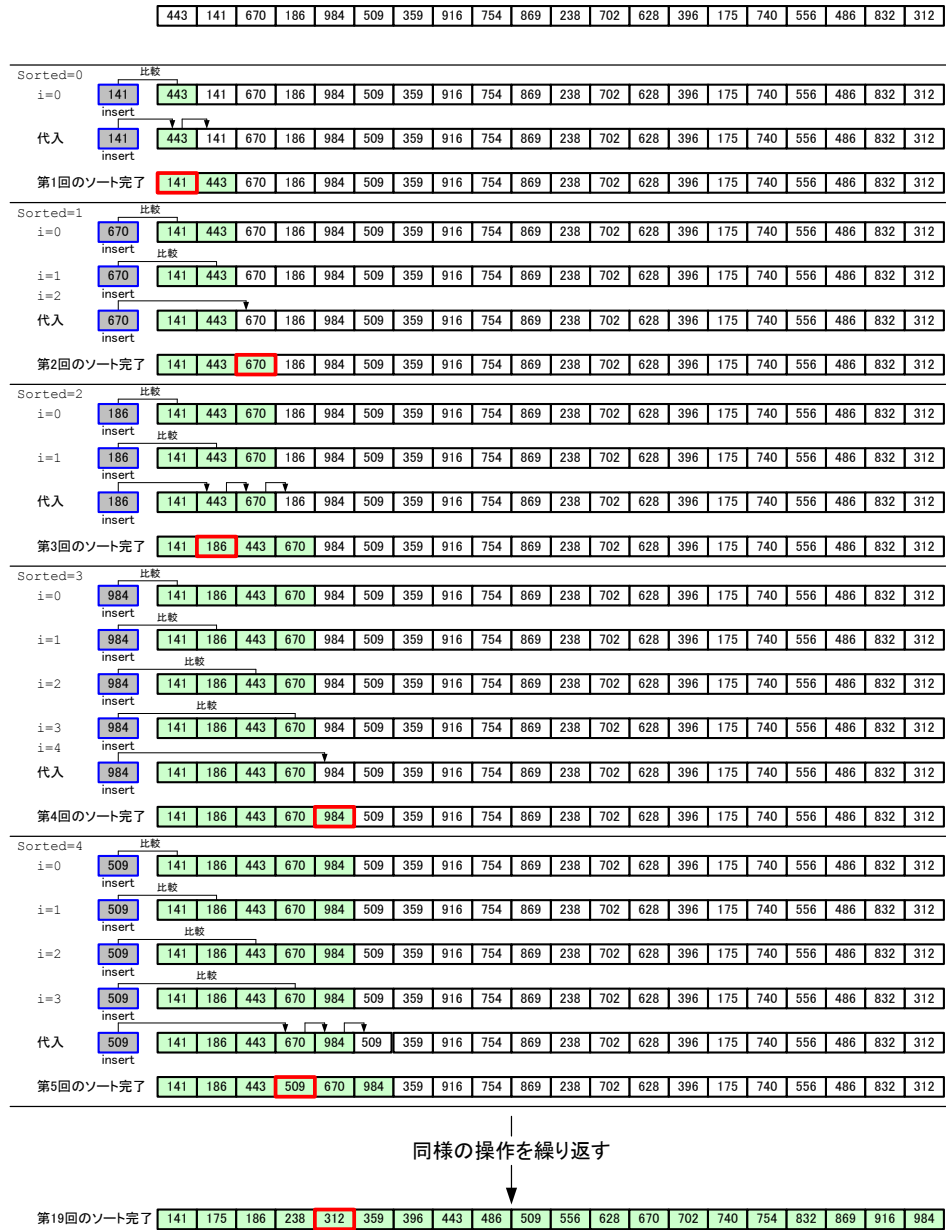


図 1: 単純挿入ソートの様子．

3 2分挿入ソート

3.1 原理

これも、教科書に分かりやすく原理が書かれているので、それを使って説明する。説明するまでもなく、教科書を一読すれば、これも直ちに理解できるであろう。

計算量のオーダーに関しては、教科書の説明は不十分である。先ほどと同様に、左から i 個、整列が完了して、 $i+1$ 個目を整列させることを考える。

- 数列の $n+1$ 番目が挿入されるべき位置を見つけるための比較の平均回数 (期待値) は、 $\log_2 i$ である。
- 位置が見つかった場合、それを挿入するために、配列は平均で $i/2$ の交換が必要である。

このことから、数列の $i+1$ 番目を正しい位置に挿入するためには、合わせて $i/2 + \log_2 i$ 回の計算が必要となる。大きさが N の数列全体では、

$$\sum_{i=1}^{N-1} \left(\frac{i}{2} + \log_2 i \right) = \frac{(N-1)(N-2)}{4} + \log_2(N-1)!$$

N が大きい場合のスターリングの近似 ($n! \simeq \sqrt{2\pi n} n^n e^{-n}$) を使うと

$$\begin{aligned} &\simeq \frac{1}{4}N^2 - \frac{3}{4}N + \frac{1}{2} + \log_2 \left[\sqrt{2\pi(n-1)} (n-1)^{(n-1)} e^{-n+1} \right] \\ &\simeq \frac{1}{4}N^2 - \frac{3}{4}N + \frac{1}{2} + \frac{1}{2} \log_2 [2\pi(n-1)] + (n-1) \log_2(n-1) + (1-n) \log_2 e \quad (2) \end{aligned}$$

となる。 N が大きい場合、 $N^2/4$ が支配的で N^2 に比例して、計算量が増加する。従って、計算量のオーダーは、 $O(N^2)$ である。

これまでの計算から、2分挿入ソートは、単純挿入ソートに比べて、2倍程度の速度の増加であることが分かる。

3.2 2分挿入ソートの様子

教科書のプログラム (List 1-8) を使った場合の2分挿入ソートの様子を図2に示す。ただし、教科書とは異なり、数列 (配列) の大きさは $N = 20$ としている。

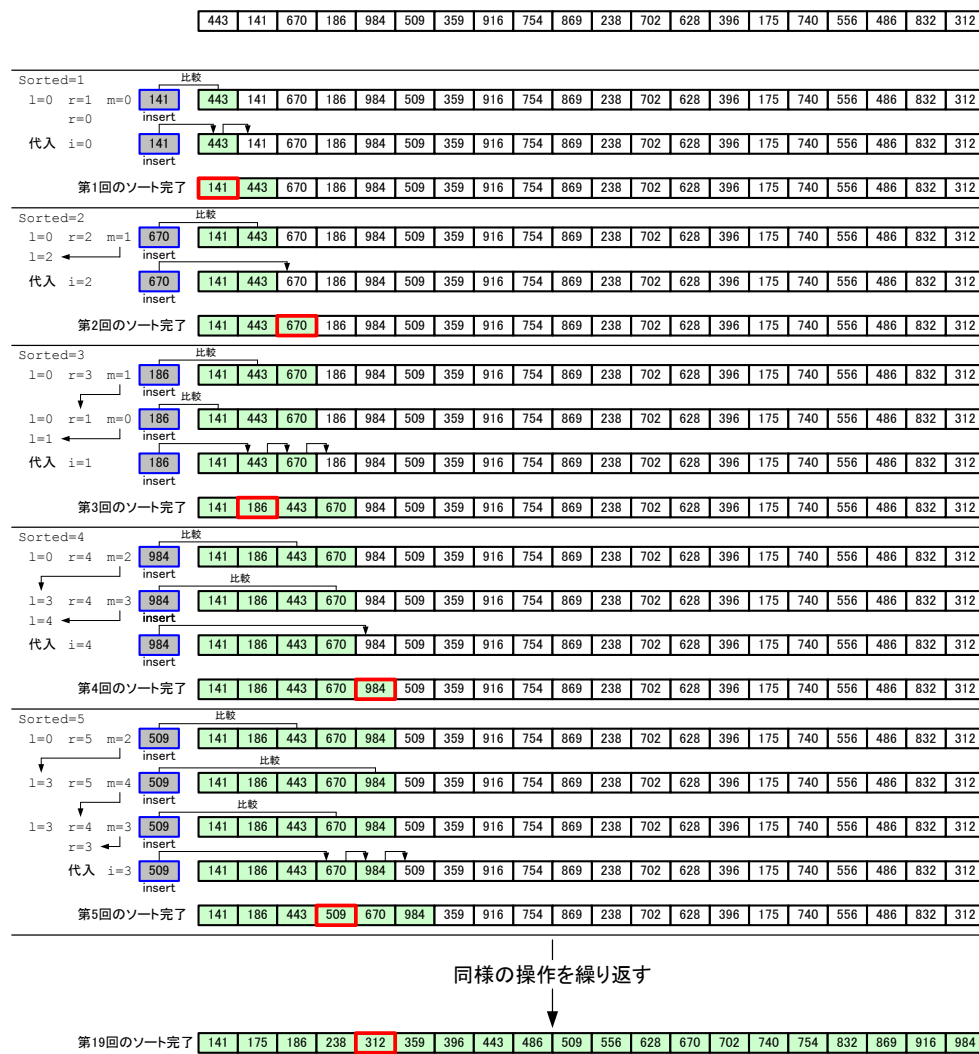


図 2: 2 分挿入ソートの様子 .

4 ソートのまとめ

コンピュータでは、大量のデータを取り扱う。その場合、データが規則どおりに並んでいることが重要である。規則どおりに並んでいると、データを探す時間を大幅に短縮できる。これは、人間が図書館であるタイトルの本を探すときや、英和辞典で単語を調べるときと同じである。

規則どおりにデータを並べることをソート (sort)² という。ここでは、整数のデータを小さい順 (昇順) に並べる方法を学習した。ここでは述べなかったが、逆に大きい順のことを降順と言う。講義では、以下のソートについて学習した。

²ソートリング、並べ替え、順序付けと言うこともある。

- バブルソート

数列の隣どうしの要素の大小を比較してそれらを交換しながらソートする方法である。交換が 1 回も生じなかったら、ソートが完了である。これは、小さい値のデータが泡 (バブル; bubble) のように浮かんで行くように見える³ことからこの名前がつけられた。

- クイックソート

ある一つの要素を基準とし基準値より大きいグループと小さいグループに分ける。分けられたグループも同様の方法で分ける。全てのグループの要素が 1 つになるまでこの作業を繰り返えし、このときソートが完了する。いろいろなソートの中で、平均的には最速であるから、この名前がついている。

- マージソート

最初に少ない個数の数列を並べ替えたグループを作る。次に、2 つのグループを併合 (マージ) して、より大きな並び替えられたグループを作る。これを繰り返すことにより、大きなグループができ、最終的には一つのグループになり並び替えが終了する。

- コームソート

バブルソートは隣との比較のため、小さい値は一つずつしか移動できない (昇順)。比較する間隔を広くして、一気に移動させる方法がコームソートである。このソートでは、適当な間隔でバブルソートを行い徐々にその間隔を狭める方法がとられる。実験から、間隔は $1/1.3$ ずつ小さくしていくのが良いと言われている。最終的には隣との比較 (間隔が 1) を行い、交換が起これなければソートの完了である。

- 単純挿入ソート

まず、最初の数列の 2 つを比較して、小さいほうを左に、大きいほうを右にする。次に数列の 3 番目を、その左にある 2 つの数列と小さいほうから順に比較し、収まる位置を探索して、挿入する。これを、4 番目、5 番目、... と順次、数列の最後まで繰り返す。最後の数列が挿入されたらソートが完了である。

- 2 分挿入ソート

単純挿入ソートでは、数列のある値を挿入する適当な位置は端から順に探している (リニアサーチ)、位置を探す数列は並べ替えられているので、真中の値から比較するバイナリサーチが可能で、それを適用して速度の向上を図った方法が、2 分挿入ソートである。

これら学習したソートの計算量と安定性については、表 1(教科書の Table 1-2) のとおりである。

³昇順にソートする場合。

表 1: ソートのアルゴリズムの特徴

アルゴリズム	計算量オーダー	安定性
バブルソート	$O(N^2)$	安定
クイックソート	$O(N \log N) \sim O(N^2)$	不安定
マージソート	$O(N \log N)$	安定
コームソート	$O(N \log N)$	不安定
単純挿入ソート	$O(N^2)$	安定
2 分挿入ソート	$O(N^2)$	安定

5 課題

5.1 課題内容

以下の数列のソートに関する問題である．これをそれぞれにソートで並び替えを行った場合，その様子を
示せ．この講義ノートの「ソートの様子」で示したように記述すれば良い．ただし，様子は手書きで書く
こと．

752 778 608 239 956 244 535 840 629 353 784 199 945 847 637 413 686 172 462 223

ただし，ソートする数列は出席番号により異なる．

- 学籍番号の下 1 桁が 1～9 の場合，その数から 10 個の数列をソートする．たとえば，下一桁が 3 の場
合は，

608 239 956 244 535 840 629 353 784 199

である．

- 学籍番号の下 1 桁が 0 の場合，その 10 番目から 10 個の以下の数列をソートする．

353 784 199 945 847 637 413 686 172 462

[問 1] 単純挿入ソートで昇順の場合

[問 2] 2 分挿入ソートで昇順の場合

5.2 レポート 提出要領

提出方法は、次の通りとする。

期限 11 月 14 日 (月) AM 10:40
用紙 A4
提出場所 山本研究室の入口のポスト
表紙 表紙を 1 枚つけて、以下の項目を分かりやすく記述すること。
授業科目名「情報工学」
課題名「課題 ソート (その 4)」
2E 学籍番号 氏名
提出日
内容 2 ページ以降に問いに対する答えを分かりやすく記述すること。