

情報工学実験II
ソーティングプログラム

イマム カイリ ルビス

2023年6月11日

目次

1	概要	2
1.1	ソートアルゴリズムとは	2
1.2	計算量	2
1.3	効率的なアルゴリズム	2
1.4	実行環境	2
1.5	対象データ	2
2	バケットソート	3
2.1	プログラム	3
2.2	動作	5
2.3	時間計算量	5
2.4	改良案	6
2.5	考察	6
3	挿入ソート	6
3.1	プログラム	6
3.2	動作	8
3.3	時間計算量	8
3.4	考察	9
4	バブルソート	9
4.1	プログラム	9
4.2	動作	11
4.3	時間計算量	11
4.4	改良案	12
4.5	考察	12
5	シェーカーソート	12
5.1	プログラム	12

1 概要

1.1 ソートアルゴリズムとは

ソートアルゴリズムとは、データの要素をある順序に並べるアルゴリズムである。最も頻繁に使用される順序は、数値順と辞書順で、昇順または降順のどちらかである。効率的なソートは、入力データがソートされたデータであることを必要とする他のアルゴリズム（検索やマージアルゴリズムなど）の効率を最適化するために重要である。また、人間が読みやすい出力を作成したりする際にもよく使われる。

1.2 計算量

アルゴリズムを実行するのにかかるコンピュータの時間を**時間計算量** (*time complexity*) と呼ばれる。時間計算量は、一般的に *big O notation*(オーダー) という書き方で記す。例えば、 $O(n)$, $O(n \log n)$, $O(2^n)$, など。

1.3 効率的なアルゴリズム

効率的なアルゴリズムとは、与えられた課題を最も早く解決するアルゴリズムであると考えられることができる。そこで、この実験では、データを数値順に並べるいくつかのソートアルゴリズムを比較し、ある種の数値データに対して、どのアルゴリズムが最も効率的かを調べる。本実験では各プログラムの実行時間は、ソート関数が呼び出される前後のシステム時間差を `clock()` 関数でカウントされる。

1.4 実行環境

本実験で使用される実行環境：

- プロセッサ：AMD Ryzen 5 5600X
- メモリー：16.0 GB
- OS：Windows 11 Pro
- コンパイラ：gcc

1.5 対象データ

この実験では、すべてのアルゴリズムが C 言語で記述されます。ショータニング対象データは、表 (1) に示すように、異なる特性を持っている。

表 1: ソーティング対象データ

データ	特徴
データ 1-3	乱数
データ 4	昇順
データ 5	降順
データ 6	バイトニック
データ 7	ジグザグ
データ 8	ランダムマイナス

2 バケットソート

バケットソートは、ソートされていない配列要素をバケットと呼ばれるいくつかのグループに分割するソートアルゴリズムです。各バケットは、適切なソートアルゴリズムを使用するか、同じバケットアルゴリズムを再帰的に適用することによってソートされます。

今回は、入力されたデータの値を配列の添字として、他の配列に格納することになります。そのため、データの個数を数える必要はなく、データの最小値と最大値を知ることが必要。

2.1 プログラム

以下はバケットソートのプログラムである。

```

1 #include <stdio.h>
2 #include <stdlib.h>
3 #include <string.h>
4 #include <time.h>
5
6 void setFileName(char *dst, int index) {
7     sprintf(dst, "data%d.dat", index);
8 }
9
10 void setOutFileName(char *dst, int index){
11     sprintf(dst, "bucket%d.dat", index);
12 }
13
14 void processTime(clock_t t) {
15     double time = ((double)t)/CLOCKS_PER_SEC;
16     printf("%.3lf ms\n", time*1000);
17 }
18
19 void bucketSort(FILE *in, int *p, int *sort, int offset, int n) {
20     int x;
21
22     rewind(in);
23
24     while (fscanf(in, "%d", &x) != EOF) {
25         p[x-offset]++;
26     }
27
28     int s = 0;
29
30     for (int i = 0; i <= n; i++) {
31         if (p[i] > 0)
32             for (int j = 0; j < p[i]; j++) {

```

```

33         sort[s] = i + offset;
34         s++;
35     }
36 }
37 }
38
39 int countLines(FILE *in) {
40     char c;
41     int count = 0;
42
43     do {
44         c = fgetc(in);
45         if(c == '\n') count++;
46     } while (c != EOF);
47
48     rewind(in);
49
50     return count;
51 }
52
53 void startSorting(int n) {
54     FILE *in;
55     FILE *out;
56
57
58     for (int i = 1; i <= n; i++) {
59         int x;
60         char *filename = malloc(10);
61         char *outname = malloc(10);
62
63         int max = 0;
64         int min = 0;
65
66         printf("data %d,", i);
67
68         setFileName(filename, i);
69
70         if((in = fopen(filename, "r"))== NULL)
71             printf("input file name wrong\n");
72
73         while(fscanf(in, "%d", &x) != EOF) {
74             max = (max < x) ? x : max;
75             min = (min > x) ? x : min;
76         }
77
78         rewind(in);
79         int lines = countLines(in);
80
81         int *p = calloc((max-min+1), sizeof(int));
82         int *s = malloc((lines) * sizeof(int));
83
84         clock_t t;
85
86         t = clock();
87         bucketSort(in, p, s, min, (max-min));
88         t = clock() - t;
89         processTime(t);
90
91         setOutFileName(outname, i);
92
93
94         out = fopen(outname, "w");
95
96         for (int i = 0; i < lines; i++) {
97             fprintf(out, "%d\n", s[i]);
98         }
99

```

```

100     free(p);
101     free(s);
102     free(filename);
103     free(outname);
104 }
105
106 fclose(in);
107 fclose(out);
108 }
109
110 int main() {
111     startSorting(8);
112     return 0;
113 }

```

2.2 動作

このアルゴリズムを動けるためには、 $(\max - \min + 1)$ の大きさの配列を用意する必要がある、この理由は後述する。

バケットソートのプログラムの主な流れは、以下の通りである：

1. バケット配列に $(\max - \min + 1)$ の大きさを設定する。
2. 出力配列に入力データの大きさを設定する。
3. 各入力データ x はバケット配列の添字 $(x - \min)$ 値を +1 増やす。
4. バケット配列から出力配列に元の値変更し、格納する。
5. ソート完了。

2.3 時間計算量

バケットソートの時間的複雑さは $O(n + k)$ であり、ここで n は要素数、 k はバケット配列の大きさである。

各データに対してバケットソートの実行時間は表 (2) のように表す。ここで表す結果は 10 回の実行結果の平均値である。

表 2: 計算時間 (バケットソート)

データ	data1	data2	data3	data4
計算時間 (ms)	7.2712	7.2169	7.1648	6.3128

データ	data5	data6	data7	data8
計算時間 (ms)	6.1445	5.9061	7.0154	8.6275

2.4 改良案

入力データをバケット配列の添字に変換するだけでは、マイナスの値に対応できなくなる。マイナスの配列添字はそもそもないからだ。そこで、バケット配列の大きさが $(max - min + 1)$ に設定しないといけない。入力データがすべて同じ数である場合、 $x - x = 0$ となり、大きさ 0 の配列はデータを格納できないため、このような場合に +1 が不可欠。

表 3: 改良結果

問題	元のアルゴリズム	改良したアルゴリズム
マイナスの対応	×	○
同じ値データ	○	○
バケット配列の大きさ	大きい	小さい

2.5 考察

実行結果から、バケットソートはどのような種類のデータでもうまく動作すると結論付けられる。その差は $1ms$ 程度であり、あまりに速すぎて認識できない。

残りの問題は、小数が持っている数字をソートしようとした場合、小数は配列のインデックスになり得ないということだ。そこで、このアルゴリズムの実装方法について、別の考え方を考えなければならない。

3 挿入ソート

挿入ソートは、トランプを並べ替えるのと同じような仕組みです。配列は事実上、ソートされた部分とソートされていない部分に分けられます。ソートされていない部分から値が選ばれ、ソートされた部分の正しい位置に配置される。

今回は、全部の入力データをソートするでなく、入力データを 1 個ずつ取ってからソートする。

3.1 プログラム

以下は挿入ソートのプログラムである。

```
1 #include <stdio.h>
2 #include <stdlib.h>
3 #include <time.h>
4
5 int countLines(FILE *in) {
6     char c;
7     int count = 0;
8
9     do {
10         c = fgetc(in);
11         if(c == '\n') count++;
12     } while (c != EOF);
13
14     rewind(in);
```

```

15     return count;
16 }
17
18
19 void store_array(FILE *in, int *p) {
20     int x;
21     int i = 0;
22
23     while (fscanf(in, "%d", &x) != EOF) {
24         p[i] = x;
25         i++;
26     }
27 }
28
29 void processTime(clock_t t) {
30     double time = ((double)t)/CLOCKS_PER_SEC;
31     // printf("%.3lf ms\n", time*1000);
32     printf("%.3lf\n", time*1000);
33 }
34
35 void insertionSort(FILE *in, int lines, int *p) {
36     int key;
37     int i = 0;
38
39     while (fscanf(in, "%d", &key) != EOF) {
40         p[i] = key;
41         int j = i - 1;
42         while(j >= 0 && p[j] > key) {
43             p[j+1] = p[j];
44             j--;
45         }
46         p[j+1] = key;
47         i ++;
48     }
49 }
50
51 void setFileName(char *dst, int index) {
52     sprintf(dst, "data%d.dat", index);
53 }
54
55 void setOutFileName(char *dst, int index){
56     sprintf(dst, "insert%d.dat", index);
57 }
58
59 void startSorting(int n) {
60     FILE *in;
61     FILE *out;
62
63     for (int i = 1; i <= n; i++) {
64         char *filename = malloc(10);
65         char *outname = malloc(10);
66
67         // printf("%d, ", i);
68
69         setFileName(filename, i);
70
71         if((in = fopen(filename, "r")) == NULL)
72             printf("input file name error\n");
73
74         int lines = countLines(in);
75         int *p = malloc(lines * sizeof(int));
76
77         clock_t t;
78         t = clock();
79         insertionSort(in, lines, p);
80         t = clock() - t;
81         processTime(t);

```



```

82         setOutFileName(outname, i);
83
84         out = fopen(outname, "w");
85
86         for (int i = 0; i < lines; i++) {
87             fprintf(out, "%d\n", p[i]);
88         }
89
90         free(p);
91         free(filename);
92         free(outname);
93     }
94
95     fclose(in);
96     fclose(out);
97 }
98
99
100 int main(void) {
101     startSorting(8);
102     return 0;
103 }

```

3.2 動作

挿入ソートは比較によりソーティングを行うアルゴリズムである。配列に格納されるデータは、1 個前の要素から最初の要素までに比較し、適切な位置に置くこと。しかし、適切な位置に置くという関数がないため、以下の通りである：

1. 入力データ x が n 番目の要素に入力する。
2. $(n - 1)$ 番目から 0 番目の要素までに比較する。
3. 左隣の要素が着目している要素より大きければ、着目している要素の位置に代入する。
4. 全部ソートされるまで繰り返す。
5. ソート完了。

3.3 時間計算量

挿入ソートの計算量は場合によって分れる。

- 最悪の計算量： $O(N^2)$
- 平均的な計算量： $O(N^2)$
- 最良の計算量： $O(N)$

各データに対してバケットソートの実行時間は表 (4) のように表す。ここで表す結果は 10 回の実行結果の平均値である。

表 4: 計算時間 (バケットソート)

データ	data1	data2	data3	data4
計算時間 (ms)	3568.585	3566.8257	3569.5971	5.7314

データ	data5	data6	data7	data8
計算時間 (ms)	7068.0738	3536.4216	3554.4499	3535.3692

3.4 考察

実行結果を見ると data4 のソートは最速で、5.7314ms ですべてのデータをソートすることができました。挿入ソートは、データがソートされているかどうかを検出することができると思います。そのため、すでにソートされたデータに対しては、アルゴリズムが代入処理を行うことはない。

最も時間がかかるのは、データ 5(降順) である。これは、データを昇順にソートするために、data5 が最も処理時間がかかるからだ。代入の回数を n とデータ個数を m とすると $(n = m + (m - 1) + (m - 2) + \dots + 1)$ 、ソートを実行すると、各反復で代入回が増加する。

4 バブルソート

バブルソートは隣接している要素を比較し、意図する順番になるまで入れ替える。水の気泡が水面上がっていくように、配列の各要素は反復するごとに末尾に移動していきます。そのため、バブルソートと呼ばれています。

しかし、今回は要素は末尾に移動するでなく、先頭の方へ移動することにする。そうすると、ソートされた配列は配列の先頭から形成される。

4.1 プログラム

以下はバブルソートのプログラムである。

```

1 #include <stdio.h>
2 #include <stdlib.h>
3 #include <time.h>
4
5 int countLines(FILE *in) {
6     char c;
7     int count = 0;
8
9     do {
10         c = fgetc(in);
11         if(c == '\n') count++;
12     } while (c != EOF);
13
14     rewind(in);
15
16     return count;
17 }
18
19 void swap(int *a, int *b) {

```

```

20     int tmp = *a;
21     *a = *b;
22     *b = tmp;
23 }
24
25 void store_array(FILE *in, int *p) {
26     int x;
27     int i = 0;
28
29     while (fscanf(in, "%d", &x) != EOF) {
30         p[i] = x;
31         i++;
32     }
33 }
34
35 void processTime(clock_t t) {
36     double time = ((double)t)/CLOCKS_PER_SEC;
37     printf("%.3lf ms\n", time*1000);
38 }
39
40 void bubbleOpt(int lines, int *p) {
41     int k = 0;
42     clock_t t;
43
44     t = clock();
45     while(k < lines - 1) {
46         int last = lines - 1;
47         for(int j = lines - 1; j > k; j--) {
48             if(p[j-1] > p[j]) {
49                 swap(&p[j-1], &p[j]);
50                 last = j;
51             }
52         }
53         k = last;
54     }
55
56     t = clock() - t;
57     processTime(t);
58 }
59
60 void bubbleSort(int lines, int *p) {
61     clock_t t;
62
63     t = clock();
64
65     for (int i = 0; i < lines; i++) {
66         for(int j = lines - 1; j > i; j--) {
67             if(p[j-1] > p[j])
68                 swap(&p[j-1], &p[j]);
69         }
70     }
71
72     t = clock() - t;
73     processTime(t);
74 }
75
76 void setFileName(char *dst, int index) {
77     sprintf(dst, "data%d.dat", index);
78 }
79
80 void setOutFileName(char *dst, int index){
81     sprintf(dst, "bubble%d.dat", index);
82 }
83
84 void startSorting(int n) {
85     FILE *in;
86     FILE *out;

```

```

87
88     for (int i = 1; i <= n; i++) {
89         char *filename = malloc(10);
90         char *outname = malloc(10);
91
92         printf("%d | ", i);
93
94         setFileName(filename, i);
95
96         if((in = fopen(filename, "r")) == NULL)
97             printf("input file name error\n");
98
99         int lines = countLines(in);
100
101         int *p = malloc(lines * sizeof(int));
102         store_array(in, p);
103
104         bubbleOpt(lines, p);
105         // bubbleSort(lines, p);
106
107         setOutFileName(outname, i);
108
109         out = fopen(outname, "w");
110
111         for (int i = 0; i < lines; i++) {
112             fprintf(out, "%d\n", p[i]);
113         }
114
115         free(p);
116         free(filename);
117         free(outname);
118     }
119
120     fclose(in);
121     fclose(out);
122 }
123
124 int main(void) {
125     startSorting(8);
126     return 0;
127 }

```

4.2 動作

バブルソートは隣接している要素を比較するに基づく。バブルソートのプログラムの主な流れは、以下の通りである：

1. 末尾の要素に着目する.
2. ソート済の末尾まで比較する.
3. 全部ソートされるまで繰り返す.
4. ソート完了.

4.3 時間計算量

バブルソートの計算量は $O(N^2)$ である.

この実験では、バブルソートの改良版も実装しています。これは後で詳しく説明する。各データに対してバケットソートの実行時間は表 (5) のように表す。ここで表す結果は 10 回の実行結果の平均値である。

表 5: 計算時間 (バケットソート)

データ	data1	data2	data3	data4
(元) 計算時間 (ms)	17405.559	17438.424	17391.166	5271.074
(改良) 計算時間 (ms)	18172.166	18163.021	18152.781	0.109

データ	data5	data6	data7	data8
(元) 計算時間 (ms)	15314.797	10281.237	16197.128	17328.02
(改良) 計算時間 (ms)	16301.372	8158.549	16487.387	18105.532

4.4 改良案

バブルソートは入力データの状態を考えず、隣接している要素を比較する。そうすると、バブルソートはすでにソートされている状態でも、比較を繰り返すことになる。一連の比較を行うにおいて、ある時点で交換がなければ、それより先頭側はソート済みである [?]. これに対して、すでにソートされたデータをチェックして、比較するのはソートされていない部分だけで行うことにする。

4.5 考察

その結果、表 (5) に示すように、data4 (昇順) と data6 (バイトニック) では処理時間を減り、残りのデータでは遅くなった。これは、データがすでにソートされているかどうかを確認するために、いくつかの条件があって、単純比較を行うより処理量は少し増やす。なので、大量のデータに対してこのトレードオフを現れる。

5 シェーカーソート

シェーカーソートは、2つのバブルソートを2つの異なる方向に行うものである。しかし、普通のバブルソートと違って、無駄な繰り返しをしないので、大きな配列でも効率よく処理できます [?]. これはシェーカーの動きと似ているため、シェーカーソートと呼ばれています。

5.1 プログラム

以下はシェーカーソートのプログラムである。

```

1 #include <stdio.h>
2 #include <stdlib.h>
3 #include <time.h>
4
5 int countLines(FILE *in) {

```

```

6     char c;
7     int count = 0;
8
9     do {
10         c = fgetc(in);
11         if(c == '\n') count++;
12     } while (c != EOF);
13
14     rewind(in);
15
16     return count;
17 }
18
19 void store_array(FILE *in, int *p) {
20     int x;
21     int i = 0;
22
23     while (fscanf(in, "%d", &x) != EOF) {
24         p[i] = x;
25         i++;
26     }
27 }
28
29 void swap(int *a, int *b) {
30     int tmp = *a;
31     *a = *b;
32     *b = tmp;
33 }
34
35 void processTime(clock_t t) {
36     double time = ((double)t)/CLOCKS_PER_SEC;
37     // printf("%.3lf ms\n", time*1000);
38     printf("%.3lf\n", time*1000);
39 }
40
41 void shacker(const int lines, int *p) {
42     int right = lines - 1;
43     int left = 0;
44
45     while(left != right) {
46         int i;
47         int j;
48         int last;
49
50         for(i = left; i < right; i++) {
51             if(p[i] > p[i+1]) {
52                 swap(&p[i], &p[i+1]);
53                 last = i;
54             }
55         }
56         right = last;
57         for(int j = right; j > left; j--) {
58             if(p[j-1] > p[j]) {
59                 swap(&p[j-1], &p[j]);
60                 last = j;
61             }
62         }
63         left = last;
64     }
65
66 }
67
68
69 void setFileName(char *dst, int index) {
70     sprintf(dst, "data%d.dat", index);
71 }
72

```

```

73 void setOutFileName(char *dst, int index){
74     sprintf(dst, "shacker%d.dat", index);
75 }
76
77 void startSorting(int n) {
78     FILE *in;
79     FILE *out;
80
81     for (int i = 1; i <= n; i++) {
82         char *filename = malloc(10);
83         char *outname = malloc(10);
84
85         // printf("%d | ", i);
86
87         setFileName(filename, i);
88
89         if((in = fopen(filename, "r")) == NULL)
90             printf("input file name error\n");
91
92         int lines = countLines(in);
93         int *p = malloc(lines * sizeof(int));
94         store_array(in, p);
95
96         clock_t t;
97         t = clock();
98         shacker(lines, p);
99         t = clock() - t;
100        processTime(t);
101
102        setOutFileName(outname, i);
103
104        out = fopen(outname, "w");
105
106        for (int i = 0; i < lines; i++) {
107            fprintf(out, "%d\n", p[i]);
108        }
109
110        free(p);
111        free(filename);
112        free(outname);
113    }
114
115    fclose(in);
116    fclose(out);
117 }
118
119 int main(int argc, char **argv) {
120     startSorting(8);
121     return 0;
122 }

```