Comparison of Sorting Algorithms

Arden Rasmussen 19-11-2016

Contents

1	Intr	$\operatorname{oduction} \hspace{1cm} 1$
	1.1	Terms
2	Alac	prithms 2
	2.1	Bubble Sort
	$\frac{2.1}{2.2}$	Bucket Sort
	2.3	Cocktail Sort
	$\frac{2.3}{2.4}$	Counting Sort
	$\frac{2.4}{2.5}$	8
	$\frac{2.5}{2.6}$	1
	$\frac{2.0}{2.7}$	
		8
	2.8	Quick Sort
	2.9	Radix Sort
	2.10	Selection Sort
		Shell Sort
		Tim Sort
	2.13	Tree Sort
Ap	peno	dices 9
_	_	
\mathbf{A}		orithms 9
	A.1	Bubble
		A.1.1 C++ 9
	A.2	Bucket
		A.2.1 C++
	A.3	Cocktail
		A.3.1 C++
	A.4	Counting
		A.4.1 C++ 13
	A.5	Insertion
		A.5.1 C++ 14
	A.6	Merge
		A.6.1 C++
	A.7	Quick
		Å.7.1 C++
	A.8	Selection
		A.8.1 C++

Abstract

This project aims to compare and contrast the some commonly used sorting algorithms. The different sorting algorithms that will be tested will be tested on value range, and value size, in order to determine the best and worst case scenarios for each algorithm.

1 Introduction

1.1 Terms

This is a set of standard terms and variables that will be referenced to for explanations of the sorting algorithms, and their methods.

n: The number of values that will be sorted by the sorting algorithms.

r: The range for n that will be sorted.

Values: The stored values to be sorted, of n length, and in r range.

t: The time in seconds that an algorithm takes to sort n values.

(i) Reference to specific line in referenced code example.

2 Algorithms

Here is a description of the algorithms that were a part of this research. In this section each algorithm will be described, and the method of its sorting will be explained in words, as well as with persudo code.

2.1 Bubble Sort

Bubble sort works off of a method where the algorithm only focus on a two terms of the list at a time. Bubble sort compares two adjacent pairs of numbers, and switches them to be in the correct order. The algorithm moves through the list until it reaches the end. However, this does not mean that the list is sorted, so the algorithm returns to the beginning and repeats the loop switching items when necessary until there are no more items that need to be switched.

This method can then be optimized through several methods. One optimization method that is implemented in this example is setting a limit to the maximum on the loop. If there are no more values that need to be swapped, then the algorithm does not need to continue to check those values. So simply the algorithm can track a limit to the maximum value that needs to be checked, and instead of looping through the entire list of values, the algorithm can loop through a select number of the values from the lists.

Here is an example of the persudo code, followed by a line by line explanation of the code. A example of bubble sort in actual code can be found in A.1.

```
function BubbleSort(Values)

n = \operatorname{length}(Values)

while n \neq 0 do

newn = 0

for i = 1 to n - 1 do

if Values[i - 1] > Values[i] then

\operatorname{Swap}(Values[i - 1], Values[i])

newn = i

end if

end for

n = newn

end while

end function
```

- (1) The definition of the function with an input of Values which is a list of values to be sorted.
- (2) The beginning of a loop the follow until Values is completely sorted.
- (3) The definition of *newn* variable which will be used to determine the maximum limit on the sorting loop.
- (4) The for loop that begins at the first two items in Values and ends at

the n-1 item in Values.

- (5) The if check to determine if the two items need to be switched.
- (6) Swaps the two compared items positions.
- (7) Sets newn to i in order to move the upper limit up to the current maximum necessary swap position.
- (10) Sets n to newn essentially setting the upper limit on the for loop.

Bubble sort is considered a very poor method for sorting algorithms, and is most often less efficient than other extremely simple sorting algorithms. Bubble sort has a complexity of $O(n^2)$. This complexity means that for an increase in the n, the com-

plexity of the algorithm, and the time requirement grow exponentially. For small n values this algorithm would work sufficiently, but as the number of values grows this algorithm becomes much less efficient, and soon becomes useless.

2.2 Bucket Sort

Bucket sort works with a method of dividing the set of values into smaller and smaller sub groups. The algorithm creates a set of "buckets" that it then sorts the values into. This method can be simplified as sorting groups at a time. For example any values that are in the first 10% will be placed in the first bucket, then values in the second 10% will be placed in the second bucket. This continues until all values have been placed into a bucket. Then the algorithm is run again with each of the bucket that were just created as the main bucket. Each of the terms that was placed in the first bucket are then again sorted into ten more buckets. This is done until all values are in their own bucket. Then the buckets are placed back together in order. This results in the sorted list of values.

Bucket sort can as well be optimized. Bucket sort is very effective

at breaking up the large data set into smaller semi-sorted sets of data, but at a point bucket sort becomes inefficient. Because going through algorithm until every value has its own bucket requires much more calculations than some optimized algorithms. One such optimization is to use bucket sort, then once the number of elements in a bucket is below a specified level, then that bucket is sorted using some other sorting algorithm. Or another option is to place all the buckets back into a single semi-sorted list, and then use Insertion Sort to complete the sorting, because Insertion Sort's runtime is based off of an elements distance from its final position.

Here is bucket sort code in persudo code, accompanied with a line by line explanation. Real examples of code can be found in A.2.

```
function BucketSort(Values, n)

Arrays < - new array of n empty lists

for i = 0 to (length(Values)-1) do

insert Values[i] into Arrays[floor(Values[i]/(rangeof(Values)/n))]
end for

clear Values

for i = 0 to n - 1 do

append BucketSort(Arrays[i], n) to Values
end for

return Values
end function
```

- (1) The definition of the function with an input of Values which is a list of values to be sorted, and n which is the number of buckets to sort into.
- (2) Creates a new list of n empty list or "buckets".
- (3) Beginning of a for loop that goes through the length of *Values* and sorts them into the *Arrays*.
- (4) The method used for determining what Array to place the item into is done by dividing the value of the item by the value range of the array, then rounding down. This will result in returning the correct Array for the item to be sorted into.
 - (6) Clears Values of any items to

Bucket sort is a moderately effective sorting algorithm, but could be greatly improved with some simple optimizations. On its own Bucket sort has a complexity of O(n), as there is a need for a bucket for every element

prepare it for rewriting with sorted results.

- (7) Beginning of a for loop that goes through every list in *Arrays*
- (8) Recursively calls bucket sort with each new list of items that was placed in each bucket. Eventually there will only be a single item in a bucket and it cannot be sorted. The returned results will be appended together into a complete and sorted list of values.
- (10) This sorted list of values is then returned for any functions higher up in the chain of being called in the recursive function.

that is being sorted. This means that it is fairly efficient, but still could be greatly improved by the implementation of a limit for the minimum size of bucket before initiation a different sorting algorithm.

2.3 Cocktail Sort

Cocktail Sort uses a method similar to bubble sort. However one improvement that Cocktail sort makes over bubble sort is that the algorithm of cocktail sort goes both ways through the list of items. Bubble sort can only move the items up the the correct position, but cocktail sort is able to also

move items down in the list, so the terms are sorted from both directions at the same time.

Here is the persudo code for an example of Cocktail sort, accompanied with a line by line explanation. Real examples of code can be found in A.3.

```
function COCKTAILSORT(Values)
```

```
Swapped = true
Start = 0
End = \operatorname{length}(Values) - 1
while Swapped = true do
Swapped = false
for i = Start to End do

if Values[i-1] > Values[i=1] then
Swap(Values[i], Values[i+1])
Swapped = true
```

```
end if
      end for
      if Swapped = false then
         break
      end if
      Swapped = false
      End-1
      for i = End to Start do
         if Values[i] > Values[i+1] then
            Swap(Values[i], Values[i+1])
            Swapped = true
         end if
      end for
      Start + 1
   end while
end function
```

- (1) The definition of the function with an input of Values which is a list of values to be sorted.
- (2) Creates variables storing the beginning and end of the sort range.
- (5) Beginning of a while loop that runs until Values is sorted.
- (7) Basic Bubble sort for loop, moving one item up to its final position.

Cocktail sort is considered a very poor method for sorting algorithms, and is most often less efficient than other extremely simple sorting algorithms. Cocktail sort has a complexity of $O(n^2)$. This complexity means that for an increase in the n, the com-

- (13) Checks if nothing was moved, which would mean that Values is sorted.
- (17) Reduces End by one, because that item is correctly sorted.
- (18) Second Bubble sort loop, but traveling backwards through Values
- (24) Increases *Start* by one, because that item is correctly sorted.

plexity of the algorithm, and the time requirement grow exponentially. For small n values this algorithm would work sufficiently, but as the number of values grows this algorithm becomes much less efficient, and soon becomes useless.

2.4 Counting Sort

Counting Sort uses a method where the algorithm counts the number of elements with the same value. Counting sort goes through all the elements in the list, and creates counters for each value in the element list range. The algorithm then goes through the list and adds to each respective counter for each element. Once all the counters are created, then the final list of elements is created by adding each counter to the list. This means that this algorithm is extremely memory efficient with data sets that have a small range, as the algorithm only needs to store a list of range number of inte-

gers. However when the data set has a wide range, then this algorithm becomes much less efficient, as there will be many unused counters, and will require storing many more values. This algorithm is very efficient for time, but for memory the algorithms complexity quickly grows, causing it to be unmanageable for data sets with a wide range of values.

This algorithm can be optimized by

reducing the number of for loops that are necessary, and limiting the counters to a lower and upper bound. Because if the values in a list begin until sever hundred, there is no reason to store counters for everything up until then.

Here is Counting sort code in persudo code, accompanied with a line by line explanation. Real examples of code can be found in A.4.

function CountingSort(Values)

```
Counters < - an array of integer values
   min < - integer for smallest value in Values
   max < - integer for largest value in Values
   for i = 0 to length(Values) - 1 do
      if Values[i] > max then
         max = Values[i]
      end if
      if Values[i] < min then
         min = Values[i]
      end if
   end for
   for i = min to max do
      append 0 to Counters
   end for
   for i = 0 to length(Values) - 1 do
      Counters[values[i] - min] + 1
   end for
   Clear Values
   for i = 0 to length(Counters) - 1 do
      for j = 0 to Counters[i] do
         append (i + min) to Values
      end for
   end for
end function
```

- (1) The definition of the function with an input of Values which is a list of values to be sorted.
- (5) For loop used to determine the minimum and maximum values in Values.
- (13) For loop to create counters for each value from min to max
- (16) For loop that runs through all elements in Values and adds to each respective counter for each value.
- (19) Clears *Values* of any items to prepare it for rewriting with sorted results.
- (20) Beginning of a for loop that goes through every counter in

Counters

(21) For loop that adds the counted

Counting sort is extremely efficient in terms of time, as there is no actual comparison between the different elements in the list. This means that counting sort runs fairly quickly, but the memory necessary grows with the range of elements. For a list of elements that has a small range, counting sort will be extremely effective, but for

number of each element back into Values based off of Counters

a list of elements with a large range of elements, counting sort will require much more memory to run. The complexity of counting sort is O(n + k), where k is the range of values. This is true for the best and worst case scenarios for counting sort, making it an extremely predictable sorting algorithm.

2.5 Heap Sort

2.6 Insertion Sort

Insertion sort is a very common sorting algorithm, as it is very simple, and still effective. The method of insertion sort is to select the first value in the list, and to then check it against

every value after it in the list. If the selected value is greater than the other value in the comparison, then the two values are switched, otherwise the loop stops.

- 2.7 Merge Sort
- 2.8 Quick Sort
- 2.9 Radix Sort
- 2.10 Selection Sort
- 2.11 Shell Sort
- 2.12 Tim Sort
- 2.13 Tree Sort

Appendices

A Algorithms

A.1 Bubble

A.1.1 C++

```
#include "../../sort_headers.h"
   #include <cmath>
   #include <ctime>
   #include <iostream>
   #include <vector>
   #include <appareo_files/appareo_headers.h>
   double sort::BubbleSort(bool display) {
     if (display == true) {
       appareo::induco::CreateProgressBar("Bubble Sort");
10
     double timeelapsed;
     appareo::induco::Timer(true);
     int n = values.size();
14
15
     while (n != 0) {
       int newn = 0;
16
        for (int i = 1; i < n; i++) {
17
          if (values[i - 1] > values[i]) {
18
            iter_swap(values.begin() + i - 1, values.begin() + i);
           newn = i;
20
         }
21
       }
22
       n = newn;
23
        if (display == true) {
24
         appareo::induco::UpdateProgressBar((values.size() - n) / (
25
        float ) values . size () );
26
27
     timeelapsed = appareo::induco::Timer();
     if (display == true) {
29
       appareo::induco::TerminateProgressBar();
30
31
     return (timeelapsed);
32
```

algorithms/bubble/bubble_sort.cpp

A.2 Bucket

A.2.1 C++

```
#include <appareo_files/appareo_headers.h>
   #include <cmath>
2
   #include <ctime>
   #include <iostream>
   #include <vector>
   #include "../../sort_headers.h"
6
   namespace sort {
   int Btotalrecursioncount = 0;
10
11
   double sort::BucketSort(int bucketcount, bool display) {
12
      if (display == true) {
        appareo::induco::CreateProgressBar("Bucket Sort");
14
15
16
      Btotalrecursioncount = 0;
      double timeelapsed;
17
      appareo :: induco :: Timer (true);
18
      if (display == true) {
19
        appareo::induco::UpdateProgressBar(0);
20
21
      BRecursion(bucketcount, &values, display, true);
22
      timeelapsed = appareo::induco::Timer();
      if (display == true) {
24
        appareo::induco::TerminateProgressBar();
25
26
      return (timeelapsed);
27
29
    void sort::BRecursion(int bucketcount, std::vector<int> *bucket,
30
        bool display,
                            bool top) {
31
      std::vector<std::vector<int>>> buckets;
32
      int max = 0, min = 0, bucketrange;
33
      for (int i = 0; i < bucketcount; i++) {
34
        std::vector<int> newbucket;
35
        buckets.push_back(newbucket);
36
37
     \max = bucket \rightarrow at(0);
38
      \min = bucket \rightarrow at(0);
39
      for (int i = 1; i < bucket->size(); i++) {
40
        int value = bucket->at(i);
41
        if (value > max) {
42
          \max = \text{value};
43
44
        if (value < min) {
45
          min = value;
46
47
48
      bucketrange = ceil((max - min) / (double)bucketcount);
49
      for (int i = 0; i < bucket \rightarrow size(); i++) {
50
        int value = bucket->at(i);
        if (value = max) {
52
```

```
buckets \, [\, buckets \, . \, size \, (\,) \,\, - \,\, 1 \, ] \, . \, push\_back \, (\, value \,) \, ;
53
54
          else {
           buckets[BBucket(value - min, bucketrange)].push_back(value);
55
        }
56
57
      bucket->clear();
for (int i = 0; i < buckets.size(); i++) {</pre>
58
59
         if (buckets[i].size() <= 1) {
60
           bucket->insert(bucket->end(), buckets[i].begin(), buckets[i].
61
        end());
        } else if (BDupes(buckets[i]) == true) {
62
           Btotalrecursioncount += buckets[i].size();
63
           bucket->insert(bucket->end(), buckets[i].begin(), buckets[i].
64
        end());
        } else {
65
           BRecursion (bucketcount, &buckets[i], display);
66
67
           bucket->insert(bucket->end(), buckets[i].begin(), buckets[i].
        end());
         if (display = true && top = true) {
69
70
           appareo::induco::UpdateProgressBar(bucket->size() /
                                                    (double) scrambled values.
71
         size());
72
      }
73
74
75
    int sort::BBucket(int value, int bucketsize) {
76
      return ((floor(value / (double)bucketsize)));
77
78
79
    bool sort::BDupes(std::vector<int> bucket) {
80
      if (bucket.size() <= 1) {
81
        return (true);
82
83
84
      int value = bucket[0];
      for (int i = 1; i < bucket.size(); i++) {
  if (bucket[i] != value) {</pre>
85
86
           return (false);
87
88
89
      return (true);
90
```

algorithms/bucket/bucket_sort.cpp

A.3 Cocktail

A.3.1 C++

```
#include "../../sort_headers.h"
   #include <appareo_files/appareo_headers.h>
2
   #include <cmath>
   #include <ctime>
   #include <iostream>
   #include <vector>
6
   double sort :: CocktailSort (bool display) {
9
     if (display == true) {
       appareo::induco::CreateProgressBar("Cocktail Sort");
10
11
     double timeelapsed;
12
     appareo :: induco :: Timer (true);
14
     bool swapped = true;
     int start = 0, end = values.size() - 1;
15
16
      while (swapped == true) {
       swapped = false;
17
        for (int i = start; i < end; i++) {
18
          if (values[i] > values[i + 1]) {
19
            iter_swap(values.begin() + i, values.begin() + i + 1);
20
21
            swapped = true;
         }
22
23
        if (swapped == true) {
24
         end --;
25
26
         swapped = false;
          for (int i = end; i >= start; i--) {
27
            if (values[i] > values[i + 1]) 
              iter\_swap(values.begin() + i, values.begin() + i + 1);
29
30
              swapped = true;
31
         }
33
          start++;
          if (display == true) {
34
           appareo::induco::UpdateProgressBar((values.size() - (end -
35
       start)) /
                                                 (double) values. size() *
36
       2);
37
         }
38
39
     timeelapsed = appareo::induco::Timer();
40
41
     if (display = true) {
       appareo::induco::TerminateProgressBar();
42
43
     return (timeelapsed);
44
45
```

algorithms/cocktail/cocktail_sort.cpp

A.4 Counting

A.4.1 C++

```
#include "../../sort_headers.h"
   #include <cmath>
2
   #include <ctime>
   #include <iostream>
   #include <vector>
   #include <appareo_files/appareo_headers.h>
6
   double sort::CountingSort(bool display) {
9
     if (display == true) {
       appareo::induco::CreateProgressBar("Counting Sort");
10
11
     double timeelapsed;
12
     appareo :: induco :: Timer (true);
     std::vector<int> counters;
14
     15
     for (int i = 1; i < values.size(); i++) {
16
       if (values[i] > max) {
17
         max = values[i];
18
19
       if (values[i] < min) {
20
21
         min = values[i];
23
     for (int i = min; i \le max; i++) {
24
       counters.push_back(0);
25
26
     for (int i = 0; i < values.size(); i++) {
27
       counters [values [i] - min]++;
       if (display == true) {
29
30
         appareo::induco::UpdateProgressBar(0.5 * (i / (double) values.
       size()));
       }
31
32
     values.clear();
33
     for (int i = 0; i \le \max; i++) {
34
       if (display == true) {
35
         appareo::induco::UpdateProgressBar((0.5 * (i / (double)max))
36
       + 0.5);
37
       for (int j = 0; j < counters[i]; j++) {
38
         values.push_back(i + min);
39
40
41
     timeelapsed = appareo::induco::Timer();
42
43
     if (display = true) {
       appareo::induco::TerminateProgressBar();
44
45
46
     return (timeelapsed);
47
```

algorithms/counting/counting_sort.cpp

A.5 Insertion

A.5.1 C++

```
#include "../../sort_headers.h"
   #include <cmath>
2
   #include <ctime>
   #include <iostream>
   #include <vector>
   #include <appareo_files/appareo_headers.h>
6
   double sort::InsertionSort(bool display) {
9
     if (display = true) {
       appareo::induco::CreateProgressBar("Insertion Sort");
10
11
     double timeelapsed;
12
13
     appareo :: induco :: Timer (true);
     for (int i = 1; i < values.size(); i++) {
14
       int j = i;
15
        while (j > 0 \&\& values[j-1] > values[j]) {
16
         iter_swap(values.begin() + j, values.begin() + j - 1);
17
18
         j --;
19
        if (display = true) {
20
          appareo::induco::UpdateProgressBar(i / (double)values.size())
21
23
     timeelapsed = appareo :: induco :: Timer();
24
     if (display == true) {
25
       appareo::induco::TerminateProgressBar();
26
27
     return (timeelapsed);
28
29
```

algorithms/insertion/insertion_sort.cpp

A.6 Merge

A.6.1 C++

```
#include "../../sort_headers.h"
   #include <cmath>
2
   #include <ctime>
   #include <iostream>
   #include <vector>
   #include <appareo_files/appareo_headers.h>
   namespace sort {
   \mathtt{std}:: \mathtt{vector} \negthinspace < \negthinspace \mathsf{int} \negthinspace > \negthinspace \mathsf{Mworkvector} \: ;
9
10
11
    double sort::MergeSort(bool display) {
12
      if (display == true) {
        appareo::induco::CreateProgressBar("Merge Sort");
14
15
16
      double timeelapsed;
      appareo::induco::Timer(true);
17
      Mworkvector = values;
18
      for (int i = 1; i < values.size(); i = 2 * i) {
19
        for (int j = 0; j < values.size(); j = j + 2 * i) {
20
21
           if (display == true) {
             appareo::induco::UpdateProgressBar(j / (double)values.size
22
        ());
23
           int vara, varb;
24
           if (j + i >= values.size()) {
25
            vara = values.size();
26
27
          } else {
             vara = j + i;
28
29
           if (j + 2 * i >= values.size()) {
30
             varb = values.size();
31
           } else {
32
             varb = j + 2 * i;
33
34
          MMerge(j, vara, varb);
35
        }
36
37
        values = Mworkvector;
38
39
      timeelapsed = appareo::induco::Timer();
      if (display = true) {
40
        appareo::induco::TerminateProgressBar();
41
42
      return (timeelapsed);
43
44
45
    void sort::MMerge(int ileft, int iright, int iend) {
46
47
      int i = ileft;
48
      int j = iright;
      for (int k = ileft; k < iend; k++) {
49
        if (i < iright && (j >= iend || values[i] <= values[j])) {
50
          Mworkvector [k] = values [i];
51
          i++;
52
```

 $algorithms/merge/merge_sort.cpp$

A.7 Quick

A.7.1 C++

```
#include "../../sort_headers.h"
   #include <cmath>
2
   #include <ctime>
   #include <iostream>
   #include <vector>
   #include <appareo_files/appareo_headers.h>
   namespace sort {
   int Qtotalrecursioncount = 0;
10
11
   double sort::QuickSort(bool display) {
12
      if (display == true) {
       appareo::induco::CreateProgressBar("Quick Sort");
14
15
16
      Qtotalrecursioncount = 0;
      double timeelapsed= appareo::induco::Timer(true);
17
      QRecursive(0, values.size(), display);
18
      timeelapsed = appareo::induco::Timer();
19
      if (display = true) {
20
21
        appareo::induco::TerminateProgressBar();
22
      return (timeelapsed);
23
   }
24
25
   void sort::QRecursive(int low, int high, bool display) {
26
      if (low < high) {</pre>
27
        int partition = QPartition(low, high);
        QRecursive (low\,,\ partition\,-\,1\,,\ display\,)\,;
29
30
        QRecursive (partition + 1, high, display);
31
      if (display == true) {
32
        Qtotalrecursioncount++;
33
        appareo::induco::UpdateProgressBar(Qtotalrecursioncount / (
34
        double)(2 * values.size()));
35
   }
36
37
   int sort::QPartition(int low, int high) {
38
      int piviot = values[high];
      int i = low;
40
      for (int j = low; j < high; j++) {
41
42
        if (values[j] <= piviot) {</pre>
          iter_swap(values.begin() + i, values.begin() + j);
43
44
          i++;
       }
45
46
      iter_swap(values.begin() + i, values.begin() + high);
47
      return (i);
48
49
```

algorithms/quick/quick_sort.cpp

A.8 Selection

A.8.1 C++

```
#include "../../sort_headers.h"
   #include <cmath>
2
   #include <ctime>
   #include <iostream>
   #include <vector>
   #include <appareo_files/appareo_headers.h>
   double sort::SelectionSort(bool display) {
9
     if (display = true) {
       appareo::induco::CreateProgressBar("Selection Sort");
10
11
     double timeelapsed;
12
     appareo::induco::Timer(true);
     for (int i = 0; i < values.size() - 1; i++) {
14
       int smallestpointer = i;
15
        for (int j = i + 1; j < values.size(); j++) {
16
         if (values[j] < values[smallestpointer]) {</pre>
17
            smallestpointer = j;
18
         }
19
20
        if (smallestpointer != i) {
21
         iter_swap(values.begin() + i, values.begin() +
       smallestpointer);
23
        if (display = true) {
24
         appareo::induco::UpdateProgressBar(i / (double)values.size())
25
26
27
28
     timeelapsed = appareo::induco::Timer();
     if (display == true) {
29
       appareo::induco::TerminateProgressBar();
30
31
     return (timeelapsed);
32
33
```

algorithms/selection/selection_sort.cpp