4.1 Performance

Introduction to Programming in Java: An Interdisciplinary Approach · Robert Sedgewick and Kevin Wayne · Copyright © 2002–2010 · 13/2/2012 9:01:40

Running Time

“As soon as an Analytic Engine exists, it will necessarily guide the future course of the science. Whenever any result is sought by its aid, the question will arise—by what course of calculation can these results be arrived at by the machine in the shortest time?” — Charles Babbage

The Challenge

Q. Will my program be able to solve a large practical problem?

Key insight. [Knuth 1970s]
Use the scientific method to understand performance.

Scientific Method

Scientific method.

- Observe some feature of the natural world.
- Hypothesize a model that is consistent with the observations.
- Predict events using the hypothesis.
- Verify the predictions by making further observations.
- Validate by repeating until the hypothesis and observations agree.

Principles.

- Experiments must be reproducible.
- Hypothesis must be falsifiable.

Reasons to Analyze Algorithms

Predict performance.

- Will my program finish?
- When will my program finish?

Compare algorithms.

- Will this change make my program faster?
- How can I make my program faster?

Basis for inventing new ways to solve problems.

- Enables new technology.
- Enables new research.

Algorithmic Successes

Sorting.

- Rearrange array of N items in ascending order.
- Applications: databases, scheduling, statistics, genomics, ...
- Brute force: \( N^2 \) steps.
- Mergesort: \( N \log N \) steps, enables new technology.

Charles Babbage (1864)

Analytic Engine

How many times do you have to turn the crank?

 compile debug solve problems in practice
Algorithmic Successes

Discrete Fourier transform.
- Break down waveform of \( N \) samples into periodic components.
- Applications: DVD, JPEG, MRI, astrophysics, ...
- Brute force: \( N^2 \) steps.
- FFT algorithm: \( N \log N \) steps, enables new technology.

Freidrich Gauss 1805

Algorithmic Successes

N-body Simulation.
- Simulate gravitational interactions among \( N \) bodies.
- Application: cosmology, semiconductors, fluid dynamics, ...
- Brute force: \( N^2 \) steps.
- Barnes-Hut algorithm: \( N \log N \) steps, enables new research.

Three-Sum Problem

Three-sum problem: Given \( N \) integers, how many triples sum to 0?

Context. Deeply related to problems in computational geometry.

Q. How would you write a program to solve the problem?

Three-Sum: Brute-Force Solution

```java
public class ThreeSum {
    public static int count(int[] a) {
        int N = a.length;
        int cnt = 0;
        for (int i = 0; i < N; i++)
            for (int j = i + 1; j < N; j++)
                for (int k = j + 1; k < N; k++)
                    if (a[i] + a[j] + a[k] == 0) cnt++;
        return cnt;
    }
    public static void main(String[] args) {
        int[] a = StdArrayIO.readInt1D();
        int[] a = StdArrayIO.readInt1D();
        StdOut.println(count(a));
    }
}
```

Empirical Analysis

Empirical analysis: Run the program for various input sizes.

<table>
<thead>
<tr>
<th>( N )</th>
<th>time ( \text{sec} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>512</td>
<td>0.03</td>
</tr>
<tr>
<td>1,024</td>
<td>0.26</td>
</tr>
<tr>
<td>2,048</td>
<td>2.16</td>
</tr>
<tr>
<td>4,096</td>
<td>17.18</td>
</tr>
<tr>
<td>8,192</td>
<td>137.76</td>
</tr>
</tbody>
</table>

† Running Linux on Sun Fire X4100 with 16GB RAM.

Caveat. If \( N \) is too small, you will measure mainly noise.
**Stopwatch**

**Q.** How to time a program?  
**A.** A stopwatch.

**Empirical Analysis**

**Data analysis.** Plot running time vs. input size $N$.

**Q.** How fast does running time grow as a function of input size $N$?

**Doubling Hypothesis**

**Doubling hypothesis.** Quick way to estimate $b$ in a power law hypothesis.

Run program, doubling the size of the input?

<table>
<thead>
<tr>
<th>$N$</th>
<th>time’</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>512</td>
<td>0.033</td>
<td>1</td>
</tr>
<tr>
<td>1,024</td>
<td>0.26</td>
<td>7.88</td>
</tr>
<tr>
<td>2,048</td>
<td>2.16</td>
<td>8.43</td>
</tr>
<tr>
<td>4,096</td>
<td>17.18</td>
<td>7.96</td>
</tr>
<tr>
<td>8,192</td>
<td>136.76</td>
<td>7.96</td>
</tr>
</tbody>
</table>

Hypothesis. Running time is about $a N^b$ with $b = \log c$.

**Performance Challenge 1**

Let $T(N)$ be running time of `main()` as a function of input size $N$.

**Scenario 1.** $T(2N) / T(N)$ converges to about 4.

**Q.** What is order of growth of the running time?

1. $N$
2. $N^2$
3. $N^3$
4. $N^4$
5. $2^N$
Performance Challenge 2

Let \( T(N) \) be running time of \( \text{main}() \) as a function of input size \( N \).

Scenario 2. \( T(2N) / T(N) \) converges to about 2.

Q. What is order of growth of the running time?
   1. \( N \)
   2. \( N^2 \)
   3. \( N^3 \)
   4. \( 2^N \)

Prediction and Validation

Hypothesis. Running time is about \( a \cdot N^3 \) for input of size \( N \).

Q. How to estimate \( a \)?
A. Run the program!

Refined hypothesis. Running time is about \( 2.5 \times 10^{-10} \cdot N^3 \) seconds.

Prediction. 1,100 seconds for \( N = 16,384 \).

Observation.

\[
\begin{array}{c|c}
N & \text{time} \\
\hline
16,384 & 1118.86 \\
\end{array}
\]

Mathematical Analysis

Donald Knuth
Turing award ’74

Mathematical Analysis

Running time. Count up frequency of execution of each instruction and weight by its execution time.

\[
\begin{align*}
\text{int count} & = 0; \\
\text{for} \ ( \text{int } i = 0; \ i < N; \ i++) \\
\quad & \text{for} \ ( \text{int } j = i+1; \ j < N; \ j++) \\
\quad & \text{if} \ (a[i] + a[j] == 0) \ \text{count++;}
\end{align*}
\]

Tilde Notation

Tilde notation.
- Estimate running time as a function of input size \( N \).
- Ignore lower order terms.
  - when \( N \) is large, terms are negligible
  - when \( N \) is small, we don’t care

Ex 1. \( 6 \cdot N^3 + 17 \cdot N^2 + 56 \sim 6 \cdot N^3 \)
Ex 2. \( 6 \cdot N^3 + 100 \cdot N^{4/3} + 56 \sim 6 \cdot N^3 \)
Ex 3. \( 6 \cdot N^3 + 17 \cdot N^{2 \log N} \sim 6 \cdot N^3 \)

\[
\begin{array}{c|c|c}
\text{operation} & \text{frequency} & \text{N} \\
\hline
\text{variable declaration} & 2 \\
\text{variable assignment} & 2 \\
\text{less than comparison} & N+1 \\
\text{equal to comparison} & N \\
\text{array access} & N \\
\text{increment} & \leq 2 \cdot N \\
\end{array}
\]

\( \sum \text{N} - 1 + \ldots + 2 + 1 + 0 = \frac{1}{2} N(N-1) \)

\[
\begin{align*}
\text{(N} - 1) & \ldots + 2 + 1 + 0 = \frac{1}{2} N(N-1) \\
\end{align*}
\]

between \( \text{to be zero} \) and \( \text{N.} \)
Mathematical Analysis

Running time. Count up frequency of execution of each instruction and weight by its execution time.

```
public static int count(int[] a) {
    int N = a.length;
    int cnt = 0;
    for (int i = 0; i < N; i++)
        for (int j = i; j < N; j++)
            if (a[i] + a[j] + a[k] == 0)
                cnt++;
    return cnt; // depends on input data
}
```

Inner loop. Focus on instructions in "inner loop."

Constants in Power Law

Power law. Running time of a typical program is $\sim a N^b$.

Exponent $b$ depends on:
- Algorithm.
- Input data.
- Caching.
- Machine.
- Compiler.
- Garbage collection.
- Just-in-time compilation.
- CPU use by other applications.

Leading constant $a$ depends on:
- System-independent effects
- Algorithm.
- Input data.
- Caching.
- Machine.
- Compiler.
- Garbage collection.
- Just-in-time compilation.
- CPU use by other applications.

Our approach. Use doubling hypothesis (or mathematical analysis) to estimate exponent $b$, run experiments to estimate $a$.

Analysis: Empirical vs. Mathematical

Empirical analysis:
- Measure running times, plot, and fit curve.
- Easy to perform experiments.
- Model useful for predicting, but not for explaining.

Mathematical analysis:
- Analyze algorithm to estimate # ops as a function of input size.
- May require advanced mathematics.
- Model useful for predicting and explaining.

Critical difference. Mathematical analysis is independent of a particular machine or compiler; applies to machines not yet built.

Order of Growth Classifications

Observation. A small subset of mathematical functions suffice to describe running time of many fundamental algorithms.

Mathematical function
```
while (N > 1) {
    N = N / 2;
}
```

$log N$ (logarithmic)
```
for (int i = 0; i < N; i++)
```

$N$ (linear)
```
for (int i = 0; i < N; i++)
for (int j = 0; j < N; j++)
```

$N^2$ (quadratic)
```
public static void g(int N) {
    if (N == 0) return;
    g(N/2);
    g(N/2);
    for (int i = 0; i < N; i++)
```

$N \log N$ (linearithmic)
```
public static void h(int N) {
    if (N == 0) return;
    if (N == 1) return;
    for (int i = 0; i < N; i++)
```

$2^N$ (exponential)

Order of Growth: Consequences

Effect of increasing problem size for a program that runs for a few seconds

<table>
<thead>
<tr>
<th>Order of growth</th>
<th>Predicted running time if problem size is increased by a factor of 100</th>
<th>Predicted factor of problem size increase if computer speed is increased by a factor of 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td>a few minutes</td>
<td>linear</td>
</tr>
<tr>
<td>logarithmic</td>
<td>a few minutes</td>
<td>logarithmic</td>
</tr>
<tr>
<td>linearithmic</td>
<td>several hours</td>
<td>quadratic</td>
</tr>
<tr>
<td>quadratic</td>
<td>a few weeks</td>
<td>cubic</td>
</tr>
<tr>
<td>cubic</td>
<td>exponential</td>
<td>exponential</td>
</tr>
<tr>
<td>exponential</td>
<td>forever</td>
<td>Effect of increasing computer speed on problem size that can be solved in a fixed amount of time</td>
</tr>
</tbody>
</table>
Binary Search

Sequential Search vs. Binary Search

- Sequential search in an unordered array.
  - Examine each entry until finding a match (or reaching the end).
  - Takes time proportional to length of array in worst case.

- Binary search in an ordered array.
  - Examine the middle entry.
  - If equal, return index.
  - If too large, search in left half (recursively).
  - If too small, search in right half (recursively).

```
// precondition: array a[] is sorted
public static int search(int key, int [] a) {
  int lo = 0;
  int hi = a.length - 1;
  while (lo <= hi) {
    int mid = lo + (hi - lo) / 2;
    if (key < a[mid]) hi = mid - 1;
    else if (key > a[mid]) lo = mid + 1;
    else return mid; // not found
  }
  return -1; // not found
}
```

Searching Challenge 1

Q. A credit card company needs to whitelist 100 million customer account numbers, processing 10,000 transactions per second.

Using sequential search, what kind of computer is needed?

A. Toaster.
B. Cell phone.
C. Your laptop.
D. Supercomputer.
E. Google server farm.

Searching Challenge 2

Q. A credit card company needs to whitelist 100 million customer account numbers, processing 10,000 transactions per second.

Using binary search, what kind of computer is needed?

A. Toaster.
B. Cell phone.
C. Your laptop.
D. Supercomputer.
E. Google server farm.