

Komplexität von Algorithmen

Martin Wirsing

in Zusammenarbeit mit
Matthias Hölzl, Piotr Kosciuzenko, Dirk Pattinson

10/04/03

Ziele

- Zeit- und Speicherplatzbedarf einer Anweisung berechnen können
- Zeit- und Speicherplatzbedarf einer Methode berechnen können
- Unterschiede der Komplexität vom schlechtesten, besten und mittleren Fall kennen
- O-Notation wiederholen
- „Praktische“ Berechenbarkeit eines Algorithmus einschätzen können

Zeit- und Speicherplatzbedarf

Der Aufwand für die Abarbeitung eines Algorithmus hängt ab von

- Art, Anzahl und Zusammensetzung der verwendeten Datentypen und algorithmischen Konzepte
- Art der Realisierung der Datentypen und algorithmischen Konzepte auf einer bestimmten Rechenanlage (z.B. Verwaltungsaufwand bei Rekursion) sowie von konkreten Maschineneigenschaften (z.B. Art und Ausführungsgeschwindigkeit der Maschinenoperationen).

Zeit- und Speicherplatzbedarf

Speicherplatzbedarf S für die Deklaration von	Summe des Speicherplatzbedarfs der Attribute
Zahl-, Char-, Boole'schen Werten	1 Maschinenwort
Feldern vom Typ <code>type</code> z.B.	Länge des Feldes * Bedarf für 1 Wert vom Typ <code>type</code>
▪ einstufigen <code>int</code> -Feldern	Länge des Feldes viele Maschinenworte
▪ zweistufigen Feldern <code>int [n][m]</code>	(<code>n * m</code>) viele Maschinenworte
Referenzvariablen (Objektreferenzen) und Objekte	1 Maschinenwort

Zeit- und Speicherplatzbedarf

Zeitbedarf T für die Deklaration von	Zeitbedarf
Konstante <code>k</code>	konstant c_{fetch}
lokale Variable <code>x</code>	konstant c_{fetch}
Zuweisung <code>x = exp;</code>	$c_{store} + \text{Zeit von } exp$
Grundoperationen:	
+	c_+
*	c_*
-1	c_{-1}
<	$c_<$
<code>exp₁ + exp₂</code>	Zeit von <code>exp₁</code> + c_+ + Zeit von <code>exp₂</code>
<code>new C(t)</code>	$c_{call} + \text{Zeit von } t + T_C + c_{store}$

Zeit- und Speicherplatzbedarf

Zeitbedarf T für die Deklaration von	Zeitbedarf
<code>S₁S₂</code>	Zeit von <code>S₁</code> + Zeit von <code>S₂</code>
<code>o.a</code>	Zeit von $o_{att} + c$ ($c = \text{Zeit für Zugriff auf Halde}$)
<code>if (B) S₁ else S₂</code>	Zeit von <code>if</code> + Zeit von <code>B</code> + Zeit von <code>S₁</code> , falls <code>B == true</code> , Zeit von <code>if</code> + Zeit von <code>B</code> + Zeit von <code>S₂</code> , falls <code>B == false</code>
<code>while (B) { S }</code>	Anzahl der Durchläufe * (Zeit von <code>B</code> + Zeit von <code>S</code>) + Zeit von <code>B</code>

Zeit- und Speicherplatzbedarf

Sei die Funktions- / Methoden type m (type x) {stms} gegeben.

Der **Zeitbedarf** eines Aufrufs $m(a)$ hängt ab von den Kosten der Parameterübergabe C_p , bestehen aus

- den Kosten der Berechnung von a T_a und
- den Kosten der Übergabe (des Wertes) von a C_{store}
(die Kosten des „organisatorischen“ Bedarfs zur Berechnung der Rücksprungadresse etc. werden vernachlässigt), sowie
- der Anzahl der Berechnungsschritte des Rumpfes $T_{stms[x=a]}$

ist definiert als $T_m(a) = C_{call} + T_a + T_{stms[x=a]} + C_{store}$

wobei $T_{stms[x=a]}$ die Anzahl der Berechnungsschritte des Rumpfes (in Abhängigkeit von R-Wert von a) angibt.

Zeit- und Speicherplatzbedarf

Der Speicherplatzbedarf eines Aufrufs $m(a)$ hängt ab von

- den Kosten S_a des aktuellen Parameters,
- den organisatorischen Kosten der Parameterübergabe call S_{call} und
- den Kosten $S_{stms[x=a]}$ für die lokalen Deklarationen.

Der Speicherbedarf von $m(a)$ ist definiert als

$$S_m(a) = S_{call} + S_a + S_{stms[x=a]}$$

Bemerkung: In aktuellen Java-Implementierungen wird der zur Berechnung von a benötigte Speicherplatz gelöscht, bevor der Rumpf ausgewertet wird; d.h man verwendet das Maximum anstelle der Summe:

$$S_m(a) = \max(S_a, S_{call} + S_{stms[x=a]})$$

Zeit- und Speicherplatzbedarf

Beispiele:

```
1. T_int result = 1; = C_store + C_fetch
   S_int result = 1; = 1
```

```
2. Sei stm ≡ result = x * result; x = x - 1;
   T_stm = 2C_fetch + C_v + C_store + C_fetch + C_i + C_store
   S_stm = 0
```

```
3. Sei stm ≡ Point p = new Point (0,1);
   T_stm = C_store + 2C_fetch + 2C_store + C_call + 2 · (C_fetch + C_store)
           ↑           ↑           ↑           ↑           ↑
           p           Halde       Punkt(0,1)
           Zuweisung Kosten zur Übergabe
           an p       Berechnung der akt. Parameter
```

Beispiele für den Zeit- und Speicherplatzbedarf des Rumpfs einer Methode

Beispiel 1: Iterative Fakultät (Zeitbedarf)

```
int fakl(int x)
{ int result = 1;
  while (x > 0) {
    result = x * result;
    x = x - 1;
  }
  return result;
}
```

$T_{fakl}(x) =$
 $C_{fetch} + C_{store}$ // Kosten der Vorbesetzung
 $\left. \begin{matrix} + C_1 \\ + C_2 \\ + C_3 \end{matrix} \right\} \cdot x +$ // Kosten der Schleife // x - Anzahl der DI.
 $+ C_{fetch} + C_{return}$ // Kosten von return

$C_1 = 2C_{fetch} + C_v$
 $C_2 = 2C_{fetch} + C_v + C_{store}$
 $C_3 = C_{fetch} + C_v + C_{store}$

Beispiele für den Zeit- und Speicherplatzbedarf des Rumpfs einer Methode

Beispiel 1: Iterative Fakultät (Speicherplatzbedarf)

$$S_{fakl}(x) = S_{call} + \underset{x}{1} + \underset{result}{1} + \underset{return}{1} \quad // \text{Kosten der lokalen Variablen}$$

Beispiele für den Zeit- und Speicherplatzbedarf des Rumpfs einer Methode

Beispiel 2: Binäre Suche

```
class SearchArray
{ char[] arr;
  boolean binSuch(char e)
  { int left = 0;
    int right = this.arr.length - 1;
    int mid;
    boolean found = false;
    while ((left <= right) & !found) {
      mid = (left + right) / 2;
      if (e < this.arr[mid]) right = mid - 1;
      else if (e > this.arr[mid]) left = mid + 1;
      else found = true;
    }
    return found;
  }
}
```

$T_{binSuch}(e) \leq$
 $2(C_{fetch} + C_{store}) + C_v + C_{fetch} + 2C_{call}$ // Vorbesetzung
 $+ C_{fetch} + C_{store}$ // Vorbesetzung
 $+ \lceil \log_2 n \rceil \cdot ((3C_{fetch} + C_v + C_v) // \text{Bedingung}$
 $+ C_{store} + C_{fetch} + C_v + C_v // \text{mid} = \dots$
 $+ 2 \cdot (3C_{fetch} + 2C_{call} + \max\{C_v, C_v\}) // 2\text{mal Bedingung}$
 $+ C_{fetch} + C_{store} + \max\{C_v, C_v\} // \text{Ja, Nein-Zweig}$
 $+ C_{return} + C_{fetch} // \text{Kosten von return}$
 $+ (3C_{fetch} + C_v + C_v + C_v)$
 $S_{binSuch}(e) = 4 // \text{Kosten der lokalen Variablen}$

Beispiele für den Zeit- und Speicherplatzbedarf des Rumpfs einer Methode

Beispiel 3: Rekursive Fakultät

Anweisung	Zeit $T_{fac}(n)$	
	$n == 0$	$n > 0$

```

1 public static int fac(int n)
2 {
3     if (n == 0)                 $c_{ife} + 2c_{fetch}$     $c_{ife} + 2c_{fetch} + c_{==}$ 
4                                  $+ c_{==}$ 
5                                  $c_{fetch} + c_{return}$        ....
6     return 1;
7 }
6     return n * fac(n - 1);    ....            $2c_{fetch} + c_c + c_{store} + c_c +$ 
7                                  $+ c_{call} + c_{return} + T_{fac}(n-1)$ 

```

Beispiele für den Zeit- und Speicherplatzbedarf des Rumpfs einer Methode

Beispiel 3: Fortsetzung

Also

$$T_{fac}(n) = \begin{cases} t_1 & \text{falls } n == 0 \\ T_{fac}(n-1) + t_2 & \text{falls } n > 0 \end{cases}$$

mit $t_1 = c_{ife} + 2c_{fetch} + c_c + c_{return}$ und $t_2 = c_{ife} + 2c_{fetch} ?? + c_{=} + c_{store} + c_c + c_{call} + c_{return}$

Beispiele für den Zeit- und Speicherplatzbedarf des Rumpfs einer Methode

Beispiel 3: Fortsetzung

Die Lösung dieser Gleichung erhält man durch wiederholte Substitution:

$$\begin{aligned}
 T_{fac}(n) &= T_{fac}(n-1) + t_2 \\
 &= T_{fac}(n-2) + t_2 + t_2 \\
 &\vdots \\
 &= T_{fac}(0) + n \cdot t_2
 \end{aligned}$$

d.h.

$$T_{fac}(n) = t_1 + n \cdot t_2$$

Im Vergleich dazu erhält man für die iterative Berechnung ???

$$T_{fak2}(n) = t'_1 + n \cdot t'_2$$

mit $t'_1 = c_{ife} + 3c_{fetch} + c_{store} + c_c + c_{return}$ und $t'_2 = c_{ife} + 5c_{fetch} + 2c_{store} + c_c + c_{=} + c_c$

D.h. Die Laufzeiten unterscheiden sich nicht prinzipiell (siehe später).

Beispiele für den Zeit- und Speicherplatzbedarf des Rumpfs einer Methode

Beispiel 3: Rekursive Fakultät

Anweisung	Speicher	
	$n == 0$	$n > 0$

```

1 public static int fac(int n)    1 + 1 + s_{call}   1 + 1 + s_{call} + S_{fac}(n-1)
2 {
3     if (n == 0)
4         return 1;
5     else
6         return n * fac(n - 1);
7 }

```

Beispiele für den Zeit- und Speicherplatzbedarf des Rumpfs einer Methode

Wieder ergibt sich

$$S_{fac}(n) = \begin{cases} s_1 & \text{falls } n == 0 \\ S_{fac}(n-1) + s_2 & \text{falls } n > 0 \end{cases}$$

Folgerung:

$$S_{fac}(n) = s_1 + n \cdot s_2$$

Während der Speicherplatzbedarf von `fac1` gleich $3 + s_{call}$, also konstant, ist, hängt der Speicherplatzbedarf der rekursiven Funktion von der Größe von n ab und wächst linear mit n .

Beispiel: Suche im (ungeordneten) Feld

```

class SearchArray
{
    int[] arr;
    ...
    boolean such(int e)
    {
        int k = this.arr.length;
        int i = 0
        while (i < k && !(arr[i] == e))
            i = i + 1;
        return (i < k);
    }
}

```

Komplexitätsarten

- Um den Zeit- und Speicherplatzbedarf verschiedener Algorithmen vergleichen zu können, abstrahiert man von der speziellen Eingabe und gibt die Kosten in *Abhängigkeit von der Größe der Eingabe* an.
- Man unterscheidet die Komplexität in *schlechtesten, mittleren und besten Fall* (engl. worst case, average case, best case)

Komplexität im schlechtesten, besten und mittleren Fall

Zeitkomplexität im

schlechtesten Fall

$$T_m^w(n) = \max\{T_m(a) \mid \text{Größe von } a = n\}$$

mittleren Fall

$$T_m^{av}(n) = \text{Durchschnitt von } \{T_m(a) \mid \text{Größe von } a = n\}$$

besten Fall

$$T_m^b(n) = \min\{T_m(a) \mid \text{Größe von } a = n\}$$

Speicherkomplexität im

schlechtesten Fall

$$S_m^w(n) = \max\{S_m(a) \mid \text{Größe von } a = n\}$$

mittleren Fall

$$S_m^{av}(n) = \text{Durchschnitt von } \{S_m(a) \mid \text{Größe von } a = n\}$$

besten Fall

$$S_m^b(n) = \min\{S_m(a) \mid \text{Größe von } a = n\}$$

Beispiele für den Zeit- und Speicherplatzbedarf des Rumpfs einer Methode

Beispiel:

$$T_{\text{binSuch}}^w(n) = \max\{T_{\text{binSuch}}(e, a) \mid \text{Größe von } a = (n-1)\}$$

da die Größe von $e = 1$.

- Für f_{ac} , f_{akl} stimmen der beste, schlechteste und mittlere Fall jeweils überein.

Beispiel:

$$T_{\text{such}}^w(\text{this}, e) \leq t_1 + n \cdot t_2$$

$$S_{\text{such}}(\text{this}, e) = \text{konstant}$$

Beispiele für den Zeit- und Speicherplatzbedarf des Rumpfs einer Methode

$$T_{\text{such}}^w(n) = \max\{T_{\text{such}}(\text{this}, e) \mid \text{Größe von } (\text{this}, e) = n\}$$

$$= t_1 + n \cdot t_2$$

$$T_{\text{such}}^{av} = \text{Durchschnitt von } (\text{this}, e) \mid \text{Größe von } (\text{this}, e) = n\}$$

$$= t_1 + \frac{\sum_{i=1}^n i}{n} \cdot t_2 = t_1 + \frac{n(n+1)}{2n} \cdot t_2$$

wenn man das arithmetische Mittel als Durchschnitt wählt, wobei Größe von (this, e) = Länge von this.arr

$$T_{\text{such}}^b(n) = \min\{T_{\text{such}}(\text{this}, e) \mid \text{Größe von } (\text{this}, e) = n\}$$

$$= t_1 + t_2$$

Größenordnung der Komplexität: Die O -Notation

Die Komplexität $T(n)$ bzw. $S(n)$ wird häufig nur bis auf konstante Faktoren untersucht. Dazu ordnen wir den Aufwandsfunktionen $T(n)$ und $S(n)$ bestimmte Funktionsklassen ihrer „Größenordnung“ zu.

Sei $f : \mathbb{N} \rightarrow \mathbb{N}$ eine Funktion. Wir definieren die Klasse:

$$O(f(n)) =$$

$$\{g(n) \mid \text{es gibt } c > 0, n_0 : 0 \leq g(n) \leq c \cdot f(n) \text{ für alle } n \geq n_0\}$$

$h(n) \in O(f(n))$ bedeutet also, daß h höchstens so schnell wächst wie f (modulo lineare Transformation).

Größenordnung der Komplexität: Die O -Notation

Satz. Sei $k > 0$ eine Konstante, $h(n) \in O(f(n))$, $g(n)$ eine beliebige Funktionen, die für $n \geq n_0$ nichtnegativ ist. Dann gilt:

- $k \cdot h(n)$, $k + h(n)$, $h(n) - k \in O(f(n))$
- $(h(n) \text{ op } g(n)) \in O(f(n) \text{ op } g(n))$ für $\text{op} \in \{+, -, *, /\}$
- O ist transitiv: für alle $f(n)$, $g(n)$, $h(n)$ gilt:
Wenn $f(n) \in O(g(n))$ und $g(n) \in O(h(n))$
so ist $f(n) \in O(h(n))$

Größenordnung der Komplexität: Die O -Notation

Man nennt f

konstant	falls	$f \in O(1)$
logarithmisch	falls	$f \in O(\log_2 n)$
linear	falls	$f \in O(n)$
quadratisch	falls	$f \in O(n^2)$
polynomial	falls	$f \in O(n^k)$ für ein $k \geq 0$
exponentiell	falls	$f \in O(k^n)$ für ein $k \geq 2$

Größenordnung der Komplexität: Die O -Notation

Beispiele:

$T_{\text{fakt}}(n) \in O(n)$	linear,	$S_{\text{fakt}}(n) \in O(1)$	konstant
$T_{\text{binSuch}}(n) \in O(\log_2 n)$	logarithmisch,	$S_{\text{binSuch}}(n) \in O(1)$	konstant
$T_{\text{fac}}(n) \in O(n)$	linear,	$S_{\text{fac}}(n) \in O(n)$	linear

Außerdem gilt:

- Die Hintereinanderausführung zweier Anweisungen mit linearer Zeitkomplexität ist linear.
- Die Zeitkomplexität zweier verschachtelter Schleifen mit linearer Terminierung ist quadratisch.

Exponentielle und polynomiale Zeitkomplexität

Eine exponentielle Zeitkomplexität hat folgendes Problem des „Handelsreisenden“ (engl. Traveling Salesman):

Gegeben sei ein Graph mit n Städten und den jeweiligen Entfernungen sowie eine Entfernung B . Gibt es eine Tour der Länge $\leq B$ durch alle Städte, so daß jede Stadt einmal besucht wird?

Bemerkung:

Für das Traveling-Salesman-Problem gibt es einen nichtdeterministisch-polynomialen Algorithmus („man darf die richtige Lösung raten“).

Das Traveling-Salesman-Problem ist NP-vollständig, d.h. falls es einen polynomialen Algorithmus zu seiner Lösung gibt, so hat jeder nichtdeterministisch-polynomiale Algorithmus eine polynomiale Lösung.

Exponentielle und polynomiale Zeitkomplexität

Algorithmen mit **polynomialer** Komplexität nennt man **praktisch berechenbar**, während **exponentielle** Algorithmen **nicht** (mehr) praktisch berechenbar sind. (Grund: bei Vergrößerung der Eingabe um 1 verdoppelt sich der Aufwand.)

Die Klasse der nichtdeterministisch-polynomialen Algorithmen (bzgl. der Zeitkomplexität) nennt man NP.

Die Klasse der polynomialen Algorithmen (bzgl. der Zeitkomplexität) nennt man P. Die bekannteste ungelöste Frage der theoretischen Informatik ist: „ $P = NP?$ “.

Zusammenfassung

- Der Zeitbedarf einer Anweisung berechnet sich aus der Anzahl der Berechnungsschritte, der Speicherplatzbedarf aus dem Bedarf an lokalen Variablen und (neuen) Objekten.
- Die Zeit- und Speicherplatzkomplexität eines Algorithmus hängt von der Größe der Eingabe ab. Im schlechtesten Fall bildet man das Maximum der Berechnungsschritte für Eingaben gleicher Größe. Analog nimmt man im mittleren Fall den Durchschnitt.
- Speicher- und Zeitkomplexität werden nach ihrer Größenordnung, der O -Notation, klassifiziert. Diese hängt im wesentlichen von der Anzahl der nötigen Schleifendurchläufe ab. Geschachtelte Schleifen erhöhen die Komplexität, im Gegensatz zu hintereinander ausgeführten Schleifen.
- Polynomial berechenbare Algorithmen heißen auch praktisch berechenbar (obwohl schon die Komplexität n^3 häufig Probleme bereitet). Exponentielle Algorithmen sind nicht praktisch berechenbar.
- Bis heute offen ist die Frage $P=NP$: ob nichtdeterministisch polynomiale Algorithmen auch polynomial sind. (Beispiel: Handelsreisender)