Complexity in grammar

Formale Komplexitätsbegriffe

Timm Lichte

HHU Düsseldorf

WS 2015/2016, 28.10.2015

SFB 991

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Komplexität?

≈ Vielschichtigkeit, Schwierigkeit — Effizienz, Ökonomie, Sparsamkeit

Komplexitäten!

- Komplexität einer formalen Grammatik
- Komplexität einer formalen Sprache
- Komplexität der Verarbeitung (Parsen/Generierung) einer Sprache anhand einer Grammatik
- Komplexität eines natürlich-sprachlichen Phänomens
- Komplexität einer natürlichen Sprache

⇒ Wie lässt sich Komplexität formal präzise charakterisieren?
Was verursacht Komplexität?

Ontologie

Wir betrachten zunächst Stringsprachen:

$$a \triangleleft w \in L_G \in \{L_G | G \in \hat{G}\}$$

a	ist ein **Buchstabe** und Element eines Alphabets Σ.
w	ist ein **Wort** und Element einer Sprache L.
L_G	ist die **Sprache** der Grammatik G.
$\hat{G}$	ist eine Menge/Klasse von Grammatiken.

Beispiele:

$$\begin{aligned}
\Sigma &= \{a, b\} \\
w &= abba \\
L &= \{abba, abab, bbb, \ldots\} \\
G &= \langle N, \Sigma, S, P\rangle && \text{(z.B. eine CFG)} \\
L_G &= \{w' | S \xrightarrow{*}_{P_G} w'\}
\end{aligned}$$

Informatische Komplexität

Intuition

Je komplexer desto schwieriger auswendig zu lernen.

$$
\begin{aligned}
K(\Sigma)&: \text{Komplexität eines Alphabets } \Sigma && = |\Sigma| \\
K(w)&: \text{Komplexität eines Wortes } w && = |w|, K(\Sigma_w), \text{Entropie} \\
K(L)&: \text{Komplexität einer Sprache } L && = |L| \\
K(G)&: \text{Komplexität einer Grammatik } G && = |G|, K(L_G) \\
K(G, w) & = |\text{Menge der möglichen Derivationen}|\ \textbf{(Ambiguität)} \\
& = \text{Summe der Längen der möglichen Derivationen} \\
& = |\text{längste Derivation}|
\end{aligned}
$$

Beschreibungskomplexität

Intuition

Je komplexer desto aufwändiger zu beschreiben.

Derivationskomplexität:

$$K(G, w) = |\text{kürzeste Derivation}|$$

Kolmogorov-Komplexität:[2]

$$K(w) = |\text{kürzeste Programm das } w \text{ ausgibt}|$$

Beispiel (Wikipedia)

Erzeugung einer Folge mit 1000 Nullen:

```
begin
  for i:= 1 to 1000
   print "0"
end
```

(Standardisierung durch die Minimal Description Length)

Relative extensionale Komplexität

Intuition

Je komplexer desto mächtiger.

Gegeben zwei Wörter w und w':

$$K(w) > K(w') \quad \longleftrightarrow \quad w \triangleright w'$$

Gegeben zwei Grammatiken G und G':

$$K(G) > K(G') \quad \longleftrightarrow \quad P_G \supset P_{G'}$$

$$K(G) > K(G') \quad \longleftrightarrow \quad K(L_G) > K(L_{G'}) \quad \longleftrightarrow \quad L_G \supset L_{G'}$$

Gegeben zwei Grammatikklassen $\hat{G}$ und $\hat{G}'$:

$K(\hat{G}) > K(\hat{G}') \quad \longleftrightarrow \quad$ Für jede $G' \in \hat{G}'$ gibt es eine $G \in \hat{G}$ und $P_{G'} = P_G$

$K(\hat{G}) > K(\hat{G}') \quad \longleftrightarrow \quad$ Für jede $G' \in \hat{G}'$ gibt es eine $G \in \hat{G}$ und $L_{G'} = L_G$

$\Rightarrow$ berühmtestes Beispiel: Chomsky(-Schützenberger)-Hierarchie[1]

Relative extensionale Komplexität von Grammatikklassen

Chomsky(-Schützenberger)-Hierarchie[1]

type 0: recursively enumerable

$\phi \to \mu$ $\qquad a^{f(n)}$

type 1: context-sensitive

$\phi \to \mu$, $|\phi| \leq |\mu|$ $\qquad a^{2^n}, a^n b^n c^n \ldots, W^k$

type 2: context-free

$A \to \mu$, $A \in N, \mu \in (N \cup T)^*$ $\qquad a^n b^m c^m d^n, WW^R$

type 3: regular

$A \to wB$ oder $A \to a$, $A, B \in N, a \in T, w \in T^*$ $\qquad a^n b^m c^k d^l$

N := die Menge der Nichtterminale
T := die Menge der Terminale (aka Σ)

Relative extensionale Komplexität von Grammatikklassen

Chomsky(-Schützenberger)-Hierarchie[1]

type 0: recursively enumerable	HPSG, TG, TM	$a^{f(n)}$
type 1: context-sensitive	LFG, LBA	a^{2^n}, $a^n b^n c^n \ldots$, W^k
type 2: context-free	CFG, PDA	$a^n b^m c^m d^n$, WW^R
type 3: regular	FSA	$a^n b^m c^k d^l$

Chomsky(-Schützenberger)-Hierarchie[1]

type 0: recursively enumerable

HPSG, TG, TM — $a^{f(n)}$

type 1: context-sensitive

LFG, LBA — a^{2^n}, $a^n b^n c^n \ldots$, W^k

mildly context-sensitive

TAG, EPDA — $a^n b^m c^n d^m$, WW

type 2: context-free

CFG, PDA — $a^n b^m c^m d^n$, WW^R

type 3: regular

FSA — $a^n b^m c^k d^l$

NL is **mildly** context-sensitive? (Joshi [3])

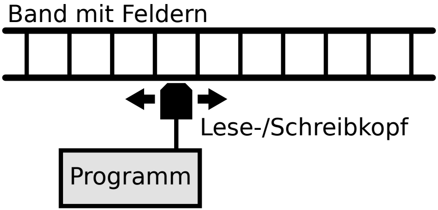
- $\supset$ CFL
- cross-serial dep.
- semi-linear
- in PTIME

Algorithmische Komplexität (Komplexitätstheorie)

Intuition

Je komplexer desto aufwändiger zu berechnen.

Modell: Turingmaschine (Turing 1936)



(aus Wikipedia)

Komponenten:

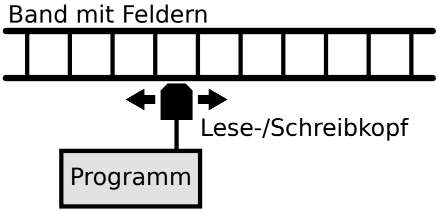
- unendlich langes **Speicherband** von linear geordneten Feldern/Zellen
- **Lese-/Schreibkopf**, der sich schrittweise auf dem Band bewegt
- **Programm** (= endlicher Automat) kontrolliert den Kopf

Algorithmische Komplexität (Komplexitätstheorie)

Intuition

Je komplexer desto aufwändiger zu berechnen.

Modell: Turingmaschine (Turing 1936)



(aus Wikipedia)

Funktionsweise:

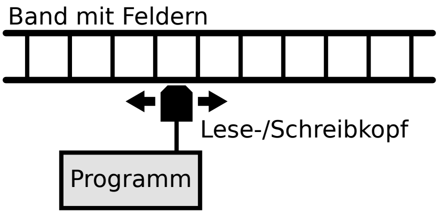
- Start: Eingabe auf dem Band, Kopf direkt davor
- Ende: Ausgabe auf dem Band, Programm im Endzustand
- $\delta(\text{Zustand}, \text{alter Zellinhalt}) = (\text{Zustand}, \text{neuer Zellinhalt}, \{R, L, N\})$

Algorithmische Komplexität (Komplexitätstheorie)

Intuition

Je komplexer desto aufwändiger zu berechnen.

Modell: Turingmaschine (Turing 1936)



(aus Wikipedia)

Wortproblem: $w \in L_G$?

$\Rightarrow$ Wie viele Schritte werden benötigt? (Zeitkomplexität)

$\Rightarrow$ Wie viele Zellen werden benötigt? (Raumkomplexität)

$\Rightarrow$ Hält die Turingmaschine? (Halteproblem)

Algorithmische Komplexität (Komplexitätstheorie)

semi-entscheidbar		
entscheidbar		
factorial time	$\mathcal{O}(n!)$	
EXPSPACE	$\mathcal{O}(2^{p(n)})$	
EXPTIME	$\mathcal{O}(2^{p(n)})$	
NPSPACE	$\mathcal{O}(n^c)$	
PSPACE	$\mathcal{O}(n^c)$	"non-efficient"
NP(TIME)	$\mathcal{O}(n^c)$	
P(TIME)	$\mathcal{O}(n^c)$	
LIN	$\mathcal{O}(n)$	"efficient"
REALTIME	$\mathcal{O}(n)$	
constant time	$\mathcal{O}(c)$	

Algorithmische × extensionale Komplexität

type 0: recursively enumerable

HPSG, TG, TM	$a^{f(n)}$	

type 1: context-sensitive

LFG, LBA	a^{2^n}, W^n	NP
	$a^n b^n c^n \ldots$, W^k	P

mildly context-sensitive

TAG, EPDA	$a^n b^m c^n d^m$, WW	$\mathcal{O}(n^6)$

type 2: context-free

CFG, PDA	$a^n b^m c^m d^n$, WW^R	$\mathcal{O}(n^3)$

type 3: regular

FSA	$a^n b^m c^k d^l$	$\mathcal{O}(n)$

Zusammenfassung

Komplexitätsbegriffe:

- informatische Komplexität
- Beschreibungskomplexität
- extensionale Komplexität
- algorithmische Komplexität

Andere Begriffe:

- Monoton versus nicht-monoton

Fragen des Seminars:

- Anwendung/Kritik in der theoretischen Linguistik?
  z.B. Rolle der Grammatikgröße in der algorithmischen Komplexität?
- Andere Komplexitätsbegriffe?

Entropie

(Shannon & Weaver 1949)

Maß für die Informationsdichte eines Wortes $w = x_1 \ldots x_n$:

$$I(x_i) = -\log_2(p_{x_i})$$

$$I_{ges} = I(x_1) + I(x_2) + \ldots + I(x_n) = \sum_{i=1}^{n} I(x_i)$$

Beispiel (aus Wikipedia)

$w = \text{Mississippi}$ mit $|w| = 11$ und $\Sigma = \{\text{i}, \text{M}, \text{p}, \text{s}\}$ und

$$p(\text{i}) = \frac{4}{11}, \;\; p(\text{M}) = \frac{1}{11}, \;\; p(\text{p}) = \frac{2}{11}, \;\; p(\text{s}) = \frac{4}{11}$$

$$\begin{aligned} I_{ges} &= 4 \cdot I(\text{i}) + 1 \cdot I(\text{M}) + 2 \cdot I(\text{p}) + 4 \cdot I(\text{s}) \\ &= 4 \cdot 1.46 bit + 1 \cdot 3.46 bit + 2 \cdot 2.46 bit + 4 \cdot 1.46 bit \\ &= 20.06 bit \approx 21 bit \end{aligned}$$

zurück

[1] Chomsky, Noam & Marcel-Paul Schützenberger. 1963. The algebraic theory of context-free languages. In P. Braffort & D. Hirschberg (eds.), *Computer programming and formal systems* (Studies in Logic and the Foundations of Mathematics 35), 118–161. Elsevier.
[2] Cover, Thomas M., Peter Gacs & Robert M. Gray. 1989. Kolmogorov's contributions to information theory and algorithmic complexity. *The Annals of Probability* 17(3). 840–865. `http://www.jstor.org/stable/2244387`.
[3] Joshi, Aravind K. 1985. Tree adjoining grammars: how much context-sensitivity is required to provide reasonable structural descriptions. In David Dowty, Lauri Karttunen & Arnold Zwicky (eds.), *Natural language parsing*, 206–250. Cambridge University Press.
[4] Shannon, Claude E. & Warren Weaver. 1949. *The mathematical theory of communication*. Urbana, IL: The University of Illinois Press.
[5] Turing, Alan M. 1936. On computable numbers, with an application to the Entscheidungsproblem. *Journal of Mathematics* 58(345-363). 5.