Biochemisch inspirierte Berechnungen

Stringoperationen

Peter Leupold

Vorlesung Wintersemester 2009/2010

Duplication

Wir haben gesehen, dass Duplikationsabschlüsse einzelner Wörter und regulärer Sprachen zum Teil regulär, zum Teil nicht regulär, abhängig von

- der Größe des Alphabets und
- der Längenbeschränkungen der Duplikationsrelation.

Für die nichtregulären Varianten ist die naheliegende Frage nun, ob sie kontextfrei sind.

Dazu müsste in gewisser Weise die linke Seite einer Regel $u \rightarrow u^2$ in ein einziges Zeichen komprimierbar sein.

Duplication – Kontextfreiheit

Für beschränkte Duplikation lässt sich eine solche Kompression bewerkstelligen.

Satz

$w^{\heartsuit \leq n}$ ist kontextfrei für beliebige n.

Beweisskizze an der Tafel...
mit Exkursion in die Komplexitätstheorie:

Satz

DSPACE(n) = DSPACE(Lin)

und

Satz

DTIME(n) = DTIME(Lin)

Duplication – Berechnungskraft

Offenes Problem

Offen ist die Frage, ob $w^{\heartsuit}$ stets kontextfrei ist, oder nicht.

Hier stellen die unbeschränkte Länge der linken Seiten sowie die unendliche Anzahl der Regeln das Problem dar.

Die Komplexität von Duplikationsabschlüssen einzelner Wörter liegt also je nach

- Größe des Alphabets und
- Längenbeschränkung

im Bereich der regulären Sprachen oder wahrscheinlich jenseits der kontextfreien Sprachen.

Duplikationswurzeln

Bislang haben wir uns damit beschäftigt, welche Wörter sich per Duplikationsregeln erzeugen lassen.

Nun stellen wir die umgekehrte Frage: aus welchen Wörtern kann ein gegebenes Wort per Duplikationsregeln erzeugt worden sein?

Von besonderem Interesse sind hierbei die minimalen Generatoren. Im Falle der Duplikation sind das die quadratfreien Wörter, die ja minimale Elemente der Halbordnung $\heartsuit^*$ darstellen.

Duplikationswurzeln

Zur Suche nach möglichen quadratfreien Ursprungswörtern verwenden wir die inverse Relation zu $\heartsuit$, bezeichnet durch $\rightarrowtail := \heartsuit^{-1}$.

Die Notationen für die längenbeschränkten Varianten werden entsprechend verwendet.

Dabei ist zu beachten, dass die Längenbeschränkung bei $\rightarrowtail_{\leq k}$ sich nun auf die rechte Seite der Regeln $u^2 \to u$ bezieht.

Das gewährleistet zum einen $\rightarrowtail_{\leq k} = (\heartsuit \leq k)^{-1}$, zum anderen hätten wir für gerades k anderenfalls $\rightarrowtail_{\leq k} = \rightarrowtail_{\leq k+1}$.

Duplikationswurzeln

Da das Wurzelsymbol häufig bei der Reduktion auf Primitive verwendet wird, tun wir dies auch hier.

Definition

Die *Duplikationswurzel* eines nichtleeren Wortes w ist

$$\sqrt[\heartsuit]{w} := IRR(\succ\!\rightarrow) \cap \{u : w \succ\!\rightarrow^{*} u\}.$$

Wie üblich dehnen wir diesen Begriff von einzelnen Wörtern auf ganze Sprachen aus:

$$\sqrt[\heartsuit]{L} := \bigcup_{w \in L} \sqrt[\heartsuit]{w}.$$

Duplikationswurzeln

Die sieben quadratfreien Wörter λ, a, b, ab, ba, aba und bab lassen sich durch folgende drei Parameter eindeutig bestimmen:

1. Anfangsbuchtaben
2. Endbuchstaben
3. Anzahl der verwendeten Buchstaben

Gleichzeitig sind diese Parameter invariant unter Duplikation, das heisst sie sind bei gegebenem Wort w für alle Wörter in $w^{\heartsuit}$ gleich.

Daher lässt sich die Wurzel über zweibuchstabigem Alphabet fast durch blosses Hinschauen finden.

Duplikationswurzeln – Beispiele

Einige Beispiele für Duplikationswurzeln:

- Durch die Anwendung von Regeln der Relation $\rightarrowtail$ erhalten wir ausgehend vom Wort $w = abcbabcbc$ die Wörter $\{abc, abcbc, abcbabc\}$.
  Die Wurzel enthält die quadratfreien dieser Wörter, also $\sqrt[\heartsuit]{abcbabcbc} = \{abc, abcbabc\}$. Durch Überprüfung aller kürzeren Wörter lässt sich feststellen, dass es keine kürzeren Beispiele mit mehrdeutiger Wurzel gibt.
- Es sind auch Mehrdeutigkeiten von höherem Grad als zwei möglich, z.B. ist für $w_3 = babacabacbcabacb$ die Wurzel
  $$\sqrt[\heartsuit]{w_3} = \{bacabacb, bacbcabacb, bacb\}.$$
- Für $w_5 = ababcbabcacbabcabacbabcab$ ist die Wurzel
  $$\sqrt[\heartsuit]{w_5} = \{abcbabcabacbabcab, abcbabcab, abcacbabcab,$$
  $$abcabacbabcab, abcab\}.$$

Duplikationswurzeln

In der Tat lässt sich eine Folge von Wörtern konstruieren, die immer mehr Duplikationswurzeln haben.

Das wirft die Frage auf, wie schnell die Funktion

$$\mathrm{duproots}(n) := \max\{|\sqrt[\heartsuit]{w}| : |w| = n\}$$

wachsen kann. Die Antwort ist: sehr schnell; genauer gesagt liegt exponentielles Wachstum vor. Es gilt

$$\frac{1}{30} 110^{\frac{n}{42}} \leq \mathrm{duproots}(n) \leq 2^n.$$

Duplikationswurzeln

Die genauen Werte der Funtion duproots(n) sind nur für eine relativ kurzes Anfangssegment bekannt.

n	: 1-8,	9-13,	14-24,	25	...
duproots(n):	1,	2,	4,	5	...

Näheres findet sich auch in der

Online Encyclopedia of Integer Sequences
`http://www2.research.att.com/~njas/sequences/A160675`

Duplikationswurzeln –Wachstum

Um das exponentielle Wachstum der Funktion duproots(n) zu beweisen, nutzen wir die einfache Zweideutigkeit

$$\sqrt[\heartsuit]{abcbabcbc} = \{abc, abcbabc\}$$

und folgende Resultate

Satz (Crochemore)

Ein uniformer Morphismus ist quadratfrei genau dann wenn er quadratfrei für alle Wörter der Länge 3 ist.

Satz (Sun)

Die Zahl $s(n)$ quadratfreier Wörter über drei Buschtaben ist mindestens $s(n) \geq 110^{\frac{n}{42}}$.

Duplikationswurzeln –Wachstum

Satz

$\frac{1}{30} 110^{\frac{n}{42}} \leq \mathit{duproots}(n) \leq 2^n$ *for all* $n > 0$.

Da die Konstruktion nur Regeln der Länge 30, d.h. Regeln aus $\rightarrowtail_{\leq 30}$ verwendet, gelten diese Schranken auch für die Funktion bduproots$_{\leq 30}$, die die maxumale Zahl der 30-beschränkten Wurzeln von Wörtern der Länge n angibt.

Satz

$\frac{1}{30} 110^{\frac{n}{42}} \leq \mathit{bduproots}_{\leq 30}(n) \leq \max\{812^{\frac{n-3}{2}}, 2^n\}$ *for all* $n > 0$.

Duplikationswurzeln

Zur Komplexität von Duplikationswurzelsprachen lässt sich unmittelbar folgendes sagen:

Satz

Wenn die Duplikationswurzel einer Sprache unendlich ist, dann ist diese Wurzel nicht kontextfrei.

Dies ergibt sich direkt aus dem Pumpinglemma für kontextfreie Sprachen.

Duplikationswurzeln sind also entweder endlich, oder relativ kompliziert.

Duplikationswurzeln

Für die beschränkte Duplikation gilt keine analoge Aussage. Es gibt zum Beispiel reguläre Wurzeln.

Für beliebiges $k \geq 1$ können wir ein zirkulär quadratfreies Wort w wählen, das länger ist als k. Solche Wörter existieren für beliebige $k \geq 18$.

Dann ist offensichtlich $\sqrt[\heartsuit \leq k]{w^+} = w^+$. Die Sprache w^+ ist also eine Art Fixpunkt.

Duplikationswurzeln

Das eben beschriebene Beispiel ist kein Einzelfall, sondern die Regel.

Satz

Die Klasse der regulären Sprachen ist abgeschlossen unter beschränkten Duplikationswurzeln, d.h.
aus $L \in REG$ *folgt* $\sqrt[\heartsuit \leq k]{L} \in REG$ *für jedes* $k > 0$.

Beweis an der Tafel...

Im Zusammenhang mit Duplikationswurzeln ergeben sich zahlreiche interessante algorithmische Probleme. Zum Beispiel:

- Berechnung der Duplikationswurzel eines gegebenen Wortes.
- Berechnung der Duplikationsgeschichte eines gegebenen Wortes.
- Berechnung eines (des kürzesten Elements) der Duplikationswurzel eines gegebenen Wortes.
- Berechnung der Funktionen duproots und bduproots.
- ...