

Biochemisch inspirierte Berechnungen

Stringoperationen III

Peter Leupold

Vorlesung Wintersemester 2009/2010

DNA Haarnadeln sind im Bereich des DNA Computing im allgemeinen ein unerwünschtes Phänomen, da die Haarnadel verhindert, dass ein Strang sich mit einem anderen zu einem Doppelstrang verbindet.

Andererseits kann die Haarnadel im Zusammenspiel mit Polymerase auch zu einer Veränderung des gegebenen Stranges führen.

Wir betrachten kurz, wie dies geschieht, und gehen dann auf eine Formalisierung dieses Mechanismus als Operation auf Strings ein.

Haarnadelvervollständigung

Definition

Bei Abstraktion von der dreidimensionalen Struktur des Moleküls zu einer einfachen Zeichenkette können wir die Haarnadelvervollständigung also folgendermaßen formalisieren:

*u**v**w**x*



*u**v**w**x**y*

*v**w**x**y*



*u**v**w**x**y*

Hierbei sind die Paare *u*–*y* und *v*–*x* jeweils komplementäre Spiegelbilder voneinander.

Beschränkte Haarnadelvervollständigung

Definition

Wie schon im Falle der Duplikation stellen sich die meisten Probleme als sehr schwierig heraus. Darum betrachten wir auch hier eine längenbeschränkte Variante.

Bei der **beschränkten** Haarnadelvervollständigung ist sowohl die Länge der komplementären Teilstränge ($v-x$) als auch die des Überhanges in ihrer Länge beschränkt.

Die Anwendung einer Haarnadelvervollständigung, bei der die Länge der komplementären Teilstränge durch p und der Überhang durch k beschränkt sind wird mit $pHC_k(w)$ bezeichnet. Die iterierte Anwendung schreiben wir $pHC_k^*(w)$. Analoge Notationen gelten für die Anwendung auf ganze Sprachen statt einzelner Wörter.

Beschränkte Haarnadelvervollständigung

Definition

Formal müssen wir unterscheiden, ob der Überhang links oder rechts liegt.

$$pHC \curvearrowright_k (w) = \{ \overline{\gamma^R} w \mid w = \alpha \beta \overline{\alpha^R} \gamma, |\alpha| = k, \alpha, \beta \in V^+, \\ \gamma \in V^*, |\gamma| \leq p \}$$

$$pHC \curvearrowleft_k (w) = \{ w \overline{\gamma^R} \mid w = \gamma \alpha \beta \overline{\alpha^R}, |\alpha| = k, \alpha, \beta \in V^+, \\ \gamma \in V^*, |\gamma| \leq p \}$$

Die beschränkte Haarnadelvervollständigung vereinigt beide Möglichkeiten.

$$pHC_k(w) = pHC \curvearrowright_k (w) \cup pHC \curvearrowleft_k (w)$$

Beschränkte Haarnadelvervollständigung

Wir betrachten ausschliesslich die iterierte Variante der Haarnadelvervollständigung.

Wie schon bei der Duplikation ist die Variante, bei der nur ein Schritt ausgeführt wird, wohl eher von theoretischem Interesse.

In der Realität liesse sich kaum kontrollieren, dass alle Stränge in einem gegebenen Zeitraum genau einen Ableitungsschritt vollzögen.

Iterated Hairpin Completion

Auch wenn die Haarnadelvervollständigung eine konzeptuell einfache Operation ist, ist sie doch mit Bezug auf die Chomsky-Hierarchie schwer zu fassen.

Das Hauptproblem hierbei ist, dass die Regeln sich auf verschiedene Teile der Zeichenkette beziehen, die unbeschränkt weit voneinander entfernt sein können.

Dagegen arbeiten Grammatikregeln sehr lokal. Ebenso können einfache Maschinenmodelle wie endliche Automaten und Kellerautomaten Anfang und Ende des Eingabewortes im wesentlichen nur einmal betrachten.

Wie die Haarnadelvervollständigung lokal werden kann

Wir verwandeln eine Operation, die an beiden Enden des Wortes agiert:

durch zirkuläre Permutation in eine lokale Operation:

$$\begin{array}{c} \text{ACTTGAGTGT} \\ \downarrow \\ \text{ACACTTGAGTGT} \\ \downarrow \\ \text{ACACTTGAGTGTGT} \end{array}$$
$$\begin{array}{c} \text{AGTGT} \odot \text{ACTTG} \\ \downarrow \\ \text{AGTGT} \odot \text{ACACTTG} \\ \downarrow \\ \text{AGTGTGT} \odot \text{ACACTTG} \end{array}$$

Die Komplementaritätsrelation hier ist die übliche für DNA: $A \cong T$ and $C \cong G$.

AGTGT ⊙ ACAC TTG



AGTGTGT ⊙ ACAC TTG

wird

AGTGT [GTGT ⊙ ACAC] ACAC TTG



AGTGTGT [GTGTGT ⊙ ACAC] ACAC TTG

und dies ist kontextfrei

Wenn wir die Ausgangssprache

- zunächst zirkulär permutieren,
- dabei das Wortende markieren,
- und schließlich mit einem GSM

dann können wir die beschränkte Haarnadelvervollständigung mit kontextfreien Regeln simulieren, und mit

- GSM,
- zirkulärer Permutation,
- und Durchschnittsbildung

die erzeugte Sprache erhalten.

Eine unmittelbare Folgerung:

Offene Frage

Ist die
Haarnadelvollständigung eines
einzelnen Wortes kontextfrei?

Satz

*Die beschränkte
Haarnadelvollständigung eines
einzelnen Wortes ist stets
kontextfrei.*

Offene Fragen

- Ist die the (beschränkte) Haarnadelvollständigung eines Wortes regulär?
- Bewahrt (beschränkte) Haarnadelvollständigung Regularität?

Haarnadelreduktion

Wie schon im Fall der Duplikationswurzel betrachten wir auch hier die umgekehrte Relation, die für ein gegebenes Wort die möglichen Ursprungswörter sucht.

Diese erhalten wir durch Haarnadelreduktion:

$uvwxy$



$uvwX$

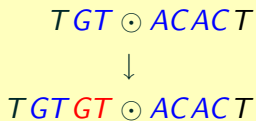
$uvwxy$



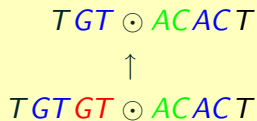
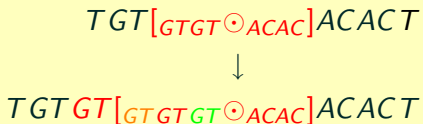
$vWxy$

Hierbei sind wieder die Paare $u-y$ und $v-x$ jeweils komplementäre Spiegelbilder voneinander.

Haarnadelreduktion



wird



wird



durch monadische
Textersetzungsregeln.

Haarnadelreduktion

Die wesentlichen benötigten Operationen, um die Haarnadelreduktion lokal zu machen und zu simulieren sind also zirkuläre Permutation und monadische Textersetzung.

Somit ist jede Sprachklasse, die unter diesen Transformationen abgeschlossen ist auch unter der Haarnadelreduktion abgeschlossen.

Satz

Die Klasse der regulären Sprachen ist abgeschlossen unter iterierter beschränkter Haarnadelreduktion.

Abschließende Bemerkungen zum DNA Computing

- In den idealisierten Formen, die wir betrachtet haben, bieten die von der DNA inspirierten Stringoperationen durchaus beachtliche Berechnungskraft, indem sie aus einzelnen Wörtern zum Teil wohl nicht kontextfreie Sprachen erzeugen können.
- Andererseits lassen sich kaum ganze Sprachklassen, d.h. zum Beispiel alle kontextfreien Sprachen, erzeugen.
- Berechnungstechnische Vollständigkeit aus endlichen oder zumindest regulären Mengen läßt sich nur mit zusätzlichen Kontrollmechanismen erreichen, die wir beim Splicing kurz angesprochen aber nicht näher behandelt haben.

Abschließende Bemerkungen zum DNA Computing

- Da weiterhin unsere Operationen im Vergleich zum Laborbetrieb stark idealisiert sind, entstünden bei einer Implementierung wohl Probleme, die die Berechnungskraft weiter schwächen würden.
- Ein vollständiger DNA Computer scheint also in weiter Ferne zu liegen.
- Koprozessoren, insbesondere für massiv parallelisierbare Probleme wie die meisten NP-vollständigen Probleme, sind hingegen bei weiteren Fortschritten in der Labortechnik durchaus denkbar.