Knuth-Bendix-Vervollständigung

Pamela Trowe

22.10.2003

1 Wortersetzungssysteme

Definition 1. *Sei Σ ein endliches Alphabet und Σ^* die Menge aller Wörter über Σ. λ ist das leere Wort, und die Konkatenation ist die Verknüpfung zweier Wörter aus Σ^*.*

Bemerkung: Die Konkatenation auf Σ^* ist assoziativ und hat λ als neutrales Element. Damit ist Σ^* das von Σ frei erzeugte Monoid.

Definition 2. *Ein Wort-/Termersetzungssystem über dem Alphabet Σ ist eine Teilmenge $S \subseteq \Sigma^* \times \Sigma^*$. Die Elemente von S werden Regeln genannt und mit (l, r) oder $l \rightarrow r$ bezeichnet. Das Wortersetzungssystem induziert die Ersetzungsrelation $\rightarrow_S \subseteq \Sigma^* \times \Sigma^*$, mit*

$$\rightarrow_S := \{(ulv, urv) | (l, r) \in S,\ u, v \in \Sigma^*\}.$$

Ihr reflexiver und transitiver Abschluss wird mit $\overset{*}{\rightarrow}_S$ und ihr reflexiver, symmetrischer und transitiver Abschluss mit $\overset{*}{\leftrightarrow}_S$ bezeichnet.

D.h. $\overset{*}{\leftrightarrow}_S$ ist ein Äquivalenzrelation. Auf der Menge der Äquivalenzklassen

$$\Sigma^* / \overset{*}{\leftrightarrow}_S := \left\{ [w]_S \,| w‘ \in \Sigma^* : w \overset{*}{\leftrightarrow}_S w‘ \right\}$$

wird die Verknüpfung $\circ$ wie folgt definiert:

$$[u]_S \circ [v]_S := [uv]_S \,.$$

Das bedeutet $\circ$ wohldefiniert, denn mit $u \overset{*}{\leftrightarrow}_S u‘$ und $v \overset{*}{\leftrightarrow}_S v‘$ folgt $uv \overset{*}{\leftrightarrow}_S uv‘ \overset{*}{\leftrightarrow}_S u‘v‘$. Damit bildet $\Sigma^* / \overset{*}{\leftrightarrow}_S$ zusammen mit der Verknüpfung $\circ$ und $[\lambda]_S$ als neutralem Element einen Monoiden über dem freien Monoiden Σ^* modulo $\overset{*}{\leftrightarrow}_S$.

Ist M isomorph zu $\Sigma^* / \overset{*}{\leftrightarrow}_S$, dann wird (Σ, S) die Monoid-Darstellung von M mit Erzeugenden Σ und Relationen S genannt.

Der einfachste Weg die Darstellung eines Monoiden zu finden ist M selbst als Erzeugnis zu nehmen und die Verknüpfungstafel als Relationen zu wählen.

Die Schwierigkeit ist es nun zu bestimmen, ob zwei Wörter u, v in der selben Äquivalenzklasse liegen. Dieses wird auch als Wortproblem bezeichnet, welches für allgemeine Systeme nicht entscheidbar ist. Wenn wir jedoch wissen, dass das System konvergent (d.h. noethersch und konfluent) ist, dann gilt $u \overset{*}{\leftrightarrow} v$, wenn sie die gleiche (eindeutige) Normalform besitzen.

Beispiel 3. *Sei $M = \{a, b, c, d\}$ und $S = \{ab \to \lambda, ba \to \lambda, cd \to \lambda, dc \to \lambda\}$.*

Dann liegen z.b. „abdacd“, „badacd“, „dcababda“ in der selben Äquivalenzklasse, da sie sich alle bis zu „da“ reduzieren lassen.

Konvergenz ist somit eine wichtige Eigenschaft für Wortersetzungssysteme. Diese werden wir im folgenden genauer untersuchen und am Ende einen Algorithmus erhalten, der dann ein gegebenes System auf (lokale) Konfluenz testet, und es gegebenenfalls zu einem solchen vervollständigen kann.

2 Ordnungen

Definition 4. *Eine Relation $\succeq$ auf einer Menge M ist eine Partialordnung, falls sie reflexiv, transitiv und anti-symmetrisch ist, d.h. aus $a \succeq b$ und $b \succeq a$ folgt $a = b$.*

Eine Partialordnung heißt total, falls für a und b aus M immer $a \succeq b$ oder $b \succeq a$ gilt. Mit $\succ$ bezeichnen wir den irreflexiv transitiven Teil von $\succeq$, d.h. $a \succ b$ genau dann wenn $a \succeq b$ und $a \neq b$. Eine Partialordung heißt wohlfundiert, falls es keine unendliche Kette der Form $a_0 \succ a_1 \succ a_2 \succ ...$ gibt.

Definition 5. *Eine Ordnung $\geq$ auf Σ^* heißt verträglich, falls für alle Wörter $u, v, x, y \in \Sigma^*$ gilt:*

1. *$u \geq \lambda$ und*
2. *$u > v$ impliziert $xuy > xvy$.*

Mit Hilfe solcher verträglichen wohlfundierten Ordnungen können wir nun ein hinreichendes Kriterium für die Terminierung eines Reduktionssystems geben.

Lemma 6. *Es sei (Σ, S) ein Reduktionssystem und $\succeq$ eine verträgliche wohlfundierte Partialordnung auf Σ^* mit $S \subseteq \succ$. Dann ist (Σ, S) noethersch.*

Bemerkung:

Die Bedingung $S \subseteq \succ$ bedeutet: $\forall (l, r) \in S$ gilt: $l \succ r$.

Definition 7. *Ein Wort $u \in \Sigma^*$ heißt „länge-lexikographisch größer“ als ein Wort v falls entweder*

1. *$|u| > |v|$ oder falls*
2. *$|u| = |v| = n$ und für ein $1 \leq k \leq n$ gilt: für $j < k$: $u(j) = v(j)$ und $u(k) > v(k)$*

(Hierbei bedeutet $|u|$ die Länge des Wortes u und $u(j)$ mit $1 \leq j \leq |u|$ die „j-ten Buchstaben“ von u.)

Bemerkung: Die länge-lexikographische Ordnung $\succ$ ist eine verträgliche wohlfundierte Ordnung auf Σ^*.

3 Kritische Paare

Bei der Untersuchung von Wort-/Termersetzungsystemen interessiert man sich besonders für die lokale Konkfluenz, da sie bei noetherschen Systemen ausreicht, die Existenz und Eindeutigkeit der Normalformen zu erhalten.

Folgendes Beispiel zeigt, in welchen Fällen ein Wort zwei direkte Nachfolger besitzt:

Beispiel 8. *Sei das Wortersetzungssystem S über dem Alphabet $\Sigma = \{a, b, c, d\}$ mit folgenden Regeln gegeben:*

$$abcd \rightarrow r_1 \tag{1}$$
$$bc \rightarrow r_2 \tag{2}$$
$$dc \rightarrow r_3. \tag{3}$$

Die rechten Seiten der Relation sind zunächst nicht von Bedeutung und r_1, r_2, r_3 seien beliebige Wörter in Σ^*.

- *nicht überlappende Regelanwendung:*

  $$br_1abca \;{}_S\!\xleftarrow{(1)} b\underline{abcd}a\overline{bc}a \xrightarrow{(2)}_S babcdar_2a.$$

  Hier ist die lokale Konfluenz nicht bedroht, da die entstandenden Wörter einen gemeinsamen Nachfolger besitzen:

  $$br_1a\overline{bc}a \xrightarrow{(2)}_S br_1ar_2a \;{}_S\!\xleftarrow{(1)} b\underline{abcd}r_2a.$$

  Das bedeutet die jeweils fehlenden Schritte zum gemeinsamen Nachfolger br_1ar_2a sind unabhängig von der vorhergehenden Regelanwendung, da die zuerst angewendeten Regeln nicht überlappende Teilwörter in w ersetzt haben.

- *überlappende Regelanwendungen - geschachtelt:*

  $$br_1a \;{}_S\!\xleftarrow{(1)} b\underline{a\overline{bc}d}a \xrightarrow{(2)}_S bar_2da.$$

  Hier zerstört die Anwendung der einen Regel, die Anwendbarkeit der anderen Regel. Ob br_1a und bar_2da einen gemeinsamen Nachfolger haben, hängt deshalb von den rechten Regelseiten und evtl. weiteren Regeln ab. Eine solche Situation ist kritisch für die lokale Konfluenz und muss daher genauer überprüft werden.

- *überlappende Relanwendung - verzahnt:*

  Genauso verhält es sich in der folgenden Situation, beide Regeln überlappen sich, ohne dass die eine die andere enthält:

  $$br_1ca \;{}_S\!\xleftarrow{(1)} b\underline{abc\overline{d}}\overline{c}a \xrightarrow{(3)}_S babcr_3a.$$

  Auch hier hängt die Existenz eines gemeinsamen Nachfolgers von weiteren Regeln ab.

Das obige Beispiel hat gezeigt, welche Situationen auftreten können, wenn zwei Regeln gleichzeitig anwendbar sind. Offenbar ist es im ersten Fall egal welche Regel zuerst angewendet wird, da die Anwendbarkeit der zweiten und damit auch die lokale Konfluenz nicht zerstört wird. Anders verhält es sich in den Fällen, bei denen eine Überlappung vorliegt. Allerdings sieht man, dass der linke und rechte Kontext der Überlappung keine Rolle spielt. Das liegt daran, dass die Relation $\rightarrow_S$ nach Definition unter linkem und rechtem Kontext abgeschlossen sind. D.h. haben die Wörter s_1 und s_2 einen gemeinsamen Nachfolger, dann auch die Wörter us_1v und us_2v. Diese Überlegungen führen zur folgender Definition:

Definition 9. *Sei (Σ, S) ein Wortersetzungssystem mit Regeln $l_1 \rightarrow r_1$ und $l_2 \rightarrow r_2 \in S$.*

1. *Gilt $l_1 = ul_2v$ mit $u, v \in \Sigma^*$, dann ist (r_1, ur_2v) ein kritisches Paar für (Σ, S).*

2. Gilt $l_1v = ul_2$ mit $u, v \in \Sigma^*$, und $|l_1| > |u|$, dann ist (r_1v, ur_2) ein kritisches Paar für (Σ, S).

Ein kritisches Paar (s_1, s_2) für S heißt harmlos, falls $s_1 \downarrow s_2$ gilt.

Für das vorangegangene Beispiel gibt es zwei kritische Paare:

- $(abcd) = a(bc)d$ mit $u = a$ und $v = d$ liefert das kritische Paar (r_1, ar_2d),
- $(abcd)c = abc(dc)$ mit $v = c$ und $u = abc$ liefert $(r_1c, abcr_3)$.

Satz 10. *Ein Wortersetzungsystem ist genau dann lokal konfluent, wenn alle seine kritischen Paare harmlos sind.*

Beweis:

„$\Rightarrow$:" Sei (s_1, s_2) ein kritisches Paar für S, dann gibt es laut Definition Regeln $l_1 \to r_1$ und $l_2 \to r_2$ in S, mit $l_1 = ul_2v$ oder $l_1v = ul_2$, $u, v \in \Sigma^*$. Dann gilt im ersten Fall $s_1 = r_1 \leftarrow_S l_1 = ul_2v \to_S ur_2v = s_2$ und im zweiten Fall $s_1 = r_1v \leftarrow_S l_1v = ul_2 \to_S ur_2 = s_2$. Mit der lokalen Konfluenz folgt dann, dass (s_1, s_2) harmlos ist.

„$\Leftarrow$:" Sei $w \in \Sigma^*$ ein Wort, für das es Regeln $l_1 \to r_1$ und $l_2 \to r_2$ in S gibt, mit $w \to_{\{l_1 \to r_1\}} s_1$ und $w \to_{\{l_2 \to r_2\}} s_2$. Aus Symmetriegründen brauchen wir nur die Fälle untersuchen in denen die zweite Regel nur rechts von der ersten angewendet werden kann und da $\xrightarrow{*}_S$ unter linkem und rechtem Kontext abgeschlossen ist, reicht es folgende Fälle zu betrachten.

<u>1.Fall:</u> $w = l_1 = ul_2v$

Dann gilt $r_1 \leftarrow_{\{l_1 \to r_1\}} w \to_{\{l_2 \to r_2\}} ur_2v$. Damit ist (r_1, ur_2v) ein kritisches Paar und laut Voraussetzung harmlos.

<u>2.Fall:</u> $w = l_1v = ul_2$

(a) $|l_1| > |u|$, dann ist (r_1v, ur_2) wie oben ein kritisches Paar, also harmlos.

(b) $|l_1| \le |u|$, dann existiert ein $w‘ \in \Sigma^*$ mit $l_1v = l_1w‘l_2 = ul_2$. Dann gilt $s_1 = r_1w‘l_2$ und $s_2 = l_1w‘r_2$ und es folgt $r_1w‘l_2 \to_S r_1w‘r_2 \leftarrow_S l_1w‘r_2$. Mit anderen Worten: die Regeln können unabhängig voneinander angewendet werden, und führen zum selben (gemeinsamen) Nachfolger.

□

Folgerung 11. *Lokale Konfluenz, und damit auch Konfluenz, ist für endliche und terminierende (noethersche) Wortersetzungssysteme entscheidbar.*

Beweis:

Für ein endliches System gibt es nur endlich viele kritische Paare, die berechnet werden können. Für jedes Paar (s_1, s_2) berechnen wir die Normalformen. Sei also $\hat{s}_1$ die Normalform von s_1 und $\hat{s}_2$ die von s_2.

Gilt für alle kritischen Paar $\hat{s}_1 = \hat{s}_2$, dann ist das System lokal konfluent. Existiert jedoch ein Paar mit $\hat{s}_1 \neq \hat{s}_2$, dann ist das System nicht (lokal) konfluent.

□

4 Knuth-Bendix-Vervollständigung

Der folgende Algorithmus untersucht ein System (Σ, S) auf lokale Konfluenz, indem er alle kritischen Paare bildet und sie auf Harmlosigkeit testet. Ist das System endlich und ist eine verträgliche wohlfundierte Ordnung gegeben, dann läßt es sich zu einem konvergenten System vervollständigen, das den selben Monoiden definiert.

Prozedur:

Knuth-Bendix-Vervollständigung

Eingabe: ein endliches WES (Σ, S) und eine totale verträgliche wohlfundierte Ordnung $\succeq$;
Definiere: $R := \{(l,r) |\ l \succ r,\ (l,r) \in S \vee (r,l) \in S\}$ und
$B := \{((l_1,r_1),(l_2,r_2)) |\ (l_1,r_1),(l_2,r_2) \in S\}$.

Solange $B \neq \emptyset$ gilt, wähle zufällig ein Element $((l_1,r_1),(l_2,r_2))$ aus B und entferne es aus dieser Menge.

Bestimme alle kritischen Paare von $((l_1,r_1),(l_2,r_2))$.
Für jedes kritische Paar (s_1, s_2) bestimme die irreduziblen Reste bezüglich S,
(reduziere dabei ohne eine bestimmte Strategie anzuwenden.)
Sei s_1‘ der irreduzible Rest von s_1 und s_2‘ der von s_2.

Falls $s_1‘ \neq s_2‘$ gilt, dann setze $l_{neu} := max_{\succ}\{s_1‘, s_2‘\}$ und $r_{neu} := min_{\succ}\{s_1‘, s_2‘\}$.
Ersetze R dann durch die Menge $R \cup \{(l_{neu}, r_{neu})\}$
und B durch $B \cup \{((l,r),(l_{neu},r_{neu})),((l_{neu},r_{neu}),(l,r)) | (l,r) \in R\}$;

Gilt $B = \emptyset$, dann erhält man (Σ, R) als neues konvergentes Ersetzungssystem.

Da das Wortproblem für allgemeine Wortersetzungssysteme nicht entscheidbar ist, terminiert diese Prozedur im allgemeinen nicht. Wenn sie jedoch abbricht, dann erhalten wir ein noethersches und konfluentes Wortersetzungssystem. Dieses repräsentiert den selben Monoiden, wie das ursprüngliche System.