

Dreibeinige PCP-Biber

MARIO SCHMIDT, HEIKO STAMER und JOHANNES WALDMANN

*Institut für Informatik, Universität Leipzig
Augustusplatz 10–11, D-04109 Leipzig, Germany
e-mail: pcp@informatik.uni-leipzig.de*

ABSTRACT

We look for small PCP instances that are hard: They consist of only a few pairs of short words, but their minimal solution is very long.

In particular, we collect information on PCP instances consisting of three pairs. We list the known record holders, then describe a family of hard instances (related to D0L systems), and finally conjecture an upper bound for hardness.

Keywords: Post Correspondence Problem, Busy Beaver

1. Einleitung

Oft kann man dadurch Informationen über eine Problemklasse gewinnen, daß man nach kleinen, aber möglichst schweren Instanzen sucht. Ein klassisches Beispiel ist die Suche nach Busy-Beaver-Turingmaschinen [Mar00].

Gelingt es zudem, die Schwere einer Instanz durch eine berechenbare Funktion ihrer Größe zu beschränken, erhält man Aussagen über Entscheidbarkeit und Komplexität.

Wir wenden diese Technik auf eingeschränkte Klassen des Postschen Korrespondenz-Problems (PCP) an.

2. Bezeichnungen und Voraussetzungen

Eine *PCP-Instanz* ist eine endliche Liste $[(u_1, v_1), \dots, (u_n, v_n)]$ von Wörtern $u_i, v_i \in \Sigma^+$, die zwei Morphismen $\phi, \psi : \Gamma^* \rightarrow \Sigma^*$ (mit $\Gamma = \{1, \dots, n\}$) durch $\phi(i) = u_i, \psi(i) = v_i$ bestimmen. Eine *Lösung* einer Instanz (ϕ, ψ) heißt jedes nichtleere Wort aus der Menge $E(\phi, \psi) = \{w : \phi(w) = \psi(w)\}$. Es ist wohlbekannt, daß die Menge der lösbaren Instanzen nicht entscheidbar ist [Pos46].

Die *Größe* einer Instanz ist die Anzahl ihrer Wortpaare, d. h. $|\Gamma|$. Wir bezeichnen mit $\text{PCP}(n)$ die Menge aller lösbaren Instanzen der Größe n . Die Menge $\text{PCP}(2)$ ist entscheidbar [EKR82], aber $\text{PCP}(7)$ nicht [MS96]. Wir interessieren uns hier für $\text{PCP}(3)$. (Im folgenden betrachten wir nur Instanzen mit Ziel-Alphabet $\Sigma = \{0, 1\}$.)

Die *Weite* einer Instanz ist die größte vorkommende Wortlänge $\max\{|u_i|, |v_i|\}$. Die Menge aller lösbaren Instanzen mit Größe n und Weite w heißt $\text{PCP}(n, w)$. (Diese Mengen sind endlich.) Wir fragen nach den Instanzen aus $\text{PCP}(3, w)$, für die die Länge einer kürzesten Lösung möglichst groß wird.

Beispiel 1 *Die Instanz*

$$P_1 = \begin{pmatrix} 110 & 0 & 1 \\ 1 & 1 & 01 \end{pmatrix}$$

hat die kürzeste Lösung AACBCCBC.

Zu jeder Instanz $P = (\phi, \psi : \Gamma \rightarrow \Sigma)$ definieren wir den unendlichen, gerichteten, kantenmarkierten Graphen $G(P)$: seine Knoten sind alle möglichen *Differenzen* $\{(p, q) : p, q \in \Sigma^*, p = \epsilon \vee q = \epsilon\}$, und er enthält eine Kante $(p, q) \xrightarrow{i} (p', q')$ genau dann, wenn $\exists c \in \Sigma^* : p \cdot u_i = c \cdot p' \wedge q \cdot v_i = c \cdot q'$. Eine Lösung von P ist dann ein nichtleerer Pfad von $\text{START}=(\epsilon, \epsilon)$ zurück zu START in $G(P)$.

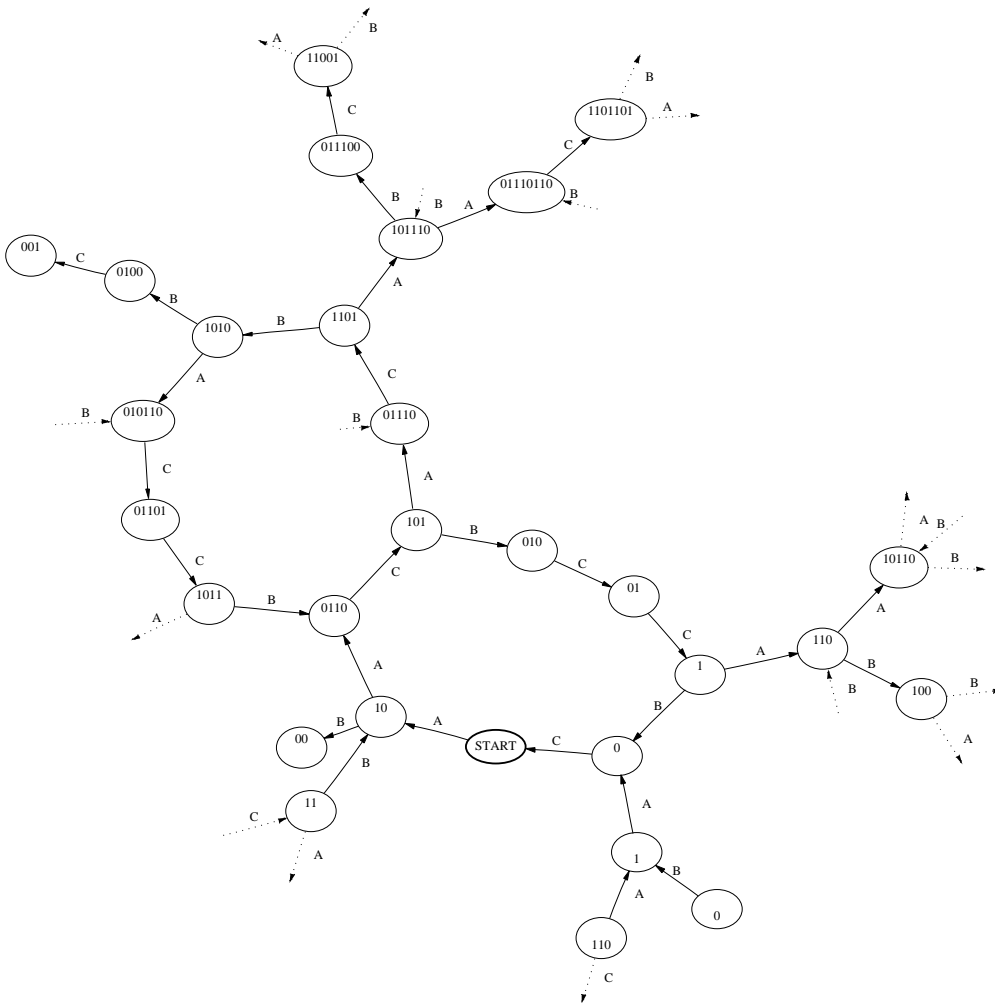


Figure 1: von START erreichbare Knoten (der Weite ≤ 6) in $G(P_1)$

Bei der maschinellen Suche nach Lösungen einer Instanz kommt es darauf an, die Menge der bereits betrachteten Knoten in $G(P)$ effizient zu verwalten.

3. Schwierige PCP-Instanzen

Wir listen hier die derzeitigen Rekordhalter. (Aktuelle Informationen sind immer auf der Webseite <http://www.informatik.uni-leipzig.de/~pcp/>)

(Größe, Weite)	Instanz	kürzeste Lösung	Autor
(3,3)	$\begin{pmatrix} 0 & 1 & 001 \\ 001 & 0 & 1 \end{pmatrix}$	75	Lorentz [Lor00], Waldmann
(3,4)	$\begin{pmatrix} 101 & 1 & 010 \\ 1 & 01 & 1101 \end{pmatrix}$	216	Zhao [Zha01]
(3,5)	$\begin{pmatrix} 11010 & 1 & 11111 \\ 11 & 10101 & 01 \end{pmatrix}$	189	Stamer
(4,3)	$\begin{pmatrix} 000 & 0 & 11 & 10 \\ 0 & 111 & 0 & 100 \end{pmatrix}$	204	Lorentz [Lor00]
(4,4)	$\begin{pmatrix} 1010 & 11 & 0 & 01 \\ 100 & 1011 & 1 & 0 \end{pmatrix}$	256	Zhao [Zha01]

4. Spezielle PCP(3)

Es ist uns aufgefallen, daß viele schwere PCP(3)-Instanzen die Form $M(u, v) = \begin{pmatrix} 0 & 1 & u \\ v & 0 & 1 \end{pmatrix}$ haben, beispielsweise das Rekord-PCP(3,3). (Dort gilt sogar $u = v$. Siehe auch [HH01] über eine Verallgemeinerung.)

Um eine Lösung für $M(u, v)$ zu suchen, können wir, von links beginnend, beliebig lange ausschließlich die ersten beiden Paare benutzen. Wir erhalten so Anfangsstücke des vom Morphismus $0 \mapsto v, 1 \mapsto 0$ erzeugten unendlichen DOL-Wortes. Wir können auch von rechts beginnen und ausschließlich die letzten beiden Paare benutzen, und erhalten (gespiegelt) Anfangsstücke des DOL-Wortes von $0 \mapsto 1, 1 \mapsto \bar{u}$. Die Frage dabei ist, ob und wie sich diese beiden Wörter schließlich treffen. Die Lösungen haben im Allgemeinen jedoch *nicht* die Form $\{A, B\}^* \{B, C\}^*$.

Ein Spezialfall dieses Musters sind Fibonacci-Instanzen $F(u) = M(u, 01)$. Die ersten beiden Paare ergeben das Fibonacci-Wort $f = 01001010010\dots$. Die Instanz $F(1001)$ besitzt die minimale Lösungslänge 78, und $F(001001)$ erreicht 120. Es ist klar, daß $F(u)$ nur dann lösbar ist, wenn u ein Faktor von f ist. Diese Bedingung ist jedoch nicht ausreichend, denn beispielweise $F(1001001)$ ist unlösbar.

5. Notwendige Bedingungen für lösbare Instanzen

Wir sind dabei, maschinell *alle* Instanzen aus PCP(3,3) auf Lösbarkeit hin zu überprüfen. Dabei kommt es natürlich darauf an, Lösungen schnell zu finden, aber auch darauf, unlösbare Instanzen schnell zu erkennen.

Bei PCP(3) kann man aus den Parikh-Vektoren der u_i, v_i zwei homogene lineare Gleichungen für die drei Elemente des Parikh-Vektors jeder Lösung bestimmen. D. h. das Verhältnis der Buchstabenanzahlen A, B, C in der Lösung steht (bis auf pathologische Fälle) von vornherein fest, und kann zur Beschleunigung der Suche benutzt werden.

6. PCP-Familien und Wachstumsraten

Lorentz [Lor00] bemerkt, daß die Familie $\begin{pmatrix} 0 & 10^n \\ 0^n 1 & 0 \end{pmatrix}$ mit minimalen Lösungen $A^n B^n$ das stärkste bekannte Wachstum (der Länge einer kürzesten Lösung, betrachtet als Funktion der Weite der Instanz) innerhalb PCP(2) erreicht. Jedoch scheint es keinen einfachen Beweis dafür zu geben, daß minimale PCP(2)-Lösungen nur linear lang sind, denn das würde einen einfachen Beweis der Entscheidbarkeit von PCP(2) implizieren.

Welches sind schnell wachsende PCP(3)-Familien? Wir erreichen quadratisches Wachstum durch $\begin{pmatrix} 0 & 0^n & 1^n 0 \\ 00 & 1 & 0 \end{pmatrix}$ mit minimalen Lösungen $A^{n^2} B^n C$, und konnten das bisher nicht übertreffen.

Es gibt einige Familien, deren Verhalten nicht so offensichtlich ist. Zum Beispiel ist die Lösungslänge von $\begin{pmatrix} 11010 & 1 & 1^n \\ 11 & 10101 & 01 \end{pmatrix}$ dann groß, wenn 2 eine primitive Wurzel modulo $4n - 1$ ist. Das ist für $n = 5$ der Fall und gibt das derzeitige Rekord-PCP(3,5). Die Lösungslänge ist jedoch selbst dann quadratisch beschränkt.

7. Zusammenfassung und Ausblick

Wir vermuten, daß PCP(3) entscheidbar ist. Die bisher vorliegenden Daten legen Beweisversuche aus zwei Richtungen nahe:

1. spezielle PCP(3):

(a) Zeige, daß die Lösbarkeit von $M(u, v) = \begin{pmatrix} 0 & 1 & u \\ v & 0 & 1 \end{pmatrix}$ entscheidbar ist.

Spezialfall: zeige, daß Fibonacci-PCPs $F(u) = M(u, 01)$ entscheidbar sind.

(b) Zeige, daß sich andere PCP(3) auf die Form $M(u, v)$ reduzieren lassen.

Zum Beispiel: welcher Zusammenhang besteht zwischen dem Rekord-PCP(3,4)

$\begin{pmatrix} 101 & 1 & 010 \\ 1 & 01 & 1101 \end{pmatrix}$ und dem Fibonacci-PCP $F(10100101) = \begin{pmatrix} 0 & 1 & 10100101 \\ 01 & 0 & 1 \end{pmatrix}$?

Beide haben eine kürzeste Lösung der Länge 216.

2. Wachstum:

(a) Zeige, daß eine kürzeste Lösung von PCP(3, w) die Länge $O(w^2)$ besitzt — oder finde eine stärker wachsende Familie.

In jedem Fall würde eine vergrößerte Sammlung von Busy-Beaver-Instanzen und -Instanzfamilien weiterhelfen. Wir laden deswegen alle Interessenten herzlich ein, diese Tiere zu jagen und an unser Museum zu schicken.

Fordern Sie auch Ihre Studenten zur Mitarbeit auf! Das Thema eignet sich nach unserer Erfahrung sehr gut zum Erlernen und Ausprobieren von implementierungs-technischen, aber auch formal-sprachlichen Methoden.

Auf unserer Webseite <http://www.informatik.uni-leipzig.de/~pcp/> finden Sie neben den aktuellen Rekorden auch ein Online-PCP-Puzzlespiel, für das täglich neue (mittelschwere) Instanzen generiert werden.

Literatur

- [EKR82] A. Ehrenfeucht, J. Karhumäki, and G. Rozenberg. The (generalized) post correspondence problem with lists consisting of two words is decidable. *Theoretical Computer Science*, 21(2):119–144, November 1982.
- [HH01] Vesa Havala and Tero Harju. Some new results on post correspondence problem and its modifications. TUCS Technical Report 338, Turku Centre for Computer Science, January 2001. <http://www.tucs.fi/publications/techreports/TR338.html>.
- [Lor00] Richard J Lorentz. Creating difficult instances of the post correspondence problem. 2nd International Conference on Computers and Games, Hamamatsu, 2000. <http://www.etl.go.jp/etl/suiron/~ianf/cg2000/>.
- [Mar00] Heiner Marxen. Busy beaver. <http://www.drb.insel.de/~heiner/BB/index.html>, 2000.
- [MS96] Yuri Matiyasevich and Géraud Sénizergues. Decision problems for semi-Thue systems with a few rules. In *Proceedings, 11th Annual IEEE Symposium on Logic in Computer Science*, pages 523–531, New Brunswick, New Jersey, 27–30 July 1996. IEEE Computer Society Press.
- [Pos46] Emil Post. A variant of a recursively unsolvable problem. *Bulletin of the American Mathematical Society*, 52:264–268, 1946.
- [Zha01] Ling Zhao. Web pages on posts correspondence problem. <http://games.cs.ualberta.ca/~zhao/PCP/intro.htm>, 2001.