Bifix Codes

Definition: set of words both a prefix and a suffix code
=antichains both for the prefix order and for the suffix order.

References:
Gilbert and Moore (1961)
Schützenberger (1956)
Berstel, Perrin (1984)

Examples

1. words of fixed length : $aaa + aab + \ldots$
2. $a + ba^*b$
3. group codes : $X^* = \varphi^{-1}(e)$ for $\varphi A^* \to G$ morphism into a group.
4. first returns to zero in coin tossing.
5. others : $aaa + ab + bba + bbb$.

The indicator

$$\begin{aligned} L_X(s) &= 1 + \text{ length of } s \\ &\quad -\text{number of factors of } s \text{ in } X \\ &= \#\text{ prefixes of } s \text{ without a suffix in } X \\ &= \#\text{ suffixes of } s \text{ without a prefix in } X \end{aligned}$$

Actually

$$\begin{aligned} L_X &= A^*(1 - X)A^* \\ &= A^*U \\ &= VA^* \end{aligned}$$

where $U = A^* - XA^*$ and $V = A^* - A^*X$.

How does L_X increase?

$$L_X(ua) - L_X(u) = \begin{cases} 0 & \text{if } ua \in A^*X \\ 1 & \text{otherwise} \end{cases}$$

Maximality

Maximal biprefix codes: for set inclusion.

For a biprefix code which is thin (i. e. there is some word w which is not a factor of X), the following are equivalent:

- X is a maximal biprefix code
- X is complete (i.e. each word is a factor of X^*).
- L_X is bounded.

The maximal value of L_X is called the *degree* of X.

Example

The following example already appears in the paper of Schützenberger of 1956 ('Une théorie algébrique du codage').

$$\begin{aligned} X &= aaa + aaba + aabb + ab \\ &\quad + baa + baba + babb + bba + bbb \end{aligned}$$

	prefix order		suffix order
1	aaa	1	aaa
2	$aaba$	5	baa
3	$aabb$	2	$aaba$
4	ab	6	$baba$
5	baa	8	bba
6	$baba$	4	ab
7	$babb$	3	$aabb$
8	bba	7	$babb$
9	bbb	9	bbb

Binary trees

Completion

Theorem (Zhang and Shen, 1994)*Any rational biprefix code is included in a maximal one.*

Solution of a conjecture by Jean Berstel and D. P. (1984).

Previous results

1. Analogous statement for general codes: Ehrenfeucht and Rosenberg, 1983
2. statement for finite biprefix codes : D. P. 1984.
3. Statement for codes with bounded deciphering delay: V. Bruyère and L. Zhang, 1990.
4. Statement for codes with bounded synchronization delay : V. Bruyère, 1997.

Example

Any biprefix code containing $a + bb$ is infinite: one may not add ba nor baa nor $baaa$ nor ...

Rational completion:

$$a + ba^*b$$

A method for finite biprefix codes

Idea: add to X a set Y such that no word of Y is a factor of another.

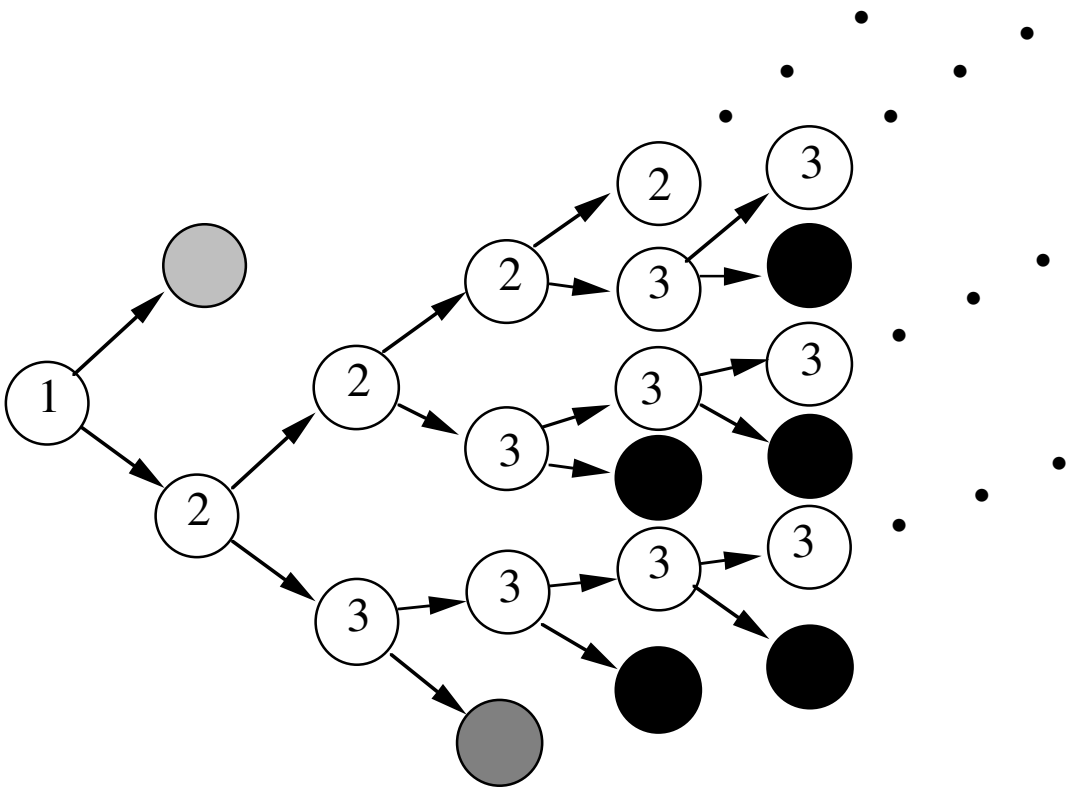
Theorem : *Let X be a finite biprefix code and let $n > \max\{L_X(x) | x \in X\}$.*

Let Y be the set of words y minimal for the factor ordering such that $L_X(y) = n$.

Then $Z = X+Y$ is a maximal rational biprefix code.

Proof:

- Z is biprefix: any $x \in X$ and $y \in Y$ are incomparable for the prefix and the suffix orders.
- Z is maximal: for any $u \in A^*$ there is a $v \in A^*$ such that $L_X(uv) > n$. Then uv has a prefix in Z.
- Z is rational: obvious.



Example

Let

$$X = a + bbb$$

Then

$$Z = a + ba^*ba^*b$$

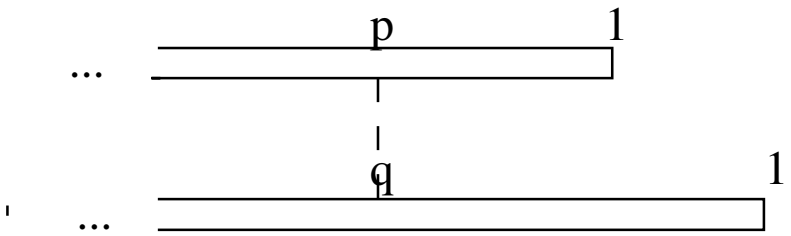
The chinese way

When X is infinite, L_X might be unbounded on X. The idea is to progressively complete X, beginning with the set M_X=$\{$words of $P(I_X)$ minimal for the factor order$\}$ where

I_X is the set of words incomparable to X for the prefix order and

$P(S) = A^* - SA^*$ is the set of *first words* of S.

Miracle : L_X is bounded M_X (by the number of states of the automaton).



Chains

Let X be a rational biprefix code recognized by an n-state automaton. Let

p= the maximal length of chains of prefixes of X for the suffix order.

s= the maximal length of chains of suffixes of X for the prefix order.

m= the maximal value of L_X on M_X

Then

$$m \leq p \leq n, \quad s \leq n$$

The algorithm

Completion in two nested loops:

while $I_X \neq \emptyset$ **do**
{
  $F := M_X$;
  $n := \max\{L_X(u) \mid u \in F\}$;
  while $F \neq \emptyset$ **do**
  {
    $Z := P\{u \in FA^* \mid L_X(u) = n\}$;
    $F = Z \cap A^{-}X$
    $Y := Z - A^{-}X$;
    $X := X + Y$;
    $n := n + 1$;
  }
}

where
$A^{-}X$=$\{$suffixes of $X\}$.

Analysis

Correction: The result is clearly a maximal biprefix code. If the algorithm terminates, it is also rational.

Termination:

- Inside loop: at most s iterations.
- External loop: at most p iterations.

Example

Let

$$X = ba^*bb$$

We have $M_X = a$ and completing M_X gives

$$X = a + ba^*bb$$

A second iteration allows to complete X as

$$X = a + ba^*ba^*b$$

My way

Let X be a biprefix code and let Y be the set of words

- incomparable with X for the prefix and the suffix order
- minimal for this property in the factor order

Then

$$X' = X + Y$$

is again biprefix. Continue...

Can this really loop for ever when X is rational?

Example

Let

$$X = a + bbb$$

Then

$$Y = ba^+b$$

A second iteration gives

$$Y' = bb(a^+b)^*b$$

A new result

Theorem (Véronique Bruyère & D.P.)
A bifix code X is a subset of a thin maximal bifix code iff

(i) the chains of prefixes for the suffix order are bounded.

(ii) the chains of suffixes for the prefix order are bounded.

Examples:

$$X = \{a^n b^n \mid n \geq 1\}$$
$$X = \{ba^n b^n a \mid n \geq 1\}$$

the second one is included in

$$X' = aa + ab + ba^+b^+a + bb$$