Applied Algorithmics COMP526 – tutorial 4

L.A. Gąsieniec and D. Cartwright

1 Questions

1.1 Non-periodicity and witness table

Fibonacci language contains words defined recursively. In particular, $w_0 = a$, $w_1 = b$, and $w_i = w_{i-1} \cdot w_{i-2}$, for all integer $i \geq 2$. E.g., $w_2 = ba$, $w_3 = bab$, $w_4 = babba$, etc. (Note that the length of w_i corresponds to the length of i^{th} Fibonacci number f_i, where similarly $f_0 = 1, f_1 = 1$, and $f_i = f_{i-1} + f_{i-2}$, for all integer $i \geq 2$.)

Construct the witness table (see notes from the lecture) for w_6.

1.2 String matching with don't care symbols

Given a pattern $P = 10 * 1$ and a text $T = 1010110010110101$. Recall the pattern matching algorithm for patterns equipped with the don't care symbol $*$ (symbol $*$ matches both 1s and 0s). Show (via computing appropriate values of convolution vectors) that P occurs at position 5 (recall that symbols are counted from position 0) in T, and that P does not occur at position 9.

2 Solutions

2.1 Non-periodicity and witness table

Recall that the witness table is stored in the array $W[0..|w_6|-1]$, where $W[i]$ stands for the position of a witness against periodicity i. We assume that when the input string, in this case w_6, has a period i then the position of the witness is 0 meaning that such a witness does not exist.

index	0 1 2 3 4 5 6 7 8 9101112	
$w_6 =$	b a b b a b a b b a b b a	$W[i]$
shift $i = 0$	$\underline{\text{b a b b a b a b b a b b a}}$	0
shift $i = 1$	b a b b a b a b b a b b a	1
shift $i = 2$	$\underline{\text{b}}$ a b b a b a b b a b b a	3
shift $i = 3$	$\underline{\text{b a b}}$ b a b a b b a b b a	6
shift $i = 4$	b a b b a b a b b a b b a	4
shift $i = 5$	$\underline{\text{b a b b a}}$ b a b b a b b a	11
shift $i = 6$	b a b b a b a b b a b b a	6
shift $i = 7$	$\underline{\text{b}}$ a b b a b a b b a b b a	8
shift $i = 8$	$\underline{\text{b a b b a}}$ b a b b a b b a	0
shift $i = 9$	b a b b a b a b b a b b a	9
shift $i = 10$	$\underline{\text{b}}$ a b b a b a b b a b b a	11
shift $i = 11$	$\underline{\text{b a}}$ b b a b a b b a b b a	0
shift $i = 12$	b a b b a b a b b a b b a	12

2.2 String matching with don't care symbols

The mechanism for fast search for patterns with don't care symbols is based on FFT (Fast Fourier Transform) and a couple of observations on how to count matched pattern's 1s and 0s at every position i in the text.

And indeed, we create two modified instances of the pattern P : (1) P_0 which is obtained from P via exchange of all $*$s by 0s. P_0 will be used to identify matched 1s; (2) P_1 which is obtained from P via exchange of all $*$s by 1s and further swap of all 1s by 0s and vice versa. Assuming that $P = 10*1$ we get $P_0 = 1001$ and $P_1 = 0100$.

In order to test every position i in the text for matching 1s in P we interpret T and $P_0^R = 1001$ (reversed P_0, in this case $P_0^R = P_0$ since they are palindromes) as integers (or polynomials) and we use fast integer (polynomial) multiplication to obtain convolution coefficients including required positions 5 and 9. The respective convolution coefficients are *two* and *zero*, see Figure 1. The first value (two) corresponds to the expected number of matched 1s in the pattern occurrence at position

5 in the text. The second value (zero) provides an evidence that there is no pattern occurrence at position 9 in the text.

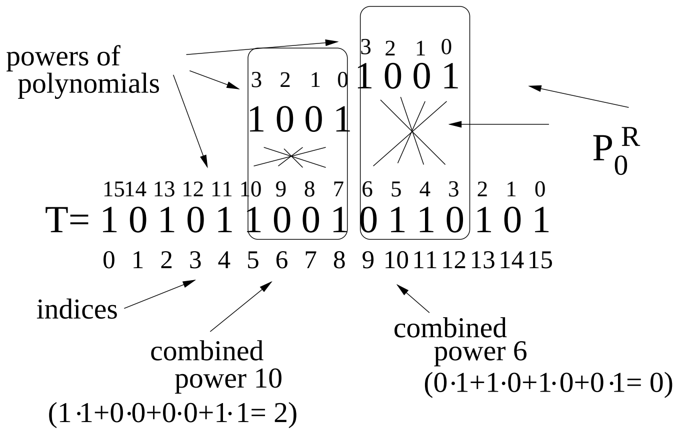

Figure 1: Numbers of matched 1s at positions 5 and 9.

Now we have to test every position i in the text for matching 0s in P. We interpret $\sim T$ and $P_1^R = 0010$ (reversed P_1) as integers (or polynomials) and we use fast integer (polynomial) multiplication to obtain convolution coefficients including required positions 5 and 9. The respective convolution coefficients are *one* and *zero*, see Figure 2. The value one at position 5 confirms that there is an occurrence of the pattern since there are two matched 1s and one matched 0. However at position 9 two matched 1s are not accompanied by one 0 thus we conclude that there is no pattern occurrence at position 9.

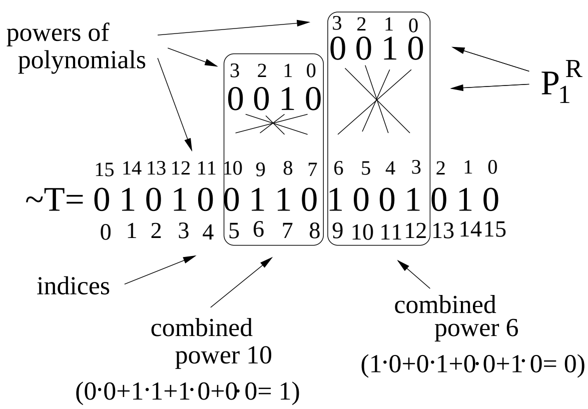

Figure 2: Numbers of matched 0s at positions 5 and 9.