Dhayalini Nagara
A20359686

Homework Assignment-10

1. Rewrite COMPUTE-TRANSITION-FUNCTION (page 1001 of CLRS 3) using Prefix[i] to compute the transition function in time $O(\#pattern|\Sigma|)$. (Remember in the text m is used for what the note calls #pattern).

We get one more match if the next character 'a' matches.

So $\delta(q, a) = q+1$

Otherwise we go back to state Prefix[q] and retrieve the state $\delta(Prefix[q], a)$.

Algorithm:

COMPUTE-TRANSITION-FUNCTION-MOD(P, Σ)

1. m = P.length
2. for q = 0 to #pattern do
3. for each character a ∈ Σ do
4. if q < #pattern and a == P[q+1] then
5. $\delta(q, a) = q+1$
6. else
7. $\delta(q, a) = \delta(Prefix[q], a)$

Scanned by CamScanner

8. end if

9. end for

10. end for

11. return δ

In line 2 and line 3, the loop executes $O(\#\text{pattern})$ and $O(|\Sigma|)$ times respectively. Thus the running time is $O(\#\text{pattern} \cdot |\Sigma|)$.

2. There are four places in the discussion of correctness in KMP lecture notes that "are left to the reader" (two on page 4, top - middle of the page; two on page 5, lower middle of the page). Fill in those details.

Scanned by CamScanner

Assertion P : $0 \leq m \leq \#\text{pattern}$

$\wedge\ 0 \leq n \leq \#\text{text}$

$\wedge$ notbegin $(0, n-m-1)$

$\wedge$ pattern $[0 \ldots m-1] = \text{text}[n-m \ldots n-1]$,

where notbegin (t, u) is an abbreviation for "pattern does not occur in text beginning in text $[t \ldots u]$". Assertion P holds trivially for $n = m = 0$.

a) Line 6 of match() : return $n-m$ // a match

From the assertion P, that pattern $[0..m-1]$ = text $[n-m \ldots n-1]$. If $m = \#\text{pattern}$, then the pattern is $[0 \ldots \#\text{pattern}-1] = \text{text}[n-\#\text{pattern} \ldots n-1]$.

That is Pattern = text $[n-\#\text{pattern} \ldots n-1]$

The string is found in the text starting at position $n-m$, and it exactly returns the same $n-m$.

Scanned by CamScanner

b) Line 8 of match(): return -1 // no match.

From the assertion P, $m \leq$ # pattern.

If line 8 is executed, then we get $m \leq$ #pattern

The condition $n =$ # text means that we have reached the end of the whole text. At this point, the number of matched characters is strictly less than that of Pattern. So we will fail to find a match of pattern in Text, so it returns -1.

c) Line 6 of prefix(): if (Prefix [i] == -1)

If Prefix [i] $\neq -1$, it means that the Prefix[i] has been precomputed and we have to retrieve it.

d) Line 8 of prefix(): Prefix [i] = 0

If $i = 1$, then we have to compute the length of the longest prefix of a single character pattern [0]. By definition it is an empty string. Therefore it returns the length of the empty string 0.

Scanned by CamScanner

3). Show how to use extend to precompute all the values in the array Prefix[i] - that is, to do the job of CLRS 3's COMPUTE-PREFIX-FUNCTION on page 1006. Prove that your algorithm terminates and is correct. Analyse the amount of time required.

We can use extent to compute all values in the array Prefix[i].

Algorithm:

COMPUTE-PREFIX-FUNCTION-MOD (String Pattern)

1. let Prefix[1.... #pattern] be a new array.
2. Prefix[1] = 0
3. for q = 2 to #pattern do
4. Prefix[q] = extend (P, Prefix[q-1], P[q-1])
5. end for
6. return Prefix.

Scanned by CamScanner

In this algorithm we are setting Prefix[1] = 0 and the array Prefix[i] is computed from i = 2 to # pattern. Since Prefix[i] is computed in ascending order, Prefix[i-1] is pre-computed when we compute Prefix[i]. Prefix[i-1] is precomputed, and the next character in the pattern, extend can be invoked to compute Prefix[i].

The algorithm terminates since extend terminates, according to the lecture notes. The total amount of time to compute values in the array Prefix telescopes to

$2 \times \#pattern + Prefix[1] - Prefix[\#pattern]$

$2 \times \#pattern + Prefix[1] - Prefix[\#pattern]$, which is $O(\#pattern)$.

The total running time is $O(\#pattern)$

Scanned by CamScanner

4. Professor Reingold claims that if we memoize the function extend (rather than prefix) you are constructing the FSA rather than the shift function. Is this claim correct? Explain.

Yes, Professor Reingold's claim is correct.

The state set Q is $\{0, 1, \ldots, m\}$.

The start state q_0 is state 0.

m is the only accepting state.

For any state q and character a, the transition function is defined by

$$\delta(q, a) = \text{extend}(\text{pattern}, q, a)$$

References:

homepages.math.uic.edu

Stackoverflow.com/questions/10444596

Staff.ustc.edu.cn

www.ics.uci.edu.

Scanned by CamScanner