Appl 4: String matching with mismatches

Given 2 strings $p_1 \cdots p_m \in \Sigma^*$ $(m \le n)$
$t_1 \cdots t_n \in \Sigma^*$ & k,

decide $\exists i$ s.t. Hamming dist between $p_1 \cdots p_m$
& $t_{i+1} \cdots t_{i+m}$
is $\le k$.

i.e. $|\{j : p_j \ne t_{i+j}\}| \le k$.

e.g. text: algorithmisfun
pattern: muffin $k=3$

naive alg'm: $O(mn)$ time

faster?
next time: $O(|\Sigma| n \log n)$ time
what if $|\Sigma|$ large?
$\tilde{O}(n\sqrt{m})$ time

TO BE CONTINUED...

First Alg'm

generalize problem: compute
$$\mu_i = |\{j : p_j = t_{i+j}\}| \text{ for all } i$$

How? for each $c \in \Sigma$, compute
$$\mu_i^{(c)} = |\{j : p_j = t_{i+j} = c\}| \text{ for all } i$$

$$\mu_i^{(c)} = \left| \{ j : p_j = t_{i+j} = c \} \right| \quad \text{for all } i$$

$\mu_i := \sum_c \mu_i^{(c)}$

$$= \sum_j \underbrace{[p_j = c]}_{f_j} \cdot \underbrace{[t_{i+j} = c]}_{g_{i+j}}$$

$[I] = \begin{cases} 1 & \text{if } I \text{ true} \\ 0 & \text{else} \end{cases}$

CONVOLUTION!

$\Rightarrow \boxed{O(|\Sigma| \, n \log n)}$ time

What if alphabet large? $(2 \le |\Sigma| \le m+1)$

Abrahamson's Alg'm '87:

idea - high vs. low frequency

$\Sigma_H = \{ c \in \Sigma : c$ occurs $\ge d$ times in pattern $\}$

$\Sigma_L = \{ c \in \Sigma : c$ occurs $< d$ times in pattern $\}$

High case:

$|\Sigma_H| \le \frac{m}{d} \Rightarrow$ by First alg'm, $O(|\Sigma_H| \, n \log n) = O(\frac{m}{d} n \log n)$

Low case:

for $l = 1$ to n $\quad (i+j=l)$
  if $t_l \in \Sigma_L$
    for each j with $p_j = t_l$
      increment μ_{l-j} $\quad$ } $O(d)$ time

increment $\mu_{\ell - j}$ } time

$\Rightarrow$ $O(dn)$ time

Total time: $O\left(\left\lceil \frac{m}{d} \right\rceil n \log n + dn\right)$

Set $d = \sqrt{m \log n}$ $\Rightarrow$ $O(n\sqrt{m \log n} + n \log n)$

$= \boxed{\tilde{O}(n\sqrt{m})}$

$\lesssim \tilde{O}(n^{1.5})$

OPEN: better in terms of n?

Rmk - Amir et al. '04 $\tilde{O}(n\sqrt{k})$

$\vdots$

Gawrychowski-Uznanski '18 $\tilde{O}\left(n + \frac{nk}{\sqrt{m}}\right)$

[C. et al. '20 $O\left(n + \frac{nk}{\sqrt{m}}\sqrt{\log m}\right)$,

also approx alg'm in $O(n)$ time]

Appl 5 Subset sum

Given set S of n positive integers & t, decide if $\exists$ subset $R \subseteq S$ that sums to t

$\sim O(2^{n/2})$

or $O(nt)$ time by DP:

or $O(nt)$ time by DP:

let $S = \{a_1, \dots, a_n\}$

Define $C[\ell, i]$ = true iff $\exists R \subseteq \{a_1, \dots, a_\ell\}$ that sums to i

$(i = 0, \dots, t, \quad \ell = 0, \dots, n)$

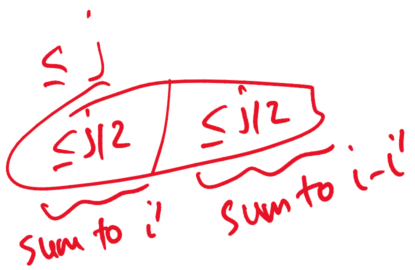
Then
$$C[\ell, i] = \underbrace{C[\ell-1, i]}_{\text{not use } a_\ell} \vee \underbrace{C[\ell-1, i-a_\ell]}_{\text{use } a_\ell}$$

Rmk: if duplicates allowed, i.e. R is a multiset, can do better:

Define $C^{(\leq j)}[i]$ = true iff $\exists$ multiset R of $\leq j$ elems of S summing to i

$(i = 0, \dots, t)$

Then
$$C^{(\leq j)}[i] = \bigvee_{i'} \left(C^{(\leq j/2)}[i'] \wedge C^{(\leq j/2)}[i-i'] \right)$$
$$= C^{(\leq j/2)} \circ C^{(\leq j/2)}$$



Boolean convolution!

$O(t \log t)$ time

$\Rightarrow$ $O(\log t)$ convols repeated squaring

$\Rightarrow$ $O(t \log^2 t)$ time.

What about orig problem where each elem could be used $\leq$ once?

Koiliaris-Xu's Det. Alg'm ('17): $\tilde{O}(\sqrt{n}\, t)$ time

$(u \le t)$

Lemma 1 If $S \subseteq [u]$,
there is an $\tilde{O}(nu)$-time alg'm.

Pf: By D&C.
Will compute $C_S(i) =$ true iff $\exists$ subset of S summing to i
for all $i = 0 \ldots nu$.

Divide S into L, R of $n/2$ elems each. arbitrarily
Solve problem for L recursively
" " R "

Combine:
$$C_S[i] = \bigvee_{i'} \left(C_L[i'] \wedge C_R[i - i'] \right)$$

S: $L \mid R$

convolution on array of size $O(nu)$. $(n \le u)$

$$\Rightarrow \quad T(n,u) = 2T\left(\frac{n}{2}, u\right) + O(nu \log(\cancel{n}u))$$

$$\Rightarrow \quad O(nu \log^2 u) \qquad \square$$

"$\le k$-SUM"

Lemma 2 If $S \subseteq [u]$ & # elems used $\le k$,
then there is an $\tilde{O}(k^2 u)$-time alg'm.

Pf: next time

idea - small vs. large elems

$\downarrow$ Lem 1 $\qquad$ $\downarrow$ k small, Lem 2.

Overall Alg'm:

Consider all $a_i \in (u/2, u]$. Set $k = \frac{t}{u/2} = \Theta(\frac{t}{u})$

Lem 1 $\Rightarrow$ $\tilde{O}(nu)$

Lem 2 $\Rightarrow$ $\tilde{O}(k^2 u) = \tilde{O}(\frac{t^2}{u})$.

$$\Rightarrow \tilde{O}\left(\min\left\{nu, \frac{t^2}{u}\right\}\right) = \tilde{O}(\sqrt{n}\,t)$$

$nu = \frac{t^2}{u}$

$u^2 = \frac{t^2}{n}$

$u = \frac{t}{\sqrt{n}}$

Try all $u = 1, 2, 4, \ldots$

Combine by $O(\log t)$ convolutions in $\tilde{O}(t)$ time

$$\Rightarrow \boxed{\tilde{O}(\sqrt{n}\,t)}$$