A String Minimum Deletetions Problem

Ştefan Vargyas

stvar@yahoo.com

Jun 17, 2021

Table of Contents

1 The Problem Definition and Its Solution

The string minimum deletions problem to which this note refers to is defined as follows:

> Given a non-empty sequence of characters, where each character is allowed to be one of
> the two letters A or B, do find the minimum number of character deletions that makes the
> given sequence become of form A^*B^*.

The algorithm that solves this problem is very simple, as shown by the following C++ and Python implementations:

```cpp
size_t min_del(const std::string& S)
{
    size_t m = 0, b = 0;
    for (auto c : S) {
        if (c == 'A')
            m = std::min(m + 1, b);
        else
            b ++;
    }
    return m;
}
```

```python
def min_del(S)
    m, b = 0, 0
    for c in S:
        if c == 'A':
            m = min(m + 1, b)
        else:
            b += 1
    return m
```

2 Mathematical Proof of the Algorithm's Correctness

1 Definition.

(1) $\quad s \in \{A, B\}^*$ is *valid* $\quad \overset{\text{def}}{\Longleftrightarrow} \quad \exists\, n, m \in \mathbb{N}$ such that $s = A^n B^m$.

2 Notation.

(2) $\quad E : \{A, B\} \ \times \{0, 1\} \to \{\lambda, A, B\}$

$\qquad E(x, 0) \overset{\text{def}}{=} \lambda, \ E(x, 1) \overset{\text{def}}{=} x$

(3) $\quad E_n : \{A, B\}^n \times \{0, 1\}^n \to \{A, B\}^*$

$\qquad E_n(s, e) \overset{\text{def}}{=} E(s_1, e_1) \cdots E(s_n, e_n), \ \text{ for } s \in \{A, B\}^n, \ e = \langle e_1, \cdots, e_n \rangle \in \{0, 1\}^n$

(4) $\quad L_n : \{0, 1\}^n \to \mathbb{N},$

$\qquad L_n(e) \overset{\text{def}}{=} \neg e_1 + \cdots + \neg e_n, \ \text{ for } e = \langle e_1, \cdots, e_n \rangle \in \{0, 1\}^n$

(5) $\quad \mathcal{L}_n : \{A, B\}^n \to \mathcal{P}(\mathbb{N})$

$\qquad \mathcal{L}_n(s) \overset{\text{def}}{=} \{ L_n(e) \mid e \in \{0, 1\}^n \text{ and } E_n(s, e) \text{ is valid} \}$

(6) $\quad M_n : \{A, B\}^n \to \mathbb{N}$

$\qquad M_n(s) \overset{\text{def}}{=} \min \mathcal{L}_n(s)$

(7) $\quad \sharp : \{A, B\}^* \times \{A, B\} \to \mathbb{N}$

$\qquad \sharp(s, x) \overset{\text{def}}{=} \text{ the number of occurrences of } x \text{ in } s$

3 Remark. For $m \in \mathbb{N}$ and $\phi \neq X \in \mathcal{P}(\mathbb{N})$, $m = \min X$ if and only if:

(8) $\quad m \in X$ and

(9) $\quad m \leq x$, for all $x \in X$

4 Proposition. *For all $X, Y \in \mathcal{P}(\mathbb{N})$ and $a \in \mathbb{N}$, we have that:*

(10) $\quad \phi \neq X \subseteq Y \implies \min X \geq \min Y$

(11) $\quad \min X + \min Y = \min X + Y, \ \text{ where } X + Y \overset{\text{def}}{=} \{ x + y \mid x \in X, \ y \in Y \} \in \mathcal{P}(\mathbb{N})$

(12) $\quad a + \min X = \min(a + X),$

Proof. For (10), we have that:

$$\min X \overset{(8)}{\in} X \overset{\text{hyp}}{\implies} \min X \in Y \overset{(9)}{\implies} \min X \geq \min Y$$

For (11), let $m \overset{\text{not}}{=} \min X + \min Y$. On one hand we have that (8) holds for m and $X + Y$:

$$\min X \overset{(8)}{\in} X, \ \min Y \overset{(8)}{\in} Y \implies m = \min X + \min Y \in X + Y$$

On the other hand, (9) holds for m and $X + Y$:

$$z \in X + Y \implies \exists\, x \in X, y \in Y \text{ such that } z = x + y$$
$$\overset{(9)}{\implies} \min X \leq x, \ \min Y \leq y$$
$$\implies m = \min X + \min Y \leq x + y = z$$

(12) is true by (11) taking $Y = \{a\}$. $\qquad\qquad \square$

5 Proposition. *For all $s \in \{A, B\}^n$ and $x \in \{A, B\}$, we have that:*

(13) $\quad \mathcal{L}_n(s) \subseteq \mathcal{L}_{n+1}(sB)$

(14) $\quad M_n(s) \geq M_{n+1}(sB)$

(15) $\quad M_n(s) \leq M_{n+1}(sx)$

Proof. For (13), we have that, for an arbitrary l:

$$l \in \mathcal{L}_n(s) \overset{(5)}{\Longrightarrow} \exists\, e \in \{0,1\}^n \text{ such that } E_n(s,\, e) \text{ is valid and } L_n(e) = l$$
$$\Longrightarrow \exists\, e' \overset{\text{def}}{=} \langle e_1,\, \cdots,\, e_n,\, 1 \rangle \in \{0,1\}^{n+1} \text{ such that}$$
$$E_{n+1}(sB,\, e') \overset{(3)}{=} E_n(s,\, e)\, E(B, e'_{n+1}) \overset{(2)}{=} E_n(s,\, e)\, B$$
$$L_{n+1}(e') \overset{(4)}{=} L_n(e) + \neg e'_{n+1} = L_n(e) = l$$
$$\Longrightarrow \exists\, e' \in \{0,1\}^{n+1} \text{ such that } E_{n+1}(sB,\, e') \text{ is valid and } L_{n+1}(e') = l$$
$$\overset{(5)}{\Longrightarrow} l \in \mathcal{L}_{n+1}(sB)$$

(14) follows then from (13) and (10).

For (15), let $x \in \{A, B\}$ be arbitrary; we have that:

$$m \overset{\text{not}}{=} M_{n+1}(sx) \overset{(8)}{\Longrightarrow} m \in \mathcal{L}_{n+1}(sx)$$
$$\overset{(5)}{\Longrightarrow} \exists\, e \in \{0,1\}^{n+1} \text{ such that } E_{n+1}(sx,\, e) \text{ is valid and } L_{n+1}(e) = m$$
$$\Longrightarrow \exists\, e' \overset{\text{def}}{=} \langle e_1,\, \cdots,\, e_n \rangle \in \{0,1\}^n \text{ such that}$$
$$E_{n+1}(sx,\, e) \overset{(3)}{=} E_n(s,\, e')\, E(x, e_{n+1})$$
$$L_{n+1}(e) \overset{(4)}{=} L_n(e') + \neg e_n \geq L_n(e')$$
$$\overset{(1)}{\Longrightarrow} \exists\, e' \in \{0,1\}^n \text{ such that } E_n(s,\, e') \text{ is valid and } L_n(e') \leq m$$
$$\overset{(5)}{\Longrightarrow} \exists\, l \in \mathcal{L}_n(s) \text{ such that } m \geq l$$
$$\overset{(9)}{\Longrightarrow} m \geq \min \mathcal{L}_n(s) \overset{(6)}{=} M_n(s)$$

$\square$

6 Proposition. *For all $s \in \{A, B\}^n$ and $x \in \{A, B\}$, we have that:*

(16) $\quad \mathcal{L}_{n+1}(sx) \supseteq \mathcal{L}_n(s) + 1$

(17) $\quad M_{n+1}(sx) \leq M_n(s) + 1$

(18) $\quad \mathcal{L}_{n+1}(sA) \ni \sharp(s,\, B)$

(19) $\quad M_{n+1}(sA) \leq \sharp(s,\, B)$

(20) $\quad M_{n+1}(sA) \geq M_n(s) + 1$ *or*
$\quad\quad M_{n+1}(sA) \geq \sharp(s,\, B)$

Proof. For (16), we have that, for an arbitrary l:

$$l \in \mathcal{L}_n(s) + 1 \overset{(5)}{\Longrightarrow} \exists\, e \in \{0,1\}^n \text{ such that } E_n(s,\, e) \text{ is valid and } l = L_n(e) + 1$$
$$\Longrightarrow \exists\, e' \overset{\text{def}}{=} \langle e_1,\, \cdots,\, e_n,\, 0 \rangle \in \{0,1\}^{n+1} \text{ such that}$$
$$E_{n+1}(sx,\, e') \overset{(3)}{=} E_n(s,\, e)\, E(x, e'_{n+1}) \overset{(2)}{=} E_n(s,\, e)$$
$$L_{n+1}(e') \overset{(4)}{=} L_n(e) + \neg e'_{n+1} = L_n(e) + 1 = l$$
$$\Longrightarrow \exists\, e' \in \{0,1\}^{n+1} \text{ such that } E_{n+1}(sx,\, e') \text{ is valid and } L_{n+1}(e') = l$$
$$\overset{(5)}{\Longrightarrow} l \in \mathcal{L}_{n+1}(sx)$$

(17) follows then from (16) and (10).

For (18), let have $e \in \{0,1\}^{n+1}$ such that it deletes all Bs from sA. Then, we have that:

$$L_{n+1}(e) = \sharp(sA,\, B) = \sharp(s,\, B)$$

3

Consequently, by (5), we obtain (18).

For (19), apply (18), (9) and (6).

For (20), taking l arbitrarily, we have that:

$$l \in \mathcal{L}_{n+1}\left(s\right) \overset{(5)}{\implies} \exists\, e \in \{0,1\}^{n+1} \text{ such that } E_{n+1}\left(sA,\, e\right) \text{ is valid and}$$
$$L_{n+1}\left(e\right) = l$$

Let $e' \overset{\mathrm{not}}{=} \langle e_1,\, \cdots,\, e_n \rangle \in \{0,1\}^n$. In case $e_{n+1} = 0$:

$$E_{n+1}\left(sA,\, e\right) \overset{(3)}{=} E_n\left(s,\, e'\right),\; L_{n+1}\left(e\right) \overset{(4)}{=} L_n\left(e'\right) + \neg e_{n+1} = L_n\left(e'\right) + 1$$
$$\overset{(5)}{\implies}\; l \in \mathcal{L}_n\left(s\right) + 1 \overset{(9)}{\implies}\; l \geq \min \mathcal{L}_n\left(s\right) + 1 \overset{(6)}{=} M_n\left(s\right) + 1$$

In case $e_{n+1} = 1$:

$$E_{n+1}\left(sA,\, e\right) \overset{(3)}{=} E_n\left(s,\, e'\right) A,\; L_{n+1}\left(e\right) \overset{(4)}{=} L_n\left(e'\right) + \neg e_{n+1} = L_n\left(e'\right)$$
$$\overset{(1)}{\implies}\; E_n\left(s,\, e'\right) \text{ does not contain } B \text{ s} \overset{(4)}{\implies}\; l \geq \sharp\left(s,\, B\right)$$

Consequently:

$$l \in \mathcal{L}_{n+1}\left(s\right) \implies\; l \geq M_n\left(s\right) + 1 \text{ or } l \geq \sharp\left(s,\, B\right)$$

Due to the arbitrareness of l, by (6), the last implication above proves (20). $\qquad\square$

7 Fact. *For all $s \in \{A, B\}^n$, we have that:*

(21) $M_{n+1}\left(sA\right) = \min\left\{\, M_n\left(s\right) + 1,\, \sharp\left(s,\, B\right)\,\right\}$
(22) $M_{n+1}\left(sB\right) = M_n\left(s\right)$

Proof. For (21), apply (17), (19) and (20).

For (22), apply (14) and (15). $\qquad\square$