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:

```
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;
}
```

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

2 Notation.

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

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

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

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

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

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

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

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

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

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

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

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

(9) $m \leq x$, 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) $\phi \neq X \subseteq Y \implies \min X \geq \min Y$

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

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

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 $X + Y$:

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

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

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

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

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

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

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

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

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

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

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

□

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

(16) $\mathcal{L}_{n+1}(sx) \supseteq \mathcal{L}_n(s) + 1$
(17) $M_{n+1}(sx) \leq M_n(s) + 1$
(18) $\mathcal{L}_{n+1}(sA) \ni \sharp(s,\, B)$
(19) $M_{n+1}(sA) \leq \sharp(s,\, B)$
(20) $M_{n+1}(sA) \geq M_n(s) + 1$ *or*
$M_{n+1}(sA) \geq \sharp(s,\, B)$

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

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

(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 and nothing more. Then:

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

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

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

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

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

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

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

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

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

Consequently:

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

Due to the arbitrareness of l, by (6), the last implication above proves (20). □

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

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

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

For (22), apply (14) and (15). □