

A String Minimum Deletions Problem

Ştefan Vargyas

stvar@yahoo.com

Jun 17, 2021

Table of Contents

1	The Problem Definition and Its Solution	1
2	Mathematical Proof of the Algorithm's Correctness	2

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:

```
1 size_t min_del(const std::string& S)
2 {
3     size_t m = 0, b = 0;
4     for (auto c : S) {
5         if (c == 'A')
6             m = std::min(m + 1, b);
7         else
8             b++;
9     }
10    return m;
11 }
```

```
1 def min_del(S):
2     m, b = 0, 0
3     for c in S:
4         if c == 'A':
5             m = min(m + 1, b)
6         else:
7             b += 1
8     return m
```

2 Mathematical Proof of the Algorithm's Correctness

1 Definition.

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

2 Notation.

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

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

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

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

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

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

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

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

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

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

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

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

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

$$(8) \quad m \in X \text{ and}$$

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

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

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

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

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

Proof. For (10), we have that:

$$\min X \stackrel{(8)}{\in} X \xrightarrow{\text{hyp}} \min X \in Y \xrightarrow{(9)} \min X \geq \min Y$$

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

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

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

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

(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:

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

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

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

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

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

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

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

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

□

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) $\mathcal{L}_{n+1}(sA) \ni \#(s, B)$
- (19) $M_{n+1}(sA) \leq \#(s, B)$
- (20) $M_{n+1}(sA) \geq M_n(s) + 1$ or $M_{n+1}(sA) \geq \#(s, B)$

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

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

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

For (18), let have $e \in \{0, 1\}^{n+1}$ such that it deletes all B s from sA and nothing more. Then:

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

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

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

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

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

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

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

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

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

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 arbitrariness of l , by (6), the last implication above proves (20). □

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

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

$$(22) \quad M_{n+1}(sB) = M_n(s)$$

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

For (22), apply (14) and (15). □