

例題 実際の問題を見てみよう

出現頻度が高いデータを短いビット列に、出現頻度の低いデータを長いビット列に符号化して圧縮する方法をハフマン符号化と呼ぶ。次のテキストデータに対して、次の問いに答えよ。ただし、1文字を3ビットとする。

AABAAAECAABADAAABC

- (1) テキストデータの出現回数をもとに、木構造(ハフマン木)で表せ。
- (2) 圧縮率を求めよ。
- (3) 圧縮前と圧縮後のデータ量の差を求めよ。

考え方

- (1) 木構造の作り方は

1:出現数の少ない数値から接点をつくり、隣同士の出現数を足して新しい接点で結んでいく。
2:接点から分かれる枝の左に0を、右に1を割り当てる。

この流れを行うと、DとEが最も少なく(出現数1)、Aが最も多い(11回)とわかるため、DとEが1。そしてその上の新しい接点は2。という流れで作成する。

- (2) 圧縮率とはデータがどれくらい圧縮されたかを示す値で、以下の式で表される。

$$\text{圧縮率} = \frac{\text{圧縮後のデータ量}}{\text{圧縮前のデータ量}} \times 100$$

今回、圧縮した後のデータ量を確認すると、Aは出現回数11回で与えている符号は0のみなので、データ量は

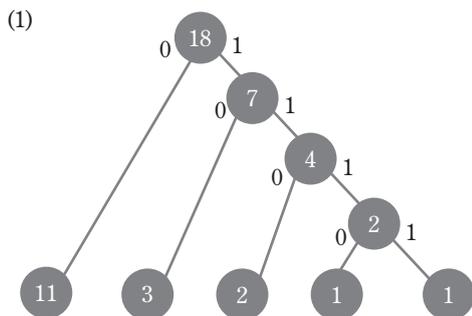
$$1 \times 11 = 11 \text{ ビット}$$

となる。続いてBは出現回数3回で符号は10の2ビットであるので、データ量は $2 \times 3 = 6$ ビット。

これを繰り返していくと、 $1 \times 11 + 2 \times 3 + 3 \times 2 + 4 \times 1 + 4 \times 1 = 31$ ビットとなる。データ圧縮前のデータ量は1文字3ビットとすると18文字あるため $18 \times 3 = 54$ ビットであるので、上記の式に入れ計算を行う。

- (3) 圧縮前のデータ量は 54ビット、圧縮後のデータ量は 31ビットのため、23ビット。

解答



- (2) 57%
- (3) 23 ビット