02.03.2020

Arithmetic coding

E3 scaling

Applying, $E_3E_2 = E_2E_1$

$E_3E_1 = E_1E_2$

Arguing similarly, one can show that

$$\underbrace{E_3E_3 \cdots E_3}_{m} E_2 = E_2 \underbrace{E_1E_1 \cdots E_1}_{m} = 1\underbrace{00 \cdots 0}_{m \text{ times}}$$

$$\underbrace{E_3E_3 \cdots E_3}_{m\text{-times}} E_1 = E_1 \underbrace{E_2 \cdots E_2}_{m} = 0\underbrace{11 \cdots 1}_{m \text{ times}}$$

If we went through three E_3 mappings at the encoder, followed by an E_2 mapping, we would transmit a '1', followed by three zeros.

Arithmetic coding : Floating point Implementation

Integer implementation

1) Word length should be determined - m

2) $\underbrace{x\,x\,x \cdots x}_{m \text{ bits}}$

$$00 \cdots 0 \xrightarrow{\text{equivalent}} 0.0$$

$$11 \cdots 1 \longrightarrow 1.0$$

$$1\underbrace{0 \cdots 0}_{m-1 \text{ bits}} \longrightarrow 0.1$$

Total number of symbols = Total count

Let n_j denote the number of times the j^{th} symbol has occured in the total-count.

$$\text{cum\_count}(k) = \sum_{j=1}^{k} n_j$$

$$F_X(k) = \frac{\text{cum\_count}(k)}{\text{total count}}$$

$$\boxed{l^{(n)} = l^{(n-1)} + \left(u^{(n-1)} + l^{(n-1)}\right) F_X(x_n - 1)}$$

$$= l^{(n-1)} + \left\lfloor \frac{\left(u^{(n-1)} + l^{(n-1)} + 1\right) \times \text{cum\_count}(x_n - 1)}{\text{total\_count}} \right\rfloor$$

$$u^{(n)} = l^{(n-1)} + (u^{(n-1)} - l^{(n-1)}) F_X(x_n)$$

$$= l^{(n-1)} + \left\lfloor \frac{u^{(n-1)} - l^{(n-1)} + 1 \times \text{cum\_count}(x_n)}{\text{Total\_count}} \right\rfloor - 1$$

In floating-point implementation,

Condition for E_1 mapping

$l^{(n)} \geq 0 \quad u^{(n)} < 0.5$ MSB's of both $l^{(n)}$ and $u^{(n)}$ is '0'

Condition for E_2

$l^{(n)} \geq 0.5 \quad u^{(n)} < 1.0$ MSB's of both $l^{(n)}$ and $u^{(n)}$ are '1'.

Condition for E_3

$l^{(n)} \geq 0.25 \quad u^{(n)} < 0.75$ MSB of $l^{(n)} = 0$ second MSB of $l^{(n)} = 1$

" $u^{(n)} = 1$ " $u^{(n)} = 0$

E_1 mapping: $\boxed{\text{Send MSB} = 0}$

$E_1(x) = 2x$

Send '0'

Shift left by '1' bit of both $l^{(n)}$ and $u^{(n)}$

Shift in '0' into LSB of $l^{(n)}$,

shift in '1' into LSB of $u^{(n)}$

E_2 mapping: $\boxed{\text{SEND MSB} = 1}$

$E_2(x) = 2(x - 0.5)$

Send '1'

Shift left by '1' bit of both $l^{(n)}$ and $u^{(n)}$

shift in '0' into LSB of $l^{(n)}$

Shift in '1' into LSB of $u^{(n)}$

E_3 mapping:

$E_3(x) = 2(x - 0.25)$

increment scale '3' by 1

Increment scale 3 by '1'

Shift out MSB of both $l^{(n)}$ and $u^{(n)}$

Shift in '0' into LSB of $l^{(n)}$

Shift in '1' into LSB of $u^{(n)}$

Complement the new MSB's of both $l^{(n)}$ and $u^{(n)}$

Eg: We will encode the sequence 1321 with parameters as follows:

count(1) = 40   count(2) = 1   count(3) = 9   Total_count = 50

Cum_count(1) = 40

Cum_count(2) = 41

Cum_count(3) = 50

$1/4 = 50$

one-fourth length of interval should be able to represent 50 numbers.

'm' must be such that, $2^m \geq \frac{4 \times 50}{\text{smallest count}}$

$$2^m \geq \frac{200}{1}$$

$$m = 8$$

$$l^{(0)} = 00000000 = 0$$

$$u^{(0)} = 11111111 = 255$$

$$l^{(1)} = l^{(0)} + \left\lfloor \frac{(u^{(0)} - l^{(0)} + 1)\ \text{cum\_count}(0)}{\text{total\_count}} \right\rfloor$$

$$= 0 + \left\lfloor \frac{(255 - 0 + 1) \times 0}{50} \right\rfloor$$

$$= \left\lfloor \frac{\overset{128 \times 0}{\cancel{256}}}{\underset{25}{\cancel{50}}} \right\rfloor = 0$$

$$u^{(1)} = l^{(0)} + \left\lfloor \frac{(u^{(0)} - l^{(0)} + 1) \times \text{cum\_count}(1)}{\text{total\_count}} \right\rfloor - 1$$

$$= 0 + \left\lfloor \frac{(255 - 0 + 1) \times 4\cancel{0}}{5\cancel{0}} \right\rfloor - 1$$

$$= \left\lfloor \frac{256 \times 4}{5} \right\rfloor - 1 = 203$$

$$l^{(1)} = 0 = 00000000$$

$$u^{(1)} = 203 = 11001011$$

$$l^{(2)} = l^{(1)} + \left\lfloor \frac{(u^{(1)} - l^{(1)} + 1)\ \text{cum\_count}(2)}{\text{total\_count}} \right\rfloor$$

$$= 0 + \left\lfloor \frac{(203 - 0 + 1) \times 41}{50} \right\rfloor$$

$$= \left\lfloor \frac{204 \times 41}{50} \right\rfloor = 167$$

$$u^{(2)} = l^{(1)} + \left\lfloor \frac{(u^{(1)} - l^{(1)} + 1) \times cum\_count(3)}{total\_count} \right\rfloor - 1$$

$$= 0 + \left\lfloor \frac{(203 - 0 + 1) \times \cancel{50}}{\cancel{50}} \right\rfloor - 1$$

$$= \lfloor 204 \rfloor - 1 = 203$$

$l^{(2)} = 167 = 10100111$

Condition for E2-mapping. $u^{(2)} = 203 = 11001011$

So, $\boxed{\text{SEND 1}}$ and apply E2 mapping to get

$l^{(2)} = 01001110 = 78$

$u^{(2)} = 10010111 = 151$

Condition for E3-mapping. So, increment $\boxed{\text{Scale } 3 = 1}$ and apply E3 mapping

$l^{(2)} = 00011100 = 28$

$u^{(2)} = 10101111 = 175$

05.03.2020

Eg: $A = \{a_1, a_2, a_3\}$ Encode the sequence $a_1 a_3 a_2 a_1$

Count(1) = 40 | cum_count(0) = 0

Count(2) = 1 | cum_count(1) = 40

count(3) = 9 | cum_count(2) = 41

Total_count = 50 | cum_count(3) = 50

scale3 = 0

Word length decided to be m=8

$l^{(0)} = 00000000 = 0$

$u^{(0)} = 11111111 = 255$

None of the mappings are applicable. So, we go on encoding the next symbol i.e.,

a_2.
$$l^{(3)} = l^{(2)} + \left\lfloor \frac{(u^{(2)} - l^{(2)} + 1) \times cum\_count(1)}{Total\_count} \right\rfloor$$

$$= 28 + \left\lfloor \frac{(175 - 28 + 1) \times 40}{50} \right\rfloor$$

$$= 28 + \left\lfloor \frac{148 \times 4}{5} \right\rfloor = 146 = 10010010$$

$$u^{(3)} = l^{(2)} + \left\lfloor \frac{(u^{(2)} - l^{(2)} + 1) \times cum\_count(2)}{Total\_count} \right\rfloor - 1$$

$$= 28 + \left\lfloor \frac{148 \times 41}{50} \right\rfloor = 148 = 10010100$$

$l^{(3)} = 10010010$

$u^{(3)} = 10010100$

Condition for E_2-mapping. So,

$\boxed{\text{SEND 1}}$

Since scale3 = 1, we $\boxed{\text{SEND 0}}$, decrement $\boxed{\text{scale 3} = 0}$ and apply E_2-mapping to get,

$l^{(3)} = 00100100 = 36$

$u^{(3)} = 00101001 = 41$

Condition for E_1 mapping. So, $\boxed{\text{SEND 0}}$ and apply E_1 mapping to get,

$l^{(3)} = 01001000 = 72$

$u^{(3)} = 01010011 = 83$

Condition for E_1 mapping. So, $\boxed{\text{SEND 0}}$ and apply E_1 mapping to get,

$l^{(3)} = 10010000 = 144$

$u^{(3)} = 10100111 = 167$

Condition for E_2 mapping. So, $\boxed{\text{SEND 1}}$ and apply E_2 mapping

$l^{(3)} = 00100000 = 32$

$u^{(3)} = 01001111 = 79$

Condition for E_1 mapping. So, $\boxed{\text{SEND } 0}$ and apply E_1 mapping

$$l^{(3)} = 01000000 = 64$$

$$u^{(3)} = 10011111 = 159$$

Conditions for E_3 mapping. So, increment $\boxed{\text{Scale}3 = 1}$ and apply E_3 mapping

$$l^{(3)} = 00000000 = 0$$

$$u^{(3)} = 10111111 = 191$$

None of the mappings are applicable. So, we go on encoding the next symbol ie, a_1.

$$l^{(4)} = l^{(3)} + \left\lfloor \frac{(u^{(3)} - l^{(3)} + 1) \times \text{cum\_count}(0)}{\text{Total\_count}} \right\rfloor$$

$$= 0 + \left\lfloor \frac{(191 - 0 + 1) \times 0}{50} \right\rfloor = 0$$

$$u^{(4)} = l^{(3)} + \left\lfloor \frac{(u^{(3)} - l^{(3)} + 1) \times \text{cum\_count}(1)}{\text{Total\_count}} \right\rfloor - 1$$

$$= 0 + \left\lfloor \frac{(191 - 0 + 1) \times 40}{50} \right\rfloor - 1$$

$$= 152$$

$$l^{(4)} = 0 = 00000000$$

$$u^{(4)} = 152 = 10011000$$

None of the mappings are applicable. So, we go on encoding but there are no symbols. So, we stop and send the current status of tag value.

Here, we send $l^{(4)}$.

$\boxed{\text{SEND } 0}$ $\boxed{\text{SEND } 1}$ $\boxed{\text{Scale}3 = 0}$

$\boxed{\text{SEND } 0000000}$

1100010000000000

Choose the initial 8 bits as the tag value

Decoder implementation

$m = 8$

$l^{(0)} = 00000000 = 0$

$u^{(0)} = 11111111 = 255$

tag $= 11000100 = 196$

$$l^{(1)} = l^{(0)} + \left\lfloor \frac{(u^{(0)} - l^{(0)} + 1) \cdot cum\_count(x_1 - 1)}{Total\_count} \right\rfloor$$

$$u^{(1)} = l^{(0)} + \left\lfloor \frac{(u^{(0)} - l^{(0)} + 1) \times cum\_count(x)}{50} \right\rfloor - 1$$

$$\left\lfloor \frac{256 \times cum\_count(x-1)}{50} \right\rfloor \leq 196 < \left\lfloor \frac{256 \times cum\_count(x)}{50} \right\rfloor - 1$$

$x = 1,$ $\quad 0 \leq 196 < \left\lfloor \frac{256 \times 4}{5} \right\rfloor - 1 = 203$ ✓

$\boxed{\text{Decode } a_1}$

$l^{(1)} = 00000000$

$u^{(1)} = 11001011$

None of the mappings are applicable. So, we go on decoding the next symbol.

$$l^{(2)} = l^{(1)} + \left\lfloor \frac{(u^{(1)} - l^{(1)} + 1) \times cum\_count(x-1)}{Total\_count} \right\rfloor$$

$$= 0 + \left\lfloor \frac{(203 - 0 + 1) \times cum\_count(x-1)}{50} \right\rfloor$$

$$u^{(2)} = 0 + \left\lfloor \frac{204 \times cum\_count(x)}{50} \right\rfloor - 1$$

$$\left\lfloor \frac{204 \times cum\_count(x-1)}{50} \right\rfloor \leq 196 < \left\lfloor \frac{204 \times cum\_count(x)}{50} \right\rfloor - 1$$

$x = 1,$ $\quad 0 \leq 196 < 162$ ✗

$x = 2,$ $\quad 163 \leq 196 < 166$ ✗

$x=3, \quad 167 \leq 196 < 203$ ✓

$\boxed{\text{Decode } a_3}$

$l^{(2)} = 167 = (10100111)_2$

$u^{(2)} = 203 = (11001011)_2$

Condition for E_2 mapping. So, pushout the MSB from the tag and bring in the next received bit.

$\text{Tag} = 10001001 = 137$

Push out the MSB.

$l^{(2)} = (01001110)_2 = 78$

$u^{(2)} = (10010111)_2 = 151$

Condition for E_3 mapping,

$l^{(2)} = (00011100)_2 = 28$

$u^{(2)} = (10101111)_2 = 175$

$\text{tag} = (10010010)_2 = 146$

None of the mappings are applicable. So, we go on decode the next symbol.

$$l^{(3)} = 28 + \left\lfloor \frac{(175-28+1) \times \text{cum-count}(x-1)}{50} \right\rfloor$$

$$= 28 + \left\lfloor \frac{148 \times \text{cum-count}(x-1)}{50} \right\rfloor$$

$$u^{(3)} = 28 + \left\lfloor \frac{148 \times \text{cum-count}(x)}{50} \right\rfloor - 1$$

$$28 + \left\lfloor \frac{148 \times \text{cum-count}(x-1)}{50} \right\rfloor \leq 146 < 28 + \left\lfloor \frac{148 \times \text{cum-count}(x)}{50} \right\rfloor - 1$$

$$\left\lfloor \frac{148 \times \text{cum-count}(x-1)}{50} \right\rfloor \leq 118 < \left\lfloor \frac{148 \times \text{cum-count}(x)}{50} \right\rfloor - 1$$

$x=1, \quad 0 \leq 118 < 117$ ✗

$x=2, \quad 118 \leq 118 < 120$ ✓

Scanned by CamScanner

$\boxed{\text{Decode } a_2}$

~~$l^{(3)} = 146 = (01110110)_2$~~

$l^{(3)} = 146 = (10010010)_2$

$u^{(3)} = 148 = (10010100)_2$

Condition for E_2,

$tag = 00100100 = 36$

$l^{(3)} = (00100100)_2 = 36$

$u^{(3)} = (00101001)_2 = 41$

Condition for E_1 mapping,

$tag = 01001000$

$l^{(3)} = 01001000$

$u^{(3)} = 01010011$

Condition for E_1 mapping,

$tag = 10010000$

$l^{(3)} = 10010000$

$u^{(3)} = 10100111$

Condition for E_2 mapping,

$tag = 00100000$

$l^{(3)} = 00100000$

$u^{(3)} = 01001111$

Condition for E_1 mapping,

$tag = 01000000$

$l^{(3)} = 01000000$

$u^{(3)} = 10011111$

Condition for E_3 mapping,

$tag = 10000000 - 00000000$

complement MSB

$l^{(3)} = (10000000 = 00000000$

complement

$u^{(3)} = 10111111 = 191$

$$l^{(4)} = 0 + \left\lfloor \frac{(191-0+1)\, cum\_count(x-1)}{50} \right\rfloor$$

$$= \left\lfloor \frac{192 \times cum\_count(x-1)}{50} \right\rfloor$$

$$u^{(4)} = \left\lfloor \frac{192 \times cum\_count(x)}{50} \right\rfloor - 1$$

$$\left\lfloor \frac{192 \times cumcount(x-1)}{50} \right\rfloor \leq 0 \leq \left\lfloor \frac{192 \times cum\_count(x)}{50} \right\rfloor - 1$$

$x = 1$, $0 \leq 0 < 152$ ✓

Dictionary techniques

Dictionary : Most frequent letters/words are stored in the dictionary.

1. Venkaiah
2. Data compression

Whenever we come across a pattern present in the dictionary, we send only the index of the pattern in the dictionary and not the pattern.

There are two types of dictionary techniques:

- Static dictionary
- Adaptive dictionary

<u>Ex:</u> $A = \{a, b, \ldots\ldots z, \cdot, ,, \ldots\}$ $|A| = 32$

Source which outputs 4 letter words.

4-letter word = 20 bits

Let us assume that 257 most frequent 4-letter words are put in the dictionary.

We'll need 256 keys to represent each of the words.

8 bits are used to represent the 256 keys.

How we do we know if the pattern is coming from a dictionary?

Send 4-letter words present in the dictionary and a flag bit (say '0') and send the corresponding index of the pattern.

If the pattern is not present, I send a flag bit (say '1') and send 20 bits corresponding to this word.

Average no. of bits required for every pattern -

Let 'p' be the prob. with which we encounter a pattern from the dictionary, then

$$\text{Average no. of bits required} = 9p + (1-p)21 < 20$$