Name: ROTHA Dapravith
ID: e20190915
Group: I5-GIC(B)

Assignment Discussion Lesson 6

1) What is Huffman coding?

2) Explain Huffman coding algorithm?

3) Why do we use Huffman tree?

4) Assume that we have 6 symbols, draw Huffman tree and find codeword of each symbol? You can choose your own values.

5) Find entropy and average codeword length from the word "lessonseemscool"?

Answer

1). Huffman coding is a compression technique used to reduce the number of bits needed

to send or store a message.

2). To do Huffman coding algorithm:

+ Scan text to be compressed and tally occurrence of all characters.

+ Sort or prioritize characters based on number of occurrences in text. + Build Huffman code tree based on prioritized list.

+ Perform a traversal of tree to determine all codewords.

+ Scan text again and create new file using the Huffman codes.

3). We use Huffman tree to find codeword of character

4). Assume we have 6 symbols: S1, S2, S3, S4, S5, S6 and

P(S1) = 0.10

P(S2) = 0.10

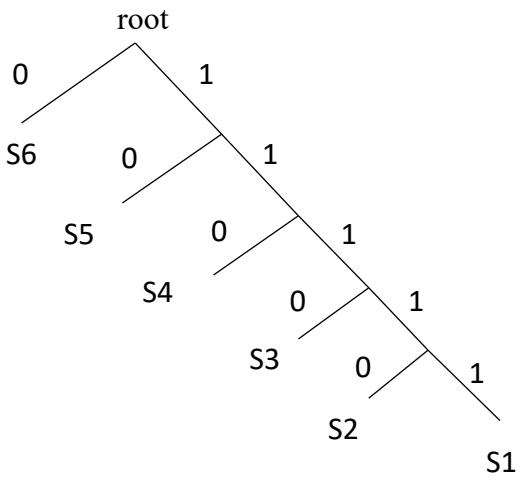
P(S3) = 0.15

P(S4) = 0.15

P(S5) = 0.25

P(S6) = 0.25

+ And we get the Huffman Tree



So, we can get codeword from tree

+ S6 = 0
+ S5 = 10
+ S4 = 110

+ S3 = 1110

+ S2 = 11110

+ S1 = 11111

5). Find entropy from the word "lessonseemscool":

- Total length of characters is: 15

- Probability of each character:
+ P(l) = 2/15 = 0.13
+ P(e) = 3/15 = 0.2

+ P(s) = 4/15 = 0.26

+ P(o) = 3/15 = 0.2

+ P(n) = 1/15 = 0.06

+ P(m) = 1/15 = 0.06

+ P(c) = 1/15 = 0.06

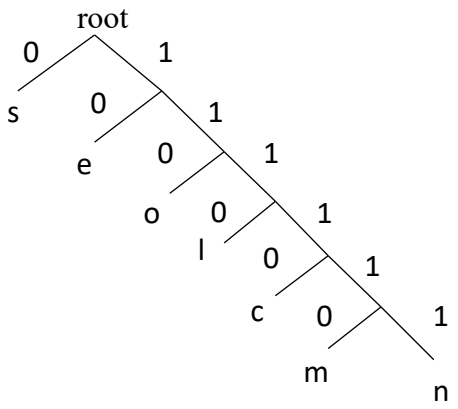
H = $0.13\log_2[1/0.13] + 0.2\log_2[1/0.2] + 0.26\log_2[1/0.26] + 0.2\log_2[1/0.2] +$

$0.06\log_2[1/0.06] + 0.06\log_2[1/0.06] + 0.06\log_2[1/0.06]$

H = 2.54 bits

- ⇨ Find average codeword length from the word "lessonseemscool":

+ We can get Huffman tree by the probability above:



So, We can get codeword from tree:

+ s = 0 => $I_s = 1$
+ e = 10 => $I_e = 2$
+ o = 110 => $I_o = 3$

+ l = 1110 => $I_l = 4$
+ c = 11110 => $I_s = 5$
+ m = 111110 => $I_s = 6$

+ n = 111111 => $I_s = 6$

- ⇨ E = 0.13x4 + 0.2x2 + 0.26x1 + 0.2x3 + 0.06x6 + 0.06x6 + 0.06x5 = 2.8 bits