

Huffman coding

Irina Prosvirina

Huffman coding

Given symbols and their frequencies, our goal is to construct a rooted binary tree where the symbols are the labels of the leaves.

The algorithm begins with a forest of trees each consisting of one vertex, where each vertex has a symbol as its label and where the weight of this vertex equals the frequency of the symbol that is its label.

Huffman coding

At each step, we combine two trees having the least total weight into a single tree by introducing a new root and placing the tree with larger weight as its left subtree and the tree with smaller weight as its right subtree.

Furthermore, we assign the sum of the weights of the two subtrees of this tree as the total weight of the tree. (Although procedures for breaking ties by choosing between trees with equal weights can be specified, we will not specify such procedures here.)

Huffman coding

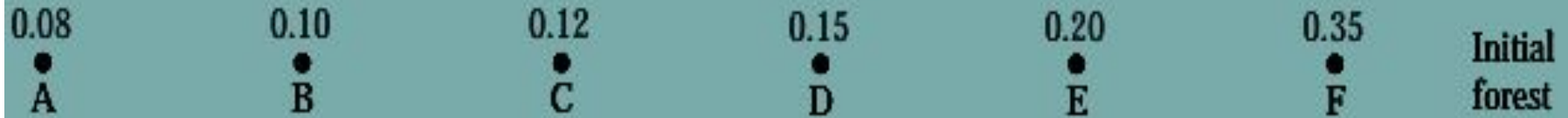
The algorithm is finished when it has constructed a tree, that is, when the forest is reduced to a single tree.

Exercise 1

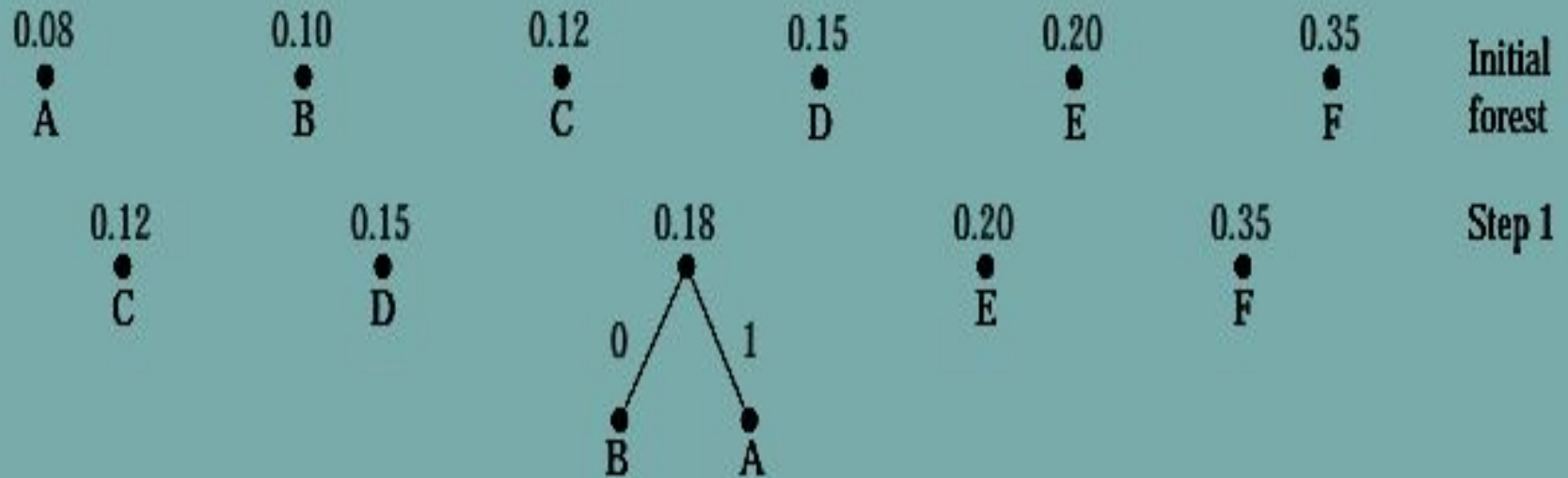
Use Huffman coding to encode the following symbols with the frequencies listed: A: 0.08, B: 0.10, C: 0.12, D: 0.15, E: 0.20, F: 0.35. What is the average number of bits used to encode a character?

Exercise 1

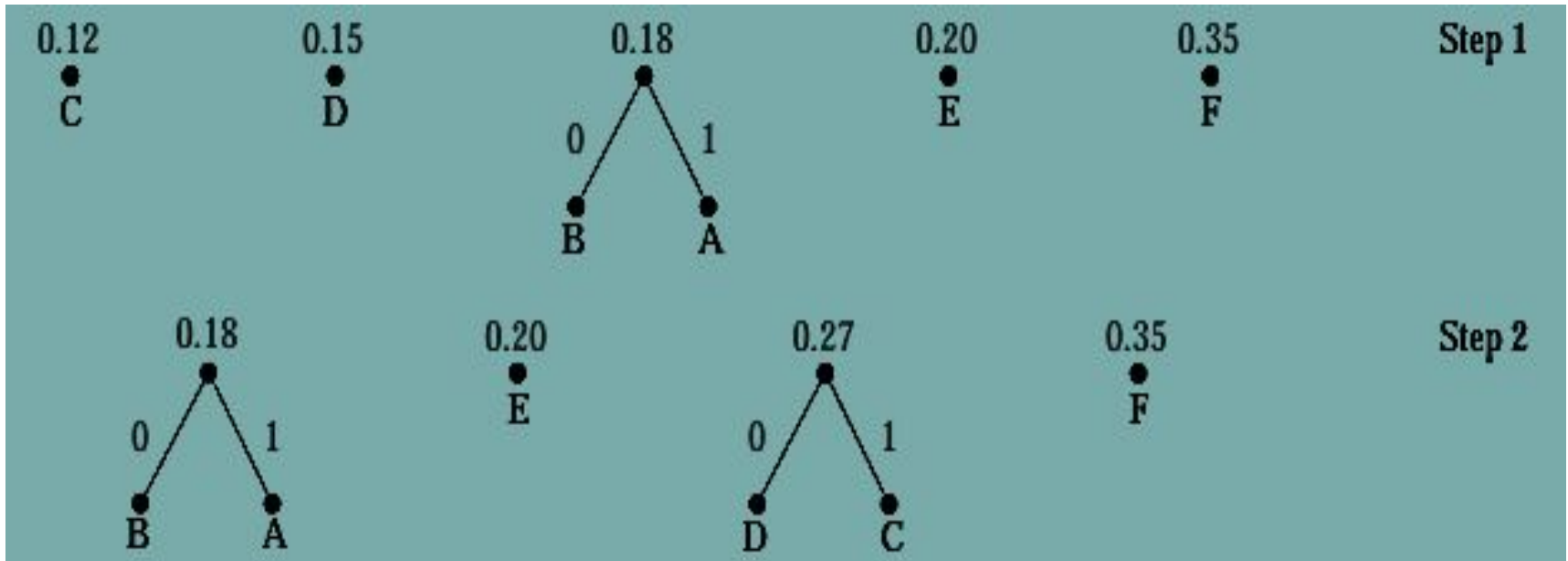
The following figure displays the steps used to encode these symbols.



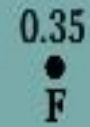
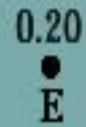
Exercise 1



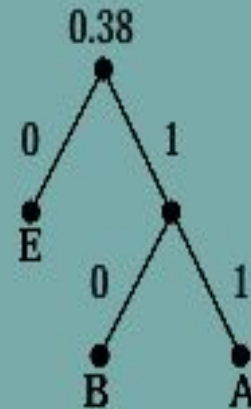
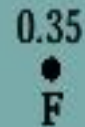
Exercise 1



Exercise 1

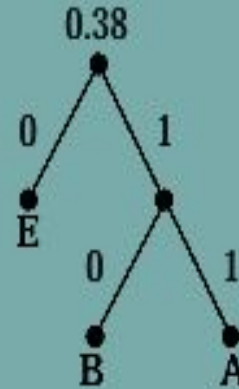
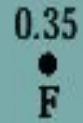


Step 2

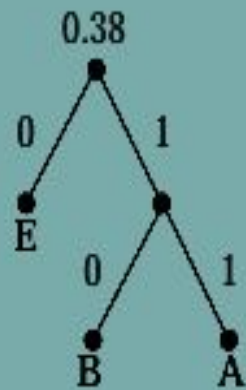


Step 3

Exercise 1



Step 3

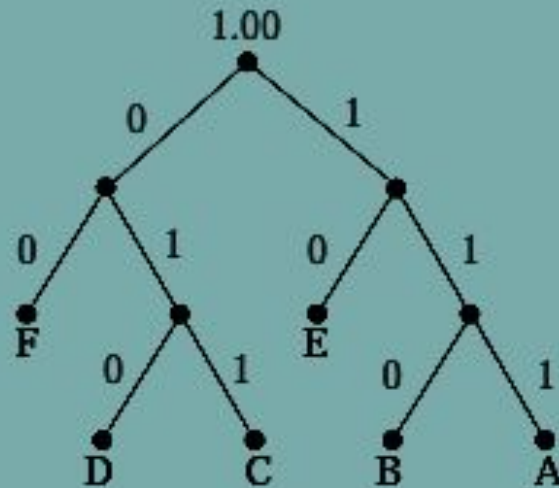


Step 4

Exercise 1



Step 4



Step 5

Exercise 1

The encoding produced encodes A by 111, B by 110, C by 011, D by 010, E by 10, and F by 00.

The average number of bits used to encode a symbol using this encoding is

$$3 \cdot 0.08 + 3 \cdot 0.10 + 3 \cdot 0.12 + 3 \cdot 0.15 + 2 \cdot 0.20 + 2 \cdot 0.35 = 2.45$$

Exercise 2

Use Huffman coding to encode these symbols with given frequencies: A: 0.10, B: 0.25, C: 0.05, D: 0.15, E: 0.30, F: 0.07, G: 0.08.

What is the average number of bits required to encode a symbol?

Exercise 3

Construct two different Huffman codes for these symbols and frequencies: $t: 0.2$, $u: 0.3$, $v: 0.2$, $w: 0.3$.

Exercise 4

A	0.0817	N	0.0662
B	0.0145	O	0.0781
C	0.0248	P	0.0156
D	0.0431	Q	0.0009
E	0.1232	R	0.0572
F	0.0209	S	0.0628
G	0.0182	T	0.0905
H	0.0668	U	0.0304
I	0.0689	V	0.0102
J	0.0010	W	0.0264
K	0.0080	X	0.0015
L	0.0397	Y	0.0211
M	0.0277	Z	0.0005

Construct a Huffman code for the letters of the English alphabet where the frequencies of letters in typical English text are as shown in this table.