ITEC201006 - Introduction to Computer Science & Engineering, Jeonghun Park

Due date: 6/10

Project 1 – Exploitation vs Exploration in the Huffman Encoding

Language: MATLAB

Check the text file named 'source_JPARK2019_vfinal.txt'

In this file, some samples of the alphabetical sources (for example, a, b, c, …) are written. The total number of the samples is 10^5.

Unfortunately, we do not know how many sources we have, and what their appearance probabilities are. For this reason, to use the Huffman encoding, we have to ***LEARN*** the appearance probability first.

Design a source coding algorithm that can minimize the used number of bits.

Two useful assumptions are given in the following:

1. We can change our source coding strategy in the middle of the source coding process. For example, we can use a source coding algorithm "A" until the x-th sample, and use a different a different source coding algorithm "B" until the end of the samples.
2. The ASCII encoding method (7bits per each alphabet) can be used without "learning" anything.

In the report (page limit = 5),

1. Explain the key idea of the algorithm.
2. Show the performance (expected bits length). Draw some performance graphs if you have one.
3. Copy & paste the Matlab code.
4. Write your lesson of exploitation vs exploration in this problem.