

EE326 Introduction to Information Theory and Coding, Homework #3

- Coverage: Lecture note 8
 - Due date: October 27, 11:59 PM through KLMS
 - Late submission: -50% for < 12 hours, -100% for >12 hours
 - TA in charge of Homework #3: Hyeonseong Im (imhyun1209@kaist.ac.kr)
-

Instruction:

- Use Matlab for the implementation of the LZ77 algorithm. In case you never used Matlab before, it is fine to use Python or C.
 - Write a report with detailed answers to each question. For the simulation results, include the screen captures.
 - Submit your report together with all the source files (.m file for Matlab).
-

- 1) (50 points) Develop LZ77 encoding and decoding programs, that compress and reconstruct an arbitrary 4-ary sequence of arbitrary length with arbitrary window size.
 - a) Verify your encoding program by encoding the sequence '1001001000110101' with window size 4, which is the example shown in the lecture.
 - b) Verify your decoding program by decoding the result in b).
- 2) (15 points) Develop a program (code) that generates a length-10000 4-ary sequence where each symbol is randomly generated according to $\mathbf{p}(q) = (\frac{q}{3}, \frac{q}{3}, \frac{q}{3}, 1 - q)$ for arbitrary q . Using this program, generate a length-10000 4-ary sequence randomly generated according to $\mathbf{p}(q)$ with $q = 0.1$. Now, encode this code by using your LZ77 encoding program with window size 100. Also, for the encoded sequence, verify that your decoding program reconstructs the original source sequence correctly.
- 3) (15 points) Develop a program that calculates the compression rate $R_c = \frac{\text{code length}}{10000}$. Here, the code length is calculated by adding the code lengths for encoded tuples, where the code length for each encoded tuple is calculated as follows:
 - $(1, i, j) \rightarrow 1 + \lceil \log(i) \rceil + \lceil \log(j) \rceil$ bits. Use base 2 for the logarithms.
 - $(0, k) \rightarrow 3$ bits for any $k \in \{0, 1, 2, 3\}$.

Use this program to calculate the compression rate in 2).
- 4) (30 points) Calculate the compression rates for length-10000 4-ary sequences, where the source symbols follow $\mathbf{p}(q) = (\frac{q}{3}, \frac{q}{3}, \frac{q}{3}, 1 - q)$ with $q = 0.02 : 0.02 : 0.2$. The window size is set to 100. Plot the compression rate $R_c(q)$ and the entropy $H(\mathbf{p}(q))$ with respect to q . Use base 2 for the entropy. Discuss the results.