

# SSIE-500: Homework 4

## Measuring Complexity of Text

Hiroki Sayama  
sayama@binghamton.edu

Deadline: November 9, 2020 (1:00pm)

Do the following tasks. You can use any programming language for computation.

1. Read a sufficiently long text data written in English. Calculate the information entropy of the frequency distribution of characters that appear in the text. You can keep spaces, punctuations, and other symbols in this calculation too.
2. Implement a computer program that generates a Huffman code for a given data set. Then apply the implemented program to the text data and encode it into the Huffman code.
3. Calculate the achieved average code word length (i.e., the average number of bits per character in the original text data) and compare it with the information entropy in the text calculated above.
4. Produce a PDF file of your report using  $\text{\LaTeX}$ , showing your program code, the results and any comments/discussions, and submit it to the Assignment page of MyCourses by the deadline.