CPSC 530
Group 17 - Project Proposal

Aiden Taylor - 30092686 - Computer Science
Noah Pinel - 30159409 - Computer Science
Ty Irving - 30105319 - Computer Science

Feb. 5th, 2023

Introduction to the Topic

The proposed project concerns the implementation and analysis of algorithms used in Data Compression. When storing, transferring, sending, or receiving large files it is important to have efficient data compression and decompression algorithms. However, this creates a problem, specifically, in the decision of which algorithm to use in different scenarios. In order to analyze if a choice of algorithm was the most efficient, one has to implement this algorithm, then compare it against the other options, which is what we plan to do. In our case, we plan to implement the lossless data compression algorithm, Bit Code Complete Binary Tree (BCCBT), proposed in our first referenced paper. We will then compare this implementation against other sophisticated compression utilities, such as Huffman encoding, the Lempel-Ziv family of algorithms, Dynamic Markov Compression, and Arithmetic Encoding. We will mainly follow the first referenced paper, since we plan to implement the algorithm proposed in this paper, however, we plan to use the other two papers as inspiration for potential tests to analyze our implementation of BCCBT. In the end, the goals of this project are to gain insight into the different factors and tests used to analyze compression algorithms, and to gain experience in implementing compression algorithms that are to be used in practical scenarios.

Papers and References

(1) Paper: A study in compression algorithms
    Author: Mattias Håkansson Sjöstrand
    Link to paper

(2) Paper: Comparison Study of Lossless Data Compression Algorithms for Text Data
    Authors: A. Bhattacharjee, T. Bej, S. Agarwal
    Link to paper

(3) Paper: Comparison of lossless data compression methods
    Authors: D. Berz, M. Engstler, M. Heindl, F. Waibel
    Link to paper

Outline of Proposed Work and Implementation

Proposed Work

The following is the work that needs to be done to complete this project:

1. An analysis of the BCCBT algorithm and its pseudocode.

2. An implementation of the BCCBT algorithm.

3. A comparison of our implementation of BCCBT against other compression utilities.

4. A compilation of the comparison results into a 7-10 page technical paper.

5. The design of three quiz questions that are relevant to our paper.

6. A presentation that summarizes the main results of our paper.

Implementation

We will first focus on implementing the BCCBT algorithm, which will be written in a group agreed upon programming language. Once we have successfully implemented the algorithm, and have a way to compress and decompress files with this algorithm, we will begin analyzing our implementation by testing it with multiple data compression factors. We will then compare these results against other commonly used compression algorithms mentioned above. Some of the factors we will be using to analyze and compare these algorithms are as follows:

1. Compression Time

2. Decompression Time

3. Saving Percentage $= \frac{Original\ File\ Size - Compressed\ File\ Size}{Original\ File\ Size}$

4. Compression Ratio $= \frac{Compressed\ File\ Size}{Original\ File\ Size}$

Using these factors we will be able to see in what scenarios certain algorithms should and should not be used, and we will be able to tell if our choice of algorithm was the most efficient, in the given scenario. During our comparison testing, we plan on using different size .txt files as our sources of information, however, if time permits, we may incorporate other file types as well.