A Review of Lossless Data Compression Algorithms

Aditya Meharia
School of Computer Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India 751024
adityameharia14@gmail.com

Junaid H Rahim
School of Computer Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India 751024
junaidrahim5a@gmail.com

March 7, 2020

Abstract

Objective

This work aims to provide an introduction to the domain of Data Compression in Information Theory and a review of the existing literature in the field of Algorithms for Lossless Data Compression. We identify and discuss the potential opportunities, barriers and the future scope of the field. We also review data compression methods used for text, image, video and audio data.

Results

Mostly about which algo performs best for what kinda field and all that. will write this later when the paper is complete. [1] thats a reference example

Keywords Data Compression · Algorithms · Lossless Compression · Information Theory

1 Introduction

A rapid growth in modern communication technology led an explosion in the amount of data we transmit and store. Extremely large files consume significant resources for transmission as well as storage. Due to this exponential increase in the size of the data we transmit, researchers developed algorithms that can be used to compress the data to save storage space and transmission time. Data Compression is a process by which we encode the input data into a representation that occupies fewer bits than the original input. This encoded representation is transmitted and decoded back to the original form at the destination. Data compression algorithms are broadly classified into two classes viz **Lossless Compression** and **Lossy Compression** algorithms. We will be only covering Lossless Compression in this article.

Lossless Compression algorithms are a class of algorithms that can reproduce the original content from the encoded respresentation without any loss of information, the data before compression and after decompression is exactly the same. Lossless compression is used in a variety of fields where it is important that the original and decompressed information be the same. The GNU tool gzip uses lossless algorithms for the ZIP file format.

1.1 Overview of Lossless Data Compression Techniques

Lossless compression algorithms usually have a two step procedure.

1. A probabilistic model of the input data is generated. This usually assigns a probability of occurrence to pieces of input data. For example, if the input data is piece of text, then the model would be the probabilities of occurance of each alphabet

2. A coding system uses this model to map the data in a way that the pieces with high probability of occurrence are assigned a shorter code than those with a low probability of occurance

JHR

References

[1] PM Parekar and SS Thakare. Lossless data compression algorithm–a review. *International Journal of Computer Science & Information Technologies*, 5(1), 2014.