A REVIEW PAPER ON COMPRESSION TECHNIQUES

Asst. Prof. Prateek Manya

G.T.B. khalsa college for women, Dasuya.

Abstract: Compression is reduction in number of bits required to represent the information. Information can be represented in the form of any multimedia component may it be text, image, audio, video or graphical data. As digital data is growing at very fast pace, compression can help us to store huge amount of data in small memory. Compression can be applied to reduce the size of data and also to reduce the time required to transmit the information. Data can be compressed using two techniques: Lossless and Lossy compression. In this paper, we review some methods of lossless data compression.

Key terms: Compression, Lossless, Lossy, Lossless compression techniques.

1. Introduction:

With the advent of Multimedia (where various media components are grouped together), file sizes are also becoming larger and larger day-by-day. Such files (if uncompressed), require more space in memory for saving and more RAM. They will also take more time to get uploaded and downloaded over the internet. Using data compression techniques, one can reduce the file size to many folds. After compression, these files will require less space in hard disk. Also, the files after compression will take much less time in getting uploaded and downloaded over the internet. Compression is further categorized into two types: Lossless and Lossy data compression. In lossless data compression, file size is reduced with no data loss and with decompression, we can retrieve whole data. We can apply this compression on spreadsheets, graphics, text documents etc. [1] In lossy data compression, some data is lost permanently and this data cannot be recovered even after performing decompression operation, which means that exact replica cannot be recovered from the file after lossy compression of data. Various algorithms that perform lossless compression are Run length encoding, Huffman encoding, Arithmetic encoding etc. and the algorithms that perform lossy compression on data are DCT (Discrete cosine transform), fractal compression etc. Lossless algorithms are used to compress text, images and sound and lossy algorithms are mainly used for Images, audio and video. [2]

- **Pros and cons of Lossless compression:**

As we have discussed earlier, compression (either lossless or lossy) reduces the file size and now the same file requires lesser storage space and will take lesser time for transmission. The advantage of using lossless compression algorithms is that no data will be lost after compression. We achieve a smaller file size and it does not degrade the quality of data. But the disadvantage is that the file sizes are not as small as size of files which are compressed using lossy data compression.

- **Pros and cons of Lossy compression:**

The main advantage of using lossy algorithms is that we achieve much smaller file sizes as compared to files compressed using lossless compression. But the disadvantage is that some data is lost which cannot be recovered even after decompression and degrades the quality of data. The process of data loss is irreversible. The lossy data compression is used when complete retrieval of the original data is not necessary after decompression. [3]

In this paper, we will discuss various lossless data compression techniques. We will give review of some techniques and examples will be based on text data of files.

Techniques for lossless data compression:

1. **Run length encoding (RLE):**

   It is simplest lossless compression algorithm. The basic concept here is to replace a long sequence of the same character by a shorter one. [4] Many file formats such as bitmap, TIFF (Tagged Image File Format) uses RLE.

   e.g. Consider a continuous sequence of 20 ‘H’ characters, which without compression would require 20 bytes storage area.

   HHHHHHHHHHHHHHHHHHHH

   After compressing it with RLE, it will become

   20H

   It will require only 2 bytes storage area, 1 byte for count (20) and 1 byte for character ‘H’. As it is lossless compression technique, no data will be lost but the compression ratio is not much high.

2. Entropy encoding:

Entropy encoding is a type of lossless data compression where frequently occurring characters are assigned fewer bits and rarely occurring characters are assigned more bits. Huffman encoding and Arithmetic encoding are types of entropy encoding.

a) Huffman encoding:

This encoding technique is named after D.A. Huffman. This is a variable length lossless compression technique where probability of occurrence of each symbol in the text is estimated. Then Huffman encoding is applied to calculate the variable length code for each input symbol. For this, we generate Huffman tree.

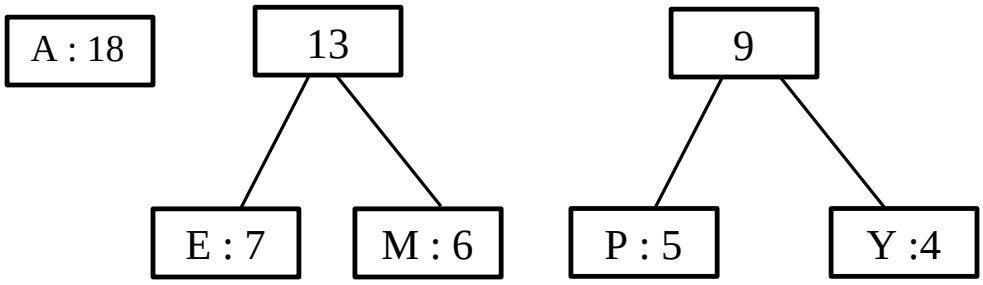
e.g. Consider the frequencies of characters.

A : 18 E : 7 M : 6 P : 5 Y : 4

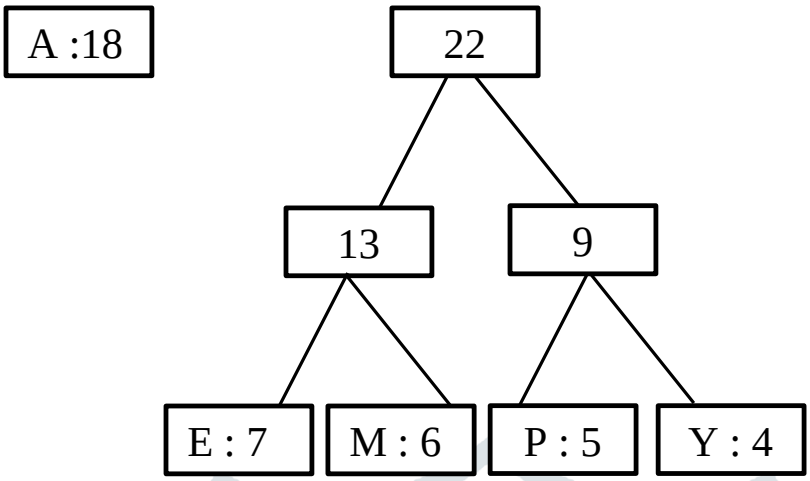
First of all, arrange these frequencies in ascending or descending order. As given frequencies are already in descending order, we will move to next step.

Combine two lowest frequencies to form a smallest subtree.

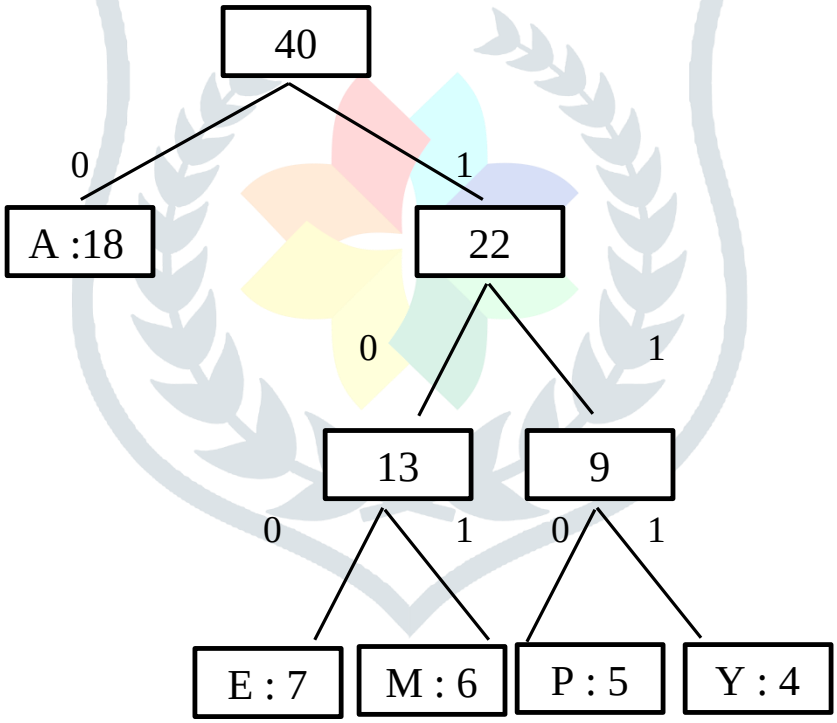
Now subtrees E and M have lowest frequencies. So we will make a subtree by joining them:



Subtrees of (E,M) and (P,Y) have lowest frequencies. So merge them:



And the last step is to make a Huffman tree



When the final tree is made, left branch of the tree will be assigned 0 and right will be assigned 1.

So following are the code:

Character	Occurrence	Code	Code length
A	18	0	1
E	7	100	3
M	6	101	3
P	5	110	3
Y	4	111	3

So where without compression total number of bits required to store the text were 320 bits ((18+7+6+5+4)*8=320). After compression the number of bits required would be:

18 * 1 = 18
7 * 3 = 21
6 * 3 = 18
5 * 3 = 15
4 * 3 = 12

84 bits

So 320-84=236 bits space would be saved (in this example).

Conclusion:

Various data compression algorithms exits some of which are lossless and other are of lossy nature. The choice of algorithm depends upon the tradeoff between quality of data and file size. If one can compromise with quality of data but requires lesser space for data storage, then one should choose lossy algorithm and vice versa.

References:

1. Prof. Dipti Mathpal, Prof. Mittal Darji "A Research Paper on Lossless Data Compression techniques" IJIRST –International Journal for Innovative Research in Science & Technology| Volume 4 | Issue 1 | June 2017 ISSN (online): 2349-6010
2. https://techdifferences.com/difference-between-lossy-and-lossless-compression.html
3. Shrusti Porwal, Yashi Chaudhary, Jitendra Joshi, Manish Jain, "Data Compression Methodologies for Lossless Data and Comparison between Algorithms", IJESIT, March 2013
4. Ruchi Gupta, Mukesh Kumar, Rohit Bathla, "Data Compression - Lossless and Lossy Techniques", IJAIEM, July 2016