Lossless Data Compression Techniques with Block Coding: An explanation

Deepak Mathur#1, Dr. Prabhat Mathur*2

#Assistant Professor, Computer Science Department LACHOO MEMORIAL COLLEGE OF SCIENCE & TECHNOLOGY (Autonomous) JODHPUR, RAJASTHAN, INDIA
*Associate Professor, Computer Science Department LACHOO MEMORIAL COLLEGE OF SCIENCE & TECHNOLOGY (Autonomous) JODHPUR, RAJASTHAN, INDIA

Abstract - Data compression is the process of encoding information using fewer bits (or other information bearing units) than an unencoded representation would use, through use of specific encoding schemes. Using this we can achieve reduced resource consumption and security of data as well. Data Compression may be categorized in following two categories.

(A) Lossless Compression algorithms usually exploit statistical redundancy in such a way as to represent the sender's data more concisely without error. Lossless compression schemes are reversible so that the original data can be reconstructed [3].

(B) Lossy Compression is possible if some loss of some fidelity is acceptable. Lossy schemes accept some loss of data in order to achieve higher compression. Lossy data compression is better known as rate-distortion theory.

Claude E. Shannon formulated the theory of data compression[6]. Lossless Data Compression Theory and Rate-Distortion Theory are known collectively as Source Coding theory.

Key words - Lossy Compression, Lossless Compression, encoding, block coding

I. INTRODUCTION

Data is defined as the raw facts, figures or anything that is unprocessed. Data compression is a technique in which the data/information is encoded using fewer bits through use of specific encoding schemes. Data compression[2] can be seen as a technique which has two folds objective. Prime objective is to minimize the amount of data to be stored or transmitted. Another objective may be treated as the security of data as well because the data is send in encoded form. Data Compression is useful because it helps in reducing the consumption of resources, such as hard disk space or transmission bandwidth[5]. Also it increase speed of data transfer from disk to memory. This paper discuss[7]s the data compression with emphasis on Lossless technique.

II. DATA COMPRESSION TECHNIQUES

Now if we talk about Data compression[1], it may be divided into following 2 categories:

A. Lossless compression

Is the one which uses the statistical redundancy that occurs in the data in such a way as to represent the sender's data more concisely without error. Lossless compression schemes are reversible so that the original data can be reconstructed [3].

B. Lossy compression

Is the one which is possible only in the case where some loss of fidelity is acceptable. Lossy data compression is better known as rate-distortion theory. Lossy schemes[11] accept some loss of data in order to achieve higher compression.

Claude E. Shannon formulated the theory of data compression[6]. Lossless Data Compression Theory and Rate-Distortion Theory are known collectively as source coding theory. Now we would discuss the Lossless Source Coding and Rate-Distortion Theory for Lossy Data Compression.

III. LOSSLESS SOURCE CODING

Lossless data compression is used in many applications. For example, it is used in the zip file format. It is also often used as a component within lossy data compression technologies.

It has its base in the block coding[6]. To show this concept, we consider a source of string in which the set of alphabets consists of only two letters namely a and b

\[ A = \{a, b\} \] (1)

But it is sure that if ‘a’ occurred in the previous character, the probability that ‘a’ occurs again in the present character is 0.9. Similarly, given that if ‘b’ occurred in the previous character, the probability that ‘b’ occurs again in the present character is 0.9.

An n-th order block code is just a mapping which assigns to each block of n consecutive characters a sequence of bits of varying length. This can be shown by following 2 examples.
A. First-Order Block Code

Here each character is mapped to single bit as shown in table below.

<table>
<thead>
<tr>
<th>B1</th>
<th>P(B1)</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>.5</td>
<td>0</td>
</tr>
<tr>
<td>b</td>
<td>.5</td>
<td>1</td>
</tr>
</tbody>
</table>

Table I  
Table of First-Order Block Code

Data Original  a a a a b b b b a a a b b b

Data Compressed 0 0 0 0 1 1 1 1 0 0 0 1 1 1

Fig 1. Showing original and compressed data (First-Order Block Code)

Using Table I we refer to Fig. 1 and note that here 14 bits are used to represent 14 characters -- an average of 1 bit/character.

B. Second-Order Block Code

Here pair of characters are mapped to either one, two, or three bits as shown in table below.

<table>
<thead>
<tr>
<th>B2</th>
<th>P(B2)</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>.45</td>
<td>0</td>
</tr>
<tr>
<td>bb</td>
<td>.45</td>
<td>10</td>
</tr>
<tr>
<td>ab</td>
<td>.05</td>
<td>110</td>
</tr>
<tr>
<td>ba</td>
<td>.05</td>
<td>111</td>
</tr>
</tbody>
</table>

Data Original  a a a a a b b b b b b b b b

Data Compressed 0 0 0 1 1 1 0 1 0 1 0 1 0 1 0

Fig 2. Showing original and compressed data(Second-Order Block Code)

Using Table II we refer to Fig. 2 and note that here 16 bits are used to represent 18 characters -- an average of 0.88 bits/character. What a decrease in consumption of bits/character!!!!

Similarly we can move to third ,fourth and so on up to n order.

C. Formula

Here we note that the rates shown in the table[6]s are calculated from

\[ R = \frac{1}{n} \sum p(B_n) l(B_n) \text{ bits/sample,} \]

\( l(B_n) \) is the length of the codeword for block \( B_n \). The higher the order, the lower the rate that is better compression is achieved.

As we move to n Order Block Code the consumption of bits / character decrease significantly.

We are only interested in lossless data compression code. That is, we are able to rederive the original data. All of the examples given above are lossless.

IV. RATE-DISTORTION THEORY

In lossy data compression[11], the decompressed data need not be exactly the same as the original data. So some amount of distortion is acceptable. A distortion measure specifies exactly how close the approximation is.

V. CONCLUSION

We surveyed the data compression technique that use block coding concept. It shows basic information about data compression, & their techniques are applied on text file. By above discussion we note that as we move gradually from 1 to n Order Block Code the consumption of bits / character decrease significantly as shown by examples.

Both lossy and lossless have their advantage in different situations. These not only minimize the amount of data to be stored or transmitted but also the security of data is strengthened.

ACKNOWLEDGMENT

Deepak Mathur is Assistant Professor in Computer Science Department of Lachoo Memorial College Of Science & Technology(Autonomous) Jodhpur,Rajasthan India. He is an MCA having teaching experience of more than 10 years and pursuing Ph.D. in Computer Applications from JNU,Jodhpur

Dr. Prabhat Mathur is Associate Professor in Computer Science Department of Lachoo Memorial College.

REFERENCES


[2] Priyanka Chouksey1 and Dr. Prabhat Patel , “Secret Key Steganography technique based on three-layered DWT and SVD


