RESEARCH ARTICLE OPEN ACCESS

Implementation of Run-Length Encoding In Examination For Lossless Data Compression

Thirunavukkarasu E [1], Karuppusami G [2], Ezhilarasu P [3]
Professor and Head [1]
Department of Computer Science and Engineering
Jairupaa College of Engineering, Tiruppur
Dean – Research and Innovations [2]
Department of Mechanical Engineering
Sri Eshwar College of Engineering, Coimbatore
Associate Professor [3]
Department of Computer Science and Engineering
Hindusthan College of Engineering and Technology, Coimbatore
Tamil Nadu - India.

ABSTRACT
In this paper, we discuss Run-length encoding data compression technique. Three type of input taken. First, more than two input characters with frequent run-length considered. Then, two input character with almost equal frequency of occurrence are considered. Finally two input character with one character having less frequency of occurrences and another one with more frequencies of occurrences are considered. Its compression ratio, space savings, and average bits also calculated. Comparison between each condition performed.
Keywords:- Run-length, Compression, Encoding, Decoding.

I. INTRODUCTION

Data compression defined as the reorganization of data in such a way that, the volume of the resultant data (compressed data) is less than that of the volume of the source data (uncompressed data). The decompression technique applied to get back the uncompressed source data. After decompression if some data destroyed because of compression and decompression, then the compression called as lossy compression. If none of the data destroyed and available in its original uncompressed form, then the compression called as lossless compression. The Run-length encoding comes under lossless compression. Time complexity and space complexity are the two important complexities in data compression.

Because of data suppression, only little amount of time needed for data transfer between source system and target system. For instance, if the original size of the source data is 4GB and the transfer rate is 512 kbps. The time need for the transfer obtained by the given equation 1.

Time needed for transfer of data = Input data / transfer rate (1)

(1 MB = 1024 KB and 1 KB = 8 kb)

Input data = 4 GB

= 4 * 1024 * 1024 * 8

= 4 * 1024 * 1024 * 8 kb

Transfer rate = 512 kbps

So time taken for transfer of data = (4 * 1024 * 1024 * 8) / 512

= 4 *2* 1024 * 8

= 65536 seconds

If the given data compressed into 2.5GB, then the time taken for transfer will be 40960 seconds.

If the destination allowed amount of storage is 2TB, then the target machine can store the following number of files by using the equation 2.

Total number of files can be stored = Total amount of storage / Volume of the file (2)

(1TB = 1024 GB)

= 2 * 1024 GB / 4 GB

= 1024 / 2

=512 files

Therefore, the destination system can store 512 files for uncompressed data.

For compressed file, it can store

= 2 * 1024 MB / 2.5 GB

= 2048/2.5

= 819 files.

The compression ratio affects both the space and time complexity. It calculated by using the following equation 3.

Compression ratio = Uncompressed original data / Compressed Data (3)

= 4 GB / 2.5 GB

= 8:5

Space savings also calculated by using compressed and uncompressed data. It obtained by using the following equation 4.

Space savings = 1- (Compressed Data/Uncompressed original data) (4)

In compression, we have the following types

1. Lossy compression.
2. Lossless compression.

In this paper, we discuss Run-length encoding.

II. RELATED WORK

Run-length encoding (RLE) is a very simple form of data compression in which runs of data (that is, sequences in which the same data value occurs in many consecutive data elements) are stored as a single data value and count, rather than as the original run. This is most useful on data that contains many such runs [1]. Khalid Sayood [2000] delineated an introduction into the various areas of coding algorithms, both lossless and lossy, with theoretical and mathematical background information [2]. Mark Nelson and Jean-loup Gailly [1995] represented the basics of data compression algorithms. It includes lossless and lossy algorithms [3].David Salomon [2000] explained many different compression algorithms altogether with their uses, limitations, and common usages. He gave an overview on lossless and lossy compression [4]. Many books [5, 6, 7, 8 and 9] published about the data compression techniques. Golomb [1966] explained about the golomb code which refers the case where the index function is f(z) = z/g where g denotes the positive integer then only each member of the set consists of g members [10].

III. RUN-LENGTH ENCODING

A.MORE THAN TWO CHARACTERS

eeeeeeeeeeeeezzzzhhhhhhhhhhhhiiiiiiiiiiillllllllll
after encoding
e13z4h12i11l10
Total number of characters =14

uuuuuuummmmmmmaaaaaddddddeeeeeeeeeeeeeeevv viiiiiii
after encoding
u7m7a5d6e15v3i7
Total number of characters =15

pppppppppaaaaaaaaaalllllllllaaaaaaaaaaaannnnniiiiii
after encoding
p8a9l10a12n5i6
Total number of characters =14

B.TWO UNIQUE CHARACTERS WITH APPROXIMATE EQUAL FREQUENCY AND COMBINATION OF LESS & MORE FREQUENCY

The consolidated result for the nine theories are as shown in the table 1.

S.NO	REG.NO	SUBJECT RESULT								
		T1-3	T2-3	T3-3	T4-3	T5-3	T6-3	L1-3	L2-3	L3-3
1.	9727K0201	P	P	P	P	P	P	P	P	P
2.	9727K0202	P	P	P	P	P	P	P	P	P
3.	9727K0203	F	P	P	F	P	P	P	P	P
4.	9727K0204	F	F	P	P	P	F	P	P	P
5.	9727K0205	P	P	F	P	P	P	P	P	P

6.	9727K0206	P	P	P	P	P	P	P	F	P
7.	9727K0207	P	P	P	P	F	P	P	P	P
8.	9727K0208	F	P	P	P	P	P	P	P	P
9.	9727K0209	P	P	P	P	P	P	P	P	P
10.	9727K0210	F	P	P	P	P	P	P	P	P
11.	9727K0211	P	P	F	P	P	P	P	P	P
12.	9727K0212	P	P	P	P	P	P	P	P	P
13.	9727K0213	F	P	P	P	P	F	P	F	P
14.	9727K0214	P	P	P	P	P	P	P	P	P
15.	9727K0215	P	P	P	P	P	P	P	P	P
16.	9727K0216	F	P	P	P	P	P	P	P	P
17.	9727K0217	P	F	P	P	P	P	P	P	P
18.	9727K0218	P	P	P	F	P	P	P	P	P
19.	9727K0219	F	P	P	P	P	P	P	P	P
20.	9727K0220	P	P	P	P	P	P	P	P	P
21.	9727K0221	P	P	P	P	P	P	P	P	P
22.	9727K0222	P	P	P	P	F	P	P	P	P
23.	9727K0223	P	P	P	P	P	P	P	P	P
24.	9727K0224	P	P	P	P	P	P	P	P	P
25.	9727K0225	F	P	P	P	P	F	P	P	P
26.	9727K0226	P	P	P	P	P	P	P	P	P
27.	9727K0227	P	P	P	P	P	P	P	P	P
28.	9727K0228	P	P	P	P	P	F	P	P	P
29.	9727K0229	P	P	P	P	P	P	P	P	P
30.	9727K0230	P	P	P	P	P	P	P	P	P
31.	9727K0231	P	P	P	P	P	P	P	P	P
32.	9727K0232	P	F	P	P	P	P	P	P	P
33.	9727K0233	P	P	P	P	P	P	P	P	P
34.	9727K0234	F	P	P	P	F	P	P	P	P
35.	9727K0235	P	P	P	P	P	P	P	P	P
36.	9727K0236	P	P	P	P	F	P	P	P	P
37.	9727K0237	P	P	P	P	P	P	P	P	P
38.	9727K0238	P	P	P	P	P	P	P	P	P
39.	9727K0239	F	P	P	P	P	F	P	P	P
40.	9727K0240	F	P	P	P	P	P	P	P	P
41.	9727K0241	F	P	P	P	P	P	P	P	P
42.	9727K0242	P	P	P	P	P	P	P	P	P
43.	9727K0243	P	P	P	P	P	P	P	P	P
44.	9727K0244	P	P	P	P	P	P	P	P	P
45.	9727K0245	P	P	P	P	P	P	P	P	P
46.	9727K0246	P	P	P	P	P	P	P	P	P
47.	9727K0247	F	F	F	F	F	F	F	F	P
48.	9727K0248	P	P	P	P	P	P	P	P	F
49.	9727K0249	F	P	P	P	P	P	P	P	P
50.	9727K0250	P	P	P	P	P	P	P	P	P

Table . 1. Consolidated results for nine subjects with six theory and three lab.

THEORY 1 RESULT
PPFFPPPFPFPPFPPFPPFPPPPPFPPPPPPPFPPPPFFF
PPPPPFPFP
THEORY 2 RESULT
PPPFPPPPPPPPPPPFPPPPPPPPPPPPPFPPPPPPPPP
PPPPPFPPP
THEORY 3 RESULT
PPPPFPPPPPFPPPPPPPPPPPPPPPPPPPPPPPPPPPP
PPPPPFPPP
THEORY 4 RESULT
PPFPPPPPPPPPPPPPPFPPPPPPPPPPPPPPPPPPPPP
PPPPPFPPP
THEORY 5 RESULT
PPPPPPFPPPPPPPPPPPPPFPPPPPPPPPPPFPFPPPPP
PPPPPFPPP
THEORY 6 RESULT
PPPFPPPPPPPPFPPPPPPPPPPPFPPFPPPPPPPPPPFPP
PPPPPFPPP
LAB 1 RESULT
PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
PPPPPFPPP
LAB 2 RESULT
PPPPPFPPPPPPPFPPPPPPPPPPPPPPPPPPPPPPPPPP
PPPPPFPPP
LAB 3 RESULT
PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
PPPPPPFPP
Replacing P by 1 and F by 0 the result will be
THEORY 1 RESULT
11001110101101101101111101111111011110001111
10101
THEORY 2 RESULT
11101111111111101111111111111101111111111111
10111
THEORY 3 RESULT
11110111110111111111111111111111111111111111
10111
THEORY 4 RESULT
11011111111111110111111111111111111111111111
10111
THEORY 5 RESULT
11111101111111111111101111111111010111111111
10111
THEORY 6 RESULT
11101111111101111111111101101111111111101111
10111
LAB 1 RESULT
11111111111111111111111111111111111111111111
10111
LAB 2 RESULT
11111011111101111111111111111111111111111111
10111
LAB 3 RESULT
11111111111111111111111111111111111111111111
11011

THEORY 1 RESULT
11001110101101101101111101111111011110001111
10101
223111212121518143511111
Total Number of Character after encoding = 23
THEORY 2 RESULT
11101111111111110111111111111110111111111111
10111
319031905190513
Total Number of Character after encoding = 15
THEORY 3 RESULT
11110111101111111111111111111111111111111111
10111
4151909090813
Total Number of Character after encoding = 13

THEORY 4 RESULT
11011111111111111011111111111111111111111111
10111
219051909090113
Total Number of Character after encoding = 15
THEORY 5 RESULT
11111101111111111111101111111111010111111111
10111
619051902111190113
Total Number of Character after encoding = 18
THEORY 6 RESULT
11101111111101111111111101101111111111101111
10111
31819021219011713
Total Number of Character after encoding = 17
LAB 1 RESULT
11111111111111111111111111111111111111111111
10111
9090909090113
Total Number of Character after encoding = 13
LAB 2 RESULT
11111011111101111111111111111111111111111111
10111
5161909090613
Total Number of Character after encoding = 13
LAB 3 RESULT
11111111111111111111111111111111111111111111
11011
9090909090212
Total Number of Character after encoding = 13

IV. RESULT AND DISCUSSION

The compression ratio, space savings and average bits calculated for the examples are

For the input :
eeeeeeeeeeeeezzzzhhhhhhhhhhhhhiiiiiiiiiiilllllllll

Compression ratio = 400/112
= 50:14
=3.57:1
Space savings =1-(112/400)
= 1-(7/25)
= 1-0.28
= 0.72
= 72%
Average bits = 112/50
= 2.24 bits per character

For the input :
uuuuuuummmmmmmaaaaaddddddeeeeeeeeeeeeeeevvviiiiiii

Compression ratio = 400/120
= 50:15
=3.66:1
Space savings =1-(120/400)
= 1-(15/50)
= 1-0.3
= 0.7
= 70%
Average bits = 120/50
= 2.4 bits per character

For the input
:pppppppppaaaaaaaaallllllllllaaaaaaaaaaaannnnniiiiii

Compression ratio = 400/112
= 50:14
=3.57:1
Space savings =1-(112/400)
= 1-(7/25)
= 1-0.28
= 0.72
= 72%
Average bits = 112/50
= 2.24 bits per character

The Consolidated Compression ratio, Space savings and Average bits for nine subjects are as shown in the table 2.

S.No	Subject	Compression Ratio	Space savings	Average bits
1	Theory1	2.173913	54%	3.68
2	Theory2	3.333333	70%	2.4
3	Theory3	3.846154	74%	2.08
4	Theory4	3.333333	70%	2.4
5	Theory5	2.777778	64%	2.88
6	Theory6	2.941176	66%	2.72
7	Lab1	3.846154	74%	2.08
8	Lab2	3.846154	74%	2.08
9	Lab3	3.846154	74%	2.08

Table . 2. Consolidated Compression ratio, Space savings and Average bits for nine subjects with six theory and three lab.

V. CONCLUSION

The obtained results depicts that the Run-Length encoding gives better compression ratio, space savings, and average bits for the input with the long Run-Length. The Run-Length Encoding gives better compression ratio, space savings, and average bits as compared with the uncompressed data.

REFERENCE

[1] https://en.wikipedia.org/wiki/Run-length_encoding
[2] Khalid Sayood, "Introduction to Data Compression", Morgan Kaufmann Publishers, Burlington, United States of America, 2nd edition, 600 pages, 2000.
[3] Mark Nelson and Jean-loup Gailly , "The Data Compression Book", M&T Books, New York, United States of America, 2nd edition, 541 pages, 1995.
[4] David Salomon, "Data Compression: The Complete Reference", Springer, New York, Berlin, Heidelberg, United States of America, Germany, 2nd edition, 823 pages, 2000.
[5] David Salomon and Giovanni Motta, "Handbook of Data Compression", Springer London, 2000.
[6] J.A. Storer, "Data Compression", Computer Science Press, Rockville, MD ,1988.

[7] E. Blelloch, "Introduction to Data Compression", Computer Science Department, Carnegie Mellon University, 2002.

[8] Lynch, J. Thomas, "Data Compression: Techniques and Applications", Lifetime Learning Publications, Belmont, CA, 1985

[9] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, Pearson Education, 2002

[10] S. W. Golomb, "Run-Length Encodings," IEEE Transactions on Information Theory, vol. IT-12, no. 3, pp. 399–401, July 1966.