Muhammad Raheel.

Design and Analysis of Algorithms.

CID#103415.

SID#62793.

Submitted to: Sir Muhammad Fahad.

Github: https:/github.com/raheel4033/DAA

The assignment’s goal was to read the papers and then propose some logical solution in terms of our own understanding which is related to that particular problem. Here are my problems with the proposed solutions.

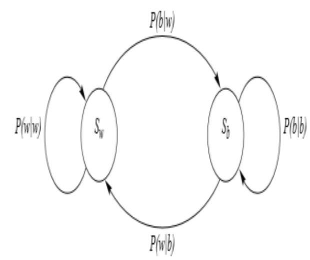
INDEX:

Problem 3 in this document is Not Related to Compression

****While Paper 3 consists of the artificial neural network which helps in solving the rice problem in some state of India****.

Mathematical Preliminaries are very important for the lossless data compression as we have to gather the compression without losing the data at any cost. So here are some models and papers discussed which are used in the compression.

Problem 1.

Problem Statement with Mathematical Model.	Dependence.	Further Work.	Advancement.
One of the most popular way of representing data is through the markovs model as it is a process of discrete time markovs chain with sequence in kth order of markovs chain. $P(x_n \| x_{n-1}, \ldots, x_{n-k}) = P(x_n \| x_{n-1}, \ldots, x_{n-k}, \ldots).$	There can be a dependence in the linear manner as output will be given by white noise having equation. $x_n = \rho x_{n-1} + \epsilon_n$	It doesn’t require to be linear only for image it can have the white and black pixel also on which one pixel is dependent on another pixel which is: 	Markov model can be used in the text compression as the probability of preceding letter is heavily depended on the next letter. The kth of the markov model is widely known as finite context model.

1) Suggested Solution

The solution can be given as the number the number of chains which are in the kth sequence must be placed in such the order the all the values received are in such order that the values dependence must be accurate to the previous one.

P(x(n)|x(n-1),x(n-2),.....x(n-k))

in my proposed solution.

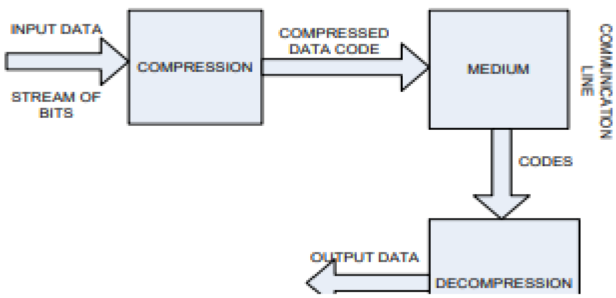
One value can repeat itself anywhere in the middle to make the preceding value heavily dependent on itself.

P(x(n)|x(n-1),x(n-2),.....,x(n-9),x(n-1),x(n-10))

here the limit must kept 10 because one choice value of compression must be selected at last to make compression value able because if used value appear somewhere in the middle it might disturb the the frequency and dependence of previous an also the repetition.

Problem 2

Data compression is most widely used technique for the representation of the data in fewer bits as compared to what it has been in its original form. It helps in reducing hard disk space and transmission bandwidth. There are two types of compression techniques Lossy and lossless. In lossless technique every bit in the file remain as original after the compression while in Lossy some redundant information can be reduced.



In this technique the data bits can be represented into it half as 128 bits will be represented in the 64 bits. Similarly 64 bits into 32 bits and going on. The truth table will be.

A	B	Z
0	0	0
0	1	$\bar{0}$
1	0	$\bar{1}$
1	1	1

Truth table of proposed technique

In table 1:

Resultant output is 0 when two input bits are 0 and 0

Resultant output is 1 when two input bits are 1 and 1

Resultant output is $\bar{0}$ when two input bits are 0 and 1

Resultant output is $\bar{1}$ when two input bits are 1 and 0

Input data sequence is checked where it is even or odd. If even then process or if odd then add a bit either 0 or 1 depend on last bit. If the last bit of input sequence is 0 then add 0 or if the last bit of input sequence is 1 then add 1.Then it feed to our proposed technique. In this method two bits are collected and converted it into one bit using table 1. So we can get compress data.

This is also one of the good data compression techniques but it must be looked into the concern with how it decodes to the further the on big data because reducing bits on large data can cause problems but it can be handled if taken care of properly.

2)Suggested Solution.

In this sequence where the data is which is 128 allowed to convert to 64 and 64 into 32 in the same way it is going like (2^n)/2 where n is 5 (2^5) = 32 and 32/2 = 16 which is some of its result and then it is placed in the truth table for the further generation of the data and then allowing the compression. So, here i suggest a formula for altering this a bit which might improve it further more.

$$((2^n)-4)/2$$

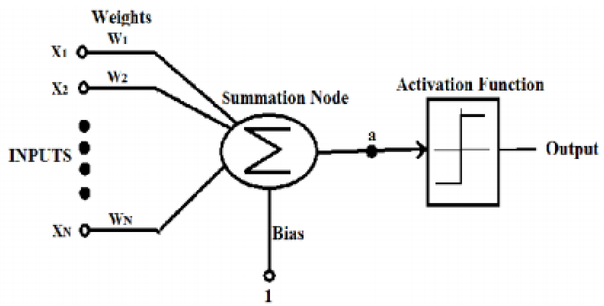
It now yield some reduced bits in the truth table and also will be easy to decompress by just doing.

$$(2^n)*2+4$$

which will simply gives the result of the table to which bits had been reduced.

Problem 3

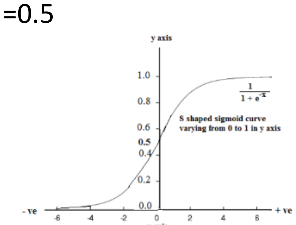
<table>
<tr>
<td>The neural network is massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for the use. The artificial neural network is similar to that of how a brain works. It has the input layer which takes some input and make summation then place the output in such a manner.



</td>
<td>Sigmoid function is a function which gives the sigmoid curve given by formula

$$S(t) = \frac{1}{1+e^{-t}}.$$

Sigmoid functions are very similar to the input-output relationships of biological neurons, although not exactly the same. Sigmoid function exhibits smoothness and has the desired asymptotic properties. The sigmoid curve is shown in below. As t goes to minus infinity, S(t) goes to 0. As t goes to infinity, S(t) goes to 1. As t =0, S(t) =0.5



</td>
<td>Error correction is a technique which is used to measure the actual output to the desired output. Error values can be used by back propagation and adjusting the weights of the input in the first layer

$$\text{error} = \frac{\text{actual output - ANN output}}{\text{actual output}}$$

If error is greater than the threshold then using the back propagation check updated weights and compute the summation.</td>
<td>The ANN can be used in the prediction of the rainfall using back propagation as well as comparison of Dynamic and Static neural network can also be made. The model can also be used for the long-range parameters and pattern recognition.</td>
</tr>
</table>

3)ANN problem for rice in an state of India.

In this problem instead of changing the inputs after error checking in the back propagation the simple solution could have been to introduce to new way in which the error is added to the actual answer which came up and then added the value of error of that particular sigmoid function to the input so the formula instead of back propagating and changing the input weights correct itself it that position only.

S(t) = 1/1+e^-t   (1) sigmoid function

error = (actual output - ANN output)/actual output (2) error

now error must be calculated in such a way that when added with the function it yields similar desired output.

Desired Output = Sigmoid result + error

Problem 4

Source output consists of 4 bit word {0,1,2,…..,15}.The source encoder encodes each value by shifting out the less significant bit. The output alphabet for the source coder is {0,1,2,……,7}. At the receiver we cannot recover the original value. Let X be random variable that takes values from source alphabet X={x0,x1,……,xn-1} and Y takes the random variable values form reconstruction variable Y={y0,y1,….,ym-1}. $$H(X) = -\sum_{i=0}^{N-1} P(x_i)\log_2 P(x_i)$$ $$H(Y) = -\sum_{j=0}^{M-1} P(y_j)\log_2 P(y_j).$$	A measure of relationship between two random variables is called conditional entropy. $$i(A) = \log \frac{1}{P(A)} = -\log P(A).$$ In a similar manner, the conditional self-information of an event A, given that another event B has occurred, can be defined as $$i(A\|B) = \log \frac{1}{P(A\|B)} = -\log P(A\|B).$$	As In case of self-information we are generally interested in the average value of the self-information. The conditional entropy with source and reconstruction is defined as: $$H(X\|Y) = -\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} P(x_i\|y_j)P(y_j)\log_2 P(x_i\|y_j)$$ $$H(Y\|X) = -\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} P(x_i\|y_j)P(y_j)\log_2 P(y_j\|x_i).$$	The amount of uncertainty About source X and reconstruct Y. The additional knowledge of Y should reduce uncertain X. $$H(X\|Y) \leq H(X)$$

4)Entropy Problem for compression.

There is a possibility that the random variable values can be placed adjacent to the reconstruction variables and the variables that has been occurred once in the reconstruction can help in mapping when that particular variable occurs in the reconstructed at once can be place in its place again for example.

X = {A,B,D,C,A,A,A}

Y = {A,B,D,C,Y0,Y0,Y0}

so in this way when A,A,A occurs in the reconstruction table again it knows that A has been occurred in Y0 so just use it in that manner and save the bits occurring from twice to reduce the similarity checks and place that for once only.

In this way uncertainty issue can also be handled for the X random to Y reconstruction as decompression will be easier and some way optimal.

References:

1) Introduction to data compression by Khalid Sayood

Markov model 2.3.3 page 23-26

2) An Improved Data Compression Method for General Data by Salauddin Mahmud

3) DESIGN AND DEVELOPMENT OF ARTIFICIAL NEURAL NETWORKING (ANN) SYSTEM USING SIGMOID ACTIVATION FUNCTION TO PREDICT ANNUAL RICE PRODUCTION IN TAMILNADU S.Arun Balaji1 and K.Baskaran2

4) Introduction to data compression by Khalid Sayood

Conditional Entropy 8.4.1 page 202-204