Information and Compression

Paul Cockshott
Faraday partnership

Introduction

- We will be looking at what information is
- Methods of measuring it
- The idea of compression
- 2 distinct ways of compressing images

Summary of Course

- At the end of this lecture you should have an idea what information is, its relation to order, and why data compression is possible

Agenda

- Encoding systems
- BCD versus decimal
- Information efficiency
- Shannons Measure
- Chaitin, Kolmogorov Complexity Theory
- Relation to Goedels theorem

Overview

- Data coms channels bounded
- Images take great deal of information in raw form
- Efficient transmission requires denser encoding
- This requires that we understand what information actually is

Connections

- Simple encoding example shows different densities possible
- Shannons information theory deals with transmission
- Chaitin relates this to Turing machines and computability
- Goedels theorem limits what we know about compression

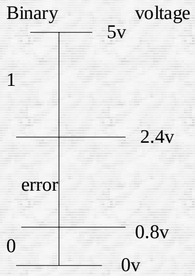
Vocabulary

- information,
- entropy,
- compression,
- redundancy,
- lossless encoding,
- lossy encoding

What is information

- Information measured in bits
- Bit equivalent to binary digit
- Why use binary system not decimal?
- Decimal would seem more natural
- Alternatively why not use e, base of natural logarithms instead of 2 or 10 ?

Voltage encoding



- Digital information encoded as ranges of continuous variables
- Each digital value takes a range
- isolation bands in between

Noise Immunity

- BINARY
- 2 values 0,1
- 1 isolation band
- 1.6 volt isolation band
- Good noise imunity, 0.5 volt noise on power no problem

- DECIMAL
- 10 values 0..9
- 9 isolation bands
- 0.32 volt isolation bands
- 0.5 volt power noise would destroy readings

Decimal Computing

- Great advantages of decimal for commerce, banking
- Burroughs, ICL went on making decimal computers until late 1970s
- Burroughs B8 series, ICL System 10
- Used Binary Encoded Decimal BCD

BCD encoding of 1998

0	0000	
1	0001	
2	0010	
3	0011	
4	0100	
5	0101	
6	0110	
7	0111	
8	1000	
9	1001	

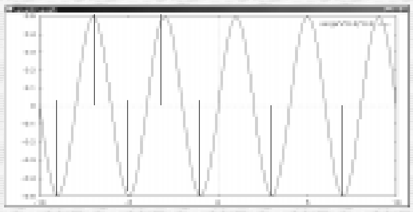
- Reliable since encoded in binary
- 1998 represented as
- 0001 1001 1001 1000
- In pure binary this is
- 0000 0111 1100 1110
- BCD 13 sig digits
- Binary 11 sig digits

BCD less dense

- BCD requires 16 bits to represent dates up to 9999 BCD : 1001 1001 1001 1001
- Binary requires 14 bits to go up to 9999 binary: 10 0111 0000 1111
- Why ?
- What is the relative efficiency of two encodings?

Nyquist Limit

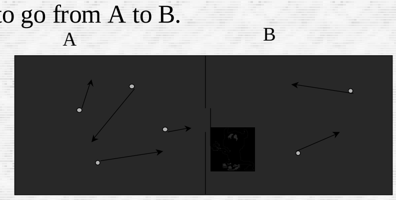
Here capture works



To capture a frequency with discrete samples the capture rate
Must be twice the frequency.

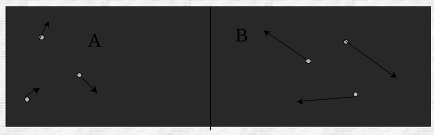
Maxwells Demon

- Daemon opens door only for fast molecules to go from A to B.



End Result – Free Heat

- Slow molecules in A, fast in B
- Thus B hotter than A, and can be used to power a machine



This apparently contradicts second law of thermodynamics
Entropy of the system has been reduced.

Szilards response

- Szilard pointed out that to decide which molecules to let through, the daemon must measure the speed of them. Szilard showed that these measurements (bouncing photons off the molecules) would use up more energy than was gained.

Topic Two: Shannon's Theory

- Shannon defines a bit as *the amount of information required to decide between two equally probable outcomes*
- **Example**: a sequence of tosses of a fair coin can be encode 1 bit per toss, such that heads are 1 and tails 0.

Information as Surprise

According to Shannon the information content of a message is a function of how surprised we are by it.

The less probable a message the more information it contains

$$H = - \log_2 p$$

H = information also called ***entropy***, *p* the probability of occurrence of a messages. The mean information content of an ensemble of messages is obtained by weighting the messages by their probability of occurence

$$-\Sigma\, p \log_2 p$$

This explains inefficiency of BCD

- Consider the frequency of occurrence of 1 and 0 in the different columns of a BCD digit.
- Only the least significant digit decides between equiprobable outcomes
- For most significant digit, 0 is four times more common than 1

8s	**4s**	**2s**	**1s**	***Value***
0	0	0	0	*0*
0	0	0	1	*1*
0	0	1	0	*2*
0	0	1	1	*3*
0	1	0	0	*4*
0	1	0	1	*5*
0	1	1	0	*6*
0	1	1	1	*7*
1	0	0	0	*8*
1	0	0	1	*9*
2	4	4	5	*total*
0.2	**0.4**	**0.4**	**0.5**	***Prob = 1***

Determining encoding efficiency

- For the most significant digit of BCD we have
- *p*= 0.2 of a 1, *p*=0.8 of a 0
- Applying Shannon's formula a 1 contributes 2.321 bits of information, but with a probability of 0.2 this contributes 0.464 bits
- Sum of information for 0 and 1 is 0.721 bits

0.800	0.4	0.6
0.322	1.322	0.737
0.258	0.529	0.442
	0.971	

Mean info of a BCD digit

Bit position	8s	4s	2s	1s
Information in bits	0.722	0.971	0.971	1
Total for BCD digit	3.664			

Thus the 4 bits of a BCD digit contain at most 3.664 bits of information. *Binary is a denser encoding.*

Channel capacity

- The amount of data that can be sent along a physical channel – for example a radio frequency channel depends upon two things

1. The frequency of the channel – follows from Nyquist
2. The signal to noise ratio of the channel

Signal to noise ratio

- As the signal to noise ratio falls, the probability that you get an error in the data sent rises. In order to protect against errors one needs to use error correcting codes which have greater redundancy

Channel capacity formula

$$C = W \log_2 (P+N)/N$$

- Where
  - C is channel capacity in bits per second
  - P is signal power (say in watts)
  - N is noise power (say watts)
  - W is the frequency of the channel
- *Note that the dimensions of P and N cancel*

Topic Three: Algorithmic Information

Chaitin defines the information content of a number (or sequence of binary digits) to be the length of the shortest computer program capable of generating that sequence.

He uses Turing machines as cannonical computers Thus the information content of a sequence is the shortest Turing machine tape that would cause the machine to halt with the sequence on its output tape.

Randomness and π

We know from Shannon that 1 million tosses of fair coin generates 1 million bits of information.

On the other hand, from Chaitin we know that π to 1 million bit precision contains much less than 1 millon bits, since the program to compute π can be encoded in much less than 1 million bits.

Randomness

Andrei Kolgomorov defined a random number as a number for which no formulae shorter than itself existed.

By Chaitin's definition of information a random number is thus incompressible.

A Random Sequence Thus Contains the Maximum Information.

Randomness goal of compression

- A fully compressed data sequence is indistinguishable from a random sequence of 0s and 1s.
- This follows directly from Kolgomorov and Chaitin's results.
- From Shannon we also have the result that for each bit of the stream to have maximal information it must mimic the tossing of a fair coin : be unpredictable, random.

Information is not meaning

- One million digits of π are more meaningful than one million random bits.
- *But they contain less information.*
- Meaningful sequences generally contain redundancy.

Information is not complexity

- One million digits from π have less information than a million random bits but they are more *complex*.
- The complexity of a sequence is measured by the number of machine cycles a computer would have to go through to generate it (Bennet's theorem of logical depth).

Topic Four: Limits to compression

- *We can not in principle come up with an algorithm for performing the maximum compression on any bit sequence.*
- 3/7 is a rule of arithmetic, an algorithm that generates the sequence 0.42857142857
- So this sequence is presumably less random than 0.32857142877 (changed 2 digits)
- But we can never be sure.

Consequence of Goedels theorem

- Showed we can not prove completeness of a consistent set of arithmetic axioms. There will be true statements that can not be proven.
- If there existed a general procedure to derive the minimal Turing machine program for any sequence, then we would have a procedure to derive any true proposition from a smaller set of axioms, *contra* Goedel.

Image sizes

- QCIF: 160 x 120 (used in video phones)
- MPEG: 352 x 288
- VGA: 640 x 480
- NTSC: 720 x 486 (US Television)
- Workstation: 1280 x 1024
- HDTV: 1920 x 1080
- 35mm slide: 3072 x 2048

Storage capacities

Floppy disk

- Floppy disk capacity = 1.44 MB
- A single 1280x1024x24 image = 3.9 MB
- A single 640x480x24 image = 922K
- Floppy disk holds only one 640x480x24 image.

Video on CD rom

- CD-ROM capacity = 600 MB
- A 160x120x16 image @30 fps (QCIF)= 1.15 MB/sec
- CD-ROM now holds 8.7 minutes of video

(still not enough), but only at about a quarter the image size of normal TV

Telecoms links

- Ethernet = 10 meg bits /sec
- ISDN = 128 K bits /sec
- ADSL = max 2 meg bit /sec

Compute max frame rates for QCIF for each of these channels

Summary

- Information = randomness = disorder = *entropy*.
- Meaningful data is generally compressible.
- Information measured by length of program to generate it.
- We will look at two techniques to apply these ideas to images: Fractal compression, and Vector Quantization.

Where to get more information

- 'The user Illusion' by Tor Norretranders , Penguin 1998.
- 'Information Randomness and Incompleteness' by Chaitin
- 'Information is Physical' by Landauer, in *Physics Today*,May 1991.
- 'Logical depth and physical complexity' by Bennet, in *The Universal Turing Machine a Half Century Survey*, Oxford 1988.