Foundations of Computer Security

Lecture 36: Entropy Odds and Ends

Dr. Bill Young
Department of Computer Sciences
University of Texas at Austin

Entropy Aside

Recall that information content of a message depends on the state of knowledge of the receiver. Hence, entropy is relative to a particular observer.

Consider the entropy of the contents of an envelope marked "Best Picture" at the Academy Awards (assuming 5 nominees):

- If all were equally likely to win, the entropy would be $\log 5 \approx 2.322$.
- For everyone who knows that the odds aren't even, it's less, though hard to compute.
- For the auditors who stuffed the envelope, it's 0 since they have no uncertainty.

Often, prior probabilities are impossible to compute.

Entropy and Randomness

Note that entropy can be used to measure the amount of "redundancy" in the encoding. If the information content of a message is equal to the length of the encoded message, there is no redundancy.

Some sources define a *random* string as one that cannot be represented any more efficiently. (I.e., no compression is possible.)

Finding a Coding

Huffman coding is guaranteed to find an efficient code for a given language *assuming* you know the probabilities of language units.

In fact, it always uses *less than one bit per symbol more than the entropy,* which is extremely efficient.

Lempel-Ziv is an "adaptive coding" algorithm used in many commercial text compression utilities. It builds an encoding on the fly according to the strings it encounters.

Lempel-Ziv is *asymptotically optimal*. That is, as the text length tends to infinity, the compression approaches optimal.

Lessons

- The information content of a message is relative to the state of knowledge of an observer.
- If an encoding's efficiency matches the entropy, there is no redundancy to compress out.
- Huffman coding and the Lempel Ziv algorithms both give highly efficient codes.

Next lecture: Cryptography