Entropy encoding

From Wikipedia, the free encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Languages
Català
Deutsch
Español
Français
한국어
Italiano
日本語
Русский
Українська
中文
Edit links

This article includes a list of references, related reading or external links, **but its sources remain unclear because it lacks inline citations**. Please improve this article by introducing more precise citations. *(December 2013)*

In information theory an **entropy encoding** is a lossless data compression scheme that is independent of the specific characteristics of the medium.

One of the main types of entropy coding creates and assigns a unique prefix-free code to each unique symbol that occurs in the input. These entropy encoders then compress data by replacing each fixed-length input symbol with the corresponding variable-length prefix-free output codeword. The length of each codeword is approximately proportional to the negative logarithm of the probability. Therefore, the most common symbols use the shortest codes.

According to Shannon's source coding theorem, the optimal code length for a symbol is $-\log_b P$, where b is the number of symbols used to make output codes and P is the probability of the input symbol.

Two of the most common entropy encoding techniques are Huffman coding and arithmetic coding. If the approximate entropy characteristics of a data stream are known in advance (especially for signal compression), a simpler static code may be useful. These static codes include universal codes (such as Elias gamma coding or Fibonacci coding) and Golomb codes (such as unary coding or Rice coding).

Entropy as a measure of similarity [edit]

Besides using entropy encoding as a way to compress digital data, an entropy encoder can also be used to measure the amount of similarity between streams of data and already existing classes of data. This is done by generating an entropy coder/compressor for each class of data; unknown data is then classified by feeding the uncompressed data to each compressor and seeing which compressor yields the highest compression. The coder with the best compression is probably the coder trained on the data that was most similar to the unknown data.

External links [edit]

- *Information Theory, Inference, and Learning Algorithms*, by David MacKay (2003), gives an introduction to Shannon theory and data compression, including the Huffman coding and arithmetic coding.
- *Source Coding*, by T. Wiegand and H. Schwarz (2011).

v · t · e	Data compression methods	[hide]
Lossless	**Entropy type**	Unary · Arithmetic · Golomb · Huffman (Adaptive · Canonical · Modified) · Range · Shannon · Shannon–Fano · Shannon–Fano–Elias · Tunstall · Universal (Exp-Golomb · Fibonacci · Gamma · Levenshtein)
	Dictionary type	Byte pair encoding · DEFLATE · Lempel–Ziv (LZ77 / LZ78 (LZ1 / LZ2) · LZJB · LZMA · LZO · LZRW · LZS · LZSS · LZW · LZWL · LZX · LZ4 · Statistical)
	Other types	BWT · CTW · Delta · DMC · MTF · PAQ · PPM · RLE
Audio	Concepts	Bit rate (average (ABR) · constant (CBR) · variable (VBR)) · Companding · Convolution · Dynamic range · Latency · Nyquist–Shannon theorem · Sampling · Sound quality · Speech coding · Sub-band coding
	Codec parts	A-law · μ-law · ACELP · ADPCM · CELP · DPCM · Fourier transform · LPC (LAR · LSP) · MDCT · Psychoacoustic model · WLPC
Image	Concepts	Chroma subsampling · Coding tree unit · Color space · Compression artifact · Image resolution · Macroblock · Pixel · PSNR · Quantization · Standard test image
	Methods	Chain code · DCT · EZW · Fractal · KLT · LP · RLE · SPIHT · Wavelet
Video	Concepts	Bit rate (average (ABR) · constant (CBR) · variable (VBR)) · Display resolution · Frame · Frame rate · Frame types · Interlace · Video characteristics · Video quality
	Codec parts	Lapped transform · DCT · Deblocking filter · Motion compensation

Theory	Entropy · Kolmogorov complexity · Lossy · Quantization · Rate–distortion · Redundancy · Timeline of information theory
	Compression formats · Compression software (codecs)

Categories: Lossless compression algorithms | Entropy and information