Information theory and coding

Data compression:
[MacKay, 2003]

Antonio Sutera
a.sutera@uliege.be

March 21, 2019

Previously on...

Symbol codes

- Shannon code
- Fano code
- Huffman code: optimal

$\bar{l}_{\mathrm{Huffman}} \leqslant \min\{\bar{l}_{\mathrm{Shannon}}, \bar{l}_{\mathrm{Fano}}\}$

but $\bar{l}_{\mathrm{Shannon}} \gtrless^{?} \bar{l}_{\mathrm{Fano}}$

Previously on...

Arithmetic coding

- Shannon-Fano-Elias:
  $\bar{l}_{\mathrm{Shannon-Fano-Elias}} < H_m(f_1, \ldots, f_m) + 2$

Dasher [MacKay, 2003]

Riddle

The string:

A good example of how dictionary based compression

can be coded as:

1/1 822/3 674/4 1343/60 928/75 550/32 173/46 421/2

How?

Riddle: Solution

The string:

A good example of how dictionary based compression

can be coded as:

1/1 822/3 674/4 1343/60 928/75 550/32 173/46 421/2

- Using the dictionary "Random House Dictionary" of the English language, 2^{nd} edition, Unabridged
- Each word is coded as x/y, where x gives the page in the dictionary, and y gives the number of the word on that page.

References

MacKay, D. J. (2003). Information theory, inference, and learning algorithms.