

Lecture(1):Entropy and Data Compression

Abdelrahman Elnamaki

September 18, 2024

Introduction

In this lecture, we delve into the fundamental concepts of entropy and data compression, central to information theory. Entropy, a measure of uncertainty or unpredictability, is pivotal in understanding how much information is conveyed by a data source. We explore Shannon's information content, which quantifies the amount of information based on the probability of events, and Shannon entropy, which provides an average measure of uncertainty across all possible outcomes. By examining these concepts, we gain insights into the theoretical limits of data compression and how they guide the development of efficient compression algorithms.

1 Shannon Information Content

Shannon's information content is a fundamental concept in information theory that measures how much *information* is gained from an event or message. It quantifies the surprise or uncertainty of an event—more surprising events carry more information. If something is **very likely** to happen (like the sun rising tomorrow), the information you gain from knowing it is **low** because it's not surprising. If something is **unlikely** to happen (like winning the lottery), the information content is **high** because it's a rare and surprising event.

Shannon's information content measures this using probability. The less probable an event, the more information it conveys when it happens.

1.1 Information Content Formula

The information content $I(x)$ of an event x , based on its probability $P(x)$, is given by:

$$I(x) = -\log_b P(x)$$

Where:

- $I(x)$ is the information content (measured in bits if the base of the logarithm is 2).
- $P(x)$ is the probability of the event x .
- b is the base of the logarithm. Typically, $b = 2$ is used, meaning the information is measured in bits.

Explanation:

- **Low probability (surprising event):** If $P(x)$ is small, $-\log_2 P(x)$ will be large, meaning high information content.
- **High probability (expected event):** If $P(x)$ is large, $-\log_2 P(x)$ will be small, meaning low information content.

For example:

- If the probability of an event x is $\frac{1}{2}$ (a 50% chance), then the information content is:

$$I(x) = -\log_2 \left(\frac{1}{2} \right) = 1 \text{ bit}$$

- If the probability of an event y is $\frac{1}{4}$ (a 25% chance), the information content is:

$$I(y) = -\log_2 \left(\frac{1}{4} \right) = 2 \text{ bits}$$

The rarer the event (lower probability), the higher the information content.

1.2 Shannon Entropy Formula

Shannon's *entropy* is the average information content of all possible events in a probability distribution. It measures the *uncertainty* of the entire system. The entropy $H(X)$ of a random variable X with possible outcomes x_1, x_2, \dots, x_n , and probabilities $P(x_1), P(x_2), \dots, P(x_n)$, is given by:

$$H(X) = - \sum_{i=1}^n P(x_i) \log_b P(x_i)$$

Where:

- $H(X)$ is the entropy (average information content).
- $P(x_i)$ is the probability of each event x_i .
- b is the base of the logarithm (typically 2 for bits).

Explanation:

- If all events are equally likely, the entropy is high because there's a lot of uncertainty.
- If one event is very likely and others are very unlikely, the entropy is low because there's less uncertainty about the outcome.

For example, if you have a fair coin (two equally probable outcomes, heads or tails):

$$H(X) = - \left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) = 1 \text{ bit}$$

This shows that flipping a fair coin has an entropy of 1 bit, which means each flip provides 1 bit of information.

1.3 Summary

- **Shannon's information content:** $I(x) = -\log_2 P(x)$ measures the surprise or information from an event.
- **Shannon entropy:** $H(X) = -\sum P(x_i) \log_2 P(x_i)$ measures the average uncertainty of all possible events.

Both these concepts are key in understanding how information can be efficiently represented and transmitted in communication systems.

2 Source Coding Theorem

The **Source Coding Theorem**, introduced by Claude Shannon, is a key result in information theory that establishes the theoretical limits of lossless data compression. It provides a way to quantify how efficiently we can compress information from a source, while still allowing perfect reconstruction of the original message.

2.1 Shannon Entropy and Compression Limit

The Shannon Entropy $H(X)$ of a source X is a measure of the average amount of information produced by the source per symbol.

The **Source Coding Theorem** states that for a source with entropy $H(X)$, it is possible to encode sequences of symbols from the source using an average of $H(X)$ bits per symbol. More formally, if a source emits N symbols, we can compress the message into approximately $N \cdot H(X)$ bits without losing information. This is the *optimal limit* of compression, and no further compression can be achieved without loss of information.

2.2 Example

Consider a source that generates four possible symbols $\{A, B, C, D\}$ with the following probabilities:

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{1}{4}, \quad P(C) = \frac{1}{8}, \quad P(D) = \frac{1}{8}$$

Using the entropy formula, the entropy $H(X)$ of this source is:

$$H(X) = - \left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{8} \log_2 \frac{1}{8} + \frac{1}{8} \log_2 \frac{1}{8} \right)$$

$$H(X) = 1 + 0.5 + 0.375 + 0.375 = 2.25 \text{ bits per symbol.}$$

This means, on average, we need 2.25 bits to represent each symbol from this source.

Now, if the source generates 1000 symbols, the total number of bits required to encode the entire message can be approximated as:

$$N \cdot H(X) = 1000 \times 2.25 = 2250 \text{ bits.}$$

Thus, it is theoretically possible to compress the message into 2250 bits while maintaining the ability to perfectly reconstruct the original sequence.

2.3 Practical Implications

In practice, compression algorithms such as **Huffman Coding** and **Arithmetic Coding** are used to approach this theoretical limit. For example, compression formats like ZIP and PNG aim to reduce file sizes by encoding data close to the source's entropy. The Source Coding Theorem ensures that the best possible compression can be achieved by encoding messages with an average number of bits per symbol equal to the entropy $H(X)$.

Conclusion

This lecture highlights the crucial role of entropy in data compression and its implications for efficient information transmission. Shannon's concepts of information content and entropy provide a theoretical framework for understanding how to minimize the number of bits required to represent information. By leveraging these principles, we can develop and refine compression algorithms to achieve optimal performance. Understanding these concepts equips us with the tools to enhance communication systems, ensuring data is stored and transmitted effectively in various applications.