UNIVERSITY OF TARTU
Institute of Computer Science
Computer Science Curriculum

Sten Sootla

Analysing information distribution in complex systems

Bachelor's Thesis (9 ECTS)

Supervisor: Raul Vicente Zafra, PhD
Supervisor: Dirk Oliver Theis, PhD

Tartu 2017

Contents

1 Introduction

Complex systems

In Chapter 1, the basics of classical information theory and partial information decomposition are covered. The chapter ends with an overview of the numerical estimator for PID.

The subsequent 3 chapters each introduce a specific complex system and the results of measuring information distribution in it.

Chapter 2 - Elementary Cellular Automata Chapter 3 - Ising model Chapter 4 - Artificial Neural Networks

In the final, concluding chapter, a summary of the contributions of this thesis is given, alongside suggestions for further work.

2 Background

2.1 Classical information theory

In order to understand partial information decomposition, which is the mathematical framework that is used in this thesis to analyse complex systems, a solid understanding of basic information theory is essential. This subsection fills that gap, giving a brief overview of the fundamental concepts of information theory. Where appropriate, the rather abstract definitions are further elaborated on by providing the reader with intuitive explanations and concrete examples.

2.1.1 Entropy

Entropy is the most fundamental quantity of information theory.

Let X be a discrete random variable with possible realizations from the set $\{x_1, x_2, ..., x_n\}$ and a probability mass function $p(x_i) = P\{X = x_i\}$ $(i = 1, ..., n)$. Shannon [Sha48] defines the *entropy* of X as follows:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \tag{1}$$

If the base of the logarithm is 2, the units the entropy is measured in are called *bits*. Another common base for the logarithm is Euler's number $e \approx 2.718$, in which case the units of measurment are called *nats*.

Intuitively, entropy can be thought of as the average amount of uncertainty one has about a random variable. It is indeed the *average* amount of uncertainty, because the uncertainty of a single relatization of x_i of a random variable X can be quantified by $-log_2 p(x_i)$. Viewed from this angle, the definition of entropy can be rewritten as an expectation of the random variable $-log_2 p(X)$:

$$H(X) = \mathbb{E}\left[-\log_2 p(X)\right] = \mathbb{E}\left[\log_2 \frac{1}{p(X)}\right].$$

To see why this intuition should correspond to the mathematical definition, it is instructive to look at a concrete example, inspired by [CT06]. Suppose we have a binary random variable X, defined as follows:

$$X = \begin{cases} 1 & \text{with probability } p, \\ 0, & \text{with probability } 1 - p. \end{cases}$$

Essentially, this random variable encodes a coin toss, where the probability of flipping heads is p and the probability of flipping tails is $1-p$. If $p = 0.5$, the coin is considered to be unbiased, otherwise it is called biased.

Using equation 1, it is straightforward to calculate the entropy of X, given some specific value of p. Figure 1 graphs the value of $H(X)$ against every possible $p \in [0, 1]$. When $p \in \{0, 1\}$, then the outcome of the coin toss is completely deterministic, and there is no uncertainty in the outcome. Accordingly, the entropy is 0 at these points. Conversely, when the coin is fair, we are completely uncertain about the outcome, unable to favour neither heads or tails. Again, the mathematical definition and intuition agree, as the entropy is indeed at its maximum when the coin is fair.

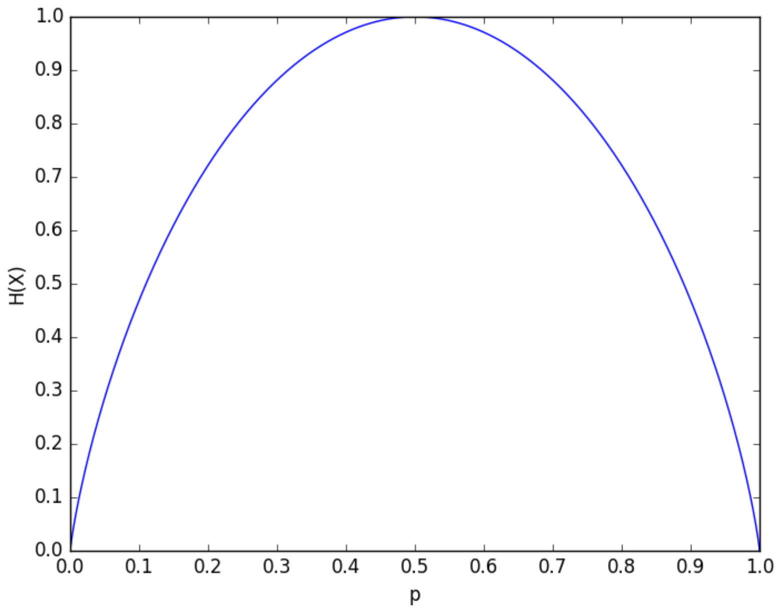

Figure 1: Entropy of X plotted against the value of p.

The fact that the entropy is maximal when the distribution is uniform is true in general. Another notable property of entropy is that it is non-negative.

2.1.2 Joint Entropy

Let X and Y be discrete random variables with realizations from the sets $\{x_1, x_2, ..., x_n\}$ and $\{y_1, y_2, ..., y_n\}$, respectively. Then the *joint entropy* [CT06] of the pair (X, Y) is defined as

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) log_2 p(x_i, y_j).$$

2.1.3 Conditional entropy

2.1.4 Kullback-Leibler distance

2.1.5 Mutual information

Given two discrete random variables X and Y, the *mutual information* [CT06] between them is given by

$$MI(X;Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}.$$

2.1.6 Conditional mutual information

The *conditional mutual information* [CT06] of discrete random variables X and Y given Z is defined by

$$MI(X;Y|Z) = H(X|Z) - H(X|Y,Z) = \mathbb{E}_p(x,y,z) log_2 \frac{p(X,Y|Z)}{p(X|Z)p(Y|Z)}.$$

2.2 Partial information decomposition

2.3 Numerical estimator

3 Elementary cellular automata

3.1 Problem description

3.2 Related work

3.3 Experimental setup

3.4 Results

3.5 Discussion

4 Ising model

4.1 Problem description

4.2 Related work

4.3 Experimental setup

4.4 Results

4.5 Discussion

5 Neural networks

5.1 Problem description

5.2 Related work

5.3 Experimental setup

5.4 Results

5.5 Discussion

6 Conclusion

References

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

Non-exclusive licence to reproduce thesis and make thesis public

I, Sten Sootla (date of birth: 17th of January 1995),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1 reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2 make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

Analysing information distribution in complex systems

supervised by Raul Vicente Zafra and Dirk Oliver Theis

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, dd.mm.yyyy