Information Theory and Coding Techniques: Chapter 1.1

What is Information Theory?

Why you should take this course?

What is “Information Theory” ?

“Information Theory” answers two fundamental questions in communication theory：

- What is the ultimate data compression (the entropy *H*)
- What is the ultimate transmission rate of communication (the channel capacity Ｃ)

It founds the most basic theoretical foundations of communication theory.

Moreover, “Information Theory” intersects

- Physics (Statistical Mechanics)
- Mathematics (Probability Theory)
- Electrical Engineering (Communication Theory)
- Computer Science (Algorithm Complexity)
- Economics (Portfolio / Game Theory)

This is why you should learn “Information Theory”.

Electrical Engineering (Communication Theory)

In the early 1940s, Shannon proved that

the error probability of transmission error could be made nearly zero for all communication rates below “Channel Capacity”.

Source : H — channel : C — Destination

$H<C$

The Capacity, C, can be computed simply from the noise characteristics (described by conditional probabilities) of the channel.

Shannon further argued that random processes (signals) such as music and speech have an irreducible complexity below which the signal cannot be compressed.

This he named the “Entropy”.

Shannon argued that if the entropy of the source is less than the Capacity of the channel, asymptotically (in probabilistic sense) error-free communication can be achieved.

Computer Science (Kolmogorov Complexity)

Kolmogorov, Chaitin, and Solomonoff put the idea that the complexity of a string of data can be defined by the length of the shortest binary computer program for computing the string.

The “Complexity” is the “Minimum description length” !

This definition of complexity is universal, that is, computer independent, and is of fundamental importance.

“Kolmogorov Complexity” lays the foundation for the theory of “descriptive complexity”.

Gratifyingly, the Kolmogorov complexity K is approximately equal to the Shannon entropy H if the sequence is drawn at random from a distribution that has entropy H.

Kolmogorov complexity is considered to be more fundamental than Shannon entropy. It is the ultimate data compression and leads to a logically consistent procedure for inference. (For example, if one can descript a picture by a program with length L, then L and the associated program can be treated as a more meaningful representation of the picture than that of the bit count does.)

One can think about computational complexity (time complexity) and Kolmogorov complexity (program length or descriptive complexity) as two axes corresponding to program running time and program length. Kolmogorov complexity focuses on minimizing along the second axis, and computational complexity focuses on minimizing along the first axis.

Little work has been done on the simultaneous minimization of the two.

Mathematics (Probability Theory and Statistics)

The fundamental quantities of Information Theory – Entropy, Relative Entropy, and Mutual Information – are defined as functionals of probability distributions.

In turn, they characterize the behavior of long sequences of random variables and allow us to estimate the probabilities of rare events and to find the best error exponent in hypothesis tests.

Computation vs. Communication.

As we build larger Computers out of smaller components, we encounter both a computation limit and a communication limit. Computation is communication limited and communication is computation limited. These become intertwined, and thus all of the developments in communication theory via information theory should have a direct impact on the theory of computation.

Real examples for the above “dual-limit” effects can be found in VLSI, Multi-Core Processor, and Cloud Computing fields.

New Trends in Information Theory.

- Compress each of many sources and then put the compressed descriptions together into a joint reconstruction of the sources — Wyner-Ziv and Slepian-Wolf theorems.
  Distributed Source Coding!

- If one has many senders sending information independently to a common receiver, what is the channel capacity of this "Multiple-Access channel"—Liao and Ahlswede theorem.

- If one has one sender and many receives and wishes to communicate (perhaps different) information simultaneously to each of the receiver, what is the channel capacity of this “Broadcasting channel”.

  ------- Scalable Video Coding!

- If one has arbitrary number of senders and receivers in an environment of interference and noise, what is the capacity region of achievable rates from the various senders to the receivers.

  -------- Networked Information Theory!