Information Theory: Basics and Applications

Presenter:
Dr. Ahmad El-Banna

OCTOBER 2016

Agenda

- Basic Concepts and Meaning
- Main Quantities of Information Theory
- Prerequisites
- Applications
- Trends and Research

About the Presenter

Dr. Ahmad EL-Banna

- B.Sc. in Telecommunications and Electronics, Fac. of Eng. at Shoubra, Benha Univ. 2005.
- 9-month Diploma in Embedded Systems, ITI, 2008.
- M.Sc. in Telecommunications and Electronics, Fac. of Eng. at Shoubra, Benha Univ. 2011.
- PhD. in Telecommunications and Electronics, E-JUST Univ., 2014.
- Visiting Researcher , Wireless Communications Lab, Osaka University, 2013-2014.
- Find more at
  - www.bu.edu.eg/staff/ahmad.elbanna

MEANING AND BASIC CONCEPTS

What is information?

- Let us consider some examples of sentences that contain some "information":
  1. The weather will be good tomorrow.
  2. The weather was bad last Sunday.
  3. The president will come to you tomorrow and will give you one million dollars.
- The second statement seems not very interesting as you might already know what the weather has been like last Sunday.
- The last statement is much more exciting than the first two and therefore seems to contain much more information.
- But on the other hand do you actually believe it?

What is information?..

Let us make some easier examples:

- You ask: “Is the temperature in Cairo currently above 30 degrees?”
- This question has only two possible answers: “yes” or “no”.
- You ask: “The president of Taiwan has spoken with a certain person from Hsinchu today. With whom?”
- Here, the question has about 400,000 possible answers (since Hsinchu has about 400,000 inhabitants).
- Obviously the second answer provides you with a much bigger amount of information than the first one.
- We conclude that:

The number of possible answers r should be linked to “information”

What is information?...

Let us have another example.

- You observe a gambler throwing a fair die.
- There are 6 possible outcomes { 1, 2, 3, 4, 5, 6 } .
- You note the outcome and then tell it to a friend.
- By doing so you give your friend a certain amount of information.
- Next you observe the gambler throwing the die three times.
- Again, you note the three outcomes and tell them to your friend.
- Obviously, the amount of information that you give to your friend this time is three times as much as the first time.
- We conclude that:

"Information" should be additive in some sense.

What is information?....

- Now we face a new problem:
- Regarding the example of the gambler before we see that in the first case we have $r = 6$ possible answers, while in the second case we have $r = 6^3 = 216$ possible answers.
- Hence in the second experiment there are 36 times more possible outcomes than in the first experiment.
- But we would like to have only a 3 times larger amount of information.
- So how do we solve this?
- A quite obvious idea is to use a logarithm.
- If we take the logarithm of the number of possible answers, then the exponent 3 will become a factor 3, exactly as we wish: $\log_b 63 = 3 \cdot \log_b 6$.
- Precisely these observations have been made by the researcher Ralph Hartley.

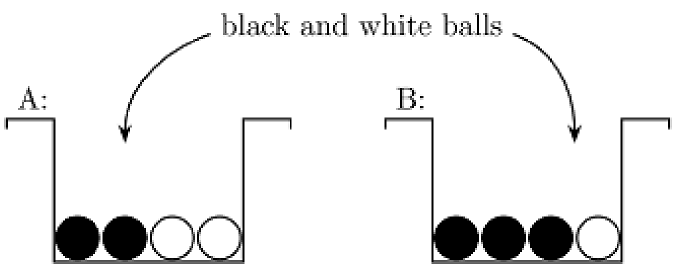
Measure of Information

- The first attempt (partially correct) was by Hartley in 1928 in Bell Labs by defining the following measure of information:

$$\tilde{I}(U) \triangleq \log_b r$$

where r is the number of all possible outcomes of a random message U & basis b of the logarithm is 2 (bit) or e (nat) or 10 (Hartley)

- But something was wrong or at least missed!
- Let's draw one ball at random from the two shown hats.



Measure of Information..

- In both hats we have r = 2 colors: black and white, i.e., $\tilde{I}(U) \triangleq \log_2 2 = 1\ bit$
- But obviously, we get less information if in hat B black shows up, since we somehow expect black to show up in the first place.
- Black is much more likely!
- And That's it ! We can now see that:

A proper measure of information needs to take into account the probabilities of the various possible events.

- This has been observed for the first time by **Claude Elwood Shannon** in 1948 in his landmark paper: "*A Mathematical Theory of Communication*"

Measure of Information…

- Shannon's measure of information is an "average Hartley information":

$$\sum_{i=1}^{r} p_i \log_2 \frac{1}{p_i} = -\sum_{i=1}^{r} p_i \log_2 p_i$$

where p_i denotes the probability of the i^{th} possible outcome.

Shannon the father of the information age !

- Before 1948, the engineering community was mainly interested in the behavior of a sinusoidal waveform that is passed through a communication system.

- Shannon, however, asked why we want to transmit a deterministic sinusoidal signal.

- Shannon had the fundamental insight that we need to consider random messages rather than deterministic messages whenever we deal with information.

- He is considered the inventor of the information theory.

Shannon the father of the digital age !

- Besides the amazing accomplishment of inventing information theory, at the age of 21.
- Shannon also "invented" the computer in his Master thesis!
- He proved that electrical circuits can be used to perform logical and mathematical operations, which was the foundation of digital computer and digital circuit theory.
- It is probably the most important Master thesis of the 20th century!
- Incredible, isn't it?

Why do we need to know information theory?

- First, who are we?
- “We” means the Telecommunication Engineers and Researchers.

- Simply the main purpose of any communication system is to properly and efficiently transfer one form of information from one side to another side at somewhere else.

- Therefore, our product is the information!

MAIN QUANTITIES OF INFORMATION THEORY

Uncertainty or Entropy

- It formally defines the Shannon measure of "self-information of a source".
- The uncertainty or entropy of a discrete random variable(RV) U that takes value in the set $\mathcal{U}$ (also called alphabet $\mathcal{U}$) is defined as

$$\mathrm{H}(U) \triangleq - \sum_{u \in \mathrm{supp}(P_U)} P_U(u) \log_b P_U(u)$$

- where $P_U(\cdot)$ denotes the probability mass function (PMF) of the RV U, and
- where the support of P_U is defined as

$$\mathrm{supp}(P_U) \triangleq \{u \in \mathcal{U}\colon P_U(u) > 0\}.$$

Entropy..

- Another, more mathematical, but often very convenient form to write the entropy is by means of expectation:

$$\mathrm{H}(U) = \mathsf{E}_U[-\log_b P_U(U)]$$

- Be careful about the two capital U: one denotes the name of the PMF, the other is the RV that is averaged over.

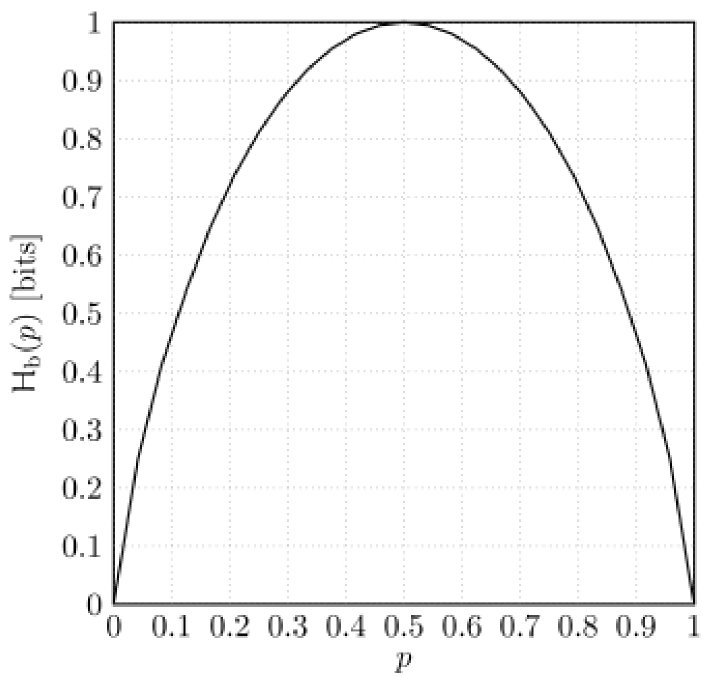
Binary Entropy Function

- If U is binary with two possible values u_1 and u_2, $\mathcal{U} = \{u_1, u_2\}$, such that $\Pr[U = u_1] = p$ and $\Pr[U = u_2] = 1 - p$, then

$$\mathrm{H}(U) = \mathrm{H}_\mathrm{b}(p)$$

- where $\mathrm{H}_\mathrm{b}(\cdot)$ is called the binary entropy function and is defined as

$$\mathrm{H}_\mathrm{b}(p) \triangleq -p \log_2 p - (1-p)\log_2(1-p), \quad p \in [0,1]$$



Conditional Entropy

- Similar to probability of random vectors, there is nothing really new about conditional probabilities given that a particular event Y = y has occurred.
- The conditional entropy or conditional uncertainty of the RV X given the event Y = y is defined as

$$
\begin{aligned}
\mathsf{H}(X|Y=y) &\triangleq -\sum_{x \in \operatorname{supp}(P_{X|Y}(\cdot|y))} P_{X|Y}(x|y) \log P_{X|Y}(x|y) \\
&= \mathsf{E}\left[-\log P_{X|Y}(X|Y) \,\middle|\, Y=y\right].
\end{aligned}
$$

Note that the definition is identical to before apart from that everything is conditioned on the event Y = y.

Conditioning Reduces Entropy

- For any two discrete RVs X and Y ,

$$\mathrm{H}(X|Y) \leq \mathrm{H}(X)$$

- with equality if, and only if, X and Y are statistically independent, $X \perp\!\!\!\perp Y$.
- Attention!
- The conditioning reduces entropy rule only applies to random variables, not to events! In particular,

$$\mathrm{H}(X|Y = y) \lesseqgtr \mathrm{H}(X)$$

Mutual Information

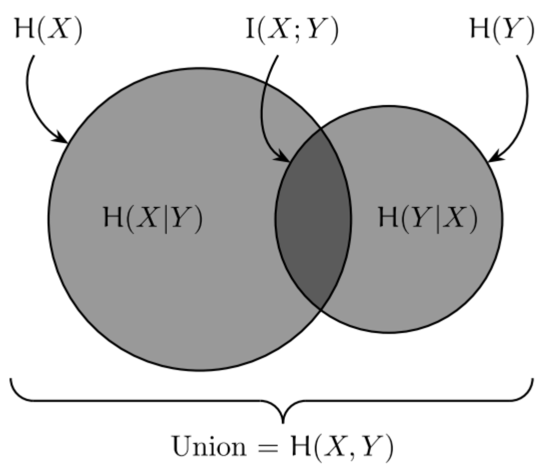
- Finally, we come to the definition of information.
- The following definition is very intuitive:
- Suppose you have a RV X with a certain uncertainty H(X).
- The amount that another related RV Y can tell you about X is the information that Y gives you about X.
- How to measure it?
- Well, compare the uncertainty of X before and after you know Y .
- The difference is what you have learned!
- And that’s the mutual information.

Mutual Information..

- The mutual information between the discrete RVs X and Y is given by $$\mathrm{I}(X;Y) \triangleq \mathrm{H}(X) - \mathrm{H}(X|Y)$$
- Note that:

1. H(X | Y) is the uncertainty about X when knowing Y .
2. It is a mutual information, not an "information about X provided by Y "! $$\mathrm{I}(X;Y) = \mathrm{I}(Y;X)$$

- A diagram depicting mutual information and entropy in a set-theory way of thinking looks like:



PREREQUISITES

Prerequisites of IT

To study the IT theory, you should at least have basic knowledge in:

- Probability Theory
- Linear Algebra
- Communication systems

APPLICATIONS

IT study field

- Information theory studies the
  - Quantification
  - Storage
  - and communication of information.
- The field is at the intersection of
  - Mathematics
  - Statistics
  - Computer science
  - Physics
  - Neurobiology
  - and electrical engineering

IT Applications

Therefore, this theory has found applications in many areas beside the communication field, including:

- Statistical inference
  - The process of deducing properties of an underlying distribution by analysis of data.
- Natural language processing
  - A field of computer science, artificial intelligence and computational linguistics concerned with the interactions between computers and human (natural) languages.
- Cryptography
  - The practice and study of techniques for secure communication in the presence of third parties called adversaries.
- Neurobiology
  - The scientific study of nervous systems.
- Thermal physics
  - The combined study of thermodynamics, statistical mechanics and kinetic theory.

IT Applications..

- Quantum computing
  - Which studies theoretical computation systems (quantum computers) that make direct use of quantum mechanical phenomena such as superposition and entanglement to perform operations on data.
- Plagiarism detection
  - the process of locating instances of plagiarism within a work or document.
- Pattern recognition
  - A branch of machine learning that focuses on the recognition of patterns and regularities in data.
- Anomaly detection
  - The identification of items, events or observations which don't conform to an expected pattern or other items in a data set.
- And many others…..

Fundamental topics applications

Applications of the fundamental topics of IT include:

- Lossless data compression
  - Algorithms that allow the original data to be perfectly reconstructed from the compressed data.
  - e.g. ZIP files.
- Lossy data compression
  - Algorithms that permit reconstruction only of an approximation of the original data but improves compression rate.
  - e.g. MP3 and JPEGs.
- Channel coding
  - Concerns with finding explicit methods, called codes, for reducing the error rate of data communication over noisy channels to near the channel capacity.
  - e.g. DSL.

IT impacts

IT impacts has been crucial to the success of

- The voyager missions to deep space
- The invention of the compact disk
- The feasibility of mobile phones
- The development of the internet
- The study of linguistics and human perception
- The understanding of black holes
- And numerous other fields…

TRENDS AND RESEARCH

Trends & Research

Important subfields of IT include:

- Coding theory
  - Source coding (data compression)
  - Channel coding (error correction)
- Algorithmic complexity theory
  - Measures of computational resources needed to specify the object.
- Algorithmic information theory
  - Concerns the relation between computation and information.
- Information-theoretic security
  - A cryptosystem with its security derived purely from IT.
- Measures of information

Main References

- Information Theory, Lecture Notes by Stefan M. Moser, ETH Z¨urich, 2014.
- Information Theory, Course lectures by Prof. Muriel Médard, MIT Univ., 2010.
- Lecture Notes in *Coding and Information Theory* — based on the book by Richard Hamming, Haverford College.
- The giant wiki.

Thank you !

For further inquires, send to: ahmad.elbanna@feng.bu.edu.eg