Publish AI, ML & data-science insights to a global community of data professionals.

Submit an Article

LATEST   EDITOR'S PICKS   DEEP DIVES   NEWSLETTER   WRITE FOR TDS
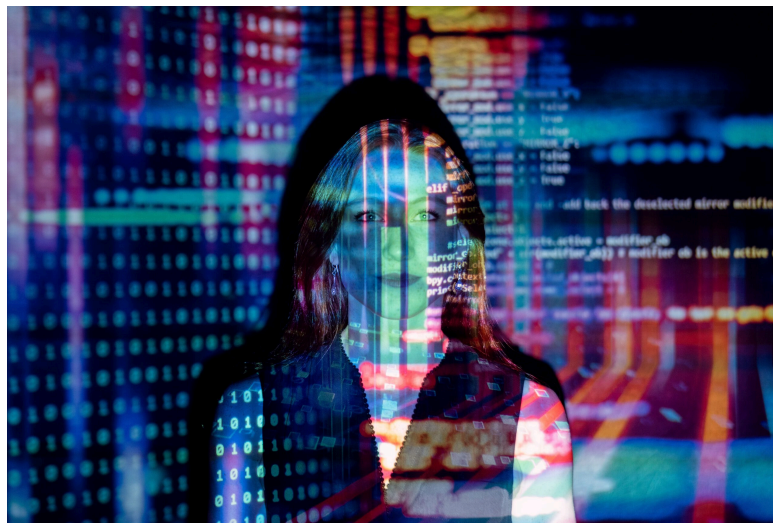
MACHINE LEARNING

Information Theory – A Introduction

At The Heart Of Data & AI

Ted Gross
Apr 15, 2022   10 min read

Understanding What Lies Behind Data & AI

Photo by ThisisEngineering RAEng on Unsplash

When I give seminars on AI or deliver papers at symposiums, I always try to lead with a few minutes on theory. Information Theory, Bayes Theorem, Linguistics, Bias, Chaos, Complexity, Innovation, Disruption, and a host of others. I do this because I have found a severe gap in what many have listed as their title in high-tech positions and what the knowledge that comes along with that title should entail.

It seems that all CVs which cross my desk for programming positions have the same buzzwords attached. AI, Big Data, and Machine Learning (ML) are among the favorites. While it may be true that these highly talented individuals do know the AI programming stack and are 'ninjas' in python and algorithms, it is painfully apparent that most do not have sufficient knowledge of the theories and constructs that lay behind what they are doing.

Lacking that knowledge creates a creativity gap. And in AI, coupled with huge amounts of data and predictive analytics, where there are so many factors to consider, the inability to exhibit creative thinking can be a huge hindrance.

I do hope the following short articles on various subjects and theories which lay at the heart of AI, help to elucidate and educate. One should not consider them by any means comprehensive, but rather a jumping-off point to delve further into the area of interest.

Boolean Functions – The Seed Tl Slowly Grew

Photo by Robert Anasch on Unsplash

It is often the case that significant discoveries and innovations require a seed from which to develop. This seed, in and of itself, is often of immense importance. Such is the case with information theory,[1] the significance of which is impossible to quantify. Digital information in any form would simply not exist were it not for information theory.

It began in 1854 with George Boole's paper on algebraic logic, _'An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities.'_[2] Boole's algebraic and logical notions are known today as a 'Boolean function'[3] and permeate our thought processes from an early age. Computer programmers are entirely reliant upon the Boolean logical operators, and without such propositions represented in code, it would prove impossible to develop any level of programming sophistication.

> 'Boole revolutionized logic by finding ways to express logical statements using symbols and equations. He gave true propositions the value 1 and false propositions a 0. A set of basic logical operations – such as and, or, not, either/or, and if/then – could then be performed using these propositions, just as if they were math equations.'[4]

It took almost 100 years for the beguiling simplicity of the and-or and true-false proposition to wield significant changes. Two choices existed: true and false. A linear progression began based upon an and-or proposition. *"In mathematics, a Boolean function*

is a function whose arguments and result assume values from a two-element set (usually {true, false}, {0,1} or {-1,1})."[5]

There are two doors in the above picture. Red and Yellow. Each has a door handle on the opposite side from the other. yet they are in all ways, except for color, identical. Think of them as switches. If one opens, and the other stays closed, [...] signify an OR situation. If both open that may signif[...] situation. Progressing through an endless series of [...] doors each pair either being opened or closed or or[...] open while the other closed, can approximate a Boo[...]

However, when the universe of true-false opened it[...] programming and-or constructs, through the mind [...] Shannon, the history of information and data chang[...]
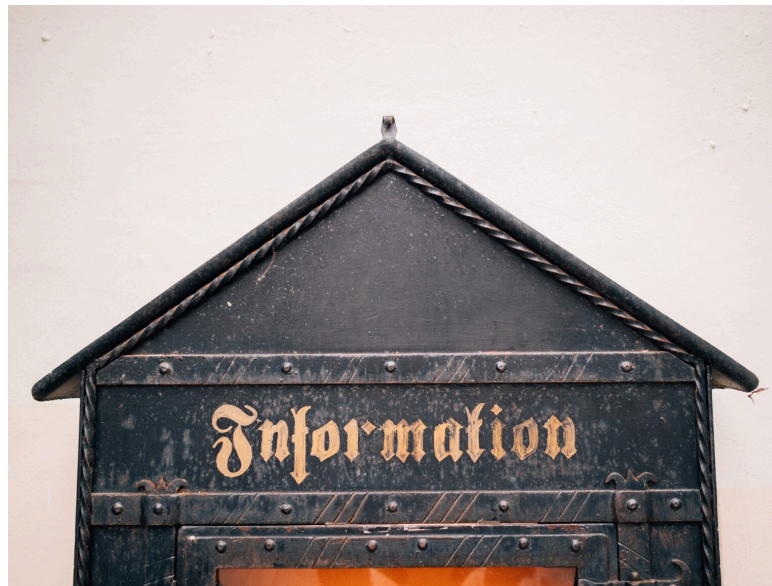
The Heart Of Information Theory

Photo by Roman Kraft on Unsplash

From the time Boole presented his mathematical proposition, the seed of information theory germinated for almost a century. In 1948, at Bell Laboratories, Claude Shannon, published a paper with an immeasurable impact on modern technology and data analysis. In *'A mathematical theory of communication',*[6] now commonly known as 'the Magna Carta of the Information Age', Shannon introduced the mind-boggling notion that information can be quantified and measured. He applied Boolean logic to a whole new cosmos while adding his personal touch of genius.

> 'Shannon figured out that electrical circuits could execute these logical operations using an arrangement of on-off switches. To perform an and function, for example, two switches could be put in sequence, so that both had to be on for electricity to flow. To perform an or function, the switches could be in parallel so that electricity would flow if either of them was on. Slightly more versatile switches called logic gates could streamline the process.'[7]

'Before Shannon's paper, information had been viewed as a kind of poorly defined miasmic fluid. But after Shannon's paper, it became apparent that information is a well-defined and, above all, measurable quantity...

... Shannon's theory of information provides a mathematical definition of information, and describes precisely [how] information can be communicated between diffe[rent parts] of a system. This may not sound like much, but S[hannon's] theory underpins our understanding of how signa[ls] are related, and why there are definite limits to t[he rate at] which information can be communicated within a[system], whether man-made or biological.'[8]

'The resulting units', wrote Shannon, 'may be call[ed binary] digits, or more briefly, bits'.[9]

'The bit now joined the inch, the pound, the quar[t, and the] minute as a determinate quantity – a fundament[al] measure. But measuring what? "A unit for measuring information", Shannon wrote, as though there were such a thing, measurable and quantifiable, as information.'[10]

Until this time, no one had ever presumed that information could be subject to a mathematical formula or computational analysis. With the publication of this seminal paper, there was the world *before* Shannon and the world *after* Shannon.

Perhaps Shannon's greatest achievement was the counter-intuitive approach to analyse 'information' independent of 'meaning'. In short, when dealing with information, one does not need to consider the meaning of the message. Indeed, 'meaning' is superfluous to that actual content. 'Meaning' is, in effect, meaningless.

As he wrote in the second introductory paragraph of 'A mathematical theory of communication':

'The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. **Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem.** The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.'[11]

What is of concern are probability and uncertainty. With the birth of the bit, Shannon took the 0,1->true/false construct of Boolean functions to a whole new stratosphere. At the heart of Shannon's theory lie 'noise' and 'surprise'. All communication – human, computer, over wires, signals, digital – as a universal fundamental – has an element of 'noise' and an element of 'surprise'. The surprise is what is left after the noise is eliminated (and there is

always noise on any channel of communication) without interfering with the original message.

> 'So, what is information? It is what remains after every iota of natural redundancy has been squeezed out of a message, and after every aimless syllable of noise has been removed. It is the unfettered essence that passes from comput[er to] computer, from satellite to Earth, from eye to bra[in, and (over many generations of natural selection) from the [...] to the collective gene pool of every species.'[12]

Shannon's information theory gave practical birth to [...] age. Without it, people would be drowning in noise [...] uncertainty regarding the veracity of the messages [...] Without it, all modes of communication would be l[...] with garbled information and incoherent meanings.

> Paradoxically, by ignoring the meaning of a messa[ge...] showing how insignificant 'meaning' is to an actu[al...] Shannon gave the world true meaning and the ability to handle massive amounts of data securely and coherently.

Simply put, information theory is fundamental to everything.

> 'But before Shannon, there was precious little sense of information as an idea, a measurable quantity, an object fitted out for hard science. Before Shannon, information was a telegram, a photograph, a paragraph, a song. After Shannon, information was entirely abstracted into bits. **The sender no longer mattered, the intent no longer mattered, the medium no longer mattered, not even the meaning mattered**: a phone conversation, a snatch of Morse telegraphy, a page from a detective story were all brought under a common code.'[13]

In his paper, and for the rest of his life, Shannon introduced the world to a whole new series of concepts:

1. He named a piece of information – binary units – a "bit."
2. He showed us that meaning had nothing to do with information. Indeed, the more we paid attention to meaning, the louder the 'noise in the message' became.
3. By ignoring noise, he was able to produce a universal method of transferring and deciphering information.
4. Finally, Shannon left us the legacy of information noise with one almost undecipherable term. This idea would also rock the foundations of science, technology, and the digital age and would leave its footprint on all that came afterward. He called it 'information entropy.'

In the next article in this series, I hope to demystify Entropy and 'Information Entropy.' Though, as anyone who has studied Entropy, the second law of thermodynamics will tell you, it is impossible to pin down one definition. Yet entropy plays a significant role in Chaos and Complexity theory, which lay at the heart of AI.

About the Author:

Ted Gross is Co-Founder & CEO of "If-What-If." Ted served as a CTO & VP of R&D for many years with expertise in database technology concentrating on NoSQL systems, NodeJS, MongoDB, Encryption, AI, Disruption, Chaos & Complexity Theory, and Singularity events. He has over 15 years of expertise [in] World Technologies & 6 years in Augmented Reality. [He has gone on] to write many articles on technological topics in pr[ofessional] academic journals and online on the Facebook If-W[hat-If,] Medium, and LinkedIn. You can also sign up for the [free] newsletter of If-What-If here or on Substack.

References:

1. Wikipedia (n.d.) 'Information theory', available a[t:] https://en.wikipedia.org/wiki/Information_theory [(accessed] 29th July, 2021).

2. Wikipedia (n.d.) 'The laws of thought', available at: https://en.wikipedia.org/wiki/The_Laws_of_Thought (accessed 19th August, 2021).

3. Wikipedia (n.d.) 'Boolean function', available at: https://en.wikipedia.org/wiki/Boolean_function (accessed 19th August, 2021).

4. Isaacson, W. (2014) 'The Innovators: How a Group of Hackers, Geniuses, and Geeks Created the Digital Revolution', Simon & Schuster, New York, NY, Kindle Edition, Location 943.

5. Wikipedia (n.d.) 'Boolean function', available at: https://en.wikipedia.org/wiki/Boolean_function (accessed 19th August, 2021).

6. Shannon, C. (1948) 'A mathematical theory of communication', *Bell System Technical Journal*, Vol. 27, July/October, pp. 379–423

7. Isaacson, ref. 4 above, Location 943

8. Stone, J.V. (2018) 'Information Theory: A Tutorial Introduction', Sebtel Press, Kindle Edition, Location 82

9. Shannon, ref. 6 above.

10. Gleick, J. (2011) 'The Information', Pantheon Books, New York, NY, Kindle Edition, Location 66.

11. Shannon, ref. 6 above.

12. Stone, ref 8 above, Location 359.

13. Soni, J. and Goodman, R. (2017) 'A Mind at Play: How Claude Shannon Invented the Information Age', Simon & Schuster, New York, NY, Kindle Edition, Location 69.

*In keeping with Medium's disclosure policy, all links to Amazon books listed above are affiliate links for If-What-If.

· · ·

WRITTEN BY

Ted Gross

See all from Ted Gross

AI   Big Data   Information Theory   Machine Learn

Twg144

Share This Article   (Facebook)   (LinkedIn)   (X)

**Towards Data Science is a community publicat
your insights to reach our global audience and c
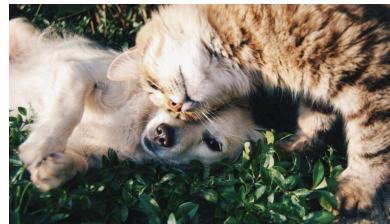the TDS Author Payment Program.**

Write for TDS

Related Articles

ARTIFICIAL INTELLIGENCE

Implementing Convolutional Neural Networks in TensorFlow

Step-by-step code guide to building a Convolutional Neural Network

Shreya Rao
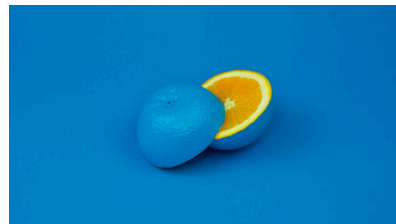August 20, 2024   6 min read

ARTIFICIAL INTELLIGENCE

How to Forecast Hierarchical Time Series

A beginner's guide to forecast reconciliation

Dr. Robert Kübler
August 20, 2024   13 min read

DATA SCIENCE

Hands-on Time Series Anomaly Detection using Autoencoders, with Python

Here's how to use Autoencoders to detect signals with anomalies in a few lines of...

Piero Paialunga
August 21, 2024   12 min read

MACHINE LEARNING

3 AI Use Cases (That Are Not a Chatbot)

Feature engineering, structuring unstructured data, and lead scoring

Shaw Talebi
August 21, 2024   7 min read

DATA SCIENCE

Back To Basics, Part Uno: Linear Regression and Cost Function

An illustrated guide on essential machine learning concepts
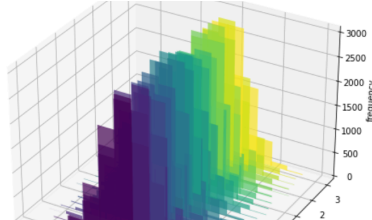
Shreya Rao

February 3, 2023   6 min read

DATA SCIENCE

Must-Know in The Bivariate N Projection Exp

Derivation and practic powerful concept

Luigi Battistoni

August 14, 2024   7 min

DATA SCIENCE

Our Columns

Columns on TDS are carefully curated collections of posts on a particular idea or category...

TDS Editors

November 14, 2020   4 min read

Recommended Articles

A Bird's-Eye View of Linear Algebra: Why Is Matrix Multiplication Like That?

Is RAG Dead? The Rise of Context Engineering and Semantic Layers for Agentic AI

MCP in Practice

The Math You Need to Pan and Tilt 360° Images