XVII Tonale Winter School in Cosmology

Working group: Data Inference and Machine Learning

by Emille E. O. Ishida

November 29, 2024

Let's build a thinking machine

The points below were taken from [1], in my personal and completely biased opinion, the best book on logic ever written.

Start from the beginning: plausible reasoning

- Question 1:

  > *Suppose some dark night a policeman walks down a street, apparently deserted. Suddenly he hears a burglar alarm, looks across the street, and sees a jewelry store with a broken window. Then a gentleman wearing a mask comes crawling out through the broken window, carrying a bag which turns out to be full of expensive jewelry. The policeman doesn't hesitate at all in deciding that this gentleman is dishonest.*

  - Do you agree with the policeman's conclusion?
  - Identify the elements which lead to the policeman's conclusion.
  - Can you think of alternative lines of thought that would lead to different judgements?

Deductive reasoning $\neq$ plausible reasoning

Deductive reasoning concerns conclusions unequivocally backed by data.

Plausible reasoning is the art of making decisions with incomplete, uncertain, messy information.

Deductive reasoning $\Longrightarrow$ strong syllogisms: (`A`): *if A is true, then B is true; if B is true, then A is true.*
Plausible reasoning $\Longrightarrow$ weak syllogisms: (`B`): *if A is true, then B is true; if B is true then A is more plausible.*

Exercise 1: List a few real life examples of situations where the strong (`A`) and weak (`B`) syllogisms can be used.

- Question 2:

  Consider these 2 statements:

  $A \equiv$ *it will start to rain by 10 AM at the latest*;
  $B \equiv$ *the sky will be cloudy before 10 AM.*

  What is the logical connection between them: $A \Longrightarrow B$ or $B \Longrightarrow A$?

Another weak silogism, (`C`): *if A is true, then B is true; if B is false than A is less plausible.*

Exercise 2: list a few real life examples of `C`.

Food for thought: what sylogism would describe the situation of the policeman?

Food for thought: Given the same description of the situation, what could lead the policeman to arrive in a different conclusion?

Goal: to build a robot equipped with common sense

Probability theory is nothing but common sense reduced to calculation.

Laplace, 1819

- Question 3:

  Given the recent technological development there has been a lot of questioning about the power of computers and how far they are able to go. In your opinion, what is the main conceptual obstacles (not hardware-related) in designing a computer which can think?

- Question 4:

  Consider that we are building a robot who follows the basic desiderata:

  I. *Degrees of plausibility are represented by real numbers.*

  II. *Qualitative correspondence with common sense.*

  III. (a) *If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result.*

  (b) *The robot always takes into account all of the evidence it has relevant to a question. It does not arbitrarily ignore some of the information, basing its conclusions only on what remains. In other words, the robot is completely nonideological.*

  (c) *The robot always represents equivalent states of knowledge by equivalent plausibility assignments. That is, if in two problems the robot's state of knowledge is the same (except perhaps for the labeling of the propositions), then it must assign the same plausibilities in both.*

  Our robot will make decisions base solely in degrees of plausibility, which by definition are real numbers. Humans, on the other hand, take into account many different aspect of a question when making a decision.

Food for thought: What class (es) of issues our robot will most closely mimic the behavior of a human?

- Question 5:

  Consider the following scenario is presented to our robot:

  $B \equiv$ *An urn contains N balls, identical in every respect except that they carry numbers (1, 2, . . . , N) and M of them are colored red, with the remaining* $(N - M)$ *white,* $0 \leq M \leq N$. *We draw a ball from the urn blindfolded, observe and record its color, lay it aside, and repeat the process until n balls have been drawn,* $0 \leq n \leq N$.

  $R_i \equiv$ *Red ball on the i-th draw.*
  $W_i \equiv$ *White ball on the i-th draw.*

Exercise 3: Suppose the experiment has not started yet. Before the first draw, can you estimate $P(R_1|B)$ and $P(W_1|b)$?

Food for thought: What does the answer from Exercise 3 tells you about the content of the urn?

Exercise 4: What is the probability of drawing a red ball in the second try, P $(R_2|B)$? In this example the robot knows that there has been one draw but it has no information about its outcome.

Food for thought: In finding the probability for red at the $k-th$ draw, knowledge of what color was found at some earlier draw is clearly relevant because an earlier draw affects the number M_k of red balls in the urn for the $k-th$ draw. Would knowledge of the color for a later draw be relevant?

- Question 6:

  The situation described in Question 5 can be identified as *sampling without replacement*. Consider now the situation of *sampling with replacement*. Meaning that every time we draw a ball from the urn we record its color and put it back before drawing again.

  – Which of the two scenarios do you consider more complex for estimating the probability of a given color in a given draw?
  – How can make the sampling with replacement simpler for this task?

- Question 7:

  In the theory we are developing, any probability assignment is necessarily 'subjective' in the sense that it describes only a state of knowledge.

Food for thought: whose state of knowledge?

Fact: It is possible to derive all commonly used statistical tools using boolean algebra.

It takes 40 pages to prove that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ... but it is possible.

...

This idea was first introduced in the *Organon*, of Aristotles, 400 BC.

...

It is also the basis upon which computers are built.

Thinking ahead: what can we realistically expect from machine learning?

This section is a summary of the ideas presented in [2].

Hypothesis: Imagine an oracle providing non-trivial predictions that are always true.

Food for thought: What the impact of such a system be in the scientific activity?

All these are based on non-trivial concepts which are difficult, if not impossible, to satisfactorily define in a finite amount of time. In order to move forward we will use the definition of [3]:

Criterion of understanding: A phenomenon P can be understood if a theory T of P exists that is intelligible.
Criterion for the intelligibility of theories: a scientific theory T is intelligible for scientists (in context C) if they can recognize qualitatively characteristic consequences of T without performing exact calculations.

Food for thought: Is scientific discovery possible without scientific understanding?

Exercise: List a few examples that illustrate your previous answer. These can be related to any scientific field (not restricted to astronomy).

Dimensions of computer assisted understanding: In [2], the authors define 3 dimensions in which computers can be a tool to sparkle scientific understanding (see Figure 1):

1. Descriptive analysis of complex data (computational microscope);
2. Resource inspiration;

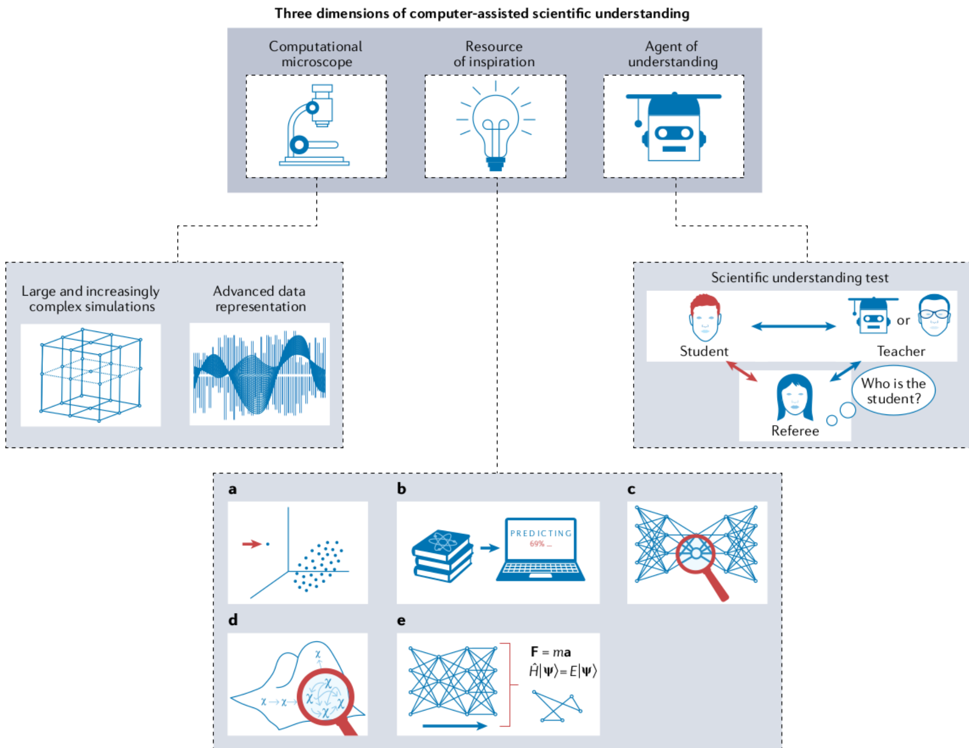

Figure 1: Aspects of computer assisted understanding. Figure from [2].

3. Agent of understanding (when knowledge is transfered to humans).

Exercise: list examples (in any science domain) in which the machine learning contribution can be identified can be described by each of the items above.

Food for thought: Suppose we have a machine which fulfills the third dimension of understanding. It is capable of learning, and transmitting its reasoning to a human agent. Do you agree with the authors that this would consist on an example of scientific understanding? What is the role played by the machine in this scenario: does it understand or does it sparkle understanding?

References

[1] Jaynes, E. (2003). *Probability Theory: The Logic of Science* (G. Bretthorst, Ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511790423

[2] Krenn *et al.* (2022), *On scientific understanding with artificial intelligence*, Nature Reviews Physics, 4, 761–769. doi:10.1038/s42254-022-00518-3

[3] De Regt, H. W. & Dieks, D. A (2005), *A contextual approach to scientific understanding.* Synthese 144, 137–170