Marcus Hutter - Can Intelligence Explode? (2012)

Tom Rochette <tom.rochette@coreteks.org>

November 2, 2024 — 36c8eb68

0.1 Context

0.2 Learned in this study

0.3 Things to explore

- Intelligence in humans increases over time by the fact we can teach, and thus have young humans know earlier than we did. Does it mean that we will be limited by how fast teaching can happen?
- It might be impossible to detect superintelligence on others planets due to the fact that increased intelligence => compressed information => indistinguishable from random noise
  - On the other hand, compressed information can be decompressed, thus it might be possible to attempt to decompress noise streams

1 Overview

2 Notes

2.1 1 Introduction

- There are many different potential paths toward a singularity:
  - Mind uploading
  - Knowledge-based reasoning and planning software
  - Artificial agents that learn from experience
  - Self-evolving intelligent systems
  - Awakening of the Internet
- Chalmers cleverly circumvents a proper discussion or definition of intelligence by arguing
  - there is something like intelligence
  - there are many cognitive capacities correlated with intelligence
  - these capacities might explode
  - intelligence might amplify or explode

2.2 2 Will there be a Singularity

- Different estimates on the computational capacity of a human brain consistently point towards $10^{15} \ldots 10^{16}$ flops/s
- If computational speeds double every two years, what happens when computer-based AIs are doing the research?
  - Computing speed doubles every two years
  - Computing speed doubles every two years of work
  - Computing speed doubles every two subjective years of work
  - Two years after Artificial Intelligence reach human equivalence, their speed doubles. One year later, their speed doubles again. Six months, 3 months, 1.5 months ... Singularity.

2.3 3 The Singularity from the Outside

- Insiders will produce improved computers ad infinitum at an accelerated pace
- Insiders may attempt to communicate with outsiders at the maximal digestible rate
- After a brief period, intelligent interaction between insiders and outsiders becomes impossible. The inside process may from the outside resemble a black hole watched from a safe distance
- Outward explosion is likely to end or convert the outsiders' existence
- People now more and more explore virtual worlds rather than new real worlds
  - Virtual worlds can be designed as one sees fit and hence are arguably more interesting
  - Outward expansion now means deep sea or space, which is an expensive endeavor
- Expansion usually follows the way of least resistance
- Inward explosion will stop when computronium is reached. Outward explosion will stop when all accessible convertible matter has been used up

2.4 4 The Singularity from the Inside

- An intelligence explosion with fixed computation, even with algorithmic improvements seems implausible
- If their (vorld inhabitants) subjective thoughts processes will be sped up at the same rate as their surroundings, nothing would change for them
  - The only difference, provided virtuals have a window to the outside real world, would be that the outside world slows down
- There seems to be no clear positive correlation between the number of individuals involved in a decision process and the intelligence of its outcome

2.5 5 Speed versus Intelligence Explosion

- A speed explosion is not necessarily an intelligence explosion
- If only the environment is sped up, this has the same effect as slowing down the agent. He will receive more information per action, and can make more informed decisions, provided he is left with enough computation to process the information
- If the agent is sped up, this has the same effect as slowing down the environment. From the agent's view, he becomes deprived of information, but has now increased capacity to process and think about his observations
- More computation only leads to more intelligent decisions if the decision algorithm puts it to good use

2.6 6 What is Intelligence

- Genetic evolution has been largely replaced by memetic evolution, the replication, variation, selection, and spreading of ideas causing cultural evolution
- If rationality is reasoning towards a goal, then there is no intelligence without goals
- What are the goals?
  - Expected utility maximization
  - Cumulative life-time reward maximization
- Who sets the goal for super-intelligence and how?
- The successful virtuals will spread, the others perish, and soon their society will consist mainly of virtuals whose goal is to compete over resources, where hostility will only be limited if this is in the virtuals' best interest
- In such evolutionary vorlds, the ability to survive and replicate is a key trait of intelligence. On the other hand, this is not a sufficient characterization, since e.g. bacteria are quite successful in this endeavor too, but not very intelligent

2.7 7 Is Intelligence Unlimited or Bounded

- The theory (AIXI) suggests that there is a maximally intelligent agent, or in other words, that intelligence is upper bounded (and is actually lower bounded too). At face value, this would make an intelligence

explosion impossible

- In the tic-tac-toe world, it is possible to reach the upper bound in practice
- In the chess world, the optimal way of playing chess is minimax tree search to the end of the game. However, unlike tic-tac-toe, this strategy is computationally infeasible in our universe. So in theory intelligence is upper-bounded in a chess vorld, while in practice we can get only ever closer but never reach the bound
- This causes two potential obstacles for an intelligence explosion:
  - We are only talking about the speed of algorithms, which do not equate with intelligence
  - Intelligence is upper bounded by the theoretical optimal chess strategy, which makes an intelligence explosion difficult but not necessary impossible
- Intelligence measure Υ, upper bounded by $\Upsilon_{max} = \Upsilon(\text{AIXI})$
  - Since AIXI is incomputable, we can never reach intelligence Υ_{max} in a computational universe
  - It might be the case that in a highly sophisticated AIXI-closed society, one agent beating another agent by a tiny epsilon on the Υ-scale makes all the difference
- Where do humans range on the Υ-scale?

2.8 9 Diversity Explosion and the Value of a Virtual Life

- Consequences of virtual life being copied/modified
  - A virtuan explosion with life becoming much more diverse
  - Life becomes less valuable
  - Life may become a disposable

2.9 10 Personal Remarks

- Hutter believes in the functionalist theory of identity
- Uploading of a human mind preserves identity and consciousness, and indeed that any sufficiently high intelligence, whether real/biological/physical or virtual/silicon/software is conscious, and that consciousness survives changes of substrate: teleportation, duplication, virtualization/scanning, etc.

3 See also

4 References

- Hutter, Marcus. “Can intelligence explode?.” Journal of Consciousness Studies 19.1-2 (2012): 143-166.
- https://arxiv.org/abs/1202.6177