Summary of Moravec's Paradox

Aditi Bhattacharya

26 July 2021

The Moravec's paradox is an idea articulated by Hans Moravec among many others, which states that it is easier to design an artificial system which can perform high-level intelligent tasks, than to make the system perform tasks which generally don't involve cognitive skills when performed by humans.

A thing to be noted here is the fact that humans, also tend to become pretty bad at these tasks once when they consciously try to perfect the task-in-hand. The intuition of something 'being a part of common sense' seems similar in the entire human society. However, if one sits to justify or deduce how a piece of information becomes a part of common-sense, s/he might end up questioning the very basics on which the entire universe works. Take the famous example of the summation of one and one, which gives a result of two. Sounds quite clear, however, there's a very famous 360 pages prove of this 'common-sense' question.

A paper published recently claimed to have built a large dataset which lets an algorithm to work with intuitive psychology. Intuitive psychology, the ability to reason about hidden mental variables that drive observable actions, comes naturally to people: even pre-verbal infants can tell agents from objects, expecting agents to act efficiently to achieve goals given constraints. They propose AGENT, a benchmark for core psychology reasoning, which consists of a large-scale dataset of cognitively inspired tasks designed to probe machine agents' understanding of key concepts of intuitive psychology in four scenarios – Goal Preferences, Action Efficiency, Unobserved Constraints, and Cost-Reward Trade-offs. The fact that it took us so long to build such a sophisticated dataset, very well verifies the Moravec's paradox.

References:

https://www.storyofmathematics.com/20th_russell.html#:~:text=Some%20idea%20of%20the%20scope,logic%20since%20Aristotle's%20%E2%80%9COrganon%E2%80%9D

Shu,Tianmin,etal."AGENT:ABenchmarkforCorePsychologicalReasoning. "arXivpreprintarXiv:2102.12321(2021).