

Think Fast, Think Slow

Joris LIMONIER

November 17, 2021

- AI lacks capabilities that humans have. See how humans can help and bring them in machine world
- Focus on human capabilities that machines do not have, exploit cognitive theories of human decision making, identify causes enabling such capabilities, [fourth point]
- Four books:
 - TFTS
 - Sapiens: historian point of view
 - Society of mind: how we interact collectively
 - theory of consciousness: how humans model the world
- System 1, 2 = TF, TS (respectively)
 - TF: automatic, fast thinking, error prone, unconscious bias, used when problem solved is easy, reactive mode, make quick decisions
 - TS: slow, deliberate decision making, effortful, requires all attention, not many parallel decisions, used when problem is cognitively difficult or we really care to solve it correctly (because many things depend on the problem). Sometimes override TF.
 - We usually go from system 2 (TS) to system 1 (TF). Example: with child, start with TS, then go to TF when they can read. But not always: arithmetic operations.
- TFTS inspires AI. AAAI 2020 panel, CERN SPARK podcast of Francesca.
- Different approach
 - Multi-agent architecture:
 - TF solvers: rely on past experience, don't look at characteristics of the problem / similar problems and what we know about the environment. Very fast, react (activate automatically) when new problems arise if their skills are relevant for the problem to be solved. Propose solution and assert confidence.
 - TS solvers: Don't rely too much on past experience, even though they have access to it. Computational complexity can be slow depending on size of input. Activate by meta-cognition.
 - Model solvers / updaters: Act in the background to update model used by agents to do their job.
- Metacognition: cognition about cognition. Improves the quality of the system's decision. Choice of author is to use a centralized meta-cognitive agent. Assert TF vs TS.
- S1 solvers do not wait to be asked to start solving a problem + give confidence. Model solvers assert quality of S1 agent and decide if activation of S2 is required. Model solvers assesses available resources, expected cost of using S2, expected reward for correct solution and solvers. 2 phases: quick assessment, then more reliable (longer one) if needed.
- Two metacognitive phases:
 - Goal: avoid using S2 when unnecessary (i.e. expected reward - expected cost > what S1 can provide)
 - 2 phases: 1) check if enough resources for S1 and S2, if not, choose TF

- Design choices:
 - S1 by default
 - S2 may not be better than S1
 - In more complex scenarios: there are several S1 and S2. MC need to choose among them
 - AI trustworthiness: take into consideration fairness, robustness, explainability, ...etc and MC must be explainable and logic-based.
- SOFAI vs neuro-symbolic
 - We do not assume that S1 are data-driven and S2 are logic-based.
- Human reasoning
 - Human deliberation (Multi dimensional theory: MDFT)
 - Results: learn the model of the world from human demonstrations, comparison of MDFT and RL
- MDFT: model how people make decisions about a set of options by choosing an option based on their preference and accumulation depending on discount factors and option similarity. Then stop criteria (number of iterations (ie time) or satisfaction with solution). Stopping times are used (with time and upper bound)
- MDFT vs RL: Machine is given a state and has to build a path towards a goal. There are penalties for violating constraints, for going for state where the agent is not supposed to go.
- S1 uses probability distribution based on history. S2 uses MDFT. MC decides between move proposed by S1 with confidence level and activating and MDFT (S2).
- Initially, system has no S1 and only S2, then at some point there is an option to use S1.

Cem's notes Thinking fast and slow AI
 Francesca Rossi IBM research

How human make decision to advance ai capability.

Currently AI lack of these slide-2

How humans can this, what causes these abilities,

Slide-3

A machine is compliantly different platform

Slide-4

Kahneman= human decide, think fast or slow, how combine each other to make decisions

Harari=Tries to understand how humans got this decision making over evolutionary. How we got here in decision making. How evolve?

Minsky= Focus on the idea our mind is not single box as collection of many specialities work together.

Graziano= How humans model world, aware in their... Metacognition?

Slide-5

System1= fast modality in terms of time. Almost automatic decision everyday life. Error prone, embeds unconscious bias.

Cognitively easy, we don't need to pay too much attention to decide

System-2

Slow?= Effortful, requires attention, we use it when the problem is cognitively difficult. We really care when we care to solve it correctly.

System2->SYSTEM1: When we don't know problem we use system2 then we get use to it and go system1.

For ex: Child reads letter then without effort read book.

Slide-6 These theory ..

Slide-7

We use multi agent architecture where we have several agents. Some of them solve problems, some use system1 some system2. Some update the world. What is solver->

System1 solver-> They just rely on past experience /similar problems. They're working constant time (very fast). They activate themselves if they think problem is relevant their skills
System-2 solver-> They cannot start working unless meta cognitive agent activate them.
System updater-> update

Slide-8

Any cog process that is reason of another meta cognition?

Main goal-> to improve the quality of decision. To assess the ...

Slide-9

System-1 solver does not wait meta cognitive agent. They start moving when problem come

Doesnot trust system1 solution so it activate system2 solver (meta cog agent)

Meta cog agent-> 1. P...

Slide-10

THE SOFAI architecture

Meta cog agent-> choose system1 solution or activate system2 solver.

Slide11

Avoid activate system2 solver because it is more costly.

Pros executed reward

Cons:

Slide12

S1 agent start immediately like human. If S1 didn't start and propose to meta agent

There is no assumption that $s1 > s2$ or $s2 > s1$.

Perception task generally $s1 > s2$

In this environment

Slide13

Slide-14

Project

Slide-15

Scnerio 1- human resoning.

Whether this system behaves better

Slide-16

MDFT

Slide17

MDFT example: 3 options, 2attributes

$w=[0.45, 0.55]$

Slide18

MDFT vs RL

Slide19

meta agent, setting weights In a way tries

... time is out.

This architecture can create ...

In begging use s2, and at some point switch s1 because it get use to solve problems.

S2 helps exploration