Collective Intelligence in Latent Imagination

Amit Kumar Singh (amitkumarsingh5@acm.org)
Cognition Researcher, Active Inference Lab
California, USA

Abstract

Intelligence is fundamentally the ability for an agent to infer causal dependencies in its environment. However, the precise conceptualization across systems and scales is a polemical question. The concept of "Intelligence" may as well refer to a quantitative measure of formal cognitive ability than to a qualitative property of skilled agency. This difficulty in understanding the concept compounds when we try to scale to descriptive and predictive models of collective behavior. While it is self-evident that groups may leverage pairwise interactions or their collective resources to tackle complex problems, is that process only the sum of individual intelligence or is the group intelligent in its own right? If the latter, what does it mean for the classical internalist conception of intelligence and agency? If the former, then what is the proper scale of analysis in systems of nested organization, such as human societies? This question can be approach rigorously through a non-reductive account of the physical processes underlying intelligence. Here I propose that the latent model framework(with active inference as intrinsic reward mechanism) framework is a promising approach that could live up to the multiple dimensions of adeptness required by any framework that would attempt to generalize cognition across scales. A statistical state model for mathematical state transitions can be built and can be used to further define cognitive model.

Keywords: Collective cognition; Mathematical modeling; Active Inference; Sate spaces; Latent spaces

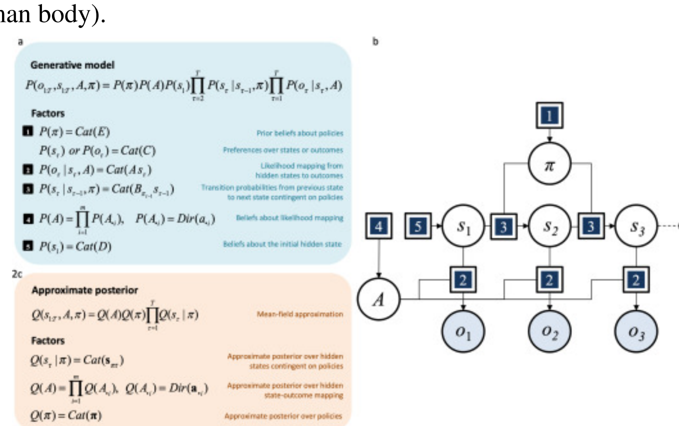
Introduction

The view of mind as an experience generating machine or a generative model goes further back than just recent Machine learning breakthroughs (like Variational Autoencoders). This paper takes a similar route. Starting out with explaining possible state-space configurations, a world-model (World Models, Ha and Schmidhuber) is mapped in the latent space. This could be enough framework to explain individual actions in an environment, just like it does with many model-based approaches to Reinforcement Learning. But, in contrast to most RL approaches where reward is extrinsic and task structure changes with it which works well in specific RL environments, where rewards are intrinsic in the complex environment itself (Reward is enough, Silver et. al) It fails to explains intelligence at the collective level where the agents apart from the environment, have formed a dynamic between themselves too. This dynamic is represented through a collective latent which can be traversed in an abstract space by the agents of the collective for inference to eventually reach cognition as a collective.

Towards a Free Energy Agent Cognitive Model

Agent's configuration at an instant is defined by its state with parameters interacting to form state variables. To establish a stable ground I invoke the Free-Energy principle. While we can argue about all derivations of intelligence, we can come to standstill that the system exists. This is the basic formulation of the Free Energy principle and everything is deduced from this assumption with agent and environment in the frame. The Active inference principle can be framed as the minimisation of surprise (Friston, 2009) by perception and action. Here, in discrete state models - agents select from different possible courses of action (i.e., policies and their gradient of preferences) in order to realise the preferences and thus minimise the surprise that they expect to encounter in the future. This enables a Bayesian formulation of the perception–action cycle (Fuster, 1990): agents perceive the world by minimising variational free energy, ensuring their model is consistent with past observations, and act by minimising expected free energy, to make future sensations consistent with their model.

Active inference describes the dynamics of systems that persist (i.e., do not dissipate), and that can be statistically segregated from their environment—conditions which are satisfied by biological systems. Mathematically, the first condition means that the system is at non-equilibrium steady-state. This implies the existence of a steady-state probability density to which the system self-organises and returns to after perturbation (i.e., the agent's preferences). The statistical segregation condition is the presence of a Markov blanket, where a set of variables through which states internal and external to the system interact (e.g., the skin is a Markov blanket for the human body).



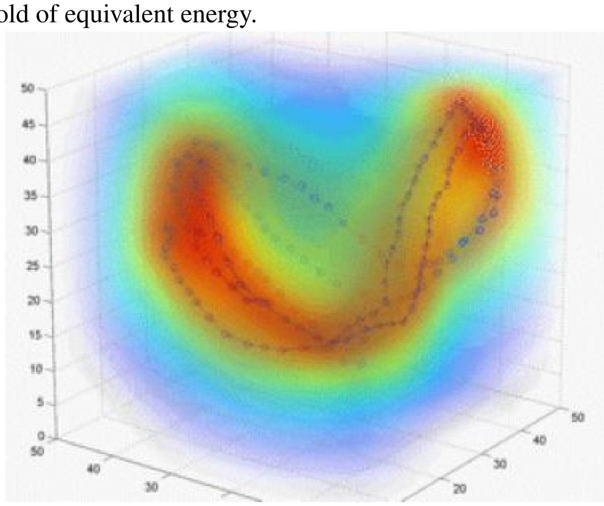
Above is an example of a discrete state-space generative model which is how the agent represents the world. The generative model is a joint probability distribution over (hidden) states, outcomes and other variables that cause outcomes. In this representation, states unfold in time causing an observation at each time-step. The likelihood matrix [A] encodes the probabilities of state–outcome pairs. The policy (Pi) specifies which action to perform at each time-step. Note that the agent's preferences may be specified either in terms of states or outcomes. It is important to distinguish between states (resp. outcomes) that are random variables, and the possible values that they can take in S (resp. in O), which we refer to as possible states (resp. possible outcomes). Note that this type of representation comprises a finite number of timesteps, actions, policies, states, outcomes, possible states and possible outcomes. The arrows represent causal relationships (i.e., conditional probability distributions). The variables highlighted in grey can be observed by the agent, while the remaining variables are inferred through approximate Bayesian inference and called hidden or latent variables. The Markov blanket of a random variable in a probabilistic graphical model are those variables that share a common factor.

Concept of an Unsupervised Loss function and Intrinsic Motivation

Since, most behaviour in individuals is directed or supervised by a goal, agents seem conformity and don't actually build on cognitive structures. Here, the agents is set to traverse in the abstract space without any prior goal or anything such, it forms a geometrical projection on its own manifold, as the process repeats, we can see what the function is being optimised for intrinsically over the timesteps.

Convergence on the Latent

Letting the agents interact with the environment unsupervised and intrinsically, they map out common latent abstract geometric manifolds (shown below). This is the moment of **Cognitive Convergence** on abstract space. The energy-based modeling view would be how a collective converges to a manifold of equivalent energy.



Results and further research directions

The Convergent latent can later manifest itself at common playground of lingual abstractions through language, implicit demographic knowledge through culture or any common cognitive structure that developed intrinsically within the collective. The framework can also be used to describe any process where goal(s) is(are) not explicit and system is set to evolve with random initial configurations.

References

Friston and Kiebel (Predictive coding under the free-energy principle), Philosophical Transactions of the Royal Society, Series B (Biological Sciences), 364

Fuster J.M (Prefrontal cortex and the bridging of temporal gaps in the perception-action cycle), Annals of the New York Academy of Sciences, 608

David Silver et. al (Reward is Enough), Artificial Intelligence - Volume 299.

Ha and Schmidhuber (World Models), NeurIPS 2018.

La Costa et. al (Active inference on discrete state-spaces: A synthesis), Journal of Mathematical Psychology - Volume 99