Self-Learning Embodied Agents: Vision Text

Jan Lemeire, VUB

May 2023

1 The research domain

Let us first have a look at the research question, after which I will outline my proposed approach.

1.1 The goal

"*To bring robots out of well-controlled environments (into the wild), they should be able to adapt and learn.*"

The goal of a *self-learning embodied agent* is to let him start with an empty brain - without prior knowledge, just with an architecture and an algorithm - learn:

- the effect of its actions on the world
- to interact with the world
- to control its actuators
- to control its environment
- to make effective action plans to achieve certain goals

1.2 Why apply self-learning on an agent?

Programming the behavior of a robot based on our scientific and mathematical understanding of the world has its successes. So, why self-learning:

- we don't have to provide the knowledge; it will be discovered autonomously if successful
- to finetune behavior (like calibration)
- to adapt to new, unseen, and uncontrolled environments (called open environments)
- to modify actuator control in case its functionality changed (due to wear or damage, for instance)
- to become versatile

1.3 Challenges for the architecture and algorithm

The main open problems are concerned with:

- achieving the same learning performance as humans, who learn generic knowledge from just a few examples, and
- finding the least amount of prior knowledge; the **first principles** which makes it possible.

1.4 Opportunities of embodiment for Machine Learning

In most ML/AI settings, one has to extract knowledge from a given fixed dataset. An embodied agent, on the other hand, can explore and decide which parts of the state space it wants to sample (active learning). A robot can select the most informative samples by considering the ambiguities and conflicts in its predictive model and history. It can test hypotheses of the world. Embodiment thus allows agents to learn incrementally by integrating exploration, exploitation, and learning.

1.5 Scientific importance

The self-learning challenge opens up a lot of questions that relate to several domains throughout the sciences, of which the most important is what we can learn about ourselves, humans. Because we are the example of good practices in self-learning; human babies are born without any skills. They even don't know how to control their muscles or limbs. They just have the skill to learn... And become experts in understanding and controlling the world. Better than animals that are preprogrammed with instincts. Or at least humans are more versatile.

The field's tentacles are touching questions of:

- **adaptiveness**: can we achieve more through learning?
- the **first principles** of learning. Learning from scratch prevents being biased toward a structure that implicitly contains knowledge.
- **developmental psychology**: how humans learn. Can we mimic it? What can we learn from it? Can we better understand the challenges of human learning? How humans learn is different than the typical brute force machine learning approaches.
- the **brain's architecture** (neurology): how knowledge (models) are stored in our brain (see 1000 brain theory).
- the development of **language** and concept learning. Self-learning leads to a bottom-up approach which contrasts with top-down approaches in which humans design and build methods to control robots, store knowledge, and let them reason. We engineer how knowledge is stored, organized, and, finally, used to control the environment. With self-learning, the challenge is to find out how an

embodied agent can learn this bottom-up: based on interactions with its environment. In our approach, the agents will learn pieces of knowledge by experience. The structural and qualitative nature of the models enables the link with language bottom-up. The advantage is that the agent knows how to link the symbolics to the real world (grounding), apply it, and verify it(!).

- **'analysis by synthesis'**: embodied agents can test hypotheses from psychology and cognitive science, provide insight, and find design principles of biological systems. Construct it to understand it.
- **intelligence, awareness, philosophy of science**...

It would be fabulous if my research could contribute to some of these topics.

2 The proposed approach: causal qualitative models.

"*Mimic the way human infants play, learn and develop.*"

I will put forward an architecture that is fundamentally different than the currently-most popular probabilistic monolithic black box neural network approaches (cf active inference). This approach is meant to be more effective and to capture better the way that humans learn.

2.1 Assumptions

This work plan relies on the following assumptions, which are not shared by the traditional approaches.

Assumption 1: known state The agent is able to deduce the world's state from its observations. Humans have good sensors which allow them to perceive the world up to the smallest detail and infer the state of affairs from it. Cats, for instance, do not have such a good sight. We assume that the agents have a complete set of state variables of the world, where the variables each represent something meaningful in a *disentangled* way. An image of an object, for instance, captures the position and pose of the objects present in the field of view. But indirectly: it has to be inferred from the pixels. Moreover, the information on position and pose is entangled. The pattern of the pixels depends on both together, while, to use this information, both should be known as separate variables. We suppose that the vision system is good at doing that.

There are a lot of possibilities to represent the state. At least the degrees of freedom should be covered. But this is not known a priori. In our approach, however, an overdefinition of the state, with a lot of redundancy among the variables, is not problematic: the algorithm will learn the relations among them and which ones are the most salient.

Assumption 2: determinism and monotonicity Based on the right set of variables (which are known or are to be discovered), the agent's world's dynamics can be modeled with a limited set of piece-wise monotonic quasi-deterministic functions. Most of what happens, happens monotonically. You push something, and it moves. You push harder, it moves faster. A monotonic increasing function, which will allow us to think qualitatively A minimal amount of force might be required to bring the object on the move, so there is a transition from one type of dynamics to another. That's a second important aspect of our solution. Moreover, and especially when you reason qualitatively, things happen deterministically. The object will always move in the direction of the push.

Assumption 3: a simple world The world in which the agents (humans, robots) live is structured and 'simple', such that there is no need for complex algorithms or complex models to understand what is going on. No need for deep mathematics or large scientific calculations. Especially with the right set of state variables (assumption 1) and the right models (coming from assumption 2).

2.2 The architecture and the learning

This links to my previous work on causal structure learning with deterministic relations and context-specific independencies.

2.2.1 Causal model

Our approach is based on first learning basic qualitative models, such as 'if I touch an object, I can change its position', before learning precise quantitative models. Understanding which actuators affect which state variables and under which conditions (e.g. I first have to grab the object before I can lift it). This will give the causal and contextual structure of the model.

How: causal influences are represented by a DAG, such as employed by causal Bayesian networks. Since it is about the dynamics, we add the previous state to the nodes (cf dynamic Bayesian network). But as we will discuss, we will quantify the model with probabilities. The construction algorithm will be similar to the independence-based learning algorithms for causal models (PC algorithm), although we cannot employ the traditional independence tests (the data is not necessarily i.i.d.). New tests (for identifying dependence and for identifying conditional independence) are being developed based on the determinism assumption. As studied intensively in my previous work [], deterministic relations give problems for the traditional algorithms. The results of my previous work have to be taken into account to handle those issues.

2.2.2 Contextual model

Context often changes the dynamics. To incorporate this effect, we depart from one global model but partition the state space into regions of similar behavior. Each such region, which we call a **subspace**, will get its own model. For each subspace, the dependencies among the variables are discovered, and a predictive behavior model is learned. This results in context-specific dependencies: edges that are present under some conditions that are defined over the current state. As we will discuss later, actually, for a given context, a model is created dynamically by gluing small submodels together. In such a way, there is no combinatorial explosion of models to be learned for the combinatorial possible subspaces.

To identify contextual changes, the (in)dependence tests have to be adapted to identify contextual changes (under some assumptions on the nature of the context).

2.2.3 The qualitative 'quantization' of the interactions

Once the relations among the variables are discovered, the models are augmented with a qualitative description of how the variables affect each other.

2.2.4 The quantitative 'quantization' of the interactions

Once the structure and qualitative influences are identified, it is rather easy to quantify the relations. Additional experimental data might be needed to finetune the quantification. But it is rather straightforward. This resembles how children learn: first they learn how to do something in a 'rough' manner, they learn what is expected to happen, and only then do they learn from rehearsal to do it right.

2.2.5 Gradual learning

Learning takes place gradually and is closely intertwined with exploration and exploitation. In a lot of circumstances, the agent is allowed to make errors. Based on wrong models, for example. This is not problematic because he will learn from it by improving the models. Incrementally learning the system's structure—dependencies, starting with simple things, is an essential ingredient for efficient and effective learning.

2.2.6 Component identification

The real proof of the power of the approach should come from being able to extrapolate, reuse and deal with large systems. A system decomposition based on reusable, monotonous components is the key, or stated vice versa, dynamically creating a model by linking submodels [Mugan].

2.3 Exploration strategy

Exploration will steer the learning, the learning will steer the exploration.

2.4 Exploitation: using the models

Good models of behavior can be turned into effective strategies for successfully performing tasks. The qualitative nature of the models makes qualitative reasoning possible. Moreover, the subspaces make it necessary to reason hierarchically: first a path through the subspaces has to be found and then a path within each subspace. This corresponds to the reasoning we often see in literature

2.5 Motivation

We focus first on the structure and patterns of the system and capture the structure explicitly in the model such that it can be exploited for reasoning and generalization. Learning a qualitative model instead of a quantitative one needs less data and allows one to generalize more easily over similar behaviors.

The field of qualitative modelling and simulation also closely connects with symbolic reasoning and logic, which allows the agent to explain the 'why' of its actions and learned model. and simulation also closely connects with symbolic reasoning and logic, which allows the agent to explain the 'why' of its actions and learned model.

In our approach the agents will learn pieces of knowledge by experience. The structural and qualitative nature of the models enables the link with language bottom-up. The advantage is that the agent knows how to link the symbolics to the real world (grounding), apply it, and verify it.

Although we take inspiration from the human example, we will not a priori stick to artificial neural networks such as is regularly done. This might be misleading because the brain's wiring and functioning implement a layer of knowledge acquisition and reasoning which might be very difficult to discern by studying the neurons and their interactions.

2.6 Advantages of approach

It bears explainability in its structure.

Allows few- or even one-shot learning (like humans). Qualitative dynamics under the assumption of determinism and monotonicity are much easier to identify than the quantitative approaches for probabilistic settings.

3 Work plan.

Current problems in consideration include: grasping objects with a robot arm (the REAL competition), learning in our simulations, and solving various AI gym challenges.