An Environment Transformation-based Framework for Comparison of Open-World Learning Agents

Matthew Molineaux, Dustin Dannenhauer
4035 Colonel Glenn Hwy, Beavercreek, OH, 45431
first.last@parallaxresearch.org

Abstract

To compare the ability of agents to learn in open worlds, we need a framework with clear definitions of open world environments and how they can vary. This paper provides such a framework, proposing clear scientific definitions for open world environments, transformations of those environments, and characterizations of those transformations. We also provide a description of the minimum necessary conditions under which learning can be expected to occur, and a discussion of how the framework can be used.

Introduction

The challenge of "open world learning", as we consider it here, requires an agent to be responsive and adapt to changes in its environment's behavior or observability. Such changes are useful for modelling a variety of difficult unforeseen challenges, such as rare environmental dynamics an agent is unaware of, like tornadoes; discontinuous, unobservable events that modify how actions work, such as a magnetic pole flip; or new sensors or effectors that change an agent's interface to the environment. While in each of these cases, there is an argument that the environment is not *actually* changing, it is nevertheless expedient to compare agents by changing their environments for the sake of unbiased experimentation.

In this paper, we seek to provide a framework that permits objective measurement and comparison of open-world learners' efficiency at adapting to changed circumstances. Tools of this framework include general formal definitions of open-world environments, *novelties*, and *novelty regions* that can be applied to any discrete-time environment, and a set of *novelty dimensions*, characteristics that can be used to group and describe novelties for the purposes of classifying open-world learning challenges. We formally define each of these concepts so as to have a solid foundation for discussion of open world challenges.

New research areas require language that differentiates the various challenges they encompass. In action selection problems, we refer to environments as "discrete-time" or "continuous", "single-agent" or "multi-agent", "deterministic" or "stochastic". It is our hope that the dimensions of novelty introduced here can be used in the same way to describe challenges in this developing field.

Unlike work that seeks to define novelty relative to an agent's experiences (e.g., Muhammad et al, 2020; Boult et al, 2021; Gamage et al, 2021), we intentionally avoid any reference to an agent's knowledge. Such frameworks are useful because they can identify how prepared an agent is for new challenges. However, they cannot usefully describe how environment challenges differ in a way that cuts across differences in knowledge representation. For historical and language reasons, both currently use the term "novelty", but in different ways.

In the rest of this paper, we review related work on open-world learning, present a novel formal description of multi-agent discrete-time environments, formally define a new set of novelty dimensions that classify environment transformations, provide a set of characteristics necessary for an environment transformation to provide agents an opportunity to learn, and conclude.

Related Work

Boult et al. (2021) defined a theory of open world novelty where novelty occurs when an agent experiences an environment sufficiently different from its prior experiences. This is fundamentally different from our definition in several ways; first, in our framework, novelty exists independent of any given agent's experiences or knowledge, for reasons given above. Second, our framework does not define novelty as occurring at a point in time, but rather as encompassing the time-independent difference between two environments. This means that novelty can exist without any changes to the observation or state space. While valuable for considering what experiences could be new to a particular agent, two agents cannot truly be compared on the "same" novelty unless they have identical prior experiences; such comparisons are a key feature of our framework.

Langley's (2020) work provides a set of broad requirements for a theory of environmental change that can help explain and measure progress in open-world learning. He suggested that such a theory would propose a formalism for environments and transformations on them. Our framework provides both, and so can be consider an example "theory" in this regard that describes environmental changes and provides specific language for characterizing how environment changes vary.

Wiggins' (2006) framework for creative system in AI describes novelty as a property of a creative output which previously did not exist. While our description of environment transformation novelties makes no restriction on the generating process, it is related: part of the purpose of our framework is to describe constraints on the generation of pairs of environments pre- and post-transformation, where the later one has properties that did not previously exist in the earlier one, so we are broadly consistent with Wiggins' definition.

Transformation-Based Novelty

Our work depends on a formal definition of novelty and an agent's interaction with the occurrence of the novelty. The formal definition allows us to describe precisely what is being measured and how the agent is impacted. We now define the formal basis of our definition of novelty from an environment-based perspective.

We describe novelty over a space of discrete-time, turn-taking environments. An environment σ can be described as a tuple $(S, A, O, Ag, Tu, \gamma, \omega)$ where S is the space of possible states, A is the space of possible actions, O is the space of observations, Ag the space of agents, Tu is a turn function $S \rightarrow Ag$ which determines which agent acts in a given state, γ is the probabilistic transition function $S \times A \times S \times Ag \rightarrow \mathbb{R}$, and ω is the probabilistic observation function $S \times O \times Ag \rightarrow \mathbb{R}$. The agent itself is an argument to the observation function, as different agents may have different perceptual interfaces and action capabilities. During environment iteration, in each state s, each actor $ag \in Ag$ receives an observation o according to probability $\omega(s, o, ag)$. Then, one environment-selected agent $Tu(s)$ selects an action a, and the environment transitions to a new state s' with probability $\gamma(s, a, s', Tu(s))$. Other agents cannot take turns, and their probability of transition is 0:

$$\begin{aligned} &\forall s, s' \in S, a \in A, ag \in Ag: \\ &\qquad \gamma(s, a, s', ag) = 0 \lor Tu(s) = ag. \end{aligned}$$

An *environment transformation*, or *novelty*, is a tuple $(\sigma, \sigma^M, \alpha)$ in which σ is an original environment as defined above, $\sigma^M = (S^M, A^M, O^M, Ag^M, Tu^M, \gamma^M, \omega^M)$ is a modified environment and α is the observing agent. Such a tuple is only a novelty if the transition functions (γ) of the original and modified environment differ or the observation function (ω) differs relative to the observing agent.

We define a "novelty-relevant region" (NR) of the modified state space which encapsulates all states where novelty can be observed or affect the behavior of the agent. As the observation function is agent-relative, the novelty region is as well. It consists of all states for which either the transition function or the observation function has a different outcome in the original and new environments. If two environments are the same, given a particular agent, the novelty region of the new environment will be empty. This can occur if, for example, the novelty only occurs in the observation function relevant to some agents.

$$\begin{aligned} &NR(\alpha, \sigma = (S, A, O, Ag, Tu, \gamma, \omega), \\ &\qquad \sigma^M = (S^M, A^M, O^M, Ag^M, Tu^M, \gamma^M, \omega^M)) \\ &\quad \equiv \{s \in S^M \mid \exists a \in A^M, s' \in S^M: \\ &\qquad\quad \gamma(s, a, s', Tu(s)) \neq \gamma^M(s, a, s', Tu(s))\} \\ &\qquad \cup \{s \in S^M \mid \exists o : \omega(s, o, \alpha) \neq \omega^M(s, o, \alpha)\} \end{aligned}$$

Outside the novelty region, the two environments will appear and react identically. Naturally, no learning can be expected to take place until an agent enters the novelty region.

In many realistic cases, we expect that particular actions or objects will precipitate the entry of an agent into the novelty region. We expect tracking these to be particularly important to the study of novelty. We will therefore refer to a set of novelty entry transitions, NET, that describes where an agent can enter the novelty region.

$$\begin{aligned} &NET(\alpha, \sigma = (S, A, O, Ag, Tu, \gamma, \omega), \sigma^M) \equiv \\ &\quad \{(s, a, s') \mid \neg\, s \in NR(\alpha, \sigma, \sigma^M) \\ &\qquad\qquad \land \gamma(s, a, s', Tu(s)) > 0 \\ &\qquad\qquad \land s' \in NR(\alpha, \sigma, \sigma^M)\} \end{aligned}$$

To permit greater specificity about how actions change in the novelty region, we use factored representations. States are therefore described as assignment functions to a finite vector of discrete state variables $\vec{S} = \{s_0, s_1, \ldots, s_n\}$. Each $s \in S$ assigns a value $s(i)$ to every variable s_i in $\vec{S}$. Similarly, O is a space of assignment functions to a vector of discrete and continuous variables $\vec{O} = \{o_0, o_1, \ldots\ o_n\}$. Each $o \in O$ assigns a value $o(i)$ to every variable $o_i \in \vec{O}$. To describe conditional transition probabilities, we use the notation $P\gamma$ to refer to the probability of changes to individual state variables under a transition function γ. For example, we refer to the probability that the ith state variable takes on a value v following an action a_t taken in state s_t by agent $\alpha \in Ag$ as $P\gamma(s_{t+1}(i) = v \mid a_t, s_t, \alpha)$. The following mathematical equations relate the conditional probabilities of state variable values before and after a transition to the transition function itself:

$$P\gamma(s_{t+1}(i) = v \mid a_t, s_t, \alpha) = \\ \Sigma_{\{s' \in S \mid s'(i) = v\}} \gamma(s_t, a_t, s', Tu(s_t))$$

$$P\gamma(s_{t+1} | a_t, s_t(i) = v) = \Sigma_{\{s \in S \mid s(i) = v\}} \gamma(s, a_t, s_{t+1}, Tu(s_t))$$

To reason about what is knowable to an agent, we introduce the concepts of state observability and trajectory, which allow us to describe the limits of what an agent can infer based on the information provided by its environment.

We say that a state *s* is "discernible" when the observation function ω gives an agent α enough information to immediately determine that it is in that state *s*. Formally, this concept means that every observation *o* received in state *s* is not received in other states.

$$\mathit{discernible}(s, \sigma, \alpha) \equiv \\ \forall s', o{:}\ (\omega(s', o, \alpha) > \epsilon \rightarrow s' = s) \lor \omega(s, o, \alpha) < \epsilon$$

In order to reason about what an agent might be able to know, we consider trajectories that describe the totality of an agent's experiences during an environment interaction. A trajectory τ is a linked list where each element is either the empty trajectory ∅ or a tuple (*a, o, τ'*). Here, *a* is the agent's selected action if it's turn is up, or *yield* if it is not; *o* is the subsequent observation received; τ' is the remaining trajectory. We define an *extension* of a trajectory τ with new action *a* and observation *o* as a copy of that trajectory with a new final element (*a, o,* ∅):

$$\mathit{extension}(\emptyset, a, o) \equiv (a, o, \emptyset)$$

$$\mathit{extension}(\tau = (a_i, o_i, \tau_i), a, o) \equiv \\ (a_i, o_i, \mathit{extension}(\tau_i, a, o))$$

The length of a trajectory τ is defined as 0 for ∅, and 1 + the length of the subtrajectory for all others:

$$\mathit{length}(\emptyset) \equiv 0$$

$$\mathit{length}(\tau = (a, o, \tau')) \equiv 1 + \mathit{length}(\tau')$$

Due to nondeterminism, an agent α with policy π and initial state *s* may experience any one of multiple possible histories τ of any given length. We define an agent's decision-making policy π: *T* → *A* as a mapping from the space of trajectories to an action to take at the end of that trajectory. Given a fixed policy π for an agent α in an environment σ, a state sequence **s**, and past trajectory τ^P, we say that the trajectory τ^F is a *possible future* for agent α when a sequence of states exists, starting with *s*, such that:

- each action in τ^F is consistent with policy π given the subtrajectory ending prior to it,
- each action in τ^F can lead from the corresponding state in the sequence to the following state, *and*
- each observation in τ^F can be received in the corresponding state in the sequence.

The empty trajectory ∅ is always a possible future.

$$\mathit{possible\text{-}future}(\mathbf{s} = [s_0], \pi, \sigma, \alpha, \tau^P, \emptyset) \equiv \mathbf{True}$$

$$\begin{aligned}&\mathit{possible\text{-}future}(\mathbf{s} = [s_0 \ldots s_n], \pi, \sigma, \alpha, \tau^P, \tau^F) \equiv \\ &\quad \sigma = (S, A, O, Ag, Tu, \gamma, \omega) \\ &\quad \land \tau^F = (a, o, \tau^{F'}) \\ &\quad \land \pi(\tau^P) = a \\ &\quad \land a = \mathit{yield} \lor \exists a'{:}\ \gamma(s_0, a', s_1, Tu(s_0)) > 0 \\ &\quad \land a \neq \mathit{yield} \lor \gamma(s_0, a, s_1, \alpha) > 0 \\ &\quad \land \omega(s_1, o, \alpha) > 0 \\ &\quad \land \mathit{possible\text{-}future}([s_1 \ldots s_n], \pi, \sigma, \alpha, \\ &\qquad\qquad \mathit{extension}(\tau^P, a, o)), \tau^{F'})\end{aligned}$$

Dimensions of Novelty

We have developed a set of dimensions for *classifying* environment-based novelty that are intended to assist in exploration of what makes novelty in the environment challenging for an agent to recognize, characterize, and respond to. We hypothesize that different types of agents will exhibit heterogeneous capabilities in robustness to different dimensions. We also expect dimension values to be correlated with the performance of the agent with respect to different novel environments.

This section is our major contribution. In each of the following eight sections, we give a description of a single dimension along which novelty can vary. This is followed by a bulleted list of possible values for each dimension, each of which is first described informally, then formally. In the same way that terms such as "deterministic", "discrete" and "multi-agent" are used to describe and differentiate environments for the purposes of action selection, we hope that this framework will serve as a point of reference for discussion of the environment transformations a learning algorithm supports.

In all formal definitions presented in this section, novelty dimension values are presented relative to a novelty (σ, σ^M, α). Unless otherwise stated, quantified variables with the following base letters are in spaces as follows: $a \in A \cup A^M$, $s \in S \cup S^M$, $o \in O \cup O^M$, $ag \in Ag \cup Ag^M$. Variables with base τ and π are trajectories and policies respectively.

Novel External Agents

In some cases, an agent's recognition or characterization of environment novelty may be confounded by other agents taking actions whose effects can mask and/or masquerade as changes in the environment. With respect to an agent α, there are two possible conditions:

- **Novel External Agents Present**: Unfamiliar observations can be caused by third parties affecting the environment.

  Formally, occurs when a third party agent has the capability to affect environment states with its actions, and adopts a policy in the modified environment that no agents took in the original environment.

$$\exists ag \in (Ag^{M} \setminus \alpha) \\ \quad \wedge\ \forall ag' \in Ag\text{: policy}(ag) \neq \text{policy}(ag')$$

- **Novel External Agents Not Present**: No third parties are present that can cause confounding sources of unfamiliar observations.

  Formally, occurs when all third-party agents either use the same policies used in the original environment, *or* have no ability to affect the state of the environment.

  $$\forall ag \in (Ag^{M} \setminus \alpha)\text{:} \\ \quad \exists ag' \in (Ag \setminus \alpha)\text{: policy}(ag) = \text{policy}(ag') \\ \quad \vee\ \forall a, a', s, s'\text{: } \gamma^{M}(s, a, s', ag) = \gamma^{M}(s, a', s', ag)$$

For example, in a self-driving vehicle domain, an experiment could add an agent with a "drunk driving" policy to the modified environment; even though weaving was possible in the original environment, weaving did not occur because no agents had a policy that caused it. We would then say that "novel external agents present" is a novel dimension of the experiment. However, as policies are not considered part of the environment itself, this can also occur independent of an environment transformation, and outside of a novelty region.

Transition Novelty Type

When a dynamic in the world changes that is observable to an agent α, novelty can be revealed in one of multiple ways, often involving a single variable. These are the simplest types of observable novelty change that can occur. In all cases below, we assume that observability of the changing state variable is high, and that's why this is the way the agent encounters the novelty. Thus, some pair of observations *o* and *o''* occur that are highly indicative of the particular (actual) states *s* and *s''* to that agent. Some agents may be biased to expect novelty to occur in the form of one of these transitions and neglect others.

- **New change**: Novel transitions cause changes to state variables that were left unchanged in the original environment

  Formally, the value of some state variable s_i is always left unchanged when taking some particular action *a* in some particular state s_t in the original environment, but whose value is sometimes changed when taking action *a* in state s_t in the modified environment.

  $$\exists s_t, a, i, \alpha\text{:} \\ \quad P\gamma(s_{t+1}(i) = s_t(i) \mid a, s_t, \alpha) = 1 \\ \quad \wedge\ P\gamma^{M}(s_{t+1}(i) = s_t(i) \mid a, s_t, \alpha) < 1$$

- **Missing change**: Novel transitions leave state variables unchanged which were modified in the original environment

  Formally, the value of some state variable s_i is sometimes changed when taking some particular action *a* in some particular state s_t in the original environment, but whose value is always left unchanged when taking action *a* in state s_t in the modified environment.

  $$\exists s_t, a, i, \alpha\text{: } P\gamma^{M}(s_{t+1}(i) = s_t(i) \mid a, s_t, \alpha) = 1 \\ \quad \wedge\ P\gamma(s_{t+1}(\mathrm{i}) = s_t(i) \mid a, s_{t+1}) < 1$$

- **Replaced Change**: Novel transitions cause different changes to a state variable than in the original environment

  Formally, for some state variable $\mathrm{s_i}$, different changes must occur when taking some particular action a in some particular state $\mathrm{s_t}$ in the original environment and modified environment. Specifically, for values c and d, $\mathrm{s_i}$ must sometimes take the value c after action a in the changed environment, but never d. In the modified environment, $\mathrm{s_i}$ must sometimes take the value d after action a, but never c.

  $$\exists s_t, a, i, c, d, \alpha\text{: } P\gamma(s_{t+1}(i) = c \mid a, s_t, \alpha) > 0 \\ \quad \wedge\ P\gamma(s_{t+1}(i) = d \mid a, s_t, \alpha) = 0 \\ \quad \wedge\ P\gamma^{M}(s_{t+1}(i) = c \mid a, s_t, \alpha) = 0 \\ \quad \wedge\ P\gamma^{M}(s_{t+1}(i) = d \mid a, s_t, \alpha) > 0$$

- **Modified change rate**: The probability of existing transitions is modified.

  Formally, there is some state variable s_i whose probability of taking on some value *c* subsequent to some particular transition is different in the original and modified environments.

  $$\exists s_t, a, i, c\text{: } P\gamma(s_{t+1}(i) = c \mid a, s_t) > 0 \\ \quad \wedge\ P\gamma^{M}(s_{t+1}(\mathrm{i}) = c \mid a, s_t) > 0 \\ \quad \wedge\ P\gamma(s_{t+1}(i) = c \mid a, s_t) \neq P\gamma^{M}(s_{t+1}(i) = c \mid a, s_t)$$

- **Increase or Decrease in Change Size**: Changes to ordinal-valued state variables are larger or smaller in the modified environment.

  Formally, there is some state variable s_i whose probability of increasing by a value greater than *c* is negligible in the original environment but non-negligible in the modified environment, *or* whose probability of increasing by a value less than *c* is negligible in the original environment but non-negligible in the modified environment.

  $$\exists s_t, a, i, c\text{:} \\ \quad P\gamma(s_{t+1}(i) - s_t(i) > c \mid a, s_t, \alpha) < \epsilon \\ \quad\quad \wedge\ P\gamma^{M}(s_{t+1}(i) - s_t(i) > c \mid a, s_t, \alpha) > \epsilon \\ \quad \vee\ P\gamma(s_{t+1}(i) - s_t(i) > c \mid a, s_t, \alpha) > \epsilon \\ \quad\quad \wedge\ P\gamma^{M}(s_{t+1}(i) - s_t(i) > c \mid a, s_t, \alpha) < \epsilon$$

Novelty Entry Determinism

This dimension concerns the agent's transition into the novelty region. Infrequent or agent-caused novelty entry may lower the likelihood that the agent encounters the novelty at all, whereas different agent's biases may allow them to more easily represent deterministic or nondeterministic transitions.

- **Deterministic**: Entry into the novelty region is deterministic.

  $$\forall(s, a, s') \in NET(\alpha, \sigma, \sigma^{M})\text{: } \gamma(s, a, s', \alpha) = 1$$

- **Frequent Nondeterministic**: Some transitions into the novelty region are non-deterministic, and at least one is high probability.

  $$\exists(s, a, s') \in NET(\alpha, \sigma, \sigma^{M})\text{: } \gamma(s, a, s') \geq 0.5 \wedge \gamma(s, a, s') < 1$$

- **Infrequent Nondeterministic**: All transitions into the novelty region are low probability, which may make the novelty region difficult to discover or characterize.

  $$\forall(s, a, s') \in NET(\alpha, \sigma, \sigma^{M})\text{: } \gamma(s, a, s') < 0.5$$

- **Agent-Caused**: The agent is in control of all entry into the novelty region.

  Formally, on the agent's turn, there must be some action that cannot transition to the novelty region; and on other agent's turns, there must be no action that can.

  $$\begin{aligned}&\forall s\ \exists a\ \forall s'\text{: } \neg\ Tu(s) \vee \neg\ (s, a, s') \in NET(\alpha, \sigma, \sigma^{M})\\ &\wedge \forall s, a, s', ag\text{: } \neg\ (s, a, s') \in NET(ag, \sigma, \sigma^{M}) \vee ag = \alpha\end{aligned}$$

Novelty Obviousness

This dimension concerns the minimum amount of observation and reasoning required to determine that the agent's present environment is not the original environment. This will affect the amount of time and concentration required for an agent to recognize novelty.

- **Alien observation**: Observations occur in the modified environment that are not present in the original

  Formally, there is some observation *o* that is in the observation space of the modified environment but not the original environment, *and* there is a non-zero probability that the agent α will receive that observation in some state *s*.

  $$\exists s, o \in O^{M}\text{: } \neg\ (o \in O) \wedge \omega^{M}(s, o, \alpha) > 0$$

- **Entry Transition Discernible:** An observation occurs on entry into the novelty region that is not consistent with the original environment.

  Formally, for each novelty entry transition (s_t, a, s_{t+1}), there is a set of corresponding non-novel transitions in the original environment that start in the same state (s_t) and take the same action (a) but lead to a different "original" state (s^{O}_{t+1}) (one not in the novelty region). Entry observability means two things: First, for every observation the probability of receiving that observation in the original environment must be arbitrarily small in either the novelty entry transition origin state (s_t) or every other. (Thus receiving that observation nearly always means the agent is in that particular state.) Second, for every observation the probability of receiving that observation must be arbitrarily small in either every original state ($s_{t+1}{}^{O}$) reached by *a* from s_t or the novelty region entry state ($s_{t+1}{}^{M}$). (Thus it is possible to tell that the transition is not one that exists in the original environment.)

  $$\begin{aligned}&\forall o \in (O \cup O^{\mathrm{M}}), (s_t, a, s_{t+1}{}^{M}) \in NET(\alpha, \sigma, \sigma^{M})\text{:}\\ &\quad discernible(s_t, \sigma^{\mathrm{M}}, \alpha)\\ &\quad \wedge \forall s_{t+1}{}^{O}\text{: } \gamma(s_t, a, s_{t+1}{}^{O}, \alpha) < \epsilon\\ &\qquad \vee \forall o\text{: } \omega(s_{t+1}{}^{O}, o, \alpha) < \epsilon\\ &\qquad\qquad \vee \omega^{M}(s_{t+1}{}^{M}, o, \alpha) < \epsilon\end{aligned}$$

- **Novel Transition Discernible**: Some novel transition available to the agent can be observed unambiguously.

  Formally, there is some modified environment transition that starts in the novelty region that cannot happen in the original environment. Both the originating state and the destination state of this transition are unambiguously recognizable, because observations received in these states are not received in other states in the original environment.

  $$\begin{aligned}&\exists s_t \in NR(\alpha, \sigma, \sigma^{M}), a, s_{t+1}\text{: } \gamma^{M}(s_t, a, s_{t+1}, \alpha) > 0\\ &\quad \wedge \gamma(s_t, a, s_{t+1}, \alpha) = 0\\ &\quad \wedge discernible(s_t, \sigma, \alpha)\\ &\quad \wedge discernible(s_{t+1}, \sigma, \alpha)\end{aligned}$$

- **Novelty Region Inferable**: Given enough time, an agent will obtain enough information to infer that it has experienced some novel transitions and/or observations.

  Formally, for every state *s* in the novelty region and every policy π, all changed environment (σ^{M}) possible futures (τ^{F}) of a length greater than *n* are trajectories that are not possible futures starting in any state *s'* in the original environment σ, no matter how the state *s* or *s'* was reached.

  $$\begin{aligned}&\forall \mathbf{s} = [s_0 \ldots s_m], \pi, \tau^{\mathrm{P}}, \exists n, \forall \tau^{\mathrm{F}}\text{:}\\ &\quad length(\tau^{\mathrm{F}}) = m\\ &\quad\ \wedge m > n\\ &\quad\ \wedge s_0 \in NR(\alpha, \sigma, \sigma^{M})\\ &\quad\ \wedge \neg\ possible\text{-}future(\mathbf{s}, \pi, \sigma^{\mathrm{M}}, \alpha, \tau^{\mathrm{P}}, \tau^{\mathrm{F}})\\ &\qquad \rightarrow \forall \mathbf{s}'\text{: } \neg\ possible\text{-}future(\mathbf{s'}, \pi, \sigma, \alpha, \tau^{\mathrm{P}}, \tau^{\mathrm{F}})\end{aligned}$$

- **Novelty Region Investigable**: Given some policy, an agent can obtain sufficient information to infer after entering the novelty region that the agent had experienced some novel transitions and/or observations.

  Formally, there is some traversal **s** starting in the novelty region of the modified environment σ^{M} whose trajectory τ^{F} is not a possible future in the original environment, for any traversal **s'**.

  $$\begin{aligned}&\exists \mathbf{s} = [s_0 \ldots s_n], \pi, \tau^{\mathrm{P}}, \tau^{\mathrm{F}}\text{:}\\ &\quad possible\text{-}future(\mathbf{s}, \pi, \sigma^{\mathrm{M}}, \alpha, \tau^{\mathrm{P}}, \tau^{\mathrm{F}})\\ &\quad\ \wedge s_0 \in NR(\alpha, \sigma, \sigma^{M})\\ &\quad\ \wedge \forall \mathbf{s'}\text{: } \neg\ possible\text{-}future(\mathbf{s'}, \pi, \sigma, \alpha, \tau^{\mathrm{P}}, \tau^{\mathrm{F}})\end{aligned}$$

- **Statistical Novelty Only**: Some changes may require statistical analysis to verify; all histories possible in the modified environment are also possible in the original.

Formally, any trajectory possible in the modified environment is also possible in the original environment (possibly visiting different states).

$$\forall \mathbf{s} = [s_0 \ldots s_n], \pi, \tau^P, \tau^F: \\ \mathit{possible\text{-}future}(s, \pi, \sigma^M, \alpha, \tau^P, \tau^F) \rightarrow \\ \rightarrow \exists \mathbf{s'}: \mathit{possible\text{-}future}(\mathbf{s'}, \pi, \sigma, \alpha, \tau^P, \tau^F)$$

- **Indistinguishable**: Some changes to environments might never be directly observable to an agent; we generally consider these to be "unfair" or "uninteresting" changes as they afford no learning opportunity.

  Formally, for every possible future τ^F, policy π, and state s, there is an equivalent conditional probability of experiencing that future τ^F in the original environment σ and from some state s' in the changed environment σ^M. (This makes the two environments indistinguishable from the agent's perspective, although the actual states and transitions may have changed.)

$$\forall \pi, s, \tau^P, \tau^F \; \exists s': \\ \Pr(\tau^F \mid s, \tau^P, \pi, \sigma, \alpha) = \Pr(\tau^F \mid s', \tau^P, \pi, \sigma^M, \alpha)$$

Novelty Observation Determinism

This dimension concerns how likely the agent is to receive a novel observation if one is possible. More infrequent observations may be harder to learn as an agent may disregard them as noise. Additionally, many agents may be biased to assume deterministic or nondeterministic observations.

- **Deterministic**: Novel observation mappings are deterministic.

  Formally, the probability of receiving any given observation o in a particular state s is the same in both environments, or is exactly 0 or 1 in the modified environment.

$$\forall s, o: \omega^M(s, o, \alpha) = \omega(s, o, \alpha) \\ \lor \omega^M(s, o, \alpha) = 0 \lor \omega^M(s, o, \alpha) = 1$$

- **Frequent Nondeterministic**: Novel observation mappings are deterministic or happen 50% more or less often in the modified environment than the original.

  Formally, the probability of receiving any given observation o in a particular state s is the same in both environments, or is exactly 0 or 1 in the modified environment, or the difference between the probabilities is greater than 0.5.

$$\forall s, o: \omega^M(s, o, \alpha) = \omega(s, o, \alpha) \\ \lor \omega^M(s, o, \alpha) = 0 \lor \omega^M(s, o, \alpha) = 1 \\ \lor |\omega^M(s, o, \alpha) - \omega(s, o, \alpha)| > 0.5$$

- **Infrequent Nondeterministic**: Some novel observation mappings occur infrequently or with close to the same frequency as in the original environment.

  Formally, there is some state s and observation o such that the probability $\omega^M(s, o, \alpha)$ of an agent α receiving that observation o in that state s in the modified environment is neither the same as in the original environment, nor exactly 0, nor exactly 1; instead, it is within 0.5 of the original environment probability $\omega(s, o, \alpha)$ of receiving the same observation o in the same state s.

$$\exists s, o: |\omega^M(s, o, \alpha) - \omega(s, o, \alpha)| < 0.5$$

Novelty Observation Localization

This dimension concerns the danger implicit in encountering a meaningful change in state before sensing that something is different; in some environments, it is possible for the agent to avoid novelty that degrades its performance, and in others not.

- **Nearby**: Agents encounter novel transitions before observations; that is to say, the environmental novelty will start to affect the agent before the agent would easily notice it.

  Formally, any possible trajectory τ in the modified environment σ^M that leads to a novel observation o, must either be also possible in the original environment (with some other state sequence) or be impossible prior to the new observation (i.e., due to a previous novel transition).

$$\forall \mathbf{s} = [s_0 \ldots s_n], \pi, \alpha, \tau, \tau', a, o: \\ \omega(s_n, o, \alpha) = 0 \\ \land \mathit{length}(\tau) = n - 2 \\ \land \tau' = \mathit{extension}(\tau, a, o) \\ \land \mathit{possible\text{-}future}(\mathbf{s}, \pi, \sigma^M, \alpha, \emptyset, \tau') \\ \land \neg \mathit{possible\text{-}future}(\mathbf{s}, \pi, \sigma, \alpha, \emptyset, \tau') \\ \rightarrow \neg \mathit{possible\text{-}future}([s_0 \ldots s_{n-1}], \pi, \sigma, \alpha, \emptyset, \tau)$$

- **Distant**: Agents encounter novel observations before novel transitions; they will observe surprising observations before being affected by the environment novelty.

  Formally, for any traversal of the modified environment that ends with a novel transition, if an agent's trajectory τ' over that traversal is not possible in the original environment, the subtrajectory τ ending just before the final transition is also not possible in the original environment.

$$\forall \mathbf{s} = [s_0 \ldots s_n], \pi, \alpha, \tau, \tau', a, o: \\ \gamma(s_{n-1}, a, s_n, Tu(s_{n-1})) = 0 \\ \land \mathit{length}(\tau) = n - 1 \\ \land \tau' = \mathit{extension}(\tau, a, o) \\ \land \mathit{possible\text{-}future}(\mathbf{s}, \pi, \sigma^M, \alpha, \emptyset, \tau') \\ \land \neg \mathit{possible\text{-}future}(\mathbf{s}, \pi, \sigma, \alpha, \emptyset, \tau') \\ \rightarrow \neg \mathit{possible\text{-}future}([s_0 \ldots s_{n-1}], \pi, \sigma, \alpha, \emptyset, \tau)$$

- **Mixed**: Sometimes novel observations are encountered first, sometimes novel transitions. We define mixed environments as those that are neither nearby nor distant.

Novel State Observation Activity

This dimension concerns whether effort is required for an agent to observe what is different about the environment.

Some agents may fail to put in effort necessary to find information that will prove beneficial later.

- **Passive**: The agent's observation of effects does not require active effort. Extra action is not required to distinguish novel states.
  Formally, for any two states in the modified environment, s and s', that cannot be distinguished by observation: every action a that leads from s to s'' with non-zero probability has an equivalent transition with identical probability from s' to s''', and s'' and s''' are indistinguishable by observation.

$$\begin{aligned}
&\forall s \in S^M / S, s', a, s'':\\
&\quad (\forall o \in O^M: \omega^M(s, o, \alpha) = \omega^M(s', o, \alpha)\\
&\quad \wedge \gamma^M(s, a, s'', \alpha) > 0)\\
&\quad\quad \rightarrow \exists s''': (\gamma^M(s, a, s'', \alpha) = \gamma^M(s', a, s''', \alpha)\\
&\quad\quad\quad\quad \wedge \forall o': \omega^M(s'', o', \alpha) = \omega^M(s''', o', \alpha))
\end{aligned}$$

- **Active**: The agent must make a special effort to detect novel transitions. Some novel states may be indistinguishable from other states without specific action.
  Formally, for some pair of transitions s, a, s'', and s', a, s''' in the modified environment, at least one of which (s) is in the modified environment state space but not the original environment state space, the two origination states (s, s') cannot be distinguished, but the two end states (s'', s''') can.

$$\begin{aligned}
&\exists s \in S^M / S, s', s'', s''', a:\\
&\quad\quad \forall o: \omega^M(s, o, \alpha) = \omega^M(s', o, \alpha)\\
&\quad\quad \wedge \gamma^M(s, a, s'', \alpha) > \epsilon \wedge \gamma^M(s', a, s''', \alpha) > \epsilon\\
&\quad\quad \wedge \exists o': \omega^M(s'', o', \alpha) \neq \omega^M(s''', o', \alpha)
\end{aligned}$$

- **Resource Expenditure Required**: There are environment actions designed for observation that are needed to distinguish novel states, such as a limited-use Geiger counter that detects novel sources of radiation.
  Formally, for any two states in the modified environment, s and s', that cannot be distinguished by observation, there are two types of outgoing actions: observation actions and non-observation actions. Non-observation actions that lead from a state s to s'' with non-zero probability have equivalent transitions with identical probability from s' to s''', and s'' and s''' are indistinguishable by observation. Observation actions allow them to be distinguished, and strictly decrease reachability. Here, we define the set of states reachable from a set of states **s**:

$$\begin{aligned}
&\mathit{adjacent}(\mathbf{s}, s', \sigma = (S, A, O, Ag, Tu, \gamma, \omega)) \equiv\\
&\quad\quad \exists a \in A, s \in \mathbf{s}: \gamma(s, a, s') > 0
\end{aligned}$$

$$\begin{aligned}
&\mathit{reach}(\mathbf{s}, \sigma) \equiv\\
&\quad\quad \mathbf{s} \cup \mathit{reach}(\mathbf{s} \cup \{s' \in S \setminus \mathbf{s} \mid \mathit{adjacent}(\mathbf{s}, s', \sigma)\}
\end{aligned}$$

$$\begin{aligned}
&\forall s \in S^M / S, s', a, s'':\\
&\quad (\forall o \in O^M: \omega^M(s, o, \alpha) = \omega^M(s', o, \alpha)\\
&\quad \wedge \gamma^M(s, a, s'', \alpha) > 0)\\
&\quad\quad \rightarrow (\exists s''': \gamma^M(s, a, s'', \alpha) = \gamma^M(s', a, s''', \alpha)\\
&\quad\quad\quad\quad \wedge \forall o': \omega^M(s'', o', \alpha) = \omega^M(s''', o', \alpha)\\
&\quad\quad\quad \vee \mathit{reach}(\{s''\}, \sigma) \subset \mathit{reach}(\{s\}, \sigma))\\
&\quad\quad\quad\quad \wedge \forall s''': \gamma^M(s', a, s'', \alpha) = 0\\
&\quad\quad\quad\quad\quad \vee \mathit{reach}(\{s'''\}, \sigma) \subset \mathit{reach}(\{s'\}, \sigma))
\end{aligned}$$

Novelty Frequency

This dimension concerns how frequently novel changes can occur. Changes that are more frequent than an agent can see, or continuous, may be significantly more difficult for some agents to track and respond to.

- **Less frequent than Observation**: Commonly, observations occur frequently enough to observe important events, so a novel change will occur at most once between observations.
- **More frequent than Observation**: Some novelty sources may have a periodicity faster than an agent's observations, such that multiple changes may occur between one observation and the next.
- **Continuous**: Novelty sources may cause change that is continuous over time, such as the water level in a reservoir.

We leave the formalization of this dimension to be addressed within a continuous time framework.

Fairness

It is easy to construct novelties that are clearly distinct but very difficult to learn about. For example, states may be different between two environments only in some subtle way that is not visible to the agent's sensors. Some unobservable novel event might cause an agent's immediate destruction, allowing no response and providing no useful information the agent could use to avoid that event in the future. Novel transitions might occur only in a distant region of the state space that the agent never encounters in practical situations. We consider some possible conditions that may be necessary to ensure fairness in novelty experimentation.

- *Relevant*: Novelty will affect the agent's performance directly
- *Noticeable*: The agent will have the opportunity to witness something different about the new environment.
- *Controllable*: The agent will have sufficient power to do something about the novelty

To formally define these qualities, we require the following definitions and assumptions. We assume that an agent is motivated by some environment-independent performance measure that is known or observable to it. This function, *Performance:* $T \times \mathbf{S} \rightarrow \mathbb{R}$ is some mapping from the space of possible trajectories and actual states traversed to the real numbers; this function is calculated over the full sequence of states and actions in the agent's past trajectory. The space of final states of an environment σ, $S_F(\sigma) \subset S$ is defined as the set of states which have no outgoing transitions:

$$S_F(\sigma = (S, A, O, Ag, Tu, \gamma, \omega)) \equiv \\ \{s \in S \mid \forall a \in A, s' \in S : \gamma(s, a, s') = 0\}$$

We then define the expected performance of a policy in an environment based on a past trajectory τ^P, state history $\boldsymbol{s}$, and policy π as the performance of the past trajectory in any final state, or otherwise as the average expected performance of the policy over each possible next state and observation as given by recursion over the next state and updated trajectory.

$$\begin{aligned} &EPP(\tau^P, \pi, \mathbf{s} = (s_0 ... s_n \in S_F(\sigma)), \alpha, \sigma) \\ &\qquad \equiv Performance(\tau^P, \mathbf{s}) \\ &EPP(\tau^P, \pi, \mathbf{s} = (s_0 ... s_n \notin S_F(\sigma)), \alpha, \\ &\qquad \sigma = (S, A, O, Ag, Tu, \gamma, \omega)) \\ &\qquad \equiv \sum_{s' \in S, o \in O} \begin{pmatrix} \gamma(s_n, \pi(\tau^P), s', \alpha) \\ * \, \omega(s', o, \alpha) \\ * \, EPP(extension(\tau^P, \pi(\tau^P), o), \\ \pi, concat(\mathbf{s}, s'), \alpha, \sigma) \end{pmatrix} \end{aligned}$$

Formally, the novelty of a modified environment σ^M relative to an original environment σ is relevant to an agent α in some starting state s_0 iff there is some policy π that produces different expected performance in σ and σ^M.

$$Relevant(\alpha, \sigma, \sigma^M, s_0) \equiv \\ \exists\pi: EPP(\emptyset,\pi,s_0,\alpha,\sigma) \neq EPP(\emptyset,\pi,s_0,\alpha,\sigma^M)$$

The novelty of a modified environment σ^M relative to an original environment σ is noticeable iff it gains enough experiences of σ^M during training to reject the hypothesis that those observations came from σ. We omit the full formal definition of noticeability to conserve space.

The novelty of an environment is considered controllable by an agent α if there is some policy π' that performs better for α in the modified environment σ^M than the optimal policy π* for the original environment σ.

$$\begin{aligned} &Controllable(\alpha, \sigma, \sigma') \equiv \\ &\quad \exists\pi^* \, \forall\pi, s \in S: \\ &\qquad EPP(\emptyset, \pi, s, \alpha, \sigma) \leq EPP(\emptyset, \pi^*, s, \alpha, \sigma) \\ &\qquad \wedge \exists\pi' \, \forall s \in S^M: \\ &\qquad\qquad EPP(\emptyset, \pi^*, s, \alpha, \sigma') \leq EPP(\emptyset, \pi', s, \alpha, \sigma') \end{aligned}$$

While these constructions ensure fairness, they also may make some problems easier than desired; future work will continue investigations into means of ensuring the fairness conditions without simplifying the target problem.

Discussion

New work on learning in open-world environments has many varied challenges. We've presented a new formal basis for describing and differentiating those challenges, that we hope will be useful in the comparison of new agent algorithms that are suited to certain types of challenges.

We have also considered the types of problems that are practically solvable. The scope of possible environment transformations is immense, and these basic "fairness" conditions start to consider what the limits of possible performance may be.

We recognize several areas of potential improvement to this framework. First, the discrete-time, turn-taking formulation of the environments is awkward, compromising the simplicity of a Markov decision process-like representation to acknowledge the presence of multiple agents. It also prevents us from formally defining novelty frequency. A future version of this framework will incorporate continuous-time change.

Second, while classifications of problems are useful, we have not yet defined any numeric measurements that can be made, which could strengthen our ability to compare.

Third, we expect refinements to this framework in the form of added dimensions, revised novelty dimension value definitions, and added values as errors are detected and affordances found. We have attempted to find orthogonal dimensions, but relationships between them are likely and should be identified.

While we believe it is important to consider environments separate from agent knowledge, this does present certain difficulties. When comparing agent's learning speed, it will be important to consider that some agents have advantages due to "accidental" model compatibilities. For example, if an environment transformation requires vehicles to drive on the left rather than the right, an agent whose model predicts negative consequences for "driving toward oncoming vehicles" will perform much better than one whose model predicts negative consequences for "driving in a left lane", even though they may have performed identically in their original environment. Considering how to measure or prevent such unfair advantages will be important to future experimentation.

References

Boult, T. E., Grabowicz, P. A., Prijatelj, D. S., Stern, R., Holder, L., Alspector, J., ... & Scheirer, W. J. (2021). Towards a Unifying Framework for Formal Theories of Novelty. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 17, pp. 15047-15052).

Gamage, C., Pinto, V., Xue, C., Stephenson, M., Zhang, P, & Renz, J. (2021). Novelty Generation Framework for AI Agents in Angry Birds Style Physics Games. *IEEE Conference on Games (IEEE-COG'21)*. Copenhagen, Denmark.

Langley, P. (2020). Open-world learning for radically autonomous agents. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 09, pp. 13539-13543).

Muhammad, F., Sarathy, V., Tatiya, G., Goel, S., Gyawali, S. Guaman, M., Sinapov, J., & Scheutz, M. (2021). A Novelty-Centric Agent Architecture for Changing Worlds. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2021).

Wiggins, G. A. (2006). A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems*, *19*(7), 449-458.