Automatic Exploration of Textual Environments with Language-Conditioned Autotelic Agents

Anonymous ACL submission

This extended abstract discusses the opportunities and challenges of studying intrinsically-motivated agents for exploration in textual environments.

Humans begin their life with very few skills, and over the course of only a few years learn complex motor coordination and locomotion capabilities, begin mastering vocalization and language, and form a rich model of their physical and social surroundings. One of the main drivers of this phenomenal knowledge acquisition is intrinsically-motivated exploration (Oudeyer and Kaplan, 2007), for instance through exploratory play (Chu and Schulz, 2020; Davidson et al., 2022). The developmental perspective on AI tries to emulate this exploratory behavior in artificial agents to achieve mastery of diverse and complex repertoires of skills (Forestier et al., 2017). When placed in open-ended environments, a successful intrinsically motivated agent will explore the space of interesting and diverse outcomes, ignoring random and unachievable subspaces of the world, reusing its previously acquired skills as stepping stones (Stanley and Lehman, 2015) to discover new ones.

One possible implementation of exploration in RL agents are so-called autotelic agents (Colas, 2021), that is, goal-conditioned Reinforcement Learning (RL) agents operating in rewardless environments that are able to choose what goal to pursue. In this case, the reward is given by a goal-satisfaction function and not extrinsically by the environment. Goal-conditioned policies have been extensively studied in the case of extrinsic goals (Schaul et al., 2015). In the case of intrinsically chosen goals, the goal-selection mechanism allows autotelic agents to form a self-curriculum, progressing from easier to increasingly harder goals until all achievable skills have been mastered. In this perspective, the goal representation is of paramount importance. Most previous works (for instance Andrychowicz et al. (2017)) have used concrete end-state representations such as raw observations, images or embeddings, which has some drawbacks. A goal should be insensitive to changes in the environment that are uncontrollable (such as the color of the sky), to avoid the agent targeting impossible goals (for instance changing the sky color), or to provide useful abstraction for goal achievement (such as considering the goal of navigating to the garden is satisfied regardless of sky color). Furthermore, the agent should ideally be able to combine known goals into novel ones. **Goals expressed as language** (Tam et al., 2022; Colas et al., 2020; Mu et al., 2022) fulfill both conditions: they are at once abstract and combinatorial (Szabó, 2020); they are thus a prime way for autotelic agents to self-specify goals to be executed in the environment.

1 A bridge between autotelic agents and text environments

The main point of this essay is the relevance of studying language autotelic agents in textual environments (Côté et al., 2018; Hausknecht et al., 2020; Wang et al., 2022), both for testing exploration methods in a context that is at once simple experimentally and rich from the perspective of environment interactions; and for transferring the skills of general-purpose agents trained to explore in an autonomous way to the predefined tasks of textual environment benchmarks. We identify three key properties, plus one additional benefit, of text worlds:

1. Depth of learnable skills: skills learnable in the world should involve multiple low-level actions and be nested, such that mastering one skill opens up the possibility of mastering more complex skills. Interactive fiction (IF) (Hausknecht et al., 2020) games usually feature an entire narrative and extensive maps, such that navigating and passing obstacles requires many successful actions (and subgoals) to be completed. While the origi-

nal TextWorld levels were not as deep as would be desirable, other non-IF text worlds such as ScienceWorld feature nested repertoires of skills (such as learning to navigate to learn to grow plants to learn the rules of Mendelian genomics);

2. Breadth of the world: there should be many paths to explore in the environment; this ensures that we train agents that are able to follow a wide diversity of possible goals, instead of learning to achieve goals along a linear path. This allows us to study generally-capable agents. Some IF games are very linear, having a clear progression from start to finish (e.g., Acorn Court, Detective; others have huge maps that an agent has to explore before it can progress in the quest (e.g., Zork, Hitchhiker's Guide to the Galaxy). Exploration heuristics are a part of some successful methods for playing IF with RL (Yao et al., 2020). ScienceWorld (Wang et al., 2022) has an underlying physical engine allowing for a combinatorial explosion of possibilities like making new objects, combining existing objects, changing states of matter, etc.

3. Niches of progress: real-world environments have both easy skills and unlearnable skills. Our simulated environments should mimic this property to test the agent's ability to focus only on highly learnable parts of the space and avoid spending effort on uncontrollable aspects of the environment. In textual environments, high depth implies that some skills are much more learnable than others, already implementing some progress niches. The combinatorial property of language goals allows us to define many unfeasible goals, goals that an autotelic agent has to avoid spending too many resources on.

4. Language representation for goals: a language-conditioned agent has to learn to ground its goal representation in its environment (Harnad, 1990; Hill et al., 2020), to know when a given observation or sequence of observations satisfies a given goal, or to know what goals were achieved in a given trajectory. This grounding is made much simpler in environments with a single modality; relating language goals to language observations is simpler than grounding language in pixels or image embeddings. This allows us to study language-based exploration in a simpler context.

2 Drivers of exploration in autotelic agents

We identify three main drivers of exploration in autotelic agents. Environments we use should support exploration algorithms that implement these principles; the resulting agents then have a chance to acquire a diverse set of skills that can be repurposed for solving the benchmarks proposed in textual environments.

1. Goal self-curriculum: automatic goal selection (Portelas et al., 2020) allows the agent to refine its skills on the edge of what it currently masters. Among metrics used to select goals are novelty/surprise of a goal (Tam et al., 2022; Burda et al., 2018), intermediate competence on goals (Campero et al., 2020), ensemble disagreement (Pathak et al., 2019), or (absolute) learning progress (Colas et al., 2019). Progress niches in textual environments support such goal curriculum;

2. Additional exploration after goal achievement: after achieving a given goal, the agent continue to run for a time to push the boundary of explored space (Ecoffet et al., 2021). The depth of text worlds makes goal chaining relevant, such that an agent that has achieved a known goal can imagine additional goals to pursue. Random exploration can also be used once a known goal has been achieved. Agents exploring in textual environments and choosing uniformly among the set of valid actions in a given state have a high chance of effecting meaningful changes in the environment, making discovery of new skills probable. This property is relevant in any environment with high depth, and both IF and ScienceWorld fit this description.

3. Goal composition: as mentioned above, this means using the compositionality afforded by language goals to imagine novel goals in the environment (Colas et al., 2020). Goal-chaining is an example of composition, but language offers many other composition possibilities, such as recombining known verbs, nouns and attributes in novel ways, or making analogies. This is relevant if there exists some transfer between the skills required to accomplish similar goal constructions (e.g., picking up an apple and picking up a carrot requires very similar actions if both are in the kitchen). This is at least partially true in textual environments where objects of the same type usually have similar affordances.

3 Challenges for autotelic textual agents

Text worlds bring a set of unique challenges for autotelic agents, among which we foresee:

1. The goal space can be very large. An agent with a limited training budget needs to focus on a subset of the goal space, possibly discovering only a fraction of what is discoverable within the environment. This calls for finer goal-sampling approaches that encourage the agent at making the most out of its allocated time to explore the environment. In addition, we need better methods to push the agent's exploration towards certain parts of the space (e.g., warm-starting the replay buffer with existing trajectories, providing linguistic commonsense knowledge);

2. The action space is also very large in textual environments making exploration potentially challenging.

3. Agents must be trajectory-efficient for a given goal; complex goals might be seen only once;

4. Catastrophic forgetting needs to be alleviated, so that learning to achieve new goals does not impair the skills previously learned.

Agents trained in such environments will learn a form of language use, not by predicting the most likely sequence of words from a large-scale dataset (Radford and Narasimhan, 2018; Brown et al., 2020) but by learning to use it pragmatically to effect changes in the environment. Of course, the limits of the autotelic agent's world will mean the limits of its language; an interesting development is to build agents that explore textual environments to refine external linguistic knowledge provided by a pretrained language model. This external knowledge repository can be seen as culturally-accumulated common sense, a perspective that is related to so-called Vygotskian AI (Colas, 2021) in which a developmental agent learns by interacting with an external social partner that imparts outside language knowledge and organizes the world so as to facilitate the autotelic agent's exploration.

To conclude, textual environments are ideal testbeds for autotelic language-conditioned agents, and conversely such agents can bring progress on text world benchmarks. There is also promise in the interaction between exploratory agents and large language models encoding exterior linguistic knowledge. We hope to foster discussion, define concrete implementations and identify challenges by bringing together the developmental perspective on AI and the textual environment community.

References

Marcin Andrychowicz, Filip Wolski, Alex Ray, Jonas Schneider, Rachel Fong, Peter Welinder, Bob McGrew, Josh Tobin, OpenAI Pieter Abbeel, and Wojciech Zaremba. 2017. Hindsight experience replay. *Advances in neural information processing systems*, 30.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Yuri Burda, Harrison Edwards, Amos Storkey, and Oleg Klimov. 2018. Exploration by random network distillation. *arXiv preprint arXiv:1810.12894*.

Andres Campero, Roberta Raileanu, Heinrich Küttler, Joshua B Tenenbaum, Tim Rocktäschel, and Edward Grefenstette. 2020. Learning with amigo: Adversarially motivated intrinsic goals. *arXiv preprint arXiv:2006.12122*.

Junyi Chu and Laura Schulz. 2020. Play, curiosity, and cognition. *Annual Review of Developmental Psychology*, 2.

Cédric Colas. 2021. *Towards Vygotskian Autotelic Agents : Learning Skills with Goals, Language and Intrinsically Motivated Deep Reinforcement Learning*. Theses, Université de Bordeaux.

Cédric Colas, Pierre Fournier, Mohamed Chetouani, Olivier Sigaud, and Pierre-Yves Oudeyer. 2019. Curious: intrinsically motivated modular multi-goal reinforcement learning. In *International conference on machine learning*, pages 1331–1340. PMLR.

Cédric Colas, Tristan Karch, Nicolas Lair, Jean-Michel Dussoux, Clément Moulin-Frier, Peter Dominey, and Pierre-Yves Oudeyer. 2020. Language as a cognitive tool to imagine goals in curiosity driven exploration. *Advances in Neural Information Processing Systems*, 33:3761–3774.

Marc-Alexandre Côté, Akos Kádár, Xingdi Yuan, Ben Kybartas, Tavian Barnes, Emery Fine, James Moore, Matthew Hausknecht, Layla El Asri, Mahmoud Adada, et al. 2018. Textworld: A learning environment for text-based games. In *Workshop on Computer Games*, pages 41–75. Springer.

Guy Davidson, Todd M Gureckis, and Brenden M Lake. 2022. Creativity, compositionality, and common sense in human goal generation.

Adrien Ecoffet, Joost Huizinga, Joel Lehman, Kenneth O Stanley, and Jeff Clune. 2021. First return, then explore. *Nature*, 590(7847):580–586.

Sébastien Forestier, Rémy Portelas, Yoan Mollard, and Pierre-Yves Oudeyer. 2017. Intrinsically motivated goal exploration processes with automatic curriculum learning. *arXiv preprint arXiv:1708.02190*.

Stevan Harnad. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346.

Matthew Hausknecht, Prithviraj Ammanabrolu, Marc-Alexandre Côté, and Xingdi Yuan. 2020. Interactive fiction games: A colossal adventure. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7903–7910.

Felix Hill, Olivier Tieleman, Tamara von Glehn, Nathaniel Wong, Hamza Merzic, and Stephen Clark. 2020. Grounded language learning fast and slow. *arXiv preprint arXiv:2009.01719*.

Jesse Mu, Victor Zhong, Roberta Raileanu, Minqi Jiang, Noah Goodman, Tim Rocktäschel, and Edward Grefenstette. 2022. Improving intrinsic exploration with language abstractions. *arXiv preprint arXiv:2202.08938*.

Pierre-Yves Oudeyer and Frederic Kaplan. 2007. What is intrinsic motivation? a typology of computational approaches. *Frontiers in neurorobotics*, 1:6–6. 18958277[pmid].

Deepak Pathak, Dhiraj Gandhi, and Abhinav Gupta. 2019. Self-supervised exploration via disagreement. In *International conference on machine learning*, pages 5062–5071. PMLR.

Rémy Portelas, Cédric Colas, Lilian Weng, Katja Hofmann, and Pierre-Yves Oudeyer. 2020. Automatic curriculum learning for deep rl: A short survey. *arXiv preprint arXiv:2003.04664*.

Alec Radford and Karthik Narasimhan. 2018. Improving language understanding by generative pre-training.

Tom Schaul, Dan Horgan, Karol Gregor, and David Silver. 2015. Universal value function approximators. page 1312–1320.

Kenneth O. Stanley and Joel Lehman. 2015. *Why Greatness Cannot Be Planned: The Myth of the Objective*. Springer Publishing Company, Incorporated.

Zoltán Gendler Szabó. 2020. Compositionality. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, Fall 2020 edition. Metaphysics Research Lab, Stanford University.

Allison C Tam, Neil C Rabinowitz, Andrew K Lampinen, Nicholas A Roy, Stephanie CY Chan, DJ Strouse, Jane X Wang, Andrea Banino, and Felix Hill. 2022. Semantic exploration from language abstractions and pretrained representations. *arXiv preprint arXiv:2204.05080*.

Ruoyao Wang, Peter Jansen, Marc-Alexandre Côté, and Prithviraj Ammanabrolu. 2022. Scienceworld: Is your agent smarter than a 5th grader? *arXiv preprint arXiv:2203.07540*.

Shunyu Yao, Rohan Rao, Matthew Hausknecht, and Karthik Narasimhan. 2020. Keep calm and explore: Language models for action generation in text-based games. *arXiv preprint arXiv:2010.02903*.