Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU33500

A Survey on Hindsight Experience Replay for Sparse Reward Problems

Yuhan Song

School of Management Science and Engineering, Central University of Finance and Economics, Beijing, China
songyuhan0208@email.cufe.edu.cn

Abstract. Sparse reward is a critical problem in reinforcement learning (RL), where agents receive sparse or delayed rewards, which can drastically affect exploration and learning efficiency. This problem results in inefficient learning, increased sample complexity and credit assignment problems. There are several proposed methods to overcome these issues, such as reward shaping, intrinsic motivation, hierarchical reinforcement learning, and imitation learning. But many of these techniques require extra information or lack scalability. Hindsight Experience Replay (HER) is a simple yet powerful alternative which relabels goals and treats unsuccessful trajectories as a success under different goals. This allows agents to increase their sample efficiency by learning from nearly all past experiences, which is particularly beneficial for goal-oriented tasks like robotics manipulation and navigation. This paper examines the sparse reward problem and its current understandings, specifically HER and its variants. It also outlines recent advances, and proposes some future directions for research. Recent works have also examined the combination of HER and other methods to enhance HER's efficacy when dealing with more complex problems.

Keywords: Sparse reward problem, Hindsight Experience Replay, Reinforcement Learning, Exploration strategies.

1. Introduction

Reinforcement Reinforcement learning (RL) has had great success in many applications such as robotics and decision making [1,2]. Markov Decision Processes (MDP) help to both analyze the RL problem and define a solution. Agents learn to take actions in their environments and receive feedback as they proceed by trial and error. RL in that sense will likely have different results depending on the design of the reward signals. This is especially true since in many cases, the feedback that the learner receives may be less frequent and may have a long delay, making reinforcement learning learning very slow and inefficient.

Due to those problems, many methods have been proposed to make using RL and sparse reward systems more efficient. Reward shaping is one of them, and it alters the reward in a system to make them less frequent in order compensate for the delay in feedback. On the other hand, novel exploration strategies with and without intrinsic motivation are proposed to stimulate discovery.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU33500

Other works involve Hierarchical Reinforcement Learning (HRL), which decompose a problem further to make it more digestible. Again, there is a challenge of applying such methods in environments with less structure that require more knowledge of the domain in order to comprehend the problems being solved.

Hindsight Experience Replay (HER) was introduced as a more efficient and effective approach to RL with sparse reward systems. HER proposes a different perspective to view and therefore learn from said uncompleted goals. Tasks from goal-related problems are generally improved by sample efficiency through HER.

2. Challenges of sparse rewards in reinforcement learning

In sparse reward environments, reinforcement learning agents encounter numerous difficulties because they do not have informative feedback cues throughout training [3]. When compared to dense reward environments that have feedback, sparse rewards do not provide the feedback required to facilitate learning and hence the learning process is highly inefficient [4].

A big problem is inefficient exploration. Sparse reward situations compel the agents to employ random exploration policies to discover rewarding states [5], which may be difficult as the size of the state space is augmented [6,7]. Combined with advanced exploration methods, including count-based exploration and intrinsic motivation, the probability of discovering any meaningful rewards in highly complex environments remains low [8-10].

Another issue is sample efficiency. The fact that the chances of encountering a reward are rare implies that, to learn good policies, it takes a number of interactions. This does not only increase the computational requirements, but also the application of the reinforcement learning to real world problems that require expensive data collection or experimentation [3].

The credit assignment problem is also added in the sparse reward case. It is difficult to make agents attribute the correct actions to success in cases where rewards are sparse. This may cause problems in learning series of actions.

Such concerns highlight the need to develop more effective solutions to reinforcement learning problems, and more effective approaches to learning using data like Hindsight Experience Replay that attempts to learn using experiences that seem to be failures in the first place.

3. Existing methods for solving sparse reward problems

Numerous techniques have been proposed in reinforcement learning to deal with issues of sparse rewards. These methods either boost exploration or provide additional information to the learning process to fill in the gaps of feedback.

3.1. Reward shaping

A prevalent example is reward shaping, which offers supplementary rewards [6]. By setting some rewards, the learning process can be swifter, as some prior information is introduced to the reward function. Still, it can be tricky to formulate, and the impacts of poor choice of rewards can be sub-optimal laws and undesirable policies.

3.2. Intrinsic motivation

Another example speaks to the exploration issue by integrating intrinsic motivation and count-based methods. These methods support the exploration of the gaps in the unknown and uncertain states by

DOI: 10.54254/2755-2721/2026.GU33500

deploying intrinsic rewards [10]. They can work on exploration in the sparse reward problems, however, careful consideration of intrinsic rewards should be taken and perhaps such methods may fail to scale to the complex problems.

3.3. Hierarchical reinforcement learning

Hierarchical reinforcement learning (HRL) can deal with sparse reward problems by subdividing the problems or the tasks into manageable units and addressing them as subtasks [11]. The segmentation is still a difficult process and may require expert insight.

3.4. Imitation learning

Alternatively, one can use imitation learning that provides demonstrations to guide the agent's behaviour [12,13]. It can significantly reduce the need for exploration, especially when exploring the optimal behaviour is difficult. But it requires expert knowledge, which may be difficult to acquire.

4. Hindsight Experience Replay (HER)

While these approaches have achieved great success, they have drawbacks. Many methods require additional reward shaping, supervision or specific structures, restricting applicability. This motivates new approaches that can accelerate learning without additional knowledge. Hindsight Experience Replay (HER) addresses this challenge through a different approach, improving experience reuse. Instead of modifying the reward function or relying on expert feedback, HER enables agents to learn from their failures by considering these experiences as successes with different goal.

4.1. Overview of HER

Hindsight Experience Replay (HER) was proposed as an effective way to address sparse reward problems in reinforcement learning through improving the reuse of experiences [4]. Rather than the traditional methods of tweaking reward functions or enhancing exploration, HER uses a different approach in that it reuses unsuccessful experiences. This allows learning algorithms to learn information even without rewards. This has contributed to the recent surge of HER, particularly in goal-conditioned reinforcement learning.

4.2. Methodology

HER builds on the experience replay, by increasing the number of transitions with goal labels in the replay buffer. In an off-policy setting, the transitions are stored as (s, a, r, s', g) where g is the goal. However, HER not only stores the transitions, but also relabels the transitions by replacing the goal with other goals from the trajectory [14].

Specifically, HER uses a goal relabeling strategy that samples the actual future states from the trajectory as the goal. The reward function is then updated to this new goal, and so failing episodes can be used for training. In particular, sampling future states from the same episode has been very effective in practice [4].

HER is typically used with off-policy reinforcement learning algorithms, such as Deep Deterministic Policy Gradient (DDPG), that can make good use of past knowledge. HER improves learning efficiency and stability by providing more relevant transitions to the replay buffer [3,4].

DOI: 10.54254/2755-2721/2026.GU33500

This allows HER to more effectively reuse samples without additional reward engineering or human intervention.

4.3. Advantages

HER offers several improvements to the above methods for sparse reward tasks. First, HER significantly improves sample efficiency by enabling agents to learn from both success and failure. By relabeling the goal, it enables the agent to learn from otherwise uninformative experiences to learn a good policy in fewer interactions [4]. Second, HER helps solve the exploration problem in sparse reward tasks. The agent can learn from unsuccessful trajectories, rather than having to rely on random exploration to find successful trajectories. This is particularly helpful in cases with sparse rewards. Third, unlike other methods, HER does not require reward engineering nor other types of shaping functions. This makes HER more task independent and does not require task specific expertise.

Finally, HER can be combined with other approaches to improve learning and is inherently off-policy. For example, HER can be used in conjunction with demonstrations to support efficient exploration by taking advantage of the experience of an expert [15]. More broadly, this means that HER can be combined with other reinforcement learning methods.

4.4. Limitations

Despite HER being successful, it is not without its flaws. One limitation is the possible bias introduced by relabeling the goals, especially in the multi-goal setting, which can affect the learning of the policy and lead to sub-optimal policies [14]. In this case, relabeled goals might not be the true task goals and thus provide incorrect signals to the learning process.

In addition, HER is designed for goal-conditioned and episodic tasks, limiting its application in other types of reinforcement learning problems. It may not be ideal in tasks with a more ambiguous goal specification or a high-dimensional state such as vision. This has prompted recent advances in HER extensions and improvements, which are outlined in the next section.

5. Extensions and integration of HER

5.1. HER with exploration-based techniques

HER aids sample efficiency through relabeling, but it does not address exploration challenges in sparse reward environments. To tackle this, some recent approaches has addressed the exploration problem in environments with sparse rewards by combining HER with intrinsic motivation and curiosity exploration. These approaches offer additional rewards for exploring previously unseen states [8,10]. Consequently, agents tend to explore beyond the boundaries of state spaces previously reached, allowing for quicker learning in sparse reward environments. Thus, a combination of HER and exploration methodologies for learning act in balance meta skills, exploiting relabeling of experiences to explore previously unseen states.

5.2. HER with hierarchical reinforcement learning and demonstrations

HER has been utilized with HRL to accomplish improved task decomposition. This allows agents to develop sub-goals for a task, and HER supplements this capability by adjusting the goals post-failure attempts [11]. Also, HER has been employed with demonstration learning, where agents can

DOI: 10.54254/2755-2721/2026.GU33500

utilize expert demonstrations despite not achieving the goals [15]. In the latter, learning is accelerated, especially in the real world, where exploration is expensive.

The versatility of HER is shown in these methods where HER can be used in isolation or in combination with other methods to increase HER's use and effectiveness in more complex environments.

6. Conclusion

One common problem within reinforcement learning (RL) concerns environments that provide sparse rewards. However, one promising solution to this is Hindsight Experience Replay (HER), which effectively allows agents to learn from both their positive and negative experiences. As a result, HER improves sample efficiency and exploration when sparse reward signals are provided. In addition, HER can also incorporate other adaptable exploration techniques, Hierarchical Reinforcement Learning (HRL), and Learning from Demonstrations (LfD).

HER does work with high-dimensional spaces, but has been constrained by the complexity of the goals. One possible direction for work would be creating better Goal Relabelling (GR) that extend HER to continuous control domains, and to other, more advanced RL techniques. HER will continue to improve the efficiency and ability of RL algorithms to operate effectively in sparse reward environments.

References

[1] Sutton, R.S., & Barto, A.G. (1998). Reinforcement Learning: An Introduction. IEEE Transactions on Neural Networks, 9, 1054-1054, doi: 10.1109/TNN.1998.712192.
[2] Li, Y. (2017). Deep Reinforcement Learning: An Overview. arXiv (Cornell University). https: //doi.org/10.48550/arxiv.1701.07274
[3] Ladosz, P., Weng, L., Kim, M., & Oh, H. (2022). Exploration in deep reinforcement learning: A survey. Ithaca. doi: https: //doi.org/10.1016/j.inffus.2022.03.003
[4] Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Abbeel, P., & Zaremba, W. (2018). Hindsight experience replay. arXiv. https: //doi.org/10.48550/arXiv.1707.01495
[5] Bellman, R. (1957). A Markovian decision process. Indiana University Mathematics Journal, 6(4), 679–684.
[6] Ng, A.Y., Harada, D., & Russell, S.J. (1999). Policy invariance under reward transformations: Theory and application to reward shaping. Proceedings of the Sixteenth International Conference on Machine Learning, 278–287.
[7] Ostrovski, G., Bellemare, M.G., van den Oord, A., & Munos, R. (2017). Count-based exploration with neural density models. Proceedings of the 34th International Conference on Machine Learning, 2721–2730.
[8] Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 488–489. https: //doi.org/10.1109/CVPRW.2017.70
[9] Machado, M.C., Bellemare, M.G., & Bowling, M.H. (2018). Count-Based Exploration with the Successor Representation. AAAI Conference on Artificial Intelligence
[10] Achiam, J., & Sastry, S. (2017). Surprise-based intrinsic motivation for deep reinforcement learning (arXiv: 1703.01732). arXiv. https: //arxiv.org/abs/1703.01732
[11] Pateria, S., Subagdja, B., Tan, A.-h., & Quek, C. (2021). Hierarchical reinforcement learning: A comprehensive survey. ACM Computing Surveys (CSUR), 54(5), Article 109. https: //doi.org/10.1145/3453160
[12] Hussein, A., Gaber, M. M., Elyan, E., & Jayne, C. (2017). Imitation learning: A survey of learning methods. ACM Computing Surveys (CSUR), 50(2), Article 21.
[13] Ding, Y., Florensa, C., Phielipp, M., & Abbeel, P. (2019). Goal-conditioned imitation learning. In Advances in Neural Information Processing Systems (NeurIPS), 15324–15335.
[14] Bai, C., Zhang, Y., Chen, Y., & Xu, W. (2023). Addressing hindsight bias in multigoal reinforcement learning. IEEE Transactions on Cybernetics, 53(1), 392-405. https: //doi.org/10.1109/TCYB.2021.3107202

DOI: 10.54254/2755-2721/2026.GU33500

[15] Nair, A., McGrew, B., Andrychowicz, M., Zaremba, W., & Abbeel, P. (2017). Overcoming exploration in reinforcement learning with demonstrations. arXiv. https: //doi.org/10.48550/arXiv.1709.10089