In: 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC),
5-8 October 2025, Vienna, Austria, pp. 5643-5648
DOI: 10.1109/smc58881.2025.11342492

Interpreting and Enhancing Decisions in Autonomous Navigation: A Belief-Desire-Intention Reinforcement Learning (BDI-RL) Approach

Adolfo Perrusquía, *Member, IEEE*, Deepak Panda, Weisi Guo, *Senior Member, IEEE*

***Abstract*— Explaining autonomy is becoming a crucial factor in the design of trustworthy autonomous platforms in both transport and smart living sectors. Interpretable reinforcement learning (RL) is an emerging research area that aims to explain why an autonomous platform adopts an action or set of actions. However, the state-of-the-art has focused on the design of explainable tools as independent modules that are not involved in the decision-making process of the RL agent. In this paper, we propose a novel belief-desire-intention RL (BDI-RL) approach that incorporates the explainable module as a belief model that enhances the learning capabilities of the RL as well as actions interpretability. To this end, we combine the merits of Dyna-Q algorithm as backbone RL model and belief maps as explainable element. The combined contribution of these models provides a robust model that emulates better the reasoning process of humans by leveraging beliefs and online agent-environment interactions. Simulations experiments are conducted in a grid environment of different sizes and obstacles. Comparisons are also provided to show the benefits of the proposed methodology.**

I. Introduction

Interpreting autonomous decisions is a key element for the development of safe and trustworthy autonomous platforms in both transportation and smart-living sectors. Here, applications of autonomous systems, e.g., in search and rescue [1], [2] and defence [3], [4], need a high-level of trust to avoid any hazards to people, national facilities, and the autonomous system itself. Most of the interpretable tools in the literature have been designed for supervised learning models that have a linear or non-linear mapping between inputs and outputs [5], [6]. However, reinforcement learning (RL) agents do not have a simple mapping between inputs and outputs, such that other approaches are required. Interpretable Reinforcement Learning [7] is an emerging research area that aims to provide the algorithmic tools to interpret RL decisions [8], [9]. Several methodologies have been developed in function of the RL algorithm design (e.g., tabular models, expert models, or deep neural models) which provide several insights of what we need to consider to explain autonomous decisions [10].

Explainable models are usually independent modules used in interrogable interfaces [11] that helps the user or designer to understand why a prediction or decision is obtained. Here, a natural question arises which consists in: how can we use the outcomes of explainable models to enhance the decision-making process? This is a natural mechanism in human reasoning where based in our interpretations of the world, we choose the actions that we believe are correct.

In this paper, we aim to contribute in answering the previous question by providing a novel interpretable RL approach. The proposed novelty is based on the belief-desired-intention (BDI) model that gives a natural mechanism to incorporate: 1) beliefs as explainable model and 2) belief-driven models, into the learning process to enhance the robustness and interpretability of actions. Convergence of the proposed approach is verified using contraction mappings. Experiments and comparisons are conducted to assess the approach and identify their benefits, challenges, and future research opportunities.

A. Related Work

Explainable AI (XAI) has been used to address the interpretation problem of supervised learning architectures to establish relations between input features and the outputs of the ML model [12]. Two different methods are identified: perturbation and gradient-based methods. Both approaches aim to uncover the local importance weights by altering the input features. These explainable methods allow the construction of a low complexity model that locally approximates the ML algorithm [13], [14]. However, RL agents are dynamic systems that do not pose a simple mapping from inputs to outputs such that it is not possible to use classical XAI tools.

RL decisions have been interpreted from three mainly sources: control policy [15], [16], reward function [17], and intended outcomes [18]. Interpretations from control policies are the most natural explainable mechanism since they allow to find a mapping between states and actions. Imitation learning algorithms are widely used to infer a desired behaviour as well to explain the adopted actions from the relation between states and actions using standard explainable techniques [19]. On the other hand, interpretations from the reward function design allows to define value mappings from rewards to actions. These mappings allow to understand why an action is adopted given by the optimisation of the reward function. Inverse Reinforcement Learning (IRL) algorithms [20], [21] have been applied to uncover an equivalent reward function that produces the same or almost same policy. Intended outcomes [18] aim to explain the actions preference of RL agents. In this approach, value function decomposition is used to generate a belief map with a similar structure to the

This work was supported by the Royal Academy of Engineering and the Office of the Chief Science Adviser for National Security under the UK Intelligence Community Postdoctoral Research Fellowship programme.

A. Perrusquía. Deepak Panda and W. Guo are with the School of Aerospace, Transport and Manufacturing, Cranfield University, Bedford, UK (e-mail: adolfo.perrusquia-guzman@cranfield.ac.uk).

Published by IEEE. This is the Author Accepted Manuscript issued with: Creative Commons Attribution License (CC:BY 4.0).
The final published version (version of record) is available online at DOI:10.1109/smc58881.2025.11342492. Please refer to any applicable publisher terms of use.

value function that are sequentially updated from the agent-environment interaction.

The aforementioned approaches are effective in designing explanations of RL decisions from different perspectives, as well as providing the mechanisms to modify the behaviour of an agent and obtain a desired performance. This has been observed in imitation learning and safety critical control approaches where the behaviours of other agents are adapted based on some expert policies or reward functions [22]. However, both reward and control policies cannot provide interpretations by themselves and require the use of external explainable models, e.g., counterfactual explanations [23]–[25], LIME, decision tree [26], [27], and SHAP values [8]. In contrast, intended outcomes are explainable by design and has a similar structure to RL update rules. However, intended outcomes have been used as an external model that does not compromise the learning phase of RL agents. We are interested in how we can embed intended outcomes into the RL architecture such that the agent is capable to understand better the environment and adopt decisions based on interpretations.

The state-of-the-art is interested in modelling the reasoning process of agents to interpret actions or plans. The belief-desire-intention (BDI) model has been used to represent the knowledge and the reasoning process of an agent in a specific application or domain [28]–[30]. This model has been used in the design of games in e-learning platforms [31], [32]. Here, the BDI model is designed as a tutoring and learning strategy where the students are able to solve complex problems with the resources available in the game. This is a clear example of how interpretations and beliefs of the problem can be used to solve a complex problem. This paper is motivated in incorporating the BDI model into the RL paradigm for decision-making enhancement and interpretability.

B. Challenges and Contributions

Two main challenges are identified: (1)Explainable models are usually external modules that does not compromise the decision-making of RL agents. (2) Interpretations from policies or reward functions cannot provide simple explanations of the actions preference - critical component for the design of trustworthy autonomous systems.

In view of the above challenges, this paper provides the following contributions

1) A novel approach that uses intended outcomes as interpretable model to update and improve the learning phase of RL agents.
2) A natural belief driven model learning is proposed to emulate the reasoning process of humans. This improves the exploration of the environment, whilst maintaining fast convergence results.
3) Actions preference interpretations are improved by adjusting the belief maps in terms of the intended goal.

II. Belief-Desire-Intention (BDI) model

The BDI model aims to describe the reasoning process based on well-defined cognitive structures (see Fig. 1). The framework is composed of six elements: 1) knowledge of the world, 2) beliefs a virtual representation of the world, 3) desires activities or goals that represent the main objective of the application, 4) intentions refer the activity plans required to achieve the desire goals [33].

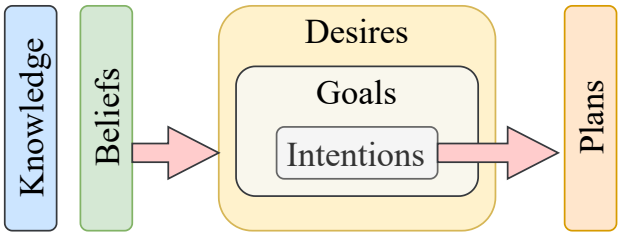


Fig. 1. Belief-Desire-Intention (BDI) model as reasoning process

The BDI model has been used by the research community to explain the decision-making process of Markov decision processes (MDPs) and machine learning algorithms, specifically in reinforcement learning. We can easily migrate from the abstract BDI model into a reinforcement learning setting by defining some analogies: i) knowledge of the world can be associated to the application task which covers state and action spaces or prior experiences. ii) Beliefs define what we observe from sensory data or a state transition probability mapping $P_{xu}^{x'} = p(x'|x,u)$. iii-iv) There can be multiple desires with specific goals encoded in a sparse reward function design, e.g., move an agent to a desired state whilst avoiding obstacles and satisfying time constraints. v-vi) Intention in a RL basis can be associated to the set of sequential actions (policy) required to optimise the reward function and complete the desire task [34]. However, the level of abstraction of the BDI model makes difficult to uncover the interplay between components, as well as the interpretation of the action plan. In this paper, we propose a novel BDI-RL framework where each element communicates and adapts from the new information gathered from the agent-environment interaction.

III. Basic Q-learning and Intended Outcomes

In this section we introduce some of the basic tools that are used in the development of the proposed BDI-RL framework.

A. Q-learning

The basic Q-learning algorithm is used as a model-free backbone model. The update rule of the Q-learning algorithm is given by

$$Q(x,u) \leftarrow (1-\alpha)Q(x,u) + \alpha\Big(r + \gamma \max_a Q(x',a)\Big), \quad (1)$$

where $\alpha \in [0,1)$ is the learning rate, $\gamma \in (0,1)$ is the discount factor, and $Q(x,u)$ is the Q-table in state x and action u. The intuition behind Q-learning is to greedily use the actions that maximise the Q-table in a specific state to complete a desired task. This allows to understand how Q-learning works in an algorithmic level, however it is unclear what it represents in the knowledge and belief spaces of a BDI model.

B. Belief maps

The concept of intended outcomes [18] has been used as a tool to interpret reinforcement learning decisions. The idea is to incorporate an additional process in the Q-learning update dedicated to learn belief maps $\boldsymbol{H}(x,u) \in \mathbb{R}^{\mathcal{X}\times\mathcal{U}}$, that is, a map of discounted expected states. The belief map update is similar to the Q-learning update rule given by

$$\boldsymbol{H}(x,u) \leftarrow (1-\alpha)\boldsymbol{H}(x,u) + \alpha\big(\mathbb{1} + \gamma \boldsymbol{H}\big(x', a_{\max}\big)\big), \quad (2)$$

where $a_{\max} = \arg\max_a Q(x',a)$. Here, the belief map allows to understand which it is the expected future state of selecting an action instead of other actions.

The use of belief maps adds a new explainable dimension to classic reinforcement learning algorithms. However, what is missing is how the belief map can be used to enhance learning capabilities of reinforcement learning algorithms using a similar mechanism to the human reasoning process, as well as the adaptive links of each learning element. To alleviate this issue and to maintain consistency with the BDI model, we propose a novel BDI-RL model that takes advantage of belief maps to update both the knowledge of the environment and plans to achieve the desired goals.

IV. Belief-Desire-Intention Reinforcement Learning (BDI-RL)

The proposed BDI-RL model is inspired in the tabular Dyna-Q algorithm [35]. This is a model-based reinforcement learning algorithm that includes two additional learning components to the classical Q-learning: 1) a model $M(x,u)$ that stores reward values and next state x' under state-action pair (x,u) and 2) a planning step that samples random next state x' and reward values r of the model $M(x,u)$ and use a simple Q-learning update rule to update the Q-table. This helps to accelerate convergence by combining experience with environment knowledge (similar to the BDI model).

The main drawback of Dyna-Q is that it uses previous visited state-action pairs to update the Q-table. This is not what we are looking for in terms of reasoning, since the learning of Dyna-Q is driven from the previous experiences rather than curiosity or beliefs such that it looks more as imitation.

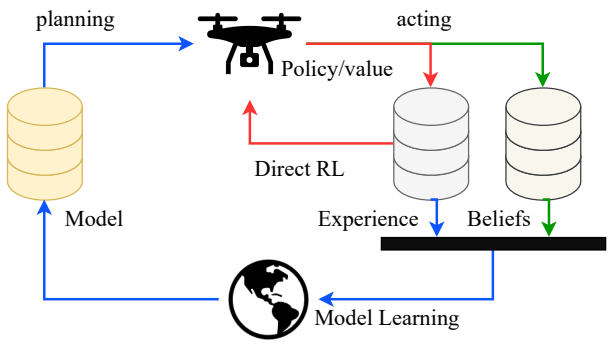


Fig. 2. BDI-RL high level scheme

To deal with the aforementioned issue, we proposed a BDI-RL model which has a similar structure to Dyna-Q, but incorporates belief maps that encourages the agent to adopt a belief driven learning. The proposed BDI-RL framework is depicted in Fig. 2.

The BDI-RL is composed of three closed-loops. The red-loop defines the standard reinforcement learning loop (in this case a Q-learning algorithm). The blue-loop is the standard Dyna-Q algorithm which incorporates the model learning as knowledge obtained by the agent-environment interaction. The green loop is our main contribution. Here, it is incorporated the belief maps as a learning mechanism to enhance the learning experience of the RL algorithm. In contrast to Dyna-Q that uses prior visited states to update the Q-table in N planning steps, the BDI-RL uses the expected states and actions of the belief map to add beliefs into the model learning step.

Algorithm 1 BDI-RL algorithm

Initialize $Q(x,u)$, $\mathcal{BM}(x,u)$, $\boldsymbol{H}(x,u)$ $\forall x \in \mathcal{X},\ u \in \mathcal{U}$
repeat {for each episode}
  $x \leftarrow$current (nonterminal) state
  $u \leftarrow \varepsilon$-greedy
  Apply action u; observe x' and reward r
  $Q(x,u) \leftarrow (1-\alpha)Q(x,u) + \alpha\big(r + \gamma \max_a Q(x',a)\big)$
  Select best action $a_{\max} = \arg\max_a Q(x',a))$
  $\boldsymbol{H}(x,u) \leftarrow (1-\alpha)\boldsymbol{H}(x,u) + \alpha\big(\mathbb{1} + \gamma \boldsymbol{H}(x', a_{\max})\big)$
  Obtain the expected action $a_e = \arg\max_a \boldsymbol{H}(x,a)$, expected state x_e and expected reward r_e
  Update belief model $\mathcal{BM}(x,a_e) \leftarrow x_e, r_e$
  for $i = 1$ **to** N **do**
    $u = \arg\max_a \mathcal{B}(x,a)$
    if $x, u \notin \mathcal{BM}(\cdot)$ **then**
      $x \leftarrow$ random previously observed state
      $u \leftarrow$ random expected action taken in x
    end if
    $x_e, r_e \leftarrow \mathcal{BM}(x,u)$
    $Q(x,u) \leftarrow (1-\alpha)Q(x,u) + \alpha\big(r_e + \gamma \max_a \mathcal{Q}(x_e,a)\big)$
    $x \leftarrow x_e$
  end for
until x' is terminal

From Fig. 2 we can observe that the proposed architecture is capable to encode the abstract BDI model of Fig. 1 into RL design, whose elements are continuously updated through the agent-environment interaction.

The BDI-RL uses the outputs of the belief map to construct a model. Specifically, in each iteration we compute the best action that maximizes the belief map, i.e., $a_e = \arg\max_a \boldsymbol{H}(x,a)$. Also from the belief map $\boldsymbol{H}(x,u)$ we can obtain the next expected state $x_e \in \mathcal{X}$ by applying the expected action $a_e \in \mathcal{U}$ and receive a expected reward $r_e(x,a_e,x_e) \in \mathbb{R}$. Then, a belief model $\mathcal{BM}(\cdot)$ is constructed and updated as $\mathcal{BM}(x,a_e) \leftarrow x_e, r_e$.

We measure the quality of our beliefs by incorporating the

following belief driven exploration function,

$$\mathcal{B}(x, a_e) = \begin{cases} \|x - x^*\|^2 & \text{if } \mathcal{BM}(x, a_e) \neq 0, \\ -\infty & \text{otherwise,} \end{cases} \quad (3)$$

The belief function (3) will weight more the expected actions that yield to a state x_e that is closer to the desired goal x^*. So, it is required to evaluate all possible actions in a specific state to determine the best expected action. Two cases are considered: 1) $x, a \in \mathcal{BM}(\cdot)$ and 2) $x, a \notin \mathcal{BM}(\cdot)$, where $a \in \mathcal{U}$ is any action in the action space.

The first case, i.e., when $x, a \in \mathcal{BM}(\cdot)$ refers to expected actions that we are more confident about its outcomes. Here, the sampling step of Dyna-Q is modified into a greedy sampling given by

$$\begin{aligned} u &= \arg\max_a \mathcal{B}(x, a), \\ x_e, r_e &\leftarrow \mathcal{BM}(x, u). \end{aligned} \quad (4)$$

Conversely, the second case, i.e., $x, a \notin \mathcal{BM}(\cdot)$ means that the action a has not been applied by the agent in the specific state x, such that we need to improve our beliefs using an inner exploration. In this case, the sampling step of Dyna-Q is modified into a belief sampling given by

$$\begin{aligned} x &\leftarrow \text{random previous observed } x \text{ of } \mathcal{BM}(\cdot) \\ u &\leftarrow \text{random expected action } a_e \text{ of } \mathcal{BM}(\cdot) \\ x_e, r_e &\leftarrow \mathcal{BM}(x, u). \end{aligned} \quad (5)$$

Notice that the second case can produce discontinuous state transitions. However these discontinuities are exclusively encoded in the planning phase to obtain a better knowledge of the environment. In other words, the belief sampling is not reflected into real actions. Then, we use the same process as Dyna-Q by computing the following Q-learning update rule in N planning steps

$$Q(x, u) \leftarrow (1 - \alpha)Q(x, u) + \alpha\big(r + \gamma \max_a \mathcal{Q}(x_e, a)\big), \quad (6)$$

where $\mathcal{Q}(x, a)$ is the Q-table generated from the belief map.

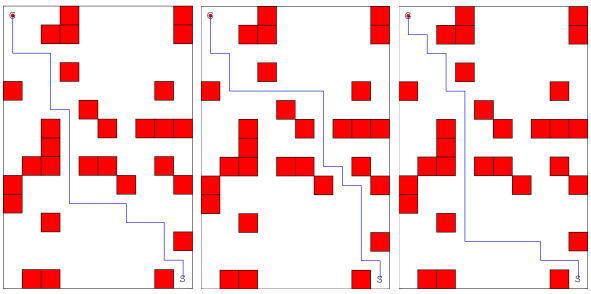
V. Simulation Studies

In this section, we evaluate the performance of the proposed BDI-RL under different classic environments and compare it against the traditional Q-learning and Dyna-Q algorithms.

A. Simple Grid Environment

Consider a user-defined grid environment of 10×15 cells. The action space consists in the typical four actions: up, right, down, and left. Here, the starting point is at the right lower corner of the grid in the point $x_0 = [9, 0]^\top$, whilst the goal is at the left upper corner of the grid in the point $x_f = [0, 14]^\top$. The blocks observed in the grid are obstacles that the agent needs to avoid to reach to the desired goal. The agent receives a reward of $r = -1$ each time step and a reward of $r = 10$ when it reaches the goal.

The same hyperparameters are used in all algorithm to ensure fair comparisons. The hyperparameters used in the experiments are: $\alpha = 0.1$, $\gamma = 0.95$, $\varepsilon = 0.9$, ε-decay$= 0.99$, planning_steps=20, iterations per episode $k = 2000$, and number of episodes $i = 500$. The trajectory followed by each agent is shown in Fig. 3. The total reward for each RL agent is shown in Fig. 4. Notice that each agent follows a different path that yields the same cumulative reward. After convergence, each agent reaches to the goal point in 23 iterations.



(a) Q-learning (b) Dyna-Q (c) BDI-RL

Fig. 3. 10×15 grid solved by RL agents

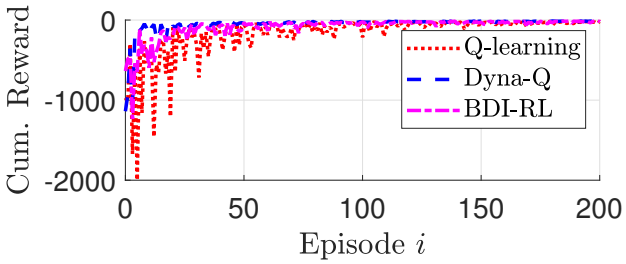
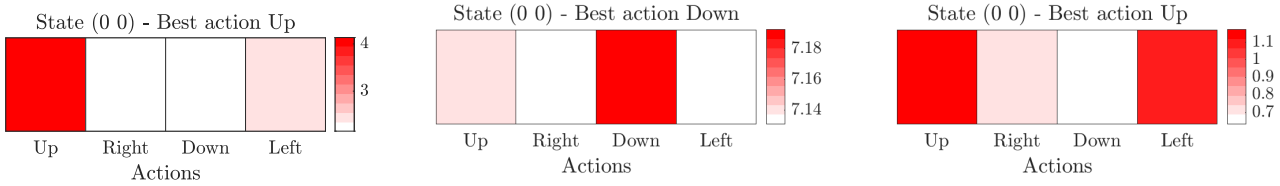


Fig. 4. Cumulative Reward of the RL gents: Simple grid

Notice that Q-learning needs more episodes to converge to a near optimal value. Conversely, Dyna-Q outperforms Q-learning and slightly the proposed BDI-RL algorithm. One can think, that Dyna-Q offers better results in terms of learning convergence. However, we will observe that this premise is not true by using the belief map $\boldsymbol{H}(x, u)$.

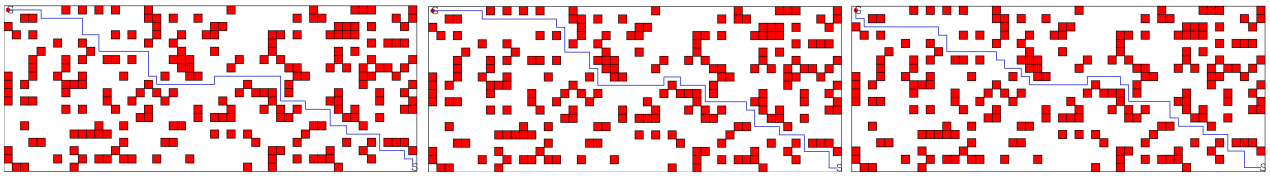
For visibility purposes, consider the state $(0, 0)$ located at the left lower corner of the grid. Fig. 5 shows the action preference of each RL agent. Notice that for this specific state different belief maps are obtained. On the one hand, the BDI-RL effectively selects the best action that can drive the agent to the desired goal. Here, the belief element plays a pivotal role to complete the exploration of the grid. The belief map of the Q-learning algorithm shows that the Q-learning algorithm fails to select the best action since it selects a forbidden action (the agent cannot move down in this state). This clearly shows that Q-learning, despite it reaches to the desired goal, it does not explore infinitely often the complete grid. Dyna-Q shows a lack of confidence in the action selection at state $(0, 0)$, this is caused due to the use of the model constructed from past experiences that causes to converge faster, but with the compromise of not exploring the grid completely.

From these results we can observe that the BDI-RL offers a better trade-off between exploration and exploitation by incorporating belief maps as a belief driven mechanism.



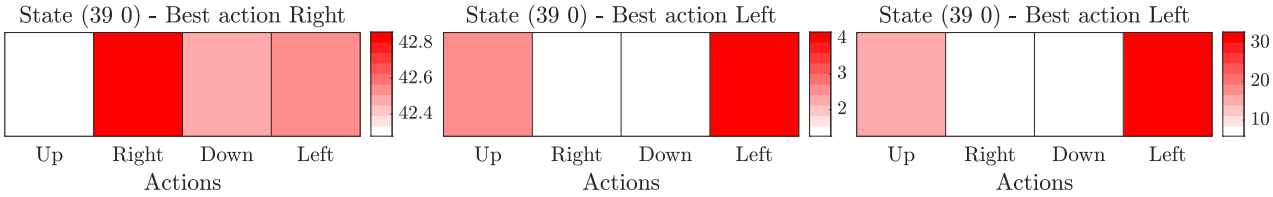
(a) BDI-RL (b) Q-learning (c) Dyna-Q

Fig. 5. Interpretation of action preference of each RL agent



(a) Q-learning (b) Dyna-Q (c) BDI-RL

Fig. 6. 50×20 grid solved by RL agents



(a) Q-learning (b) Dyna-Q (c) BDI-RL

Fig. 7. Interpretation of actions preference

In contrast to previous interpretable reinforcement learning models that uses the belief map as an independent system, the BDI-RL creates a natural mapping from beliefs to planning that enhances the exploration and model creation.

B. Complex grid environment

To show the robustness and challenges of the proposed approach, we test the RL agents (Q-learning, Dyna-Q, BDI-RL) in a relatively larger and complex grid of dimension 50×20 with the same action space.

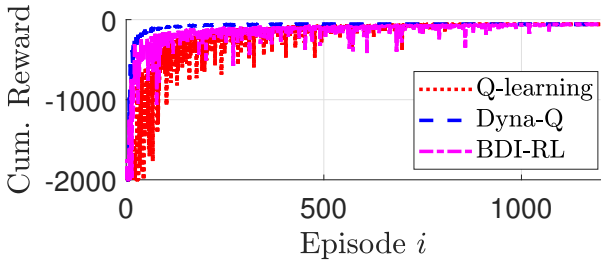


Fig. 8. Cumulative Reward of the RL agents: Complex grid

The starting point is $x_0 = [49, 0]^\top$ and the goal point is $x_f = [0, 19]^\top$. The hyperparameters used in the experiments are: no. of episodes $i = 1000$, steps per episode $k = 2000$, planning_steps $= 100$, learning rate $\alpha = 0.5$, discount factor $\gamma = 0.95$, and an exploration factor of $\varepsilon = 0.9$.

Fig. 8 shows the cumulative reward results of each RL agent under the new complex grid. Similarly to the simple grid, Q-learning needs more episodes to obtain a relative good solution until the ε-greedy algorithm starts to only exploiting previous knowledge. Dyna-Q outperforms all models by having a fast convergence to the solution by exploiting the model created from previous visited state-action pairs. The BDI-RL shows a slightly slower convergence performance compared to Dyna-Q. This is caused due to exploration enhancement using the belief function.

Fig. 6 shows the obtained results at episode 1,000. Notice that each agent follows a different path, which makes evident the need of interpretations to uncover why the agent follows a different action plan. All the agents reach the goal point in 70 steps. Fig. 7 shows the final belief map in a random testing point $[39, 0]^\top$. As expected, the Q-learning algorithm fails to cover the environment effectively such that the expected action at the testing state yields to an invalid state (obstacle). The belief maps of both Dyna-Q and BDI-RL select the best action in the testing state, however we can observe that Dyna-Q explores few times the state. Conversely, the BDI-RL explores more the testing state such that the action selection confidence is improved.

VI. Conclusions

In this paper a belief-desire-intention reinforcement learning (BDI-RL) architecture is proposed. The approach is inspired in the human reasoning process that includes beliefs in the action selection of the decision-making process. The algorithm uses belief maps as interpretable model to modify

the nature of the classical Dyna-Q algorithm. Here, the model of Dyna-Q is modified into a belief model that encourages the exploration of the environment. This accelerates the convergence of the algorithm, whilst enhancing the action selection interpretation. Convergence of the proposed model is assessed using the contraction property of Q-tables. Simulations experiments are conducted under grid environments with different sizes and obstacles. The proposed approach achieves competitive results with enhanced actions selection interpretability which is a crucial element for the design of trustworthy autonomous systems. Dyna-Q converges faster than the proposed BDI-RL approach, however it exhibits poor interpretations capabilities and confidence for the action selection of paths with low exploration. In contrast, the proposed BDI-RL shows enhanced interpretation capabilities with a minor cost of convergence time. This trade-off is acceptable due to the new benefit provided by this approach.

Future work will consider the extension of the approach for large and continuous state-action spaces, as well as including new interpretable models to cover other RL architectures.

REFERENCES

[1] J. Xu, D. Panagopoulos, A. Perrusquía, W. Guo, and A. Tsourdos, "Generalising rescue operations in disaster scenarios using drones: A lifelong reinforcement learning approach," *Drones*, vol. 9, no. 6, p. 409, 2025.

[2] C. Gruffeille, A. Perrusquía, A. Tsourdos, and W. Guo, "Disaster area coverage optimisation using reinforcement learning," in *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2024.

[3] G. Lykou, D. Moustakas, and D. Gritzalis, "Defending airports from uas: A survey on cyber-attacks and counter-drone sensing technologies," *Sensors*, vol. 20, no. 12, p. 3537, 2020.

[4] A. Singh, G. Triulzi, and C. L. Magee, "Technological improvement rate predictions for all technologies: Use of patent data and an extended domain description," *Research Policy*, vol. 50, no. 9, p. 104294, 2021.

[5] M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

[6] K. KN, A. Perrusquia, and D. Ignatyev, "Comprehensive exploration and development of explainable ai for robust aircraft landing gear fault diagnosis," in *AIAA SCITECH 2025 Forum*, 2025, p. 1928.

[7] C. Glanois, P. Weng, M. Zimmer, D. Li, T. Yang, J. Hao, and W. Liu, "A survey on interpretable reinforcement learning," *Machine Learning*, pp. 1–44, 2024.

[8] X. Kong, Y. Xing, A. Tsourdos, Z. Wang, W. Guo, A. Perrusquia, and A. Wikander, "Explainable interface for human-autonomy teaming: A survey," *arXiv preprint arXiv:2405.02583*, 2024.

[9] A. Mott, D. Zoran, M. Chrzanowski, D. Wierstra, and D. Jimenez, "Towards interpretable reinforcement learning using attention augmented agents," *Advances in neural information processing systems*, 2019.

[10] A. Verma, V. Murali, R. Singh, P. Kohli, and S. Chaudhuri, "Programmatically interpretable reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2018, pp. 5045–5054.

[11] J. Mugabe, M. Wisniewski, A. Perrusquía, and W. Guo, "Enhancing situational awareness of helicopter pilots in unmanned aerial vehicle-congested environments using an airborne visual artificial intelligence approach," *Sensors*, vol. 24, no. 23, p. 7762, 2024.

[12] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[13] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, "Evaluating the visualization of what a deep neural network has learned," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 11, pp. 2660–2673, 2016.

[14] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[15] J. Ramírez, W. Yu, and A. Perrusquía, "Model-free reinforcement learning from expert demonstrations: a survey," *Artificial Intelligence Review*, vol. 55, pp. 3213–3241, 2022.

[16] M. Zou, A. Perrusquía, and W. Guo, "Explaining data-driven control in autonomous systems: A reinforcement learning case study," in *International Conference on Control, Decision and Information Technologies (CoDIT'24)*, 2024.

[17] R. M. Annasamy and K. Sycara, "Towards better interpretability in deep q-networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 4561–4569.

[18] H. Yau, C. Russell, and S. Hadfield, "What did you think would happen? explaining agent behaviour through intended outcomes," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[19] S. Choi, J. Kim, and H. Yeo, "Trajgail: Generating urban vehicle trajectories using generative adversarial imitation learning," *Transportation Research Part C: Emerging Technologies*, vol. 128, p. 103091, 2021.

[20] A. Perrusquia and W. Guo, "Drone's objective inference using policy error inverse reinforcement learning." *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[21] J. Sun, L. Yu, P. Dong, B. Lu, and B. Zhou, "Adversarial inverse reinforcement learning with self-attention dynamics model," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1880–1886, 2021.

[22] A. Perrusquía and W. Guo, "Uncovering reward goals in distributed drone swarms using physics-informed multiagent inverse reinforcement learning," *IEEE Transactions on Cybernetics*, 2024.

[23] H. Andreas, M. Armgardt, and M. Gunther, "Counterfactuals for causal responsibility in legal contexts," *Artificial intelligence and law*, vol. 31, no. 1, pp. 115–132, 2023.

[24] T. Huber, M. Demmler, S. Mertes, and M. Olson, "GANterfactual-RL: Understanding reinforcement learning agents' strategies through visual counterfactual explanations," *arXiv preprint arXiv:2302.12689*, 2023.

[25] A. Perrusquía, M. Zou, and W. Guo, "Explainable data-driven q-learning control for a class of discrete-time linear autonomous systems," *Information Sciences*, vol. 682, p. 121283, 2024.

[26] R. Paleja, L. Chen, Y. Niu, A. Silva, Z. Li, S. Zhang, C. Ritchie, S. Choi, K. C. Chang, H. E. Tseng *et al.*, "Interpretable reinforcement learning for robotics and continuous control," *arXiv preprint arXiv:2311.10041*, 2023.

[27] C. Zhao, C. Deng, Z. Liu, J. Zhang, Y. Wu, Y. Wang, and X. Yi, "Interpretable reinforcement learning of behavior trees," in *Proceedings of the 2023 15th International Conference on Machine Learning and Computing*, 2023, pp. 492–499.

[28] A.-H. Tan, Y.-H. Feng, and Y.-S. Ong, "A self-organizing neural architecture integrating desire, intention and reinforcement learning," *Neurocomputing*, vol. 73, no. 7-9, pp. 1465–1477, 2010.

[29] A.-H. Tan, Y.-S. Ong, and A. Tapanuj, "A hybrid agent architecture integrating desire, intention and reinforcement learning," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8477–8487, 2011.

[30] D. Sacharny, T. C. Henderson, M. Cline, and B. Russon, "Reinforcement learning at the cognitive level in a belief, desire, intention uas agent," in *International Conference on Intelligent Autonomous Systems*. Springer, 2021, pp. 458–470.

[31] L. M. Giraffa, M. Móra, and A. Zamberlam, "Working on student models (really) based on mental states," in *Proceedings Fourth International Conference on Computational Intelligence and Multimedia Applications. ICCIMA 2001*. IEEE, 2001, pp. 379–383.

[32] L. Braubach, A. Pokahr, and W. Lamersdorf, "Jadex: A BDI-agent system combining middleware and reasoning," in *Software agent-based applications, platforms and development kits*. Springer, 2005.

[33] M. Georgeff, B. Pell, M. Pollack, M. Tambe, and M. Wooldridge, "The belief-desire-intention model of agency," in *Intelligent Agents V: Agents Theories, Architectures, and Languages: 5th International Workshop*. Springer, 1999.

[34] A. Perrusquía, W. Guo, B. Fraser, and Z. Wei, "Uncovering drone intentions using control physics informed machine learning," *Communications Engineering*, vol. 3, no. 1, p. 36, 2024.

[35] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.



https://dspace.org/

Interpreting and enhancing decisions in autonomous navigation: a belief-desire-intention reinforcement learning (BDI-RL) approach

Perrusquía, Adolfo; Panda, Deepak Kumar; Guo, Weisi

2025-10-05
Institute of Electrical and Electronics Engineers (IEEE)
https://doi.org/10.1109/smc58881.2025.11342492
Perrusquía A, Panda D, Guo W. (2025) Interpreting and enhancing decisions in autonomous navigation: a belief-desire-intention reinforcement learning (BDI-RL) approach. In: Proceedings of the 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 5-8 Oct 2025, Vienna, Austria, pp. 5643-5648

Downloaded from a DSpace repository. https://dspace.org/