Supplementary material to the article Identification of Unexpected Decisions in Partially Observable Monte-Carlo Planning: a Rule-Based Approach

Giulio Mazzi, Alberto Castellini, Alessandro Farinelli
Dipartimento di Informatica, Università degli Studi di Verona, Verona, Italy

1 Analysis of a specific trace

We show the results of XPOMCP on the analysis of a specific trace generated by POMCP using a wrong value of W (i.e., $W = 40$). As shown in Table 1, POMCP selects several wrong actions with this value of W. The rule generated by the MAX-SMT solver from this trace is:

$$\begin{aligned} &r_L : \texttt{select}\ Listen\ \texttt{when}\ (p_{right} \leq 0.847 \wedge p_{left} \leq 0.847); \\ &r_{OR} : \texttt{select}\ Open_R\ \texttt{when}\ p_{right} \geq 0.966; \\ &r_{OL} : \texttt{select}\ Open_L\ \texttt{when}\ p_{left} \geq 0.966; \end{aligned} \tag{1}$$

It is a compact summary of the policy that highlights the important details in a structured way. There is a gap between the value of the action rule r_L (i.e., 0.847) and that of the action rules r_{OR}, r_{OL} (i.e., 0.966). This is because the trace does not contain any belief in this gap and XPOMCP cannot build a rule to describe how to act in this belief interval. Among the total 2659 trace steps, 1601 satisfy the rule and 1058 does not satisfy it. For all steps not satisfying the rule, we computed their Hellinger distance using the procedure described in the article. The H^2 distance values are displayed in descending order in Figure 1.

To classify unexpected decisions, we use the optimal threshold of $\tau = 0.045$ (Section 4.2 of the main paper presents the process used to compute the optimal threshold). We colored in green steps having a H^2 distance below the threshold and in red steps with H^2 above the threshold. Notice that only the steps of the second group (which contains 637 unsatisfiable steps in total) are considered unexpected decisions by our approach. Three main levels of H^2 can be observed in the chart, one close to zero, one around 0.15, and the third around 0.4. We have selected one step for each level (see points p_1, p_2 and p_3) and shown the related belief probabilities, action, and H^2 distance, to characterize the errors made by POMCP in those steps. This analysis shows that in step p_2 POMCP decides to open the left door with a probability 0.860 to have the treasure behind that door, while the rule says that a probability of at least 0.966 is needed to select this action. In step p_3 the error is even more evident since the agent opens the left door with a probability of only 0.5 to find the treasure there. On the other hand in step, p_1 the agent listens with a probability 0.8471 which is slightly higher than the minimum value (i.e., 0.8470) needed to perform that action. Notice that this belief is not assigned to any action rule and it is very close to the boundary of the listen rule, hence it makes sense to assign this step to that rule and not to consider it as an exception. In this case, we achieve an F1-score of 1.0 and an accuracy of 1.0 using the proposed approach, and it takes XPOMCP $12,509s$ to compute this solution. This confirms the ability to detect unexpected decisions even when state-of-the-art anomaly detection methods have degraded performance (i.e., IF reaches an F1-score of 0.59 in this test, but it only takes $0,791s$ to do so).

Table 1: Quantitative performance comparison (AUC and AP) with uniform sampling of thresholds and different values of W

W	% errors	XPOMCP AUC	XPOMCP AP	IF AUC	IF AP
110	0.0(±0.0)	–	–	–	–
85	0.0004(±0.0003)	**0.993**(±0.041)	**0.986**(±0.082)	0.964(±0.024)	0.057(±0.1076)
65	0.0203(±0.0021)	**0.999**(±0.001)	**0.999**(±0.002)	0.992(±0.001)	0.539(±0.0520)
40	0.2374(±0.0072)	**0.995**(±0.034)	**0.987**(±0.084)	0.675(±0.020)	0.333(±0.0153)

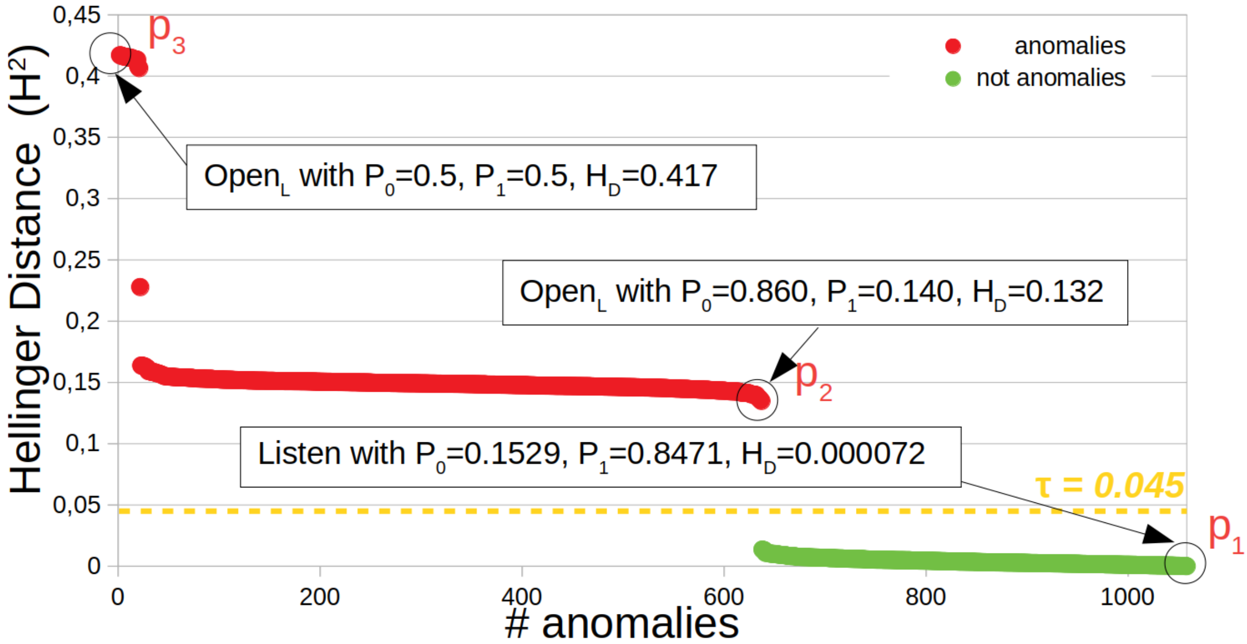

Figure 1: Values of Hellinger distance of steps that fail to satisfy the rule for *Tiger*.

2 Larger t-SNE Figure

In Section 4.3 of the main paper, we use *T-distributed Stochastic Neighbor Embedding (t-SNE)* to present in a two dimensions image belief of the *velocity regulation* problem (that is a probability distribution over 3^8 states). To improve the readability of this dense image, we provide a higher resolution version in Figure 2. Figure 3 provides the same t-SNE but it does not contain any extra annotation. We also include the two images as *png* files in the supplementary material.

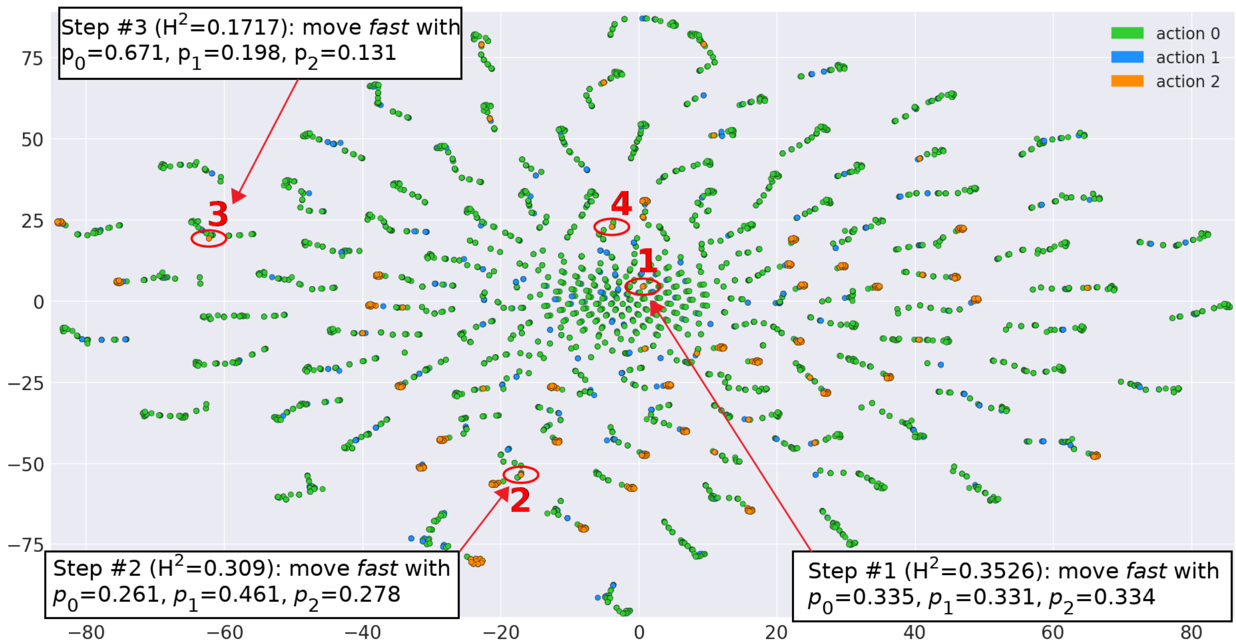

Figure 2: An higher resolution version of Figure 6 of the main paper

3 XPOMCP implementation

In the `code` folder of the supplementary material we provide an implementation of the XPOMCP methodology (in the `xpomcp` subfolder), a single trace that can be used to test XPOMCP (`example_velreg_trace`),

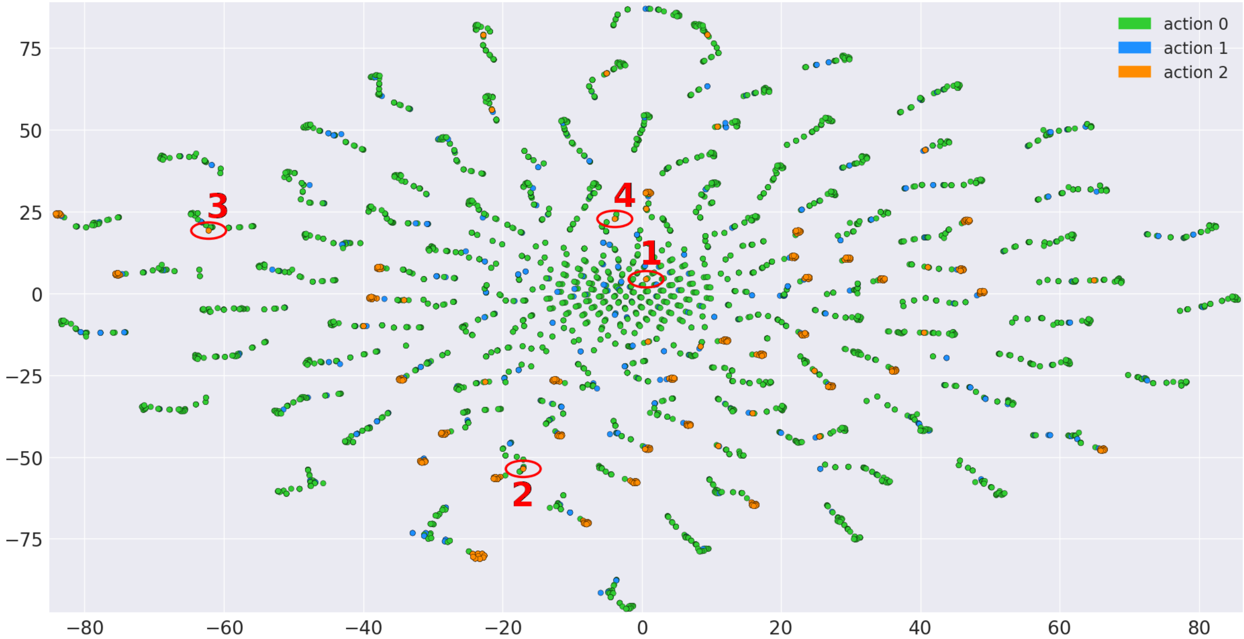

Figure 3: The t-SNE without any extra annotation

and an implementation of pomcp that can be use to generate more traces (in subfolder `pomcp`). The subfolders contain a `README.md` file that explains how to use them.

We presents the result of XPOMCP using the same rule presented in Section 4.3 of the main paper but a smaller trace. The ouput is:

```
Import experiments
Find maximum number of satisfiable step in rule 0
fail to satisfy 3 steps out of 350
Check Formulas
rule: go at speed 2 if: P_0 >= 0.900 OR P_2 <= 0.017 OR
                        (P_0 >= 0.714 AND P_1 >= 0.113)
Unsatisfiable steps:
ANOMALY: run example_velreg_trace/Run_5 step 0:
         action 2 with belief P_0 = 0.335 P_1 = 0.331 P_2 = 0.334
         --- Hellinger = 0.27355734175747 36

run example_velreg_trace/Run_9 step 12:
action 2 with belief P_0 = 0.678 P_1 = 0.228 P_2 = 0.094
--- Hellinger = 0.029123804891139585

run example_velreg_trace/Run_1 step 26:
action 2 with belief P_0 = 0.880 P_1 = 0.099 P_2 = 0.022
--- Hellinger = 0.017102605324761978
```

The output report the number of steps that we fail to satisfy (3 out of 350 in this case), the final *rule* (i.e., an instantiation of the rule template provided in Section 4.3 of the main paper with the number of the trace), and a list of the unsatisfiable steps. Only the first one is identified as an anomaly (using a threshold $\tau = 0.1$ as in the main paper).