Adaptation from Sparse Observations through Detection and Correction of Model Anomalies

(Doctoral Consortium)

Juan Pablo Mendoza
Supervisors: Manuela Veloso and Reid Simmons
The Robotics Institute, Carnegie Mellon University
5000 Forbes Avenue. Pittsburgh, PA, 15213
jpmendoza@ri.cmu.edu

ABSTRACT

To make intelligent decisions, robots often use models of the effects of their actions on the world. Unfortunately, it is often infeasible to have the perfect knowledge and computational resources required to create globally accurate models. This may lead to divergence between planned and actual execution, often resulting in suboptimal task performance. We propose an execution monitoring approach that enables robots to detect unmodeled modes of the system and correct their models accordingly. In particular, we address the problem of finding and adapting to *regions of a state-action feature space* in which outcomes of actions observed during execution deviate from the *stochastic expectations* used to select those actions. We evaluate our approach in the adversarial domain of autonomous robot soccer.

Categories and Subject Descriptors

I.2.9 [**Robotics**]

Keywords

Multi-robot systems; Robot planning and plan execution; Single and multi-agent learning; Fault tolerance and resilience.

1. BACKGROUND

The problem of adapting intelligently to the unavailability of perfect models has been a widely studied problem.

One approach to improving robot performance without relying on perfect models of the world is to do away with models entirely [1]. These methods allow robots to improve task performance based on an observed reward only, without need to explicitly model the effects of actions. While this generality is appealing and necessary in situations where any modeling is impractical, model-free learning tends to be less data-efficient and is not generalizable to different tasks [2]. We wish to improve performance by taking advantage domain knowledge through model learning.

System identification techniques focus on learning a model, out of a candidate set, that best fits the data [3]. While we are also interested in learning to correct models based on observations, we focus on domains in which the available model accurately describes most situations, but is inaccurate in others. That is, we want to detect situations in which the available model is anomalously inaccurate, and correct it. Focusing on this more specific problem enables our approach to use resources more efficiently, as learning efforts can focus on the anomalous situations.

At the core of our algorithms is the detection of unmodeled modes of behavior through anomaly detection techniques [4]. In particular, our algorithms find statistical anomalies in collections of data that is spatially related in some space of features of state-action space.

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

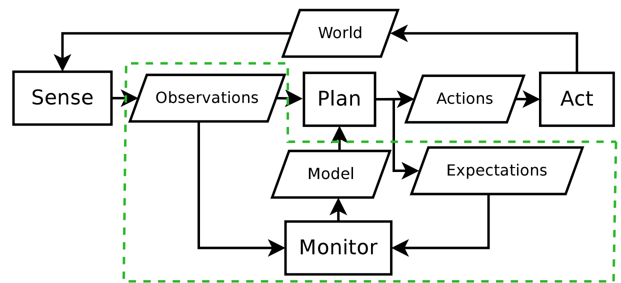

Figure 1: Framework for planning and execution. We focus on the elements within the dashed line: the monitor uses expectations generated by the planner, and observations from the world, to improve the model used for planning.

2. APPROACH

The high level representation of our completed framework is shown in Figure 1. Our focus has been on the *Monitoring* component of the framework, and its interactions, through inputs and outputs, with the *Planning* component. This interaction can be summarized in the following looping steps:

1. Given a model, create a plan to perform the desired task, along with model-based *expectations* about the results of those actions.
2. During execution, monitor whether the expectations were met in the real world by comparing them to *observations* obtained from sensing.
3. Find the conditions in which the real world is not well

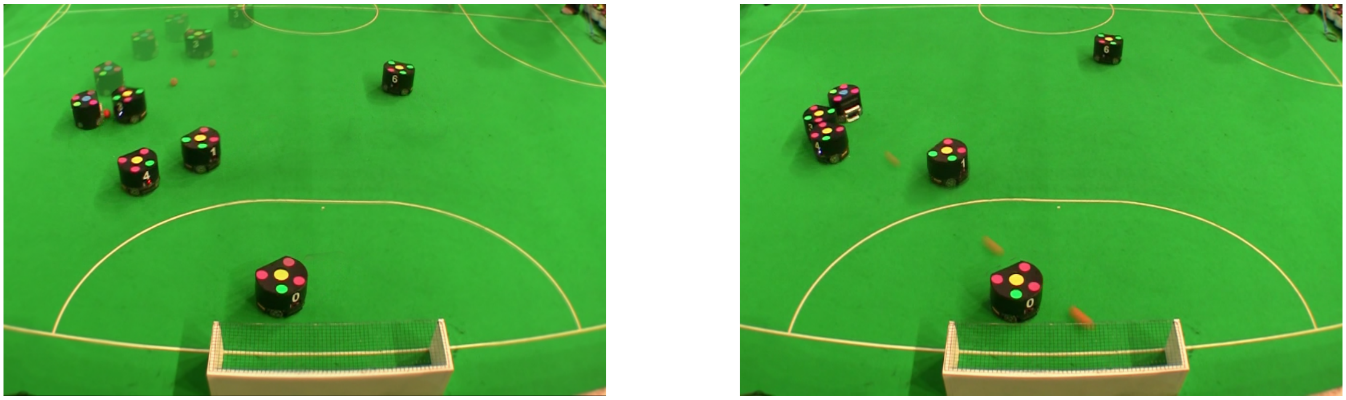

Figure 2: Goal scored by CMDragons during the quarter finals of RoboCup 2013. Accurate ball interception models of itself and of opponents allows our single robot (blue center dot) to gain control of a moving loose ball among several opponents (yellow center dots) to be able to score. (Thanks to J. Biswas [8] for the ball interception algorithm.)

represented by the model, as *regions of a feature space* in which observations deviate from expectations.

4. Based on such findings, make corrections to the model in situations determined to be inadequately modeled.

We have defined a planning framework that, apart from generating sequences of actions, also generates *expectations* $\boldsymbol{\theta}$ about such actions. During execution, these expectations are monitored by measuring the likelihood of observed values $\boldsymbol{z}$ given the expected behavior $P(\boldsymbol{z}|\boldsymbol{\theta})$.

To find the conditions in which execution is poorly modeled, we have developed the FARO anomaly detection algorithm [5, 6], which finds regions of state-action feature space of anomalous behavior using optimization techniques. Briefly, FARO takes as input a set of data points of the form $(\boldsymbol{s}_i, \boldsymbol{z}_i, \boldsymbol{\theta}_i)$, where $\boldsymbol{s}_i \in S$ is a point in state-action feature space, $\boldsymbol{z}_i \in \boldsymbol{Z}$ is an observation made at that point, and $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}$ is the expected distribution of $\boldsymbol{z}_i$; from this, FARO finds a set of regions $\mathcal{R} \subseteq S$ in which the observed data is highly unlikely given the expected distribution. It does so by conducting a parallel optimization on a few promising parameterized regions of feature space S (e.g., ellipsoids), in order to find the one most likely to be a statistically significant anomaly in behavior.

Once the unmodeled modes of a system are autonomously found, a new model must be created for the planner to be able to predict action outcomes in such modes accurately. At its core, this is a function approximation problem (we wish to approximate the function defining the outcomes of actions). We therefore approximate the model in the detected anomalous subspace using maximum likelihood parametric models derived from the observations in such subspaces [6].

So far, we have addressed the problem of passively detecting and correcting anomalies in robot planning models. For the remainder of the thesis, we will address the problem of active exploration to find potential anomalies in action outcome predictions. Robots have the ability to influence the way in which data points are sampled from its space. There may be situations in which sampling from suboptimal regions to gain more information would be more beneficial in the long term performance. The exploration-exploitation trade-off has been studied [7], but we wish to explore it in the context of detection of unmodeled modes of a system.

3. ROBOT DOMAIN

We demonstrate our work on the the CMDragons [8] team of soccer robots. The adversarial nature of this domain makes it an ideal testbed for our algorithms: Usually, our robots only have incomplete information about the opponent's strategy and capabilities. While we may be able to create a generally accurate models of opponent behavior, they are likely to be poor predictors of opponent behavior in some situations. Our goal is to enable these robots to actively detect, during the span of a soccer game, that some of their actions are not having the expected results, and to improve the predictions about these results, without needing to understand why their expectations are not being met.

REFERENCES

[1] C. Gaskett, D. Wettergreen, and A. Zelinsky, "Q-learning in continuous state and action spaces," in *Australian Joint Conference on Artificial Intelligence.* Springer-Verlag, 1999, pp. 417–428.

[2] C. G. Atkeson and J. C. Santamaria, "A comparison of direct and model-based reinforcement learning," in *Proceedings of ICRA*, 1997.

[3] L. Ljung, Ed., *System Identification (2Nd Ed.): Theory for the User.* Prentice Hall PTR, 1999.

[4] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, pp. 1–72, September 2009.

[5] J. P. Mendoza, M. Veloso, and R. Simmons, "Focused optimization for online detection of anomalous regions," in *Proceedings of ICRA*, Hong Kong, China, June 2014.

[6] ——, "Detecting and correcting model anomalies in subspaces of robot planning domains," in *Proceedings AAMAS (to appear)*, Istambul, Turkey, May 2015.

[7] J.-Y. Audibert, R. Munos, and C. Szepesvari, "Exploration-exploitation tradeoff using variance estimates in multi-armed bandits," *Theoretical Computer Science*, vol. 410, no. 19, pp. 1876 – 1902, 2009, algorithmic Learning Theory.

[8] J. Biswas, J. P. Mendoza, D. Zhu, B. Choi, S. Klee, and M. Veloso, "Opponent-driven planning and execution for pass, attack, and defense in a multi-robot soccer team," in *Proceedings of AAMAS*, January 2014.