Two Perspectives on Representation Learning

Joseph Modayil

Reinforcement Learning and Artificial Intelligence Laboratory
University of Alberta

Reasoning & Learning:

Two perspectives on knowledge representation

- For reasoning with a model:
  - Expressiveness of the model (e.g. space, objects, ...)
  - Planning with the model is useful for a robot
- For learning to predict the consequences of a robot's behaviour:
  - Semantics defined by the robot's future experience
  - Online, scalable learning during normal robot operation

An Analogy with Scientific Knowledge

- Reasoning and learning have complementary strengths that are analogous to scientific theories and experiments.
  - Scientific theories enable broad generalization within a limited domain. Scientific theories enable effective reasoning even when inaccurate.
  - Experiments measure the world without needing model assumptions. Many experiments are needed to understand the world.
- Two approaches for connecting theories and experiments.
  - Top-down: Theories have experimentally verifiable predictions.
  - Bottom-up: Many verifiable predictions can generalize to a single theory.
    - Note: A single prediction a (very) partial model of the world.

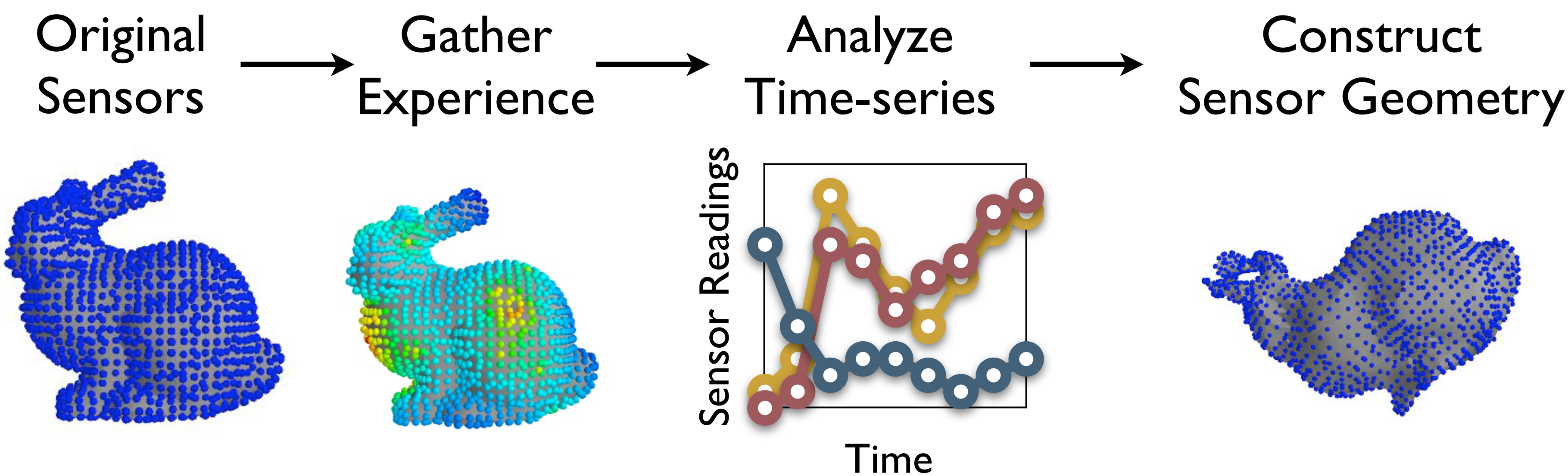
Rich representations that support reasoning

Reasoning with rich representations

- Useful analogs to human-scale abstractions can be constructed from robot experience.
  - The robot constructs models from its sensorimotor experience by searching for particular statistical structures.
  - The models describe spaces and objects.
  - The robot reasons within these models to achieve goals.

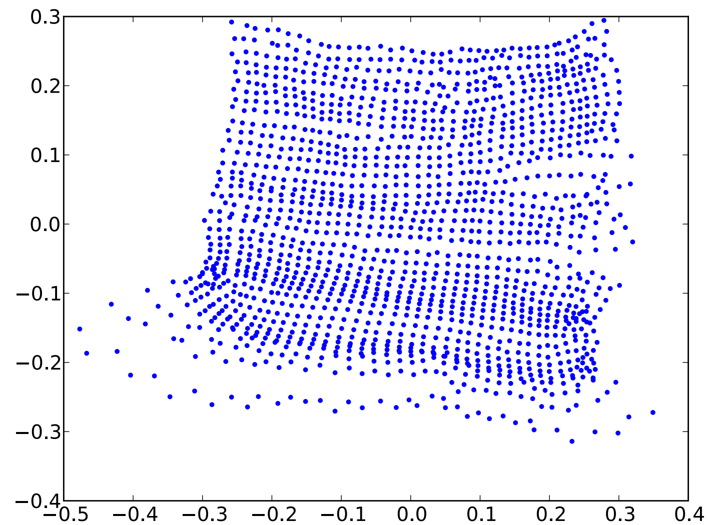
Representing sensor configurations (Modayil, 2010)

- Sensors in similar physical configurations yield highly correlated time-series data. (e.g. GP assumption)
- Invert this: use time-series data to construct a manifold of sensor configurations.



Learned geometry from real robot data

Cosy Localization Database



Method:

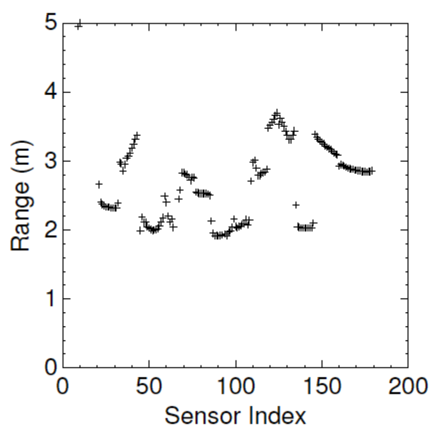
1. Define local distances between strongly correlated sensors
2. Use the fast maximum variance unfolding algorithm to construct a manifold

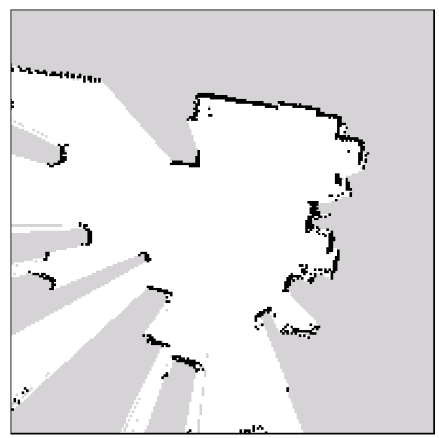
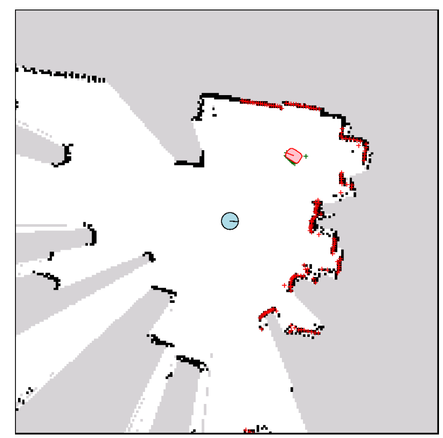
Conclusion: A robot's experience can contain enough information to recover approximate local sensor geometry (and perhaps global geometry).

Representing Objects

(Modayil & Kuipers, 2007)

- Intuition: Moving objects can be distinguished from a static world.
- Approach: Use violations of a stationary background model to perceive moving objects.

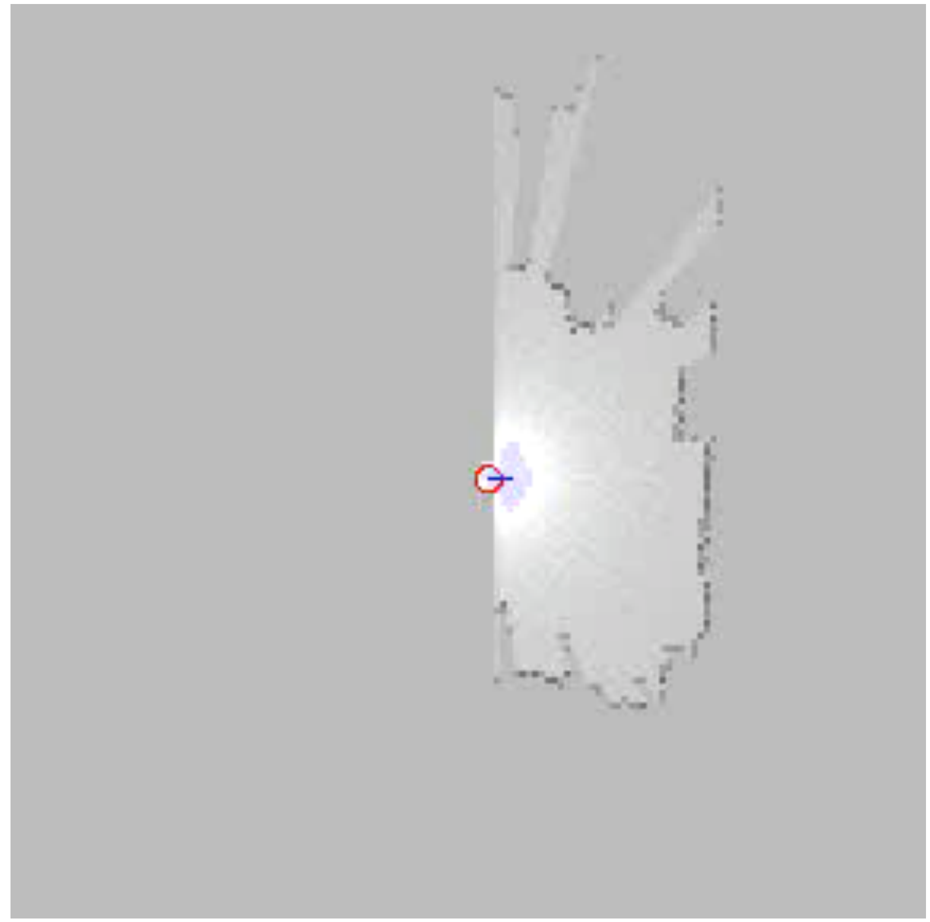



Objects:
Background Model

The agent has a model of the static environment

- Occupancy grid
- Observation model
  (pose,map) → observation
- Operators to move the robot to a target pose
- Update of the map and robot pose at each time-step



Objects: Perception

Method

1. Consider sensor readings that violate expectations of a static model.

2. Cluster them in space and then time.

3. Compute new perceptual features from the clusters.
   distance = average sensor reading
   angle = average sensor location

Objects: Learned Shapes

Note: shape models have size information

Objects:
Learning Operators

Method:

1. Perform motor babbling to collect data.

2. Use batch learning to find
contexts and motor outputs
that reliably change an attribute every
timestep (one second timesteps).

3. Evaluate the learned operators.

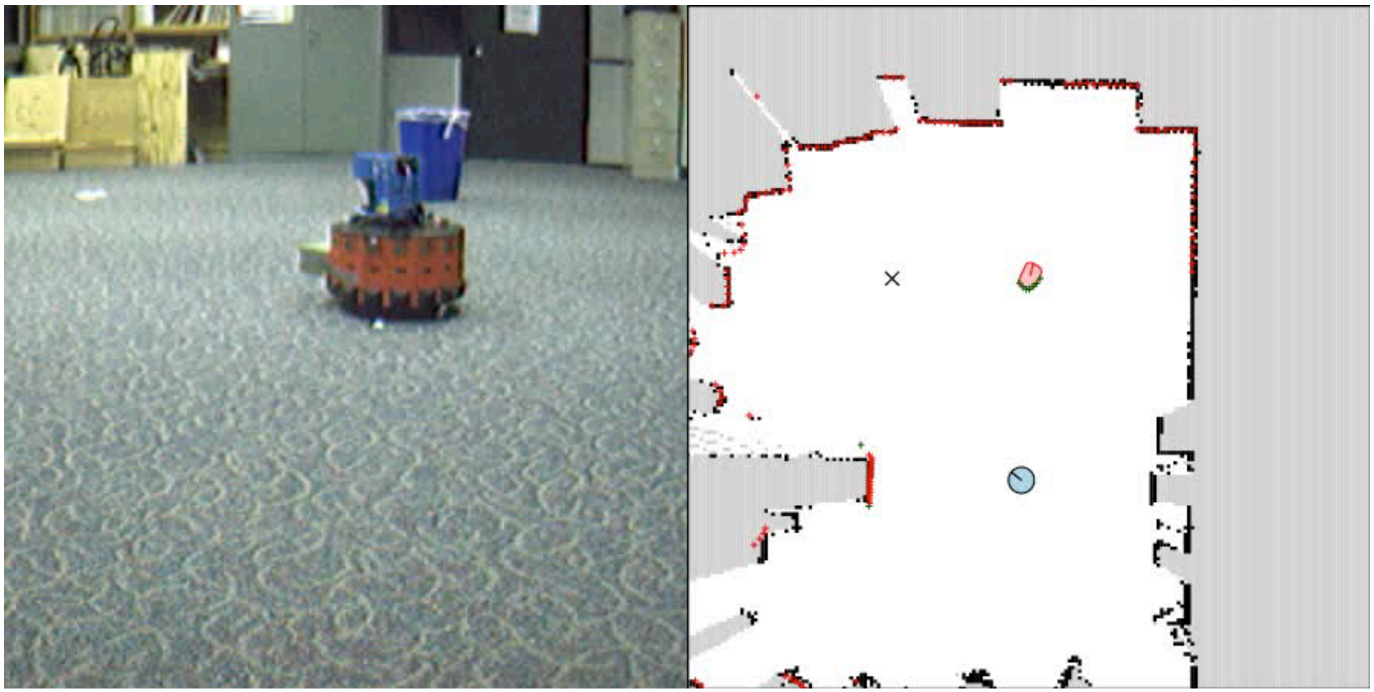
Operator 4: Decrease distance to object

Description: distance(τ), decrease, $\delta < -0.19$

Context: distance(τ) ≥ 0.43
angle(τ) ≤ 132
angle(τ) ≥ 69

Motor outputs: (0.2 m/s, 0.0 rad/s)

Objects: Using Operators



location(τ) dir[robot-heading]

distance(τ) decreasing

angle(τ) increasing

angle(τ) decreasing

Learning models that support reasoning

- Representations that support human-scale abstract reasoning can be learned from sensorimotor experience.
  - Is a robot's sensorimotor stream sufficient for learning all useful knowledge?
- How can the learning process be improved?
  - Simple unified semantics with broad applicability
  - Clarify assumptions
  - Incremental learning algorithms
  - Remove need for human oversight

Rich representations that support learning

Learning to make predictions

- A prediction is a claim about a robot's future experience.
  - Predictions verified by experiments are the foundation of scientific knowledge.
  - Thus, the semantics of experimentally verifiable predictions could be a useful foundation for a robot's knowledge.
  - An efficient online, incremental algorithm would enable the robot to make and learn many such predictions in parallel.
  - e.g. Temporal-difference reinforcement learning algorithms.

General value functions (GVF)

$$V^{\pi,\gamma,r,z}(s) = \mathbb{E}[r(s_1) + \ldots + r(s_k) + z(s_k) | s_0 = s, a_{0:k} \sim \pi, k \sim \gamma]$$

these four functions define the semantics of an experimentally verifiable prediction

policy $\pi : \mathcal{A} \times \mathcal{S} \longrightarrow [0, 1]$

pseudo reward $r : \mathcal{S} \longrightarrow \mathbb{R}$

termination $\gamma : \mathcal{S} \longrightarrow [0, 1]$

terminal reward $z : \mathcal{S} \longrightarrow \mathbb{R}$

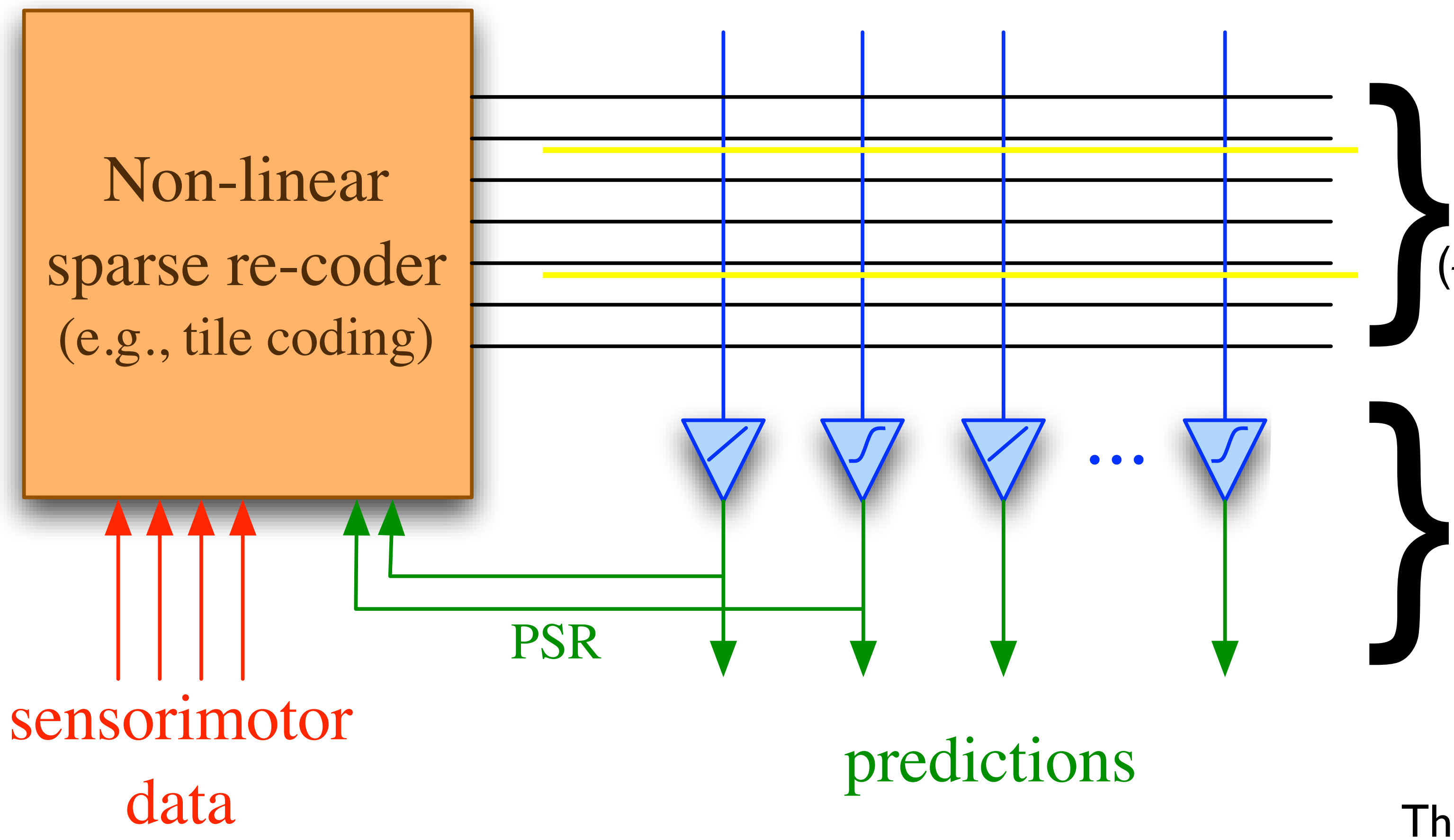
The Experimental Question
By selecting actions with the policy, how much reward will be received before termination?

Note 1: A GVF is a value function, but with a generic reward and termination.
Note 2: A constant termination probability corresponds to a timescale.

The Horde Architecture

(Sutton et al, 2011)

GVF predictions can be learned in parallel and online.



Sparsely activated binary features ϕ_t.
(#active << #features)

Each computed prediction (p) is a linear function of the features $p = \langle \theta_t, \phi_t \rangle$

The weights (θ) can be learned incrementally in O(#features) time/step by TD(λ) or related algorithms.

The firehose of experience

Normalized Sensor Values

Timesteps (0.1 second)

Predictions of a Light Sensor

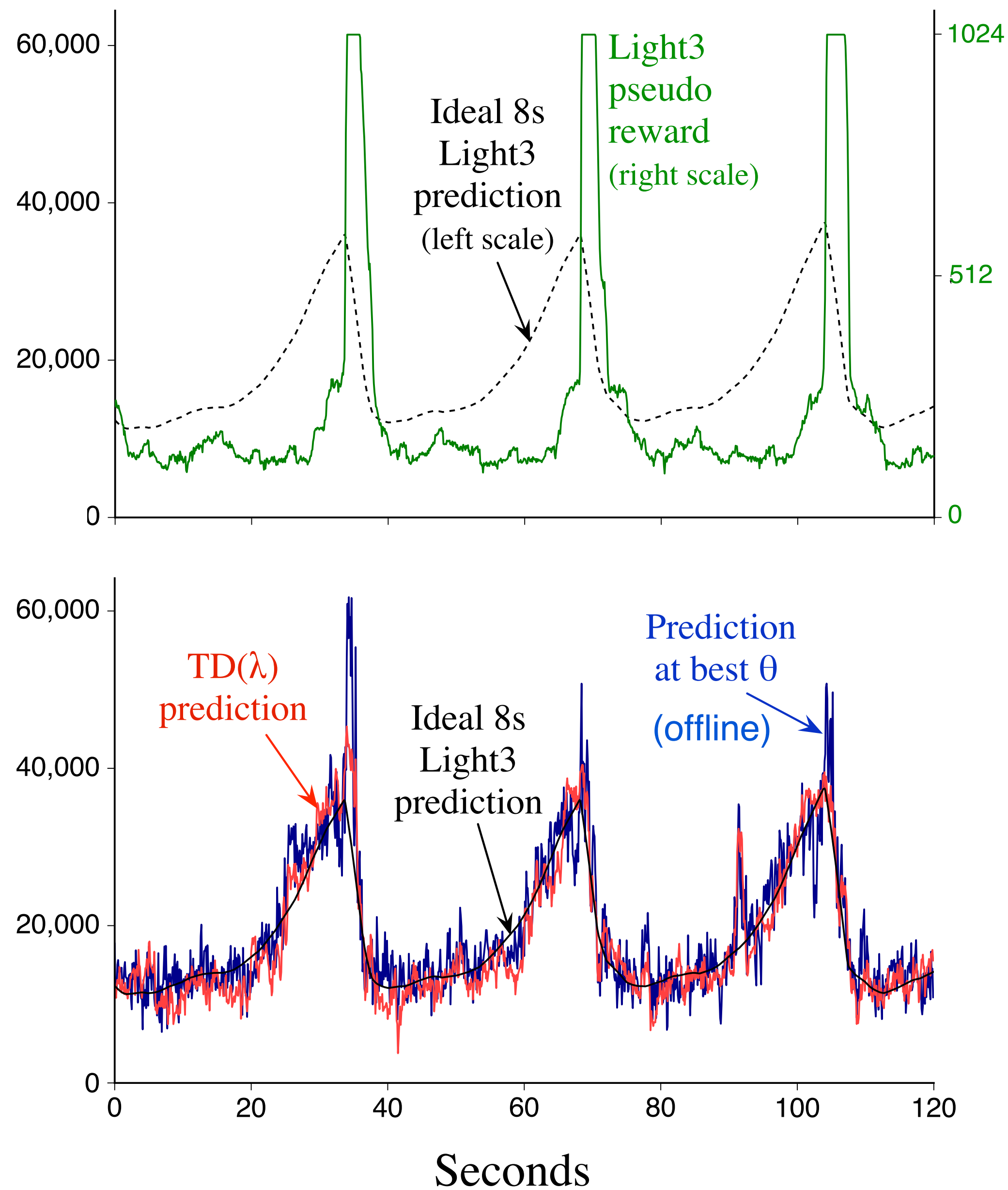
$r = Light3$

$\gamma = 0.9875$

$\pi = \text{Robot behaviour}$

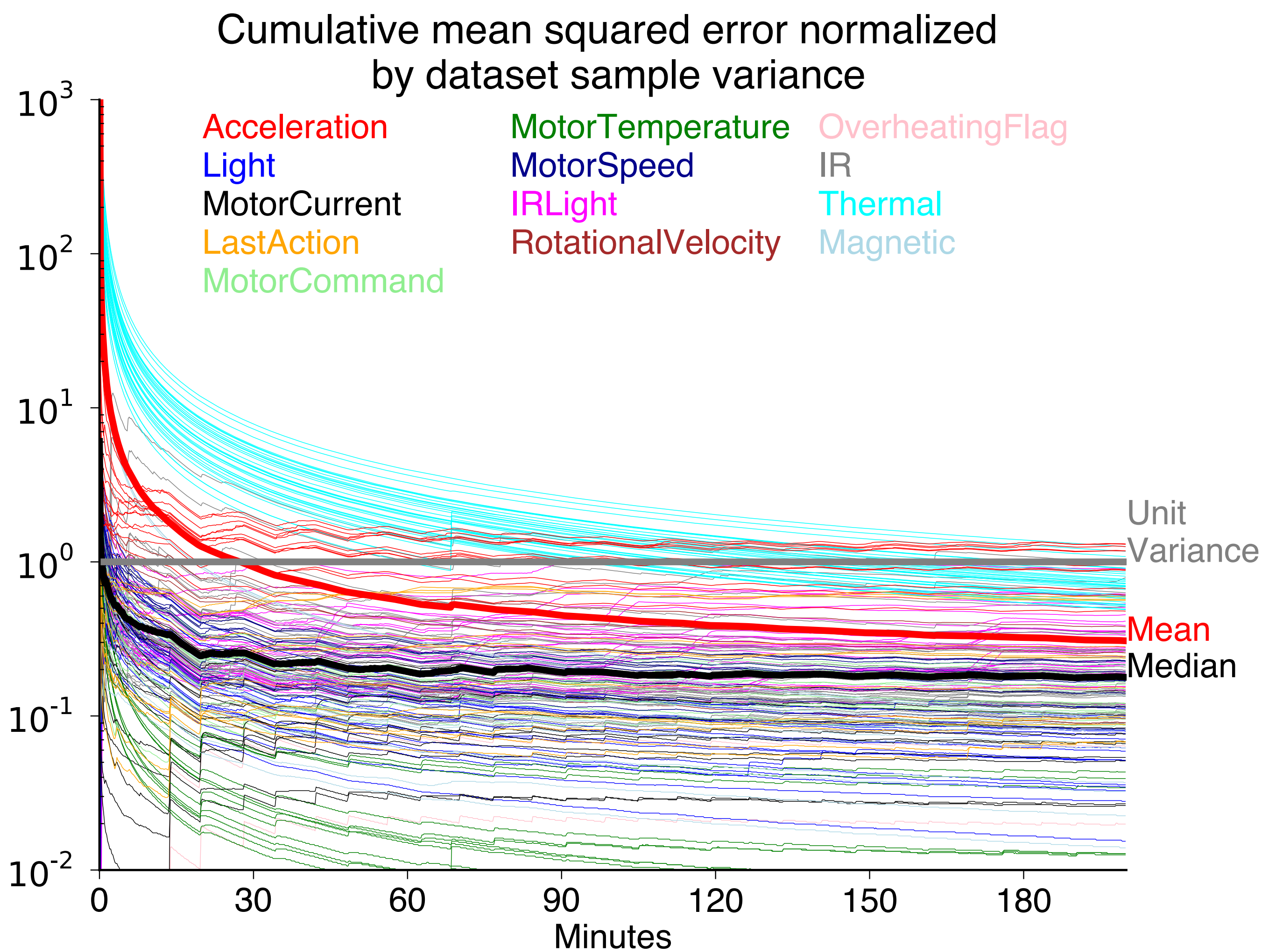
$z = 0$

The predictions learned online by TD(λ) are comparable to the ideal predictions and approach the accuracy of the best weight vector.
(shown after 3 hours of experience)



Scales to thousands of predictions

(Modayil, White, Sutton, 2012)



The 2000+ predictions

use 6000+ shared features,

shared parameters,

cover all sensors & many state bits,

cover 4 timescales (0.1, 0.5, 2, and 8 seconds),

and update every 55ms

All experience & learning performed within hours!

Learning predictions about different policies

- Off-policy learning enables the robot to learn the consequences of following different policies from a single stream of experience.
- Gradient temporal-difference algorithms provide stable, incremental, off-policy learning. (Maei & Sutton, 2009)
- Works at scale with robots. (White, Modayil, Sutton, 2012)

Summary

- Abstract models can be learned from sensorimotor experience.
  - Learned models of sensor space and objects that support goal-directed planning.
- A broad class of predictive knowledge can be learned at scale.
  - General value function predictions express an expected consequence of a precise experiment.
  - Temporal-difference algorithms can learn to make such predictions incrementally during normal robot experience.
- Robots could benefit from a tighter integration between learning from experience and reasoning with models.

Bibliography

- Model Learning

- Modayil, J., and Kuipers, B. J. 2007. Autonomous development of a grounded object ontology by a learning robot. In Proc. 22nd National Conf. on Artificial Intelligence (AAAI-2007).

- Modayil, J. 2010. Discovering sensor space: Constructing spatial embeddings that explain sensor correlations. In IEEE 9th International Conference on Development and Learning (ICDL).

- Prediction Learning

- Maei, H. R., and Sutton, R. S. 2010. GQ(λ): A general gradient algorithm for temporal-difference prediction learning with eligibility traces. In Proceedings of the Third Conference on Artificial General Intelligence.

- Modayil, J.; White, A.; and Sutton, R. S. 2012. Multi-timescale nexting in a reinforcement learning robot. SAB 2012. Springer. 299–309.

- Sutton, R. S.; Modayil, J.; Delp, M.; Degris, T.; Pilarski, P. M.; and Precup, D. 2011. Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. In Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS).

- White, A., Modayil, J., and Sutton, R.S. 2012. Scaling Life-long Off-policy Learning. In Second Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob).