
# Tools for learning RL representations

Carlos Diuk

Rutgers University

Department of Computer Science

Rutgers Laboratory for Real-Life Reinforcement Learning (RL^3)

Joint work with Lihong Li, Bethany Leffler, István Szita, Tom Walsh and Michael Littman

Barbados 2009

# Introduction


- Rich representations introduce new, representation-specific learning challenges.

- Sometimes these are not hard problems per se, except that existing methods:
  - Assume iid samples, and/or
  - Assume batch data, and/or
  - Have no formal guarantees

# In this talk

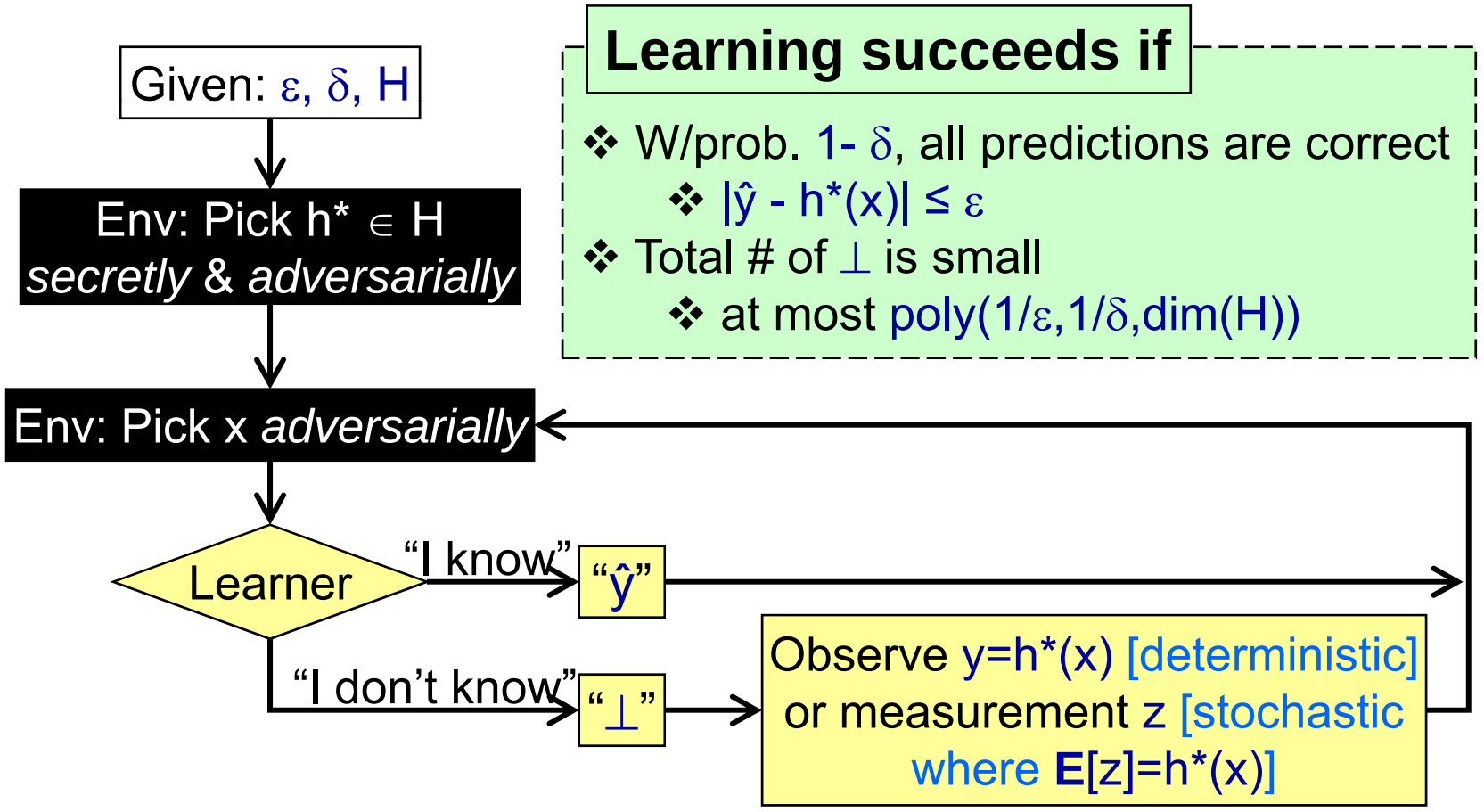

- Present online, non-iid versions of two tools that we found useful in a number of model-learning situations:
  - Probabilistic Concepts [Kearns et al, 1993]
  - Linear regression

  (with formal guarantees in the KWIK framework)

- Present applications

# KWIK Definition (slide by Lihong Li)

## "Knows What It Knows" [Li, Littman & Walsh 08]



# Tools I


- The Adaptive k-Meteorologists
  - Problem: find the correct model from a set of possible models.
    - Compromise between being provided with the correct model as input and learning it from a possibly unbounded class.
  - Applications:
    - Learning DBN structure
    - Feature selection / Choosing right classifier
    - Condition learning of OOMDP conditional effects
    - Choosing right function approximator?

# The k-Meteorologists

Meterologists make probabilistic predictions

p(rain)=0.1

p(rain)=0.6

p(rain)=0.9

Each day, you observe the predictions and then observe the true outcome (rain=0 or 1)

Which meteorologist should you trust?

# The k-Meteorologists

Meterologists make probabilistic predictions

p(rain)=0.1

$h_1$

p(rain)=0.6

$h_2$

p(rain)=0.9

$h_3$

h* is the correct concept

If $h_i$=h*, then squared error of $h_i$ is smallest

# Some results


- Adaptive version where each meteorologist *learns* its concept and everything remains KWIK-learnable.

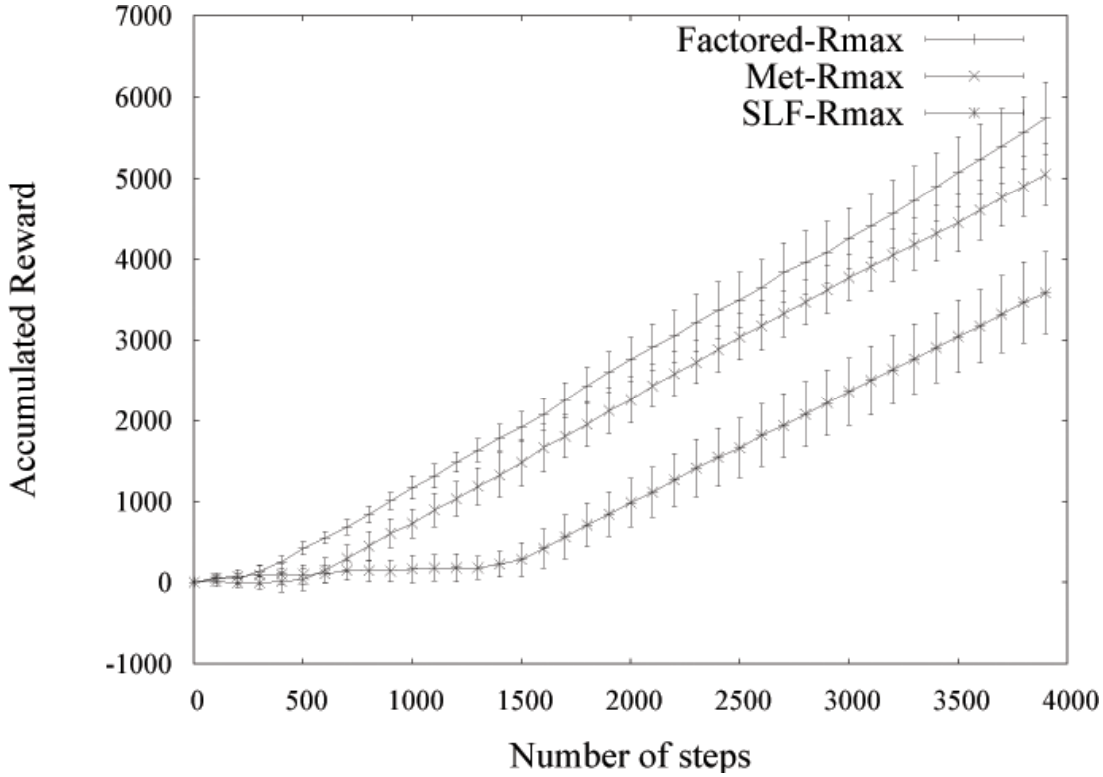
- Polynomial convergence (matching lower and upper bounds):

$$\Omega(k/\varepsilon^2 \ln k/\delta)$$

- MetRMax: Meteorologists provide predictions, if a meteorologist responds IDK we assume optimistic prediction.

# Application I

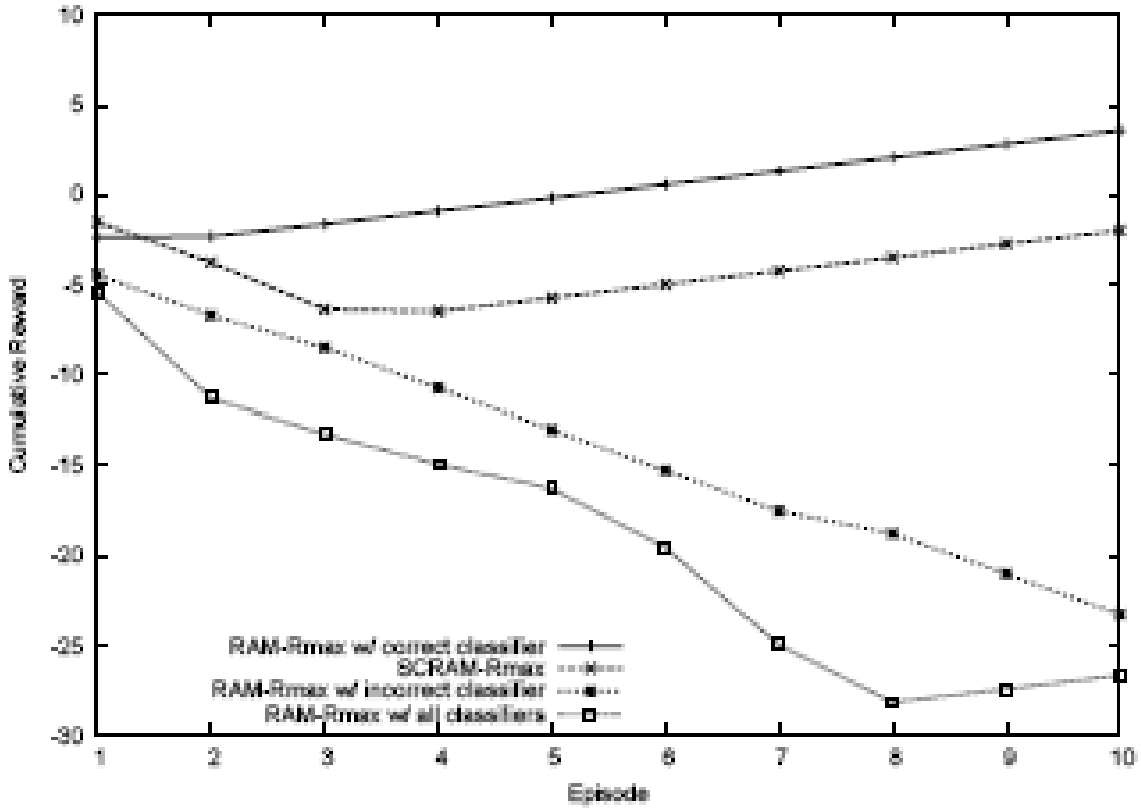
- Learning DBN structure
  - For each factor, there is a set of meteorologists, each considering a different set of parents.



# Application II (Bethany Leffler's work)

- Choosing the right classifier:
  - Robot provided with terrain classifiers based on:
    - Color
    - Texture
    - Color and texture

# Application II



# Other applications


- Object(??) Oriented MDPs
  - Each meteorologists pays attention to a different set of features (different condition) and tries to predict effect. Winner represents the condition that best explains the effect.
  - Has allowed me to eliminate deterministic assumptions in prior work
- Having multiple function approximators compete?

# Tool II


- Linear-KWIK (István Szita's work):
  - Online linear regression.
  - **Very** simple algorithm that satisfies KWIK assumptions.
  - Significant bound improvement over previous KWIK result (Strehl et al, 2008) (and it's actually a readable algorithm!)
  - Applications:
    - Learn factored rewards
    - Learn from ambiguous observations in STRIPS and OOMDP models

# Application I

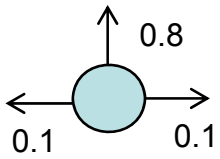

- Learning factored rewards: assumes total reward is linear combination of per-factor rewards.
- Linear-KWIK attributes reward to the right factors.

# Application II

- Ambiguous effects in stochastic STRIPS and O(?)OMDP representations

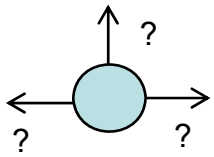
True dynamics

Action Up

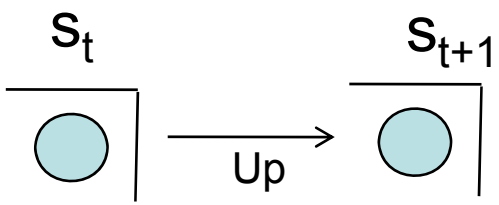



Learning Problem

Action Up




Observation




We can look at the outcome as a linear combination of effects.

Linear-KWIK allows us to learn the right probabilities.

# Conclusions


- "Learning representations is hard"
- Rich representations introduce many learning sub-problems.
- We can exploit wealth of ML learning methods, but first we need to make them RL-friendly (online, non-iid).
- Probabilistic Concept learning and online linear regression have proven very useful in RL^3, maybe they can help you too.

# Current/Future work


- Adaptive k-Meteorologist works with a number of prediction types:
  - Multinomial predictions.
  - Continuous distributions.

- Extending (heuristically) STRIPS/OOMDP effect-learning to eliminate the assumption that possible outcomes ("bins") are known using sparsification in linear-KWIK.

- Combining (pre-)condition and effect learning in a single algorithm for stochastic STRIPS and OOMDPs.