Informed Reinforcement Learning

An overview

Tom Croonenborghs

`Tom.Croonenborghs@cs.kuleuven.ac.be`

Department of Computer Science, Katholieke Universiteit Leuven

Celestijnenlaan 200A, B-3001 Leuven, Belgium

KATHOLIEKE UNIVERSITEIT LEUVEN

Overview

- Introduction
  - Reinforcement Learning
  - Relational Reinforcement Learning
  - Informed Reinforcement Learning
- The IRL framework
- A starting point
- Conclusions

KATHOLIEKE UNIVERSITEIT LEUVEN

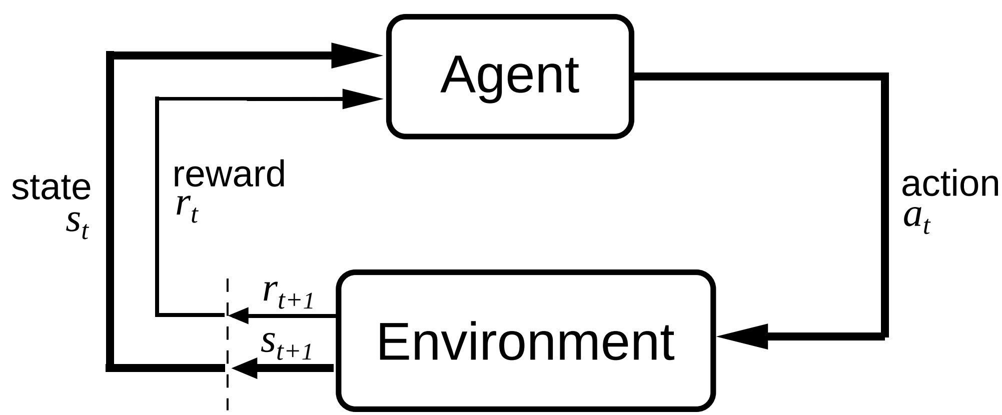
Reinforcement Learning

Based on psychological principles

- Observe behavior
- Reward desired behavior
- Improvement in behavior

Computer science

- online
- *trial and error*
- interaction
- state based world
- optimal policy



KATHOLIEKE UNIVERSITEIT LEUVEN

RL - formal

Given

- a set of possible *states* S.
- a set of possible *actions* A.
- a - for the agent unknown - *transition function* $\delta : S \times A \rightarrow S$.
- a - for the agent unknown - *reward function* $r : S \times A \rightarrow R$.

Find a policy $\pi^* : S \rightarrow A$, that maximizes

$$V^{\pi}(s_t) = \sum_{i=0}^{\infty} \gamma^i r_{t+i}$$

KATHOLIEKE UNIVERSITEIT LEUVEN

Relational Reinforcement Learning

A relation representation to represent states, actions and policies

- Allows use of
  - objects
  - properties of objects
  - relations between objects
- Allows generalisation
  - over states, actions, goals
  - re-use previous experience

KATHOLIEKE UNIVERSITEIT LEUVEN

Informed Reinforcement Learning

- a RRL-agent doesn't know what he is doing
  - just tries to maximize his future reward
- reason about actions, states, goals

The idea of IRL is

- inspired by modelbased reinforcement learning
- to generate or learn extra information
- use this information to accelerate convergence

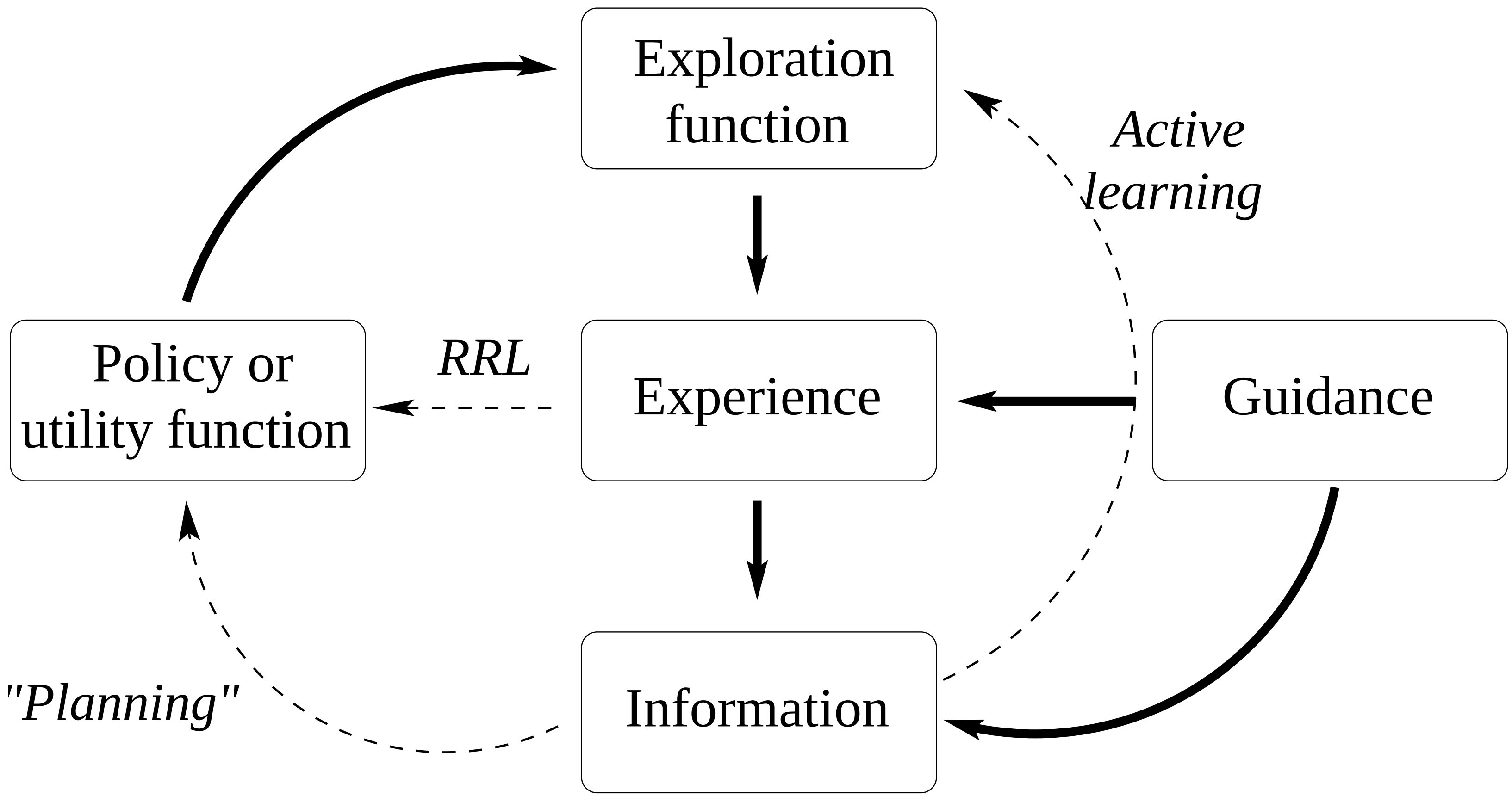
KATHOLIEKE UNIVERSITEIT
LEUVEN

Overview

- Introduction
  - Reinforcement Learning
  - Relational Reinforcement Learning
  - Informed Reinforcement Learning
- The IRL framework
- A starting point
- Conclusions

KATHOLIEKE UNIVERSITEIT LEUVEN

Informed Reinforcement Learning



KATHOLIEKE UNIVERSITEIT LEUVEN

The *I* in IRL

What to learn?

- properties of the environment and his objects
  - goal
  - subgoals (interesting properties to achieve)
- possible actions
  - preconditions
  - postconditions

What do with it?

Use this extra information in a goal-directed way to accelerate convergence

- utility function
- action selection
- lookahead
- planning

A note

Hierarchical abstraction

- allow to define action sequences

Overview

- Introduction
  - Reinforcement Learning
  - Relational Reinforcement Learning
  - Informed Reinforcement Learning
- The IRL framework
- A starting point
- Conclusions

KATHOLIEKE UNIVERSITEIT LEUVEN

A starting point

Reason about goals and “subgoals”

Accelerate convergence, using goal-oriented reasoning

- start at the end of an epsiode
- reason backwards
- search for “interesting” properties / conditions

KATHOLIEKE UNIVERSITEIT LEUVEN

The blocks world

A planning problem

- agent only receives reward when goal is reached

3 goals

- stacking
- unstacking
- on(A,B)

Each state is a set of facts, e.g. clear(a), clear(b), on(c,d), on(d,floor)

KATHOLIEKE UNIVERSITEIT LEUVEN

The blocks world

A planning problem

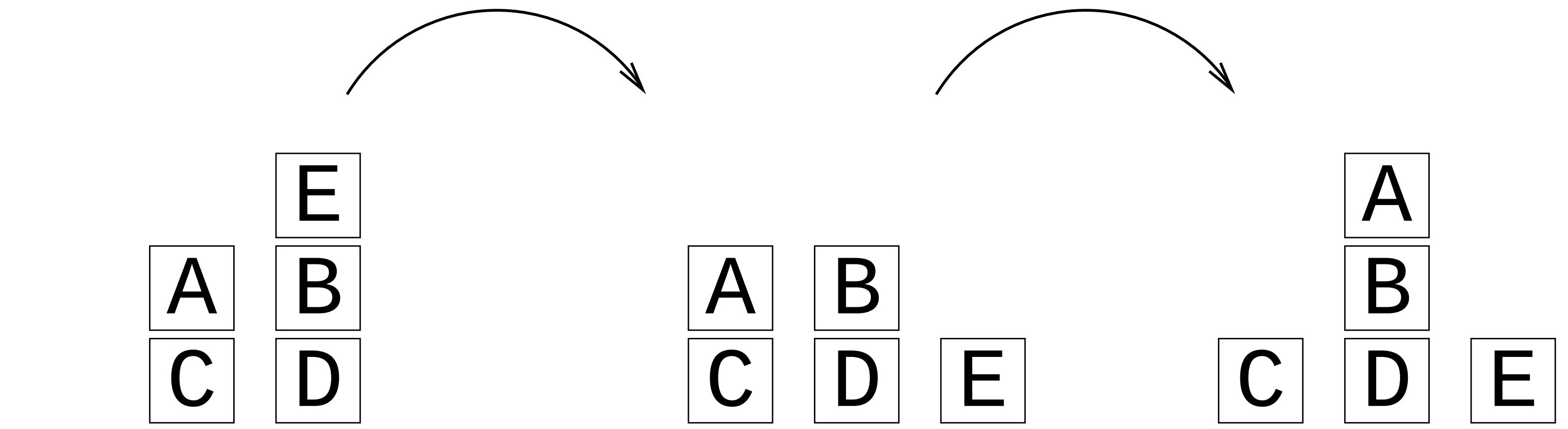
- agent only receives reward when goal is reached

3 goals

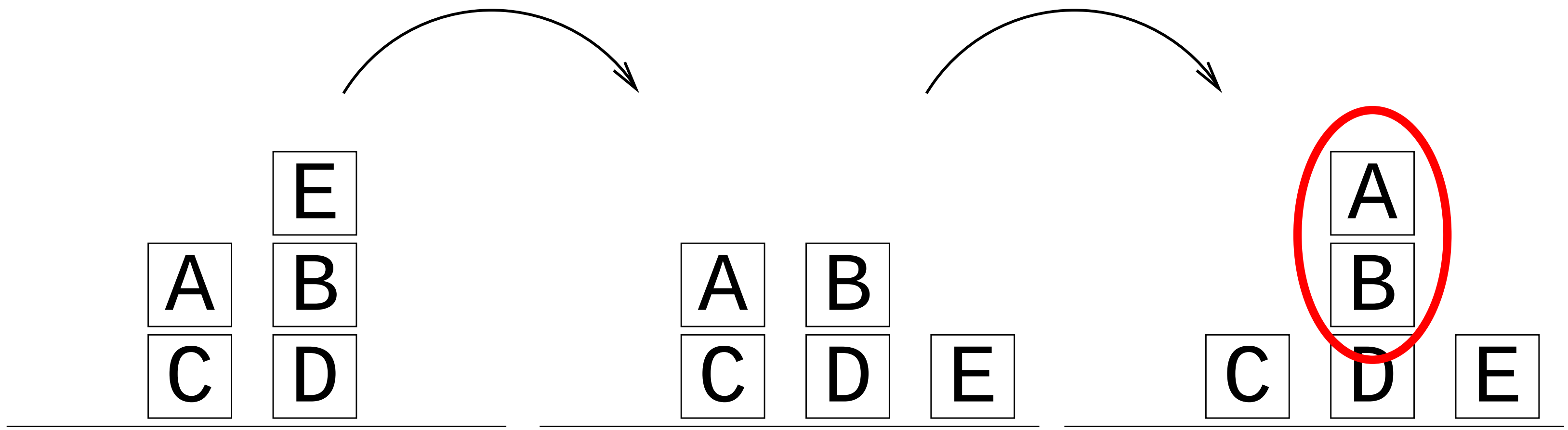
- stacking
- unstacking
- on(A,B)

Each state is a set of facts, e.g. clear(a), clear(b), on(c,d), on(d,floor)

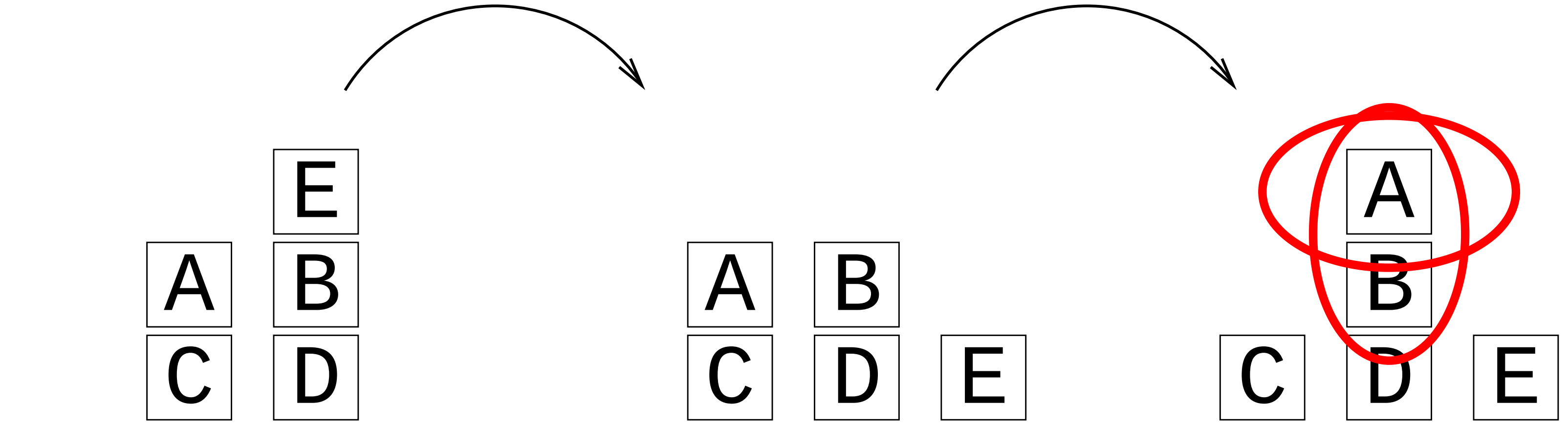
The on(A,B) goal



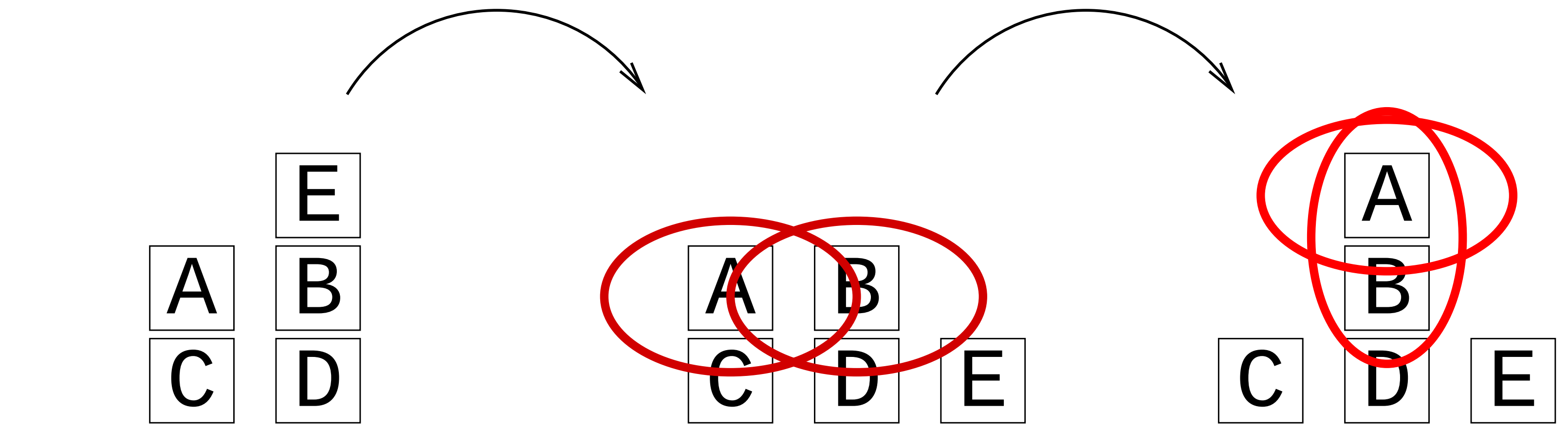
The on(A,B) goal



The on(A,B) goal



The on(A,B) goal



A starting point

Approaches to find interesting properties

- frequent pattern mining
- probabilities

Approaches to use this information

- action selection by more advanced agent
- standard RRL + extended reward function
- something in between, e.g. Q-function

KATHOLIEKE UNIVERSITEIT LEUVEN

Using probabilities

After an episode, update probabilities

- if goal reached
- if goal reached in the next state
- if goal not reached

property	$Goal$	NSR	Avg
on(a,b)	1	0	0
clear(a)	1	1	0.3
clear(b)	0	1	0.3
on(e,floor)	0.28	0.28	0.31
...	...	...	...

KATHOLIEKE UNIVERSITEIT LEUVEN

Using probabilities

After an episode, update probabilities

- if goal reached
- if goal reached in the next state
- if goal not reached

property	$Goal$	NSR	Avg
on(a,b)	1	0	0
clear(a)	1	1	0.3
clear(b)	0	1	0.3
on(e,floor)	0.28	0.28	0.31
...	...	...	...

KATHOLIEKE UNIVERSITEIT LEUVEN

Using probabilities

After an episode, update probabilities

- if goal reached
- if goal reached in the next state
- if goal not reached

property	$Goal$	NSR	Avg
on(a,b)	1	0	0
clear(a)	1	1	0.3
clear(b)	0	1	0.3
on(e,floor)	0.28	0.28	0.31
...	...	...	...

KATHOLIEKE UNIVERSITEIT LEUVEN

Using probabilities

After an episode, update probabilities

- if goal reached
- if goal reached in the next state
- if goal not reached

property	$Goal$	NSR	Avg
on(a,b)	1	0	0
clear(a)	1	1	0.3
clear(b)	0	1	0.3
on(e,floor)	0.28	0.28	0.31
...	...	...	...

KATHOLIEKE UNIVERSITEIT LEUVEN

Extended exloration function

Interesting preconditions

- properties with a significant difference between $P(Goal|Property)$ and $P(Avg|Property)$

Extended exploration function

- Use an agent that can explore the environment in a more goal-directed way.
- more information is needed

But, the agent needs to know how to accomplish these properties.

KATHOLIEKE UNIVERSITEIT LEUVEN

Extended reward function

- Extend the reward function with probability of reward in next state
- State S = clear(a), on(a,floor), clear (b) $\ldots$

$$R'(S,A) = R(S,A) + \omega P(RNS|S)$$

Compute $P(NSR|S)$ with e.g. Naive Bayes

$$\begin{aligned} P(RNS|S) &= P(RNS|clear(a)).P(RNS|on(a,floor)) \\ &\quad .P(RNS|clear(b)) \ldots \end{aligned}$$

More advanced techniques (BLP, $\ldots$)

A Bayesian approach

- Learning structure is guided by goal.
- Example in BLP format:

reward(t) | goal(on(A,B)),on(A,B,t).

on(A,B,t) | action(move(A,B),t-1).

on(A,B,t) | on(A,B,t-1).

success(move(A,B),t) | clear(A,t),clear(B,t).

clear(A,t) | clear(A,t-1).

clear(A,t) | on(B,A,t-1),action(move(B,C),t-1),B$\neq$C.

Overview

- Introduction
  - Reinforcement Learning
  - Relational Reinforcement Learning
  - Informed Reinforcement Learning
- The IRL framework
- A starting point
- Conclusions

KATHOLIEKE UNIVERSITEIT LEUVEN

Conclusions

Conclusions

- a short overview of Informed Reinforcement Learning
- a starting point
  - discovering subgoals
    - frequent pattern mining (warmr)
    - probabilities
  - extending reward function

Future work

- see previous slides

KATHOLIEKE UNIVERSITEIT LEUVEN

Questions?

Questions?