Goal Recognition over POMDPs: Inferring the Intention of a POMDP Agent

Miquel Ramirez, Hector Geffner
DTIC Universitat Pompeu Fabra
Barcelona, Spain
6/2011

Planning

- Planning is the **model-based** approach to action selection: behavior obtained from **model** of the **actions**, **sensors**, **preferences**, and **goals**

$$\textit{Model} \Longrightarrow \boxed{\textit{Planner}} \Longrightarrow \textit{Controller}$$

- Many **planning models**; many **dimensions**: uncertainty, feedback, costs, . . .

Basic Model: Classical Planning

- finite and discrete state space S
- a **known initial state** $s_0 \in S$
- a set $S_G \subseteq S$ of goal states
- actions $A(s) \subseteq A$ applicable in each $s \in S$
- a **deterministic transition function** $s' = f(a, s)$ for $a \in A(s)$
- positive **action costs** $c(a, s)$

A **solution** is a sequence of applicable actions that maps s_0 into S_G, and it is **optimal** if it minimizes sum of action costs ($\#$ of steps)

Other **models** obtained by relaxing assumptions in **bold** . . .

Uncertainty and Full Feedback: Markov Decision Processes

MDPs are **fully observable, probabilistic** state models:

- a state space S
- initial state $s_0 \in S$
- a set $G \subseteq S$ of goal states
- actions $A(s) \subseteq A$ applicable in each state $s \in S$
- **transition probabilities** $P_a(s'|s)$ for $s \in S$ and $a \in A(s)$
- action costs $c(a, s) > 0$

– **Solutions** are **functions (policies)** mapping states into actions

– **Optimal** solutions minimize **expected cost** to goal

Uncertainty and Partial Feedback: Partially Observable MDPs (POMDPs)

POMDPs are **partially observable, probabilistic** state models:

- states $s \in S$
- actions $A(s) \subseteq A$
- transition probabilities $P_a(s'|s)$ for $s \in S$ and $a \in A(s)$
- observable goal states $S_G \subseteq S$

- initial **belief state** b_0
- **sensor model** given by probabilities $P_a(o|s)$, $o \in O$, $s \in S$

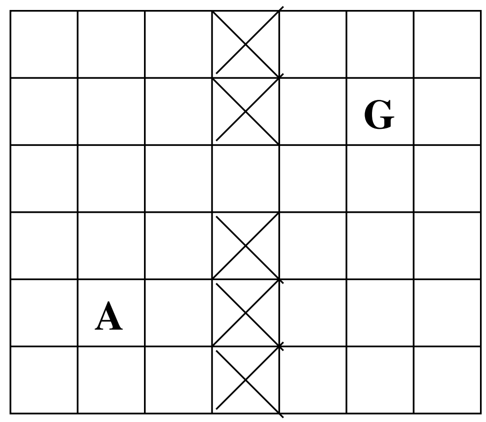
– **Belief states** are probability distributions over S

– **Solutions** are policies that map belief states into actions

– **Optimal** policies minimize **expected** cost to go from b_0 to S_G

Example

Agent **A** must reach **G**, moving one cell at a time in **known** map



- If actions deterministic and initial location known, planning problem is **classical**
- If actions stochastic and location observable, problem is an **MDP**
- If actions stochastic and location partially observable, problem is a **POMDP**

Different combinations of uncertainty and feedback: three problems, three models

From Planning to Plan Recognition

- **Plan Recognition** related to Planning (Plan Generation), but hasn't built on it; rather addressed using Grammars, Bayesian Networks, etc.

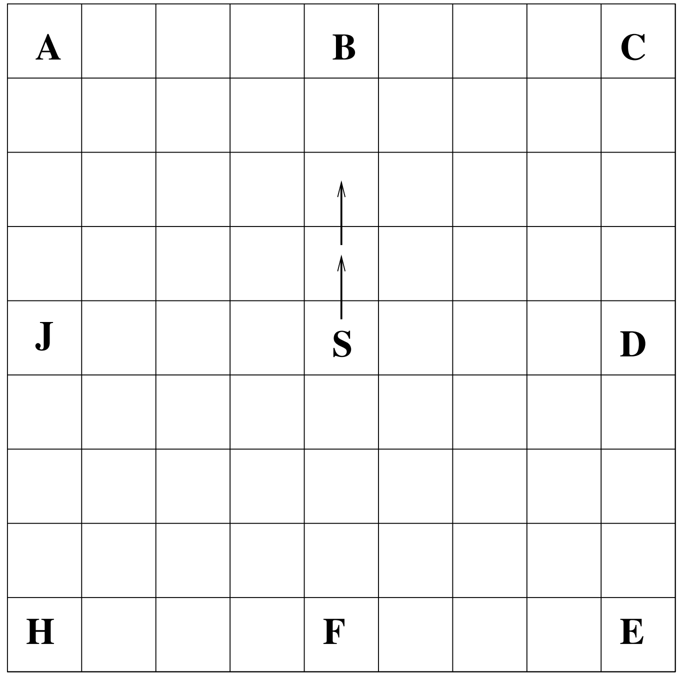
- Recent efforts to formulate and solve **plan recognition** using **planners**:

  ▷ Plan Recognition as Planning, *M. Ramirez and H. Geffner, Proc. IJCAI-2009*
  ▷ Probabilistic Plan Recognition using off-the-shelf Classical Planners, *M. Ramirez and H. Geffner, Proc AAAI-2010*
  ▷ Goal Inference as Inverse Planning, *C. Baker, J. Tenenbaum, R. Saxe. Cog-Sci 2007*
  ▷ Action Understanding as Inverse Planning. *C. Baker, R. Saxe, and J. Tenenbaum. Cognition, 2009*

- **General idea:** solve **plan recognition** problem over **model** (classical, MDP, POMDP) using **planner** for that **model**.

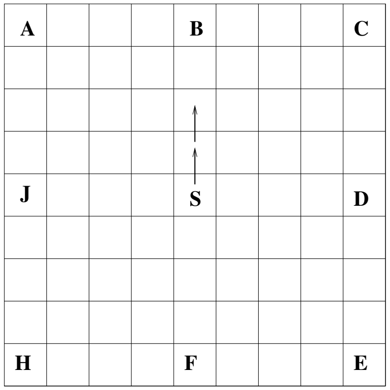
How/why can this be done?

Example



- Agent can **move** one unit in the four directions
- Possible **targets** are A, B, C, . . .
- Starting in S, he is **observed** to move up twice
- **Where** is he going? Why?

Example (cont'd)



- From Bayes, **goal posterior** is $P(G|O) = \alpha\, P(O|G)\, P(G)$, $G \in \mathcal{G}$
- If **priors** $P(G)$ given for each goal in $\mathcal{G}$, the question is what is $P(O|G)$
- $P(O|G)$ measures **how well goal** G **predicts observed actions** O
- In **classical** setting,
  - ▷ G predicts O **worst** when needs to get off the way **to comply with** O
  - ▷ G predicts O **best** when needs to get off the way **not** to comply with O

Posterior Probabilities from Plan Costs

- From Bayes, **goal posterior** is $P(G|O) = \alpha\, P(O|G)\, P(G)$,
- If **priors** $P(G)$ given, set $P(O|G)$ to

$$\textbf{function}(c(G + \overline{O})\ -\ c(G + O))$$

  - ▷ $c(G + O)$: cost of achieving G **while complying with** O
  - ▷ $c(G + \overline{O})$: cost of achieving G **while not complying with** O

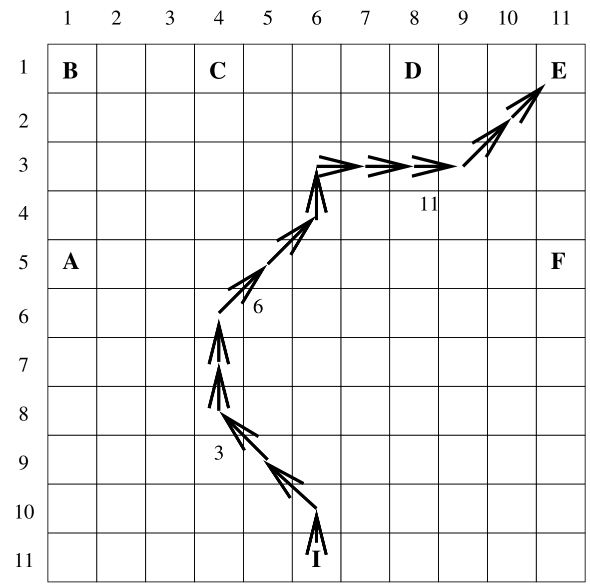
– Costs $c(G + O)$ and $c(G + \overline{O})$ **computed** by **classical planner**

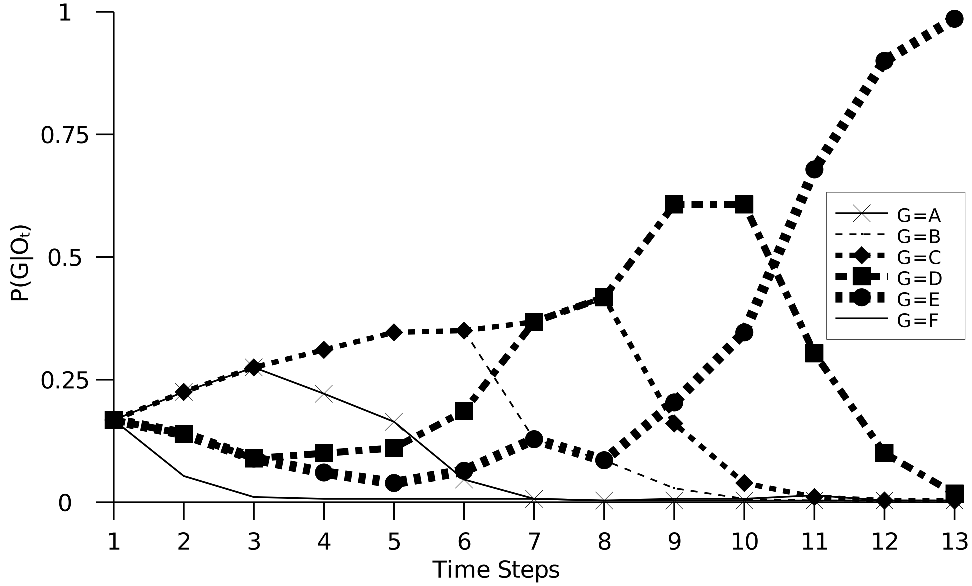
– Goals of **complying** and **not complying** with O translated into normal goals

– **Function** of cost difference set to **sigmoid**; follows from assuming $P(O|G)$ and $P(\overline{O}|G)$ are Boltzmann distributions $P(O|G) = \alpha'\, exp\{-\beta\, c(G, O)\}$, . . .

– Result is that posterior probabilities $P(G|O)$ **computed in** $2|\mathcal{G}|$ **classical planner calls**, where $\mathcal{G}$ is the set of possible goals (Ramirez and G. 2010)

Illustration: Noisy Walk





Graph on left shows 'noisy walk' and possible targets; curves on right show resulting **posterior probabilities** $P(G|O)$ of each possible target G as a function of time

Plan Recognition for MDP Agents

- **MDP planner** provides costs $Q_G(a, s)$ of achieving G from s starting with a
- Agent assumed to act 'almost' greedily following Boltzmann distribution

$$P(a|s, G) = \alpha \, exp\{-\beta \, Q_G(a, s)\}$$

- Likelihood $P(O|G)$ for observations $O = a_0, s_1, a_1, s_2, \ldots$ given G obeys recursion

$$P(a_i, s_{i+1}, a_{i+1}, \ldots | s_i, G) = P(a_i|s_i, G) \, P(s_{i+1}|a_i, s_i) \, P(a_{i+1}, \ldots | s_{i+1}, G)$$

- Assumptions in this model (Baker, Tenenbaum, Saxe, Cog-Sci 07):
  - ▷ **MDP is fully solved** with costs $Q(a, s)$ for all a, s
  - ▷ States **fully observable** by both agent and observer
  - ▷ Observation sequence is **complete**; no action is missing

Assumptions in these Models

- Ramirez and G. infer goal distribution $P(G|O)$ assuming that
  - ▷ O is a sequence of **some** of the actions done by agent, and that
  - ▷ agent and observer share **same classical model**, except for agent goal that is replaced by set of possible goals
- Baker *et al.* infer goal distribution $P(G|O)$ assuming that
  - ▷ O is the **complete** sequence of **actions** and **observations** done/gathered by agent, and that
  - ▷ agent and observer share **same MDP model**, except for agent goal that is replaced by set of possible goals
- In this work, we generalize Baker *et al.* to **POMDPs** while dropping the assumption that **all agent actions and observations visible to observer**

Example: Plan Recognition over POMDPs

- Agent is looking for item A or B which can be in one of three drawers 1, 2, or 3
- Agent doesn't know where A and B are, but has **priors** $P(A@i)$, $P(B@i)$
- He can move around, open and close drawers, look for an item in open drawer, and grab an item from drawer if known to be there
- The sensing action is not perfect, and agent may fail to see item in drawer
- Agent observed to do

$$O = \{open(1), open(2), open(1)\}$$

- If possible goals G are to have A, B, or both, and priors given, what's posterior $P(G|O)$?

Formulation: Plan Recognition over POMDPs

- **Bayes:** $P(G|O) = \alpha P(O|G)P(G)$, priors $P(G)$ given
- **Likelihoods:** $P(O|G) = \sum_\tau P(O|\tau)P(\tau|G)$ for the possible executions τ for G
- **Approximation:** $P(O|G) \approx m_O/m$, where m is total $\#$ of **executions sampled** for G, and m_O is $\#$ that **comply with** O
- **Sampling:** executions sampled assuming that agent does action a in belief b for goal G with Boltzmann distribution:

$$P(a|b, G) = \alpha' exp\{-\beta\, Q_G(a, b)\}$$

where $Q_G(a, b)$ is expected cost from b to G starting with a:

▷ $Q_G(a, b) = c(a, b) + \sum_{o \in O} b_a(o) V_G(b_a^o)$, and
▷ $V_G(b)$ precomputed by **planner**

Experiments

- Formulation tested using POMDP solver **GPT** (Bonet and Geffner)
- Three POMDP domains analyzed: **Office, Drawers, Kitchen** (features below)
- Large dataset generated varying **hidden goal** and **observations** randomly
- Resulting **binary goal classifier** evaluated according to standard measures (TPR, FPR, Accuracy, Precision)
  - ▷ G classified as **positive/negative** given O if G most likely given O
  - ▷ Classification is **true/false positive** if O generated from hidden G', $G' = G / G' \neq G$, . . .

Name	$\|S\|$	$\|A\|$	$\|Obs\|$	$\|b_0\|$	$\|\mathcal{G}\|$	$\|T\|$
OFFICE	2,304	23	15	4	3	3.4
DRAWERS	3,072	16	16	6	3	4.5
KITCHEN	69,120	29	32	16	5	10.1

Results

Domain	Obs %	L	T	ACC	PPV
office	30	4.9	24.6	0.99	0.97
	50	7.6	24.7	1.00	1.00
	70	10.8	24.8	1.00	1.00
kitchen	30	3.8	95.2	0.86	0.73
	50	5.8	95.1	0.93	0.85
	70	8.3	95.2	0.98	0.95
drawers	30	2.9	38.8	0.84	0.77
	50	3.9	38.8	0.87	0.80
	70	6.0	38.8	0.96	0.93

- **Columns:** domains, observation ratio, avg length of obs sequence (L), avg time (T), avg accuracy (ACC), precision (PPV)
- **Definitions:** Accuracy, and Precision given by $TP + TN/P + N$, and $TP/TP + FP$ . . .
- **Parameters:** $m = 10,000$ ($\#$ of sampled executions), $\beta = 40$ (noise level)

Summary

- New formulation of goal recognition for settings where agent has **partial info about environment** and observer has **partial info about actions done by agent**

- Posterior goal probabilities $P(G|O)$ computed from Bayes rule using given priors $P(G)$ and likelihoods **approximated** in three steps:

  - ▷ POMDP planner produces **expected costs** $V_G(b)$ from beliefs to goals
  - ▷ **Stochastic simulations** with action probabilities $P(a|b,G)$ computed from these costs
  - ▷ $P(O|G)$ set to ratio of simulations for G that **comply with** O

- Several **extensions** discussed in paper

  - ▷ belief recognition, failure to observe and actions that must be observed; observing what agent observes, noise in agent-observation channel