Decision-theoretic planning via probabilistic programming

Vaishak Belle
University of Edinburgh
Alan Turing Institute
References

- https://dtai.cs.kuleuven.be/problog
What’s on for today?

Automated planning in non-trivial stochastic domains

Transparent specification language with generic solution scheme

Computational outlook on open issues
Automated planning

Shakey the robot [Fikes & Nilsson, 1971]

Synthesize **action sequence** to achieve goals
A world of blocks

<table>
<thead>
<tr>
<th>act:</th>
<th>Pickup(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre:</td>
<td>OnTable(x), Clear(x), Handempty</td>
</tr>
<tr>
<td>add:</td>
<td>Holding(x)</td>
</tr>
<tr>
<td>del:</td>
<td>OnTable(x), Clear(x), Handempty</td>
</tr>
</tbody>
</table>
Planning in the real (noisy) world

Actions may be **stochastic**

Actions and states may be **continuous/discrete/mixed**

States may be defined over **unknown (number of) objects**
A set of initial conditions, which may involve uncertainty about continuous quantities like temperature, energy available, solar flux, and position.

A set of possible actions.

A set of certain and uncertain effects that describe the world following the action. Uncertain effects on continuous variables are characterized by probability distributions.

The problem that we have just described is essentially a decision-theoretic planning problem.
Markov Decision Processes

Underlying decision-theoretic framework in game theory, recommendation systems, robotics, etc.

Compute **policy**: maps states and time steps to actions

**Objective**: maximise expected reward over horizon $t$
Maximising expected reward

\[ V^\pi_d(s_t) = E \left[ \sum_{k=0}^{d} R(s_{t+k}, a_{t+k}) | s_t, \pi \right] \]

\[ V^*_d(s_t) = \max_a \left( R(s_t, a_t) + \gamma \int_{s_{t+1}} p(s_{t+1} | s_t, a_t) V^*_d(s_{t+1}) ds_{t+1} \right) \]
Additional complications

Unknown current state; estimate by noisy observation: partially observable MDPs that can be reduced to belief MDPs

Probabilities/rewards unknown: reinforcement learning
1. Elegant mathematical framework, but solving the general case notoriously hard
2. How easy to describe domain with complex relationships and discovery?
Exploit structure

Monte Carlo planners that work in arbitrary MDPs are very slow in practice

Why? Only access sample traces, but do not exploit:

- Probabilities of transitions
- Structure of the planning model

States, actions more than abstract entities: instantiated over structured domain theories that express relationships and dependencies
Desiderata

A rich modelling language that allows transparent domain axiomatisation in the presence of unknowns and stochasticity

Solution scheme that leverages inherent structure
Probabilistic programming

Languages to model structured probability distributions

Make machine learning modular and enable descriptive clarity

Programming languages with stochastic primitive

Many proposed to date: Church, BLOG, Anglican, ProbLog, IBAL, etc.
ProbLog
Two coin tosses in a sequence

% Probabilistic facts:
0.5::heads1.
0.6::heads2.

% Rules:
twoHeads :- heads1, heads2.

% Queries:
query(heads1).
query(heads2).
query(twoHeads).
Knowledge graphs

1. `0.6::edge(1,2).`
2. `0.1::edge(1,3).`
3. `0.4::edge(2,5).`
4. `0.3::edge(2,6).`
5. `0.3::edge(3,4).`
6. `0.8::edge(4,5).`
7. `0.2::edge(5,6).`
8. `path(X,Y) :- edge(X,Y).`
9. `path(X,Y) :- edge(X,Z), Y \== Z, path(Z,Y).`
10. `query(path(1,5)).`
father(bart, stijn).
father(bart, pieter).
father(luc, soetkin).

mother(katleen, stijn).
mother(katleen, pieter).
mother(lieve, soetkin).

parent(bart, stijn).
parent(bart, pieter).
parent(luc, soetkin).

female(alice).
female(an).
female(esther).

male(bart).
male(etienne).
male(leon).

grandmother(esther, soetkin).
grandmother(esther, stijn).
grandmother(esther, pieter).
...
Candidates for iteration 4:

grandmother(A,B) :- parent(C,B), parent(A,C), parent(D,A) 0.503462603878
grandmother(A,B) :- parent(C,B), parent(A,C), parent(B,D) 0.457063711911
grandmother(A,B) :- parent(C,B), parent(A,C), male(C) 0.432528690146
grandmother(A,B) :- parent(C,B), parent(A,C), male(B) 0.432528690146

RULE LEARNED: grandmother(A,B) :- parent(C,B), parent(A,C), \+male(A) 1.0

Learning a relation continued
Q1: In a group of 10 people, 60 percent have brown eyes. Two people are to be selected at random from the group. What is the probability that neither person selected will have brown eyes?
Unknowns, continuous distributions and dynamics
Unknown color

color(X) \sim \text{uniform([black, brown])} \leftarrow \text{material}(X) \sim= \text{wood}.
material(X) \sim \text{finite}([0.3:\text{wood}, 0.7:\text{metal}]) \leftarrow \text{between}(1, N, X).

size(X) \sim \text{beta}(2, 3) \leftarrow \text{material}(X) \sim= \text{metal}.

size(X) \sim \text{beta}(4, 2) \leftarrow \text{material}(X) \sim= \text{wood}.
Unknown numbers

\( n \sim \text{poisson}(6). \)

(Infinite valued discrete distribution)

\[
\text{material}(X) \sim \text{finite}(\{0.3: \text{wood}, 0.7: \text{metal}\}) \leftarrow n \sim N, \text{between}(1, N, X).
\]
Continuous distributions, dynamics

$$\text{pos}_{t+1}(ID)_x \sim \text{gaussian}(\sim(\text{pos}_t(ID)_x), \sigma^2) \leftarrow \sim(\text{move}_t) = ID.$$ 

$$\text{obsPos}_{t+1}(ID) \sim \text{gaussian}(\sim(\text{pos}_{t+1}(ID)), \text{cov}).$$
Clauses in action (object tracking)
Key inference ideas

Relevant variables: SLD resolution

Informed search: importance sampling

Avoid invalid regions: constraint propagation

\[ \Pr(q \mid Y = 3X+1) \]
From inference to planning
Specifying MDPs

State transition model: \( \text{Var}_{t+1} \sim \text{Distribution} \leftarrow \text{Conditions}_t \)

Applicable actions: \( \text{applicable}(\text{Action})_t \leftarrow \text{Conditions}_t \)

Reward: \( \text{reward}(R)_t \leftarrow \text{Conditions}_t \)

Terminal state: \( \text{stop}_t \leftarrow \text{Conditions}_t \)

\[
\begin{align*}
\text{stop}_t & \leftarrow \sim(\text{type}(X)_t) = \text{can.} \\
\text{reward}(20)_t & \leftarrow \text{stop}_t. \\
\text{reward}(-1)_t & \leftarrow \text{not} (\text{stop}_t).
\end{align*}
\]
Can be over unknowns (e.g., find red can)

Removing blue box

Removing yellow can
An additional function

\[ V_d^\pi(s_t) = E \left[ \sum_{k=0}^{d} R(s_{t+k}, a_{t+k}) \mid s_t, \pi \right] \]

\[ Q_d^\pi(s_t, a_t) = E \left[ \sum_{k=0}^{d} R(s_{t+k}, a_{t+k}) \mid s_t, a_t, \pi \right] \]
Computing an optimal policy

\[ \pi(s) \leftarrow \arg\max_a Q^\pi(s, a) \]
HYPE = Hybrid episodic planner

for each applicable action $a$ in $s_t^m$ do
  $\tilde{Q}_d^m(s_t^m, a) \leftarrow R(s_t^m, a) + \frac{\sum_{i=0}^{m-1} w^i \tilde{V}_d^{i-1}(s_{i+1}^m)}{\sum_{i=0}^{m-1} w^i}$
end for

$\pi^m \leftarrow \text{policy}(\{\tilde{Q}_d^m(s_t^m, \cdot)\})$

Sample next state

Sample $s_{t+1}^m \sim p(s_{t+1} \mid s_t^m, a_t^m)$

$G_d^m \leftarrow R(s_t^m, a_t^m) + \text{SAMPLE_EPISTODE}(d-1, s_{t+1}^m, m)$

Recursive call

$\tilde{V}_d^m(s_t^m) \leftarrow G_d^m$

Store $(s_t^m, \tilde{V}_d^m(s_t^m), d)$

Return $\tilde{V}_d^m(s_t^m)$

end function
## Evaluations

<table>
<thead>
<tr>
<th>Domain</th>
<th>Planner</th>
<th>d</th>
<th>Param.</th>
<th>Reward</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>game1</td>
<td>HYPE</td>
<td>5</td>
<td>M = 1200</td>
<td>0.87 ± 0.11</td>
<td>662</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>5</td>
<td>C = 1</td>
<td>0.34 ± 0.15</td>
<td>986</td>
</tr>
<tr>
<td></td>
<td>HYPE</td>
<td>4</td>
<td>M = 1200</td>
<td><strong>0.89 ± 0.07</strong></td>
<td>312</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>4</td>
<td>C = 2</td>
<td>0.79 ± 0.08</td>
<td>1538</td>
</tr>
<tr>
<td>game2</td>
<td>HYPE</td>
<td>5</td>
<td>M = 1200</td>
<td>0.67 ± 0.18</td>
<td>836</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>5</td>
<td>C = 1</td>
<td>0.14 ± 0.20</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>HYPE</td>
<td>4</td>
<td>M = 1200</td>
<td><strong>0.76 ± 0.19</strong></td>
<td>582</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>4</td>
<td>C = 2</td>
<td>0.27 ± 0.22</td>
<td>1528</td>
</tr>
<tr>
<td>sysadmin1</td>
<td>HYPE</td>
<td>5</td>
<td>M = 1200</td>
<td>0.94 ± 0.07</td>
<td>422</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>5</td>
<td>C = 1</td>
<td>0.47 ± 0.13</td>
<td>1068</td>
</tr>
<tr>
<td></td>
<td>HYPE</td>
<td>4</td>
<td>M = 1200</td>
<td><strong>0.98 ± 0.06</strong></td>
<td>346</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>4</td>
<td>C = 2</td>
<td>0.66 ± 0.08</td>
<td>1527</td>
</tr>
<tr>
<td>sysadmin2</td>
<td>HYPE</td>
<td>5</td>
<td>M = 1200</td>
<td><strong>0.87 ± 0.11</strong></td>
<td>475</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>5</td>
<td>C = 1</td>
<td>0.31 ± 0.12</td>
<td>1062</td>
</tr>
<tr>
<td></td>
<td>HYPE</td>
<td>4</td>
<td>M = 1200</td>
<td>0.86 ± 0.11</td>
<td>392</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>4</td>
<td>C = 2</td>
<td>0.46 ± 0.12</td>
<td>1532</td>
</tr>
</tbody>
</table>
## Evaluations (2)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>M</th>
<th>Loss</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>simplerover2</td>
<td>HYPE</td>
<td>8</td>
<td>11.8 ± 0.2</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>8</td>
<td>11.4 ± 0.3</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>HYPE</td>
<td>9</td>
<td>11.7 ± 0.2</td>
<td>195</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>9</td>
<td>11.3 ± 0.3</td>
<td>238</td>
</tr>
<tr>
<td></td>
<td>HYPE</td>
<td>10</td>
<td>11.9 ± 0.3</td>
<td>218</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>10</td>
<td>11.2 ± 0.3</td>
<td>1043</td>
</tr>
<tr>
<td>marsrover</td>
<td>HYPE</td>
<td>6</td>
<td>249.8 ± 33.5</td>
<td>985</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>6</td>
<td>227.7 ± 27.3</td>
<td>787</td>
</tr>
<tr>
<td></td>
<td>HYPE</td>
<td>7</td>
<td>269.0 ± 29.4</td>
<td>983</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>7</td>
<td>N/A</td>
<td>Timeout</td>
</tr>
<tr>
<td></td>
<td>HYPE</td>
<td>10</td>
<td>296.3 ± 19.5</td>
<td>1499</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>≥ 8</td>
<td>N/A</td>
<td>Timeout</td>
</tr>
</tbody>
</table>

Cf. paper on results with relational abstraction
Outlook: open issues

Difficulty handling low probability observations

Guessing “good” proposal distributions hard

Bounds on computed values? (E.g., Safety-critical applications)
SAT and \#SAT

Given a CNF formula,

- SAT: find a satisfying assignment
- \#SAT: count satisfying assignments

\[ (x \lor y) \land (y \lor \neg z) \]

- 5 models: (0,1,0), (0,1,1), (1,1,0), (1,1,1), (1,0,0)
- Equivalently: satisfying probability = \( \frac{5}{2^3} \)
Weighted #SAT

Polytime reduction from exact inference in discrete graphical models to weighted #SAT

Think of (1,0,0) as sequence of one heads, two tails

Exact algorithms with strong runtime bounds

Approximate algorithms with strong certificates

*ProbLog reduces inference computation to Weighted #SAT*
Weighted #SMT (IJCAI-15)

Constraint propagation capabilities

Exact and approximate methods have been identified

Dealing with countably infinite values (UAI-17)

Use these methods to provide tight correctness characterisations?

\[
\begin{align*}
(2 + u)^3 / 6 & \quad -3 \leq u \leq 3 \\
(4 - 6u^2 - 3u^3) / 6 & \quad -2 < u \leq -1 \\
(4 - 6u^2 + 3u^3) / 6 & \quad -1 < u \leq 0 \\
(2 - u)^3 / 6 & \quad 0 < u \leq 1 \\
\end{align*}
\]
Summary

HYPE works in a wide range of domains: discrete, continuous, hybrid, growing vs shrinking state spaces

Systematically handles discovery of unknown objects

Exploits the probabilistic model and relational structure to provide fast solutions

*Enables transparency and modularity of intricate stochastic specifications (e.g., MDP part of larger pipeline)*