INTRODUCTION À LA PLANIFICATION

Tutoriel 10 – PFIA 2023

Frédéric Maris[1,5] Aïdin Sumic[2] Thierry Vidal[2] Bruno Zanutini[3,5] Tiago de Lima[4,5]

[1]IRIT, Université Toulouse 3 – Paul Sabatier

[2]LGP, École nationale d'ingénieurs de Tarbes

[3]GREYC, Université de Caen Normandie

[4]CRIL, Université d'Artois

[5]CNRS

Plate-forme Intelligence Artificielle, Strasbourg, France, 5 juillet 2023

Partie 3

Epistemic Reasoning for Contingent Planning

Contents

Section 10

Setting and Motivation

Ontic goal : set of states of the environment

- any state with B above A
- $p \wedge \neg q$ (at home and not fuel tank empty)...

Ontic goal : set of states of the environment

- any state with B above A
- $p \wedge \neg q$ (at home and not fuel tank empty)...

Ontic or epistemic goals ?

- Mastermind : goal = proposed $\leftrightarrow$ hidden , $\neq K(\text{hidden})$
- Hanabi, Cluedo... : ontic as well
- epistemic : intelligence (know whether terrorist), info. diffusion (gossip)

Ontic goal : set of states of the environment

- any state with B above A
- $p \wedge \neg q$ (at home and not fuel tank empty)...

Ontic or epistemic goals ?

- Mastermind : goal = proposed $\leftrightarrow$ hidden , $\neq K(\text{hidden})$
- Hanabi, Cluedo... : ontic as well
- epistemic : intelligence (know whether terrorist), info. diffusion (gossip)

Do we care ?

- ontic $\hookrightarrow$ epistemic : $p \wedge \neg q \hookrightarrow K(p \wedge \neg q)$
- epistemic $\hookrightarrow$ ontic : $K(\text{terr.}) \vee K(\neg\text{terr.}) \hookrightarrow (\text{said-terr.} \wedge \text{terr.}) \vee (\text{said-not-terr.} \wedge \neg\text{terr.})$
- negative epistemic goals : becomes adversarial

Ontic preconditions, ontic effects :

- precondition : clear(B) , effect : $+$on-table(B), $-$on(B, A), $-$on(B, C)
- effect : tank-empty?$\emptyset$: $+$at(dest), $-$at(loc), $\pm$tank-empty

Ontic preconditions, ontic effects :

- precondition : $\text{clear}(B)$, effect : $+\text{on-table}(B), -\text{on}(B, A), -\text{on}(B, C)$
- effect : $\text{tank-empty?}\emptyset : +\text{at}(\text{dest}), -\text{at}(\text{loc}), \pm\text{tank-empty}$

Epistemic actions, really ?

- recall : one effect of $!\varphi$ is $K_A\varphi$ for all agents A
- $\textsc{search-dict}_A(\text{“trhouought”})$ (!spelling(“trhouought”)) : effect $K_A(\text{spelling}(\text{“trhouought”}))$
- $\textsc{teach}_{A\to\text{pupil}}(a^2 = b^2 + c^2)$: effect $K_{\text{pupil}}(a^2 = b^2 + c^2)$
- $!(\neg\text{Earth-flat})$: effect $K_A(\neg\text{Earth-flat})$ for all agents A

Ontic preconditions, ontic effects :

- precondition : $\text{clear}(B)$, effect : $+\text{on-table}(B), -\text{on}(B, A), -\text{on}(B, C)$
- effect : $\text{tank-empty?}\emptyset : +\text{at}(\text{dest}), -\text{at}(\text{loc}), \pm\text{tank-empty}$

Epistemic actions, really ?

- recall : one effect of $!\varphi$ is $K_A\varphi$ for all agents A
- $\textsc{search-dict}_A(\text{“trhouought”})$ (!spelling(“trhouought”)) : effect $K_A(\text{spelling}(\text{“trhouought”}))$
- $\textsc{teach}_{A\to\text{pupil}}(a^2 = b^2 + c^2)$: effect $K_{\text{pupil}}(a^2 = b^2 + c^2)$
- $!(\neg\text{Earth-flat})$: effect $K_A(\neg\text{Earth-flat})$ for all agents A

Perfect reasoners, really ?

- recall : $K_A\varphi$ and $\varphi \models \psi$ entail $K_A(\psi)$
- $\textsc{tell}(y = a^3 \wedge a = \log\log q \wedge q = 47,323)$: then $K(y < 64)$
- $!(\text{E.U. history})$: then $K(\text{voting-far-right-dangerous})$

Ontic *vs* epistemic :

- most natural problems have ontic actions, ontic goals
- epistemic actions (P.A., etc.) : strong assumptions

Should we `rm -f tutorial_tiago.pdf`?

- no, agents *do* acquire knowledge and reason about it
- rather capture in execution model than in action model

And anyway :

- existing models of imperfect reasoning
- reason with belief bases [18], with B instead of K...
- epistemic planning still makes sense in some settings

Coming next : purely ontic planning problems with epistemic reasoning

18. [Lorini, 2020, Fernandez Davila, 2022]

Section 11

Single-Agent Setting

Tiger problem :

- two doors, one with tiger, one with gold
- ontic actions : open left/right door (+10 or -100)
- sensing action : listen roar , yields good/bad clue .9/.1
- initial belief : tiger left/right .5/.5
- timestep costs 1

Intuitively : listen enough to have strong belief where tiger is

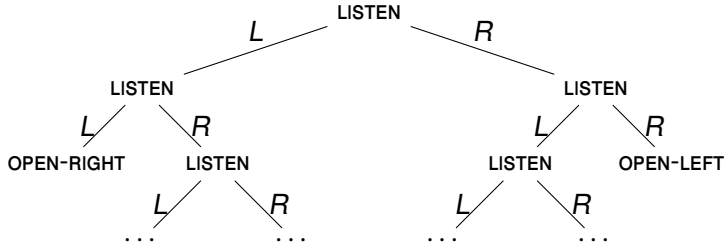
Definition (POMDP)

Partially Observable Markov Decision Problem :

- sets S (states), A (actions), Ω (observations)
- transition function : $T : S \times A \to \Delta(S)$
- reward function : $R : S \to \mathbb{R}$
- observation function : $O : S \times A \times S \to \Delta(\Omega)$
- initial belief : $B_0 \in \Delta(S)$

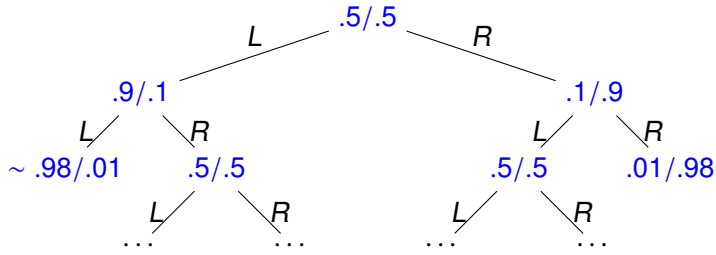
Solution/policy :

- depends on whole history : mapping $\pi : \omega \in \Omega^* \rightarrow A$
- value : expectation of cumulated reward
- note : undecidable at indefinite horizon



Execution : just follow path

Recall : B_0 is left/right .5/.5, listen gives clue .9/.1, reward +10/-100



Maintained by Bayes rule :

$$B(s') \leftarrow \eta\Big(\sum_{s} B(s)T(s' \mid s,a)O(\omega \mid s,a,s')\Big)$$

One action left :

- open left gives $-100B(\mathrm{l}) + 10B(\mathrm{r}) - 1$ $\hookrightarrow$ $v^1(\textsc{open-left}) = (-100, 10, -1)$
- open right gives $10B(\mathrm{l}) - 100B(\mathrm{r}) - 1$ $\hookrightarrow$ $v^1(\textsc{open-right}) = (10, -100, -1)$
- listen gives $0B(\mathrm{l}) + 0B(\mathrm{r}) - 1$ $\hookrightarrow$ $v^1(\textsc{listen}) = (0, 0, -1)$
- choose $\mathrm{argmax}_a \left[(B(\mathrm{l}), B(\mathrm{r}), 1) \cdot v^1(a) \right]$

Two actions left :

- open left, open right : same
- listen gives
  - $aB(\mathrm{l}) + bB(\mathrm{r}) + c$ if open right on observation L and open left on R $\hookrightarrow$ $v^2_{or,ol} = (a, b, c)$
  - $dB(\mathrm{l}) + eB(\mathrm{r}) + f$ if listen left on observation L and open right on R $\hookrightarrow$ $v^2_{l,or} = (d, e, f)$
  - ...
- choose $\mathrm{argmax}_a \left[\max_{i,j} [(B(\mathrm{l}), B(\mathrm{r}), 1) \cdot v^2_{i,j}(a)] \right]$

Choose action as a function of current belief state

Planning time ; compute α-vectors :

- set $v^0(\_) := \{R\}$
- for $i = 1, 2, \ldots$: set $v^i(a) := \{\text{reg}(v^{i-1}, o_1 : v_1, \ldots, o_k : v_k) \mid v_1, \ldots, v_k \in v^{i-1}\}$
- until ε-convergence/stopping criterion

Execution time, given α-vectors $\forall a, v(a)$:

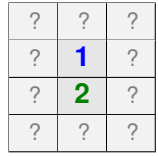
- set $B := B_0$
- perform $a := \text{argmax}_a \, B \cdot v(a)$
- observe ω
- update B using a, ω and Bayes rule
- iterate

Contingent planning instance :

- sets S (states), A (actions), Ω (observations)
- transition function : $T : S \times A \to \mathcal{P}(S)$
- reward function : $R : S \to \mathbb{R}$
- observation function : $O : S \times A \times S \to \mathcal{P}(\Omega)$
- initial belief : $B_0 \in \mathcal{P}(S)$

Strong cyclic policy :

- mapping $\pi : \Omega^* \to A$
- value : 1 if $\forall \omega_1, \omega_2, \ldots$, policy π reaches goal , else 0
- note : decidable (finite space)



Instance :

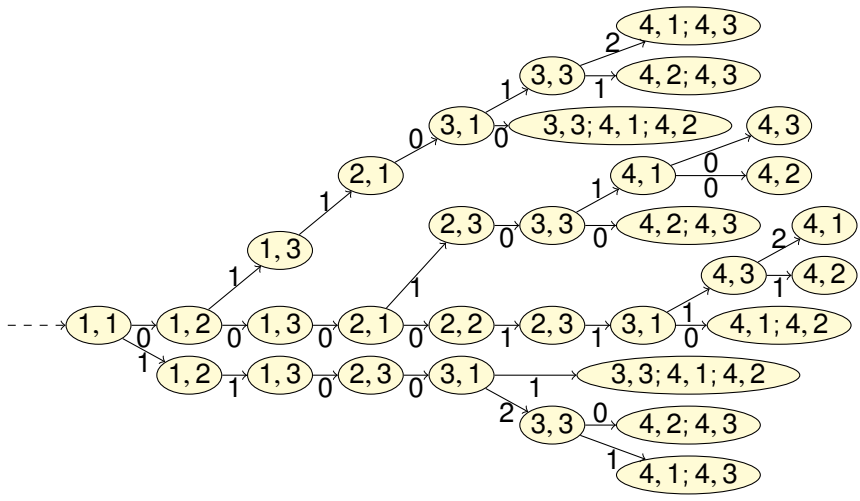
- states = all possible grids with 2 mines and consistent with numbers revealed
- actions = $\{\textsc{click}(i,j) \mid i,j\}$
- observations = $\{0, 1, 2, \ldots, 8\} \cup \{$  $\}$

Note : adversarial/robust version

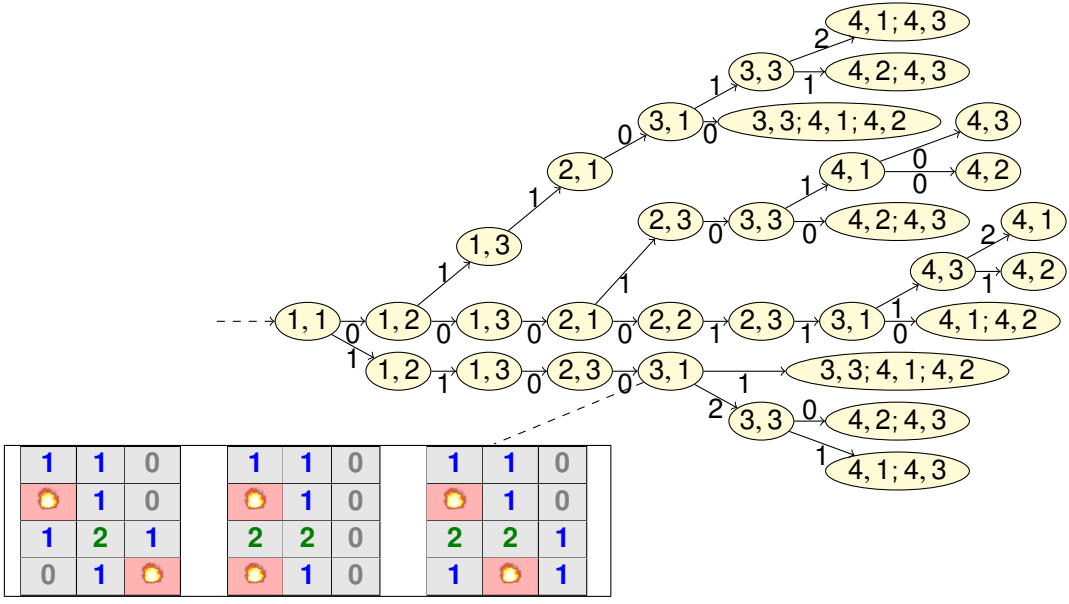
Finding strong policy for contingent planning = and/or search :

- root = B_0
- or-nodes = possible actions
- and-node = possible observations
- leaves = goal states
- policy = strategy in And/Or graph

?	?	?
?	**1**	?
?	**2**	?
?	?	?


1, 1
1, 2
1, 3
2, 1
3, 1
3, 3
4, 1; 4, 3
4, 2; 4, 3
3, 3; 4, 1; 4, 2
4, 3
4, 1
4, 2
2, 3
3, 3
4, 2; 4, 3
4, 1
4, 3
4, 2
2, 2
2, 3
3, 1
4, 1; 4, 2
1, 2
1, 3
2, 3
3, 1
3, 3; 4, 1; 4, 2
3, 3
4, 2; 4, 3
4, 1; 4, 3

Belief States are Again Here



Progression : $\text{prog}(B, a, \omega) := \{s' \in S \mid \exists s \in B : s' \in T(s, a), \omega \in O(s, a, s')\}$

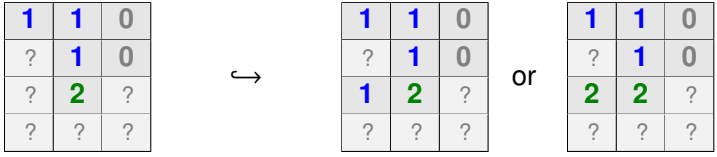
Progression : $\text{prog}(B, a, \omega) := \{s' \in S \mid \exists s \in B : s' \in T(s, a), \omega \in O(s, a, s')\}$

Belief space transformation $\cdot^{\mathcal{B}}$ for contingent instance $I = (S, A, T, R, \Omega, O, B_0)$:

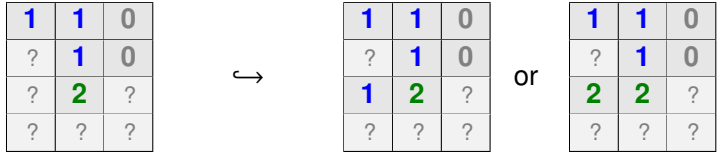
- $S^{\mathcal{B}} := \mathcal{P}S$
- $A^{\mathcal{B}} := A$
- $T^{\mathcal{B}}(B, a) := \{\text{prog}(B, a, \omega) \mid \exists s' \in T(s, a) : \omega \in O(s, a, s')\}$
- $R^{\mathcal{B}}(B) := \min_{s \in B} R(s)$
- belief state fully observed : $\Omega := S^{\mathcal{B}}$, $O(B, a, B') := \{B'\}$
- policy for $I^{\mathcal{B}} \equiv$ policy for I

Fully observable nondeterministic planning

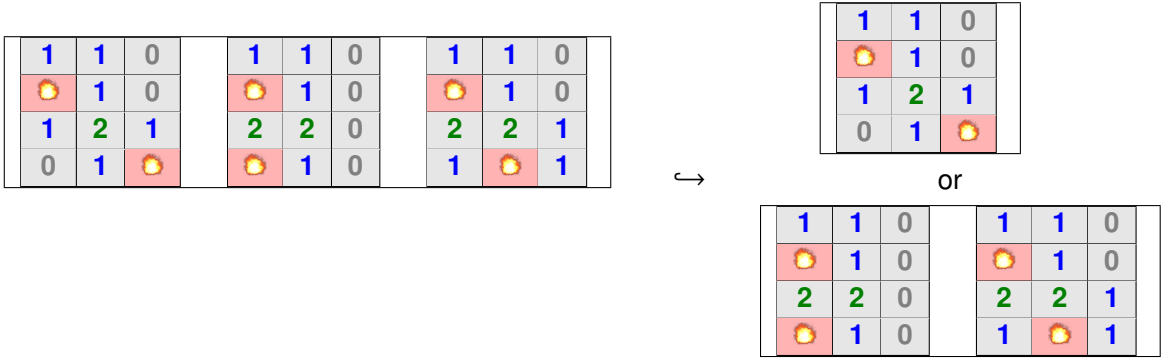
History-based, $a = \textsc{click}(3, 1)$:



History-based, $a = \textsc{click}(3, 1)$:



In belief space :



Direct approaches :

- CMBP [19] : conformant planning (no sensing), regression-based
- AO* : contingent planning [20]
- belief states are huge → symbolic representations using BDDs
- other representations : DNF, CNF, Prime Implicates [21]

Known literals [22] :

- conformant planning
- store only $K\ell$ for relevant known literals in current B
- avoids storing B

19. [Cimatti and Roveri, 2000]
20. [Bonet and Geffner, 2000]
21. [Son Thanh To, 2017]
22. [Palacios and Geffner, 2009]

Intuition :

- recall : α-vectors $v_{i,j}(\text{OPEN-LEFT})$, $v_{i,j}(\text{OPEN-RIGHT})$, $v_{i,j}(\text{LISTEN})$
- $(B(\text{l}), B(\text{r}), 1) \cdot v(\text{OPEN-LEFT}) > (B(\text{l}), B(\text{r}), 1) \cdot \text{OPEN-RIGHT}, (B(\text{l}), B(\text{r}), 1) \cdot \text{LISTEN}$
  $\rightarrow$ compact representation of set of belief states
- let's generalize

Knowledge-Based Policy [23] for Minesweeper :

---

while $\neg K(goal)$ **do**

  if $K\neg\text{mine}(1,1)$ **then** CLICK$(1,1)$ **else** ε **fi ;**

  if $K\neg\text{mine}(1,2)$ **then** CLICK$(1,2)$ **else** ε **fi ;**

  ...

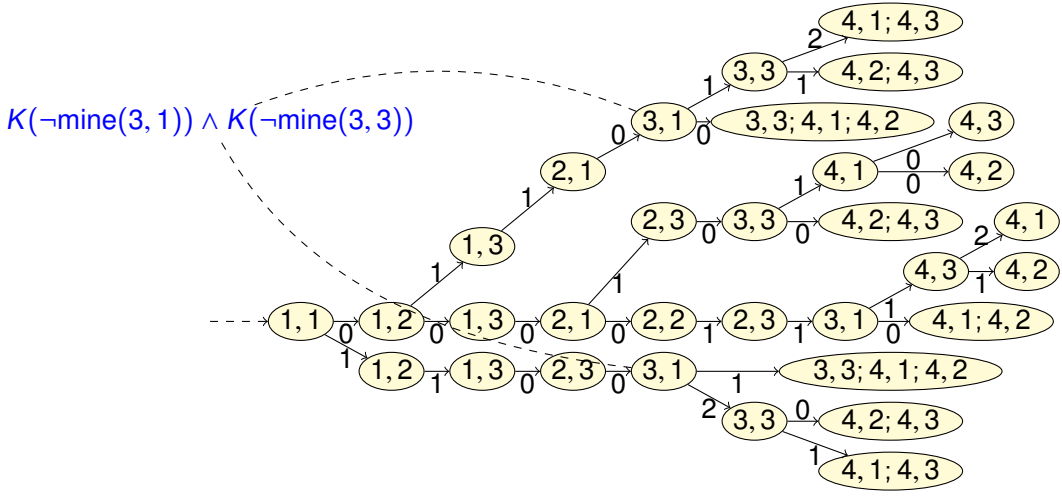
  if $K\neg\text{mine}(4,3)$ **then** CLICK$(4,3)$ **else** ε **fi**

od

---

23. [Zanuttini et al., 2020]

Intuition : several histories lead to same sufficient knowledge



- Proved : KBP always as succinct as reactive policy ; possibly exponentially more
- KBP explainable

Executing a KBP :

- maintain knowledge
- decide branching conditions
- this is (single-agent) epistemic logic !
- no free lunch : execution is Θ_2^P-complete
- computing plans mostly open

Not pretending that there are no other approaches for POMDPs/contingent :

- dedicated algorithms
- forward, backward, heuristic, complete...
- machine learning...

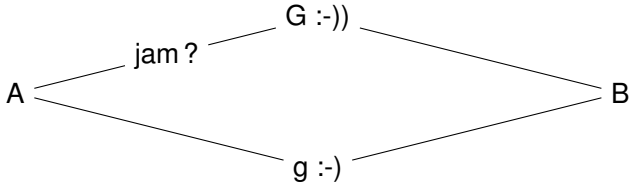
Section 12

Multi-Agent Setting

Setting :

- multi-agent, collaborative
- offline planning, centralized
- online execution, decentralized, no explicit communication

Example :



radio but no cell phone

Definition (Dec-POMDP)

Decentralized POMDP :

- sets of agents I, states S, actions A, observations Ω
- transition function $T : S \times A^I \to \Delta(S)$
- reward function $R : S \to \mathbb{R}$
- observation function $O : S \times A^I \times S \to \Delta(\Omega^I)$
- initial common belief state $B_0 \in \Delta(S)$

Definition (Dec-POMDP)

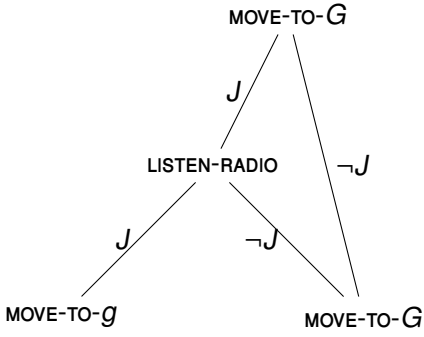
Decentralized POMDP :

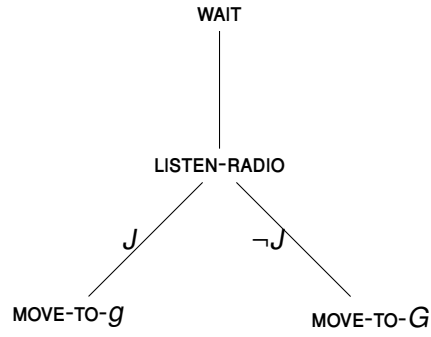
- sets of agents I, states S, actions A, observations Ω
- transition function $T : S \times A^I \to \Delta(S)$
- reward function $R : S \to \mathbb{R}$
- observation function $O : S \times A^I \times S \to \Delta(\Omega^I)$
- initial common belief state $B_0 \in \Delta(S)$

Joint policy :

- policy π for each agent
- policy of A = function from observation history of A
- value = expected reward of joint policy

Example Policy





Natural generalization of single-agent case :

- maintain belief over state : $B \in \Delta(S)$
- not sufficient !
- should distinguish :
  - there is a traffic jam and B knows this
  - there is a traffic jam and B does not know

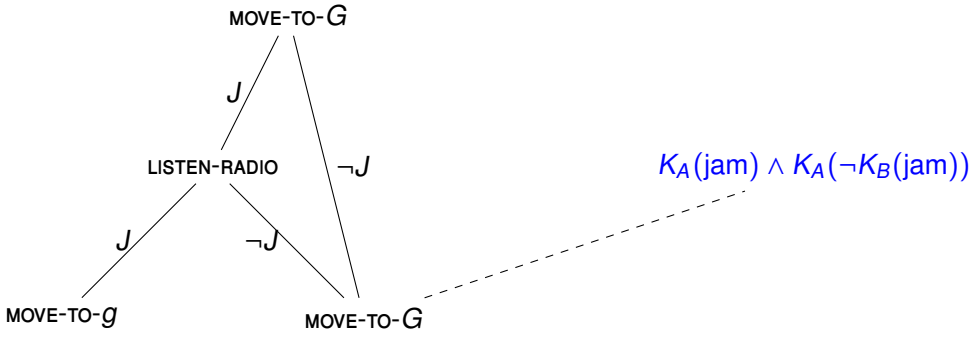
Natural generalization of single-agent case :

- maintain belief over state : $B \in \Delta(S)$
- not sufficient !
- should distinguish :
  - there is a traffic jam and B knows this
  - there is a traffic jam and B does not know

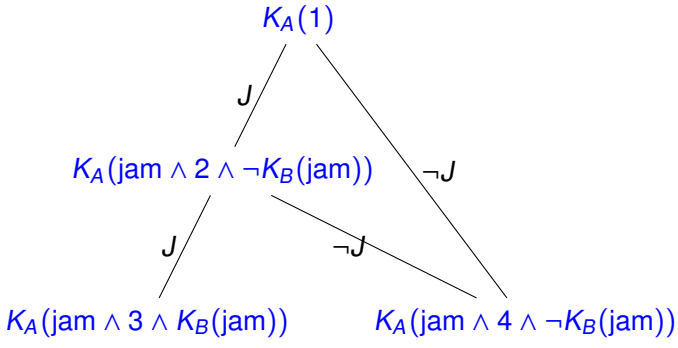
Each agent must maintain multi-agent knowledge !

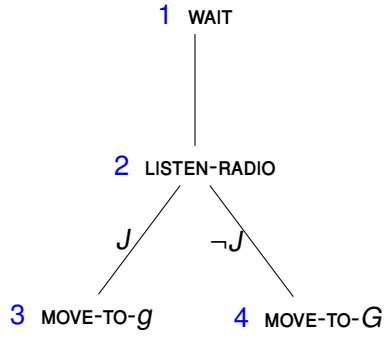
- up to any depth
- this is reasoning in DEL

Implicit anyway :

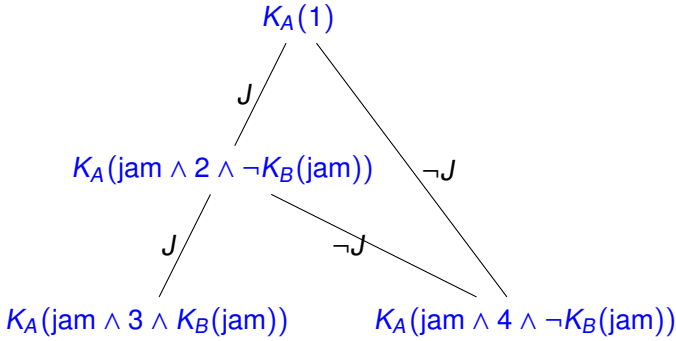


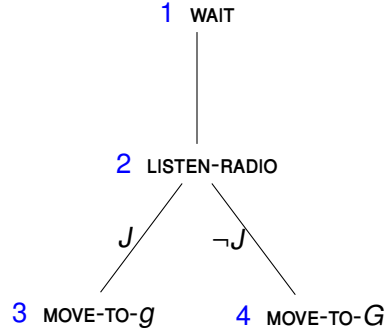
Maintain knowledge about state + other agents' "program counters"





Maintain knowledge about state + other agents' "program counters"





Notes :

- centralized planning is crucial
- knowledge about B's program counters may be imprecise, like $K_A(1 \vee 3)$

Multi-Agent KBP[24] for A :

---

while $\top$ **do**

if $K_A(\neg \text{jam}) \vee \left(\neg K_A(\text{jam}) \wedge \neg K_A(\neg \text{jam})\right)$ **then** MOVE-TO-G

else if $K_A(\text{jam}) \wedge \neg K_A(K_B(\text{jam})) \wedge \neg K_A(\neg K_B(\text{jam}))$ **then** LISTEN-RADIO

else if $K_A(\text{jam}) \wedge K_A(K_B(\text{jam}))$ **then** MOVE-TO-g

else if $K_A(\text{jam}) \wedge K_A(\neg K_B(\text{jam}))$ **then** MOVE-TO-G

od

---

and similar for B

24. [Saffidine et al., 2018]

Multi-Agent KBP [24] for A :

---

while $\top$ **do**

if $K_A(\neg\text{jam}) \vee \left(\neg K_A(\text{jam}) \wedge \neg K_A(\neg\text{jam})\right)$ **then** MOVE-TO-G

else if $K_A(\text{jam}) \wedge \neg K_A(K_B(\text{jam})) \wedge \neg K_A(\neg K_B(\text{jam}))$ **then** LISTEN-RADIO

else if $K_A(\text{jam}) \wedge K_A(K_B(\text{jam}))$ **then** MOVE-TO-g

else if $K_A(\text{jam}) \wedge K_A(\neg K_B(\text{jam}))$ **then** MOVE-TO-G

od

---

and similar for B

Properties :

- as succinct and possibly exponentially more than reactive policies
- execution : maintain multi-agent belief, incl. program counters

24. [Saffidine et al., 2018]

Observation histories [25] :

- each agent maintains belief over joint histories
- implicitly designates (distribution over) states and program counters
- used at planning time
- very efficient approaches to planning

25. [Dibangoye et al., 2016]
26. [Nebel et al., 2019]

Observation histories [25] :

- each agent maintains belief over joint histories
- implicitly designates (distribution over) states and program counters
- used at planning time
- very efficient approaches to planning

Implicitly coordinated policies [26] :

- relaxes centralized planning
- *A* takes action if *A* knows that *B* has enough knowledge to find a corresponding plan

25. [Dibangoye et al., 2016]
26. [Nebel et al., 2019]

Section 13

Wrap-Up and Further Topics

Planning under partial observability :

- knowledge is acquired, progressed through execution
- policies embed knowledge implicitly or explicitly, cf belief tracking [27]
- explicit representations trade exec. efficiency for succinctness
- execution appeals to (single-agent) epistemic reasoning

Planning for decentralized, collaborative multi-agent execution :

- essentially same remarks/results
- but knowledge over multi-agent knowledge + others' execution
- execution appeals to (general) DEL

27. [Geffner and Bonet, 2013]

Reinforcement learning :

- unknown model
- learn model or policy by exploring belief space

Games :

- adversarial setting
- similar to contingent if single-agent, robust/minmax
- reasoning with explicit knowledge/opponent models might help

Bonet, B. and Geffner, H. (2000).
Planning with incomplete information as heuristic search in belief space.
In *Proc. AIPS 2000*.

Cimatti, A. and Roveri, M. (2000).
Conformant planning via symbolic model checking.
Journal of Artificial Intelligence Research, 13 :305–338.

Dibangoye, J. S., Amato, C., Buffet, O., and Charpillet, F. (2016).
Optimally solving dec-pomdps as continuous-state mdps.
J. Artif. Intell. Res., 55 :443–497.

Fernandez Davila, J. L. (2022).
Planification cognitive basée sur la logique : de la théorie à l'implémentation.
PhD thesis, Université Toulouse III — Paul Sabatier, France.

Geffner, H. and Bonet, B. (2013).
A concise introduction to models and methods for automated planning.
Synthesis Lectures on Artificial Intelligence and Machine Learning, 8(1) :1–141.


Lorini, E. (2020).
Rethinking epistemic logic with belief bases.
Artificial Intelligence, 282.


Nebel, B., Bolander, T., Engesser, T., and Mattmüller, R. (2019).
Implicitly coordinated multi-agent path finding under destination uncertainty : Success guarantees and computational complexity.
J. Artif. Intell. Res., 64 :497–527.


Palacios, H. and Geffner, H. (2009).
Compiling uncertainty away in conformant planning problems with bounded width.
Journal of Artificial Intelligence Research.


Saffidine, A., Schwarzentruber, F., and Zanuttini, B. (2018).
Knowledge-based policies for qualitative decentralized pomdps.
In McIlraith, S. A. and Weinberger, K. Q., editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 6270–6277. AAAI Press.


Son Thanh To, Tran Cao Son, E. P. (2017).
A generic approach to planning in the presence of incomplete information : Theory and implementation (extended abstract).
In *Proc. IJCAI 2017*.


Zanuttini, B., Lang, J., Saffidine, A., and Schwarzentruber, F. (2020).
Knowledge-based programs as succinct policies for partially observable domains.
Artif. Intell., 288 :103365.