Planification temporelle dans l'incertain et actions paramétriques

Emmanuel Rachelson [1]

Frédérick Garcia [2]
Florent Teichteil [1]

[1]ONERA-DCSD — Toulouse

[2]INRA-BIA —Toulouse

04 juillet 2007

Plan

Plan

Effets incertains et actions paramétriques

Exemple 1 : Exemple jouet, naviguer dans un labyrinthe

“avancer”, d’un pas, de deux ?

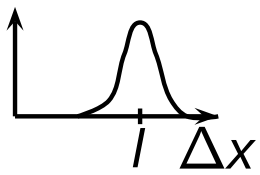
$\rightarrow$ *Dissocier en autant d’actions*

Mais avancer de 1,34 pas ?

Effets incertains et actions paramétriques

Exemple 2 : $avancer(l)$

$$P(x'|x, avancer(l)) = P_{avancer}(x' - x|l) =$$

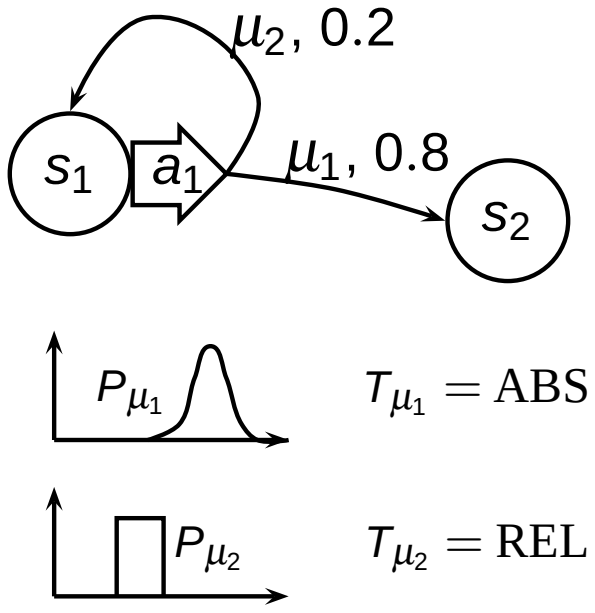


action paramétrique à paramètre continu et à effets probabilistes.

Effets incertains et actions paramétriques

Exemple 3 : Cadre TMDP

- S, A
- $L(\mu|s,t,a)$
- $P_\mu(t'), P_\mu(t'-t)$
- $r(\mu,t,t')$
- $K(s,\theta)$



a_1 → action sans paramètre à effets probabilistes continus

attendre → action à paramètre continu à effets déterministes

Plan

Cadre de représentation ?

Pas de cadre existant pour exprimer l'existence d'actions discrètes à paramètre continu.

On va proposer un cadre d'écriture générique pour les problèmes à actions continues

. . . afin de prouver que les méthodes classiques de recherche de stratégies optimales sont toujours valables.

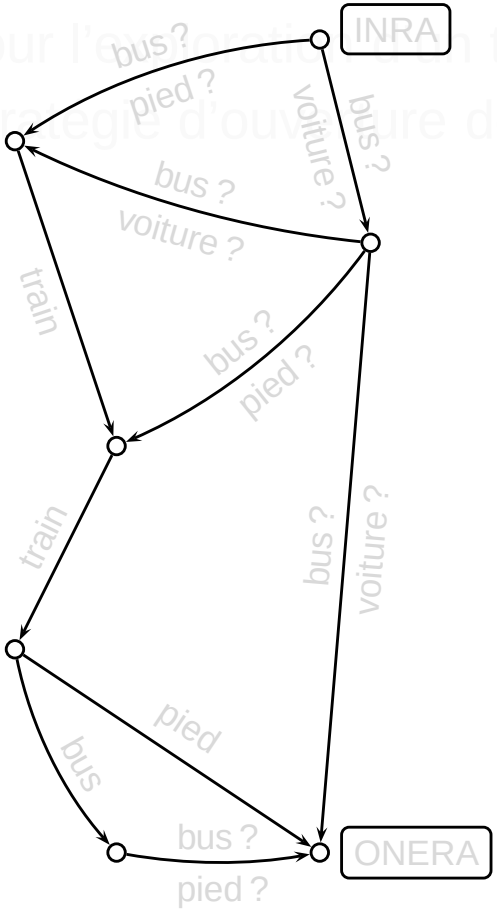
Plan

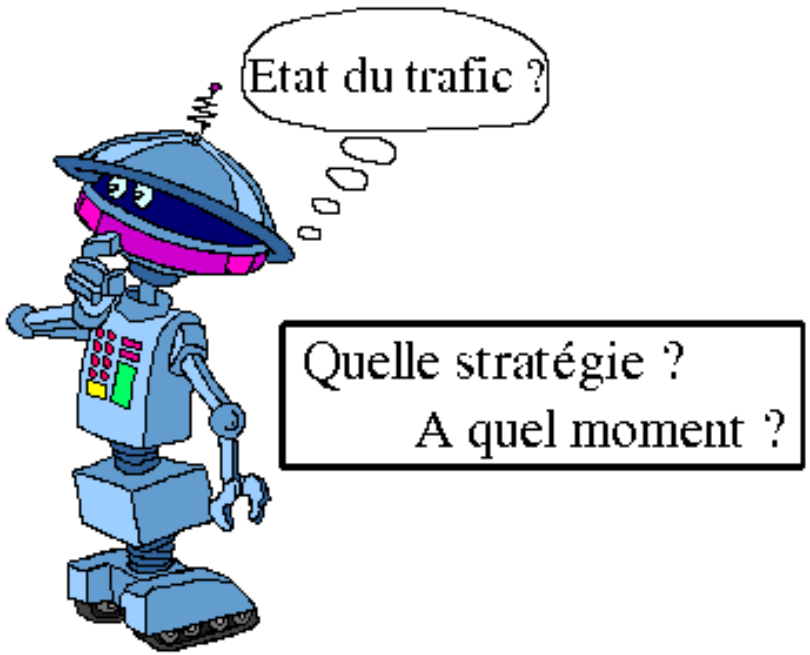
Problèmes types

- Gestion des activités d'un satellite
- Optimisation de la statégie de déplacement ONERA → INRA
- Coordination des actions d'un binôme d'agents aéroterrestre pour l'exploration d'un terrain
- Stratégie d'ouverture des guichets d'un aéroport

Problèmes types

- Gestion des activités d'un satellite
- Optimisation de la statégie de déplacement ONERA $\longrightarrow$ INRA
- Coordination des actions d'un binôme d'agents aéroterrestre pour l'exploration d'un terrain
- Stratégie d'ouverture des guichets d'un aéroport





Problèmes types

- Gestion des activités d'un satellite
- Optimisation de la statégie de déplacement ONERA $\longrightarrow$ INRA
- Coordination des actions d'un binôme d'agents aéroterrestre pour l'exploration d'un terrain
- Stratégie d'ouverture des guichets d'un aéroport



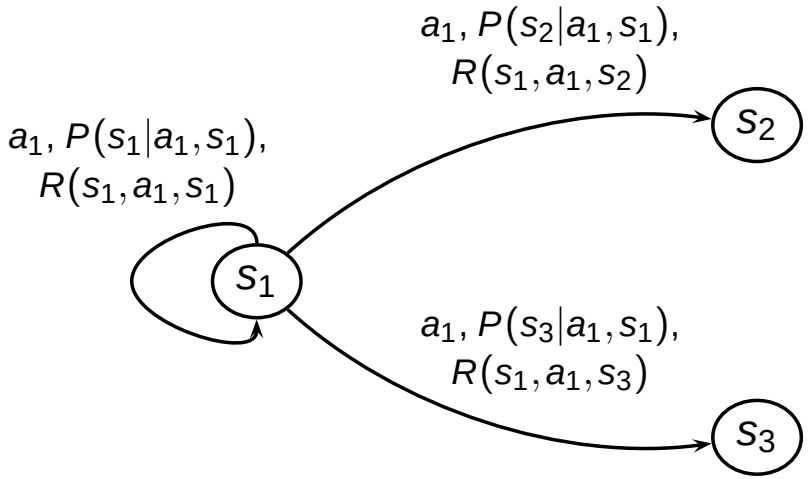
Problèmes types

- Gestion des activités d'un satellite
- Optimisation de la statégie de déplacement ONERA $\longrightarrow$ INRA
- Coordination des actions d'un binôme d'agents aéroterrestre pour l'exploration d'un terrain
- Stratégie d'ouverture des guichets d'un aéroport

Plan

Rappels MDP

$$\text{MDP} : < S, A, P, r, T >$$



$$\pi(s) \leftrightarrow V^{\pi}(s)$$

$$V^*(s) = \max_{a \in A} \left\{ r(s,a) + \gamma \sum_{s' \in S} P(s'|s,a) V^*(s') \right\}$$

Définition d'un XMDP

MDP à actions paramétriques ou XMDP : $S, A(X), p, r, T$

- Espace d'états S, variables continues et discrètes, incluant la variable temporelle.
- Espace d'actions paramétriques $a(x) \in A(X)$ dénombrable.
- Modèle de transition $p(s'|s, a(x))$.
- Modèle de récompense $r(s, a(x))$.

Définition d'un XMDP

MDP à actions paramétriques ou XMDP : $S, A(X), p, r, T$

- Espace d'états S, variables continues et discrètes, incluant la variable temporelle.
- Espace d'actions paramétriques $a(x) \in A(X)$ dénombrable.
- Modèle de transition $p(s'|s, a(x))$.
- Modèle de récompense $r(s, a(x))$.

Définition d'un XMDP

MDP à actions paramétriques ou XMDP : $S, A(X), p, r, T$

- Espace d'états S, variables continues et discrètes, incluant la variable temporelle.
- Espace d'actions paramétriques $a(x) \in A(X)$ dénombrable.
- Modèle de transition $p(s'|s, a(x))$.
- Modèle de récompense $r(s, a(x))$.

Définition d'un XMDP

MDP à actions paramétriques ou XMDP : $S, A(X), p, r, T$

- Espace d'états S, variables continues et discrètes, incluant la variable temporelle.
- Espace d'actions paramétriques $a(x) \in A(X)$ dénombrable.
- Modèle de transition $p(s'|s, a(x))$.
- Modèle de récompense $r(s, a(x))$.

Politique XMDP

On cherche des solutions au problème de décision sous la forme de politiques :

$$\pi : \begin{cases} S & \rightarrow & A(X) \\ s & \mapsto & a(x) \end{cases}$$

En mettant la variable t en évidence :

$$\pi : \begin{cases} S \times \mathbb{R} & \rightarrow & A(X) \\ s,t & \mapsto & a(x) \end{cases}$$

Hypothèses

- r est bornée,
- r est semi-continue supérieure,
- les durées de transition sont minorées par un réel strictement positif.

Plan

Critère γ-pondéré

$$V^{\pi}(s) = E\left(\sum_{i=1}^{\infty} \gamma^{t_i} r(s_i, t_i, \pi(s_i, t_i))\right)$$

$$L^{\pi}V(s,t) = r(s,t,\pi(s,t)) + \int_{\substack{t' \in \mathbb{R} \\ s' \in S}} \gamma^{t'-t} p(s',t'|s,t,\pi(s,t)) V(s',t') ds' dt'$$

$$V^{\pi} = L^{\pi}V^{\pi}$$

Plan

Equation d'optimalité

$$V^* = LV^*$$

$$V^*(s,t) = \sup_{\pi} \left\{ r_\pi(s,t) + \int_{\substack{t' \in \mathbb{R} \\ s' \in S}} \gamma^{t'-t} p_\pi(s',t'|s,t) V^*(s',t') ds' dt' \right\}$$

$$V^*(s,t) = \sup_{a(x) \in A(X)} \left\{ r(s,t,a(x)) + \int_{\substack{t' \in \mathbb{R} \\ s' \in S}} \gamma^{t'-t} \, p(s',t'|s,t,a(x)) \, V^*(s',t') ds' dt' \right\}$$

Avec $x = \tau$ (cas TMDP), on a :

$$V^*(s,t) = \max_{a \in A} \sup_{\tau \in \mathbb{R}^+} \left\{ r(s,t,a(\tau)) + \int_{\substack{t' \in \mathbb{R} \\ s' \in S}} \gamma^{t'-t} \, p(s',t'|s,t,a(\tau)) \, V(s',t') ds' dt' \right\}$$

ONERA THE FRENCH AEROSPACE LAB INRA

Plan

Boyan and Littman, 01

$$U(\mu,t) = \begin{cases} \int_{-\infty}^{\infty} P_\mu(t')[R(\mu,t,t') + V(s'_\mu,t')]dt' & \text{si } T_\mu = \text{ABS} \\ \int_{-\infty}^{\infty} P_\mu(t'-t)[R(\mu,t,t') + V(s'_\mu,t')]dt' & \text{si } T_\mu = \text{REL} \end{cases} \quad (1)$$

$$Q(s,t,a) = \sum_{\mu \in M} L(\mu|s,t,a) \cdot U(\mu,t) \quad (2)$$

$$\overline{V}(s,t) = \max_{a \in A} Q(s,t,a) \quad (3)$$

$$V(s,t) = \sup_{t' \geq t} \left(\int_t^{t'} K(s,\theta)d\theta + \overline{V}(s,t') \right) \quad (4)$$

Un cas particulier de XMDP

$$p(s',t'|s,t,a) = \sum_{\mu} \delta_{s'}(s'_{\mu})\delta_{ABS}(T_{\mu})L(\mu|s,t,a)P_{\mu}(t') + \sum_{\mu} \delta_{s'}(s'_{\mu})\delta_{REL}(T_{\mu})L(\mu|s,t,a)P_{\mu}(t'-t)$$

$$p(s',t'|s,t,attendre(\tau)) = \delta_s(s')\delta_{t+\tau}(t')$$

Les équations 1 à 4 sont équivalentes à l'équation de Bellman pour les XMDP.

Plan

Extension de TMDP

(Boyan & Littman, 01) :

- P_μ densité d'une distribution discrète
- $L(\mu|s,t,a)$ constante par morceaux
- $r(\mu,t)$, $r(\mu,t')$, $r(\mu,t'-t)$ linéaires par morceaux.

$\Rightarrow$ résolution exacte.

En fait, la résolution exacte est possible si $deg(r) < 5$.

Hypothèses trop restictives :

Extension à des fonctions P_μ, L et r polynômiales par morceaux

Itération de la valeur sur TMDP

$$\text{Si } \begin{cases} deg(P_\mu) = a \\ deg(r_i) = b \\ deg(L) = c \\ deg(V_n) = d > b \end{cases} \quad \text{alors } deg(V_{n+1}) = a + d + c + 1.$$

On ajoute une équation aux équations 1 à 4 pour projeter V_{n+1} sur les polynômes définis par morceaux de degré d.

$\rightarrow$ algorithme approché d'itération de la valeur.

Plan

Idée

La politique solution s'écrit sous forme de *timelines* en chaque état.

Recherche d'un découpage de t par état et de l'action à entreprendre sur chaque intervalle.

Cette approche exploite le fait que l'action attendre a les deux propriétés :

- action déterministe sur la variable t
- $attendre(t') \Leftrightarrow$ *ne rien faire jusqu'à t'*

Méthode

1. Projeter le XMDP en un MDP discret où les intervalles des timelines courantes sont des états discrets.
2. Calculer une politique optimale pour le MDP discret
3. Evaluer dans chaque état la date où on peut le plus améliorer la politique courante
4. Modifier les timelines de la politique trouvée
5. Itérer jusqu'à ce que l'amélioration de la phase 3 soit inférieure à ε

Généralisation : itération de la politique approchée.

Plan

Itération de la politique approchée

1. Evaluation approchée de la valeur de la politique courante
2. Amélioration sur un coup

Intérêt : converge vers l'optimum même avec une evaluation approchée.

Etape 1 : Evaluation approchée de V^{π_n}

Variantes pour l'étape 1 :

- Evaluation constante par morceaux par projection en MDP discret
- Evaluation approchée polynômiale par morceaux
- Résolution par programmation linéaire de la projection de V sur une base de fonctions

Etape 2 : Amélioration sur un coup

Variantes pour l'étape 2 :

- Calcul de l'erreur de Bellman pour trouver un nouveau couple (t, a)
- Echantillonnage dans chaque intervalle des timeline, évaluation et choix de la pire valeur pour amélioration (approche heuristique).

Conclusion

Travaux en cours

- Premiers résultats sur TMDPpoly
- Implantation des algorithmes SMDP+ et itération de la politique approchée pour comparaison
- Application aux problèmes de coordination dans l'incertain : binôme d'agents, satellite d'observation de la Terre, ...

Merci de votre attention !