Algorithme approché d'optimisation d'un modèle de Processus Décisionnel de Markov sur Graphe

Nathalie Peyrard Régis Sabbadin

INRA-MIA Avignon et Toulouse
E-Mail: {peyrard,sabbadin}@toulouse.inra.fr

Réseau MSTGA, Avignon, 20-21 Octobre 2007

Processus Décisionnels de Markov (1)

Définition (Processus Décisionnel de Markov)

Un PDM est défini par un quadruplet $< \mathcal{X}, \mathcal{A}, p, r >$ *:*

- $\mathcal{X} = \{x^1, \ldots, x^{|\mathcal{X}|}\}$. *Etats possibles du système*
- $\mathcal{A} = \{a^1, \ldots, a^{|\mathcal{A}|}\}$. *Actions applicables*
- $p(x'|x, a)$. *Probabilité de transition entre états*
- $r(x, a)$. *Fonction de récompense "immédiate".*

- Politique : $\delta : \mathcal{X} \rightarrow \mathcal{A}$
- Trajectoire : $\tau =< x_0, \delta(x_0), x_1, \delta(x_1), \ldots, x_t, \delta(x_t), \ldots >$, $t \in H \subseteq \{1 \ldots + \infty\}$
- Valeur d'une politique :

$$V_\delta(x_0) = E\left[\sum_{t \in H} \gamma^t r_t(x_t, \delta(x_t))\right], 0 \leq \gamma \leq 1.$$

Processus Décisionnels de Markov (2)

Définition (Solution d'un PDM)

Une politique δ^* est solution d'un PDM $< \mathcal{X}, \mathcal{A}, p, r >$ ssi

$$V_{\delta^*}(x) \geq V_{\delta}(x), \forall \delta, \forall x$$

Proposition (Existence d'une politique optimale)

Une politique optimale δ^* existe et est solution du système (non linéaire) :

$$V_{\delta^*}(x) = \max_{a \in \mathcal{A}} \left\{ r(x,a) + \gamma \cdot \sum_{x' \in \mathcal{X}} p(x'|x,a) \cdot V_{\delta^*}(x) \right\}, \forall x \in \mathcal{X}$$

$$\delta^*(x) = \arg\max_{a \in \mathcal{A}} \left\{ r(x,a) + \gamma \cdot \sum_{x' \in \mathcal{X}} p(x'|x,a) \cdot V_{\delta^*}(x) \right\}, \forall x \in \mathcal{X}$$

Résolution d'un PDM

Trouver une politique optimale δ^*, telle que $V_{\delta^*}(x) \geq V_\delta(x), \forall x$

Algorithme (Itération de la Politique)

Alterne évaluation *et* amélioration *d'une politique courante*

- *Évaluation d'une politique δ : Système linéaire*

$$V_\delta(x) = r(x, \delta(x)) + \gamma \cdot \sum_{x' \in \mathcal{X}} p(x'|x, \delta(x)) \cdot V_\delta(x'), \forall x$$

- *Amélioration de la politique :*

$$\delta'(x) = argmax_{a \in \mathcal{A}}(r(x, a) + \gamma \cdot \sum_{x' \in \mathcal{X}} p(x'|x, a) \cdot V_\delta(x')), \forall x$$

Propriété : $V_{\delta'}(x) \geq V_\delta(x), \forall x$ et $V_{\delta'} = V_\delta \Rightarrow \delta$ optimale.

Evaluation de la politique et mesure d'occupation

Définition (Mesure d'occupation)

Soit une politique $\delta : \mathcal{X} \to \mathcal{A}$ *et* $x \in \mathcal{X}$ *un état initial.*
La mesure d'occupation $P_{x,\delta,\gamma} : \mathcal{X} \to [0,1]$ *est définie par :*

$$\forall y \in \mathcal{X}, P_{x,\delta,\gamma}(y) = (1-\gamma) \sum_{t=0}^{+\infty} \gamma^t \cdot P(X^t_{x,\delta} = y)$$

Proposition (Fonction de valeur et mesure d'occupation)

$$\forall x \in \mathcal{X}, V_\delta(x) = \frac{1}{1-\gamma} \cdot \sum_{y \in \mathcal{X}} P_{x,\delta,\gamma}(y) \cdot r(y, \delta(y))$$

Limites de l'Itération de la Politique

Complexité trop élevée lorsque $|\mathcal{X}|$ est grand :

- Evaluation. Calcul de $P_{x,\delta,\gamma}(y)$ en tout x, y.
- Amélioration Calcul d'un *argmax* sur $\mathcal{A} \times \mathcal{X}$.

$\Rightarrow$ Approximation de $P_{x,\delta,\gamma}(y)$ en champ moyen

Processus Décisionnels de Markov sur Graphes

Définition (Processus Décisionnel de Markov sur Graphes)

Un Processus Décisionnel de Markov sur Graphes est un PDM défini par un quintuplet $< \mathcal{X}, \mathcal{A}, p, r, G >$ *:*

- $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ *: espace d'états multidimensionnel*
- $\mathcal{A} = \mathcal{A}_1 \times \ldots \times \mathcal{A}_n$ *: espace d'actions multidimensionnel*
- *p et r : fonctions de transition et de récompense locales*
- $G = (V, B)$ *: graphe orienté de sommets* $V = \{1, \ldots, n\}$ *et d'arêtes* $B \subseteq V^2$ *exprimant les dépendances entre variables*

On utilisera aussi la notion de voisinage d'un sommet i :

$$N(i) = \{j \in \{1, \ldots, n\}, (j, i) \in B\}$$

Localité d'un PDMG

Définition (Processus local)

Le processus est dit local si $p(x'|x,a)$ *s'écrit :*

$$\forall x, x' \in \mathcal{X}^2, \forall a \in \mathcal{A}, p(x'|x,a) = \prod_{i=1}^{n} p_i(x'_i|x_{N(i)}, a_i)$$

Définition (Politique locale)

Une politique $\delta : \mathcal{X} \to \mathcal{A}$ *est dite locale ssi* $\delta = (\delta_1, \ldots, \delta_n)$ *où* $\delta_i : \mathcal{X}_{N_{(i)}} \to \mathcal{A}_i$ *et* $\delta_i(x_{N(i)}) = a_i \in \mathcal{A}_i$, $\forall x_{N(i)} \in \mathcal{X}_{N_{(i)}}$.

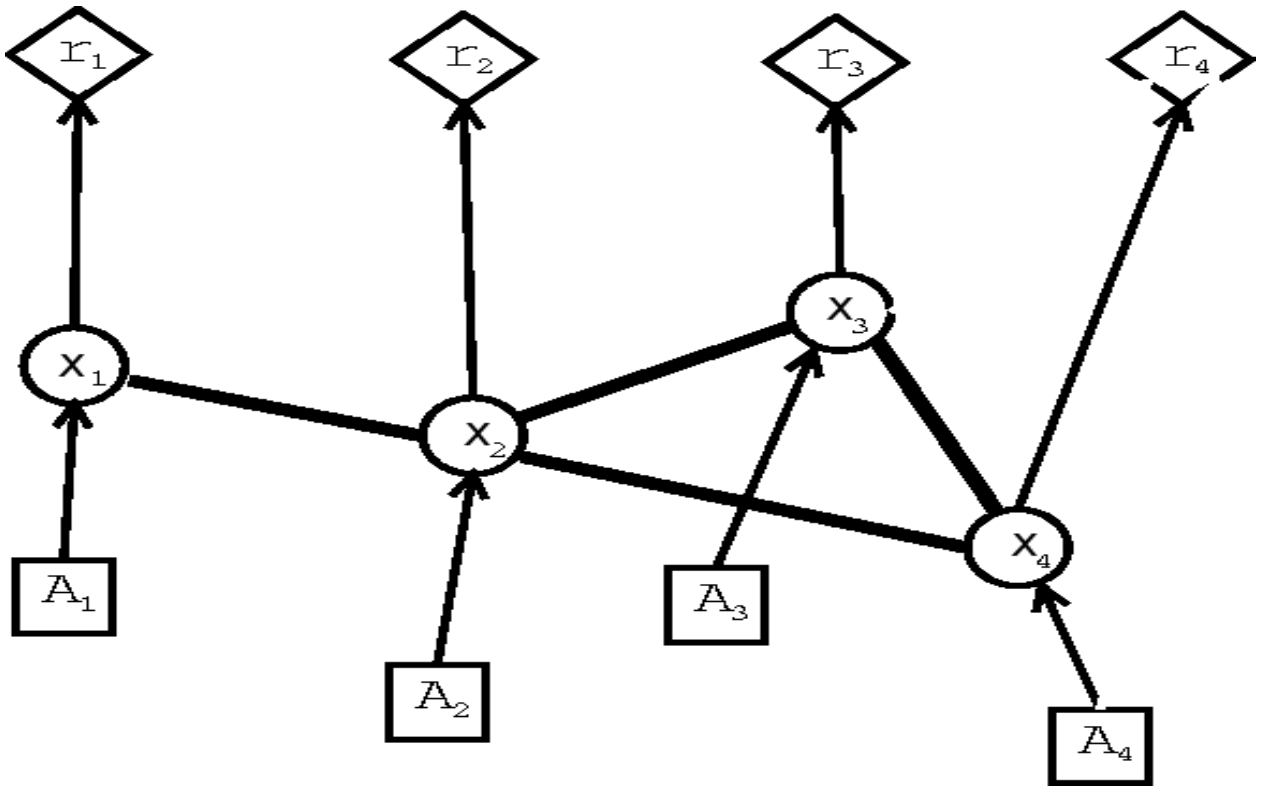
Définition (Récompense locale)

Une fonction de récompense $r : \mathcal{X} \times \mathcal{A} \to R$ *est dite locale ssi*

$$\forall x, a \in \mathcal{X} \times \mathcal{A}, r(x,a) = \sum_{i \in V} r_i(x_{N(i)}, a_i).$$

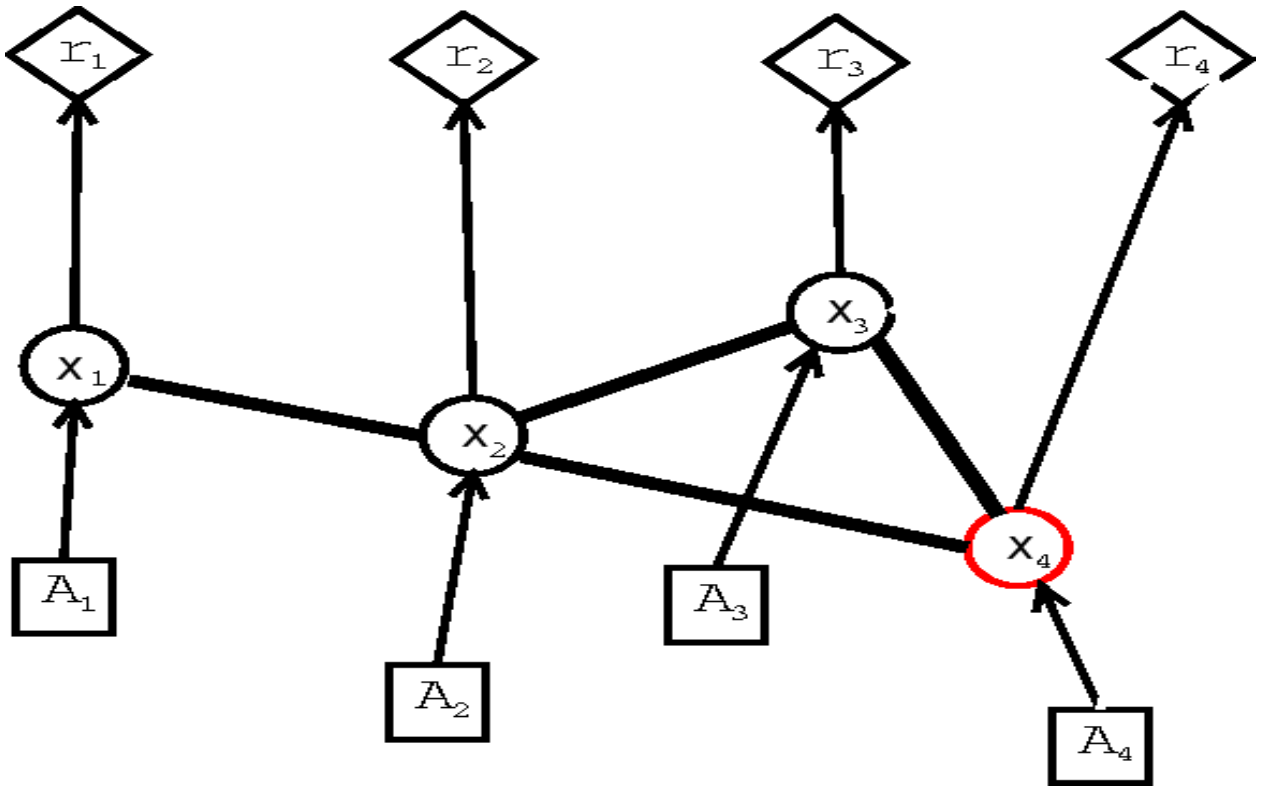
Exemple

Un Processus Décisionnel Markovien sur Graphe



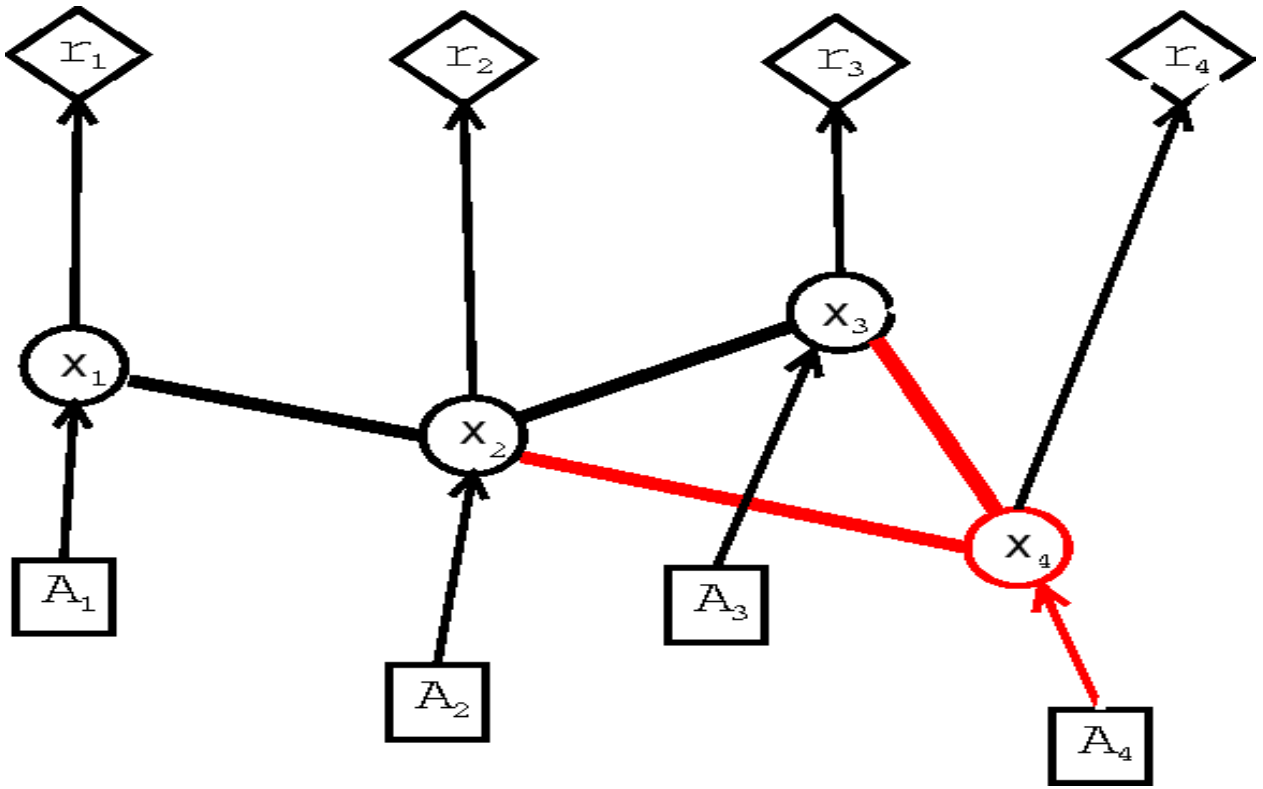
Exemple

Localité autour du noeud 4



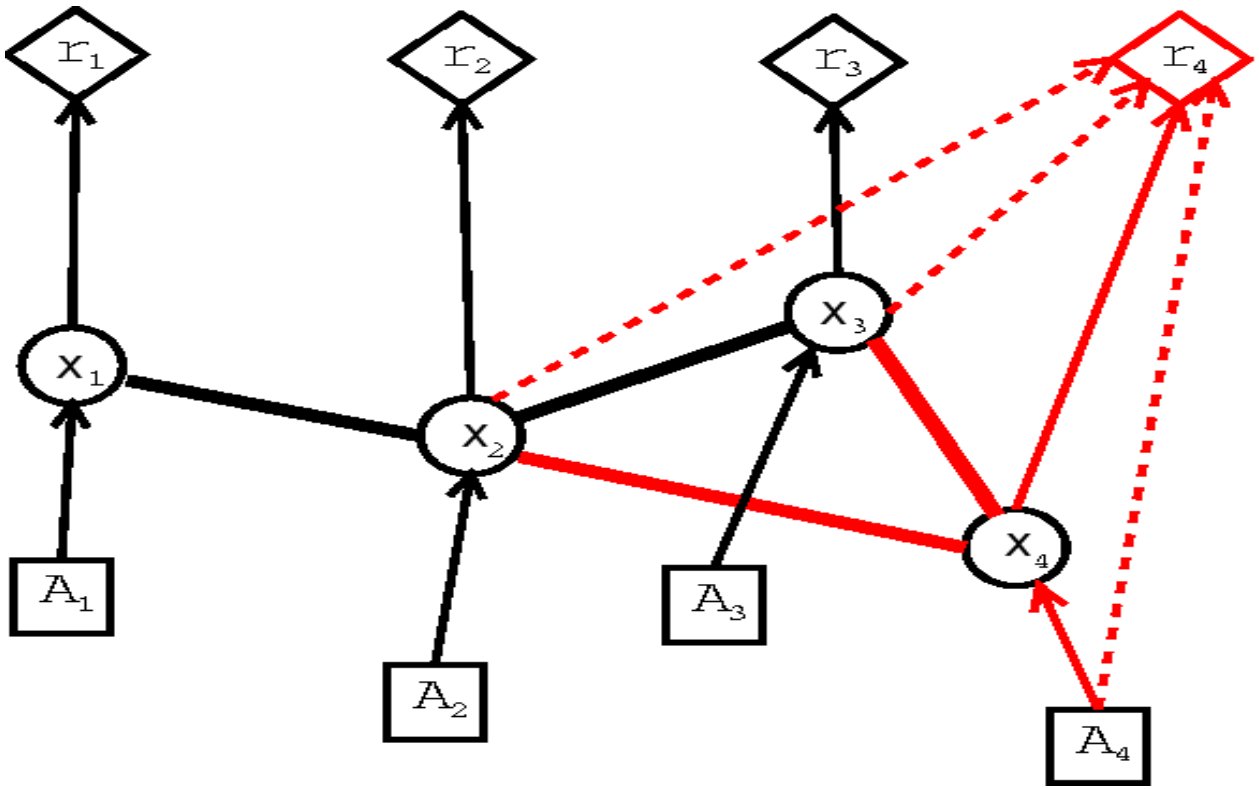
Exemple

Probabilité de transition locale



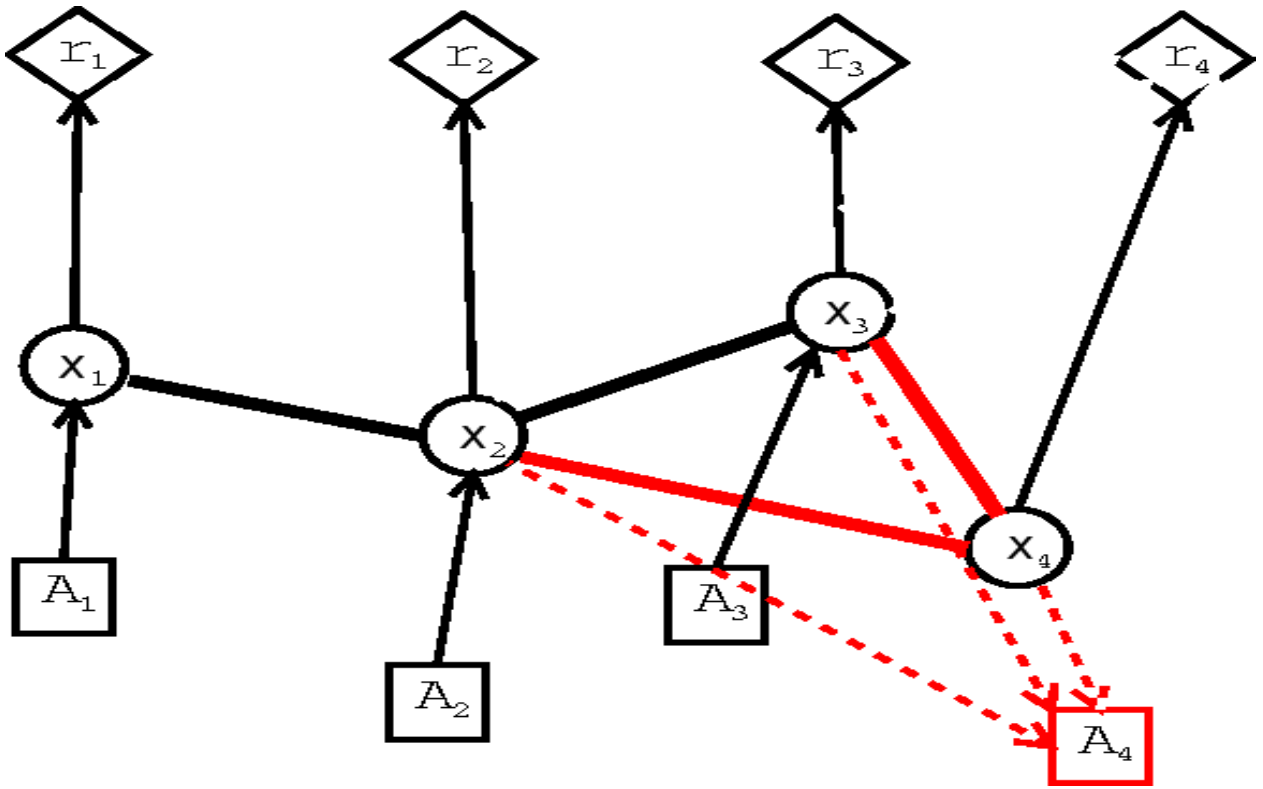
Exemple

Récompense locale

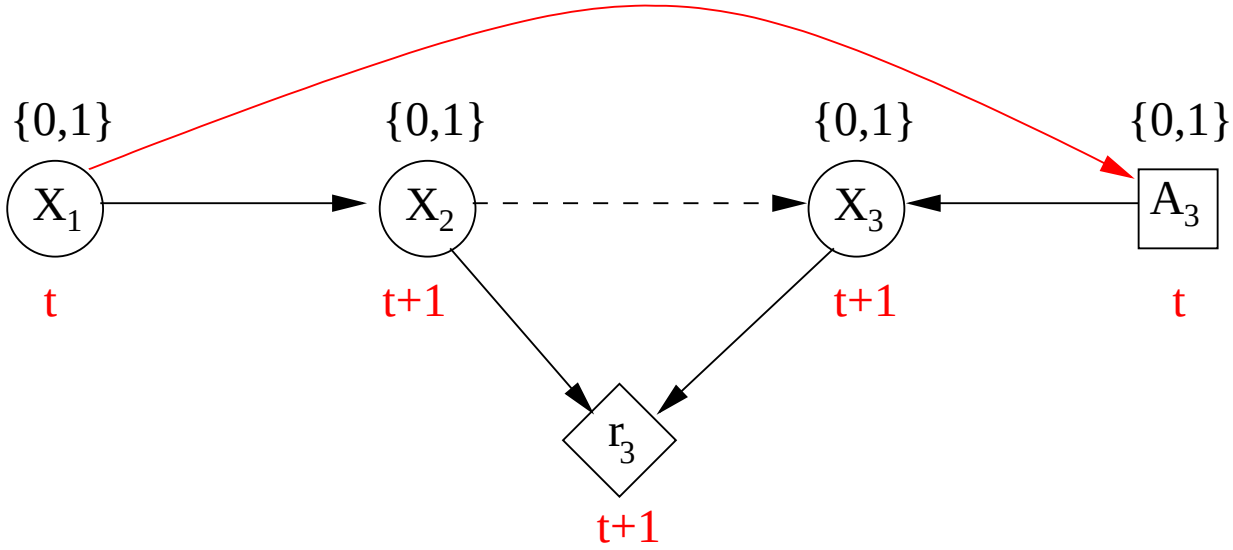


Exemple

Politique locale



Non-optimalité des politiques locales



- $p_1(x_1^{t+1} = 0) = p_1(x_1^{t+1} = 1) = \frac{1}{2}$,
  $p_2(x_2^{t+1} = x_1^t | x_1^t) = p_3(x_3^{t+1} = z | a_3^t = z) = 1$
- $r_3(x_2^t, x_3^t) = 1$ si $x_2^t = x_3^t$ et 0 sinon.
- Politique locale : $\delta_3^l : (x_2^t, x_3^t) \rightarrow a_3^t$ $\quad V_{\delta_3^l}(x) = \frac{1}{2(1-\gamma)}, \forall x$
- Politique optimale : $a_3^t = \delta_3^*(x_1^t) = x_1^t$ $\quad V_{\delta_3^*}(x) = \frac{1}{(1-\gamma)}, \forall x$

Recherche d'une "bonne" politique locale

La "meilleure" politique locale d'un PDMG peut être arbitrairement loin de la politique optimale :
Si $|\mathcal{X}_i| = n$, $V_{\delta_3^l}(x) = \frac{1}{n} \cdot V_{\delta_3^*}(x), \forall x$, mais

- les politiques locales ne nécessitent qu'un espace "raisonnable" pour être décrites.
- La meilleure politique locale est "empiriquement" bonne.
- L'ensemble des politiques locales est de taille "raisonnable".

Questions :

- Algorithme de type *champ moyen* pour le calcul de "bonnes" politiques locales ?
- Evaluation de l'erreur liée à l'approximation ?

Décomposition de la fonction de valeur

Proposition

$$V_\delta(x) = \frac{1}{1-\gamma} \cdot \sum_{y \in \mathcal{X}} P_{x,\delta,\gamma}(y) \cdot \Big(\sum_{i=1}^{n} r_i(y_{N(i)}, \delta_i(y_{N(i)})) \Big)$$

$$\text{Soit, } V_\delta(x) = \sum_{i=1}^{n} V_\delta^i(x) \text{ où}$$

$$V_\delta^i(x) = \frac{1}{1-\gamma} \cdot \sum_{y \in \mathcal{X}} P_{x,\delta,\gamma}(y) \cdot r_i(y_{N(i)}, \delta_i(y_{N(i)}))$$

Problème ! V_δ^i dépend toujours de x en entier !

Décomposition par approximation en champ moyen

Rappel

$$P_{x,\delta,\gamma}(y) = (1-\gamma)\sum_{t=0}^{+\infty}\gamma^t \cdot P(X^t_{x,\delta} = y)$$

Définition (Approximation en CM de la mesure d'occupation)

$$P(X^t_{x,\delta} = y) = P_\delta(X^t = y | X^0 = x) \approx \prod_i \hat{C}^{i,t}_\delta(X^t_i = y_i | X^0_i = x_i)$$

Décomposition par approximation en champ moyen

Définition

$$\hat{C}_{\delta}^{N(i),t}(x_{N(i)}, y_{N(i)}) = \prod_{j \in N(i)} \hat{C}_{\delta}^{j,t}(X_j^t = y_j | X_j^0 = x_j)$$

En utilisant l'approximation de la mesure d'occupation :

Proposition (Fonction de valeur approchée dans un PDMG)

$$V_{\delta}(x) \approx \sum_{i=1}^{n} \sum_{t=0}^{+\infty} \gamma^t \cdot \left(\sum_{y_{N(i)}} \hat{C}_{\delta}^{N(i),t}(x_{N(i)}, y_{N(i)}) \cdot r_i(y_{N(i)}, \delta_i(y_{N(i)})) \right)$$

$$V_{\delta}(x) \approx \sum_{i=1}^{n} \hat{V}_{\delta}^{i}(x_{N(i)})$$

Approximation des fonctions de transition locales

$\hat{Q}_\delta^{i,1}(y_i|x_i)$: approximation de $p_i^\delta(y_i|x_{N(i)}) = p_i(y_i|x_N(i), \delta(x_{N(i)}))$

$$\hat{Q}_\delta^{i,1}(y_i|x_i) = \frac{1}{|\mathcal{X}_{N(i)-i}|} \sum_{\mathcal{X}_{N(i)-i}} p_i^\delta(y_i|x_{N(i)})$$

Soit, si on pose $P^0(x) = \prod_{i=1}^{n} P^{0,i}(x_i)$ où $P^{0,i}$ uniforme,

$$\hat{Q}_\delta^{i,1}(y_i|x_i) = E_{P^0}\left[p_i^\delta(y_i|x_i, X_{N(i)-i})\right]$$

Approximation des fonctions de transition locales

Définition (Approximation de type champ moyen)

Une distribution $\mathcal{P}(u_1, \ldots, u_n)$ est approchée par une distribution indépendante $\prod_{i=1}^{n} \mathcal{Q}^i(u_i)$ minimisant la divergence de Kullback-Leibler :

$$KL(\mathcal{Q}|\mathcal{P}) = E_{\mathcal{Q}}\left[log\left(\frac{\mathcal{Q}}{\mathcal{P}}\right)\right].$$

Proposition (Minimisation de divergence de Kullback-Leibler)

$$\hat{Q}_{\delta}^{1} \approx \arg\min_{Q^1} KL\left(Q^1(X^1|X^0).P^0(X^0)|p^{\delta}(X^1|X^0).P^0(X^0)\right)$$

Remarque : minimisation sur la loi jointe de (X^0, X^1) !

Calcul itératif de $\hat{C}_{\delta}^{i,t}(X_i^t = y_i | X_i^0 = x_i)$

- $P_{\delta}^{t,i}(x_i)$ est la probabilité approchée que $X_i^t = x_i$
- $\hat{Q}_{\delta}^{i,t+1}(x_i^{t+1}|x_i^t)$ est la probabilité approchée de passer de x_i^t à x_i^{t+1} (à l'instant t)
- $\hat{C}_{\delta}^{i,t}(x_i^t|x_i^0)$ est la probabilité approchée de passer de x_i^0 à x_i^t en t pas de temps

Définition (Calcul de $\hat{Q}_{\delta}^{i,t}(x_i^t|x_i^{t-1})$, $\hat{C}_{\delta}^{i,t}(x_i^t|x_i^0)$, $P_{\delta}^{t,i}(x_i)$)

$$\begin{aligned}
\hat{Q}_{\delta}^{i,t}(x_i^t|x_i^{t-1}) &= E_{P^{t-1}}\left[p_i^{\delta}(x_i^t|x_i^{t-1}, X_{N(i)-1}^{t-1})\right] \\
\hat{C}_{\delta}^{i,t}(x_i^t|x_i^0) &= \sum_{x_i^t} \hat{Q}_{\delta}^{i,t}(x_i^t|x_i^{t-1}) \cdot \hat{C}_{\delta}^{i,t-1}(x_i^{t-1}|x_i^0) \\
P_{\delta}^{t,i}(x_i^t) &= \hat{C}_{\delta}^{i,t}(x_i^t|x_i^0) \cdot P_{\delta}^{0,i}(x_i^0)
\end{aligned}$$

Remarques

- $P_{x,\delta,\gamma}(y)$ remplacé par $\prod_{i=1}^{n} \hat{P}^{i}_{x_i,\delta,\gamma}(y_i)$, via l'approximation de $p_i^{\delta}(x_i^t | x_{N(i)}^{t-1})$ par $\hat{Q}_{\delta}^{i,t}(x_i^t | x_i^{t-1})$.
  Un processus interdépendant stationnaire est approché par un ensemble de processus indépendants non-stationnaires
- Deux facteurs d'approximation :
  - $\hat{Q}_{\delta}^{i,t}(x_i^t | x_i^{t-1})$ est obtenu via une "presque-minimisation" d'une divergence (approximation itérée à chaque t)
  - Approximation initale (choix d'une distribution initiale P^0)
- Questions liées à ces approximations
  - Borner l'erreur de $\hat{Q}_{\delta}^{i,t}(x_i^t | x_i^{t-1}) \rightarrow$ borner l'erreur de $\hat{V}_{\delta}(x)$
  - Choix judicieux de $P^0(x)$, entre loi uniforme, Dirac (si x^0 fixé) et approximation d'une mesure d'occupation simulée

Amélioration de la politique approchée

Amélioration de la politique :

$$\delta'(x) = \arg\max_{a \in \mathcal{A}}(r(x,a) + \gamma \cdot \sum_{x' \in \mathcal{X}} p(x'|x,a) \cdot \hat{V}_\delta(x')), \forall x$$

On peut montrer :

$$\delta'(x) = \arg\max_{a_1 \ldots a_n} \sum_{i=1}^{n} r_i(x_{N(i)}, a_i) + \gamma \sum_{x'_{N(i)}} P(x'_{N(i)}|x_{N(N(i))}, a_{N(i)}) \cdot \hat{V}^i_\delta(x'_{N(i)})$$

$$\text{Soit } \delta'(x) = \arg\max_{a_1 \ldots a_n} \sum_{i=1}^{n} f_i(x_{N(N(i))}, a_{N(i)})$$

$\Rightarrow$ Même si δ est locale, δ' ne l'est pas !

Amélioration de la politique approchée

$$\hat{\delta}'_i(x) = \arg\max_{a_i} r_i(x_{N(i)}, a_i) + \gamma \sum_{x'_{N(i)}} P^{\delta}(x'_{N(i)} | x_{N(N(i))}, a_i) \cdot \hat{V}^i_{\delta}(x'_{N(i)})$$

où $P^{\delta}(x'_{N(i)} | x_{N(N(i))}, a_i) = p_i(x'_i | x_{N(i)}, a_i) \cdot \prod_{j \in N(i)-i} p_j(x'_j | x_{N(j)}, \delta(x_{N(j)}))$,

mais $\hat{\delta}'_i(x)$ dépend de $x_{N(N(i))}$. On pose donc :

$$\hat{P}^{\delta}(x'_{N(i)} | x_{N(i)}, a_i) = E_{P^0}\left[P^{\delta}(x'_{N(i)} | x_{N(i)}, X_{N(N(i))-N(i)}, a_i)\right]$$

et $\hat{\delta}'_i(x) = \arg\max_{a_i} r_i(x_{N(i)}, a_i) + \gamma \sum_{x'_{N(i)}} \hat{P}^{\delta}(x'_{N(i)} | x_{N(i)}, a_i) \cdot \hat{V}^i_{\delta}(x'_{N(i)})$

Remarques

Encore deux approximations :

- On “améliore” tous les $\hat{\delta}'_i$ indépendament en considérant les δ_j fixés $\rightarrow$ Amélioration séquentielle ?
- Encore une approximation en espérance de $P^{\delta}(x'_{N(i)}|x_{N(N(i))}, a_i)$ par $\hat{P}^{\delta}(x'_{N(i)}|x_{N(i)}, a_i) \rightarrow P^0$ non uniforme ?

Itération de la politique approchée

L'algorithme d'itération de la politique approchée alterne :

- phase d'évaluation approchée
- et phase d'amélioration approchée (basée également sur le Champ Moyen)

Jusqu'à trouver deux politiques locales successives δ et δ' telles que

$$\hat{V}_{\delta} = \hat{V}_{\delta'}.$$

Conclusion provisoire

Un formalisme (PDMG) et une méthode de résolution approchée basée sur le *Champ Moyen* pour résoudre des problèmes de PDM factorisés.

- Une méthode approchée de complexité “linéaire” en le nombre de variables
- Bien que sans garantie, la méthode converge vers des politiques globales “apparement” de bonne qualité
- Autres méthodes de type “Programmation Linéaire Approchée” (Forsell et Sabbadin, 2006)

Processus de contact contrôlé

Définition (Processus de contact contrôlé)

- $\mathcal{X}_i = \{S, I, R\}$: *sain, infecté, retiré*
- $\mathcal{A}_i = \{i, \emptyset\}$: *isoler, ne rien faire*
- $p_i(x_i^{t+1} = R | x_i^t = R) = 1$; $p_i(x_i^{t+1} = S | x_i^t = S, a_i^t = i) = 1$
  $p_i(x_i^{t+1} = R | x_i^t = I) = \alpha$; $p_i(x_i^{t+1} = I | x_i^t = I) = 1 - \alpha$
  $p_i(x_i^{t+1} = S | x_{N(i)}^t, a_i^t = \emptyset) = \prod_{j \in N(i), x_j^t = I} (1 - \beta_{ji})$
  $p_i(x_i^{t+1} = 1 | x_{N(i)}^t, a_i^t = \emptyset) = 1 - \prod_{j \in N(i), x_j^t = I} (1 - \beta_{ji})$
- $r_i(x_i^t, a_i^t) = c_I(x_i^t) + c_i(a_i^t)$: *coûts d'infection et d'isolation*

Conjecture (Optimalité des politiques locales)

Il existe une politique locale globalement optimale

Processus de contact contrôlé

Remarques :

- Résoudre un processus de contact contrôlé est plus “facile” que résoudre un PDMG
- Un processus de contact contrôlé (PCC) admet (peut-être) une politique optimale locale
- On peut montrer qu’on ne peut pas résoudre exactement un PCC en temps polynomial

Questions :

- est-il plus simple de trouver des bornes de qualité pour l’algorithme de champ moyen appliqué à un PCC ?
- L’algorithme de champ moyen se simplifie-t-il dans le cas d’un PCC ?