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, 22-23 Mai 2006

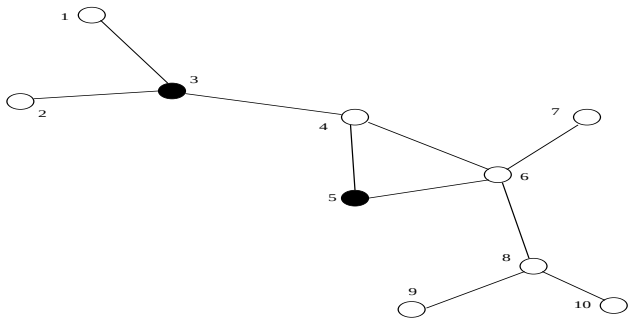
Exemple 1 : Sélection dynamique de sites de réserves face à la déforestation

- Problème : Réduire la perte de biodiversité liée à la déforestation
  - contrainte budgétaire : seul un nombre limité de sites peut être réservé chaque année
  - nécessité d'une politique dans le temps
  - la déforestation évolue comme un "processus de contact"

$\Rightarrow$ Comment construire une politique dynamique de réserve qui conduit au maximum du nombre d'espèces conservées (en espérance) à un horizon fixé ?

Exemple 2 : Contrôle de l'impact d'une épidémie sur un ensemble de parcelles

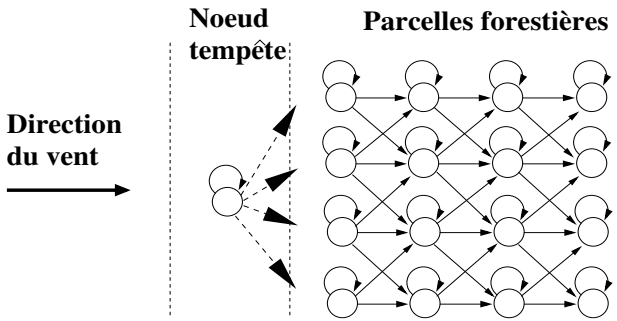
- Problème : Comment réduire la perte de récolte sur la durée ?
  - contamination à grande distance ou par les champs voisins
  - à une date donnée un champ est soit cultivé, soit en jachère et traité



$\Rightarrow$ Comment choisir une politique de gestion qui optimise le total des récoltes sur la durée ?

Exemple 3 : Gestion forestière sous risque de tempête

- Problème : Comment agencer dans l'espace et le temps les rotations de coupes de parcelles forestières ?
  - Risques de tempêtes, de direction a peu près fixée
  - L'impact des tempêtes sur les parcelles forestières génère des dégâts et des pertes de récoltes
  - L'impact des tempêtes sur une parcelle est fonction de son âge ainsi que de celui des parcelles voisines



$\Rightarrow$ Comment choisir une politique de gestion de coupes ?

Modéliser la gestion de processus spatio-temporels

Processus stochastiques spatio-temporels contrôlés :

- en Écologie et biodiversité
  - Conservation de la biologie et gestion du paysage
  - Conception de réserves
- en Gestion des risques naturels et industriels
  - Lutte contre les incendies
  - Lutte contre la diffusion de polluants
- en Épidémiologie
  - Lutte contre l'envahissement de parasites, mauvaises herbes
  - Maîtrise des épidémies végétales, animales

Caractéristiques communes des applications visées

- Processus (temporels)
- Aléatoires
- Contrôlés
- Aspect spatial (variables d'états et d'actions multidimensionnelles)

$\Rightarrow$ Comment modéliser de tels problèmes ?

Processus Décisionnels de Markov sur Graphe

$\Rightarrow$ Comment prescrire des stratégies d'action ?

Itération de la Politique approchée

$\Rightarrow$ Comment réduire la complexité ?

Champ Moyen

Processus Décisionnels de Markov

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 : $\tau =< x_0, \delta(x_0), x_1, \delta(x_1), \ldots, x_t, \delta(x_t), \ldots >, t \in H \subseteq \{1 \ldots + \infty\}$
- Critère : $V_\delta(x_0) = E\left[\sum_{t \in H} \gamma^t r_t(x_t, \delta(x_t))\right], 0 \leq \gamma \leq 1$

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 $< \mathcal{X}, \mathcal{A}, p, r >$ *un PDM stationnaire et* γ *un facteur d'amortissement.*
Soit $\delta : \mathcal{X} \to \mathcal{A}$ *une politique donnée et* $x \in \mathcal{X}$ *un état initial.*
La mesure d'occupation $P_{x,\delta,\gamma} : \mathcal{X} \to [0,1]$ *est définie par :*

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

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* $\delta_i(x_{N(i)}) = a_i \in \mathcal{A}_i$, $\forall x_{N(i)} \in \mathcal{X}_{N_{(i)}}$.

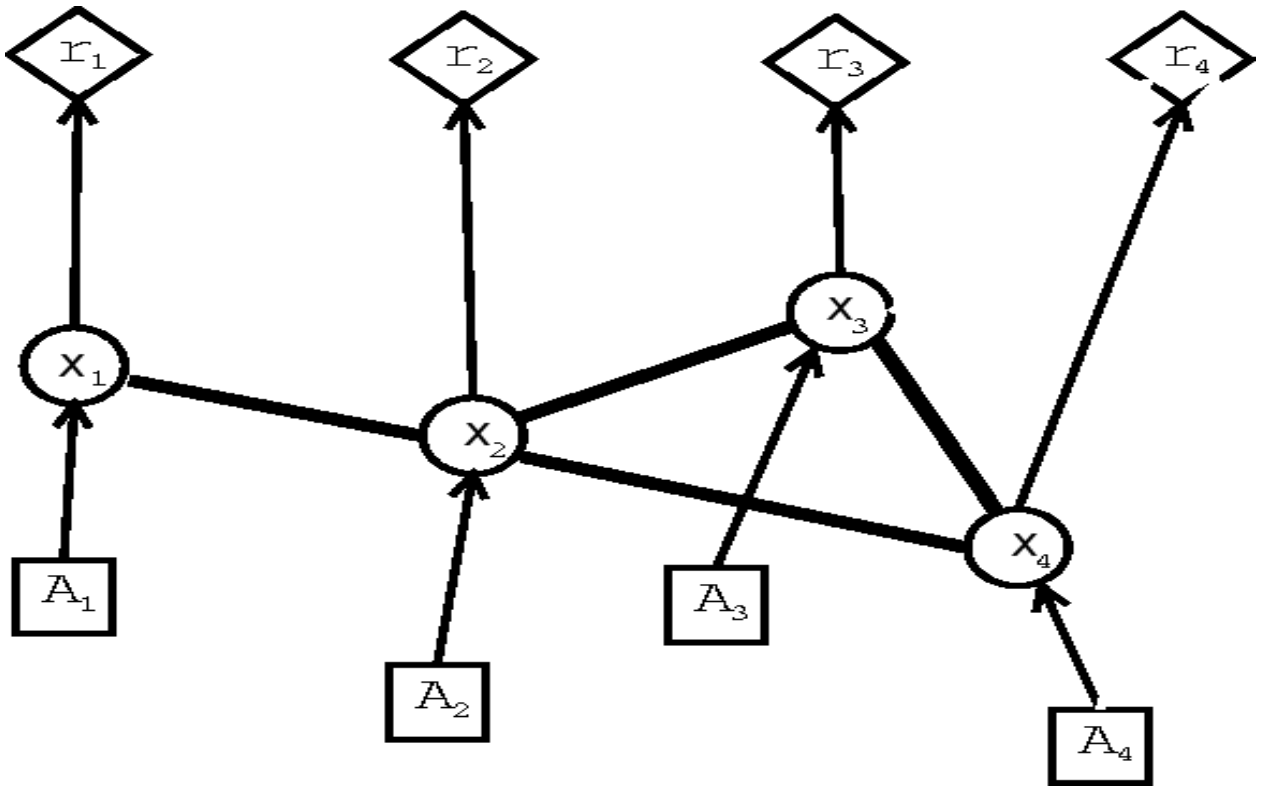
Définition (Récompense locale)

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

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

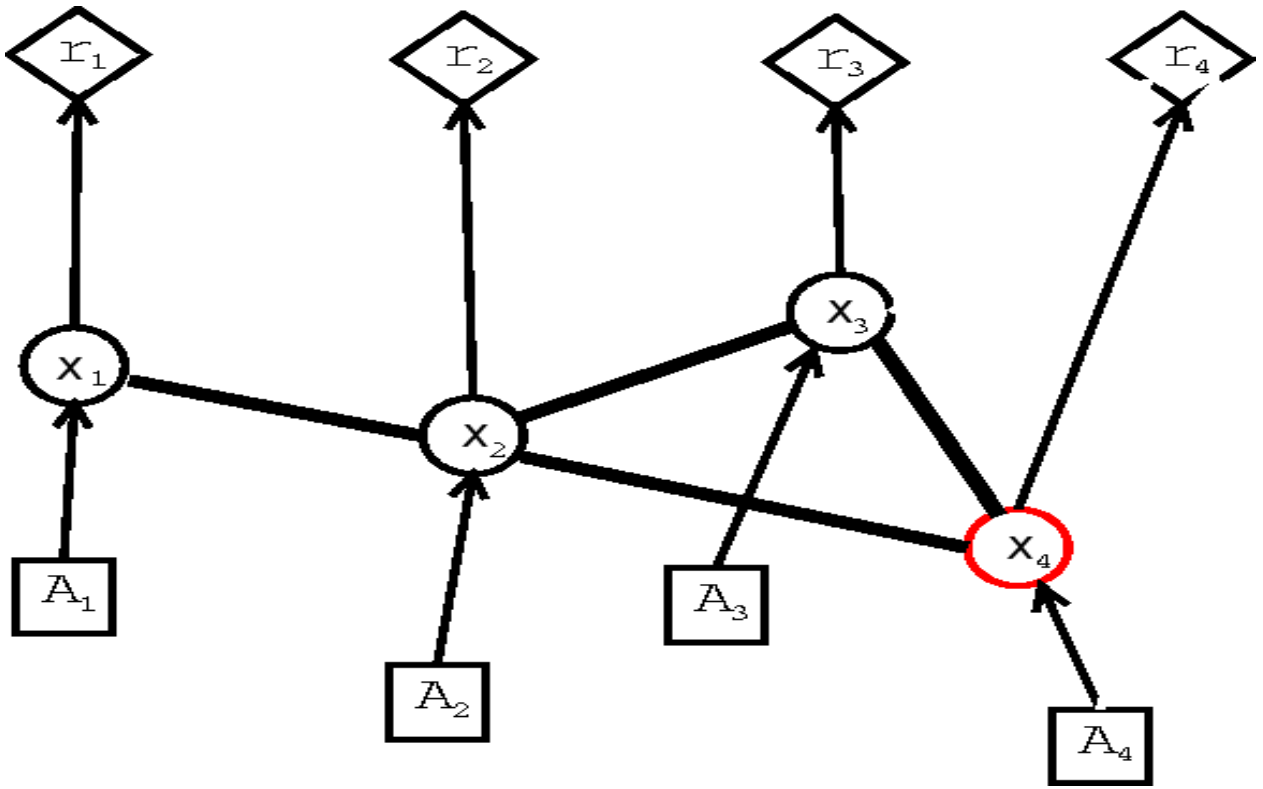
Exemple

Un Processus Décisionnel Markovien sur Graphe



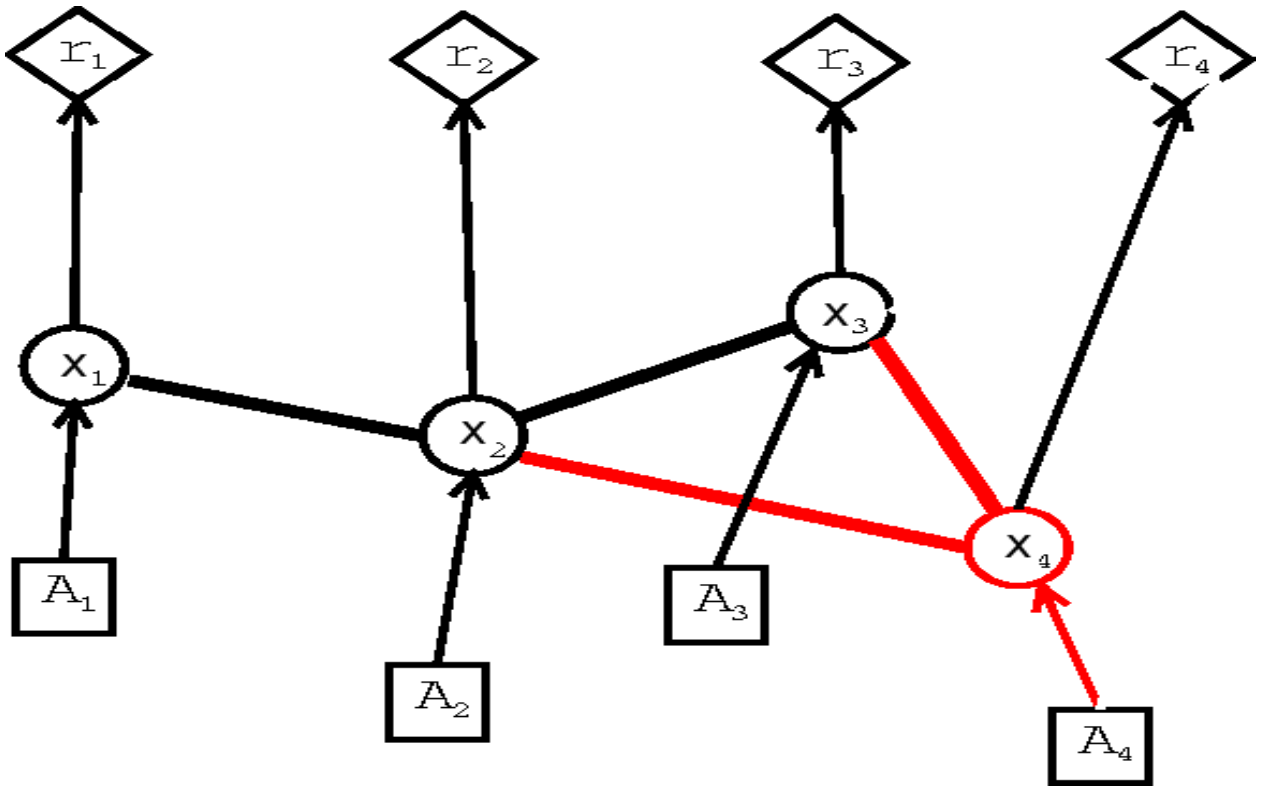
Exemple

Localité autour du noeud 4



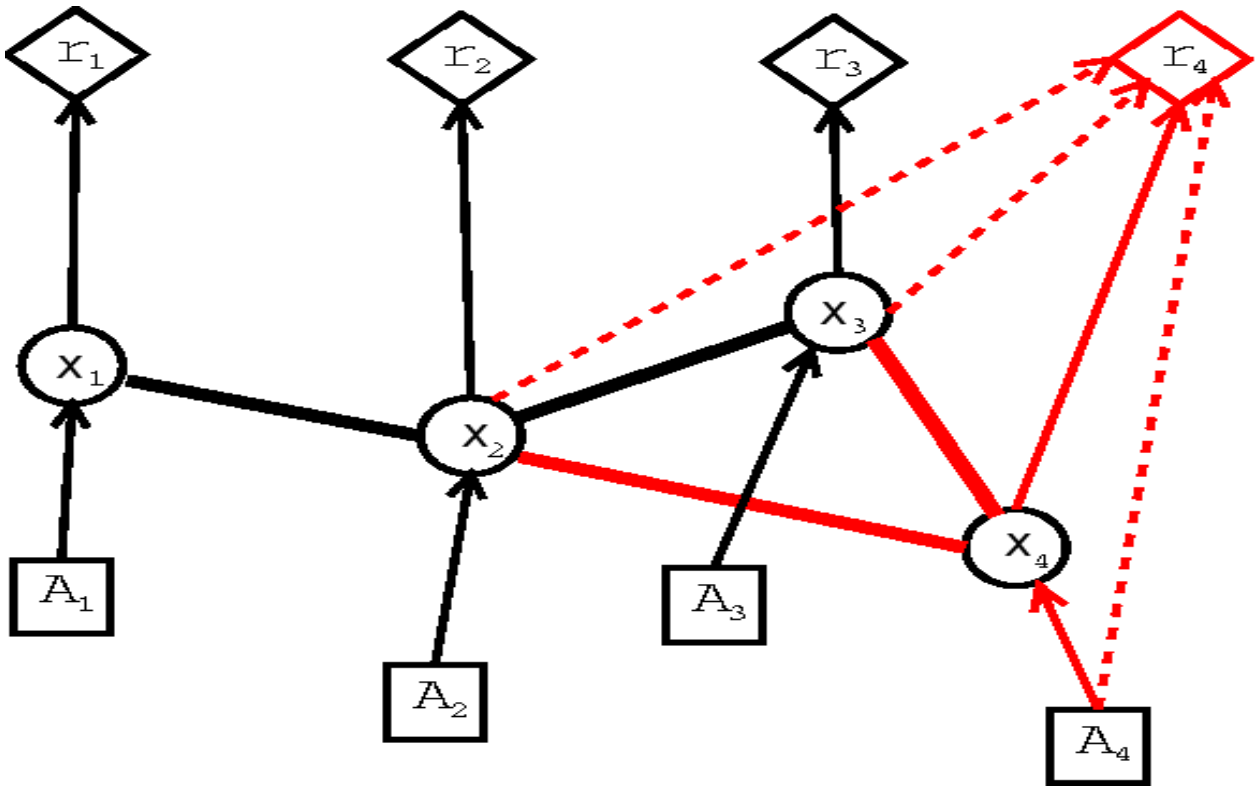
Exemple

Probabilité de transition locale



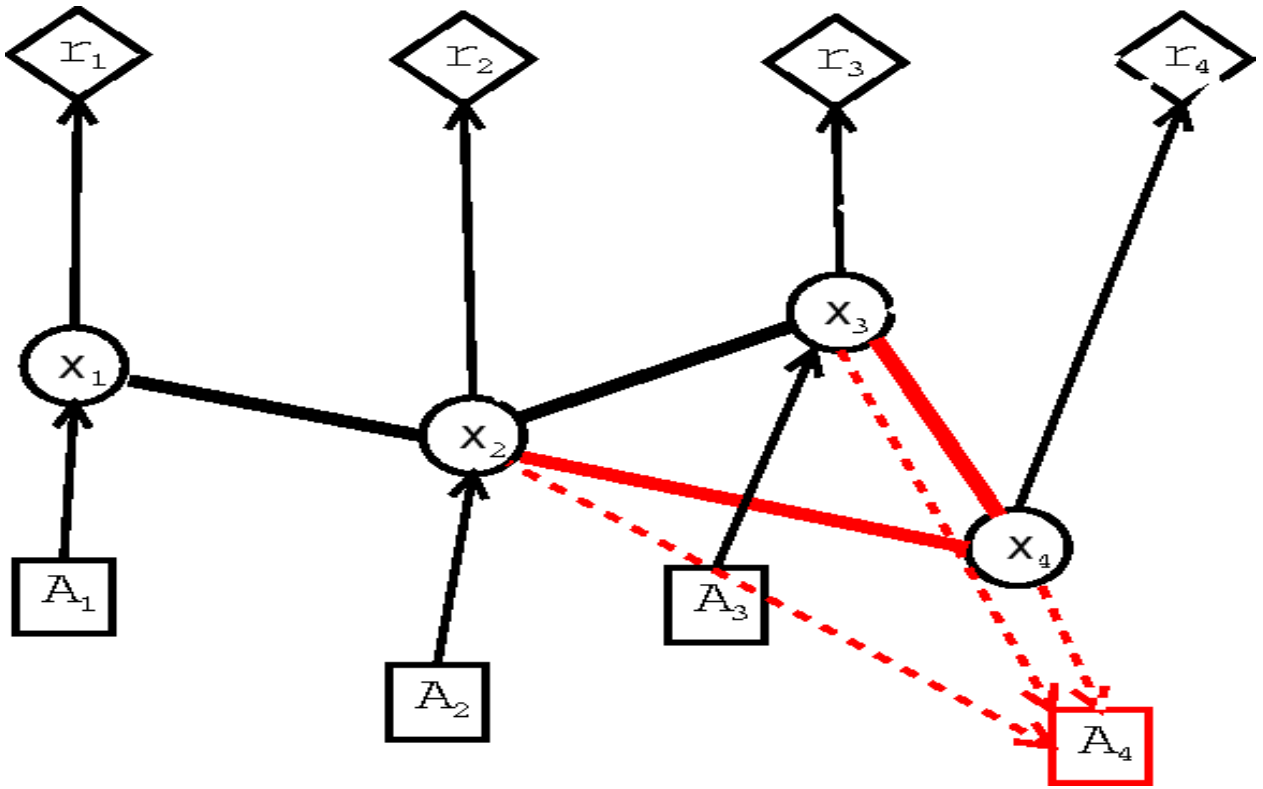
Exemple

Récompense locale



Exemple

Politique locale



Complexité de la résolution d'un PDMG

Complexité "naïve" de la résolution d'un PDMG

- n noeuds, $|\mathcal{X}_i| \leq \sigma$ et $|\mathcal{A}_i| \leq \alpha$, $\max_i |N(i)| = k$
- Complexité spatiale :
  - Traduction "naïve" en PDM : $|\mathcal{X}| \leq \sigma^n$ et $|\mathcal{A}| \leq \alpha^n$
  - $p(x'|x,a) : O((\sigma^2 \cdot \alpha)^n)$
    $r(x,a) : O((\sigma \cdot \alpha)^n)$
  - Fonction de valeur $V_\delta : O(\sigma^n)$
  - Politique globale $\delta : O(\sigma^n)$
- Complexité temporelle par itération : $O((\sigma^2 \cdot \alpha)^n)$

Complexité spatiale de l'expression du PDMG :

- Modèle : $\{p_i\} : n \times |\mathcal{A}_i| \times |\mathcal{X}_i|^{k+1}$, $\{r_i\} : n \times |\mathcal{A}_i| \times |\mathcal{X}_i|^k$, $\Gamma : n \times k \Rightarrow O(n \cdot \sigma^{k+1} \cdot \alpha)$
- Politiques locales $\{\delta_i\} : n \times |\mathcal{X}_i|^k \Rightarrow O(n \cdot \sigma^k)$
- Mais, fonction de valeur $V_\delta : |\mathcal{X}| = O(\sigma^n)$

Recherche d'une "bonne" politique locale

Dans un PDMG, la politique optimale n'est pas forcément locale, mais :

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

Question : Recherche d'un algorithme de calcul de "bonne" politique locale de complexités temporelle et spatiale raisonnables ?

Décomposition de la fonction de valeur

Proposition (Fonction de valeur dans un PDMG)

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

V_δ^i dépend toujours de x en entier !
$\Rightarrow$ Approximation en Champ Moyen de $P_{x,\delta,\gamma}(y) = P_\delta(X^t = y | X^0 = x)$:

Définition (Approximation de la mesure d'occupation)

$$P_\delta(X^t = y | X^0 = x) \approx \prod_i \hat{C}_\delta^i . (X_i^t = y_i | X_i^0 = x_i)$$

Décomposition de la fonction de valeur (suite)

En utilisant l'approximation de la mesure d'occupation, la fonction de valeur s'écrit :

Proposition (Fonction de valeur approchée dans un PDMG)

$$V_\delta(x) \approx \sum_{i=1}^{n} \Big(\sum_{y_{N(i)}} \Big(\prod_{j \in N(i)} \hat{C}^j_\delta(X^t_j = y_j | X^0_j = x_j) \Big) . r_i(y_{N(i)}, \delta_i(y_{N(i)})) \Big)$$

Le problème revient maintenant au choix des mesures d'occupation locales :

$$\hat{C}^j_\delta(X^t_j = y_j | X^0_j = x_j).$$

Choix de la mesure d'occupation locale

Définition (Approximation de la mesure d'occupation locale)

$$\begin{aligned}
\hat{C}_\delta^t(y|x^0) &= \sum_{x^{t-1}} \hat{Q}_\delta^t(y|x^{t-1}).\hat{C}_\delta^{t-1}(x^{t-1}|x^0) \\
&\approx \prod_{i=1}^{n} \Big(\sum_{x_i^{t-1}} \hat{Q}_\delta^{t,i}(y_i|x_i^{t-1}).\hat{C}_\delta^{t-1,i}(x_i^{t-1}|x_i^0) \Big) \\
&\approx \prod_{i=1}^{n} \hat{C}_\delta^{t,i}(y_i|x_i^0).
\end{aligned}$$

$\Rightarrow$ Approximations $\hat{Q}_\delta^{t,i}$ des fonctions de transition locales ?

Approximation des fonctions de transition locales

Définition (Approximation de type Kullback-Leibler)

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

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

$$\begin{aligned}\hat{Q}_\delta^{t+1,i}(x_i^{t+1}|x_i^t) &\propto \exp E_{\tilde{C}_\delta^t}\left[\log p_i(x_i^{t+1}|x_i^t, X^t_{N(i)\setminus\{i\}}, \delta_i(x_i^t, X^t_{N(i)\setminus\{i\}}))\right] \\ &\approx E_{\tilde{C}_\delta^t}\left[p_i(x_i^{t+1}|x_i^t, X^t_{N(i)\setminus\{i\}}, \delta_i(x_i^t, X^t_{N(i)\setminus\{i\}}))\right]\end{aligned}$$

where

$$\tilde{C}_\delta^t(x^t) = \prod_{j=1}^{n} \hat{C}_\delta^{t,i}(x_i^t|x_i^0) \cdot P^{0,i}(x_i^0).$$

Evaluation de la politique approchée

Algorithme

1. *On alterne pour tout t le calcul des $\hat{C}_\delta^{t,j}(y_j|x_j)$ $(\forall j, x_j, y_j)$*
2. *et un des termes de la somme :*

$$\begin{aligned}\hat{V}_\delta^i(x) &\approx \hat{V}_\delta^i(x_{N(i)}) \\ &\approx \sum_{t=0}^{+\infty} \gamma^t \cdot \sum_{y_{N(i)}} \left(\prod_{j \in N(i)} \hat{C}_\delta^{t,j}(y_j|x_j) \right) \cdot r_i(y_{N(i)}, \delta_i(y_{N(i)}))\end{aligned}$$

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

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)

Perspectives

- Comparaison avec des méthodes approchées basées sur la PL existant (en particulier (Guestrin 2003), quadratique en n)
- Restriction sur les combinaisons d'actions admissibles ($\mathcal{A} \subset \mathcal{A}_1 \times \ldots \times \mathcal{A}_n$)
- Apprentissage par renforcement et fonctions de valeurs paramétrées