Notes de Cours MTH 6415

Olivier Sirois

2018-01-01

Contents

1 Programmation Dynamique **2**

1.1 Modèle de programation dynamique probabiliste (Processus de décision Markoviens) 3

Chapter 1

Programmation Dynamique

C'est un ensemble d'outils :

- Model
- Algorithmes
- Corpus théorique..

qui sert à résoudre des problèmes de décisions séquentiels.

Problème de décision séquentiels Problème qui prend une suite (avec un ordre spécifique) de décision ou ces décision sont reliées les unes aux autres avec un objectif qui est d'optimiser (min ou max) d'un certain critère de performance.

l'information qui sert à prendre la décision n'est pas toujours disponible. Alors sans avoir la totalité de l'information nécessaire, sa nous amène à faire la différence entre :

- **Problème à boucle ouverte**. Problème ou toute l'info est nécessaire et disponible. On peut faire des choix au départ et en insistant, (((ordinairement))), on va être capable d'interpréter le problème comme étant déterministe.
- **Problème en boucle fermé**. Problèmes ou on découvre au fur et à mesure l'information utile/importantes pour la prise de décision. Dans cette famille, on à les **problème dynamique stochastique**, qui est l'objectif du cours.

l'objectif est souvent une espérance mathématique. La plupart des concepts stochastique sont applicable au problème déterministe avec une simple arrangement. Par contre, en raison des contraintes temporel, on saute directement dans le stochastique.

Politique (policy) Règle de prise de décision, de façon imprécsie, qui décrit pour chaque situation possible (état du système) la meilleur décision à prendre pour optimiser une fonction objective globale. Cette fonction objective étant souvent une espérance mathématique.

Souvent, on essaie de charactérisé les propriété de la politique optimale. on formule ensuite les modeles d'une certaine façon, on essaie après de les résoudre et finalement on caractérise les solutions.

1.1 Modèle de programation dynamique probabiliste (Processus de décision Markoviens)

Normalement, nous avons:

- Un processus de décision séquentiel qui est découpé dans ce que nous appellons des étapes (N étapes). Habituellement numérotée de O à $N-1$.
- À l'étape x_k, on observe les caractéristiques du système et on prend une décision $u_k \in U_k(x_k)$
- Nous avons une variable aléatoire que nous appelons ω_k qui est généré selon une loi de probabilité $P_k(*|x_k, u_k)$. Cette loi de probabilité peut dépendre de k, x_k, u_k. Mais ce qui est important c'est qu'elle ne dépend pas des valeurs précédentes, c-à-d, qu'elle ne dépend pas de x_m, u_m, ω_m pour $m < k$.

Normalement, on observe ω_k, on paye un cout $g_k(x_k, u_k, \omega_k)$. et à l'état prochaine, x_{k+1} ets donnée par une fonction $f_k(x_k, u_k, \omega_k)$.

Le cout totale: $g_n(x_n) + \sum_{k=0}^{N-1} g_k(x_k, u_k, \omega_k)$. il y a l'état finale x_n, sont cout associé g_n ainsi que les cout de tous les états précédent (la somme sur 0 à N-1).

Exemples:

- Gestion d'inventaire
- Gestion de réservoire (production hydroélectrique)
- autres.. (Deep Q Learning??)

Une politique admissible est une suite de **fonctions** $\pi = (\mu_1, \mu_2...)$ tel que $\mu_k : X_k \to U_k$ ou $\mu_k(x_k) \in U_k(x)$ pour tout $x \in X_k$ et k = 0,.... N-1. **Pour chaque états, j'ai une fonction qui nous donne la bonne décision à prendre**.

Pour résumé:

- X_k = Ensemble des états:
- $U_k(x)$ = ensemble des décision admissible dans l'état x
- D_k = ensemble de perturbations ω_k
- G_k = fonctions de couts
- f_k = fonction de transisition
- x_k = état à l'étape k
- u_k = décision prise à l'étape k
- ω_k = perturbation aléatoire à l'étape k

Bellman Les équations de Bellman (voir INF8215). Le principe d'optimalité est la base de la programmation dynamique.

Pour $0 \leq k \leq N-1$ et $x \in x_k$, posons:

> $J_{\pi,k}$ est le cout espéré total de l'étape k à la fin. Si on est dans l'état x à l'étape k et si on utilise la politique π.
>
> $J_{\pi,k} = E_{\pi,k}[g_n(x_n) + \sum_{m}^{N-1} g_n(x_m, u_m, \omega_m)]$ ou E est l'espérance lorsque $x_k = x$, $u_m = \mu_m(x_m)$ pour m = k,... N-1.

Pour une politique π donnée, on a une équation de récurrence $J_{\pi,M}(x) = g_N(x)$ pour $\forall x \in X_N$

$J_{\pi,k}(x) = E[g_k(x, \mu_k(x), \omega_k) + E_{\pi,K+1}[J_{K+1}(f_k(x, \mu_k, \omega_k))]]$ pour $0 \leq k \leq N-1$ et $x \forall X_k$

ou LHS est le cout immédiat et RHS est le cout future.

typiquement, on cherche une politique π qui va minimiser/maximiser $J_{\pi,0}$, l'espérance de la somme des couts de l'étape 0 jusqu'à l'étape N.

Notons que $\pi^* = (\mu_0^*, \mu_1^*, ...)$ un telle politique optimale. Posons maintenant $J_k^*(x)$ étant le cout espéré total de l'étape k à la fin si on est dans l'état x à l'étape k et donc en fait $J^*(x) = \min_\pi J_{\pi,k}(x)$.

Théorème

1. On a $J_k^* = J_k$ ou les fonctions J_k sont défini par les équations de récurrences (équations de bellman).
2. $J_N(x) = g_N(x), \forall x \in X_N$

3. $J_k(x) = \inf_{u \in U_k(x)} E[g_k(x, \mu_k, \omega_k) + J_{k+1}(f_k(x, \mu_k, \omega_k))]$ pour $0 \leq k \leq N-1$ et $x \in X_k$ et ou l'espérance E est fais par rapport à ω_k qui suit la loi $P_k(*|x, u)$

4. infimum = plus grande borne inferieur de la fonction. Si quelques chose tend vers l'infini, on est obliger de prendre l'infini quand le problème est fini.

5. une valeur de u qui fait atteindre l'infimum est une décision optimale à prendre lorsqu'on est dans l'état x à l'étape k.

6. Donc, on peut définir une politique optimal π^* (dans le cas ou elle existe) par la relation: $\mu_k^*(x) \in arg\min_{u \in U_k(x)} + E[..voirplushaut...]$ et on va voir que $J_k = J_{\pi,k}(x)$ pour tout k et x.

On peut résoudre les équations de récurrence et calculer en même temps une politique optimale par ce qu'on appelle **chaînage arrière**. qui est fait en calculant $J_N(x)$ pour tout $x \in X_N$, pour calculer $J_{N-1}(x) et \mu_{N-1}^*(x), \forall x \in X_{N-1}$ puis ensuite $J_{N-2}, \mu_{N-2}^*...$ ainsi de suite

Les valeurs optimales rechercher est $J_0(x_0)$ ou x_0 est l'etat initiale.
Comment formuler un problème de programmation dynamique Stochastique à horizon fini :

- On spécifie les étapes
- On spécifie les états, les ensemble $x_k, k = 0...N-1$
- On spécifie les décision et les ensemble $X_k(x)$
- On spécifie aussi les perturbations aléatoires ω_k et leurs distributions
- Donner les fonctions g_n et g_k