Chaînes de Markov avec coûts associés

(tous les états persistants et apériodiques)

Coût d'une transition: $A_{i,j}$ pour la transition de i à j (ne dépend pas du temps ni de l'histoire)

Coût moyen d'une transition: de l'état i:

$$a_i = \sum_{j=1}^{N} P_{i,j} A_{i,j}$$

Tendance vers la moyenne:

Le coût moyen par transition tend vers g où g est le coût moyen d'une transition dans la distribution limite Π

On peut facilement calculer g:

$$\sum_{i=1}^{N} \Pi_i a_i = g$$

Décisions qui peuvent changer les probabilités de transition

La possibilité de choix: supposons que dans l'état i on peut choisir entre plusieurs *décisions* et que pour la décision d, la probabilité d'aller en état j est $P_{i,j}^d$ avec un coût $a_{i,j}^d$

on peut calculer l'espérance du coût d'une transition:

$$a_i^d = \sum_{j=1}^{N} P_{i,j}^d a_{i,j}^d$$

mais minimiser ca (la politique gloutonne) n'est pas forcément la bonne politique

Le nombre de politiques possibles peut être trop grand pour permettre de calculer le g de chacune
et peut-être ce n'est pas la politique possédant le plus petit g qu'il convient de choisir...

Actualisation des coûts

Un compromis entre la politique gloutonne et minimiser g

Actualisation: Un euro aujourd'hui est plus utile qu'un euro l'année prochaine;
Un coût d'un euro aujourd'hui est plus grave qu'un pareil coût l'année prochaine;

On veut calculer les coûts sur la supposition qu'un euro au temps t vaut ρ euros au temps $t+1$; $0 < \rho < 1$.

Rapport avec le taux d'intérêt: si le taux réel d'intérêt (payé ou reçu) est τ, on a $\rho = 1/(1+\tau)$.

Calcul des coûts

Les équations pour les coûts totaux: On a toujours un coût immédiat (moyen) en l'état i de a_i et une probabilité $P_{i,j}$ de transition de l'état i à j:

$$v_i \ = \ a_i + \rho \sum_{j=1}^{N} P_{i,j} v_j$$

Système non singulier: on peut calculer les v_i de ces équations parce que $\rho < 1$.

Le coût dans la distribution limite: Si le coût moyen d'une transition dans la distribution limite (sans actualisation) est g, le coût (total) si on commence dans cette distribution est $g(1 + \rho + \rho^2 + \rho^3 \cdots) \ = \ g/(1 - \rho)$. Donc $g\rho^t/(1-\rho)$ est une approximation au coût à partir de temps t (t grand et la chaîne régulière).

Calcul d'une politique optimale avec actualisation

On considère (enfin) le cas où on peut choisir entre plusieurs décisions en chaque état et on veut minimiser le coût total avec actualisation. C'est évident que la décision optimale en chaque état ne dépend ni de l'histoire ni du temps.

On peut utiliser l'algorithme d'amélioration d'une politique initiale: calculer le coût total (avec actualisation) v_j à partir de chaque état j; choisir à chaque état i la décision $\hat{d}$ qui minimise $a_i^{\hat{d}} + \rho \sum_{j=1}^{N} P_{i,j}^{\hat{d}} v_j$; répéter tant qu'il y a une réduction;
(Ca équivaut à minimiser le coût qu'on aurait si on choisissait une nouvelle politique pour la première étape et gardait l'ancienne politique après, ce qui explique pourquoi il y a amélioration)

Termination de l'algorithme

- Réduction des coûts à chaque itération
- Un nombre fini de politiques possibles
- Donc, termination garantie
- En principe pourrait être très lente
  (nombre de politiques très élevé)
- Mais "en pratique" l'algorithme trouve la politique optimale après un nombre raisonnable d'itérations

Démonstration plus “mathématiqu’ que l’algorithme réduit les coûts à chaque étape, sauf si on est déjà à une politique optimale:

D’abord avec les décisions d on a:

$$v_i^d \ = \ a_i^d + \rho \sum_{j=1}^{N} P_{i,j}^d v_j^d$$

On choisit $\hat{d}$ tel que

$$a_i^{\hat{d}} + \rho \sum_{j=1}^{N} P_{i,j}^{\hat{d}} v_j^d \leq v_i^d$$

(et on continue si on a $<$ pour au moins un i)

Les vrais nouveaux coûts sont donnés par:

$$\hat{v}_i^{\hat{d}} = a_i^{\hat{d}} + \rho \sum_{j=1}^{N} P_{i,j}^{\hat{d}} \hat{v}_j^{\hat{d}}$$

Donc:

$$v_i^d - \widehat{v}_i^{\widehat{d}} \geq \rho \sum_{j=1}^{N} P_{i,j}^{\widehat{d}}(v_j^d - \widehat{v}_j^{\widehat{d}})$$

(et on a $>$ pour au moins un i)

Considérons le i qui donne le minimum de $v_i^d - \widehat{v}_i^{\widehat{d}}$: on a

$$v_i^d - \widehat{v}_i^{\widehat{d}} \geq \rho \sum_{j=1}^{N} P_{i,j}^{\widehat{d}}(v_i^d - \widehat{v}_i^{\widehat{d}})$$

$$v_i^d - \widehat{v}_i^{\widehat{d}} \geq \rho(v_i^d - \widehat{v}_i^{\widehat{d}})$$

$$v_i^d - \widehat{v}_i^{\widehat{d}} \geq 0$$

et, donc, $v_j^d - \widehat{v}_j^{\widehat{d}} \geq 0$ pour tout j

Finalement on ne peut pas avoir $= 0$ pour tout j; et donc on a une amélioration pour tout j...

Le cas sans actualisation

Ce cas est plus compliqué parce que les équations:

$$v_i \;=\; a_i + \rho \sum_{j=1}^{N} P_{i,j} v_j$$

ne sont pas utilisables si $\rho = 1$; (soit aucune solution, soit plusieurs)

On peut définir $v_i^{(n)}$ le coût moyen d'une suite de n transitions à partir de l'état i. On trouve que

$$v_i^{(n)} \;\to\; ng + v_i$$

où, maintenant, v_i est le coût d'avoir commencé en l'état i.

On a
$v_i = a_i - g + \sum_{j=1}^{N} P_{i,j} v_j$

mais ces équations ne suffisent pas (ajout de la même constante k à chaque v_i donne une nouvelle solution) Il faut ajouter

$$\sum_{i=1}^{N} \Pi_i v_i = 0$$

On essaie d'optimiser la politique en utilisant la même idée que dans le cas avec actualisation:

- Choisir une politique initiale
- Calculer Π, g et les v_i
- Utiliser ces v_i pour chercher une meilleure décision en chaque état
- Itérer tant qu'il y a amélioration

Un exemple: assurance d'un véhicule

Le coût de l'assurance d'un véhicule est 100, 200, 400 ou 600.
Après un an sans aucun accident déclaré, le prix est réduit au niveau inférieur
Après un an avec un accident déclaré, le prix est augmenté au niveau supérieur
La probabilité d'un accident grave est 5% par an
La probabilité d'un accident mineur est 15% par an et le coût de sa réparation est 500

Devrait-on déclarer un accident mineur ou payer la réparation?
(nous prenons $\rho = 0.9$ et supposons les coûts actualisés au début de l'année)

Calcul de la politique optimale

Je commence avec la politique de déclarer tout incident

$$\begin{aligned} v_1 &= 100 + 0.9(.8v_1 + .2v_2) \\ v_2 &= 200 + 0.9(.8v_1 + .2v_3) \\ v_3 &= 400 + 0.9(.8v_2 + .2v_4) \\ v_4 &= 600 + 0.9(.8v_3 + .2v_4) \end{aligned}$$

et je calcule $\mathbf{v} = [1315, 1489, 1905, 2405]$
Donc une meilleur politique pour les niveau 200 et 400 est de payer les réparations (par exemple pour v_2, $200 + 0.9(.8v_1 + .2v_3) > 275 + 0.9(.95v_1 + .05v_3)$):

$$\begin{aligned} v_1 &= 100 + 0.9(.8v_1 + .2v_2) \\ v_2 &= 275 + 0.9(.95v_1 + .05v_3) \\ v_3 &= 475 + 0.9(.95v_2 + .05v_4) \\ v_4 &= 600 + 0.9(.8v_3 + .2v_4) \end{aligned}$$

et $\mathbf{v} = [1304, 1472, 1839, 2347]$

Maintenant on voit qu'une meilleur politique à v_2 est de déclarer l'accident
(parce que maintenant
$$200 + 0.9(.8v_1 + .2v_3) < 275 + 0.9(.95v_1 + .05v_3))$$

Donc, une troisième politique:

$$\begin{aligned}
v_1 &= 100 + 0.9(.8v_1 + .2v_2) \\
v_2 &= 200 + 0.9(.8v_1 + .2v_3) \\
v_3 &= 475 + 0.9(.95v_2 + .05v_4) \\
v_4 &= 600 + 0.9(.8v_3 + .2v_4)
\end{aligned}$$

et $\mathbf{v} = [1299, 1465, 1833, 2341]$

la solution optimale!

Le modèle n'est pas très réaliste (tous les accidents mineurs n'ont pas le même coût) mais ce calcul peut aider à prendre de bonnes décisions, par exemple en l'état 2 déclarer un sinistre si le coût est supérieur à $\rho \times (1833 - 1299)$.