mdp

May 2023

1 Application

On considère le problème suivant : On observe la position et la vitesse d'une voiture se déplaçant en ligne droite, la position et la vitesse initiale sont 50m et 50m/s. On suppose dans un premier temps que l'accélération est fixée à $1m.s^{-2}$.

On suppose également qu'il y a une erreur due au processus (par exemple une bosse) et une erreur d'observation (erreur due à la prise de mesure par le GPS par exemple).

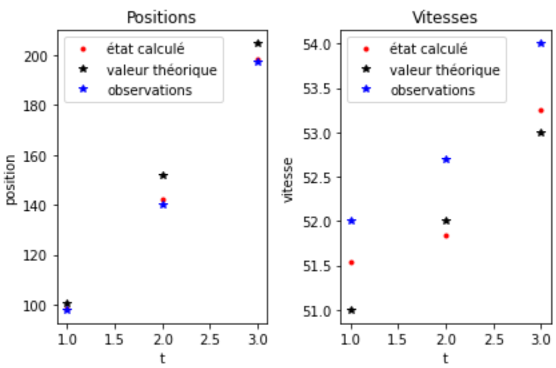
Disposant de mesures jusqu'au temps n, on souhaite estimer la position réelle de l'objet grâce au filtre de Kalman. Les équations de mouvement nous donnent le modèle suivant :

$$\begin{aligned} X_{t_n} &= AX_{t_{n-1}} + Bu + w_n \text{ avec } w_n \sim \mathcal{N}(0, R) \\ X_n &= CX_{t_n} + e_n \text{ avec } e_n \sim \mathcal{N}(0, S) \end{aligned}$$

X_{t_n} représente la valeur théorique du vecteur (position,vitesse) au temps n, X_n la valeur réelle, $A = \begin{pmatrix} 1 & dt \\ 0 & 1 \end{pmatrix}$, $B = \begin{pmatrix} \frac{dt^2}{2} \\ dt \end{pmatrix}$, $C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $u = 1$, S=$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, R=$\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$

En utilisant alors le code de résolution joint avec le rapport avec les paramètres A, B, C, u, R et S ainsi que les observations précédentes on peut alors obtenir les vecteurs d'esperance et matrices de variance covariance de la loi normale que suit le vecteur de (position, vitesse) à chaque pas de temps.

Illustration :



.

En associant cette fois-ci des actions possibles sur l'accélération, comme par exemple $\{-1, 0, 1\}$ qui correspondent au fait de choisir une de ces accélérations dans le modèle à la place de u et en ajoutant des récompenses liées au choix de l'action, par exemple $\{5, 0, -10\}$on pourrait reformuler cet exemple en un POMDP où l'objectif serait de maximiser les récompenses au bout de 5 secondes (on considère un pas de temps de 1 seconde avant de faire un choix):

1. $N = 5$
2. $E := E_X \times E_Y$, E_X pour la partie observable et E_Y pour la partie non observable du processus. On a $E_X = (\mathbb{R} \times \mathbb{R})$ et $E_Y = (\mathbb{R} \times \mathbb{R})$.
3. $Ac = \{-1, 0, 1\}$. Un élément de Ac est noté a.
4. $D \subset E_X \times Ac$. Il est défini par $D(x) = Ac$ pour tout $x \in E$. Or ici, il n'y a pas d'action interdite, donc $D = E_X \times Ac$.
5. Q correspond aux transitions qui sont définies par les équations de mouvement et le bruit du modèle, valant $S = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$, $R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
6. Q_0 est la distribution initiale de Y_0, on suppose ici : $\mathcal{N} \sim \left(\begin{pmatrix} 50 \\ 50 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \right)$
7. On définit les récompenses intermédiaires $r(x, y, a) = \begin{cases} r(x, y, 1) = -10 \\ r(x, y, 0) = 0 \\ r(x, y, -1) = 5 \end{cases}$
8. On définit la récompense finale à $g(x, y) = x_{position}$ (la première composante de x).
9. On fixe $\beta = 1$.

En utilisant le filtre de Kalman à chaque itération on peut alors transformer le problème en un Processus de Decision Markovien classique $(E, A, D, Q', r', g', \beta)$:

1. $E := E_X \times \mathcal{P}(E_Y)$, E_X pour la partie observable et $\mathcal{P}(E_Y)$ pour l'ensemble des mesures de probabilité sur E_Y. On a $E_X = (\mathbb{R} \times \mathbb{R})$. Un élément de E est noté (x, ρ).
2. $Ac = \{-1, 0, 1\}$. Un élément de Ac est noté a.
3. $D \subset E_X \times A$. est défini par $D(x) = Ac$ pour tout $x \in E$. Ici, il n'y a pas d'action interdite, donc $D = E_X \times Ac$.
4. Q' : les transitions sont définies par le filtre de Kalman à chaque étape.
5. $r'(x, \rho, a) = \begin{cases} r'(x, \rho, 1) = -10 \\ r'(x, \rho, 0) = 0 \\ r'(x, \rho, -1) = 5 \end{cases}$
6. $g'(x, \rho) = x_{position}$.
7. On fixe $\beta = 1$.