Modèles de Markov cachés

Anisa CAJA

ISFA

16 Mars 2012

Introduction

Qu'est-ce qui est caché dans un MMC ?

Les états d'un MMC ne sont pas observables, ils sont cachés.
Ces états produisent des observations et à partir de ces observations on peut déduire des choses sur le processus de Markov sous-jacent.

Les éléments d'un MMC

1. N, le nombre d'états dans le modèle. L'ensemble des états est $E = \{E_1, ..., E_N\}$ et l'état en t est noté e_t
2. M, est le nombre d'observations différents du modèle. Les observations correspondent aux sorties du modèle. On les note $V = \{v_1, ..., v_M\}$.
3. Les probabilités de transition entre états $P(E_i|E_j)$ pour $1 \leq i, j \leq N$. Elles sont groupées dans la matrice de transition du modèle $A = \{a_{ij}\}$
4. La loi des observations une fois dans l'état j, noté $b_j(k) = P(O_k|E_j)$, pour $1 \leq i, j \leq N$ et $1 \leq k \leq M$. On définit $B = \{b_j(k)\}$.
5. La loi initiale $\pi = \{\pi_i\}$ telle que $\pi_i = P(e_1 = E_i)$

Les éléments d'un MMC

1. N, le nombre d'états dans le modèle. L'ensemble des états est $E = \{E_1, ..., E_N\}$ et l'état en t est noté e_t
2. M, est le nombre d'observations différents du modèle. Les observations correspondent aux sorties du modèle. On les note $V = \{v_1, ..., v_M\}$.
3. Les probabilités de transition entre états $P(E_i|E_j)$ pour $1 \leq i, j \leq N$. Elles sont groupées dans la matrice de transition du modèle $A = \{a_{ij}\}$
4. La loi des observations une fois dans l'état j, noté $b_j(k) = P(O_k|E_j)$, pour $1 \leq i, j \leq N$ et $1 \leq k \leq M$. On définit $B = \{b_j(k)\}$.
5. La loi initiale $\pi = \{\pi_i\}$ telle que $\pi_i = P(e_1 = E_i)$

Générateur d'une séquence d'observations

Avec "les bonnes valeurs" pour les paramètres N, M, A, B, π, le MMC peut être utilisé comme générateur d'une séquence $O = O_1, ..., O_T$.

1. Choix de l'état initial $e_1 = E_i$ selon la loi π
2. Poser t=1
3. Choisir l'observation obtenue en t, $O_t = v_k$, avec $b_i(k)$
4. Passer à un autre état E_j avec la probabilité a_{ij}
5. t=t+1 et passer à l'étape 3 si $t < T$, sinon fin.

Cette procédure peut aussi servir de modèle pour décrire comment une séquence d'observations a été générée par un MMC approprié.

Hypothèses du modèle de Markov caché

- Conditionnellement au présent, le futur est indépendant du passé (comme le modèle de Markov classique)

Hypothèses du modèle de Markov caché

- Conditionnellement au présent, le futur est indépendant du passé (comme le modèle de Markov classique)
  Pourquoi on dit qu'on a besoin des observations passées pour prendre des décisions ?

Hypothèses du modèle de Markov caché

- Conditionnellement au présent, le futur est indépendant du passé (comme le modèle de Markov classique)
  Pourquoi on dit qu'on a besoin des observations passées pour prendre des décisions ?
  Car les observations passées nous permettent d'avoir une estimation de l'état présent. Si on avait connaissance de cet état, nous n'aurions pas besoin de l'historique des observations.

Hypothèses du modèle de Markov caché

- Conditionnellement au présent, le futur est indépendant du passé (comme le modèle de Markov classique)
  Pourquoi on dit qu'on a besoin des observations passées pour prendre des décisions ?
  Car les observations passées nous permettent d'avoir une estimation de l'état présent. Si on avait connaissance de cet état, nous n'aurions pas besoin de l'historique des observations.
- Les observations dépendent uniquement de l'état caché présent

Sommaire

1. Les problèmes liés aux MMC
   - Problème 1 - Filtrage
   - Problème 2 - Evaluation
     - Algorithme backward
   - Problème 3- Séquence d'états "optimale"
     - Algorithme de Viterbi
2. Les modèles de Markov à changement de régime

Problème 1 - Filtrage

Etant donnée la séquence d'observations $O = O_1, ...O_T$, quelle est la loi de distribution des états latents ?

Problème 2 - Evaluation

Etant donnée la séquence d'observations $O = O_1, ...O_T$ et le modèle A, B, π, quelle est la probabilité d'observer cette séquence ?

Problème 3 - Séquence d'états "optimale"

Etant donnée la séquence d'observations $O = O_1, ...O_T$ et le modèle A, B, π, quelle est la séquence d'états qui explique au mieux les observations ?
Ici on essaie de découvrir la partie cachée du modèle, trouver la "bonne" séquence d'états qui a produit les observations.

Problème 4 - Paramètres du modèle

Etant donnée la séquence d'observations $O = O_1, ...O_T$ et le modèle A, B, π, quels sont les paramètres A, B, π qui maximisent $P(O|A, B, \pi)$?

Filtrage

A partir d'une séquence d'observations, quelle est la distribution des états latents ?

$$
\begin{aligned}
P(e_{t+1}|o_1,\dots,o_t,o_{t+1}) &= \frac{P(o_{t+1}|e_{t+1},o_1,\dots,o_t)P(e_{t+1}|o_1,\dots,o_t)}{P(o_{t+1}|o_1,\dots,o_t)} && (1)\\
&\propto \mathbb{P}(o_{t+1}|e_{t+1},o_1,\dots,o_t)\mathbb{P}(e_{t+1}|o_1,\dots,o_t) && (2)\\
&= \mathbb{P}(o_{t+1}|e_{t+1})\mathbb{P}(e_{t+1}|o_1,\dots,o_t) && (3)\\
&= \mathbb{P}(o_{t+1}|e_{t+1})\sum_{e_t\in E}\mathbb{P}(e_{t+1}|e_t,o_1,\dots,o_t) && (4)\\
&\quad \mathbb{P}(e_t|o_1,\dots,o_t) && (5)\\
&= \mathbb{P}(o_{t+1}|e_{t+1})\sum_{e_t\in E}\mathbb{P}(e_{t+1}|e_t)\mathbb{P}(e_t|o_1,\dots,o_t) && (6)
\end{aligned}
$$

Evaluation

Comment calculer efficacement $\mathbb{P}(O) = \mathbb{P}(o_1, ..., o_t)$, la probabilité d'observer cette séquence ?

Evaluation

Comment calculer efficacement $\mathbb{P}(O) = \mathbb{P}(o_1, ..., o_t)$, la probabilité d'observer cette séquence ?
Autre interprétation :
Scoring : dans quelle mesure un modèle est adapté à une certaine séquence d'observations ?

Evaluation

Comment calculer efficacement $\mathbb{P}(O) = \mathbb{P}(o_1, ..., o_t)$, la probabilité d'observer cette séquence ?
Autre interprétation :
Scoring : dans quelle mesure un modèle est adapté à une certaine séquence d'observations ? La solution au *problème d'évaluation* nous permet de choisir le meilleur modèle parmi plusieurs candidats.

Evaluation

Comment calculer efficacement $\mathbb{P}(O) = \mathbb{P}(o_1, ..., o_t)$, la probabilité d'observer cette séquence ?
Autre interprétation :
Scoring : dans quelle mesure un modèle est adapté à une certaine séquence d'observations ? La solution au *problème d'évaluation* nous permet de choisir le meilleur modèle parmi plusieurs candidats.
Résolution avec la programmation dynamique.
Deux algorithmes : forward et backward.

Algorithme backward

On définit $f_t(e_t) = \mathbb{P}(o_t, ...o_T|e_t)$
Partons du dernier état $f_T(e_T) = \mathbb{P}(o_T|e_T)$
A chaque pas :

$$\begin{aligned}
f_t(e_t) &= \mathbb{P}(o_t, ...o_T|e_t) && (7)\\
&= \mathbb{P}(o_t|e_t)\mathbb{P}(o_{t+1}, ...o_T|e_t, o_t) && (8)\\
&= \mathbb{P}(o_t|e_t)\mathbb{P}(o_{t+1}, ...o_T|e_t) && (9)\\
&= \mathbb{P}(o_t|e_t) \sum_{e_{t+1}\in E} \mathbb{P}(e_{t+1}, o_{t+1}, ...o_T|e_t) && (10)\\
&= \mathbb{P}(o_t|e_t) \sum_{e_{t+1}\in E} \mathbb{P}(e_{t+1}|e_t)\mathbb{P}(o_{t+1}, ...o_T|e_t, e_{t+1}) && (11)\\
&= \mathbb{P}(o_t|e_t) \sum_{e_{t+1}\in E} \mathbb{P}(e_{t+1}|e_t)\mathbb{P}(o_{t+1}, ...o_T|e_{t+1}) && (12)\\
&= \mathbb{P}(o_t|e_t) \sum_{e_{t+1}\in E} \mathbb{P}(e_{t+1}|e_t)f_{t+1}(e_{t+1}) && (13)
\end{aligned}$$

Pas final :

$$\begin{aligned}
\mathbb{P}(o_1, ...o_T) &= \sum_{e_1 \in E} \mathbb{P}(o_1|e_1)\mathbb{P}(o_2, ...o_T|e_1) && (14)\\
&= \sum_{e_1 \in E} \mathbb{P}(o_1|e_1)f_1(e_1) && (15)\\
& && (16)
\end{aligned}$$

Utile pour modéliser :

- Language : les phonèmes
- Chaînes génétiques ou du texte
- Notes musicales
- Prix observés sur un marché financier

Exemple pour la reconnaissance vocale

Elements du MMC pour la reconnaissance vocale

- Observations : Séquences de phonèmes
- Classes : mots
- Logiciel : proba d'une séquence de phonèmes pour un mot donné et la probabilité a priori pour chaque mot
- Le logiciel calcule pour chaque mot dans quelle mesure il s'accorde avec la séquence de phonèmes en maximisant $\mathbb{P}(m)\mathbb{P}(o_1, ..., o_T|m)$

Maximiser la probabilité d'observer la séquence obtenue

On a $o_1, ..., o_T$

Quelle séquence d'états a produit ces observations ?

Cela revient à trouver la séquence $e_1, ..., e_T$ qui maximise $\mathbb{P}(e_1, ..., e_T | o_1, ..., o_T)$

Combien de séquences possibles ? $|E|^T$

$$arg\max \mathbb{P}(e_1, ..., e_T | o_1, ..., o_T) = arg\max \frac{\mathbb{P}(e_1, ..., e_T, o_1, ..., o_T)}{\mathbb{P}(o_1, ..., o_T)} \quad (17)$$

$$= arg\max \mathbb{P}(e_1, ..., e_T, o_1, ..., o_T) \quad (18)$$

Algorithme de Viterbi

On définit : $\mathrm{v}_i(e_i) = max\mathbb{P}(e_{i+1}, ..., e_T, o_{i+1}, ..., o_T | e_i)$
La meilleure probabilité future qu'on puisse obtenir si on part de e_i et on choisit les états futurs de façon optimale.
En T, pas de futur donc $v_T(e_T) = 1$

Algorithme de Viterbi

$$v_t(e_t) = \max_{e_{t+1}\ldots e_T} \mathbb{P}(e_{t+1},\ldots,e_T,o_{t+1},\ldots,o_T|e_t) \tag{19}$$

$$= \max_{e_{t+1}\ldots e_T} \mathbb{P}(e_{t+2},\ldots,e_T,o_{t+1},\ldots,o_T|e_t,e_{t+1})\mathbb{P}(e_{t+1}|e_t) \tag{20}$$

$$= \max_{e_{t+1}\ldots e_T} \mathbb{P}(e_{t+2},\ldots,e_T,o_{t+1},\ldots,o_T|e_{t+1})\mathbb{P}(e_{t+1}|e_t) \tag{21}$$

$$= \max_{e_{t+1}\ldots e_T} \mathbb{P}(e_{t+2},\ldots,e_T,o_{t+2},\ldots,o_T|e_{t+1},o_{t+1})\mathbb{P}(o_{t+1}|e_{t+1})\mathbb{P}(e_{t+1}|e_t) \tag{22}$$

$$= \max_{e_{t+1}\ldots e_T} \mathbb{P}(e_{t+2},\ldots,e_T,o_{t+2},\ldots,o_T|e_{t+1})\mathbb{P}(o_{t+1}|e_{t+1})\mathbb{P}(e_{t+1}|e_t) \tag{23}$$

$$= \max_{e_{t+1}} \left[\mathbb{P}(o_{t+1}|e_{t+1})\mathbb{P}(e_{t+1}|e_t) \max_{e_{t+2}\ldots e_T} \mathbb{P}(e_{t+2},\ldots,e_T,o_{t+2},\ldots,o_T|e_{t+1})\right] \tag{24}$$

Algorithme de Viterbi

On définit ensuite $w_t(e_t) = arg\ \max_{e_{t+1}} \mathbb{P}(o_{t+1}|e_{t+1})\mathbb{P}(e_{t+1}|e_t)v_{t+1(e_{t+1})}$

$$
\begin{aligned}
e_1* &= arg\ \max_{e_1} \max_{e_2,\ldots,e_T} \mathbb{P}(e_1, \ldots, e_T, o_1, \ldots, o_T) && (26)\\
&= arg\ \max_{e_1} \max_{e_2,\ldots,e_T} \mathbb{P}(e_1)\mathbb{P}(e_2, \ldots, e_T, o_1, \ldots, o_T|e_1) && (27)\\
&= arg\ \max_{e_1} \max_{e_2,\ldots,e_T} \mathbb{P}(e_1)\mathbb{P}(o_1|e_1)\mathbb{P}(e_2, \ldots, e_T, o_2, \ldots, o_T|e_1, o_1) && (28)\\
&= arg\ \max_{e_1} \mathbb{P}(e_1)\mathbb{P}(o_1|e_1) \max_{e_2,\ldots,e_T} \mathbb{P}(e_2, \ldots, e_T, o_2, \ldots, o_T|e_1, o_1) && (29)\\
&= arg\ \max_{e_1} \mathbb{P}(e_1)\mathbb{P}(o_1|e_1)v_1(e_1) && (30)
\end{aligned}
$$

On trouve alors $e_2*, \ldots, e_T*$ avec $e_t = w_{t-1}(e_{t-1})$.
$e_1*, e_2*, \ldots, e_T*$ maximise donc $\mathbb{P}(e_1, \ldots, e_T|o_1, \ldots, o_T)$.

Problème 4 - Paramètres du modèle

Pour une référence sur l'algorithme de Baum-Welch qui permet de trouver les paramètres du modèle, regarder Rabiner (1989).

Sommaire

1 Les problèmes liés aux MMC

2 Les modèles de Markov à changement de régime

Modéliser un changement de régime

Une série économique peut observer un changement de régime à cause d'un changement de politique gouvernementale, en cas de crise financière etc.
La variable y_t suit le processus autorégressif d'ordre 1 suivant :

$$y_t = c_1 + \phi y_{t-1} + \epsilon_t$$

où $\epsilon_t \sim N(0, \sigma^2)$ pour $t = 1, ..., t_0$. Après t_0 le niveau moyen de la série change et le processus qu'elle suit devient :

$$y_t = c_2 + \phi y_{t-1} + \epsilon_t$$

pour $t = t_0 + 1, t_0 + 2....$
Modélisé comme ceci le changement de processus devient déterministe alors qu'on veut que ce soit aléatoire. Hamilton (1989) propose un modèle qui englobe les deux cas et qui est aléatoire : le modèle de Markov à changement de régime.

Hamilton (1989)

$$y_t = c_{s_t} + \phi y_{t-1} + \epsilon_t$$

La variable s_t est **une variable d'état** (inobservable) qui suit une chaîne de Markov à 2 états telle que :

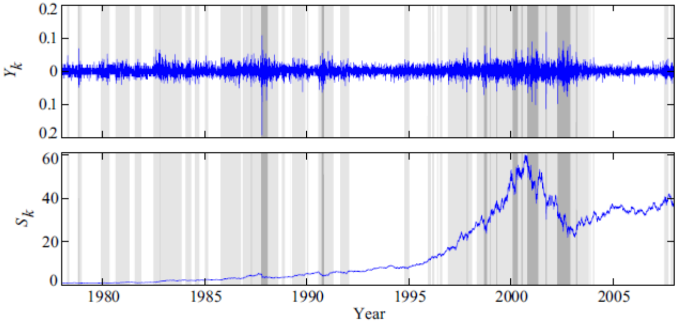
$$\mathbb{P}(s_t = j | s_{t-1} = i, s_{t-2} = k, ...) = \mathbb{P}(s_t = j | s_{t-1} = i) = p_{ij}$$

La variable s_t est inobservable, mais on peut observer y_t.
Pour modéliser y_t il faut : les constantes c_1 et c_2, la variance σ, ϕ, ainsi que p_{11} et p_{22}.
A partir du processus observable y_t on peut donc dire des choses sur s_t, en particulier la probabilité pour le processus d'être sur un état j.

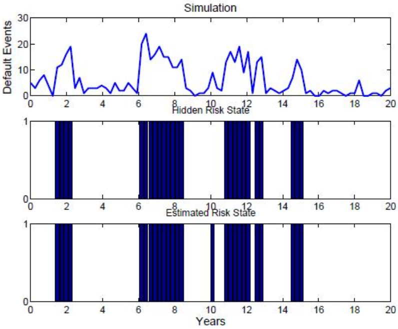
$$\gamma_{jt} = \mathbb{P}(s_t = j | O; \sigma, \phi, c_1, c_2, p_{11}, p_{22})$$

Le rendement des actions

Exemple tiré de Hidden Markov Models- Lecture notes (Princeton), R. van Handel.



Défauts sur un portefeuille- Crowder, Davis, Giampieri (2005)



Quelques références

1] CROWDER M., DAVIS M. , and GIAMPIERI G. [2005], "Analysis of Default Data Using Hidden Markov Models", Quantitative Finance, Vol. 5, No. 1, (February 2005), pp. 27-34

2] HAMILTON, J. D. [1989], "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle " Econometrica 57, 357-384.

3] HAMILTON, J. D. [2005], Palgrave Dictionary of Economics.

4] HARDY M. R., "A regime-switching model of long-term stock returns", North American Actuarial Journal, vol. 5, n˚2, 41-53.

5] RABINER L.R.,[1989], "A tutorial on hidden Markov models and selected applications in speech recognition." , Proceedings of the IEEE, Vol.77, No.2, Feb. 1989.

6] Hidden Markov Models Handout, Harvard, http ://isites.harvard.edu/fs/docs/icb.topic540049.files/cs181$_lec19_handout.pdf$

7] VAN HANDEL R. , Hidden Markov Models- Lecture notes 2008, Princeton.