



Evaluation de la pertinence des paramètres stochastiques sur des Modèles de Markov Cachés

Bernard Robles, Manuel Avila, Florent Duculty, Pascal Vrignat, Frédéric
Kratz

► To cite this version:

Bernard Robles, Manuel Avila, Florent Duculty, Pascal Vrignat, Frédéric Kratz. Evaluation de la pertinence des paramètres stochastiques sur des Modèles de Markov Cachés. 2010. hal-00601983

HAL Id: hal-00601983

<https://hal.science/hal-00601983>

Submitted on 21 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluation de la pertinence des paramètres stochastiques sur des Modèles de Markov Cachés

Bernard Roblès*, Manuel Avila*, Florent Duculty*, Pascal Vrignat*, Frédéric Kratz**

** Institut PRISME (EA 4229), Equipe MCDS, IUT de l'Indre - Dpt GEII, 2 av. F. Mitterrand, 36000 Châteauroux*

*** Institut PRISME (EA 4229), Equipe MCDS, Ecole Nationale Supérieure d'Ingénieurs 88 boulevard Lahitolle 18020 Bourges cedex*

Sections de rattachement : 61
Secteur : secondaire

RÉSUMÉ. Les modèles de Markov Cachés ou HMM¹ peuvent être utilisés pour la modélisation de processus complexes. Dans cette étude, nous souhaitons évaluer la pertinence des paramètres d'un modèle de façon objective sans connaissance a priori. Dans un premier temps, nous présentons les critères de sélection des modèles les plus utilisés dans la littérature actuelle. Nous présentons dans une deuxième partie une méthode permettant d'évaluer la pertinence d'événements stochastiques issus de Modèles de Markov Cachés. Nous étayons notre étude en nous appuyant sur l'exemple concret d'un processus industriel. Nous évaluons alors les paramètres de sortie des différents modèles testés sur ce processus, pour finalement s'orienter vers le modèle le plus pertinent.

MOTS-CLÉS : modèles de Markov cachés, sélection de modèles, caractérisation, pertinence, entropie.

¹ Hidden Markov Models

1. Introduction

Selon Lebarbier [1], la problématique de la sélection d'un modèle est basée sur la minimisation d'un critère de pénalité. Les premiers critères qui apparaissent dans la littérature sont l'AIC : l'Akaike Information Criterion [2], le BIC : Bayésian Information Criterion [3], le MDL : Minimum Description Length [4] et le Cp de Mallows [5]. De nombreux travaux théoriques ont été réalisés sur leurs propriétés statistiques afin de les adapter à des modèles spécifiques. Comme par exemple les versions corrigées du critère AIC d'Hurvish [6] avec l'AICc, Sugira [7] avec le c-AIC pour les petites tailles d'échantillons par rapport au nombre de paramètres à estimer.

L'objet de notre étude est de mesurer la pertinence de modèles de Markov cachés. Nous étayons nos propos par l'étude d'un cas concret d'un processus industriel. Dans la partie 2, nous introduisons les problèmes et les notations utilisées. Dans la partie 3, nous décrivons comment estimer la pertinence des paramètres de modèles de Markov cachés. Enfin en partie 4, nous faisons l'étude d'un cas concret en donnant le modèle le plus pertinent.

2. Modèle de Markov Caché ou HHM

Un modèle de Markov caché (Rabiner [8]) est un automate à états cachés qui est constitué d'une variable non observable. Celle-ci représente l'état du système à modéliser. Seule la variable de sortie est observable. Cela nous permet d'avoir une séquence d'observations en sortie de l'automate ; à partir de maintenant, nous parlerons simplement de symboles représentant ces observations.

3. Méthodes de mesures de pertinence de modèles

3.1. Mesure de l'entropie de Shannon

Selon [9], l'entropie de Shannon, est une fonction mathématique qui permet de mesurer la quantité d'informations contenue dans une source d'informations. Cette source peut être un texte écrit dans une langue donnée, un signal électrique ou encore un fichier informatique quelconque...

3.1.1. Définition de l'entropie

L'entropie de Shannon est définie dans [10] comme suit :

$$H(S) = - \sum_{i=1}^n P_i \log_2 P_i \quad (1)$$

P_i est la probabilité moyenne de voir apparaître le symbole i dans S

3.1.2. Propriétés formelles de l'entropie de Shannon

La minimalité : une distribution « pure » a une entropie nulle : si un seul symbole est représenté, l'impureté est nulle.

La maximalité : à l'inverse, une distribution mélangée est « impure » : la valeur de l'entropie, lorsque plusieurs symboles sont représentés, doit donc être élevée.

3.1.3. Principes de l'entropie maximale

Les deux principes de maximisation d'entropie de Jaynes [11] sont les suivants :

- Principe d'assignation des probabilités à une distribution lorsque nous ne disposons pas d'informations complètes sur elle
- De toutes les distributions de probabilité qui satisfont les contraintes, choisir celle qui présente l'entropie maximale au sens de Shannon.

Venkat [12] utilise ce 2^{ème} principe pour la sélection de modèles, ainsi que [13] pour construire des modèles de plus en plus précis simplement en rajoutant de l'information. Notre démarche consiste à comparer la moyenne des entropies des différents modèles. La valeur d'entropie moyenne serait alors maximale pour les modèles les plus pertinents.

3.2. Commentaires sur les méthodes de mesures

Attention au choix du critère de sélection du modèle. Il doit être conditionné par l'objectif de l'analyse et la connaissance des données. Les données d'apprentissage sont déterminantes dans la construction du modèle. Il n'existe pas de critère universellement meilleur. En pratique, seule une parfaite connaissance du milieu à analyser permet de donner un sens à la notion de supériorité d'un critère sur un autre.

4. Expérimentations

4.1. Description des processus industriels

Tous les processus de l'usine étudiés sont liés entre eux de manière séquentielle. L'arrêt d'un processus engendre donc l'arrêt des éléments en amont. Une maintenance préventive est donc indispensable. Pour ce faire, les agents de maintenance doivent consigner leurs actions ou observations dans une base de données centralisée. Les données sont issues d'une peseuse volumétrique sur une ligne de production continue de pain de mie. Toutes les 6 heures, nous insérons dans cette base de données le

symbole RAS (= tout va bien), si aucune donnée n'est saisie. Nous utilisons ensuite ces séquences de symboles (voir tableau 1) pour modéliser le niveau de dégradation du processus. Nous modélisons cette « signature » à l'aide de modèles de Markov.

Symboles ou événements observés									
VEP	VEP	TEP	TEP	SEP	TEP	TEP	DEP	AU	

Jour/heure/... →

Tableau 1 : *Séquence d'un message issue des données de maintenance*

4.2. Modélisation des processus industriels

Dans le cadre d'activités de maintenance, Vignat et al. [14] modélise ces dysfonctionnements à l'aide de modèles de Markov. Nous rappelons dans le tableau 2, la signification des symboles choisis issus des observations. Ces symboles définissent les actions de maintenance menées sur le processus. Par exemple le symbole DEP correspond à un arrêt de la production. C'est un état critique qu'il faut minimiser. L'étude de [14] considère deux modèles différents avec deux corpus d'apprentissage, l'un sur l'année 2005 et l'autre sur les deux années 2005-2006. Les symboles RAS sont insérés pour avoir un échantillonnage à la journée ou toutes les 6 heures. Pour la suite nous adopterons les dénominations suivantes :

- 05M1 1j	Corpus d'apprentissage 2005 modèle 1 / 1 donnée par jour
- 0506M1 1j	Corpus 2005-2006 modèle 1 / 1 donnée par jour
- 05M2 1j	Corpus 2005 modèle 2 / 1 donnée par jour
- 0506M2 1j	Corpus 2005-2006 modèle 2 / 1 donnée par jour
- 05M1 6h	Corpus 2005 modèle 1 / 1 donnée toutes les 6 heures
- 0506M1 6h	Corpus 2005-2006 modèle 1 / 1 donnée toutes les 6 heures
- 05M2 6h	Corpus 2005 modèle 2 / 1 donnée toutes les 6 heures
- 0506M2 6h	Corpus 2005-2006 modèle 2 / 1 donnée toutes les 6 heures

Nature des interventions	
DEP (Dépannage / arrêt de la production)	
RM (Réglage Machine)	
AU (Autre)	
OBS (Observation)	
TEP (Travaux Entretien Préventif pas d'arrêt de prod)	
SEC (Sécurité)	
RAN (Remise A Niveau / planifié)	
NAT (Netoyage Machine)	
VEP (Visite Entretien Préventif)	
RAS (pas d'intervention)	

Tableau 2 : *Codification symbolique des interventions de maintenance (source : [14]).*

4.3. Description des modèles de Markov utilisés

Les modèles se présentent sous la forme d'un automate stochastique. Les états représentent les niveaux de dégradation du processus. Les symboles de l'automate représentent les observations du processus (voir figure 1).

Le modèle donne la probabilité d'être dans un des quatre états (S1-S4) en fonction des symboles (noté « Cd » dans le tableau 3). Les probabilités des quatre états sont données par la variable Forward et les différents niveaux par l'algorithme de Viterbi (voir étude de [14]).

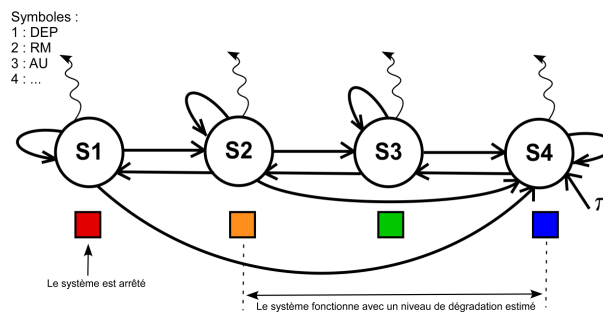


Figure 1 : Modèle à topologie orientée à quatre états d'après [14].

N°	DATE	Cd	Cd. Symb.	S1	S2	S3	S4	Niveau
1	11/01/07	RAS	10	0,00%	0,0%	0,00%	100,00%	4
2	11/01/07	RAS	10	0,00%	0,0%	0,10%	99,90%	3
3	12/01/07	SEC	6	0,00%	49,5%	48,80%	1,70%	2
4	12/01/07	RAS	10	0,00%	98,8%	0,00%	1,20%	2
5	13/01/07	DEP	1	100,00%	0,0%	0,00%	0,00%	1

Tableau 3 : Modèle à topologie orientée à quatre états d'après [14].

Les séquences de symboles ainsi produites, nous donnent selon le modèle, une estimation du taux de dégradation du processus. Nous pouvons alors quantifier ces informations au moyen de différents critères et ainsi établir une évaluation des modèles, selon ces critères.

4.4. Mesure de l'entropie de Shannon

Dans un premier temps, nous calculons les entropies moyennes des symboles seuls (colonne « 1 symb » du tableau 4), pour chaque mode d'échantillonnage (1 jour ou 6 heures). Puis nous mesurons l'entropie des modèles de Markov d'ordre 1, 2 et 3. Rappelons que pour une chaîne de Markov d'ordre k : l'état suivant dépend des k , états précédents. Nous voyons alors tableau 4 que les modèles avec un échantillonnage de 1 jour sont les plus pertinents (où l'entropie est maximale).

Nous calculons ensuite les entropies de chaque symbole par rapport aux états S1, S2,

S3 et S4 du modèle de Markov présentés au paragraphe 4.3 Nous établirons ainsi des relations symboles / états en fonction des différentes valeurs de l'entropie. Les résultats sont présentés dans le tableau 5. Le tableau montre qu'il y a moins de dispersion de l'information dans le modèle avec un échantillonnage à la journée.

- Une entropie nulle permet de discriminer totalement un symbole par rapport à un état. Ce qui nous donne trivialement, que le symbole DEP est entièrement discriminé pour l'état S1 voir tableau 5.
- Une entropie maximale indique qu'une majorité des symboles est présente dans une majorité de niveaux de dysfonctionnement du système. Elle ne permet pas de discriminer un symbole particulier contrairement à une entropie nulle, mais au contraire, elle permet d'incriminer un symbole à plusieurs états. Dans notre cas, le symbole RAS possède l'entropie maximale, il est donc le symbole le plus représentatif des états S2, S3 et S4. Ce symbole est aussi le plus utilisé, il correspond à une observation d'un état stable des processus où rien ne se passe.

Echantillonnage	Entropie moyenne			
	1 symb	Ordre 1	Ordre 2	Ordre 3
1 jour	0,292	0,112	0,084	0,072
6 heures	0,158	0,074	0,06	0,051

Tableau 4 : Entropie moyenne des modèles de Markov sur les 2 types d'échantillonnage.

Symboles	S1	S2	S3	S4	Total	Entropie
AU		4	3	1	8	1,097
DEP	9				9	0,000
OBS		3			3	0,184
RAN			2	1	3	0,817
RAS		31	5	6	42	1,418
RM		4			4	0,224
SEC		6		1	7	0,641
TEP		12			12	0,420
VEP		16			16	0,473

Tableau 5 : Mesure de l'entropie.

Le but de notre démarche est de calculer l'entropie moyenne de chaque modèle. Selon le 2^{ème} principe de l'entropie maximale, nous choisirons alors le modèle dont l'entropie moyenne est maximale. Par contre, les valeurs extrêmes de l'entropie peuvent perturber le calcul de l'entropie moyenne d'un modèle. Il est donc préférable d'éliminer ces valeurs pour ce calcul. Afin d'évaluer au plus juste notre modèle, nous proposons donc d'éliminer les symboles totalement discriminés d'entropie nulle et les symboles à large spectre comme RAS où l'entropie est maximale. Cette opération d'élimination sera nommée par la suite : « **filtre entropique** ». Nous calculons alors l'entropie moyenne du modèle pour évaluer la pertinence des séquences d'observations. Le meilleur modèle sera celui dont l'entropie moyenne est maximale après filtrage entropique.

Le tableau 6 présente les résultats des mesures d'entropies moyennes des modèles avant et après le filtre entropique. Nous effectuons un filtrage de niveau 1 i.e. nous éliminons un seul symbole dont l'entropie est maximale ainsi que toutes les entropies nulles au travers de ID3 [15].

Modèles	Entropie moy. sans filtre	Entropie moy. avec filtre
05M1 1j	0,586	0,695
0506M1 1j	0,699	0,881
05M2 1j	0,698	0,857
0506M2 1j	0,268	0,365
05M1 6h	0,524	0,741
0506M1 6h	0,397	0,828
05M2 6h	0,264	0,393
0506M2 6h	0,200	0,394

Tableau 6 : Résultats de l'entropie moyenne avec et sans filtre entropique.

Le modèle le plus pertinent apparaît alors : « 0506M1 | 1 jour » où l'entropie moyenne est maximale.

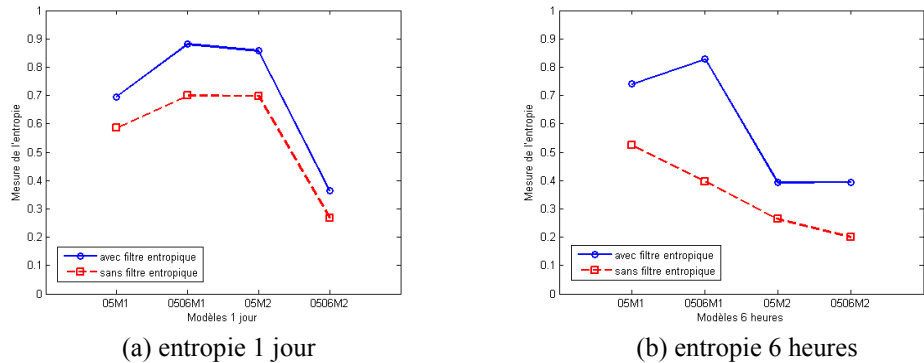


Figure 2 : Entropies moyennes des modèles.

Sur la figure 2 (a), nous voyons que le filtre entropique permet de mettre en valeur un maximum pour le modèle 0506M1 sur la base de temps de 1 jour. La figure 2 (b) nous montre de manière beaucoup plus flagrante que le modèle 0506M1 sur la base de temps de 6 heures est le plus pertinent après application du filtre entropique. D'après [14], c'est bien le modèle 0506M1 qui fournit les meilleurs résultats de prédiction de pannes.

5. Conclusions et perspectives

Dans notre étude, nous avons présenté une démarche visant à évaluer la pertinence de symboles dans des modèles de Markov cachés. Nous avons appliqué l'entropie de

Shannon, sur les symboles isolés, sur les bigrammes et trigrammes selon différents échantillonnages pour pouvoir comparer les différents découpages temporels. De la même façon, l'entropie a été évaluée selon les états des modèles afin de vérifier que le plus pertinent obtient également un bon score « entropique ». Nous illustrons ainsi que sans connaissance à priori, le modèle 0506M1 | 1 jour est le plus pertinent. Nos calculs corroborent ainsi les résultats de [16] qui montrent que le modèle 0506M1 | 1 jour est celui qui appréhende le mieux la réalité.

Dans nos travaux futurs, nous allons mesurer la pertinence de ces modèles avec d'autres critères comme le maximum de vraisemblance ou celui de Akaike (AIC) [2]. Notre objectif est de pouvoir valider de façon objective, un choix de modèle : topologie, ordre, symbole,... sans connaissance à priori sur les résultats.

Références

- [1] Emilie Lebarbier and Tristan Mary Huard. Le critère BIC : fondements théoriques et interprétation. Research Report RR-5315, INRIA, 2004.
- [2] Akaike H. Information theory and an extension of the maximum likelihood principle. 2nd inter. symp. on information theory. 2nd Inter. Symp. on Information Theory, pages 267–281, 1973.
- [3] Schwarz G. Estimating the dimension of a model. *Annals of Statistics*, pages 461–464, 1978.
- [4] Rissanen J. Modelling by the shortest data description. *Automatica* 14, pages 465–471, 1978.
- [5] Mallows J. Some comments on cp. *Echnometrics*, 15, pages 661–675, 1973.
- [6] Mallows C.M. and Tsai C.L. Regression and time series model selection in small samples. *Biometrika* 76, pages 297–307, 1989.
- [7] Sugira N. Further analysis of the data by akaike's information criterion and the finite corrections. *Communications in Statistics, Theory and Methods*, pages 13–26, 1978.
- [8] Rabiner L.R. A tutorial on hidden markov models and selected applications in speech recognition. *Proceeding of the IEEE*, 77(2) *SIAM interdisciplinary journal*, pages 257–286, 1989.
- [9] Bouquet F. Méthode et outils pour l'intelligence artificiel. Cours de master informatique MOIA, 2003.
- [10] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley, 2 edition, 1991.
- [11] Jaynes E. Information theory and statistical mechanics. *Physical Review*, vol. 16, no. 4, pages 620–630, 1957.
- [12] Chandrasekaran V., Johnson J.K., and Willsky A.S. Maximum entropy relaxation for graphical model selection given inconsistent statistics. *Laboratory for Information and Decision Systems*, Massachusetts Institute of Technology Cambridge, MA 02139, 2007.
- [13] Arminjon M. and Imbault D. Maximum entropy principle and texture formation. *Zeitschrift für angewandte Mathematik und Mechanik*, 80, Suppl. N°1, pages 13–16, 2000.
- [14] Vignat P. et Avila M. et Duculty F. et Kratz F. Modélisation des dysfonctionnements d'un système dans le cadre d'activités de maintenance. 16^{ème} Congrès de Maîtrise des Risques et de Sécurité de Fonctionnement, Avignon, Communication 4A-1, 2008.
- [15] Quinlan J.R. Algorithme de base (id3). *Expert Systems in the Microelectronic*, pages 168–201, 1979.
- [16] Vignat P. et Avila M. et Duculty F. et Roblès B. Kratz F. Utilisation des chaînes de markov cachées pour une évaluation des activités de maintenance dans le cadre d'un processus industriel pour l'agroalimentaire. CNRIUT 2009.