Estimation d'un modèle générique pour un parc de machines

Farah Ankoud, Gilles Mourot, Roger Chevalier, Nicolas Paul, José Ragot

Institut National Polytechnique de Lorraine

Centre de Recherche en Automatique de Nancy

Electricité de France

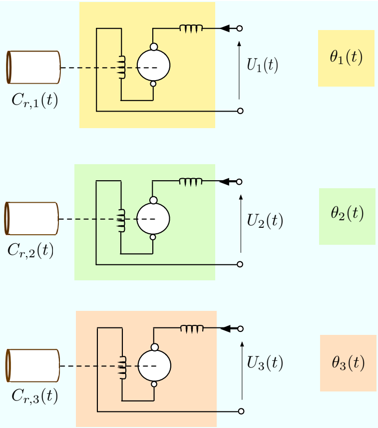
Nancy-Université INPL

CRAN

EDF

Définition et objectifs

- Parc de machines : collection de machines a priori identiques
- Estimer un a modèle générique d'un parc de machines identiques
- Déduire une stratégie générique pour le diagnostic d'un parc de machines



Motivations

- Réduction du coût d'estimation des machines du parc
- Facilité de construire le modèle d'une nouvelle machine
- Facilité de remplacement d'une machine par une autre
- Réduction du coût de maintenance du parc

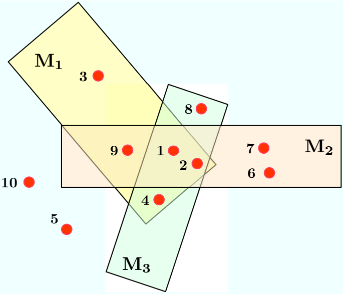
Plan

1. Introduction
2. Méthode
   - Identification des modèles de chaque machine
   - Identification des coefficients identiques des machines
   - Estimation les paramètres en considérant les parties communes
   - Validation du choix des coefficients identiques
3. Simulation
4. Conclusions et perspectives

Introduction

Introduction : modèle générique

- Le problème consiste à établir un modèle générique traduisant le comportement de toutes les machines du parc.
- Un modèle générique comporte deux parties :
  - une partie commune aux variables des machines du parc
  - une partie distincte relative aux variables environementales de chaque machine.
- Le problème consiste en :
  - classification des variables : extraction des variables communes
  - identification des modèles partageant des parties communes
- Exemple : 3 machines M_1, M_2 et M_3 avec 10 variables



	1	2	3	4	5	6	7	8	9	10
M_1	×	×	×	×	.	.	.	.	×	.
M_2	×	×	.	.	.	×	×	.	×	.
M_3	×	×	.	×	.	.	.	×	.	.

Variables 1 and 2
Variables 4 and 9
Variables 3, 5, 6, 7, 8

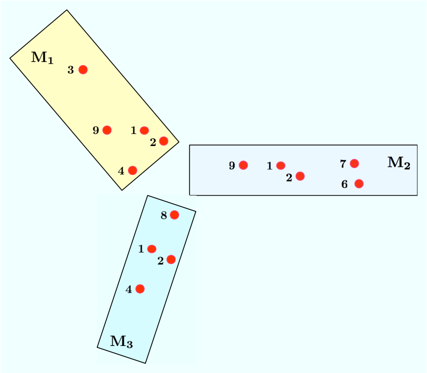
Point de vue historique

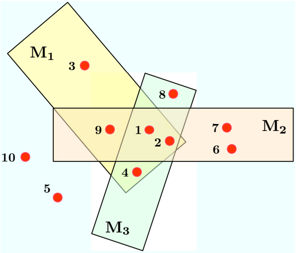
- Multitask learning
  http://books.nips.cc/papers/files/nips19/NIPS2006-0251.pdf
- Fleet Maintenance Systems
  http://www.serco-na.com/Download.aspx?ID=288&Type=Story
- Fleet Inventory Tracking
  http://www.mex.com.au/Products/FleetMEX.aspx
- Patents on fleet of machines
  http://www.freepatentsonline.com/5737215.html

Identification des modèles avec leur partie commune

Identification des modèles et de leurs parties communes

1. Identifier séparément le modèle de chaque machine
2. Identifier les variables communes dans les modèles des différentes machines
3. identifier les paramètres potentiellement identiques → partie commune
4. estimer les paramètres des modèles compte tenu de leurs parties communes
5. valider le choix de la partie commune





Identifier les modèles de chaque machine

- Considérons la kème machine, avec :

  - y^{*k} variable à expliquer
  - W^k variables pouvant expliquer y^{*k}
  - X^k variables sélectionnées pour expliquer y^{*k}
  - $\hat{\theta}^k$ paramètres du modèle

- Modèle

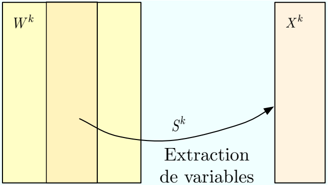
$$\hat{y}^k = X^k \hat{\theta}^k$$

$$X^k = W^k S^k$$

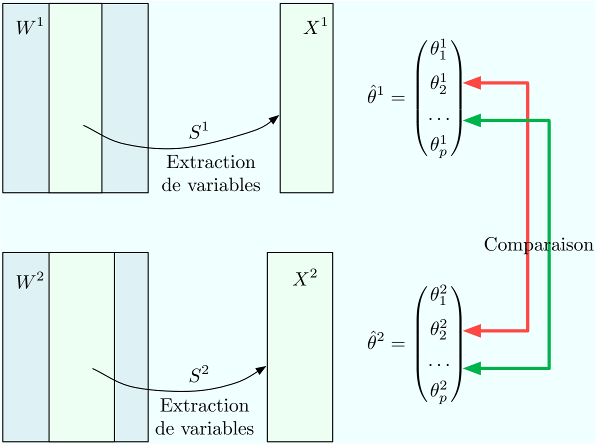
$$\hat{\theta}^k = ({X^k}^T X^k)^{-1} {X^k}^T y^k$$

- S^k et une matrice de sélection. Par exemple, la matrice ci-dessous permet de sélectionner les variables 2 et 4 dans un ensemble de 5 variables :

$$S^k = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}^T$$



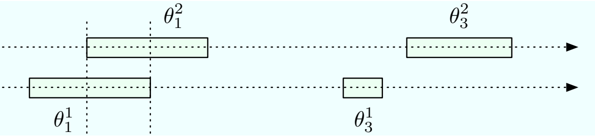
Identifier les modèles de plusieurs machines



- Identifier séparément les paramètres des modèles
- Analyser les proximités des coefficients des variables $\hat{\theta}^1$ et $\hat{\theta}^2$
- Décider quels coefficients doivent être forcés à être identiques

Identification des coefficients identiques des machines

- Principe :



- Intervalles de confiance des paramètres

L'écart-type $\hat{\sigma}_i^k$ est évalué de la façon suivante :

$$\tilde{y}^k = y^k - \hat{y}^k, \quad \hat{\sigma}^2 = \frac{1}{n_k - p_k} \parallel \tilde{y}^k \parallel^2 \quad \rightarrow \quad \boxed{\hat{\Sigma}_\theta^k = \hat{\sigma}^2 ({X^k}^T X^k)^{-1}}$$

$\hat{\sigma}_i^k$ est le i^{th} terme de la diagonale de $\hat{\Sigma}_\theta^k$.

$$I_i^k = [\hat{\theta}_i^k - 2.32\hat{\sigma}_i^k \ ; \ \hat{\theta}_i^k + 2.32\hat{\sigma}_i^k]$$

- Les coefficients $\hat{\theta}_i^k$ $(\forall k)$ sont considérés identiques si il existe intersection non vide entre leurs intervalles de confiance I_i^k :

$$I_i^{k_1} \cap I_i^{k_2} \neq \varnothing \rightarrow \hat{\theta}_i^{k_1} \ \textit{and} \ \hat{\theta}_i^{k_2} \ \textit{are identical}$$

Estimer les paramètres compte tenu des parties communes

- Modèle de la machine k

$$y^k = X^k \theta^k$$

La partie commune des coefficients est constituée des p coefficients α :

$$\theta^k = \begin{pmatrix} \alpha \\ \beta^k \end{pmatrix}, \quad X^k = \begin{pmatrix} U^k \\ V^k \end{pmatrix}$$

Les deux matrices U^k et V^k sont sélectionnées avec S_p^k etb $S_{\bar{p}}^k$, from X^k :

$$U^k = X^k S_p^k \qquad \text{and} \qquad V^k = X^k S_{\bar{p}}^k$$

- Modèle du parc de machines

$$\underbrace{\begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^K \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} U^1 & V^1 & 0 & \cdots & 0 \\ U^2 & 0 & V^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ U^K & 0 & \cdots & 0 & V^K \end{bmatrix}}_{Z} \underbrace{\begin{bmatrix} \alpha \\ \beta^1 \\ \beta^2 \\ \vdots \\ \beta^K \end{bmatrix}}_{\theta}$$

- Estimer les nouveaux coefficients

$$\hat{\theta} = \left(Z^T Z\right)^{-1} Z^T Y \quad \Longrightarrow \quad \hat{\alpha}, \hat{\beta}^1, \ldots, \hat{\beta}^K$$

Validation du choix des coefficients identiques

▷ Première estimation $\hat{y}^k$ de y^k sans effet de couplage

$$\tilde{y}^k = y^k - \hat{y}^k, \qquad \hat{y}^k = X^k \hat{\theta}^k, \qquad \boxed{\Phi_1 = \sum_{k=1}^{K} \| \tilde{y}^k \|^2}$$

▷ Seconde estimation $\hat{\hat{y}}^k$ de y^k avec effet de couplage

$$\tilde{\tilde{y}}^k = y^k - \hat{\hat{y}}^k, \qquad \hat{\hat{y}}^k = \left[U^k \quad V^k \right] \left[\begin{array}{c} \hat{\alpha} \\ \hat{\beta}^k \end{array} \right], \qquad \boxed{\Phi_2 = \sum_{k=1}^{K} \| \tilde{\tilde{y}}^k \|^2}$$

▷ Comparison des sommes des carrés des erreurs

$$\boxed{\tau = \frac{N-P}{(K-1)p} \cdot \frac{\Phi_2 - \Phi_1}{\Phi_1}} \qquad \left\{ \begin{array}{lcll} N & = & \sum_{k=1}^{K} n_k, & \textit{Nombre de données} \\ P & = & \sum_{k=1}^{K} p_k, & \textit{Nombre de paramètres} \end{array} \right.$$

$$\boxed{Si \quad \tau \leq \mathcal{F}_a((K-1)p, N-P) \rightarrow pas\ de\ perte\ significative\ d'information}$$

Remarque : une formulation directe ?

Formulation initiale

- Construction des modèles individuellement

$$\Phi_i = \parallel y_i - X_i\theta_i \parallel^2, \qquad i = 1, \ldots, K$$

- Détermination de la partie commune des différents modèles
- Analyse de la structure des modèles
- Construction d'un modèle global prenant en compte l'identité de certains coefficients

Une formulation en une étape

- Objectif global

$$\Phi = \sum_{i=1}^{K} \frac{1}{2} \parallel y_i - X_i\theta_i \parallel^2 + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \frac{1}{2} \gamma_{i,j} (\theta_i - \theta_j)^T W_{i,j}^2 (\theta_i - \theta_j)$$

- Estimation simultanée

$$\theta_i, W_{i,j}$$

- des valeurs des poids mettant directement en évidence directement les liens entre paramètres

Exemple (simulation)

Exemple

Trois modèles, trois bases de données de 250 observations chaque :

$$y^1 = x_1^1 + 5x_3^1 + 5.5x_5^1 - 10 + \varepsilon^1$$

$$y^2 = x_2^2 + 5x_3^2 + 0.6x_4^2 + 5.68x_5^2 - 12 + \varepsilon^2$$

$$y^3 = 0.5x_1^3 + 1.2x_2^3 + 5.1x_3^3 + 0.7x_4^3 + 5.3x_5^3 - 14 + \varepsilon^3$$

où ε^k est un signal de valeur moyenne nulle

Etape 1 : Identification séparée des modèles

Les estimés de y^k évalués sans l'effet de couplage :

$$\hat{y}^1 = 0.89x_1^1 + 5.09x_3^1 + 5.66{x_5}^1 - 10.29$$

$$\hat{y}^2 = 1.15x_2^2 + 4.89x_3^2 + 0.54x_4^2 + 5.40x_5^2 - 11.52$$

$$\hat{y}^3 = 0.58x_1^3 + 1.18x_2^3 + 4.98x_3^3 + 0.66x_4^3 + 5.31x_5^3 - 13.84$$

Exemple

Etape 2 : Recherche des coefficients identiques

Coef.	Interval
$\hat{\theta}_0^1$	[-11.13 ; -9.46]
$\hat{\theta}_1^1$	[0.62 ; 1.15]
$\hat{\theta}_3^1$	[4.86 ; 5.34]
$\hat{\theta}_5^1$	[5.41 ; 5.91]

Coef.	Interval
$\hat{\theta}_0^2$	[-12.38 ; -10.66]
$\hat{\theta}_2^2$	[0.89 ; 1.39]
$\hat{\theta}_3^2$	[4.66 ; 5.12]
$\hat{\theta}_4^2$	[0.31 ; 0.78]
$\hat{\theta}_5^2$	[5.19 ; 5.61]

Coef.	Interval
$\hat{\theta}_0^3$	[-14.87 ; -12.81]
$\hat{\theta}_1^3$	[0.34 ; 0.81]
$\hat{\theta}_2^3$	[0.87 ; 1.49]
$\hat{\theta}_3^3$	[4.73 ; 5.22]
$\hat{\theta}_4^3$	[0.39 ; 0.93]
$\hat{\theta}_5^3$	[5.09 ; 5.53]

Les coefficients des variables x_3 et x_5 sont unique pour toutes les base de données : une intersection non nulle de [4.86 ; 5.12] et [5.41 ; 5.53] existe entre leurs intervalles de confiance.

Exemple

Etape 3 : Validation du choix des coefficients identiques

Expressions des nouveaux estimés de y^k :

$$\tilde{y}^1 = 0.94x_1^1 + 4.98x_3^1 + 5.44x_5^1 - 9.72$$

$$\tilde{y}^2 = 1.13x_2^2 + 4.98x_3^2 + 0.54x_4^2 + 5.44x_5^2 - 11.75$$

$$\tilde{y}^3 = 0.58x_1^3 + 1.16x_2^3 + 4.98x_3^3 + 0.66x_4^3 + 5.44x_5^3 - 14.09$$

Avant couplage : $\Phi_1 = 127.95$, après couplage : $\Phi_2 = 129.31$

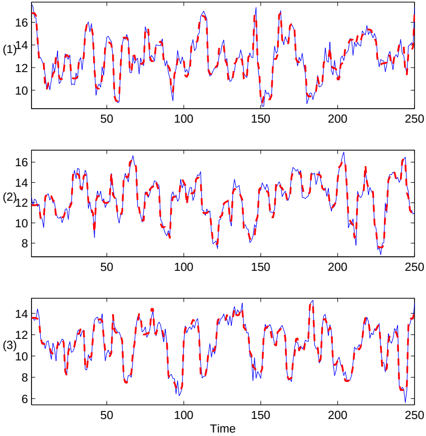
$$\boxed{\tau = \frac{N-P}{(K-1)p} \cdot \frac{\Phi_2 - \Phi_1}{\Phi_1} = 1.95}$$

avec : $N = 750$, $P = 15$, $K = 3$, $p = 2$ et pour un niveau de confiance de 99%

$$\boxed{F_a = 3.34}$$

Donc $\tau \leq F_a$ et ainsi l'effet de couplage peut être pris en compte sans perte significative d'information.

Exemple



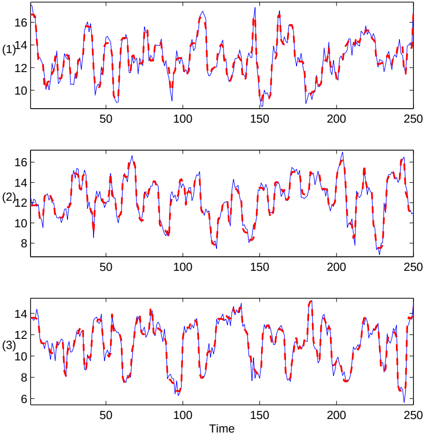


FIGURE: y^k (bleu plein) and $\hat{y}^k$ (pointillé rouge) dans chaque base de données

FIGURE: y^k (bleu plein) and $\tilde{y}^k$ (pointillé rouge) dans chaque base de données

Interpretation :

Conclusions et perspectives

Conclusions et perspectives

Conclusions

- Méthode pour identifier la partie commune dans le comportement des machines d'un parc, en utilisant des données collectées sur ce parc.
- Modèle générique caractérisant le fonctionnement normal des machines établi à partir de leurs caractéristiques communes.

Perspectives

- La méthode peut être complétée pour identifier la partie des modèles à associer à l'environnement des machines.
- Quand une nouvelle machine est acquise, comment reconnaître si elle peut être associée au parc existant ?
- Quand une nouvelle machine est ajoutée au parc, comment établir son modèle compte tenue du modèle générique du parc ?
- Test de la méthodologie sur un parc de pompes de tranches de centrales nucléaire (EDF).

Merci pour votre attention