

DÉTECTION D'ÉVÉNEMENTS ATYPIQUES DANS DES DONNÉES FONCTIONNELLES

Clémentine Barreyre ¹ & Béatrice Laurent ²
& Jean-Michel Loubes ³ & Bertrand Cabon ⁴

¹ *Airbus Defence and Space, clementine.barreyre@airbus.com*

² *INSA Toulouse, beatrice.laurent@insa-toulouse.fr*

³ *Université Paul Sabatier, jean-michel.loubes@math.univ-toulouse.fr*

⁴ *Airbus Defence and Space, bertrand.cabon@airbus.com*

Résumé. Dans de nombreux domaines de l'industrie, le fonctionnement d'un appareil peut être suivi par la production de courbes, et donc de données fonctionnelles. Une courbe atypique peut être associée à un dysfonctionnement de l'appareil. Il est donc important de développer des méthodes statistiques pour détecter ces anomalies. Nous disposons d'observations de ces courbes, et l'objectif est de détecter les anomalies à partir de ces observations. Les courbes sont approximées par leur projection sur un espace de dimension finie engendré par un nombre fini de fonctions de bases. Ensuite, nous proposons d'utiliser une méthode de type "one-class SVM" afin de trouver les courbes atypiques. Ces méthodes seront appliquées à des courbes de mesures de réponse en fréquence provenant des tests satellites.

Mots-clés. Analyse de données fonctionnelles, One-class SVM, Analyse en composantes principales fonctionnelle, méthodes à noyau

Abstract. In many fields of industry, the behaviour of a system can be monitored by the production of curves, thus functional data. An outlier curve can be due to an anomaly in the system. It is then important to develop statistical methods to detect these anomalies. We have observations of these curves, and the goal is to detect anomalies from those observations. The first step is to estimate the coefficients of these functions' projection onto a set of finite dimension. Then, we propose to use a one-class SVM method in order to find the outlier curves. These methods will be applied to frequency responses curves coming from the satellite tests data.

Keywords. Functional Data analysis, One-class SVM, principal component analysis, kernel methods

1 Introduction

Dans de nombreux domaines de l'industrie, les données contrôlant le fonctionnement d'un appareil sont enregistrées au cours du temps, donnant ainsi naissance à des données

fonctionnelles. Une courbe atypique peut être associée à un dysfonctionnement de l'appareil. Il est donc important de développer des méthodes statistiques pour détecter ces anomalies.

Nous disposons de données échantillonnées de courbes, ce qui correspond au modèle d'observation suivant :

$$Y_{j,i} = f_j(x_i), \quad j = 1 \dots J, \quad i = 1 \dots n.$$

Nous supposons lors de cette étude que $x_i \in [0, 1]$, $\forall i = 1, \dots, n$, et que les fonctions $f_j \in \mathbb{L}^2([0, 1])$, $\forall j = 1, \dots, J$. Notre objectif est de détecter les courbes atypiques, à partir de ces observations. Pour cela, nous proposons dans un premier temps d'estimer les coefficients de la projection des fonctions f_j sur des espaces de dimension finie. Dans un second temps, nous proposons d'appliquer sur ces coefficients estimés une méthode de type one class SVM, introduite par Schölkopf et al. [5] afin de détecter les courbes atypiques. Nous présentons les résultats de cette méthodologie sur des courbes gain-fréquence correspondant à des tests réalisés sur des satellites.

Auder et Fischer [1] se sont également appuyés sur des méthodes de projection afin de classifier des données fonctionnelles.

2 Réduction de dimension

2.1 Projection sur une base fonctionnelle issue d'un noyau

Lorsqu'on dispose d'une base $(\phi_l)_{l \geq 1}$ de fonctions orthonormées dans $\mathbb{L}^2([0; 1])$, alors toute fonction f_j peut s'exprimer de la manière suivante :

$$f_j(x) = \sum_{l=1}^{\infty} \alpha_{jl} \phi_l(x).$$

Les $(\alpha_{jl})_{l \geq 1; j=1 \dots J}$ sont les coefficients de f_j dans cette base. Dans ce travail, les fonctions $(\phi_l)_{l \geq 1}$ que nous choisissons sont les fonctions propres du noyau gaussien K_γ pour un certain $\gamma > 0$.

$$K_\gamma(s, t) = \exp(-\gamma(s - t)^2).$$

Le noyau gaussien est universel et permet de représenter des fonctions régulières. Voir Steinwart [7]. De cette manière, il est possible de réduire la dimension des observations en ne retenant que les d premières fonctions propres avec $1 < d < n$.

La classification de courbes avec ce type de projections a été décrite par Gonzalez et Muñoz [2]. Le noyau gaussien K_γ est un noyau de Mercer (une fonction continue symétrique définie positive). D'après le théorème de Mercer [3], K_γ admet la représentation suivante :

$$K_\gamma(x, y) = \sum_{j=1}^{\infty} \lambda_j \phi_j(x) \phi_j(y).$$

Les ϕ_j sont donc les fonctions propres du noyau K_γ , et les λ_j forment une suite décroissante de nombres positifs, tels que $\sum_{j=1}^{\infty} \lambda_j < +\infty$. D'après Steinwart [7], nous savons que les valeurs propres $(\lambda_l)_{l \geq 1}$ décroissent de manière exponentielle. Cette propriété justifie le fait de ne retenir qu'un nombre fini d de vecteurs propres.

Les fonctions f_j , $j = 1, \dots, J$, peuvent donc être approximées grâce aux d premières fonctions propres (voir Schölkopf et Smola [6] pour plus de précisions) :

$$f_{jd}(x) = \sum_{l=1}^d \alpha_{jl} \phi_l(x), \quad (1)$$

où $\alpha_{jl} = \int_0^1 f_j(x) \phi_l(x) dx$. En pratique, $\forall j = 1 \dots J$ nous estimerons les coefficients $(\alpha_{jl})_{l=1, \dots, d}$ par $(\hat{\alpha}_{jl})_{l=1, \dots, d}$, avec :

$$\hat{\alpha}_{jl} = \frac{1}{n} \sum_{i=1}^n Y_{ji} \phi_l(x_i).$$

Nous appliquerons la One-class SVM sur ces coefficients $\hat{\alpha}_{jl}$.

2.2 Analyse en composantes principales fonctionnelle

L'analyse en composantes principales fonctionnelle (ou FPCA) permet de projeter les courbes sur une base de fonctions orthonormées dans \mathbb{L}^2 de manière à maximiser la variance expliquée par ces projections. Ramsay et Silverman [4] proposent une introduction détaillée de la méthode.

N'ayant pas voulu spécifier de modèle probabiliste pour nos fonctions f_j , nous allons utiliser la matrice de covariance empirique, notée $\mathbf{\Gamma}_J$. Soit $\mathbf{Y}_j = (Y_{j1}, \dots, Y_{jn})^T$ et $\bar{\mathbf{Y}} = (Y_{\cdot 1}, \dots, Y_{\cdot n})^T$, avec $Y_{\cdot i} = \frac{1}{J} \sum_{j=1}^J Y_{ji}$, $i = 1, \dots, n$. Nous avons alors :

$$\mathbf{\Gamma}_J = \frac{1}{J} \sum_{j=1}^J (\mathbf{Y}_j - \bar{\mathbf{Y}})(\mathbf{Y}_j - \bar{\mathbf{Y}})^T.$$

On se donne $d < n$ et on note (Φ_1, \dots, Φ_d) une famille orthonormée formée des vecteurs propres de $\mathbf{\Gamma}_J$ associés aux d plus grandes valeurs propres. Nous supposons que cette famille est orthonormée pour le produit scalaire $\langle \cdot, \cdot \rangle_n$ sur \mathbb{R}^n défini par :

$$\forall u, v \in \mathbb{R}^n, \langle u, v \rangle_n = \frac{1}{n} \sum_{i=1}^n u_i v_i.$$

Nous projetons les vecteurs Y_j sur l'espace engendré par (Φ_1, \dots, Φ_d) , opérant ainsi une réduction de dimension :

$$\hat{Y}_j = \sum_{l=1}^d \hat{\alpha}_{jl} \Phi_l$$

où $\hat{\alpha}_{jl} = \frac{1}{n} \sum_{i=1}^n Y_{ji} \Phi_{li}$. Comme dans le cas précédent, nous travaillerons sur les coefficients $\hat{\alpha}_{jl}$ lors de la prochaine étape.

3 One-class SVM

3.1 Quelques rappels sur la one-class SVM

La méthode “one-class SVM” (ou OCSVM) a été introduite par Schölkopf et al. [5]. Nous disposons d’un échantillon $(x_1, \dots, x_n) \in \mathbb{R}^d$ et d’un noyau gaussien k .

L’idée de la OCSVM est de développer un algorithme qui définit une fonction h qui retourne 1 si la donnée considérée est dans une région de forte densité, -1 sinon. Les données sont donc transformées via le noyau k . Après transformation, le problème revient à trouver l’hyperplan séparateur entre les régions de forte densité et de faible densité de manière à maximiser la marge ρ entre les deux ensembles. Le problème se ramène à la résolution du problème d’optimisation suivant :

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j k(x_i, x_j) \\ \text{t.q.} \quad & 0 \leq \alpha_i \leq \frac{1}{\nu n}, \quad \sum_{i=1}^n \alpha_i = 1. \end{aligned} \quad (2)$$

Le paramètre $\nu \in]0, 1[$ est un paramètre fixé par l’utilisateur. Il représente le pourcentage minimum de valeurs nominales qui sera mis en évidence par la one-class SVM. La fonction de décision relative à ce problème est égale à :

$$h(x) = \text{sgn}\left(\sum_i \alpha_i k(x_i, x) - \rho\right).$$

Elle renvoie 1 si l’individu x se situe dans la région de plus forte densité. Plus ν sera proche de 0, plus on détectera de données atypiques.

Les anomalies trouvées sont donc les données situées au delà de l’hyperplan défini par (2).

$$\{x_j, 1 \leq j \leq n, \sum_{i=1}^n \alpha_i k(x_i, x_j) - \rho < 0\}.$$

La performance de la OCSVM dépend du choix du noyau. Pour une même valeur de ν , on peut obtenir des ensembles contenant les valeurs nominales de tailles différentes. Dans

un objectif de détecter des courbes atypiques, il sera pertinent de faire en sorte que le volume de l'ensemble nominal $G = \{x, \sum_{i=1}^n \alpha_i k(x, x_i) - \rho \geq 0\}$ soit le plus petit possible. Il existe des méthodes pour affiner le choix du noyau, telle que l'ont fait Thomas et al. [8] pour sélectionner la fenêtre optimale dans un noyau gaussien.

3.2 Application sur les données fonctionnelles

On observe des données associées à J courbes échantillonnées sur n points : Y_1, \dots, Y_n où $Y_{ji} = f_j(x_i)$, $j \in 1 \dots J$, $i = 1 \dots n$. Lors de la section 2, nous avons vu comment résumer ces fonctions en un nombre restreint de coefficients $\hat{\alpha}_{jl}$, $l = 1 \dots p$ grâce à la décomposition du noyau gaussien ou de la FPCA. Ces coefficients sont répertoriés dans une matrice Λ .

$$\Lambda = \begin{pmatrix} \hat{\alpha}_{11} & \dots & \hat{\alpha}_{1p} \\ \vdots & \ddots & \vdots \\ \hat{\alpha}_{J1} & \dots & \hat{\alpha}_{Jp} \end{pmatrix} = \begin{pmatrix} \Lambda_1 \\ \vdots \\ \Lambda_J \end{pmatrix}$$

Une fois la décomposition faite et le noyau k désigné, la fonction

$$h(\Lambda) = \text{sgn}\left(\sum_{i=1}^J \alpha_i k(\Lambda_i, \Lambda) - \rho\right)$$

va déterminer si la courbe correspondante est standard ou atypique.

4 Application : courbes gain-fréquence sur les tests satellites

4.1 Contexte industriel

Pour les satellites de télécommunication, le principe général de validation est d'envoyer des signaux Radio-Fréquence (RF) de référence à l'entrée du répéteur. Ces signaux (ou porteuses pures) représentent les signaux envoyés par une station sol et captés par le satellite au travers de son antenne de réception. On mesure alors le signal de sortie qui, après amplification, est renvoyé vers la zone de couverture terrestre, au travers de l'antenne de transmission.

Le test Gain-Fréquence réalise un balayage en fréquence et permet d'observer le gain entre le signal d'entrée et le signal de sortie sur une bande passante donnée. L'objectif est de vérifier la stabilité du gain dans la bande passante et de détecter les comportements atypiques (non-linéarité, oscillation, pentes...).

Dans la pratique, le gain est un bon indicateur de performance du répéteur, il s'agit de vérifier l'impact du chemin parcouru par le signal au travers des équipements passifs (guides d'onde, câbles coaxiaux...) ainsi que l'impact des équipements actifs traversés sur

cette performance attendue. La réponse idéale du Gain-Fréquence serait une indicatrice sur la performance. Le but est donc de trouver les résultats les plus dégradés parmi les courbes résultant du Gain-Fréquence.

L'application mise en place va s'appuyer sur une série de tests pour laquelle on connaît les résultats fournis par l'ingénieur. Il est important de savoir que cela permettra de valider les méthodes et non d'appliquer des méthodes supervisées.

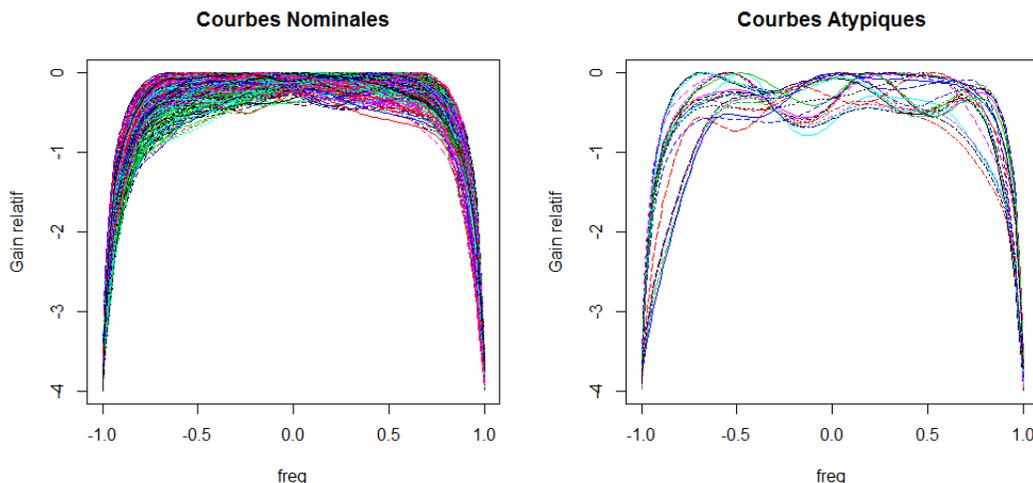


FIGURE 1 – Courbes normales et atypiques validées par l'expert validation

Comme on peut l'observer sur le graphique, les courbes atypiques sont les courbes comprenant des ondulations ou une perte plus rapide du gain. Les courbes représentent le gain relatif sur la bande passante, normalisée sur $[-1; 1]$.

4.2 Résultats

Il est important d'avoir un taux global de bien classés élevé, mais avant tout, il est nécessaire de bien détecter toutes les anomalies les plus significatives. A cette fin, nous réalisons plusieurs simulations en jouant sur plusieurs paramètres :

- La réduction de dimension utilisée. Dans le cas de la décomposition en fonctions propres du noyau K_γ , il s'agit de fixer le paramètre γ , que l'on fera varier entre 0.1 et 10. Nous étudierons également les coefficients issus de la FPCA.
- La dimension p retenue que l'on fera varier de 2 à 10.
- Le noyau k utilisé dans le cadre de la one-class SVM. Nous étudierons les noyaux gaussiens : $k(x, y) = \exp(-\delta(x - y)^2)$, en faisant varier δ .
- Le niveau ν de la one-class SVM, qui sera déterminant pour définir la marge ρ . Plus ν sera petit, plus on aura d'anomalies détectées.

Après avoir affiné les paramètres, on obtient le meilleur modèle pour $p = 3$ et en choisissant les coefficients de la FPCA. Le noyau de la One-class SVM est un noyau gaussien de paramètre $\delta = 1/p$. Les résultats pour $\nu = 0.82$ sont :

Résultats, $\nu = 0.82$	Corrects	Atypiques
Déectés valides	271	4
Déectés atypiques	30	21

Le taux de bonne classification est de 89,6%. On se restreint à la validation de seulement 15.6% des courbes, ce qui représente un gain de temps considérable. On remarque que 4 courbes n’ont pas été identifiées par la OCSVM. Afin de trouver toutes les anomalies, il faut opter pour $\nu < 0.735$, on trouve dans ce cas :

Résultats, $\nu = 0.735$	Corrects	Atypiques
Déectés valides	238	0
Déectés atypiques	63	25

On arrive tout de même à pré-valider 76,2% des courbes, ce qui nous donne un résultat satisfaisant.

5 Conclusion

Les méthodes de one-class clustering sont très efficaces pour détecter des courbes atypiques. Elles peuvent facilement être utilisées sur des fonctions, décomposées dans des bases fonctionnelles. Les méthodes à noyau sont des méthodes qui permettent d’explorer un grand nombre de bases différentes, mais la difficulté réside dans le choix du noyau le plus approprié au problème étudié. L’analyse des coefficients de la FPCA par une one-class SVM à noyau gaussien donne de très bons résultats, car toutes les courbes les plus atypiques ont été mises en évidence par la méthode. Une phase d’auto-adaptation du noyau telle que décrite par Thomas et al. [5] peut être un plus pour la détection. En effet les courbes détectées sont les plus atypiques de l’échantillon, mais parmi elles, on trouve également les courbes les plus “ parfaites ”.

D’autres méthodes auraient pu être utilisées : local outlier factor, iForest... D’autres bases fonctionnelles vont être testées, comme les bases d’ondelettes qui sont susceptibles de donner de bons résultats.

Références

- [1] B. Auder and A. Fischer. Projection-based curve clustering. February 2011.

- [2] Javier González and Alberto Muñoz. Representing functional data in reproducing kernel hilbert spaces with application to clustering and classification. *Statistics and Econometrics 013*, 2010.
- [3] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philos. Trans. Roy. Soc. London*, 1909.
- [4] J.O Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, 2005.
- [5] B. Schölkopf, J. Platt, A.J. Shawe-Taylor, J. Smola, and R.C. Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 2001.
- [6] B. Schölkopf and A. J. Smola. *Learning with Kernels*. MIT Press, 2000.
- [7] I. Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of Machine Learning Research 2*, 2001.
- [8] A. Thomas, V. Feuillard, and A. Gramfort. Calibration of one class svm for mv set estimation. August 2015.