

Détection de Comportements Anormaux dans des Trajectoires de Navires avec One-Class SVM et Dynamic Time Warping

Valérien MANGÉ^{1,5}, Jean-Yves TOURNERET^{2,5}, François VINCENT³, Laurent MIRAMBELL¹, Fábio MANZONI VIEIRA¹,
Barbara PILASTRE⁴

¹Hensoldt Nexeya France, Toulouse (FR), firstname.lastname@hensoldt.fr,

²Université de Toulouse, ENSEEIHT-IRIT, Toulouse (FR), firstname.lastname@toulouse-inp.fr,

³ISAE-Supaéro, Toulouse (FR), firstname.lastname@isae-supaero.fr,

⁴DGA, Rennes (FR),

⁵Laboratoire TéSA, Toulouse (FR), firstname.lastname@tesa.prd.fr

Recherche soutenue financièrement par le Ministère des Armées - Agence de l'innovation de défense



Résumé – La détection d’anomalies dans le domaine de la surveillance maritime est un enjeu majeur pour la sécurité des navires et des nations. De nombreux algorithmes de détection d’anomalies sont disponibles dans la littérature mais ils ne sont pas tous adaptés à l’analyse de séries temporelles comme les trajectoires de navires, en particulier lorsque ces trajectoires n’ont pas la même longueur. Cet article présente une manière d’adapter l’algorithme One-Class SVM à la détection d’anomalies dans des séries temporelles associées à des trajectoires de navires à l’aide d’un noyau basé sur une mesure de similarité de type “dynamic time warping”.

Abstract – Anomaly detection in the field of maritime surveillance is of major importance for the safety of ships and nations. Numerous anomaly detection algorithms are available in the state-of-the-art but they are not necessarily adapted to the analysis of time series such as ship trajectories, in particular when these trajectories have different lengths. This article studies an algorithm allowing the One-Class SVM method to be adapted to time series associated with ship trajectories by means of a kernel based on a dynamic time warping similarity measure.

1 Introduction

Contexte. De nombreuses opérations se font par les voies maritimes. Ces opérations sont susceptibles d’être perturbées ou peuvent être sources d’activités illégales. Ainsi, il convient d’effectuer des missions de surveillance maritime pour assurer la sécurité des intérêts nationaux [1]. Cependant, du fait de leur nombre important, l’analyse des comportements des navires ne peut se faire à la main en un temps raisonnable et doit donc être automatisée. Différents capteurs existent pour suivre le mouvement des bateaux, comme l’AIS (Automatic Identification System), le radar, les images à synthèse d’ouverture (images SAR pour “Synthetic Aperture Radar”) et d’autres encore, qui peuvent être combinés. La détection d’anomalies à partir de ces données, même après association, reste un sujet de recherche important, notamment car les trajectoires de bateaux sont des séries temporelles de longueurs généralement différentes.

État-de-l’art. Plusieurs algorithmes de détection d’anomalies ont été proposés dans la littérature en fonction de l’application visée [2]. L’algorithme One-Class SVM [3] fait partie de ceux les plus utilisés. Cet algorithme doit généralement travailler dans un espace de même dimension pour toutes les données d’apprentissage et de test. Ainsi, les données sont représen-

tées sous forme de vecteurs qui nourrissent cet algorithme. Ceci pose problème avec des séries temporelles dont la longueur et l’échelle du temps peuvent varier d’une mesure à une autre. Un algorithme de l’état de l’art permettant de comparer des séries temporelles de longueurs différentes a été proposé par Laxhammar [4]. Il utilise une détection d’anomalies conforme avec la distance de Hausdorff et sera notre méthode de référence.

Objectifs, contributions et organisation. L’objectif de cet article est de présenter un moyen d’adapter la méthode One-Class SVM à la détection d’anomalies dans des séries temporelles issues de trajectoires de navires dans un contexte de surveillance maritime. La méthode proposée sera comparée à une méthode de l’état de l’art basée sur la détection d’anomalies conforme. L’article est organisé comme suit : la section 2 rappelle le fonctionnement théorique de l’algorithme One-Class SVM. La section 3 explique comment adapter la méthode One-Class SVM à des séries temporelles de longueurs potentiellement différentes tandis que la section 4 s’intéresse à l’estimation de ses hyperparamètres. La méthode proposée dans cet article et la méthode issue de [4] sont comparées dans la section 5, pour la détection d’anomalies dans des trajectoires synthétiques. Les conclusions de cet article sont résumées dans la section 6.

2 Détection d'anomalies

La détection d'anomalies consiste à identifier des données qui ne sont pas conformes à un comportement normal [2]. En réalité, il n'existe pas de définition précise d'une anomalie et l'hypothèse souvent retenue dans ce domaine est que contrairement aux cas normaux, les anomalies se produisent rarement. Ceci rend l'entraînement de modèles compliqué car les deux classes sont déséquilibrées (il y a moins d'anomalies que de données normales) et cet entraînement doit se faire de manière non supervisé car il est difficile d'identifier a priori toutes les anomalies potentielles. Une méthode permettant de faire cet apprentissage est la méthode One-Class SVM dont le fonctionnement est rappelé dans la section suivante.

2.1 Méthode One-Class SVM

Nous rappelons dans cette section le fonctionnement de la méthode One-Class SVM. Cette méthode suppose que les données représentant la classe normale, sont linéairement séparables par rapport à l'origine et va chercher la frontière séparatrice la plus proche possible des données, i.e., maximisant la marge de l'ensemble des données d'entraînement. Considérons un ensemble de données d'entraînement $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ avec $N \geq 1$ et $\forall i \in \{1, \dots, N\}$, $\mathbf{x}_i \in \mathbb{R}^D$ avec D la longueur des données. La séparation linéaire des données de l'origine se traduit par le problème d'optimisation suivant :

$$\begin{aligned} \underset{\boldsymbol{\omega}, \rho, \xi_i}{\text{minimize}} \quad & \frac{1}{2} \|\boldsymbol{\omega}\|^2 - \rho + \frac{1}{\nu N} \sum_{i=1}^N \xi_i \\ \text{avec} \quad & \langle \boldsymbol{\omega}, \mathbf{x}_i \rangle \geq \rho - \xi_i, \forall i \in \{1, \dots, N\} \end{aligned} \quad (1)$$

où $\langle \boldsymbol{\omega}, \mathbf{x}_i \rangle = \boldsymbol{\omega}^T \mathbf{x}_i$ et ν est un majorant de la proportion de données anormales dans l'ensemble \mathcal{X} .

Supposons de plus avoir une transformation Φ qui projette nos données dans un espace de dimension supérieure telle que $\forall i \in \{1, \dots, N\}$, $\Phi(\mathbf{x}_i) \in \mathbb{R}^F$ avec $F \geq D$. Les contraintes de (1) dans \mathbb{R}^F s'écrivent :

$$\langle \boldsymbol{\omega}, \Phi(\mathbf{x}_i) \rangle \geq \rho - \xi_i, \forall i \in \{1, \dots, N\},$$

ce qui donne la fonction de décision

$$f(\mathbf{x}) = \text{sgn}(\langle \boldsymbol{\omega}, \Phi(\mathbf{x}) \rangle - \rho), \quad (2)$$

où sgn est la fonction "signe". Si on pose $\boldsymbol{\omega} = \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_i)$, $0 \leq \alpha_i \leq \frac{1}{\nu N}$ et $\sum_{i=1}^N \alpha_i = 1$ et si κ est le noyau défini par $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^D$, $\kappa(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$, on obtient :

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^N \alpha_i \kappa(\mathbf{x}_i, \mathbf{x}) - \rho \right). \quad (3)$$

L'astuce du noyau permettant d'obtenir (3) est utilisée dans la méthode One-Class SVM dès que les données ne sont pas linéairement séparables dans l'espace d'origine, afin de projeter les données dans un espace où elles deviennent linéairement séparables. Un noyau classique utilisé dans cet article est le noyau

gaussien défini par $\kappa(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right)$ où $\sigma > 0$ est un hyperparamètre à estimer. Il est possible de démontrer que l'application Φ associée à ce noyau permet de projeter les données dans un espace de dimension infinie.

2.2 Hyperparamètre de décalage

Si on utilise le noyau gaussien, on doit estimer les deux hyperparamètres ν et σ . Un troisième hyperparamètre $\delta \geq 0$ appelé "paramètre de décalage" est introduit dans [5] de manière à obtenir la fonction de décision

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^N \alpha_i \kappa(\mathbf{x}_i, \mathbf{x}) - \rho + \delta \right).$$

Ce paramètre de décalage (qui est fixé après détermination de (ν, σ)) permet d'éloigner la frontière des données afin de diminuer le nombre de faux positifs avec le risque d'augmenter le nombre de faux négatifs.

3 Dynamic Time Warping (DTW)

La méthode One-Class SVM s'applique à des vecteurs de données \mathbf{x}_i qui ont la même dimension, ce qui n'est pas le cas avec des séries temporelles de longueurs différentes. Une mesure de similarité, appelée "Dynamic Time Warping (DTW)", permettant de quantifier la ressemblance entre deux séries temporelles a été utilisée pour les signaux audio [6]. Soient $\mathbf{t} = (t_1, \dots, t_n)$, $n \geq 1$ et $\mathbf{u} = (u_1, \dots, u_m)$, $m \geq 1$ deux séries temporelles éventuellement de longueurs différentes. DTW définit une correspondance entre les éléments des deux séries de telle sorte à rapprocher les éléments similaires. Si $s(i, j)$ est le score de similarité cumulé jusqu'aux éléments i de \mathbf{t} et j de \mathbf{u} , le score associé à DTW est défini de manière récursive par :

$$s(i, j) = \min\{s(i-1, j-1), s(i, j-1), s(i-1, j)\} + \|\mathbf{t}_i - \mathbf{u}_j\|^2,$$

avec $1 \leq i \leq n, 1 \leq j \leq m$ et avec la contrainte que les premiers et derniers éléments de chaque série correspondent. Le score de similarité entre les deux séries temporelles renvoyé par DTW est alors $s(n, m)$. La figure 1 illustre cette correspondance et montre que DTW est mieux adapté que la distance euclidienne pour comparer les deux séries échantillonnées différemment. On rappelle que DTW ne possède pas toutes les propriétés pour définir une distance car l'inégalité triangulaire n'est pas vérifiée mais que cette mesure peut être utilisée pour des trajectoires de longueurs différentes.

Inspiré de [7] et [8], nous proposons de considérer un noyau défini à l'aide de DTW par :

$$\kappa(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\text{DTW}^2(\mathbf{x}, \mathbf{y})}{2\sigma^2}\right), \quad (4)$$

ce qui permet d'appliquer la méthode One-Class SVM à des séries temporelles de longueurs différentes. On notera que ce

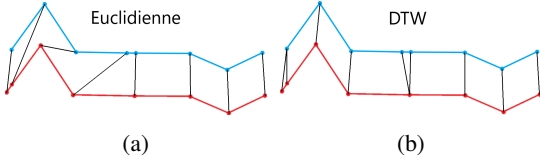


FIGURE 1 – Norme euclidienne et DTW pour deux trajectoires identiques échantillonnées différemment.

noyau a été utilisé dans [7] et [8] pour des problèmes de classification supervisée. Cet article étudie l'intérêt potentiel de ce noyau pour la détection non supervisée de trajectoires anormales de bateaux. L'objectif final est de développer un outil de surveillance maritime utilisant des données de plusieurs capteurs, notamment des capteurs AIS et SAR. Dans ce cas, nous pouvons définir des noyaux κ_{AIS} et κ_{SAR} (pouvant utiliser DTW) et des constantes positives λ_{AIS} et λ_{SAR} tels que

$$\kappa(\mathbf{x}, \mathbf{y}) = \lambda_{\text{AIS}}\kappa_{\text{AIS}}(\mathbf{x}, \mathbf{y}) + \lambda_{\text{SAR}}\kappa_{\text{SAR}}(\mathbf{x}, \mathbf{y}) \quad (5)$$

définisse un noyau conjoint adapté aux données AIS et SAR.

4 Estimation des hyperparamètres

La méthode One-Class SVM nécessite de régler trois hyperparamètres, à savoir ν , σ et δ . L'hyperparamètre ν représente la fraction maximale de données anormales dans l'ensemble d'entraînement qui est généralement choisie par l'utilisateur. On cherche en général à avoir un faible nombre d'anomalies dans \mathcal{X} d'où $\nu \ll 1$. L'hyperparamètre σ associé au noyau gaussien et δ le terme de décalage doivent être estimés pour optimiser la détection d'anomalies. Dans cette section, nous considérons un ensemble de données d'entraînement $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ qui est utilisé pour déterminer σ et δ .

4.1 Optimisation du paramètre σ

Pour l'optimisation de σ , Trinh et al. [5] proposent de maximiser la fonction objectif définie par :

$$J(\sigma) = \frac{2}{N} \sum_{i=1}^N \left[\exp\left(-\frac{P(\mathbf{x}_i)}{2\sigma^2}\right) - \exp\left(-\frac{L(\mathbf{x}_i)}{2\sigma^2}\right) \right] \quad (6)$$

avec, pour $i = 1, \dots, N$,

$$L(\mathbf{x}_i) = \max_{1 \leq j \leq N} \text{DTW}^2(\mathbf{x}_i, \mathbf{x}_j),$$

$$P(\mathbf{x}_i) = \min_{1 \leq j \leq N, i \neq j} \text{DTW}^2(\mathbf{x}_i, \mathbf{x}_j).$$

Le maximum de J est calculé par descente de gradient. Pour le point initial, nous avons utilisé la méthode de [9, p. 83-88], qui consiste à choisir comme première estimation de σ la médiane des distances entre les trajectoires (prises deux à deux).

4.2 Choix du paramètre δ

La valeur de δ dépend des besoins de l'utilisateur. Ce paramètre peut augmenter le nombre de détections au prix de moins

de fausses alarmes et inversement. Nous proposons de choisir le paramètre δ à partir d'une valeur particulière de la courbe Précision-Rappel. Dans cet article, nous avons choisi la valeur de δ qui maximise le score F_1 ¹. Les métriques Précision, Rappel, F_1 et Exactitude (Exact) sont définies dans (7), où VP, VN, FP et FN désignent les nombres de vrais positifs, de vrais négatifs, de faux positifs et de faux négatifs :

$$\begin{aligned} \text{Précision} &= \frac{\text{VP}}{\text{VP} + \text{FP}}, \quad \text{Rappel} = \frac{\text{VP}}{\text{VP} + \text{FN}}, \\ F_1 &= 2 \left(\frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \right), \quad \text{Exact} = \frac{\text{VP} + \text{VN}}{N}. \end{aligned} \quad (7)$$

5 Expérimentations

La méthode One-Class SVM avec DTW a été implantée sous Python avec les modules **NumPy** pour la manipulation de matrices, **tdaidistance** pour DTW, **scikit learn** pour **One-Class SVM** et **SciPy** pour l'estimation de σ . Nous avons réalisé deux ensembles d'expériences, avec des trajectoires complètes (Section 5.1) puis avec des données manquantes (Section 5.2).

5.1 Trajectoires complètes

Les premières simulations ont été effectuées à l'aide des données synthétiques de [11] qui rassemblent $N_T = 1000$ ensembles de trajectoires notés \mathcal{E}_n , $n = 1, \dots, N_T$ possédant chacune 16 couples de positions notés $(x_i(1), x_i(2))$ (d'où $D = 16$). Chaque ensemble \mathcal{E}_n est constitué de 260 trajectoires réparties en 5 rails de 50 trajectoires normales et 10 anomalies. Ces données, ainsi que les résultats issus de [4], servent de référence pour évaluer les performances.

Pour cette première expérience, nous calculons toutes les valeurs de noyaux entre les différents couples de trajectoires pour un ensemble donné \mathcal{E}_n et déterminons l'hyperparamètre σ_n minimisant (6). Pour cette valeur de σ_n , et $\nu = 0.12$ (fixé par l'utilisateur), on détermine le paramètre δ_n maximisant F_1 avec 100 ensembles normaux pris au hasard. Les 900 ensembles restants sont utilisés pour déterminer les performances de la méthode One-Class-SVM. La table 1 résume les résultats obtenus avec la méthode One-Class SVM avec un noyau DTW. Ces résultats sont comparés à ceux obtenus avec un noyau TWED (Time Warp Edit Distance) et la distance de Hausdorff (HD) qui définissent deux mesures de similarité entre séries temporelles de longueurs différentes (voir [7] et [4]). On observe que pour les trajectoires utilisées, le noyau TWED est légèrement meilleur que le noyau DTW mais la détection d'anomalies avec TWED nécessite 20 fois plus de temps qu'avec DTW. Enfin, ces résultats sont très prometteurs par rapport à ceux issus de [4] fournis dans la table 2.

1. On notera que les courbes précision-rappel sont mieux adaptées à la détection d'anomalies que les courbes COR car les classes "normale" et "anormale" sont déséquilibrées [10] (voir Section 5.2).

	Précision	Rappel	F1	Exact
DTW $\delta = 0$	27.82%	86.80%	42.12%	90.82%
DTW $\delta > 0$	96.79%	69.02%	79.35%	98.70%
DTW $\delta > 0$ données manquantes	93.97%	64.39%	75.21%	98.47%
TWED $\delta > 0$	95.57%	78.06%	85.14%	99.00%
HD $\delta > 0$	93.83%	66.32%	76.31%	98.51%

TABLE 1 – Mesures de performance pour la détection d’anomalies avec la méthode One-Class SVM et différents noyaux.

	SHNN-CAD	One-Class SVM	Discords
Exact	97.09%	96.3%	97.06%

TABLE 2 – Métriques “exactitude” issues de [4].

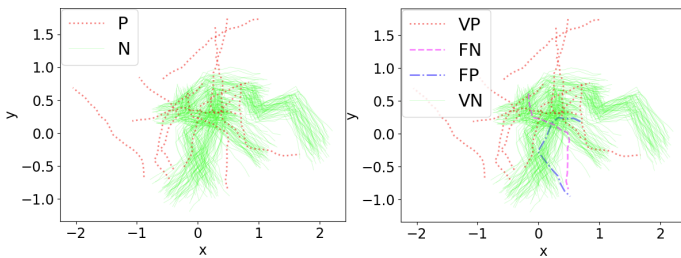


FIGURE 2 – Trajectoires avant (gauche) et après (droite) détection d’anomalies (avec un faux positif et un faux négatif).

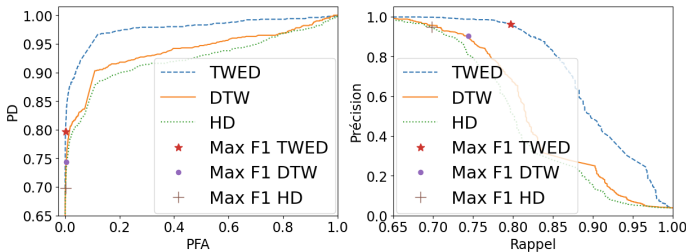


FIGURE 3 – Courbes ROC et Précision-Rappel pour évaluer δ .

5.2 Trajectoires avec données manquantes

Nous répétons l’expérience précédente en retirant au hasard quelques points par trajectoire pour avoir des trajectoires de longueurs différentes. Plus précisément, pour chaque trajectoire, on choisit un nombre de points manquants uniformément dans $\{0, \dots, 5\}$ et on enlève ensuite de manière équiprobable ces points dans la trajectoire. Les résultats obtenus sont donnés dans la table 1. Le fait d’avoir des données manquantes engendre une légère baisse de performance mais le modèle est toujours utilisable pour des trajectoires de longueur variable. Notons que dans certaines applications pratiques, on souhaitera prendre une décision avec une portion de trajectoire observée par un opérateur, ce qui induit des trajectoires courtes et un traitement séquentiel qui fera l’objet de travaux futurs.

6 Conclusion

Ces travaux illustrent l’intérêt du Dynamic Time Warping (DTW) comme outil de comparaison de séries temporelles de longueurs potentiellement variables dans des algorithmes de détection d’anomalies traditionnels. Plus précisément, l’intérêt d’utiliser un noyau basé sur DTW dans la méthode One-Class SVM a été montré pour la détection de trajectoires anormales. On peut noter que DTW pourrait aussi être utilisé comme score de similarité dans d’autres algorithmes comme Local Outlier Factor, Local Outlier Probability (LoOP) ou Isolation Forest [2], ce qui est en cours d’étude. Une autre perspective est d’évaluer la méthode proposée pour des données réelles multi-capteurs associées à la surveillance maritime.

Références

- [1] M. Guerriero, P. Willett, S. Coraluppi, and C. Carthel, “Radar/AIS data fusion and SAR tasking for maritime surveillance,” in *Proc. Int. Conf. Information Fusion (FUSION)*, Cologne, Germany, Aug. 2008, pp. 1–5.
- [2] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection : A survey,” *ACM Computing Surveys*, vol. 41, no. 3, 2009.
- [3] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the support of a high-dimensional distribution,” *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [4] R. Laxhammar and G. Falkman, “Online learning and sequential anomaly detection in trajectories,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1158–1173, 2014.
- [5] V. V. Trinh, K. P. Tran, and T. T. Huong, “Data driven hyperparameter optimization of one-class support vector machines for anomaly detection in wireless sensor networks,” in *Proc. Int. Conf. Advanced Tech. for Commun. (ATC)*, Quy Nhon, Vietnam, 2017, pp. 6–10.
- [6] E. Keogh and C. Ratanamahatana, “Exact indexing of dynamic time warping,” *Knowl. Inf. Syst.*, vol. 7, pp. 358–386, 2005.
- [7] M. Badiane and P. Cunningham, “An empirical evaluation of kernels for time series,” *Artificial Intelligence Review*, vol. 55, pp. 1803–1820, July 2021.
- [8] H. Lei and B. Sun, “A study on the dynamic time warping in kernel machines,” in *Proc. Int. Conf. Signal-Image Technologies and Internet-Based System*, Shanghai, China, 2007, pp. 839–845.
- [9] C. C. Aggarwal, *Outlier Analysis*, Springer Verlag, New York, 2016.
- [10] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” *PLOS One*, vol. 10, no. 3, pp. 1–21, 2015.
- [11] C. Piciarelli, C. Micheloni, and G. L. Foresti, “Trajectory-based anomalous event detection,” *IEEE Trans. Circuits Sys. Video Technol.*, vol. 18, no. 11, pp. 1544 – 1554, 2008.