

# Comparaison de méthodes d'apprentissage automatique pour détecter les anomalies dans l'activité des animaux

Nicolas Wagner<sup>\*,\*\*,\*\*\*</sup>, Violaine Antoine<sup>\*</sup>, Jonas Koko<sup>\*</sup>, Marie-Madeleine Mialon<sup>\*\*</sup>,  
Romain Lardy<sup>\*\*</sup>, Isabelle Veissier<sup>\*\*</sup>

<sup>\*</sup>UCA, LIMOS, UMR 6158, CNRS, Clermont-Ferrand, France

<sup>\*\*</sup>UCA, INRAE, UMR Herbivores, F-63122 Saint-Genès-Champanelle, France

<sup>\*\*\*</sup> nicolas.wagner@uca.fr

## 1 Introduction

De plus en plus de domaines scientifiques collectent des données et cherchent à en extraire de l'information pertinente. C'est par exemple le cas pour l'élevage de précision. Les éleveurs doivent en effet détecter le plus rapidement possible toute anomalie chez les animaux afin d'assurer l'efficacité de la production et le bien-être des animaux. Une telle détection semble possible grâce à l'analyse des activités des animaux (Veissier et al., 2017). Dans cette étude, nous travaillons sur des données de géolocalisation des animaux qui illustrent leur comportement sous la forme de séries temporelles. Ces données sont caractérisées par un cycle de 24 heures, un bruit important lié à leur acquisition comme par exemple l'imprécision des capteurs, la récupération manuelle des étiquettes, etc. et une fluctuation plus ou moins important due essentiellement à l'environnement comme par exemple la proximité d'autres animaux. Nous proposons une méthode basée sur les transformées de Fourier (**FBAT**) pour détecter les anomalies de cycles dans les séries temporelles. Nous comparons FBAT avec les meilleures méthodes d'apprentissage actuelles pour la classification des séries temporelles.

## 2 Fourier Based Approximation with Thresholding

La méthode FBAT (Wagner et al., 2020) permet de détecter les anomalies dans les séries temporelles en mesurant les variations des composantes cycliques. Tout d'abord, l'algorithme extrait deux sous-séries  $A$  et  $B$  de taille  $p$  et décalées de  $q$ . Une transformée de Fourier est appliquée sur les deux sous-séries pour extraire leur décomposition harmonique. Avec ces harmoniques, un nouveau modèle est calculé pour chaque sous-série selon la formule :

$$m(t) = \sum_{f=-z}^z |h_f| \cos(2\pi f \frac{t}{p} + \arg(h_f)), \quad z = 0 \dots \lceil \frac{p-1}{2} \rceil, \quad (1)$$

avec  $h_f \in \mathbb{C}$  l'harmonique correspondant à la fréquence  $f$  et  $z$  le nombre d'harmoniques à conserver dans le modèle. Les deux sous-séries  $A$  et  $B$  sont retardées de  $q$ . Par conséquent, il est nécessaire de synchroniser les modèles de  $A$  et de  $B$  en appliquant un décalage temporel

de  $-\frac{q}{p}2\pi$  au modèle de  $B$ . La distance euclidienne  $d_{L2}$  entre les deux modèles est ensuite calculée, elle reflète la variation de la composante cyclique de la série temporelle. Enfin, un seuil  $\tau$  est calculé par optimisation sur l’ensemble d’apprentissage (voir Wagner et al. (2020)) de telle sorte que  $d_{L2} > \tau$  signifie que la série est anormale.

### 3 Expérimentations et Résultats

Un jeu de données composé de séries temporelles issues de capteurs fixés sur 28 vaches Holstein est utilisé pour comparer les meilleurs algorithmes de classification des séries temporelles : BOSS, DTW, Hive-Cote FCN et ResNet (Fawaz et al., 2019) avec la méthode FBAT. Le jeu de données comprend les étiquettes *normale* et *anormale* pour décrire l’état d’une vache (e.g. saine ou malade).

Le tableau 1 présente pour chaque classifieur la précision et le rappel pour les classes positives (étiquette *anormale*) et négatives (étiquette *normale*). Le temps CPU (en heure) est également récupéré des expériences (temps d’apprentissage + temps de test). Pour les méthodes d’apprentissage profond, un mode GPU est disponible.

|                               |     | DTW  | Hive-Cote | BOSS        | FBAT        | FCN   | ResNet |
|-------------------------------|-----|------|-----------|-------------|-------------|-------|--------|
| <i>precision</i> <sub>-</sub> |     | 0.31 | 0.38      | <b>0.43</b> | 0.38        | 0.40  | 0.40   |
| <i>rappel</i> <sub>-</sub>    |     | 0.68 | 0.73      | 0.36        | <b>0.90</b> | 0.73  | 0.71   |
| <i>precision</i> <sub>+</sub> |     | 0.82 | 0.87      | 0.80        | <b>0.94</b> | 0.88  | 0.87   |
| <i>rappel</i> <sub>+</sub>    |     | 0.49 | 0.60      | <b>0.84</b> | 0.50        | 0.64  | 0.65   |
| temps (h)                     | CPU | 1h10 | 28h       | 0h38        | <b>0h06</b> | 19h28 | 16h36  |
|                               | GPU | -    | -         | -           | -           | 1h16  | 2h13   |

TAB. 1 – Résultats de tous les classifieurs.

Deux algorithmes se distinguent : BOSS qui a un *rappel*<sub>+</sub> élevé mais un *rappel*<sub>-</sub> assez bas et FBAT qui donne des résultats inverses. Autrement dit, BOSS détecte la plupart des séries anormales mais produit un nombre élevé de fausses alertes alors que FBAT, qui est le plus rapide des algorithmes, détecte peu de séries anormales mais produit très peu de fausses alertes. La détection de série anormale par FBAT peut être améliorée en prenant en compte la détection d’au moins un jour anormal dans une plage horaire fixée. Ces résultats semblent adéquats pour une utilisation à la ferme. Un test par les agriculteurs est nécessaire pour valider cette conclusion.

### Références

- Fawaz, H. I., G. Forestier, J. Weber, L. Idoumghar, et P.-A. Muller (2019). Deep learning for time series classification : a review. *Data Mining and Knowledge Discovery* 33(4), 917–963.
- Veissier, I., M.-M. Mialon, et K. H. Sloth (2017). Early modification of the circadian organization of cow activity in relation to disease or estrus. *Journal of dairy science* 100(5), 3969–3974.
- Wagner, N., V. Antoine, J. Koko, M.-M. Mialon, R. Lardy, et I. Veissier (2020). Comparison of machine learning methods to detect anomalies in the activity of dairy cows. In *International Symposium on Methodologies for Intelligent Systems*, pp. 342–351. Springer.