

ELSCP pour la détection d'anomalies dans les flux de données issues de l'agriculture

Juliet Chebet Moso*,*** Stéphane Cormier*
Cyril de Runz** Hacène Fouchal* John Mwangi Wandeto***

*CRESTIC EA 3804, Université de Reims Champagne-Ardenne, 51097 Reims, France
juliet-chebet.moso@etudiant.univ-reims.fr, stephane.cormier@univ-reims.fr,
hacene.fouchal@univ-reims.fr

**BDTLN, LIFAT, University of Tours, place Jean Jaurès, 41000, Blois, France
cyril.derunz@univ-tours.fr

***Computer Science, Dedan Kimathi University of Technology, Private Bag -10143,
Dedan Kimathi, Nyeri, Kenya
john.wandeto@dkut.ac.ke

1 Introduction

Cet article est une version poster de (Moso et al., 2021). L'agriculture bénéficie des développements récents de la technologie des capteurs, de la science des données et des techniques d'apprentissage automatique. Le but de cet article est de trouver des anomalies qui ont une influence sur l'efficacité de la récolte et peuvent être liées à la santé et à l'état des cultures pendant la récolte. En raison de la spécialisation des modèles pour différencier des caractéristiques des données, il est pertinent de combiner les diverses capacités des différentes approches de détection pour générer une décision de consensus (Zimek et al., 2014).

Nous considérons une approche orientée données avec l'objectif de détecter des anomalies à la volée en utilisant des approches de détection non supervisées pour la détection d'anomalies contextuelles locales. Nous proposons un détecteur d'anomalies par ensemble appelé Enhanced Locally Selective Combination in Parallel outlier ensembles (ELSCP). ELSCP est adapté au contexte du streaming en utilisant un système qui convertit les données en un flux et les transmet à ELSCP en utilisant un modèle de fenêtre de référence qui met en œuvre une technique de fenêtre glissante. Cette adaptation permet le traitement des données sous forme de flux, ce qui facilite l'évaluation de notre algorithme dans le contexte du streaming.

La méthode *Locally Selective Combination in Parallel outlier ensembles (LSCP)* (Zhao et al., 2019) construit une petite zone autour d'une instance de test en utilisant le consensus de ses voisins les plus proches. Nous proposons une extension de LSCP nommée ELSCP qui améliore la manière dont la région locale est extraite et la sélection de détecteurs adaptés. Dans ELSCP, pour chaque instance de test, la définition de la région locale est effectuée à l'aide d'une méthode de type ball tree k-plus proche voisin avec une métrique basée sur la distance de Harvesine. Une pseudo vérité terrain locale est construite à l'aide de la région locale, et la corrélation de Kendall est calculée entre les scores de valeurs aberrantes d'apprentissage