

Apprentissage d'ensemble d'opérateurs de projection orthogonale pour la détection de nouveauté

Fatma Hamdi*, Younès Bennani*

* LIPN-UMR 7030, Université Paris 13,
99, av. J-B Clément, 93430 Villetaneuse, France,
Prénom.Nom@lipn.univ-paris13.fr

Résumé

Dans ce papier, nous proposons une approche de détection de nouveauté fondée sur les opérateurs de projection orthogonale et l'idée de double bootstrap (bi- bootstrap). Notre approche appelée Random Subspace Novelty Detection Filter (RS-NDF), combine une technique de rééchantillonnage et

l'idée d'apprentissage d'ensemble. RS-NDF est un ensemble de filtres NDF (Novelty Detection Filter), induits à partir d'échantillons bootstrap des données d'apprentissage, en utilisant une sélection aléatoire des variables pour l'apprentissage des filtres. RS-NDF utilise donc un double bootstrap, c'est à dire un rééchantillonnage avec remise sur les observations et un rééchantillonnage sans remise sur les variables. La prédiction est faite par l'agrégation des prédictions de l'ensemble des filtres. RS-NDF présente généralement une importante amélioration des performances par rapport au modèle de base NDF unique. Grâce à son algorithme d'apprentissage en ligne, l'approche RS-NDF est également en mesure de suivre les changements dans les données au fil du temps. Plusieurs métriques de performance montrent que l'approche proposée est plus efficace, robuste et offre de meilleures performances pour la détection de nouveauté comparée aux autres techniques existantes.

Summary

In this paper we propose a novelty detection framework based on the orthogonal projection operators and the bootstrap idea. Our approach called Random Subspace Novelty Detection Filter ($RS - NDF$) combines the sampling technique and the ensemble idea. $RS - NDF$ is an ensemble of NDF , induced from bootstrap samples of the training data, using random feature selection in the NDF induction process. Prediction is made by aggregating the predictions of the ensemble. $RS - NDF$ generally exhibits a substantial performance improvement over the single NDF . Thanks to an online learning algorithm, the $RS - NDF$ approach is also able to track changes in data over time. The $RS - NDF$ method is compared to single NDF and other novelty detection methods with tenfold cross-validation experiments on publicly available datasets, where the method superiority is demonstrated. Performance metrics such as precision and recall, false positive rate and false negative rate, F-measure, AUC and G-

Détection de nouveauté

mean are computed. The proposed approach is shown to improve the prediction accuracy of the novelty detection, and have favorable performance compared to the existing algorithms.