

Méthode basée sur les ensembles approximatifs pour l'apprentissage incrémental en présence des données déséquilibrées

Sarra Bouzayane* ***, Inès Saad* **

*Université de Picardie Jules verne, Amiens
{sarra.bouzayane, ines.saad}@u-picardie.fr

**Ecole supérieure de commerce, Amiens

***Institut Supérieur d'Informatique et de Multimédia, Sfax

Résumé. Ce papier propose une méthode basée sur la théorie des ensembles approximatifs et dédiée à l'apprentissage supervisé incrémental dans un contexte de données déséquilibrées. Cette méthode consiste en trois phases : la construction d'une table de décision, l'inférence d'un ensemble de règles de décision et la classification de chaque action potentielle dans l'une des classes de décision prédéfinies. La méthode *MAI2P* est validée dans le contexte des MOOCs (*Massive Open Online Courses*).

1 Introduction

Lorsque les exemples d'apprentissage sont fournis de manière séquentielle, l'apprentissage incrémental pour une prise de décision s'avère une obligation (Greco et al., 2004). Généralement, la phase d'apprentissage est traitée par les techniques conventionnelles de l'apprentissage machine. Cependant, ces techniques demeurent sensibles au problème des données déséquilibrées qui résulte de la répartition inégale entre les instances des classes de décision. Cette inégalité affecte considérablement la qualité de la décision en particulier quand il s'agit de données massives. Ce problème peut, toutefois, être surmonté par l'approche DRSA (*Dominance-based Rough Set Approach*) (Greco et al., 2001) qui repose sur les préférences et l'expertise des décideurs humains pour la construction d'un ensemble d'apprentissage afin de garantir la répartition égale des instances sur l'ensemble de classes de décision.

Ce travail propose une méthode *MAI2P* (*Multicriteria Approach for the Incremental Periodic Prediction*) basée sur l'approche DRSA pour la classification multicritère incrémentale et périodique. La méthode *MAI2P* se compose de trois phases. La première vise à construire une table de décision et repose sur trois étapes : l'identification d'un ensemble d'apprentissage ; la construction d'une famille cohérente de critères pour la caractérisation des actions ; et la classification de chaque action d'apprentissage dans l'une des classes de décision. La deuxième phase est basée sur notre algorithme *DRSA-Incremental* (Bouzayane et Saad, 2017) pour l'inférence et la mise à jour de l'ensemble de règles de décision. La troisième consiste à la classification des "Actions potentielles", en utilisant les règles précédemment inférées. L'approche *MAI2P* est validée sur le contexte des MOOCs (*Massive Open Online Courses*).