



HAL
open science

Apprentissage de règles de classification à partir de données provenant de systèmes similaires

Si Tuc Ngo, Pierre Beuseroy, Edith Grall-Maës, André Smolarz, Gilles Mourot, Didier Maquin, José Ragot

► To cite this version:

Si Tuc Ngo, Pierre Beuseroy, Edith Grall-Maës, André Smolarz, Gilles Mourot, et al.. Apprentissage de règles de classification à partir de données provenant de systèmes similaires. 4ème Workshop du Groupement d'Intérêt Scientifique "Surveillance, Sûreté, Sécurité des Grands Systèmes", Oct 2011, Valenciennes, France. pp.CDROM. hal-00653174

HAL Id: hal-00653174

<https://hal.science/hal-00653174v1>

Submitted on 24 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**4^{ème} Workshop du Groupement d'Intérêt Scientifique
«Surveillance, Sûreté, Sécurité des Grands Systèmes» (GIS-3SGS'11)
Valenciennes, 12-13 Octobre, 2011**

**T.Ngo¹, P.Beuseroy¹, E.Grall-Maës¹, A.Smolarz¹
G. Mouro², D. Maquin² & J. Ragot²**

¹Laboratoire de Modélisation et Sûreté des Systèmes (LM2S) STMR/ICD/UTT, UMR CNRS 6279
12 rue Marie Curie - B.P. 2060 - 10010, Troyes, France
{si_tuc.ngo, pierre.beuseroy, edith.grall, andre.smolarz}@utt.fr

²Centre de Recherche en Automatique de Nancy, CNRS UMR 7039 - Université Henri Poincaré, Nancy-I
2, avenue de la forêt de Haye, 54500 Vandœuvre-lès-Nancy, France
{xiyan.he, gilles.mouro, didier.maquin, jose.ragot}@ensem.inpl-nancy.fr

Titre : Apprentissage de règles de classification à partir de données provenant de systèmes similaires

Projet : PARDI

Résumé :

Il s'agit d'étudier une ou plusieurs méthodes de classification de données dans le but de définir un système de diagnostic ou de surveillance générique d'un parc de machines ou plus généralement de systèmes. Dans le cas considéré, les données expertisées viennent de plusieurs éléments du parc. Elles constituent l'ensemble d'apprentissage de départ.

Dans l'étude menée, il est supposé que le parc est caractérisé par deux classes et que des données sont disponibles pour chaque élément et pour chaque classe. L'approche consiste à définir un classifieur par élément et à le fusionner avec un second classifieur commun à tous les systèmes (il s'agira ici du classifieur générique). La règle spécifique à chaque élément et celle commune à l'ensemble du parc sont déterminées de façon classique selon la méthode d'apprentissage choisie. Plusieurs méthodes de classification (discrimination linéaire et SVM linéaire) ont été considérées. Pour chaque élément, la fusion est obtenue par combinaison linéaire des sorties du classifieur spécifique et de la règle générique. L'importance de la règle spécifique par rapport à la règle générique dépend d'un coefficient qu'il convient de choisir.

Une étude expérimentale a été menée afin de comparer la règle composée aux règles spécifiques et à la règle générique et d'analyser la performance de classification en fonction du coefficient utilisé dans la règle composée ainsi que du nombre d'échantillons de la population d'apprentissage. Un cas simple a été considéré, où les deux modes de fonctionnement sont représentés par deux classes gaussiennes de même variance (dans ce cas, le discriminateur optimal est linéaire). Les variances et moyennes des classes changent légèrement pour chaque élément.

Les travaux réalisés permettent de montrer l'intérêt de la fusion et l'influence du nombre d'échantillons par classe sur les poids relatifs des règles spécifiques par rapport à la règle générique. De très nombreuses questions restent ouvertes et sont en cours d'investigation (détermination de la règle composée optimale, détermination de sous-groupes d'éléments...).

