HAL open science

Apprentissage de règles pour la détection de fraudes

Marine Collery, Philippe Bonnard, François Fages

To cite this version:

Marine Collery, Philippe Bonnard, François Fages. Apprentissage de règles pour la détection de fraudes. RJCIA 2021 - Dix-neuvièmes Rencontres des Jeunes Chercheurs en Intelligence Artificielle, Jul 2021, Bordeaux / Virtual, France. pp.77-78. hal-03298747

HAL Id: hal-03298747

https://hal.science/hal-03298747

Submitted on 23 Jul 2021

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.

Apprentissage de règles pour la détection de fraudes

M. Collery[1, 2], P. Bonnard[1], F. Fages[2]

[1] IBM France Lab
[2] Inria Saclay Ile-de-France

marine.collery@ibm.com

Résumé

Dans le domaine bancaire, la détection de fraudes doit être explicable et le modèle de décision compréhensible. Cela justifie un apprentissage direct de règles. CN2 et RIPPER sont comparés sur différents datasets.

Mots-clés

Apprentissage de règles, détection de fraudes, explicabilité, règles métier.

Abstract

In financial environment, fraud detection must be explainable and the decision model understandable. This justifies a rule-based learning approach. CN2 and RIPPER are compared on different datasets.

Keywords

Rule learning, fraud detection, explainability, business rules.

1 Introduction

Les fraudes bancaires ont aujourd'hui un impact financier important et nécessitent d'être détectées au plus vite. Ces difficultés s'accompagnent de complexités additionnelles dues au caractère temporel et très déséquilibré des données ainsi qu'à l'évolution constante de patterns ou motifs de fraudes. De plus, dans ce domaine particulier non seulement les décisions doivent pouvoir être expliquées, mais le modèle derrière la prise de décision doit être compréhensible. Les modèles d'apprentissage dits "black boxes" même expliqués à posteriori, ne peuvent donc être envisagés. C'est pourquoi, l'utilisation d'un langage symbolique, via un apprentissage supervisé de règles de décision est privilégiée.

$$\text{if } oldbalanceorigin - amount - newbalanceorigin = 0 \text{ and } newbalanceorigin = 0 \text{ then } fraud = true \quad (1)$$

L'objectif de ce projet est d'induire un ensemble de règles métier dites *business rules* (exemple Eq. 1), à partir de données labellisées (exemple partiel de données inspiré de *Synthetic Financial Datasets For Fraud Detection [1]* (SFD) ci-dessous) et de modèles de machine learning. Ceci permettra d'avancer dans la détection et prédiction de fraudes dans le contexte challenging de la finance et du monde bancaire où une complète transparence est nécessaire.

amount	type	old balance origin	new balance origin	fraud (label)
5324	payment	5324	0	true
695	cashout	6870	6175	false

2 Détection de fraudes bancaires

La détection de fraudes est le problème d'apprentissage qui vise à classifier automatiquement des événements comme frauduleux ou honnêtes. Dans le cadre d'un travail de recherche, la limitation principale réside dans le fait qu'il n'y a pas de données réelles disponibles publiquement pour étudier la détection de fraudes [12]. Dans le cadre d'une application réelle, ce sont les contraintes opérationnelles qui prennent le dessus. Comme relevé par [12], dans ce cas, le choix de la technique d'apprentissage dépend plus de contraintes techniques issues d'obligations opérationnelles et computationelles que de contraintes techniques liées aux données.

3 Induction de règles

Comme expliqué par J. Fürnkranz dans [9], il existe deux familles principales de méthodes pour induire des ensembles de règles : l'extraction de règles à partir d'arbre de décision (CART [2], C4.5 [13]) et le recouvrement séquentiel, c'est-à-dire l'apprentissage de règles à partir de données (CN2 [3], RIPPER [4]). Des extensions et de nouvelles approches basées sur ces modèles ont été proposés depuis. Nous pouvons mentionner FURIA, extension de RIPPER, qui construit des conditions floues ("fuzzy") apportant de la flexibilité aux règles, au détriment d'une augmentation du temps d'apprentissage [10]. Sood et al. présente en 2020, BiLevCSS, une approche basée sur deux phases d'apprentissage rassemblant règles d'association et RIPPER qui surpasse en terme de justesse (*accuracy*) les modèles existants [14]. Les algorithmes sur lesquels nous nous concentrons sont CN2 et RIPPER.

CN2 [3]. Cet algorithme a pour idée générale de créer une nouvelle règle tant que les données d'entraînement ne sont pas toutes couvertes par les règles trouvées.

RIPPER [4]. Cet algorithme dont la particularité est de proposer une étape de post-processing des règles, est toujours aujourd'hui l'état de l'art pour l'apprentissage de règles génériques. Il présente certaines limites en ce qui concerne la complexité du langage des règles générées. Plus

le nombre de conditions possibles est élevé, plus l'apprentissage sera long. La génération de conditions comportant des agrégats ou des structures plus complexes est donc compliquée dans un temps raisonnable.

4 Expériences et résultats

Framework. Pour mener à bien ce projet, un environnement de travail spécifique basé sur des conteneurs est mis en place. Il comprend des modules distincts permettant :

- le tracking des expériences (MLFlow)
- l'apprentissage de modèles (Orange [7], scikit-learn [11], module personnalisés etc),
- l'utilisation d'un langage de règle commun (module R2L implémenté pour le projet)
- une utilisation interactive (Jupyter Notebook)

Données. Avant toutes expérimentations sur des données réelles, un dataset open-source et fictif de données bancaires intitulé : *Synthetic Financial Datasets For Fraud Detection* (SFD) [1] est choisi. Ce dataset comprend 6 millions d'observations dont 8213 transactions frauduleuses ($\approx 1\%$). Découpage training/testing choisi : 10/90%. Pour valider nos résultats sur des datasets de références, 9 datasets complémentaires disponibles dans la base de données UCI [8] sont sélectionnés (*iris*, *adult*, *wine*, *heart disease cleveland* (Robert Detrano, M.D., Ph.D.), *breast cancer wisconsin*, *car evaluation*, *abalone*, *forest fires* [6], *wine quality (red & white)* [5]). Seuls les datasets adaptés pour la classification sont considérés dans la suite de ce document. Découpage training/testing choisi : 67/33%. Les premières expériences consistent à observer quelles sont les performances des algorithmes de l'état de l'art sur ces datasets.

Pre-processing. Des premiers tests sont réalisés sans et avec une étape de pré-processing manuel des données pour SFD. Celle-ci se révèle être indispensable, un module automatique de *feature generation* et *selection* est créé avec un accent particulier sur la conservation de la compréhensibilité des règles. Ce module est basé sur une analyse dimensionnelle des données et sur l'information mutuelle apportée avec la classe objectif.

Premiers résultats. De manière générale RIPPER dépasse CN2, comme le montrent les métriques présentées dans la Table ci-dessous. De plus, le nombre de règles générées est comparable pour la plupart des datasets, excepté pour *adult* où CN2 a généré 916 règles contre 35 pour RIPPER. Cependant une analyse plus poussée sur la qualité de l'ensemble des règles générées ainsi que sur la complexité des règles est nécessaire. Moyennant un coût computationnel, la qualité des résultats pourraient être améliorée en modifiant le protocole expérimental avec de la validation croisée. Par ailleurs, travailler à partir d'un dataset synthétique généré avec des règles connues (avec et sans ajout d'erreurs) permettrait d'évaluer plus précisément la qualité de l'apprentissage.

dataset	model	acc	bal_acc	f1	precision	recall
raw SFD	cn2	0.999	0.798	0.691	0.822	0.596
raw SFD	ripper	1	0.873	0.839	0.956	0.748
processed SFD	cn2	1	0.994	0.993	0.997	0.988
processed SFD	ripper	1	0.998	0.996	0.997	0.996
car evaluation	cn2	0.816	0.52	0.828	0.884	0.816
car evaluation	ripper	0.837	0.699	0.855	0.874	0.837
breast cancer wisconsin	cn2	0.65	0.65	0.788	1	0.65
breast cancer wisconsin	ripper	0.965	0.959	0.965	0.965	0.965
heart disease cleveland	cn2	0.515	0.318	0.525	0.567	0.515
heart disease cleveland	ripper	0.556	0.478	0.574	0.619	0.556
wine	cn2	0.889	0.914	0.889	0.902	0.889
wine	ripper	0.933	0.929	0.933	0.934	0.933
adult	cn2	0.782	0.747	0.839	0.934	0.782
adult	ripper	0.843	0.815	0.857	0.884	0.843
iris	cn2	0.98	0.979	0.98	0.981	0.98
iris	ripper	0.98	0.979	0.98	0.981	0.98

Remerciements

Ce travail a été soutenu par IBM France Lab / PSPC AIDA 2019-PSPC-09.

Références

[1] Synthetic Financial Datasets For Fraud Detection. Library Catalog : www.kaggle.com.

[2] L. Breiman et al. *Classification and Regression Trees*. Taylor & Francis, January 1984.

[3] P. Clark and T. Niblett. The CN2 Induction Algorithm. *Machine Learning*, 3(4) :261–283, March 1989.

[4] W. W. Cohen. Fast Effective Rule Induction. In *In Proceedings of the Twelfth International Conference on Machine Learning*, pages 115–123. Morgan Kaufmann, 1995.

[5] P. Cortez et al. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4) :547–553, November 2009.

[6] P. Cortez and A. Morais. A Data Mining Approach to Predict Forest Fires using Meteorological Data. page 12, 2007.

[7] J. Demšar et al. Orange : Data mining toolbox in python. *Journal of Machine Learning Research*, 14 :2349–2353, 2013.

[8] D. Dua et al. UCI machine learning repository, 2017.

[9] J. Fürnkranz et al. *Foundations of Rule Learning*. Springer Science & Business Media, November 2012.

[10] J. Hühn et al. FURIA : an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery*, 19(3) :293–319, December 2009.

[11] F. Pedregosa et al. Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.

[12] C. Phua et al. A Comprehensive Survey of Data Mining-based Fraud Detection Research. *Computers in Human Behavior*, 28(3) :1002–1013, May 2012. arXiv : 1009.6119.

[13] J. R. Quinlan. *C4.5 : Programs for Machine Learning*. Elsevier, June 1993.

[14] N. Sood et al. Bi-Level Associative Classifier Using Automatic Learning on Rules. In *Database and Expert Systems Applications*, pages 201–216. Springer International Publishing, 2020.