Engineering Applications of Artificial Intelligence 72 (2018) 21–29

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

A customized classification algorithm for credit card fraud detection

Alex G.C. de Sá *, Adriano C.M. Pereira, Gisele L. Pappa

Computer Science Department, Universidade Federal de Minas Gerais (UFMG), 31270-010, Belo Horizonte, Minas Gerais, Brazil

ARTICLE INFO

Keywords:
Credit card fraud
Bayesian network classifiers
Hyper-heuristic

ABSTRACT

This paper presents Fraud-BNC, a customized Bayesian Network Classifier (BNC) algorithm for a real credit card fraud detection problem. The task of creating Fraud-BNC was automatically performed by a Hyper-Heuristic Evolutionary Algorithm (HHEA), which organizes the knowledge about the BNC algorithms into a taxonomy and searches for the best combination of these components for a given dataset. Fraud-BNC was automatically generated using a dataset from PagSeguro, the most popular Brazilian online payment service, and tested together with two strategies for dealing with cost-sensitive classification. Results obtained were compared to seven other algorithms, and analyzed considering the data classification problem and the economic efficiency of the method. Fraud-BNC presented itself as the best algorithm to provide a good trade-off between both perspectives, improving the current company's economic efficiency in up to 72.64%.

1. Introduction

In 2016, a report by *CyberSource* (CyberSource, 2016) pointed out that the volume of fraudulent e-commerce credit card transactions (chargeback) in Latin America corresponds to 1.4% of the total net of the sector. Automatically identifying these transactions has many open challenges. Among them are the high volume of transactions that needs to be processed in almost real-time and the fact that frauds do not occur frequently, generating very imbalanced datasets. Furthermore, accepting a fraud as a legitimate transaction has a much higher cost than identifying a legitimate transaction as a fraud, as the seller economic losses are much higher in the first case, which generates chargeback.

There are different ways of modeling the credit card fraud detection problem, and among the most common approaches are those created to identify anomalies (Halvaiee and Akbari, 2014) and those based on classical methods for data classification (Hens and Tiwari, 2012). This paper focuses on the latter, and models the problem as a classification task, where a classifier is conceived to distinguish fraudulent from legitimate transactions.

In particular, we are interested in algorithm that generate *interpretable* models (classifiers), such as decision trees, classification rules or Bayesian network classifiers (Kotsiantis, 2007). This is because decision makers are more comfortable in accepting automatic decisions they can understand (Freitas, 2014). Although it is well-known that in some domains these methods present lower accuracy than black-box models such as Support Vector Machines, sacrificing accuracy to gain interpretability is a worth trade-off in alert systems.

There is a variety of classification algorithms that can generate interpretable models in the literature. According to the *No Free Lunch Theorem* (Wolpert and Macready, 1997), the choice of which of these algorithms is the best for a given dataset is still an open problem. The areas of meta-learning and hyper-heuristics have offered different solutions for automatically testing different types of algorithms (Pappa et al., 2014). While the literature of meta-learning has focused on selecting the best algorithm according to the characteristics of the target problem (Brazdil et al., 2008), hyper-heuristic methods have proposed different ways of generating customized algorithms for different datasets, which we considered more interesting for this work.

A hyper-heuristic is a high-level approach that, given a particular problem instance and a number of low-level heuristics, can select and apply an appropriate low-level heuristic at each decision point. Hyper-heuristics methods have been already conceived for building algorithms to solve specific classification problems (Pappa and Freitas, 2009; de Sá and Pappa, 2014). These methods help experts and practitioners in the following task: given a new classification dataset, which is the most suitable combination of the learning algorithms' components to solve this new problem? In this paper, we take advantage of one of these methods, and use Hyper-Heuristic Evolutionary Algorithm (HHEA) (de Sá and Pappa, 2014) to create a customized Bayesian Network Classifier (BNC) algorithm, named Fraud-BNC, specifically for detecting frauds in a dataset of interest.

We chose to work with BNC algorithms for fraud detection because they are robust statistical methods to classify data. They are based on

* Corresponding author.
E-mail addresses: alexgcsa@dcc.ufmg.br (A.G.C. de Sá), adrianoc@dcc.ufmg.br (A.C.M. Pereira), glpappa@dcc.ufmg.br (G.L. Pappa).

https://doi.org/10.1016/j.engappai.2018.03.011
Received 9 October 2017; Received in revised form 22 January 2018; Accepted 13 March 2018

the theoretical foundations of Bayesian networks (Bielza and Larrañaga, 2014) and produce a classification model that assumes cause–effect relations among all data attributes (including the class) (Cheng and Greiner, 1999). These relationships can be used to gain understanding about a problem domain as the output BNC model is represented by a directed acyclic graph (DAG). In the DAG, each node maps an attribute and edges define probabilistic dependencies among them. Each node is also associated with a conditional probability table, which represents the network parameters.

The literature presents several BNC algorithms (Bielza and Larrañaga, 2014; Sacha, 1999; Witten et al., 2011). Instead of choosing one of them, HHEA builds a customized BNC algorithm, which has the best combination of the essential modules (components) of the aforementioned algorithms for the dataset at hand. It is important to emphasize that HHEA produces a general BNC algorithm, even being specialized for a particular one.

Fraud-BNC was conceived by HHEA for a real-world credit card fraud detection problem. This problem is associated to a classification dataset, provided by *UOL PagSeguro*,[1] which is a popular online payment service in Brazil. The performances of Fraud-BNC and other baselines were evaluated using a classification metric (F_1) and a measure of the company economic loss, named economic efficiency. Besides, given the challenges of learning from class-imbalanced data (Sundarkumar and Ravi, 2015; Haixiang et al., 2016), we considered two strategies for dealing with cost-sensitive classification: instance reweighing and analysis of the class probability threshold.

The results showed that the best algorithm built in terms of F_1 is usually not the same that obtains the best values of economic efficiency. This happens because the latter is highly influenced by the monetary value of the transaction. Our analysis also showed that using Fraud-BNC with class probability threshold obtains the best results. Furthermore, as Fraud-BNC returns the probability of a transaction being a fraud, it can be used together with its monetary value of the transaction to help in the decision make process.

The main contributions of this paper are: (i) the generation of a customized BNC algorithm for a real-world credit card fraud detection dataset; (ii) the evaluation of how the algorithm performs in terms of both classification metrics and those used by finance specialists to evaluate fraud levels; (iii) the complete analysis of the customized BNC algorithm in terms of strategies to deal with imbalance data; (iv) the improvement of the current techniques currently used by the company to quantify fraud detection in PagSeguro in up to 72.64%; (v) the concept of how to use the produced BNC model in the auditing system to verify the inconsistent classifications.

The remainder of this paper is organized as follows. Section 2 presents related work on fraud detection modeled as a classification problem. Section 3 describes HHEA, the method used to automatically generate a customized BNC algorithm for the PagSeguro dataset, which is described in Section 4. The produced algorithm, Fraud-BNC, is presented in Section 5, followed by the definition of the metrics used to evaluated the algorithms, introduced in Section 6. Finally, Section 7 presents the experimental results, while conclusions and directions of future work are described in Section 8.

2. Related work

The problem of fraud detection has been extensively studied in the literature. This section reviews works that follow a classification approach to solve the problem. Among the methods already explored are artificial neural networks, decision trees, logistic regression, random forests, artificial immune systems, support vector machines (SVM) and hybrid methods (Chandola et al., 2009; Adewumi and Akinyelu, 2016; Alvarez and Petrovic, 2003; Lindqvist and Jonsson, 1997; Ngai et al., 2011; West and Bhattacharya, 2016), among others. Note that all these techniques follow a supervised learning approach, as they assume the existence of labeled data to generate these models.

Table 1
Summary of the six main characteristics of related works when compared to the customized algorithm Fraud-BNC.

Methods	Characteristics					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Fraud-BNC	Y	Y	Y	Y	Y	Y
Halvaiee and Akbari (2014)	Y	Y	N	Y	N	Y
Ravisankar et al. (2011)	Y	N	Y	N	N	N
Caldeira et al. (2012)	Y	Y	Y	N	N	Y
Sahin et al. (2013)	Y	Y	N	Y	Y	Y
Guo and Li (2008)	N	Y	N	N	Y	N
Fu et al. (2016)	Y	Y	Y	N	Y	N
Duman and Ozcelik (2011)	Y	Y	N	Y	Y	Y
Gadi et al. (2008)	Y	Y	N	Y	Y	Y
Vlasselaer et al. (2015)	Y	Y	N	N	N	N

Table 1 presents a comparison between a set of previously proposed methods and Fraud-BNC. Six main characteristics were analyzed: (i) if the method works with real-world data (if not, the work uses artificial data), (ii) whether the data reflects the real-world severe class imbalance, (iii) if feature selection is performed, (iv) if cost-sensitive methods are used to address the class imbalance problem or (v) if (under-)sampling techniques are used with this intention, and (vi) whether a financial analysis was taken into account when looking at the results. These six characteristics are referred in Table 1. The table indicates if the method in the row presents the characteristic defined in the column. '*Y*' indicates that the method has that characteristic, and '*N*' the opposite.

Note that most works deal with real-world unbalanced data, and use at least one strategy to deal with it. About half of the methods look beyond the results of classification, and the majority disregards any type of feature selection — although the features describing the data may differ significantly.

Concerning the learning techniques used by these methods, they encompass a large variety of algorithms. Ravisankar et al. (2011), for instance, employed six machine learning techniques, including SVM and logistic regression. Guo and Li (2008) proposed to combine confidence values, artificial neural network algorithms and receiver operating characteristic (ROC) curves for detecting credit card frauds. They performed undersampling to deal with class imbalance, resulting in a distribution of 100 legitimate transactions for each fraudulent. This is the only work that uses synthetic data.

Caldeira et al. (2012) also applied artificial neural networks and random forests to identify frauds in online transactions coming from the same data source we work with. Apart from other standard classification measures, they looked at the economic efficiency of the model, improving the results of the current company policy in 43%. However, they did not account for class imbalance or different classification costs for different classes. Fu et al. (2016), in turn, solved the problem with convolutional neural network (CNN). CNN was applied to a bank data to find a set of latent patterns for each transaction and identify frauds. The issue of data imbalance was tackled by a cost-based sampling method, which involved creating synthetic fraudulent samples from the real frauds.

Duman and Ozcelik (2011), on the other hand, developed a hybrid approach based on genetic algorithm (GA) and the scatter search (SS), named GASS, to take into consideration a classification cost function when dealing with fraud detection. GASS was applied to data from a major bank in Turkey, and used 20% of randomly chosen legitimate transactions for training due to time complexity.

Looking at works focusing on interpretable models, Sahin et al. (2013) is the only one *i* in this category, and developed a cost-sensitive decision tree algorithm. The authors self-referred their work as the pioneer at taking the misclassification costs into account while performing fraud classification. The authors used stratified sampling — i.e., they kept the class imbalance during the sampling process to help

[1] http://pagseguro.uol.com.br.