Expert Systems with Applications 39 (2012) 11583–11592

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Bagging k-dependence probabilistic networks: An alternative powerful fraud detection tool

Francisco Louzada [a,*], Anderson Ara [b]

[a] Universidade de São Paulo, Instituto de Matemática e Ciências da Computação, São Carlos, SP, Brazil
[b] Universidade Federal de São Carlos, Departamento de Estatística, São Carlos, SP, Brazil

ARTICLE INFO

Keywords:
Fraud detection
Probabilistic networks
Bayesian networks
Classification models
Bagging
Predictive performance

ABSTRACT

Fraud is a global problem that has required more attention due to an accentuated expansion of modern technology and communication. When statistical techniques are used to detect fraud, whether a fraud detection model is accurate enough in order to provide correct classification of the case as a fraudulent or legitimate is a critical factor. In this context, the concept of bootstrap aggregating (bagging) arises. The basic idea is to generate multiple classifiers by obtaining the predicted values from the adjusted models to several replicated datasets and then combining them into a single predictive classification in order to improve the classification accuracy. In this paper, for the first time, we aim to present a pioneer study of the performance of the discrete and continuous k-dependence probabilistic networks within the context of bagging predictors classification. Via a large simulation study and various real datasets, we discovered that the probabilistic networks are a strong modeling option with high predictive capacity and with a high increment using the bagging procedure when compared to traditional techniques.

1. Introduction

Fraud rates in various areas, such as financial, commercial, technological, internal accounting and others, have been growing in an accentuated manner with the expansion of modern technology and global communication (Bolton & Hand, 2002; Kou, Lu, Sirwongwattana, & Huang, 2004). According to the ESB (2011), there have been significant financial losses due to fraud in online business recently, which increased from US$5.2 billion in 2008 to US$8.6 billion in 2009.

An effective methodology for fraud detection may help companies to offer their consumers a safe and reliable online environment, which encourages loyalty to their services. Therefore, it is essential that prevention technologies and fraud detection methods are developed and updated continuously, preventing ways to circumvent such measures. In this sense, a fraud detection involves identifying fraud cases as quickly as possible one it has been perpetrated (Bolton & Hand, 2002).

There are statistical methods in the areas of Knowledge Discovery in Databases (KDD), Data Mining and Machine Learning with applicable and successful solutions in different areas of fraud crimes. These methods use a database of cases with information type fraudulent or legitimate to build a model that results in a score, usually called of suspected score, to predict new cases of fraud. Among these methods, we mention the traditional statistical classification methods, such as logistic regression, probit regression and discriminant analysis, more powerful tools, such as neural nets and rule-based algorithms (Bolton & Hand, 2002; Hand, 1981; Ngai, Y Hu, Chen, & Sun, 2011; Ripley, 1996).

Some papers have addressed the theory and the application of these tools. Wilson (2009) and Maranzato, Pereira, Naubert, and Lago (2010) used the logistic regression method as a tool to discriminate fraudulent actions from legitimate actions for insurance companies and e-commerce. Field and Hobson (1997) present a neural network based fraud management technique based on profiling techniques. Fawcett and Provost (1997) present a rule-based tool for fraud detection using a series of machine learning methods.

Alternatively, the method of probabilistic networks introduced by Pearl (1988) and disseminated in literature by the name of Bayesian networks, also known as causal networks, belief network or probabilistic dependence graphic, emerged in the 80's and has been applied in a wide variety of real-world activities (Bobbio, Portinale, Minichino, & Ciancarmela, 2001). The method is based on conditional probability distributions between variables and their causal relationship.

Geiger and Heckerman (1994), Sahami (1996), Cheng, Bell, and Liu (1997) and Friedman, Geiger, and Goldszmidt (1997) suggest classification models based on probabilistic network structures, such as naive Bayes networks, tree augmented networks and

* Corresponding author.
E-mail address: louzada@icmc.usp.br (F. Louzada).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.04.024

k-dependence probabilistic networks (KDB). Probabilistic networks can be seen as particular structures and are also known as Bayesian classifiers.

Fraud detection with Bayesian networks has been presented in Ezawa (1995, 1996, 1996), the authors use a Bayesian network as a normative expert system. They focus on the unbalanced ratio of fraud cases to legitimate cases with different misclassification costs.

A critical factor is whether a fraud detection model is accurate enough in order to provide correct classification of a case as a fraudulent or legitimate, since fraud detection tools with largest predictive capability are always required in practice. In this context, recently, some authors have performed small comparison amongst the mentioned techniques as well as proposed combination of some of the mentioned techniques. Shen, Tong, and Deng (2007) compared logistic regression, neural networks and regression trees. Their results show that the proposed model of neural networks and logistic regression approaches outperform decision tree in solving the problem under investigation. Taniguchi, Haft, HollmTn, and Tresp (1998) applied the techniques neural networks, Gaussian mixture and Bayesian networks, reaching similar results and suggests a combination of three methods. Widder, Ammon, Schaeffer, and Wolff (2008) work with a combination of discriminant analysis and neural networks that running successfully for a small set of events with two relevant attributes (fraudulent/legitime). Bhowmik (2011) applied the techniques Bayesian networks and decision tree, reaching a very similar results.

In this paper, we present a pioneer investigation on a combining of k-dependence Bayesian network, which generalizes the naive Bayes and tree augmented networks, by considering the bootstrap aggregating (bagging) procedure discussed by Breiman (1996). The basic idea is to generate multiple classifiers by obtaining the predicted values from the adjusted models to several replicated datasets and then combining them into a single predictive classification in order to improve the classification accuracy (Optiz & Maclin, 1999). Via a large simulation study and various real datasets, we discovered that the probabilistic networks are a strong modeling option with high predictive capacity and with a high increment using the bagging procedure when compared to traditional techniques, such as logistic regression, probit regression, neural networks and discriminant analysis.

The paper is organized as follows. In Section 2 we present the bagging probabilistic networks methodology. In Section 3 the metrics used to compare the models are displayed. Sections 4 shows results of the comparison of these methodologies via a large simulation study. The proposed methodology is illustrated in various real fraud detection datasets in Section 5, where we also compare it with some usual approaches generally considered as fraud detection models. The paper ends with Section 6, which displays some final comments.

2. Bagging probabilistic networks

According to Neapolitan (2004), the technique of probabilistic networks emerged within a context of large numbers of variables where the interest to verify the non-direct probabilistic influence of a variable upon another emerged consequently.

The probabilistic networks theory is built considering directed graphs, connected and acyclic, often referred to by DAG (directed acyclic graph).

In general, the discrete case was considered for this paper, the modeling structure based on the fact that X follows a multinomial distribution and, as for the continuous case, that X follows a normal distribution. Hence, a probabilistic network is determined by the trio (ξ,θ,X), where ξ is a DAG structure and θ is a set of specific parameters for conditional probabilities distributions concerning a X set of random variables purely discrete or purely continuous.

2.1. Simple probabilistic network

The construction of a simple probabilistic network is based upon the calculation of the a posteriori probability distribution $P(Y|X)$, where $Y=(y_1,y_2,\ldots,y_k)$ is a random variable to be classified featuring k categories, and $X=(X_1,X_2,\ldots X_p)$ is a set of p explanatory variables.

For calculating the conditional probability $P(Y|X)$, this method assumes probabilistic independence between the explanatory variables, given the classification variable, facilitating the application of the method computationally.

As for the case where X is a set of purely discrete explanatory variables, $P(Y|X)$ is given by,

$$P(Y=y_k|x_1,x_2,\ldots,x_p)=\frac{P(Y=y_k)\prod_{i=1}^{p}P(x_i|Y=y_k)}{\sum_j P(Y=y_j)\prod_{i=1}^{p}P(x_i|Y=y_j)} \quad (1)$$

When X is a set of purely continuous explanatory variables and, supposedly, with normal distribution, $P(Y|X)$ is given by

$$P(Y=y_k|x_1,x_2,\ldots,x_p)=\frac{P(Y=y_k)\prod_{i=1}^{p}f(x_i|Y=y_k)}{\sum_j P(Y=y_j)\prod_{i=1}^{p}f(x_i|Y=y_j)} \quad (2)$$

where

$$f(x_i|Y=y_k)\sim N(\mu_{i|y_k},\sigma^2_{i|y_k})$$

being $\mu_{i|y_k}$ and $\sigma^2_{i|y_k}$ the average and the variance of the variable x_i conditioned to the category y_k, respectively.

The method is based upon the calculation of the probability of an observation belonging to each of the categories and then classifies the observation in the most plausible category. If the classification in focus is binary, the ROC curve may be use d (Zweig & Campbell, 1993) to infer upon it.

However, in most cases, the assumption of independence between the explanatory variables does not match the reality, that is, the method does not take into account the possible probabilistic dependence relation between the explanatory variables.

Thus, other probabilistic network structures must be used. A possible alternative is presented below.

2.2. K-dependence probabilistic network

This method, unlike the previous one, considers possible dependence relations between the explanatory variables. A k-dependence probabilistic network (KDB) is a simple probabilistic network which allows within its structure for each explanatory variable X_i to have a maximum of K parent explanatory variables, i.e., for each explanatory varible X_i,$parents(X_i)$ is a set with a maximum of K other explanatory variables for every $i=1,\ldots,p$.

Also note that K can vary from 0 to $p-1$, where p is the number of explanatory variables considered.

For KDB networks, the a posteriori probabilities in the purely discrete case are calculated by,

$$P(Y=y_k|x_1,x_2,\ldots,x_p)=\frac{P(Y=y_k)\prod_{i=1}^{p}P(x_i|parents(X_i),Y=y_k)}{\sum_j P(Y=y_j)\prod_{i=1}^{p}P(x_i|parents(X_i),Y=y_j)} \quad (3)$$

While in the purely continuous case, which admits a normal distribution for the explanatory variables, are calculated by (PTrez, Larrañaga, & Inza, 2006)

$$P(Y=y_k|x_1,x_2,\ldots,x_p)=\frac{P(Y=y_k)\prod_{i=1}^{p}f(x_i|parents(X_i),y_k)}{\sum_j P(Y=y_j)\prod_{i=1}^{p}f(x_i|parents(X_i),y_j)} \quad (4)$$

where

$$f(x_i|parents_i,y_k)\sim N\left(\mu_{i|parents_i,y_k},\sigma^2_{i|parents_i,y_k}\right)$$