# A Discrete Wavelet Transform Approach to Fraud Detection

Roberto Saia(✉)

Department of Mathematics and Computer Science,
University of Cagliari, Via Ospedale 72, 09124 Cagliari, Italy
roberto.saia@unica.it

**Abstract.** The exponential growth in the number of operations carried out in the e-commerce environment is directly related to the growth in the number of operations performed through credit cards. This happens because practically all commercial operators allow their customers to make their payments by using them. Such scenario leads toward an high level of risk related to the potential fraudulent activities that the fraudsters can perform by exploiting this powerful instrument of payment illegitimately. A large number of state-of-the-art approaches have been designed to address this problem, but they must face some common issues, the most important of them are the imbalanced distribution and the heterogeneity of data. This paper presents a novel fraud detection approach based on the Discrete Wavelet Transform, which is exploited in order to define an evaluation model able to address the aforementioned problems. Such objective is achieved by using only legitimate transactions in the model definition process, an operation made possible by the more stable data representation offered by the new domain. The performed experiments show that our approach performance is comparable to that of one of the best state-of-the-art approaches such as random forests, demonstrating how such proactive strategy is also able to face the cold-start problem.

**Keywords:** Business intelligence · Fraud detection · Pattern mining · Wavelet

## 1 Introduction

A study performed by *American Association of Fraud Examiners*[1] shows that the credit card frauds (i.e., purchases without authorization or counterfeits of credit cards) are the *10–15%* of all the fraud cases, for a financial value close to *75–80%*. Only in the USA, such frauds lead toward an estimated average loss per fraud case of *2* million of dollars, and for this reason in recent years there was an increase in the researchers' efforts, aimed to define effective techniques for the fraud detection. Literature presents several state-of-the-art techniques for this

[1] http://www.acfe.com.


Z. Yan et al. (Eds.): NSS 2017, LNCS 10394, pp. 464–474, 2017.
DOI: 10.1007/978-3-319-64701-2_34

task, but all of them have to face some common problems, e.g., the imbalanced distribution of data and the heterogeneity of the information that compose a transaction. Such scenario is worsened by the scarcity of information that usually characterizes a transaction, a problem that leads toward an overlapping of the classes of expense.

The core idea of the proposed approach is the adoption of a new evaluation model based on the data obtained by processing the transactions through a *Discrete Wavelet Transformation* ($DWT$) [1]. Considering that such process involves only the previous legitimate transactions, it operates proactively by facing the *cold-start* issue (i.e., scarcity or absence of fraudulent examples during the model definition), reducing also the problems related to the data heterogeneity, since the new model is less influenced by the data variations.

The scientific contributions given by this paper are as follows:

(i) definition of the *time series* to use as input in the $DWT$ process, in terms of sequence of values assumed by the features of a credit card transaction;
(ii) formalization of the process aimed to compare the $DWT$ output of a new transaction with those of the previous legitimate ones;
(iii) classification of the new transactions as *legitimate* or *fraudulent* through an algorithm based on the previous comparison process.

The paper is organized into several sections: Sect. 2 introduces the background and related work of the fraud detection scenario; Sect. 3 reports the formal notation adopted in this paper and defines the faced problem; Sect. 4 gives all details about our approach; Sect. 5 describes the experimental environment, the used datasets and metrics, the adopted strategy and competitor approach, ending with the presentation of the experimental results; the last Sect. 6 provides some concluding remarks and future work.

## 2 Background and Related Work

**Fraud Detection Techniques:** The strategy adopted by the fraud detection systems can be of two types: *supervised* or *unsupervised* [2]. By following the *supervised* strategy it uses the previous *fraudulent* and *non-fraudulent* transactions in order to define its evaluation model. This is a strategy that needs a set of examples related to both classes of transactions, and its effectiveness is usually restricted to the recognition of patterns present in the training set. By following the *unsupervised* strategy, the system analyzes the new transactions with the aim to detect anomalous values in their features, where as anomaly we mean a value outside the range of values assumed by the feature in the set of previous legitimate cases.

The *static approach* [3] represents the most common way to operate in order to detect fraudulent transactions related to a credit card activity. By following such approach, the data stream is divided into blocks of equal size and the model is trained by using only a limited number of initial and contiguous blocks. Differently from the *static approach*, the *updating approach* [4] updates its model

at each new block, performing this activity by using a certain number of latest and contiguous blocks. A *forgetting approach* [5] can be also followed, and in this case the model is updated when a new block appears, performing this operation by using all the previous fraudulent transactions, but only the legitimate transactions present in the last two blocks. The models defined on the basis of these approaches can be used individually or they can be aggregated in order to define a bigger model of evaluation. Some of the problems related to the aforementioned approaches are the inability to model the users behavior (*static approach*), the inability to manage small classes of data (*updating approach*), and the computational complexity (*forgetting approach*), plus the common issues described in the following.

**Open Problems:** A series of problems, reported below, make the work of researchers operating in this field harder.
*(i) Lack of public real-world datasets:* this happens for several reasons, the first of them being the restrictive policies adopted by commercial operators, aimed to not reveal information about their business, for privacy, competition, or legal issues [6].
*(ii) Non-adaptability:* caused by the inability of the evaluation models to classify the new transactions correctly, when these have patterns different to those used during the model training [7].
*(iii) Data heterogeneity:* this problem is related to the incompatibility between similar features resulting in the same data being represented differently in different datasets [8].
*(iv) Unbalanced distribution of data:* it is certainly the most important issue [9], which happens because the information available to train the evaluation models is usually composed by a large number of legitimate cases and a small number of fraudulent ones, resulting in a data configuration that reduces the effectiveness of the classification approaches.
*(v) Cold-start:* another problem is related to those scenarios where the data used for the evaluation model training does not contain enough information on the domain taken into account, leading toward the definition of unreliable models [10]. Basically, this happens when the data available for the model training does not contain representative examples of all classes of information.

**Proposed Approach:** The core idea of this work is to move the evaluation process from the canonical domain to a new domain by exploiting the *Discrete Wavelet Transformation* (*DWT*) [11]. In more detail, we use the *DWT* process in a *time series* data mining context, where a *time series* usually refers to a sequence of values acquired by measuring the variation in the time of a specific data type (i.e., temperature, amplitude, etc.).

The *DWT* process transforms a *time series* by exploiting a set of functions named *wavelets* [12], and in literature it is usually performed in order to reduce the data size or the data noise (e.g., in the image compression and filtering tasks). The *time-scale multiresolution* offered by the *DWT* allows us to observe the original *time series* from different points of view, each of them containing interesting information on the original data. The capability in the new domain to

observe the data by using multiple scales (multiple resolution levels) allows our approach to define a more stable and representative model of the transactions, with regard to the canonical state-of-the-art approaches.

In our approach we define *time series* as the sequence of values assumed by the features of a credit card transaction, *frequency* represents the number of occurrences of a value in a *time series* over a unit of time, and as *scale* we refer to the time interval that characterize a *time series*.

Formally, a *Continuous Wavelet Transform* (*CWT*) is defined as shown in Eq. 1, where $\psi(t)$ represents a continuous function in both the time and frequency domain (called *mother wavelet*) and the $*$ denoting the complex conjugate.

$$X_w(a,b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} x(t)\psi^* \left( \frac{t-b}{a} \right) dt \tag{1}$$

Given the impossibility to analyze the data by using all *wavelets* coefficients, it is usually acceptable to consider a discrete subset of the upper half-plane to be able to reconstruct the data from the corresponding *wavelets* coefficients The considered discrete subset of the half-plane are all the points $(a^m, na^m b)$, where $m, n \in Z$, and this allows us to define the so-called *child wavelets* as shown in Eq. 2.

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a^m}} \psi \left( \frac{t - nb}{a^m} \right) \tag{2}$$

The use of small scales (i.e., that corresponds to large frequencies, since the scale is given by the formula $\frac{1}{frequency}$) compress the data, giving us an overview of the involved information, while large scales (i.e., low frequencies) expand the data, offering a detailed analysis of the information. On the basis of the characteristics of the *wavelets* transformation, although it is possible to use many basis functions as *mother wavelet* (e.g., *Daubechies*, *Meyer*, *Symlets*, *Coiflets*, etc.), for the scope of our approach we decided to use one of the simplest and oldest *wavelets* formalization, the *Haar wavelet* [13]. It is shown in Eq. 3 and it allows us to measure the contrast directly from the responses of low and high frequency sub-bands.

$$\psi(t) = \begin{cases} 1, & 0 \leq t > \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1 \\ 0, & otherwise \end{cases} \tag{3}$$

**Competitor Approach:** Considering that the most effective fraud detection approaches in literature need both the fraudulent and legitimate examples to train their model, we have chosen not to compare our approach to many of them, limiting the comparison to only one of the most used and effective ones, being *Random Forests* [14]. Our intention is to demonstrate the capability of the proposed approach to define an effective evaluation model by using a single class of transactions, overcoming some well-known issues.

*Random Forests* represents one of the most effective state-of-the-art approaches, since in most of the cases reported in literature it outperforms the other ones in this particular field [15,16]. It works by following an ensemble

learning method for classification and regression based on the construction of a number of randomized decision trees during the training phase and the classification is inferred by averaging the obtained results.

## 3 Notation and Problem Definition

Given a set of classified transactions $T = \{t_1, t_2, \ldots, t_N\}$, and a set of features $V = \{v_1, v_2, \ldots, v_M\}$ that compose each $t \in T$, we denote as $T_+ = \{t_1, t_2, \ldots, t_K\}$ the subset of legitimate transactions (then $T_+ \subseteq T$), and as $T_- = \{t_1, t_2, \ldots, t_J\}$ the subset of fraudulent ones (then $T_- \subseteq T$). We also denote as $\hat{T} = \{\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_U\}$ a set of unevaluated transactions. It should be observed that a transaction only can belong to one class $c \in C$, where $C = \{legitimate, fraudulent\}$. Finally, we denote as $F = \{f_1, f_2, \ldots, f_X\}$ the output of the $DWT$ process.

Denoting as $\Xi$ the process of comparison between the $DWT$ output of the *time series* in the set $T_+$ (i.e., the sequence of feature values in the previous legitimate transactions) and the $DWT$ output of the *time series* related to the unevaluated transactions in the set $\hat{T}$ (processed one at a time), the objective of our approach is the classification of each transaction $\hat{t} \in \hat{T}$ as *legitimate* or *fraudulent*. Defining a function $Evaluation(\hat{t}, \Xi)$ that performs this operation based on our approach, returning a boolean value $\beta$ (*0 = misclassification, 1 = correct classification*) for each classification, we can formalize our objective function (Eq. 4) in terms of maximization of the results sum.

$$\max_{0 \leq \beta \leq |\hat{T}|} \beta = \sum_{u=1}^{|\hat{T}|} Evaluation(\hat{t}_u, \Xi) \tag{4}$$

## 4 Proposed Approach

**Step 1 of 3 - `Data Definition`:** A *time series* is a series of events acquired during a certain period of time, where each of these events is characterized by a value. The set composed by all the acquisitions refers to a single variable, since it contains data of the same type. In our approach we consider as *time series* ($ts$) the sequence of values assumed by the features $v \in V$ in the sets $T_+$ (previous legitimate transactions) and $\hat{T}$ (unevaluated transactions), as shown in Eq. 5.

$$T_+ = \begin{vmatrix} v_{1,1} & v_{1,2} & \ldots & v_{1,M} \\ v_{2,1} & v_{2,2} & \ldots & v_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ v_{K,1} & v_{K,2} & \ldots & v_{K,M} \end{vmatrix} \quad \hat{T} = \begin{vmatrix} v_{1,1} & v_{1,2} & \ldots & v_{1,M} \\ v_{2,1} & v_{2,2} & \ldots & v_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ v_{U,1} & v_{U,2} & \ldots & v_{U,M} \end{vmatrix}$$
$$ts(T_+) = (v_{1,1}, v_{1,2}, \ldots, v_{1,M}), (v_{2,1}, v_{2,2}, \ldots, v_{2,M}), \cdots, (v_{K,1}, v_{K,2}, \ldots, v_{K,M})$$
$$ts(\hat{T}) = (v_{1,1}, v_{1,2}, \ldots, v_{1,M}), (v_{2,1}, v_{2,2}, \ldots, v_{2,M}), \cdots, (v_{U,1}, v_{U,2}, \ldots, v_{U,M}) \tag{5}$$

**Step 2 of 3 - `Data Processing`:** The *time series* previously defined are here used as input in the $DWT$ process. Without going deeply into the formal properties of the *wavelet transform*, we want to exploit the following two:
*(i) Dimensionality reduction:* the $DWT$ process can reduce the *time series* data, since its orthonormal transformation reduces their dimensionality, providing a compact representation that preserves the original information in its coefficients. By exploiting this property a fraud detection system can reduce the computational complexity of the involved processes;
*(ii) Multiresolution analysis:* the $DWT$ process allows us to define separate *time series* on the basis of the original one, distributing the information in them in terms of *wavelet* coefficients. The orthonormal transformation carried out by $DWT$ preserves the original information, allowing us to return to the original data representation. A fraud detection system can exploit this property in order to detect rapid changes in the data under analysis, observing the *data series* under two different points of view, one approximated and one detailed. The first provides an overview on the data, while the second provides useful information for the data changing evaluation.

Our approach exploits both the aforementioned properties, transforming the *time series* through the *Haar wavelet* process. The approximation coefficients at level $\frac{N}{2}$ was preferred to a more precise one in order to define a more stable evaluation model, less influenced by the data heterogeneity.

**Step 3 of 3 - `Data Classification`:** a new transaction $\hat{t} \in \hat{T}$ is evaluated by comparing the output of the $DWT$ process applied on each *time series* extracted by the set $T_+$ (previous legitimate transactions) to the output of the same process applied on the *time series* of the transaction $\hat{t}$ to evaluate.

The comparison is performed in terms of *cosine similarity* between the output vectors (i.e. values in the set $F$), as shown in Eq. 6, where $\Delta$ is the similarity, $\alpha$ is a threshold experimentally defined, and $c$ is the resulting classification. We repeat this process for each transaction $t \in T_+$, evaluating the classification of the transaction $\hat{t}$ on the basis of the average of all the comparisons.

$$\Delta = Cosim(F(t), F(\hat{t})), \quad with\, c = \begin{cases} \Delta \geq \alpha, & \text{legitimate} \\ \Delta < \alpha, & \text{fraudulent} \end{cases} \tag{6}$$

The Algorithm 1 takes the past legitimate transactions in $T_+$ as input, the transaction $\hat{t}$ to evaluate, and the threshold $\alpha$, returning a boolean value that indicates the $\hat{t}$ classification (i.e., *true* = legitimate or *false* = fraudulent) as output.

**Algorithm 1.** *Transaction evaluation*

**Require:** $T_+$=Legitimate previous transactions, $\hat{t}$=Unevaluated transaction, $\alpha$=Threshold
**Ensure:** $\beta$=Classification of the transaction $\hat{t}$
1: **procedure** TRANSACTIONEVALUATION($T_+$, $\hat{t}$)
2: $ts1 \leftarrow getTimeseries(\hat{t})$
3: $sp1 \leftarrow getDWT(ts1)$
4: **for each** $t$ **in** $T_+$ **do**
5: $ts2 \leftarrow getTimeseries(t)$
6: $sp2 \leftarrow getDWT(ts2)$
7: $cos \leftarrow cos + getCosineSimilarity(sp1, sp2)$
8: **end for**
9: $avg \leftarrow \frac{cos}{|T_+|}$
10: **if** $avg > \alpha$ **then** $\beta \leftarrow true$ **else** $\beta \leftarrow false$
11: **return** $\beta$
12: **end procedure**

## 5 Experiments

### 5.1 Environment

The proposed approach was developed in Java, by using the *JWave*[2] library for the *Discrete Wavelet Transformation*. The competitor approach (i.e., *Random Forests*) and the metrics used for its evaluation have been implemented in *R*[3], by using *randomForest*, *DMwR*, and *ROCR* packages. For reproducibility reasons, the *R* function *set.seed()* has been used, and the *Random Forests* parameters were tuned by finding those that maximize the performance. Statistical differences between the results were calculated by the independent-samples *two-tailed Student's t-tests* ($p < 0.05$).

### 5.2 DataSet

The public real-world dataset used for the evaluation of the proposed approach is related to a series of credit card transactions made by European cardholders[4] in two days of September 2013, for a total of *492* frauds out of *284,807* transactions. It is an highly unbalanced dataset [17], since the fraudulent cases are only the *0.0017%* of all the transactions.

For confidentiality reasons all dataset fields have been made public in anonymized form, except the *time*, the *amount*, and the *classification* ones.

### 5.3 Metrics

**Cosine Similarity:** It measures the similarity (*Cosim*) between two non-zero vectors $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ in terms of cosine angle between them, as shown in the Eq. (7).

[2] https://github.com/cscheiblich/JWave/.
[3] https://www.r-project.org/.
[4] https://www.kaggle.com/dalpozz/creditcardfraud.

It allows us to evaluate the similarity between vectors of values returned by the $DWT$ processes.

$$Cosim(\boldsymbol{v_1}, \boldsymbol{v_2}) = cos(\boldsymbol{v_1}, \boldsymbol{v_2}) = \frac{\boldsymbol{v_1} \cdot \boldsymbol{v_2}}{\| \boldsymbol{v_1} \| \cdot \| \boldsymbol{v_2} \|} \tag{7}$$

**F-score:** It represents the weighted average of the *Precision* and *Recall* metrics, a largely used metric in the statistical analysis of binary classification that returns a value in a range $[0, 1]$, where 0 is the worst value and 1 the best one. More formally, given two sets $T^{(P)}$ and $T^{(R)}$, where $T^{(P)}$ denotes the set of performed classifications of transactions, and $T^{(R)}$ the set that contains the actual classifications of them, it is defined as shown in Eq. 8.

$$\begin{aligned} &F\text{-}score(T^{(P)}, T^{(R)}) = 2 \cdot \tfrac{Precision \cdot Recall}{Precision + Recal} \\ &\text{with} \\ &Precision(T^{(P)}, T^{(R)}) = \tfrac{|T^{(R)} \cap T^{(P)}|}{|T^{(P)}|}, \quad Recall(T^{(P)}, T^{(R)}) = \tfrac{|T^{(R)} \cap T^{(P)}|}{|T^{(R)}|} \end{aligned} \tag{8}$$

**AUC:** The *Area Under the Receiver Operating Characteristic* curve ($AUC$) is a performance measure used to evaluate the predictive power of a classification model. Its result is in a range $[0, 1]$, where 1 indicates the best performance. More formally, given the subsets of previous legitimate transactions $T_+$ and previous fraudulent ones $T_-$, its formalization is reported in the Eq. 9, where $\Theta$ indicates all possible comparisons between the transactions of the two subsets $T_+$ and $T_-$. The result is obtained by averaging over these comparisons.

$$\Theta(t_+, t_-) = \begin{cases} 1, & if\ t_+ > t_- \\ 0.5, & if\ t_+ = t_- \\ 0, & if\ t_+ < t_- \end{cases} \qquad AUC = \frac{1}{|T_+| \cdot |T_-|} \sum_{1}^{|T_+|} \sum_{1}^{|T_-|} \Theta(t_+, t_-) \tag{9}$$

### 5.4 Strategy

**Cross-validation:** In order to improve the reliability of the obtained results and reduce the impact of data dependency, the experiments followed a *k-fold cross-validation* criterion, with $k = 10$, where each dataset is divided in $k$ subsets, and each $k$ subset is used as test set, while the other $k - 1$ subsets are used as training set, and the final result is given by the average of all $k$ results.

**Threshold Tuning:** According to the Algorithm 1 we need to define the optimal value of the $\alpha$ parameter, since the classification process depends on it (Eq. 6). It is the average value of *cosine similarity* calculated between all the pairs of legitimate transactions in the set $T_+$ ($\alpha = 0.91$ in our case).

### 5.5 Competitor

The state-of-the-art approach chosen as our competitor is *Random Forests*. It was implemented in *R* language by using the *randomForest* and the *DMwR* packages. The *DMwR* package was used to face the class imbalance problem through the *Synthetic Minority Over-sampling Technique* (*SMOTE*) [18], a popular sampling method that creates new synthetic data by randomly interpolating pairs of nearest neighbors.

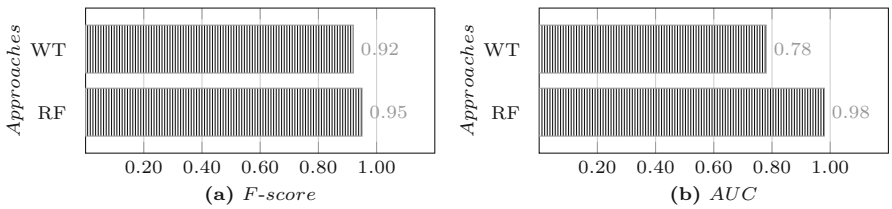
### 5.6 Results

Analyzing the experimental results, we can do the following considerations:

(i) the first set of experiments, which results are shown in Fig. 1a, was focused on the evaluation of our approach (denoted as $WT$) in terms of *F-score*. We can observe how it gets performance close to that of its competitor *Random Forests*, despite the adoption of a proactive strategy (i.e., not using previous fraudulent transactions during the model training), demonstrating its ability to define an effective model by exploiting only a class of transaction (i.e., the legitimate one);

(ii) the second set of experiments, which results are shown in Fig. 1b, was instead aimed to evaluate the performance of our approach in terms of *AUC*. This metric measures the predictive power of a classification model and the results indicate that our approach, also in this case, offers performance levels close to those of its competitor $RF$, while not using previous fraudulent cases to define its model.

(iii) summarizing all the results, the first consideration that arises is related to the capability of our approach to face the *data imbalance* and the *cold-start* problems, adopting a proactive strategy that only needs a transaction class for the model definition. The last but not least important consideration is that such proactivity allows a fraud detection system to operate without the need to have previous examples of fraudulent cases, with all the advantages that derive from it.




**(a)** *F-score* **(b)** *AUC*

**Fig. 1.** *F-score* and *AUC* performance

## 6 Conclusions and Future Work

Nowadays, credit cards represent an irreplaceable instrument of payment and such scenario obviously leads towards an increasing of the related fraud cases, making it necessary to design effective techniques for the fraud detection.

Instead of aiming to outperform the existing state-of-the-art approaches, with this paper we want to demonstrate that through a new data representation is possible to design a fraud detection system that operates without the need of previous fraudulent examples. The goal was to prove that our evaluation model,

defined by using a single class of transactions, is able to offer a level of performance similar to one of the best state-of-the-art approaches based on a model defined by using all classes of transactions (i.e., *Random Forests*), overcoming some important issues such as the *data imbalance* and the *cold-start* ones.

We can consider the obtained results to be very interesting, given that our competitor, in addition to use both classes of transactions to train its model, adopts a data balance mechanism (i.e., *SMOTE*).

For the aforementioned considerations, a future work will be focused on the definition of an hybrid fraud detection approach able to combine the advantages of the non-proactive state-of-the-art approaches with those of our proactive alternative.

**Acknowledgments.** This research is partially funded by *Regione Sardegna* under project *Next generation Open Mobile Apps Development* (*NOMAD*), *Pacchetti Integrati di Agevolazione* (*PIA*) *Industria Artigianato e Servizi* (2013).

## References

1. Chaovalit, P., Gangopadhyay, A., Karabatis, G., Chen, Z.: Discrete wavelet transform-based time series analysis and mining. ACM Comput. Surv. **43**(2), 6:1–6:37 (2011)
2. Bolton, R.J., Hand, D.J.: Statistical fraud detection: a review. Stat. Sci. **17**, 235–249 (2002)
3. Pozzolo, A.D., Caelen, O., Borgne, Y.L., Waterschoot, S., Bontempi, G.: Learned lessons in credit card fraud detection from a practitioner perspective. Expert Syst. Appl. **41**(10), 4915–4928 (2014)
4. Wang, H., Fan, W., Yu, P.S., Han, J.: Mining concept-drifting data streams using ensemble classifiers. In: Getoor, L., Senator, T.E., Domingos, P.M., Faloutsos, C. (eds.) Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, pp. 226–235. ACM, 24–27 August 2003
5. Gao, J., Fan, W., Han, J., Yu, P.S.: A general framework for mining concept-drifting data streams with skewed distributions. In: Proceedings of the Seventh SIAM International Conference on Data Mining, Minneapolis, Minnesota, USA, pp. 3–14. SIAM, 26–28 April 2007
6. Phua, C., Lee, V., Smith, K., Gayler, R.: A comprehensive survey of data mining-based fraud detection research (2010)
7. Sorournejad, S., Zojaji, Z., Atani, R.E., Monadjemi, A.H.: A survey of credit card fraud detection techniques: data and technique oriented perspective. CoRR abs/1611.06439 (2016)
8. Chatterjee, A., Segev, A.: Data manipulation in heterogeneous databases. ACM SIGMOD Rec. **20**(4), 64–68 (1991)
9. Japkowicz, N., Stephen, S.: The class imbalance problem: a systematic study. Intell. Data Anal. **6**(5), 429–449 (2002)
10. Donmez, P., Carbonell, J.G., Bennett, P.N.: Dual strategy active learning. In: Kok, J.N., Koronacki, J., Mantaras, R.L., Matwin, S., Mladenič, D., Skowron, A. (eds.) ECML 2007. LNCS, vol. 4701, pp. 116–127. Springer, Heidelberg (2007). doi:10.1007/978-3-540-74958-5_14

11. Chernick, M.R.: Wavelet methods for time series analysis. Technometrics **43**(4), 491 (2001)
12. Percival, D.B., Walden, A.T.: Wavelet Methods for Time Series Analysis, vol. 4. Cambridge University Press, Cambridge (2006)
13. Mallat, S.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**(7), 674–693 (1989)
14. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
15. Lessmann, S., Baesens, B., Seow, H., Thomas, L.C.: Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research. Eur. J. Oper. Res. **247**(1), 124–136 (2015)
16. Brown, I., Mues, C.: An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Syst. Appl. **39**(3), 3446–3453 (2012)
17. Dal Pozzolo, A., Caelen, O., Johnson, R.A., Bontempi, G.: Calibrating probability with undersampling for unbalanced classification. In: 2015 IEEE Symposium Series on Computational Intelligence, pp. 159–166. IEEE (2015)
18. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)