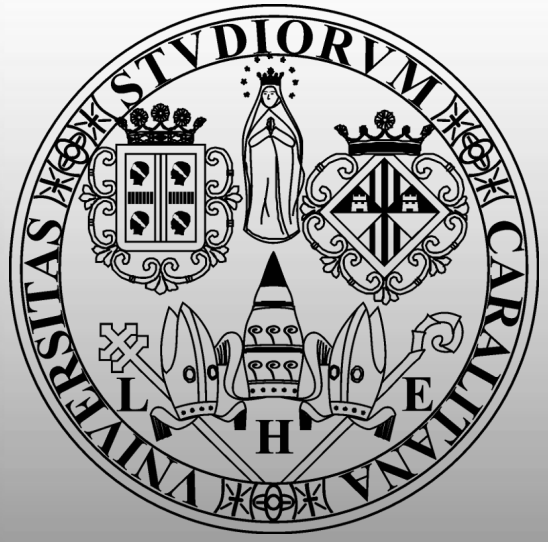


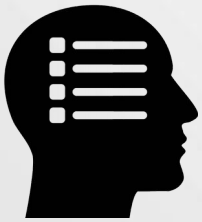
Proactivity or Retroactivity?
Evaluating the Benefits of Using Proactive Transformed-domain-based Techniques in Fraud Detection Tasks

Roberto Saia and **Salvatore Carta**
Department of Mathematics and Computer Science
University of Cagliari, Italy

IoTBDS 2018, 3nd International Conference on Internet of Things, Big Data and Security
High Quality Journal Forum - 19÷21 March 2018 – Funchal, Madeira - Portugal

Outline

- **Overview**
  Introduction
  State of the Art
  Limits
- **Intuition**
  Idea
  Formalization
  Advantages
- **Approach**
  Data Definition
  Data Evaluation
  Data Classification
- **Conclusions**
  Experimental Results
  Conclusions
  Future Work



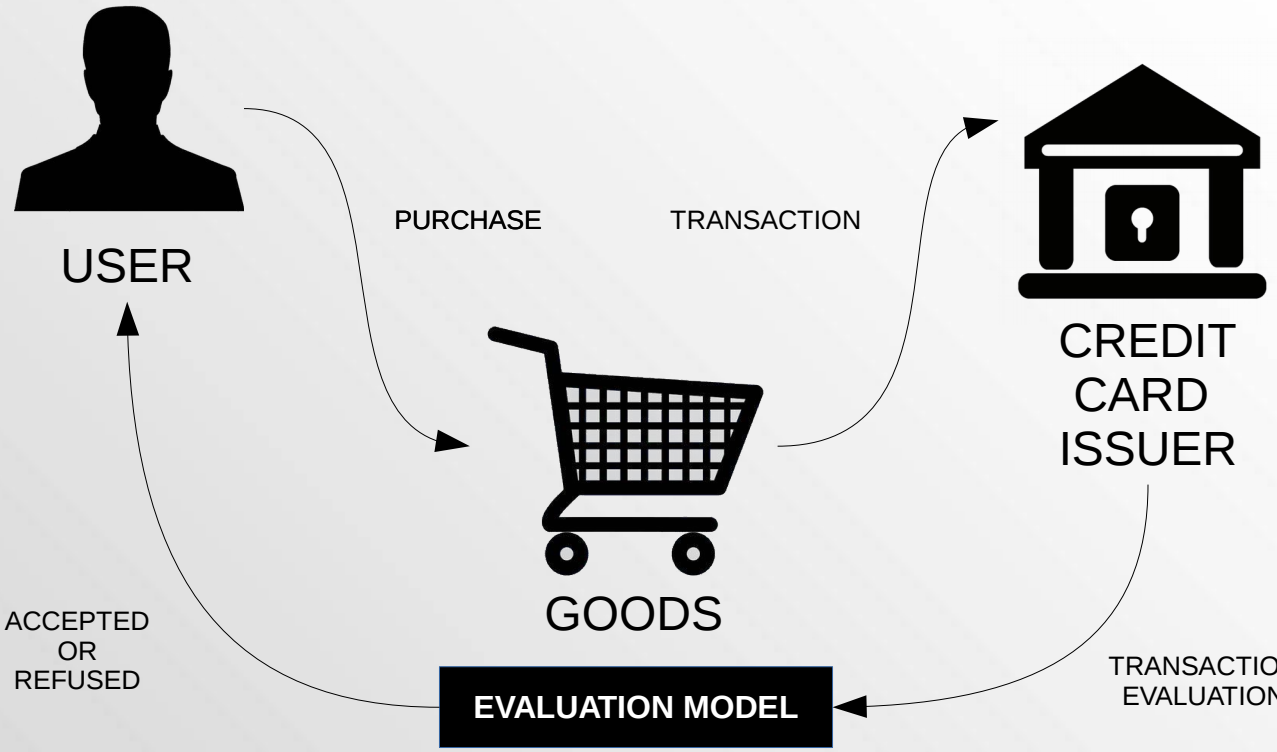
OVERVIEW

Introduction

A **Fraud Detection** task is aimed to classify a financial transaction as *legitimate* or *fraudulent*

Introduction

Such operation is performed by defining an **evaluation model** on the basis of the previous user transactions



State of the Art

A large number of **state-of-the-art techniques** have been designed to address this problem

Data Mining

Artificial Intelligence

Machine Learning

And so on . . .

State of the Art

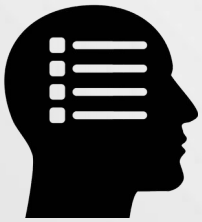
- Regardless of the adopted technique, the definition of effective *Fraud Detection* models represents a **hard challenge** for several factors
- The most important of which is the **Data Imbalance**

Limits

- **Data Imbalance** refers to a classification problem given by the fact that the classes are *not represented equally*
- Such configuration reduces the **effectiveness** of the state-of-the-art approaches

FRAUDULENT CASES

LEGITIMATE CASES



INTUITION

Idea

- Overcome the **imbalanced class distribution** issue by using a single class of data (*legitimate transactions*) during the model definition
- Perform this operation through a model able to well characterize the legitimate transactions, defining it in a **new domain**

We move the data analysis from the canonical domain to a new **frequency** or **time-frequency** domain, in order to obtain a **more stable** model

Formalization

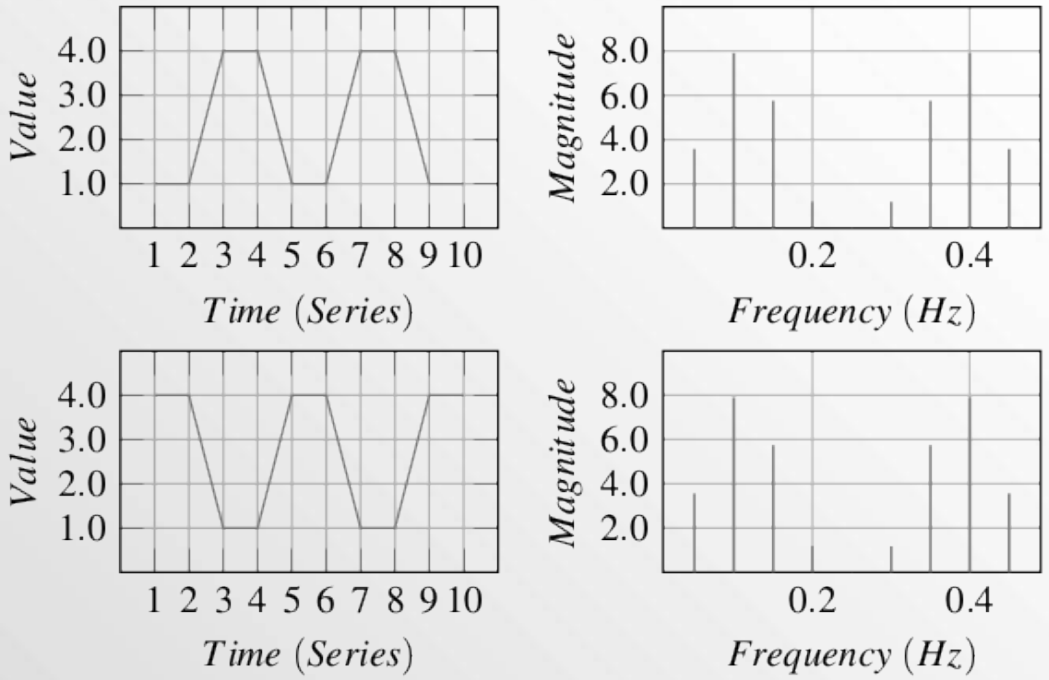
- We propose a **frequency-domain-based** or a **frequency-time-domain-based pattern mining** process able to evaluate a new transaction in a novel way
- It is made by exploiting the **Fourier** or the **Wavelet transformation**, applying it on the values assumed by the **features*** of a transaction
- We perform this by considering this series of **values** as a **time series**

* It contains several information about the credid card transaction, e.g., **place**, **amount**, **date**, etc.

Formalization

FOURIER TRANSFORM

- The first interesting property of the new data representation is the **phase invariance**

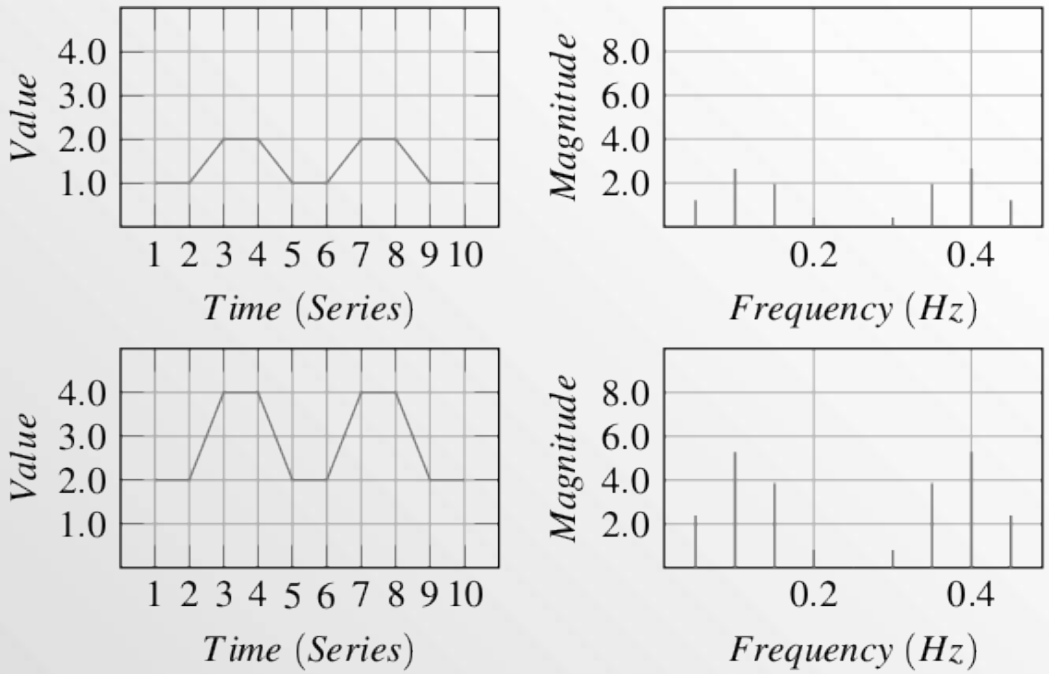


- This allows us to detect **specific patterns**, also when they shift along the transaction features

Formalization

FOURIER TRANSFORM

- The second interesting property of the new data representation is the **amplitude correlation**



- This grants that our model is able to **differentiate** same patterns with different feature values

Formalization

WAVELET TRANSFORM

- The first interesting property of the new data representation is the **dimensionality reduction**
- The DWT process can reduce the time-series data, offering a **compact representation** that preserves the original information
- This allows a system to reduce the **computational complexity** of the involved processes

Formalization

WAVELET TRANSFORM

- The second interesting property of the new data representation is the **multiresolution analysis**
- The DWT process allows us to define **several time series** on the basis of the original one, preserving the information
- A system can exploit this property to obtain two **different point of views** on data:
  - **approximated** (*data overview*)
  - **detailed** (*data changing*)

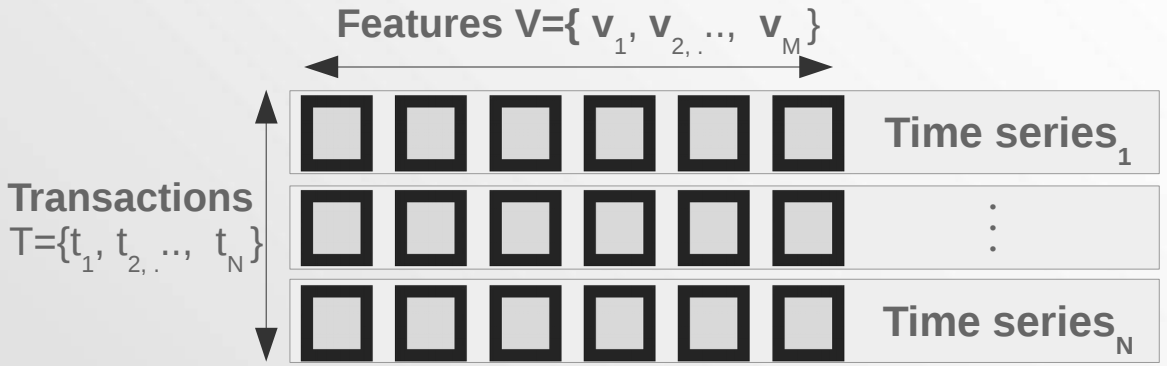
Advantages

- The adoption of a model based on the **frequency** or **time-frequency** pattern leads toward some advantages:
  - It **overcomes** the imbalance class distribution issue by using only a class of data (previous legitimate transactions)
  - The use of only *legitimate* cases, allows us to operate in a **proactive** way
  - It also overcomes the problems related to the **cold-start** issue

APPROACH

Data Definition

Definition of the **time series** in terms of values assumed by the features of a transaction



Each time series is processed by using a **Fourier Transform** or a **Wavelet Transform** process

Data Evaluation

- Comparison of the **frequency** or **time-frequency** patterns of two transactions
- It is performed in terms of **cosine similarity** between the output vectors of the *Fourier* or *Wavelet* processes

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

α is a threshold experimentally defined

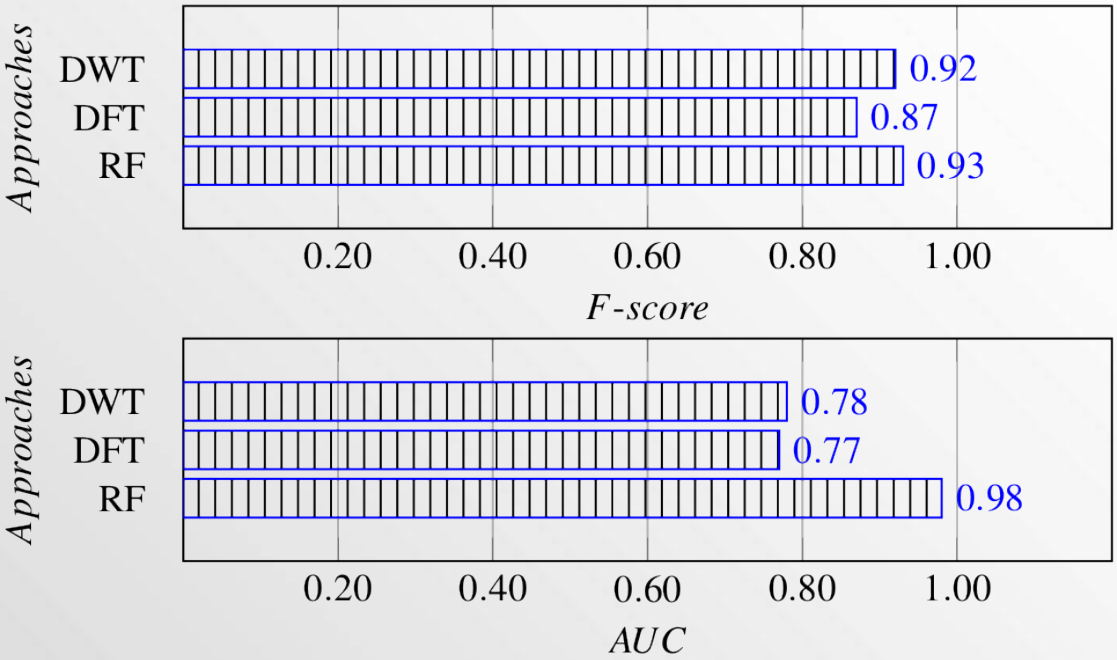
Data Classification

- Definition of an **algorithm** able to classify a new transaction as *legitimate* or *fraudulent*, on the basis of the *previous comparison process*
- *It takes into account the **cosine similarity** between the new transaction to evaluate and each previous legitimate transaction*
- *The **classification** of the new transaction is based on the average of these results*

CONCLUSIONS

Experimental Results

Our **DFT** and **DWT** strategies get performance close to that of our state-of-the-art competitor Random Forests (**RF**)



Considering that our approach exploits **only a class** of data in the model definition process

Conclusions

- Such **proactive** approach does not replace the existing **retroactive** state-of-the-art approaches
- It is aimed to face some well-known problems, such as the **cold-start** and the **data unbalance** ones
- It can be used to define **hybrid** fraud detection approaches able to operate in all real-world scenarios

Future Work

- A future work will be the definition and evaluation of a **hybrid fraud detection approach** that exploits both previous *legitimate* and *fraudulent* cases
- We also consider to evaluate our approach in **heterogeneous scenarios** that involve different types of financial data, such as those that characterize the **E-commerce** environment

Proactivity or Retroactivity?

Evaluating the Benefits of Using Proactive Transformed-domain-based Techniques in Fraud Detection Tasks

THANK YOU FOR YOUR ATTENTION

IoTBDS 2018, 3nd International Conference on Internet of Things, Big Data and Security
High Quality Journal Forum - 19÷21 March 2018 – Funchal, Madeira - Portugal