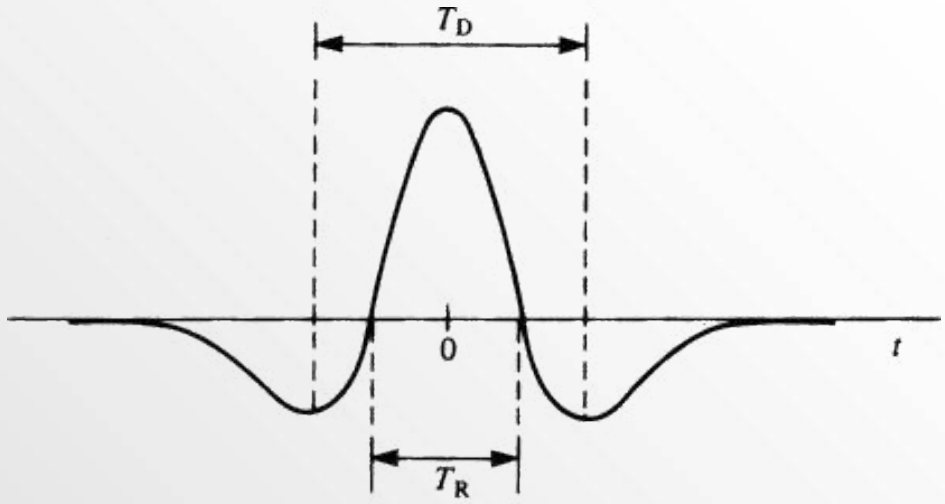


Exploiting Wavelet Analysis in Fraud Detection Process

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

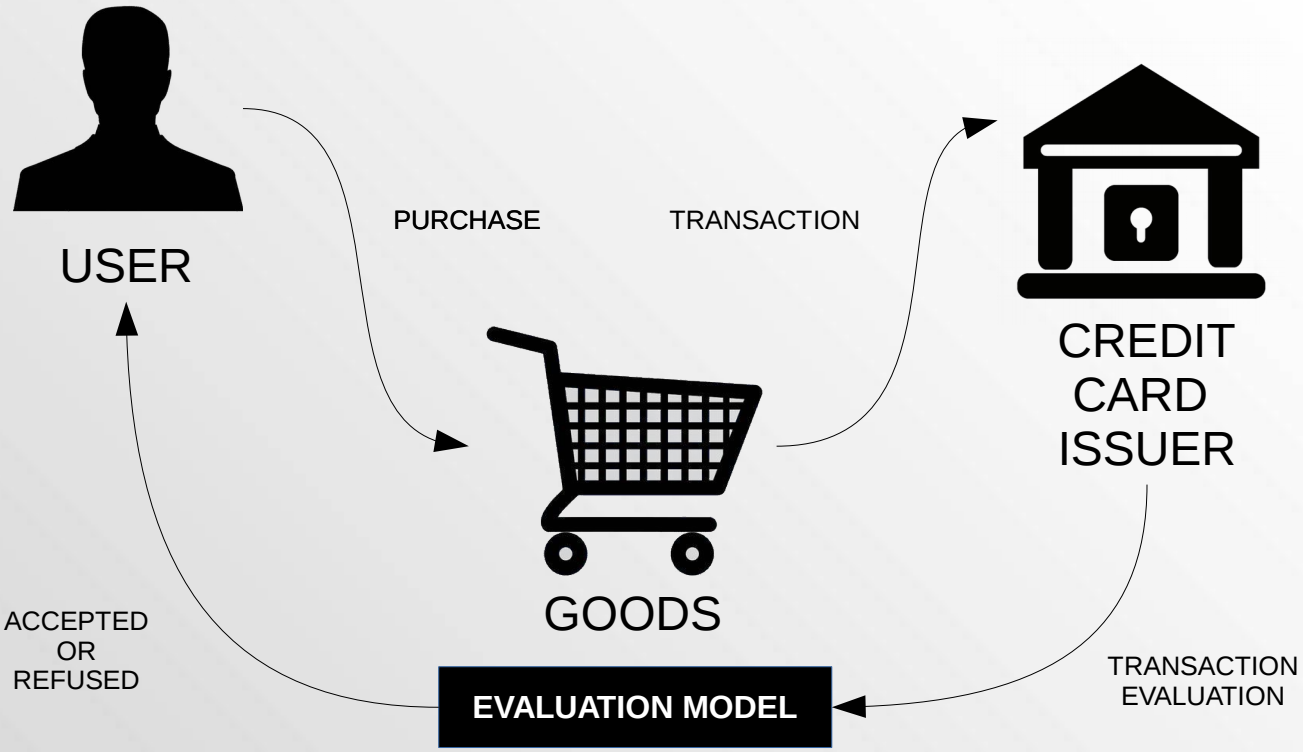
GARR 2018 Conference – Data (R)evolution: *Cybersecurity and AI Session*
03÷05 Ottobre 2018 – University of Cagliari, Italy

Introduction

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

Introduction

Such a 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 **Imbalanced class distribution**

LEGITIMATE CASES **vs** FRAUDULENT CASES

Idea

- Overcome the **imbalanced class distribution** issue by using a single class of data (*legitimate transactions*) during the model definition
- Define this model in a **new domain** that allows us to well characterize this class of data
- This goal is achieved thanks to the **different point of view** on data offered by the new domain

Formalization

- We propose a **transformed-domain-based pattern mining** process able to evaluate a new transaction in a novel way
- It is made by exploiting the **Discrete Wavelet Transformation**, applying it on the values assumed by the transaction **features***
- We perform this operation by considering these values as a **time series**

* They contain several information about a credit card transaction, e.g., **place**, **amount**, **date**, etc.

Formalization

- 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 Fraud Detection system to reduce the **computational complexity** of the involved processes

Formalization

- 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 Fraud Detection system can exploit this property to obtain two **different point of views** on data:
  - **approximated** (*data overview*)
  - **detailed** (*data changing*)

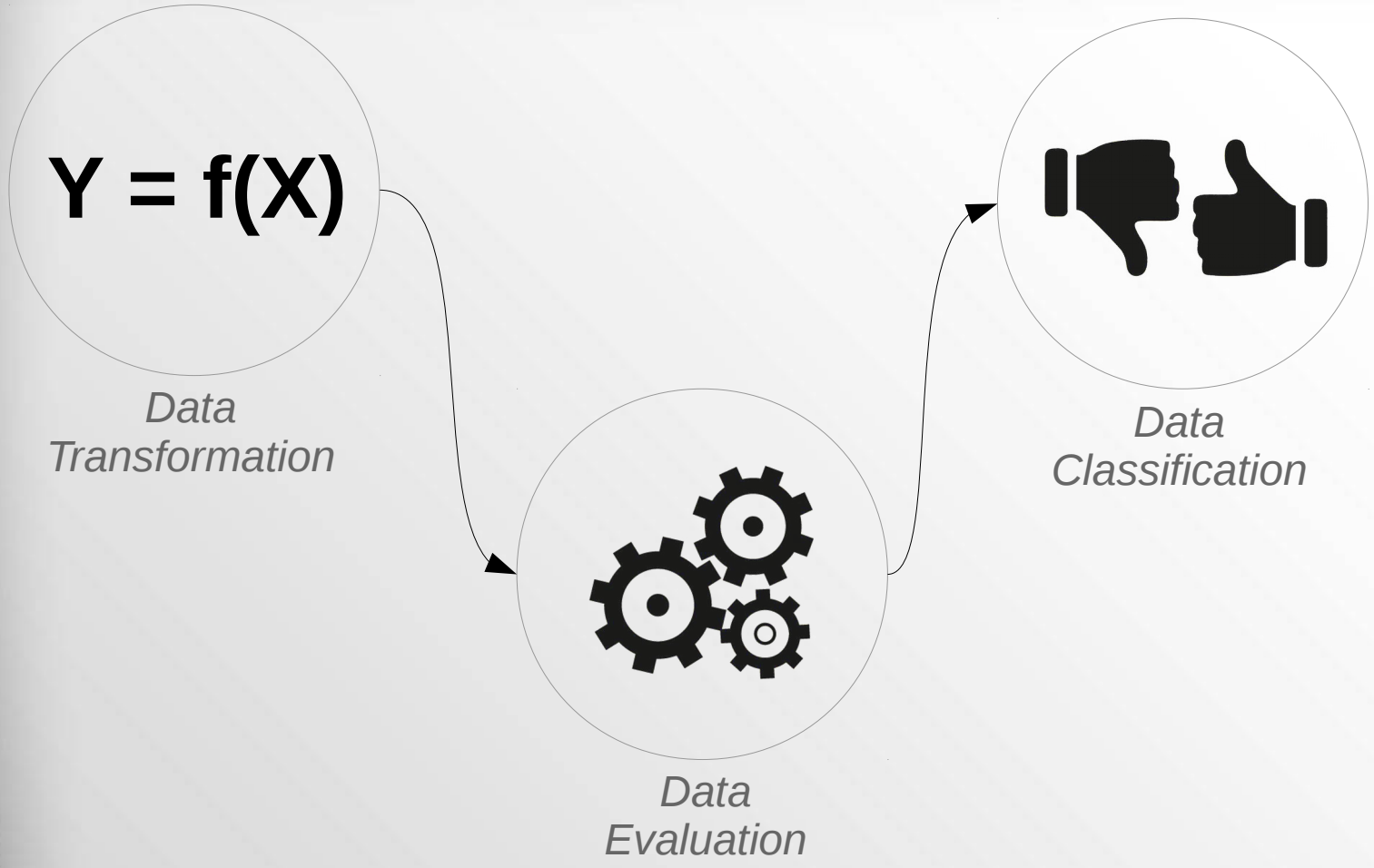
Formalization

- Our approach exploits **both** the aforementioned properties
- The original time series are transformed by using the **Haar wavelet**

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2}, \\ -1 & \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \qquad \varphi(t) = \begin{cases} 1 & 0 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases}$$

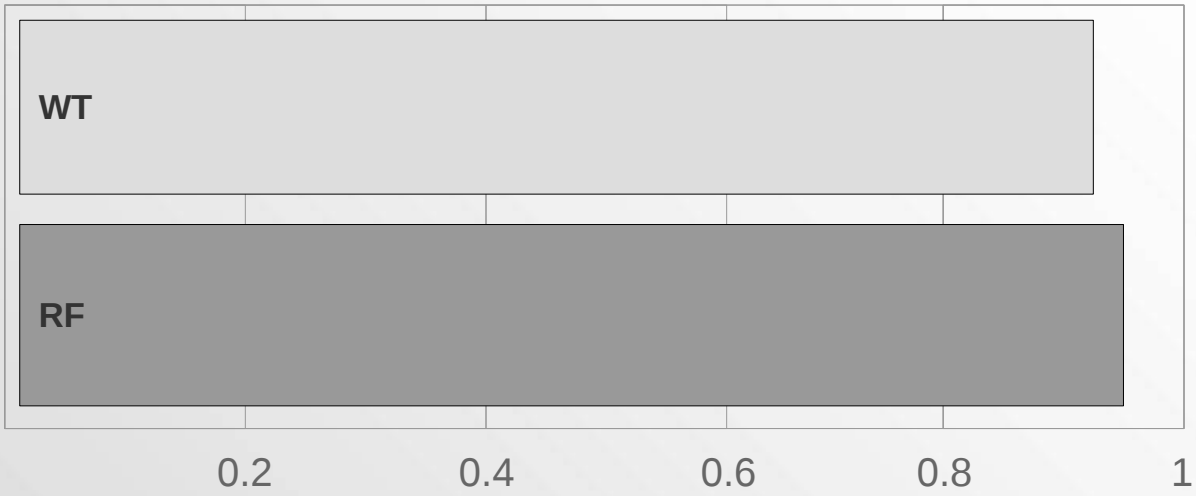
- It offers us an approximated point of view on data that allows us to **model** the considered class of information

Implementation



Experiments

- The general performance of our **WT** approach in terms of ***F-score*** is close to that of *Random Forests* (**RF**), which is one of the most performing state-of-the-art approach



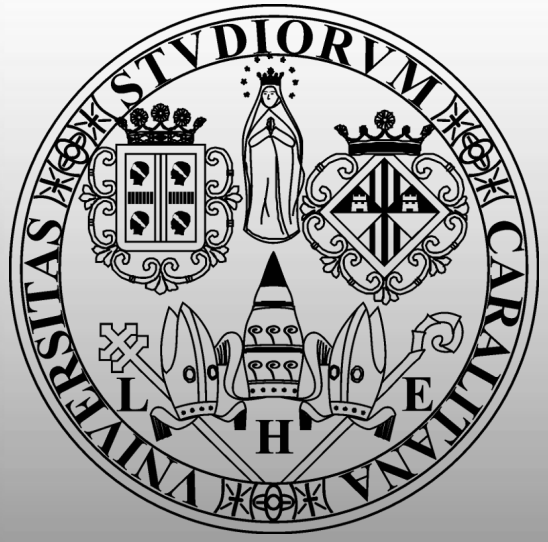
- In spite of the fact that we did **not use** any past default instance to train the evaluation model

Advantages

- The adoption of a model based on the **wavelet** gives us 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

Conclusions

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



THANK YOU FOR YOUR ATTENTION