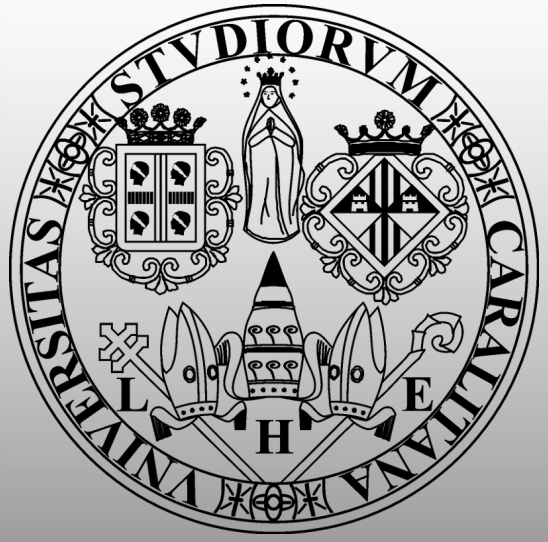


A Fourier Spectral Pattern Analysis to Design Credit Scoring Models

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

IML 2017, International Conference on Internet of Things and Machine Learning
17÷18 October 2017, Liverpool John Moores University – Liverpool, United Kingdom

Credit Scoring

- A ***Credit* Scoring** system is aimed to classify the new loan applications (from now on named as ***instances***) as *reliable* or *unreliable,* on the basis of the previous ones

State of the Art

- The state of the art offers **many approaches** designed to perform *Credit Scoring* tasks
- *However, the definition of effective evaluation models represents a **hard challenge** for several factors*
- The most important of them is the **data imbalance** in the information available for their definition

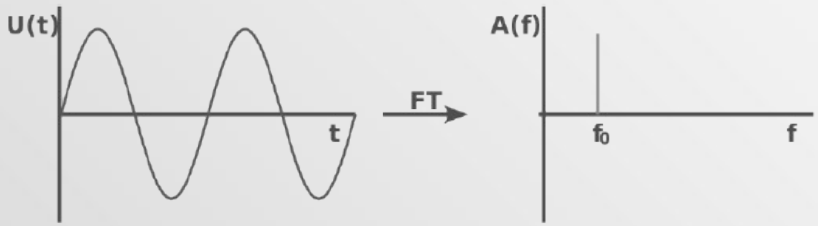
Open Problems

- The **data imbalance** exists because the number of previous un*reliable* instances (*non-refunded loans*) is usually **much smaller** than that of the *reliable* ones (*refunded loans*)

- This reduces the **effectiveness** of the *machine learning* techniques

Idea

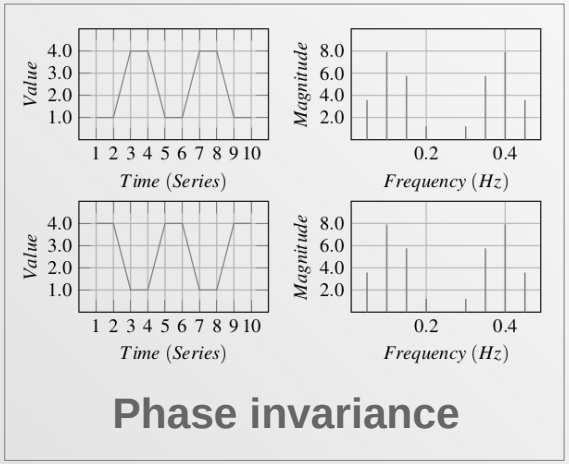
- Overcoming the **data imbalance** issue by using a single class of data (*reliable instances*) during the model definition
- Performing this operation through a model able to well characterize this *class of instances*, defining it in the **frequency domain**
- Exploiting the **Fourier transformation** to make this, applying it on the series of values that compose each instance (features*)

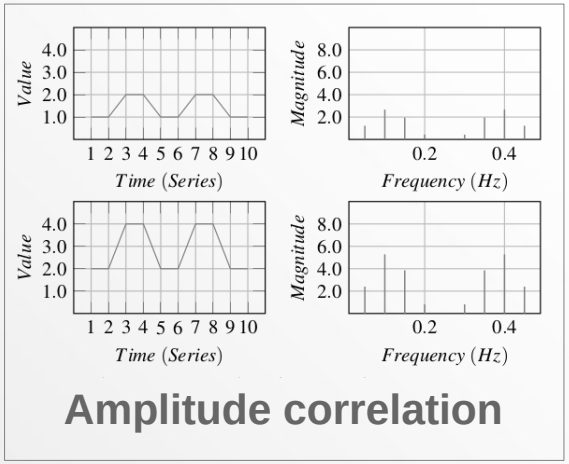


* They contain information about the instance, e.g., applicant age, job, loan purpose, etc.

Formalization

- In the new domain, our approach exploits two properties: **phase invariance** and **amplitude correlation**





- The first property allows it to detect **specific patterns**, also when they shift along the instance features
- The second property allows it to **differentiate** same patterns with different feature values

Implementation

1. Definition of the **time series** to use as input in the Fourier process, in terms of values assumed by instance features

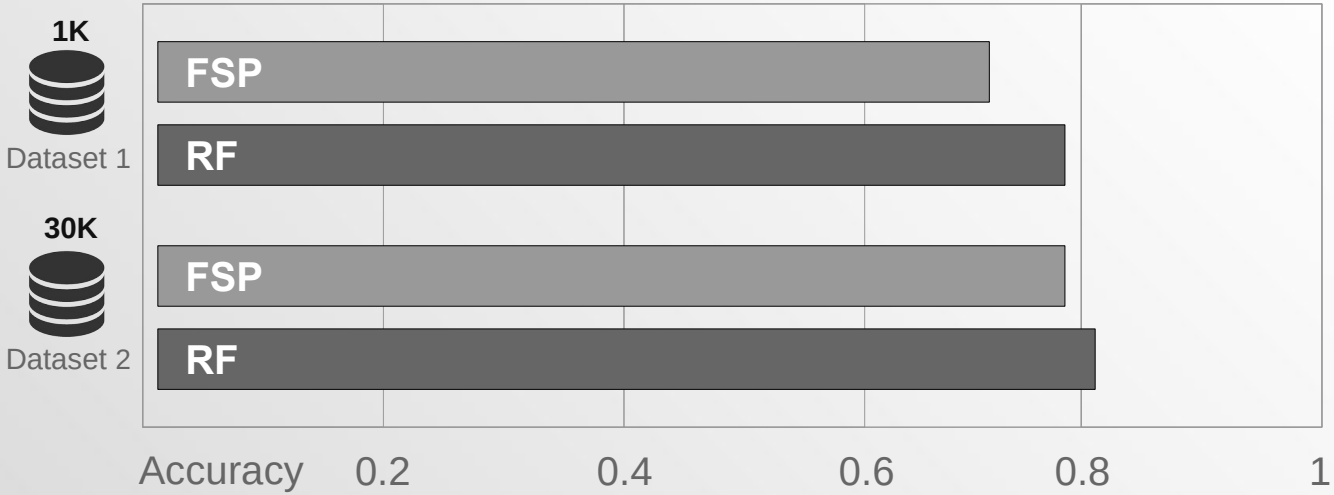
2. Comparison of the **frequency patterns** of two instances in terms of their frequency components **magnitude**

3. Definition of an **algorithm** able to **classify** a new instance as *reliable* or *unreliable*, by comparing it to all the previous legitimate ones

The **experiments** have been conducted on two imbalanced real-world datasets, using as competitor one of the most performing state-of-the-art approach, i.e., **Random Forests**

Accuracy

- The general performance of our **FSP*** approach in terms of *Accuracy* is very close to those of its competitor (**RF**)

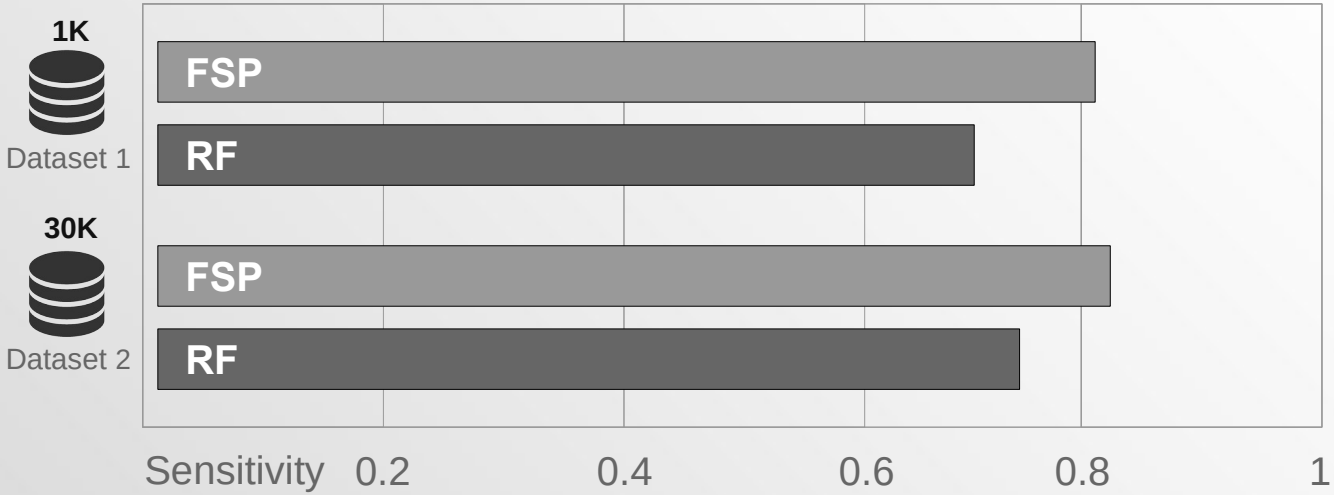


- In spite of the fact that we did **not use** any past default instance to train our model

** Fourier Spectrum Pattern*

Sensitivity

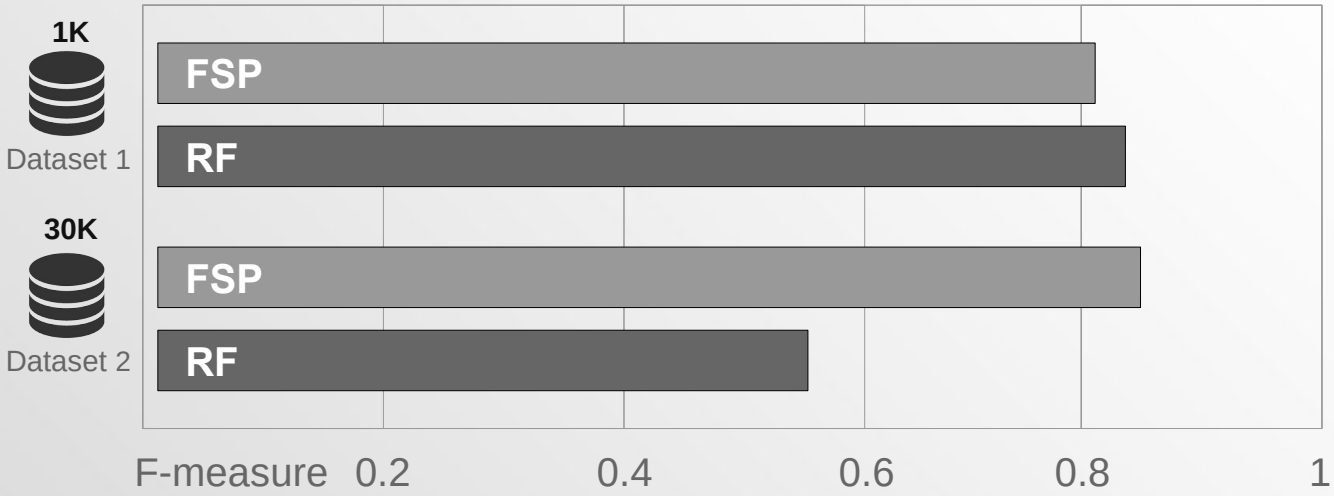
- In terms of Sensitivity, the general performance of our **FSP** approach is better than that of its competitor (**RF**)



- This indicates a better capability to detect the *reliable* instances (wrt to **RF)**

F-measure

- The last set of experiments measure the performance in terms of **F-measure**



- Differently from **RF**, the effectiveness of **FSP** increases with the number of instances

Conclusions

- The proposed approach leads toward some **advantages**:
  - It overcomes the **data imbalance** issue by using only a class of data to define its evaluation model (*previous reliable instances*)
  - Using only reliable instances it allows a Credit Scoring system to operate in a **proactive** way
  - Consequently, such strategy is able to face the **cold-start** issue

- It is not designed to replace the **retroactive** state-of-the-art approaches
- Its aim is to address some well-known problems, such as the **data unbalance** and the **cold-start**
- It allows us to define **hybrid Credit Scoring approaches** able to operate in different real-world scenarios

A Fourier Spectral Pattern Analysis to Design Credit Scoring Models

THANK YOU FOR YOUR ATTENTION!

IML 2017, International Conference on Internet of Things and Machine Learning
17÷18 October 2017, Liverpool John Moores University – Liverpool, United Kingdom