

Harbinger

Uma estrutura unificada para detecção
de eventos em séries temporais

Eduardo Ogasawara
eduardo.ogasawara@cefet-rj.br
<https://eic.cefet-rj.br/~eogasawara>

Harbinger: Uma estrutura unificada para detecção de eventos em séries temporais



Harbinger



Stars 22 downloads 18K/month

Harbinger is a framework for event detection in time series. It provides an integrated environment for anomaly detection, change point detection, and motif discovery. Harbinger offers a broad range of methods and functions for plotting and evaluating detected events.

For anomaly detection, methods are based on: - Machine learning model deviation: Conv1D, ELM, MLP, LSTM, Random Regression Forest, and SVM - Classification models: Decision Tree, KNN, MLP, Naive Bayes, Random Forest, and SVM - Clustering: k-means and DTW - Statistical techniques: ARIMA, FBIAD, GARCH

For change point detection, Harbinger includes: - Linear regression, ARIMA, ETS, and GARCH-based approaches - Classic methods such as AMOC, ChowTest, Binary Segmentation (BinSeg), GFT, and PELT

For motif discovery, it provides: - Methods based on Hashing and Matrix Profile

Harbinger also supports **multivariate time series analysis** and **event evaluation** using both traditional and soft computing metrics.

The architecture of Harbinger is based on **Experiment Lines** and is built on top of the [DAL Toolbox](#). This design makes it easy to extend and integrate new methods into the framework.

Por que uma estrutura unificada?

- Existem vários métodos, mas eles são focados em tipos específicos
- A escolha apropriada depende da natureza da série cronológica
- As aplicações do mundo real exigem comparação sistemática, combinação de detectores e avaliação rigorosa



	Anomalies	Change point	Motifs
Methods	50	10	9

Features

Metrics

Hard (traditional) and Fuzzy

Models

Ensemble and Ensemble Fuzzy

Thresholds customization

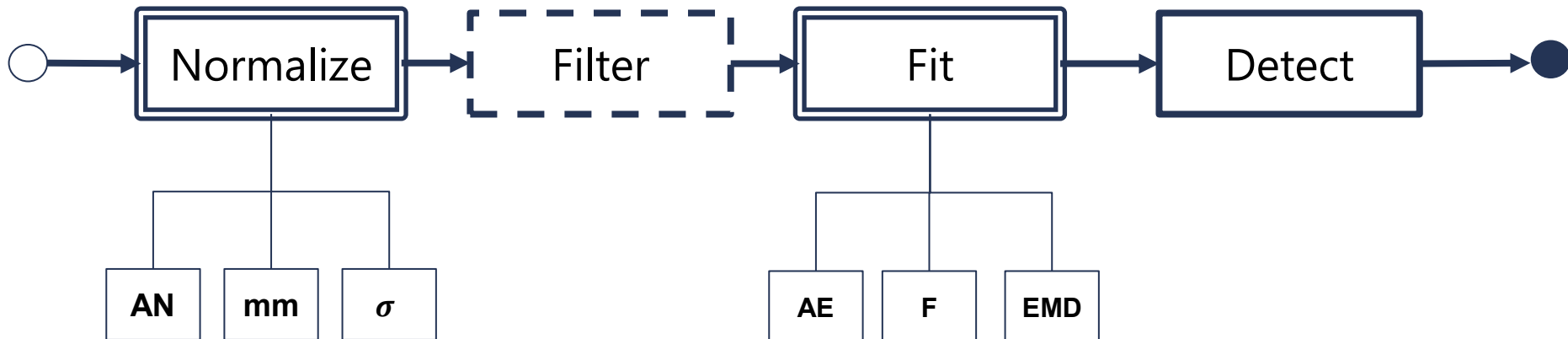
Deviation, Filters, and Candidate selection

Visualization

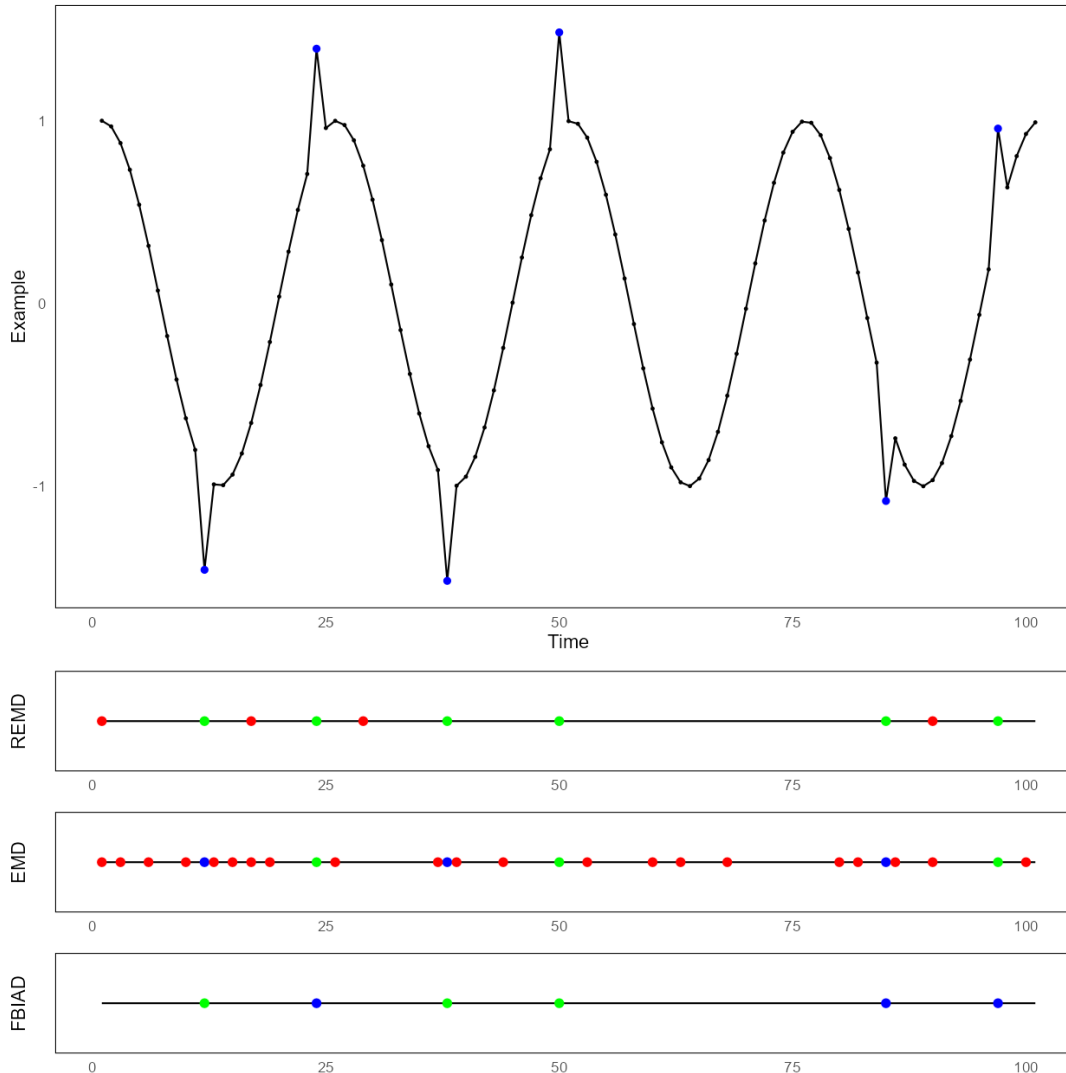
Time Series, Comparison, Residuals, and thresholds

Arquitetura modular

- Detecção, Avaliação, Combinação, Comparação
- Cada módulo pode ser usado sozinho ou integrado
- Interface rígida (baseada em álgebra de fluxo de trabalho e linhas de experimento)
- Construído sobre o DAL Toolbox
- <https://cran.r-project.org/web/packages/daltoolbox>
- Inspirado no Scikit-learn: funções fit() e detect()



Harbinger permite a avaliação e benchmark de detectores



REMD: A Novel Hybrid Anomaly Detection Method Based on EMD and ARIMA

Rosine Souza CHETEM rosine.souza@ufpe.br	Elton Paixão CHETEM elton.paixao@ufpe.br	Fernando Fraga CHETEM fernando.fraga@ufpe.br
Luis Baroni CHETEM luis.baroni@ufpe.br	Ronaldo Fernando Santos Alves Fiscomp ronaldo.alves@ufpe.br	Keli Belloze CHETEM keli.belloze@ufpe.br
Joel Santos CHETEM joel.santos@ufpe.br	Edwards Bezerra CHETEM edwards@ufpe.br	Fabio Porto LACC fporto@ufpe.br
		Edwards Ogasawara CHETEM ogawasawa@ufpe.br

Abstract—Anomalies are defined as behavioral deviations from expected patterns and pose challenges to identify them. Anomaly detection is a fundamental activity of time series analysis. It is often defined as identifying, in noisy, temporal and multidimensional data, events or observations that deviate from the expected behavior, such as healthcare, cyber security, system diagnosis, and so on. Many anomaly detection methods exist, but choosing the appropriate method is complex due to the intrinsic nature of the time series. There is a demand for robust and adaptable anomaly detection methods. This paper introduces Retimed Empirical Mode Decomposition (REMD) as a hybrid approach, addressing the need for integrating Empirical Mode Decomposition (EMD) and Autoregressive Integrated Moving Average (ARIMA) models. REMD's design aims to optimize the synergy of both methods and overcome their limitations. It is evaluated against state-of-the-art methods on diverse datasets. It demonstrates superior performance, with up to three times better F1 score.

1. INTRODUCTION

Time series analysis plays a fundamental role in various fields, providing valuable insights into the behavior of data over time [1]. Through this analysis, it is possible to identify seasonal and trend components, patterns, and anomalies, enabling the prediction of future directions and making informed decisions [2]. Among its applications, anomaly detection plays a highly relevant role in sectors such as healthcare [3], water quality [4], system reliability analysis [5], and in defining and exploring [6]. Anomaly detection enables proactive decision-making, contributing to operational efficiency, safety, and the prevention of potential issues [7].

Anomalies refer to behavioral changes that deviate from the expected pattern [8, 9]. Identifying and classifying them is challenging, and choosing the appropriate method for such a purpose becomes complex. Characteristics such as how the changes in behavior manifest in the original time series and the duration of the phenomenon contribute to the specialization of methods for distinct types of problems, leading to a propensity to generate false positives in divergent situations [10]. Therefore, the complexity lies not only in the intrinsic

nature of anomalies in time series but also in the need to choose detection approaches that appropriately consider the specificity of each scenario.

In this dynamic context, the demand for robust and highly adaptable methods becomes essential, regardless of the time series's origin under analysis [11]. Furthermore, these methods should be able to be applied in many different time series without knowing their intrinsic nature beforehand.

Retimed Empirical Mode Decomposition (REMD) was designed to address this problem. It is a hybrid method that integrates the benefits of time series decomposition using Empirical Mode Decomposition (EMD) [12] and time series modeling using Autoregressive Integrated Moving Average (ARIMA) [13]. The REMD was conceived with the synergistic vision of combining two relatively adaptable methods to optimize their strengths and overcome limitations.

REMD was evaluated against state-of-the-art (SOTA) statistical and machine learning methods based on different datasets with diverse properties to be tested, such as volatility, trend, and the presence or absence of seasonality. The results indicate that REMD outperformed SOTA methods with up to three times better F1 score.

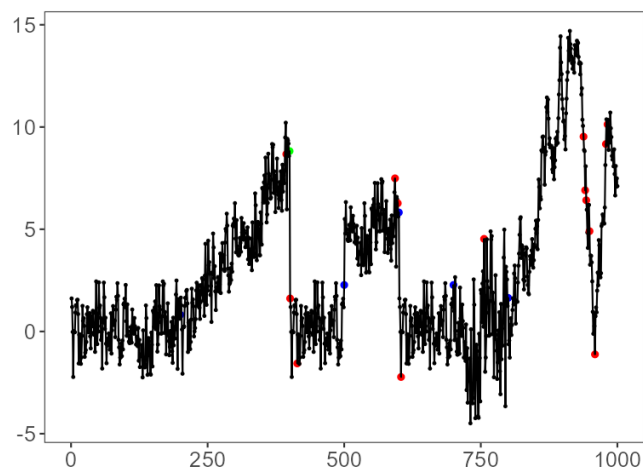
Besides this introduction, the paper is organized into five main sections. Section II provides general background on anomaly detection. Section III discusses related works. Section IV introduces the REMD. Section V presents the experimental evaluation and its discussion, and finally, Section VI offers concluding remarks.

II. BACKGROUND

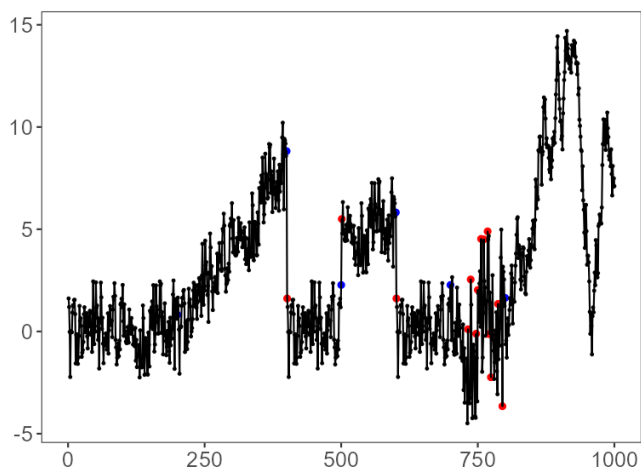
A. Time series

A time series is a collection of observations made chronologically. Formally, a time series Y is described as a sequence of observations y_1, y_2, \dots, y_n , where $n \in \mathbb{N}$ and t represents the number of observations. Here, n represents the entire observation, while n_t denotes the most recent one. This

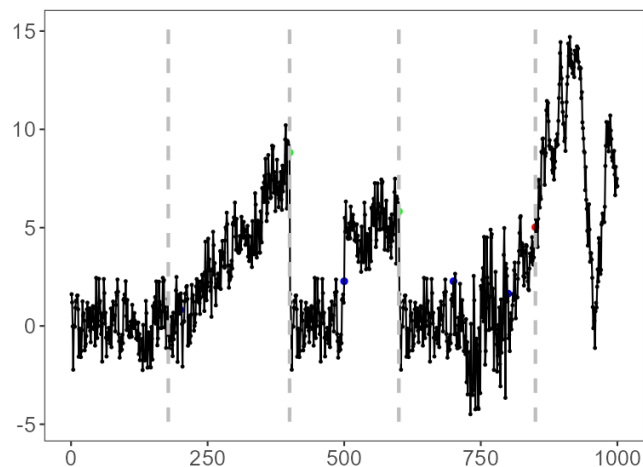
Visualização combinada de detectores em uma série temporal



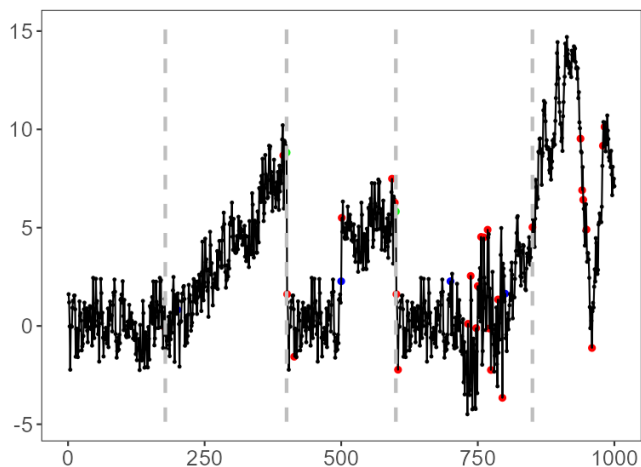
(a) FBIAD



(b) LSTM



(c) GFT



(d) Integrated View (FBIAD+LSTM+GFT)

Exemplo simples de detecção de anomalias

- O Harbinger vem com algumas séries de exemplos
- Séries temporais rotuladas
- Funções de plotagem
- Muitos detectores, como o ARIMA

```
library(daltoolbox)
```

```
library(harbinger)
```

```
data(examples_anomalies)
```

```
dataset <- examples_anomalies$simple
```

```
model <- hanr_arima()
```

```
# fitting the model
```

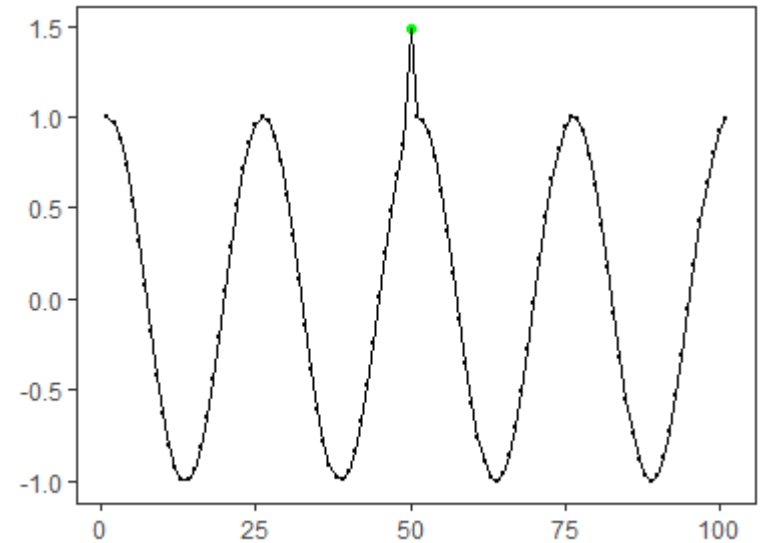
```
model <- fit(model, dataset$serie)
```

```
# making detections
```

```
detection <- detect(model, dataset$serie)
```

```
# plotting the results
```

```
har_plot(model(), dataset$serie, detection)
```



Entendendo a detecção

- Explorando detecções
- Visualizando os resíduos
- Avaliando os resultados

filtering detected events

```
print(detection |>  
      dplyr::filter(event==TRUE))
```

```
##   idx event   type  
## 1   50  TRUE anomaly
```

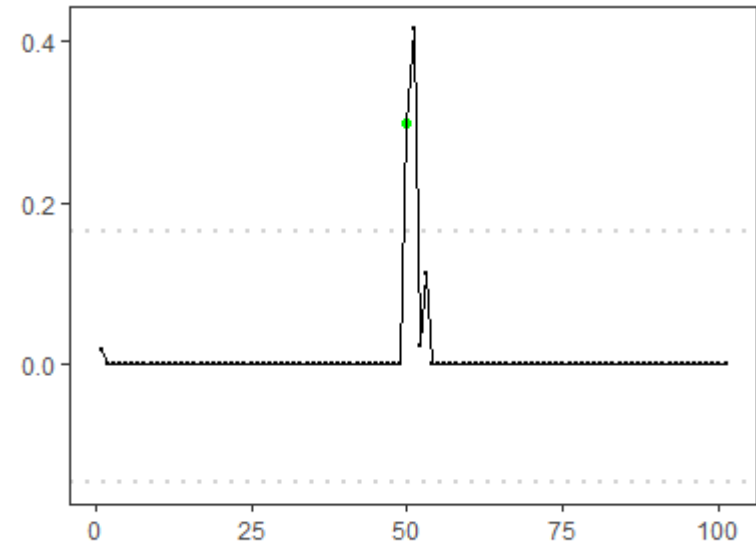
understating detected events

```
har_plot(model, attr(detection, "res"), detection,  
          dataset$event, yline = attr(detection, "threshold"))
```

evaluating detected events

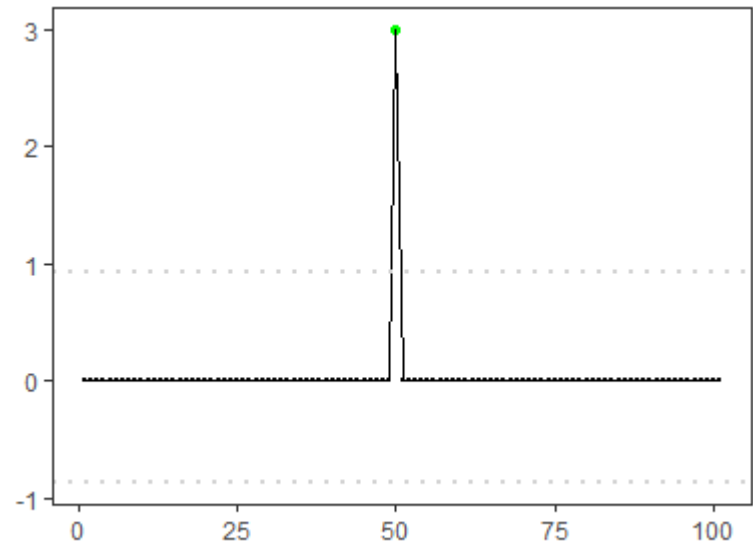
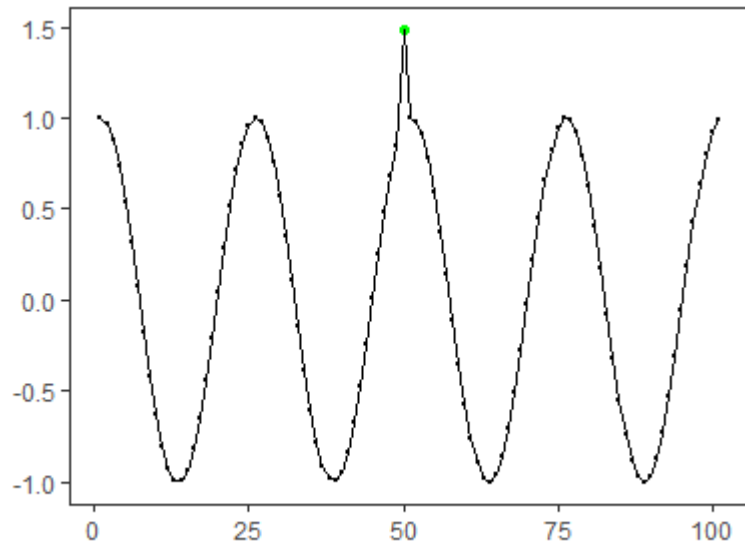
```
evaluation <- evaluate(model, detection$event, dataset$event)  
print(evaluation$confMatrix)
```

```
##           event  
## detection TRUE  FALSE  
## TRUE      1     0  
## FALSE     0    100
```



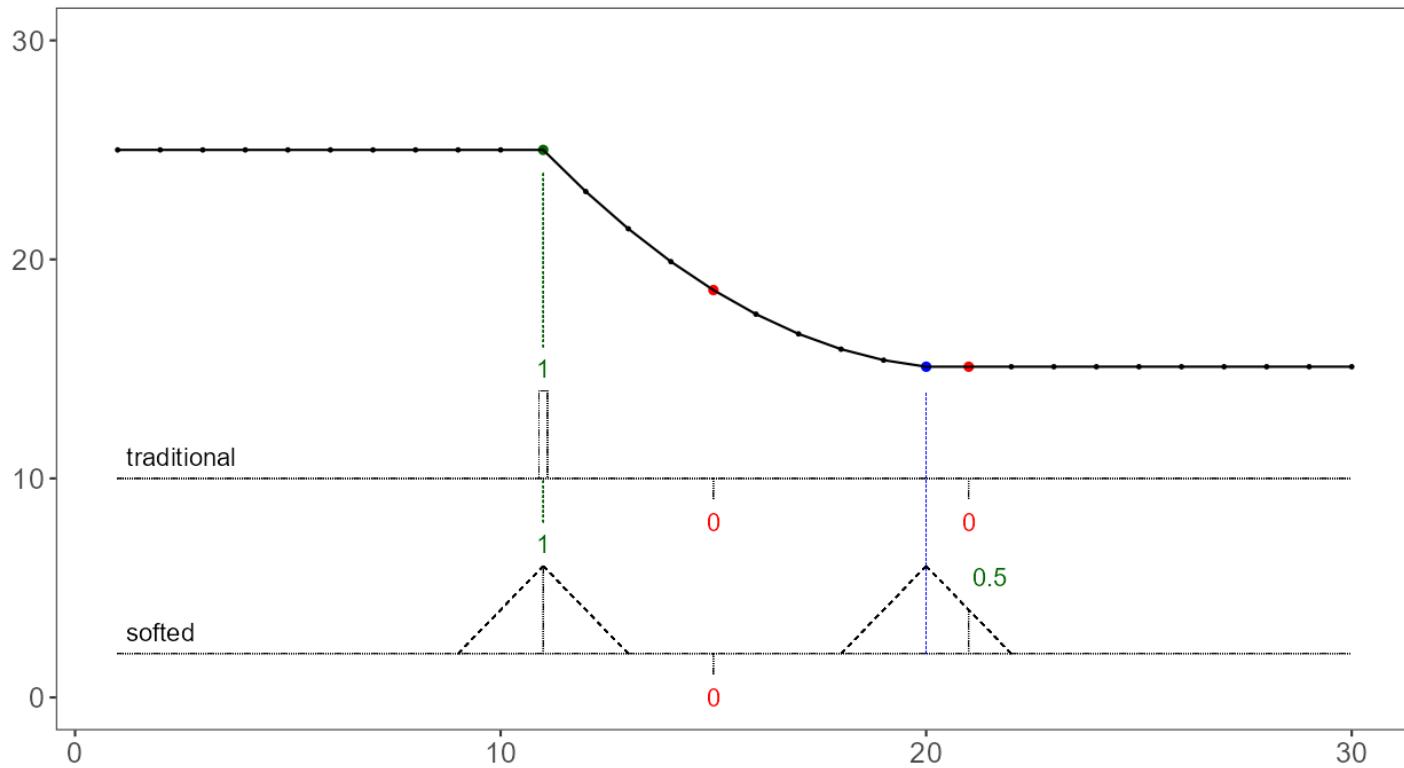
Comitê de modelos

```
model <- har_ensemble(hanr_fbiad(), hanr_arima(), hanr_emd())  
model <- fit(model, dataset$serie)  
detection <- detect(model, dataset$serie)  
• har_plot(model, dataset$serie, detection, dataset$event)  
• har_plot(model, attr(detection, "res"), detection,  
  dataset$event, yline = attr(detection, "threshold"))
```

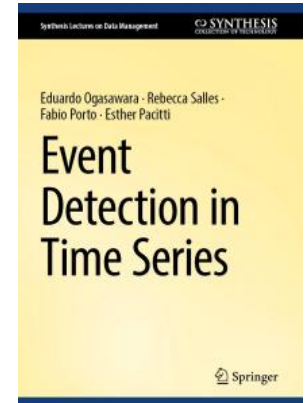
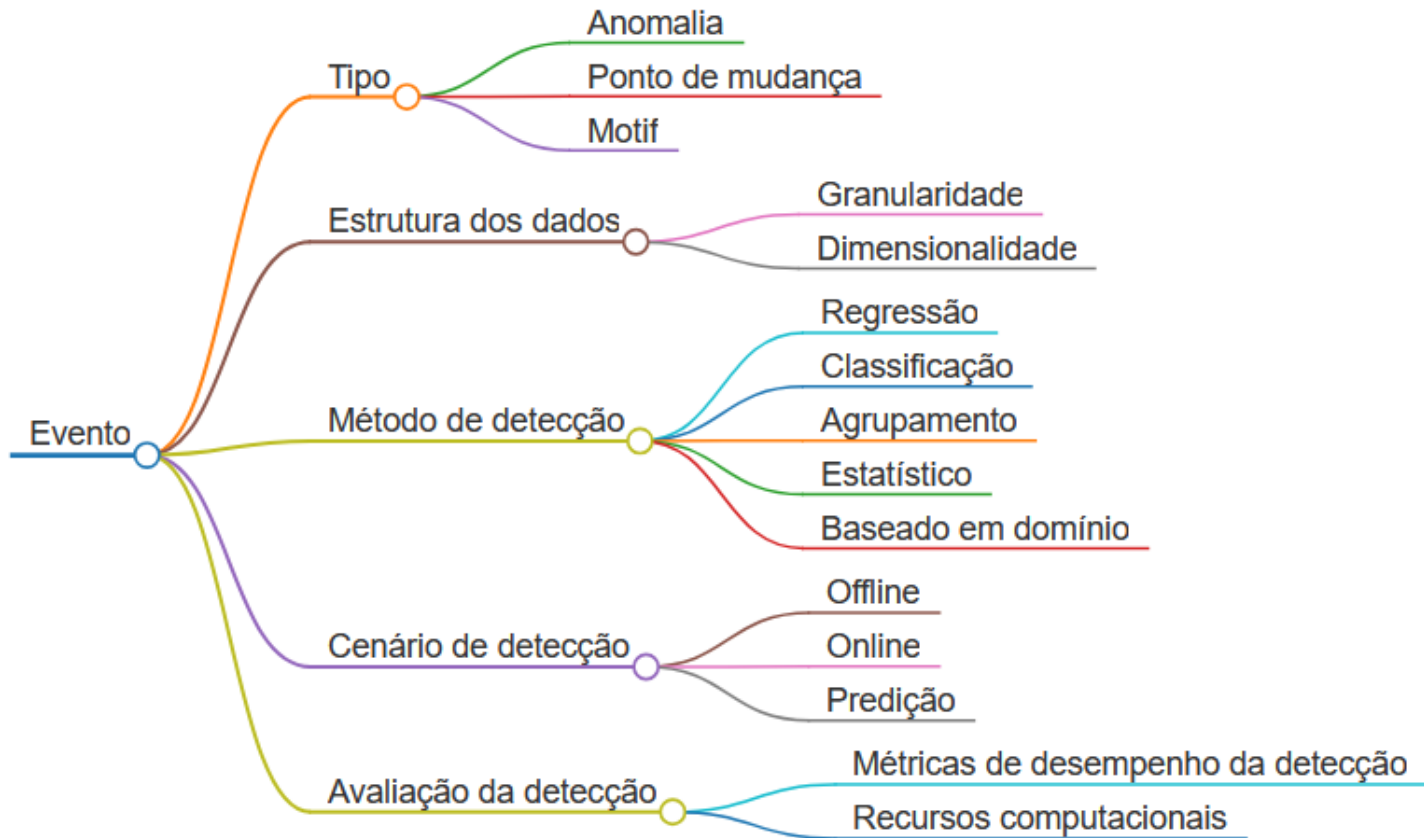


Fuzificando eventos

- Ao contrário das métricas tradicionais, o SoftED tolera pequenos desalinhamentos temporais, mas impõe penalidades mais fortes em padrões de detecção fragmentados ou inconsistentes



Taxonomia da detecção de eventos em séries temporais



Harbinger

Uma estrutura unificada para detecção
de eventos em séries temporais

Eduardo Ogasawara
eduardo.ogasawara@cefet-rj.br
<https://eic.cefet-rj.br/~eogasawara>