

Harbinger: Um framework para integração e análise de métodos de detecção de eventos em séries temporais*

**Rebecca Salles¹, Luciana Escobar¹, Lais Baroni¹, Roccio Zorrilla²,
Artur Ziviani², Vinicius Kreischer², Flavia C. Delicato³, Paulo F. Pires³,
Luciano Maia⁴, Rafaelli Coutinho¹, Laura Assis¹, Eduardo Ogasawara¹**

¹CEFET/RJ - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca

²LNCC - Laboratório Nacional de Computação Científica

³UFF - Universidade Federal Fluminense

⁴Petrobras

{rebecca.salles,luciana.vignoli,lais.baroni}@eic.cefet-rj.br

{ziviani,roccio.zorrilla,vinicius.kreischer}@lncc.br

{fdelicato,paulo.pires}@ic.uff.br, luciano.maia@petrobras.com.br

{rafaelli.coutinho,laura.assis}@cefet-rj.br, eogasawara@ieee.org

Abstract. *When analyzing time series, it is possible to observe significant changes in the behavior of its observations that frequently characterize the occurrence of events. Events may appear as anomalies, change points, or frequent patterns. In literature, there are several methods for event detection. However, the search for a suitable method for a time series is not a simple task, especially considering that the nature of the events is often not known. In this context, this work presents Harbinger, a framework for integration and analysis of event detection methods. Harbinger was evaluated with synthetic and real data, where it was possible to verify that its functionalities foster the selection of methods and the understanding of detected events.*

Resumo. *Ao analisar séries temporais é possível observar mudanças significativas no comportamento das observações que frequentemente caracterizam a ocorrência de eventos. Eventos se apresentam como anomalias, pontos de mudança, ou padrões frequentes. Na literatura existem diversos métodos para detecção de eventos. Entretanto, a busca por um método adequado para uma série temporal não é uma tarefa simples, principalmente considerando-se que a natureza dos eventos muitas vezes não é conhecida. Neste contexto, este trabalho apresenta Harbinger, um framework para integração e análise de métodos de detecção de eventos. O Harbinger foi avaliado em dados sintéticos e reais, onde foi possível constatar que suas funcionalidades promovem a seleção de métodos e a compreensão dos eventos detectados.*

1. Introdução

Na análise de séries temporais, frequentemente é possível observar a ocorrência de uma mudança significativa em seu comportamento em um certo ponto ou intervalo de tempo.

*Os autores agradecem a Petrobras, FAPERJ, CAPES (cód. 001) e CNPq pelo financiamento do projeto.

Tal mudança de comportamento geralmente caracteriza a ocorrência de um evento [Guralnik and Srivastava, 1999]. Um evento pode representar um fenômeno com significado definido em um domínio de conhecimento. Assim, o problema de detecção de eventos se mostra relevante para aplicações baseadas em análise de dados de sensores. Exemplos de tais aplicações são observados nos domínios da química, da sísmica de reflexão e da perfuração e exploração de petróleo, em que o monitoramento das operações é essencial.

A literatura apresenta muitos métodos para detecção de eventos. A escolha de um método adequado para uma série temporal não é uma tarefa simples. A seleção e parametrização de um método estão diretamente relacionadas às suposições iniciais quanto ao comportamento e às propriedades estatísticas dos dados. A complexidade desta tarefa é agravada considerando-se que a natureza dos eventos observados em uma série temporal muitas vezes é desconhecida. Eventos detectados em séries temporais comumente se apresentam como anomalias, pontos de mudança ou padrões frequentes. Neste caso, métodos especializados na detecção de um tipo de evento podem negligenciar a ocorrência de outro, ou ainda, identificá-los erroneamente. Falhas na identificação de eventos podem afetar a tomada de decisão ou levar a falsos positivos. Com isso, tem-se uma perda de credibilidade no uso das técnicas de controle e um possível prejuízo às aplicações.

Neste contexto, este trabalho apresenta o *Harbinger*, um *framework* para integração e análise de métodos de detecção de eventos. Para isso, são uniformizados e integrados alguns dos principais métodos de detecção de eventos disponíveis na literatura, podendo também ser incluídos pelo usuário. Tal integração viabiliza uma análise comparativa qualitativa e quantitativa de suas detecções, bem como o uso combinado de métodos para caracterizar de modo mais assertivo os eventos detectados. A proveniência dos resultados se dá via sistemas de gerenciamento de ciclo de vida de modelos.

O *Harbinger* foi avaliado com base em dados sintéticos e reais. Constatou-se que as suas funcionalidades ajudam a nortear a escolha de métodos de detecção apropriados para uma determinada série temporal e contribuem para a compreensão da semântica dos eventos detectados. A Seção 2 apresenta o referencial teórico e trabalhos relacionados. Na Seção 3 são descritos o *framework* e suas funcionalidades. A Seção 4 apresenta a avaliação experimental e sua discussão, enquanto a Seção 5 tece considerações finais.

2. Séries temporais e detecção de eventos

Uma série temporal é uma sequência de observações coletadas ao longo do tempo. Geralmente, considera-se uma série temporal y como um processo estocástico, *i.e* uma sequência de n variáveis aleatórias $\langle y_1, y_2, \dots, y_n \rangle$ [Carmona, 2014]. Uma observação de uma série temporal é referenciada como y_i , indexada no tempo por $i = 1, \dots, n$, onde y_1 representa a primeira observação e y_n a mais recente. Uma subsequência de tamanho p de y na posição i , representada por $seq_{i,p}(y)$, é uma sequência contínua de valores $\langle y_{i-(p-1)}, y_{i-(p-2)}, \dots, y_i \rangle$, onde $|seq_{i,p}(y)| = p$ e $p \leq i \leq |y|$. A média móvel $\bar{y}_{i,p}$ na posição i , de p termos para y é computada pela média das t_k observações nas subsequências definidas por $seq_{i,p}(y)$. A Eq. 1 descreve $\bar{y}_{i,p}$.

$$\bar{y}_{i,p} = p^{-1} \cdot \sum_{k=1}^p t_k \mid t_k \in seq_{i,p}(y), 1 \leq k \leq p \quad (1)$$

2.1. Detecção de anomalias de tendência

Anomalias são observações que se destacam por parecerem não terem sido geradas pelo mesmo processo que as demais observações na série temporal. Neste caso, anomalias podem ser modeladas como observações isoladas dos dados restantes com base em funções de similaridade ou distância [Gupta et al., 2014]. Podem ser identificadas como $a(y)$ por meio da Eq. 2, onde $Q_1(y)$ e $Q_3(y)$ são respectivamente o primeiro e terceiro quartil e IQR é a distância interquartil [Gupta et al., 2014].

$$a(y) = \{i, \forall i \mid y_i \notin [Q_1(y) - 1.5 \cdot IQR(y), Q_3(y) + 1.5 \cdot IQR(y)]\} \quad (2)$$

No contexto de séries temporais, existe um interesse particular na detecção de anomalias que podem representar a ocorrência de um evento que foge à tendência inerente ao processo gerador de y . Seja \hat{y} uma estimativa do processo gerador de y , produzido pelo ajustamento de um modelo α , tendo $\hat{y}_i = \alpha(y)_i$. Sendo ϵ uma série temporal de resíduos (*white noise*) obtidos após a remoção da tendência \hat{y} , tem-se que as anomalias de tendência de y são geralmente identificadas como $at(y)$ por meio da Eq. 3.

$$at(y) = a(\epsilon), \epsilon_i = y_i - \hat{y}_i \quad (3)$$

A literatura apresenta diversos métodos para a detecção de anomalias de tendência. Entre eles estão os baseados em decomposição, normalização adaptativa (AN), e o KNN-CAD. O método de decomposição adota uma abordagem que consiste em decompor a série temporal em três componentes: tendência, sazonalidade e o restante, sobre o qual é feita a busca por anomalias [Gupta et al., 2014]. No AN, é feito uso de inércia para abordar séries não estacionárias heterocedásticas através do cálculo da média móvel [Salles et al., 2019].

2.2. Detecção de anomalias de volatilidade

A maioria das séries temporais financeiras exibe propriedades não lineares que não podem ser capturadas pelos modelos lineares existentes, dado que a volatilidade dessas séries varia bastante ao longo do tempo. Surge assim uma demanda para o estudo da volatilidade das séries temporais. Modelos econométricos aparecem para tratar a não linearidade dos dados, incluindo volatilidade estocástica, como o ARCH e GARCH, sendo o último o mais conhecido e aplicado [Carmona, 2014]. Na área financeira, a volatilidade está associada a risco, que no contexto de séries temporais pode indicar um evento.

Os modelos do tipo GARCH envolvem a estimativa da volatilidade com base em observações anteriores. GARCH é um modelo de série temporal não linear, onde uma série y_i é explicada conforme a Eq. 4, sendo μ_i o componente médio. A sequência de ruído w_i é i.i.d. $N(0, 1)$, de modo que a distribuição condicional de $\tilde{y}_i = y_i - \mu_i$, dado $\tilde{y}_{i-1}, \tilde{y}_{i-2}, \dots$ é $N(0, \sigma_i^2)$ [Carmona, 2014]. Tal modelo pode ser utilizado como α para detecção de anomalias na Eq. 3, ou ainda, suas estimativas de volatilidades instantâneas podem ser sujeitas a uma detecção de anomalias (Eq. 2).

$$y_i = \mu_i + \sigma_i w_i \quad (4)$$

2.3. Detecção de pontos de mudança

Os métodos de detecção de pontos de mudança objetivam encontrar na série temporal os pontos ou intervalos no tempo que representam uma transição entre diferentes estados

em um processo que gera os dados da série temporal [Takeuchi and Yamanishi, 2006]. A detecção de ponto de mudança pode ser definida como um problema de teste de hipóteses, onde a hipótese nula H_0 caracteriza a inexistência de pontos de mudança e a hipótese alternativa H_A nega H_0 . Seja $seq_{i,p}(y)$ uma subsequência de observações de uma série temporal e $t, k \in \{i, \dots, i + p\}$, sendo $t \leq k$, tem-se formalmente que $H_0 : \forall t, k (t \neq k) \mid \mathbb{P}_{y_t} = \mathbb{P}_{y_k}$ e $H_A : \forall t \exists k (t \neq k) \mid \mathbb{P}_{y_t} \neq \mathbb{P}_{y_k}$, onde \mathbb{P}_{y_i} é a função densidade de probabilidade da subsequência e k é um ponto de mudança [Chen and Zhang, 2015].

Uma das propostas desenvolvidas para a detecção de pontos de mudança em séries temporais é baseada na Média Móvel Exponencialmente Ponderada (EWMA) [Raza et al., 2015]. Entretanto, o método seminal de detecção de pontos de mudança (SCP), tornou-se referência na literatura [Guralnik and Srivastava, 1999] e promoveu o desenvolvimento de inúmeras abordagens como o *ChangeFinder* (CF) [Takeuchi and Yamanishi, 2006], composto por duas fases. Dada uma série temporal y , na primeira fase, um modelo α é ajustado, resultando em \hat{y}_i , e são identificadas suas anomalias a partir dos resíduos contidos na série s definida na Eq. 5. Na segunda fase é produzida uma nova série \bar{s}_p composta pelas médias móveis de s com p termos. A detecção de pontos de mudança é reduzida ao problema de detecção de anomalias em \bar{s}_p .

$$s_i = (\hat{y}_i - y_i)^2, \hat{y}_i = \alpha(y)_i \quad (5)$$

2.4. Trabalhos relacionados

A literatura apresenta *frameworks* para detecção de anomalias, como o desenvolvido por De Paepe et al. [2020], que descreveram o *framework* SDM que utiliza distância entre subsequências através da *Matrix Profile*, e por Eriksson et al. [2010] que propuseram *BasisDetect*, um *framework* para detecção de anomalias em dados de redes. Além disso, propostas para detecção de anomalias em dados de fluxo contínuo são apresentadas. Calikus et al. [2020] implementaram detectores de eventos obtidos através da combinação de doze métodos em um *framework* chamado SAFARI e Talagala et al. [2020] propuseram um *framework* que utiliza a teoria dos valores extremos para a detecção de anomalias. Em contrapartida, para a detecção de pontos de mudança, Lu et al. [2016] desenvolveram um *framework* baseado no teste estatístico de anomalias, enquanto Xiong et al. [2015] apresentaram um *framework* para detectar pontos de mudança em séries multivariadas.

No entanto, os *frameworks* disponíveis na literatura geralmente se especializam na detecção de eventos de semântica específica, podendo negligenciar ou mal interpretar eventos de diferentes tipos contidos em uma série temporal. Além disso, são comumente dependentes do número limitado de métodos de detecção que implementam, que podem ser inadequados a determinadas aplicações. Segundo o conhecimento dos autores, existe uma demanda por um *framework* flexível para integração e análise de diversos métodos de detecção de eventos, permitindo assim a identificação de eventos de diferentes tipos pela combinação de métodos personalizados e viabilizando uma análise comparativa de suas detecções. O *Harbinger* visa atender a essa demanda.

3. Harbinger

A estrutura geral do *Harbinger* é ilustrada na Fig. 1(a), sendo formada por quatro principais módulos de funcionalidade: detecção de eventos (em azul), avaliação de qualidade

de detecções (em verde), combinação de resultados de detecção (em laranja), e comparação de desempenhos de detecção (em roxo). A descrição dos parâmetros de execução do *Harbinger*, assim como de seus artefatos produzidos é apresentada na Fig. 1(b).

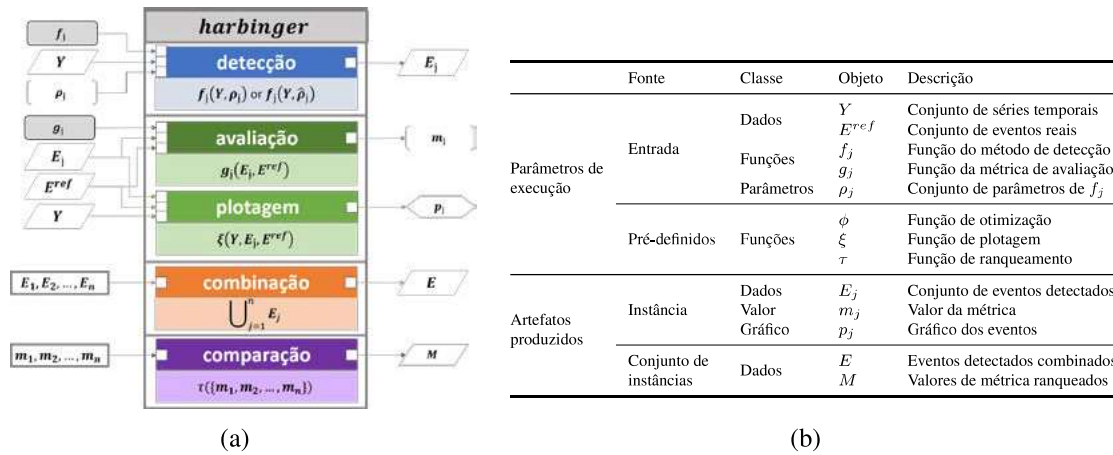


Figura 1. (a) Diagrama geral dos módulos; (b) Parâmetros de execução e artefatos produzidos.

O processo de uso do *Harbinger* inclui inicialmente a j -ésima instanciação de um processo de detecção de eventos, $O_j = \langle Y, E^{ref}, f_j, g_j, \rho_j, E_j, m_j, p_j \rangle$. O_j é implementada pelos módulos de detecção e avaliação ilustrados no diagrama na Fig. 1(a). Para esta instanciação são passados como parâmetros de entrada um conjunto de séries temporais Y , uma lista opcional de seus respectivos eventos reais previamente classificados E^{ref} , um método de detecção f_j , seus parâmetros necessários ρ_j e uma métrica selecionada para avaliação da qualidade de suas detecções, g_j (caso E^{ref} tenha sido informado).

Como resultado da execução de uma instância O_j , *Harbinger* retorna o conjunto de eventos detectados $E_j = f_j(Y, \rho_j)$. É também permitida a otimização de ρ_j pela função ϕ . Neste caso, $E_j = f_j(Y, \hat{\rho}_j)$, com $\hat{\rho}_j = \phi(Y, f_j, \rho_j)$. E_j é apresentado ao usuário contendo os instantes de tempo em que eventos foram detectados, as séries temporais a que os eventos pertencem e os tipos dos eventos detectados. Em seguida, a qualidade das detecções em E_j é avaliada por g_j utilizando E^{ref} como base de comparação. O valor da métrica computada, $m_j = g_j(E_j, E^{ref})$, é retornado. m_j é então utilizado para a análise quantitativa da qualidade das detecções de O_j . Além disso, *Harbinger* gera um gráfico p_j através da função ξ para análise visual de E_j tendo Y e E^{ref} como referências. p_j pode então ser utilizado para uma análise qualitativa das detecções de O_j .

Enfim, o processo de uso do *Harbinger* permite a combinação dos resultados das execuções de um conjunto $O = \{O_1, \dots, O_n\}$ de n instâncias de um processo de detecção de eventos definidas no *Harbinger*. Esta tarefa é ilustrada no módulo de combinação do diagrama da Fig. 1(a), onde *Harbinger* une os conjuntos de eventos detectados em cada instância j , contidos em E_j , com $j = 1, \dots, n$. O conjunto união E é retornado contendo todos os eventos detectados em Y por diferentes métodos de detecção e/ou diferentes parâmetros. Em consequência, E é capaz de englobar eventos detectados de diferentes tipos apresentando diferentes semânticas que são devidamente identificadas de acordo com a metodologia característica de seus respectivos métodos de detecção.

Da mesma forma, o módulo de comparação do diagrama na Fig. 1(a) ilustra a

comparação das métricas de qualidade m_j , com $j = 1, \dots, n$, computadas durante as execuções das instâncias de O . O conjunto de valores $\{m_1, m_2, \dots, m_n\}$ é submetido a um processo de ranqueamento τ . O conjunto M de valores ranqueados de métricas de qualidade é retornado. M pode então ser utilizado como base para avaliação das n instâncias definidas cujos desempenhos de detecção de eventos são ordenados conformemente. Essa ordenação norteia a escolha dos métodos de detecção e suas parametrizações mais adequadas para uma determinada aplicação de detecção de eventos e pode ser registrada em sistemas de gerenciamento de ciclo de vida de modelos [Silva et al., 2019].

3.1. Implementação do *framework*

Desenvolvido em R, *Harbinger* disponibiliza uma biblioteca de funções que implementam os módulos de detecção, avaliação, combinação e comparação descritos anteriormente e ilustrados na Fig. 1(a). O *framework* permite ao usuário a definição de métodos de detecção de eventos personalizados em f_j . Neste caso, a otimização de ρ_j por ϕ é baseada em uma abordagem *grid search*. Mesmo assim, *Harbinger* também conta com a implementação de diversos métodos de detecção de eventos. Os métodos implementados são descritos na Seção 2, sendo que o método de decomposição foi disponibilizado pelo pacote-R *anomalize*, o GARCH pelo pacote-R *rugarch*, e os métodos KNN-CAD e EWMA pelo pacote-R *otsad*. A implementação do método SCP no *Harbinger* utiliza o modelo de regressão linear para o ajuste nos dados das séries temporais, tendo os erros médios quadráticos (MSE) como métrica para avaliação dos ajustamentos. A implementação do CF, entretanto, tem seu modelo de aprendizado passado como parâmetro. Com o objetivo de facilitar o processo de avaliação, *Harbinger* conta com a implementação das principais métricas utilizadas para avaliar a qualidade da detecção de eventos. Neste caso, g_j pode ser definida como uma das seguintes métricas: matriz de confusão, sensibilidade, especificidade, valor preditivo positivo, valor preditivo negativo, precisão, revocação, F1, prevalência, taxa de detecção, prevalência de detecção, ou acurácia equilibrada. Mais detalhes sobre o *Harbinger* são disponibilizados em sua página dedicada.¹

4. Avaliação dos módulos de funcionalidade *Harbinger*

O *Harbinger* foi submetido à uma avaliação experimental com o objetivo de analisar as contribuições de seus módulos no contexto do problema de detecção de eventos em séries temporais. São avaliados os módulos de detecção e combinação, e em seguida os módulos de avaliação e comparação apresentados na Fig. 1(a). Para isso foi construída uma série temporal não estacionária sintética e foram selecionados conjuntos de dados utilizados como referência para a avaliação da qualidade dos métodos de detecção de eventos.

4.1. Conjuntos de dados

Para evidenciar as mudanças de comportamento de uma série temporal, foi desenvolvida uma série temporal sintética y com propriedades não estacionárias (NE) produzida a partir do trabalho de Salles et al. [2019]. Esta série é composta por 1000 observações onde a subsequência $\langle y_1, \dots, y_{200} \rangle$ representa uma série estacionária. As demais subsequências ilustram diversas formas de não estacionariedade. Em $\langle y_{201}, \dots, y_{400} \rangle$ observa-se a estacionariedade de tendência, em $\langle y_{401}, \dots, y_{600} \rangle$ a estacionariedade de nível,

¹<https://eic.cefet-rj.br/~dal/harbinger/>

em $\langle y_{601}, \dots, y_{800} \rangle$ a heteroscedasticidade, e em $\langle y_{801}, \dots, y_{1000} \rangle$ a estacionariedade por diferenciação. Foi também utilizado para análise o conjunto de dados criado para o *GECCO Challenge 2018* [Rehbach et al., 2018]. Para compor este conjunto, foram selecionadas 1500 observações coletadas a cada minuto contendo 72 eventos identificados. Foram coletadas 9 variáveis relacionadas à qualidade da água, das quais 5 foram selecionadas por apresentar maior variabilidade em suas observações. São elas Tp , pH , $Redox$, $Leit$ e $Trueb$, representando a temperatura, PH, potencial Redox, condutividade elétrica e turvação, respectivamente. Por fim, foi selecionado o conjunto de dados criado por *Yahoo Labs*. Este consiste de observações coletadas por hora contendo anomalias identificadas manualmente por editores [Webscope, 2015]. Uma parte dos dados é sintética e outra é baseada no tráfego real dos serviços do *Yahoo*. O conjunto de dados é dividido em quatro, sendo $A1$ composto pelos dados reais e $A2$, $A3$ e $A4$ por dados sintéticos.

4.2. Configurações da avaliação

Os conjuntos de dados selecionados, daqui por diante referenciados por NE, GECCO e *Yahoo*, respectivamente, foram utilizados como base para avaliação das contribuições dos módulos do *Harbinger*. Processos de detecção de eventos, O_j , foram instanciados para cada uma das séries temporais em estudo. Para isso, os métodos descritos na Seção 2 foram passados como parâmetro em f_j , abrangendo a detecção de eventos de diferentes tipos. No caso do CF, exceto quando dito o contrário, o modelo de regressão linear é dado em ρ_j para viabilizar comparação justa com resultados produzidos por SCP. A avaliação quantitativa da qualidade das detecções E_j se dá através das métricas da matriz de confusão e F1, enquanto a avaliação qualitativa das detecções se dá através de análise gráfica. Nos gráficos produzidos são mostradas as observações da série temporal, assim como marcações referentes aos resultados de detecção de eventos. Marcações vermelhas indicam eventos detectados. Além disso, são marcados em azul os eventos reais em E_{ref} , e em verde a coincidência entre o evento detectado e o evento real.

4.3. Avaliação dos módulos de detecção e combinação

Para avaliação dos módulos de detecção de eventos e de combinação de detecções, selecionamos a série temporal contida em NE. Apesar de não conter eventos previamente classificados, esta apresenta claras mudanças de comportamento e anomalias de tendência e de volatilidade que evidenciam os eventos detectados pelas instâncias do *Harbinger* e facilitam a sua avaliação por análise gráfica. A Fig. 2 apresenta NE com marcações correspondentes aos seus eventos detectados através do módulo de detecção do *Harbinger*. São apresentados nas Figs. 2(a) a 2(c) eventos de diferentes tipos detectados e identificados utilizando métodos de detecção diferentes: AN, CF e GARCH. Além disso, na Fig. 2(d), é apresentada a contribuição do módulo de combinação do *Harbinger* que engloba de maneira unificada as detecções produzidas pelos métodos citados. Marcações em forma de pontos indicam anomalias de tendência, triângulos indicam anomalias de volatilidade e linhas verticais tracejadas indicam pontos de mudança detectados.

Na Fig. 2(a) são observadas anomalias de tendência detectadas por AN. Sendo especializado em séries temporais não estacionárias e heterocedásticas, este foi capaz de detectar eventos nas subsequências de NE em que ocorreram mudanças significantes em sua média ou variância. Também é possível observar o efeito inercial da média móvel utilizada por AN sobre suas detecções. Anomalias de tendência são também detectadas

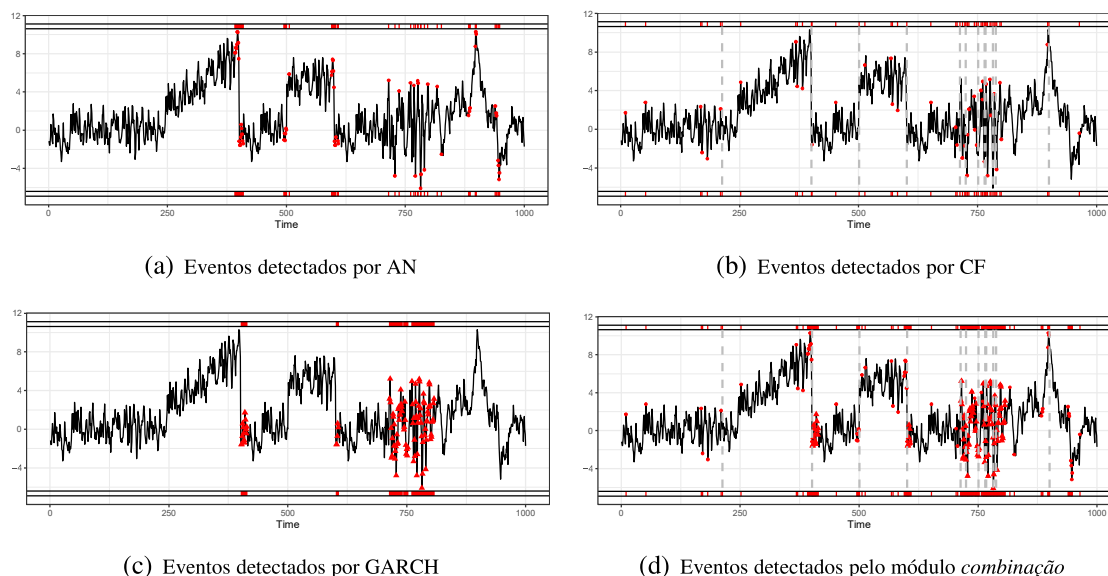


Figura 2. Detecções de eventos em NE por diversas instâncias O_j do *Harbinger*.

pelo método CF e são apresentadas na Fig. 2(b). Tendo o modelo ARIMA como base para estimativa de tendência, CF detecta as anomalias com base em seus desvios (Eq. 5). Por sua vez, diferentemente de AN, CF é capaz de detectar pontos de mudança, evidenciando mudanças bruscas de comportamento. É possível observar que os pontos de mudança detectados tendem a demarcar as subsequências não estacionárias de NE. Ao se comparar as detecções nas Figs. 2(a) e 2(b) observa-se que, apesar de adotarem abordagens diferentes, AN e CF apresentam detecções coincidentes. Neste caso, AN foi capaz de detectar pontos de mudança erroneamente identificados como anomalias. Tem-se que a análise dos resultados de CF contribui, portanto, para a melhor compreensão da natureza dos eventos detectados por AN. O mesmo pode ser dito ao se comparar as detecções de CF com as produzidas pelo método GARCH na Fig. 2(c). Analogamente, a análise dos resultados de GARCH possibilita a compreensão de que o conjunto de pontos de mudança detectados por CF em torno da observação y_{750} são na verdade anomalias de volatilidade.

Os métodos apresentados se mostram complementares no que diz respeito à detecção de eventos abrangendo diferentes tipos e à compreensão adequada de suas diferentes semânticas. Isso indica uma demanda pela combinação dos resultados de diferentes métodos de detecção de eventos. Tal combinação visa unificar informações relevantes extraídas sobre os eventos em uma série temporal. Essas informações, por sua vez, podem se encontrar compartimentalizadas pela aplicação e análise isoladas dos resultados de cada método de detecção ou de um número inadequadamente limitado de métodos.

Harbinger atende a essa demanda através da execução de seu módulo de combinação, cujo resultado pode ser observado na Fig. 2(d). Nesta podemos observar de maneira unificada todos os eventos detectados pelos métodos discutidos, assim como suas respectivas semânticas. Neste sentido, *Harbinger* propõe uma visão completa e abrangente das detecções, de maneira a evitar problemas como a negligência ou a identificação errônea de eventos de diferentes tipos. Tais problemas podem afetar o processo de tomada de decisão e trazer prejuízo a aplicações que dependem do controle e monitoramento de eventos.

4.4. Avaliação dos módulos de avaliação e comparação

Para apreciação dos módulos de avaliação e comparação de resultados de detecção do *Harbinger*, selecionamos os conjuntos de dados GECCO e *Yahoo*. Para isso, primeiramente foram definidas e executadas instâncias do *Harbinger* para detecção de eventos das séries temporais referentes ao GECCO. Eventos foram detectados nestas séries pelo método SCP, tendo seus parâmetros definidos tanto manualmente quanto por otimização. A Tab. 1 apresenta resultados produzidos pelo módulo de avaliação da qualidade das detecções.

Tabela 1. Avaliação de eventos detectados no GECCO usando SCP (manual versus otimização)

	Tp		pH		Redox		Leit		Trueb	
	Man.	Otim.	Man.	Otim.	Man.	Otim.	Man.	Otim.	Man.	Otim.
VP	6	<u>17</u>	<u>72</u>	48	<u>72</u>	56	<u>12</u>	0	<u>72</u>	48
FP	105	<u>82</u>	192	<u>98</u>	208	<u>85</u>	192	<u>77</u>	191	<u>71</u>
F1	0.94	<u>0.95</u>	0.93	<u>0.96</u>	0.92	<u>0.96</u>	0.91	<u>0.95</u>	0.93	<u>0.97</u>
Precisão	0.93	<u>0.94</u>	0.86	<u>0.93</u>	0.85	<u>0.94</u>	0.86	<u>0.95</u>	0.87	<u>0.95</u>
Revocação	0.95	<u>0.96</u>	<u>1.00</u>	0.98	<u>1.00</u>	0.99	0.95	<u>0.95</u>	<u>1.00</u>	0.98

Na Tab. 1, podemos avaliar a qualidade das detecções de SCP nas séries selecionadas de GECCO através de diferentes métricas: verdadeiros positivos (VP), falsos positivos (FP), F1, precisão e revocação. A escolha das métricas mais adequadas para avaliação da qualidade das detecções dependem diretamente do objetivo de uma determinada aplicação. Por exemplo, se o principal interesse de um especialista é que nenhum evento deixe de ser detectado, as métricas VP e revocação podem ser mais adequadas, levando a conclusão de que o SCP obteve melhores resultados nas séries *pH*, *Redox* e *Trueb*. Por outro lado, se o seu principal interesse é fazer detecções mais precisas, as métricas FP e precisão podem ser mais adequadas, levando a conclusão de que melhores resultados foram obtidos nas séries *Tp* e *Trueb*. Geralmente, é necessário balancear e analisar diferentes métricas para avaliar de maneira mais abrangente a qualidade de detecções de eventos. O módulo de avaliação do *Harbinger* facilita essa análise, permitindo que diferentes métricas sejam passadas por parâmetro, possibilitando a sua comparação.

Particularmente, são apresentadas na Tab. 1 métricas para avaliação de eventos detectados pelo SCP, tendo seus parâmetros definidos tanto manualmente quanto por otimização. Esta otimização foi conduzida com base na métrica F1, que é calculada considerando tanto a precisão quanto a revocação. Como esperado, as métricas F1 avaliadas com a otimização dos parâmetros de SCP foram maximizadas assim como os falsos positivos foram reduzidos, resultando em detecções de eventos mais precisas quando comparadas a execução do método com a definição manual dos parâmetros.

Mesmo assim, é possível que a avaliação quantitativa das detecções de eventos através de métricas de qualidade não seja suficiente para a compreensão do potencial de desempenho de um método de detecção. Neste caso, uma avaliação qualitativa por visualização gráfica dos seus resultados pode contribuir, trazendo uma perspectiva complementar. O módulo de avaliação do *Harbinger* possibilita a avaliação qualitativa do potencial de desempenho de um método de detecção através da produção de gráficos como os apresentados na Fig. 3. São apresentados os eventos detectados na série *Trueb* pelo SCP com parâmetros definidos manualmente (Fig. 3(a)) e por otimização (Fig. 3(b)).

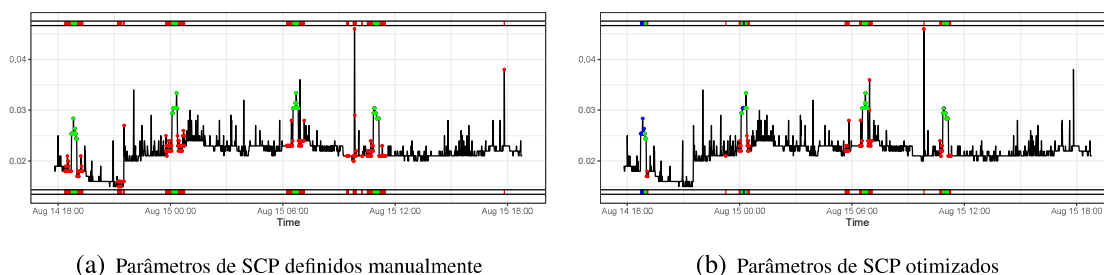


Figura 3. Detecções de eventos pelo método SCP na série *Trueb* em GECCO.

Analisando-se a Fig. 3(a) junto com suas respectivas métricas na Tab. 1, temos que a quantidade de eventos detectados por SCP corresponde ao total de eventos presentes na série, representando uma revocação perfeita. No entanto, a definição manual dos parâmetros resultou numa quantidade relativamente grande de falsos positivos, sendo superior a 10% do total de observações presentes na série. Mesmo assim, podemos perceber que os falsos positivos detectados se encontram nos arredores dos eventos reais, indicando que o método SCP é sensível não só às mudanças de comportamento em séries temporais, mas também aos seus efeitos inerciais. Por sua vez, com base na Fig. 3(b), observamos que de fato a otimização dos parâmetros resultou em maior precisão e menor número de falsos positivos. Entretanto, nem todos os eventos foram detectados.

Frequentemente, entretanto, especialistas precisam lidar com o problema de escolher o método de detecção de eventos mais adequado a uma série temporal e à aplicação. Essa escolha pode ser uma tarefa complexa, dado que existem diversos métodos de detecção na literatura, cada um especializando-se em séries temporais que apresentam propriedades estatísticas diversas ou fazendo suposições sobre a distribuição dos seus dados. Além disso, como discutido na Seção 4.3, muitas vezes a natureza dos eventos contidos na série temporal não é conhecida. Neste caso, é possível que eventos de determinados tipos sejam negligenciados ou identificados erroneamente como resultado da escolha inadequada de um método de detecção. Isso indica, portanto, uma demanda por uma análise comparativa dos resultados de diferentes métodos de detecção de eventos. Tal comparação visa nortear a escolha de métodos de detecção de eventos adequados a uma série temporal e a uma aplicação.

Harbinger atende a essa demanda através da execução de seu módulo de comparação cujo resultado pode ser observado na Tab. 2. É apresentada uma comparação das métricas de qualidade F1 das detecções de eventos produzidas por diferentes métodos de detecção. Podemos observar de maneira unificada os desempenhos dos diversos métodos discutidos com base nas séries temporais contidas nos conjuntos de dados GECCO e *Yahoo*. Do conjunto de dados do *Yahoo* foram selecionados dados reais contidos em *A1* e dados sintéticos com sazonalidade contidos em *A3*.

Ao analisar a Tab. 2, observa-se que para o conjunto de dados GECCO, os métodos AN, decomposição, KNN-CAD e CF apresentaram os melhores desempenhos de detecção. Considerando a metodologia característica destes métodos, isso indica que os eventos em GECCO são em sua maioria anomalias de tendência. Neste caso, métodos especializados em outros tipos de eventos podem não ser os mais adequados. Mesmo assim, é possível que a combinação entre as detecções dos métodos citados apresentem informa-

Tabela 2. Comparação da qualidade das detecções de eventos usando F1

		Decomp.	AN	KNN-CAD	GARCH	EWMA	SCP	CF
GECCO	Tp	0.88	0.88	<u>0.96</u>	0.95	0.68	0.94	0.94
	pH	<u>1.00</u>	<u>1.00</u>	0.96	0.94	0.99	0.93	<u>1.00</u>
	Redox	<u>1.00</u>	0.99	0.96	0.97	0.99	0.92	<u>1.00</u>
	Leit	0.88	0.90	<u>0.96</u>	0.89	-	0.91	0.88
	Trueb	<u>0.99</u>	<u>0.99</u>	0.96	0.91	0.78	0.93	0.95
Yahoo A1	real_1	-	<u>0.99</u>	<u>0.99</u>	0.92	<u>0.99</u>	0.95	0.98
	real_2	-	<u>0.98</u>	<u>0.98</u>	0.93	0.56	0.92	0.93
	real_3	-	<u>1.00</u>	0.98	0.96	<u>1.00</u>	0.99	<u>1.00</u>
	real_4	-	0.98	0.98	0.90	<u>0.99</u>	0.92	<u>0.99</u>
	real_5	-	<u>0.99</u>	0.98	0.96	<u>0.99</u>	0.96	0.98
Yahoo A3	saz_1	<u>1.00</u>	0.99	0.99	0.99	0.76	0.96	0.99
	saz_2	<u>1.00</u>	0.99	0.98	0.92	0.39	0.97	0.99
	saz_3	<u>0.99</u>	<u>0.99</u>	<u>0.99</u>	0.96	0.47	0.93	<u>0.99</u>
	saz_4	-	<u>0.99</u>	<u>0.99</u>	0.96	0.64	0.91	<u>0.99</u>
	saz_5	<u>1.00</u>	-	<u>0.99</u>	0.98	0.59	0.95	<u>0.99</u>

ções mais relevantes, incluindo a semântica de pontos de mudança detectados pelo CF. Uma análise análoga pode ser feita para o conjunto de dados *Yahoo A3*.

Para o conjunto de dados *Yahoo A1*, melhores desempenhos de detecção foram obtidos por AN e EWMA. Como ambos são baseados em médias móveis, isso pode indicar que os eventos em *Yahoo A1* são afetados pela inércia do fenômeno gerador dos dados. Além disso, os dois métodos são especializados na detecção de tipos diferentes de eventos, sugerindo que estes podem ser complementares e a sua combinação pode gerar melhor compreensão dos eventos. Observamos também que o método de decomposição encontrou dificuldades ao detectar eventos nos dados reais do *Yahoo*. Isso pode indicar que não foi possível identificar uma tendência linear clara nos dados em *Yahoo A1*, o que muitas vezes pode ocorrer em dados reais oriundos de fenômenos de alta complexidade.

Os resultados sugerem que a análise comparativa dos métodos de detecção possibilitada pelo módulo de comparação do *Harbinger* é capaz de nortear a escolha de métodos mais adequados. Além disso, pode indicar a necessidade de combinação de métodos determinados com diferentes metodologias. O resultado de comparação apresentado na Tab. 2, assim como as suas respectivas instâncias e as combinações de instâncias resultantes podem ser registrados em sistemas de gerenciamento do ciclo de vida de modelos.

5. Conclusão

Foi apresentado *Harbinger*, um *framework* para integração e análise de métodos de detecção de eventos. Este foi avaliado com base em três conjuntos de dados sintéticos e reais. Foi possível constatar que o *Harbinger* auxilia a compreensão da natureza dos eventos detectados, provendo uma visão abrangente e unificada das detecções de diferentes métodos combinadas de maneira a evitar problemas como a negligência ou a identificação errônea de eventos que podem trazer prejuízo à aplicações que dependem do monitoramento de eventos. Além disso, os resultados sugerem que o *Harbinger* é capaz não só de auxiliar na escolha e aplicação de métodos adequados a um conjunto de dados, como também de indicar a necessidade de combinação das detecções de métodos determinados.

A extensão deste estudo nos leva ao problema de predição de eventos que se baseia não só em técnicas de detecção, mas também em soluções nas áreas de análise e

predição de séries temporais e aprendizado de máquina. Dessa forma, *Harbinger* visa auxiliar futuros esforços que devem ser direcionados ao desenvolvimento de um modelo classificador das observações coletadas. Este modelo pode então ser adequado à análise de séries temporais coletadas em fluxo contínuo.

Referências

- E. Calikus, S. Nowaczyk, A. Sant’Anna, and O. Dikmen. No free lunch but a cheaper supper: A general framework for streaming anomaly detection. *Expert Systems with Applications*, 155: 113453, October 2020. ISSN 0957-4174.
- R. Carmona. *Statistical Analysis of Financial Data in R*. Springer-Verlag New York, 2014.
- H. Chen and N. Zhang. Graph-based change-point detection. *The Annals of Statistics*, 43(1): 139–176, February 2015. ISSN 0090-5364.
- D. De Paepe, S. V. Haute, B. Steenwinckel, F. De Turck, F. Ongena, O. Janssens, and S. V. Hoecke. A generalized matrix profile framework with support for contextual series analysis. *Engineering Applications of Artificial Intelligence*, 90, 2020. ISSN 0952-1976.
- B. Eriksson, P. Barford, R. Bowden, N. Duffield, J. Sommers, and M. Roughan. Basisdetect: A model-based network event detection framework. In *Proceedings of the 10th ACM SIGCOMM*, page 451–464, New York, NY, USA, 2010. Association for Computing Machinery.
- M. Gupta, J. Gao, C. Aggarwal, and J. Han. Outlier Detection for Temporal Data: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 2014. ISSN 1041-4347.
- V. Guralnik and J. Srivastava. Event Detection from Time Series Data. In *Proceedings of the Fifth ACM SIGKDD*, KDD ’99, pages 33–42, New York, NY, USA, 1999. ACM. ISBN 978-1-58113-143-7.
- G. Lu, Y. Zhou, C. Lu, and X. Li. A novel framework of change-point detection for machine monitoring. *Mechanical Systems and Signal Processing*, 83, 07 2016.
- H. Raza, G. Prasad, and Y. Li. EWMA Model Based Shift-Detection Methods for Detecting Covariate Shifts in Non-Stationary Environments. *Pattern Recogn.*, 48(3):659–669, March 2015. ISSN 0031-3203.
- F. Rehbach, S. Moritz, S. Chandrasekaran, M. Rebolledo, M. Friese, and T. Bartz-Beielstein. GECCO 2018 Industrial Challenge: Monitoring of drinking-water quality. pages 1–7, 2018.
- R. Salles, K. Belloze, F. Porto, P.H. Gonzalez, and E. Ogasawara. Nonstationary time series transformation methods: An experimental review. *Knowledge-Based Systems*, 164:274–291, 2019.
- D. Silva, A. Simões, C. Cardoso, D. E. M. Oliveira, Y. Souto, L. E. G. Vignoli, R. Salles, H. S. C. Jr, A. Ziviani, E. Ogasawara, F. C. Delicato, P. F. Pires, H. L. C. P. Pinto, L. Maia, and F. Porto. A conceptual vision toward the management of machine learning models. In *Proceedings of the ER Forum 2019, Salvador, Bahia, Brazil*, volume 2469, pages 15–27, 2019.
- J.-I. Takeuchi and K. Yamanishi. A unifying framework for detecting outliers and change points from time series. *IEEE Transactions on Knowledge and Data Engineering*, 18(4):482–492, 2006.
- P. D. Talagala, R. J. Hyndman, K. Smith-Miles, S. Kandanaarachchi, and M. Muñoz. Anomaly Detection in Streaming Nonstationary Temporal Data. *Journal of Computational and Graphical Statistics*, 29(1):13–27, 2020.
- Yahoo! Webscope. *Labeled anomaly detection dataset*. March 2015.
- L. Xiong, C. Jiang, C. Xu, K. Yu, and S. Guo. A framework of change-point detection for multivariate hydrological series. *Water Resources Research*, 51, 09 2015.