

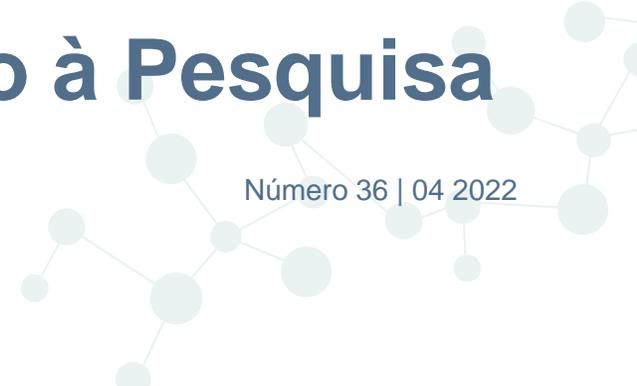
Série dos Seminários de Acompanhamento à Pesquisa

Deviance mining of online processes with non-atomic events

Autor:

Lucas Seixas Jazbik

Série dos Seminários de Acompanhamento à Pesquisa



Número 36 | 04 2022

Deviance mining of online processes with non-atomic events

Autor:

Lucas Seixas Jazbik

Orientadora: Fernanda Baião



CRÉDITOS:

SISTEMA MAXWELL / LAMBDA
<https://www.maxwell.vrac.puc-rio.br/>

Organizadores: Fernanda Baião / Soraida Aguilár

Layout da Capa: Aline Magalhães dos Santos



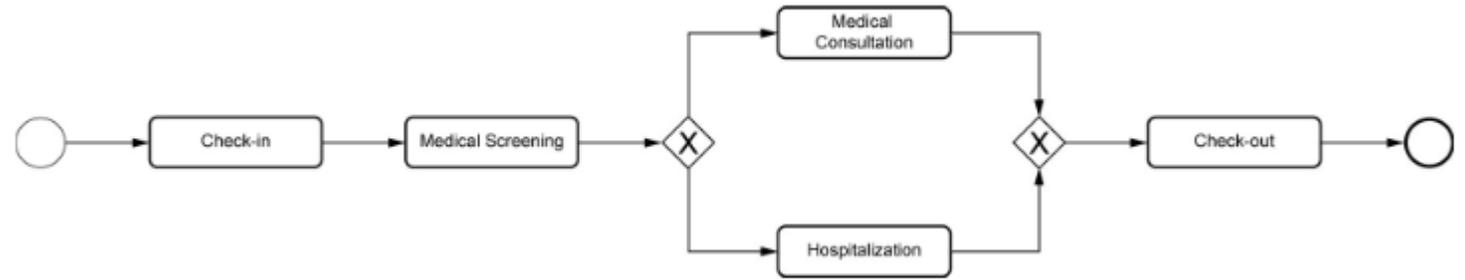
PUC
RIO

Deviance mining of online processes with non-atomic events

- Mestrado;
- Início: 2020.1;
- Orientadora: Fernanda Baião;
- Área de Concentração: Operações e Negócios em Engenharia.

DEI
DEPARTAMENTO
DE ENGENHARIA
INDUSTRIAL

Introdução

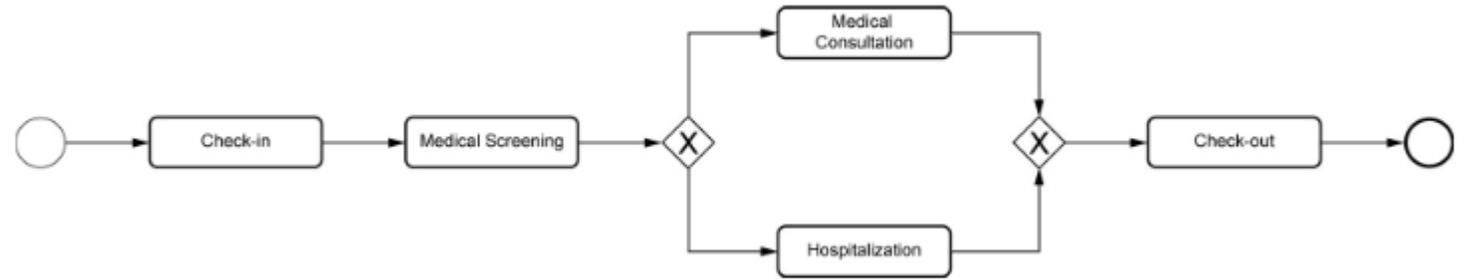


- Modelo de um processo;
 - É possível entender todo o processo apenas observando ele?
 - É uma boa representação?
 - Há flexibilidade nas execuções do processo, em relação ao modelo?
 - Mineração de Processos;
- Indicadores de Performance;
 - Avaliar o resultado do processo;
 - Variável de Classe;
 - Ex: Tempo de permanência, pacientes que não retornaram...
- Desvios na execução do processo;
 - Deviance Mining [1];
- Eventos não atômicos [2];
 - Mais um desafio para **análise do processo** e dos **desvios**;
 - Literatura costuma ignorar essa característica.

[1] - Nguyen, H., Dumas, M., La Rosa, M., Maggi, F. M., & Suriadi, S. (2016). Business process deviance mining: review and evaluation.

[2] - Bernardi, M. L., Cimitile, M., Francescomarino, C. D., & Maggi, F. M. (2014, August). Using discriminative rule mining to discover declarative process models with non-atomic activities. In *International Symposium on Rules and Rule Markup Languages for the Semantic Web*.

Introdução



- Processo com início e fim bem definido;
 - Existem exceções → online scenarios [3];
- Janelas de Análise;
 - Recortes do processo para aplicação de process mining;
- A maior parte dos trabalhos de Mineração de Processo utilizam processos com início e fim bem definidos.

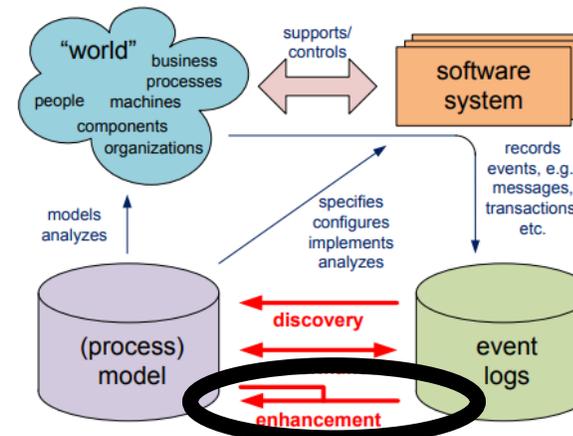
Objetivo

- Identificar desvios e explicar o efeito desses desvios;
 - Foco em cenários online com eventos não atômicos;
- Criação de um método que permita a aplicação de deviance mining em cenários **online** com eventos **não atômicos**;
 - Diferenciais em relação à literatura atual.

Revisão Bibliográfica

- Mineração de Processos [4];
 - Se situa entre data mining e a modelagem e análise de processos;
- Extrair conhecimento, utilizando logs de eventos;
 - Log de evento → Gravações das execuções reais do processo;
 - A mineração de processos se baseia nas execuções reais para gerar conhecimento;
 - É o processo como ele é e não como imaginamos que seja;

- Instâncias;
- Eventos;
- Atributos;
- Modelos.



Revisão Bibliográfica

- Deviance Mining [1];
 - Análise dos desvios em processos e suas consequências;
- Sob a ótica da performance;
 - Esse desvio afeta positivamente ou negativamente o resultado do processo?
- Identificar qual evento (ou conjuntos de eventos) que caracterizam um desvio e qual a consequência desse desvio;
 - A utilização de treatment learning ajuda nesse objetivo.

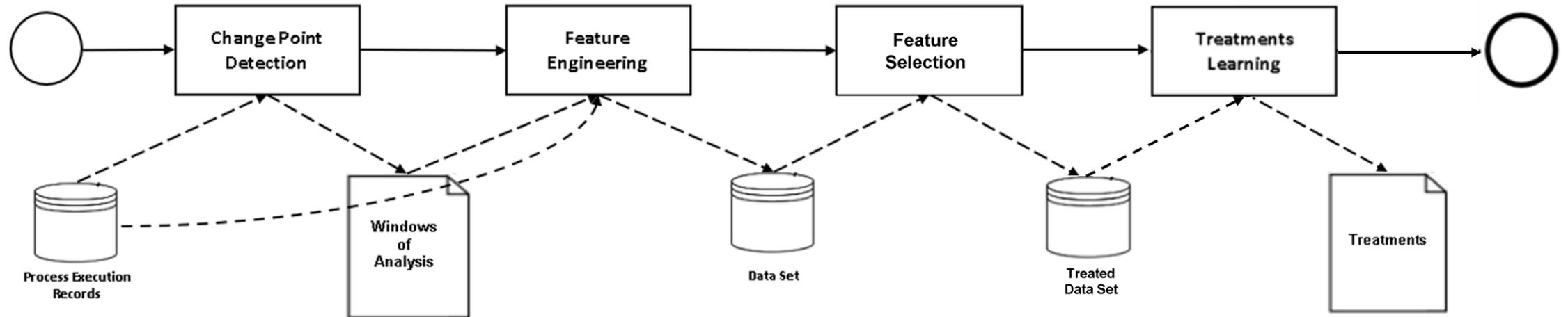
Revisão Bibliográfica

- Treatment Learning [5];
 - Descoberta da menor cesta de atributos (tratamento) que diferencia os casos de uma classe A dos casos de outra classe B;
- O peso das classes pode ser definido de acordo com a variável de classe;
 - Classe de peso maior → Bons resultados;
 - Classe de peso menor → Maus resultados;

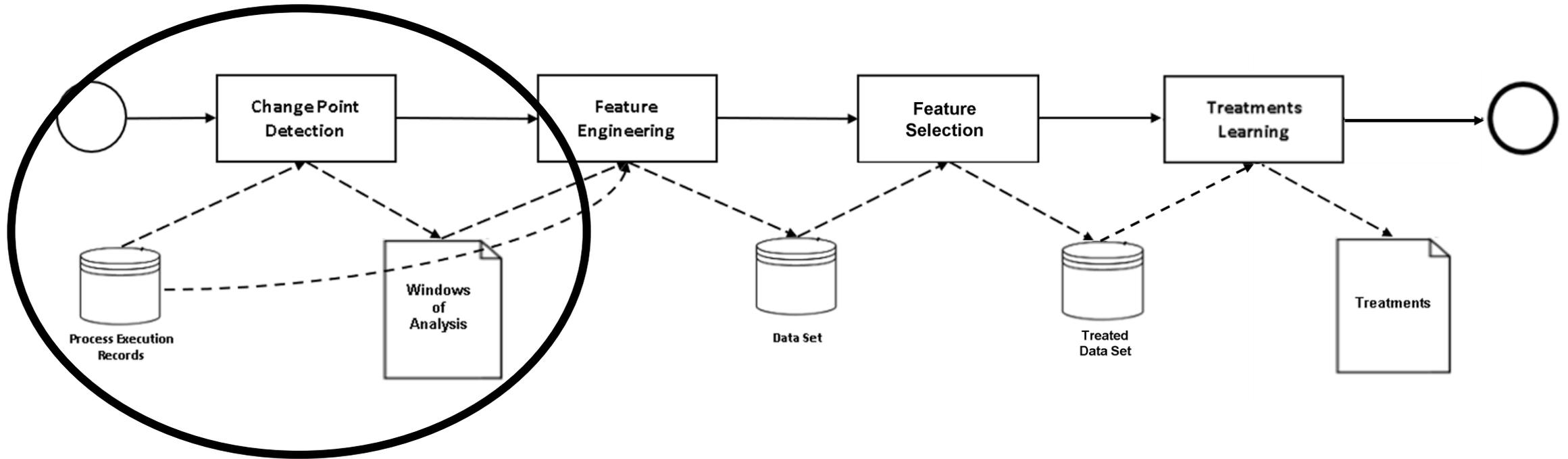
Revisão Bibliográfica

- Para aplicação de treatment learning é necessário que os dados estejam organizados como um data set;
 - Dados de um processo estão organizados em log de eventos;
- Limitação das propostas na literatura
 - Aplicação de treatment learning em processos;
 - Tratar cenários online;
 - Tratar eventos não atômicos;

Metodologia proposta



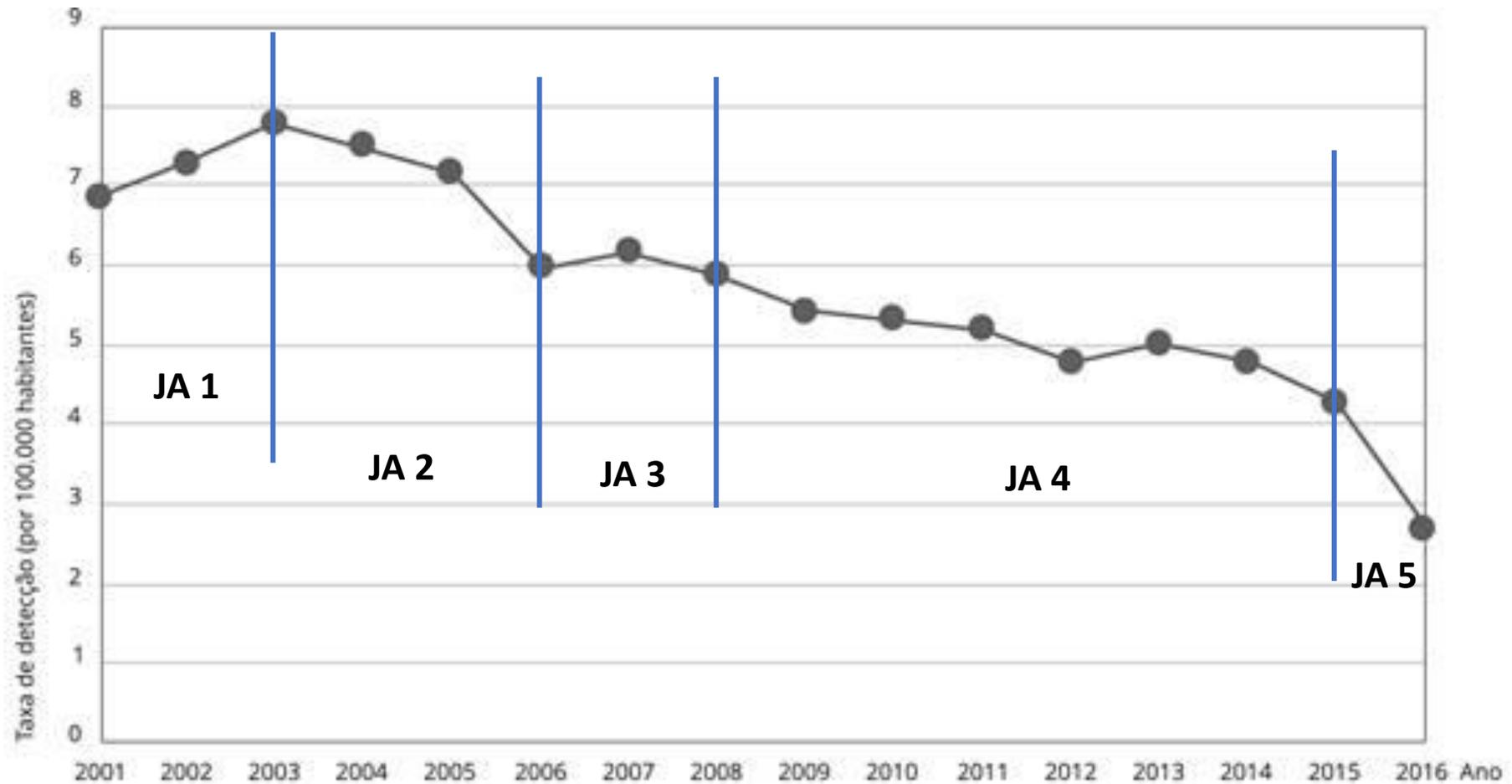
Metodologia proposta



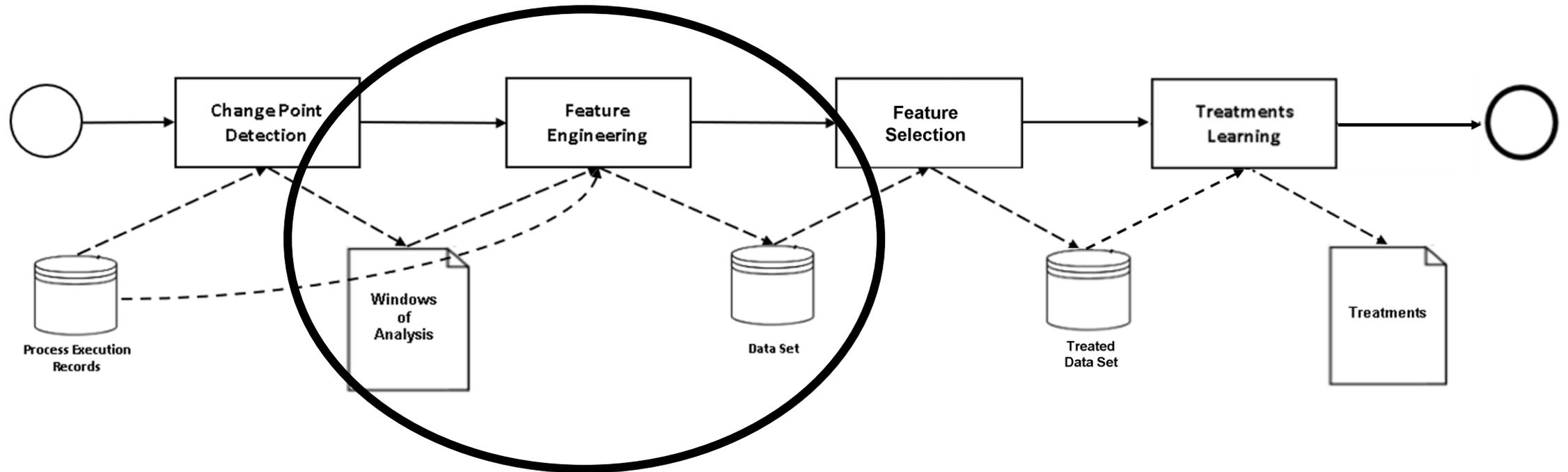
Change Point Detection

- Online Scenarios → Necessidade de definir janelas de análise;
 - Como o processo não tem fim, é necessário realizar cortes no tempo para entender como o que ocorreu afetou o resultado até aquele momento;
- Change Point Detection;
 - Identificar um ponto de mudança estrutural na curva da variável de classe;
 - Definir as janelas de análise, a partir desses pontos;
- As janelas de análise se tornam as instâncias do processo;
 - Tudo que ocorreu dentro de uma janela de análise será analisado para entender o resultado no change point que determina seu fim.

Change Point Detection



Metodologia proposta

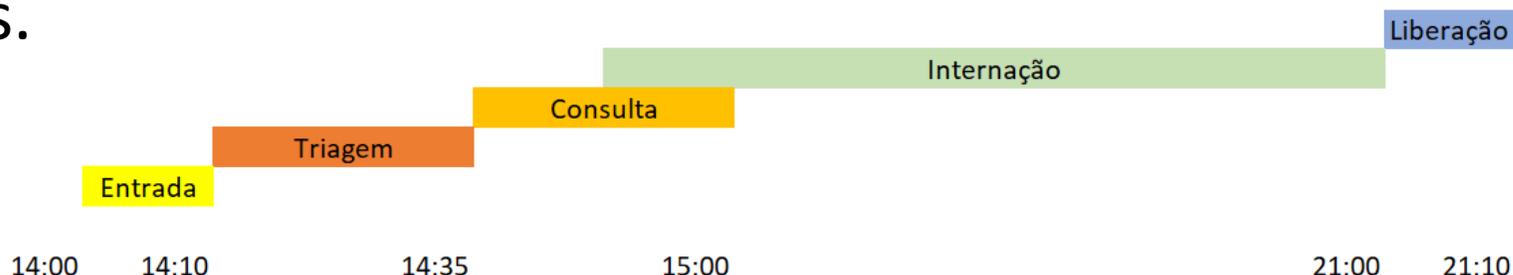


Feature Engineering

- Treatment learning requer um dataset com atributos representativos;
- Eventos não atômicos → Dimensão temporal é importante;
 - Criar atributos que representem a **duração** e a **simultaneidade** de ocorrência entre cada par de eventos;

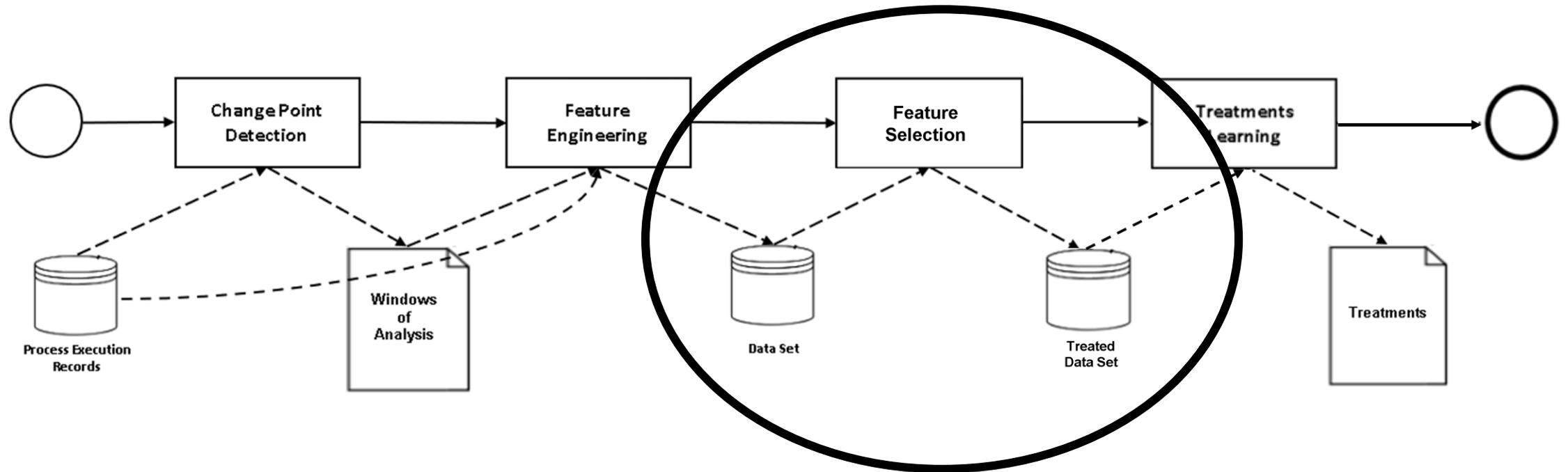
Feature Engineering

- Exemplos:
 - Como o tempo de consulta interfere na volta dos pacientes? Uma triagem rápida afeta o retorno dos pacientes?
 - Uma consulta ocorrendo simultaneamente à internação aumenta as chances do paciente não precisar retornar?
- Essas perguntas só serão respondidas se tivermos variáveis que representem a duração e a simultaneidade de ocorrência dos eventos.



Paciente	Tempo de Triagem	Tempo de Consulta	Coocorrência Consulta-Internação	...	Retorno
104	5 min	25min	0 min	...	Não
105	10 min	35 min	15 min	...	Não
106	7 min	5 min	0 min	...	Sim

Metodologia proposta



Feature Selection

- A etapa de Feature Engineering pode formar uma grande quantidade de atributos;
- O algoritmo de treatment learning utilizado apresenta perda de efetividade quando aplicado a uma quantidade muito grande de variáveis;
 - Se faz necessário a etapa de Feature Selection;
- O problema de feature selection é custoso para ser resolvido por uma busca exaustiva;
 - Se temos m atributos, precisamos buscar $\binom{m}{d}$ combinações, para todo $d \leq m$;
 - É comum a utilização de metaheurísticas para solucionar esse problema;

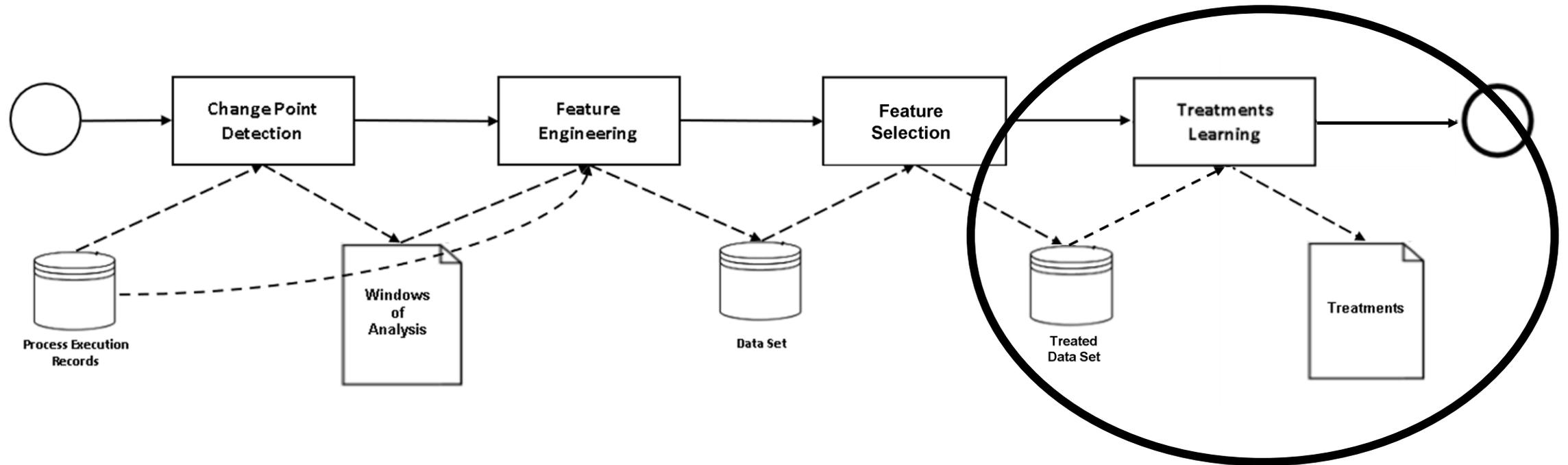
Feature Selection

- Não há exemplos na literatura da solução do problema de feature selection para aplicações de treatment learning;

Meta-heurística Proposta:

1. Do treatment learning with data set x and $r = (1, \dots, d)$, until $J(x)_{r+1} < J(x)_r$. Or, until $J(x)_{r+1} = J(x)_r$ and the large treatment with $r+1$ is smaller than $r+1$;
2. Record $\text{best_lift} = J(x)$ and $\text{best_r} = r$;
3. Delete from data set x the attributes that are not present in the treatments discovered with the best_r ;
4. Repeat until find a iteration that $J(x)_r < \text{best_lift}$ for all $r = (1, \dots, d)$ or until find a best_r which the treatments encompass all attributes presents in data set x .

Metodologia proposta



Treatment Learning

- Após a codificação do processo em um data set tratado, se torna viável a realização do treatment learning;
- Tal etapa irá identificar quais tratamentos levam a um melhor (ou pior) desfecho;
 - Identificação dos desvios do processo que levam às mudanças mais significativas do desfecho.

Cenário de Aplicação

- O método proposto foi aplicado a um cenário de combate à COVID-19;
- O processo estudado foi: A aplicação das medidas não farmacêuticas de combate à COVID-19 utilizadas por diversos países ao longo do tempo;
 - Eventos: medidas não farmacêuticas;
 - Variável de classe (desfecho): aceleração (segunda derivada) do número de contaminados;
 - Instâncias do processo: As janelas de análise de cada país;
 - Os pontos de quebra utilizados são identificados na curva de aceleração de contaminados de cada país;

Pais	Janela de análise	No School closing	Recommend school closing	Partial school closing	Required school closing	...	Taxa de variação ao final do período de análise
BRA	JA1	60%	25%	10%	5%	...	Ruim
EUA	JA1	30%	30%	20%	20%	...	Médio
BRA	JA2	0%	0%	40%	60%	...	Bom

Avaliação do Método Proposto;

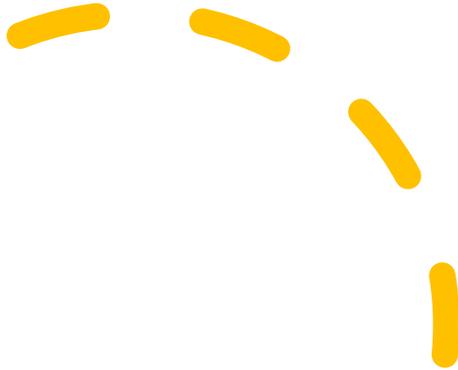
- A viabilidade do método proposto foi avaliada por meio de um estudo de caso;
 - Aplicação do método ao cenário da COVID-19;
 - Experimento utilizando a variável tempo de duração e experimento utilizando a variável de simultaneidade;
 - Análise da efetividade da meta-heurística;
 - Análise da importância do uso de atributos não-atômicos;
 - Avaliação de um especialista.

Publicações

- Richetti, P. H. P., Jazbik, L. S., Baião, F. A., & Campos, M. L. M. (2022). Deviance mining with treatment learning and declare-based encoding of event logs. *Expert Systems with Applications*, 187, 115962.
 - Desenvolvimento e avaliação da codificação, via linguagem *Declare*, de processos visando a realização de deviance mining por treatment learning;
- Artigo em R1: “Deviance mining of online processes with non-atomic events” – Periodico: “International Journal of Data Science and Analytics” – Percentual Scopus: 83%



PUC
RIO



Muito Obrigado!