

Detecção de Anomalias em Robôs Industriais

1st Andre

Centro de Informática - UFPE

Recife, Brazil

@cin.ufpe.br

2nd Bianca

Centro de Informática - UFPE

Recife, Brazil

@cin.ufpe.br

3rd Caio

Centro de Informática - UFPE

Recife, Brazil

@cin.ufpe.br

4rd Rodrigo

Centro de Informática - UFPE

Recife, Brazil

@cin.ufpe.br

Abstract—Breve apresentação do contexto do trabalho, problema/tema a ser abordado, soluções existentes, método proposto e resultados obtidos.

I. INTRODUÇÃO

Deverá abordar a motivação, justificativa e principais contribuições do trabalho em questão. Alguns pontos que devem ser abordados de forma **breve**, são:

- **Contextualização do problema:** Qual o problema/tema que está sendo investigado?
- **Relevância prática do problema:** Por que ele é interessante?

II. ANÁLISE DE DADOS E FEATURE ENGINEERING

A. Análise Exploratória dos Dados

1) *Análise Exploratória Estrutural:* A análise exploratória estrutural foi conduzida com o objetivo de compreender a organização e os tipos de atributos presentes no conjunto de dados. Para garantir a integridade da avaliação durante as etapas de *feature selection* e modelagem de forma a prevenir a tomada de decisão enviesada e o vazamento de dados, adotou-se uma estratégia de particionamento dos dados a priori. A divisão em subconjuntos de Treino, Validação e Teste foi realizada imediatamente após a coleta. O Conjunto de Validação, composto por 10% dos dados Normais e 50% dos dados de Falha (de cada cenário), foi utilizado para a Análise Exploratória de Dados (EDA) e demais etapas subsequentes.

O dataset é composto por séries temporais de sensores inerciais, contendo variáveis numéricas contínuas associadas às medições de acelerômetro, giroscópio e magnetômetro, nos eixos X, Y e Z; uma coluna temporal que representa o instante de coleta de cada amostra e uma variável indicativa da condição de operação do robô, utilizada como rótulo para diferenciar entre comportamento normal e anômalo. Particularmente, os dados anômalos são separados de acordo com o tipo de falha, sendo elas:

- Hitting Platform, que lida com colisão contra a plataforma;
- Hitting Arm, colisão contra o próprio braço robótico;
- Extra Weight, esforço mecânico dado por peso extra;
- Earthquake, vibração estrutural externa.

Cada observação corresponde a uma leitura dos sensores em um determinado instante de tempo, sendo os dados organizados de forma sequencial. A variável temporal é representada por valores inteiros de alta resolução, indicando registros em escala de nanosegundos, reforçando o caráter temporal do problema. No entanto, com o objetivo de facilitar a interpretação e a análise de integridade temporal, os valores foram convertidos para milissegundos, mantendo a proporcionalidade entre as amostras e reduzindo a magnitude numérica dos registros.

Durante a análise estrutural, foi identificada a presença de uma coluna textual associada à identificação do sensor, que por apresentar valor constante em todas as observações e não agregar informação discriminativa ao problema de detecção de anomalias, essa variável foi removida do conjunto de dados. Adicionalmente, foi realizada a verificação de duplicidade nos registros temporais, não sendo identificadas amostras com timestamps repetidos, o que indica consistência na indexação temporal das observações.

2) *Informações básicas:* O conjunto de dados normais analisado é composto por 874.937 amostras e 11 atributos após a etapa inicial de pré-processamento. Todas as variáveis de sensores são numéricas contínuas do tipo *float64*, enquanto a variável alvo é representada por valores inteiros binários. O volume total de memória ocupado por esse dataset é de aproximadamente 73 MB.

Por outro lado, o conjunto de dados anômalos é composto por 49.185 amostras com os 11 atributos contínuos numéricos e a variável alvo binária. Com um volume total de memória de aproximadamente 10MB. Adicionalmente, o dataset possui um atributo *scenario* que indica o tipo de falha a qual esse registro é associado, segundo a proporção destacada na tabela abaixo.

TABLE I
QUANTIDADE DE ANOMALIAS POR TIPO DE FALHA

Cenário	Quantidade	Proporção
Hitting Platform	14.967	0.3043
Hitting Arm	11.924	0.2424
Earthquake	11.409	0.2320
Extra Weight	10.885	0.2213

3) *Análise de Duplicatas, Valores Faltantes e Outliers*: A integridade dos dados foi avaliada sob três perspectivas: redundância de registros, continuidade temporal e presença de valores extremos.

a) *Duplicatas e Valores Nulos Explícitos*: Uma varredura inicial no conjunto de dados brutos não identificou linhas duplicadas ou valores nulos (*NaN*) explícitos nas leituras dos sensores, tanto nos dados normais quanto anômalos. O sistema de aquisição registrou continuamente as 9 variáveis do IMU (acelerômetros, giroscópios e magnetômetros) sem falhas de escrita evidentes.

b) *Análise de Continuidade Temporal*: Embora não houvesse *Nan*s no arquivo original, a análise do intervalo entre amostras revelou um problema crítico de integridade temporal. A frequência nominal de coleta, era de 10 Hz (com *Sampling Rate* de 100 ms). No entanto, observou-se um *Jitter* (desvio padrão do intervalo de tempo) relevante de aproximadamente 22 ms, com 429 *Gaps* temporais significativos variando entre 2 ms e 342 ms, onde essa quantidade de perda de pacotes foi considerada quando o valor do *Gap* foi maior que 2x a média, sendo categorizada como quebra de continuidade.

Para padronizar a frequência e facilitar o janelamento das séries temporais nas etapas seguintes, foi utilizada a técnica de *Resampling*, impondo uma grade temporal rígida de 100 ms. Nesse contexto, a irregularidade original foi exposta sob a forma de *Gaps* temporais. No conjunto de dados Normal, o alinhamento gerou cerca de 44.766 lacunas onde não havia dados registrados no timestamp esperado.

Para corrigir os problemas mencionados anteriormente sem descartar dados, optou-se pela **interpolação linear** dos valores faltantes gerados pelo *Resampling*. Essa abordagem preservou a tendência do movimento entre os pontos conhecidos, restaurando a continuidade necessária para a extração de *features* de janela deslizante.

Para os conjuntos de dados anômalos, a etapa de regularização temporal com *Resampling* e Interpolação foi aplicada de forma iterativa e independente para cada cenário de falha em detrimento a aplicação direta sobre o conjunto anômalo completo. Essa abordagem foi necessária para prevenir as descontinuidades temporais, visto que os experimentos de falha foram registrados em contextos diferentes e com *Gaps* de minutos a horas entre si e a imposição de uma grade temporal contínua de 10Hz no dataset completo forçaria a criação de milhares de linhas vazias para preencher esses intervalos e a interpolação linear geraria dados sintéticos falsos, o que tornaria inviável alimentar o modelo com esses dados, mesmo com etapas subsequentes de pré-processamento.

c) *Análise de Outliers*: A análise de distribuição, por meio de Boxplots e Histogramas, detectou uma quantidade massiva de *outliers* estatísticos, especialmente nos eixos do acelerômetro. Nos dados de colisão (*Hitting Arm*), a curtoze do eixo Z atingiu valores extremos (> 100), com picos de aceleração de até 10g, muito superiores à faixa média normal de operação ($\pm 1g$). Entretenato, diferente de problemas de regressão clássicos, onde *outliers* são ruídos a serem removidos, neste projeto eles constituem o próprio **sinal de**

interesse (a falha mecânica). Consequentemente, optou-se por **não remover** os *outliers*. Para mitigar o impacto desses valores extremos na normalização dos dados mais a frente, substituiu-se o *StandardScaler* (sensível à média/desvio padrão) pelo *RobustScaler*. Este escalonador utiliza a mediana e o intervalo interquartil (IQR), garantindo que os picos de colisão permaneçam destacados na escala transformada, preservando a assinatura da anomalia para o modelo.

4) *Análise Univariada*: Executar análises entre atributos e alvo de forma univariada com descrição estatística e visualizações de apoio.

5) *Análise Bivariada*: Executar análises semelhantes a II-A4 considerando também a correlação entre atributos.

6) *Análise Multivariada*: Aplicar técnicas de análise multivariada como *pairplots*, análise de grupos de interesse e criação de variáveis derivadas além de visualizações para apoio.

B. Pré-processamento dos dados

Apresentar ações referentes ao pré-processamento de dados a exemplo de:

1) *Tratamento de Valores Faltantes*: Em caso de valores faltantes, descrever o tratamento executado para sanar os problemas com justificativas bem fundamentadas.

2) *Tratamento de Outliers*: Em caso de *outliers*, apresentar o tratamento executado com devidas justificativas.

3) *Detecção e Tratamento de Duplicadas*: Avaliar se há presença de duplicadas no dataset utilizado e tratá-las de acordo.

4) *Feature Scaling*: Utilizar ferramentas de normalização como *StandardScaler* ou *MinMaxScaler* para normalizar os atributos.

5) *Encoding de Variáveis Categóricas*: Utilizar ferramentas adequadas para tratamento de variáveis. Indicar qual técnica foi utilizada e a motivação.

C. Divisão dos Dados

Justificar e apresentar a divisão de dados entre Treino, Validação e Teste com justificativas a respeito da técnica utilizada.

D. Feature Engineering

Deverá abordar a seleção e extração de *features* caso cabível. Em caso de utilização de técnicas de redução de dimensionalidade, justificar a escolha do método e configurações.

III. MODELAGEM

Nesta seção deverá ser feita uma breve revisão dos algoritmos selecionados. Além disso, justificar a escolha do algoritmo juntamente com hiperparâmetros a serem testados. Importante explicitar qual(is) foi(ram) o(s) espaço(s) de busca utilizado, bem como a técnica escolhida para tunagem de hiperparâmetros.

A. Modelo A

1) *Conceitos Básicos*:

- 2) Justificativa:
- 3) Espaço de Busca:
- 4) Hiperparâmetros Selecionados:

B. Modelo B

- 1) Conceitos Básicos:
- 2) Justificativa:
- 3) Espaço de Busca:
- 4) Hiperparâmetros Selecionados:

C. Modelo C

- 1) Conceitos Básicos:
- 2) Justificativa:
- 3) Espaço de Busca:
- 4) Hiperparâmetros Selecionados:

IV. ANÁLISE E COMPARAÇÃO DE RESULTADOS

Deverá conter as métricas que foram utilizadas para a análise juntamente com revisão de conceito e justificativas. É fundamental comparar os resultados obtidos entre os diferentes modelos treinados. Além disso, é interessante utilizar ferramentas estatísticas e/ou testes de hipótese quando cabível.

V. CONCLUSÃO E DISCUSSÃO

Explicar os principais achados ao longo do trabalho bem como vantagens e limitações de métodos e/ou algoritmos selecionados. Além disso, apresentar principais *insights* extraídos e potenciais trabalhos futuros.

MATERIAL DE APOIO

TABLE II
EXEMPLO DE TABELA

Trabalhos	Método	Vantagens	Desvantagens
Trabalho A	Método X	A	D
Trabalho B	Método Y	G	J

VI. MODELO DE AMEAÇA (SE APLICÁVEL)

Utilizar figuras e/ou algoritmos e/ou equações para descrever os comportamentos da análise.

Na Eq. (1) é apresentado um exemplo de equação com a equação de uma reta.

$$y = ax + b \quad (1)$$

Na Fig. 1 é apresentado um exemplo de figura com a logo do centro de informática.

No Alg. 1 é apresentado um exemplo de algoritmo.

REFERÊNCIAS

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.



Fig. 1. Logo do centro de informática.

Algorithm 1 Algoritmo com legenda

Require: $n \geq 0$

Ensure: $y = x^n$

```

 $y \leftarrow 1$ 
 $X \leftarrow x$ 
 $N \leftarrow n$ 
while  $N \neq 0$  do
    if  $N$  is even then
         $X \leftarrow X \times X$ 
         $N \leftarrow \frac{N}{2}$  ▷ Exemplo de comentário
    else if  $N$  is odd then
         $y \leftarrow y \times X$ 
         $N \leftarrow N - 1$ 
    end if
end while

```
