

Detecção de anomalias em tubulações de transporte de petróleo e gás utilizando long short-term memory networks

Marcus da Silva Carr

Carlos Mauricio Ruiz Diaz

Oscar Mauricio Hernandez Rodriguez

Laboratório de Escoamento Multifásicos Industriais (LEMI) - EESC/USP

marcus_carr@usp.br; carlosruiz978@usp.br; oscarmhr@sc.usp.br.

Objetivos

As plataformas de petróleo offshore são monitoradas por operadores em centros de controle. Com o aumento contínuo do número de poços em exploração, há uma demanda crescente para otimizar as operações de gerenciamento e aumentar a robustez desses sistemas. Uma abordagem promissora para enfrentar esse desafio é a implementação de sistemas inteligentes, que podem oferecer ferramentas avançadas de análise para os operadores. Este estudo tem como objetivo avaliar a eficácia das Long short-term memory networks (LSTM) na detecção de eventos indesejados em operações offshore. Utilizaremos dados de temperatura, pressão e vazão de um único poço real como base para este estudo.

Métodos e Procedimentos

A base de dados 3W (VARGAS et al., 2019), fornecido pela Petrobras, foi escolhida como fonte de dados para este estudo. A versão 1.1.1 do dataset abrange um total de 1.978 eventos, categorizados em 1.019 de fonte real, 939 de fonte simulada e 20 desenhados manualmente. Cada evento consiste em uma série temporal composta por nove variáveis, oito delas representando características numéricas do escoamento (pressão, temperatura e vazão) e a nona servindo como

rótulo de classificação. Esse rótulo indica se o momento no tempo está em um regime normal, em um regime transitório de anomalia ou em um regime permanente de anomalia, abrangendo nove classificações distintas que capturam diferentes padrões de fluxo. O processo de extração de dados preservou as características autênticas da informação original, que inclui a presença de valores ausentes, congelados e discrepantes, atribuídos a problemas nos sensores instalados ou na comunicação com o sistema de controle.

A presença de dados nulos e a variação de amplitude entre diferentes variáveis e tipos de anomalias demandam um processo de pré-processamento para adaptá-los aos modelos de aprendizado de máquina a serem utilizados. Neste estudo, foram aplicadas quatro etapas para tratar os dados: (1) redução da taxa de amostragem, através da média, para filtrar e reduzir a quantidade de dados para a modelagem; (2) imputação mediante interpolação linear para lidar com valores nulos; (3) estandardização para aproximar os dados a uma escala semelhante; e (4) janelamento das séries temporais no formato adequado para alimentar o modelo de detecção.

O modelo preditivo foi estruturado com 3 camadas, a primeira denominada camada de entrada LSTM com 50 unidades (HOCHREITER; SCHMIDHUBER, 1997),

seguida por uma camada densa com 100 unidades, seguida por uma camada densa de saída com 9 unidades (número de classes). A função de perda foi a "categorical_crossentropy" e o otimizador foi o "Adam". Esse modelo foi então treinado com dados de 1 único poço, que foi escolhido de modo a possuir boa disponibilidade de dados e equilíbrio entre as suas classes. 80% dos dados foram usados para treinamento, e 20% para validação. Ditos valores foram selecionados após o desenvolvimento de modelos preliminares.

Na etapa da validação do desempenho do modelo foram utilizados os dados escolhidos para validação para calcular a matriz de confusão, a fim de determinar o nível de exatidão do modelo desenvolvido para realizar a classificação de diferentes classes.

Resultados

O poço escolhido para a realização da modelagem foi o de número identificador igual a 5. Esse poço tem disponível 119 arquivos para análise, onde 81 são da classe normal (272 horas de medidas) e 38 são da classe instabilidade de fluxo (75 horas de medidas). A escolha desse poço foi feita devido ao fato do poço 5 possuir uma grande quantidade de dados, e um melhor equilíbrio entre classes comparado aos outros.

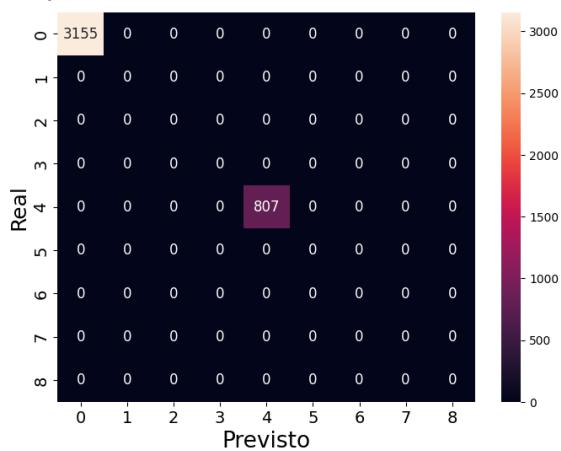


Figura 1: Matriz de confusão obtida.

Em seguida, se aplicou as 4 etapas de processamento: redução da taxa de

amostragem, imputação, estandardização, janelamento das séries temporais. Esses dados transformados foram utilizados para treinar o modelo composto por camadas LSTM e densas.

Finalmente, os dados de validação foram usados para calcular uma matriz de confusão, visível na Figura 1. Os resultados indicam boa performance com acurácia de 100% para a classificação dos dados normais e com anomalia de instabilidade de fluxo.

Conclusões

Este estudo alcançou seu objetivo ao avaliar o desempenho das long short-term memory networks (LSTM) na detecção de eventos indesejados em operações offshore, utilizando dados de temperatura, pressão e vazão de um único poço de petróleo real. Os resultados obtidos mostram que a LSTM conseguiu identificar as diferentes classes com êxito. Contudo, os dados analisados representam um cenário idealizado pela etapa de seleção, tendo em conta que a dinâmica dos eventos pode variar ao longo do tempo e ter nuances em sua intensidade que não foram representadas no dataset.

Para futuras pesquisas, está previsto aplicar esta metodologia em casos que envolvam tipos de anomalias mais complexas, buscando avaliar a sua capacidade de detecção em cenários com condições ainda mais críticas. Além disso, a exploração de diferentes modelos, como variações da LSTM, e a otimização dos hiperparâmetros podem ser estratégias para aprimorar ainda mais o desempenho do modelo.

Referências

- VARGAS, Ricardo Emanuel Vaz et al. A realistic and public dataset with rare undesirable real events in oil wells. *Journal of Petroleum Science and Engineering*, v. 181, p. 106223, 2019.
- HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long Short-Term Memory. *Neural Computation*, v. 9, n. 8, p. 1735–1780, nov. 1997.