



Detecção de anomalias em poços produtores de petróleo usando aprendizado de máquina

Fernandes Júnior, W.

Vargas, R.E.V.

Komati, K. S.

Gazolli, K. A. S.

23-26/nov/2020

Introdução

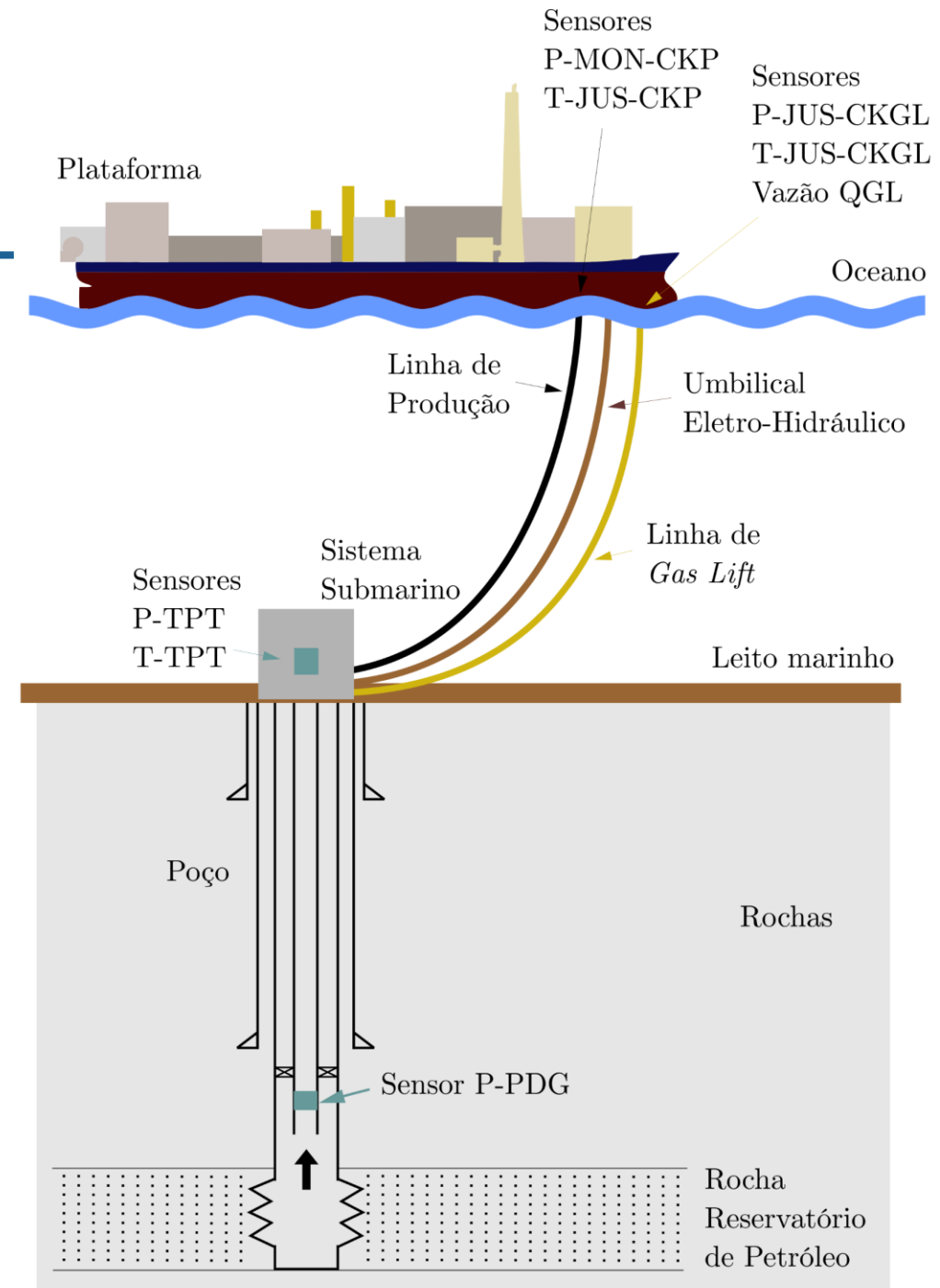


- No Brasil, a produção de petróleo e gás natural em janeiro de 2020 foi de 3,168 MMbbl/d e 139 MMm³/d, respectivamente. Essa produção foi proveniente de 7.227 poços, sendo 649 marítimos e 6.558 terrestres (ANP, 2020).
- A produção do pré-sal correspondeu a 66,4% desse total e foi oriunda de 119 poços marítimos, o que equivale a uma média de produção de aproximadamente 18 Mbbl/d.



Introdução

- Um poço de petróleo é uma estrutura perfurada no solo em etapas que formam um telescópio invertido e munida com equipamentos e sensores que permitem o fluxo de petróleo e gás da rocha reservatório de petróleo até a superfície (Guo, 2011)
- O óleo e o gás fluem para a plataforma através de tubulações e equipamentos submarinos.



Introdução



- Durante a produção de petróleo é possível a ocorrência de eventos indesejados denominados anomalias, que podem provocar impactos financeiros significativos (Vargas, 2019).
- É importante que o processo de produção de petróleo seja monitorado a fim de detectar e classificar anomalias. Uma possível solução é a aplicação de estatísticas multivariadas e métodos de aprendizado de máquina para detecção e classificação de anomalias (Qin, 2012).



Abordagem Proposta



- Um processo de classificação pode ser dividido em cinco etapas: coleta de dados, pré-processamento, extração de características, aplicação da técnica de classificação e avaliação de desempenho (Kadhim, 2019).
- A base de dados utilizada neste trabalho foi publicada por Vargas et al. 2019 e é intitulada 3W. Cada instância é composta por oito variáveis (oito séries temporais) provenientes de sensores de sistemas de produção de petróleo. Para instância existe uma variável adicional que é um vetor de rótulos no nível de observação que estabelece até três períodos em cada instância de qualquer tipo: normal, transiente de anomalia e estado estável de anomalia.

Abordagem Proposta

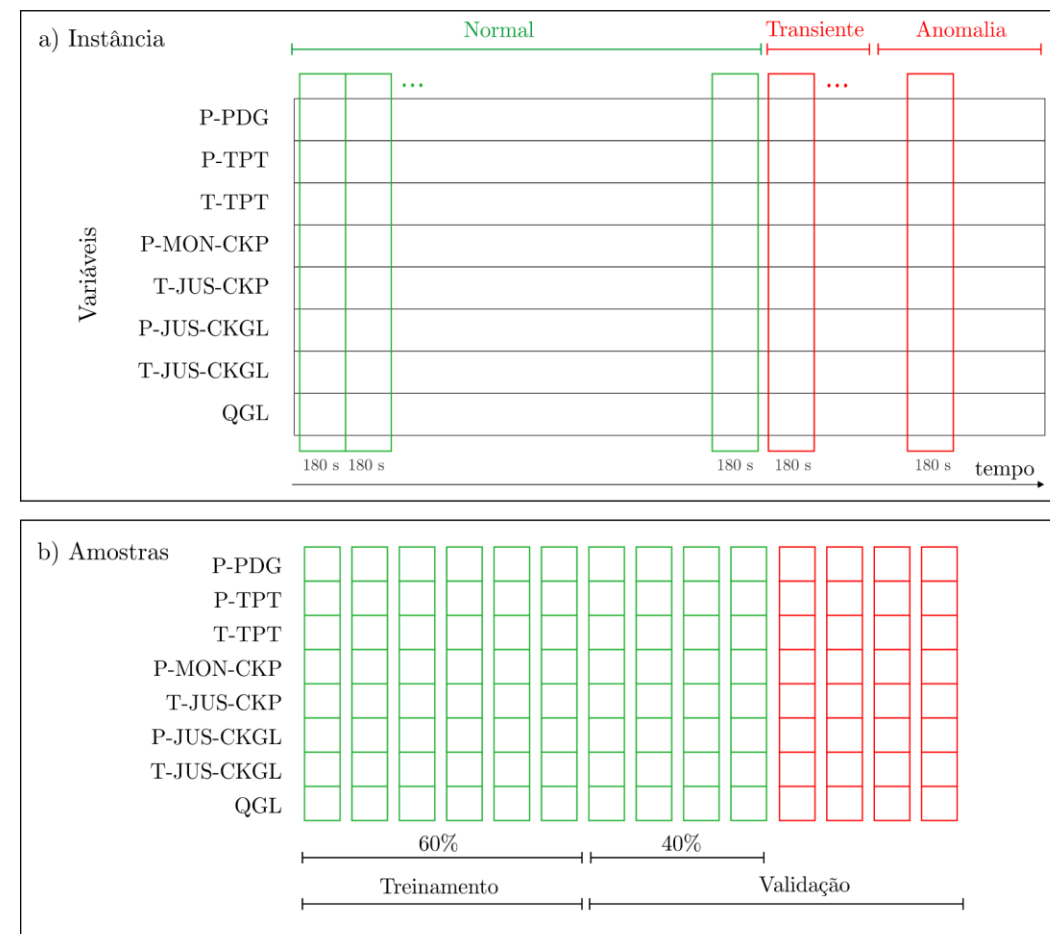


- A base de dados utilizada é pública e contém 1.984 instâncias de séries temporais da produção de poços de petróleo.
- Essas instâncias estão separadas em operação em condições normais e anomalias (organizadas em oito classes).
- A base pode ser utilizada tanto para detecção quanto para classificação de anomalias em poços de petróleo. Nesse trabalho foi realizada a detecção de anomalias.



Abordagem Proposta

- Na etapa de pré-processamento foi realizada amostragem das instâncias com janela deslizante com geração de até 15 amostras com 180 observações cada. Dos períodos normais as primeiras observações foram utilizadas para treinamento (60%) e as últimas, para teste (40%). Dos períodos rotulados como anomalias, as observações foram utilizadas apenas para teste. Os dados foram normalizados.



Abordagem Proposta

- A partir de cada amostra de séries temporal, foram extraídas e utilizadas como características a mediana, média, desvio padrão, variância, máximo e mínimo para cada variável.

Extração de Características

Usando a biblioteca **tsfresh** foram extraídas 6 características para cada série temporal de cada amostra:

P-PDG	→	Mediana	Média	Desvio Padrão	Variância	Máximo	Mínimo
P-TPT	→	Mediana	Média	Desvio Padrão	Variância	Máximo	Mínimo
T-TPT	→	Mediana	Média	Desvio Padrão	Variância	Máximo	Mínimo
P-MON-CKP	→	Mediana	Média	Desvio Padrão	Variância	Máximo	Mínimo
T-JUS-CKP	→	Mediana	Média	Desvio Padrão	Variância	Máximo	Mínimo
P-JUS-CKGL	→	Mediana	Média	Desvio Padrão	Variância	Máximo	Mínimo
T-JUS-CKGL	→	Mediana	Média	Desvio Padrão	Variância	Máximo	Mínimo
QGL	→	Mediana	Média	Desvio Padrão	Variância	Máximo	Mínimo

180 s

O vetor de características é formado pelos 48 valores (6 características em cada uma das 8 séries temporais).

Abordagem Proposta



- Classificadores de classe única podem ser utilizados na detecção de padrões raros. Em geral utiliza-se apenas a classe comum (normalidade) no treinamento (*novelty detection*) e nos testes há uma mistura de instâncias normais e anormais (Khan and Madden, 2014).
- Foram escolhidas as seguintes técnicas:
 - *One-class* SVM (OCSVM) – baseado em otimização;
 - Floresta de Isolamento (Iforest) – baseado em busca;
 - *Local Outlier Factor* (LOF) – baseado em densidade;
 - Envelope Elíptico – baseado em estatística.



Abordagem Proposta



- Na última etapa do processo de classificação, é necessário avaliar o desempenho obtido. Essa avaliação é feita por meio de métricas obtidas a partir de uma matriz de confusão, tais como acurácia, revocação, precisão e/ou medida F1.
- Nesse trabalho utilizou-se a medida F1 para comparação de resultados com o trabalho de Vargas (2019).

Experimentos e Resultados



- Os experimentos seguiram as seguintes regras estabelecidas no benchmark de detecção de anomalia:
 - Apenas instâncias reais com anomalias que têm períodos normais maiores ou iguais a vinte minutos foram utilizadas;
 - Múltiplas rodadas de treinamento e teste foram realizadas, sendo o número de rodadas igual ao número de instâncias. Em cada rodada, as amostras utilizadas para treinamento ou teste foram extraídas de apenas uma instância. O conjunto de teste foi composto pelo mesmo número de amostras de cada classe (normalidade e anormalidade);
 - Em cada rodada, precisão, revocação e medida F1 foram computadas (valor médio e desvio padrão de cada métrica) (Vargas, 2019)



Experimentos e Resultados

- Para calibração dos classificadores gerou-se 1.716 diferentes combinações entre classificadores e hiperparâmetros. Na Tabela 1 abaixo constam as métricas de medida F1 e desvio padrão para os melhores casos de cada classificador. Na Tabela 2 constam os resultados obtidos apresentam melhorias em relação aos resultados obtidos originalmente por Vargas (2019).

Tabela 1 – Resultados obtidos

Classificador	F1 (Média)	Desvio Padrão
<i>Local Outlier Factor</i>	0,882	0,126
Floresta de Isolamento	0,743	0,179
Envelope Elíptico	0,664	0,157
<i>One-class SVM</i>	0,567	0,162
<i>Dummy</i> (ingênuo)	0,500	0,000

Tabela 2 – Resultados do benchmark

Classificador	F1 (Média)	Desvio Padrão
Floresta de Isolamento	0,727	0,182
<i>One-class SVM</i>	0,532	0,075

Conclusão



- Os resultados obtidos apresentaram melhoria em relação ao benchmark e estimulam a continuação do trabalho com a experimentação de outras técnicas - tanto para extração de características quanto para utilização de outras famílias de algoritmos para detecção de anomalias, tais como os baseados em redes neurais.



Obrigado!

