

Previsão da produção de um reservatório por série temporal utilizando o algoritmo Random Forest

Isabel F. A. Gonçalves¹, Sinesio Pesco²

Abelardo B. Barreto³

PUC-Rio, Rio de Janeiro, RJ

Thiago M. D. Silva⁴

SLB, Rio de Janeiro, RJ

A previsão precisa do desempenho de um campo é essencial no planejamento e gerenciamento de reservatórios de petróleo. No entanto, prever produção de petróleo ainda é um problema complexo e altamente não linear devido suas propriedades geofísicas e petrofísicas. Isto significa que com poucas variações, as observações podem ser muito diferentes. Na engenharia de reservatórios, a tomada de decisão de um projeto de exploração precisa levar em consideração diferentes algoritmos eficientes para simulação dados, fornecendo cenários robustos e confiáveis, com objetivo de reduzir os riscos. Para administrar a incerteza no processo de caracterização e estimar dados futuros, pesquisadores tem aplicado algoritmos de aprendizado de máquina, capazes de resolver problemas de engenharia com eficiência, extraindo o máximo de informações dos dados observados. Como exemplo, podemos citar o trabalho de Goncalves *et al.* [1], que utilizou o algoritmo Random Forest para prever a produção de óleo de um campo de petróleo um dia a frente. A estratégia utilizada foi similar a de previsão em séries temporais. Neste estudo foram observados bons resultados no conjunto de dados Volve [2]. Este reservatório é localizado no Mar do Norte, Noruega, foi descoberto em 1993, com profundidade de cerca de 2900 m e produção iniciando em 2008, sendo desativado em 2016, com produção acumulada de 63 milhões de barris.

Neste estudo, buscamos expandir os resultados apresentados por Goncalves *et al.* [1]. O algoritmo Random Forest será utilizado para previsão de produção de óleo em um reservatório, utilizando os dados reais do campo Volve. A estratégia utilizada foi semelhante à apresentada por [1], similar a previsão de séries temporais. Tal como em [1], utilizaremos o conjunto de dados do campo Volve para avaliar a robustez e a eficiência do método na previsão de dados dinâmicos. Utilizamos os dados de produção referente ao poço 15/9-F-1 C, contido no campo Volve. Devido ao grande número de erros e imprecisões nas informações disponíveis, foi necessário pré-processar todos os dados, garantindo a não ocorrência de valores nulos ou não numéricos. Os dados finais formam uma sequência ininterrupta de 746 dias corridos de produção de óleo.

Neste estudo, avaliamos diferentes valores para os parâmetros do algoritmo Random Forest, tais como os seguintes: $n_estimators$ como 10, 25 ou 50 e $max_samples$ como 50%, 80% ou 100% do conjunto de dados original. Todos os outros parâmetros são usados como padrão definido pela biblioteca *Scikit-Learn*, disponível em linguagem *Python*. Além disso, definimos dois novos parâmetros: i) *look-back*, que indica o número de passos de tempo passados que o algoritmo utilizará para previsão futura e ii) *forward-days*, que indica o número de passos de tempo a frente que queremos prever. Estrutturamos o conjunto de dados com três valores diferentes para *look-back*: 10, 25 e 50. Assim, para cada *look-back*, usamos GridSearch para encontrar os melhores

¹isabelfagoncalves@gmail.com

²sinesio@puc-rio.br

³abelardo.puc@gmail.com

⁴thiagoomenez@gmail.com

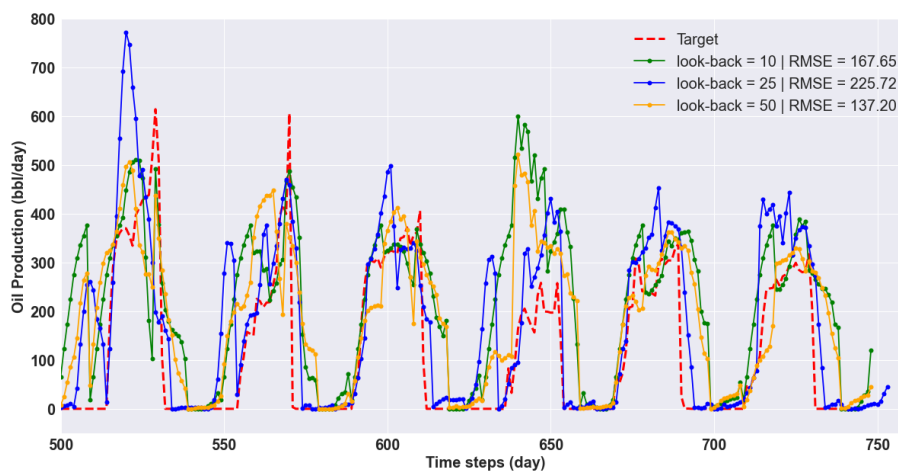


Figura 1: Resultado no conjunto de teste para $look-back = \{10, 25, 50\}$ e $forward-days = 10$

$n_estimadores$ e $max_samples$. Assim, para cada valor $look-back$, o GridSearch constrói nove modelos diferentes e retorna o melhor, de acordo com a métrica $RMSE$. Observe que quando $look-back$ é definido como t , são necessários t passos de tempo passados para fazer uma previsão de um passo à frente, então valores diferentes para $look-back$ geram diferentes quantidades de previsões mesmo que o conjunto de teste sempre tenha o mesmo tamanho.

A principal diferença para o presente trabalho e o apresentado por Goncalves *et al.* [1] é a abordagem metodológica para previsão de dias a frente. Nós avaliamos a sensibilidade da metodologia a diferentes valores de $look-back$, avaliando a capacidade do método em dados reais, disponíveis no conjunto de dados do campo Volve. Nós testamos o desempenho do método aplicando diferentes valores para o parâmetro $look-back$, e prevemos 10 passos de tempo a frente ($forward-days = 10$). Os resultados sugerem que o algoritmo Random Forest pode obter previsões razoáveis ao prever o desempenho do campo de petróleo. O algoritmo não obteve resultados precisos, como comprovam os altos valores de $RMSE$, mas é possível observar que o comportamento de produção, mesmo sendo não linear, foi reproduzido nos resultados, como esperado em uma abordagem de série temporal. Além disso, o valor $look-back = 50$, obteve o melhor resultado. A taxa de amostragem $max_samples = 50\%$ mostrou-se suficiente para o treinamento do algoritmo.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico e Petrobras como parte do projeto TCBR 485.

Referências

- [1] I. F. A. Gonçalves, T. M. D. Silva, A. Barreto e S. Pesco. “Predicting oil field production using the Random Forest algorithm”. Em: **Anais Estendidos do XXXV Conference on Graphics, Patterns and Images**. Natal/RN: SBC, 2022, pp. 134–139.
- [2] Equinor. **Volve dataset**. <https://www.equinor.com/en/news/14jun2018-disclosing-volve-data.html>. [Online; accessed 13-April-2022]. 2008.