

Aprendizado de Máquina para Cálculo de Equilíbrio de Fases em Simulações Composicionais de Reservatórios de Petróleo

Antonioni B. Campos

Programa de Engenharia Elétrica
Universidade Federal do Rio de Janeiro - UFRJ

Junho de 2023



Engenharia
Elétrica
UFRJ

Conteúdo

- 1 Introdução
- 2 Problema
- 3 Trabalhos Relacionados
- 4 Método
- 5 Resultados
- 6 Conclusões
- 7 Referências

Conteúdo

1 Introdução

2 Problema

3 Trabalhos Relacionados

4 Método

5 Resultados

6 Conclusões

7 Referências

Introdução

- A simulação de reservatórios é uma área da engenharia de reservatórios em que modelos de computador são usados para prever o fluxo de fluidos através de meios porosos.
- Os modelos são usado por empresas de petróleo e gás para o desenvolvimento de campos novos ou desenvolvidos onde as previsões de produção são necessárias para ajudar a tomar decisões de investimento [1].

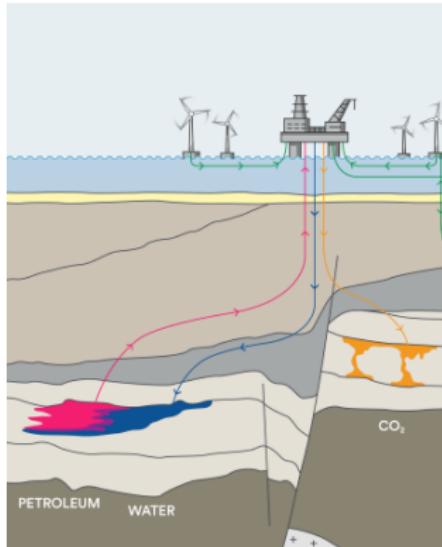


Fig.: Ilustração do caminho do petróleo entre o reservatório até a plataforma [Referência: <https://cssr.no/>]

Introdução

- Uma das principais aplicações da termodinâmica na indústria de petróleo tem é a previsão do comportamento de fases em fluidos de petróleo [2].
- Simulação numérica de escoamento multifásico multicomponente é uma ferramenta essencial para muitas aplicações de subsuperfície [3].
- Um elemento central para tais aplicações é determinar a distribuição de fase de uma dada mistura de fluidos em equilíbrio, também conhecido como **cálculo de flash** [3].

Conteúdo

1 Introdução

2 Problema

3 Trabalhos Relacionados

4 Método

5 Resultados

6 Conclusões

7 Referências

Equilíbrio de Fase

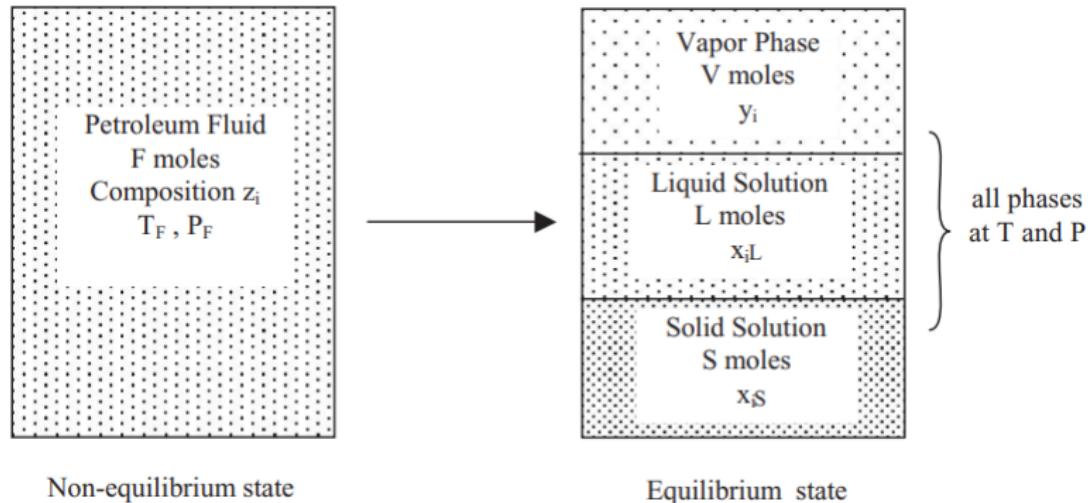
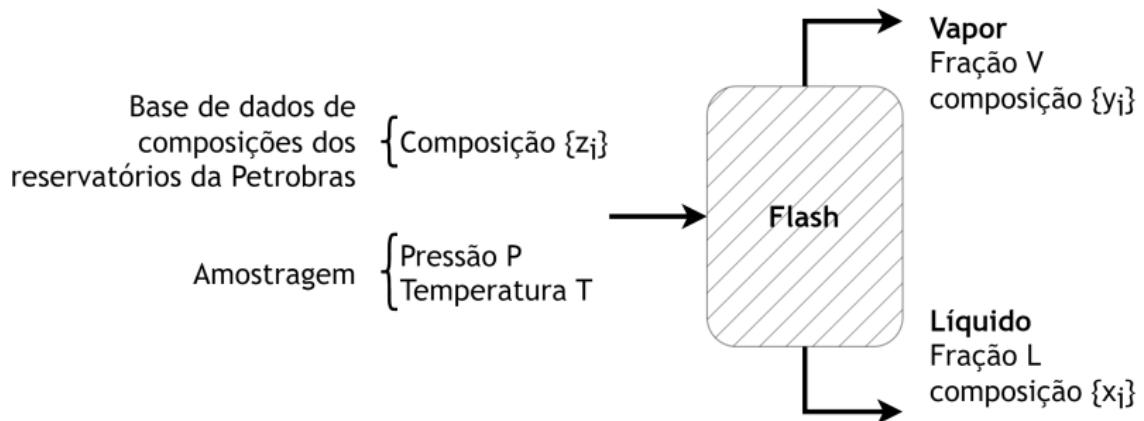


Fig.: Equilíbrio vapor-liquido-sólido [2]

Cálculo de Flash

Equilíbrio Vapor-Líquido (VLE¹)



P	T	N2	CO2	C1	C2	C3	IC4	NC4	IC5	NC5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
12.29	166.33	0.28	2.18	50.33	8.8	5.71	0.96	2.47	0.75	1.25	1.56	2.03	2.46	2.14	1.81	1.56	1.39	1.42	1.19	1.12	0.84	0.75	0.74	0.71	7.54
25.06	775.3	0.28	2.18	50.33	8.8	5.71	0.96	2.47	0.75	1.25	1.56	2.03	2.46	2.14	1.81	1.56	1.39	1.42	1.19	1.12	0.84	0.75	0.74	0.71	7.54
326.34	623.63	0.28	2.18	50.33	8.8	5.71	0.96	2.47	0.75	1.25	1.56	2.03	2.46	2.14	1.81	1.56	1.39	1.42	1.19	1.12	0.84	0.75	0.74	0.71	7.54
313.98	623.63	0.28	2.18	50.33	8.8	5.71	0.96	2.47	0.75	1.25	1.56	2.03	2.46	2.14	1.81	1.56	1.39	1.42	1.19	1.12	0.84	0.75	0.74	0.71	7.54
10.43	775.3	0.28	2.18	50.33	8.8	5.71	0.96	2.47	0.75	1.25	1.56	2.03	2.46	2.14	1.81	1.56	1.39	1.42	1.19	1.12	0.84	0.75	0.74	0.71	7.54
10.9	160.40	0.26	2.22	45.21	7.16	6.02	1.0	3.73	0.93	1.55	1.09	2.5	2.96	2.16	1.78	1.40	1.25	1.4	1.12	1.07	0.93	0.69	0.74	0.67	10.00

Fig.: Flash Contínuo [2]

Cálculo de Flash

- ① Teste de estabilidade de fase pode dizer se uma composição específica em determinada temperatura e pressão é estável em uma única fase ou se dividirá em duas ou mais fases.
- ② Se a composição for instável, os cálculos de divisão de fase devem ser realizadas para calcular as frações molares de cada fase e a composição de cada componente em cada fase.

Os cálculos de equilíbrio de fase geralmente precisam ser executados em cada bloco de grade em cada iteração de cada passo de tempo em uma simulação composicional. Portanto, os custos computacionais dos cálculos de equilíbrio de fase podem ser um grande parte do tempo total de simulação [4].

Diagrama de Fases

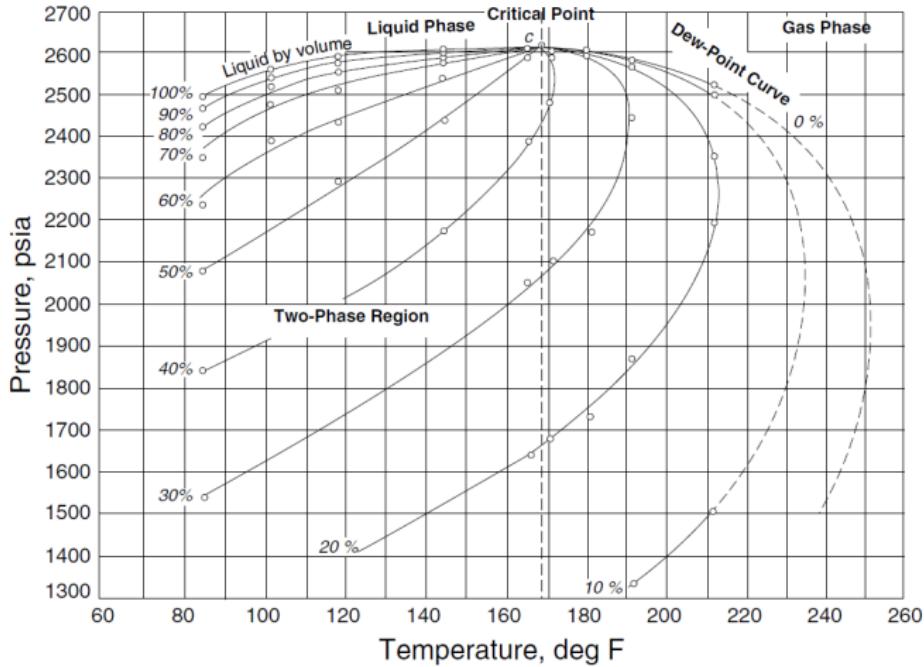


Fig.: Diagrama p-T típico para um sistema multicomponente [5]

Conteúdo

1 Introdução

2 Problema

3 Trabalhos Relacionados

4 Método

5 Resultados

6 Conclusões

7 Referências

Artigo: PTFlash: A vectorized and parallel deep learning framework for two-phase flash calculation [3]

- Foram projetadas duas Redes Neurais para tarefas específicas, uma para inferir a estabilidade da mistura (*classifier*) e a outra para inicializar os coeficientes K_i para o cálculo das composições (*initializer*).
- Arquitetura da rede com três (3) camadas escondidas com 32 neurônios cada. Entrada com P, T e nove (9) componentes e saída binária (estável ou instável).
- A performance final do classificador no conjunto de teste foi de $bce = 0.002$ e $accuracy = 99.93\%$

Trabalhos Relacionados II

Artigo: Practical application of machine learning on fast phase equilibrium calculations in compositional reservoir simulations [4]

- Modelo de Rede Neural MLP (IANN-STAB), várias camadas ocultas são empregadas e a camada de saída inclui tanto o pressão de saturação superior e a pressão de saturação inferior.
- Para o cálculo da divisão de fase, um modelo ANN (IANN-SPLIT) com várias camadas ocultas é empregado para fornecer a valores iniciais de K e a fração molar da fase gasosa para as iterações não lineares.

Model	Hidden units	Training time	Prediction accuracy
Two hidden layers	10–10	222.2 secs	99.57
Single hidden layer	20	395.8 secs	99.50
Single hidden layer	40	415.2 secs	99.52

Fig.: Comparando diferentes arquiteturas do IANN-STAB para três (3) componentes [4]

Conteúdo

1 Introdução

2 Problema

3 Trabalhos Relacionados

4 Método

5 Resultados

6 Conclusões

7 Referências

Dataset

- A Petrobras possui um banco de dados de informações de reservatórios possuindo amostra de composições de diferentes campos. Desse banco de dados, foram extraídos um total de **475 composições**.
- A Equinor possui uma biblioteca *open source* para simulação de processo chamada **neqsim**¹ que foi utilizada para criação dos diagramas de fase para determinação das classes (mistura, líquido e gás).

¹<https://equinor.github.io/neqsimhome/>

Datasets

- Foram calculados os diagramas de fase para cada uma das 475 composições distintas.
- Para cada um dos diagramas de fase associados as composições foram amostrados valores de **pressão** e **temperatura** nas três regiões distintas de fases (mistura, líquido, gás).
- Foram criados **dois datasets** distintos. O primeiro amostrando arbitrariamente 2 (dois) pontos (pressão e temperatura) por região, por composição e o segundo amostrando arbitrariamente 20 (vinte) pontos (pressão e temperatura) por região, por composição.

Pré-Processamento

- Para cada dataset, foram criados subconjuntos de treinamento e teste usando a estratégia de validação cruzada k-fold ($k = 10$) estratificada pelas classes e para o subconjunto de treinamento ainda formaram-se subconjuntos de treinamento e validação.

	Treinamento	Validação	Teste
2 amostras de P, T por composição	2.293	255	284
20 amostras de P, T por composição	22.939	2.549	2.832

Tabela: Quantidade de amostras por subconjunto

- Os dados foram normalizados entre zero (0) e um (1) pelo minimo/máximo de cada variável dos datasets.

Modelos

③ Modelos avaliados

- Classificadores de Vetores de Suporte (SVC)
- Redes Neurais do tipo *Multilayer Perceptron* (MLP)

Parâmetro	
C	0,1
	1,0
	10,0
Kernel	
$\langle x, x' \rangle$	
$(\gamma \langle x, x' \rangle)^d$	
$\exp(-\gamma \ x - x'\ ^2)$	

Tabela: Parâmetros utilizados para o SVC

Modelos

Parâmetro	
Função de Ativação	ReLU Sigmoid Tangente Hiperbólica
Neurônios	8 16 32 64
Camadas	1 ou 2
Fator de Regularização α	0,1 0,0001
Otimizador	Adam SGD
Iterações Máximas	400
Taxa de Aprendizagem	Iniciando com 0,01 e reduzindo numa razão de 10 para cada iteração sem melhora
Critério de Parada	Máxima quantidade de iterações atingida ou três iterações sem melhora

Tabela: Parâmetros utilizados para o MLP

Resultados

- ④ Como figura de mérito para a avaliação do desempenho dos modelos, foi utilizado o **Índice SP** [6] (sum-product index), que é o produto entre a média aritmética e geométrica da sensibilidade (True Positive Rate - TPR) e especificidade (True Negative Rate - TNR).

$$\text{SP Index} = \sqrt{\text{TPR} \cdot \text{TNR}} \cdot \frac{(\text{TPR} + \text{TNR})}{2}$$

- ⑤ Foram calculadas as **matrizes de confusão** para o melhor modelo SVC e MLP baseados no valor do Índice SP para avaliar o desempenho por classe.

Resultados

- ⑥ Em seguida, foi feita uma **análise de relevância** para todas as variáveis de entrada para determinar o impacto de cada variável no desempenho de classificação dos modelos.

- ⑦ Por fim, foi feita uma análise, para alguns componentes selecionados aleatoriamente, dos **diagramas de fase gerados pelos modelos** confrontando com o diagrama de fase gerado pelos cálculos termodinâmicos do neqsim.

Conteúdo

1 Introdução

2 Problema

3 Trabalhos Relacionados

4 Método

5 Resultados

6 Conclusões

7 Referências

Figuras de Mérito

kernel	degree	gamma	C	sensitivity	specificity	sp index
poly	3	-	10,0	72,43 ± 2,55	86,21 ± 1,27	63,62 ± 2,91
	4	-	10,0	71,78 ± 2,48	85,89 ± 1,24	63,02 ± 2,82
	5	-	10,0	70,86 ± 2,47	85,43 ± 1,23	62,07 ± 2,75
			1,0	71,09 ± 2,37	85,54 ± 1,19	62,02 ± 2,70
	6	-	1,0	70,75 ± 2,32	85,37 ± 1,16	61,80 ± 2,59

Tabela: Máximo valores de índice SP, para modelos SVC e para dataset com **6 pares** pressão/temperatura por composição

kernel	degree	gamma	C	sensitivity	specificity	sp index
poly	5	-	10,0	96,09 ± 0,40	98,04 ± 0,20	94,22 ± 0,58
	4	-	10,0	96,07 ± 0,40	98,03 ± 0,20	94,19 ± 0,58
	6	-	10,0	96,05 ± 0,40	98,03 ± 0,20	94,17 ± 0,57
	7	-	10,0	96,04 ± 0,40	98,02 ± 0,20	94,16 ± 0,58
			1,0	95,92 ± 0,41	97,96 ± 0,21	93,97 ± 0,60

Tabela: Máximo valores de índice SP, para modelos SVC e para dataset com **60 pares** pressão/temperatura por composição

Figuras de Mérito

hidden layer sizes	activation	optimizer	alpha	sensitivity	specificity	sp index
32, 16	tanh	Adam	0,000001	$80,45 \pm 2,45$	$90,22 \pm 1,23$	$73,40 \pm 2,98$
64, 16	tanh	Adam	0,000001	$80,34 \pm 3,48$	$90,17 \pm 1,74$	$73,18 \pm 4,14$
	sigmoid	Adam	0,000001	$79,03 \pm 4,54$	$89,52 \pm 2,27$	$71,64 \pm 5,08$
64, 32	tanh	Adam	0,000001	$78,88 \pm 4,22$	$89,44 \pm 2,11$	$71,43 \pm 4,93$
64, 16	tanh	Adam	0,001000	$78,95 \pm 2,23$	$89,47 \pm 1,11$	$71,40 \pm 2,71$

Tabela: Máximo valores de índice SP, para modelos MLP e para dataset com **6 pares** pressão/temperatura por composição

hidden layer sizes	activation	optimizer	alpha	sensitivity	specificity	sp index
32, 16	sigmoid	Adam	0,000001	$97,29 \pm 0,38$	$98,64 \pm 0,19$	$95,98 \pm 0,56$
16, 16	sigmoid	Adam	0,000001	$97,29 \pm 0,34$	$98,64 \pm 0,17$	$95,98 \pm 0,49$
32, 32	sigmoid	Adam	0,000001	$97,19 \pm 0,39$	$98,59 \pm 0,20$	$95,82 \pm 0,58$
64, 16	sigmoid	Adam	0,000001	$97,13 \pm 0,38$	$98,56 \pm 0,19$	$95,73 \pm 0,56$
64, 64	sigmoid	Adam	0,000001	$97,11 \pm 0,37$	$98,55 \pm 0,19$	$95,70 \pm 0,55$

Tabela: Máximo valores de índice SP, para modelos MLP e para dataset com **60 pares** pressão/temperatura por composição

Índice SP

Multilayer Perceptron vs. Support Vector Classifier

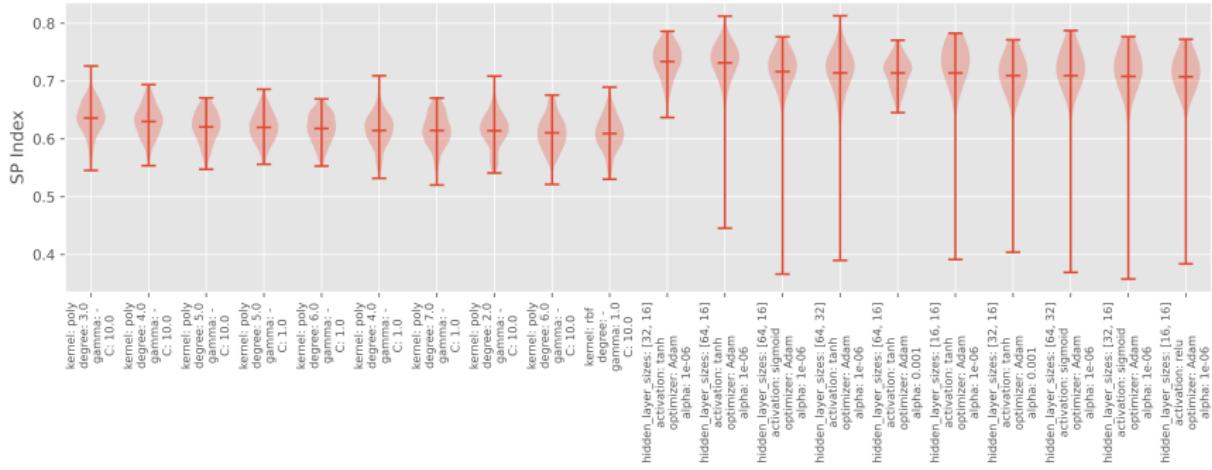


Fig.: Resultados dos índices SP para dataset com **6 pares** pressão/temperatura por composição

Índice SP

Multilayer Perceptron vs. Support Vector Classifier

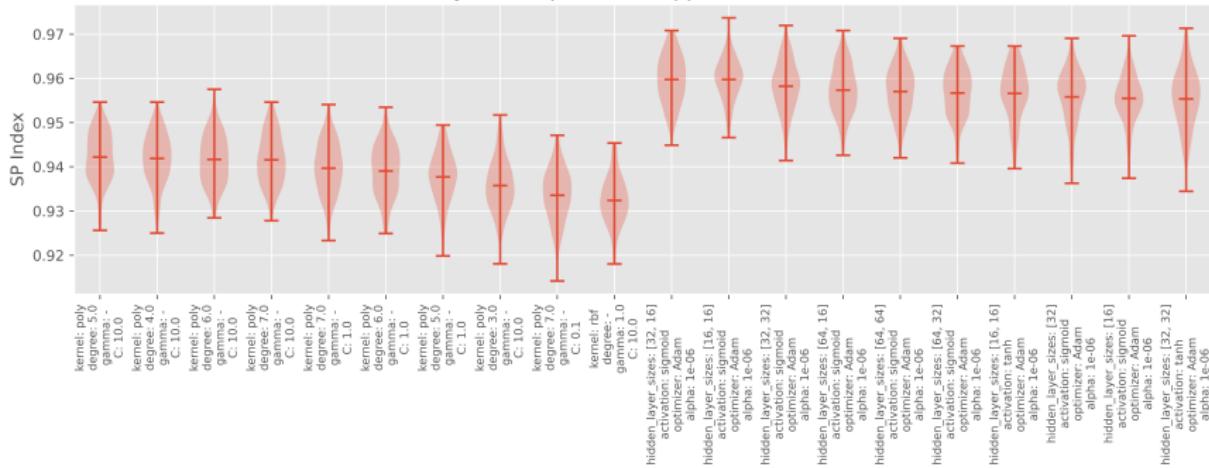


Fig.: Resultados dos índices SP para dataset com **60 pares** pressão/temperatura por composição

Matriz de Confusão

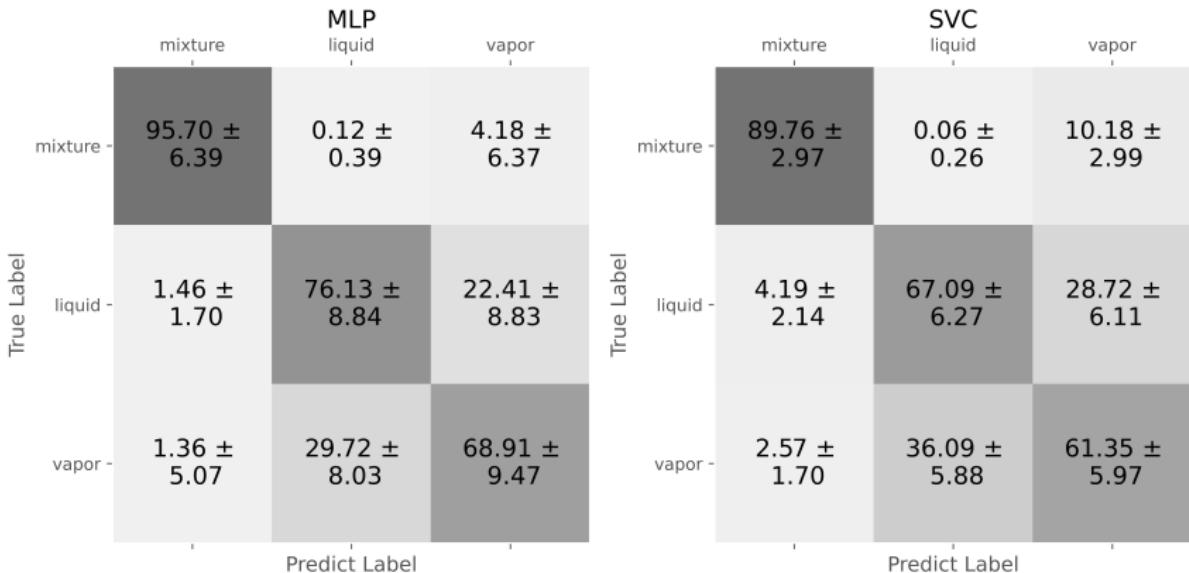


Fig.: Matriz de Confusão para modelos com maior índice SP na média e dataset com **6 pares** pressão/temperatura por composição

Matriz de Confusão

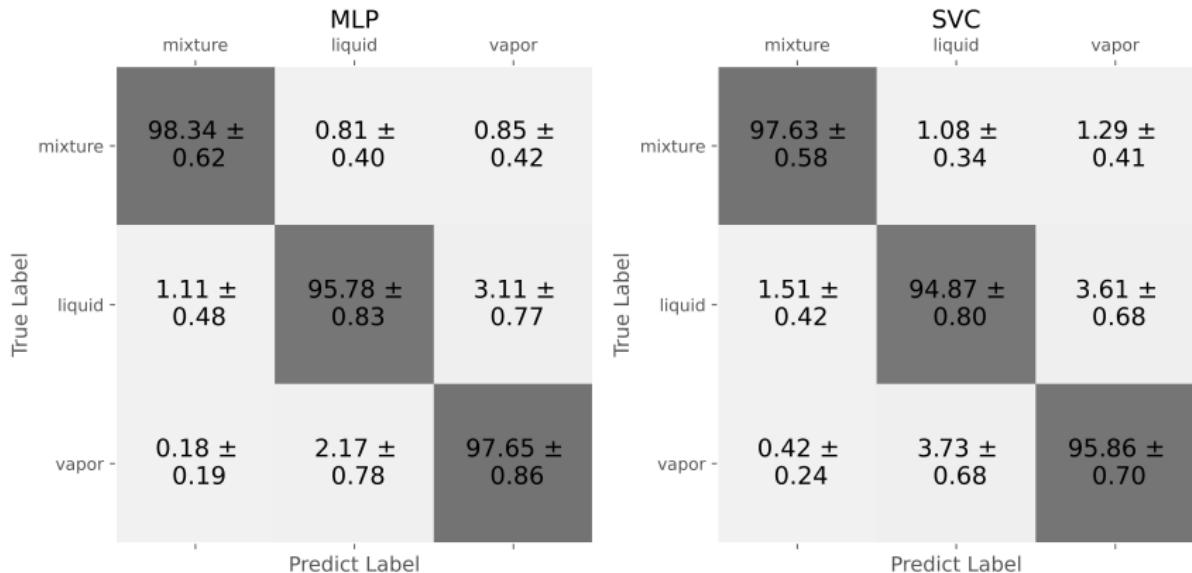


Fig.: Matriz de Confusão para modelos com maior índice SP na média e dataset com **60 pares** pressão/temperatura por composição

Análise de Relevância

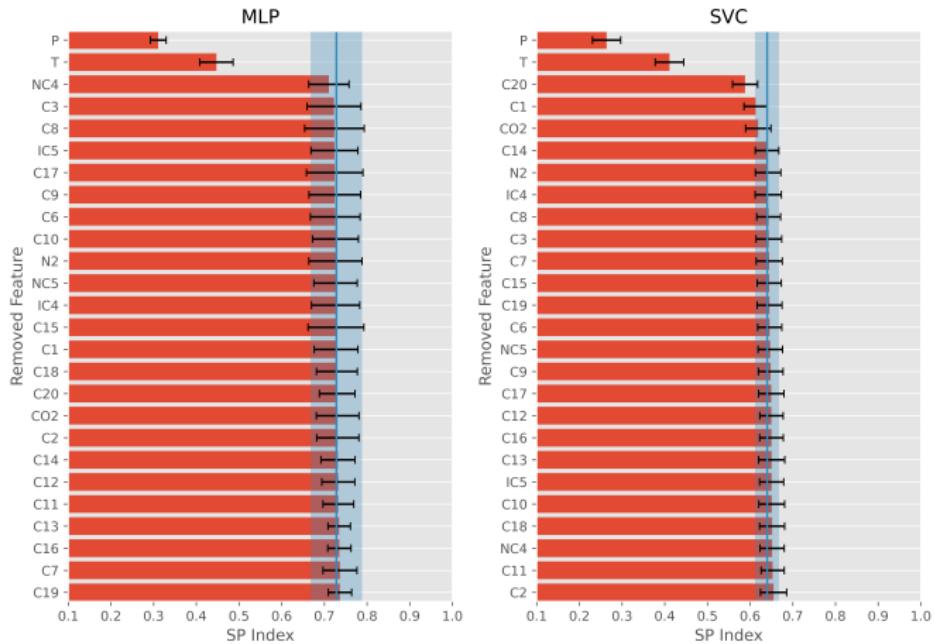


Fig.: Análise de relevância para modelos com maior índice SP na média e dataset com 6 pares pressão/temperatura por composição

Análise de Relevância



Fig.: Análise de relevância para modelos com maior índice SP na média e dataset com **60 pares** pressão/temperatura por composição

Conteúdo

1 Introdução

2 Problema

3 Trabalhos Relacionados

4 Método

5 Resultados

6 Conclusões

7 Referências

Conclusões

- As composições z_i no dataset gerado tem pouca diversidade comparado aos valores gerados de temperatura (T) e pressão (p). Pode ser melhorado estimando uma distribuição $p(z_1, \dots, z_{26})$ das composições encontradas nos reservatórios da Petrobras.
- Os algoritmos clássicos tem dificuldade (precisam de mais iterações para chegar ao resultado) quando a combinação z_i , p e T estão próximas ao ponto crítico. Como esperado, os modelos tiveram mais dificuldade para classificar líquidos e gases nessa região.
- Pressão e Temperatura tem uma influencia bem mais relevante em comparação com as composições do petróleo na classificação de fase.

Conteúdo

1 Introdução

2 Problema

3 Trabalhos Relacionados

4 Método

5 Resultados

6 Conclusões

7 Referências

Referências I

- [1] Nasima Begum, Mohsen Dadashpour, and Jon Kleppe.
Subchapter 1.6 - a case study of reservoir parameter estimation in norne oil field, norway by using ensemble kalman filter (enkf).
In A.K. Moitra, J.R. Kayal, Biplab Mukerji, Jayanta Bhattacharya, and A.K. Das, editors, *Innovative Exploration Methods for Minerals, Oil, Gas, and Groundwater for Sustainable Development*, pages 61–78. Elsevier, 2022.
- [2] M.R. Riazi.
Characterization and Properties of Petroleum Fractions.
ASTM manual series MNL 50. ASTM International, 2005.

Referências II

- [3] Jingang Qu, Thibault Faney, Jean-Charles de Hemptinne, Soleiman Yousef, and Patrick Gallinari.
Ptflash: A vectorized and parallel deep learning framework for two-phase flash calculation.
Fuel, 331:125603, 2023.
- [4] Kun Wang, Jia Luo, Yizheng Wei, Keliu Wu, Jing Li, and Zhangxin Chen.
Practical application of machine learning on fast phase equilibrium calculations in compositional reservoir simulations.
Journal of Computational Physics, 401:109013, 2020.
- [5] T. Ahmed.
Reservoir Engineering Handbook.
Elsevier Science, Houston, Texas, 2010.

Referências III

- [6] Júlio de Castro Vargas Fernandes, Natanael Nunes de Moura Junior, and José Manoel de Seixas.
Deep learning models for passive sonar signal classification of military data.
Remote Sensing, 14(11), 2022.
- [7] Tahsin Ahmed.
Equations of State and PVT Analysis: Applications for Improved Reservoir Modeling: Second Edition.
Elsevier, 03 2016.

Obrigado!