

Estratégia adversarial para análise de similaridade de poços

Henrique M O Carvalho

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil



Renato Assunção

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil



I. INTRODUÇÃO

A indústria do petróleo depende fortemente de pesquisas petroquímicas que obtêm características físicas e químicas de rochas e suas interações com fluidos internos. O processo de extração do petróleo envolve análises primárias de um terreno por meio ferramentas chamadas de sísmicas. Quando as sísmicas entregam informações promissoras sobre uma região, as empresas petrolíferas perfuram poços em busca de petróleo.

Nesses poços, são colhidas informações complementares às sísmicas, conhecidas como perfis do poço. Esses dados são obtidos por ferramentas que percorrem toda a profundidade de um poço e registram dados sobre características dele em intervalos pré-definidos. Esse conjunto de informações viabiliza uma análise minuciosa sobre atributos geológicos do local. Em particular, uma tarefa crucial para a indústria é identificar similaridades entre os poços perfurados utilizando esses perfis.

Essa tarefa tem profunda relevância para a busca de terrenos novos que podem conter petróleo. Comparar um poço em que foi encontrado petróleo a um recentemente perfurado permite inferir, mais precisamente, a probabilidade de esse novo poço também conter o hidrocarboneto. Essa informação auxilia o processo de decisão em diferentes estágios da exploração de um poço, desde perfuração até abandono, uma vez que é um indicativo da viabilidade econômica dele. Investir em pesquisa para a tarefa de similaridade de poços pode ajudar a empresa a poupar gastos, considerando o preço extremamente elevado de cada perfuração, do uso de ferramentas extras para a análise de um poço ou mesmo da escolha de não abandoná-lo.

Os dados de perfis de poço são comumente representados em formato tabular, em que as linhas são medidas individuais feitas por uma série de sensores e essas medidas são feitas para cada 10 ou 15cm de profundidade do poço. Um poço contém, em geral, poucos milhares de metros de profundidade, então a tabela com os dados de um poço pode ter de 6000 linhas até cerca de 50000, dependendo da profundidade. A Figura 1 mostra uma visualização de perfis para um poço. Cada uma das colunas nos plots da Figura 1 representa uma ferramenta que foi descida pelo poço e capturou informações que indicam alguma característica geofísica do conteúdo do poço.

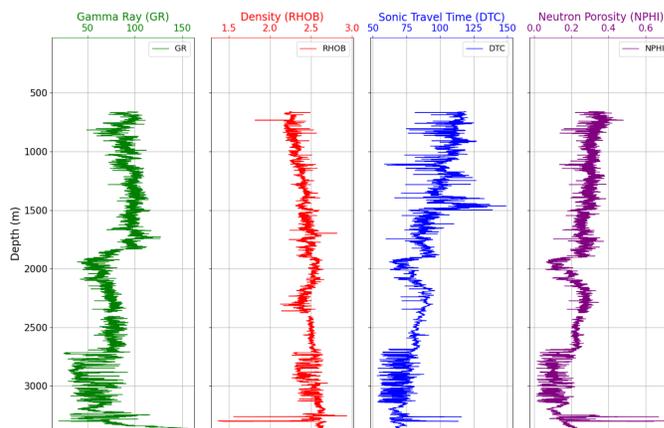


Fig. 1. Poço Ahuroa South B-1 localizado na Nova Zelândia. Dados provenientes do dataset Taranaki [W.S.R. de Carvalho et al. 2020]. Cada coluna representa um perfil do poço. Dados para raios gama (GR), densidade (RHOB), onda compressional (DTC) e porosidade neutrão (NPHI), respectivamente.

A tarefa de similaridade, no caso de perfis de poço, pode ser vista como uma tarefa de similaridade em séries temporais, tendo em vista que o tempo, nesse caso, pode ser representado como a profundidade do poço. No aspecto geológico, essa comparação pode ser confirmada visto que as rochas contidas em cada poço representam tempos geológicos diferentes. Quanto mais fundo no poço, mais antigas são as rochas que se assentaram naquela posição. Essas rochas mais antigas, em maior profundidade, foram sendo sobrepostas por outras a medida que o tempo passou. Desse modo, a noção de tempo geológico é um indício de que a comparação da profundidade do poço com o tempo em séries temporais é relevante e propícia.

Dessa forma, o objetivo deste projeto é propor um método de identificação de similaridade entre poços, mas que possa, também, ser expandido a outras tarefas em séries temporais. O método de resolução será aplicar uma rede neural auto-supervisionada com a capacidade não só de classificar pares positivos e negativos de poços no que tange à similaridade, mas também fornecer os pares negativos por meio de um processo generativo. O modelo será baseado em GANs [Goodfellow et al. 2014] e será treinado de modo adversarial.

Durante a primeira etapa do projeto (POC I), o objetivo era compreender os dados relacionados a esta tarefa e propor uma versão inicial do modelo, com experimentos introdutórios que indicassem a possibilidade de continuar a pesquisa no método proposto. Os resultados obtidos pela GAN modificada para o propósito de similaridade, com a geração automatizada de exemplos negativos, mostraram que o modelo é capaz de visualmente produzir bons resultados e identificar aspectos de similaridade entre os perfis. Como os resultados foram obtidos apenas com um perfil (raio gama - GR) em uma avaliação puramente visual de amostragem de exemplos, ainda é necessário validá-lo de outras formas. Por isso, essa etapa do projeto envolve propor novas formas de avaliação do projeto, para, então, otimizar arquiteturas e parâmetros. Foi necessário, ainda, estabelecer baselines com os quais se pudesse comparar o modelo proposto.

Esse documento está dividido em cinco seções. A primeira, esta, introduz o problema e apresenta a forma de abordagem proposta para o projeto, além de sumarizar os objetivos e resultados obtidos na primeira etapa do trabalho (POC I). A segunda inclui a revisão literária, a fim de explicitar os métodos de solução ao problema apresentados por outros autores. A terceira apresenta a metodologia proposta, com detalhes da modelagem e da forma de implementação. A quarta seção aborda os resultados esperados com o projeto. Por fim, a quinta propõe as etapas e o cronograma a serem seguidos durante o período de duração deste projeto.

II. REFERENCIAL TEÓRICO

Tendo em vista a relação entre perfis de poços e séries temporais, existem métodos clássicos de similaridade em séries temporais aplicados ao ramo de perfis de poços. Desses métodos, os mais bem sucedidos são os baseados em deformação da série temporal, como o Dynamic Time Warping (DTW). O método DTW visa transformar uma série finita de forma que ela se assemelhe ao máximo à outra e foi inicialmente projetado para usar a métrica de correlação cruzada na vizinhança dos pontos das primeira e segunda sequências. No entanto, em algumas aplicações geofísicas é possível encontrar as distâncias Euclidiana (norma L2) ou Manhattan entre as duas séries temporais.

Esse algoritmo foi aplicado para estimar a semelhança entre perfis de poços e séries sísmicas por [Lineman et al. 1987] e [Zoraster et al. 2004], de modo que a informação contida nos registros de poços pudesse ser usada para inferir características geológicas comuns. Desde os trabalhos que derivaram seu algoritmo original [Bellman and Kalaba 1959], o DTW e outros algoritmos baseados em deformação receberam melhorias e modificações ao longo do tempo. [Ezenkwu et al. 2023] apresentaram uma revisão abrangente de diferentes técnicas de correspondência aplicadas à análise de perfis de poços com o gamma ray, na qual o desempenho de algoritmos como DTW, Constrained Dynamic Time Warping (CDTW) e Correlation Optimised Warping (COW) foi investigado.

[Verma et al. 2014] propõem métodos alternativos ao DTW. Os autores utilizam dois métodos para detectar sincronização

generalizada: probabilidade de sincronização (synchronization likelihood - SL) [Stam and Van Dijk 2002] e semelhança do gráfico de visibilidade (visibility graph similarity - VGS) [Ahmadlou and Adeli 2012]. A ideia principal de ambos é dividir os dados do poço em vários blocos menores e determinar a semelhança entre dois poços calculando uma métrica de semelhança em cada bloco do primeiro poço e os blocos do segundo. Em seguida, eles extraem a média dos resultados para ver o quão semelhantes são os poços.

Assim como em outros domínios, aprendizado de máquina também foi aplicado com sucesso ao ramo de análise de dados geológicos. A exemplo disso, artigos como [Choubey and Karmakar 2021], [Koroteev and Tekic 2021], [Sircar et al. 2021] escrevem sobre a aplicação de aprendizado de máquina na indústria petrolífera em diversas tarefas. De modo complementar, existem trabalhos mais específicos que propõem métodos de aprendizado mais focados à tarefa de similaridade com dados de perfis de poço.

[Brazell et al. 2019] propõem uma ferramenta capaz de correlacionar perfis de poços. Um topo de poço (uma profundidade inicial de formação geológica) é marcado em um poço e uma rede convolucional (CNN) tenta detectar esse mesmo topo em outros poços. Essa tarefa se assemelha a similaridade, visto que a rede deve ser capaz de encontrar seções similares em poços, de forma a compreender onde está o início da formação. Para isso, os autores usaram perfis de gamma ray e resistividade, assim como a coordenada geográfica dos poços.

[Ali et al. 2021] propõem um método de similaridade a fim de recuperar informações de perfis ausentes. Os autores propõem uma "footmark" - uma geometria no formato 2D de dois perfis (sônico e densidade) - para calcular a semelhança entre dois poços com base na intersecção de suas geometrias. Esse "footmark" é usado para computar a distância de Jaccard entre cada poço, criando uma matriz de similaridade baseada nessas distâncias, em que cada posição (i,j) da matriz representa a distância de Jaccard entre os poços i e j. Com essa informação, eles aplicam o algoritmo de PageRank a fim de criar uma métrica que organize os poços de acordo com a significância na matriz, e isso mostra o quanto cada poço explica os outros no dataset. Usaram três poços para treino e dois para teste.

[Romanenkova et al. 2022] propuseram uma rede recorrente baseada em blocos LSTM com treinamento não supervisionado e auto-supervisionado para calcular a similaridade entre intervalos de poços. A similaridade foi estimada com base na distância euclidiana e de cosseno entre as representações latentes obtidas das redes (Figura 2).

Mesmo com as propostas de estado da arte para a tarefa, é possível perceber como a literatura carece de modelos mais aprofundados e aplicações mais específicas para o problema de similaridade em perfis de poços. A abordagem proposta em [Romanenkova et al. 2022] é uma abordagem padrão de transformação em estado latente, que é aplicada desde o surgimento das redes, inclusive com um modelo muito simplificado baseado em LSTMs. Dessa forma, a proposta desse projeto é diferencial para a área, uma vez que aborda um

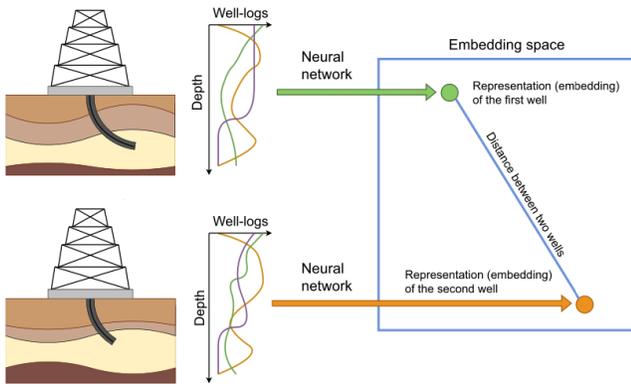


Fig. 2. Esquema proposto por [Romanenkova et al. 2022] em que seqüências a serem comparadas são convertidas para representações latentes antes de serem efetivamente comparadas. Fonte da imagem: [Romanenkova et al. 2022].

problema em aberto e introduz uma técnica nova e adaptada para o problema de similaridade. O aprendizado adversarial do modelo proposto neste trabalho é vantajoso por ser capaz de propor pares negativos automaticamente para a rede de similaridade, em um processo muito semelhante a um augmentation dos dados.

III. METODOLOGIA

O framework a ser desenvolvido gerará uma medida de similaridade entre dois poços por meio dos dados de perfis deles. Existem duas partes principais da arquitetura: geração e similaridade. Na geração, usamos uma GAN Bidirecional (BiGAN) [Donahue et al. 2016]. A rede geradora é encarregada de gerar perfis de poço falsos a partir de vetores de ruído, enquanto a rede discriminadora diferencia dados gerados de dados reais. A maior diferença da GAN bidirecional é a adição do Encoder, que transforma dados em formato de perfis de poço em vetores de ruído no mesmo espaço dos vetores fornecidos como entrada para o gerador.

Na similaridade, usamos uma rede treinada de modo autossupervisionado. Para isso é usada a triplet loss, que tem como objetivo minimizar a distância entre representações de dados âncora e positivos, enquanto maximiza a distância entre dados âncora e negativos. Dados âncora são provenientes do dataset, isto é, dados reais de poço. Neste trabalho, dados positivos são coletados por meio de augmentation, enquanto dados negativos são obtidos com o gerador da BiGAN.

Para os dados positivos, foram aplicados quatro métodos de augmentation baseados em augmentation de séries temporais. Nem todas as abordagens de séries temporais se adaptam bem aos dados de perfil de poço de maneira que a seqüência resultante permaneça semelhante à original, então selecionamos:

- Jittering: Adição de ruído Gaussiano.
- Scaling: Multiplicação do perfil com um valor aleatório de uma distribuição Gaussiana centrada em 1.
- Magnitude Warping: Multiplicação por uma curva spline cúbica. A Figura 4 mostra um exemplo de tal curva.
- Time Warping: Esticando e comprimindo o perfil.

Cada abordagem implementada é mostrada na Figura 4. Durante a fase de treinamento, a amostra positiva foi criada a partir de qualquer subconjunto desses augmentations, exceto do conjunto nulo, o que significa que não há amostras positivas que não passem por nenhuma técnica.

No caso dos dados negativos, eles devem ser dissimilares às âncoras. Considerando a variabilidade dos dados, é mais provável que o dado gerado pela GAN seja dissimilar. No entanto, isso não é garantia e, para confirmar isso, implementamos a GAN bidirecional. Como o Encoder E retorna dados para o estado vetorial passado ao Gerador G , para garantir a dissimilaridade entre os dois, passamos a âncora x por E , obtendo $E(x)$ e, com $E(x)$, selecionamos um vetor z espacialmente distante para gerar o dado negativo $G(z)$. A proposta final está representada na Figura 3 e um exemplo da geração com um z distante de $E(x)$ está na Figura 5.

A. Baselines

Para avaliar a qualidade da nossa metodologia, comparamos com cinco abordagens da literatura. Essas abordagens incluem o método não-contrastivo BYOL usado por [?], o método contrastivo SimCLR usado por [?], a estratégia de triplet loss proposta por [Romanenkova et al. 2022], o algoritmo Dynamic Time Warping (DTW), comumente usado para avaliar a similaridade de séries temporais, e o variational autoencoder usado tanto por [Romanenkova et al. 2022] quanto por [?], que é uma abordagem autoassociativa para aprendizado autossupervisionado. O objetivo com os baselines é passar por uma ampla gama de alternativas em diferentes técnicas e comparar novas técnicas usadas principalmente em outros campos, como o BYOL e o SimCLR.

A abordagem de triplet loss de [Romanenkova et al. 2022] é realizada com um ResNet50 1D como backbone seguido por uma camada totalmente conectada. O BYOL e o SimCLR também utilizam o mesmo modelo como backbone. Para o VAE, o codificador é composto por cinco seqüências de duas convoluções seguidas por uma camada de pooling máximo. O decodificador contém cinco seqüências de duas convoluções seguidas por uma camada de convolução transposta.

IV. EXPERIMENTOS E RESULTADOS

Os experimentos realizados para validação da proposta foram feitos com dois bancos de dados. O primeiro é da competição FORCE 2020 [Bormann et al. 2020], fornecidos pelo Norwegian Petroleum Directorate (NPD) - empresa que coordena a perfuração de poços e extração de petróleo na Noruega. O dataset contém perfis de 118 poços. O segundo é da bacia de Taranaki, na Nova Zelândia, com 407 poços. Todos os poços foram usados a fim de garantir a qualidade no processo de validação dos resultados obtidos pelo modelo.

Para este trabalho, selecionamos quatro perfis destacados na Figura 6. O primeiro foi o GR porque é o perfil mais comum para ambos os conjuntos de dados e mede informações relevantes sobre a radiação natural emitida por diferentes tipos de rochas. A porosidade de nêutrons (NPHI) também foi selecionada para auxiliar na avaliação da porosidade e saturação de

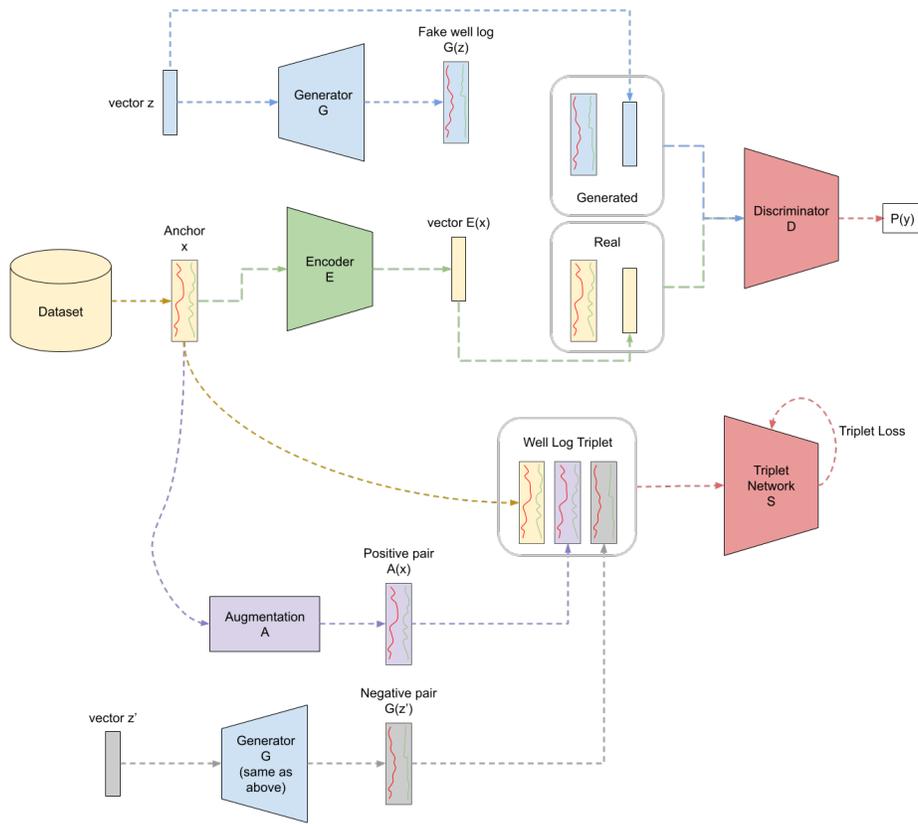


Fig. 3. Representação em alto nível da arquitetura

fluidos do poço. O perfil de densidade (RHOB) foi selecionado para fornecer informações sobre a densidade total da formação. Por fim, a medição das propriedades acústicas (DTC) por meio de perfis sônicos também foi selecionada para ajudar a determinar as velocidades das ondas compressivas, o que auxilia na avaliação das propriedades mecânicas das rochas.

Todos os dados passaram por um pré-processamento que consistiu na extração de anomalias e redimensionamento dos valores das séries. A extração de anomalias foi feita por meio do corte de valores superiores ao percentil 99 e inferiores ao percentil 1 para cada um dos poços. O motivo é para evitar que o modelo receba dados em que as máquinas falharam ou em que os sensores captaram ruídos excessivos, processos que ocorrem com baixa frequência, mas afetam todos os dados. O redimensionamento foi feito por meio dos valores mínimos e máximos dos dados (min-max scaling), de acordo com a equação 1:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Assim como já mencionado, o modelo recebe sequências de 256 valores como entrada. É necessário, também, desconsiderar os dados faltantes no momento de fornecer os dados para o modelo. Para isso, são selecionadas todas as sequências

possíveis que não possuem dados faltantes, o que totaliza 15431 sequências em 525 poços.

A. Ranking de Senoides

O primeiro experimento serve como um experimento motivacional para demonstrar o desempenho do nosso modelo quando aplicado a dados estruturados como sequências, similares aos dados de perfis de poço. Este experimento teve como objetivo classificar senoides com base em sua frequência. A principal vantagem deste experimento é a existência de um ranking ground-truth, o que facilita a comparação entre os modelos.

Para cada senoide usada como query, espera-se que o modelo ordene o conjunto de dados com base na proximidade de seus vetores latentes ou, no caso do DTW, no valor de alinhamento. Comparado ao ranking ground-truth, que é criado avaliando os valores de frequência mais próximos, métricas podem ser extraídas para comparar os modelos. As amostras são geradas como na Figura 7, adicionando-se um ruído para diferenciar as curvas.

Os resultados podem ser vistos na Tabela I. Este experimento sugere que nosso modelo tem grande potencial para a tarefa de similaridade, especialmente quando comparado a modelos mais antigos como VAE e DTW. No entanto, modelos avançados de visão computacional, como BYOL,

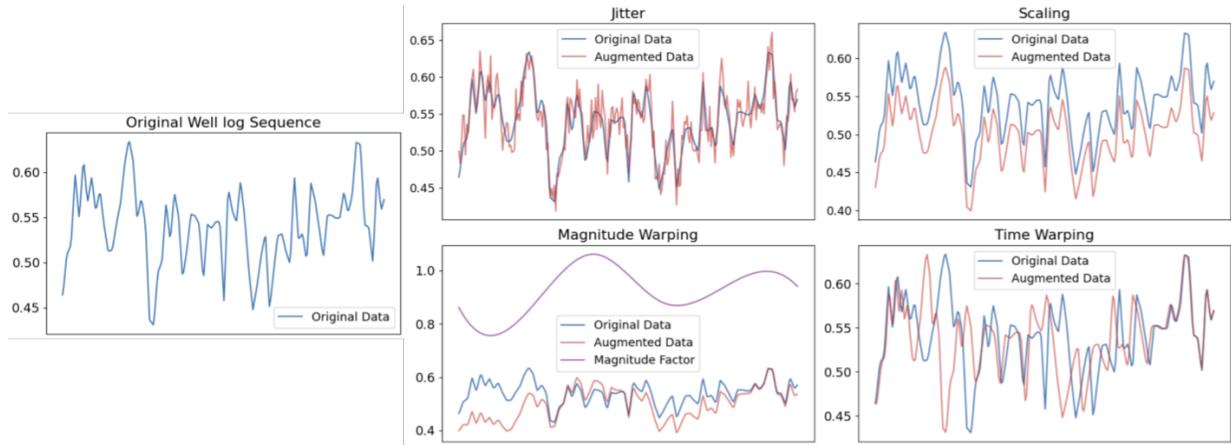


Fig. 4. Exemplos das técnicas de augmentation selecionadas para este estudo. A sequência original foi retirada do perfil GR do poço Ahuroa South B-1 da bacia de Taranaki [W.S.R. de Carvalho et al. 2020].

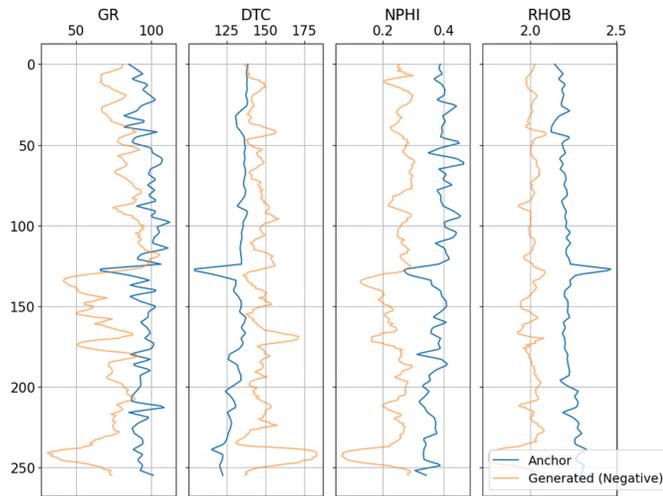


Fig. 5. Exemplo de sequência gerada com espaço latente z distante do encoding $E(x)$ da sequência âncora.

Method	MAP1	MAP5	MAP10	MRR10
DTW	0.03	0.29	0.46	0.52
VAE	0.05	0.33	0.52	0.59
BYOL	0.23	0.81	0.92	0.97
GAN(Ours)	0.16	0.67	0.81	0.88

Methods — VAE: Variational Autoencoder; BYOL: Bootstrap Your Own Latent. **Performance** — MAPk: Mean Average Precision at k (MAP@k); MRRk: Mean Reciprocal Rank at k (MRR@k).

TABLE I

PERFORMANCE MÉDIA DOS MODELOS NO EXPERIMENTO DE RANKING DE SENOIDES.

podem ser mais eficientes nessa tarefa. Nosso modelo captura corretamente a similaridade nas senoides e o faz melhor do que modelos tradicionais em todas as métricas utilizadas.

B. Ranking de Perfis de Poço com Augmentation

Como o augmentation está relacionado à nossa meta de similaridade, ou seja, queremos que a âncora e seu augmentation

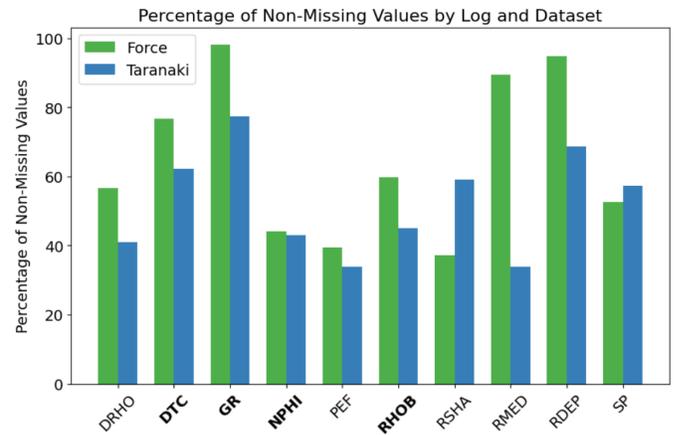


Fig. 6. Percentual de dados não faltantes no dataset Force [Bormann et al. 2020] e Taranaki [W.S.R. de Carvalho et al. 2020]. Quanto maior, melhor (significa mais dados disponíveis para cada perfil). Perfis destacados no eixo das abscissas são usados neste trabalho.

sejam semelhantes, outra abordagem para avaliar os modelos é classificando todo o conjunto de dados com um aumento da imagem de consulta nele. O objetivo é estabelecer onde os modelos classificariam o aumento, considerando que é isso que se espera da similaridade dos perfis de poço. De acordo com a posição do augmentation, métricas são estabelecidas para avaliar o desempenho de todos os modelos.

Os augmentations usados para este experimento foram uma combinação aleatória de todas as quatro técnicas mencionadas anteriormente, cada uma com uma chance de 50% de ser selecionada. Para cada consulta, seu augmentation foi adicionado ao conjunto de dados, e a consulta foi comparada a todos os dados. Os resultados são vistos na Tabela II.

Assim como no experimento com senoides, nosso modelo alcança bons resultados neste experimento de ranking de perfis de poço, superando os modelos tradicionais para a tarefa de similaridade, como VAE e DTW. Do mesmo modo, os modelos atuais usados em visão computacional (BYOL e

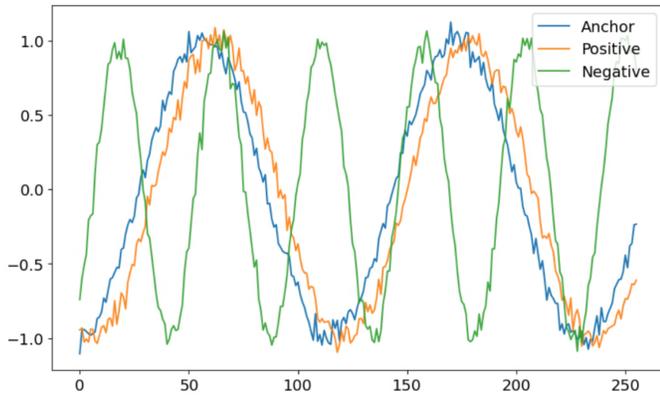


Fig. 7. Exemplo de seqüências no experimento das senoides.

Method	Acc1	Acc5	Acc10	MRR
DTW	0.41	0.49	0.54	0.46
Triplet-Rom	0.15	0.21	0.24	0.18
VAE	0.50	0.55	0.57	0.52
BYOL	0.85	0.93	0.95	0.88
SimCLR	0.80	0.92	0.94	0.85
GAN(Ours)	0.60	0.77	0.82	0.68

Methods — Triplet-Rom: Triplet model from [Romanenkova et al. 2022]; VAE: Variational Autoencoder; BYOL: Bootstrap Your Own Latent.

Performance — Acck: Accuracy at k (Acc@k); MRR: Mean Reciprocal Rank.

TABLE II

PERFORMANCE MÉDIA DOS MODELOS NO EXPERIMENTO DE RANKING DE DADOS DE POÇO COM AUGMENTATION.

SimCLR) obtiveram melhores resultados neste experimento.

Em particular, a métrica Acc1 é menos relevante do que a Acc5 ou Acc10. Isso porque, embora o augmentation seja uma seqüência semelhante à consulta e deva ser altamente classificado, não é garantido que seja a mais semelhante no conjunto de dados, pois pode haver seqüências mais semelhantes do que o augmentation. Assim, Acc5 e Acc10 indicam melhor a qualidade dos modelos ao mostrar com que frequência o augmentation está no topo, onde deveria estar, mas não necessariamente em primeiro lugar.

C. Resultados Qualitativos

Este último experimento visa fornecer uma confirmação visual da qualidade de cada método. Para isso, uma seqüência de consulta do conjunto de dados foi selecionada aleatoriamente e comparada com todas as outras seqüências para ver as melhores classificações que cada modelo daria. Para ilustrar a similaridade entre poços diferentes, o que é mais apropriado ao perfurar um novo poço para busca de petróleo, excluimos temporariamente todas as seqüências do mesmo poço que a query selecionada. Dessa forma, apenas seqüências de poços diferentes são classificadas no topo do experimento. A Figura 8 mostra o resultado desse experimento para cada um dos modelos.

Observa-se que nosso modelo baseado no treinamento com triplet loss e GAN obtém um resultado muito fidedigno. A seqüência apresentada por ele é mais semelhante a original por

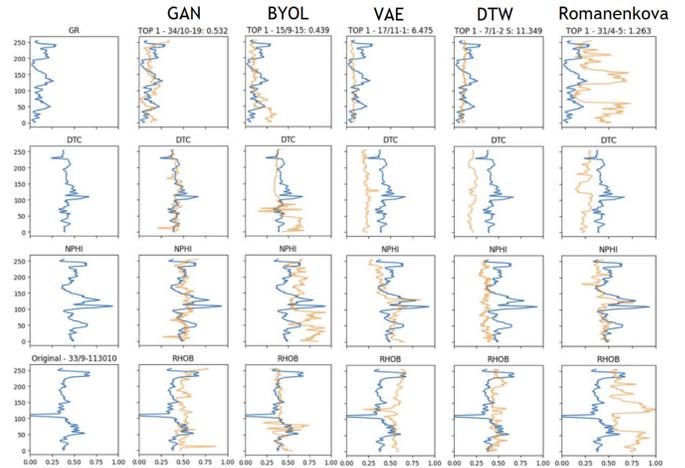


Fig. 8. Example of sequences in the sinusoidal experiment.

estar em uma faixa de valores similar e com padrões também parecidos. É interessante notar que o modelo prioriza bem uma faixa semelhante de valores na seqüência, o que é muito importante para a análise geológica de perfis de poço. Desse modo, os experimentos mostrados nesse artigo demonstram a funcionalidade do modelo e são motivações para a pesquisa na indústria de extração de petróleo.

V. CONCLUSÃO

Os resultados obtidos foram promissores e refletem a qualidade do modelo para a similaridade de perfis de poços. Ainda que os resultados tenham sido inferiores a modelos da classe de contrastivos e não contrastivos, a validade do modelo apresentado neste trabalho é o largo espaço de melhoria, já que é uma primeira versão apresentada. Ainda assim, o método foi capaz de superar modelos tradicionais e se aproximar de modelos atuais, o que indica um bom resultado.

A arquitetura com um modelo adversarial treinada de modo auto-supervisionado é efetiva na solução desse problema, que possui grande relevância para o mercado do petróleo. Essa ferramenta pode, portanto, ser usada como suporte em análises manuais de similaridade de poços. A proposta com um modelo gerador que fornece seqüências dissimilares se mostrou relevante e eficiente nos experimentos amostrados, de forma que o modelo de similaridade obteve resultados consistentes com o desejado.

Este trabalho ainda trouxe novos métodos de experimentação que podem ser utilizados para comparação de modelos aplicados na área. O benchmark mostrado é importante para o estabelecimento dos melhores modelos voltados à similaridade de perfis de poço.

Ainda que alcance resultados convincentes, o modelo tem a desvantagem comum entre abordagens de aprendizado de máquina de aceitar seqüências de tamanho fixo. Os próximos passos devem ser direcionados a lidar com entradas de tamanho variável, isto é, seqüências de poços de tamanho diferente, já que poços possuem profundidades variáveis.

REFERENCES

- [Ahmadlou and Adeli 2012] Ahmadlou, M. and Adeli, H. (2012). Visibility graph similarity: A new measure of generalized synchronization in coupled dynamic systems. *Physica D: Nonlinear Phenomena*.
- [Ali et al. 2021] Ali, M., Jiang, R., Ma, H., Pan, H., Abbas, K., Ashraf, U., and Ullah, J. (2021). Machine learning-A novel approach of well logs similarity based on synchronization measures to predict shear sonic logs. *Journal of Petroleum Science and Engineering*.
- [Bellman and Kalaba 1959] Bellman, R. and Kalaba, R. (1959). On adaptive control processes. *IRE Transactions on Automatic Control*.
- [Bormann et al. 2020] Bormann, P., Aursand, P., Dilib, F., Manral, S., and Dischington, P. (2020). FORCE 2020: Well log and lithofacies dataset for machine learning competition. <https://www.npd.no/en/force/events/machine-learning-contest-with-wells-and-seismic/>.
- [Brazell et al. 2019] Brazell, S., Bayeh, A., Ashby, M., and Burton, D. (2019). A Machine-Learning-Based Approach to Assistive Well-Log Correlation. *Petrophysics*.
- [Choubey and Karmakar 2021] Choubey, S. and Karmakar, G. (2021). Artificial intelligence techniques and their application in oil and gas industry. *Artificial Intelligence Review*.
- [Donahue et al. 2016] Donahue, J., Krähenbühl, P., and Darrell, T. (2016). Adversarial feature learning. *arXiv preprint arXiv:1605.09782*.
- [Ezenkwu et al. 2023] Ezenkwu, C. P., Guntoro, J., Starkey, A., Vaziri, V., and Addario, M. (2023). Automated well-log pattern alignment and depth-matching techniques: An empirical review and recommendations. *Petrophysics*.
- [Goodfellow et al. 2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.
- [Koroteev and Tekic 2021] Koroteev, D. and Tekic, Z. (2021). Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future. *Energy and AI*.
- [Lineman et al. 1987] Lineman, D., Mendelson, J., and Toksoz, M. N. (1987). Well to well log correlation using knowledge-based systems and dynamic depth warping. In *SPWLA 28th Annual Logging Symposium*.
- [Romanenkova et al. 2022] Romanenkova, E., Rogulina, A., Shakirov, A., Stulov, N., Zaytsev, A., Ismailova, L., Kovalev, D., Katterbauer, K., and AlShehri, A. (2022). Similarity learning for wells based on logging data. *Journal of Petroleum Science and Engineering*.
- [Sircar et al. 2021] Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., and Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research*.
- [Stam and Van Dijk 2002] Stam, C. J. and Van Dijk, B. W. (2002). Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets. *Physica D: Nonlinear Phenomena*.
- [Verma et al. 2014] Verma, A., Routray, A., and Mohanty, W. (2014). Assessment of Similarity Between Well Logs Using Synchronization Measures. *IEEE Geoscience And Remote Sensing Letters*.
- [W.S.R. de Carvalho et al. 2020] W.S.R. de Carvalho, B., Oliveira, M., Avalone, M., Hoffmann, J., Szwarcman, D., Guevara Diaz, J., and Zadrozny, B. (2020). Taranaki basin curated well logs. <https://zenodo.org/record/3832955>.
- [Zoraster et al. 2004] Zoraster, S., Paruchuri, R., and Darby, S. (2004). Curve alignment for well-to-well log correlation. In *SPE Annual Technical Conference and Exhibition*.