

Hugo Fabiano Alves Cunha

Estimação da Importância de Atributos com Base em Mecanismo de Atenção para Atributos Sísmicos

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós–graduação em Informática, do Departamento de Informática da PUC-Rio .

Orientador: Prof. Marcelo Gattass

Rio de Janeiro Junho de 2024



Hugo Fabiano Alves Cunha

Estimação da Importância de Atributos com Base em Mecanismo de Atenção para Atributos Sísmicos

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós–graduação em Informática da PUC-Rio . Aprovada pela Comissão Examinadora abaixo:

Prof. Marcelo Gattass Orientador Departamento de Informática – PUC-Rio

> Prof. Alberto Barbosa Raposo PUC-Rio

Dr. Jan José Hurtado Jauregui PUC-Rio

Rio de Janeiro, 12 de Junho de 2024

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, do autor e do orientador.

Hugo Fabiano Alves Cunha

Graduado em ciência da computação pela Pontificia Universidade Católica do Rio de Janeiro (PUC-RIO), onde trabalha em projetos no grupo de Geofísica Computacional do Instituto Tecgraf.

Ficha Catalográfica
Cunha, Hugo Fabiano Alves
Estimação da Importância de Atributos com Base em Mecanismo de Atenção para Atributos Sísmicos / Hugo Fabiano Alves Cunha; orientador: Marcelo Gattass. – 2024.
48 f: il. color. ; 30 cm
Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2024.
Inclui bibliografia
 Informática – Teses. 2. Aprendizado Profundo. 3. Redes Neurais. 4. Mecanismo de atenção. 5. Multi-atributo sísmico. I. Gattass, Marcelo. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

Aos meus pais, irmão e amigos por todo suporte e encorajamento.

Agradecimentos

Ao meu orientador Professor Marcelo Gattass por todo o auxilio e apoio que me permitiram realizar este trabalho.

A minha equipe que me deu todo o suporte e ajuda, com agradecimento especial a Luiz Fernando Trindade Santos por todo auxílio e orientação. Sem vocês, esse trabalho não seria possivel.

Aos meus amigos que me apoiaram nessa jornada. Em especial Arthur, Bianca, Eduardo, Gabriel T., Henrique, Hugo, Jonathan, Jonny, Pietro, Rodrigo, Suellen e Yan.

A ENEVA S.A por me disponibilizar o dataset usado para esse projeto de pesquisa.

Ao CNPq e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Cunha, Hugo Fabiano Alves; Gattass, Marcelo. Estimação da Importância de Atributos com Base em Mecanismo de Atenção para Atributos Sísmicos . Rio de Janeiro, 2024. 48p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A reflexão sísmica é o método geofísico mais empregado na indústria de petróleo e gás para estudar as camadas do subsolo. Com base nos padrões de reflexão das ondas sísmicas, os geocientistas podem inferir a estrutura e a composição das camadas geológicas abaixo da superfície, identificando potenciais reservatórios de petróleo e gás. No entanto, a interpretação dessas informações é desafiadora devido à ambiguidade inerentes dos dados, ou seja, eventos distintos podem ter respostas sísmicas similares. Com a intenção de direcionar e auxiliar esse processo, especialistas frequentemente empregam um grande conjunto de atributos sísmicos. No entanto, o uso de mais informação, em um contexto de aprendizado de máquina, não garante melhoria nos resultados e, em alguns casos, muita das features podem não ser aproveitadas pelo modelo. Sendo assim, a seleção de quais features apresentam maior relevância torna-se essencial. Contudo, uma seleção manual entre centenas de atributos pode apresentar um desafio exponencial. Este trabalho propõe uma abordagem que incorpora o uso de uma camada de atenção customizada para lidar com múltiplas features em conjunto a um modelo Long Short-Term Memory (LSTM). Essa abordagem visa ponderar automaticamente os atributos sísmicos, pré-selecionados por especialistas da área, para avaliar quais são aqueles que apresentam para o modelo uma maior importância no processo de detecção de gás natural. Para avaliar a metodologia foram empregados levantamentos sísmicos 2D e 3D onshore e aplicado a técnica de K-fold. Para os resultados de forma quantitativa, foi avaliado a métrica F1-score atingindo uma melhora de até 13,94%.

Palavras-chave

Aprendizado Profundo; Redes Neurais; Mecanismo de atenção; Multiatributo sísmico.

Abstract

Cunha, Hugo Fabiano Alves; Gattass, Marcelo (Advisor). Feature Importance Estimation Based in Attention Mechanism for Seismic Attributes . Rio de Janeiro, 2024. 48p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Seismic reflection is the most widely used geophysical method in the oil and gas industry to study subsurface layers. Based on the reflection patterns of seismic waves, geoscientists can infer the structure and composition of geological layers beneath the surface, identifying potential oil and gas reservoirs. However, interpreting this information is challenging due to the inherent ambiguity of the data, meaning distinct events can have similar seismic responses. In order to guide and assist this process, experts often employ a large set of seismic attributes. However, the use of more information in a machine learning context does not guarantee improvement in results, and in some cases, many of the features may not be utilized by the model. Therefore, selecting which features are most relevant becomes essential. However, manual selection among hundreds of attributes can pose an exponential challenge. This work proposes an approach that incorporates the use of a customized attention layer to handle multiple features in conjunction with a Long Short-Term Memory (LSTM) model. This approach aims to automatically weigh the seismic attributes, preselected by domain experts, to evaluate which ones are most important for the model in the natural gas detection process. To evaluate the methodology, 2D and 3D onshore seismic surveys were employed, and the K-fold technique was applied. For quantitative results, the F1-score metric was evaluated, achieving an improvement of up to 13.94%.

Keywords

Deep Learning; Neural Networks; Attention mechanism; Seismic Multi-Attribute.

Sumário

1 Introdução	13
 2 Trabalhos Relacionados 2.1 Aprendizado de máquina aplicado a geociência 2.2 Mecanismos de atenção 	15 15 16
3 Fundamentação Teórica	19
3.1 Aquisição e Processamento de Dados Sísmicos	19
3.2 Atributos Sísmicos	21
3.3 Redes Neurais Recorrentes	22
3.4 Long Short-Term Memory, LSTM	24
3.5 Método de Seleção de Parâmetros Supervisionados Base	eado em
Multiatenção para Séries Temporais Multivariadas	25
3.6 Métricas	28
4 Metodologia	29
4.1 Visão Geral	29
4.2 Conjunto de Dados	30
4.3 Extração de atributos	30
4.4 Pré-processamento do Dado	31
4.5 Visualização da Predição	33
4.6 Visualização da atenção nos atributos	34
5 Resultados	38
5.1 Resultados 3D na Bacia do Amazonas	40
5.2 Resultados 2D, Campo C no Complexo de Parnaiba	42
6 Conclusão e Trabalhos Futuros	46
7 Referências bibliográficas	47

Lista de Figuras

Figura 3.1 Princípio da aquisição terrestre e marítima de dados sísmicos. Figura obtida de (GERHARDT, 1998) Figura 3.2 Traço sísmico.	20 20
Figura 3.3 Seção sísmica, com agrupamento de traços sísmicos individu- ais, em posições adjacentes e intervalo (espaçamento) regular. Figura 3.4 Cubo sísmico de uma aquisição 3D. São mostradas as princi- pais direções em que seções são obtidas: <i>inline</i> (paralela à direção de aquisi- cão) cresclina (ertegenel à esce direção) e 7 clica (também chamado time	21
cao), <i>crossime</i> (ortogonal a essa direção) e 2- <i>silce</i> (também chamado <i>time</i> - ou <i>depth-slice</i>), que é uma seção horizontal. Figura 3.5 — Estrutura conceituais da m-AFS. Figura obtida de (CAO et	21
al., 2021) Figura 3.6 Estrutura da atenção para Variáveis. Figura obtida de (CAO	25
et al., 2021) Figura 3.7 Estrutura da atenção para Faixas Temporais. Figura obtida de (CAO et al., 2021)	26 27
Figura 4.1 Síntese do <i>pipeline</i> da metodologia	30
Software Opendtect, (c) Atributos. Figura 4.3 Visualização da Inline 96 do 3D na Bacia do Amazonas antes	31
da definição da janela de tempo baseada na Rol. Distância entre <i>xlines</i> de 30 m. Figura 4.4 Visualização da Inline 96 do 3D na Bacia do Amazonas antes e após definição da janela de tempo baseada na Rol. Distância entre <i>xlines</i> da 20 m	32
Figura 4.5 Visualização da predição para Inline 72 do 3D na Bacia do Amazonas. Observar ótima correlação entre a interpretação do especialista (em vermelho) com a predição (em verde). Distância entre xlines de 30 m.	34
Figura 4.6 Visualização da Inline 96 do 3D na Bacia do Amazonas por atributo (esquerda) e seu peso da atenção (direita). Figura 4.7 Visualização da média da saída da camada AoV com a Inline 96 do dado 3D na Bacia do Amazonas como entrada. (a) Amplitude, (b) Cosseno da Fase Instantânea, (c) Fase Instantânea, (d) Frequência Instantânea, (e) Energia. Quanto mais azul, mais próximo o peso dado ao atributo se aproxima de 0. Quanto mais vermelho, mais próximo o peso dado ao atributo se aproxima de 1. A soma de todos os pesos por amostra somam 1.	36 37
Figura 5.1 Divisão por <i>folds</i> da região do 3D na Bacia do Amazonas. O mapa mostra a interpretação sísmica para a extensão da ocorrência de gás. A dimensão aproximada de cada fold é 1,8 km na direção NW por 12,9 km	
na direção NE. Figura 5.2 Região no Complexo de Parnaiba. Mapa em coordenadas UTM (zona 23S), sistema de projeção SIRGAS 2000. Escala transalada para (0,0) por confidencialidade.	39 40

Figura 5.3 Pesos da atenção para o 3D na Bacia do Amazonas. (a) Amplitude, (b) Cosseno da Fase, (c) Fase Instantânea, (d) Frequência Instantânea, (e) Energia. Quanto mais azul, mais próximo o peso dado ao atributo se aproxima de 0. Quanto mais vermelho, mais próximo o peso dado ao atributo se aproxima de 1. A soma de todos os pesos por amostra somam 1.

Figura 5.4 Pesos da atenção para o Campo C. (a) Amplitude, (b) Cosseno da Fase, (c) Fase Instantânea, (d) Frequência Instantânea, (e) Energia. Quanto mais azul, mais próximo o peso dado ao atributo se aproxima de 0. Quanto mais vermelho, mais próximo o peso dado ao atributo se aproxima de 1. A soma de todos os pesos por amostra somam 1.

42

44

Lista de Tabelas

Tabela 5.1Resultados comparativo de usar multi-atributos e atributo único41Tabela 5.2Experimento 1: resultados comparativos entre não usar aten-
ção e apenas um atributo, não usar atenção e todos os atributos selecionados,
e usar atenção junto de todos os atributos selecionados43Tabela 5.3Experimento 2: resultados comparativos entre não usar aten-
ção e apenas um atributo, não usar atenção e todos os atributos menos a fase
instantânea, e usar atenção junto de todos os atributos selecionados menos
a fase instantânea45

Lista de Abreviaturas

AoT – Attention over Time
AoV – Attention over Variates
BLSTM – Bideractional Long Short-Term Memory
GANs – Generative Adversarial Networks
LSTM – Long Short-Term Memory
M-AFS – MultiAttention-based supervised Feature Selection
MLP – Multi-Layer Perceptron
MTS – Multivariable Time Series
RNC – Rede Neural Convolucional
RNN – Recurrent Neural Network
RNP – Redes Neurais Profunda
RoI – Region of Interest
SAGAN – Self-Attention and Generative Adversarial Networks

1 Introdução

A reflexão sísmica é utilizada na indústria de petróleo e gás para obter informações sobre o subsolo. Dentro desse contexto da indústria de petróleo e gás, os geofísicos usam esses dados para identificar características estruturais e estratigráficas e, juntamente com outras fontes de dados geofísicos, localizar potenciais reservas de hidrocarbonetos. No entanto, os dados obtidos a partir da reflexão sísmica podem às vezes ser ambíguos, tornando a interpretação desafiadora.

Os atributos sísmicos surgiram com o objetivo de fornecer dados complementares para interpretação sísmica, auxiliando no entendimento de subsuperfície. Essa abordagem visa acelerar e objetivar o processo, especialmente ao lidar com conjuntos volumosos de dados sísmicos 3D (REIS, 2022). Esses atributos representam uma ferramenta analítica eficaz, para auxiliar o intérprete ao realçar, identificar e correlacionar automaticamente eventos nos dados sísmicos com as estruturas geológicas reais (CHOPRA; MARFURT, 2005). Contudo, no contexto de aprendizado de máquina, apesar desses atributos sísmicos contribuirem para problemas como o de Reis (2022), que classifica faces sísmicas utilizando atributos sísmicos, à medida que dezenas de atributos sísmicos distintos estão disponíveis, a seleção desses atributos torna-se um desafio significativo. O aumento exponencial nas possibilidades de testes para escolher os atributos adequados pode se tornar um obstáculo à eficiência e à precisão do processo interpretativo (CHOPRA; MARFURT, 2005).

Diversas arquiteturas têm sido propostas para abordar uma variedade de desafios na área da geofísica. Modelos clássicos como U-net e ResNet, foram usadas em aplicações específicas, como a identificação de domos de sal e falhas, conforme discutido por Alfarhan, Deriche e Maalej (2020). Em um contexto semelhante, Pochet et al. (2018) explorou uma abordagem baseada em modelos de convolução aplicados a dados sísmicos reais. Especificamente, o estudo concentrou-se no bloco F3 offshore dos Países Baixos, localizado no Mar do Norte com intuito de detectar falhar sísmicas.

Já numa linha detecção de hidrocarbonetos, temos trabalhos como o de Santos e Gattass (2020), que propuseram uma arquitetura baseada em camadas Long Short-Term Memory (LSTM) para identificar reservatórios de gás. Essa estratégia busca capturar dependências temporais complexas nos dados sísmicos, oferecendo uma abordagem mais dinâmica para a detecção de hidrocarbonetos. Outra contribuição significativa na área de identificação de reservatórios de gás foi apresentada por Andrade et al. (2021), que propôs a utilização de um LSTM-AutoEncoder em dados sísmicos 2D terrestres. Os resultados obtidos foram positivos, indicando a eficácia dessa abordagem para identificar padrões e características associadas à presença de hidrocarbonetos em dados sísmicos.

Na linha de multi-atributos sísmicos, temos o trabalho de Cunha et al. (2023), em que foi proposto o uso de cinco atributos sísmicos (Amplitude, Cosseno de Fase, Fase instantânea, Frequência instantânea e Energia) para realizar a identificação de estruturas de gás natural em dados sísmicos 3D terrestres. Os resultados obtidos por ele, quando utilizam multi-atributos sísmicos, superam os valores de F1-Score quando comparado ao uso de apenas a amplitude sísmica. Por sua vez, Reis (2022) propôs utilizar multi-atributos sísmicos junto a uma arquitetura de Rede Neural Convolucional (RNC) seguida de uma deconvolução para classificação automática de fácies sísmicas, também obtendo melhora com relação a utilização de apenas a amplitude. Seguindo a linha de classificação de fácies sísmicas, temos o trabalho de AlSalmi e Elsheikh (2024), que incorpora mecanismos de atenção junto a arquitetura U-Net.

Esse trabalho apresenta uma metodologia para realizar a ponderação automatica dos atributos mais relevantes através de uma camada de atenção especializada em multiplas séries temporais, inspirada por Cao et al. (2021). A camada de atenção apresentada neste trabalho será utilizada em conjunto a uma rede com arquiteturas de *Long Short-Term Memory* (LSTM) que será treina para caracterizar regiões que apresentem assinaturas de gás em dados sísmicos utilizando diversos atributos sísmicos. Para treinar e avaliar este modelo, foi usada uma base de dados é provida pela empresa Eneva S.A. em dados reais com observação em campo de gás natural.

Esta dissertação deu origem a um artigo aceito na European Association of Geoscientists & Engineers (EAGE) e apresentado no 84th EAGE Annual Conference & Exhibition.

Esse documento está organizado em 6 capítulos. O capítulo 2 consiste dos trabalhos relacionados, que serviram como fundação para o desenvolvimento deste trabalho. O capítulo 3 apresenta a fundamentação teórica do trabalho, abrangendo a aquisição e processamento de dados sísmicos, atributos sísmicos, redes neurais recorrentes, long short-term memory e camadas de multi-atenção. O capitulo 4 descreve a metodologia do trabalho, definindo o fluxo de execução do modelo, desde o dado bruto e o pré-processamento do dado, até a predição e visualização do dado, seguido pela conclusão. O capitulo 5 apresenta os resultados do trabalho. Por fim o capítulo 6 apresenta a conclusão e possiveis trabalhos futuros.

2 Trabalhos Relacionados

Está seção será dividida em duas partes, a primeira se trata de trabalhos relacionas majoritamente a área de computação aplicada à geociência, a segunda seção irá tratar em particular sobre mecanismos de atenção e a relação com este trabalho.

2.1 Aprendizado de máquina aplicado a geociência

Diversas arquiteturas têm sido propostas para abordar uma variedade de desafios na área da geofísica. Modelos consagrados, como U-net e ResNet, foram usados em aplicações específicas, como a identificação de domos de sal e falhas, conforme discutido por Alfarhan, Deriche e Maalej (2020). O uso dessas arquiteturas tem se mostrado eficaz em várias tarefas relacionadas à interpretação de dados geofísicos.

Em um contexto semelhante, Pochet et al. (2018) explorou uma abordagem baseada em modelos de convolução aplicados a dados sísmicos reais com intuito de detectar falhar sísmicas. Especificamente, o estudo concentrou-se no bloco F3 offshore dos Países Baixos, localizado no Mar do Norte.

Para reconstrução de dados sísmicos temos propostas como a de Yoon, Yeeh e Byun (2021) que, com o objetivo de mitigar problemas e irregularidades no processo de aquisição de dados sísmicos, reconstruir seções faltantes da aquisição sísmica utilizando redes *Bideractional Long Short-Term Memory* (BLSTM). Já Yu e Wu (2021) aborda este problema utilizando Redes Neurais Convolucionais junto a um mecanismo de atenção e funções de perda híbrida para guiar o processo de treinamento. Este processo se mostrou especialmente funcional em casos cujas seções faltantes são consecutivas quando comparados a modelos que não utilizam atenção ou a perda híbrida.

Já numa linha detecção de hidrocarbonetos, temos trabalhos como o de Santos e Gattass (2020), que propuseram uma arquitetura baseada em camadas *Long Short-Term Memory* (LSTM) para identificar reservatórios de gás. Essa estratégia busca capturar dependências temporais complexas nos dados sísmicos, oferecendo uma abordagem mais dinâmica para a detecção de hidrocarbonetos.

Outra contribuição significativa na área de identificação de reservatórios de gás foi apresentada por Andrade et al. (2021), que propôs a utilização de um *LSTM-AutoEncoder* em dados sísmicos 2D terrestres. Os resultados obtidos

foram positivos, indicando a eficácia dessa abordagem para identificar padrões e características associadas à presença de hidrocarbonetos em dados sísmicos.

Na linha de multi-atributos sísmicos, temos o trabalho de Cunha et al. (2023), em que foi proposto o uso de cinco atributos sísmicos (Amplitude, Cosseno de Fase, Fase instantânea, Frequencia instantânea e Energia) para realizar a identificação de estruturas de gás natural em gás em dados sísmicos 3D terrestres. Os resultados quando utilizam multi-atributos sísmicos superam os valores de F1-Score quando comparado ao uso de apenas a amplitude sísmica.

Por sua vez, Reis (2022) propós utilizar multi-atributos sísmicos junto a uma arquitetura de Rede Neural Convolucional (RNC) seguida de uma deconvolução para classificação automática de fácies sísmicas, também obtendo melhora com relação a utilização de apenas a amplitude como atributo sísmico.

No tópico de classificação de fácies sísmicas, temos o trabalho de AlSalmi e Elsheikh (2024). Para mapeamento de fácies sísmicas em cubos sísmicos tridimensionais, ele propõe uma arquitetura U-Net baseada em RNC. O método incorpora blocos residuais no *encoder*, emprega *attention gates* semelhantes a *transformers* (VASWANI et al., 2017) após cada bloco residual e inclui dados de espectro de frequência além da amplitude sísmica como entrada. Os resultados superam *benchmarks* recentes, evidenciando maior precisão na segmentação de fácies sísmicas. A validação em dois conjuntos de dados tridimensionais, F3 *offshore* dos Países Baixos e Penobscot, destaca a eficácia do método, proporcionando uma segmentação de alta qualidade com redução significativa de artefatos, detecção de bordas aprimorada e maior consistência lateral.

Além do trabalho de AlSalmi e Elsheikh (2024), mecanismos de atenção também tem sido empregados para auxliar em problemas relacionados a sísmica. Como o trabalho de Ding, Zhou e Chi (2024), este aborda as complexidades na aquisição de dados sísmicos, propondo o uso de *Self-Attention* and Generative Adversarial Networks (SAGAN), uma abordagem que combina mecanismo de atenção e Generative Adversarial Networks (GANs) para reconstrução e remoção de ruídos em sinais sísmicos. Experimentos demonstram a superioridade do SAGAN em interpolação e remoção de ruídos, estabelecendo o SAGAN como uma escolha líder para recuperação de sinais sísmicos.

2.2 Mecanismos de atenção

Especificamente no que diz respeito à atenção, atenção é um poderoso mecanismo que permite aos modelos de aprendizado de máquina focalizarem diferentes partes das entradas durante o processo de treino, essencialmente permitindo uma abordagem mais flexível e eficaz para modelar dependências de longo alcance em sequências de dados. Esse mecanismo funciona ao atribuir pesos adaptativos a cada elemento da sequência de entrada, permitindo que o modelo priorize informações mais relevantes para a tarefa em questão, o que é fundamental para a qualidade do processamento e da geração de saídas precisas (VASWANI et al., 2017). Essa abordagem de atenção tem sido amplamente utilizada em uma variedade de aplicações de aprendizado de máquina, incluindo tradução automática (VASWANI et al., 2017), geração de imagens a partir de texto (KIM et al., 2023), sumarização de documentos (LING, 2017), reconhecimento de fala (CHOROWSKI et al., 2015), entre outros.

Em problemas cuja entrada é de texto, as palavras e suas posições em uma frase são interpretados como uma sequencia que varia ao longo do tempo e se relaciona de forma dependente das palavras que vieram antes e depois, essas relações podem ser usadas para gerar uma saída especifica, seja uma outra frase em outro idioma (VASWANI et al., 2017), seja uma imagem (KIM et al., 2023).

No caso em que a entrada envolve voz e sons a fala de uma pessoa é interpretada como um sinal sonoro, e, no caso do trabalho de Chorowski et al. (2015), esse sinal corresponderá a um texto. Este sinal sonoro é interprado como uma série temporal pelo modelo, porém diferentemente de problemas onde o texto é a entrada, esses sinais sonoros muitas vezes podem possuir bastante ruído, frequentemente não possuim uma estrutura clara e podem ser longos. Essa peculiaridade do sinal sonoro torna o uso de mecanismo de atenção altamente atrativo para auxiliar o modelo a dar peso apenas as partes mais relevantes do dado (CHOROWSKI et al., 2015).

As abordagens citadas acima tratam de apenas uma série temporal, e seus mecanismos de atenção utilizados estão preparados para receber apenas uma série temporal como entrada. Para o dominio do problema que queremos tratar neste trabalho, estes mecanismos de atenção não serão suficientes para lidar com multiplas séries temporais e discriminar entre elas. Este trabalho visa utilizar multiplos atributos sísmicos, onde cada atributo por sua vez é uma série temporal, e ponderar sua relevância durante o processo de treinamento.

Em um contexto de seleção de atributo, para superar a limitação de atenção para um única série temporal, temos o trabalho de Cao et al. (2021), que apresenta uma abordagem chamada de *multiple attention-based supervised feature selection method* (m-AFS). Este metodo utilizada um mecanismo de atenção usado para identificar a importancia de atributos em *multivariate time series* (MTS). Para isso, são utilizados duas camadas distintas para lidar com

o espaço temporal do dado e o espaço de atributos do dado. Essas camadas são chamadas de Aov (*Attention over Variates*), que lida com o espaço de atributos, e AoT (*Attention over Time*), que lida com o espaço de tempo.

3 Fundamentação Teórica

Nesta seção, serão apresentados alguns temas que serviram como fundamentação teórica. Os estudos selecionados demonstram o valor dessas abordagens em tarefas de processamento de dados e análise de sistemas complexos.

3.1 Aquisição e Processamento de Dados Sísmicos

Esta seção se baseia fortemente nas seções "Aquisição de dados" e "Processamento sísmico" da dissertação de mestrado de Gerhardt (1998) além das Notas de Curso Interno Tecgraf de Rodrigues (2019).

O método de reflexão sísmica destaca-se como a abordagem mais prevalente na indústria do petróleo atualmente, devido à sua capacidade de fornecer a melhor definição possível, entre todos os métodos (geofísicos ou não) indiretos das características geológicas na subsuperfície, tornando-se propício para a detecção de acumulações de hidrocarbonetos. Investimentos significativos, superando os 90% do total dedicado à geofísica de prospecção, são direcionados para a aplicação da sísmica de reflexão (THOMAS et al., 2001). Os resultados finais podem englobam, entre outras informações, imagens das estruturas e camadas geológicas na subsuperfície, apresentadas em diversas formas, e que são disponibilizadas para orientar o trabalho dos intérpretes. Segundo Robinson e Treitel (2000), o processo de exploração de hidrocarbonetos (petróleo e gás) fundamentado na sísmica, pode ser desmembrado em três fases distintas: aquisição, processamento e interpretação. Abaixo, forneceremos uma breve síntese de cada etapa.

Convencionalmente, as fontes sísmicas são provenientes de explosões de dinamite ou vibradores possantes em caso de aquisição terrestre, e canhões pneumáticos de ar em caso de aquisição marítima. Após o uso desses instrumentos, para medir o tempo de trânsito do pulso sísmico pelo meio, são usados receptores chamados geofones, no caso de aquisição por terra, ou hidrofones, no caso de aquisição marítma. Ambos receptores funcionam como transdutores que convertem as amplitudes das ondas sísmicas em valores de tensão. Os valores de amplitudes, que podem ser negativos ou positivos, estão associados aos contrastes de impedância (produto entre a velocidade de propagação da onda e a densidade) encontrado pelo pulso sísmico em cada interface (topo ou base) de cada camada ou sequência geológica. Esse princípio pode ser observado simplificadamente na figura 3.1:



Figura 3.1: Princípio da aquisição terrestre e marítima de dados sísmicos. Figura obtida de (GERHARDT, 1998)

Após a aquisição, ocorre o processamento, cujos objetivos principais são restaurar perdas de energia do pulso, atenuar ruídos (coerentes e aleatórios) e obter a posição em que as ondas foram refletidas para superfície.

O traço sísmico é a forma em que o dado pode ser observado de forma 1D, variando ao longo do tempo. Para este trabalho é utilizado em milisegundos (ms) ao invés de profundidade em metros.



Figura 3.2: Traço sísmico.

Uma vez adquirida uma série de traços sísmicos é possivel formar uma seção sísmica 2D posicionando, lado a lado, traços com intervalo (espaçamento) regular. Essa seção sísmica é pode ser visualizada como uma imagem, aplicando-se uma escala de cores aos valores de amplitudes. Uma escala comumente usada é a cinza, em que o menor valor de amplitude aparece na cor branca e o maior valor amplitude aparece na cor preta (Figura 3.3).



Figura 3.3: Seção sísmica, com agrupamento de traços sísmicos individuais, em posições adjacentes e intervalo (espaçamento) regular.

Se o dado sísmico for 3D, ele também pode ser chamado de cubo sísmico. Neste caso, seções em diferentes direções podem se obtidas, com as mais comuns sendo as paralelas à direção de aquisição (*inline*), perpendicular a essa direção (*crossline*) e uma seção horizontal (*time*- ou *depth-slice*)



Figura 3.4: Cubo sísmico de uma aquisição 3D. São mostradas as principais direções em que seções são obtidas: *inline* (paralela à direção de aquisição), *crossline* (ortogonal à essa direção) e *Z-slice* (também chamado *time-* ou *depth-slice*), que é uma seção horizontal.

3.2 Atributos Sísmicos

A definição do que é um atributo sísmico pode variar dependendo do autor, mas usaremos aqui a estabelecida por Taner (2001). Segundo ele, atributos sísmicos são definidos como toda e qualquer informação extraída do dado sísmico, seja por medições diretas, lógicas ou por análises empiricas, segundo. Já Barnes (2001) entende que atributos sísmicos descrevem dados sísmicos e são referências específicas que quantificam características dos dados, representando subconjuntos da informação total. Na prática, os cálculos de atributos decompõem os dados sísmicos em atributos constituintes. Na verdade, qualquer quantidade calculada a partir de dados sísmicos pode ser considerada um atributo.

Contudo, é amplamente aceito na comunidade científica que os atributos sísmicos manifestam-se em diversas categorias, apresentando distintas qualidades e funcionalidades. Essa diversidade tem como objetivo facilitar a identificação de litologias e cenários geológicos específicos, além proporcionar suporte ao intérprete na formulação de modelos geológicos para áreas particulares. Tal abordagem contribui significativamente para um refinamento mais aprofundado da região de interesse, permitindo a extração de informações geológicas, estruturais e petrofísicas da subsuperfície (JANNUZZI, 2020). Os atributos promovem uma decomposição, ou avaliação alternativa, dos dados sísmicos originais de amplitude. Seu desenvolvimento contribui para a compreensão da subsuperfície, reduzindo as incertezas e riscos, apesar das falhas inerentes ao método. Entretanto, a interpretação dos dados sísmicos, a partir da perspectiva fornecida por um atributo, pode potencializar significativamente as informações do dado, revelando detalhes previamente "escondidos"antes da aplicação do atributo. Essa revelação, proporcionada pelos atributos, ocorre por meio da análise do sinal sísmico e da aplicação de técnicas de extração de informações, além da experiência do geocientista que realiza a interpretação dos dados.

3.3 Redes Neurais Recorrentes

As redes neurais apresentam diferentes arquiteturas que impactam significativamente seu desempenho em tarefas específicas. As redes neurais tradicionais, exemplificadas pelo perceptron multicamadas (MLP), operam no formato *feedforward*, caracterizando-se pela ausência de ciclos e retroalimentação. Esse modelo propaga a informação unidirecionalmente, da camada de entrada à camada de saída, gerando uma única saída para um dado vetor de entrada, independente do contexto. Essas redes são cruciais em diversas aplicações, sendo particularmente eficazes em tarefas que não exigem a consideração do contexto sequencial.

Por outro lado, as redes neurais recorrentes (*Recurrent Neural Networks*, RNN), ao considerarem o contexto em informações sequenciais, tornam-se essenciais em tarefas complexas como previsão de séries temporais ou reconhecimento de fala. Ao contrário das redes *feedforward*, as recorrentes executam a mesma tarefa para cada elemento sequencial, gerando saídas dependentes de cálculos anteriores. Esta abordagem assemelha-se a uma memória que registra computações prévias, permitindo, teoricamente, a utilização de informações em sequências de comprimento arbitrário. Na prática, no entanto, a análise costuma ser limitada a alguns passos anteriores. (LEVIN, 1990)

Além de seu papel em tarefas sequenciais, as redes neurais recorrentes

destacam-se em clusterização e detecção de anomalias. Sua capacidade de identificar similaridades e dissimilaridades ao agrupar exemplos no espaço vetorial as torna valiosas em diversas aplicações.

Nos modelos de redes recorrentes, o valor alcançado na etapa de tempo t é diretamente influenciado pelos valores da etapa de tempo t - 1. Dessa forma, essas redes possuem duas fontes de entrada: o presente e o passado recente, cuja combinação determina suas respostas. Uma representação visual das redes neurais recorrentes seria imaginar múltiplas cópias da mesma rede, cada uma transmitindo informações à sua subsequente. Essa configuração em cadeia evidencia a estreita relação das redes neurais recorrentes com sequências e listas, tornando essa arquitetura indispensável para esse tipo de dado.

A preservação da ideia de informação sequencial ocorre nas camadas escondidas da rede recorrente, possibilitando que a informação percorra vários intervalos de tempo à medida que avança no modelo, impactando o processamento de cada novo exemplo. Essa capacidade permite que o modelo identifique correlações entre eventos separados por longos períodos, sendo essas correlações referidas como "dependências de longo prazo". Em outras palavras, um evento no tempo torna-se dependente e é uma função de um ou mais eventos anteriores. (LEVIN, 1990)

3.3.1

Limitações das Redes Recorrentes Tradicionais

Como mencionado na seção anterior, as RNNs são amplamente utilizadas para modelar dependências temporais em dados sequenciais. Apesar que, em princípio, as RNNs possuem plena capacidade para lidar com "dependências de longo prazo", na aplicação prática, observa-se uma dificuldade aparente das RNNs em aprender essas dependências. Uma análise detalhada desse problema foi conduzida por Hochreiter et al. (2001). Outra dificuldade que essas redes enfrentam são desafios críticos durante o treinamento, notavelmente relacionados aos gradientes: os fenômenos de "*Exploding Gradients*"e "*Vanishing Gradients*". O "*Exploding Gradients*"ocorre quando os gradientes crescem exponencialmente, resultando em instabilidades durante o treinamento. Em contrapartida, o "*Vanishing Gradients*"ocorre quando os gradientes tornam-se extremamente pequenos, causando estagnação no aprendizado.

Esses fenômenos estão ligados à propagação da informação por múltiplos estágios nas RNNs. A regra da cadeia na diferenciação mostra que o gradiente final é o produto das derivadas em cada estágio. Se esse produto for menor que 1, elevado a uma potência alta, ocorre o "*Vanishing Gradients*". Se for maior que 1, tende a infinito, resultando no "*Exploding Gradients*". Essa propriedade

torna as RNNs sensíveis à escolha inadequada de parâmetros, impactando a estabilidade e eficiência do treinamento.

Soluções para esses desafios foram exploradas, incluindo métodos como truncamento e esmagamento de gradientes para lidar com o "*Exploding Gradients*". No entanto, o "*Vanishing Gradients*"é mais complexo de abordar, exigindo estratégias como arquiteturas de memória especializadas, entre as quais, para este trabalho, se destaca a arquitetura *Long-Term Short Memory* (LSTM).

3.4 Long Short-Term Memory, LSTM

As redes de *Long Short-Term Memory* (LSTM) foram desenvolvidas para solucionar as limitações de modelos mais simples ao lidar com dados sequenciais. Desenvolvidas para solucionar o problema de "*Vanishing Gradients*" das RNNs, as LSTMs destacam-se por sua capacidade de capturar dependências temporais mais longas (HOCHREITER; SCHMIDHUBER, 1997).

O elemento central das LSTMs é a célula de memória, uma estrutura projetada para armazenar informações ao longo de extensos intervalos temporais. São um total de três portões: de entrada, de esquecimento e de saída. Esses portões regulam o fluxo de informações na célula de memória. (HOCH-REITER; SCHMIDHUBER, 1997)

O portão de entrada decide quais informações da entrada atual devem ser armazenadas, é utilizada uma função de ativação sigmoide, produzindo valores entre 0 e 1. Valores próximos de 0 indicam que a informação não deve ser incluída, enquanto valores próximos de 1 indicam relevância. Enquanto o portão de esquecimento determina quais dados antigos devem ser descartados, assim como o portão de entrada, é usado uma função de ativação sigmoide para decidir o que esquecer e o que reter. Por fim, o portão de saída controla a parte da célula de memória que será utilizada como saída. Uma função de ativação sigmoide para determinar partes da célula de memória a serem enviadas como saída, e uma função de ativação tanh para normalizar os valores.

A célula de memória é atualizada multiplicando a entrada pelo portão de entrada e adicionando a informação mantida após a aplicação do portão de esquecimento. Esse processo possibilita que as LSTMs capturem e preservem informações relevantes ao longo de sequências temporais complexas.

3.5

Método de Seleção de Parâmetros Supervisionados Baseado em Multiatenção para Séries Temporais Multivariadas

No contexto da seleção de atributos, o artigo do pesquisador Cao et al. (2021) apresenta uma abordagem única chamada método de seleção de atributos supervisionados baseado em múltipla atenção (m-AFS). Este método vai além das técnicas convencionais de seleção de atributos, utilizando mecanismos de atenção para identificar a importância dos atributos em dados de séries temporais multivariadas (MTS). Esta seção discute o funcionamento e a aplicação deste mecanismo para o uso neste trabalho.

O m-AFS aborda atenção de multiplas séries semporais por meio da geração de pesos de atributos e pesos de faixas temporais. Sua implementação envolve duas camadas distintas, conhecidas como AoT (*Attention over Time*), que gera os pesos em função do tempo e AoV (*Attention over Variates*), que gera os pesos em função dos atributos. Uma vez que esses pesos são calculados é possivel gerar uma matriz de pesos global.

A figura 3.5 mostra a aquitetura desta camada de atenção para um problema de multi-atributo em séries temporais:



Figura 3.5: Estrutura conceituais da m-AFS. Figura obtida de (CAO et al., 2021)

Como apresenta na Figura 3.6, a camada AoV concentra-se em atribuir pesos para a dimensão de atributos, avaliando quais deles são mais relevantes para a tarefa de aprendizado. Esta etapa de avaliação dos atributos ocorre ao longo das faixas temporais, de tal forma que, para cada amostra de tempo, temos que $t_j = \{x_1(j), x_2(j), ..., x_m(j)\}$ como entrada, onde j é o passo de tempo, x_i são os atributos e m é o total de atributos. Nessa arquitetura temos uma primeira camada densa (denominada E). Essa camada E costuma ter um tamanho menor de dimensionalidade do que m. Com essa compressão é esperado que certas variáveis redundantes sejam descartadas durante este processo. Em seguida, extraindo a saída da camada E passando por uma ativação de tangente hiperbólica e usando como entrada, temos uma rede rasa U. Essa rede U corresponde ao número de variáveis. A saída desta rede U após a ativação de *softmax* representa os pesos das variáveis para o passo de tempo j. Por fim, é possivel concatenar todas as saídas dessas camadas AoV para cada faixa temporal formando uma matriz da atenção das variáveis (A_v) (CAO et al., 2021).



Figura 3.6: Estrutura da atenção para Variáveis. Figura obtida de (CAO et al., 2021)

Já camada AoT, por sua vez, lida com a dimensão temporal dos dados. Como apresenta na Figura 3.7, a camada AoT concentra-se em atribuir pesos para a dimensão de tempo, avaliando qual é mais relevantes para a tarefa de aprendizado. Esta etapa de avaliação dos atributos ocorre ao longo dos atributos, de tal forma que, para cada atributos, temos que $x_i =$ $\{x_i(1), x_i(2), ..., x_i(d)\}$ como entrada, onde *i* é o atributo que dado foi cortado, *x* são os atributos, *d* é o total de passos temporais. Nessa arquitetura temos uma primeira camada densa (denominada E). Essa camada E costuma ter um tamanho menor de dimensionalidade do que *d*. Com essa compressão é esperado que certas faixas temporais redundantes sejam descartadas durante este processo. Em seguida, extraindo a saída da camada E passando por uma ativação de tangente hiperbólica e usando como entrada, temos uma rede rasa U. Essa rede U corresponde ao número de passos temporais. A saída desta rede U após a ativação de *softmax* representa os pesos de cada passo temporal para cada atributo *i*. Por fim, é possivel concatenar todas as saídas dessas camadas AoT para cada atributo formando uma matriz da atenção temporal (A_t)(CAO et al., 2021).



Figura 3.7: Estrutura da atenção para Faixas Temporais. Figura obtida de (CAO et al., 2021)

O próximo passo é a geração a matriz de pesos globais A_g . Para isso, realizamos o produto de Hadamar, denotado por \odot , com as matrizes $A_v \in A_t$

$$A_g = A_v \odot A_t$$

A matriz A_g pode ser multiplicada sobre o dado original X para gerar um dado ponderado G. Esse dado ponderado por sua vez é dado como entrada para se treinar o modelo desejado.

$$G = A_q \odot X$$

3.6 Métricas

Para analisar modelos relacionados a problemas de classificação, a literatura utiliza diversas métricas. Estas métricas possibilitam a comparação entre modelos gerados por diferentes algoritmos ou por um mesmo algoritmo com diferentes configurações de parâmetros. Dessa forma, essas métricas são cruciais para identificar a configuração mais eficaz de um modelo de classificação.

No contexto da avaliação da metodologia proposta neste trabalho, a segmentação é feita a cada amostra do traço sísmico, assim para validar, cada predição de amostra do traço sísmico se encaixa em uma de quatro categorias: Positivo Verdadeiro (*True Positive, TP*), Negativo Verdadeiro (*True Negative, TN*), Positivo Falso (*False Positive, FP*), ou Negativo Falso (*False Negative, FN*).

Como a quantidade de traços com gás é relativamente menor do que a quantidade de traços sem gás e a extensão rotulada com presença de gás, por traço, é significativamente pequena quando comparada ao traço completo, haverá uma baixa presença de positivos. Esse desbalanceamento nos dados pode indicar um grande números de casos TN, o que pode tornar algumas métricas, como a acurácia, pouco confiaveis. Portanto para esta avaliação¹, o objetivo será reduzir os NF e os FP, assim como maximizar os TP . Duas métricas convencionais que avaliam esses casos são *Precision* e *Recall*.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3-1}$$

$$Precision = \frac{TP}{TP + FP}$$
(3-2)

Porém, como o intuito é de tentar maximizar essas duas métricas simultaneamente, a principal métrica avaliada é o F1-Score, que por sua vez é a média harmônica entra essas duas métricas.

$$F1-Score = \frac{2 |Recall * Precision|}{|Recall | + |Precision|}$$
(3-3)

¹Para uma explicação sobre a interpretação probabilística das métricas de avaliação consultar (GOUTTE; GAUSSIER, 2005)

4 Metodologia

Nesta seção será detalhado o dado sísmico usado, seus rótulos e o fluxo para se obter predições e uma visualização sobre os atributos para os quais o modelo deu um maior peso.

4.1 Visão Geral

A abordagem segue a ideia proposta por Santos e Gattass (2020), onde um modelo LSTM é usado para analizar linhas sismicas 1D para segmentação de gás natural. Diferenciando deste método, é usado como entrada multiplos atributos sismicos ao invés de apenas amplitude. Além disso é acrescentado um mecanismo de atenção customizado para atribuir, automaticamente, pesos aos atributos de entrada, distinguindo entre os mais relevantes para segmentação de gás natural.

O fluxograma da metodologia pode ser resumido em: geração dos atributos, pré-processamento do dado, definição da região de interesse, normalização do traço sísmico, entrada na camada de atenção, entrada no modelo e predição. Uma versão simplificada desse fluxograma pode ser visualizada na Figura 4.1 abaixo:



Figura 4.1: Síntese do *pipeline* da metodologia

4.2 Conjunto de Dados

O dado utilizado pertence a empresa brasileira de energia Eneva S.A. . Este dado sísmico é com migração em tempo antes do empilhamento. O dado fornecido é a amplitude e servirá como um dos parâmetros de entrada para o modelo. O dado é composto por elementos 2D e 3D. Os elementos 3D são localizados na Bacia do Amazonas, os 2D no Complexo de Parnaiba.

4.3 Extração de atributos

Para trabalhar com atributos sísmicos é preciso gera-los. Para este proposito é usado o *software Opendtect* versão 6.6, que é uma ferramente especializada e amplamente utilizada na industria de óleo e gás para diversas tarefas de interpretação. A partir dos dados de amplitude são gerados um conjunto de atributos sísmicos. Os atributos selecionados para serem gerados são: Amplitude, Cosseno da Fase Instantânea, Fase Instantânea, Frequência Instantânea e Energia. Essa escolha de atributos foi feita baseada no trabalho Cunha et al. (2023) e na recomendação de experts na área de identificação de gás natual da Eneva S.A., e foi aplicado para dados 2D e 3D. Portanto, geramos cinco volumes distintos como dado de entrada, cada representando um desses atributos.



Figura 4.2: Processo de extração de atributos. (a) Volume sísmico, (b) *Software Opendtect*, (c) Atributos.

4.4 Pré-processamento do Dado

Uma vez que temos o dado sísmico e todos seus atributos gerados, precisamos realizar algumas etapas antes de alimentar o modelo para treino. Todas as etapas devem ser realizadas para cada um dos atributos sísmicos.

4.4.1 Normalização do Dado Sísmico

Existem diversas formas de se tratar o dado sísmico: estandardização, normalização, aumento de dados, entre outras transformações. Neste processo aplicamos apenas a normalização do dado para que ele fique entre os intervalos de [-1, 1]. Esta normalização ocorre de tal forma que o valor do zero permaneça igual tanto antes quanto depois da normalização, isso é importante pois o valor zero indica mudança do material na camada litológica. Para que isso ocorra, é selecionado o maior valor absoluto no dado sísmico e em seguida os valores do dado sismico são divididos por este maior valor. O dado é normalizado no volume sísmico completo antes de se aplicar a RoI no dado. Já o dado 2D, cada linha é normalizada individualmente após a aplicação do corte de RoI. Todas essas etapas de normalização seguem de forma individual para cada atributo sísmico.

4.4.2 Região de Interesse

A primeira etapa é definição de uma região de interesse (RoI). A RoI é definida por um geocientista e será a região vertical do dado sísmico da qual será realizado um corte em termos de tempo. É um aspecto crucial para aprimorar a qualidade dos dados. Essa delimitação da RoI tem como objetivo primordial reduzir o ruído presente nos dados e concentrar a análise em áreas com maior probabilidade de conter gás natural. Através desse delineamento estratégico, é possível direcionar a atenção para as áreas com maior relevância geológica, onde a presença de gás natural é mais provável.



Figura 4.3: Visualização da Inline 96 do 3D na Bacia do Amazonas antes da definição da janela de tempo baseada na RoI. Distância entre xlines de 30 m.

Para efetuar essa análise, é crucial a definição da janela de tempo apropriada para um dado instante, representado por t. A definição da janela de tempo é definida durante a criação da RoI e segue uma fórmula precisa: Dada o menor valor de tempo t de um horizonte definido por um especialista, escolhemos subir um total de X amostras para servir como inicio da RoI e em seguidas seguimos outras Y amostras, resultando em uma estrutura de dados com dimensões [Y, número de atributos].

Esse formato é essencial para garantir a correta representação das características temporais e de atributos, levando em consideração o contexto da detecção de gás natural.

Para este trabalho, O valor de X equivale a 25 amostras, ou 100ms para o caso do dado utilizado. Para o Y definimos 200 amostras, ou 800ms para o caso do dado utilizado, de tal forma gerando uma janela de com o total de 200 amostras começando de 25 amostras do topo da RoI para baixo. A amostragem do dado sísmico utilizado é de 1 amostra a cada 4ms.



Figura 4.4: Visualização da Inline 96 do 3D na Bacia do Amazonas antes e após definição da janela de tempo baseada na RoI. Distância entre xlines de 30 m.

Após todos estes passos, o formato do dado fica [200, 5], onde temos 200 amostras e 5 atributos.

4.5 Visualização da Predição

Para cada linha 2D, ou inline no caso do volume sísmico, é feita a predição do modelo traço a traço. Para tornar essa predição visivel, gera-se um mapa para sobrepor aos dados sísmicos originais e criar visualizações para as regiões em questão. Na Figura 4.5 pode-ser ver um exemplo de predição realizado num dado de teste 3D na Bacia do Amazonas. A linha vermelha indica a interpretação (marcação) para a ocorrência de gás realizada por um geocientista com experiencia na região. A linha verde indica a predição do model, e as linhas em roxo os limites superior (topo) e inferior (base) da RoI.



Figura 4.5: Visualização da predição para Inline 72 do 3D na Bacia do Amazonas. Observar ótima correlação entre a interpretação do especialista (em vermelho) com a predição (em verde). Distância entre xlines de 30 m.

4.6 Visualização da atenção nos atributos

Em muitos casos, aumentar a quantidade de informação não necessariamente leva a um desempenho melhor dos modelos de deep learning. Pelo contrário, além de aumentar o tempo de treinamento, isso pode piorar os resultados. Por outro lado, estimar a importância de quais features realmente ajudam a resolver o problema pode ser uma ferramenta poderosa para melhorar o desempenho do modelo, maximizando a influência das features importantes enquanto minimiza o impacto das features redundantes.

No contexto da seleção de features, Cao et al. (2021) apresenta uma abordagem chamada método de seleção de features supervisionado com múltiplas atenções (m-AFS). Este mecanismo é explicado em detalhes na seção 3.5.

Para o caso deste trabalho, no intuito de concentrar o foco apenas na relevância dos atributos, utilizamos uma versão customizada deste mecanismo de atenção. Para esta customização, usamos apenas matriz de pesos para as variáveis A_v . Assim gerando exclusivamente os pesos de contribuição para cada atributo. Outra mudança realizada, na camada E do bloco AoV, foi utilizado um total de 100 neuronios densos ao invés do que o autor recomenda, que seria inferior ao número de atributos, neste caso, deveriam ser menos de 5 neuronios.

Para visualizarmos esses pesos, primeiro treinamos o modelo em um conjunto de dados, em seguida extraimos e congelamos os pesos da camada de atenção do modelo. A seguir, pegamos a matriz de peso A_v e fazemos o produto de Hadamar com o dado de entrada I gerando a matriz de dado ponderado W.

$$A_v \odot I = W$$

A partir dessa matrizes, é possivel produzir visualizações para o mecanismo de atenção. Neste trabalho exploramos duas formas: o dado ponderado em forma de mapa de calor e o peso atribuido a cada atributo.

A Figura 4.6 abaixo apresenta a Inline 96 do dado 3D na Bacia do Amazonas. Cada imagem à esquerda apresenta o atributo do dado sísmico, enquanto cada imagem à direita mostra o peso dado pela camada de atenção a cada traço de cada atributo sísmico de forma individual, equivalente a matriz W. A indicação vertical é número da amostra (e não tempo em milisegundos) e a vertical número da *Crossline*.

Essa visualização tem a vantagem de poder ser interpretada por um especialista, uma vez que ela apresenta regiões em particular que tem um maior foco em um atributo, como por exemplo o cosseno de fase e energia tem um impacto mais relevante próximo de uma estrutura presente no intervalo vertical de 0-25 da Figura 4.6.



Figura 4.6: Visualização da Inline 96 do 3D na Bacia do Amazonas por atributo (esquerda) e seu peso da atenção (direita).

Alternativamente, podemos realizar a média dessa atenção por atributo para uma inline, ou linha sísmica, ou qualquer conjunto de traços que for desejado, seria o equivalente a uma média de matrizes A_v para entrada desejada. Como podemos ver na figura 4.7 abaixo, essa visualização é uma média da saída camada AoV da atenção, resultando na matriz de pesos para a inline 96 do 3D na Bacia do Amazonas.



Figura 4.7: Visualização da média da saída da camada AoV com a Inline 96 do dado 3D na Bacia do Amazonas como entrada. (a) Amplitude, (b) Cosseno da Fase Instantânea, (c) Fase Instantânea, (d) Frequência Instantânea, (e) Energia. Quanto mais azul, mais próximo o peso dado ao atributo se aproxima de 0. Quanto mais vermelho, mais próximo o peso dado ao atributo se aproxima de 1. A soma de todos os pesos por amostra somam 1.

A vantagem dessa visualização com relação a Figura 4.6 é a concentração de informação passada ao usuário. Isso permite uma interpretação mais objetiva a relevância dos atributos para o modelo.

5 Resultados

Neste trabalho analizamos duas regiões sísmicas: na Bacia do Amazonas, cujo dado é 3D e no Complexo de Parnaiba, cujo dado é 2D. Os resultados dos testes realizados foi usando um total de cinco atributos sísmicos: Amplitude, Cosseno de Fase Instantânea, Frequencia Instantânea, Fase Instantânea e Energia. Essa escolha de atributos foi feita baseada no trabalho de Cunha et al. (2023) e foi aplicado para o 2D e 3D. Do mesmo trabalho, foi usado os resultados como *baseline* para o caso do dado 3D na Bacia do Amazonas. No caso do 2D, estaremos usando como *baseline* o caso de usar apenas a amplitude sísmica como dado de entrada sem camada de atenção.

Para o 3D, o dado é dividido em cinco *folds* (sub-divisões) diferentes para que seja realizado *cross-validation*. Esses *folds* são divididos por *inline* em intervalos de aproximadamente 20 *inlines* começando na *inline* 33 até a *inline* 137. Para cada *fold* testado, são usados os outros *folds* para treinar e validar o modelo. Podemos ver esse processo ilustrado na Figura 5.1 abaixo.



Figura 5.1: Divisão por *folds* da região do 3D na Bacia do Amazonas. O mapa mostra a interpretação sísmica para a extensão da ocorrência de gás. A dimensão aproximada de cada fold é 1,8 km na direção NW por 12,9 km na direção NE.

No caso do Complexo de Parnaiba (2D), usamos um total de três campo, dois desses campos rotulados são usados para treinar os modelos em dados sísmicos 2D, esses campos podem ser vistos na Figura 5.2 abaixo. Esses campos foram selecionada devido à sua proximidade geografica e por possuir dados rotulados. Para garantir a integridade do treino, parte dos dados dos Campos A e B foram usado para testes, porém seus resultados não entraram para as métricas.

Na Figura 5.2, pode-se observar os campos usados e as linhas sísmicas 2D empregadas. As linhas em verde foram usada para o processo de treino e as linhas em vermelho foram usadas para testar os modelos.



Figura 5.2: Região no Complexo de Parnaiba. Mapa em coordenadas UTM (zona 23S), sistema de projeção SIRGAS 2000. Escala transalada para (0,0) por confidencialidade.

5.1 Resultados 3D na Bacia do Amazonas

Para avaliar o 3D na Bacia do Amazonas foi utilizado os resultados do trabalho de Cunha et al. (2023) apresentados na tabela 5.1 abaixo. Inicialmente foram realisados testes utilizando o mecanismo de atenção utilizando todos os atributos disponiveis. Neste experimento foi obtido uma melhora significativa quando comparada ao U-NET ou aos resultados de Cunha et al. (2023).

Uma vantagem do mecanismo de atenção é a capaciade de verificação da relevânciade algum atributo específico. Nesse caso, podemos ver na Figura 5.3 abaixo que o atributo da terceira coluna, Fase Instantânea, é o atributo cuja camada de atenção atribui menor peso. Fazendo um novo experimento removendo a Fase Instantânea podemos observar uma melhora significativa quando comparado a qualquer um dos outros resultados. Comparando os resultados do baseline, do mecanismo de atenção sem o atributo de menor peso, pode-se ver uma melhora de 13,94% na métrica.

Tabela 5.1: Resultados comparativo de usar multi-atributos e atributo único

F1 Score	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
U-Net (Amplitude)	0.6323	0.8079	0.5537	0.7275	0.7402	0.6923
Cunha et al. (2023) (Amplitude) - Baseline	0.4413	0.7187	0.7306	0.7285	0.5349	0.6308
Cunha et al. (2023) (All Attributes)	0.5060	0.7177	0.7415	0.7556	0.5661	0.6574
Cunha et al. (2023) (All Attributes without Instantaneous Phase)	0.5272	0.7198	0.7316	0.6872	0.5502	0.6432
Model with MTS Attention Layer (All Attributes)	0.7287	0.7945	0.6645	0.7809	0.7322	0.7402
Model with MTS Attention Layer (All Attributes without Instantaneous Phase)	0.7740	0.7977	0.6796	0.8265	0.7733	0.7702



Figura 5.3: Pesos da atenção para o 3D na Bacia do Amazonas. (a) Amplitude, (b) Cosseno da Fase, (c) Fase Instantânea, (d) Frequência Instantânea, (e) Energia. Quanto mais azul, mais próximo o peso dado ao atributo se aproxima de 0. Quanto mais vermelho, mais próximo o peso dado ao atributo se aproxima de 1. A soma de todos os pesos por amostra somam 1.

5.2 Resultados 2D, Campo C no Complexo de Parnaiba

Para avaliar o Campo C, foi se realizado um experimento para servir como *baseline*. Dentro desse experimento, o modelo não utiliza camada de atenção e treina apenas com a amplitude como atributo. Para verificarmos se o uso de multiplos atributos contribui para o problema, temos também um experimento onde o modelo não utiliza camada de atenção, porém passa a usar todos os 5 atributos durante seu treino. Por fim, temos um experimento onde o modelo treina com camada de atenção e todos os atributos afim de verificar se a camada gera algum ganho no processo de aprendizado. Tabela 5.2: Experimento 1: resultados comparativos entre não usar atenção e apenas um atributo, não usar atenção e todos os atributos selecionados, e usar atenção junto de todos os atributos selecionados

Metrics	Model NO Attention Layer	Model NO Attention Layer	Model with MTS Attention Layer
	(Amplitude) - Baseline	(Amplitude + Attributes)	(Amplitude + Attributes)
F1-Score	0.3521	0.3723	0.3720

Similarmente ao caso do 3D, podemos observar uma melhora com o uso do multi-atributos, porém sem aumentos com relação ao uso da camada de atenção.

Para realizar a verificação da relevância dos atributos, para o caso do Campo C, é feita a média de todas as linhas da região para conseguirmos produzir a Figura 5.4 abaixo:



Figura 5.4: Pesos da atenção para o Campo C. (a) Amplitude, (b) Cosseno da Fase, (c) Fase Instantânea, (d) Frequência Instantânea, (e) Energia. Quanto mais azul, mais próximo o peso dado ao atributo se aproxima de 0. Quanto mais vermelho, mais próximo o peso dado ao atributo se aproxima de 1. A soma de todos os pesos por amostra somam 1.

Para este experimento, optamos por remover a Fase Instantânea uma vez que é um dos atributos de menor relevância, uma vez que o Cosseno da fase é muitas vezes utilizado para verificar continuidade lateral (RODRIGUES, 2019). Dois modelos foram treinados com todos os atributos menos o cosseno da fase. Um dos modelos vai ter a camada de atenção, enquanto o outro não. Os resultados deste experimento podem ser vistos na tabela 5.3 abaixo: Tabela 5.3: Experimento 2: resultados comparativos entre não usar atenção e apenas um atributo, não usar atenção e todos os atributo menos a fase instantânea, e usar atenção junto de todos os atributos selecionados menos a fase instantânea

Metrics	Model NO Attention Layer (Amplitude) - Baseline	Model NO Attention Layer (Amplitude + Cossine Phase + Insta Frequency + Energy)	Model with MTS Attention Layer (Amplitude + Cossine Phase + Insta Frequency + Energy)
F1-Score	0.3521	0.3729	0.3730

Comparando os resultados do Experimento 1 com o Experimento 2, podemos verificar que não há uma melhora mesmo removendo um atributo de baixa relevância. Porém, é ainda possível de se obter resultados comparativos com menos informação passada para o modelo.

6 Conclusão e Trabalhos Futuros

Neste trabalho, exploramos a detecção de acúmulos de gás natural em dados sísmicos por meio do uso de um modelo baseado em múltiplos atributos. Utilizamos dados 3D na Bacia do Amazonas e 2D no Complexo de Parnaiba e aplicamos uma abordagem que incorpora cinco atributos sísmicos: Amplitude, Cosseno de Fase, Fase Instantânea, Frequência Instantânea e Energia.

Os melhores resultados foram visto no caso do 3D, onde foi possivél melhorar sobre o trabalho de Cunha et al. (2023) o valor do *F1-Score* em 13,94%, utilizando a camada de atenção e realizando um *fine tunning* manual dos atributos utilizados. Os objetivos estimar a relevância dos atributos foi atingida assim como sua seleção, apesar de manual.

Enquanto no 2D obtivemos um entendimento melhor a respeito dos atributos por meio de sua visualição, não possivel melhorar os resultados. Isso pode ser devido a uma série de fatores, tais quais, uma escolha não ideal dos atributos selecionados, um volume inferior na quantidade de traços, uma qualidade inferior do dado sismico, foi usado regiões não exploradas no cálculo das métricas, podendo ter atrapalhado assim as métricas calculadas.

Como principal contribuição do trabalho destacamos a incorporação de uma camada de atenção no modelo, permitindo verificar a relevância de cada atributo em diferentes partes dos dados sísmicos além de sua seleção de pesos automática. A análise da atenção revelou que certos atributos, como a Fase Instantânea, têm menos relevância, sugerindo potenciais ajustes na escolha dos atributos.

Mesmo com resultados satisfatórios, acreditamos que na metodologia proposta ainda existem algumas evoluções que podem alcançar melhores resultados. Para trabalhos futuros pode-se investigar e experimentar com diferentes atributos sísmicos e alterar quantos atributos são utilizados de uma vez. Alternativamente pode-se aprofundar a análise da camada de atenção e considerar variações no tamanho de suas redes internas. Uma outra possível evolução é explorar o aspecto temporal da M-AFS com relação ao dado sísmico. Buscar uma forma automatica de escolher quais atributos são usados para treinar, uma vez que este processo continua manual.

7 Referências bibliográficas

ALFARHAN, M.; DERICHE, M.; MAALEJ, A. Robust concurrent detection of salt domes and faults in seismic surveys using an improved unet architecture. **IEEE Access, 10, 39424-39435.**, 2020. Citado 2 vezes nas páginas 13 e 15.

ALSALMI, H.; ELSHEIKH, A. H. Automated seismic semantic segmentation using attention u-net. **Geophysics**, Society of Exploration Geophysicists, v. 89, n. 1, p. WA247–WA263, 2024. Citado 2 vezes nas páginas 14 e 16.

ANDRADE, F. et al. Gas reservoir segmentation in 2d onshore seismics using lstmautoencoder. **First International Meeting for Applied Geoscience & Energy**, Society of Exploration Geophysicists, p. 1651–1655, 2021. Citado 2 vezes nas páginas 14 e 15.

BARNES, A. E. Seismic attributes in your facies. **CSEG recorder**, v. 26, n. 7, p. 41–47, 2001. Citado na página 21.

CAO, L. et al. A multiattention-based supervised feature selection method for multivariate time series. **Computational Intelligence and Neuroscience**, ACM, v. 2021, p. 10, 2021. Citado 7 vezes nas páginas 9, 14, 17, 25, 26, 27 e 34.

CHOPRA, S.; MARFURT, K. J. Seismic attributes—a historical perspective. **Geophysics**, 70.5: 3SO-28SO, 2005. Citado na página 13.

CHOROWSKI, J. K. et al. Attention-based models for speech recognition. **Advances in neural information processing systems**, v. 28, 2015. Citado na página 17.

CUNHA, H. et al. Multi-attribute seismic data for natural gas detection using lstm. **84th EAGE Annual Conference & Exhibition**, European Association of Geoscientists & Engineers, v. 2023, n. 1, p. 1–5, 2023. Citado 7 vezes nas páginas 14, 16, 30, 38, 40, 41 e 46.

DING, M.; ZHOU, Y.; CHI, Y. Self-attention generative adversarial network interpolating and denoising seismic signals simultaneously. **Remote Sensing**, MDPI, v. 16, n. 2, p. 305, 2024. Citado na página 16.

GERHARDT, A. L. B. Aspectos da visualização volumétrica de dados sísmicos. **Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ**, 1998. Citado 3 vezes nas páginas 9, 19 e 20.

GOUTTE, C.; GAUSSIER, E. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. **Lecture Notes in Computer Science**, p. 345–359, 2005. Citado na página 28.

HOCHREITER, S. et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. 2001. Citado na página 23.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural com-putation**, n. 8, p. 1735–1780, 1997. Citado na página 24.

JANNUZZI, J. Aplicação e análise de atributos para a interpretação sísmica 3d no pré-sal do novo campo de jubarte, parque das baleias, bacia de campos. **Monografia (Graduação em Geofísica) — Universidade Federal Fluminense**, 2020. Citado na página 22.

KIM, Y. et al. Dense text-to-image generation with attention modulation. In: **Proceedings of the IEEE/CVF International Conference on Computer Vision**. [S.I.: s.n.], 2023. p. 7701–7711. Citado na página 17.

LEVIN, E. A recurrent neural network: Limitations and training. **Neural Networks, 3(6), 641-650.**, 1990. Citado 2 vezes nas páginas 22 e 23.

LING, J. **Coarse-to-fine attention models for document summarization**. Tese (Doutorado), 2017. Citado na página 17.

POCHET, A. et al. Seismic fault detection using convolutional neural networks trained on synthetic poststacked amplitude maps. **IEEE Geoscience and Remote Sensing Letters 16**, n. 3, p. 352–356, 2018. Citado 2 vezes nas páginas 13 e 15.

REIS, N. C. Classificação de fácies sísmicas utilizando multiatributos sísmicos. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 2022. Citado 3 vezes nas páginas 13, 14 e 16.

ROBINSON, E. A.; TREITEL, S. Geophysical signal analysis. Society of Exploration Geophysicists, 2000. Citado na página 19.

RODRIGUES, C. Notas de curso interno tecgraf. 2019. Citado 2 vezes nas páginas 19 e 44.

SANTOS, L. F.; GATTASS, M. Natural gas detection in onshore data using transfer learning from a lstm pre-trained with offshore data. **SEG Technical Program Expanded Abstracts 2020**, Society of Exploration Geophysicists, p. 1190–1195, 2020. Citado 3 vezes nas páginas 13, 15 e 29.

TANER, M. T. Seismic attribute. **CSEG recorder**, v. 26, n. 7, p. 48–56, 2001. Citado na página 21.

THOMAS, J. E. et al. Fundamentos de engenharia de petróleo. **Interciência**, Petrobrás, Rio de Janeiro, Brasil, 2001. Citado na página 19.

VASWANI, A. et al. Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017. Citado 2 vezes nas páginas 16 e 17.

YOON, D.; YEEH, Z.; BYUN, J. Seismic data reconstruction using deep bidirectional long short-term memory with skip connections. **IEEE Geoscience and Remote Sensing Letters**, v. 18, n. 7, p. 1298–1302, 2021. Citado na página 15.

YU, J.; WU, B. Attention and hybrid loss guided deep learning for consecutively missing seismic data reconstruction. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 60, p. 1–8, 2021. Citado na página 15.