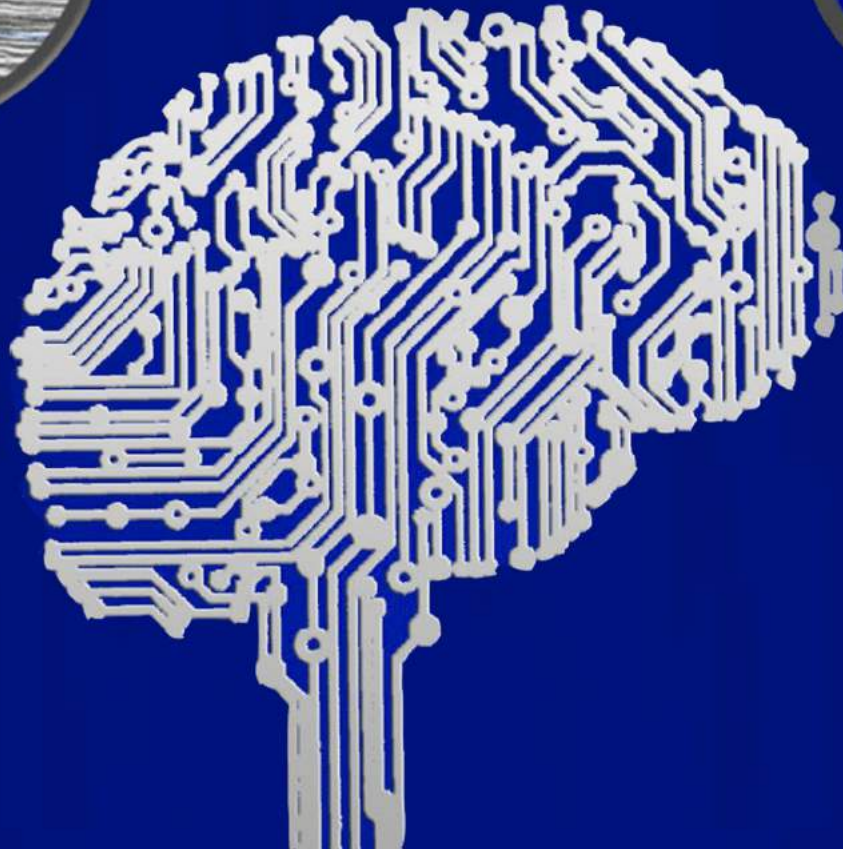
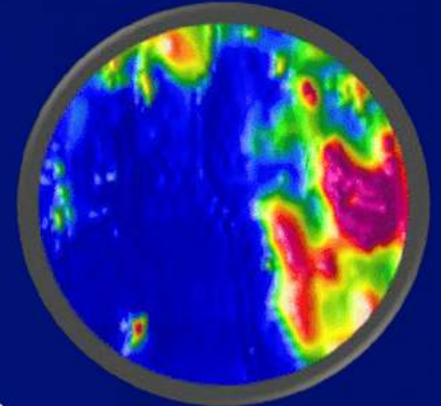
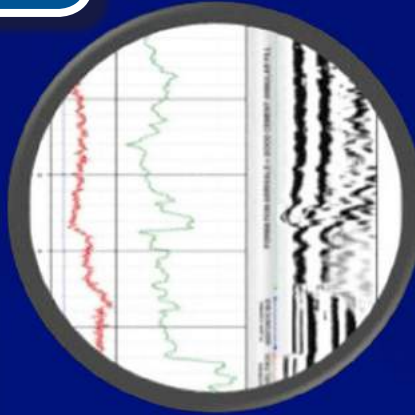
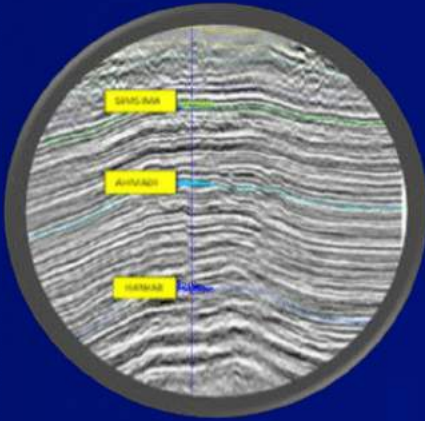


# boletim **SBGf**

Publicação da Sociedade Brasileira de Geofísica  
Número 104 – Julho/Agosto 2018 ISSN 2177-9090



# MACHINE LEARNING

# Machine Learning: O avanço das técnicas de Inteligência Artificial

Por Alan Cunha

Diversas áreas do conhecimento têm sido impactadas pelo avanço das técnicas de Inteligência Artificial (AI), porém, entre a vitória histórica do supercomputador "Deep Blue" sobre o enxadrista Gary Kasparov e os dias atuais, o conceito de AI tem sido expandido rapidamente. O "Machine Learning" (ML), ou aprendizado de máquina, em bom português, tenta "emular" a capacidade humana de aprendizado. Ainda no universo dos neologismos, o "Deep Learning" (DL ou aprendizado profundo) tem revolucionado a interação entre "máquinas" e sistemas complexos, e tarefas antes impensadas para execução "robótica" são agora feitas por mais um APP na tela de seu smartphone. Nossa rotina já experimenta a interação com algoritmos ainda que nossas profissões não tenham relação direta com a indústria de tecnologia. E como será impactada a Geofísica com a introdução dessas novas tecnologias? Somos uma indústria altamente tecnológica, trabalhamos com grandes volumes e variedades de dados, e decisões tomadas a partir de dados geofísicos são sempre altamente complexas. Serão elas facilitadas? Que tarefas são hoje executadas por geofísicos que num futuro próximo poderão ser substituídas por algoritmos, modificando também o mercado de trabalho? Há setores ou especialidades mais ou menos protegidas dentro das geociências? Quantos robôs serão necessários para uma reunião de locação de um poço exploratório? Quantos algoritmos precisarão ser escritos e executados para um menor impacto de nossas atividades ao meio ambiente? A única certeza que temos é que o caminho em direção ao uso intensivo destas tecnologias parece irreversível e que vivemos uma janela única de oportunidade de substituição de afazeres humanos subsidiada por processadores ultra rápidos e capacidades de armazenamento robustas para digitalização de nosso meio físico. Em outras palavras, estamos no pico da tempestade perfeita desta revolução e este Boletim tenta no meio dela municiar o leitor com diferentes olhares sobre como o "cérebro eletrônico" (sic) tem afetado e afetará a forma de fazer geofísica. Boa leitura.

## ADMINISTRAÇÃO DA SBGf

Presidente

Neri João Boz

Vice-presidente

José Agnelo Soares

Secretário-Geral

Guilherme Sidou Canha

Secretário de Finanças

Simplicio Lopes de Freitas

Secretário de Relações Institucionais

Ricardo Augusto Rosa Fernandes

Secretário de Relações Acadêmicas

George Sand Leão Araújo de França

Secretário de Publicações

Luiz Fernando Santana Braga

Conselheiros

Ellen de Nazaré Souza Gomes

Adalene Moreira Silva

Adriana Perpétuo Socorro da Silva

Augusto Cesar Bittencourt Pires

Marco Cesar Schinelli

Marco Antonio Pereira de Brito

Renato Cordani

Rosângela Correa Maciel

Sergio Luiz Fontes

Secretários Regionais

Josibel Gomes Junior (Nordeste Setentrional)

Carolina Barros da Silva (Norte)

Pedro Mário Cruz e Silva (Centro-Sul)

Welitom Rodrigues Borges (Centro-Oeste)

Maximilian Freis (Sul)

Susana Silva Cavalcanti (Nordeste-Meridional)

Editor-chefe da Revista Brasileira de Geofísica

Cleverson Guizan Silva

Assistente de Diretoria

Luciene Victorino de Carvalho

Assistente Administrativo

Ivete Berlice Dias

Coordenadora de Eventos

Renata Vergasta

Editora de publicações científicas

Adriana Reis Xavier

Técnico de Informática

Gabriel Nunes Dias

BOLETIM SBGf

Editor-chefe

Alan Cunha

Jornalista Responsável

Rosimeri Figueiredo

Registro: MTb16045/RJ

Edição gráfica

Juliana Lima de Souza

Tiragem: 1.000 exemplares

Distribuição restrita

Também disponível no site [www.sbgf.org.br](http://www.sbgf.org.br)

Sociedade Brasileira de Geofísica – SBGf

Av. Rio Branco, 156 sala 2.509

20040-901 – Centro – Rio de Janeiro – RJ

Tel./Fax: (55-21) 2533-0064

[sbgf@sbgf.org.br](mailto:sbgf@sbgf.org.br) | [www.facebook.com/sbgf.org](http://www.facebook.com/sbgf.org)

## CONFIRA NESTA EDIÇÃO

### 3 NOTÍCIAS

- Inscrições abertas para os cursos durante o VIII Simpósio

### 4 EVENTOS

- Eventos apoiados pela SBGf

### 5 NOTAS

- Lançada 2ª edição do livro "Análise do Sinal Sísmico" de André Luiz Romanelli

### 7 ENTREVISTA ESPECIAL

- VIII SimBGf em Salinópolis

### 8 MEMÓRIA DA GEOFÍSICA

Estórias da Amazônia - Transglobe

### 10 PUBLICAÇÕES

- Livros SBGf

### 12 ESPECIAL

- Machine Learning

### 18 ARTIGO TÉCNICO 1

Deteção de estruturas em dados sísmicos com Deep Learning

### 22 ARTIGO TÉCNICO 2

Classificação litológica com dados incompletos

## FUNDO SBGf

### DIAMANTE



PETROBRAS

### OURO



HALLIBURTON

### PRATA



### BRONZE



Spectrum





## NOTÍCIAS

## Anuidade 2018: Evite acréscimo pagando nos prazos definidos



Aproveite os valores desse período para pagamento da sua anuidade:

De 01/07 a 31/12:  
 Associado: R\$ 120,00  
 Estudante: R\$ 55,00

## Inscrições abertas para os Cursos durante o VIII Simpósio Brasileiro de Geofísica



Interessados em participar dos cursos oferecidos durante o Simpósio poderão se inscrever no dia 18/09. Vagas limitadas.

Os temas do curso são:

- Petrofísica Digital, Prof. Dr. Agnelo Soares, Universidade de Campina Grande;
- Geofísica Rasa, Dr. Luiz Antonio Pereira Silva, IPT/SP;
- Exploração Mineral, Prof<sup>a</sup> Dr<sup>a</sup> Adalene Moreira Silva, Universidade de Brasília;
- Imageamento, Dr. Martin Tygel, Unicamp.

Informações no site [www.sbgf.org.br](http://www.sbgf.org.br).

Atualize seu cadastro  
no site

[www.sbgf.org.br](http://www.sbgf.org.br)

Curta a SBGf  
no facebook!



[www.facebook.com/sbgf.org](https://www.facebook.com/sbgf.org)

**SERVIÇOS E PRODUTOS**  
 no setor de **ÓLEO E GÁS** e mineração

+55 21 3535 - 9664 | [www.geohub.com.br](http://www.geohub.com.br) | [contato@geohub.com.br](mailto:contato@geohub.com.br)

## Workshop on Least Squares Migration



A SBGf, junto à EAGE, realizará entre os dias 27 e 29/11, no Grand Mercure Rio Copacabana, o *Workshop on Least Squares Migration*, que propõe discussões e atualizações sobre o estado da arte em técnicas de imageamento sísmico na área de óleo e gás.

O evento reunirá importantes representantes destas empresas e discutirá novas tecnologias para melhoria do imageamento do dado sísmico, com a finalidade de abrir caminho para novos investimentos em técnicas de aquisição sísmica e computação de alto desempenho.

## Diretoria da SBGf no 49º CBG



O 49º CONGRESSO BRASILEIRO DE GEOLOGIA, realizado de 20 a 24 de agosto deste ano, no Centro de Convenções Sul América, no Rio de Janeiro teve como objetivo debater temas relevantes das geociências. Simultaneamente ocorreram o 9º Simpósio do Cretáceo do Brasil e o VII Simpósio de Vulcanismo e Ambientes Associados.

Foram selecionadas 16 sessões temáticas de geofísica, coordenadas por: Marco Latgé (Petrobras/SBGf), Luiz Braga (SBGf) e Ricardo Rosa (Petrobras). Nas Sessões Temáticas, foram discutidos os avanços recentes na Geofísica no Brasil em diferentes escalas e os avanços metodológicos alcançados por diferentes grupos de pesquisa. Três grandes temas foram abordados: (1) Estrutura e dinâmica da Placa Sul-Americana e de suas bacias sedimentares; (2) Avanços metodológicos e integração de métodos Geofísicos de Exploração; (3) Aplicações em Geofísica Rasa e Geofísica de Poço, abrangendo as áreas de Geofísica Ambiental, Biogeofísica e Petrofísica.

A SBGf participou ativamente como apoiadora do evento e na exposição de livros em seu estande.

## II SBPA e o XI Encogercos acontecem em outubro



O II Simpósio Brasileiro de Praias Arenosas (II SBPA) e o XI Encontro Nacional de Gerenciamento Costeiro (XI ENCOGERCO), eventos promovidos pelas instituições UFSC, UFPE E UFRG, estão programados para acontecer entre os dias 15 e 10 de outubro, no SESC CACUPÉ, em Florianópolis.

Os seminários terão quatro atividades principais: palestras, mesas redondas, trabalhos técnico-científicos e saída de campo. Informações sobre datas para envio de trabalhos e prazo para inscrições, acesse o site <http://www.praiaegestao.com.br/site>

## Evento Rio Oil & Gas 2018



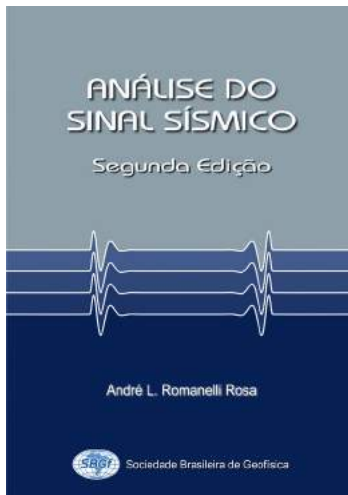
Sob o lema: "Transformando Desafios em Oportunidades", a 32ª Edição da Rio Oil & Gas Expo and Conference será realizada entre os dias 24 e 27 de setembro de 2018, no Rio de Janeiro.

Terá como palestrantes internacionais o CEO global da TechnipFMC, Doug Pferdehirt, no dia 24/9, o presidente e CEO da Murphy Oil Corporation, Roger W. Jenkins no dia 26/9 e o presidente da ExxonMobil Exploration Company e Vice-Presidente da ExxonMobil Corporation, Stephen Greenlee, no dia 27/9.

Não perca esta oportunidade de fazer parte do maior evento da indústria na América Latina. Conheça a programação no site <http://www.riooilgas.com.br/#info>

## NOTAS

## Lançada 2ª edição do livro "Análise do Sinal Sísmico" de André Luiz Romanelli



O geofísico André Luiz Romanelli Rosa é Consultor Master na Petrobras, graduado em Geologia pela UnB com mestrado em Geofísica pela Universidade de Houston.

A SBGf, com o patrocínio da Petrobras, lançou a segunda edição do seu livro "Análise do Sinal Sísmico". O objetivo principal desta edição,

uma evolução da primeira, continua sendo o de preencher a lacuna frequentemente observada entre os textos teóricos publicados e a atividade rotineira dos geofísicos da indústria do petróleo. A apresentação dos temas possibilita diferentes níveis de leitura, desde a mais básica, de caráter eminentemente intuitivo, até a que favorece, em alguns casos, a elaboração de programas de computador.

Perguntado sobre qual público este livro atinge, Romanelli afirma que, em tese, ele pode ser útil para geofísicos em geral e professores, além dos estudantes de graduação e pós-graduação em Geofísica. Outros profissionais interessados na teoria da elasticidade e sua relação com os fluidos que saturam as rochas também podem reutilizar dos ensinamentos da obra. Para Romanelli, o que difere a primeira da segunda edição são correções de pequenos erros, um aprofundamento de vários temas, alguns temas novos, tudo isso representado por um acréscimo de 45 páginas. As diferenças mais significativas dizem respeito à anisotropia, física de rocha, AVO e propagação de ondas, tanto do ponto de vista teórico, quanto do prático. A bibliografia também teve sua forma substancialmente melhorada, com a introdução do DOI (sigla em inglês para Identificador Digital de Objetos).

Indagado sobre a nova versão em inglês publicada pela SEG, cita que desde a sua primeira edição, publicado pela SBGf, em 2010, passou a atualizar o texto em português e em inglês, atendendo a uma sugestão de Fred Hilterman. O livro, publicado há poucos dias, teve seu título modificado a pedido da SEG, porque *Seismic Signal Analysis* soa muito similar a *Seismic Data Analysis*, livro de Oz Yilmaz publicado pela SEG no início da década passada. O novo título é *The Seismic Signal and Its Meaning*.

"Do meu ponto de vista, espero que ele seja realmente tão útil quanto minhas pretensões nesse sen-

tido", afirma Romanelli, se referindo às expectativas para esta segunda edição.

Sobre a repercussão da 1ª edição, explica que não houve uma preocupação específica sobre qual área abordada ele deveria se aprofundar para atender aos seus leitores, uma vez que o material do livro vem sendo continuamente atualizado para atender às necessidades da Petrobras, seja para os cursos internos, seja para consulta por seus profissionais. Acrescenta ainda que no Prefácio, muitos colegas contribuíram direta ou indiretamente para a construção e revisão do livro. Uma pessoa em particular exerceu papel especial, ao incentivá-lo a publicar tanto esta, quanto a primeira edição do livro: Paulo Siston. Destacou também o papel da própria Petrobras, que vem promovendo de diversas formas o crescimento e a disseminação do conhecimento em geociências.

**INVISION**  
geophysics

**TOCANDO O INVISÍVEL**

TECNOLOGIA GEOFÍSICA APLICADA AO E&P

- Processamento sísmico
- Pré-condicionamento
- Caracterização de reservatório
- Inversão sísmica

EQUIPE ESPECIALIZADA EM PROJETOS DE P&DI

[www.invisiongeo.com.br](http://www.invisiongeo.com.br)





VIII  
**SimBGf**  
novos desafios da Geofísica

# Simpósio Brasileiro de Geofísica

18 a 20/SETEMBRO/2018

Praia do Atalaia  
Salinópolis - Pará  
Hotel Clube Privê do Atalaia

Patrocínio  
Institucional:



Universidade  
Federal do Pará

Realização:



Patrocínio:



Prefeitura  
Municipal de  
Salinópolis

<http://simposio.sbgf.org.br>

# Entrevista

Por Rosimeri Figueiredo



O Simpósio Brasileiro de Geofísica (SimBGf) é o evento nacional promovido pela SBGf que ocorre nos anos intercalados aos do Congresso Internacional da SBGf. A realização do VIII SimBGf em Salinópolis, no Pará, segue o caráter itinerante deste evento e ressalta a dinâmica da geofísica na região. A Coordenadora do VIII SimBGf, **Carolina Barros da Silva**, concedeu entrevista ao Boletim e nos revelou como está sendo os últimos preparativos.

## Boletim SBGf - Qual o objetivo do Simpósio?

Eu vejo o Simpósio como um evento cujo principal objetivo é contribuir para o desenvolvimento da pesquisa em Geofísica no País, pois ele oportuniza a interação entre alunos, professores-pesquisadores e profissionais de Geofísica.

O Simpósio, além da apresentação de resultados de trabalhos de pesquisa, gera um ambiente de reflexão e discussão além de possibilitar novas parcerias. É o único encontro nacional desse porte na área de Geofísica, sendo dessa forma fundamental para pesquisa na área.

## Boletim SBGf - Fale um pouco de sua atividade como geofísico e de sua atuação na SBGf.

Minha formação é em Geofísica de Poço e trabalho atualmente com a parte de Petrofísica, além disso sou a Coordenadora do Programa de Pós-Graduação em Geofísica da UFPA e tutora do PET-Geofísica, programa financiado pelo MEC com a participação de 12 alunos de graduação.

Na SBGf atuo como secretária da Regional Norte. As atividades da Secretaria consistem na divulgação da Geofísica e das diversas atividades da SBGf, fomentar ações da regional norte no sentido de integrar seus associados, além de trabalhar em ações para agregar o maior número de alunos, professores e profissionais de Geofísica à SBGf. Teremos uma melhor representatividade, como Geofísicos, se atuamos juntos e a SBGf vem de encontro a esses interesses de nossa classe.

## Boletim SBGf - Como você recebeu o convite para presidir o VIII Simpósio Brasileiro de Geofísica?

Fiquei feliz. A região norte está geograficamente muito afastada das principais atividades de Geofísica, que estão em geral no sudeste, assim, trazer o simpósio para cá foi com a intenção de aproximar mais a Sociedade e seus associados de outras partes do país aos grupos daqui que fazem Geofísica. A Sociedade deve

ter sempre em foco esse trabalho de agregar seus associados de diferentes regiões do país. Principalmente aqueles grupos mais afastados do sudeste, pois estes já contam com o Congresso Internacional que já acontece por lá.

## Boletim SBGf - Quais são as suas expectativas para este evento?

Antes de mais nada espero que o Simpósio cumpra seu papel de facilitador sendo um espaço adequado para conversarmos sobre Geofísica. Espero também poucos *no show* e, por fim, que as pessoas possam também aproveitar e curtir um pouco de Salinópolis.

## Boletim SBGf - Qual a importância de organizar o Simposio em Salinópolis?

Em Salinópolis foi criado recentemente o Curso de Engenharia do Petróleo, com ele o estado do Pará conta com três cursos que envolvem Geofísica (os outros dois de Graduação em Geofísica). Assim realizar o Simpósio aqui facilita a participação dos alunos dessas três graduações, que teriam pouco contato com outros grupos de fora do Estado.

## Boletim SBGf - Como é Salinópolis? O que as pessoas que nunca foram ao local encontrarão no evento e no local?

É um dos balneários mais importantes da Estado do Pará, uma cidade bucólica com alto potencial turístico e que recentemente recebeu um Campus da UFPA. As pessoas encontrarão belas praias e bons restaurantes com comidas típicas de praia. Além de ter o contato com a natureza bem típica dessa região.

## Boletim SBGf - Há uma expectativa de receber quantas pessoas?

Tendo em vista a quantidade de trabalhos submetidos, temos a expectativa de receber em torno de 200 pessoas.



Cadastre-se no LinkedIn  
da SBGf



## Estórias da Amazônia - Transglobe

Por Roberto Breves Vianna, geólogo, geofísico desde 1963.

brevevianna.vianna2@gmail.com

Certamente todos os geólogos e geofísicos que trabalharam na Amazônia, nas décadas de 60 e 70 conheceram os famosos rádios Transglobe da Philco, cuja foto reproduzo abaixo:



Mais do que tudo, obter os rádios era o sonho de consumo de todos os trabalhadores, tanto os da equipe de geologia, quanto os de sísmica, que adentravam a mata e acampavam, geralmente à beira de igarapés ou rios maiores, dependendo do programa.

Eventualmente, o esturro da onça, felizmente à distância, podia ser ouvido, mas certo mesmo era o silêncio quebrado pela algazarra dos guaribas ao nascer do dia e pelo som do rádio Transglobe, geralmente lançando ao ar o vozeirão do Valdick Soriano... Não era fácil acordar ao som de “eu não sou cachorro não”.

À noite os trabalhadores se reuniam em torno de quem tinha os famosos tijolões, preferencialmente os que tinham a caixa plástica de cor preta, para ouvir notícias de seus familiares, transmitidas via radiofônica.

O rádio no Amazonas prestava um serviço de utilidade pública, divulgando as viagens dos ribeirinhos que saíam ou chegavam às suas cidades. Horários de barcos, internamento ou alta hospitalar de familiares, nascimentos e falecimentos, tudo se podia ouvir através do Transglobe Philco. Eram nove faixas de ondas entre AM e FM, possibilitando a escuta de emissoras situadas a longas distâncias, mas o importante mesmo era sintonizar as rádios amazonenses.

Pois bem, a estória verdadeira que conto agora aconteceu com uma equipe sísmica que se encontrava lá por meados dos anos 60, realizando um levantamento 2D, ao longo da fronteira Brasil-Peru, acompanhando o curso do rio Javari, ao sul de Tabatinga. (vide mapa abaixo).



Corria o boato, nas equipes sísmicas, que as maiores sacanagens eram praticadas por topógrafos ou enfermeiros.

Minha tendência é acreditar, dadas às oportunidades que tive de comprovar presepadas oriundas destes profissionais. Nada contra, eram excelentes profissionais, mas não perdiam a oportunidade de aproveitarem da ingenuidade de seus parceiros.

Já estávamos na região há algum tempo, quando chegou à equipe um auxiliar de topografia orgulhosamente exibindo um Transglobe, luzindo de novo e comprado em Manaus, antes que o Catalina o trouxesse para a base.

À noite, o auxiliar todo prosa, reuniu os colegas e ligou o Transglobe. Ficou tentando em todas as nove faixas do aparelho, mas nada de conseguir sintonia com rádios do Brasil. O aparelho só sintonizava rádios do Peru e Colômbia, em alto e bom som espanhol.

Desacoroçado, o auxiliar perguntou ao topógrafo o que poderia ter havido. Foi sua perdição...

Fingindo que entendia de eletrônica, o topógrafo ficou sintonizando as várias frequências, aumentando e diminuindo o volume e após alguns minutos de escuta deu seu veredito:

- Companheiro, onde você comprou este rádio?

- Foi em uma loja de Manaus, respondeu o auxiliar já se sentindo passado para trás.

- Pois meu camarada, tu foste lesado, te venderam um aparelho que só fala espanhol, você devia ter comprado radio brasileiro, que transmite em português, redarguiu o topógrafo.

Ouviram-se algumas imprecações e até ameaças de porrada no vendedor, além das gozações da peãozada que classificou o novato de “muito burro”. Se fosse nos dias de hoje, iam dizer que era bullying, mas naqueles tempos era gozação mesmo e todos entendiam assim.

Mas o nosso amigo topógrafo, de caso pensado e contando com a cumplicidade dos demais, propôs um negócio para seu auxiliar.

- Olha, eu amanhã saio de folga, daqui vou para Tabatinga e de lá pego o avião da Varig para Iquitos, onde tenho uma namorada peruana. Me interessa comprar seu rádio que fala espanhol. Você aceitaria a metade do que pagou, te dou o dinheiro agora mesmo!!

Vislumbrando a possibilidade de diminuir o que julgava ser seu prejuízo, o auxiliar aceitou na hora e o negócio se concretizou a vista de dezenas de testemunhas.

No dia seguinte, no retorno do PT-AXL Catalina da Pannair, a serviço da Petrobras, o topógrafo embarcou, mas rumo a Manaus, de onde iria para Belém, curtir as rádios que transmitiam em português...

Quanto ao auxiliar, ao descobrir o engodo em que havia caído, jurou pegar de porrada o topógrafo, mas acho que eles nunca mais cruzaram seus caminhos, na imensidão dos projetos na Bacia Amazônica.



## Boletim SBGf tem novo editor-chefe



O Boletim SBGf, a partir desta edição, passará das mãos do dr. Renato Silveira - que o conduziu por quase 20 anos - para Alan Cunha, nomeado novo editor chefe do periódico. Cunha é Físico e Engenheiro de Petróleo com mestrado e doutorado em Geociências na UFRJ. Atuando no mercado como geofísico desde 2006.

Atualize seu cadastro  
no site  
[www.sbgf.org.br](http://www.sbgf.org.br)

### Make Better Decisions on your Brazil Obligations

#### Santos Vision, Brazil

- Area 1: Now Available
- Areas 2-5 Early Out: Now Available

Better pre-salt images using advanced processing and imaging techniques.

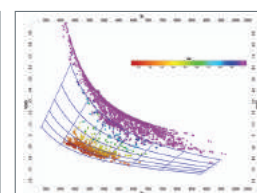
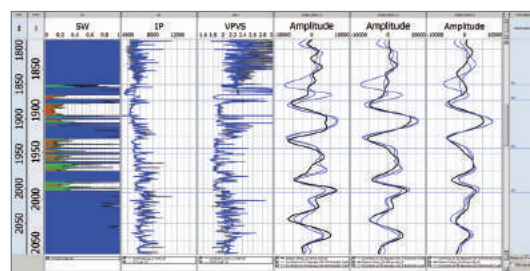
Email: [nsa.info@pgs.com](mailto:nsa.info@pgs.com)

A Clearer Image | [www.pgs.com/SantosVision](http://www.pgs.com/SantosVision)



AT EVERY STAGE

## Connect Seismic and Rock Properties With Rock Physics



Example of a rock physics template on a crossplot of P-impedance versus  $V_p/V_s$  ratio.

Input well curves compared to synthetic and seismic response at the well location.

**RockSI™** explores the link between rock properties and seismic data, and offers modeling for inputs into geomechanics and seismic inversion:

- Generate elastic property logs - even in wells with no measured sonic logs
- Use rock physics templates to graphically understand the reservoir and rocks
- Employ Monte-Carlo simulations and uncertainty analysis for time-lapse feasibility studies
- Create robust interpretations for AVO 1D/2D modeling, attribute analysis and inversion

Rock physics technology from CGG GeoSoftware connects geology, petrophysics and geophysics to gain a clearer understanding of the reservoir - a powerful advantage throughout the field lifecycle.

[cgg.com/rockphysics](https://www.cgg.com/rockphysics)

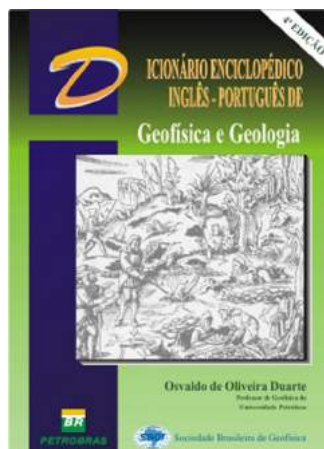


## PUBLICAÇÕES

# Livros SBGf



SBGf: três décadas promovendo a geofísica



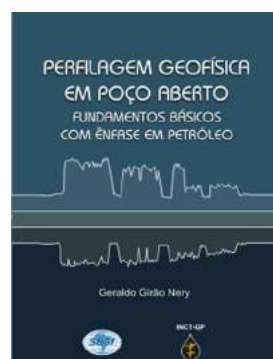
Dicionário Enciclopédico Inglês-Português de Geofísica e Geologia



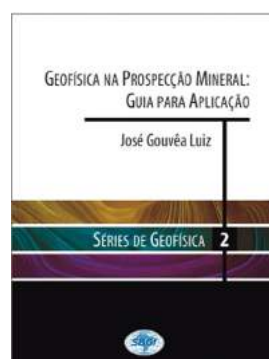
Fundamentos de Física para Geociências



Fundamentos do Método Magnetotelúrico na Exploração de Hidrocarbonetos



Perfilagem Geofísica em Poço Aberto: Fundamentos básicos com ênfase em petróleo



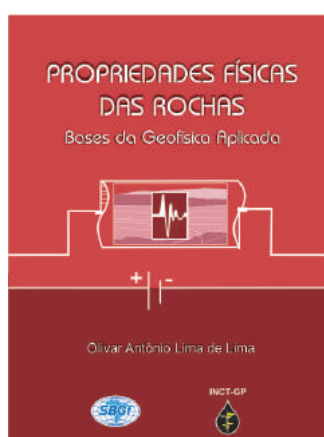
Geofísica na Prospecção Mineral: Guia para Aplicação



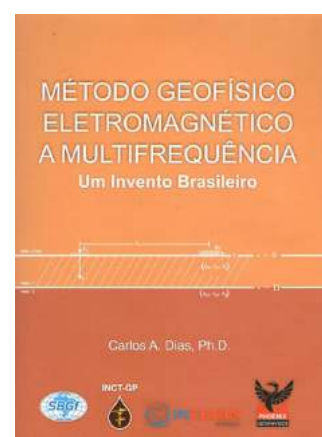
Séries de Geofísica 1 "Considerações sobre a Aquisição Sísmica Multiclientes no Brasil: Aspectos Legais"



Análise do Sinal Sísmico



Propriedades físicas das rochas



Método Geofísico Eletromagnético a Multifrequência - Um Invento Brasileiro



A Sociedade Brasileira de Geofísica agradece  
às empresas que participam do Fundo SBGf  
de Apoio à Ciência Geofísica

2018

Diamante



Ouro



Prata



Bronze



# Machine Learning

Bruno Pereira Dias / Edson Borin / Telma Aisengart / Fernando Bordignon / Ricardo Rosa / Ghassan AlRegib / Matt Hall  
Por Alan Cunha

O tema *machine learning* discutido por Bruno Pereira Dias, Edson Borin, Telma Aisengart, Fernando Bordignon, Ricardo Rosa, Ghassan AlRegib e Matt Hall nos leva a refletir sobre os impactos das técnicas Inteligência Artificial na geofísica. As aplicações do *machine learning* têm revolucionado diversas áreas de conhecimento, e suas atuações relacionadas à geofísica favorecem a compreensão de relacionamentos complexos entre um conjunto grande e diversificado de variáveis.

Indagados sobre como a Geofísica será ou está sendo alterada pelas técnicas de *machine learning*, Bruno Pereira define a geofísica como uma área que



trabalha com gigantescas quantidades de dados fazendo uso intensivo de computação em larga escala, apresentando assim, muitas oportunidades de melhorias dos processos. A automatização dos fluxos de processamentos e inversão, através da determinação automática de parâmetros, assim como facilidades para interpretação, serão as primeiras aplicações a aparecer na Geofísica. Complementa que a grande mudança nesta área virá com a melhor integração de dados estruturados e não estruturados de diferentes métodos (sísmicos e não sísmicos) e em diferentes escalas (rocha, perfil e sísmica).



Por sua vez, Edson Borin diz que muito do trabalho realizado por geofísicos e geólogos ainda é baseado em um *know-how* que é difícil de se codificar em forma de equações e métodos, o que inviabiliza a automação deste trabalho com algoritmos e programas de computador. Por outro lado, as técnicas de *machine learning* permitem que o computador aprenda a partir de exemplos (dados), o que permitirá que este *know-how*, que pode ser visto como “a experiência do geofísico”, seja aprendido e aplicado automaticamente por sistemas computacionais. Ainda que de forma bem pontual, técnicas de *machine learning*, por exemplo *k-means*, máquinas de vetores de suporte e redes neurais artificiais, já eram utilizadas na Geofísica. De fato, algumas atividades de processamento sísmico mais simples, como *picking* de primeira quebra, já são parcialmente automatizadas em ferramentas comerciais com o auxílio destas técnicas. Espera-se que os avanços recentes na área, principalmente em *deep learning*, permitam a automação de atividades de processamento cada vez mais sofisticadas.



No parecer de Telma Aisengart, a geofísica tem uma tendência natural para o *machine learning* já que lida com grande volume de dados – o tão falado *Big Data* – que é a base para aplicação do *machine learning*. Enquanto que na indústria

de petróleo o maior desafio é relacionado ao grande volume de dados, na mineração a maior dificuldade é quanto à variedade de dados. Com o aumento da capacidade computacional, o acesso a algoritmos de inversão de dados gerando modelos 3D de propriedades físicas como densidade, vetor de magnetização, resistividade, cargabilidade entre outros tem um grande impacto na capacidade humana para analisar e obter um modelo único da Terra. A aplicação de algoritmos de aprendizado de máquina para a redução de parâmetros é adequada neste ambiente, permitindo a identificação e classificação de anomalias que ajudam na busca de novos depósitos minerais.



Fernando Bordignon diz que a revolução que se vê atualmente em *machine learning* é devido ao uso de redes neurais com aprendizado profundo (*deep learning*). Essas redes precisam de um conjunto extensivo de dados para serem treinadas, diferente das técnicas citadas anteriormente, as quais trabalham com registros de poucos poços, por exemplo. Assim como a área de interpretação está sendo afetada pela chamada interpretação quantitativa, acredito que a geofísica também adotará modelos construídos baseados em menos interpretação e mais nos dados. O uso das técnicas de *deep learning* irá trazer essa possibilidade, uma vez que extrair certos parâmetros ou certas relações era inviável com metodologias tradicionais.



Para Ricardo Rosa todas as áreas do conhecimento estão sendo alteradas pela inteligência artificial e uso de sistemas que aprendem com os dados. Na Geofísica, os métodos tradicionais de definição de alvos exploratórios, de redução de risco e mesmo de solução de sistemas complexos (sistemas com muito mais variáveis que equações) estão sendo impactados positivamente pelas rotinas que aprendem com os dados e permitem chegarmos a soluções únicas pelo conhecimento acumulado em muitos outros lugares semelhantes. Essas soluções podem ser muito mais rápidas e precisas que as soluções convencionais e demandam menor influência humana, reduzindo custos e acelerando as decisões de investimento.

Ghassan AlRegib considera que, com a AI, os Geofísicos terão uma ferramenta mais eficaz que pode aju-





dá-los a tomar decisões com mais precisão e rapidez. Esta tem sido nossa visão desde que lançamos nosso centro de pesquisa em AI para geofísica na *Georgia Tech* há cinco anos. Além disso, ao longo das últimas décadas, muitos avanços ocorreram nos aplicativos que usam processamento vetorial e reconhecimento de padrões. Há conexão entre essas tecnologias e a geofísica ajudará a todos. Finalmente, acredito que, no futuro próximo, começaremos a ver a comunidade mais aberta para compartilhar conhecimento, dados e experiências. Assistimos a essas mudanças nos últimos cinco anos e esperamos mudanças mais drásticas nos próximos anos.



Matt Hall concorda que vai mudar de várias maneiras que não podemos prever, e é por isso que este é o momento mais emocionante para ser um geofísico desde o início do processamento de dados digitais. Se você aceitar que o FWI é basicamente *deep learning*, então, claramente, a AI já mudou o processamento sísmico. Mas, além disso, a IA está mudando a maneira como pensamos sobre a coleta de dados e o gerenciamento de dados, e como estruturamos os problemas empresariais e científicos. E com certeza vai mudar a forma como abordamos a interpretação e análise de dados. E a IA está definitivamente mudando a maneira como as empresas pensam sobre *marketing*! (Isso é uma piada.)

Em relação aos novos requisitos para o profissional que queira trabalhar com Geofísica no futuro, Bruno Pereira revela que a evolução tecnológica e das metodologias continuará avançando em um ritmo acelerado. Assim, um requisito primordial ao geofísico é ser capaz e estar aberto ao aprendizado contínuo. A disponibilidade de informação hoje em dia é bastante abundante. Portanto, o profissional que souber absorver esse conhecimento de modo autodidata ou a partir da interação com outros profissionais se posicionará com um diferencial para adoção de novas tecnologias. Especificamente relacionado às aplicações de *machine learning*, conhecimento de estatística e programação serão cada vez mais importantes para os futuros geofísicos.

Edson Borin considera que o trabalho de um geofísico é muito semelhante ao de um cientista de dados no sentido de que ambos têm que constantemente formular hipóteses sobre os dados e projetar e executar experimentos para corroborar ou refutar estas hipóteses. Com a popularização e proliferação das técnicas de *machine learning* para processamento sísmico, o geofísico do futuro terá à sua disposição mais ferramentas para executar estes experimentos. Além disso, métodos de aprendizado supervisionado ou não, auxiliarão o geofísico no processo de formulação das hipóteses através da detecção automática de padrões no dado. Para que o geofísico possa fazer uso eficaz dessas novas tecnologias, ele terá que aprender como utilizá-las, selecionando as técnicas mais adequadas para cada situação e evitando as armadilhas que podem levar a conclusões erradas.

Telma Aisengart afirma que será ainda mais importante a compreensão dos algoritmos aplicados aos dados: "O entendimento dos algoritmos de *machine learning* permitirá a escolha do algoritmo mais adequado a ser aplicado, assim como para a validação dos conjuntos de dados a serem considerados e dos resultados encontrados", afirma Aisengart.

Fernando Bordignon explica que o profissional deve entender os principais modelos de *machine learning* e *deep Learning*, quais são suas capacidades e limitações, além dos requisitos necessários para utilizá-los, como custo computacional e volume de dados necessário para treinamento. Com essa visão geral, o geofísico poderá enxergar os problemas que poderiam ser resolvidos com as novas ferramentas de *machine learning*.

"Os profissionais com conhecimento em programação também levam vantagem na hora de encarar os desafios técnicos que estão sendo construídos pelas rotinas que envolvem inteligência artificial e *machine learning*", complementou Ricardo Rosa.

Para Ghassan AlRegib a ciência de dados é fundamental: "embora a definição de ciência de dados seja muito ampla, acredito que um Geofísico precisa estar equipado com a mais recente teoria e ferramentas de *machine learning*, a fim de enfrentar os desafios atuais e futuros. Isso requer que universidades e escolas se concentrem mais em programas multidisciplinares. Defendo programas de ciência e engenharia em que um aluno de Geofísica, por exemplo, é obrigado a se inscrever em vários cursos de *machine learning*", conclui.

Em maio de 2018 foi feita uma pesquisa para 120 pessoas na AAPG, em *Salt Lake City*, diz Matt Hall. "É claro que muitas pessoas mencionaram que queriam aprender mais sobre *machine learning* e *deep learning*", esclarece Hall. Foram listadas os dez principais assuntos que as pessoas disseram que precisavam aprender: *Python*, programação em geral, estatística, matemática, análise econômica, visualização de dados, física de rochas, análise de séries temporais, visão computacional, escrita e comunicação.

Sobre como será a integração entre analistas, cientistas dos dados e o geofísico, Bruno Pereira afirma que o desenvolvimento da *pilha de software* (conjunto integrado de programas que realizam uma determinada tarefa) assim como as análises de dados tem sido feitos de maneira cada vez mais rápida e dinâmica. Isso permite uma integração entre analistas, cientistas de dados e geofísicos quase que em tempo real. Em conjunto, os especialistas dessas áreas poderão aproveitar o máximo de desenvolvimento e aprimoramento das ferramentas de *machine learning*. Porém, é crucial que haja a manutenção de canais de comunicação e incentivo a *feedback* permanente.

Edson Borin destaca que o cientista de dados é um especialista que faz uso de diversas ferramentas de análise de dados, incluindo técnicas de *machine learning*, para extrair informações de valor dos dados. A tendência é que estas ferramentas sejam cada vez mais aplicadas para extrair informações de dados sísmicos. Neste contexto, a experiência do geofísico será um fator fundamental no desenvolvimento e validação das novas técnicas de processamento sísmico, o que exi-



girá muita interação entre os profissionais destas áreas. Além disso, a aplicação de técnicas de *machine learning* para o processamento sísmico produzirá, em muitos casos, resultados inesperados. O processo de investigação para determinar se estes resultados são novos insights no campo da geofísica ou se são apenas artefatos produzidos por problemas na aplicação ou nas técnicas de *machine learning* em si, também exigirá uma interação próxima entre estes profissionais.

Para **Telma Aisengart**, o trabalho em equipes multidisciplinares já faz parte do dia-a-dia do geofísico de exploração, que deve trabalhar próximo ao geólogo, engenheiro de minas (ou de reservatório), entre outros profissionais. No futuro, com a inclusão de cientistas de dados e analistas na equipe, o papel do especialista – no caso o geofísico, de fazer as perguntas corretas e validar os resultados, ou seja, o conhecimento do significado dos dados e do negócio é o que permite a formulação do problema. Ao cientista de dados cabe testar e escolher o algoritmo mais adequado para solucionar o problema, enquanto o analista é responsável pela garantia de acesso a dados de qualidade.

Já para **Fernando Bordignon**, tendo geofísicos versados nas capacidades gerais das técnicas mais comuns de *machine learning*, será gerada demanda aos cientistas de dados e analistas, por modelos que resolvam os problemas trazidos pelos geofísicos. Posteriormente, se iniciará um trabalho conjunto para identificar melhor o problema, definir o resultado esperado e iniciar a coleta e tratamento dos dados. O resultado seria um *workflow* para solução desse problema específico, que precisa ser implantado como *software* para o uso do geocientista. Este cenário contempla soluções específicas, para um estágio inicial de adoção de *machine learning*. Eventualmente, partes deste processo serão automatizadas a medida que mais ferramentas são implementadas.

**Ricardo Rosa** conclui ainda que em todas as áreas do conhecimento, a quantidade de dados tem crescido enormemente, de forma exponencial, tornando o manuseio e interpretação desses dados uma rotina cada vez mais dependente de análise prévia com seleção de domínios, processamento e classificação. Dessa forma, analistas, cientistas de dados e geofísicos deverão trabalhar em conjunto para atacar os problemas de forma organizada, tratando da taxonomia dos dados, de sua organização e padronização, da sequência de processamento e de sua interpretação e classificação final.

**Ghassan AlRegib** pressupõe que isso já está acontecendo em alguns lugares ao redor do mundo e vamos testemunhar mais em breve.

**Matt Hall** explica que essa integração precisa acontecer em várias escalas. No nível pessoal, temos que começar a pensar em programação e uso de ferramentas de análise de dados como coisas tais incorporados intensivamente à sua rotina de trabalho. Eu ainda ouço pessoas falando que escrever códigos “não é meu trabalho real” – isso está errado. Programação é absolutamente parte de ser um cientista, assim como o trabalho de campo, o trabalho de laboratório e a escrita. A demanda será diferente para todos. No nível de equipe, precisamos começar a incluir dados e suporte a programação em nossas equipes – como

costumávamos incluir cartógrafos e desenhistas. Não apenas as equipes precisam ser compostas de maneira diferente, mas elas precisarão colaborar de maneiras novas e às vezes desconfortáveis – e os gerentes precisarão negociar essa realidade em mudança. No nível organizacional, precisamos dar suporte a todas essas mudanças. Isso significa explorar novas maneiras de criar equipes, construir infraestrutura e colaborar com outras organizações e até setores. Nada disso será fácil, e é improvável que acertemos da primeira vez. Então, se você ainda não está experimentando essas coisas, você tem trabalho a fazer! Finalmente, no nível da comunidade, também precisamos nos adaptar, as organizações técnicas precisam encontrar novas maneiras de conectar outras organizações e apoiar a comunidade profissional em mudança. Eu não estou vendo isso acontecer rápido o suficiente.

Pela experiência de outras áreas que estão se aproveitando do desenvolvimento em *machine learning* (exemplo, serviços *online*), para garantir a rápida adoção na indústria será necessário inicialmente uma verticalização das iniciativas e esforços em *machine learning*. Agregar os especialistas será essencial para se desenvolver as melhores soluções com tempo de resposta adequado para a indústria. Porém, somente isso não basta. Também é primordial que haja uma horizontalização nos acessos às bases de dados em diferentes áreas geográficas e de diferentes especialidades. Não é por acaso que a onda do *Big Data* precedeu a onda do *machine learning*, pois sem a organização e integração de bases de dados com extração das informações relevantes fica inviável o treinamento dos sistemas de aprendizado automáticos, revela **Bruno Pereira**, sobre como a indústria (que usa geofísica diretamente, como a indústria de óleo e gás ou mineração) deve se adaptar.

Na opinião de **Edson Borin**, em um horizonte curto, a indústria deverá investir na contratação de profissionais de computação com expertise em *machine learning* para facilitar a incorporação e validação das novas técnicas de processamento sísmico. A experiência no *High Performance Geophysics lab*, na Unicamp, mostra que os cientistas e engenheiros da computação não têm dificuldade para assimilar os conceitos básicos de geofísica e conseguem trabalhar de forma bem integrada com os geofísicos do grupo. A tendência é que, a médio prazo, a indústria invista na formação de profissionais que tenham conhecimento em geofísica e *machine learning*.

**Telma Aisengart**, concorda com os colegas de profissão e acrescenta que o crescimento do volume e da variedade dos dados tornam imprescindível a adoção de uma estratégia para busca e acesso aos dados dentro de uma organização. A análise dos dados e aplicação de algoritmos de aprendizado de máquina pressupõe que existem dados de qualidade a serem analisados e que sejam encontrados em tempo para as análises. A velha regra de dados ruins de entrada acarretando resultados ruins segue sendo a regra básica de qualquer análise ou processamento com estes dados.

As organizações que mais consomem e produzem ferramentas para *machine learning*, hoje em dia, são as



que possuem mais dados de seus produtos ou processos. **Fernando Bordignon** cita como exemplo óbvio o *Facebook*, empresa na qual regulamentações sobre o uso de dados produzidos pelos seus usuários afetam diretamente sua habilidade de gerar receita e seu valor de mercado. Portanto, a principal adaptação que deverá acontecer, antes mesmo de vencer a resistência dos profissionais mais conservadores, será a criação da cultura de manter os dados e conhecimentos gerados em um formato acessível para máquinas.

**Ricardo Rosa** também comenta que as maiores empresas de energia, seja na área de óleo & gás, mineração ou em energias alternativas, tem procurado ampliar a formação dos seus profissionais através de convênios com entidades que oferecem treinamento avançado em análise de dados, inteligência artificial e aprendizado de máquina. Aquelas com recursos internos já estão oferecendo essas disciplinas em seus cursos de formação, de forma a permitir que os profissionais exerçam todo potencial produtivo com utilização de conhecimento nessas áreas.

Cinco anos atrás, a indústria se recusou a ouvir muitas propostas para mudar a atitude em relação à ciência de dados. Nos últimos dois anos, isso mudou drasticamente, revela **Ghassan AlRegib**. Essa mudança pode ser depreendida das muitas discussões que tivemos com a indústria, bem como com as decisões de contratação nos últimos meses. Acredito que, para receber o máximo benefício, o setor precisa fornecer aos pesquisadores o que realmente precisam para desenvolver tecnologias de inteligência artificial: dados de código aberto, liberdade de compartilhar experiências e conhecimentos e, ainda, um foco menor em recursos computacionais, que já estão disponíveis. Uma maneira de fazê-lo é investir em novos tipos de consórcios que permitam aos membros inovar livremente sem muitas limitações.

**Matt Hall** afirma que a curto prazo, a indústria precisa contratar especialistas em *machine learning* e, em seguida, realmente pregar como eles os integram no negócio das Geociências. É muito importante que a experiência no assunto - isto é, conhecimento geocientífico e de engenharia - esteja presente em todos os experimentos, em todos os projetos. Então, a maioria dos gerentes, especialmente aqueles que, como eu, começaram suas carreiras antes de 2000, precisam se reinventar. Uma coisa que aparece muito nessa conversa são dados abertos, porque o *machine learning* e especialmente o *deep learning* precisam de muitos dados. Quase ninguém está levando a sério este problema - sendo a Equinor a única exceção que posso imaginar, com o recente lançamento do conjunto de dados da *Volvo*. Espero que o resto da indústria esteja prestando atenção a esta questão.

Na maioria das universidades já existem grupos de pesquisa na área de ciência de dados, com linhas de pesquisa em *machine learning*. Os outros grupos de pesquisa, em particular, na área de geofísica deverão fomentar integração entre os departamentos para aproveitar ao máximo da expertise já construída ao longo dos anos. Além disso, projetos de *machine learning* necessitam

de grande capacidade computacional e banco de dados, destacou **Bruno Pereira**, sobre como as universidades devem se adaptar para essas novas vertentes. Complementa dizendo que, possivelmente, as soluções de computação na nuvem (*cloud computing*) resolvem o problema de capacidade computacional, porém com relação aos dados, seria importante que as universidades e institutos de pesquisa brasileiros se organizassem e unissem esforços para montar um banco de dados comum com o objetivo de acelerar o avanço nessa área.

Segundo **Aisengart**, existe uma cultura crescente em Serviços Geológicos, Agências de Governo e algumas universidades de disponibilização gratuita de dados de levantamentos aerogeofísicos regionais, além de levantamentos geofísicos terrestres e, em alguns casos, de dados sísmicos e de sondagem. Para ela, a universidade deve incentivar seus alunos a utilizarem estes dados em toda a sua potencialidade.

**Fernando Bordignon** revela que hoje já existem grupos, como o dos Professores Martin Tygel e Edson Borin da Unicamp e do Professor Mauro Roisenberg da UFSC, os quais possuem estudantes e docentes especialistas em geociências e computação cognitiva. Outro ponto relevante é que, muitas vezes, o custo computacional para o treinamento das redes neurais com *deep learning* é grande, gerando a necessidade de treinamento na nuvem. Este problema ainda tem sido contornado utilizando dados sintéticos e modelos mais enxutos para os desenvolvimentos iniciais e treinamento em *workstations*. Além disso, estão sendo iniciados investimentos em infraestrutura de *clusters* computacionais especialmente desenhados para aplicações de *deep learning*, os quais possuem placas gráficas (GPUs) para acelerar o treinamento das redes neurais.

Segundo **Ricardo Rosa** o ajuste para oferecerem cursos específicos de tratamento de grandes dados (*Big Data*) e aprendizado de máquina é pequeno, e deve ser feito em cada universidade respeitando suas especificidades.

Como havia mencionado, **Ghassan AlRegib** completa dizendo que as universidades devem investir em programas multidisciplinares, onde os alunos aprendem sobre ciência de dados dentro de seu domínio. Uma maneira de fazer isso é ter um *minor* em ciência de dados em todas as escolas de geofísica.

Para **Matt Hall** a geo-inteligência artificial é um campo tão novo que hoje tudo é de ponta. Isso mudará, é claro, e as universidades precisarão prestar atenção às tendências se quiserem manter o foco nos problemas mais difíceis e de alto valor.

Indagados sobre a possibilidade de haver atividades mapeadas e praticadas pelos geofísicos que sejam desenvolvidas a partir do *machine learning*, os entrevistados fora unânimes que futuramente esta será uma realidade.

**Bruno Pereira** acredita que a passagem de conhecimento em projetos de processamento e interpretação poderão ser sintetizados de maneira automática permitindo garantia de controle de qualidade de maneira muito mais rápida. Sistemas de recomendação automatizados serão capazes de entender as particularida-

## ESPECIAL

des intrínsecas à atividade do geofísico, abrangendo os diferentes métodos. Assim, esses sistemas poderão supervisionar e homogeneizar o conhecimento tanto dos profissionais novos quanto dos profissionais que necessitam de atualização nas metodologias do estado da arte.

Edson Borin explica que tarefas repetitivas e tediosas em grandes volumes de dados, como detecção de padrões em dados sísmicos e dados de poços, classificação de sinais e interpolação de propriedades físicas, por exemplo, são custosas e só são realizadas quando sabe-se que o resultado terá valor. Em função do alto custo, este tipo de atividade não é realizado de forma exploratória, quando se deseja testar uma hipótese. A expectativa é que *machine learning* permita a automação de diversas destas tarefas, viabilizando assim a exploração de mais hipóteses sobre os dados.

Telma Aisengart presume que possa contribuir na previsão de terremotos. Ressaltou que hoje ainda não é possível prever a hora e localização dos terremotos, mas com a crescente coleta de dados relacionados a esses eventos acredita que será possível ter melhores previsões que ajudem a diminuir os danos causados às comunidades por eles afetadas.

Fernando Bordignon diz que no encontro anual da sociedade de geofísicos de exploração (SEG) de 2017, diversas empresas participaram de debates e apresentações em uma *workshop* pós conferência sobre análise de dados para geociências. Vários objetivos são traçados, entre eles o mais ambicioso foi proposto pela Google. Eles estão experimentando com *deep learning* para *full waveform inversion*, com o objetivo de ter uma rede neural que resolva o problema. Argumentam que como a modelagem direta da FWI é conhecida, eles podem gerar um conjunto de dados com muitos modelos de subsuperfície e suas respostas sísmicas AVO. Portanto, com uma rede neural grande o suficiente, seria possível treiná-la para realizar a função inversa, ou seja, a entrada da rede seria a sísmica e a saída desejada os modelos de velocidades.

Nos diversos ramos da ciência Geofísica existem vários problemas técnicos que exigem solução teórica difícil (Aplicada - óleo e gás, mineração e engenharia, por exemplo, Espacial, Teórica, etc.) extremamente dependente de condicionamento dos dados, que envolvem premissas limitantes da solução. Para esses casos, entendemos que o desenvolvimento maior do aprendizado de máquina pode vir a contribuir fortemente na construção de soluções sem condicionamento a priori, ampliando o leque de conhecimento,” relata Ricardo Rosa.

Ghassan AlRegib diz que a visão na *Georgia Tech* para estudar estruturas de subsuperfície foi a seguinte: para implantar uma equipe de pesquisa, é possível implantar milhares de drones e robôs que pesquisam uma determinada região, fazem análises preliminares dos dados no campo, modificam o *layout* de aquisição e repetem até que os dados adquiridos sejam a mais alta resolução possível para a estrutura interna. Então, GANs, NNs e muitos outros modelos podem executar todos os tipos de processamento, análise e interpretação. Todos os resultados intermediários e resultados finais são distribuídos aos geofísicos, geólogos e petrofísicos, que podem estudar o desempenho dos algoritmos, os dados e os resultados. Então, eles podem ajustar

os algoritmos e o modelo para alcançar um resultado final. Eu chamo isso de Inteligência Aumentada.

”Eu vejo duas coisas: gerenciamento de dados e o tratamento significativo da incerteza”, diz Matt Hall. ”Eu estou sendo um pouco obcecado com o gerenciamento de dados. E não quero dizer apenas que o *machine learning* melhorará o gerenciamento de dados por meio da descoberta de anomalias e assim por diante. Quero dizer que finalmente temos uma necessidade irrefutável de gerenciamento de dados incrível. Concluí dizendo que fluxos de trabalho de interpretação humana toleraram dados horríveis por décadas, mas as máquinas serão menos tolerantes. No caminho para esta utopia, muitos erros serão cometidos, e alguns deles podem até ser descobertos, como resultado de coordenadas incorretas, arquivos ausentes ou unidades erradas.”

**Ao comentar sobre o impacto no mercado de trabalho para geofísicos, uma vez que algumas funções podem desaparecer no futuro, Bruno Pereira Dias revela que o cenário econômico e o impacto da Geofísica nos projetos na indústria de óleo e gás continuará sendo o principal fator para o balanço de empregos. Em épocas de alta lucratividade demandará profissionais altamente treinados e experientes capazes de julgar e corrigir sistemas de recomendação automatizados. Por outro lado, em cenários de restrição orçamentária, acredito que projetos de aquisição, processamento e interpretação do tipo “fast track” exigirão menos profissionais da geofísica que atualmente. Isso fará que os efeitos no balanço de empregos serão mais agudos em épocas de crise.**

Na concepção de Edson Borin as ferramentas de *machine learning* substituirão o trabalho manual do ser humano em diversas áreas. No caso da Geofísica, a evolução das técnicas de processamento com auxílio de *machine learning* fará com que as atividades de processamento, principalmente as mais simples e tediosas, sejam realizadas automaticamente por computadores. Consequentemente, profissionais que atuam de forma mais mecânica terão cada vez menos espaço no mercado de trabalho. Por outro lado, estas técnicas de processamento oferecerão mais informações ao Geofísico, que terá mais condições de derivar informações relevantes, de valor no processamento. Neste contexto, o Geofísico terá um papel mais importante na cadeia de produção e será cada vez mais reconhecido pela sua expertise.

Fernando Bordignon vê isso com muito otimismo, pois a eficiência das empresas aumentará muito, ou seja, a produtividade de um geofísico será maior do que hoje em dia. Tarefas repetitivas serão automatizadas, liberando o profissional para realização de análises que ainda exigirão intervenção humana. Portanto, a capacidade das empresas de realizar projetos aumentará, utilizando a mesma quantidade de profissionais. As empresas que souberem aproveitar essa vantagem competitiva irão utilizar este cenário para realizar mais projetos, levando à contratação de mais profissionais.

Para Ricardo Rosa praticamente todas as profissões que envolvem rotinas de trabalho automatizáveis, e na Geofísica existem rotinas de trabalho claramente automatizáveis (interpretação de dados geofísicos que envolvam picagem de horizontes ou superfícies, por

exemplo), estão sujeitas à serem despriorizadas pelas empresas. Essa despriorização pode levar a redução de sua importância e futura extinção. Portanto, os Geofísicos não estão imunes aos processos de transformação digital e a possível automatização de funções. Para ampliar o escopo de atuação do Geofísico e permitir que o profissional continue sendo importante, é fundamental a formação ampla e conhecimento, não apenas nos temas diretamente ligados a Geofísica, mas também a ciência de dados, análise e programação.

Ghassan AlRegib completa dizendo que embora seja verdade para qualquer indústria, a automação e a IA podem ter impacto em alguns trabalhos, mas para os especialistas em domínio sempre serão necessários. No entanto, a natureza do trabalho diário que um Geofísico executa terá que mudar. Em vez de gastar horas frustrantes lidando com dados e recursos limitados, os geofísicos podem causar impacto diariamente, obtendo o máximo dos dados e da tecnologia usando inteligência artificial. "Eu gosto de pensar nisso como uma grande oportunidade para todos os especialistas de domínio, incluindo geofísicos, onde eles podem fazer coisas que eram impossíveis há alguns anos. Naturalmente, isso significa que os Geofísicos precisam estar equipados com um novo conjunto de habilidades e ferramentas, o que não é um problema, pois eles sempre adquiriram as habilidades necessárias para realizar o trabalho. A propósito, eu sou um engenheiro electricista e um matemático aplicado, mas sempre admirei o trabalho que os geofísicos fazem, especialmente desde que comecei a trabalhar nesta aplicação há mais de cinco anos. Tiramos o chapéu de fato!"

"É tremendamente importante que geofísicos estejam presentes nesta jornada rumo a uma abordagem mais orientada a dados para a ciência da terra e a indústria", afirma Matt Hall. "Nosso trabalho, sem dúvida, mudará à medida que nos libertarmos de algumas das tarefas mais mundanas ou repetitivas (gerenciamento de dados, análise de logs, detecção de falhas, rastreamento de horizonte e assim por diante). Quando isso acontecer, seremos capazes de perguntar - e tentar responder - algumas das questões mais urgentes para as quais ainda não tivemos tempo. Somente geocientistas podem delimitar essas questões, coletar os conjuntos de dados de que precisamos, projetar as funções de perda, visualizar os resultados e entender as implicações. Matt Hall finaliza dizendo que longe de estar em risco, nossos empregos estão prestes a ficar muito mais interessantes!"

Curta a SBGf  
no facebook!



[www.facebook.com/sbgf.org](http://www.facebook.com/sbgf.org)



FUNDAÇÃO GORCEIX  
DEPARTAMENTO DE GEOLOGIA DO PETRÓLEO



Área de Atuação ..... Geologia do Petróleo  
Sistemas Petrolíferos ..... Remasterização de Dados Sísmicos  
Vetorização ..... Linhas Sísmicas  
Rede de Dados Geofísicos ..... RDG

CONTATO: Fundação Gorceix - DEPETRO  
Rua Carlos Walter Marinho Campos, nº57, Vila Itacolomy  
CEP:35400-000 Ouro Preto - Minas Gerais - Brasil  
Telefones: (31) 3559 - 7100/ 3559 - 7144

Uma grande  
conquista  
para a nossa  
história de  
sucesso.



A QGEP – Queiroz Galvão Exploração e Produção – é a única empresa privada 100% brasileira a operar no polígono do pré-sal em águas ultraprofundas da Bacia de Santos, no campo de Atlanta.

Essa conquista nos motiva a seguir adiante. Para nós, o futuro é hoje.

<http://www.qgep.com.br/>





# Detecção de estruturas em dados sísmicos com Deep Learning

Lucas de M. Araújo | Fabíola M. C. de Oliveira | Jorge H. Faccipieri |

Tiago A. Coimbra | Sandra Avila | Martin Tygel

Centro de Estudos de Petróleo (CEPETRO) e Instituto de Computação (IC) da Universidade Estadual de Campinas (Unicamp).

## INTRODUÇÃO

Aplicações acadêmicas e comerciais de *Deep Learning* apresentaram grande crescimento em tempos recentes. Modelos para reconhecimento de fala e interpretação de linguagem natural potencializam funções assistivas de *smartphones* enquanto veículos autônomos estão em fase avançada de teste e consultores-robôs operam diariamente no mercado financeiro (Hatcher & Yu, 2018). Estas técnicas estão permeando campos tão diversos quanto agricultura, medicina diagnóstica, ciência forense, física teórica e aplicada, administração e gerenciamento.

No contexto da Geofísica, técnicas de *Machine Learning* já são utilizadas há décadas (Sandham & Leggett, 2003). A aplicação de *Deep Learning* nesta área é um tema recente, porém publicações em eventos e principais revistas de Geofísica mostram que *Deep Learning* é um assunto que está chamando cada vez mais a atenção da comunidade. Em maio de 2018, ocorreu, no Rio de Janeiro o *Joint SBGf/SEG Workshop on Machine Learning*, com apresentação de trabalhos tanto do âmbito acadêmico quanto da indústria. Este evento teve oito dos vinte e nove trabalhos apresentados sobre *Deep Learning*. Realizamos um levantamento em três periódicos de destaque na área da Geofísica: *Geophysics*, *Interpretation* e *Geophysical Journal International*, entre janeiro de 2016 até junho de 2018, e encontramos vinte artigos que exploram técnicas de *Machine Learning*, sendo três destes na área de *Deep Learning*. Estes três últimos foram publicados há menos de um ano. Este levantamento ilustra a tendência crescente do tema de *Machine Learning* nas pesquisas em Geofísica e, mais recentemente, o aparecimento de técnicas de *Deep Learning*.

Uma arquitetura de *Deep Learning* é uma sequência de várias camadas compostas de módulos simples, nas quais cada módulo é capaz de aprender relações não-lineares entre entrada e saída. Cada módulo desta sequência se torna sensível a certas características de sua entrada, selecionando aquelas que são relevantes e invariantes para a associação entre um dado de entrada e à categoria a qual o dado pertence (LeCun et al., 2015). Diversos tipos de arquiteturas de *Deep Learning* (Goodfellow et al., 2016) foram desenvolvidas e aplicadas nos mais variados contextos. Por exemplo, *Deep Neural Networks (DNN)* são utilizadas para problemas gerais de classificação ou regressão nos quais os dados de entrada são heterogêneos. *Convolutional Neural Networks (CNN)* têm tido enorme sucesso na classificação, detecção de objetos e segmentação em imagens. *Recurrent Neural Networks (RNN)* são utilizadas em problemas em que o tempo ou sequência são fatores importantes, como previsão de texto, análise de vídeo e previsão do mercado financeiro. *Autoencoders* são empregados para comprimir dados, criar dados sintéticos realistas, tratar valores faltantes e lidar com grande quantidade de amostras não rotuladas.

Neste artigo, apresentaremos um fluxo que permite a prototipagem rápida de experimentos de *Deep Learning* em dados que podem ser representados como imagens, incluindo dados sísmicos. Por fim, mostraremos um estudo de caso no qual este fluxo é aplicado para resolver o problema de detecção de ápices de difrações, de interesse para o imageamento sísmico.

## APRENDIZAGEM SUPERVISIONADA COM DEEP LEARNING

Aprendizagem supervisionada se refere a problemas nos quais amostras rotuladas são usadas para treinamento e avaliação de modelos (Goodfellow, 2016). Por exemplo, suponha que temos imagens de três classes de veículos: *carro*, *moto*, *caminhão*. Se cada imagem estiver anotada com seu rótulo correspondente, podemos treinar um modelo para reconhecer estas classes em novas imagens (LeCun et al., 2015).

A Figura 1 apresenta um fluxo padrão para aprendizagem supervisionada que permite realizar rapidamente experimentos usando arquiteturas de *Deep Learning*.

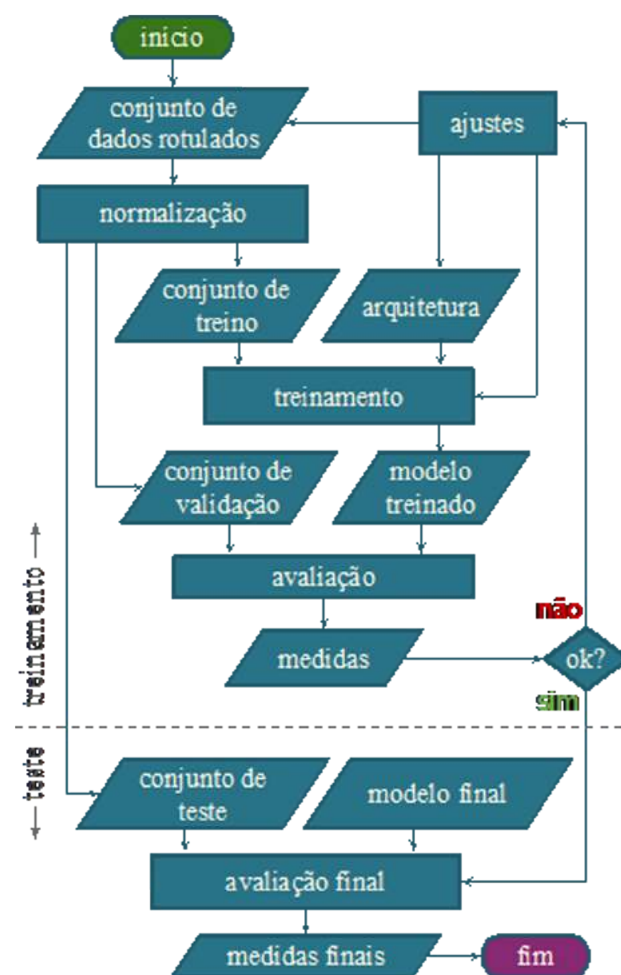


Figura 1: Fluxo para aprendizagem supervisionada com arquiteturas de Deep Learning.

A primeira etapa da aprendizagem supervisionada é obter um conjunto com dados anotados, ou seja, um conjunto em que cada amostra está associada à sua classe correspondente. Existem bases de dados nos quais a anotação já foi realizada, ou pode ser necessário anotar manualmente. Em alguns problemas, é possível também gerar dados sintéticos, anotados automaticamente. Apesar da abundância de dados na Geofísica, há escassez de dados rotulados. Portanto, a etapa de anotação de dados ou geração de dados sintéticos é essencial para experimentos de *Deep Learning* neste domínio.

Este conjunto inicial deve ser particionado em três subconjuntos: treino, validação e teste. Uma proporção típica é reservar 10–20% dos exemplos para teste, 10–20% para validação e o restante para treino. O conjunto de teste não deve ser usado em nenhum momento durante o processo de ajuste, treinamento e validação do modelo e será usado somente para avaliar a versão final do sistema. Esta separação é importante para avaliar se o modelo produzido generaliza o problema, ou seja, se o modelo classifica corretamente dados que não foram vistos no processo de treinamento. Dessa forma, uma condição necessária é que não haja duplicações entre amostras do conjunto de teste e dos conjuntos de treinamento e validação, para evitar que a avaliação final do modelo treinado seja enviesada. Pode ser necessário *normalizar* os dados para que todas as amostras estejam representadas de maneira consistente e conveniente para a arquitetura utilizada.

Para a etapa de treinamento é necessário escolher a arquitetura, como por exemplo DNNs, CNNs, RNNs. Uma vez que o treinamento é concluído, usamos o conjunto de validação para avaliar a qualidade do modelo treinado, medindo-se a acurácia, por exemplo. Enquanto os valores medidos não estiverem adequados, o treinamento deve ser realizado novamente, ajustando-se o conjunto de treino (aumentar o número de amostras, por exemplo), parâmetros da arquitetura utilizada ou parâmetros de treinamento, como taxa de aprendizagem, tamanho do *batch* etc. Idealmente, queremos encontrar a combinação de parâmetros que minimize o erro do modelo.

Finalmente, quando obtemos o modelo com medidas adequadas no conjunto de validação, utilizamos o conjunto de teste para avaliação final deste modelo.

## DETECÇÃO DE ESTRUTURAS EM DADOS SÍSMICOS

Problemas relacionados ao processamento e interpretação de dados sísmicos podem ser tratados como problemas de detecção de estruturas (*picking*) ou de classificação de eventos por um intérprete.

Em problemas de detecção de estruturas, o objetivo é identificar pontos, ou coordenadas, do dado que contém estruturas de interesse enquanto que, em problemas de classificação, o objetivo é determinar a qual classe um dado pertence. Uma forma de se resolver o problema de detecção de estruturas é realizar a classificação de janelas associadas a cada ponto do dado. Neste caso, para se construir um conjunto de treinamento, um intérprete pode anotar alguns dados (*picking*), rotulando coordenadas que contém estruturas de interesse. A partir destas coordenadas, podemos traçar *janelas* de forma que o evento associado ao rótulo esteja no *centro* da janela. Denominamos este processo de *janelamento*.

Podemos então construir conjuntos, compostos de janelas associadas a classes, e aplicar o fluxo proposto para treinamento, validação e teste de um modelo. Determinar o tamanho adequado da janela se torna um dos parâmetros importantes a serem ajustados durante o processo de treinamento.

De posse de um modelo treinado, podemos processar um dado de tamanho arbitrário como uma *sequência de janelas deslizantes*, na qual o centro da janela passa por cada *pixel*, ou ponto, do dado original.

Dessa forma, o processo de construir conjuntos usando janelamento e processar o dado com uma janela deslizante reduz o problema de detecção de estruturas em um dado ao problema de classificação de janelas.

A seguir, apresentamos um estudo de caso que aplica o fluxo para o problema de detecção automática de ápices de difração.

## ESTUDO DE CASO

Utilizamos *Deep Learning* para automatizar um processo de construção de modelos de velocidades para migração em tempo a partir de informações de difrações contidas no dado aberto. Inicialmente, utilizamos o método de extração de eventos de difração, via um empilhamento, que usa uma equação de tempo de trânsito do tipo *Double Square Root*, ou DSR (Coimbra et al., 2015; Facciopieri et al., 2016). Este método, além de realizar a separação de sinais de difração, também gera painéis com os parâmetros relacionados à velocidade de migração em tempo e à inclinação desses eventos.

Os parâmetros de velocidade associados ao ápice das difrações oferecem uma estimativa confiável da velocidade de migração naquele ponto. Dessa forma, estas velocidades podem ser interpoladas para se produzir um modelo de velocidade suave para migração em tempo (Yilmaz, 2001).

O processo de interpolação é amplamente utilizado na área de processamento de imagens e no processamento de dados sísmicos com técnicas bem estabelecidas de implementação. Entretanto, a detecção dos ápices de difrações não é um processo trivial. Neste trabalho, usamos o fluxo apresentado na seção anterior para treinar um modelo de *Deep Learning* capaz de identificar ápices nos painéis de difrações.

Em nossos experimentos, utilizamos dados empilhados de difrações reais de três bacias brasileiras. Para representar o dado sísmico como imagem, mapeamos todas as amostras presentes em seus traços para pixels. Ademais, como o valor absoluto da amplitude de cada dado pode variar consideravelmente, é importante normalizar os dados. Para este experimento, re-escalamos as amostras pelo inverso do valor *Root Mean Square* (RMS) das amplitudes de cada dado.

As anotações foram realizadas por um geofísico, que selecionou e classificou amostras em duas classes: *ápice* e *não-ápice*. Amostras da classe *ápice* contêm um *ápice* de difração sem interferências no centro da janela, enquanto que amostras da classe *não-ápice* apresentam ausência de eventos,

## ARTIGO TÉCNICO

caudas de difrações ou mesmo ápices com interferência de outras difrações.

O geofísico selecionou cerca de 400 amostras de cada dado sísmico, sendo 200 amostras da classe ápice e 200 da classe não-ápice. Os conjuntos de treino, validação e teste foram construídos pelo janelamento ao redor dos pontos anotados, cada janela com dimensão de 64 X 64 *pixels*. Neste experimento, usamos dois dados para treinamento e validação e o terceiro foi reservado como teste cego para avaliação do modelo.

Como classificador, utilizamos uma versão adaptada da LeNet5 (LeCun et al., 1998), implementada em TensorFlow ([www.tensorflow.org](http://www.tensorflow.org)). A taxa média em nossos conjuntos de teste foi de 84,0% de acerto na classe *ápices* e 86,6% na classe *não-ápices*.

A Figura 2 mostra o resultado da detecção de ápices no dado sísmico de teste. Pontos preditos como ápices estão marcados em verde. Observe que o modelo foi capaz de identificar o ápice da maior parte das difrações mesmo sem utilizar nenhuma amostra deste dado no processo de treinamento.

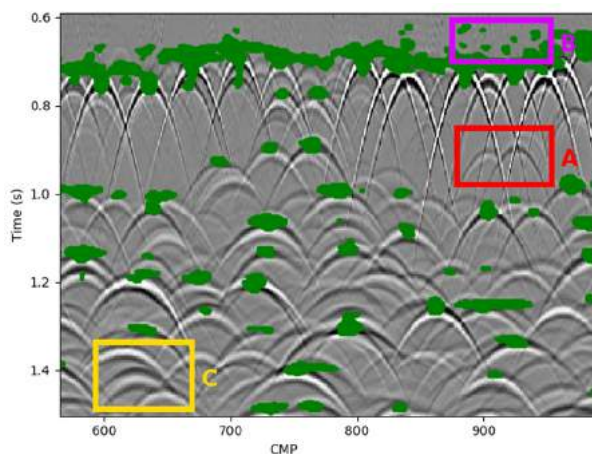


Figura 2: Exemplo de detecção de ápices (em verde) em um empilhamento de difrações. As regiões A, B e C representam, respectivamente, verdadeiros negativos, falsos positivos e falsos negativos.

Além dos ápices das difrações, a Figura 2 destaca três regiões. A região A mostra exemplos de verdadeiros negativos de difícil detecção, que são ápices com interferência de outras difrações. A região B aponta exemplos de falsos positivos, enquanto a região C mostra falsos negativos, ou seja, ápices que não foram detectados.

Apesar dos falsos positivos detectados, como os que podemos observar na Figura 2, conseguimos produzir um modelo de velocidades de migração satisfatório. Como trabalho futuro, pretendemos explorar ajustes nos dados e nos parâmetros para minimizar a taxa de falsos positivos, já que estes levam a erros que podem comprometer o modelo de velocidades para migração em tempo. Além disso, uma pequena taxa de falsos negativos (não detectar alguns ápices) pode ser tolerada.

Por fim, a Figura 3 mostra um empilhamento de difrações de um dado marinho, com detecção automática de ápices (superior), e dois campos de velocidades para migração em tempo da mesma seção (meio e inferior).

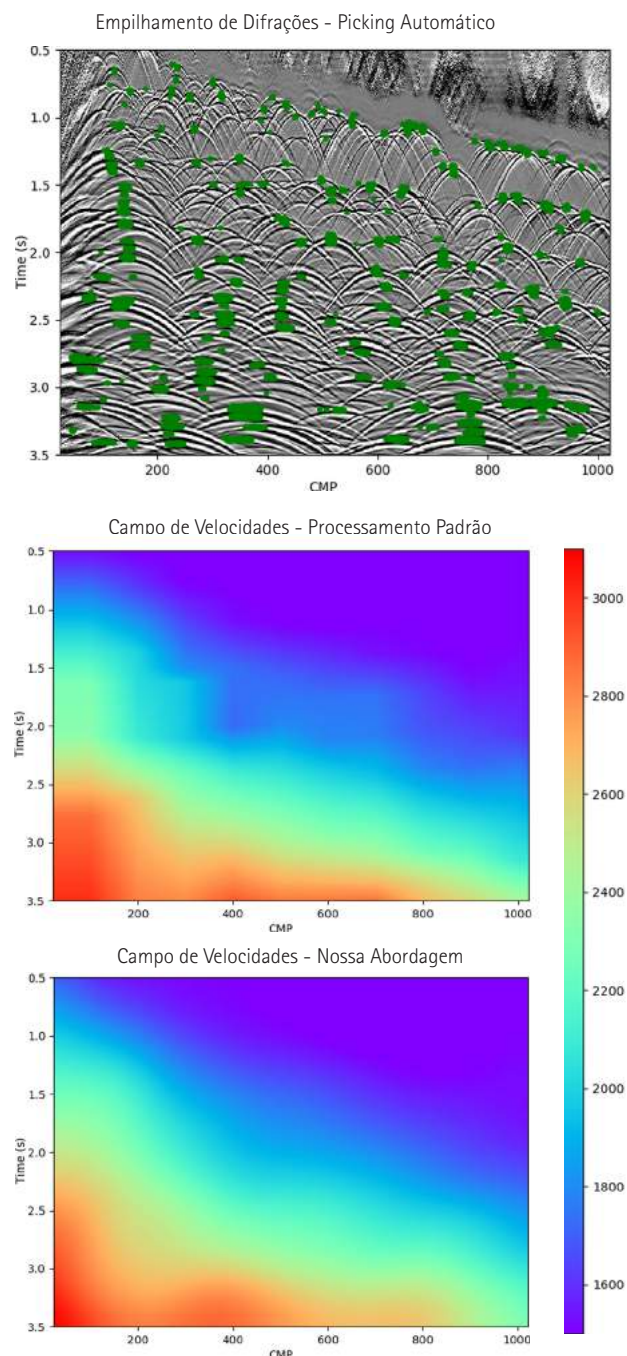


Figura 3: Empilhamento de difrações com picking automático (superior), painéis de velocidade para migração em tempo gerados pelo fluxo padrão de processamento (intermediário) e pela nossa abordagem (inferior). Velocidades variam de 1500m/s (azul) a 3100m/s (vermelho).

No painel do meio, o campo foi obtido a partir do fluxo de processamento padrão, realizado por um intérprete. No painel inferior, o campo foi gerado interpolando-se as velocidades dos ápices detectados. Percebe-se a mesma tendência geral nos dois painéis, porém o gerado com a nossa abordagem tem comportamento mais suave, possivelmente pelo método de interpolação escolhido (*natural neighbor interpolation*, Sibson, 1981). Consideramos a qualidade dos dois painéis comparáveis e o tempo de processamento para este dado com a abordagem proposta foi de cerca de 15 minutos. Ressaltamos que o modelo conseguiu detectar os ápices das difrações em um dado marinho, mesmo



tendo sido treinado somente com dados terrestres.

## DISCUSSÃO E CONCLUSÕES

Neste trabalho nós discutimos o aumento de interesse na aplicação de *Deep Learning* para resolver problemas geofísicos, apresentamos um fluxo para realização de experimentos de aprendizagem supervisionada com arquiteturas de *Deep Learning* e mostramos como este fluxo foi usado para resolver o problema de detecção automática de ápices de difrações para geração de modelos de velocidades em dados sísmicos.

Em nossos experimentos, gerar um campo de velocidades a partir das difrações demorou cerca de 15 minutos com a utilização do modelo treinado. Este método pode ser utilizado por um intérprete para se obter rapidamente um campo de velocidades para migração em tempo ou para comparar a um campo produzido pelo fluxo tradicional.

O tamanho da rede LeNet5 nos permitiu treinar um modelo a partir de um conjunto de treino com poucas amostras e atingir em torno de 85% de acurácia. Esperamos que as redes de classificação mais modernas, como a *Inception-v4* e a *ResNet*, ou redes especializadas em detecção, como a *YOLO-v3*, apresentem taxas de acurácia maiores do que as obtidas com a LeNet5. Entretanto, por serem redes maiores, estas exigirão um conjunto maior de dados para que os modelos possam ser treinados com sucesso. Como o processo de anotação manual dos dados é tedioso, técnicas automáticas e semiautomáticas de geração de amostras para treinamento terão um papel importante na exploração destas novas arquiteturas. Neste contexto, uma das possíveis soluções é a utilização do modelo construído em nossos experimentos para geração automática de dados anotados.

Finalmente, esperamos que o fluxo de processamento divulgado aqui possa ser utilizado por outros pesquisadores e impulsionar a exploração das técnicas de *Deep Learning* em problemas da Geofísica.

## AGRADECIMENTOS

Os autores agradecem o apoio da Petrobrás, do CNPq, da CAPES e do Centro de Pesquisa em Ciência e Engenharia Computacional (Fapesp/Cepid #2013/08293-7-Brasil). Agradecemos também a Pedro Mário Cruz e Silva, Arquiteto de Soluções da NVIDIA, pelas valiosas contribuições, incluindo as discussões sobre o conceito de janelamento utilizado neste trabalho.

## REFERÊNCIAS

COIMBRA, T.A., FACCIPIERI, J.H., GELIUS, L.-J., & TYGEL, M. 2015. Enhancement of stacked sections using ZO CRS parameters. In: Congresso Internacional da Sociedade Brasileira de Geofísica & EXPOGEF, 14, Rio de Janeiro, p. 1251-1255.

FACCIPIERI, J.H., COIMBRA, T.A., GELIUS L.-J., &

TYGEL, M. 2016. Stacking apertures and estimation strategies for reflection and diffraction enhancement. *GEOPHYSICS*, 81(4): 271-282.

GOODFELLOW, I., BENGIO, Y., & COURVILLE, A. 2016. *Deep Learning*. MIT Press, Cambridge, 775 p.

HATCHER, W.G., & YU, W. 2018. A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends. *IEEE Access*, (6): 24411-24432.

LECUN, Y., BOTTOU, L., BENGIO, Y., & HAFFNER, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278-2324.

LECUN, Y., BENGIO, Y., & HINTON, G. 2015. Deep learning. *Nature*, 521: 436-444.

SANDHAM, W., & LEGGETT, M. 2003. *Geophysical Applications of Artificial Neural Networks and Fuzzy Logic*. Modern Approaches in Geophysics (21), Springer Netherlands, Dordrecht, 325 p.

SIBSON, R. 1981. A Brief Description of Natural Neighbor Interpolation. In: BARNETT V. (Ed.). *Interpreting Multivariate Data*, John Wiley & Sons, New York, p. 21-36.

YILMAZ, Ö. 2001. *Seismic data analysis: Processing, inversion and interpretation of seismic data*. Society of Exploration Geophysicists, Tulsa, 1000 p.

## Estabelecendo o Ritmo com Liderança de Desempenho na Indústria

Para saber mais sobre o nosso histórico comprovado, contacte-nos através de [lbraga@sbgs.com](mailto:lbraga@sbgs.com)

Dados cortesia da Maersk Oil. Resultados obtidos de Dan Field Ocean Bottom Node (OBN) Survey - Um Estudo De Caso De Águas Rasas. Zasko et al. EAGE Conference (2014)

**Seabed**  
GEOSOLUTIONS

[seabed-geo.com/Performance](http://seabed-geo.com/Performance)

## Classificação litológica com dados incompletos

Erick Costa e Silva Talarico – Petrobras

Ruy Luiz Milidiú – PUC-Rio

Anderson Rafael Rezende Alves – Petrobras

### RESUMO

Perfis de poço geralmente são incompletos: diferentes conjuntos de perfis são usados em diferentes poços e, mesmo em um poço, cada perfil cobre um intervalo diferente. No presente trabalho, examinamos o problema de classificação de fácies, quando o registro dos dados é incompleto. Desta forma, as amostras do conjunto de treino e teste não possuem todas as variáveis necessárias para o treinamento e previsão. Deve o geocientista eliminar todas amostras de treinamento com informação incompleta? Como prever a litologia para amostras sem a medida de todos perfis? Como a integração de dados incompletos beneficia a classificação de fácies? O presente artigo aborda tais questões.

### INTRODUÇÃO

Há inúmeras publicações na área de classificação de fácies com base em perfis de poço (Xie, et al., 2018). Porém, em geral, estas publicações consideram que há um conjunto pré-definido de perfis de entrada comum a todas as amostras do conjunto de treino e ao conjunto de teste, o que limita a aplicação do modelo treinado. Por outro lado, há publicações na área de inteligência artificial e estatística que estudam como utilizar dados incompletos no treino e na previsão (Laencina, Gómez, & Vidal, 2009). No presente trabalho apresentamos três possíveis estratégias para fazer uso de dados incompletos. Como consequência obtemos um modelo com no mínimo 5 perfis, o qual consegue prever 5 vezes mais dados, mantendo uma acurácia de 85%, enquanto que o modelo treinado com 7 perfis obtém acurácia de 91%. Uma redução pequena na acurácia em relação ao poder de previsão obtido. O presente estudo não pretende exaurir o assunto ou apresentar o estado da arte, mas estimular a discussão deste tópico, visto a limitação que a perfilagem irregular dos poços impõe aos estudos petrofísicos.

### CARACTERIZAÇÃO DOS DADOS

O dado utilizado neste estudo é proveniente do Serviço Geológico Holandês (SGH), disponível no site <http://www.nlog.nl/>. O SGH disponibiliza informações de 6.000 poços perfurados desde 1.900 no território Holandês *onshore* e *offshore*. Esse conjunto de dados é interessante por caracterizar de forma compreensiva a bacia sedimentar daquele país. Cada poço possui, dentre outros dados, arquivos de perfilagem, para os quais naturalmente existe grande variabilidade de ferramentas. Para o download automático deste grande volume de dados utilizamos as bibliotecas Python *bs4*

*requests*. Cada empresa de serviço oferece ferramentas com diferentes nomes para um mesmo tipo de medida, e a tecnologia varia ao longo do tempo. Logo, uma etapa de pré-processamento consiste na identificação das ferramentas de perfilagem utilizadas, priorização das mesmas, da mais confiável para a menos, a emenda das diferentes curvas e controle de qualidade.

Definimos um conjunto padrão de 7 perfis de atributos como a base para a classificação de fácies: raios gama, sônico compressional, sônico cisalhante, porosidade neutrão, densidade, resistividade e fator fotoelétrico. Dos 6.000 poços listados pelo SGH, somente 32 possuem todos os perfis básicos e passaram no controle de qualidade. A lista dos poços utilizados no estudo encontra-se na Tabela 1.

BHM-06	BRW-04	F16-A-03-S1	G17-S-01
K04-A-05-S2	K05-12	K05-13	K06-GT-04-S1
K09-12	K18-07	K18-08	K18-KOTTER-14
L04-06	L05-09	L05-10	L05-11
L05-12-S1	L05-12	L05-B-03	L05-C-02-S1
L05-C-02-S2	L05-C-02-S3	L06-06	L08-16-S1
L11-13	MDZ-02	MSG-03-S1	OPK-03-S1 *
Q04-C-01	Q05-A-01	TVN-01 *	WIT-03

Tabela 1: Lista de poços utilizados no estudo. Todos são disponibilizados pelo Serviço Geológico Holandês (SGH).

Mesmo nos poços da Tabela 1, o intervalo de perfilagem de cada ferramenta é variável. A Tabela 2 mostra a cobertura de cada perfil, ou seja, o percentual de amostras medidas com cada perfil.

Perfil	Cobertura (%)
Raios Gama	82%
Sônico compressional	79%
Sônico cisalhante	73%
Porosidade neutrão	74%
Densidade	74%
Resistividade	52%
Fator fotoelétrico	42%

Tabela 2: Percentual de amostras medidas para cada perfil, ordenadas da maior cobertura para a menor.

O SGH fornece as descrições litológicas como imagens em relatórios de perfilagem (*composites*). Para a realização do presente estudo, transcrevemos alguns trechos de imagens, resultando em 60.000 amostras anotadas, das quais 20.000 possuem todas as 7 medidas de perfil.

As classes litológicas consideradas são: arenitos, arenitos com gás, folhelhos, halita, anidrita, marga, carbonato, dolomito e carvão.

Para a extrapolação deste conjunto de treino, o geocientista deve usualmente decidir o número de medidas: quanto mais perfis utilizar, maior o poder discriminativo do classificador, porém mais limitado é o

intervalo previsto, como mostra a Figura 1.

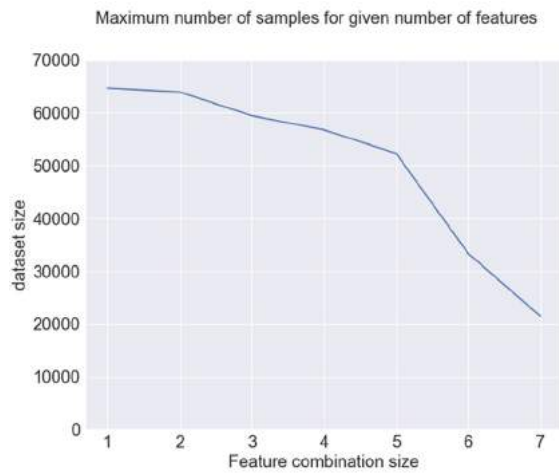


Figura 1: Máximo número de amostras medidas por número de perfis. Apesar de não discriminado, o gráfico captura a combinação de n perfis com maior cobertura e desenha contra a absissa n.

As classes são desbalanceadas. Por exemplo, há tão poucos exemplos de margá e não há margas medidas com todos os perfis. Utilizamos a técnica *bootstrap* (Efron, 1979) para melhorar a detecção das classes mais raras. A Figura 2 mostra o histograma de litologias após o *bootstrap*. Note que, mesmo com *bootstrap*, não há amostras de margá com todos os 7 perfis.

Dataset “full-feature” + Bootstrap

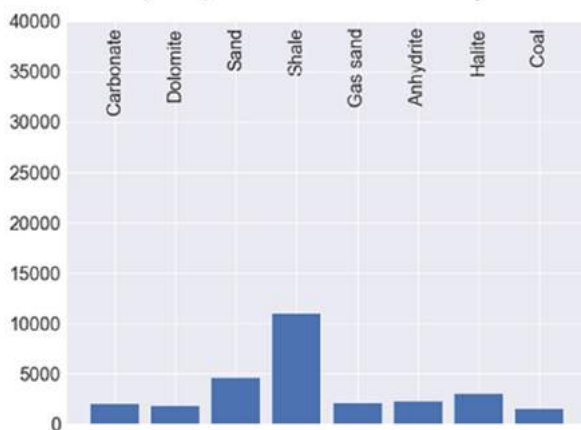


Figura 2: Histograma de amostras por litologia. Só estão considerados no gráficos amostras com todas as 7 medidas.

### TÉCNICAS DE CLASSIFICAÇÃO

A Figura 3 ilustra o problema de classificação com dados incompletos. A litologia determina a distribuição de valores possíveis para as medidas. A classificação pretende inferir a litologia através dos valores de um subconjunto das medidas.

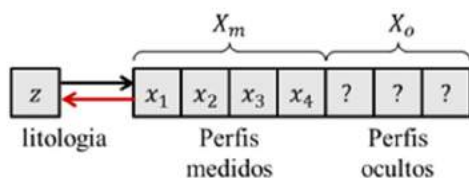
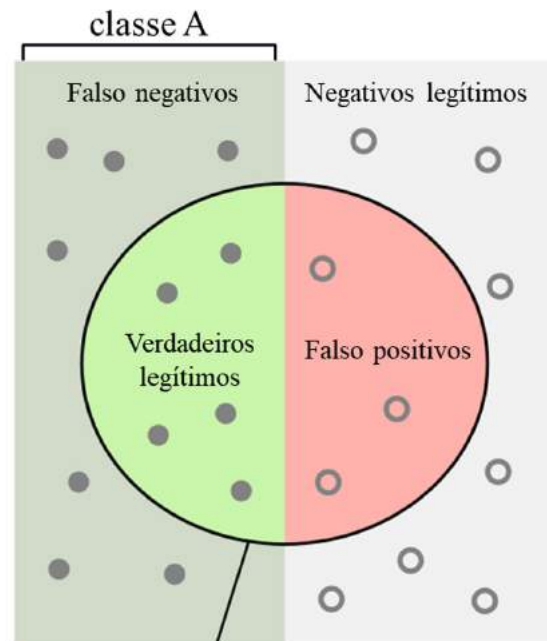


Figura 3: Representação esquemática da medida de poço. A litologia determina a distribuição das propriedades (seta preta). Na classificação pretende-se inferir a litologia a partir de um subconjunto de perfis medidos.

Para explorar o tema de classificação com dados incompletos, utilizamos três algoritmos de aprendizado de máquina, a saber: Naïve Bayes, Misturas Gaussianas e Redes Neurais. Adaptamos cada um deles para lidar com dados incompletos, tanto na etapa de treino como na etapa de predição.

Para tornar o estudo robusto, treinamos cada classificador 20 vezes, onde em cada iteração 30% dos dados são para treino e 70% para validação. As estatísticas de acerto são calculadas com base nos 20 experimentos.

Para avaliar quantitativamente a performance dos algoritmos, utilizamos as métricas: *accuracy* (acurácia), *precision* (precisão) e *recall* (detecção). A acurácia mede a taxa de acerto global do classificador, enquanto que precisão e detecção são métricas de acerto específicas para cada classe litológica. A Figura 4 ilustra as métricas.



Amostras previstas como A

Quantas amostras previstas como A estão corretas?

$$\text{Precision} = \frac{\text{Verdadeiros legítimos}}{\text{Verdadeiros legítimos} + \text{Falso positivos}}$$

Quantas amostras da classe A foram detectadas?

$$\text{Recall} = \frac{\text{Verdadeiros legítimos}}{\text{Verdadeiros legítimos} + \text{Falso negativos}}$$

$$\text{accuracy} = \frac{\text{Previsões corretas}}{\text{Total de previsões}}$$

Figura 4: Esquema adaptado de (Walber, 2017) para ilustrar as métricas de classificação.





BASELINE: NAÏVE BAYES

Nesta técnica, descrita em (Bishop, 2007), a probabilidade da litologia, dadas as medidas, é fornecida pela equação 1. Nesta equação,  $p(z)$  é a frequência da litologia  $z$  no conjunto de treino, e  $\mathcal{N}(x|\mu, \sigma^2)$  representa uma distribuição Gaussiana com média  $\mu$  e desvio padrão  $\sigma$ . Devido à fatorabilidade da equação 1, os parâmetros do modelo  $\mu_i(z)$  e  $\sigma_i(z)$  cada perfil  $i$  e litologia  $z$  podem ser estimados independentemente. Para a previsão de novas amostras, a fatorabilidade permite ignorar os fatores correspondentes às medidas ausentes, o que equivale a marginalizar em relação a estas variáveis, conforme descrito na equação 2.

$$p(z|X) \propto p(z) \prod_{i=1}^7 \mathcal{N}(x_i|\mu_i(z), \sigma_i(z)^2) \quad 1$$

$$p(z|X_m) \propto p(z) \prod_{i=1}^5 \mathcal{N}(x_i|\mu_i(z), \sigma_i(z)^2) \quad 2$$

Na Figura 5, mostramos as métricas deste classificador para dois casos: nenhum perfil faltando (a) ou exatamente 2 perfis faltando (b).

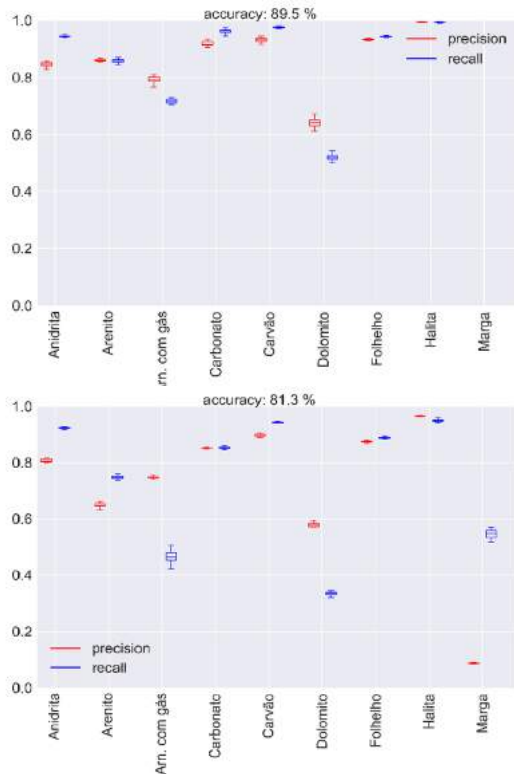


Figura 5: Métricas de classificação do modelo Naive Bayes. (a) com todos 7 perfis, e (b) com 5 perfis.

A Figura 6 mostra o resultado da classificação para o poço L05-B-03. As previsões recuperam bem as amostras anotadas, mas prevê uma intercalação entre as classes arenito e arenito com gás, que não é fisicamente possível.

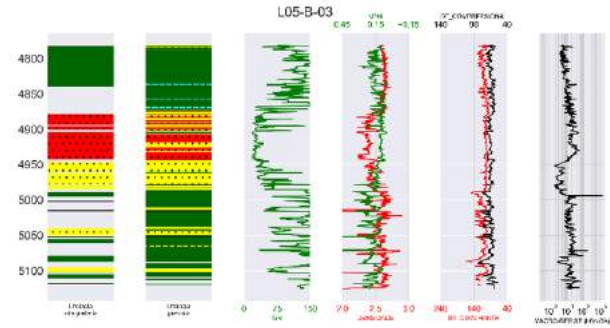


Figura 6: Previsões no Naive Bayes para o poço L05-B-03. Nota-se uma recuperação consistente das amostras anotadas, mas a extrapolação falha ao prever intercalações entre as litologias arenito e arenito com gás, o que não é fisicamente razoável.

MISTURA GAUSSIANA

Nesta técnica, descrita em (Bishop, 2007), a probabilidade das litologias é dada pela equação 3. Neste caso, a probabilidade não é fatorável, já que é necessário a estimativa da matriz de covariância  $\Sigma(z)$ , a qual guarda as interrelações entre os diferentes perfis.

$$p(z|X) \propto p(z) \mathcal{N}(X|\mu(z), \Sigma(z)) \quad 3$$

Para um conjunto de dados completo, a estimativa de  $\Sigma(z)$  é simples e garantidamente resulta em uma matriz positiva semi-definida, e, em geral, positiva definida, condição necessária para a distribuição Gaussiana. Porém, quando há dados incompletos, essa garantia é perdida, logo devemos buscar técnicas aproximadas para estimar  $\Sigma(z)$ . Para possibilitar o treino deste classificador, usamos três adaptações. Primeiramente, adotamos uma equação de estimativa bayesiana para os parâmetros (Bishop, 2007), considerando a matriz de covariância global  $\Sigma_0$  e o vetor de média global  $\mu_0$  como conhecimento a priori, conforme descrito nas equações 4 a 6. Nestas equações,  $N$  é o número de amostras,  $\mu$  é o número de perfis, e  $\nu$  é o hiperparâmetro que controla o quanto a informação a priori influencia na estimativa final.

Em segundo lugar, uma projeção, conhecida como projeção semi-positiva é aplicada para garantir semi-positividade da matriz de covariância.

Por fim, para combater o viés da proporção de litologias anotadas, aplicamos um algoritmo EM (expectation-maximization) (Bishop, 2007).

$$\mu_{MAP}(z) = \frac{N\mu(z) + v\mu_0}{N + v} \quad 4$$

$$\Sigma_{MAP}(z) = \frac{1}{N + v - M + 1} \times \{(v - M - 1)\Sigma_0 + N\Sigma(z) + A\} \quad 5$$

$$A = (\mu(z) - \mu_0)(\mu(z) - \mu_0)^T \quad 6$$

Durante a previsão de amostras, marginalizamos as variáveis correspondentes aos perfis ocultos, visto que a distribuição Gaussiana possui uma fórmula analítica para a marginalização de variáveis, conforme as equações 7 a 9.

$$p(z|X_m) \propto p(z)\mathcal{N}(X|\mu_m(z), \Sigma_{mm}(z)) \quad 7$$

$$\mu(z) = \begin{bmatrix} \mu_m \\ \mu_o \end{bmatrix} \quad 8$$

$$\Sigma(z) = \begin{bmatrix} \Sigma_{mm} & \Sigma_{mo} \\ \Sigma_{om} & \Sigma_{oo} \end{bmatrix} \quad 9$$

Os resultados da classificação usando mistura gaussiana estão ilustrados na Figura 7. O gráfico superior mostra as métricas para as amostras com todos 7 perfis, e o gráfico inferior mostras as métricas para todas amostras com algum perfil ausente.

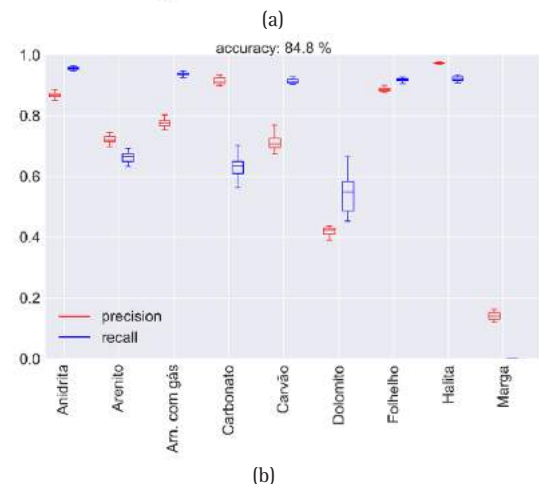
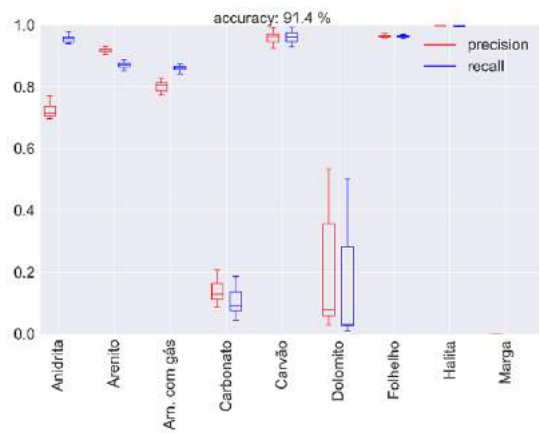


Figura 7: Métricas de classificação do modelo de Mistura Gaussiana. (a) com todos 7 perfis, e (b) sem todos perfis

A mostra as previsões da mistura gaussiana no poço L05-B-03. As previsões são melhores do que as previsões do modelo Naïve Bayes, e não apresentam tão intensamente intercalação entre arenitos e arenitos com gás, porque a mistura gaussiana é capaz de modelar interrelações entre os perfis. Os arenitos e arenitos com gás podem ser diferenciados, por exemplo, pelas diferentes correlações entre os perfis sônicos compressional e cisalhante.

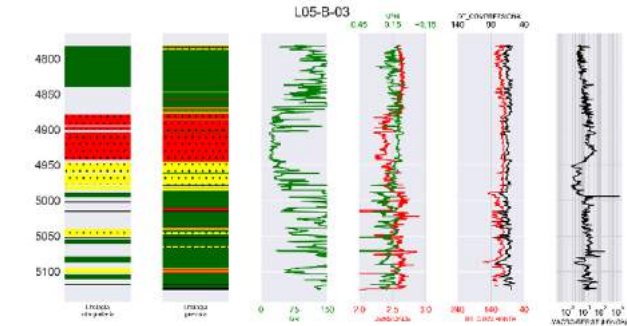


Figura 8: Previsões da mistura gaussiana para o poço L05-B-03. Nota-se uma recuperação consistente das amostras anotadas, e melhor determinação das litologias arenito e arenito com gás do que o modelo Naïve Bayes.

## REDES NEURAIS

Nas redes neurais (Bishop, 2007) a probabilidade de cada litologia é calculada por meio de uma recursão de K passos. Em cada etapa da recursão as medidas de entrada são transformadas em outro conjunto de medidas, através de combinações lineares mais uma operação não-linear. As probabilidades são computadas finalmente pela normalização da última camada da recursão. As equações 10 a 12 descrevem o modelo de rede neural.

$$X^0 = X \quad 10$$

$$X_j^k = \sigma \left( \sum_i w_{i,j} X_i^{k-1} \right) \quad 11$$

$$p(z|X) = \frac{X_z^K}{\sum_{z'} X_{z'}^K} \quad 12$$

Diferentemente da mistura gaussiana ou do modelo Naïve Bayes, as redes neurais não são analiticamente marginalizáveis, portanto, utilizamos uma técnica aproximada para lidar com as variáveis ocultas: a imputação, ou seja, substituição dos valores ausentes por algum valor representativo (Laencina, Gómez, & Vidal, 2009). Durante o treinamento, substituímos os valores ocultos pela média global daquele perfil (imputação simples). Durante a previsão, diferentes resultados são obtidos quando os valores ocultos são substituídos pelos valores médios de cada litologia, e o resultado é então computado como uma média das previsões (imputação múltipla).

A Figura 9 mostra a qualidade da classificação gerada pelas redes neurais. Essa técnica apresenta resul-

## ARTIGO TÉCNICO

tados melhores que as demais nos dados completos, porém apresenta resultados piores nos dados incompletos. Provavelmente, esse resultado pode ser explicado pela imputação que é uma aproximação pobre para a marginalização das variáveis ocultas.

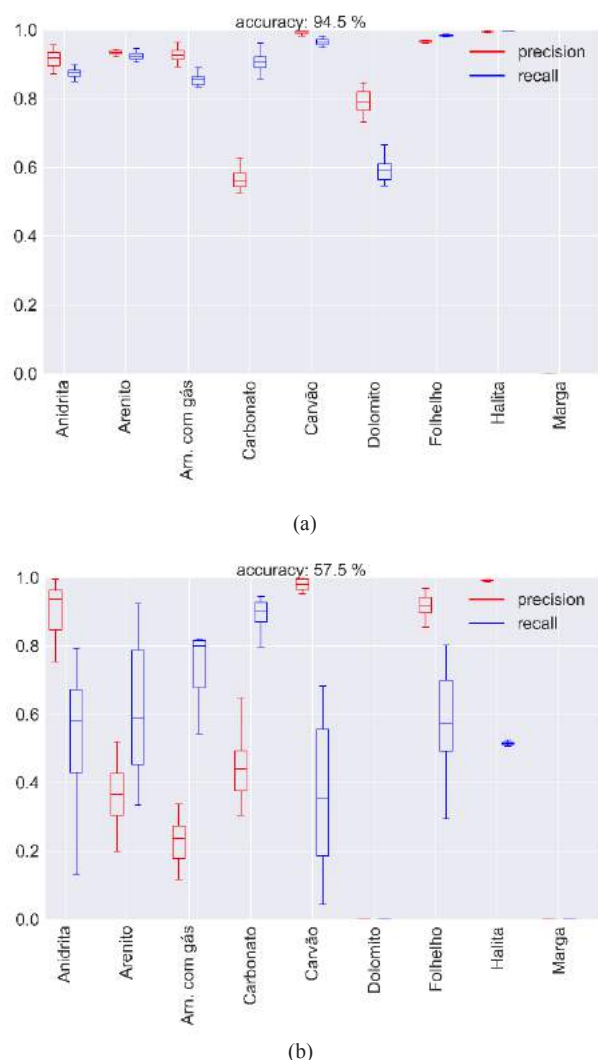


Figura 9: Métricas de classificação do modelo de redes neurais. (a) com todos 7 perfis, e (b) sem todos perfis.

A Figura 10 mostra a previsão da rede neural para o poço L05-B-03. A rede neural apresenta resultado inferior aos modelos anteriores, já que este poço possui perfis ocultos. Nem as amostras anotadas são bem recuperadas pela rede neural.

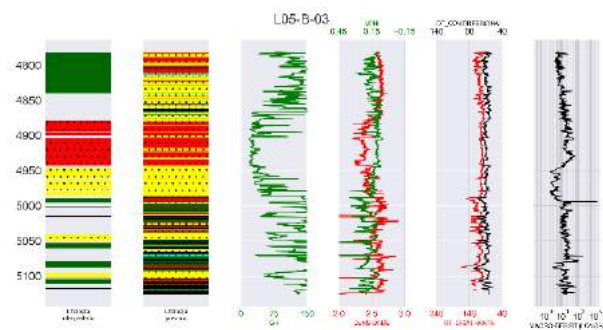


Figura 10: Previsões da rede neural para o poço L05-B-03. O modelo não consegue recuperar bem nem as amostras anotadas, provavelmente à aproximação aplicada para a marginalização das medidas ausentes.

## DISCUSSÃO DOS RESULTADOS

A comparação das métricas de acerto para os diferentes classificadores, na Tabela 3, mostra que a mistura gaussiana apresenta as melhores acurácias dentre as amostras com dados incompletos (84%). Apesar desta acurácia ser menor que o melhor classificador nos dados completos (redes neurais, com 95% de acurácia), o tratamento de dados incompletos permite quintuplicar o intervalo previsto.

(a)	Modelo	acurácia	Número de previsões
	Naïve Bayes	90%	21460
	Mistura Gaussiana	91%	21460
	Rede Neural	95%	21460
(b)	Modelo	acurácia	Número de previsões
	Naïve Bayes	81%	41181
	Mistura Gaussiana	85%	109664
	Rede Neural	58%	109664

Tabela 3: Comparação das acurácias médias por modelo (a) para as amostras com os 7 perfis, e (b) para as amostras com ao menos um perfil ausentes (exceção para o Naïve Bayes, com número fixo de 5 perfis).

As redes neurais, apesar de tenderem a ser uma técnica mais robusta, apresentam resultados bem inferiores aos outros classificadores. Provavelmente isso se deve à imputação não ser uma boa forma de se lidar com dados incompletos (marginalização é o modo mais teoricamente correto), e ao fato de termos utilizado duas estratégias diferentes de imputação, uma para o treino e outra para o teste.

## CONCLUSÕES

No presente trabalho discutimos a importância de utilizar técnicas de classificação que levem em consideração dados incompletos. Mostramos que tais técnicas podem aumentar em muito o intervalo predito, sem grande perda em acurácia.

Discutimos algumas estratégias para lidar com dados incompletos, e seus resultados no conjunto de dados fornecido pelo SGH. No entanto, futuros desenvolvimentos são encorajados para melhorar as estratégias utilizadas. A expectativa é que o modelo de redes neurais, se melhor adaptado ao problema de dados incompletos, apresente melhores resultados haja visto sua flexibilidade e grande sucesso em aplicações diversas na literatura.

## REFERÊNCIAS

- Bishop, C. (2007). *Pattern Recognition and Machine Learning*. Springer.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 1-26.
- Laencina, P., Gómez, J., & Vidal, A. (2009). *Pattern classification with missing data: a review*. Springer.
- Walber. (2017). *Precision and Recall*. Retrieved from Wikipedia: <https://commons.wikimedia.org/w/index.php?curid=36926283>
- Xie, Y., Zhu, C., Zhou, W., Li, Z., Liu, X., & Tu, M.



(2018). Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. *Journal of Petroleum Science and Engineering*, 182-193.



CALL FOR PAPERS

Curta a SBGf  
no facebook!



[www.facebook.com/sbgf.org](http://www.facebook.com/sbgf.org)

multi-client seismic  
**BRAZIL**

## Offshore Brazil

2D & 3D Seismic Coverage for License Rounds

**Legend**

- Spectrum Multi-Client 3D
- Spectrum Multi-Client 2D library
- In Progress
- Open Door 1st Cycle
- Open Door 2nd Cycle

Spectrum has the data you need to evaluate upcoming Brazil offshore licensing rounds. Extensive Multi-Client seismic coverage along the coast of Brazil totalling over 300,000 km<sup>2</sup> 2D and 19,000 km<sup>2</sup> 3D data.

spectrumgeo.com  
mc-us@spectrumgeo.com  
+1 281 647 0602



© 2017 Halliburton. All Rights Reserved.

## Increase Production. Maximize Results.

### THROUGHOUT THE LIFE OF YOUR ASSETS

Whatever your production challenge, Halliburton offers a full range of engineered solutions. From real-time diagnostic well interventions to customized specialty chemicals, reliable artificial lift systems, and pipeline and process pre-commissioning and maintenance solutions, we're ready to help keep your production high and costs low. Contact us to learn more.

[halliburton.com/production](http://halliburton.com/production)

## 2018-2019

### VIII Simpósio Brasileiro de Geofísica

18 a 20 de setembro de 2018

Salinópolis - Pará

Informações: <http://35.184.7.207/simposio/>

### First EAGE/SBGf Workshop on Least Squares Migration

27 e 28 de novembro de 2018 - Rio de Janeiro - Brasil

Informações: [www.sbgf.org.br](http://www.sbgf.org.br)

### GEM 2019 Xi'an: International Workshop on Gravity, Electrical, & Magnetic Methods and Their Applications

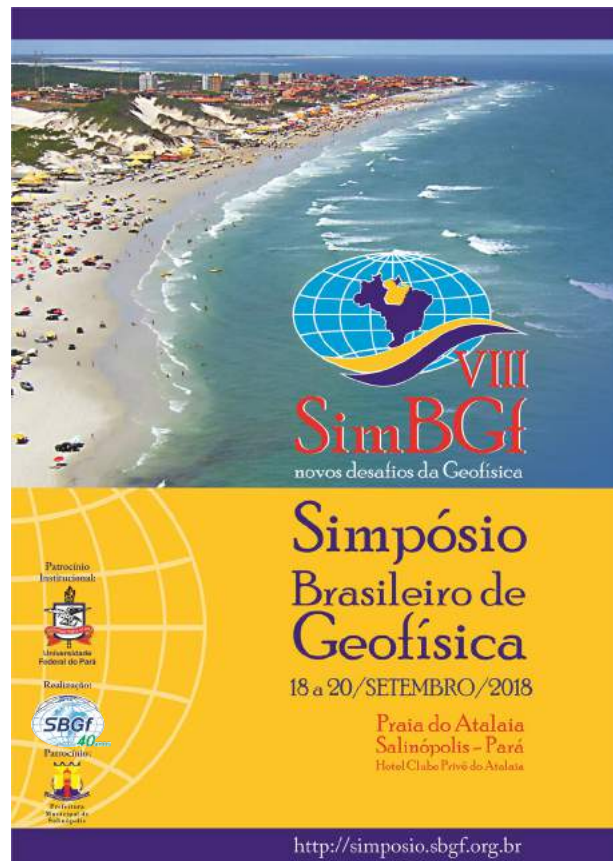
19 a 22 de maio de 2019 - Xi'an - China

Informações: <https://seg.org/Events/Events-Calendar/GEM-2019-Xian>

### 16th International Congress of the Brazilian Geophysical Society - Call For Papers

19 e 22 de agosto de 2019 - Rio de Janeiro - Brasil

Informações: [www.sbgf.org.br](http://www.sbgf.org.br)



[jornadapeloconhecimento.com.br](http://jornadapeloconhecimento.com.br)

BR PETROBRAS

MINISTÉRIO DE MINAS E ENERGIA

