

## Seleção de Características para Classificação de Fluidos por Meio da Análise de Perfis Elétricos de Poços de Petróleo da Amazônia

Joacir M. de Oliveira Júnior<sup>1</sup>, Eulanda M. dos Santos<sup>1</sup>, José Reginaldo Carvalho<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Informática  
Universidade Federal do Amazonas (UFAM) – Manaus – AM

{joacir,emsantos,reginaldo}@dcc.ufam.edu.br

**Abstract.** *Among several steps which are necessary in order to lead oil to be marketed, the analysis of well logs plays an important role to estimate the capacity of a well. One goal of this analysis is to classify fluids into oil, gas or water. This analysis is traditionally conducted as a semi-automated process which has a system to generate graphs of curves. We propose in this work to carry out well logs classification as an automatic pattern recognition process. In addition, we compare two classifiers in order to identify the best method for automatic analysis of well logs data which are obtained from the same reservoirs. Finally, we conduct an exhaustive feature subset selection process focusing on pointing out the best feature set for fluids classification. Our results indicate that kNN achieved the best recognition rates, while the feature subset composed of six features, namely DT, GR, ILD, MSFL, NPFI e RHOB, is the most relevant information for fluids classification. Our feature subset selection strategy helped to reduce the number of features while the recognition rate was increased.*

**Resumo.** *Dentro das várias etapas que são necessárias até o petróleo ser comercializado, a análise de perfis elétricos representa papel de grande importância para se estimar a capacidade produtiva de um poço. Um dos objetivos da análise é a classificação de fluidos em água, óleo, gás e não reservatório. Essa análise atualmente é semi-automatizada, onde um sistema gera gráficos de curvas a partir dos quais geólogos com conhecimentos específicos, analisam e realizam a caracterização do reservatório. Outro ponto importante para os geólogos é a escolha dos perfis a serem utilizados como determinantes à classificação. Neste trabalho busca-se realizar o reconhecimento automático dos padrões destes dados, demonstrar e comparar classificadores capazes de realizar análise automática de dados de perfis de poços de petróleo com índice de acerto igual ou superior à análise de especialistas humanos e identificar os melhores perfis para a classificação de fluidos por meio de testes exaustivos com diversas combinações de perfis. Nossos resultados indicam que o classificador com melhor desempenho foi o kNN, e o melhor subconjunto de características é composto pelos perfis DT, GR, ILD, MSFL, NPFI e RHOB, sendo que o número de perfis foi reduzido enquanto ocorreu o aumento na taxa de acerto do classificador.*

## 1. Introdução

A constatação de que a Amazônia é uma região petrolífera ocorreu no dia 13 de março de 1955, segundo relatos históricos de Garcia (2008). O poço pioneiro foi batizado de NO-1-AZ e localizava-se a 125 quilômetros de Manaus, a uma profundidade de 2.744 metros. Esse evento produziu muita euforia, especialmente na população do estado do Amazonas, pois a perspectiva de produção do poço NO-1-AZ depois dos primeiros testes realizados era de 600 a 700 barris por dia. Essa perspectiva de produção era suficiente para estimular mais pesquisas e a abertura de novos poços na região. Além disso, na época acreditava-se que em apenas três anos o Brasil estaria livre da importação de combustíveis, tão grande era o otimismo dos engenheiros da Petrobras.

Atualmente, a indústria de petróleo brasileira cresce vigorosamente. Conforme mostra a Tabela 1, o Brasil produz mais de 2,5 milhões de barris por dia e o estado do Amazonas é responsável por cerca de 107.496 barris diários. Segundo o plano de negócios da Petrobras (2011), a empresa espera duplicar suas reservas provadas até 2020. A realidade da indústria de petróleo brasileira é diferente da situação da indústria mundial, pois a produção de petróleo declina mundialmente e seus preços disparam no mercado internacional. Portanto, a pesquisa na área de petróleo apresenta elevada demanda no Brasil.

**Tabela 1. Distribuição da Produção de Petróleo e Gás Natural no Brasil por Estado. Ano de 2011. FONTE: ANP/SDP/SIGEP. Siglas: bbl = oil barrel, Mm<sup>3</sup> = metros cúbicos, boe = barris de óleo equivalente.**

Estado	Petróleo (bbl/d)	Gás Natural (Mm <sup>3</sup> /d)	Produção Total (boe/d)
Rio de Janeiro	1.551.597	25.360	1.711.114
Espírito Santo	319.518	11.901	394.373
Amazonas	35.110	11.508	107.496
Bahia	42.880	6.565	84.172
São Paulo	42.225	5.169	74.739
Rio Grande do Norte	58.639	1.701	69.338
Sergipe	42.365	2.428	57.640
Alagoas	5.198	1.539	14.878
Ceará	7.209	86	7.752
<b>Total geral</b>	<b>2.104.743</b>	<b>66.257</b>	<b>2.521.502</b>

Antes de um poço começar a produzir, existe um longo e caro processo que envolve o estudo e a análise de dados geológicos e geofísicos. Esta fase é denominada **avaliação de formações**. Trata-se de atividades e estudos que visam definir em termos qualitativos e quantitativos o potencial de uma jazida petrolífera, isto é, sua capacidade produtiva e a valoração de suas reservas de óleo e gás (Thomas et al., 2001). Essa avaliação é baseada principalmente nos dados coletados na perfilagem de um poço aberto. Conforme mostra a Figura 1 e segundo Thomas et al. (2001), a perfilagem é uma imagem visual em relação à profundidade, de uma ou mais características ou perfis das rochas perfuradas tais como resistividade elétrica, potencial eletroquímico natural, tempo de trânsito das ondas mecânicas, radioatividade natural ou induzida. Outra definição pode ser encontrada em (An-nan and Lu, 2009). Esse autor afirma que perfilagem é o método que busca identificar propriedades das

rochas e fluidos medindo informações elétricas, acústicas, termais e radioativas. Portanto, com a perfilagem é possível extrair dados importantes das formações pelas quais o poço atravessa como a litologia (tipo de rocha), espessura, porosidade, presença de óleo, gás ou água. Dependendo da análise dos perfis e das informações complementares, pode se decidir pelo abandono do poço ou no caso de se provar interesse econômico, os testes de formação para se estimar a capacidade produtiva do poço são iniciados.

A classificação de perfis de poços de petróleo pode claramente ser tratada como um problema de reconhecimento automático de padrões. Não é surpresa a existência de trabalhos na literatura que realizam a classificação dos perfis de poços com ferramentas de reconhecimento de padrões (Cunha, 2002), (Flexa et al., 2004), (Wong et al., 2005) e (An-nan and Lu, 2009). Porém, apesar da possibilidade de resolução de diferentes problemas de classificação, tais como: litologia, porosidade, fluido, etc., a maioria dos trabalhos publicados na literatura realiza classificação de litologia, também conhecida como classificação de litofácies. Uma das principais razões desse consenso é que a classificação de litologia pode ser caracterizada apenas com a leitura dos perfis, fato que torna essa tarefa mais simples de ser automatizada do que a classificação de fluidos, por exemplo, que é mais complexa por depender da análise de um especialista (geólogo) com conhecimento sobre a reserva. Fluidos são classificados em óleo, gás ou água.

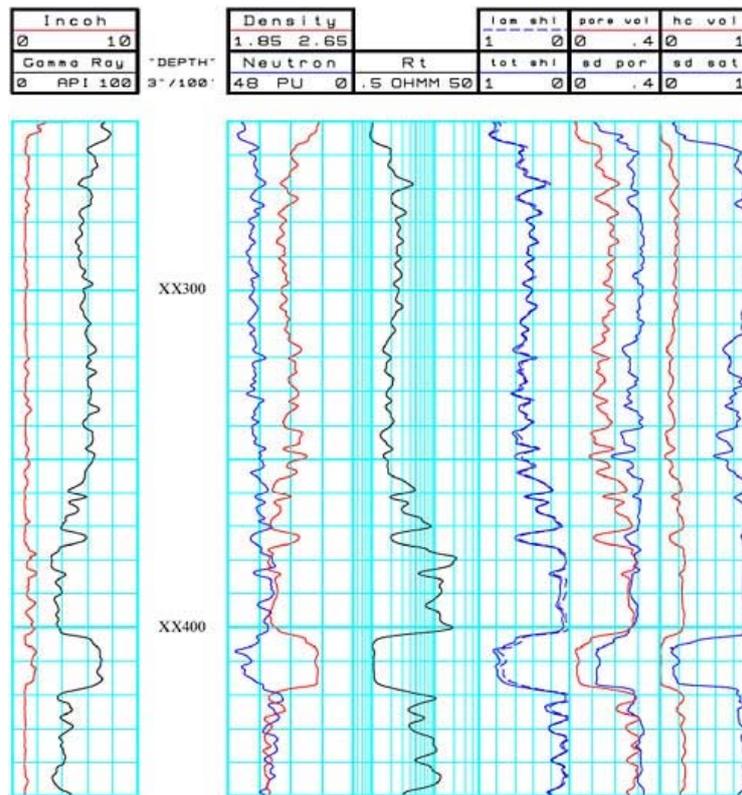
Porém, a identificação de fluidos de um reservatório de petróleo é um processo intensivo que envolve o investimento de uma quantidade considerável de tempo por parte de um especialista experiente e com conhecimento da reserva em estudo. O problema se torna muito mais difícil à medida que aumenta o número de perfis simultâneos a serem analisados (Cunha, 2002). É papel do geólogo determinar quais os melhores perfis a serem analisados para a reserva e fazer a caracterização do reservatório que é de fundamental importância para justificar investimentos da continuidade de exploração de um poço ou o seu abandono. Portanto, a automação do processo de seleção de perfis e da identificação de fluidos pode acelerar a etapa de caracterização de reservatórios.

A seleção de perfis para a classificação de fluidos pode ser visto como um processo de otimização conhecido na literatura como seleção de características ou atributos. Segundo Kudo and Sklansky (2000), a seleção de características contribui para: (1) reduzir os custos de extração de características; (2) aumentar a acurácia da classificação; e (3) aumentar a confiança da estimativa de desempenho. O primeiro aspecto é uma consequência da redução da dimensão do espaço de entrada obtida por meio da eliminação de características irrelevantes. Os dois últimos aspectos estão relacionados com o fato de que métodos de classificação são ajustados para tomar decisões com base nas características mais relevantes. Logo, características irrelevantes ou redundantes podem confundir o classificador.

O processo de seleção de subconjuntos de perfis pode ser realizado por algoritmos de busca, normalmente não exaustivos, especialmente quando há muitos perfis a serem testados. Dada uma amostra  $\mathcal{X}$ , representada por um vetor de características (perfis), onde  $\mathcal{X} = \{x_1, x_2, \dots, x_d\}$ , algoritmos de busca são usados com frequência quando  $\mathcal{X}$  possui muitos perfis devido à complexidade exponencial que

uma busca exaustiva demandaria, uma vez que, sendo  $\mathcal{P}(\mathcal{X})$  o conjunto de todos os subconjuntos possíveis dos perfis de  $\mathcal{X}$ , o tamanho de  $\mathcal{P}(\mathcal{X})$  é  $2^d$ , onde  $d$  representa o número de perfis em  $\mathcal{X}$ .

Na análise de perfis, entretanto, há oito principais tipos de perfis elétricos: Raios Gama (GR), Neutrônico (NPHI), Indução (ILD), Sônico (DT), Densidade (RHOB), Caliper (Cal), Micro SFL (MSFL) e Índice Fotoelétrico (PEF). A descrição de cada perfil é mostrada na Tabela 2. Portanto, um processo de busca exaustiva pode ser utilizado devido ao pequeno número de perfis que representam cada amostra  $\mathcal{X}$ . É importante também destacar que esses perfis são os mesmos utilizados para a classificação de litologias na maioria dos trabalhos publicados na literatura. Apesar do papel fundamental da etapa de seleção de características em problemas de classificação, normalmente não é realizada uma análise de impacto de cada perfil nos trabalhos relacionados com o estudo de perfis elétricos (Cunha, 2002), (Flexa et al., 2004), (Wong et al., 2005) e (An-nan and Lu, 2009). Logo, todos os perfis são utilizados para treinar os algoritmos de classificação.



**Figura 1. Perfis elétricos coletados durante a perfilagem de um poço.**

Diante desse contexto, este trabalho aplica métodos de classificação ao problema de identificação automática de fluidos utilizando dados de perfis de um poço de petróleo da Amazônia. Para que a taxa de acerto dos métodos de classificação seja maximizada, é realizada uma etapa de seleção de características para o problema investigado. O processo de otimização é realizado de forma exaustiva devido ao pequeno número de perfis investigados, isto é, oito perfis. Os objetivos deste trabalho são: (1) identificar os perfis mais relevantes para a tarefa de classificação

**Tabela 2. Principais Tipos de perfis elétricos e suas propriedades.**

Perfil	Descrição
Raios Gama (GR)	Detecta a radioatividade total da formação geológica. Utilizado para a identificação da litologia, a identificação de minerais radioativos e para o cálculo do volume de argilas ou argilosidade.
Neutrônico (NPHI)	São utilizados para estimativas de porosidade, litologia e detecção de hidrocarbonetos leves ou gás.
Indução (ILD)	Fornece leitura aproximada da resistividade da rocha contendo hidrocarbonetos, através da medição de campos elétricos e magnéticos induzidos nas rochas.
Micro SFL (MSFL)	Similar ao ILD.
Índice Fotoelétrico (PEF)	É uma medida complementar de ferramentas de medição de densidade que registra a absorção de raios gama.
Sônico (DT)	Mede a diferença nos tempos de trânsito de uma onda mecânica através das rochas.
Densidade (RHOB)	Além da densidade das camadas, permite o cálculo da porosidade e a identificação das zonas de gás.
Caliper	Fornece o diâmetro do poço. É aplicado no cálculo do volume de cimento para tampões ou cimentação do revestimento, apoio a operações de teste de formação, controle de qualidade de perfis e indicações das condições do poço em um determinado intervalo.

automática de fluidos; (2) mostrar que métodos de classificação automática podem alcançar índice de acerto igual ou superior à análise de especialistas humanos; e (3) comparar dois métodos de classificação, kNN (*k Nearest Neighbors*) e SVM (*Support Vector Machines*), a fim de também identificar o método com maior taxa de classificação correta para o problema.

O restante deste artigo está organizado da seguinte forma: na seção 2, é feito um levantamento do estado da arte através da descrição das soluções propostas nos principais trabalhos encontrados na literatura para o problema de análise de perfis elétricos. Na seção 3 é descrita a metodologia que será empregada no desenvolvimento do trabalho, bem como a base de dados, normalização dos dados, descrição dos experimentos realizados e dos resultados obtidos com os classificadores kNN e SVM. A seção 3.2 descreve a análise de características realizada através de um método exaustivo o qual seleciona os principais perfis que permitem atingir melhores taxas de acerto na classificação por fluidos. As discussões sobre os resultados obtidos são apresentadas na seção 4. Por fim, a seção 5 resume as conclusões deste trabalho.

## 2. Trabalhos Relacionados

Algumas tentativas de automatização do processo de análise de perfis elétricos têm sido publicadas na literatura. Nesta seção são discutidos alguns desses trabalhos. Conforme mencionado na introdução, a maioria dos trabalhos realiza classificação de litologia.

Cunha (2002) utilizou dados de perfis e de testemunhos de poços para treinar redes neurais para a classificação de litofácies. A rede neural foi treinada com dados

de 8 poços de petróleo e atingiu uma taxa de acerto de 80%. A autora relata que uma solução para o problema só foi possível após a incorporação de uma estratégia para agrupamento prévio das litofácies e tratamento de padrões problemáticos (regiões de conhecimento incerto nos conjuntos de treinamento e de teste). Outro ponto a destacar é que a autora utilizou apenas redes neurais, que têm apresentado um desempenho inferior quando comparadas com SVM (Al-Anazi and Gates, 2010a), (Wong et al., 2005), por exemplo.

Flexa et al. (2004) propuseram uma metodologia baseada na técnica de análise discriminante para a identificação de litologias nos perfis de poços. Os autores pesquisaram a associação entre as variações das propriedades físicas e os tipos litológicos em profundidade. Os perfis geofísicos utilizados nesse trabalho foram raios gama (GR), sônico (DT), porosidade neutrônica (PHIN), densidade (RHOB) e resistividade (ILD). Segundo Flexa et al. (2004), a análise discriminante é uma técnica da estatística multivariada com grande eficiência na solução de problemas geofísicos. A técnica é utilizada nos casos em que se deseja realizar a separação estatística entre dois ou mais grupos de dados, que devem ser definidos previamente pela caracterização de suas variáveis discriminantes. Entretanto, no trabalho de Flexa et al. (2004), a técnica de análise discriminante não é comparada com outros classificadores. Os autores consideram que bons resultados foram obtidos, porém o desempenho real não foi apresentado.

Wong et al. (2005) investigaram o desempenho do classificador SVM para a caracterização de reservatórios como permeabilidade e porosidade. Essas características são fundamentais para a localização de hidrocarbonetos, pois, sem uma boa porosidade não há como a rocha conter fluidos e sem uma permeabilidade adequada, o fluido não tem como escoar entre os poros da rocha, podendo assim, ser retirado. Nos experimentos, o classificador SVM teve uma taxa de erro de 2.6% enquanto o classificador baseado em rede neural do tipo *Backpropagation* (BPNN) atingiu uma taxa de erro superior, 7.4%. Apesar desse trabalho não ter como objetivo a classificação de litofácies, é realizada a classificação de características das rochas como porosidade e permeabilidade, em que os mesmos perfis utilizados para classificar litologias e fluidos são empregados.

An-nan and Lu (2009) propuseram um trabalho de identificação de litologias a partir de dados de perfilagem utilizando DE-SVM com uma taxa de acerto de 90%. DE é o acrônimo de diferença evolutiva, um algoritmo baseado em algoritmos genéticos. Os autores observam que SVM tem sido usado com sucesso para identificação de litologias, mesmo com uma amostra pequena de dados. O artigo apresentou um método de otimização baseado na diferença aritmética-evolutiva (DE) para selecionar os parâmetros. Além disso, como o objetivo principal dos autores foi desenvolver um método para encontrar os melhores parâmetros para SVM, nesse trabalho não houve comparações entre classificadores como kNN e redes neurais.

Al-Anazi and Gates (2010a) comentam sobre a necessidade de novas técnicas de predição de litologia em reservatórios que sejam mais eficientes que os métodos estatísticos. Os autores citam a aplicação de técnicas como redes neurais e, mais atualmente, a aplicação de SVM com bons resultados mesmo em reservatórios heterogêneos. Eles comparam então esses três métodos e revelam melhores resultados

com SVM frente à análise estatística discriminante e redes neurais. Os experimentos foram realizados a partir de dois poços provenientes de reservatórios distintos. Em um outro trabalho mais recente, Al-Anazi and Gates (2010b) aplicam SVM em um reservatório bastante heterogêneo com dados de três poços para classificar eletrofácies e prever distribuição de permeabilidade. Novamente, os autores comparam o desempenho de SVM com redes neurais e com a técnica de análise discriminante, e também confirmam a superioridade de SVM frente às duas outras técnicas.

Conforme pode ser observado nesta seção, existem diversos trabalhos sobre classificação de litofácies utilizando ferramentas de reconhecimento de padrões. Durante as pesquisas de trabalhos de percepção de padrões utilizando perfis elétricos, não foram encontrados trabalhos sobre a classificação de fluidos. Como foi mencionado na introdução, o motivo é que a classificação de fluidos em óleo, gás ou água exige do geólogo um profundo conhecimento da área de exploração. Outro fato que pode ser observado, é que os trabalhos mais recentes para o problema de classificação de litofácies têm demonstrado a superioridade do classificador SVM frente a classificadores como redes neurais e análise estatística discriminante ((Al-Anazi and Gates, 2010a), (Wong et al., 2005)). Porém, kNN não foi investigado. Por fim, nenhum trabalho relacionado nesta seção realizou a seleção dos perfis elétricos. Essa seleção pode contribuir para aumentar a taxa de acerto do classificador, conforme será mostrado na próxima seção.

### 3. Experimentos e Resultados

Esta seção descreve os experimentos e os resultados obtidos. Inicialmente, porém, descrevemos os detalhes do protocolo experimental utilizado nos experimentos. Em seguida, é feita uma análise de características através de um método exaustivo para descobrir quais perfis são mais decisivos para a classificação de fluidos. Por fim, são apresentados os resultados da comparação entre os classificadores SVM e kNN.

#### 3.1. Protocolo Experimental

Os experimentos foram realizados com uma base de dados composta por 283 amostras de perfis de um único poço, sendo que cada amostra representa uma profundidade específica do poço. Foi realizada uma etapa de ajuste de parâmetros, onde no caso de kNN, o melhor valor de  $k$  encontrado foi  $k=1$ . Para o classificador SVM, o parâmetro de regularização  $C$  foi variado de 0.001 a 1000 sem alteração na classificação. O kernel escolhido deste classificador foi o polinomial, sendo que grau do polinômio=2 obteve melhores taxas de acerto. Os dados estão divididos em quatro classes de fluidos: água, gás, petróleo e não reservatório, e foram rotulados por um especialista geofísico. A Tabela 3 mostra um exemplo das amostras da base de dados.

Como pode ser observado na tabela 3, os valores obtidos para cada perfil estão em escalas diferentes. Essa não normalização dos dados pode influenciar negativamente o desempenho dos classificadores, especialmente kNN, que é baseado em medida de distância. Diante desse fato, um processo de normalização dos dados foi realizado. Aksoy and Haralick (2000) apresentam um estudo comparativo entre seis procedimentos de normalização que possibilitam que os dados fiquem dentro

**Tabela 3. Exemplo de amostras da base de dados**

n <sup>o</sup>	Depth	Cal	DT	GR	ILD	MSFL	NPHI	Pef	RhOB
1	2435,0471	6,7227	56,2182	106,4375	27,6483	35,1158	10,8416	3,168	2,6629
2	2435,1995	6,7227	57,9455	146,5	25,7482	53,9805	12,8258	3,1602	2,6624
278	2477,262	6,7099	71,9797	75,8699	0,6001	1,6795	11,4636	2,9233	2,5086
283	2478,0239	6,582	68,1109	63,4688	5,0932	7,0145	13,3006	2,5742	2,4555

de um intervalo  $[0,1]$ . Neste trabalho foram utilizados os dois primeiros procedimentos investigados em Aksoy and Haralick (2000). No primeiro procedimento de pré-processamento é feito um escalonamento linear normalizando os dados, sendo que é dado um limite inferior  $l$  e um limite superior  $u$  para uma característica  $x$ ,

$$\tilde{x} = \frac{x - l}{u - l}, \quad (1)$$

resultando em  $\tilde{x}$  dentro de um intervalo  $[0,1]$ .

No segundo procedimento de pré-processamento os dados são padronizados, consistindo em transformar uma característica  $x$  para uma variável aleatória com média zero e variância igual a 1 como:

$$\tilde{x} \cong \frac{\mu - l}{\sigma}, \quad (2)$$

onde  $\mu$  e  $\sigma^2$  são a média da amostra e a variância igual a 1 da característica  $x$ , respectivamente.

A estratégia de validação cruzada conhecida como *leave-one-out* foi utilizada para a divisão dos dados em base de treinamento e teste. Essa estratégia consiste em treinar o classificador com todas as amostras, exceto uma amostra que é usada para o teste. Em seguida, uma nova amostra é selecionada para teste, enquanto as demais são usadas para treinamento. Esse processo é repetido até que todas as amostras sejam usadas como amostra de teste. O resultado da classificação é obtido por meio do cálculo da razão entre o número de acertos e o número total de amostras. *Leave-one-out* foi utilizado devido à pequena quantidade de amostras na base de dados. Os parâmetros de SVM e kNN também foram ajustados.

Os métodos de normalização são comparados na tabela 4, com base na taxa de acerto dos classificadores kNN e SVM. Como pode ser observado na tabela 4, o classificador kNN foi mais beneficiado pela normalização dos dados do que SVM. O método kNN melhorou seu desempenho com ambos os procedimentos de normalização, fato que não foi observado com o classificador SVM, que não alterou a taxa de acerto de 94.34%. Esses resultados preliminares indicam que kNN, com  $k=1$ , superou SVM (polinomial de grau 2) em taxa de acerto. Entretanto, a maioria das classificações incorreta, tanto para SVM quanto kNN, ficou dentro da zona de transição entre água, óleo e gás. É possível que a própria classificação feita pelo especialista esteja incorreta, pois o limite entre essas zonas se confunde. A comparação entre kNN e SVM é discutida com mais detalhes na seção 3.2.

**Tabela 4. Comparação dos classificadores kNN e SVM com e sem normalização.**

Classificador	Sem Normalização	Normalização Tipo 1	Normalização Tipo 2
SVM	94.34%	94.34%	94.34%
kNN	95.75%	96.46%	96.11%

**Tabela 5. Taxas de erro de perfis analisados individualmente com o classificador kNN com k=1.**

Perfis Utilizados	Cal	DT	GR	ILD	MSFL	NPHI	PEF	RHOB
Tx de Erro	67.12	49.82	55.12	<b>19.43</b>	<b>28.62</b>	48.4	57.24	43.46

### 3.2. Análise de características

A seleção de características no contexto de classificadores baseados em percepção de padrões tem três objetivos: (1) reduzir os custos de extração de características, (2) aumentar a acurácia da classificação e (3) aumentar a confiança da estimativa de desempenho (Kudo and Sklansky, 2000).

Para descobrir qual ou quais as melhores combinações de perfis para a classificação, testamos todas as combinações possíveis para os 9 perfis, desconsiderando a profundidade uma vez que a mesma varia para cada poço e reserva e tende mais a ser prejudicial do que benéfica para a classificação, quando comparado com mais de um poço. No trabalho de Wong et al. (2005) a profundidade também foi retirada pois o autor trabalhou com reservatórios heterogêneos. Restaram assim 8 perfis (Cal, DT, GR, ILD, MSFL, NPHI, PEF e RHOB) conforme mostrado na Tabela 5, e estes geraram 244 combinações distintas. Combinações onde os perfis eram os mesmos e apenas a ordem era distinta, foram rejeitadas por não apresentarem alteração na taxa de acerto.

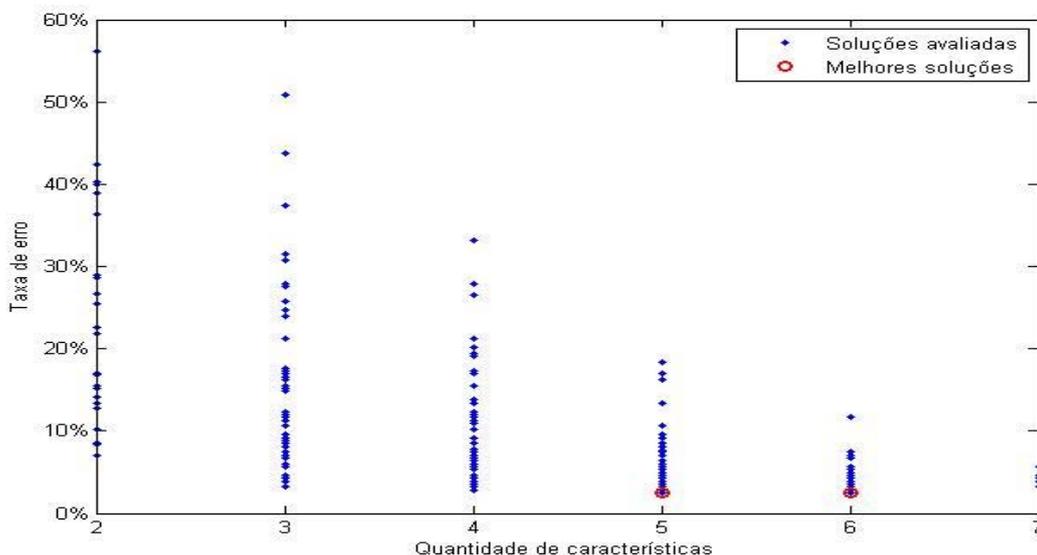
Inicialmente, o impacto individual de cada perfil foi analisado, ou seja, o classificador foi treinado com apenas uma característica por vez. As taxas de erro obtidas estão resumidas na Tabela 5. É possível observar nessa tabela que os perfis ILD e MSFL produziram os melhores resultados individuais, sendo que o classificador obteve taxas de erro iguais a 19.43% e 28.62%, respectivamente para cada perfil. Por outro lado, o perfil CAL foi o menos relevante para a classificação de fluidos, dado que o classificador apresentou o pior resultado (67.12% de taxa de erro) ao ser treinado apenas com esse perfil. Entretanto, conforme esperado, apenas um perfil não fornece informação suficiente para o classificador produzir elevada capacidade de generalização. Logo, a seleção de subconjuntos de perfis busca aumentar o poder de discriminação das informações fornecidas para treinar o método de classificação.

Na tentativa de identificar o melhor subconjunto de características, foram feitos testes exaustivos com todas as combinações de perfis, com o objetivo de aumentar a taxa de acerto e diminuir o número de características. Essa tarefa seria proibitiva caso o número de perfis fosse muito elevado. Os resultados dessa busca são sintetizados na Tabela 6. Pode ser observado que as melhores taxas de acerto foram obtidas com a seguinte combinação de perfis (DT, GR, ILD, MSFL, NPHI, RHOB), tanto com SVM quanto com kNN. Portanto, o número de características foi reduzido, de 08 para 06 perfis, enquanto a taxa de acerto aumentou em 0.34%

**Tabela 6. Melhores Combinações de Perfis.**

Classif.	Perfis utilizados	Tx. Acerto	Normalização
SVM	GR, ILD, MSFL, NPHI, RHOB	94.34%	tipo 2
SVM	DT, GR, ILD, NPHI, RHOB	94.34%	tipo 2
SVM	DT, ILD, MSFL, PEF, RHOB	92.22%	tipo 2
SVM	DT, GR, ILD, MSFL, NPHI, RHOB	94.69%	tipo 1
SVM	DT, GR, ILD, MSFL, NPHI, RHOB	94.69%	tipo 2
kNN	GR, ILD, MSFL, NPHI, RHOB	96.81%	tipo 2
kNN	DT, GR, ILD, NPHI, RHOB	97.17%	tipo 2
kNN	DT, ILD, MSFL, PEF, RHOB	95.40%	tipo 2
kNN	DT, GR, ILD, MSFL, NPHI, RHOB	97.52%	tipo 1
kNN	DT, GR, ILD, MSFL, NPHI, RHOB	97.52%	tipo 2

com SVM e 1.06% com kNN, confirmando a importância do processo de seleção de características. Outro fator de destaque é que combinações com 05 perfis também apresentaram bons resultados. Os seguintes subconjuntos de cinco perfis produziram os melhores resultados: (DT, GR, ILD, NPHI, e RHOB) com SVM - reduziu o número de características e manteve a taxa de acerto do classificador; e (GR, ILD, MSFL, PEF e RHOB) com kNN que também reduziu o número de características e aumentou a taxa de acerto. Embora as combinações sejam diferentes, os perfis GR, ILD e RHOB estão presentes nas duas combinações. O perfil ILD também foi o perfil mais relevante individualmente. A Figura 2 mostra o comportamento das soluções avaliadas durante o processo de busca.



**Figura 2. Evolução do Método Exaustivo.**

Analisando a matriz de confusão na Tabela 7 e a Figura 1 concluímos que a maioria dos erros ocorreu dentro das zonas de transição entre não reserva e água. Conforme mencionado anteriormente, é possível que os próprios rótulos dos dados nessas zonas de transição estejam incorretos, pois o limite entre essas zonas se confunde.

**Tabela 7. Matriz de confusão.**

	Não Reservatório	Gás	Óleo	Água
Não Reservatório		1		
Gás	1		1	
Óleo		1		2
Água	1			

#### 4. Discussão

Respondemos com sucesso à seguinte pergunta:

- É eficaz o aprendizado dentro de um poço de petróleo para as demais profundidades do mesmo poço?

A resposta é sim, o classificador obtém uma alta taxa de acerto para as demais profundidades de um poço, além de utilizar uma representação reduzida dos dados, obtida por meio do processo de seleção de subconjuntos de perfis. No entanto, devido ao fato da base pertencer a apenas um poço, não foi possível verificar se o aprendizado gerado com dados de um único poço é suficientemente representativo para prever água, gás ou óleo de poços distintos, na mesma reserva e em diferentes reservatórios.

Foram entrevistados 03 geólogos sobre essa questão, sendo dois da Petrobras e 01 de uma empresa de perfilagem dos Estados Unidos. Todos dizem que a classificação de fluidos em água, óleo ou gás, não pode ser feita somente com os dados da perfilagem e que necessitam de um grande conhecimento da reserva pelos geólogos. Isso acontece porque, em muitos casos, os mesmos parâmetros que em uma reserva indicam a presença de óleo, em outra reserva podem indicar a presença de gás ou nenhuma presença de fluidos. Assim, o aprendizado de um poço, dificilmente poderia ser aproveitado para outros poços em reservas diferentes. Porém, para poços vizinhos ou com características similares do reservatório, o aprendizado poderia ser generalizado. Para responder esta pergunta com confiabilidade, serão necessários testes com bases de outros poços da mesma reserva.

#### 5. Conclusão

Neste trabalho foi realizado um estudo de seleção de características para o problema de classificação de fluidos por meio de perfis elétricos. A seleção dos perfis foi realizada através de um método exaustivo em que foram testadas todas as combinações possíveis (excluindo as mesmas combinações em ordens diferentes). Os testes foram feitos com somente 1 perfil até os 8 perfis combinados. Nossos resultados indicaram que 6 perfis (DT, GR, ILD, MSFL, NPHI, RHOB) são as informações mais relevantes tanto para o aprendizado do classificador kNN quanto de SVM, provando assim a importância da escolha das melhores características.

Foi também apresentada uma comparação entre os classificadores kNN e SVM. Em nossos experimentos realizados com uma pequena base de 283 amostras, kNN obteve melhores taxas de acerto. A comparação foi feita com dados de apenas um poço por meio da técnica *leave-one-out* para garantirmos uma melhor confiabilidade dos dados apresentados. No entanto, para a classificação de fluidos, é possível que o aprendizado obtido com dados de um poço somente seja relevante

para poços vizinhos ou poços com as mesmas características da base de treino. Ao se trabalhar em uma nova reserva, será necessário que o geólogo analise o novo poço para que poços vizinhos possam ser classificados automaticamente. Nossa proposta como trabalho futuro é incorporar à base dados de outros poços da mesma reserva e propor um sistema em que o geólogo não precise classificar todos os poços e sim, apenas classifique um poço para que os poços vizinhos a este possam ser classificados automaticamente.

## Referências

- Aksoy, S. and Haralick, R. M. (2000). Effects of feature normalization on image retrieval. *IAPR International Conference on Pattern Recognition*, 22:563 – 582.
- Al-Anazi, A. and Gates, I. (2010a). On the capability of support vector machines to classify lithology from well logs. *Natural Resources Research*, 19:125 – 139.
- Al-Anazi, A. and Gates, I. (2010b). A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. *Engineering Geology*, 114(3-4):267 – 277.
- An-nan, J. and Lu, J. (2009). Studying the lithology identification method from well logs based on de-svm. In *Control and Decision Conference, 2009. CCDC '09. Chinese*, pages 2314 –2318.
- Cunha, E. S. (2002). Identificação de litofácies de poços de petróleo utilizando um método baseado em redes neurais artificiais. Master's thesis, Universidade Federal de Campina Grande.
- Flexa, R. T., Andrade, A., and Carrasquilla, A. (2004). Identificação de litotipos nos perfis de poço do campo de namorado (bacia de campos, brasil) e do lago de maracaibo (venezuela) aplicando estatística multivariada. *Revista Brasileira de Geociências*, 34:571-578.
- Garcia, E. N. (2008). *A PETROBRAS NA AMAZÔNIA: a história da pesquisa de combustíveis fósseis*. Norma Editora, Manaus, 1 edition.
- Kudo, M. and Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33(1):25 – 41.
- Petrobras (2011). Conheça nosso plano de negócios 2011-2015.
- Thomas, J. E., Triggia, A. A., and Correia, C. A. (2001). *Fundamentos de Engenharia de Petróleo*. Interciência.
- Wong, K., Ong, Y., Gedeon, T., and Fung, C. (2005). Reservoir characterization using support vector machines. In *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, volume 2, pages 354 –359.