



Copyright 2004, Instituto Brasileiro de Petróleo e Gás - IBP

Este Trabalho Técnico Científico foi preparado para apresentação no 3º Congresso Brasileiro de P&D em Petróleo e Gás, a ser realizado no período de 2 a 5 de outubro de 2005, em Salvador. Este Trabalho Técnico Científico foi selecionado e/ou revisado pela Comissão Científica, para apresentação no Evento. O conteúdo do Trabalho, como apresentado, não foi revisado pelo IBP. Os organizadores não irão traduzir ou corrigir os textos recebidos. O material conforme, apresentado, não necessariamente reflete as opiniões do Instituto Brasileiro de Petróleo e Gás, Sócios e Representantes. É de conhecimento e aprovação do(s) autor(es) que este Trabalho será publicado nos Anais do 3º Congresso Brasileiro de P&D em Petróleo e Gás

CLASSIFICAÇÃO DE CROMATOGRAMAS GASOSOS DE ÓLEOS BASEADA EM REDES NEURAIIS

Rosa Crithyna de Oliveira Vieira Paes¹, Ana Cristina da Silva Serra²

^{1, 2} Universidade Federal do Rio de Janeiro, LAB2M/COPPE/UFRJ - Bloco I-2000/Anexo,
Ilha do Fundão, Rio de Janeiro - RJ - Brasil - 21949-900, rosa.paes@lab2m.coppe.ufrj.br,
ana.serra@lab2m.coppe.ufrj.br

Resumo – Atualmente, devido ao fato de diversas bacias sedimentares serem alvo de estudos geoquímicos detalhados, a obtenção de uma ferramenta computacional que permita classificar diferentes tipos de óleos de maneira rápida e precisa possibilitaria uma agilização para caracterização de sistemas petrolíferos. O objetivo desse estudo consiste em aplicar a técnica de redes neurais e avaliar seu desempenho para classificação de diferentes tipos de amostras de óleo, de acordo com os perfis de distribuição dos *n*-alcanos em cromatogramas gasosos. O principal resultado deste trabalho foi a obtenção de um algoritmo que permite uma classificação de diferentes tipos de óleos com um alto nível de confiabilidade. Em virtude da grande quantidade de dados a ser analisada, procura-se tornar o processo o mais automatizado possível, com geração de resultados rápidos e precisos, cujos benefícios concentram-se na redução do tempo de análise do tipo de óleo.

Palavras-Chave: cromatogramas gasosos; óleo; classificação; redes neurais

Abstract – Nowadays, since several sedimentary basins have been extensively studied, using geochemical analysis, to find a computational tool that make possible to classify different types of oil fast and accurately would become more agile the characterization of petroleum systems. The purpose of this study consists of apply neural network techniques and evaluate its performance to classify different types of oil samples, according to the *n*-alkanes distribution profiles in gas chromatograms. The main result of this study was to get algorithm that provide a classification of oils with a high level of accuracy. Since there are a plenty of data to be analyzed, the aim is to become the process automated, to generate accurate results as fast as possible, that would reduce the duration of oil analyses.

Keywords: gas chromatograms, oil, classification, neural network

1. Introdução

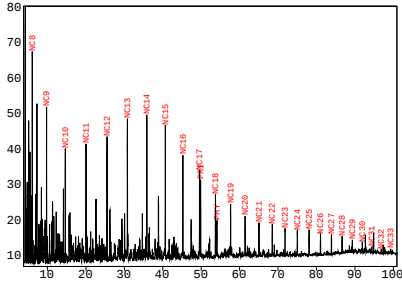
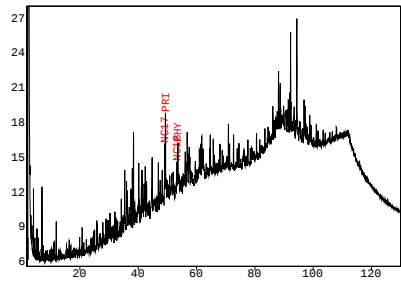
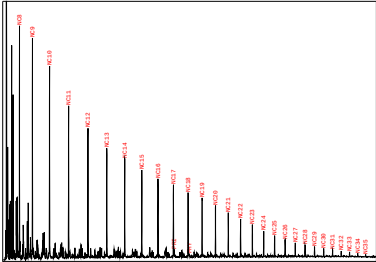
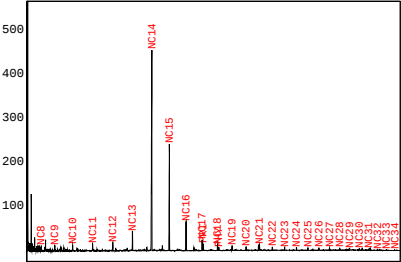
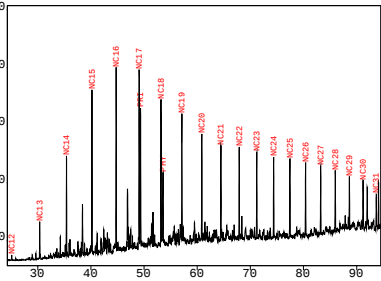
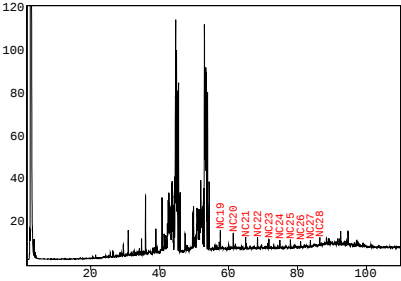
Nos últimos anos diversas bacias sedimentares tem sido alvo de estudos geoquímicos detalhados, tanto através de interpretação de teores, como mediante a integração de diferentes ferramentas prospectivas de óleos, a exemplo da sísmica 3D e do geoprocessamento de imagens de satélites. Não obstante a utilização destas ferramentas complementares, o desenvolvimento de técnicas que permitam a caracterização mais imediata e conclusiva de diferentes tipos de óleos, com ênfase nas peculiaridades geoquímicas é um anseio das áreas operacionais que necessitam de ferramentas computacionais para geração de resultados rápidos e precisos.

Há algum tempo a análise visual de cromatogramas tem revelado a existência de similaridades geométricas entre cromatogramas de óleos pertencentes a um mesmo tipo. A detecção de similaridades entre cromatogramas de diferentes campos de uma mesma bacia sedimentar, permitiria avaliar efeitos de compartimentalização de reservatórios, juntamente com informações geológicas da região.

Os perfis dos cromatogramas variam de acordo com a distribuição dos *n*-alcanos nestas amostras: Dentro desse conjunto de amostras existem 6 tipos de cromatogramas de petróleo com perfis bem definidos.

A Tabela 1 mostra os tipos de cromatogramas existentes no banco de dados a ser analisado.

Tabela 1- Tipos de cromatogramas presentes no banco de dados

Classe	Cromatograma Típico	Classe	Cromatograma Típico
1		4	
2		5	
3		6	

2. Objetivos

Este trabalho tem por objetivo aplicar a técnica de classificação por redes neurais e avaliar seu desempenho para um problema de classificação de diferentes tipos de amostras de petróleo, de acordo com os perfis de distribuição dos *n*-alcanos em cromatogramas gasosos.

3. Descrição do Método de Redes Neurais

Segundo Haykin, Fausset e Kovács, nas redes neurais artificiais, a idéia é realizar o processamento de informações tendo como princípio a organização de neurônios do cérebro. Assim, uma rede neural pode ser interpretada como um esquema de processamento capaz de armazenar conhecimento baseado em aprendizagem (experiência) e disponibilizar este conhecimento para a aplicação em questão.

Existem várias formas de se desenvolver uma rede neural. Ela deve ser montada de acordo com o problema a ser resolvido. Em sua arquitetura são determinados o número de camadas usadas (as camadas são formadas por neurônios), a quantidade de neurônios em cada camada, o tipo de sinapse utilizado, etc.

Resumidamente, uma rede neural consiste em um conjunto de unidades de entrada e saída, onde cada conexão tem um peso associado à ela. A Figura 1 indica a topologia da rede neural utilizada neste trabalho.

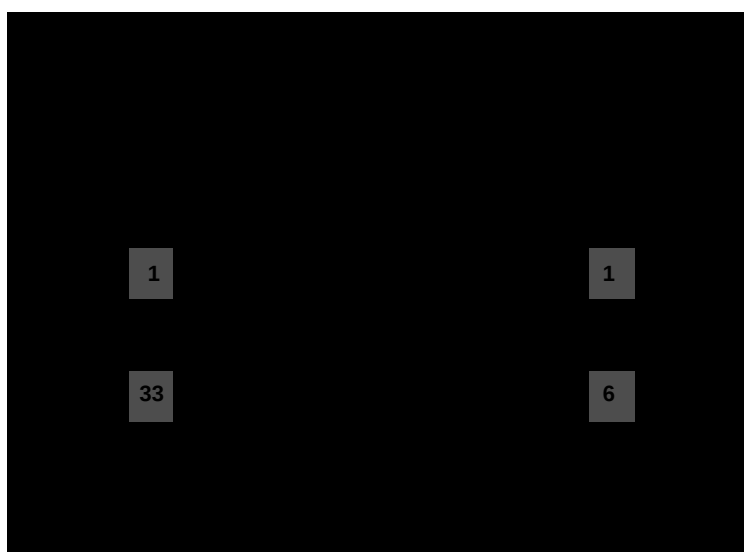


Figura 1. Configuração da rede neural - tamanho das camadas: entrada - 33, intermediária - 16, saída - 6

Durante a fase de aprendizagem (treinamento), enquanto a rede aprende, ela ajusta os pesos até ser capaz de prever a correta classe das amostras de entrada. Ou seja, depois da rede treinada, ela é utilizada para fazer previsões.

Existem, basicamente, 3 tipos de aprendizado nas redes neurais artificiais. Estaremos trabalhando com o aprendizado supervisionado. Neste tipo, a rede neural recebe um conjunto de entradas padronizadas e seus correspondentes padrões de saída, onde ocorrem ajustes nos pesos sinápticos até que o erro entre os padrões de saída gerados pela rede tenham um valor desejado.

4. Metodologia

O banco de dados consiste em uma sequência dos valores das alturas dos *n*-alcanos homólogos para cada amostra determinados em pico ampére (pA), obtidos como resultados da análise de cromatografia gasosa conforme descrito por Serra. O banco de dados original apresenta 33 variáveis como ponto de partida que consistem nos *n*-alcanos dos perfis cromatográficos. A princípio foram utilizadas 300 amostras contendo 6 tipos diferentes de distribuições para serem classificadas por redes neurais. Cada tipo de distribuição foi considerado uma classe, portanto as amostras foram divididas em seis classes, onde 77 amostras eram de Classe 1, 46 da Classe 2, 47 da Classe 3, 64 de Classe 4, 38 de Classe 5 e 68 de Classe 6.

Esses dados foram normalizados, dividindo-se os valores de cada variável pelo respectivo valor máximo de cada uma delas, ficando os valores num intervalo entre 0 e 1. Foram selecionadas aleatoriamente cerca de 70% dos dados para treinamento e 30% para teste.

Além deste banco de dados, outro banco foi utilizado na rede neural. Este segundo banco é derivado do original, onde amostras do conjunto de treino foram duplicadas aleatoriamente por classe até que todas as classes ficassem com a mesma população, ou seja, todas apresentassem um mesmo peso.

A rede neural foi implementada no *software Matlab*® versão 6.5, utilizando o método do *backpropagation*.

Para implementar a rede, diferentes topologias e algoritmos de treinamento foram testados até se chegar ao melhor resultado. Quanto ao critério de parada da rede, na maioria das vezes o treinamento parou devido ao erro no conjunto de validação ter começado a aumentar.

As matrizes de confusão tem por objetivo mostrar a quantidade de amostras classificadas corretamente para cada classe, portanto mede a eficiência. Através dela também pode ser verificado, no caso de percentual de classificação incorreta, com qual classe houve a confusão.

As matrizes foram calculadas para os conjuntos de treino e teste separadamente. Além disso, calculou-se a eficiência média da classificação do conjunto de teste, que consiste na soma da diagonal principal da matriz de confusão dividida pelo número de classes. Tanto as matrizes de confusão quanto as eficiências médias também foram calculadas no programa *Matlab*. Os bancos utilizados foram: original e treino duplicado como mencionado anteriormente.

5. Resultados

Diversos algoritmos de treinamento foram testados e utilizados para implementação da rede. Os principais foram: Gradiente descendente (GD); Gradiente descendente com momento (GDM); Gradiente descendente com taxa de aprendizado adaptativo (GDA); Gradiente descendente com momento e taxa de aprendizado adaptativo (GDA); *Resilient Propagation* (RP); Levenberg-Maquardt (LM).

Da análise dos diferentes algoritmos utilizados nos bancos de dados, observou-se que aquele que apresentou maior confiabilidade e acurácia no treinamento e teste foi o algoritmo *Resilient Backpropagation*, utilizando-se 16 neurônios na camada intermediária, funções de ativação do tipo tangente hiperbólica e função de treinamento rp. Neste caso, a convergência se deu em média por volta de 50 épocas.

As matrizes de confusão dos conjuntos de treino (MCTR) e de teste (MCTS), calculadas para o banco de dados original e modificado, podem ser vistas nas Tabela 2 e Tabela 3, respectivamente.

Tabela 2. MCTR e MCTS da base de dados original

MCTR	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
Classe 1	100	0	0	0	0	0
Classe 2	10,256	89,744	0	0	0	0
Classe 3	0	0	100	0	0	0
Classe 4	0	2,2727	0	95,455	2,2727	0
Classe 5	0	4	0	0	96	0
Classe 6	0	0	12,5	0	0	87,5

MCTS	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
Classe 1	95,652	4,3478	0	0	0	0
Classe 2	14,286	85,714	0	0	0	0
Classe 3	0	0	100	0	0	0
Classe 4	0	0	0	95	5	0
Classe 5	0	0	0	0	100	0
Classe 6	0	0	16,667	0	0	83,333

Tabela 3. MCTR e MCTS da base de dados modificada, onde houve duplicação de amostras do conjunto de treino

MCTR	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
Classe 1	100	0	0	0	0	0
Classe 2	7,547	92,453	0	0	0	0
Classe 3	0	7,547	92,453	0	0	0
Classe 4	0	0	1,8868	96,226	1,8868	0
Classe 5	0	3,7736	3,7736	0	92,453	0
Classe 6	0	0	11,321	0	0	88,679

MCTS	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
Classe 1	100	0	0	0	0	0
Classe 2	14,286	85,714	0	0	0	0
Classe 3	0	0	100	0	0	0
Classe 4	0	0	0	82,353	17,647	0
Classe 5	0	0	10	0	90	0
Classe 6	0	0	20	0	0	80

Obtivemos para eficiência média no treinamento um percentual de 94,78 % e 93,71 % e no teste 93,28 % e 89,678 % para as bases de dados original e modificada, respectivamente. O gráfico da Figura 2 mostra a variação do erro em função do número de épocas para os conjuntos de treinamento (em azul) e teste (verde).

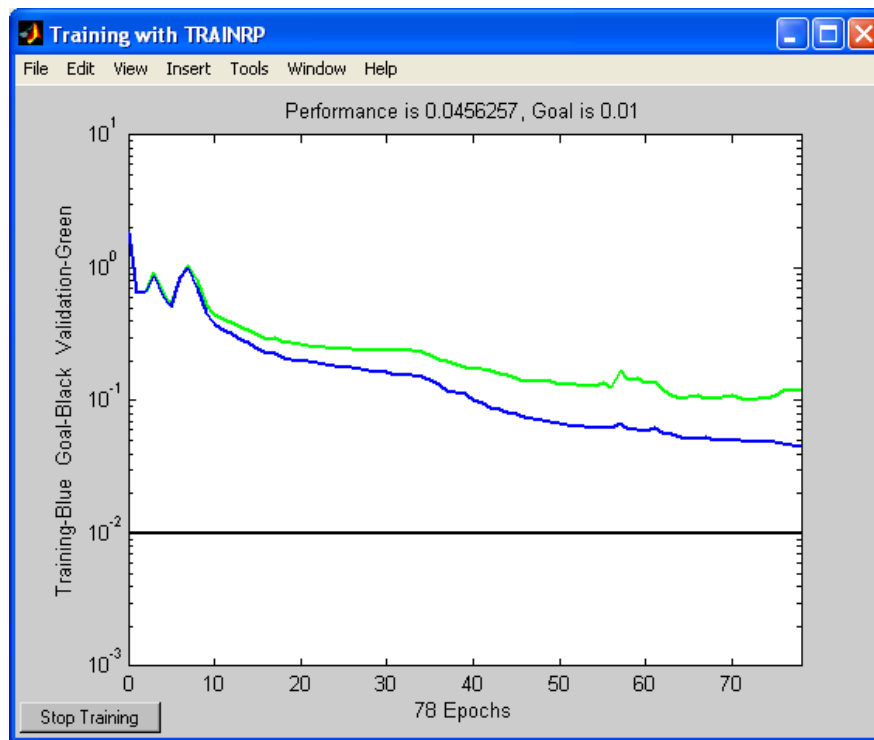


Figura 2. Variação do erro em função do número de épocas

6. Conclusão

Avaliar a eficiência de um classificador não consiste apenas em avaliar sua acurácia. Outros fatores como escalabilidade dos algoritmos, tempo de processamento, tolerância com dados ruidosos, facilidade para integrar, analisar e interpretar os dados também são fatores relevantes para considerar um classificador eficaz.

Neste trabalho, em todas as funções o tempo de processamento não foi problema, exceto no algoritmo **lm** (Levenberg-Marquardt *backpropagation*), já que centenas ou milhares de épocas podem ser necessárias antes que os pesos convirjam (o que pode tornar o treinamento da rede demasiadamente demorado).

A Classificação neural de diferentes tipos de amostras de óleo baseada em cromatogramas gasosos apresentou um resultado excelente num tempo de processamento pequeno. Neste caso, o algoritmo mais eficiente foi o 'rp' (*resilient backpropagation*), que apresentou uma eficiência média de 94,78 % e 93,28 % nos conjuntos de treino e de teste, respectivamente.

7. Agradecimentos

Ao Laboratório de Modelagem de Bacias Sedimentares (LAB2M), a Agência Nacional do Petróleo (ANP) e a Petrobrás (Cenpes) pelo suporte técnico e financeiro.

8. Referências

- HAYKIN, S., 2001, REDES NEURAIAS:Princípios e Prática, 2ª edição, Bookman.
- FAUSSET, L., 1984, Fundamentals of Neural Networks, Prentice Hall.
- KOVÁCS, Z., 1996, Redes Neurais Artificiais, Acadêmica.
- SERRA, A. C. S. 2003. A influência de aditivos de lamas de perfuração sobre as propriedades geoquímicas de óleos. Tese de M. Sc., COPPE/UFRJ, Rio de Janeiro, Brasil.