



Identificação de reservatórios utilizando Learning Vector Quantization

Michelle C. Kuroda, Alexandre Campana Vidal, Juliana Finoto Bueno, Ancilla M. Carvalho, Universidade Estadual de Campinas.

Copyright 2010, SBGf - Sociedade Brasileira de Geofísica

Este texto foi preparado para a apresentação no IV Simpósio Brasileiro de Geofísica, Brasília, 14 a 17 de novembro de 2010. Seu conteúdo foi revisado pelo Comitê Técnico do IV SimBGf, mas não necessariamente representa a opinião da SBGf ou de seus associados. É proibida a reprodução total ou parcial deste material para propósitos comerciais sem prévia autorização da SBGf.

Resumo

A caracterização de reservatórios é tarefa crucial para exploração de óleo e gás e requer a utilização de ferramentas que diminuam o custo de extração, riscos e incertezas. Com este objetivo o atual trabalho propõe o uso de um método matemático computacional de Redes Neurais, chamado Learning Vector Quantization (LVQ), utilizado para identificação de padrões em amostras supervisionadas. A ferramenta foi aplicada em uma base de dados que combina informações de perfis elétricos de cinco poços e dados de testemunho advindos do Campo de Namorado, Bacia de Campos –RJ, com os quais foram obtidos níveis de acerto de até 77,95% de acerto.

Introdução

Para a identificação das propriedades petrofísicas com o uso de ferramentas computacionais é importante considerar a sensibilidade do método à presença de ruído e interferências externas, comuns neste tipo de dado. Mediante tais características, o método escolhido para reconhecimento e classificação foi LVQ (Learning Vector Quantization), modelo matemático derivado do método SOM (Self-Organizing Maps), ambos elaborados por Kohonen (2001), que demonstram maior índice de sucesso que outros métodos de estatística clássica (Wong et al., 1995).

Os métodos aprendem ou treinam através de processo competitivo, no qual reconhecem padrões e correlacionam os dados de entrada, agrupando os similares.

A grande diferença entre os métodos mencionados é o tipo de amostra a que se destinam: SOM é aplicado a conjunto de dados não supervisionado e LVQ é utilizado em amostras supervisionadas (quando há informações sobre identificação de parte ou todo o conjunto de dados analisado). Dessa maneira é possível escolher previamente quais e quantas são as classes a serem identificadas. Em relação ao método SOM, LVQ tem a vantagem de ser um método de conversão rápida e de alta precisão (Fung et al., 1997).

Diferente dos outros métodos de decisão ótima para reconhecimento de padrões estatísticos baseados em corolários do Teorema da Probabilidade Total, como o Teorema de Bayes, LVQ apresenta estrutura de

aproximação baseada em uma filosofia totalmente diferente (Kohonen, 2001; Biehl et al., 2006).

A rede LVQ apresenta duas camadas, uma competitiva e outra linear, formadas por unidades, chamadas de neurônios. A camada competitiva aprende a classificar os dados de entrada, de maneira muito similar à camada competitiva do método SOM, na qual apenas uma unidade ganha o direito de agregar o vetor de entrada a si. A segunda camada linear transforma a camada competitiva em classes identificadas previamente, como mostra a Figura 1.

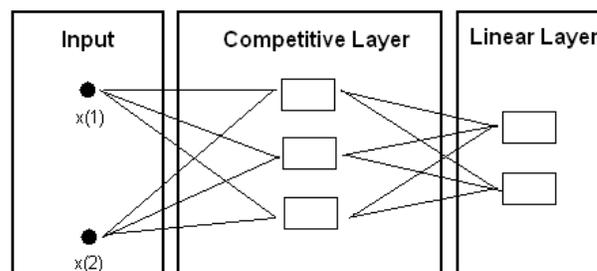


Figura 1 – Representação das camadas da estrutura LVQ.

O diferencial do funcionamento deste método é atribuído ao treinamento direcionado dos valores de entrada através das atribuições de classe, no início do processo. O agrupamento é baseado em critérios de maior similaridade, em geral representado pela distância Euclidiana entre os vetores de entrada, no qual quanto menor a distância entre eles maior a sua correlação e, provavelmente, serão definidos pela mesma classe e pertencerão ao mesmo grupo.

O algoritmo não faz aproximações por função de densidade, mas delimita fronteiras de vizinhanças entre vetores de referência chamados de *codebooks*, processados na primeira camada. Trata-se de vetores peso dos grupos formados pelas classes pré-estabelecidas, assumindo um vetor para cada classe com mesma dimensão do conjunto de entrada e com valores iniciais aleatórios, que se ajustam a cada iteração em direção ao vetor de entrada (Kohonen, 2001).

Além destas componentes, o algoritmo também permite a escolha da taxa de aprendizagem, que representa a taxa de mudança dos parâmetros a cada treinamento do programa (Kohonen, 2001). De acordo com as variações desta variável encontram-se respostas diferentes para o mesmo conjunto de dados e, por isso, tal *input* foi analisado para diferentes valores.

Metodologia/ Problema Investigado

O método descrito foi utilizado em um conjunto de dados de perfis elétricos (com 4733 amostras), referentes a 5 poços (NA1A, NA2, NA4, NA7 e RJS42) e seus dados de testemunho, advindos do Campo de Namorado, Bacia de Campos, disponibilizados pela Agência Nacional de Petróleo. Dentre os perfis foram analisados: raios gama (GR), sônico (DT), densidade (RHOB) e porosidade neutrônica (NPHI).

A base de dados analisada apresenta 29 litologias distintas e, baseada nestas informações, as classificações escolhidas previamente foram: reservatório, possível reservatório e não-reservatório, definidas através dos dados de testemunho (Tabela1).

Tabela 1: Classificação utilizada para treinamento de LVQ.

Fácies	Descrição	Classificação
1	Interlaminado Lamoso Deformado	Não reservatório
2	Conglomerados e Brechas Carbonáticas	Não reservatório
3	Diamictito Arenoso Lamoso	Não reservatório
4	Conglomerados Residuais	Possível reservatório
6	Arenito Grosso, Amalgamado	Reservatório
7	Arenito Médio Fino Laminado	Possível reservatório
8	Arenito Médio Gradado ou Maciço	Reservatório
9	Arenito Médio Cimentado	Não reservatório
10	Arenito/Folhelho Interestratificado	Possível reservatório
11	Arenito/Folhelho Finamente Interestratificado	Possível reservatório
12	Siltito Argiloso Estratificado	Não reservatório
13	Interlaminado Siltito Argiloso e Marga	Não reservatório
14	Folhelho Radioativo	Não reservatório
15	Interlaminado Arenoso Bioturbado	Não reservatório
16	Interlaminado de Siltito e Folhelho, Deformado, Bioturbado	Não reservatório
17	Marga Bioturbada	Não reservatório
18	Ritmito	Não reservatório
19	Arenito Glauconítico	Não reservatório
20	Folhelho Siltico com Níveis de Marga Bioturbada	Não reservatório
21	Arenito Cimentado, com Feições de Escorregamento	Não reservatório
22	Siltito Argiloso/Arenito Deformado	Não reservatório
23	Arenito Médio/Fino Laminado Cimentado	Não reservatório
24	Interestratificado Siltito/Folhelho Intensamente Bioturbados	Não reservatório
25	Marga Bioturbada Outra	Não reservatório
26	Folhelho Carbonoso	Não reservatório
27	Arenito Maciço Muito Fino	Não reservatório
28	Siltito Arenó-Argiloso	Não reservatório
29	Interlaminado Siltito/Folhelho	Não reservatório

Das três possíveis opções de LVQ (LVQ1, LVQ2 e LVQ3), o treinamento do conjunto de dados foi realizado com LVQ1, pois apesar dos métodos apresentarem precisão de reconhecimento de padrões estatístico similar, LVQ1 tem processo mais robusto, em que os vetores de referência assumem valores fixos, mesmo após longos períodos de aprendizagem e Kohonen (2001) sugere que o treinamento seja inicializado com ele. No algoritmo LVQ2 dois vetores de referências mais próximos do vetor de entrada são atualizados simultaneamente e esta distância é otimizada sem garantia de preservação de topologia, corrigida no método LVQ3 (Kohonen, 2001; Hammer & Villmann, 2002).

Para ajuste do algoritmo os parâmetros escolhidos foram determinados de acordo com as considerações de Kohonen (2001), dessa forma o treinamento foi realizado com 150 e 500 iterações e taxa de treinamento igual a 0.01, 0.025 e 0.001.

O conjunto de dados foi separado em supervisionado (ou seja, nos intervalos com testemunho, cerca de 40% do total amostrado) e não supervisionado.

Para treinamento do algoritmo dois métodos foram analisados. No primeiro caso, o conjunto de dados com testemunho foi utilizado para treinamento e a base de dados total foi identificada posteriormente, mediante resultado de classificação do outro conjunto (Validação Cruzada 1).

No segundo caso os dados testemunhados foram separados em pares e ímpares. Os pares foram utilizados para treinamento do algoritmo e os outros adicionados ao restante do conjunto e validados em segunda etapa (Validação Cruzada 2).

Primeiramente os dados foram normalizados, porém verificou-se que a rede convergia mais facilmente quando não havia tratamento de dados de normalização, como também foi constatado por Souza et al. (1998) e, por isso, não foi utilizada nesse processo.

Resultados Preliminares e Discussões

Para cada um dos critérios de validação cruzada escolhido, foram criados três cenários, diferentes segundo o taxa de aprendizagem: 0.1, 0.025 e 0.01. Dos quais foram obtidos os resultados apresentados na Tabela 2.

Tabela 2: Média de acerto de LVQ para diferentes taxas de treinamento.

Média de Acertos					
Validação cruzada 1			Validação cruzada 2		
Raio de Treinamento	150 iterações	500 iterações	Raio de Treinamento	150 iterações	500 iterações
0.1	0.7544	0.7451	0.1	0.7795	0.7087
0.025	0.6821	0.74	0.025	0.752	0.7303
0.01	0.6821	0.741	0.01	0.7421	0.7461

Com a aplicação do método descrito, foi possível identificar se os dados analisados eram referentes a reservatórios, possíveis ou não-reservatórios, com melhor índice de acerto correspondente a 77,95% e erro quadrado médio de 21%, referente ao treinamento dos dados com raio de treinamento igual a 0.1, 150 iterações e treinamento de testemunho com índice par e validação de ímpares.

Tais resultados são referentes à técnica de validação cruzada, que combina dados de treinamento com valores de testemunho para avaliar a robustez do modelo e dessa forma identificar dentre todos os cenários construídos o que melhor define a simulação mediante base estatística.

Vale ressaltar que como a inicialização dos vetores de referência é aleatória, os resultados são ligeiramente diferentes a cada treinamento, ainda que apresente mesmas variáveis e atributos de entrada.

A classificação de perfis de rochas não reservatórios, mediante cenário descrito, demonstrou desempenho de 91% de acerto, enquanto a classificação entre litologias de reservatórios e possíveis reservatórios apresentaram média de acerto de 60% devido à estreita semelhança de valores (Tabela 3).

Tabela 3: Análise estatística dos grupos testemunhados

Não Reservatório		RHOB	DT	NPHI	GR
	min	1,7556	52,5322	0,6355	21,1875
max	2,7333	111,8711	37,7302	108,4648	
mediana	2,4722	82,67045	20,3164	53,99759	
média	2,455045	82,54976	19,95449	57,23727	
desvio padrão	0,115688	8,391241	4,783761	15,36022	
Possível Reservatório		RHOB	DT	NPHI	GR
	min	1,7336	67,453	12,0859	46,974
max	2,5796	108,188	28,8765	109,6797	
mediana	2,2958	91,4531	22,949	62,2734	
média	2,321208	89,7255	22,5273	68,05252	
desvio padrão	0,952291	35,05213	13,16057	33,19265	
Reservatório		RHOB	DT	NPHI	GR
	min	2,095	58,0625	6,125	32,0586
max	2,5637	103,922	29,707	97,6484	
mediana	2,249847	89,9922	23,10345	53,4882	
média	2,262842	88,4777	22,77073	54,15359	
desvio padrão	0,79	31,01	10,47	30,43	

Os valores de desvio padrão da Tabela 3 são baixos, demonstrando que os grupos estão bem concentrados, entretanto observa-se que os valores médios entre os

grupos de possíveis reservatórios e reservatórios são muito próximos, devido ao grande conjunto de litologias que englobam cada uma destas classificações.

Essa característica torna a classificação confusa, uma vez que não há definições precisas que diferenciem tais grupos.

Através da Figura 2 é possível observar que a classificação realizada mediante LVQ foi muito próxima da interpretação visual dos perfis, indicando o bom desempenho do método.

Apesar da alta correlação entre os dados, o método obteve resultados satisfatórios e pode ser utilizado como uma alternativa para identificação de reservatórios, em detrimento de outras técnicas de estatística clássica.

Conclusões

O crescente interesse na área de caracterização de reservatórios de petróleo tem motivado diversas pesquisas a respeito de como detectar, de forma eficiente, a distribuição das propriedades litológicas e petrofísicas do reservatório. O desafio maior desse processo é a obtenção de dados que confirmem os dados de testemunho, presente em alguns casos. O algoritmo LVQ utilizado neste trabalho mostrou ser um caminho promissor a ser explorado para diagnóstico de identificação de reservatórios a partir de dados de perfis. A Rede Neural proposta conseguiu detectar padrões de *clusters*, com taxa de acerto de até 77,95%. Entretanto será necessário, para trabalhos futuros, avaliar a necessidade de ajustes da operação do algoritmo para obtenção de maiores taxas de acerto.

Para o trabalho atual, a identificação de reservatórios baseada no método LVQ proposta resultou na classificação satisfatória das classes escolhidas, portanto pode ser utilizada como um método alternativo em adição aos métodos tradicionais utilizados, com a vantagem de ser rápido e preciso.

Agradecimentos

Os autores agradecem à Agência Nacional de Petróleo - ANP e à PETROBRAS pela disponibilidade de dados do Campo de Namorado.

Referências

Biehl, M.; Hammer, B.; Schneider, P. 2006. Matrix Learning in Learning Vector Quantization, Technical Report, Institute of Informatics, Clausthal University of Technology.

Fung, C. C.; Wong, K. W.; Eren, H. 1997. Modular Artificial Neural Network for Prediction of Petrophysical Properties From Well Log Data. IEEE Transactions on Instrumentation and Measurement, 46:1295-1299.

Kohonen, T., 2001. *Self-Organizing Maps*. Springer Verlag, New York, 501pp.

Hammer, B. & Vilmann. 2002. Generalized Relevance Learning Vector Quantization, *Neural Network*, 15:1059-1068.

Souza, O. M. de; Pezzi, S.; Bastos L.; Zouza, M. S. S., 1998. Avaliação de Desempenho de Redes LVQ em Comparação com Redes MLP com Backpropagation na Classificação de Cromossomos. Associação Brasileira de Engenharia de Produção. Disponível em: <<http://www.abepro.org.br>>. Acesso em: Junho de 2010.

Wong, P. M.; Jian, F. X.; Taggart, I. J. 1995. A critical comparison of neural networks and discriminant analysis in lithofacies, porosity and permeability predictions. *J. Petroleum Geol.*, 18:191-206.

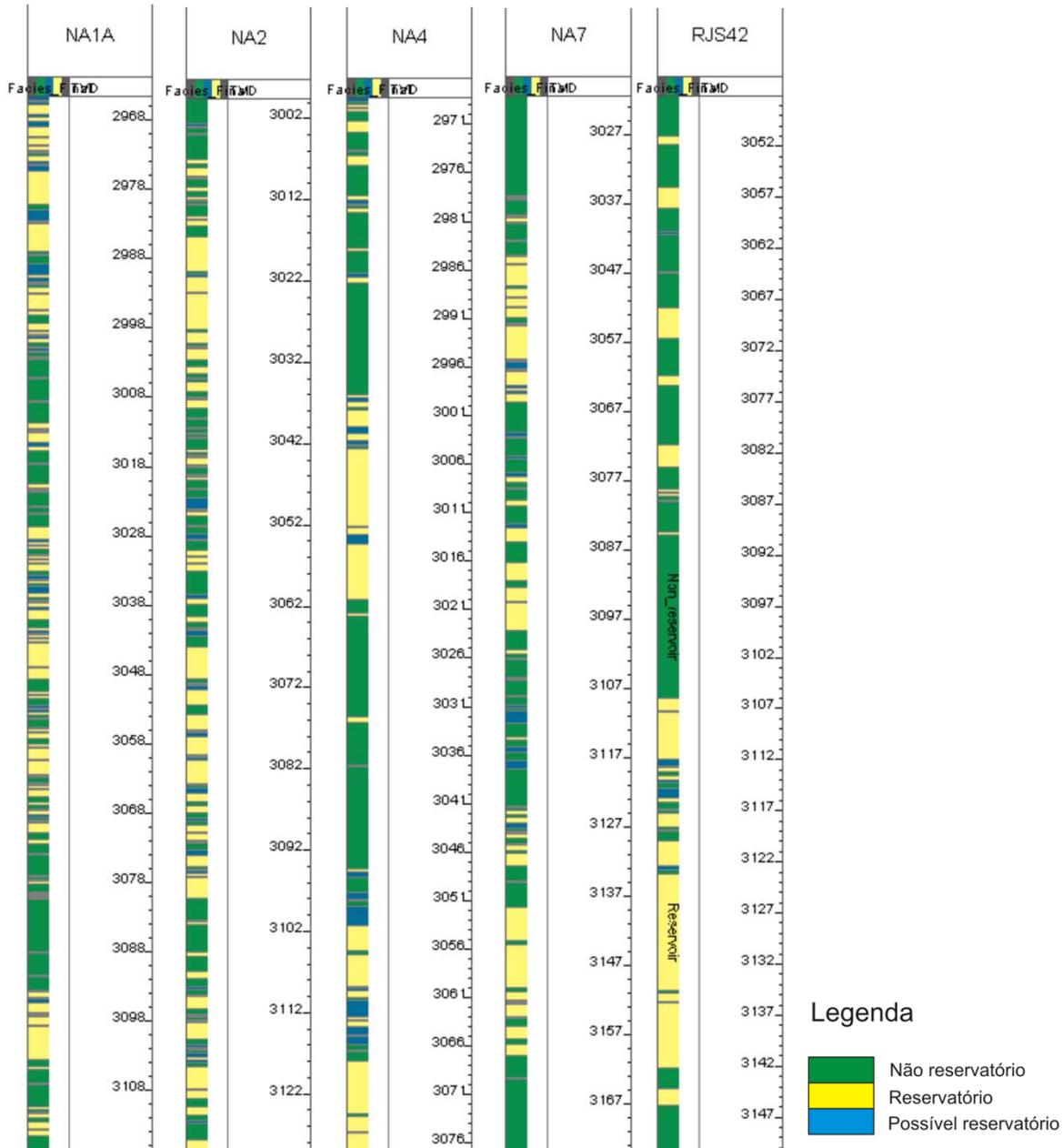


Figura 2 – Imagem ilustrativa da classificação de rochas reservatórios, possíveis reservatórios e não reservatórios, mediante resultados obtidos através do método LVQ, dos poços NA1A, NA2, NA4, NA7 e RJS42.