



AVALIAÇÃO DE PREÇO DE POSTOS DE COMBUSTÍVEIS COM REGRESSÃO LINEAR EM SITUAÇÕES DE PRAZO REDUZIDO (URGENTES) E COMPARAÇÃO DE RESULTADOS COM REDE NEURAL

Paulo César Vidal Morato¹, Jackson Carvalho Guedes²

Copyright 2006, Instituto Brasileiro de Petróleo e Gás - IBP

Este Trabalho Técnico foi preparado para apresentação na *Rio Oil & Gas Expo and Conference 2006*, realizada no período de 11 a 14 de setembro de 2006, no Rio de Janeiro. Este Trabalho Técnico foi selecionado para apresentação pelo Comitê Técnico do evento, seguindo as informações contidas na sinopse submetida pelo(s) autor(es). O conteúdo do Trabalho Técnico, 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, seus Associados e Representantes. É de conhecimento e aprovação do(s) autor(es) que este Trabalho Técnico seja publicado nos Anais da *Rio Oil & Gas Expo and Conference 2006*.

Resumo

Mostra o emprego da análise de regressão e de redes neurais na avaliação de postos de combustíveis em situações urgentes, nas quais necessita-se obter um laudo rápido e preciso mas não se dispõe de tempo suficiente para o desenvolvimento de uma avaliação normal. Etapas:

- Levantamento de laudos anteriores feitos em situações de prazo normal.
- construção do modelo de regressão.
- construção do modelo de rede neural.
- Comparação regressão x rede neural.

Conclusões:

- Permite a obtenção de laudos com boa precisão e curto espaço de tempo, típico de situações urgentes.
- Pode ser aplicada para avaliações patrimoniais ou econômicas.
- Para avaliações patrimoniais, o modelo foi sensível à: valor do terreno, área construída, quantidade de equipamentos e loja de conveniência.
- Na análise por redes neurais foram usados os mesmos dados que alimentaram a regressão, empregando-se o programa EasyNN.
- Neste trabalho o desempenho da rede neural foi superior ao da análise de regressão linear, resultando em menores desvios e na constatação que a distribuição dos resíduos também é normal, possibilitando a construção de intervalos de confiança com facilidade.

Abstract

It shows the job of the analysis of regression and neural nets in the evaluation of fuel ranks in urgent situations, in which it is needed to get a necessary finding fast but it is not made use of enough time for the development of a normal evaluation. Stages:

- Survey of made previous findings in situations of normal stated period.
- construction of the regression model.
- construction of the model of neural net.
- Comparison regression x neural net.

Conclusions:

- It allows to the attainment of findings with good precision and short space of time, typical of urgent situations.
- Can be applied for patrimonial or economic evaluations.
- For patrimonial evaluations, the model was sensible to: value of the land, constructed area, amount of equipment and store of convenience.
- In the analysis for neural nets had been used the same data that had fed the regression, using itself the EasyNN program.
- In this work the performance of the neural net was superior to the one of the analysis of linear regression, resulting in lesser shunting lines and that the distribution of the residues also is normal, making possible the

¹ Engenheiro Mecânico – Petrobras S.A.

² Mestre, Engenheiro Civil – Petrobras S.A.

reliable construction of intervals with easiness.

1- Análise de Regressão Linear.

1.1- Introdução.

No caso da avaliação de postos de combustíveis, a Petrobras tem catalogado mais de 600 laudos de avaliação para compra/venda em um período de 20 anos. Tais trabalhos se desenvolveram em condições normais de prazo, consumindo aproximadamente entre 5 à 10 dias úteis cada posto. Porém, existem situações em que tais prazos normais não podem ser respeitados, e o avaliador é obrigado a apresentar um trabalho rapidamente. Tais situações seriam: estudos de viabilidade, troca de ativos, atualização contábil, garantia dada pelo dono do posto para compra de combustíveis. Através do emprego da análise de regressão linear, o presente trabalho tem por objetivo oferecer uma opção melhor para estas situações de emergência nas quais não se dispõe de tempo suficiente. A figura 1 indica um posto típico de estrada e a figura 2 um posto típico de cidade.



Figura 1: posto típico de estrada.



Figura 2: posto típico de cidade.

1.2- Etapas do Uso de Regressão: O procedimento proposto é composto pelas seguintes etapas:

1.2.1- Pesquisa de Mercado.

No caso da Petrobras, levantamos diversos trabalhos de avaliação de postos desde 1999 em todo o Brasil (sem restrição de local).

1.2.2- Construção do Modelo Inferencial: com os dados da amostra, parte-se para a construção do modelo inferido.

1.2.3- Aplicação do Modelo ao Posto que se quer Avaliar:

1.3- Aplicação Prática.

Para exemplificar o emprego do uso de regressão, relatamos os resultados do primeiro modelo feito na Petrobras para avaliação patrimonial de postos, em situação urgente, em qualquer local do Brasil. Para a modelagem estatística levantamos diversos trabalhos de avaliação normal (com vistoria) de postos realizados pela Petrobras desde 1999 em todo o Brasil (sem restrição de local). Desta forma obtivemos uma amostra total com 37 postos, sendo levantadas as seguintes variáveis independentes: valor do terreno(em dólar), área construída do posto(m²), idade do posto(anos), quantidade de equipamentos, existência ou não de loja de conveniência, galonagem de combustíveis vendidos mensalmente, receita bruta do posto exceto combustíveis(em dólar), tipo de imagem/lay-out do posto (antiga, média, moderna), tipo de posto(cidade ou estrada). Ao final, foram selecionadas quatro delas, as quais explicaram, razoavelmente, a formação dos preços: valor do terreno, área construída, quantidade de equipamentos, existência ou não de loja de conveniência. As figuras 3, 4, 5 e 6 tentam detalhar melhor essas quatro variáveis.



Figura 3: variável valor do terreno



Figura 4: variável área construída (prédios e coberturas)



Figura 5: variável quantidade de equipamentos.



Figura 6: variável loja de conveniência.

A equação inferida selecionada apresentou a seguinte forma:

$$Vp = A + B \times Vt + C \times Ac + D \times Equip + E \times @Conv, \text{ onde}$$

Vp = valor patrimonial do posto, em US\$ (dólares americanos).

Vt = valor do terreno, em US\$ (dólares americanos).

Ac = área construída do posto (apenas prédios e cobertura de bombas), em m^2 .

$Equip$ = quantidade de equipamentos fixos no posto, em unidades.

$@Conv$ = dicotômica que assume valor 0 para postos sem loja e valor 1 para postos com loja.

A, B, C, D, E = valores numéricos.

Tal equação responde por 89,40% da formação do valor (r^2 ou coeficiente de determinação), e através da Validação Cruzada testou-se a mesma com dados da pesquisa que não foram utilizados, comprovando-se a sua consistência. Os resultados de alguns postos estão presentes na Tabela 2:

Tabela 2: resultados da validação cruzada para alguns postos.

VALIDAÇÃO CRUZADA					
Município	Valor do Laudo	Tipo de Posto	Limite Inferior	Valor Central	Limite Superior
	(US\$)				
Pelotas/RS	314684	Estrada	228409	283728	343744
Rio de Janeiro/RJ	1455144	Cidade	1388190	1545015	1708484
Barueri/SP	167650	Cidade	154000	181000	209000
Florianópolis/SC	767000	Cidade	664000	725000	788000
Taguatinga/DF	308000	Cidade	269000	302000	336000
Vitória da Conquista/BA	347000	Estrada	309000	349000	391000
Duque de Caxias/RJ	1415000	Estrada	1221000	1355000	1496000
São Borja/RS	231000	Cidade	179000	219000	262000
Porto Alegre/RS	237000	Cidade	177000	208000	240000
Belo Horizonte/MG	579000	Cidade	522000	573000	626000
Belo Horizonte/MG	275000	Cidade	260000	292000	326000
Porto Velho/RO	149000	Cidade	146000	173000	203000
Curitiba/PR	517000	Cidade	472000	520000	570000
Curitiba/PR	336000	Cidade	341000	377000	415000

A figura 7 mostra um resumo do software que modelou a regressão (Programa Regre).

REGRESSAO MULTIPLA				

MODELO :				
$Vp = \{ A + B * Vt + C * Ac + D * Equip + E * @conv \} ^{1.666667}$				

ERRO PADRAO DA REGRESSAO = 333.5148 C.V. = .1567715				
COEFICIENTES : DETERMINACAO = .8940283 CORRELACAO = .9455307				
R2 AJUSTADO = .8807818				
TESTES DE HIPOTHESES				

VARIAVEL	REGRESSORES	ERRO PADRAO	T OBSERVADO	VALOR-PROB.
Vt	B1= 3.55973E-03	2.191762E-04	16.24141	5.073868E-09
Ac	B2= 4.836734E-02	1.784179E-02	2.710902	5.188489E-03
equip	B3= 29.0595	9.410927	3.087846	2.17959E-03
@conv	B4= 497.3874	119.4819	4.162868	2.119999E-04

RELACOES ENTRE AS VARIAVEIS				

VARIAVEIS	CORRELACAO	T OBSERVADO	VALOR-PROB.	
Vp^6xVt	.8712401	10.50066	6.819301E-08	
Vp^6xAc	-2.256929E-02	-.1335557	.4472595	
Vp^6xequip	9.058848E-02	.5381413	.2969432	
Vp^6x@conv	.1302904	.7774354	.2210633	

Figura 7: Cálculos do Programa Regre.

2- Análise Por Redes Neurais.

Pode-se treinar as redes neurais apresentando a elas fatos, isto é, pares de dados de entrada e saída, permitindo que elas façam associações, descobrindo assim a existência de algum padrão de comportamento. O programa utilizado nesta aplicação, chamado EasyNN, também emprega um tipo específico de rede neural, chamada de rede de retro-propagação ou de encadeamento para trás; a rede é treinada apresentando-se a ela um conjunto de fatos (entradas/saídas) repetidas vezes. Cada vez que os dados de entrada são apresentados ela retorna uma resposta com resultados que ela pensa ser o correto, comparando-a com o fato real ou padrão. Quando sua resposta é incorreta ela se corrige internamente. Após percorrer toda a lista de fatos, o programa revê todo o rol recursivamente, até que todas as respostas sejam consideradas aceitáveis. Durante o treinamento pode-se acompanhar graficamente o aprendizado da rede neural e identificando quando os erros se tornam menor do que aquele estabelecido como limite superior, conforme se vê na figura 8:

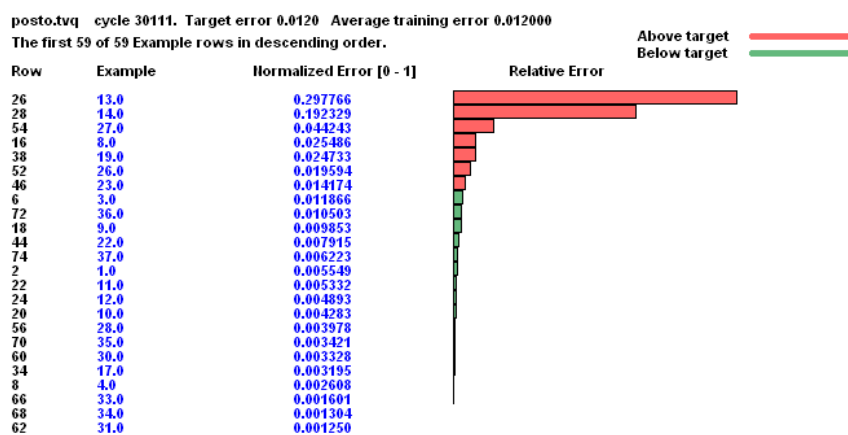


Figura 8: evolução dos acertos e erros.

Em sua forma mais simples uma rede neural consiste de três camadas: uma camada de dados de entrada, uma camada oculta e uma camada de dados de saída, conforme representado na figura 9:

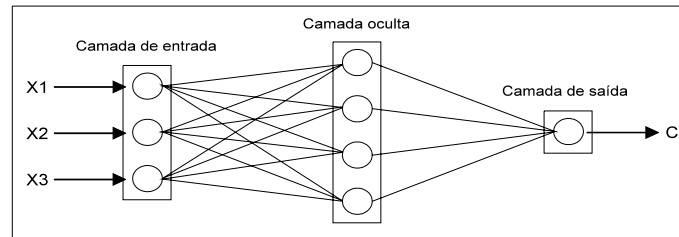


Figura 9: rede neural simples.

A rede neural utilizada nesse exercício, com uma camada de entrada, uma camada oculta e uma camada de saída, é a apresentada na figura 10:

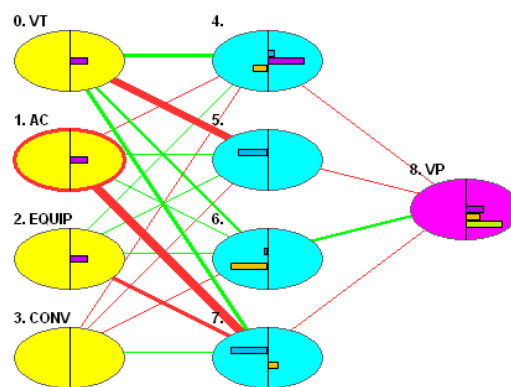


Figura 10: rede neural utilizada neste trabalho.

O programa EasyNN® mostra graficamente a influência dos atributos, ou variáveis independentes, conforme se vê na figura 11:

posto.tvq 30111 cycles. Target error 0.0120 Average training error 0.012000
The first 4 of 4 Inputs in descending order.

Column	Input Name	Importance	Relative Importance
1	VT	122.8473	
2	AC	80.2679	
3	EQUIP	51.5385	
4	CONV	2.9539	

Figura 11: influência das variáveis independentes na formação do preço do posto.

Em alguns casos, dependendo da complexidade do modelo, é mais fácil analisar as influências dos atributos, expressos pelas variáveis independentes, fazendo-se simulações, provocando variações de valores do atributo em estudo, mantendo-se as demais variáveis constantes. As redes neurais, porém, não expressam sua função ou funções de transferência de maneira simples para que o usuário possa compreender e quantificar de imediato as influências dos atributos do bem avaliado. Isto não significa que as equações não possam ser explicitadas, mas apenas são mais complexas. Frequentemente as funções de transferência das redes neurais são matematicamente bastante sofisticadas. O programa BrainMaker®, por exemplo, usa na maioria das vezes uma função de transferência sigmóide, podendo entretanto, à vontade do projetista da rede, empregar funções lineares, em degrau, gaussiana, etc. A melhor e mais acessível maneira de se analisar o desempenho de uma rede neural é por meio de simulações dos resultados. Evitaremos neste trabalho descrever a matemática utilizada no treinamento de uma rede neural.

3- Conclusões.

3.1-Quanto à Análise de Regressão:

3.1.1- Esta metodologia permite a obtenção de laudos de avaliação de postos com bom grau de precisão e em curto espaço de tempo, típico de situações urgentes: estudos de viabilidade, troca de ativos, atualização contábil, garantia dada pelo dono do posto para compra de combustíveis.

3.2-Quanto à Análise por Redes Neurais (Comparação de Resultados):

Para comparar as duas técnicas foram utilizados os dados apresentados por Vidal, referentes a 37 postos, e os resultados da análise de regressão inferida anteriormente. O erro quadrático médio (EQM) foi calculado para os 37 postos usados na equação obtida pelo REGRE, conforme equação 1:

$$EQM = \frac{\sum (Y_i - Y_{real})^2}{N} = 54.270.353.143,84 \quad (1)$$

onde:

Y_i = valor do posto i calculado pela equação inferida,

Y_{real} = valor do posto i, na pesquisa,

N = número total de dados da pesquisa.

Rede neural

As mesmas informações, sem modificações na forma de entrada dos dados, foram usadas para avaliação, empregando-se o programa EasyNN-Plus®, programado para a criação de redes neurais, e calculou-se, também, o erro quadrático médio:

$$\text{Com a rede neural: } EQM = \frac{\sum (Y_i - Y_{real})^2}{N} = 32.089.411.786,69 \quad (1)$$

$$\text{Com a análise de regressão: } EQM = \frac{\sum (Y_i - Y_{real})^2}{N} = 54.270.353.143,84 \quad (1)$$

Conclusão: o desempenho da rede neural treinada foi superior ao da análise de regressão linear, pois apresentou um EQM menor. No caso em estudo os resultados confirmaram com um caso prático o que a teoria afirmava a priori: que as predições via redes neurais deveriam ser melhores daquelas obtidas com a análise de regressão linear utilizando mínimos quadrados ordinários ou modelos lineares generalizados. No caso em estudo, o erro quadrático médio obtido via regressão linear foi maior cerca de 69% do que aquele gerado pela rede neural treinada com os mesmos dados. De fato, em nossa experiência, na maioria das vezes o desempenho da rede neural tem sido superior ao da análise de regressão. O melhor desempenho da rede neural deve-se, fundamentalmente, ao fato de os fenômenos sociais e relações do mundo real, expressos por variáveis, não serem necessariamente lineares. Mesmo quando lineariza-se a função, transformando-se as variáveis para melhor captar essa relação não retilínea, continua havendo a possibilidade de existir um melhor estimador não linear. Neste trabalho partiu-se do pressuposto que o melhor modelo é aquele que apresenta os menores desvios em relação aos verdadeiros valores pesquisados. No caso em estudo, a distribuição dos resíduos é aderente à distribuição normal; o que pode ser comprovado pelo teste de aderência não paramétricos, como o Kolmogorov-Sminov. Tal fato facilita o cálculo dos intervalos de confiança para cada uma das estimativas feitas utilizando a rede neural treinada e a Comparação de Resultados está presente na Tabela 4.

Tabela 4: Comparação de Resultados Regressão x Rede Neural (Valores em preto: regressão)

COMPARAÇÃO DE RESULTADOS REGRESSÃO X REDE					
Município	Valor do Laudo	Tipo de Posto	Limite Inferior	Valor Central	Limite Superior
	(US\$)		(US\$)		
Pelotas/RS	314684	Estrada	228409 221065	283728 251742	343744 284612

Rio de Janeiro/RJ	1455144	Cidade	1388190 1357820	1545015 1449795	1708484 1544483
Barueri/SP	167650	Cidade	154000 154844	181000 170765	209000 187072
Florianópolis/SC	767000	Cidade	664000 748118	725000 789643	788000 831999
Taguatinga/DF	308000	Cidade	269000 270258	302000 290069	336000 310228
Vitória da Conquista/BA	347000	Estrada	309000 239607	349000 258096	391000 277269
Duque de Caxias/RJ	1415000	Estrada	1221000 1272206	1355000 1356020	1496000 1443123
São Borja/RS	231000	Cidade	179000 183655	219000 207322	262000 232450
Porto Alegre/RS	237000	Cidade	177000 299651	208000 330430	240000 361810
Belo Horizonte/MG	579000	Cidade	522000 528493	573000 559624	626000 591576
Belo Horizonte/MG	275000	Cidade	260000 208370	292000 223691	326000 239769
Porto Velho/RO	149000	Cidade	146000 130008	173000 144060	203000 159481
Curitiba/PR	517000	Cidade	472000 417616	520000 443184	570000 469489
Curitiba/PR	336000	Cidade	341000 340629	377000 362249	415000 384788

Referências

- BARBOSA, E.P. e BIDURIN, C.P. – Seleção de Modelos de Regressão para Predição via Validação Cruzada: Uma Aplicação em Avaliação de Imóveis – Revista Brasileira de Estatística, 52 (N°s 197/198) pág. 105-120 – Jan/Dez-1991.
- DANTAS, R. A. – Engenharia de Avaliações: Uma Introdução à Metodologia Científica – Pini, 1998.
- DRANG, D. E. , EDELSON, B. , LEVINE, R. I. - Inteligência Artificial e Sistemas Especialistas, tradução de Maria Claudia Santos Ribeiro, São Paulo, McGraw-Hill, 1988.
- GARZA, J. M., ROUHANA, K., - Neural Networks Versus Parameter-Based Applications in Cost Estimating, Cost Engineering, vol. 37, fevereiro 1995, p. 15-17.
- GUEDES, J. C. - Avaliação de Bens Utilizando Metodologia Científica - Tese de Mestrado - COPPE/UFRJ, Rio de Janeiro, 1992.
- HARMAN, P. - Expert Systems tools and applications, New York, John Willey, 1988.
- HOFFMAN, R. , VIEIRA, S. - Análise de Regressão: Uma Introdução à Econometria, São Paulo, Hucitec, 1977.
- LAWRENCE, J. - Introduction to Neural Networks – Design, Theory, and Applications – California Scientific Software Press, 6th edition, 1994.
- MOREIRA, A. L. – Princípios da Engenharia de Avaliações – Pini, 1994.
- NELDER, I.A. and WEDDERBURN, R.W.M., Generalized Linear Models. JRSS A 135. P. 370-384.
- NETER, J. , WASSERMAN, W. and KUTNER, M. H. - Applied Linear Statistical Models, Boston, Irwin, 1990.
- NORMA NB-901 – Avaliações de Máquinas, Equipamentos, Instalações e Complexos Industriais – ABNT (Associação Brasileira de Normas Técnicas), 1984.
- NORMA NB-502 – Avaliações de Imóveis Urbanos - ABNT (Associação Brasileira de Normas Técnicas), 1989.