

6

Aplicações

Nesse capítulo o método RCRI será utilizado para resolver dois problemas diferentes. O primeiro é na área de atuária, apresentado na seção 6.1, e consiste em melhorar a estimativa da reserva de IBNR. O segundo, apresentado na seção 6.2, é uma aplicação em geologia que busca estimar dados faltantes dentro da amostra de um poço. Dessa forma é possível avaliar o desempenho do método em aplicações reais.

6.1

Cálculo de Reservas de Sinistros Ocorridos e Ainda não Avisados

O mercado segurador nacional realiza mensalmente a provisão para sinistros ocorridos e não avisados, mais conhecido pela sua sigla em inglês IBNR, *incured but not reported*. Esta provisão tornou-se obrigatória a partir do ano de 1998, através da Resolução CNSP nº 18 de 25 de agosto de 1998. Trata-se da soma dos valores de todos os sinistros ocorridos no passado e ainda não avisados até a data presente. Para determinar essa quantia é necessário promover uma estimação, uma vez que os sinistros ocorridos e não avisados são desconhecidos pela seguradora.

Até hoje já foram publicados muitos trabalhos com diferentes métodos para estimar a reserva do tipo IBNR. Um dos mais antigos foi desenvolvido em 1934 por Tarbell (46). Também podem ser citados Doray (13), de 1996, que sugere um modelo de regressão linear lognormal; Herbst (21), de 1999, que basea-se em um modelo de dados aleatoriamente truncados; de Jong (12), de 2004, que propõe um método baseado em séries temporais; e Kubrusly (27), de 2008, que utiliza não só os sinistros ocorridos como também outros dados da apólice para determinar essa reserva. Esse último modelo ainda teve algumas modificações que podem ser vistas nos trabalhos (28) e (26).

Mesmo com tantas formas de se estimar o IBNR, sem dúvida alguma, o método *Chain Ladder* ainda é o mais utilizado na prática. Uma breve apresentação desse método é feita a seguir. Para mais detalhes sobre sua modelagem recomenda-se os trabalhos de England e Verrall (15) e Mack (36).

O método *Chain Ladder* é baseado no triângulo de desenvolvimento, que corresponde à matriz cuja entrada $c_{i,j}$ é definida pelo total de sinistros ocorridos no mês i e avisados em até j meses. Observe que se o triângulo contém as informações dos últimos m meses, as entradas com $j + i \leq m$ representam os sinistros ocorridos e avisados até a data presente. Enquanto as entradas com $j + i > m$ representam sinistros ocorridos e ainda não avisados, por isso estas são entradas incompletas do triângulo de desenvolvimento.

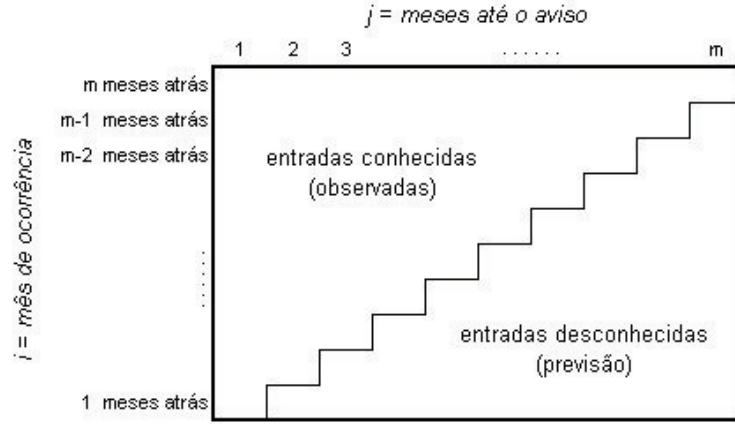


Figura 6.1: Triângulo de desenvolvimento para calcular o IBNR.

O objetivo do método *Chain Ladder* é preencher todas as entradas incompletas do triângulo, prevendo o fator de crescimento entre as colunas. Existem diversas formas de estimar esse fator, por exemplo, para Mack em (35) o fator de crescimento entre as colunas j e $j + 1$ é uma soma ponderada entre todas as razões $c_{i,j+1}/c_{i,j}$, como mostra a equação (6-1).

$$\hat{f}_j = \frac{\sum_{i=1}^{m-j} c_{i,j+1}}{\sum_{i=1}^{m-j} c_{i,j}} = \sum_{i=1}^{m-j} \frac{c_{i,j+1}}{c_{i,j}} \times \frac{c_{i,j}}{\sum_{k=1}^{m-j} c_{k,j}} \quad (6-1)$$

O método *Chain Ladder* segue da seguinte maneira: depois de computados os fatores de crescimento \hat{f}_j , $j = 1 \dots m - 1$ pela equação acima, são estimados primeiro as entradas desconhecidas do triângulo e em seguida o valor do IBNR, de acordo com a seguinte expressão:

$$\text{IBNR}_{MCL} = \sum_{i=2}^m c_{i,m-i+1} \times \left(\hat{f}_{m-i+1} \times \dots \times \hat{f}_{m-1} - 1 \right) \quad (6-2)$$

Mack (35) também mostrou que o estimador acima é não tendencioso quando são satisfeitas as hipóteses: $\mathbf{E}[c_{i,j+1}/c_{i,j} \mid c_{i,1}, c_{i,2}, \dots, c_{i,j}] = \hat{f}_j$, $\mathbf{Var}[c_{i,j+1}/c_{i,j} \mid c_{i,1}, c_{i,2}, \dots, c_{i,j}] = \sigma_j^2$ e para $i \neq k$ $\{c_{i,1}, c_{i,2}, \dots, c_{i,j}\}$ e $\{c_{k,1}, c_{k,2}, \dots, c_{k,j}\}$ são conjuntos independentes.

Modelagem

Sejam $f_{i,j} = \frac{c_{i,j+1}}{c_{i,j}}$ e \hat{f}_j como mostra a equação (6-1). Estudando o comportamento do IBNR ao longo do triângulo de desenvolvimento observou-se que a medida de erro do fator $f_{i,j}$, dada pela razão $\frac{f_{i,j}}{\hat{f}_j}$, está correlacionado com a mesma medida de erro para os fatores $f_{i-1,j}$ e $f_{i,j-1}$, que correspondem à $\frac{f_{i-1,j}}{\hat{f}_j}$ e $\frac{f_{i,j-1}}{\hat{f}_{j-1}}$, respectivamente.

Baseado nessa análise, o método RCRI foi usado para estimar o erro no fator $f_{i,j}$ a partir dos erros nos fatores $f_{i-1,j}$ e $f_{i,j-1}$. Mais precisamente, a equação (6-3) e a figura 6.2 mostram como a amostra de dados do método RCRI é construída a partir do triângulo de desenvolvimento. Para $2 \leq i \leq m-2$ e $2 \leq j \leq m-i$,

$$\mathbf{x} = \left(\frac{f_{i-1,j}}{\hat{f}_j}, \frac{f_{i,j-1}}{\hat{f}_{j-1}} \right) \quad \text{e} \quad y = \frac{f_{i,j}}{\hat{f}_j} . \quad (6-3)$$

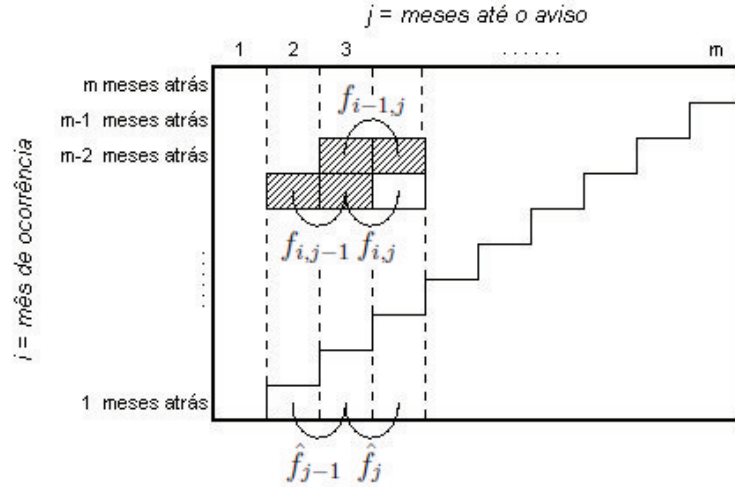


Figura 6.2: Criação dos dados de entrada para a aplicação em IBNR.

Observe que o tamanho N da amostra, formada pelos conjuntos $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \subset \mathbb{R}^2$ e $Y = \{y_1, y_2, \dots, y_N\} \subset \mathbb{R}$, depende do tamanho do triângulo de desenvolvimento, ou seja, $N = N(m)$. Como na equação (6-3) os índices i e j variam de forma que $2 \leq i \leq m-2$ e $2 \leq j \leq m-i$, o número de elementos na amostra pode ser determinado por:

$$N(m) = \sum_{i=2}^{m-2} \sum_{j=2}^{m-i} 1 = \sum_{i=2}^{m-2} (m-i-1) = \frac{m^2 - 5m + 6}{2} \quad (6-4)$$

O desejado é que o método RCRI seja capaz de reconhecer quando um fator $f_{i,j}$ está sub-estimado ou super-estimado e fazer a devida correção nele. Não espera-se que o domínio tenha divisões mais elaboradas, por isso o parâmetro d_{lim} foi escolhido de forma que a árvore tenha no máximo um nó interno e duas folhas, ou seja, $d_{lim} = 1$.

Além disso, acredita-se que em média metade dos fatores estejam sub-estimados e a outra metade super-estimados. Para evitar que a árvore faça uma divisão nos dados que não retrate esse fato, será exigido que no mínimo 40% dos dados de entrada estejam em uma das folhas, ou seja, o parâmetro $n_{lim} = \lfloor 0.4N \rfloor$, onde $\lfloor a \rfloor$ representa a parte inteira do número a e N é o número de linhas da matriz X e do vetor Y .

O parâmetro que indica a probabilidade do erro do tipo I para o teste de hipótese do 4º critério de parada, α , foi aumentado para 0.5. Não que desejassem aumentar a probabilidade desse erro, mas nessa aplicação é interessante aumentar as chances de que ocorra a divisão da raíz. Sem contar que nesse caso, dividir um nó sem necessidade não é algo tão ruim, uma vez que o fator será apenas multiplicado por um número muito próximo de 1. Para os demais parâmetros foram escolhidos seus fatores padrão: $g = 2$, $p_\lambda = 0.5$.

Dados e Comparações

O método RCRI, com a modelagem descrita acima, foi rodado para três conjuntos diferentes de dados, ou seja, três triângulos de desenvolvimento. Os dois primeiros, chamados de ABC e RAA, são triângulos de desenvolvimento disponíveis na plataforma **R** dentro do pacote do *Chain Ladder* (18). Ambos são triângulos incompletos, como na figura 6.1, construídos com dados anuais. O primeiro deles, ABC, é formado por dados de uma carteira de compensação de trabalhadores de uma grande empresa durante o período de 1977 até 1987. Já o segundo, RAA, é composto por dados de passivos em geral no período de 1981 até 1990.

O terceiro conjunto, que será chamado por TRV, é constituído por sinistros no ramo vida com datas de ocorrência de janeiro de 1999 até dezembro de 2002, ou seja, quatro anos de observação. O triângulo incompleto foi montado com observações mensais e por isso ele tem 48 linhas e colunas. A fim de facilitar as comparações com outro métodos seria muito importante saber qual o valor real que se deseja estimar. Para isso foram consideradas apenas as primeiras 24 linhas e colunas do triângulo incompleto, formando assim um triângulo completo de 24 meses como mostra a figura 6.3.

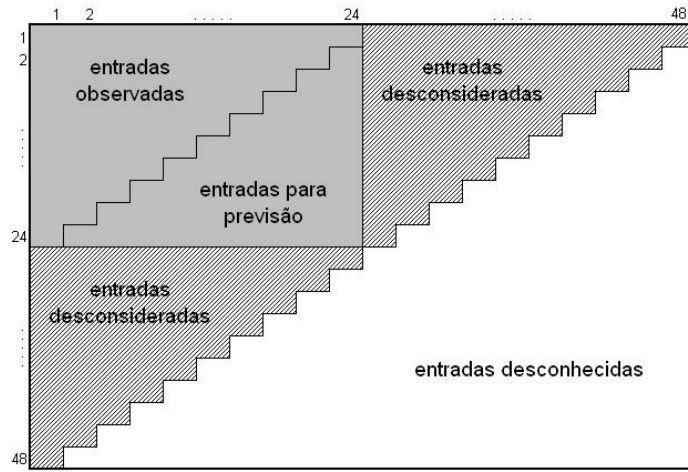


Figura 6.3: Dados utilizados na estimativa do IBNR.

Dessa forma, para os dados TRV, será possível determinar o chamado IBNR_{OBS} , que é formado pela soma dos valores de todos os sinistros ocorridos antes do ano representado pela linha 24 e avisados depois desta data. É claro que o seu valor não é exatamente o valor real do IBNR, uma vez que podem existir sinistros com intervalo de tempo entre a ocorrência e o aviso maior que 24 meses, mesmo assim é um bom valor para comparações. Sua expressão está definido na equação (6-5).

$$\text{IBNR}_{OBS} = \sum_{i=2}^{24} (c_{i,24} - c_{i,25-i}) \quad (6-5)$$

Para saber se a aplicação o método RCRI no problema de estimar o valor do IBNR teve bons resultado, além do valor do IBNR_{OBS} , sua saída foi comparada com as saídas de outros quatro modelos. O primeiro deles foi IBNR_{MCL} , definido na equação (6-2).

O segundo é o *Bootstrap Chain Ladder*, chamado nesse trabalho por IBNR_{BCL} . Trata-se de um método que também procura corrigir o *Chain Ladder* realizando os seguintes passos: 1º) calcula-se o IBNR_{MCL} ; 2º) calcula-se os resíduos de *Pearson*; 3º) estima-se o incremento nos pagamentos de sinistros; 4º) simula-se o erro do processo. Esse método está disponível na plataforma **R** dentro do pacote do *Chain Ladder* (18). Mais detalhes desse modelo encontram-se no artigo do England e Verrall (15).

O terceiro, chamado por IBNR_{CLGPR} , é o modelo apresentado no trabalho (25). Este também busca estimar o erro gerado pelo *Chain Ladder*, mas sua correção é feita através da ferramenta de aprendizagem computacional *Gaussian Process Regression*. Nesse caso, diferente do RCRI, a aprendizagem é feita diretamente pelas entradas do triângulo em vez de pelos fatores.

A última comparação será com a resposta dada pelo método RCRI caso a raiz não seja dividida. Essa comparação é muito importante para que se possa verificar que não só a modelagem está passando informação sobre o erro, como também a divisão nos dados feita pela árvore reconhece padrões na amostra. Chamada de **IBNR_{ACL}** (*Adjusted Chain Ladder*), essa saída é simplesmente o resultado do *Chain Ladder* quando todos os fatores são multiplicados pelo média de $\frac{f_{i,j}}{f_j}$ para $i = 2 \dots m - 2$ e $j = 2 \dots m - i$.

Resultados

Os primeiros resultados estão na tabela 6.1 e representam as saídas de todos os métodos citados acima para os dados dos triângulos ABC e RAA.

	IBNR_{MCL}	IBNR_{BCL}	IBNR_{CLGPR}	IBNR_{ACL}	IBNR_{RCRI}
ABC	5.277.760,36	5.279.419,00	5.301.731,00	5.261.861,20	5.787.129,73
RAA	52.135,23	54.360,00	54.002,76	86,059.35	86,059.35

Tabela 6.1: Valores das saídas para os dados ABC e RAA.

Como esses triângulos são muito pequenos, não foi possível reduzir os dados para computar o **IBNR_{OBS}**. Além disso, o número de elementos na amostra do método RCRI para o triângulo ABC é $N = 36$ e para o triângulo RAA é $N = 28$, de acordo com a equação (6-4) considerando $m = 11$, para o ABC, e $m = 10$ para o RAA. Por se tratarem de amostras muito pequenas, o resultado não tem como ser muito significativo. Mas mesmo assim podem ser feitos alguns comentários das saídas obtidas.

Para os dados ABC o **IBNR_{ACL}**, que representa a saída caso a árvore não faça divisão alguma, ficou abaixo de **IBNR_{MCL}** e o **IBNR_{RCRI}** ficou acima. Isso indica que o método RCRI reconheceu um padrão no domínio e assim realizou uma partição, na tentativa de melhorar a estimativa final. Já para os dados RAA o **IBNR_{ACL}** e o **IBNR_{RCRI}** foram iguais, mostrando que nesse caso não foi possível reconhecer um padrão no domínio por parte do método RCRI.

Para esses dois primeiros triângulos não foi possível fazer uma análise mais profunda do desempenho dos métodos, uma vez que são poucos os dados. Já para o último banco de dados, TRV, a amostra é um pouco maior, o que possibilita o cálculo do **IBNR_{OBS}**. Isso torna essa comparação muito mais rica, como pode ser visto a seguir.

Para que se possa ter uma noção da estimativa do IBNR ao longo do tempo, os métodos de estimação foram rodados para sub-triângulos de tamanho 8 até de tamanho 24, que geram amostras para o RCRI de tamanhos 15 até 231, de acordo com a equação (6-4). A fim de esclarecimento, o sub-triângulo de tamanho m é formado pelos dados da semi-diagonal pra cima da matriz formada pelas primeiras m linhas e colunas. Os resultados estão expostos na tabela 6.2 e no gráfico apresentado da figura 6.4.

	IBNR_{MCL}	IBNR_{BCL}	IBNR_{CLGPR}	IBNR_{ACL}	IBNR_{RCRI}	IBNR_{OBS}
8	2.929,94	2.950,00	2.917,67	2.677,52	2.677,52	2.198,06
9	2.992,19	3.025,00	2.877,16	2.731,58	2.731,58	2.403,73
10	2.765,33	2.793,00	2.711,66	2.528,26	2.528,26	2.702,46
11	2.035,77	2.065,00	1.943,91	1.800,01	1.800,01	3.033,43
12	2.371,65	2.358,00	2.438,32	2.152,28	2.730,53	3.026,79
13	2.554,99	2.572,00	2.460,65	2.370,50	2.999,69	3.433,79
14	3.084,32	3.101,00	3.097,57	2.915,44	2.915,44	2.858,02
15	3.271,88	3.277,00	3.226,43	3.095,54	3.202,44	2.854,45
16	3.214,53	3.225,00	3.257,16	3.035,44	3.060,57	3.095,06
17	3.191,01	3.206,00	3.207,12	2.989,82	3.549,23	3.332,77
18	3.126,63	3.151,00	2.940,38	2.935,95	3.571,60	3.622,55
19	3.512,36	3.490,00	3.512,79	3.299,73	4.096,49	4.239,08
20	3.931,84	3.938,00	3.932,26	3.713,42	4.374,33	4.252,08
21	4.020,38	4.029,00	4.020,43	3.840,30	4.322,01	4.637,02
22	4.740,65	4.746,00	4.740,85	4.502,60	4.502,60	4.506,90
23	4.012,00	4.010,00	4.012,23	3.760,20	4.765,70	4.446,52
24	3.591,91	3.584,00	3.592,35	3.350,87	3.997,54	4.626,28

Tabela 6.2: Valores das saídas para os dados TRV.

A primeira observação a ser feita é que em 13 das 17 vezes que os métodos foram rodados o RCRI obteve a melhor saída. Isso é um dado significativo que mostra o bom desempenho desse método para essa aplicação.

Outra observação importante é que nem sempre o método reconhece alguma padrão no domínio e nesse caso ele fica idêntico ao ACL, como aconteceu para os triângulo de tamanho 8, 9, 10, 11, 14 e 22. O interessante em ser destacado é que, a não ser pelo triângulo de tamanho 15, todas as vezes que houve uma partição no domínio o método RCRI supera o ACL. Isso mostra que está havendo uma aprendizagem por parte do método quando a raiz da árvore é subdividida.

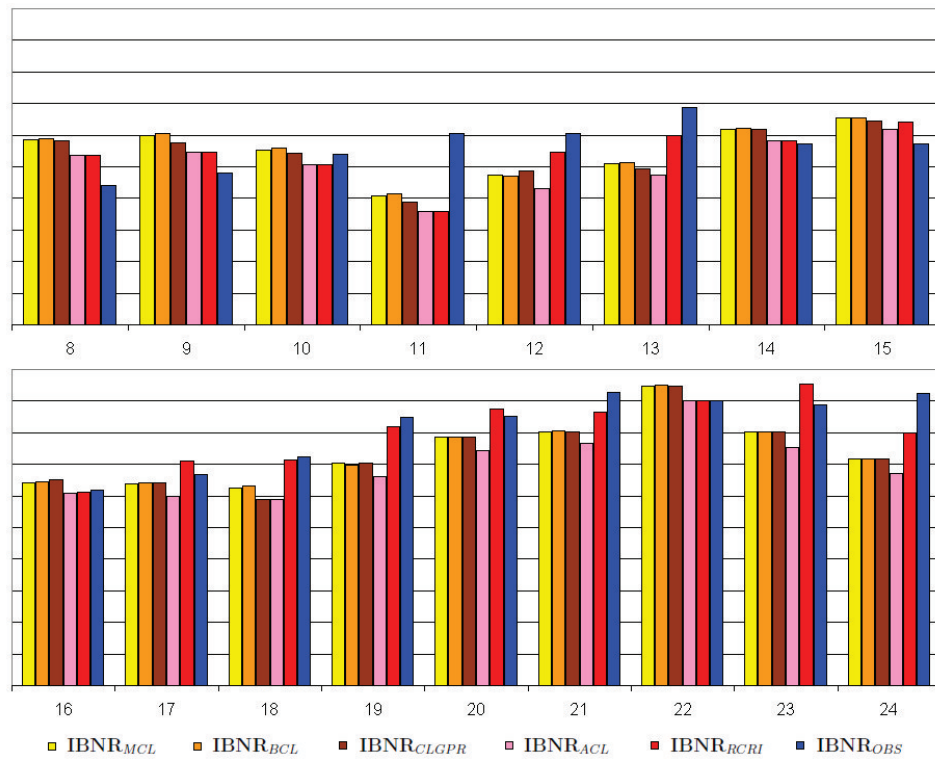


Figura 6.4: Resultado para dados TRV.

Se para cada tamanho de triângulo for calculado o erro quadrático, definido por $\text{Erro}^2 = (\text{IBNR}_X - \text{IBNR}_{OBS})^2$, onde X representa qualquer um dos métodos, é possível averiguar o comportamento das saídas de uma forma geral. Para isso foram computados os 17 erros quadráticos, um para cada tamanho de triângulo, e dentro dessa amostra as estatísticas média, mediana, MAD, mínimo e máximo estão apresentadas na tabela 6.3 abaixo.

	IBNR_{MCL}	IBNR_{BCL}	IBNR_{CLGPR}	IBNR_{ACL}	IBNR_{RCRI}
média	$3,48 \times 10^5$	$3,50 \times 10^5$	$3,68 \times 10^5$	$4,91 \times 10^5$	$1,75 \times 10^5$
mediana	$2,46 \times 10^5$	$2,22 \times 10^5$	$2,24 \times 10^5$	$2,90 \times 10^5$	$8,78 \times 10^4$
MAD	$1,95 \times 10^5$	$2,05 \times 10^5$	$1,98 \times 10^5$	$2,87 \times 10^5$	$7,28 \times 10^4$
mínimo	$3,95 \times 10^3$	$8,20 \times 10^3$	$8,46 \times 10^1$	$1,85 \times 10^1$	$1,85 \times 10^1$
máximo	$1,07 \times 10^6$	$1,09 \times 10^6$	$1,19 \times 10^6$	$1,63 \times 10^6$	$1,52 \times 10^6$

Tabela 6.3: Estatísticas para os erros quadráticos do IBNR.

Mais uma vez o resultado retrata o bom desempenho do método RCRI, que teve a menor média, a menor mediana, o menor MAD e o menor mínimo erro quadrático.

6.2

Geologia de Petróleo

Nos últimos anos a indústria do petróleo tem investido muito em pesquisa na área de geologia. Com os altos custos na obtenção de amostras, cresce a necessidade de aprimorar as técnicas de mineração de dados, a fim de se obter uma melhor qualidade em sua interpretação. Uma entre as diversas possibilidades de pesquisa nessa área é o estudo estatístico que busca a estimação de dados faltantes dentro de uma amostra já obtida.

Em muitos bancos de dados, com informações sobre poços, existem algumas variáveis com valores desconhecidos. A proposta é realizar uma regressão onde a variável de predição é composta por esses dados faltantes, enquanto as variáveis explicativas são os dados completos. Dessa forma, busca-se aproveitar ao máximo as informações que custaram muito para serem obtidas.

A realização dessa regressão possui algumas fragilidades, entre elas a não linearidade entre as variáveis explicativas e de predição. Dessa forma torna-se valiosa a existência de uma ferramenta estatística não linear, supervisionada, que consiga trabalhar com o maior número de atributos possíveis. Por se tratar de uma ferramenta que satisfaz todas essas exigências, é grande a motivação em usar o método de Regressão Construtiva em Regiões Implícitas (RCRI).

Modelagem

São vários os dados dos poços que podem ser estudados, como por exemplo, rhob, phic e dt. Essa aplicação será feita apenas para rhob, mas poderia ser estendida para os outros também. Rhob é um dado petrofísico, com valores reais, que mede a densidade da rocha. Ele é usado para determinar sua porosidade e assim caracterizar o tipo de rocha no local, como por exemplo, identificar a formação de areias, areias com óleo, argila ou sal. Dessa forma, conhecendo seu valor, é possível inferir melhor o volume poroso, ou o volume de óleo no reservatório, e com isso auxiliar nas decisões de exploração.

As figuras 6.5 e 6.6 a seguir foram retiradas da página do professor Lars Ole Boldreel, do departamento de Geografia e Geologia da Universidade de Copenhagen. Elas ilustram a relação entre o valor de rhob e os diferentes tipos de rocha.

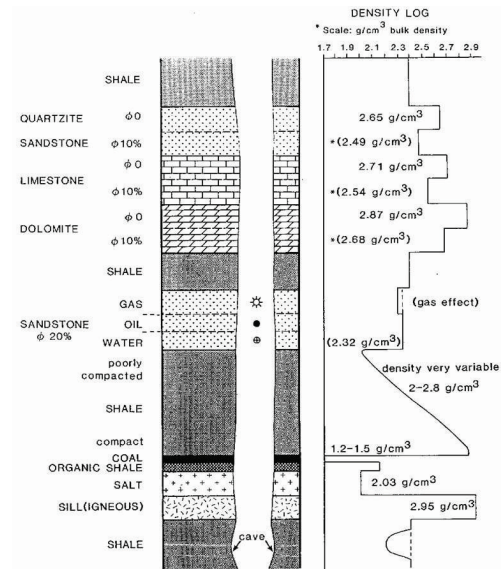


Figura 6.5: Relação entre rhob e os diferentes solos.

Em 6.5 rhob é indicado por "DENSITY LOG". A figura representa o corte transversal ao longo de um poço e mostra como varia o valor de rhob quando muda o tipo do rocha. Já a figura 6.6 representa o gráfico de rhob em função da profundidade. Ele mostra como é possível perceber a mudança das diferentes fácies observando os valores de rhob. No gráfico, quando mais para a direita no eixo horizontal, mais profundo está o ponto amostrado.

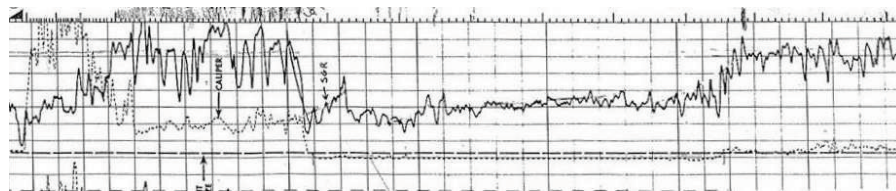


Figura 6.6: Exemplo de uma amostra de rhob.

A idéia é que o método RCRI utilize os atributos geológicos conhecidos para inferir no valor de rhob desconhecido, ou seja, para identificar o tipo de rocha através de uma partição do domínio. A única manipulação nos dados de entrada feita foi a normalização, a fim de padronizá-los evitando assim escalas muito diferentes. Mas isso não modificou a escala do resultado final, uma vez que as contas inversas forma feitas antes da comparação.

Praticamente todos os parâmetros foram escolhidos de acordo com os seus valores padrão: $g = 2$, $n_{lim} = \frac{N}{2^{d_{lim}}}$, $\alpha = 0.05$ e $p_\lambda = 0.5$. Como a amostra de dados é muito grande foi possível permitir que a árvore seja mais profunda, por isso foi escolhido $d_{lim} = 6$.

Dados e Comparações

A amostra utilizada nessa aplicação é de domínio público e pode ser encontrada no endereço <http://nerslweb.cr.usgs.gov>. Tratam-se de informações geológicas sobre poços localizados no Alaska, EUA. Foram considerados 4 poços em diferentes localidades: norte, sul, leste e oeste. Cada um dos poços possui uma amostra de tamanho diferente, variando de 7000 a 20000 pontos, que são formadas por 10 variáveis explicativas e 1 de predição, rhob.

Cada poço foi rodado separadamente. Em cada um deles 90% da amostra foi usada como dados de aprendizagem enquanto os outros 10% como dados para testes fora da amostra. Para garantir uma boa amostragem, os dados de aprendizagem foram escolhidos de forma aleatória. Para descartar a possibilidade de que o sorteio possa favorecer ou prejudicar o modelo, o problema foi rodado 10 vezes, em cada uma das vezes uma diferente amostra foi escolhida aleatoriamente.

Com o propósito de comparação essa regressão não será testada apenas pelo RCRI. Uma alternativa natural para resolver problemas com grande quantidade de informações, que é esse caso, é utilizar ferramentas de inteligência artificial. Introduzidos por Vapnik (47), Métodos Baseados em Núcleos (*Kernel Based Methods*) e as Máquinas de Suporte Vetorial (*Support Vector Machines*, ou simplesmente SVM's) são algoritmos de aprendizagem estatística que têm tido bastante sucesso nas áreas de geofísica e geologia do petróleo, como mostram os artigos (3), (7), (29), (31), (32), (43) e (49). Quando o problema é de regressão, em vez de classificação, utiliza-se a extensão contínua para o SVM, chamada por Regressão de Suporte Vetorial (*Support Vector Regression* - SVR).

Por esse motivo a ferramenta escolhida para comparação foi o SVR. Um pequeno resumo sobre o SVR encontra-se no apêndice B, mais detalhes sobre esse método podem ser encontrados no livro de Cristianini e Shawe-Taylor (10). Para rodar o SVR foi utilizada a biblioteca LIBSVM (8), implementada para Java. Um pseudo código desse algoritmo pode ser encontrado em (16). Aos parâmetros usados para esse modelo foram atribuídos os valores padrões para regressão: `svm_type = epsilon-SVR`, `kernel_type = RBF` (*radial basis function*), `d = 3`, `gamma = 1.0/d`, `coef0 = 0`, `nu = 0.5`, `cache_size = 40`, `C = 1`, `eps = 0.01`, `p = 0.1`, `shrinking = 1`, `probability = 0`, `nr_weight = 0`.

Resultados

Em cada uma das 10 vezes foi calculado o erro médio quadrático, utilizando os dados fora da amostra. Entre esses 10 erros médios quadrático foram determinados algumas estatísticas, apresentadas na tabela 6.4 abaixo.

	RCRI (Local)	RCRI (Global)	SVR	RCRI (Local)	RCRI (Global)	SVR
mediana ($\times 10^{-3}$)	2,10	2,05	4,09	4,83	4,83	5,59
MAD ($\times 10^{-4}$)	0,589	0,515	1,93	6,80	6,87	5,28
mínimo ($\times 10^{-3}$)	1,96	1,90	3,76	3,76	3,74	4,32
máximo ($\times 10^{-3}$)	2,48	2,44	5,16	5.68	5.62	6.43
Norte			Sul			

	RCRI (Local)	RCRI (Global)	SVR	RCRI (Local)	RCRI (Global)	SVR
mediana ($\times 10^{-3}$)	4,54	4,63	5.15	3,55	3,52	4,06
MAD ($\times 10^{-4}$)	1,98	1,76	3,46	0,966	0,931	1,92
mínimo ($\times 10^{-3}$)	3,89	4,11	4,77	3,33	3,34	3,52
máximo ($\times 10^{-3}$)	4.85	4,89	6,11	3,92	3,99	4,50
Oeste			Leste			

Tabela 6.4: Estatísticas, dos quatro poços, para os erros médios quadráticos nas estimativas de ρ_{hob} .

A tabela 6.4 mostra o bom desempenho do método RCRI. Como pode ser visto, foi testado tanto a forma local quanto a forma global de se estimar os parâmetros nas folhas. Veja na equação (3-8) que para determinar a estimativa global é preciso resolver um sistema $N \times l$, onde N é o número de dados de entrada e l o número de folhas. Por isso, a estimativa global é mais lenta que a local, principalmente quando N e l são razoavelmente grandes.

Como as amostras utilizadas nessa aplicação tem $N > 7.000$, a diferença no tempo de execução entre a forma local e global fica notória. Por exemplo, para o caso do poço ao norte, $N = 17.765$, o tempo gasto para se construir a árvore global é mais de 10 vezes o tempo gasto com a árvore local. No caso local, em média, foram gastos 57 segundos, enquanto no caso global a média ficou em 13 minutos e 29 segundos. A fim de comparação, o modelo SVR demorou, em média, 2 minutos e 34 segundos. Todos esses valores dizem respeito ao tempo gasto em um processador Pentium(R) D CPU 2.80GHz com 1G de RAM na linguagem java.

Então, mesmo a forma global obtendo melhores resultados, é preferível à forma local, que é bem mais rápida e com saídas praticamente tão boas quanto. Além disso, seus resultados são melhores que os do SVR e seu tempo de execução é menor do que a metade do tempo gasto por ele.