

2

Uma Metodologia para Classificação do Risco de Eventos Adversos em LLA

Neste capítulo serão expostos os aspectos relativos a material e métodos.

Problemas com um número elevado de variáveis e um número de casos reduzidos exigem, mais do que em nenhum outro, um tratamento adequado dos dados. É importante que se escolha criteriosamente o conjunto de variáveis a serem efetivamente utilizadas. Cuidado especial há que se ter também com a validação do modelo usado, no presente caso, Redes Neurais. O reduzido número de casos impede qualquer alternativa de particionamento do banco de dados para testes na-amostra e fora-da-amostra, forçando a utilização do método *'leave-one-out'* para que a estimativa obtida possa ser considerada como testada fora-da-amostra.

Este capítulo se inicia com a descrição do banco de dados; em seguida é apresentada a proposta de estratégia utilizada para seleção do conjunto de variáveis de cada paciente; e finaliza-se apresentando o modelo de Redes Neurais usado para estimar o risco de evento adverso.

2.1. Descrição e Tratamento da Base de Dados

2.1.1. Descrição da Base de Dados

O banco de dados utilizado nesta tese é composto por pacientes no Setor de Hematologia do Instituto de Puericultura e Pediatria Martagão Gesteira (IPPMG) da UFRJ e no Serviço de Hematologia do Hospital Universitário Pedro Ernesto (HUPE) da UERJ. Estas duas instituições universitárias recebem pacientes com doenças malignas hematológicas provenientes principalmente do Município do Rio de Janeiro e são também centros de referências para tratamento de crianças com LLA no Brasil. O período do estudo envolve pacientes diagnosticados desde

abril de 1993. O banco foi atualizado, no que concerne à recidiva, até abril de 2004.

As variáveis de entrada utilizadas para tratar o problema são compostas por um conjunto de dados clínicos e biológicos obtidos quando do diagnóstico. O banco de dados original é composto por 189 variáveis e de 152 crianças com leucemia linfoblástica aguda (LLA) diagnosticadas e tratadas nos hospitais IPPMG/UFRJ e HUPE/UERJ. Estas variáveis (candidatas à entrada do modelo) podem ser organizadas basicamente nos seguintes grupos: a) dados clínicos ao diagnóstico - idade; sexo; raça; data do diagnóstico; data de nascimento, presença ou ausência de perda de peso; febre; anemia; artrite; hemorragia; aumento de linfonodos; aumento do fígado e baço; foco infeccioso; dor abdominal; dor óssea; massa mediastinal; massa abdominal, dentre outras; b) dados laboratoriais ao diagnóstico - concentração de hemoglobina no sangue periférico; contagem de leucócitos do sangue periférico; contagem de blastos do sangue periférico; contagem de plaquetas no sangue periférico; LDH sérica; classificação FAB; imunofenótipo dos blastos (T ou não T) e anormalidades citogenética em blastos, dentre outras; c) dados relacionados à resposta ao tratamento - número absoluto de blastos no sangue periférico no 8º dia de tratamento; remissão completa no 33º dia de tratamento, entre outras.

O trabalho de tese envolveu um tratamento criterioso destas variáveis de forma a não só selecionar àquelas relevantes para a estimativa do risco de evento adverso pelo modelo proposto, mas como também tornar viável o tratamento de um banco de dados com tão poucas observações e tantas variáveis.

São comuns as situações onde nem todas as variáveis são igualmente importantes para a realização de uma tarefa específica. Algumas das variáveis podem ser redundantes ou irrelevantes, e mais do que isso é conhecido que, encontrar uma combinação apropriada de variáveis explicativas pode melhorar significativamente o desempenho do modelo (Fukunaga, 1972). Esta questão é ainda mais delicada quando as variáveis alimentam um modelo não-linear. O tratamento das variáveis se dará em dois níveis, primeiramente serão explorados aspectos de redundância e capacidade de transmitir informação ao desfecho, em seguida, já explorando os resultados numéricos, diferentes conjuntos de variáveis serão testadas frente ao modelo proposto.

2.1.2. Tratamento da Base de Dados

Uma das contribuições importantes desta tese é a definição de uma estratégia cuidadosa para o tratamento da base de dados. Esta estratégia pode ser dividida em três etapas descritas a seguir:

1) Eliminação de variáveis com dados faltantes.

Nesta etapa eliminaram-se as variáveis para as quais existia um excesso de dados faltantes e também variáveis redundantes (variáveis que são fatores de outras¹). Este critério é importante pois, como visto anteriormente, o banco é composto por “apenas” 152 pacientes e a não exclusão de variáveis com muitos dados faltantes envolveria redução do número de casos o que poderia inviabilizar as análises. Nesta fase o banco de dados foi reduzido de 189 para 51 variáveis (redução de 73%).

2) Avaliação do poder discriminante e do valor de predição positiva das variáveis.

Nesta etapa selecionou-se variáveis através do cruzamento com a variável “desfecho²” (utilizada no sistema de estimativa do risco) utilizando os seguintes critérios:

- “poder discriminante” da variável: o objetivo principal desta análise é explorar relações entre a presença ou ausência de um determinado sinal ou sintoma na presença ou ausência de recaída. Por exemplo, ao analisar a variável “palidez”, verifica-se que dos 145 pacientes, apenas 17 não apresentaram este sinal, sendo que destes, 41% (ou seja, apenas 7 pacientes) não tiveram recaída e 59% (ou seja, apenas 10 pacientes) tiveram recaída. Isto é, estar-se-ia tirando conclusões a partir de um pequeno número de pacientes (17).
- “valor da predição positiva (VPP)”: probabilidade do paciente ter recaída condicionado ao fato dele apresentar determinados sintomas e/ou sinais. Adotou-se o critério de selecionar variáveis que

¹ Por exemplo, uma variável categórica (em faixas) criada a partir da variável idade (Categoria 1: < 1 ou > 10 anos; Categoria 2: entre 1 e 10 anos, etc).

² Variável dicotômica que indica se o paciente teve recaída (assume valor 1) ou não (assume valor 0).

apresentassem probabilidade condicional acima de 50% de ter ou não recaída.

Como consequência dos critérios desta etapa, o número de variáveis foi reduzido para 20 (redução acumulada de 89%). As variáveis excluídas foram: palidez, emagrecimento, artrite, artralgia, dor em vários locais, gânglios em vários locais, aplasia de medula óssea, uso prévio de corticoide, lesão dos ossos, sangramento em vários locais e presença de foco infeccioso. As variáveis que continuaram a serem consideradas foram: sexo, astenia, dor, dor óssea, tamanho do fígado, tamanho do baço, gânglio, gânglio cervical, gânglio inguinal, sangramento, febre, infecção, leucócitos, resposta ao corticóide no oitavo dia, idade, classificação FAB, raça, blastos, hemoglobina e imunologia T. Para essas 20 variáveis, o número de casos foi reduzido de 152 para 144 pacientes pois ainda existiam alguns dados faltantes.

3) Seleção das variáveis através de Informação Mútua.

Sobre o conjunto de variáveis obtida na etapa anterior utilizou-se o algoritmo de Seleção de Variáveis sob Informação Mútua com Distribuição Uniforme de Informação (MIFS-U) (Kwak & Choi, 2002). Trabalhos anteriores usando informação mútua para seleção de variáveis no contexto de Redes Neurais inclui Battiti (1994), Daberllay (1997,2000) e Setiono & Liu (1997). O resultado desta etapa é uma seleção das variáveis por ordem de importância. Este procedimento é resumido a seguir (e descrito com mais detalhes no apêndice A):

Seja F o conjunto de variáveis selecionadas previamente (no caso, as 20 variáveis que foram selecionadas pela etapa 2) e S o conjunto de variáveis a serem selecionadas por ordem de importância (resultado desta etapa). S é inicialmente um conjunto vazio.

Calcula-se então, a informação mútua (IM) entre cada uma dessas variáveis com a variável desfecho. A informação mútua é uma medida de dependência entre variáveis aleatórias. É verdade (Cover, 1991) que $IM \geq 0$ e que $IM = 0 \Rightarrow$ independência estatística. Um valor alto (ou pequeno) de IM significa que as variáveis são muito (ou pouco) relacionadas.

Na sequência, seleciona-se (f_i), a variável que apresenta a maior informação mútua com o desfecho. Então $F \leftarrow F - \{f_i\}$ e $S \leftarrow \{f_i\}$. O processo de cálculo da informação mútua deveria se repetir considerando-se agora (1) o desfecho, (2) a variável selecionada e (3) as variáveis candidatas a serem

selecionadas. Este processo torna-se rapidamente complexo, uma vez que se teria de calcular uma distribuição conjunta de três variáveis aleatórias, o que implicaria na necessidade de se estimar a função densidade de probabilidade conjunta destas 3 variáveis. Este procedimento fica ainda mais difícil no caso presente, por causa da pequena quantidade de dados. O algoritmo MIFS-U (Kwak & Choi, 2002), e os resultados obtidos, descritos no apêndice A, reduz o problema a cálculos de informação mútua entre duas variáveis aleatórias.

Obtém-se então uma ordenação por significância das 20 variáveis selecionadas nas etapas anteriores. Esta ordenação está apresentada no quadro 03 a seguir.

Quadro 03 Resultado Preliminar de Seleção e Ordenação de Variáveis

RESULTADO		
Ordem de Importância	Variáveis	IM
1	BLASTOS	0.203
2	IDADE	0.086
3	Imunofenótipo T	0.032
4	RESP. ao TTO no 8º DIA	0.036
5	SEXO	0.024
6	Classificação FAB	0.000
7	RAÇA	0.012
8	ASTENIA	0.004
9	FEBRE	0.004
10	G. CERVICAL	0.010
11	G. VIRILHA	0.001
12	DOR OSSEA	0.002
13	SANGRAMENTO	0.002
14	INFECÇÃO	0.001
15	DOR	0.001
16	HEMOGLOBINA	0.069
17	GANGLIO	0.001
18	BAÇO	0.033
19	FIGADO	0.063
20	LEUCÓCITOS	0.197

As variáveis “blastos” e “leucócitos” têm um grau de associação (coeficiente de correlação linear de Pearson (Larson, 1982)) muito forte (0,99)³. Nesta etapa, a indicação da variável ‘leucócitos’ como a última (menos relevante no grupo) se deu claramente pelo seu caráter de redundância com a variável selecionada como principal que foi ‘blastos’. Do ponto de vista de informação mútua (IM) com o desfecho, os valores relativos às duas variáveis são muito semelhantes indicando que há pouca diferença entre usar uma ou a outra. Claro, que pelo aspecto de redundância, ao se considerar uma delas, a outra é colocada no final da lista, como de fato ocorreu no resultado apresentado no quadro 03. Optou-se então por considerar a ‘contagem de leucócitos’ (aqui denominada simplesmente de ‘leucócitos’) em detrimento da ‘contagem de blastos’ (denominada por simplicidade de ‘blastos’) uma vez que a primeira é quantificada precisamente de forma automatizada, enquanto a segunda depende de um observador treinado para examina-los ao microscópio óptico. Além disso, a maioria dos sistemas de classificação de risco, como por exemplo, o BFM95, utiliza a variável contagem de leucócitos no sangue periférico.

No quadro 04 apresenta-se o resultado final de seleção e ordenação obtido pelas etapas anteriormente descritas.

³ O coeficiente de correlação pode variar de -1 (relação inversa) a 1 (relação direta) e quanto mais próximo dos extremos, maior a associação entre as variáveis. Obviamente, um alto grau de associação entre duas variáveis pode indicar que apenas uma seja necessária nas análises.

Quadro 04 Resultado de Seleção e Ordenação de Variáveis Após as Três Etapas do Tratamento

RESULTADO	
Ordem de Importância	Variáveis
1	LEUCÓCITOS
2	IDADE
3	Imunofenótipo T
4	RESP. ao TTO no 8º DIA
5	SEXO
6	Classificação FAB
7	RAÇA
8	ASTENIA
9	FEBRE
10	G. CERVICAL
11	G. VIRILHA
12	DOR OSSEA
13	SANGRAMENTO
14	INFECCÃO
15	DOR
16	HEMOGLOBINA
17	GANGLIO
18	BAÇO
19	FIGADO

Claramente incluem-se, dentre estas variáveis pré-selecionadas, algumas com razoável grau de subjetividade e de difícil interpretação do ponto de vista biológico. Entretanto, como será visto no capítulo seguinte, após algumas etapas de refinamento estas variáveis biologicamente inconsistentes também se mostrarão irrelevantes na metodologia proposta. Note-se que o processamento realizado até este ponto sobre estas variáveis independeu do modelo de estimação de risco. Na seqüência apresenta-se o modelo proposto para estimação do risco de recidiva, primeiramente na sua forma básica sendo em seguida propostos refinamentos ao mesmo.

2.2. Estimativa do Risco de Evento Adverso

Nesta seção apresenta-se a proposta de modelagem para estimativa de risco de evento adverso. Primeiramente, na subseção a seguir, resume-se um dos critérios mais importantes em uso, o BFM95, que será usado no capítulo de

resultados, com fins de análise e comparação com a metodologia proposta. Ainda nesta seção mostra-se a estratificação de risco pelo critério BFM95 no banco de dados utilizado nesta tese.

2.2.1. Estimativa pelo Critério BFM95

A estratificação terapêutica é feita através do enquadramento dos pacientes em três ramos terapêuticos: risco padrão (SR), risco médio (MR) e alto risco (HR). Essa distribuição baseia-se em critérios onde o número de leucócitos/ μl no sangue ao diagnóstico é um fator importante (Riehem, 1995).

Na figura 01 apresenta-se um diagrama das possibilidades da relação risco e desfecho dos pacientes.

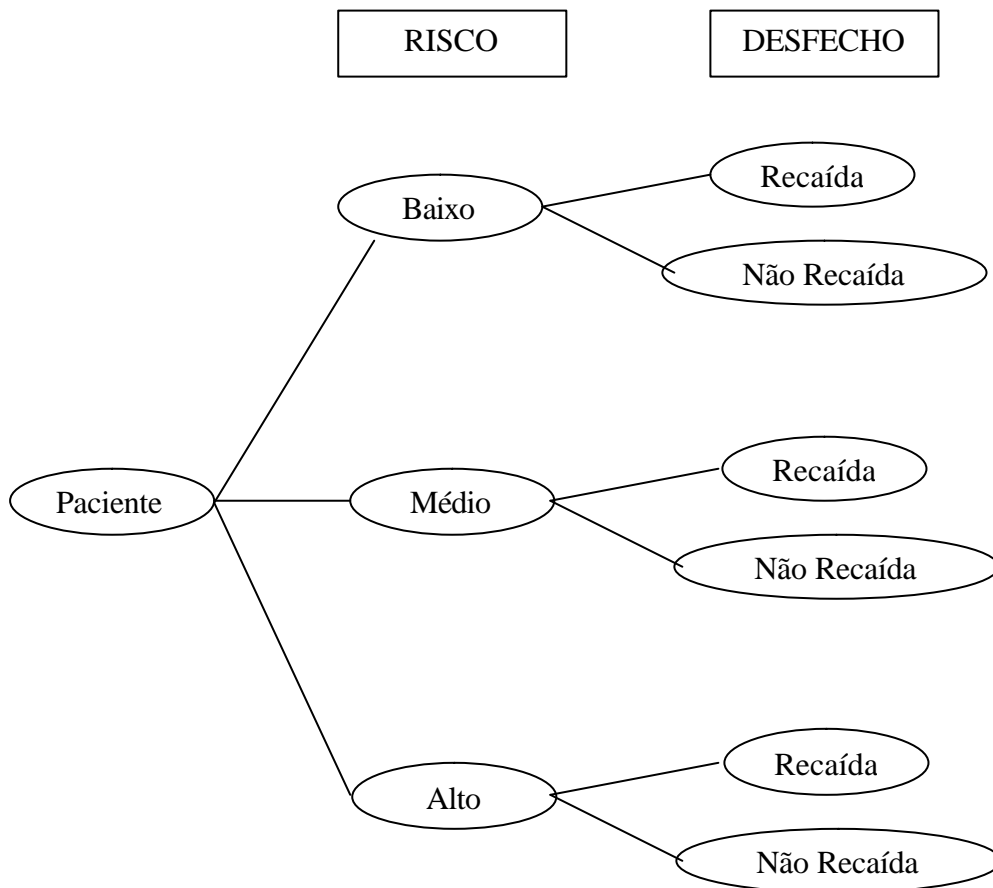


Figura 01. Representação do Desfecho em Função do Risco BFM95

No quadro 05 encontra-se o cenário de estratificação de risco versus desfecho do critério BFM95 no banco de dados utilizado.

Quadro 05: Cruzamento Desfecho *versus* Risco – Avaliação pelo Critério BFM95

	Desfecho				Total	
	Não recaída		Recaída		%	Casos
	Casos	%	Casos	%		
Risco Baixo	24	26.7%	5	9.3%	20.1%	29
Médio	46	51.1%	24	44.4%	48.6%	70
Alto	20	22.2%	25	46.3%	31.3%	45
Total	90	100.0%	54	100.0%	100.0%	144

Uma parcela razoavelmente importante dos pacientes, 44.4%, foi avaliada como sendo de médio risco de recidiva.

2.2.2.

Uma Proposta de Modelagem para a Estimativa de Risco

O risco de evento adverso será estimado usando uma Rede Neural *feedforward*. Este tipo de modelo não linear tem sido usado com sucesso em uma gama extensiva de aplicações desde o final da década de 80. Referências clássicas em Redes Neurais incluem Haykin, 1988, Bishop, 1995 e Príncipe *et al.*, 2000.

Redes Neurais Artificiais (ou simplesmente ‘Redes Neurais’) é um modelo distribuído composto por unidades (chamadas na literatura de ‘neurônios’) constituídas de funções não lineares (tipicamente sigmóides e tangentes hiperbólicas). A combinação destas unidades, através de parâmetros estimados a partir dos dados, é o que confere a capacidade deste modelo de inferir relações não lineares de complexidade arbitrária. Na forma como será usada nesta tese estas unidades são arrumadas em camadas, incluindo uma camada oculta, que não está diretamente conectada a saída do modelo. Estas conexões entre as unidades, ou neurônios são chamadas de pesos (originalmente a terminologia era ‘pesos sinápticos’). Estes pesos são os parâmetros do modelo que serão ajustados por um algoritmo iterativo através dos dados. Uma vez ajustados os pesos à rede tem a capacidade de representar a relação dos dados (variáveis de entrada) com a variável de saída, neste caso a estimativa do risco. A capacidade de aprender através de ‘exemplos’ ou dados (na-amostra) e de generalizar (fora-da-amostra) informação gerada em ambientes não-lineares complexos, é sem dúvida a grande vantagem das Redes Neurais.

As variáveis selecionadas ao final da seção 2.1.2, quadro 04, serão utilizadas, paciente a paciente, como entrada da Rede Neural, e se espera após o processo de treinamento ter como saída da Rede à estimativa do risco de evento adverso. Assim, na fase de treinamento, a saída da rede assume valores 0 ou 1 para cada uma de duas possibilidades, não recaída ou recaída. Usa-se uma função de ativação sigmóide na unidade de saída de forma que a saída da rede varie sempre entre 0 e 1 já que esta função satura nestes valores. Já na fase de teste, e posteriormente de utilização do modelo, valores próximos a 1 indicam risco alto de recaída e próximos de 0 indicam baixo risco de recaída. O ponto de corte adotado para diferenciar alto de baixo risco foi 0.5 e desta forma vai-se considerar apenas dois tipos de risco, alto ou baixo.

Nas implementações do próximo capítulo, as variáveis foram normalizadas, por razões numéricas no intervalo [-1;1].

Na Rede implementada usou-se o algoritmo de Regularização Bayesiana (MacKay, 1992) (vide Anexo B, seção B.2.2.1.1). Neste algoritmo, assume-se que os parâmetros da Rede são variáveis aleatórias com distribuições especificadas. Os parâmetros de regularização são variâncias desconhecidas associadas a estas distribuições e pode-se calcular estes parâmetros usando então técnicas estatísticas. Portanto o modelo não é especificado de uma forma arbitrária.

O aprendizado ou treinamento de uma rede neural tem tipicamente por objetivo reduzir a soma dos quadrados dos erros (Foresee & Hagan, 1977):

$$\hat{\mathbf{y}} = \arg \min_{\mathbf{y}} Q_1(\mathbf{y}) = \arg \min_{\mathbf{y}} \sum_{t=1}^N (y_t - G(x, \mathbf{y}))^2 \quad (2.1)$$

Neste ponto y_i é a saída alvo da rede e $G(x, \mathbf{y})$ é a saída estimada pela rede. Assim como outros modelos flexíveis não lineares, as Redes Neurais podem sofrer de *overfitting*. Este problema ocorre quando utilizamos um número excessivo de neurônios na camada oculta, que levarão a uma perda da capacidade de generalização (fora da amostra). Em contrapartida, se reduzirmos o número de neurônios em excesso, teremos perda da capacidade de aproximar o processo gerador dos dados (Medeiros e Pedreira, 2001).

Atualmente, diversas metodologias são utilizadas para solucionar o problema de *overfitting* (Haykin, 1998). Nesta tese utilizar-se-á o procedimento

desenvolvido por Mackay (1992) chamado de Regularização Bayesiana, que consiste em adicionar um termo de penalização (regularização) à função objetivo, de forma que o algoritmo de estimação faça com que os parâmetros irrelevantes converjam para zero, reduzindo assim o número de parâmetros efetivos utilizados no processo.

Seguindo a notação utilizada por Medeiros e Pedreira (2001), o problema de estimação passa a ser definido como:

$$\hat{\mathbf{y}} = \arg \min_{\mathbf{y}} Q_T(\mathbf{y}) = \arg \min_{\mathbf{y}} \sum_{t=1}^N (\mathbf{h}Q_1(\mathbf{y}) - \mathbf{f}Q_2(\mathbf{y}))^2 \quad (2.2)$$

onde a função de penalização é a soma do quadrado dos parâmetros (\mathbf{a}, \mathbf{g}) :

$$Q_2(\mathbf{y}) = \sum_{h=0}^H \mathbf{a}_h^2 + \sum_{h=0}^H \sum_{i=0}^I \mathbf{g}_{hi}^2 \quad (2.3)$$

O problema de regularização é otimizar a função objetivo de forma a encontrar valores para os parâmetros \mathbf{f} e \mathbf{h} . Este problema de otimização requer o cálculo da matriz Hessiana como pode ser visto em (Mackay, 1992). O algoritmo desenvolvido por (Foresee & Hagan, 1997) propõe a aproximação da matriz Hessiana pelo algoritmo de Levenberg-Marquardt (1963), reduzindo o custo computacional.

Todos os modelos utilizados nesta tese tiveram como arquitetura da rede neural uma camada de entrada, uma camada escondida com dez neurônios e uma camada de saída com um neurônio. A função de ativação sigmóide foi utilizada em todos os neurônios, inclusive no de saída. Os pesos e os *bias* foram inicializados através do algoritmo de Nguyen-Widrow (1989).

Uma ferramenta padrão da estatística conhecida como validação cruzada fornece um princípio orientador atraente (Stone, 1974, 1978). Primeiramente o conjunto de dados disponível é dividido aleatoriamente em um conjunto de treinamento e em um conjunto de teste. O conjunto de treinamento é dividido adicionalmente em dois subconjuntos distintos: subconjunto de estimação (usado para selecionar o modelo) e subconjunto de validação (usado para testar ou validar

o modelo). O desempenho de generalização do modelo selecionado é medido sobre o conjunto de teste, que é diferente do subconjunto de validação.

Quando o número de exemplos rotulados disponíveis, N , for severamente limitado, pode-se usar a forma extrema de validação cruzada múltipla conhecida como o “método deixe um de fora” (*leave-one-out method*, Haykin, 1998 e Bishop, 1995⁴). Neste caso, $N-1$ exemplos são usados para treinar o modelo, e o modelo é validado testando-o sobre o exemplo deixado de fora. O experimento é repetido para um total de N vezes, cada vez deixando de fora um exemplo diferente para a validação. O erro quadrado na validação é então a média sobre as N tentativas do experimento (Haykin, 1998, Bishop, 1995). No caso do banco de dados desta tese fica mais do que justificado o uso do critério de *leave-one-out*.

Desta forma, dos 144 pacientes retira-se 1 e treina-se a Rede com os 143 casos restantes (conjunto de treinamento). Em seguida, a performance da Rede é verificada apenas com a amostra retirada. Este procedimento é repetido para os 144 pacientes do conjunto de dados. O desempenho do modelo é avaliado pela média do erro quadrado obtido na validação sobre todas as tentativas do experimento. Este resultado é uma excelente aproximação da medida fora-da-amostra (Haykin, 1998 e Bishop, 1995).

2.2.2.1.

Um Refinamento no Modelo: Inversão Dirigida da Saída-Alvo da Rede

É importante notar que objetivo proposto é estimar o risco de recidiva do paciente e não prever se este paciente irá recair ou não. Na fase de treinamento, quando os parâmetros da Rede Neural estão sendo ajustados, usa-se o desfecho real para cada paciente (recidiva ou não) como uma estimativa preliminar do seu risco. Obviamente, este valor de saída usado para treinar a Rede é apenas uma indicação preliminar do risco, mas tem de ser usada por ser a única informação relativa ao risco, independente das variáveis de entrada, que se tem em mãos.

É interessante considerar três situações entre os pacientes que não recaíram e que foram estimados como de risco alto pelo modelo (saída maior 0.5): a) Os pacientes realmente são de “alto risco” e não recaíram porque receberam um tratamento adequado; b) Os pacientes realmente são de “alto risco” e poderão

⁴ Anexo B, seção B.2.2.1.2.

recair no futuro; c) Erro do modelo, que estaria estimando como alto risco um paciente que na verdade não o é. A situação (b) evidentemente não pode ser prevista e será inevitavelmente misturada a situação (c). Desta forma, pode-se considerar que o erro neste caso será sempre um limite superior do verdadeiro, ou seja, o erro real é no máximo o valor estimado. Claramente a situação (a) também não pode ser em principio prevista, mas é possível identificar alguns pacientes para os quais se tem uma forte indicação desta ocorrência. Estes pacientes, que provavelmente são de alto risco, mas não recaíram, inspiraram uma nova modificação no procedimento de estimação do risco, apresentada a seguir. Como será visto no capítulo de resultados esta modificação vai produzir uma melhora significativa na performance.

Basicamente, o que se vai considerar é que um paciente que não recaiu mas: 1) foi avaliado pelo critério BFM95 como risco alto; e 2) foi avaliado pelo modelo claramente⁵ como risco alto; apresenta alta probabilidade de ser efetivamente de risco alto. Nestes casos, inverte-se a ‘saída-alvo’ da Rede e volta-se a estimar os parâmetros da mesma, isto é, torna-se a treinar a Rede.

O quadro 06 a seguir exemplifica a inversão da saída alvo da rede:

Quadro 06: Inversão da Saída Alvo da Rede

Risco Modelo	Risco BFM95	Saída Alvo
Alto	Padrão Intermediário Alto	Manter Manter Inverter

Este procedimento, que é repetido até que não existam mais pacientes que satisfaçam as condições 1 e 2, encontra-se ilustrado na figura 02 a seguir.

⁵ Entende-se por ‘claramente’ avaliado como risco alto pelo modelo situações onde a saída da Rede Neural apresenta um valor alto, maior que 0.7.

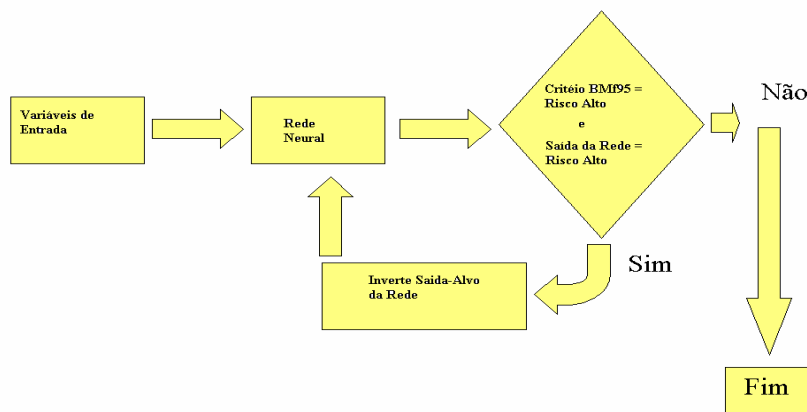


Figura 02. Representação da Inversão da Saída-Alvo da Rede

2.2.2.2.

Um Segundo Refinamento: Considerando o Acompanhamento do Tempo de Tratamento Decorrido

Nesta seção propõe-se mais uma etapa no refinamento do método ao se levar em conta o tempo que um paciente está sob tratamento. Usa-se então para ajuste dos parâmetros, treinamento da Rede, apenas os pacientes com mais de 36 meses de tratamento e os que recidivaram em menos de 36 meses. Desse modo, dentre os pacientes com menos de 36 meses de acompanhamento, estão incluídos todos aqueles pacientes que apresentaram recidivas precoces. Na fase de testes, os pacientes que foram retirados (os que têm menos de 36 meses de tratamento e não recaíram) são testados não pelo método *leave-one-out*, mas sim com base nos resultados de treinamento dos pacientes mais antigos. Desta forma se esta utilizando a partição clássica para medir performance fora-da-amostra. Para os pacientes com mais de 36 meses de tratamento o teste para medida de desempenho fora-da-amostra continua sendo realizada pelo método *leave-one-out*. Em todos os casos o procedimento apresentado na seção anterior é usado também aqui.