Redes Neurais Artificiais (Parte Final)

Uma abordagem revolucionária em Inteligência Artificial

Antonio Augusto Gorni

PROBLEMAS MAIS COMUNS

Talvez a falha mais freqüente que ocorra durante a implementação de redes neurais seja seu mau desempanho durante sua avaliação com o sub-conjunto de teste, apesar do proceso de treinamento ter sido aparentemente bem sucedido. O que ocorre neste caso é que a rede neural memorizou os dados fornecidos ao invés de estabelecer relações gerais entre eles. Assim, ao se apresentar à rede assim treinada dados novos, ela simplesmente não consegue calcular as respostas certas. Este caso é muito semelhante ao de um aluno que decorou muito bem os exemplos de operações aritméticas em sua cartilha, que fatalmente não irão constar em seu exame final!

As prováveis causas deste problema são:

- Incompatibilidade entre os sub-conjunto utilizados nas fases de treinamento e teste;
- Sub-conjunto de treinamento mau dimensionado e/ou concebido;. Rede neural super-dimensionada;
- Apresentação seqüencial de dados à rede neural durante sua fase de treinamento;
- Excesso de treinamento.

As quatro primeiras causas implicam em falta de condições para que a rede neural consiga generalizar as relações entre as variáveis de entrada e saída. Nas demais ocorre o contrário: há excesso de recursos para tal. Em conseqüência, ela passa a memorizar os dados.

As contra-medidas para esses problemas são simples, mas eventualmente muito trabalhosas devido à falta de embasamento teórico para se estabelecer diretrizes seguras:

- Assegurar a qualidade dos dados fornecidos;
- Repartir adequadamente os dados entre os sub-conjunto de treinamento e teste;
- Garantir que haja no sub-conjunto de dados para treinamento uma participação eqüitativo de todos os casos que ocorrem no universo a ser modelado;
- Dimensionar criteriosamente a rede neural, efetuando-se testes com várias arquiteturas;
- Efetuar apresentação em ordem aleatória dos registros de dados à rede neural durante sua fase de treinamento;
- Evitar excesso de treinamento, interrompendo-o a intervalos regulares e avaliando-se o grau de ajuste conseguido até então, utilizando-se o sub-conjunto de teste. Desse modo, é possível determinar a partir de que ponto a rede passou a ficar demasiadamente treinada.

Há casos em que a rede neural falha já na fase de treinamento, ou seja, ela não converge para um estado estável e eficaz do ponto de vista preditivo. As prováveis causas para essa perturbação podem ser:

- Rede neural subdimensionada;
- Coeficiente de aprendizado inicial muito alto;
- Dados com magnitude incompatível com a função de ativação dos neurônios;

As soluções sugeridas neste caso são:

- Dimensionamento correto da rede neural, procurando-se testar diversas arquiteturas;
- Redução na magnitude do coeficiente de aprendizado α, promovendo-se seu declínio ao longo da fase de treinamento, de modo a evitar que os valores dos pesos das sinapses atinjam valor máximo e trunquem a capacidade de aprendizado da rede;
- Se necessário, comprimir a faixa de magnitude dos dados, e garantir que seus valores sejam adequados ao processamento pela rede neural.

Outro problema sério no desenvolvimento das redes neurais é o excessivo tempo que por vezes é requerido em seu treinamento. Tal problema, felizmente, vem sendo resolvido através do desenvolvimento de algoritmos de

aprendizado cada vez mais precisos e rápidos, e com a disseminação de microcomputadores cada vez mais poderosos [1,3,9].

PROGRAMAS DISPONÍVEIS

Há informação suficiente na literatura para se desenvolver bons programas para a implementação de redes neurais. De fato, um trabalho nesse sentido é um exercício acadêmico excelente para se entender toda a teoria que rege esta técnica.

Contudo, do ponto de vista estritamente utilitário, essa abordagem é muito questionável quanto à sua eficácia. Atualmente há dezenas de programas comerciais destinados ao desenvolvimento de redes neurais, e certamente surgirão muitos outros no futuro. Eles apresentam as seguintes vantagens:

- Boa documentação sobre a técnica;
- Interface gráfica amigável com o usuário;
- Disponibilidade em vários ambientes: microcomputadores IBM (DOS/Windows), Macintosh, Workstations (Sun, HP, DEC), "Mainframes";
- Uso automático de recursos especiais do microcomputador, cujo acesso não é elementar ao usuário comum, como a expansão de memória e o co-processador aritmético nos microcomputadores IBM;- Incorporação de várias topologias de redes neurais e as respectivas estratégias para aprendizado;
- Alguns produtos dimensionam e treinam a rede neural de forma automática;
- Uso de algoritmos de aprendizagem extremamente rápidos. Alguns são até considerados propriedade industrial e, portanto, protegidos por patentes;
- Disponibilidade de todo um ferramental para monitorar o que se passa na rede neural durante sua fase de treinamento, permitindo ainda intervenções em neurônios ou sinapses selecionados;
- Possibilidade de se traduzir a rede neural treinada em código-fonte (em linguagem C, por exemplo), gerando sub-rotinas que podem ser chamadas a partir de programas externos;
- Implementação de sistemas que tenham capacidade de se auto-treinar periodicamente "on-line", o que é particularmente útil em redes neurais que controlam processos industriais.

Assim, é pouco provável que programas para desenvolvimento de redes neurais elaborados por usuários não-especializados venham a se igualar em termos de desempenho e custo com programas comerciais similares. Contudo, a multiplicidade de opções disponíveis no mercado requer que se faça uma análise cuidadosa para que se selecione o produto mais adequado para um dado usuário.

No tocante à configuração de microcomputador necessária, velocidade de processamento nunca será demais. Eventualmente, em aplicações críticas, o tempo de aprendizado poderá chegar alguns dias, pois esse processo efetua cálculos de forma intensiva. No caso de microcomputadores da linha IBM, a configuração mínima deverá ser um AT-286 com co-processador aritmético 80287. Uma expansão de memória de pelo menos 1 MB também é um requisito mínimo.

UM EXEMPLO SIMPLES: cálculo de $y = \sqrt{x}$

Nada melhor que um exemplo simples para se assimilar os conceitos expostos até o momento. A título de exercício, pode-se desenvolver uma rede neural para se calcular a raiz quadrada de um número, ou seja, $y = \sqrt{x}$. Certamente há outros métodos bem mais simples e precisos para se efetuar esse cálculo, mas a simplicidade deste exemplo é bastante oportuna.

O primeiro passo é dimensionar a rede neural requerida. Uma vez que se deve fornecer apenas um dado e se obtém apenas um resultado, obviamente deve haver um neurônio na camada de entrada e outro na camada de saída da rede neural. A presença de uma camada oculta é vital, uma vez que redes neurais com um total de três camadas são ideais para o cálculo de funções contínuas. De acordo com o teorema de HECHT-KOLMOGOROV, a camada oculta deve conter $2 \cdot 1 + 1 = 3$ neurônios. Conforme já foi explicado anteriormente, é recomendável que seja adicionado um neurônio unitário de "bias" na camada de entrada para melhorar a precisão do modelamento pela rede neural. A figura 4 mostra esquematicamente a configuração desta rede neural específica.

A seguir, deve-se estabelecer os conjuntos de dados a serem utilizados durante as fases de treinamento e teste. A rede neural deve ser treinada para efetuar cálculos dentro de uma faixa de magnitude limitada. Assim, por exemplo, pode-se gerar um arquivo de dados global que contenha cem dados, de 1 a 100, com espaçamento unitário entre eles, e a respectiva raiz quadrada. Logo, a rede neural obtida será válida para $1 \leq x \leq 100$.

Desses dados, deve-se separar aleatoriamente dez (10% do total), por exemplo, para formar o arquivo de dados a ser utilizado na fase de teste após o treinamento, que avaliará o grau de precisão da rede neural treinada. Obviamente, os noventa elementos restantes constituirão o conjunto que será utilizado na etapa de treinamento.

O programa de treinamento proposto para esta rede neural foi o seguinte:

a) Início do aprendizado, sendo aplicadas 5000 iterações à rede neural;

b) Avaliação da rede neural treinada, comparando-se os resultados calculados a partir do sub-conjunto de teste com os valores reais da função $y = \sqrt{x}$, anotando-se o coeficiente de correlação r de Pearson. Este parâmetro define quantitativamente o grau de ajuste dos valores calculados pela rede neural em relação aos dados reais.

c) Continuação da fase de aprendizado, sendo aplicadas mais 5000 iterações.

d) Nova avaliação da rede neural. Se o valor de r agora obtido for maior que o anterior, ou seja, o nível de aprendizado melhorou em relação ao estágio anterior,

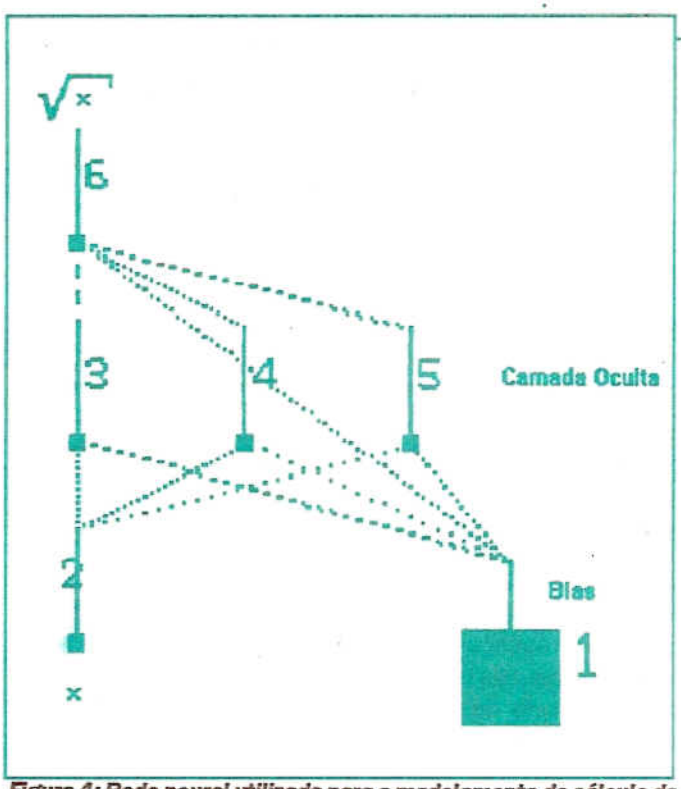

Figura 4: Rede neural utilizada para o modelamento do cálculo da função $y = \sqrt{x}$.

voltar ao passo (c) para refinar ainda mais a rede neural. Caso o valor de r for igual ao anterior, interromper o treinamento. Se, ao contrário, ele for menor, ocorreu excesso de treinamento, ou seja, a rede neural começou a memorizar os dados fornecidos ao invés de estabelecer uma relação entre eles. Neste caso, deve-se reinicializar o processo de aprendizado e aplicar o número de iterações correspondente ao máximo valor de r obtido.

Neste caso específico, a rede neural convergiu após 30.000 iterações de aprendizado, apresentando r igual a 0,9983. Fazendo-se x = 36, por exemplo, a rede neural calculará como sua raiz quadrada o valor de 5,942, ou seja, um erro de 0,96% em relação à resposta certa. A figura 5

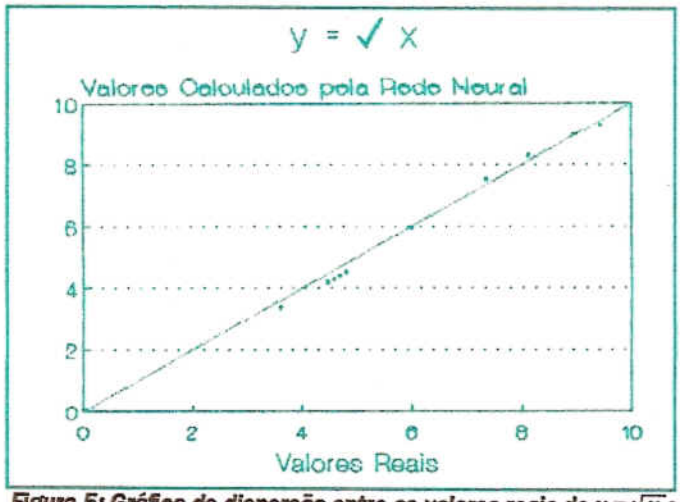

Figura 5: Gráfico de dispersão entre os valores reais de $y = \sqrt{x}$ e os calculados pela rede neural.

mostra o ajuste gráfico entre os resultados reais e o calculado pela rede neural ora desenvolvida.

Um ponto que deve ficar bem claro é que a fase mais complexa e demorada no desenvolvimento de uma rede neural é seu treinamento. De fato, a execução da rede neural já treinada é simples. A seqüência abaixo mostra os cálculos efetuados na rede neural aqui obtida, considerando-se que ela tinha como função-transferência a tangente hiperbólica:

a) Conversão do valor de x dentro da faixa operacional da rede neural. No caso de redes que usem a função tangente hiperbólica como função transferência, os valores que a rede neural processa devem estar dentro da faixa entre −1 e +1:

$$x' = \frac{x}{50.5} - 1 \qquad (2)$$

onde x' é o valor atribuído ao neurônio de entrada.

b) Cálculo do valor dos neurônios da camada oculta:

$$n_1 = Tanh\,(0{,}2909 + 0{,}5956x') \qquad (3)$$

$$n_2 = Tanh\,(-0{,}3587 - 0{,}7611x') \qquad (4)$$

$$n_3 = Tanh\,(-0{,}610 - 0{,}4313x') \qquad (5)$$

Note-se que a primeira parcela dentro do argumento da função Tanh se refere ao peso relativo ao neurônio de "bias".

c) Cálculo do valor do neurônio de saída, y':

$$y' = Tanh\,(-0{,}1285 + 0{,}524n1 - 0{,}6317n2 - 0{,}3904n3) \qquad (6)$$

d) Conversão do valor do neurônio de saída y' ao valor real y, considerando que, por definição, y' deve variar entre −0,8 e 0,8:

$$y = 5{,}5 + 4{,}5 \qquad (7)$$

Este exemplo prova que a execução da rede neural é simples e direta, ou seja, não requer iterações. A grosso modo, seu cálculo é muito semelhante ao de um polinômio de interpolação complexo, onde as funções de potenciação foram substituídas por tangentes hiperbólicas. O único passo realmente complicado – o treinamento da rede neural – é, na verdade, transparente para o usuário final do modelo pronto.

CONCLUSÕES

A neurocomputação, técnica que emula sistemas nervosos biológicos em programas ou circuitos de computador, foi concebida há mais de cinqüenta anos atrás, mas só recentemente experimentou ampla utilização prática, graças à intensa pesquisa teórica nesse campo e ao enorme desenvolvimento na área dos microprocessadores.

A grande característica das redes neurais é sua capacidade

de aprendizado, ou seja, a possibilidade de estabelecer, de forma precisa, relações complexas entre diversas variáveis numéricas, sem que seja imposto qualquer modelo preconcebido. É uma abordagem revolucionária, que contrasta com a filosofia hoje consagrada para o tratamento de dados, a computação programada, que requer algoritmos rigorosamente detalhados para processá-los. Ela apresenta diversas vantagens em relação a outras técnicas de modelamento e controle, como regressão estatística, sistemas especialistas e, talvez, modelos matemáticos.

Os campos de aplicação para as redes neurais são vastos: análise e processamento de sinais, controle de processos, classificação de dados, reconhecimento de padrões, análise de imagens, diagnósticos médicos, etc. Na área industrial, destacam-se as utilizadas na prevenção de desvios de processo e em sistemas híbridos, associados a técnicas de lógica difusa e sistemas especialistas, para a detecção de problemas de manutenção. Trata-se normalmente de problemas com quantificação matemática difícil, ineficaz ou impossível.

Apesar de ser uma técnica que já encontrou inúmeras aplicações na vida real, ela ainda apresenta alguns aspectos obscuros. A pesquisa sobre elas ainda continua em ritmo febril, e é de se esperar num futuro próximo que seu uso se torne ainda mais fácil, mesmo para usuários inexperientes. Em conseqüência, pode-se prever prever sua ampla disseminação a curto prazo.

BIBLIOGRAFIA

1. HECHT-NIELSEN, R. Neurocomputing. Addison-Wesley Publishing Company, New York, 1991. 433 p.
2. HEIDEPRIEM, J. Stahl und Eisen, 112(9):53-59, September 1992.
3. BLUM, A. Neural Networks in C++. John Wiley & Sons, Inc., New York, 1992, 214 p.
4. HATTORI, S. e outros. Hitachi Review, 41(1):31-38, 1992.
5. ANON. Integrating a Neural Network and Expert System for Machine Diagnosis. Carnegie Group Report, 1991. 4 p.
6. BULSARI, A. & SAXÉN, H. In: SAIS WORKSHOP. Proceedings. SAIS, Stockholm, 1992.
7. BULSARI, A. e outros. Scandinavian Journal of Metallurgy, 21(3):146-150, 1992.
8. REUTER, M.A. & VAN DEN VENTER, J.S.J. A ser publicado em Chemical Engineering Science, 1993.
9. MAREN, A. e outros. Handbook of Neural Computing Applications. Academic Press, San Diego, 1990. 670 p.
10. REISENER, J. e outros. Chemical Engineering Science, 48(6): 1089-1101, 1992.
11. GORNI, A.A. A ser apresentado na I Semana de Conformação da Associação Brasileira de Metalurgia e Materiais, Joinville, Setembro de 1993.
12. OLIVEIRA, J.B. e outros. In: SEMINÁRIO DE LAMINAÇÃO. Anais. Associação Brasileira de Metalurgia e Materiais, Porto Alegre, 1992, 319-335.
13. REUTER, M.A. e outros. Metallurgical Transactions B, 23B(10):643-650, October 1992.
14. LIPPMAN, R.P. IEEE ASSP Magazine, April 1987, 4-22.

ANTONIO AUGUSTO GORNI é Engenheiro da Divisão de Pesquisas Tecnológicas da Companhia Siderúrgica Paulista e Professor-Assistente do Departamento de Metalurgia da Faculdade de Engenharia Industrial - FEI

ANTONIO AUGUSTO GORNI é Engenheiro da Divisão de Pesquisas Tecnológicas da Companhia Siderúrgica Paulista e Professor-Assistente do Departamento de Metalurgia da Faculdade de Engenharia Industrial – FEI.

PC FORÇA

NOVIDADES PARA PC AT

ETERNAN - 5HD
REALMS OF ARKANIA - 3HD
RIDERS OF ROHAN - 2HD
RAGS TO RICHES - 7HD
ONE MUST FALL - 1HD
MONTER BASH - 1HD
ELETRO BOBY - 1HD
JACK IN THE DARK - 1HD
BOBOCOP 3 - 4HD
JURASSIC PARK - 3HD
RETURN TO ZORK - 12HD
COACHES CLUB FOOTBALL - 3HD
PRIVATEER + SPEECH PACK - 10HD
DARK SUN - 5HD
BUZZ ALDRIN RACE - 8HD
SILVERBALL - 1HD
PROSTAR - 4HD
SHADOWORLDS - 2HD
OSCAR - 1HD

GALACTIX - 1HD
COOL WORLD - 1HD
SINK OR SWIN - 1HD
CAESAR DE LUXE - 2HD
SIMFARM - 2HD
HARRIER ASSAULT - 2HD
EIGHT BALL DELUXE - 2HD
MIG 29 FOR FALCON 3.0 - 3HD
MICKEY - 3HD
FOOKEEPER - 5HD
WING COMMANDER 3 - 3HD
THE LOST VICKINGS - 1HD
NIGEK MANSEL - 2HD
LAND OF LORE - 8HD
SYNDICATE - 5HD
STRICK COMMANDER - 8HD
PRINCE OF PÉRCIA 2 - 6HD
STREET FIGHTER 2 - 5HD

E MUITOS OUTROS...

- TEMOS GRANDE VARIEDADE DE APLICATIVOS PARA DOS E WINDOWS
- NO RIO DE JANEIRO ENTREGAMOS EM SUA CASA OU EM SEU TRABALHO
- PARA RECEBER CATÁLOGO GRATUITO, ENVIE DISCO HD

PREÇO DOS JOGOS
(Disco não incluso)
5 1/4 DD ..... CR$ 150,00 POR DISCO
5 1/4 HD ..... CR$ 200,00 POR DISCO

PREÇO DOS DISCOS
5 1/4 DD ..... US$ 0.6 (PARALELO)
5 1/4 HD ..... US$ 1.0 (PARALELO)

POSTAGEM: A CADA 16 DISCOS...CR$ 450,00
- PEDIDOS ACIMA DE **CR$ 8.000,00**, PAGAMENTO EM DUAS VEZES COM CHEQUE PRÉ-DATADO, UM À VISTA E OUTRO PARA 15 DIAS.
- CASO SEJA SUA PREFERÊNCIA, VOCÊ PODE ENVIAR SEUS PRÓPRIOS DISCOS, MAS SÓ ACEITAREMOS DISCOS NOVOS.
- PEDIDOS A CADA 10 DISCOS GRAVADOS, GANHE 2 GRAVAÇÕES (ESCOLHA).

COMO FAZER SEU PEDIDO:
PEÇA POR TELEFONE OU RELACIONE NUMA FOLHA DE PAPEL OS JOGOS QUE DESEJA E A QUANTIDADE DE DISCOS OCUPADOS E ESCREVA SEU NOME, ENDEREÇO, CIDADE, ESTADO E CEP E ENVIE PARA O NOSSO ENDEREÇO.

FORMAS DE PAGAMENTO:
A) CHEQUE NOMINAL E CRUZADO À **ROGÉRIO VICENTE PEIXOTO GOMES**. VOCÊ DEVERÁ ANEXAR O CHEQUE AO SEU PEDIDO.
B) DEPÓSITO EM CONTA: BANCO NACIONAL AGÊNCIA 0001, CONTA 176412 EM NOME DE ROGÉRIO VICENTE PEIXOTO GOMES. ENVIE O COMPROVANTE JUNTO COM O PEDIDO.
C) VALE POSTAL EM NOME DE ROGÉRIO VICENTE PEIXOTO GOMES.

PC FORÇA • CAIXA POSTAL 16048 - RIO DE JANEIRO - RJ - CEP 22222-970 • **TEL.: (021) 265-9265**