

# 1º Congresso Brasileiro de Redes Neurais

Escola Federal de Engenharia de Itajubá  
Itajubá, 24 a 27 de outubro de 1994

## PROCESSAMENTO NEURAL-ADAPTATIVO DE SINAIS

João Batista Destro Filho e João Marcos Travassos Romano

Departamento de Comunicações/FEE - UNICAMP - Caixa Postal 6101  
Cidade Universitária - Campinas - SP - 13081-970  
e-mail: destro@decom.fee.unicamp.br

**RESUMO** - Este trabalho propõe uma abordagem unificada das técnicas de redes neurais e da filtragem adaptativa, com o objetivo de desenvolvê-las simultaneamente. Tal abordagem está baseada na inter-relação existente entre as redes neurais e a filtragem adaptativa, a qual pode ser evidenciada por uma série de analogias matemáticas e conceituais entre os dois campos. Apresentam-se alguns resultados preliminares já alcançados, que consistem na análise matemática do processamento paralelo distribuído associado a uma rede Perceptron multi-camadas e na proposição de um algoritmo alternativo para a adaptação da cascata de filtros adaptativos transversais, cujo desempenho é avaliado experimentalmente para um caso simples.

### 1. INTRODUÇÃO.

As redes neurais correspondem, atualmente, a uma das técnicas mais promissoras em diversas áreas (por exemplo, na robótica [1]), graças às suas propriedades coletivas emergentes. Por outro lado, segundo Amari [1], "a matemática das redes neurais encontra-se ainda em sua infância". De fato, constata-se atualmente uma certa dificuldade de compreensão matemática do processamento paralelo distribuído [2] e da

auto-organização [3], princípios básicos do processamento de informação neural. Isto explica, de certa forma, o caráter empírico do projeto de redes neurais, bem como seu aprendizado complexo e lento.

As técnicas de processamento de sinais (e, dentre estas, a filtragem adaptativa), já são aplicadas industrialmente em diversas áreas, por exemplo, nas telecomunicações e na geofísica [4]. Isto pode ser justificado pelo seu sólido embasamento matemático, que compreende, dentre várias teorias, o controle automático e a teoria estatística do sinal. Entretanto, a filtragem adaptativa também possui algumas limitações, por exemplo, baixo desempenho em aplicações que envolvam ruído não-gaussiano e não-aditivo.

Conclui-se, portanto, que as redes neurais e a filtragem adaptativa correspondem a técnicas com vantagens e fraquezas complementares. De um lado, as propriedades coletivas emergentes neurais, de difícil análise matemática; de outro, os filtros adaptativos com seu sólido formalismo matemático e limitações práticas.

O objetivo deste trabalho é propiciar elementos para a pesquisa unificada das duas áreas, bem como evidenciar como esta pode ser útil para ambas. Tal pesquisa é aqui denominada de "Processamento Neural-Adaptativo de Sinais". Corresponde

Para analisar simultaneamente, de forma teórica e experimental, estruturas matematicamente equivalentes de redes neurais e de filtros adaptativos. Desta forma, objetiva-se alcançar contribuições para ambas as áreas, que representem resultados alternativos àqueles da pesquisa tradicional.

Na seção 2, apresentam-se várias analogias conceituais e matemáticas, que evidenciam a existência de uma inter-relação entre as redes neurais e a filtragem adaptativa. Na seção 3, analisa-se simultaneamente um Perceptron multi-camadas linear, parcialmente interconectado, e a cascata de filtros adaptativos, propondo-se uma expressão matemática para o processamento paralelo distribuído desta rede e um algoritmo alternativo para o treinamento da cascata. Tais hipóteses são verificadas experimentalmente na seção 4. As principais conclusões deste trabalho são apresentadas na seção 5.

2. A INTER-RELAÇÃO EXISTENTE ENTRE AS REDES NEURAIS E A FILTRAGEM ADAPTATIVA.

ANALOGIA 1: Filtro não-linear adaptativo e neurônio Perceptron.

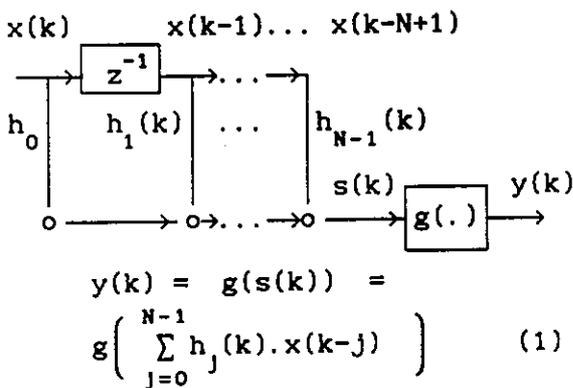
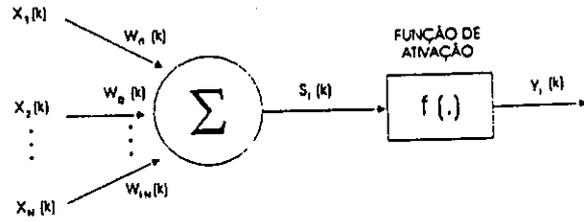


Figura 1: Filtro não-linear adaptativo - ordem N.



$$y_1(k) = f(s_1(k)) = f\left(\sum_{j=1}^N w_{1j}(k) \cdot x_j(k)\right) \quad (2)$$

Figura 2: Neurônio Perceptron.

A grandeza g(.) (fig. 1) representa a não-linearidade do filtro adaptativo. A comparação das figs. 1-2 e das eqs. (1) - (2) permite concluir que as duas estruturas são análogas, sendo possível estabelecer uma equivalência entre g(.) e a função de ativação f(.), bem como entre os coeficientes h\_j do filtro e os pesos sinápticos w\_{1j} do neurônio.

ANALOGIA 2: Equalizador adaptativo de Bussgang e neurônio Perceptron treinado de forma não-supervisionada.

Esta analogia representa um caso especial da analogia 1. O equalizador de Bussgang [4] corresponde a um filtro não-linear adaptativo, idêntico àquele mostrado pela fig. 1. Sua relação de entrada-saída é exatamente a mesma da eq. (1), sendo que a não-linearidade g(.) será aqui denominada de "estimador de Bussgang".

O propósito de aplicação desta estrutura está mostrado na fig. 3.

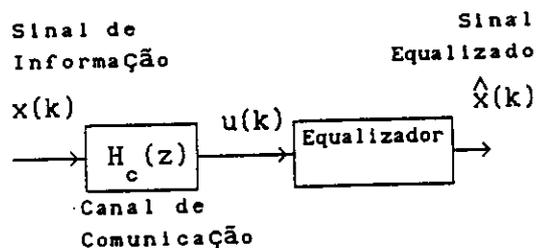


Fig. 3: Operação de um equalizador.

O objetivo é inverter a função de transferência do canal de comunicação (representada por  $H_c(z)$ ), para gerar a melhor estimativa possível do sinal de informação  $x(k)$  transmitido. Isto é realizado de forma auto-organizativa. Portanto, através da analogia 1, conclui-se que o equalizador de Bussgang corresponde a um neurônio Perceptron treinado de forma não-supervisionada.

Deve-se notar que o estimador de Bussgang pode ser deduzido matematicamente [4] e depende da relação sinal-ruído do sistema de comunicação. Se este parâmetro possuir amplitude reduzida, o estimador de Bussgang assume a forma de uma sigmóide, exatamente como a função de ativação de um neurônio Perceptron.

ANALOGIA 3: Filtro de erro de predição linear treinado pelo algoritmo LMS e neurônio de Kohonen [3].

Estas estruturas estão respectivamente mostradas nas figs. 4 e 5.

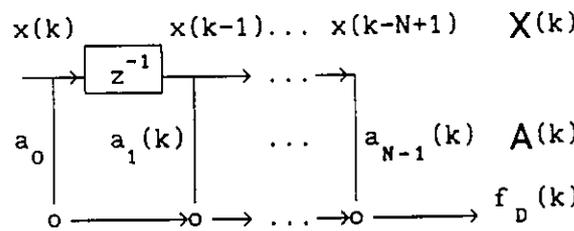


Figura 4: Filtro de erro de predição linear - ordem N.

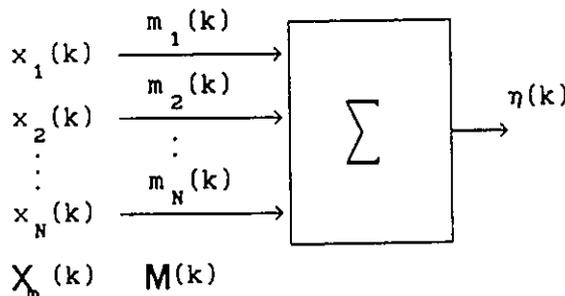


Fig. 5: Neurônio de Kohonen [3].

Apresentam-se a seguir as equações de filtragem e de adaptação de cada estrutura.

Filtro de erro de predição:

$$f_D(k) = \sum_{j=0}^{N-1} a_j(k) \cdot x(k-j) \quad (3)$$

$$A(k+1) = A(k) - \mu \cdot f(k) \cdot X(k) \quad (4)$$

$$X(k)^T = [x(k) \ x(k-1) \ \dots \ x(k-N+1)] \quad (5)$$

$f_D(k)$ : erro de predição linear.

$A(k)$ : vetor de coeficientes.

$\mu$  : passo de adaptação.

$X(k)$ : vetor de dados de entrada.

A eq.(4) corresponde ao algoritmo do gradiente estocástico (ou LMS).

Neurônio de Kohonen:

$$\eta(k) = \sum_{j=1}^N m_j(k) \cdot x_j(k) \quad (6)$$

$$M(k+1) = M(k) - \alpha \cdot \eta(k) \cdot X_m(k) \quad (7)$$

$M(k)$ : vetor de pesos sinápticos.

$\alpha$  : constante real positiva.

As equações de filtragem (3) e (6), bem como as equações de treinamento (4) e (7), são claramente idênticas. Além disso, demonstra-se em [5] a igualdade das condições de convergência dos algoritmos expressos pelas eqs. (4) e (7).

ANALOGIA 4: Cascata de filtros adaptativos transversais e Perceptron multi-camadas (ou MLP).

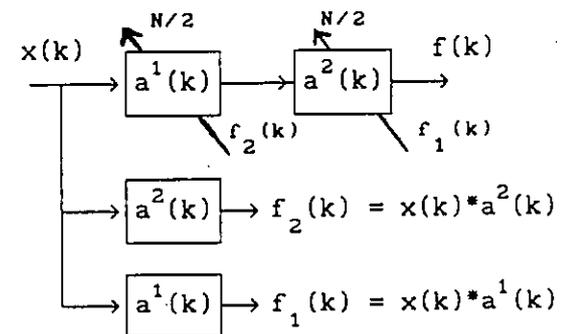


Fig. 6: Filtro de erro de predição linear, ordem N - FORMA CASCATA.

A fig. 6 apresenta dois filtros de erro de predição linear, cada

qual de ordem  $N/2$ , conectados em série. Tal estrutura, denominada "realização cascata", desempenha a mesma função que o filtro de erro de predição linear da fig. 4, de ordem  $N$ , denominado de "forma direta".

O símbolo "\*" representa a operação de convolução discreta, enquanto  $a^i(k)$  é a resposta ao impulso do filtro  $i$ . Os termos  $f_1(k)$  e  $f_2(k)$  são denominados "sinais de gradiente", pois são utilizados no treinamento da cascata através do algoritmo do gradiente estocástico na forma cascata (ou LMSc) [6].

Algoritmo LMSc

$$A_1(k+1) = A_1(k) - \mu \cdot f(k) \cdot F_2(k) \quad (8a)$$

$$A_2(k+1) = A_2(k) - \mu \cdot f(k) \cdot F_1(k) \quad (8b)$$

$A_1(k)$ : vetor de coeficientes do filtro  $i$  ( $i = 1, 2$ ).

$\mu$  : passo de adaptação.

$$F_i(k)^T = [f_i(k-1) \ f_i(k-2) \ \dots \ f_i(k-(N/2))]; \quad i = 1, 2 \quad (9)$$

$$f_i(k) = A_i(k)^T \cdot X_p(k); \quad i = 1, 2 \quad (10)$$

$$X_p(k)^T = [x(k) \ x(k-1) \ \dots \ x(k-(N/2))] \quad (11)$$

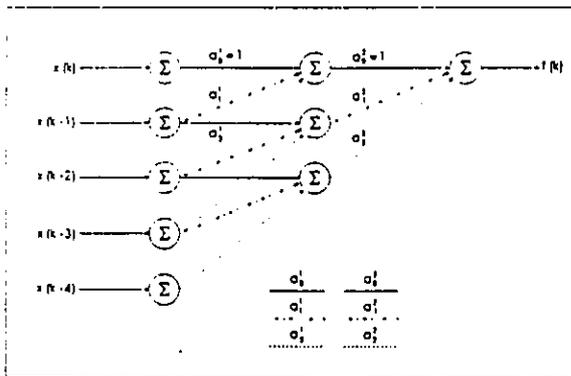


Fig.7: Perceptron multi-camadas linear e parcialmente conectado [6].

A rede neural da fig. 7, proposta em [6], possui função de ativação linear, é parcialmente interconectada e seus vetores de treinamento  $X(k)$  possuem componentes que representam amostras adjacentes de uma série temporal  $x(k)$ :

$$X(k)^T = [x(k) \ x(k-1) \ \dots \ x(k-N)] \quad (12)$$

Obs.:  $N = 4$  na figura 7.

Pode-se demonstrar [6] que as equações de filtragem e do algoritmo LMSc (eqs.(8a-b)), para a cascata da fig. 6, são respectivamente iguais às equações de filtragem e do algoritmo de retropropagação [2], para a rede neural mostrada na fig. 7.

Ou seja, as figs 6 e 7 constituem estruturas matematicamente equivalentes. Em particular, deve-se notar que a primeira e a segunda camadas do Perceptron multi-camadas da fig. 7 podem ser respectivamente associadas aos filtros 1 e 2 da cascata (fig. 6). Isto porque os pesos sinápticos dos neurônios situados em uma mesma camada são iguais aos coeficientes do respectivo filtro associado.

ANALOGIA 5: Sinapse de Hebb e algoritmo LMS para a predição linear.

A fig.8 apresenta em detalhe a sinapse biológica existente entre a terminação  $j$  do axônio de um neurônio A e o dendrito de seu vizinho B.

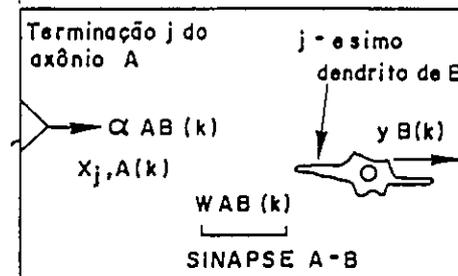


Fig. 8: Sinapse de Hebb.

A lei de Hebb pode ser assim enunciada [7]: "A sinapse A-B será cada vez mais eficiente à medida que um processo simultâneo de excitação intensa, proveniente de A, gerando atividade neural de saída intensa em B, persistir repetidamente".

"Eficiência" significa aqui que, por menor que seja a amplitude da excitação proveniente do neurônio A, esta influencia cada vez mais a atividade neural de saída de B, em

relação às demais excitações externas que atingem o neurônio B.

Definem-se agora algumas grandezas matemáticas (vide fig. 8).

$w_{AB}(k)$  : eficiência da sinapse A-B no instante k.

$x_{j,A}(k)$  : amplitude da excitação externa, proveniente de A.

$y_B(k)$  : amplitude da atividade neural de saída de B.

$\alpha_{AB}$  : coeficiente de plasticidade da sinapse A-B.

$\alpha_{AB}$  é uma constante real que evidencia o quanto os parâmetros bioquímicos, controladores do mecanismo de sinapse entre os neurônio A e B, são susceptíveis a variações.

Desta forma, a lei de Hebb pode ser expressa matematicamente pela seguinte equação [3,5]:

$$w_{AB}(k+1) = w_{AB}(k) + \alpha_{AB} \cdot x_{j,A}(k) \cdot y_B(k) \quad (13)$$

Reescrevendo o algoritmo do gradiente estocástico, aplicado à predição linear (eq. (4)), considerando-se apenas um único coeficiente:

$$a_j(k+1) = a_j(k) - \mu \cdot x_j(k) \cdot f(k) \quad (14)$$

As equações (13) e (14) são claramente análogas.

### 3. ANÁLISE MATEMÁTICA DO PROCESSAMENTO PARALELO DISTRIBUÍDO (ou PDP)

Analizam-se nesta seção as estruturas mostradas nas figs. 6 e 7, no contexto da analogia 4.

O processamento paralelo distribuído (PDP) do Perceptron multi-camadas da fig.7 corresponde à interação entre as duas camadas desta estrutura. Ou seja, de acordo com a analogia 5, o PDP corresponde à interação dos dois filtros da cascata da fig.6. Tal interação, por sua vez, é função dos sinais de gradiente. Isto pode ser concluído a partir da análise das equações do algoritmo LMSc (eqs. (8)). Por exem-

plo, reescrevendo a eq. (8a), tem-se que:

$$\Delta A_1(k) = A_1(k+1) - A_1(k) \quad (15)$$

$$\Delta A_1(k) = -\mu \cdot f(k) \cdot F_2(k) \quad (16)$$

De acordo com a eq. (16), a variação nos coeficientes do filtro 1 é diretamente proporcional ao sinal de gradiente 2.

Denotando o PDP da rede neural pelo sinal  $p(k)$ , uma primeira aproximação para este poderia ser [5]:

$$p(k) \cong f_1(k) + f_2(k) \quad (17)$$

Desde que a estrutura da fig.6 é análoga ao Perceptron multi-camadas, conclui-se que o PDP constitui elemento intrínseco da cascata de filtros adaptativos. Desta forma, considerando que este sistema é variante no tempo e que realiza PDP, pode-se considerar que a saída da cascata é função não apenas da convolução discreta entre a entrada  $x(k)$  e as respostas ao impulso dos filtros 1 e 2, como também de  $p(k)$ . Neste contexto, uma primeira aproximação para a saída da cascata poderia ser [5]:

$$f_p(k) \cong f(k) + p(k) \quad (18)$$

Onde  $f_p(k)$  denota o sinal de saída da fig.6, sob a ótica de redes neurais, a ser considerado durante a adaptação da cascata.

Definindo-se uma função de custo apropriada com base nas eqs. (17) e (18), é possível deduzir um algoritmo alternativo para o treinamento da cascata. Isto é realizado em [5] e tal algoritmo, denominado gradiente estocástico neural na forma cascata (ou NLMS) é apresentado logo a seguir.

$$A_1(k+1) = A_1(k) - \mu \cdot f(k) \cdot F_2(k) - \mu \cdot f(k) \cdot X_p(k) \quad (19)$$

A eq. (18) representa um sistema que não obedece ao princípio da superposição [5], justamente devido

ao termo  $p(k)$ . Desta forma, o PDP pode ser considerado uma espécie de não-linearidade sistêmica, independentemente da função de ativação.

4. RESULTADOS DE SIMULAÇÃO

Com o objetivo de avaliar a validade das aproximações das eqs. (17)-(18), simularam-se o filtro direto da fig.4 (treinado pelo algoritmo LMS - eq.(4) -) e a cascata da fig.6, para  $N = 4$  (treinada pelo algoritmo LMSc - eqs. (8a-b) - e pelo NLMS - eq.(19)-). Tais estruturas foram utilizadas para a predição linear de um processo estocástico auto-regressivo (AR), obtido pela filtragem de um ruído branco gaussiano por uma estrutura recursiva de ordem 4.

A fig. 9 apresenta as curvas de erro quadrático médio para o três algoritmos simulados. Tais curvas correspondem à média para vários modelos AR e para várias condições iniciais impostas ao sistema. A tabela 1 apresenta o desvio padrão  $s(.)$  da grandeza IT, definida como sendo a quantidade média de iterações para o algoritmo considerado convergir.  $s(IT)$  caracteriza, portanto, a dependência do algoritmo considerado relativamente à inicialização do sistema. Quanto menor  $s(IT)$ , menor esta dependência.

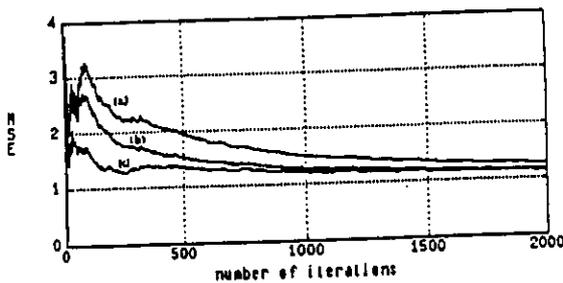


Fig. 9: Dinâmica dos algoritmos. (a) LMSc; (b) NLMS; (c) LMS.

TABELA 1

Algoritmo	LMS	LMSc	NLMS
$s(IT)$	248	518	439

O algoritmo NLMS, além de conver-

gir mais rapidamente que sua versão original (o LMSc), é menos dependente das condições iniciais, de acordo com a tabela 1.

5. CONCLUSÃO

Evidenciou-se neste trabalho a existência de uma inter-relação entre as redes neurais e a filtragem adaptativa. Através do estudo de uma das analogias apresentadas, foi possível, simultaneamente, analisar o processamento paralelo distribuído de um Perceptron multi-camadas linear e propor um algoritmo alternativo para o treinamento da cascata de filtros adaptativos, o qual se mostrou mais eficiente que sua versão original para a simulação de um caso simples. Estes resultados preliminares evidenciam novos horizontes através da pesquisa conjunta de redes neurais e da filtragem adaptativa.

6. REFERÊNCIAS

- [1] S.-I. Amari; "Mathematical Foundations of Neurocomputing", Proc. IEEE, vol.78, no. 9, pp. 1443 - 1463, Sep. 1990.
- [2] D.E. Humelhart, J.L. McClelland and the PDP Research Group; "Parallel Distributed Processing: Explorations in the Microstructure of Cognition", vol.1, MIT Press, 1989.
- [3] T. Kohonen; "Self-Organization and Associative Memory", Springer-Verlag, Berlin, 1989.
- [4] S. Haykin; "Adaptive Filter Theory", Prentice-Hall Inc., New Jersey, 1991.
- [5] J.B. Destro Filho; "Base Teórica para o Processamento Neural-Adaptativo de Sinais", tese de mestrado, FEE-UNICAMP, Junho 1994.
- [6] S. Marcos et alli; "A Unified Framework for Gradient Algorithms used for Filter Adaptation and Neural Network Training", Int. J. of Circuit Theory and Applications, vol. 20, pp. 159 - 200, 1992.
- [7] D.O. Hebb; "The Organization of Behavior", John Wiley & Sons, New York, 1949.