

REDES NEURAIS ARTIFICIAIS COM TREINAMENTO BIOLÓGICAMENTE PLAUSÍVEL

Télvio Orrú*

RESUMO

Pesquisas recentes mostram que os modelos conexionistas baseados em neurociência são ~~considerados a próxima geração de Redes Neurais Artificiais (RNAs)~~ (MEDLER, 1998). Nestes estudos, o termo "biologicamente plausível" é empregado para representar a aproximação do "artificial" ao "real". O todo de algoritmos de treinamento biologicamente plausíveis é muito motivador, não somente para restaurar a história e os princípios das Redes Neurais Artificiais, mas principalmente, embora tão pouco se conheça sobre o modelo natural (cérebro), sabe-se que quanto maior a aproximação ao modelo real, maior a possibilidade de se terem ganhos, inclusive em relação à eficiência computacional.

A comparação entre Redes Neurais Artificiais que utilizam algoritmo de treinamento conexionista biologicamente plausível e as Redes Neurais Artificiais convencionais, que utilizam algoritmo de treinamento conexionista biologicamente implausível, como o backpropagation, é a principal motivação deste trabalho.

Palavras-chave: Neurociência, eficiência computacional, treinamento de redes neurais artificiais.

ABSTRACT

Recent research reveals that the connective models based on Neurology are considered the next generation of the Artificial Neural Webs (MEDLER, 1998). In these studies, the expression "biologically plausible" is employed for representing the approach between 'artificial' and 'real'. It's exciting to have a set of algorithm training which is biologically plausible, not only to restore the history and the principles of Artificial Neural Webs, but especially because the larger approach to the real model makes the possibility of earnings bigger, concerning to computational effectiveness.

The comparison between the Artificial Neural Webs that employ biologically plausible algorithms of connective training and the conventional Artificial Neural Webs that employ biologically not plausible algorithms of connective training, as the Backpropagation, is the main motivation of this paper.

Key words: Neurology; computational effectiveness; training of Artificial Neural Webs.

* Mestrando em Engenharia Elétrica (UNICAMP), Pós-graduado em Administração de Empresas (USF), Graduado em Análise de Sistemas (PUC-Campinas); Membro do Grupo de Pesquisa de Automação e Sistemas Inteligentes (UNICAMP); Membro do Grupo de Pesquisa de Sistemas Inteligentes (PUC-Campinas); Professor no Centro Universitário Padre Anchieta e no Colégio Divino Salvador.

1. REDES NEURAIS ARTIFICIAIS

As RNAs tiveram sua origem a partir do esforço de pesquisadores em entender detalhadamente o funcionamento do cérebro humano e então desenvolver modelos matemáticos para representá-lo. A intenção seria de que estas redes operassem de modo similar ao cérebro humano, podendo tomar decisões, aprender e lembrar de forma semelhante (ou melhor) ao cérebro humano.

As principais áreas envolvidas na implementação de RNAs são: estatística, teoria de informação, teoria de sistemas lineares e não-lineares, teoria da computação, álgebra linear, aproximação de funções, processamento de sinais, controle de processos e otimização de sistemas.

As RNAs artificiais são baseadas na interconexão de unidades de processamento simples e similares, denominadas neurônios artificiais.

As RNAs podem apresentar soluções para problemas que possuam multidimensionalidade e variáveis sujeitas a interações não-lineares, desconhecidas ou matematicamente intratáveis (VON ZUBEN; CASTRO, 2003). Além disto, uma interessante característica das RNAs é sua capacidade de “descobrir soluções gerais”. Considerando que tenham aprendido através de um conjunto de exemplos, as RNAs podem produzir saídas corretas através de entradas que não tinham sido apresentadas às mesmas.

O tratamento fornecido por uma RNA através de entradas que geram saídas é feito por meio de treino iterativo, comumente com exemplos de pares de entrada/saída. Um aprendizado é reconhecido como “ótimo” quando o erro médio da rede foi minimizado. A minimização do erro é obtida por meio da modificação dos pesos das conexões, assim pode-se obter a reprodução das saídas desejadas quando uma nova entrada é recebida pela RNA.

Há vários modelos de RNAs e neste trabalho utilizar-se-á perceptron de múltiplas camadas (“MLP”).

Considerando o modelo de rede que irá ser utilizado (“multicamada”), deve-se definir a maneira como a rede será treinada e para isto existem vários algoritmos de treinamento supervisionado, entre eles:

- detalhes deste algoritmo serão comentados no capítulo 2;
- método do gradiente (GRAD) o erro deve apresentar um mínimo em função do parâmetro que o causa;
- gradiente conjugado (GC) efetua a busca local diferenciando-se do backpropagation pelo cálculo dos gradientes e conseqüentemente pelas correções de pesos;
- são variações do algoritmo do gradiente conjugado, diferenciando-se na fórmula usada para atualizar os pesos;
- gradiente conjugado escalonado de Moller (1993) (SCG) desenvolvido para evitar o consumo de tempo para a convergência. A idéia básica é combinar o modelo de Levenberg-Marquardt (descrito na seqüência) com o gradiente conjugado;

- O algoritmo de Levenberg-Marquardt é uma modificação do método de Newton. Este algoritmo faz sucessivas iterações dos parâmetros calculados com os dados experimentais, e com base na resposta obtida localiza o mínimo da superfície;
- O algoritmo de Davidon-Fletcher-Powell (DFP) recebe a denominação de quase-Newton ou métrica variável. Possui métodos de armazenagem, cálculo de derivada e subminimização unidimensional (PRESS; FLANNERY; TEUTOLSKY, 1987);
- O algoritmo de segunda ordem para treinar o backpropagation (BATTITI, 1992). Esse método mostrou ser eficiente para dados gerados, sendo usado em ambas as redes para cálculo da média e da diferença da estimativa;
- O algoritmo de Levenberg-Marquardt A idéia básica deste algoritmo é usar a segunda derivada da função do erro para tentar direcionar ao erro mínimo. Para isto, assume-se que a) todas as derivadas de ordem alta são zero ou próximas a zero; b) cada função de erro é independente das funções de erros dos outros nós da rede.

O treinamento de redes neurais com várias camadas pode ser entendido como um caso especial de aproximação de funções, no qual não é levado em consideração nenhum modelo explícito dos dados (SHEPHERD, 1997).

Não se deve esquecer que alguns dos métodos de treinamento de RNAs citados acima são conhecidos como "Métodos de Segunda Ordem". Estes métodos são considerados a maneira mais eficiente de se fazer o treinamento de redes neurais do tipo MLP (SHEPHERD, 1997). Recorrem a um rigor matemático baseado em modelos de otimização não-linear irrestrita bem definidos, não apresentando assim um vínculo natural com a inspiração biológica inicialmente proposta para as RNAs. (CASTRO, 2003).

Neste trabalho dar-se-á foco ao algoritmo padrão (Backpropagation) e a algoritmos de treinamento biologicamente plausíveis.

2. UM ALGORITMO DE TREINAMENTO BIOLÓGICAMENTE IMPLAUSÍVEL (BACKPROPAGATION)

2.1. CONCEITOS E CARACTERÍSTICAS

O algoritmo *backpropagation* é utilizado amplamente nos dias atuais como o algoritmo conexionista supervisionado mais eficiente computacionalmente.

Basicamente o procedimento do *backpropagation* está descrito nas seguintes etapas:

Etapa 1: Um padrão é apresentado à camada de entrada da rede. A atividade resultante flui através da rede, camada por camada, até que a resposta seja produzida pela camada de saída.

Etapa 2: A saída obtida é comparada à saída desejada para esse padrão particular. Se estas não forem iguais, o erro (diferença entre ambas) é calculado. O erro é propagado de volta a partir da camada de saída até a camada de entrada. Os

pesos das conexões das unidades da camada de saída e das camadas intermediárias vão sendo modificados conforme o erro é retropropagado.

2.2. POR QUE O BACKPROPAGATION É CONSIDERADO BIOLÓGICAMENTE IMPLAUSÍVEL ?

O algoritmo *backpropagation* é considerado biologicamente implausível (CRICK, 1989).

Justificativas:

1. Supondo que o *backpropagation* ocorresse no cérebro, o erro seria propagado através do dendrito do neurônio pós-sináptico para o axônio e então para o dendrito do neurônio pré-sináptico. Isto é improvável (ROSA, 2002).

2. Pesquisas mostram que os pesos sinápticos são modificados para permitir o aprendizado, porém não desta maneira. Acredita-se que os pesos mudam usando apenas informação local da sinapse. Assim o *backpropagation* mostra-se fisiologicamente implausível.

A figura 1 representa, de acordo com O'Reilly, criador do GeneRec (1996), o funcionamento do algoritmo *backpropagation*.

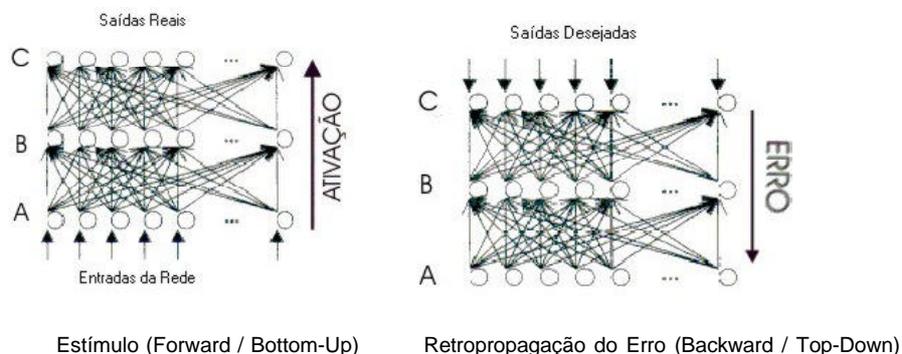


Figura 1(O'REILLY; MUNAKATA, 2000)

3. UM ALGORITMO DE TREINAMENTO BIOLÓGICAMENTE PLAUSÍVEL (GENEREC)

3.1. CONCEITOS E CARACTERÍSTICAS

O algoritmo GeneRec foi desenvolvido em 1996 por O'Reilly, respeitando as

propriedades de um algoritmo de treinamento de rede biologicamente plausível.

É fundamental nos modelos biologicamente plausíveis que a representação adotada seja distribuída (ou seja, o conceito é representado através de várias unidades da arquitetura conexionista).

É possível reescrever as equações do backpropagation para se obter o GeneRec, ou seja, o GeneRec é derivado do backpropagation.

O procedimento de aprendizado do GeneRec é inspirado no algoritmo Recirculation, desenvolvido por Hinton and McClelland em 1988.

O GeneRec faz uso de duas fases: “menos” e “mais” (Figura 2).

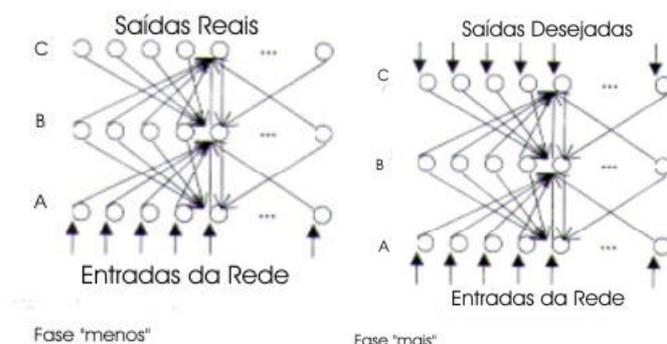


Figura 2 - Duas fases de GeneRec (O'REILLY, 1996):

1. Fase “menos”

Quando as entradas x_i são apresentadas para a camada de entrada “A” existe a propagação desse estímulo para a camada escondida (propagação bottom-up). Existe também a propagação da saída anterior o_k para a camada escondida (propagação top-down). Então a ativação escondida “menos” – fase menos – (h_j^-) é gerada (soma das propagações bottom-up e top-down). A função de ativação é sigmoideal (s). Na equação 1 tem-se a representação desta fase.

Finalmente, a saída real o_k é gerada através da propagação da ativação da camada “menos” para a camada de saída, representada na equação 2.

Nota-se que a arquitetura é bidirecional.

$$h_j^- = s \left(\sum_{i=0}^A w_{ij} \cdot x_i + \sum_{k=1}^C w_{jk} \cdot o_k \right) \quad (1)$$

$$o_k = s \left(\sum_{j=1}^B w_{jk} \cdot h_j^- \right) \quad (2)$$

x_i => entradas
 o_k => saída atual anterior
 w_{ij} => pesos sinápticos (camada entrada e camada escondida)
 w_{jk} => pesos sinápticos (camada escondida e camada saída)
 y_k => saída desejada
 h_j^- => fase negativa
 (O'REILLY, 1996)

2. Fase "mais"

As entradas x_i são apresentadas novamente para a camada de entrada; existe a propagação para a camada escondida (bottom-up). Depois disto, existe a propagação da saída desejada y_k para a camada de saída (top-down). Em seguida a ativação escondida "mais" é gerada, somando essas duas propagações (equação 3).

$$h_j^+ = s \left(\sum_{i=0}^A w_{ij} \cdot x_i + \sum_{k=1}^C w_{jk} \cdot y_k \right) \quad (3)$$

x_i => entradas
 w_{ij} => pesos sinápticos (camada entrada e camada escondida)
 w_{jk} => pesos sinápticos (camada escondida e camada saída)
 y_k => saída desejada
 (O'REILLY, 1996)

Para tornar o aprendizado possível, ocorre a atualização dos pesos sinápticos w , baseados sobre x_i , h_j^- , h_j^+ , o_k e y_k (equações 4 e 5).

$$\Delta w_{jk} = h \cdot (y_k - o_k) \cdot h_j^- \quad (4)$$

$$\Delta w_{ij} = h \cdot (h_j^+ - h_j^-) \cdot x_i \quad (5)$$

h => Taxa de Aprendizagem
 (O'REILLY, 1996)

É importante lembrar que a arquitetura bidirecional em alguns casos – principalmente em problemas simples - pode fazer com que o GeneRec demore mais para convergir que o backpropagation ("sobrecarga"). Porém em problemas cognitivos mais complexos, por exemplo, em Processamento de Línguas Naturais, há compensação, conforme constatado em Rosa (2002).

3.2. POR QUE O GENEREC É CONSIDERADO BIOLÓGICAMENTE PLAUSÍVEL ?

Conforme O'Reilly e Munakata (2000), existem evidências de que o córtex cerebral seja conectado de forma bidirecional e a representação distribuída é vista

como predominante no córtex cerebral. Sendo assim, modelos conexionistas biologicamente plausíveis devem possuir as seguintes características:

Representação distribuída: Através deste item pode-se obter generalização e redução do tamanho da rede, visto que as conexões entre um conjunto de unidades são capazes de suportar um amplo número de diferentes padrões e criar novos conceitos sem alocação de novo hardware;

Competição inibitória: uma espécie de winner-takes-all; aqueles neurônios que estão próximos ao “vencedor” recebem um estímulo negativo destacando desta maneira o neurônio vencedor. Durante uma inibição lateral, um neurônio excita um interneurônio inibitório que faz uma conexão de realimentação sobre o primeiro neurônio (O’REILLY, 1998).

Propagação de ativação bidirecional: As camadas escondidas recebem estímulos das camadas de entrada e das camadas de saída;

Aprendizado de tarefa dirigido a erros: No GeneRec o erro é calculado através da diferença local nas sinapses, baseado em propriedades neurofisiológicas (O’REILLY, 1998), ao contrário do backpropagation, que requer a propagação de sinais de erros.

4. IMPLEMENTAÇÕES

4.1. IMPLEMENTAÇÕES EXISTENTES

Comentam-se na seqüência algumas comparações existentes entre métodos de treinamento (plausível & implausível):

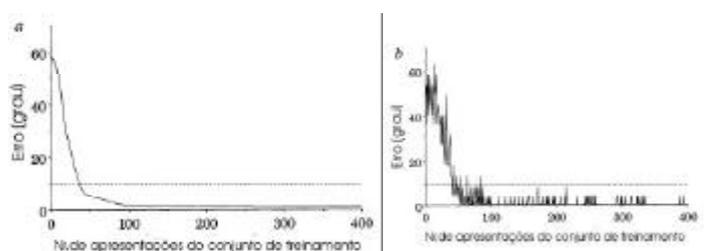
4.1.1 VARIAÇÃO DO $A_{R,P}$

Mazzoni (1991) propõe uma comparação (com aparente interesse em provar que há algoritmos com eficiência *semelhante* e, “talvez”, melhor que o tradicional algoritmo implausível) entre o backpropagation e um algoritmo mais biologicamente plausível, descrito como uma variação do $A_{R,P}$ “associative reward-penalty” de Barto e Jordan (1987). As principais alterações feitas são: todas as unidades da rede recebem um sinal de reforço, cujo valor depende da comparação sobre a saída corrente da rede com a saída desejada; b) usa apenas a informação local para realizar a sinapse ou o ajuste do peso da conexão.

O treino nesta rede é similar ao da arquitetura do modelo de Zipser e Andersen (1989), porém usando um algoritmo de treinamento mais biologicamente plausível. Possui três camadas completamente conectadas com arquitetura feed-forward. A camada de entrada (sensorial) foi modelada de acordo com as características dos neurônios da área 7a¹ utilizada em estudos de Zipser e Andersen (1989). As cama-

das escondidas e a camada de saída consistem em elementos estocásticos binários, que produzem a saída "1" com uma probabilidade dada pela função logística da soma dos pesos de entrada e uma saída "0" em outros casos.

Em um dos experimentos relatados em seu artigo, o autor exhibe graficamente o comportamento da rede neural quando utilizado o A_{r-p} e quando utilizado o Backpropagation para o treinamento. Na figura 3a mostra-se a curva do aprendizado com o uso do backpropagation, enquanto em 3b mostra-se a curva do aprendizado com o uso do A_{r-p} .



(Backpropagation) (A_{r-p})

Figura 3 (MAZZONI et al., 1991)

Nota-se que a curva de aprendizado em A_{r-p} é mais ruidosa do que no backpropagation devido à natureza estocástica das unidades escondidas.

Conforme comentado, a intenção do trabalho deste autor parece ser provar a existência de algoritmos biologicamente plausíveis com grau de satisfação semelhante ao do backpropagation. A conclusão do autor foi que, no mínimo, há uma igualdade na eficiência, com um pequeno ganho quando usado algoritmo biologicamente plausível para treinamento de redes neurais artificiais, como pode ser mostrado na figura 4.

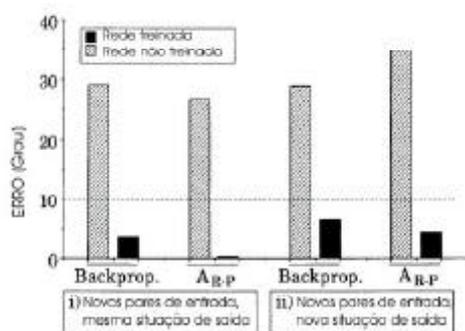


Figura 4 (MAZZONI et al., 1991)

A figura 4 mostra o erro produzido com o uso do backpropagation e com o uso da rede A_{r-p} antes e depois de treinar com 40 novos pares de entrada. Quando apresentado (i) 40 novas entradas randômicas com a mesma situação de saída do conjunto de treinamento, e (ii) 40 novas entradas randômicas para 40 novas saídas randômicas.

É possível notar na situação (i) - na qual as novas entradas randômicas possuem as mesmas situações de saída de outras entradas empregadas anteriormente - que o erro é menor quando utilizado o algoritmo A_{r-p} comparado ao Backpropagation (independente do fato de a rede estar treinada). Já na situação (ii) - na qual se têm novas entradas randômicas com novas situações de saída - o Backpropagation apresenta menor erro (comparado ao A_{r-p}) quando a rede ainda não estava treinada devido a sua natureza matemática. A situação opõe-se quando a rede é treinada, oferecendo menor taxa de erro quando empregado o A_{r-p} . Sendo assim, é possível concluir através deste experimento que a rede treinada com o algoritmo mais biologicamente plausível (neste caso, A_{r-p}) possui tendência à convergência mais rápida.

O autor conclui que a rede A_{r-p} aprendeu a desempenhar o ajuste da tarefa para qualquer exatidão desejada.

4.1.2. BIO-PRED

Rosa (2002) elaborou um sistema (denominado Bio-Pred) que faz a previsão da próxima palavra em uma sentença. Abaixo algumas características e comparações extraídas de seu artigo:

- Bio-Pred1-> GeneRec com 24,000 ciclos de treinamento
- Bio-Pred2-> GeneRec com 4,057 ciclos de treinamento
- Pred-DR -> Backpropagation com 24,000 ciclos de treinamento. Para o treinamento foram apresentadas cem (100) sentenças inteiras diferentes entre si.

Parâmetros utilizados:

Entradas.: Palavras de uma sentença (uma de cada vez)

Saídas.....: A próxima palavra

Taxa Aprendizagem (h) -> 0,25

Erro máximo aceitável (e) -> 0,02

Caso 1: Saída em relação à sentença

The wolf frightened the girl.

Quando informada a palavra wolf os índices de acerto para a previsão de frightened como próxima palavra serão:

Bio-Pred1=> 81,3%

Bio-Pred2=> 62,8%

Pred-DR => 78,0%

Quando inserida a segunda palavra, frightened, os resultados obtidos são (que

a próxima palavra será girl):

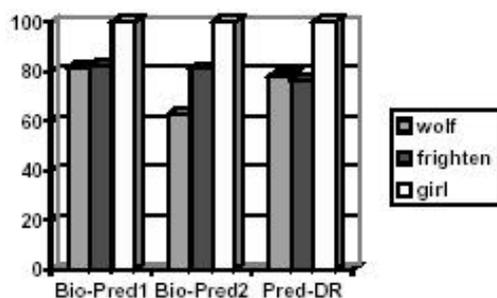
Bio-Pred1 => 82,6%

Bio-Pred2 => 81,2%

Pred-DR => 76,7%

Quando inserida a palavra girl todos retornaram com 100% de exatidão "fim da sentença".

Graficamente:



Comparação entre Bio-Pred1, Bio-Pred2 e Pred-DR (ROSA, 2002)

Caso 2: Saída em relação à sentença

The stone broke the vase.

Quando informada a palavra stone os resultados obtidos são (que a próxima palavra será break):

Bio-Pred1 => 66,1%

Bio-Pred2 => 74,5%

Pred-DR => 84,9%

Neste caso, a eficiência menor no algoritmo usando GeneRec deve-se ao fato de que a palavra break não é facilmente processada, pois admite um, dois ou três operadores (ROSA, 2002).

Quando informada a segunda palavra, break, os resultados obtidos são (que a próxima palavra será vase):

Bio-Pred1 => 91,7%

Bio-Pred2 => 44,8%

Pred-DR => 67,9%

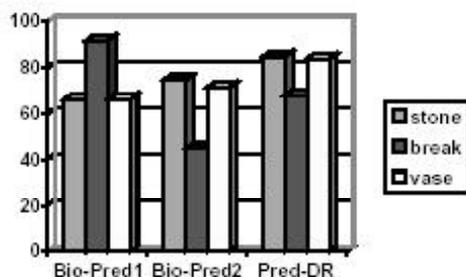
Verbos com possibilidade de diferentes números de argumentos fazem com que sistemas não biologicamente plausíveis tenham melhores resultados. A predição do final da sentença não foi dada com 100% em nenhum dos casos, visto que "break" permite mais do que um operador. Por exemplo, o verbo "quebrar" permite: "Quem quebrou o que" ou "Quem quebrou o que com o que".

Bio-Pred1 => 66,2%

Bio-Pred2 => 71,5%

Pred-DR => 83,8%

Graficamente:



Comparação entre Bio-Pred1, Bio-Pred2 e Pred-DR (ROSA, 2002)

4.1.3. EXPERIMENTOS DO AUTOR

Na seqüência, quatro implementações foram efetuadas na disciplina de Redes Neurais no curso de Mestrado da Universidade Estadual de Campinas, onde compara-se a performance dos algoritmos de treinamento GeneRec e Backpropagation, representados por GR e BP, respectivamente.

4.1.3.1. FUNÇÃO LÓGICA OR

Parâmetros utilizados:

Taxa de Aprendizagem => 0.01

Erro Mínimo => 0.01

Quantidade de Neurônios na camada intermediária => 5

Iteracoes	OR	
	GR	BP
100	0,2566	0,1849
500	0,1722	0,1802
1000	0,1714	0,1735
5000	0,1687	0,1693
10000	0,1675	0,1692
20000	0,1673	0,1689

A tabela representa o MSE / EQM (Erro Quadrático Médio) por iterações nos

dois algoritmos de treinamento citados. Nota-se que o GeneRec obtém melhor performance.

4.1.3.2. FUNÇÃO LÓGICA AND

Parâmetros utilizados:

Taxa de Aprendizagem => 0.01

Erro Mínimo => 0.01

Quantidade de Neurônios na camada intermediária => 5

AND		
Iteracoes	GR	BP
100	0,1714	0,1262
281	ErroMin	
311		ErroMin
500		
1000		
5000		
10000		
20000		

A tabela representa o MSE / EQM (Erro Quadrático Médio) por iterações nos dois algoritmos de treinamento citados. Nota-se que utilizando o GeneRec ocorre convergência na iteração 281, enquanto com o Backpropagation ocorre a convergência na iteração 311.

4.1.3.3. APROXIMAÇÃO DA FUNÇÃO => $\text{SEN}(X) * \text{COS}(2 * X)$ COM 36 AMOSTRAS

Taxa de Aprendizagem => 0,25

Erro Mínimo => 0,001

Quantidade de Neurônios na camada intermediária => 5

Saída Desejada => $\text{sin}(x) * \text{cos}(2 * x)$

Pesos sinápticos iniciais e bias => aleatórios entre -0.1 e +0.1

$\text{sen}(x) + \cos(2 \cdot x)$		
Iterações	GR	BP
100	0,1607	0,1282
500	0,0733	0,1183
1000	0,0731	0,1181
5000	0,0731	0,1170
10000	0,0729	0,1185
20000	0,0729	0,1209

A tabela representa o MSE / EQM (Erro Quadrático Médio) por iterações nos dois algoritmos de treinamento citados. Nota-se que o GeneRec obtém melhor performance. Importante ressaltar que o menor EQM (0,0729) foi obtido com o GeneRec (iteração 10000).

CONCLUSÕES

O foco deste projeto está na comparação entre Redes Neurais Artificiais com treinamento conexionista biologicamente plausível e Redes Neurais Artificiais com treinamento conexionista biologicamente implausível. Em todo o projeto evidenciou-se a maneira pela qual os algoritmos biologicamente plausíveis funcionam, comparando-os com algoritmos biologicamente implausíveis.

Acredita-se que quanto maior a aproximação ao modelo real ("mais biologicamente plausível"), maior a possibilidade de se terem ganhos, inclusive em relação à eficiência computacional.

REFERÊNCIAS BIBLIOGRÁFICAS

- BATTITI, R. (1992). First and second order methods for learning: Between steepest descent and Newton's method. *Neural Computation* 4, pp. 141-166.
- CHALMERS, D. (1990). Why Fodor and Pylyshyn were wrong: the simplest refutation. Proceedings of the Twelfth Annual Conference of the Cognitive Science Society, Cambridge, MA. 340-347.
- CASTRO, L. N. de (1998). Análise e Síntese de Estratégias de Aprendizado para Redes Neurais Artificiais, Dissertação de Mestrado, DCA – FEEC/ UNICAMP, Campinas/SP, Brasil.

- CASTRO, L. N. de; VON ZUBEN, F. J. (1998). Optimized Training Techniques for Feedforward Neural Networks. Technical Report – RT DCA 03/98, FEEC/UNICAMP, Brasil.
- CASTRO, L. N. de. (2003). Fundamentals of Neurocomputing; Technical Report - RT DCA 01/03. In: <http://www.dca.fee.unicamp.br/~lnunes>.
- CASTRO, L. N. de; VON ZUBEN, F. J. (1998). Optimized Training Techniques for Feedforward Neural Networks. Technical Report – RT DCA 03/98, FEEC/UNICAMP, Brasil, 41 p. In: <http://www.dca.fee.unicamp.br/~lnunes>.
- FAHLMAN, S. E. (1988). Faster-learning variations on back-propagation: an empirical study. Connectionist Models Summer School, Morgan-Kaufmann, Los Altos, CA.
- FODOR, J.A.; PYLYSHYN, W. (1988). Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, 28:3-71.
- MAZZONI, P; ANDERSEN, R.A.; JORDAN, M.I. (1991). A more biologically plausible learning rule for neural networks. Massachusetts Institute of Technology, Cambridge, MA 02139.
- MEDLER, D. A. A Brief History of Connectionism . *Neural Computing Surveys* 1, 1998, pp. 61-101.
- MINSKY, M.L.; PAPERT, S. A. (1969). *Perceptrons*. Cambridge, MA. MIT Press1.
- O'REILLY, R.C. (1996). Biologically Plausible Error-driven Learning using Local Activation Differences: The Generalized Recirculation Algorithm. *Neural Computation*, 8:5, pp.895-938.
- _____. (1998). Six Principles for Biologically-Based Computational Models of Cortical Cognition. *Trends in Cognitive Science*, 2, pp. 455-462.
- PARDO, T.A.S; RINO, L.H.M.; NUNES, M.G.V. (2003b). NeuralSumm: Uma Abordagem Conexcionista para a Sumarização Automática de Textos. Anais do IV Encontro Nacional de Inteligência Artificial. Campinas-SP.
- PRESS, W. H.; FLANNERY, B.P.; TEUTOLSKY, S.A.; VETTERLING, W.T. *Numerical recipes the art of scientific computing*. Cambridge University

Press, 1987.

ROHDE, D.L.T.; PLAUT, D.C. (1999). Language Acquisition in the Absence of Explicit Negative Evidence: How Important is Starting Small?. *Cognition* 72, pp. 67-109.

ROSA, J.L.G. (2002). A Biologically Motivated Connectionist System for Predicting the Next Word in Natural Language Sentences. Proceedings of the 2002 IEEE International Conference on Systems, Man and Cybernetics – SMC'2002, October 6-9, 2002, Hammamet, Tunisia.

_____. An Artificial Neural Network Model Based on Neuroscience: Looking Closely at the Brain. In: KURKOVÁ, V.; STEELE N.C.; NERUDA R.; KARNY, M. (Ed.). *Artificial Neural Nets and Genetic Algorithms - Proceedings of the International Conference in Prague, Czech Republic - ICANNGA-2001*. April 22-25, Springer-Verlag, pp. 138-141.

_____. A Biologically Inspired Connectionist System for Natural Language Processing. Proceedings of the VII Brazilian Symposium on Neural Networks (SBRN'02), Brasil.

SCHRAUDOLPH, N.N.; SEJNOWSKI, T.J. (1996). Tempering Backpropagation Networks: Not All Weights are Created Equal. *Advances in Neural Information Processing Systems* 8, MIT Press, Cambridge.

HAYKIN, S. (1999). *Neural Networks - A Comprehensive Foundation*, 2nd edition. Prentice Hall, Upper Saddle River, New Jersey.