

Rede Neural Artificial com Aprendizagem Híbrida para Problemas de Classificação de Padrões

Franciele A. S. Alves¹, Evelyne L. Ferreira¹, Maury M. Gouvêa Jr.¹

¹ Programa de Pós Graduação em Engenharia Elétrica – Pontifícia Universidade Católica de Minas Gerais – Belo Horizonte – MG – Brazil

fran.aalves92@gmail.com, evelynelf@hotmail.com, maury@pucminas.br

Abstract. *Artificial Neural Network is a mathematical model able to store and recover informations by association from partial and noisy knowledge. One of the most important properties of an artificial neural network is the ability to learn either from its environment or with examples. The learning takes place through a set of rules, namely training algorithms, in order to solve a given problem. This paper analyzes the performance of three learning algorithms for one layer neural network: (i) evolutionary algorithm; (ii) descendent gradient; and (iii) hybrid algorithm. The results showed that the hybrid algorithm outperformed the other ones with the highest hit rate in the pattern classification problem.*

Resumo. *Rede Neural Artificial é um modelo matemático capaz de armazenar e recuperar informações por associação a partir de conhecimento parcial ou ruidoso. Uma das propriedades mais importantes de uma rede neural artificial é a habilidade de aprender com seu ambiente ou com exemplos. O aprendizado ocorre por meio de um conjunto de regras, denominado algoritmo de treinamento, para a solução de um determinado problema. Este artigo compara o desempenho de três algoritmos de aprendizagem para rede neural artificial com camada única, a saber: (i) algoritmo evolucionário; (ii) gradiente descendente; e (iii) algoritmo híbrido. Os resultados mostraram que o algoritmo híbrido obteve taxas de acerto superior aos demais métodos no problema de classificação de padrões.*

1. Introdução

As redes neurais artificiais (RNAs) são modelos computacionais, inspirados no neurônio biológico, capazes de reconhecer e classificar padrões a partir de um modelo de aprendizagem baseado no aprendizado humano [Meireles et al. 2003]. O processo de aprendizagem da RNA se dá através de um algoritmo de treinamento que deve ser capaz de ajustar iterativamente os parâmetros que descrevem o comportamento do modelo [Gouvêa 2012]. Na literatura, vários modelos têm sido propostos ao longo das últimas décadas para solução de diferentes problemas. Por exemplo, [Karabatak and Ince 2009] apresentam um sistema de diagnóstico automático para detecção do câncer de mama, utilizando a base de dados *Breast Cancer* [Wolberg and Mangasarian 1990], com base em regras de associação e uma rede neural artificial *multilayer perceptron*. [Shekhawat and Dhande 2011] propõem um classificador de dados utilizando uma rede neural artificial como memória associativa com algoritmo de aprendizagem *Backpropagation* utilizando a base de dados *Iris Flower* [Fisher 1936].

Uma maneira de determinar qual o método de treinamento mais eficiente para a aprendizagem da RNA a ser aplicada em um determinado problema de classificação de padrões é o estudo qualitativo de suas características, focando no que cada método pode fazer de melhor, identificando suas vantagens e limitações.

Este trabalho tem por objetivo avaliar e comparar o desempenho de três algoritmos de aprendizagem para rede neural artificial de camada única, a saber: (i) algoritmo evolucionário; (ii) gradiente descendente; e (iii) aprendizagem híbrida, mesclando (i) e (ii). Para análise do desempenho dos algoritmos apresentados, utilizou-se a tarefa de classificação de padrões utilizando as bases de dados *Iris Flower* e *Breast Cancer Wisconsin*.

O restante deste artigo está organizado como segue. A Seção 2 apresenta o conceito de rede neural artificial. A Seção 3 descreve os algoritmos de aprendizagem utilizados para treinamento da RNA. A Seção 4 descreve os experimentos realizados, assim como, os resultados alcançados para os diferentes métodos de treinamento. Por fim, a Seção 5 apresenta as considerações finais.

2. Rede Neural Artificial Multilayer Perceptron

As RNAs *multilayer perceptron* (MLP) são sistemas de neurônios ligados por conexões sinápticas que são divididos em neurônios de entradas, que recebem estímulos do meio externo; em neurônios internos, responsáveis por interligar os neurônios das camadas de entrada e saída; e em neurônios de saídas, que se comunicam com o exterior [Haykin 2001]. A RNA possui pesos sinápticos, uma junção de somatórios e uma função de ativação.

Um sinal de entrada x_i na entrada de um neurônio I é multiplicado pelo peso sináptico w_{ij} e, após o cálculo, o valor é enviado à entrada do neurônio J . Cada neurônio J realiza o somatório de todos os sinais aplicados à sua entrada, conforme Equação (1), e aplica este resultado em uma função de ativação.

$$u = \sum x_i w_{ij} \quad (1)$$

Neste artigo foi utilizada a função de ativação sigmoide, conforme Equação (2). A função de ativação é responsável por definir o limiar de ativação do neurônio.

$$f(u) = \frac{1}{1 + e^{-ku}} \quad (2)$$

sendo k uma constante. A saída y_j é igual ao valor da função de ativação em um determinado instante, dada pela Equação (3).

$$y_j = f(u) \quad (3)$$

3. Algoritmo de Aprendizado

Define-se algoritmo de aprendizado como um conjunto de regras, que tem por objetivo encontrar a melhor solução para um determinado problema, através do reajuste dos pesos [Haykin 2001]. Encontra-se, na literatura, uma grande variedade de algoritmos, que se diferem principalmente pelo modo como os pesos são modificados. Neste trabalho, são abordados o algoritmo evolucionário, o gradiente descendente, e o algoritmo híbrido, que mescla os dois primeiros.

3.1. Algoritmo Evolucionário

O algoritmo evolucionário (AE) é um algoritmo baseado nos princípios da evolução biológica e pode ser utilizado como algoritmo de aprendizagem de uma rede neural artificial, para encontrar os pesos sinápticos que representam a melhor solução do problema [Calvete et al. 2013].

No AE cada possível solução do problema representa um indivíduo. A partir de uma população de indivíduos é esperado que, ao se aplicar operadores genéticos de mutação e cruzamento nos indivíduos, os mesmos evoluam e representem melhores soluções para o problema em questão.

Uma implementação de um AE inicia-se com uma população inicial gerada aleatoriamente, cuja aptidão de cada indivíduo é avaliada. A aptidão dá a cada indivíduo uma pontuação baseada no quão ele se aproxima da resposta desejada. Nesta pesquisa, o cálculo da aptidão de um indivíduo da população é medido pela função do erro médio quadrático, conforme Equação (4).

$$E = \frac{1}{p} \sum_p \sum_{i=1}^n (D_i - Y_i)^2 \quad (4)$$

sendo p o número de padrões, D_i a saída desejada e Y_i a saída da rede neural artificial.

Após o cálculo da aptidão, é gerada uma nova população de indivíduos (filhos), através do processo de seleção e cruzamento. No processo de seleção por torneio são escolhidos dois indivíduos (pais) para o processo de reprodução, garantindo a evolução dos indivíduos. Em seguida, é aplicado o operador genético de cruzamento, onde ocorre a troca de material genético, entre os pais selecionados, permitindo que os filhos herdem características dos pais. É então aplicado o operador genético de mutação na nova população de filhos, onde é alterado aleatoriamente o valor de um gene do indivíduo, com uma probabilidade mínima de mutação. Por fim, a nova população passa pelo processo de elitismo, no qual o pior indivíduo é substituído pelo melhor indivíduo da população anterior. Este processo continua através de iterações sucessivas do algoritmo até o critério de parada [Affenzeller et al. 2009].

3.2. Gradiente Descendente

O gradiente descendente (GD) é um método de otimização que parte de um ponto inicial aleatório e realiza a busca direcionada do ótimo local através do cálculo do gradiente da função-objetivo. O GD permite encontrar, através do anti-gradiente da função-objetivo, a direção que produz o maior decréscimo da função. Ou seja, como o gradiente aponta uma direção onde a função é crescente, busca-se caminhar em direção contrária [Santos 2012].

Este método tem como objetivo usar o gradiente descendente para minimizar o erro E em função dos pesos w [Zeybeka et al. 2006]. Seja uma função erro obtida pela média dos erros quadráticos, dada pela Equação (5):

$$E = \sum_p E^p = \frac{1}{2} \sum_p (d_p - a_p)^2 \quad (5)$$

sendo p o padrão de entrada. Desta forma E^p representa o erro, d_p é a saída desejável e a_p é a saída obtida. Calcula-se a influência de cada peso da rede neural artificial a partir da derivada do erro. Uma vez calculada a derivada parcial, o próximo passo é calcular o gradiente descendente para minimizar a função Erro, através da Equação (6):

$$w_{ij}(t+1) = w_{ij}(t) - \gamma \frac{dE}{dw_{ij}} \quad (6)$$

3.3. Algoritmo Híbrido

Aprendizagem híbrida é um método que combina duas ou mais técnicas de aprendizagem, para ajuste dos pesos sinápticos de uma rede neural artificial, com o objetivo de unir as vantagens e superar limitações individuais de cada técnica. Este método permite construir sistemas mais robustos, resolver problemas mais complexos, dentre outras vantagens.

Nesta pesquisa, o treinamento híbrido da rede neural artificial combina as vantagens de busca global do algoritmo evolucionário com as vantagens de busca local do gradiente descendente. O AE é utilizado para encontrar boas regiões no espaço de peso, enquanto o GD realiza uma busca local nessa região [Prudêncio 2002].

Na aprendizagem híbrida, o algoritmo de busca local gradiente descendente é inserido no processo de evolução do AE, isto é, o GD é utilizado para definir os pesos iniciais do AE, conforme Figura 1. Este método pode se tornar complexo por ser necessário definir tanto os parâmetros do algoritmo de treinamento local como os do algoritmo evolucionário [Prudêncio 2002].

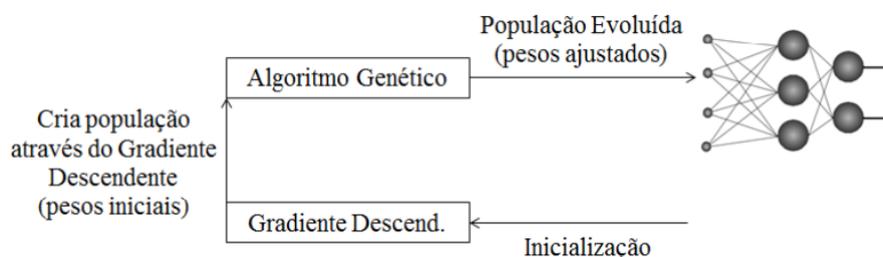


Figura 8. Esquema do Algoritmo Híbrido proposto

4. Estudo Experimental

O desenvolvimento deste estudo comparativo é dividido em três implementações, a saber, (i) MLP com aprendizagem por algoritmo evolucionário, (ii) MLP com aprendizagem por GD e (iii) MLP com Aprendizagem Híbrida. Para implementação da rede neural artificial foi utilizado o *software* MATLAB.

4.1. Base de Dados

Neste estudo foram realizados os testes dos modelos propostos em dois problemas de classificação: a *Iris Flower* e *Breast Cancer Wisconsin* [Lichman 2013].

A primeira base de dados, *Iris Flower*, contém 150 amostras de dimensões de sépalas e pétalas da flor *Iris*, classificadas em três categorias distintas de flores (*Iris Setosa*, *Iris Versicolour* e *Iris Verginica*). Assim, a topologia da RNA fica definida como 4 entradas representando os valores de largura e altura das pétalas e sépalas da flor e 3 saídas, onde cada neurônio representando uma categoria distinta da flor.

A segunda base de dados, *Breast Cancer Wisconsin*, contém 699 amostras e classifica o câncer de mama como benigno ou maligno, de acordo com nove atributos da biópsia: *clump thickness*, *uniformity of cell size*, *uniformity of cell shape*, *marginal adhesion*, *single epithelial cell size*, *bare nuclei*, *bland chromatin*, *normal nucleoli*, *mitoses*. Assim, a topologia da RNA fica definida como 9 entradas representando os nove atributos da biópsia e 2 saídas que classificam o câncer como benigno ou maligno.

4.2. Ambiente Experimental

A metodologia para análise dos diferentes métodos de treinamento da RNA, com camada única, adotada neste estudo comparativo consistiu em vários treinamentos realizados, variando-se os parâmetros de treinamento. Para facilitar na comparação das diferentes abordagens propostas, os códigos foram executados por 100 épocas.

Os dados de entrada foram normalizados e embaralhados e, em seguida, a base de dados foi dividida em conjuntos de treinamento, com 60% das amostras, validação, com 20% das amostras e teste, com 20% das amostras, sendo utilizado o conjunto de treinamento no aprendizado da RNA em todas as abordagens propostas. Os demais parâmetros específicos de cada método são explicitados abaixo:

- (i) MLP com aprendizagem por Algoritmo Evolucionário:
 - Tamanho da população de pesos sinápticos: 50
 - Taxa de seleção: 10%
 - Tipo de seleção: Seleção por torneio
 - Taxa de cruzamento: 60%
 - Taxa de mutação: 10%
- (ii) MLP com aprendizagem por Gradiente Descendente:
 - Taxa de aprendizado = 0,5
- (iii) MLP com aprendizagem Híbrida:
 - Número de repetições: 10
 - Número de gerações do Gradiente Descendente: 10
 - Número de épocas do Algoritmo Evolucionário: 100
 - Tamanho da população de pesos sinápticos: 50
 - Taxa de seleção: 10%
 - Tipo de seleção: Seleção por torneio
 - Taxa de cruzamento: 60%
 - Taxa de mutação: 10%
 - Taxa de aprendizado = 0,5

O conjunto de validação da base de dados foi utilizado para verificar o poder de generalização da RNA. Já na fase de teste, foi utilizado o conjunto de teste, para determinar a performance da RNA com dados que não foram previamente utilizados. Primeiramente, utilizou-se os valores dos pesos sinápticos ajustados, na execução da rede neural artificial, para obter os valores de saída da RNA. Em seguida, os mesmos foram comparados com a saída desejada para determinar a taxa de acerto da RNA.

4.3. Resultados Experimentais

Neste tópico apresenta-se, para cada uma das abordagens propostas, o erro médio quadrático (EQM) do treinamento e a taxa de acerto para cada um dos conjuntos das bases de dados, objetivando verificar a efetividade dos algoritmos de aprendizado a serem aplicados no modelo para classificação de padrão das bases de dados *Iris Flower* e *Breast Cancer Wisconsin*.

Após realizar testes experimentais variando-se o intervalo dos pesos sinápticos, de forma a alcançar o intervalo que a rede neural artificial apresentasse o melhor desempenho, observou-se que uma escolha inicial inadequada para os valores dos pesos pode fazer com que o treinamento não conduza a modelos de aproximação satisfatórios. Sendo assim, definiu-se para a abordagem (i) e (iii) o intervalo dos pesos w de -25 a 25 e para a abordagem (ii) o intervalo dos pesos w de -5 a 5.

A Figura 2 mostra a evolução do erro médio quadrático na fase de treinamento de cada um dos métodos propostos para a base de dados *Iris Flower*. Observa-se que nas abordagens (ii) e (iii) o desempenho é superior, uma vez que em poucas épocas o EQM já sofre uma queda considerável em torno de 0,2; enquanto que, na abordagem (i), o erro tende a diminuir visivelmente apenas em 70 épocas. Nas três abordagens os

EQMs da base de treinamento, representados pela linha pontilhada, são estáveis durante todo o treinamento. Nas abordagens (i) e (iii) os EQMs da base de validação, representados pela linha contínua, oscilam muito e, portanto, não garantem à RNA uma boa capacidade de generalização.

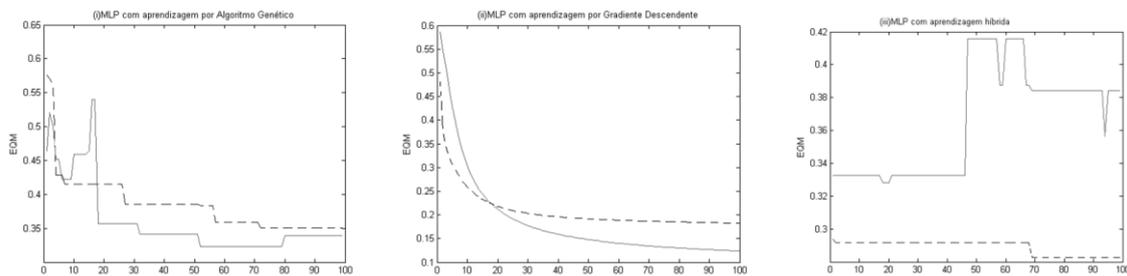


Figura 2. EQM dos dados de treinamento e validação para base Iris Flower

A Figura 3 mostra a evolução do EQM na fase de treinamento de cada uma das abordagens propostas para a base de dados *Breast Cancer Wisconsin*. Observa-se que na abordagem (iii) o desempenho é superior, uma vez que no início do treinamento, o EQM já sofre uma queda considerável em torno de 0,2; enquanto que, na abordagem (ii), o erro tende a diminuir visivelmente apenas em 50 épocas e na abordagem (i), o erro tende a diminuir apenas em 80 épocas. Nas três abordagens os EQMs da base de treinamento, representados pela linha pontilhada, são estáveis durante todo o treinamento. Porém, só nas abordagens (ii) e (iii), o EQM da base de validação, representados pela linha contínua, permanecem estáveis, garantindo a capacidade de generalização da rede neural artificial.

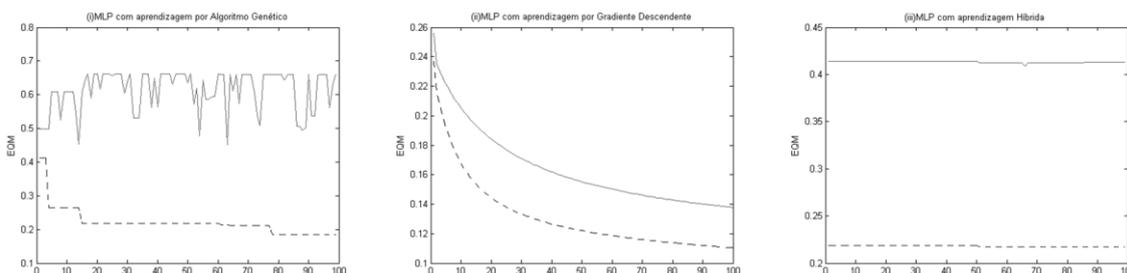


Figura 3. EQM dos dados de treinamento e validação para base Breast Cancer Wisconsin

A Tabela 1 mostra os resultados obtidos de cada abordagem proposta, para as duas bases de dados, sendo na primeira coluna a abordagem adotada. Os valores são obtidos através da média do total de execuções do mesmo. Os dados apresentados comprovam a análise dos gráficos da Figura 2 e da Figura 3. Na base de dados *Iris Flower*, os valores do EQM das abordagens (ii) e (iii) são estatisticamente iguais, melhores que o valor da abordagem (i). Para a base de dados *Breast Cancer Wisconsin*, os valores do EQM da abordagem (ii) e (iii) são próximos, melhores que o valor da abordagem (i). As taxas de acerto, quanto à capacidade de generalização da RNA, teve valores aceitáveis, uma vez que no treinamento, rodou-se apenas 100 épocas os algoritmos de aprendizagem, para efeito de comparação de desempenho entre os métodos. Observa-se que dentre os treinamentos para a MLP, o método de treinamento que mais se destacou, para as duas bases de dados, é o (iii) MLP com aprendizagem híbrida, apresentando o menor EQM e melhores valores das taxas de acerto. O bom desempenho de generalização é confirmado analisando a base de teste, onde a taxa de acerto média obtida é de 77,99%, para a base de dados *Iris Flower* e 86,00%, para a

base de dados *Breast Cancer Wisconsin*, ambos os valores são superiores aos demais métodos estudados.

Tabela 1. Resultados do treinamento para diferentes modelos de MLP

Modelo	Treinamento		Validação		Teste	
	EQM	Taxa de acerto	EQM	Taxa de acerto	EQM	Taxa de acerto
<i>Iris Flower Data Set</i>						
i.AE	0,23	67,77%	0,14	76,00%	0,23	68,66%
ii.GD	0,18	63,99%	0,13	71,33%	0,17	71,33%
iii.Híbrido	0,18	71,55%	0,14	76,66%	0,17	77,99%
<i>Breast Cancer Wisconsin Data Set</i>						
i.AE	0,22	73,69%	0,24	70,66%	0,22	74,86%
ii.GD	0,13	83,51%	0,15	80,29%	0,14	80,10%
iii.Híbrido	0,11	87,59%	0,15	81,61%	0,11	86,00%

5. Conclusão

Neste artigo foi proposto avaliar e comparar três métodos distintos de treinamento de rede neural artificial. Os problemas de classificação de padrões das bases de dados *Iris Flower* e *Breast Cancer Wisconsin* foram utilizados para demonstrar qualitativamente as características de cada método.

Os experimentos mostraram que o método híbrido apresentou os melhores resultados, para ambos os problemas estudados. A taxa de acerto média obtida nos testes foi de 77,99%, para a base de dados *Iris Flower* e 86,00%, para a base de dados *Breast Cancer Wisconsin*, ambos os valores foram superiores aos demais métodos estudados. Assim, concluiu-se que o método híbrido, que combina algoritmos de busca local e global, é uma opção válida para problemas de classificação de padrões.

Como sugestão para trabalhos futuros, propõe-se a utilização de parâmetros adaptativos no algoritmo híbrido, combinando as vantagens da busca global e local a fim de encontrar soluções ainda mais efetivas.

References

- Affenzeller, M., Wagner, S., Winkler, S., and Beham, A. (2009). Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications.
- Calvete, H. I., Galé, C., and Iranzo, J. A. (2013). An efficient evolutionary algorithm for the ring star problem. *European Journal of Operational Research*, 246:343.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annual Eugenics*, 7(2):179–188.
- Gouvêa, M. M. (2012). A neural-based gradient optimization in large power systems. *WSEAS Transaction on Systems*, 11(8).
- Haykin, S. (2001). *Redes Neurais – Princípios e Práticas*. Porto Alegre.
- Karabatak, M. and Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. 36:3465–3469.
- Lichman, M. (2013). UCI machine learning repository.

- Meireles, M. R. G., Almeida, P. E. M. and Simões, M. G. (2003). A comprehensive review for industrial applicability of artificial neural networks. *IEEE Transaction on Industrial Electronics*, 50(3):585–601.
- Prudêncio, R. (2002). Projeto híbrido de redes neurais. Master's thesis, Centro de Informática, Universidade Federal de Pernambuco.
- Santos, T. F. (2012). Operador para distribuição de soluções para algoritmos evolutivos via adaptação da matriz de covariância. PhD thesis, UFMG, Belo Horizonte.
- Shekhawat, P. B. and Dhande, P. S. S. (2011). A classification technique using associative classification. *International Journal of Computer Applications*, 20(5):0975 – 8887.
- Wolberg, W. H. and Mangasarian, O. L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proc Natl Acad Sci U S A*, 87:9193–9196.
- Zeybeka, Z., Cetinkayab, S., Hapoglua, H., and Albaza, M. (2006). Generalized delta rule (gdr) algorithm with generalized predictive control (gpc) for optimum temperature tracking of batch polymerization. *Chemical Engineering Science*, 61:6691–6700.