

A MELHORIA NO DESEMPENHO DE MLP COM O USO DE NOVOS VETORES ALVO

José Ricardo Gonçalves Manzan, Shiguo Nomura, Keiji Yamanaka

Faculdade de Engenharia Elétrica (FEELT) - Universidade Federal de Uberlândia (UFU)

joserickardo@iftm.edu.br, shiguo@feelt.ufu.br, keiji@ufu.br

Resumo – O presente trabalho realiza uma análise experimental por meio da utilização de vetores bipolares ortogonais como vetores alvo das Redes Neurais Artificiais (RNA) do tipo *multilayer perceptron* (MLP). O emprego de tais vetores é feito tanto na fase de treinamento quanto na fase de aplicação. A análise experimental consiste na utilização de dados de dígitos manuscritos que se encontram armazenados na UCI Machine Learning Repository de acesso livre e internacional. O treinamento foi realizado com o intuito de obter as melhores taxas de reconhecimento, possibilitando dessa forma a comparação de resultados dos vetores convencionais com os vetores ortogonais sugeridos como uma nova abordagem de treinamento. Os resultados experimentais têm mostrado que o desempenho da MLP melhora significativamente quando são usados os vetores bipolares ortogonais.

Palavras – chave – Reconhecimento de padrões, redes neurais multicamadas, vetores alvo, vetores bipolares ortogonais

1 Introdução

O campo da inteligência computacional tem possibilitado grandes avanços para a ciência, respondendo a diversas questões que há muito tempo não podiam ser respondidas. Juntamente com os avanços conquistados esta área da ciência tem permitido que muitos problemas que antes eram considerados praticamente insolúveis sejam resolvidos em um curto espaço de tempo. Nesse cenário vários pesquisadores [1][2][3] tem se esforçado para melhorar o desempenho das Redes Neurais Artificiais (RNA) ante as amplas possibilidades de aplicação aos quais elas se destinam. A relevância destas pesquisas é incomensurável, por que elas podem tornar as redes mais robustas, eficientes e, sobretudo mais aplicáveis a situações novas que vem surgindo.

É possível perceber que as tentativas de aperfeiçoar as RNA são variadas e ocorrem em diferentes pontos. Em alguns casos os esforços se destinam a trabalhar com a organização dos padrões de entrada, permitindo que a aprendizagem seja mais efetiva. Por outro lado há casos em que a idéia se destina a melhorar a arquitetura da rede encontrando uma topologia equilibrada que seja capaz de produzir resultados melhores. Também são identificados estudos que buscam melhorias no algoritmo de aprendizagem. Sem ter a mínima pretensão de esgotar o rol de estudos realizados na melhoria do desempenho das Redes, podemos ainda citar os casos em que são procurados modelos de minimização do erro. Nessa direção se faz necessária a determinação de alguma técnica de otimização. Entretanto realizar a modelagem onde há um grande número de variáveis, não é uma tarefa fácil.

De um modo geral, é perceptível que não têm sido encontradas contribuições de pesquisas no sentido de trabalhar com os vetores alvo da rede. Nesse sentido este trabalho propõe uma nova metodologia no treinamento de Redes Perceptron Multicamadas (MLP). Além de selecionar adequadamente os padrões de entrada e de determinar uma topologia eficiente com a escolha correta do número de neurônios da camada intermediária, é preciso definir os vetores alvo da rede de forma a garantir uma aprendizagem com resultados mais eficientes. Os vetores alvo influenciam significativamente na capacidade da RNA em reconhecer padrões.

A nova metodologia mostra que se os vetores alvo forem bipolares e ortogonais (VBO), é possível reduzir a similaridade entre eles, o que permitirá no caso específico de redes MLP, diferenciarem com mais facilidade os padrões de entrada que são propagados na rede. É de conhecimento científico que a não similaridade entre os padrões de saída facilita a aprendizagem da memória humana. Como as RNA são baseadas no funcionamento dos neurônios humanos, é de se esperar que a teoria defendida pela abordagem proposta tenha grande relevância.

Esse estudo já foi apresentado em [4][5] com resultados que comprovam a eficácia dessa nova metodologia para as MLP. Neste trabalho, mostramos mais resultados experimentais sobre a melhora na aprendizagem das redes com o uso de Vetores Bipolares Ortogonais (VBO) como alvo da rede no reconhecimento de dígitos manuscritos.

Nas seções 2 e 3 são apresentados os antecedentes do reconhecimento de padrões e a motivação para este trabalho. As seções 4 e 5 versam sobre a motivação e a importância dessa nova abordagem. As seções 6 e 7 tratam da definição dos vetores alvo do experimento bem como o algoritmo para a geração de vetores do tipo VBO. Procedimentos e resultados experimentais são encontrados nas seções 8 e 9 comprovando a melhora no desempenho da rede. Finalmente a conclusão é feita por meio da seção 10.

2 Antecedentes

Pesquisas envolvendo o reconhecimento de padrões tanto para descrição, quanto para classificação, tem se tornado um campo de grande destaque dentro da área de processamento da informação. A princípio, estas pesquisas utilizavam a abordagem estatística, mais conhecida como decisão teórica, e também a abordagem sintática, também conhecida como estrutural [7]. Atualmente esse panorama tem sofrido algumas alterações, no sentido de que a nova abordagem de reconhecimento de padrões degradados ocorre com o uso das Redes Neurais Artificiais. Nessa linha, os resultados tem sido promissores e demonstram uma grande capacidade de flexibilização e robustez, o que faz das RNA uma ferramenta amplamente utilizada nesse campo de aplicação. O grande mérito das RNA se dá pela capacidade que a rede possui em realizar sua aprendizagem com dados de treinamento supervisionado ou não supervisionado.

As RNA realizam sua aprendizagem com o uso de modelos para processamento distribuído e paralelo, o que não ocorre com as técnicas da inteligência artificial clássica onde os símbolos são processados sequencialmente. Nos neurônios biológicos a aprendizagem se dá por meio de exemplos, onde a tentativa e o erro vão permitindo a aquisição da habilidade de diferenciar padrões. Como as RNA são baseadas nesse comportamento, seu mapeamento corresponde a uma interação de um grande número de neurônios que enviam sinais inibitórios ou excitatórios a outros neurônios da rede.

3 Motivação

O sistema de cognição biológico tem a capacidade de reconhecer e diferenciar padrões mesmo com alto grau de degradação. No caso das RNA essa habilidade é possível se os parâmetros de treinamento forem adequadamente ajustados. Redes que são treinadas com uma abordagem destinada a modelos mais genéricos, serão hábeis a reconhecer um padrão degradado, mesmo que ele não tenha sido utilizado na fase de treinamento. Assim, modelos mais flexíveis, são extremamente úteis para esse tipo de aplicação. Por outro lado, uma rede treinada de forma rígida, com excesso, poderá gerar resultados indesejados deixando de reconhecer aqueles padrões possuidores de grande quantidade de ruído.

Trabalhos destinados a otimização do treinamento de RNA vem sendo realizados sempre na tentativa de melhorar a capacidade de reconhecimento da rede. É possível encontrar propostas destinadas ao tratamento adequado dos vetores de entrada [10] para atingir a melhoria esperada. Contudo, são raras as propostas de abordagens que se destinem ao tratamento dos vetores alvo da rede. Este trabalho propõe esta nova metodologia, mostrando que a utilização de vetores bipolares e ortogonais (VBO) influencia de forma satisfatória no desempenho da rede em reconhecer padrões. Os estudos [4][5][11] já realizados têm mostrado um promissor aumento na taxa de reconhecimento de padrões com os vetores alvo não convencionais.

4 Justificativa

O produto interno de vetores alvo convencionais aumenta conforme eles aumentam de tamanho. Esse produto representa a similaridade entre os vetores. Quanto maior for o valor do produto interno, menor será o ângulo formado por eles, o que ocasiona um aumento de similaridade. No caso de vetores alvo ortogonais, a similaridade é sempre nula, independente de seus tamanhos. Isso é possível pelo fato de que vetores ortogonais têm produto interno nulo. Outro ponto interessante de ser analisado é a questão da distância euclidiana. As equações (1) e (2) representam dois possíveis vetores alvo, a equação (3) ilustra o produto interno entre eles e a equação (4) sua distância euclidiana.

$$\vec{V}_i = (v_1, v_2, \dots, v_n) \quad (1)$$

$$\vec{W}_i = (w_1, w_2, \dots, w_n) \quad (2)$$

$$\vec{V}_i \cdot \vec{W}_i^T = v_1 \cdot w_1 + v_2 \cdot w_2 + v_3 \cdot w_3 + \dots + v_n \cdot w_n \quad (3)$$

$$d_{v,w} = \sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2 + (w_3 - v_3)^2 + \dots + (w_n - v_n)^2} \quad (4)$$

Se \vec{v} e \vec{w} forem ortogonais de tamanho n , haverá $n/2$ parcelas cujo produto é positivo e $n/2$ parcelas cujo produto é negativo. As parcelas cujo produto é positivo correspondem a aquelas em que os termos têm mesmo sinal. Estes termos têm diferença nula, não interferindo na soma compreendida pela raiz quadrada da fórmula da distância euclidiana (4). Os termos que tem sinais diferentes terão diferença cujo módulo é igual a 2. O quadrado dessas diferenças contribui para a soma compreendida pela raiz quadrada da fórmula da distância euclidiana. Logo, quanto maior, for o valor de n , maior será sua distância euclidiana.

Vetores convencionais tem um termo igual a 1 e os demais iguais a -1. Dessa forma se \vec{v} e \vec{w} forem convencionais, as diferenças entre os termos correspondentes serão sempre iguais a 0, exceto nos dois termos, um de cada vetor, que são iguais a 1. Conclui-se que a distância euclidiana de vetores convencionais é sempre menor que a de vetores bipolares ortogonais.

Esse aspecto dos VBO proporciona a obtenção de um número maior de neurônios ativos. Essa distância ampliada entre os vetores alvo pode ser comparada a diferença entre padrões, o que nos mostra que a MLP melhora seu desempenho graças à similaridade nula entre esses vetores. O que não ocorre com os vetores convencionais.

5 Importância

Seria muita presunção ter a intenção de citar todas as possibilidades de aplicações para RNA. Sem a mínima intenção de cometer injustiças, podemos citar que as MLP têm sido utilizadas, por exemplo, no reconhecimento de letras e dígitos, no reconhecimento de sinais biológicos, de imagens de radiografia, tomografia, sinais sonoros, sinais elétricos e outros. Em específico, pode-se citar o reconhecedor automático de códigos de placas de veículos de imagens reais captadas por câmeras de controle de tráfego.

A abordagem deste trabalho consiste na utilização de vetores alvo não convencionais, a saber, VBO, na classificação de dígitos manuscritos. Trata-se de uma rede do tipo MLP. Esse sistema serve como prova para a hipótese levantada, consistindo num problema que representa um grande desafio.

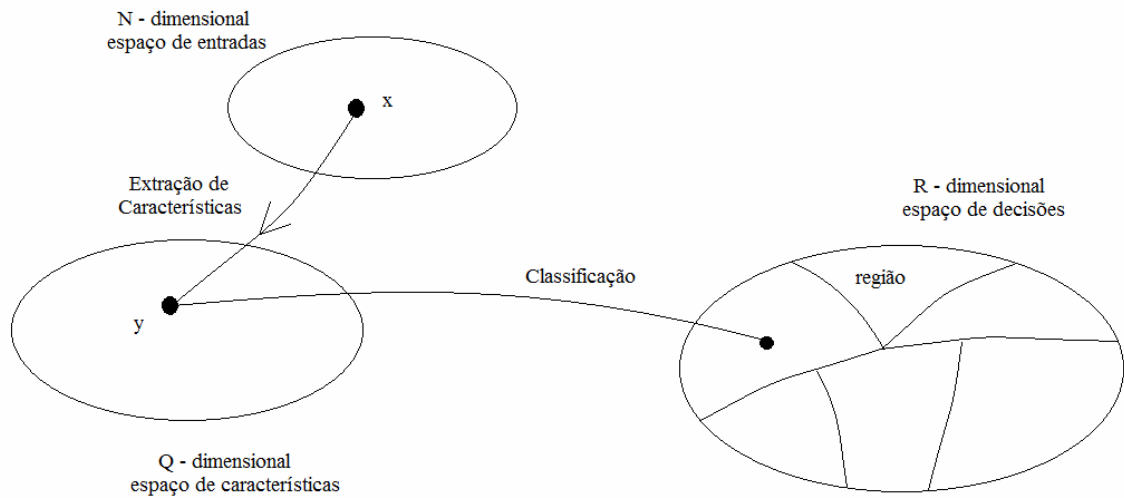


Figura 1 – Processo de extração de características seguido de classificação no espaço de decisões.

A Figura 1 mostra o processo de treinamento como um modelo que extrai características, classifica e finalmente realiza decisões.

6 Vetores Alvo na aprendizagem de uma MLP

Os vetores bipolares convencionais (VBC) (5) são aqueles em que o elemento 1 na posição i corresponde ao dígito i do vetor V_i que representa esse dígito. A exceção ocorre para o dígito 0, cujo elemento positivo se encontra na posição 10.

$$\vec{V}_i \stackrel{def}{=} (-1, -1, \dots, 1, \dots, -1) \quad (5)$$

Já os vetores ortogonais têm por propriedade o produto interno nulo, ou seja, para dois vetores v e w pertencentes ao espaço R^n , o produto interno dado por (25) tem seu resultado igual a zero.

$$\vec{V} \cdot \vec{W} = x_1 y_1 + x_2 y_2 + \dots + x_n y_n \quad (6)$$

Onde:

$$\vec{V} \stackrel{def}{=} (x_1, x_2, \dots, x_n) \quad (7)$$

$$\vec{W} \stackrel{def}{=} (y_1, y_2, \dots, y_n) \quad (8)$$

Os vetores usados nessa abordagem, além de terem produto interno nulo, são bipolares, ou seja, seus elementos assumem os valores 1 ou -1 . A esses vetores nós denominamos por VBO.

Abordam-se ainda os vetores bipolares não ortogonais VNO que tem o mesmo tamanho dos VBO, mas possuem produto interno não nulo. O produto interno dá a medida do grau de similaridade entre os vetores. Devido ao fato de que estes vetores terão apenas um dígito igual a 1 e os demais iguais a -1 , o produto interno tende a crescer a medida que o tamanho dos vetores aumenta. Nesse sentido, a taxa de reconhecimento da rede diminui à medida que o grau de similaridade dos vetores alvo aumenta.

7 Algoritmo para Geração de VBO

Esse método [6] usa um vetor de sementes na geração de VBO. O número de componentes de cada VBO é calculado por:

$$n = 2^k m \quad (9)$$

Em (9) 2^k é o número de vetores ortogonais para k maior que 0 e m é o número de componentes em um vetor de sementes. Assim, um conjunto com 2^k VBO é construído com $2^k m$ componentes. O algoritmo é realizado a partir dos seguintes passos:

1) *Passo 1 - Inicialização de m e k* - Os valores de m e k devem ser determinados de acordo com a necessidade da aplicação.

2) *Passo 2 - Inicialização do vetor de sementes* - O vetor de sementes é obtido através de (10).

$$V_m^0 = \overbrace{(1, 1, \dots, 1)}^m \quad (10)$$

3) *Passo 3 - Cálculo de número de componentes em um VBO* - O cálculo do número de componentes do VBO é feito através da equação (9).

4) *Passo 4 - Construção de vetores* - São construídos os vetores $V_{2m}^1 = fcc(V_m^0, V_m^0)$ e $V_{2m}^2 = fcc(V_m^0, -V_m^0)$, onde $fcc(V_1, V_2)$ é a função de concatenação dos vetores V_1 e V_2 .

5) *Passo 5 - Construção de vetores* - São construídos os vetores $V_{4m}^1 = fcc(V_{2m}^1, V_{2m}^1)$, $V_{4m}^2 = fcc(V_{2m}^1, -V_{2m}^1)$, $V_{4m}^3 = fcc(V_{2m}^2, V_{2m}^2)$ e $V_{4m}^4 = fcc(V_{2m}^2, -V_{2m}^2)$.

6) *Passo 6 - Sequência da Concatenação* - São realizadas concatenações até que se consiga 2^k vetores ortogonais com n componentes: $V_n^1, \dots, V_n^{2^k}$.

8 Procedimento Experimental

A RNA utilizada neste trabalho é uma MLP com 256 neurônios da camada de entrada. Para a camada intermediária foram realizadas simulações com a utilização de 100 e 200 neurônios escondidos. Na camada de saída foram feitos testes com 10 e 64 neurônios de acordo com o tipo de vetor alvo escolhido para treinamento. Numa primeira etapa foram utilizados VBO com 64 elementos. Num segundo momento foram utilizados VNO também com 64 elementos. Finalmente a rede foi simulada com VBC contendo 10 elementos. A taxa de aprendizagem foi sempre mantida com o valor de 0,01 com critério de parada efetuado através do alcance do erro tolerado ou pelo limite máximo de 30000 ciclos. Para todos os tipos de simulação foram mantidos os pesos sinápticos que foram gerados de forma aleatória.

Os dados de treinamento e teste são dígitos manuscritos obtidos através do repositório internacional conhecido como Semeion Handwritten Digit do Machine Learning Repository. Tratam-se de padrões obtidos de um grupo de cerca de 80 pessoas que foram solicitadas a escrever duas vezes os algarismos de 0 a 9. Na primeira solicitação as pessoas foram solicitadas a escrever os dígitos de forma tranquila primando pela perfeição dos manuscritos. Na segunda solicitação as pessoas eram induzidas a escrever os algarismos de forma rápida sem preocupação com a legibilidade dos mesmos.

Cada algarismo foi digitalizado em uma imagem contendo 256 pixels, no formato de 16 linhas e 16 colunas. Cada imagem foi processada numa escala de resolução de 256 níveis de cinza. Posteriormente uma matriz para representação das mesmas foi obtida de forma a posicionar cada linha de pixels a direita da linha abaixo, formando então matrizes contendo uma linha e 256 colunas. Para cada pixel correspondente ao fundo da imagem foi atribuído o valor 0 e para os pixels do algarismo foi atribuído o valor 1. No treinamento desse trabalho os pixels com valor 0 passaram a receber o valor - 1 para que fosse possível realizar o treinamento com vetores bipolares.

9 Resultados Experimentais

Embora a MLP tenha sido projetada com 200 e com 100 neurônios escondidos na camada intermediária, as melhores taxas de reconhecimento ocorreram usando 200 neurônios. Várias combinações de parâmetros foram realizadas, sempre no intuito de obter as melhores taxas de reconhecimento para cada tipo de vetor alvo utilizado. Assim há a garantia de que as comparações sejam as mais justas possíveis.

Na tabela 1 são apresentados os resultados da rede treinada com vetores alvo VBC.

Tabela 1 – Análise do desempenho da MLP usando VBC

Nº da Simulação	Tolerância	Nº de Ciclos	Taxa de Reconhecimento
213	$5 \cdot 10^{-3}$	804	80%
350	$5 \cdot 10^{-4}$	7243	<u>81,67%</u>
351	$3 \cdot 10^{-4}$	11928	<u>81,67%</u>

Na tabela 2 são mostrados os resultados da rede treinada com vetores VNO.

Tabela 2 – Análise do desempenho da MLP usando VNO

Nº da Simulação	Tolerância	Nº de Ciclos	Taxa de Reconhecimento
212	$5 \cdot 10^{-3}$	2622	81,33%
300	$5 \cdot 10^{-4}$	25808	81,67%
301	$3 \cdot 10^{-4}$	30000	<u>82,00%</u>

Finalmente a tabela 3 apresenta os resultados da rede treinada com vetores VBO.

Tabela 3 – Análise do desempenho da MLP usando VBO

Nº da Simulação	Tolerância	Nº de Ciclos	Taxa de Reconhecimento
211	$5 \cdot 10^{-3}$	2750	82,67%
250	$5 \cdot 10^{-4}$	26965	83,00%
251	$3 \cdot 10^{-4}$	30000	<u>83,33%</u>

Os valores sublinhados e em negrito simbolizam as melhores taxas obtidas. Os resultados mostram que em todos os casos, o uso dos VBO propicia um melhor desempenho no reconhecimento de padrões. Em todos os experimentos a taxa de aprendizagem foi mantida constante. O critério de parada do treinamento da rede foi realizado através de dos critérios do erro e do limite de 30000 ciclos. O critério que ocorre primeiro indica o encerramento do treinamento. No caso de parada por erro consideramos os seguintes cálculos:

- $E = \frac{1}{2} \sum_{j=1}^{Ns} (d_j - y_j)^2$ onde E é o erro quadrático para um padrão; Ns é o número de neurônios de saída; d_j é a saída desejada para o neurônio j ; y_j é o valor de saída do neurônio j ;
- $E_m = \frac{1}{N} \sum_{p=1}^N E_p$ onde E_m é o erro quadrático médio de todos os padrões para cada ciclo de treinamento; N é o número de padrões; E_p é o erro quadrático para o padrão p calculado pela equação acima;

Dessa forma, o treinamento é finalizado quando a condição $E_m < \varepsilon$ for satisfeita. O valor de ε corresponde ao valor de *Tolerância* fornecido na simulação. Se a condição do erro não for satisfeita e o número de ciclos for igual a 30000, o treinamento será finalizado.

10 Conclusão

Os resultados mostram claramente a melhora da capacidade de reconhecimento da MLP no caso da utilização de VBO. Isso vem de encontro com a hipótese levantada nesse trabalho e também com resultados já mostrados em pesquisas anteriores [4] e [5].

Nesse sentido a utilização destes vetores não convencionais propicia benefícios para o processo de treinamento da rede de acordo com as questões levantadas inicialmente, como o aumento da distância euclidiana, o aumento de neurônios ativos na camada de saída da MLP, e a redução da similaridade entre os vetores.

11 Agradecimentos

Os autores agradecem à Pró-Reitoria de Pesquisa e Pós-Graduação da Universidade Federal de Uberlândia pelo apoio na execução do projeto 72/2010 do qual o presente trabalho faz parte.

12 Referências

- [1] X. Wang, C. Chang, F. Du, Achieving a More Robust Neural Network Model for Control of a MR Damper by Signal Sensitivity Analysis, **Neural Computing & Applications**, 13(2002), 330-338.
- [2] M. A. Costa, A. P. Braga, B. R. Menezes, Improving Neural Networks Generalization With New Constructive and Pruning Methods, **Journal Of Intelligent & Fuzzy Systems**, 13(2003), 75-83.
- [3] C. M. Lee, S. S. Yang, C. L. Ho, Modified back-propagation algorithm applied to decision-feedback equalization, **Iee Proceedings – Vision, Image & Signal Processing**, 153(2006), 805-809.
- [4] S. Nomura, K. Yamanaka, O. Katai, H. Kawakami, T. Shiose, Improving MLP learning via Orthogonal Bipolar Target Vectors, **Journal of Advanced Computacional Intelligence and Intelligent Informatics**, 9(2005), 580-589.
- [5] S. Nomura, K. Yamanaka, O. Katai, H. Kawakami, T. Shiose, Uma Nova Abordagem Matemática para Melhoria do Desempenho de Redes Neurais Artificiais, **VIII Simpósio Brasileiro de Redes Neurais**, (2004).
- [6] L. Fausset, Fundamentals os Neural Networks: Architecture, Algorithms, and Applications, **Prentice-Hall**, (1994).
- [7] A. Browne, Neural Network Analysis, Architetures, and Applications, **Institute of Physics Pub** (1997).
- [8] R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification, **John Wiley & Sons** (2001).
- [9] T. Kohonen, Associative Memory: A System-Theoretical Approach, **Springer-Verlag** (1977).
- [10] R.O. Duda, P. E. Hart, Pattern Classification and Scene Analysis, **Wiley** (1973).
- [11] K. S. Fu, Syntactic Methods in Pattern Recognition, **Academic Press** (1974).