

Treinamento Multi-Objetivo de Redes Neurais Artificiais Baseado em Coordenadas Esféricas

Honovan Paz Rocha*, Marcelo Azevedo Costa* e Antônio Pádua Braga*

*Programa de Pós-Graduação em Engenharia Elétrica - Universidade Federal de Minas Gerais - Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brasil

Abstract—Artificial neural networks have been successfully applied in solving problems such as function approximation and pattern classification. The search for a model that best represents the problem makes the generalization ability the main concern in the training of artificial neural networks. In this context, this paper proposes a new technique for multi-objective training of neural networks, where the formulation of the network is modified so that the optimization problem is represented in the spherical coordinate system.

Keywords—Multi objective training, spherical coordinate, fixed norm, unconstrained optimization.

Resumo—As Redes neurais artificiais tem sido aplicadas com sucesso na resolução de problemas como aproximação de funções e classificação de padrões. A busca por um modelo que melhor represente o problema torna a habilidade de generalização o principal objetivo no treinamento de redes neurais artificiais. Neste contexto, este trabalho propõe uma nova técnica para o treinamento multi-objetivo de redes neurais, onde a formulação da rede é modificada de forma que o problema de otimização é representado no sistema de coordenadas esféricas.

Keywords—Treinamento multi-objetivo, coordenadas esféricas, norma fixa, otimização irrestrita.

I. INTRODUÇÃO

A teoria do Aprendizado Estatístico formalizou as bases do princípio da minimização do Risco Estrutural (MRE) através do qual se busca pela menor estrutura capaz de minimizar o Risco Empírico [1], [2]. Através desta teoria foi possível analisar o problema de generalização de Redes Neurais Artificiais (RNAs) onde foi demonstrado que a obtenção de soluções eficientes pode ser alcançada a partir da minimização de dois objetivos conflitantes: o erro de treinamento e a capacidade (ou complexidade) da classe de funções fornecida pela máquina de aprendizagem. Desta maneira, uma boa capacidade de generalização da rede neural pode ser obtida através do equilíbrio entre estes dois objetivos.

Pode-se realizar o controle da complexidade da rede através da manipulação direta da estrutura da rede, alterando as conexões e quantidade de neurônios em suas camadas. De forma alternativa, é possível manipular a magnitude dos pesos da rede, não levando em consideração sua estrutura física. Em [3] é mostrado que a variância é proporcional à magnitude dos pesos e em [4] é visto que a magnitude dos pesos é mais importante que a quantidade destes na rede, quando se busca uma rede neural com maior poder de generalização. A seleção de modelos visando maior capacidade de generalização tem sido obtida através de técnicas como métodos de validação [5], algoritmos de poda ou construtivos

[6] e regularização [7]. Com base no controle da complexidade a partir de restrições na magnitude dos pesos, algoritmos de aprendizado baseados em técnicas de otimização multiobjetivo (MOBJ) têm sido aplicados ao aprendizado de RNAs [8], [9], [10], [11], [12], [13]. Estes métodos visam à seleção de modelos através da minimização de dois objetivos conflitantes: o erro de treinamento e a norma do vetor de pesos. Após o processo de otimização, um conjunto de soluções eficientes (não dominadas) é obtida, constituindo uma estimativa do conjunto Pareto-Ótimo [14]. Dentre as soluções encontradas, aquela que apresenta complexidade compatível com a tarefa de aprendizagem em questão deve ser selecionada.

O objetivo deste trabalho é utilizar uma nova formulação para representação do problema de otimização inerente ao treinamento da rede, de maneira a diminuir a complexidade do problema de otimização levando o foco à formulação do problema de aprendizado. Desta maneira, torna-se possível a exploração de novas características do treinamento com foco na geração de uma estimativa do Pareto-Ótimo. Com base na função norma euclidiana a formulação modela o problema de otimização representando os pesos da rede no sistema de coordenadas esféricas, fazendo com que um dos objetivos seja tratado de forma intrínseca. Nesta formulação o problema de otimização torna-se mono-objetivo irrestrito onde visa-se encontrar o mínimo valor de erro no espaço limitado pela norma. Com o treinamento a partir desta formulação obteve-se resultados promissores quanto à estimativa do Pareto e à qualidade das soluções encontradas em problemas de regressão e classificação, comparando-se às soluções obtidas pelos métodos baseados no MOBJ.

O restante do trabalho está organizado da seguinte forma. A seção 2 detalha o treinamento multiobjetivo de RNAs, na seção 3 é definida a formulação proposta e, nas duas últimas seções são apresentados os resultados e conclusões do trabalho.

II. APRENDIZADO MULTI OBJETIVO DE REDES NEURAIS ARTIFICIAIS

O ajuste simultâneo da complexidade do modelo e da minimização do erro busca encontrar o modelo que permite equilibrar o ajuste aos dados e o sobreajuste ao ruído. O ajuste entre o erro de treinamento e a complexidade do modelo é conhecido na literatura como o dilema entre o vício e a variância [3]. Considerando-se uma rede MLP, conhecendo-se a relação entre a capacidade de aprendizagem e o tamanho dos pesos da rede, pode-se definir a norma dos pesos da rede $\|w\|$ como uma medida de controle da variância [4]. Desta maneira pode-se definir um algoritmo que controle erro e complexidade em termos de um modelo de otimização definido

pela minimização conjunta das funções objetivo apresentadas nas equações 1 e 2, sendo J_1 o erro médio quadrático e J_2 a norma dos pesos da rede.

$$J_1 = \frac{1}{2} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (1)$$

$$J_2 = \sum_{i=1}^{n_p} w_i^2 \quad (2)$$

Em 1, y_i e \hat{y}_i são os valores para a saída desejada e a saída da rede para i -ésima amostra, respectivamente, sendo que N é o número total de amostras do conjunto de treinamento. Em 2 w_i é o i -ésimo peso da rede e n_p é o número total de pesos da rede.

Uma das possíveis soluções do problema de otimização caracterizado pela minimização conjunta das funções J_1 e J_2 é a utilização do método ϵ -restrito, como descrito em [8]. Nesta abordagem o problema multi-objetivo é descrito por vários problemas de otimização mono-objetivo em que a função J_1 é otimizada tendo $J_2 < \epsilon$ como restrição, conforme representado em 3.

$$\arg \min_{\mathbf{w}} J_1 \quad (3)$$

$$J_2 \leq \epsilon$$

A Figura 1 mostra de forma esquemática o princípio da minimização do Risco Estrutural e a sua analogia com o Aprendizado Multiobjetivo. A solução dos múltiplos problemas de otimização caracterizados em 3 é análogo à minimização do Risco Estrutural através de estruturas aninhadas, uma vez que a restrição $J_2 \leq \epsilon$ define regiões hiperesféricas concêntricas no espaço de parâmetros. O valor da restrição ϵ determina o raio da hiperesfera e controla, portanto, o volume do espaço de soluções em que o problema de otimização será resolvido. Quanto maior o raio da hiperesfera maior será o seu volume e, consequentemente, maior a capacidade da família de soluções envolvida. Desta forma, a definição do raio através do controle da norma do vetor de pesos utilizando a função J_2 leva a uma aproximação do conjunto Pareto-ótimo, já que a função J_1 é minimizada para valores discretos de ϵ , conforme representado de maneira esquemática na Figura 1.

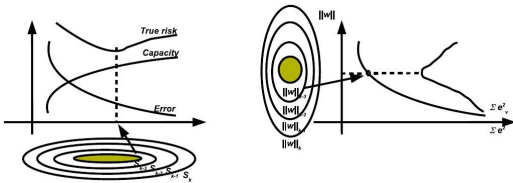


Figura 1. Visão esquemática do princípio de minimização do Risco Estrutural [1].

A formulação proposta visa explorar novas características do treinamento de redes neurais a partir da geração de uma estimativa do conjunto Pareto-ótimo de forma que se mantenha

a simplicidade do problema de otimização da rede, realizando um treinamento mono-objetivo irrestrito. Um dos objetivos é tratado de forma intrínseca, onde dado um valor de norma o treinamento busca pelo menor erro no espaço limitado por esta norma nas condições da nova formulação.

III. FORMULAÇÃO DO PROBLEMA DE OTIMIZAÇÃO NO SISTEMA DE COORDENADAS ESFÉRICAS

A utilização do método ϵ -restrito para a solução do problema multiobjetivo conforme sendo que mesmo se tornando mais complexa possui um mesmo formato 3 requer que o erro seja minimizado dentro de cada um das hiperesferas $J_2 \leq \epsilon$. Assim, do ponto de vista de resolução do problema de otimização, esta restrição deve ser verificada e mantida enquanto a função J_1 é minimizada. Na prática, a resolução pelo ϵ -restrito resulta em uma aproximação da solução Pareto-Ótima de erro mínimo e norma igual ao raio da hiperesfera, conforme representação esquemática da Figura 2. A solução final é equivalente, do ponto de vista numérico, àquela que seria encontrada através da minimização do erro com restrição de igualdade, ou seja, $\arg \min_{\mathbf{w}} J_1$ para $J_2 = \epsilon$, conforme Figura 2.

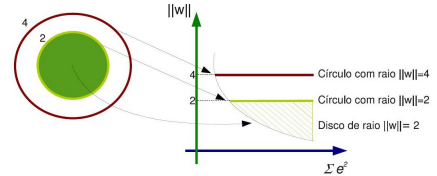


Figura 2. Representação no espaço dos objetivos $\sum e^2$ e $\|\mathbf{w}\|$ das restrições de desigualdade e igualdade para o problema de minimização do erro nas regiões $J_2 \leq \epsilon$ e $J_2 = \epsilon$.

A. Representação dos Pesos por Coordenadas Esféricas

Do ponto de vista da resolução do problema numérico de otimização, ambas as abordagens com restrições de igualdade e desigualdade requerem um esforço computacional para que a restrição seja obedecida durante a minimização do erro. É possível, no entanto, representar os pesos da rede através de coordenadas esféricas e, assim, garantir que a restrição seja atingida antes do problema de otimização ser resolvido. Desta forma, o problema de otimização resultante passa a ser mono-objetivo irrestrito, conforme será descrito a seguir.

A restrição $J_2 = \epsilon$ equivale a $\|\mathbf{w}\| = \epsilon$ ou $\sqrt{\sum_i w_i^2} = \epsilon$ e resulta na superfície da hiperesfera de raio ϵ , que pode ser definida através dos ângulos que compõem o vetor em relação a cada um dos eixos, conforme o conjunto de equações que se seguem.

$$\begin{aligned} w_0 &= \epsilon \sin \theta_1 \\ w_1 &= \epsilon \cos \theta_1 \sin \theta_2 \\ w_2 &= \epsilon \cos \theta_1 \cos \theta_2 \sin \theta_3 \\ &\vdots \\ w_{n-1} &= \epsilon \cos \theta_1 \cdots \cos \theta_{n-2} \sin \theta_{n-1} \\ w_n &= \epsilon \cos \theta_1 \cdots \cos \theta_{n-2} \cos \theta_{n-1} \end{aligned}$$

Segundo esta representação, os ângulos $\theta_1 \cdots \theta_{n_p-1}$ terão uma correspondência direta com os valores dos pesos $w_0 \cdots w_{n_p}$ da rede. Assim, o problema passa a ser o de encontrar os valores dos ângulos que minimizam o erro caracterizado pela função J_1 , definindo o novo problema de otimização como mostrado em 4.

Com a geração da estrutura de uma esfera n-dimensional obtida pela formulação, tem-se $n - 1$ ângulos representando uma rede com n pesos. Nesta forma de representação do problema de otimização garante-se que os pesos da rede gerem soluções sobre a superfície da norma, além de se garantir também um limite no domínio das variáveis, onde um dos ângulos terá uma variação na faixa de 0 a 2π e os demais ângulos ficarão entre 0 e π , limitando assim a dimensão do espaço de busca no problema.

$$\arg \min_{\theta_1 \cdots \theta_{n_p-1}} J_1 \quad (4)$$

Com esta nova formulação, o problema de treinamento de redes neurais é descrito como vários problemas de minimização de erro com norma fixa visando gerar soluções aproximadas ao Pareto-Ótimo, conforme representado de forma esquemática na Figura 3. Uma outra abordagem para a aproximação do conjunto Pareto através do controle da trajetória no espaço de objetivos utilizando Controle por Modos Deslizantes e otimização com o algoritmo Levenberg-Marquadt foi descrita em trabalho anterior [11].

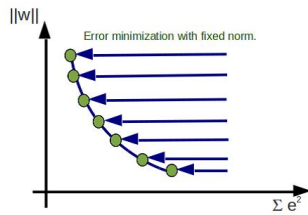


Figura 3. Aproximação da fronteira de Pareto através da minimização de erro com norma fixa.

B. Metodologia de Otimização na Nova Superfície de Busca

Considerando-se uma superfície de erro para um neurônio com função de ativação não linear, onde pode ser selecionada uma curva de erro em que o valor da norma dos seus parâmetros seja constante, obtém-se o esquema mostrado na Figura 4 em que são mostradas duas curvas de erro com valor constante de norma. O objetivo do treinamento é encontrar, no espaço limitado por cada uma destas curvas, o valor de erro mínimo. Denominamos esta metodologia como *busca na superfície da norma* (em inglês, *Norm Surface Search - NSS*).

Na Figura 4 a visualização da superfície de erro ocorre em função dos pesos da rede. Analisando sob a ótica da nova formulação, através dos ângulos que definem os valores dos pesos, tem-se uma superfície com uma dimensão a menos como pode ser visto na Figura 5.

Como pode ser visualizado, a superfície de erro torna-se mais complexa na medida em que são utilizados maiores valores de norma, onde a superfície inicial é suave e possui

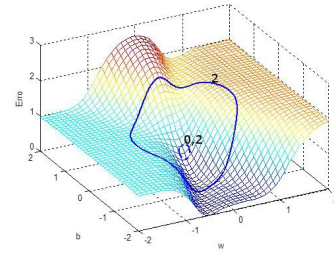


Figura 4. Curvas de erro para valores constantes de norma.

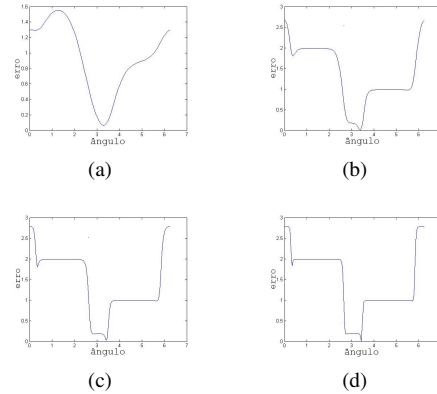


Figura 5. Superfície de erro em função dos ângulos. (a) Norma = 0.5, (b) Norma = 2, (c) Norma = 4 e (d) Norma = 8.

uma grande região convexa que vai cedendo lugar à platôs e mínimos locais mais íngremes à medida que o valor da norma aumenta. A superfície de erro se tornará ainda mais complexa na medida em que se aumentar a quantidade de dimensões do problema, o que dificultará o processo de otimização quando se utilizar algoritmos determinísticos baseados em direção de busca. Devido a esta dificuldade propõe-se uma estratégia para que seja possível manter as soluções iniciais do processo na região de interesse. Gera-se aleatoriamente valores de ângulos como solução inicial para o processo de otimização e, em seguida, inicia-se o processo de busca com um valor pequeno de norma. No passo seguinte, utiliza-se um valor maior de norma, mas mantem-se os valores de ângulos encontrados como solução final do processo anterior. Desta maneira, a solução encontrada numa superfície mais suave é projetada, no passo seguinte, para uma região próxima ao mínimo na nova superfície um pouco menos suave, tornando fundamental a utilização de pequenos passos para incremento da norma durante o processo. Como pode ser visto na Figura 5 a superfície geral de erro mantém um padrão, sendo que mesmo se tornando mais complexa possui um mesmo formato, fazendo com que com as projeções propostas mantenham as soluções encontradas para valores pequenos de norma na mesma região de interesse quando se utilizar valores maiores de norma. A Figura 6 ilustra este processo para diferentes valores de norma, onde o erro mínimo foi encontrado para o primeiro valor de norma, e as demais curvas mostram a projeção desta solução.

A utilização da formulação baseada em coordenadas esféricas torna o problema de otimização da rede irrestrito, dado que o controle de complexidade está acoplado à função objetivo necessitando-se apenas utilizar o processo para diferentes valo-

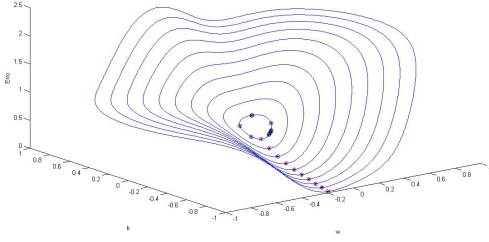


Figura 6. Valor de erro mínimo projetado para diversos valores de norma.

res de norma. Desta forma, pode-se utilizar qualquer algoritmo para otimização não linear para resolução do problema. Com o objetivo de gerar uma estimativa do conjunto Pareto, foi utilizado o método ϵ -restrito para o treinamento multi-objetivo da rede, de maneira semelhante à utilizada em [8].

A utilização de algoritmos baseados em gradiente permite ajustar uma rede com esta formulação apenas acrescentando a derivada dos pesos em relação aos ângulos, devido ao emprego da regra da cadeia. Pode-se verificar que as derivadas acrescentadas são facilmente calculadas dado que a formulação do problema consiste basicamente na utilização das funções trigonométricas seno e cosseno, que têm derivadas simples e conhecidas. Desta forma a direção de ajuste dos ângulos é definida através de:

$$d = \frac{\partial e}{\partial \vec{\sigma}} \quad (5)$$

onde e é o erro médio quadrático da rede e $\vec{\sigma}$ representa o vetor de ângulos. Com isso tem-se a seguinte equação para se encontrar a i -ésima coordenada do vetor direção d :

$$\frac{\partial e}{\partial \sigma_i} = \frac{\partial e}{\partial \vec{w}} \frac{\partial \vec{w}}{\partial \sigma_i} \quad (6)$$

onde \vec{w} é o vetor de pesos da rede. A busca para o novo valor de um determinado ângulo de forma que se minimize o erro quadrático da rede é realizada da seguinte forma:

$$\sigma_i^{k+1} = \sigma_i^k - \alpha d \quad (7)$$

onde k é a iteração corrente do treinamento e α é o tamanho do passo na direção de busca.

C. Validação da Formulação

Com o objetivo de verificar a consistência da nova formulação foram testados dois algoritmos de natureza distinta para otimização: o *Back-propagation* [15] e o Evolução Diferencial (DE) [16]. Estes métodos foram testados no problema de aproximação da função seno, onde verificou-se que o *Back-propagation* não gerou uma estimativa adequada para o Pareto além de demonstrar instabilidade no processo de otimização, como pode ser visualizado pela Figura 7. Esta instabilidade pode ser devido à grande interconexão entre os parâmetros da rede nesta formulação além das dificuldades da nova superfície. O DE demonstrou estabilidade no treinamento, gerando resultados promissores além de obter uma estimativa

aceitável do conjunto Pareto, cujo modelo selecionado pode ser visualizado na Figura 8(b). A Figura 8 mostra a aproximação obtida pelos métodos a partir da solução selecionada pelo decisor.

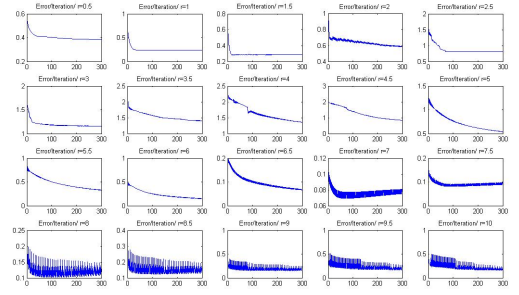


Figura 7. Comportamento do erro durante as iterações do *Back-Propagation* para valores crescentes de norma.

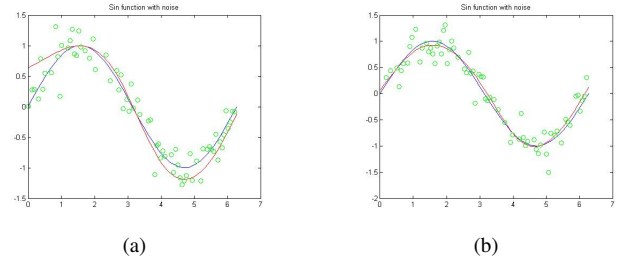


Figura 8. Aproximação da função seno obtida pelo modelo selecionado. (a) *Back-propagation*, (b) DE.

Devido ao melhor desempenho, o DE será utilizado no restante do trabalho como método para treinamento da rede com a formulação proposta. Os parâmetros utilizados no DE são os mesmos descritos em [17].

IV. RESULTADOS

Foram realizados testes para verificação da eficiência do método proposto, consistindo na comparação deste com dois outros métodos para treinamento multi-objetivo de RNAs: o MOBJ [8], pioneiro do treinamento multi-objetivo no domínio da complexidade aparente e o DEANN [17], desenvolvido para o treinamento multi-objetivo através da utilização do algoritmo evolução diferencial [16]. Estes métodos foram aplicados a problemas de regressão e classificação, onde medidas de estatística descritiva foram obtidas para avaliação do desempenho.

A solução selecionada no conjunto Pareto estimado é aquela que gera o menor erro para o conjunto definido como validação, sendo que a qualidade desta solução é avaliada com base no erro para um conjunto de testes. Estes dois conjuntos contêm dados desconhecidos ao treinamento. Os critérios utilizados para avaliação foram o erro médio quadrático (EMQ) e acurácia para os problemas de regressão e classificação respectivamente e o tempo médio de execução (TME) em ambos os tipos de problema, definido em segundos. Foram realizadas 10 execuções de cada algoritmo para validação dos métodos, onde tomou-se ao fim a média e o desvio padrão.

Nas simulações realizadas foram utilizados problemas de regressão tipicamente encontrados na literatura [11], [17] para

este propósito. As funções que descrevem cada problema são:

$$f1(x) = 4,26(e^{-x} - 4e^{-2x} + 3e^{-3x})$$

$$f2(x) = (x - 2)(2x + 1)/(1 + x^2)$$

As amostras de dados dos problemas de regressão foram obtidas a partir das funções geradoras com o acréscimo de um ruído gaussiano com média 0 e desvio padrão 0,2. Foram utilizadas três bases de dados para classificação: diabetes, cancer e coração com 8, 32 e 13 atributos por amostra, respectivamente, sendo todas obtidas em [18]. Nas simulações realizadas para as funções de regressão gerou-se resultados que demonstram o grau de aproximação obtido para cada um dos métodos avaliados.

Nas Figuras 9(a) e 9(b) são mostrados os modelos obtidos para as funções $f1$ e $f2$ para cada método testado, sendo também apresentados na figura a função geradora e os dados de treinamento utilizados no processo. As curvas exibidas são resultantes de soluções escolhidas pelo decisor de validação.

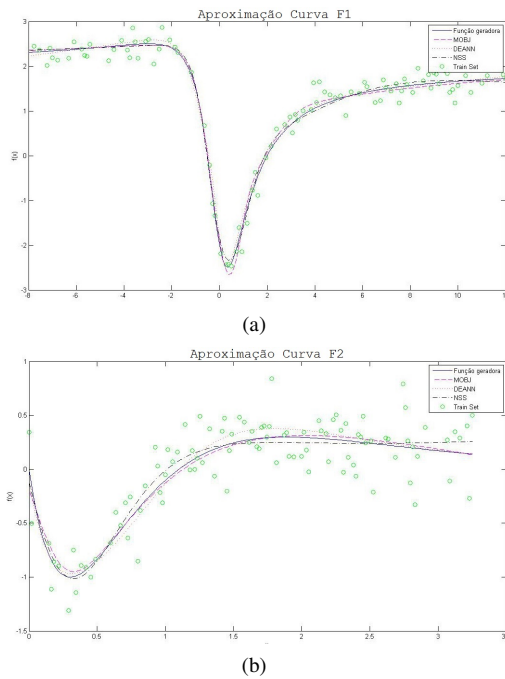


Figura 9. Aproximação das funções obtida pelos métodos. (a) $f1$, (b) $f2$.

A Tabela I mostra o erro médio quadrático para os dados de teste obtidos em cada problema de regressão, considerando-se todos os métodos avaliados. Pode-se visualizar nesta tabela a média(Md) do erro, o desvio padrão(Dp) e os valores mínimo(Mn) e máximo(Mx), dadas as execuções realizadas. Na Tabela II pode ser visualizado o tempo médio em segundos para execução dos métodos avaliados para as funções de regressão, considerando-se as mesmas medidas estatísticas utilizadas na avaliação do erro.

As Tabelas III e IV exibem os indicadores relativos à acurácia e tempo médio de execução para todas as bases de classificação utilizando cada método.

Através da Tabela I é verificado que a média do erro para regressão foi bem semelhante em todos os métodos, mostrando

Tabela I. EMQ OBTIDO PARA OS PROBLEMAS DE REGRESSÃO UTILIZANDO CADA MÉTODO

		MOBJ	DEANN	NSS
$f1$	Md	0,0330	0,0402	0,0373
	Dp	0,0030	0,0015	0,0073
	Mn	0,0295	0,0382	0,0314
	Mx	0,0395	0,0424	0,0549
$f2$	Md	0,0369	0,0375	0,0368
	Dp	0,0003	0,0005	0,0012
	Mn	0,0363	0,0367	0,0342
	Mx	0,0378	0,0384	0,0377

Tabela II. TME OBTIDO PARA OS PROBLEMAS DE REGRESSÃO UTILIZANDO CADA MÉTODO

		MOBJ	DEANN	NSS
$f1$	Md	8,1956	9,5036	160,5251
	Dp	0,4759	30,0530	0,4773
	Mn	7,5986	1,2313	159,8014
	Mx	8,9573	95,0398	161,4952
$f2$	Md	8,0325	9,6034	161,0197
	Dp	0,3023	30,3687	0,4333
	Mn	7,6093	1,1287	160,1391
	Mx	8,5076	97,1342	161,7486

uma equivalência dos métodos avaliados no quesito eficácia. Em relação ao tempo computacional, verifica-se através da Tabela II que o método DEANN foi equivalente ao MOBJ, ambos obtiveram tempos bem próximos, no entanto o DEANN teve um desvio padrão elevado, o que pode ter ocorrido devido à natureza estocástica do treinamento. O NSS neste caso foi o pior método devido ao alto custo computacional encontrado. Um possível fator que pode ter influenciado para o alto custo computacional do NSS, foi o aumento de complexidade no cálculo de função devido à camada de complexidade adicional inerente à formulação introduzida neste método. Através das Figuras 9(a) e 9(b) verifica-se que os métodos também foram equivalentes nos modelos escolhidos. Todos se aproximam da função geradora, analisando-se apenas de forma visual.

Nas simulações para as tarefas de classificação, pode-se ver através da Tabela III que os métodos propostos obtiveram maior valor de acurácia do que o MOBJ, exceto para a base do câncer. Este resultado pode ter ocorrido devido à utilização do algoritmo DE no treinamento do DEANN e do NSS, que torna o problema mais difícil de ser resolvido devido à base do câncer ter a maior quantidade de dimensões, gerando elementos maiores e mais difíceis de se trabalhar nos algoritmos evolutivos. Em relação ao maior valor de acurácia obtido o NSS foi melhor em duas das três bases de dados. Em

Tabela III. ACURÁCIA OBTIDA PARA AS BASES DE CLASSIFICAÇÃO UTILIZANDO CADA MÉTODO

		MOBJ	DEANN	NSS
Diabetes	Md	0,7491	0,7783	0,7765
	Dp	0,0036	0,0065	0,0062
	Mn	0,7435	0,7696	0,7609
	Mx	0,7522	0,7913	0,7826
Câncer	Md	0,9882	0,9512	0,9819
	Dp	0,0000	0,0028	0,0033
	Mn	0,9882	0,9471	0,9766
	Mx	0,9882	0,9529	0,9883
Coração	Md	0,8259	0,8642	0,8889
	Dp	0,0039	0,0058	0,0101
	Mn	0,8148	0,8519	0,8765
	Mx	0,8272	0,8765	0,9012

Tabela IV. TME OBTIDO PARA AS BASES DE CLASSIFICAÇÃO UTILIZANDO CADA MÉTODO

		MOBJ	DEANN	NSS
Diabetes	Md	2.0582	18.8480	64.5260
	Dp	0.1607	59.6027	12.2158
	Mn	1.8571	1.1254	46.2044
	Mx	2.4168	189.5804	84.2729
Câncer	Md	9.8125	23.6872	168.1766
	Dp	1.0284	74.9055	4.4819
	Mn	8.1842	2.3471	165.4678
	Mx	11.3090	239.4719	178.5135
Coração	Md	1.5940	15.3864	75.4833
	Dp	0.2328	48.6560	0.3841
	Mn	1.3763	2.3710	75.0370
	Mx	2.0017	156.1638	76.0353

relação ao tempo computacional o MOBJ superou o DEANN e o NSS, sendo que a utilização do DE para o treinamento dos métodos pode ter contribuído para este resultado.

A saída dos métodos é demonstrada na Figura 10 onde pode-se visualizar a estimativa do Pareto obtida para cada método em um problema de regressão e outro de classificação.

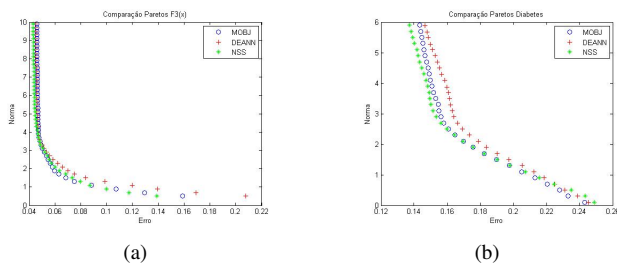


Figura 10. Estimativa do Pareto com cada método. (a) f_2 , (b) diabetes.

V. CONCLUSÕES

Neste trabalho foi apresentado um novo método para o treinamento multi-objetivo de RNAs, que foi comparado aos métodos MOBJ e DEANN. O método demonstra uma nova forma de treinamento que evita o tratamento de restrições de forma direta no algoritmo de otimização utilizado para o treinamento da rede. Os resultados obtidos com os testes foram promissores. Mesmo com custo computacional maior que as abordagens similares, as soluções apresentam alta generalização.

O NSS além de ter obtido bons resultados nas simulações trouxe novas características para o treinamento de RNAs, onde devido à utilização de coordenadas esféricas para representação dos pesos da rede gerou-se uma nova região de busca que deve ser melhor explorada, sendo que qualquer método de otimização não linear está habilitado a esta tarefa. A não obtenção de bons resultados durante a validação do método com o *back-propagation* não elimina a possibilidade da aplicação de outros métodos determinísticos, pois devido à facilidade em se obter derivadas com a nova formulação, o objetivo torna-se buscar um método adequado à superfície de busca do problema gerado. Em relação à estimativa do Pareto, o NSS mostrou-se tão consistente quanto os outros métodos analisados, sendo que a resolução do Pareto em relação a um dos objetivos é facilmente controlada devido à hiperesfera gerada.

Trabalhos futuros poderiam propor novos algoritmos para o treinamento da rede baseada em coordenadas esféricas. Mé-

todos determinísticos deveriam ser melhor explorados devido à facilidade de se obter derivadas e, possíveis dificuldades geradas pela camada de complexidade adicionada à superfície de busca, poderiam ser minimizadas com a utilização de métodos que visam a convexificação desta superfície como visto em [19], [20], o que facilitaria a utilização de métodos que pressupõem que a região de busca seja convexa para que o processo de otimização seja eficiente.

REFERÊNCIAS

- [1] V. N. Vapnik, *The nature of statistical learning theory*. Springer Verlag New York, 1995.
- [2] —, *Statistical Learning Theory*. Wiley-Interscience, 1998.
- [3] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Comput.*, vol. 4, no. 1, pp. 1–58, 1992.
- [4] P. Bartlett, "For valid generalization the size of the weights is more important than the size of the network," *Advances in Neural Information Processing Systems*, vol. 9, pp. 134–140, 1997.
- [5] M. Stone, "Cross-validated choice and assessment of statistical predictions," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 111–147, 1974.
- [6] R. Reed, "Pruning algorithms: A survey," vol. 4, no. 5, pp. 740–746, 1993.
- [7] G. E. Hinton, "Connectionist learning procedures," *Artificial intelligence*, vol. 40, no. 1-3, pp. 185–234, 1989.
- [8] R. H. T. R. A. Teixeira, A. P. Braga and R. R. Saldanha, "Improving generalization of mlps with multi-objective optimization," *Neurocomputing*, vol. 35, no. 1, pp. 189–194, 2000.
- [9] M. A. Costa, A. P. Braga, and B. R. de Menezes, "Improving neural networks generalization with new constructive and pruning methods," *J. Intell. Fuzzy Syst.*, vol. 13, no. 2-4, pp. 75–83, 2002.
- [10] M. A. Costa, A. P. Braga, B. R. de Menezes, R. Teixeira, and G. Parma, "Training neural networks with a multi-objective sliding mode control algorithm," *Neurocomputing*, vol. 51, pp. 467–473, 2002.
- [11] A. P. B. M. A. Costa and B. R. Menezes, "Improving generalization of mlps with sliding mode control and the levenberg-maquardt algorithm," *Neurocomputing*, vol. 70, no. 7-9, pp. 1342–1347, 2007.
- [12] Y. Jin and B. Sendhoff, "Pareto-based multiobjective machine learning: An overview and case studies," *IEEE Transactions on Systems Science and Cybernetics*, vol. 39, no. 3, p. 373, 2009.
- [13] I. Kokshenev and A. P. Braga, "An efficient multi-objective learning algorithm for rbf neural network," *Neurocomputing*, vol. 37, no. 16-18, pp. 2799–2808, 2010.
- [14] E. G. Horta, A. P. Braga, and R. R. Saldanha, "Acelerando o treinamento multiobjetivo de rnas pelo método de gradiente projetado," in *Congresso Brasileiro de Automática, Setembro, 2008*.
- [15] D. Rumelhart and J. McClelland, *Parallel Distributed Processing, vol1: Foundations*. The MIT Press, 1986.
- [16] R. M. Storn and K. V. Price, "Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces," *Technical Report TR-95-012, International Computer Science Institute*, p. 22, 1995.
- [17] H. P. Rocha, C. L. Castro, and A. P. Braga, "Seleção de modelos neurais utilizando evolução diferencial através do controle de erro e norma do vetor de pesos," in *X Congresso Brasileiro de Inteligência Computacional (CBIC2011)*, November 2011.
- [18] C. L. Blake and C. J. Merz. (1998) Uci repository of machine learning databases. Irvine, CA: University of California, Dept. of Information and Computer Science. [Online]. Available: <http://www.ics.uci.edu/mllearn/MLRepository.html>
- [19] D. P. Bertsekas, "Convexification procedures and decomposition methods for nonconvex optimization problems," *Journal of Optimization Theory and Applications*, vol. 29, no. 2, pp. 169–197, 1979.
- [20] D. Li, "Zero duality gap for a class of nonconvex optimization problems," *Journal of Optimization Theory and Applications*, vol. 85, no. 2, pp. 309–324, 1995.