

ESTRATÉGIA DE DECISÃO BASEADA EM MARGEM PARA O APRENDIZADO MULTI OBJETIVO DE REDES NEURAIIS

Luiz C.B. Torres, Cristiano L. Castro, Antonio P. Braga

Programa de Pós-Graduação em Engenharia Elétrica

Universidade Federal de Minas Gerais

Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brasil

{luizbambirra,crislcastro}@gmail.com, apbraga@ufmg.br

Resumo – O presente artigo apresenta uma nova estratégia de decisão para o aprendizado multiobjetivo de redes neurais artificiais. O objetivo é encontrar no conjunto pareto-ótimo, a solução que fornece a melhor capacidade de generalização. A abordagem proposta para a tomada de decisão é baseada em uma estimativa geométrica para a margem (distância) máxima de separação entre as classes, que é obtida através das seguintes etapas: modelagem dos padrões de entrada com o grafo de gabriel, detecção das bordas de separação das classes e síntese de padrões junto à região de margem máxima. Essa metodologia permite que modelos suaves (que ignoram ruído) e bem ajustados sejam selecionados de forma transparente para o usuário, ou seja, sem a necessidade da definição de parâmetros ou do uso de um conjunto representativo de validação. Resultados preliminares com dois *benchmarks* conhecidos na literatura mostraram que o decisor proposto, aliado ao treinamento multiobjetivo, foi eficiente no controle da generalização de modelos neurais.

Palavras-chave – tomada de decisão, aprendizado de máquina multiobjetivo, grafo de gabriel, classificação.

Abstract – This paper presents a new decision-making strategy to the multiobjective learning of artificial neural networks. The objective is to find the solution within the pareto-optimal set that has the best generalization performance. The proposed decision-making approach is based on a geometric approximation to the maximum margin (distance) of class separation, which is estimated through the following steps: modeling of input patterns using the gabriel graph, detection of class separation borders and synthesis of patterns along the maximum margin region. This methodology allows the selection of smooth (that ignore noise) and well-fitting models in a straightforward manner, i.e., without the need of the tuning of parameters by the user or the use of a representative validation data set. Preliminary results on two benchmarks in literature showed that our decision-making method, combined with multiobjective training, was efficient to control the generalization of neural models.

Keywords – decision-making, multiobjective machine learning, gabriel graph, classification.

1. INTRODUÇÃO

O problema da generalização em Redes Neurais Artificiais (RNAs) foi analisado formalmente por V. Vapnik [1, 2], que demonstrou, com o princípio indutivo de minimização estrutural do risco (MER), que para se obter uma solução eficiente para o problema do aprendizado é necessário minimizar dois objetivos conflitantes: o erro de treinamento e a capacidade (ou complexidade) da classe de funções fornecida pela máquina de aprendizagem. Tais objetivos não devem ser apenas minimizados, mas também equilibrados; caso contrário, o modelo resultante não generalizará bem. Essa idéia é análoga ao conhecido dilema entre a polarização e a variância, descrito inicialmente em [3].

Grande parte das técnicas para controle de generalização, tais como as redes de regularização [4], o método *weight decay* [5] e os algoritmos de poda (*pruning*) [6], minimizam ambos o erro e a complexidade implicitamente através de um único funcional custo, determinando seu equilíbrio através do ajuste de um ou vários parâmetros (como, por exemplo, o parâmetro que controla a regularização).

Não obstante, a formulação multiobjetivo do aprendizado (MOBJ) fornece uma abordagem alternativa para a implementação do princípio MER através da minimização explícita (separada) do erro de treinamento e de uma medida que reflete a complexidade da rede [7–10]. Sabe-se porém, que na abordagem MOBJ não é possível minimizar esses objetivos simultaneamente, pois o ótimo de um funcional raramente corresponde ao ótimo do outro. Assim, não existe um único ótimo, mas um conjunto deles, que formam o conjunto Pareto-Ótimo (*PO*). A formulação MOBJ para o treinamento de RNAs resulta então em conjunto de soluções (*PO*) que corresponde aos melhores compromissos entre os funcionais erro e complexidade.

Uma vez obtido o conjunto *PO*, a escolha da solução final (através de um decisor) constitui uma etapa mais crítica do algoritmo MOBJ. De acordo com o princípio MER, a solução escolhida deve fornecer um equilíbrio adequado entre os funcionais minimizados, para evitar (sub)sobreajuste do modelo aos dados de treinamento. Estratégias para seleção da solução final no aprendizado MOBJ têm sido propostas, tais como o decisor por mínimo erro de validação [8] e o decisor baseado em conhecimento prévio [11]. Cabe ressaltar, no entanto, que a eficiência desses decisores pode ficar limitada em situações em que o

conjunto de dados é muito pequeno e informação a priori sobre o processo de amostragem dos dados não se encontra disponível. Infelizmente, essas características são frequentes em problemas reais de aprendizado.

Para superar essas dificuldades, esse artigo apresenta uma nova estratégia de decisão direcionada a problemas de classificação de padrões. Usando ferramentas da Geometria Computacional [12], o decisor proposto busca pela solução do conjunto PO que possui a maior margem (ou distância) de separação entre as classes. Um estudo experimental preliminar com dois *benchmarks* conhecidos na literatura mostram que a abordagem é promissora. Os resultados obtidos são comparados ao decisor por mínimo erro de validação, comumente usado por algoritmos MOBJ.

O restante do artigo encontra-se organizado da seguinte forma. A Seção 2 descreve os conceitos teóricos que fundamentam o problema da seleção de modelos no treinamento multiobjetivo de RNAs: Minimização Estrutural do Risco (MER), Aprendizado Multiobjetivo e o Problema de Decisão Multiobjetivo. Em seguida, na Seção 3, nossa estratégia de decisão baseada em margem é apresentada. Na Seção 4, são descritos a metodologia adotada na condução dos experimentos e os resultados obtidos com a aplicação do novo decisor. Finalmente, a Seção 5 traz as conclusões.

2. REFERENCIAL TEÓRICO

2.1 Minimização Estrutural do Risco

A teoria do aprendizado estatístico, através do princípio indutivo de minimização estrutural do risco (MER) estabelece condições matemáticas que permitem definir, com probabilidade de pelo menos $1 - \epsilon$ sobre N (tamanho do conjunto de treinamento), um limite superior para o risco esperado (ou erro de generalização) de uma máquina de aprendizagem [1, 2],

$$R \leq R_{emp} + \sqrt{\frac{h \left(\ln \frac{2N}{h} + 1 \right) - \ln \left(\frac{\epsilon}{4} \right)}{N}} \quad (1)$$

Analisando (1), observa-se que o limite superior de R é uma função inversa de N e direta de dois termos: o primeiro, denominado risco empírico (R_{emp}), representa o erro de treinamento e o segundo, conhecido como capacidade (Ω), depende da complexidade h da classe de funções implementada pela máquina de aprendizagem. A minimização do limite superior de R pode então ser obtida através do aumento do número de exemplos N e/ou do decréscimo simultâneo de: R_{emp} e Ω .

2.2 Aprendizado Multiobjetivo

O princípio de minimização estrutural do risco (MER) pode ser interpretado como um problema de otimização multiobjetivo que busca encontrar o melhor compromisso entre dois objetivos conflitantes,

$$(MER) : \min \begin{cases} J_1 = R_{emp} \\ J_2 = \Omega \end{cases} \quad (2)$$

onde R_{emp} corresponde a uma estimativa para o erro de treinamento e Ω é uma medida de complexidade da máquina de aprendizagem. No caso particular de redes *MultiLayer Perceptron* (MLP), uma medida de complexidade (Ω) comumente usada é a norma euclidiana do vetor de pesos da rede [5, 8, 9].

Dentre os algoritmos de treinamento multiobjetivo para redes MLP, destaca-se o método MOBJ que foi projetado para resolver o problema MER descrito em (2) [8]. Dado o conjunto de dados $T = \{\mathbf{x}_i, y_i \mid i = 1 \dots N\}$, a Equação (3) a seguir, fornece a formulação biobjetivo, proposta em [8], para o aprendizado de redes MLP,

$$\min \begin{cases} J_1(\mathbf{w}) = \sum_{i=1}^N \left(y_i - \hat{f}(\mathbf{x}_i, \mathbf{w}) \right)^2 \\ J_2(\mathbf{w}) = \|\mathbf{w}\| \end{cases} \quad (3)$$

onde $\hat{f}(\mathbf{x}_i, \mathbf{w})$ é saída estimada pela rede para o i -ésimo padrão de entrada, y_i é a saída esperada (rótulo), \mathbf{w} é o vetor que armazena todos os pesos da rede e, $\|\cdot\|$ é o operador que fornece a norma euclidiana de um vetor.

Ao final do aprendizado, o algoritmo MOBJ gera uma estimativa para o conjunto de soluções não dominadas¹ denominado conjunto Pareto-ótimo (PO). Tais soluções representam o *trade-off* entre o erro de treinamento e a complexidade da rede.

Na ausência de qualquer informação a priori referente aos objetivos $J_1(\mathbf{w})$ e $J_2(\mathbf{w})$, todas as soluções pertencentes ao conjunto PO são candidatas à solução do problema descrito em (3). Uma etapa de decisão (decisor) é então necessária para a escolha da solução que fornece o melhor compromisso entre o erro de treinamento e a complexidade da rede. Tal solução constitui a melhor aproximação para o mínimo absoluto do funcional risco esperado R (ou erro de generalização).

¹Em um problema de otimização multiobjetivo, uma solução é dita ser não dominada, se não existe nenhuma solução com desempenho superior a ela em todos os objetivos.

2.3 Problema de Decisão Multiobjetivo

O problema geral de decisão multiobjetivo deve obedecer o esquema dado pela Equação (4) [11],

$$\mathbf{w}^* = \arg \max_{\mathbf{w} \in \mathcal{W}} f_e \quad (4)$$

onde f_e é um funcional capaz de classificar as soluções componentes do conjunto Pareto-Ótimo segundo um critério especificado; f_e deverá fornecer uma medida de “qualidade” entre uma dada solução (função) particular, $\hat{f}(\mathbf{x}, \mathbf{w})$, obtida com o aprendizado multiobjetivo (MOBJ) e a função desconhecida que representa o mínimo absoluto do risco esperado, $f_0(\mathbf{x})$.

Dentre as estratégias de decisão multiobjetivo propostas na literatura destacam-se: (i) o decisor por mínimo erro de validação [8], onde a tomada de decisão é feita através de um conjunto de validação apresentado a todas as soluções pertencentes ao conjunto PO . A rede que apresentar o menor erro aos padrões de validação é escolhida como solução final. A desvantagem desse método é a necessidade de se separar parte do conjunto de dados para posterior validação dos modelos. Isso representa um problema para tarefas de aprendizado baseadas em um número muito limitado de exemplos. (ii) o decisor com conhecimento prévio [11] que realiza um teste estático para quantificar a probabilidade de um modelo de classificação ser o melhor comparado com os outros do conjunto PO . A formulação desse decisor depende de uma informação a priori sobre a distribuição do ruído presente nos dados. Na maioria dos problemas reais de aprendizado, no entanto, essa informação não se encontra disponível.

3. DECISOR BASEADO EM MARGEM

O decisor proposto nesse artigo foi projetado para problemas de classificação. Nesse contexto, parte-se da premissa que, dentre as soluções do conjunto Pareto-Ótimo (PO), deve-se selecionar aquela que ignora o ruído presente nos dados e maximiza a margem (ou distância) de separação entre as classes.

A implementação dessa estratégia de decisão é baseada nas seguintes etapas: eliminação de dados ruidosos, detecção da borda das classes e síntese de padrões junto à região que representa a margem de separação. Isso é obtido modelando-se os padrões de entrada como um grafo de proximidade, conhecido como Grafo de Gabriel [12]. Antes de detalhar o algoritmo de decisão proposto, é necessário definir o Grafo de Gabriel com base nos fundamentos da Geometria Computacional [12].

3.1 Diagrama de Voronoi

O Diagrama de Voronoi representa a decomposição de um espaço em regiões de acordo com a distância entre determinados pontos [13]. Dados dois pontos $p, q \in S$, onde S representa um conjunto de pontos em um plano, a bissetriz $B(p, q)$ corresponde a uma reta perpendicular que atravessa o centro do segmento de reta \overline{pq} ; $\delta(\cdot)$ é o operador que fornece a distância euclidiana entre dois pontos (vetores). O diagrama de Voronoi ($Vor(S)$) pode ser considerado como a divisão do plano em m polígonos convexos P [12, 14]. A Figura 1(a) ilustra um exemplo. Um polígono $P(x_i)$ é chamado de polígono de Voronoi relativo a x_i e é formado através da intersecção do conjunto das bissetrizes $B(x_i, P(x_i))$. Um ponto $p \in S$ pertence a $P(x_i)$ se somente se a seguinte desigualdade for satisfeita,

$$\delta(p, x_i) \leq \delta(p, x_j), x_i, x_j \in S, j \neq i \quad (5)$$

3.2 Triangulação de Delaunay

No diagrama de Voronoi $Vor(S)$ cada elemento $v \in S$ está associado a um polígono de $Vor(S)$. O grafo dual de $Vor(S)$ tem por vértices os elementos de S e por arestas os pares de elementos de S cujos polígonos de $Vor(S)$ são vizinhos [15, 16]. Tal diagrama resultante é chamado de Diagrama de Delaunay [12, 14], conforme ilustrado na Figura 1(b). A Triangulação de Delaunay (TD) é uma subdivisão de um objeto geométrico em triângulos. A TD pode ser modelada através de um grafo planar G_p composto por um conjunto de vértices V e um conjunto de arestas E não ordenadas, como pode ser visto na Figura 1(c). Uma aresta $(v_i, v_j) \in G_p$ é definida se e somente se existir um círculo contendo o par (v_i, v_j) e que todos os outros vértices de V sejam exteriores a este círculo. Essa característica é ilustrada na Figura 1(d).

3.3 Grafo de Gabriel

O grafo de Gabriel G_G é um subconjunto de pontos do Diagrama de Voronoi e também um subgrafo da Triangulação de Delaunay [17], ou seja, $G_G \subseteq T_D \subseteq Vor$. Segundo [12], o Grafo de Gabriel G_G de um conjunto de pontos S , é um grafo cujo conjunto de vértices $V = S$ e seu conjunto de arestas E deve obedecer à seguinte definição,

$$(v_i, v_j) \in E \leftrightarrow \delta^2(v_i, v_j) \leq [\delta^2(v_i, z) + \delta^2(v_j, z)] \quad \forall z \in V, v_i, v_j \neq z \quad (6)$$

o que implica que para (v_i, v_j) constituir uma aresta de G_G , não pode haver nenhum outro vértice dentro do círculo cujo o diâmetro é a distância euclidiana entre v_i e v_j .

O KNN (*K-Nearest Neighbors*) é um método de classificação cuja idéia principal é classificar um elemento no espaço n -dimensional de acordo com a classe dominante de seus K vizinhos mas próximos obtidos no conjunto de treinamento. Um dos problemas encontrados no método é encontrar o melhor valor para o parâmetro K . O Grafo de Gabriel e o Diagrama de Voronoi

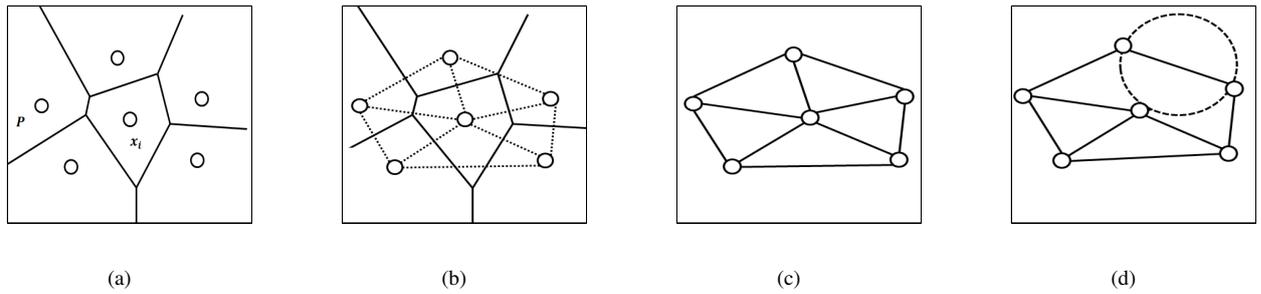


Figura 1: (a) Diagrama de Voronoi. (b) Grafo Dual do Diagrama de Voronoi. (c) Triangulação de Delaunay resultante. (d) Par de pontos de Voronoi que possuem uma aresta em comum representado em uma Triangulação de Delaunay.

também podem ser usados para esse propósito, mas diferente do KNN não necessitam do parâmetro K . Como mencionado na Seção 3.1, um ponto no Diagrama de Voronoi representa um polígono convexo cujas fronteiras representam seus vizinhos mais próximos. Assim, a fronteira que separa dois pontos de diferentes classes pode ser usada como uma fronteira de decisão para classificação.

3.4 Algoritmo de Decisão Proposto

O decisor proposto nesse artigo busca pela solução w^* do conjunto PO mais próxima a pontos distribuídos dentro da margem de separação entre as classes. Seu algoritmo pode ser formulado de acordo com os seguintes passos:

1. A partir de um conjunto de dados $T = \{\mathbf{x}_i, y_i \mid i = 1 \dots N\}$, é realizado o treinamento multiobjetivo (MOBJ) da rede MLP, conforme descrito na Seção 2.2, e o conjunto Pareto-Ótimo (PO) é obtido.
2. Obtém-se o Grafo de Gabriel G_G com o conjunto de vértices sendo formado por todos os padrões de entrada, i.e., $V = \{\mathbf{x}_i \mid i = 1 \dots N\}$, e o conjunto de arestas E satisfazendo a condição estabelecida em (6). Veja exemplo na Figura 2(a).
3. Este passo do algoritmo é responsável por detectar e remover o ruído dos dados. Para todo $\mathbf{x}_i \in V$, analisa-se o subgrafo induzido pelo vértice \mathbf{x}_i , ou seja, o subgrafo formado pelas arestas que possuem \mathbf{x}_i como uma das extremidades. Se a maioria dos vizinhos (vértices adjacentes de \mathbf{x}_i) possui rótulo diferente de y_i , então \mathbf{x}_i é considerado como ruído e eliminado de V . Esse passo deve ser repetido até que não haja mais exclusão. A Figura 2(b) ilustra o conjunto de dados após a eliminação dos dados ruidosos.
4. A borda B_r das classes é encontrada da seguinte forma: seja $\mathbf{x}_i \in V$ e, $\forall \mathbf{x}_j \in V$ com $j \neq i$, caso a aresta $(\mathbf{x}_i, \mathbf{x}_j)$ for formada por vértices de classes distintas (com rótulos diferentes), então ela é incluída no conjunto B_r , conforme ilustrado pela Figura 2(c).

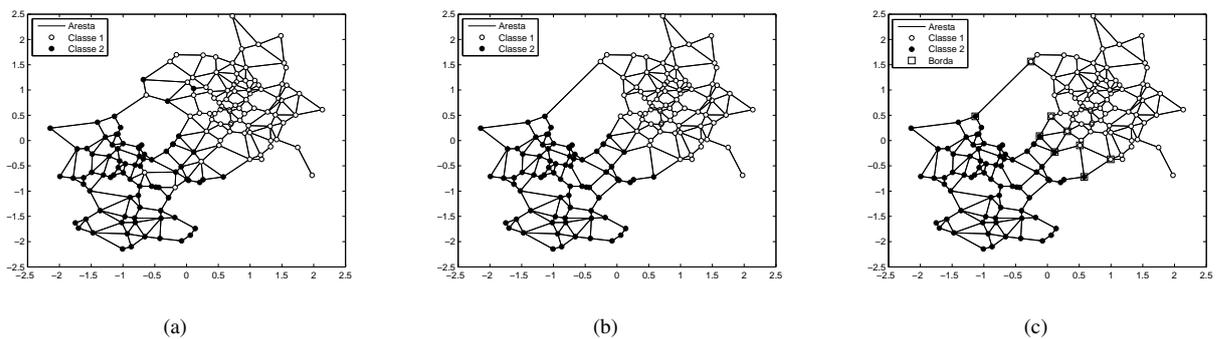


Figura 2: (a) Conjunto de padrões de entrada modelado com o Grafo de Gabriel. (b) Conjunto resultante após a eliminação dos dados ruidosos. (c) Detecção da borda das classes.

5. Neste passo são calculados os pontos médios entre os exemplos que compõem as bordas das classes. Para cada aresta $(\mathbf{x}_i, \mathbf{x}_j)$ pertencente ao conjunto B_r , calcula-se o ponto médio entre os vértices \mathbf{x}_i e \mathbf{x}_j de acordo com a seguinte expressão,

$$\bar{\mathbf{x}}_{ij} = \sum_{t=1}^n \mu(\mathbf{x}_i(t), \mathbf{x}_j(t)) \quad (7)$$

onde n é número de atributos (características) dos padrões de entrada e $\mu(\cdot)$ é o operador que calcula a média para o t -ésimo atributo. Após o cálculo dos pontos médios para todas as arestas de B_r , obtém-se um conjunto de ponto médios P_M relativo à bordas das classes, conforme ilustrado pela Figura 3(a).

6. Seja $W = \{\mathbf{w}_k \mid k = 1 \dots L\}$ o conjunto de soluções do conjunto PO. Para se obter a solução que maximiza a margem de separação entre as classes, o decisor deve escolher a solução $\mathbf{w}_k \in W$ mais próxima dos pontos médios (P_M) calculados no passo anterior. Isso é obtido calculando-se, para cada \mathbf{w}_k , a distância entre os vetores de entrada que se encontram exatamente na superfície de decisão s_k (estimada a partir de \mathbf{w}_k) e os vetores pertencentes ao conjunto P_M , conforme as expressões a seguir,

$$d_k(j) = \min \delta(s_k, P_M(j)), \quad \forall j \in P_M \quad (8)$$

$$D_k = \sum_{j=1}^{|P_M|} d_k(j) \quad (9)$$

onde $d_k(j)$ fornece a distância entre o j -ésimo ponto médio ($P_M(j)$) e o ponto pertencente à superfície de decisão s_k que se encontra mais próximo de $P_M(j)$; $|P_M|$ é o número de pontos médios nas bordas das classes e $\delta(\cdot)$ é o operador que fornece a distancia euclidiana entre dois vetores. Finalmente, a solução escolhida pelo decisor proposto deve ser,

$$\mathbf{w}^* = \min D_k \quad (10)$$

onde \mathbf{w}^* é a solução mais próxima dos pontos médios, ou seja, a solução que possui a maior margem de separação entre as classes. A Figura 3(b) ilustra algumas superfícies de decisões geradas a partir de diferentes soluções do conjunto PO. A superfície de decisão mais próxima dos pontos médios é mostrada pela Figura 4(a). A solução correspondente no conjunto PO é marcada na Figura 4(b).

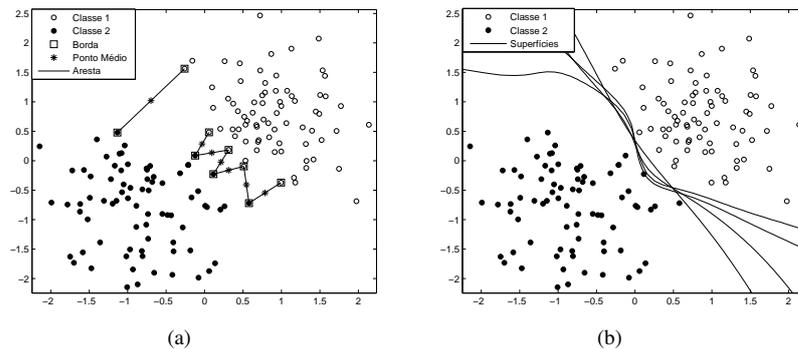


Figura 3: (a) Pontos médios relativo às bordas das classes. (b) Superfícies de decisão geradas a partir de soluções de PO.

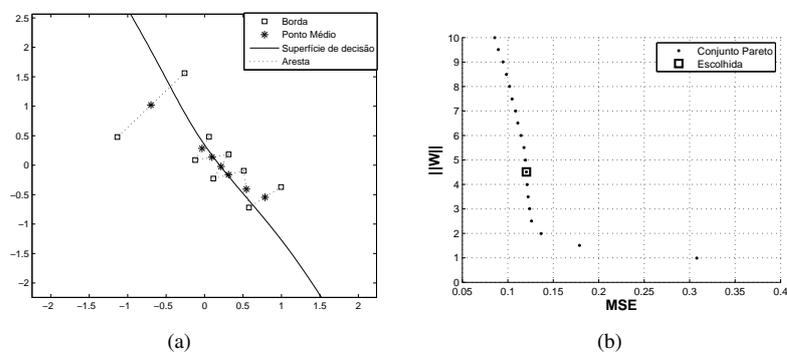


Figura 4: (a) Superfície de decisão mais próxima dos pontos médios. (b) Conjunto Pareto-Ótimo.

4 EXPERIMENTOS E RESULTADOS

Com o objetivo de verificar a eficiência da estratégia de decisão proposta no aprendizado multiobjetivo de redes MLP, experimentos foram conduzidos com dois problemas (*benchmarks*) de classificação conhecidos na literatura: “tabuleiro de xadrez” e “duas luas”.

4.1 Tabuleiro de Xadrez

O “tabuleiro de xadrez” é um problema multimodal em que as classes (círculos preenchidos e não-preenchidos) encontram-se distribuídas na forma de um tabuleiro 4x4, apresentando sobreposição, conforme ilustrado pela Figura 5. O número total de padrões de entrada é igual a 1600.

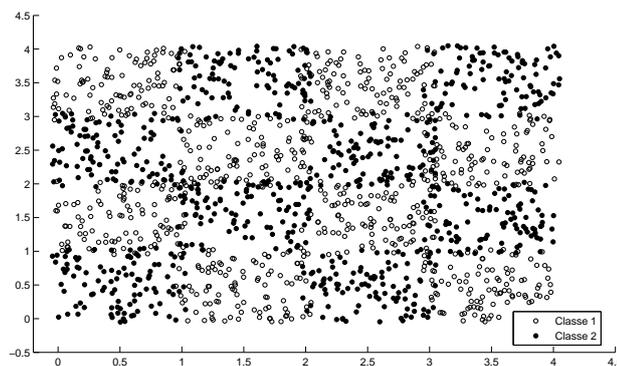


Figura 5: Problema do tabuleiro de xadrez.

O algoritmo de treinamento MOBJ [8] teve seus parâmetros ajustados da seguinte forma: o conjunto PO de soluções foi composto por 189 redes MLP com 30 neurônios na camada oculta. Foi utilizada a função de transferência tangente hiperbólica para a camada oculta e de saída. A diferença de norma euclidiana (complexidade) entre as soluções foi de $\Delta\|w\| = 0.5$.

O decisor baseado em margem foi aplicado para a escolha da solução PO . Tal decisor não possui parâmetros para ajuste e também não necessita do uso de um conjunto de novos exemplos para validação. Assim, todos os padrões puderam ser usados no treinamento MOBJ da rede MLP. A Figura 6 mostra a fase inicial do algoritmo decisor proposto com o conjunto de padrões de entrada modelado através do Grafo de Gabriel. O ruído foi eliminado e as bordas de separação entre as classes detectada. A Figura 7 mostra a superfície de decisão (estimada a partir de w^*) selecionada pelo decisor baseado em margem. Note a partir da Figura 7, que apesar da dificuldade do problema do “tabuleiro de xadrez”, a superfície de decisão escolhida apresenta certa suavidade (complexidade baixa), sendo capaz de separar corretamente as classes. Além disso, não se observa comportamento de *under/overfitting* em relação aos dados de treinamento.

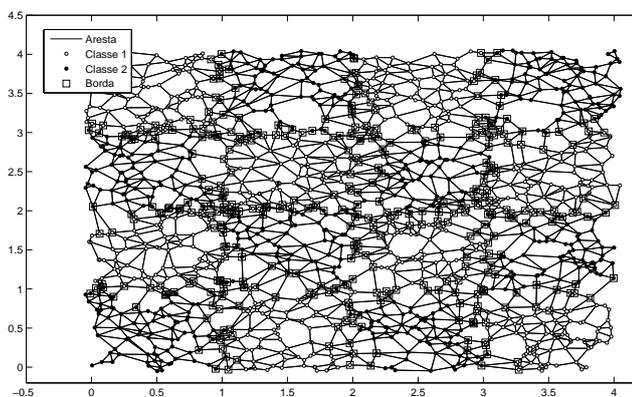


Figura 6: Modelagem do problema com o Grafo de Gabriel. O ruído dos dados foi eliminado e as bordas entre as classes foi detectada.

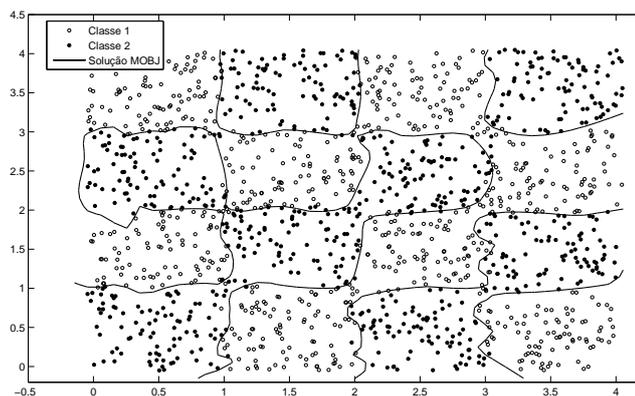


Figura 7: Solução (superfície de decisão) escolhida pelo decisor baseado em margem.

4.2 Duas Luas

A base de dados “duas luas” é não linearmente separável e possui 400 padrões de entrada divididos, igualmente, em duas classes (círculos preenchidos e não preenchidos), conforme ilustrado pela Figura 8(a).

Nesse experimento, os resultados obtidos com nossa estratégia de decisão foram confrontados com o decisor por mínimo erro de validação [8], comumente usado por algoritmos de aprendizado multiobjetivo. Para a aplicação desse decisor, o conjunto de dados original teve que ser particionado em 70% para treinamento e 30% para validação. O decisor baseado em margem pôde contar com 100% do conjunto de dados para treinamento da rede MLP.

No aprendizado com o algoritmo MOBJ, o conjunto PO de soluções foi composto por 19 redes MLP com 5 neurônios na camada oculta. Similarmente ao problema do “tabuleiro de xadrez”, foi utilizada a função de transferência tangente hiperbólica para a camada oculta e de saída. A diferença de norma euclidiana (complexidade) entre as soluções foi de $\Delta\|w\| = 0.5$.

A Figura 8(b) ilustra a aplicação do decisor proposto com os padrões entrada sendo modelados através do Grafo de Gabriel. A superfície de decisão selecionada (linha pontilhada) é mostrada na Figura 8(c). A Figura 8(d) compara as superfícies de decisão obtidas, respectivamente, pelo decisor de validação (linha contínua) e pelo decisor baseado em margem (linha pontilhada). Nota-se, pela Figura 8(d), que a solução escolhida por nosso decisor apresenta melhor margem (distância) de separação entre as classes que a solução obtida por validação. Esse resultado é importante e sugere que o decisor proposto, aliado ao algoritmo MOBJ, é eficiente no controle da capacidade de generalização da rede MLP.

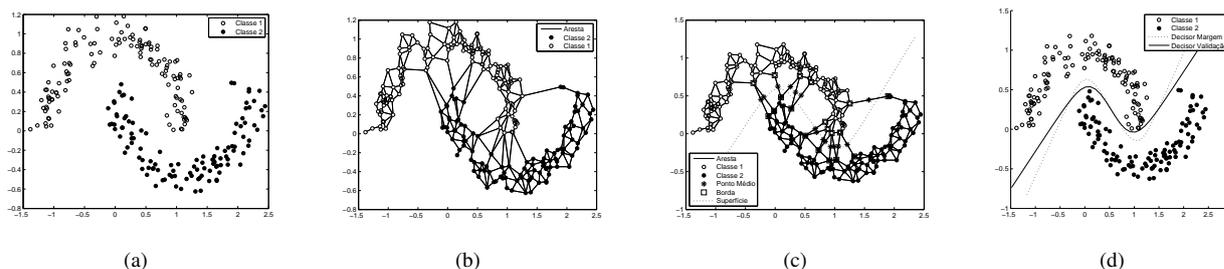


Figura 8: (a) Problema das duas luas. (b) Modelagem do problema “duas luas” com o Grafo de Gabriel. (c) Solução encontrada pelo decisor baseado em margem. (d) Decisor por validação (linha contínua) vs. decisor baseado em margem (linha pontilhada).

5. CONCLUSÃO

Conforme argumentado nas seções iniciais desse artigo, o problema do aprendizado de RNAs e de máquinas de aprendizagem de uma maneira geral é intrinsecamente multiobjetivo. As funções custo relativas aos ajustes do modelo aos dados (erro) e da complexidade do modelo a uma dada tarefa de aprendizagem são conflitantes. Como não existe um único mínimo global que satisfaça à minimização simultânea dessas funções, um processo de decisão que determinará o equilíbrio entre o erro e a complexidade é necessário. Em outras máquinas de aprendizagem como as *Support Vector Machines* (SVMs), por exemplo, a tomada de decisão ocorre quando o usuário define previamente os parâmetros de margem (regularização) e *kernel*. A definição desses parâmetros por si só é considerado um processo de tomada de decisão multiobjetivo que, geralmente, é feito através de uma busca exaustiva com base em um conjunto representativo de validação (*crossvalidation*).

No caso particular desse trabalho, partimos do princípio de que a solução se encontra no conjunto Pareto-Ótimo (PO) e que uma delas deve então ser escolhida. O processo de escolha aqui proposto se baseia em uma estimativa geométrica para a margem de separação entre as classes. A solução que se encontra mais próxima da margem máxima é selecionada. Com essa metodologia, os processos de treinamento (geração do conjunto PO) e escolha da solução eficiente ocorrem de forma transparente para o usuário, o que contrasta com outros modelos de aprendizado, tais como as SVMs e as redes de regularização, em que a solução é obtida com base em parâmetros fornecidos pelo usuário.

Resultados experimentais com *benchmarks* bidimensionais mostraram que o método de decisão proposto, aliado ao algoritmo MOBJ, foi eficiente para a obtenção de modelos suaves (que ignoram ruído) e bem ajustados. Acreditamos que a generalização dessa metodologia para problemas de maior dimensão pode abrir caminho para uma concepção mais geral e única na construção de modelos neurais artificiais.

REFERÊNCIAS

- [1] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.
- [2] V. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.
- [3] S. Geman, E. Bienenstock and R. Doursat. “Neural Networks and the Bias/Variance Dilemma”. *Neural Computation*, vol. 4, pp. 1–58, 1992.
- [4] F. Girosi, M. Jones and T. Poggio. “Regularization Theory and Neural Networks Architectures”. *Neural Computation*, vol. 7, pp. 219–269, 1995.
- [5] G. Hinton. “Connections Learning Procedures”. *Artificial Intelligence*, vol. 40, pp. 185–234, 1989.
- [6] R. Reed. “Pruning Algorithms: A Survey”. vol. 4, no. 5, pp. 740–746, 1993.
- [7] Y. Jin and B. Sendhoff. “Pareto-Based Multiobjective Machine Learning: An Overview and Case Studies”. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 39, no. 3, pp. 373, 2009.
- [8] R. A. Teixeira, A. P. Braga, R. H. C. Takahashi 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. Menezes. “Improving generalization of MLPs with sliding mode control and the Levenberg-Marquardt algorithm”. *Neurocomputing*, vol. 70, no. 7-9, pp. 1342–1347, 2007.
- [10] I. Kokshenev and A. P. Braga. “An efficient multi-objective learning algorithm for RBF neural network”. *Neurocomputing*, vol. 73, no. 16-18, pp. 2799 – 2808, 2010.
- [11] T. H. Medeiros, Takahashi and H. C. R. . A. P. Braga. “A Incorporação do Conhecimento Prévio na Tomada de Decisão do Aprendizado Multiobjetivo. In: Congresso Brasileiro de Redes Neurais”. *Congresso Brasileiro de Redes Neurais - Inteligencia Computacional*, vol. 9, pp. 25–28, 2009.
- [12] M. de Berg, M. van Kreveld, M. Overmars and O. Schwarzkopf. *Computational Geometry: Algorithms and Applications*. Springer-Verlag, second edition, 2000.
- [13] F. Aurenhammer and R. Klein. “Voronoi Diagrams”. In *Handbook of Computational Geometry*, pp. 152–159. Elsevier, 1990.
- [14] L. H. Figueiredo. *Introdução a Geometria Computacional*. Impa, Rio de Janeiro, 1991.
- [15] J. Li and C.-C. Kuo. “A dual graph approach to 3D triangular mesh compression”. In *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, volume 2, pp. 891–894 vol.2, oct 1998.
- [16] H. Zhang and X. He. “On simultaneous straight-line grid embedding of a planar graph and its dual”. *Information Processing Letters*, vol. 99, no. 1, pp. 1 – 6, 2006.
- [17] W. Zhang and I. King. “A study of the relationship between support vector machine and Gabriel graph”. In *Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on*, volume 1, pp. 239–244, 2002.