

Otimização do desempenho de um classificador por modificação no processo de seleção de características

Cristiano Luiz Silva Tavares^{1,2}, Adriano Vilela Barbosa²
cristianot@ifes.edu.br , adrianovilela@ufmg.br

¹ Núcleo de Pesquisa em Eletrônica e Sinais
Instituto Federal do Espírito Santo – Campus São Mateus
São Mateus, Brasil

² Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil

Resumo—Este trabalho apresenta um algoritmo de otimização para o processo de reconhecimento de padrões atuando diretamente na etapa de seleção de características. O algoritmo proposto é baseado em computação evolucionária e conta com o método ReliefF para seleção de características aliado ao método *Support Vector Machine* (SVM) para classificação. As variáveis de busca para otimização são os K-Vizinhos Mais Próximos utilizado no ReliefF e o número de características selecionadas para o classificador, ou seja, a dimensão trabalhada pelo SVM. O algoritmo proposto produziu melhorias de performance no classificador SVM. Testes foram realizados em duas bases de dados e os resultados obtidos, comparados com o resultado de outros três algoritmos, comprovam o bom desempenho do algoritmo proposto.

Palavras-chave—Algoritmo Genético, Otimização, Reconhecimento de Padrões, ReliefF, *Support Vector Machine*.

I. INTRODUÇÃO

O reconhecimento de padrões é a ciência com objetivo de classificar amostras em um número de categorias ou classes. Geralmente, estas amostras possuem longas cadeias de características, sendo que apenas algumas destas podem ser suficientes para uma boa classificação. O uso demorado de características pode, inclusive, prejudicar o processo de classificação [1]. Desta forma, uma das etapas primordiais no processo de reconhecimento de padrões é a seleção, ou extração, de características que resultem em uma boa classificação.

Além de eleger as características mais relevantes, o processo de seleção pode ter outras motivações, tais como [2]:

- Redução geral dos dados: para diminuir os requisitos de armazenamento e aumentar a velocidade de resposta dos algoritmos;
- Melhoria de desempenho: para gerar ganhos de acurácia;
- Compreensão de dados: para adquirir conhecimento sobre o processo que gerou os dados ou, simplesmente, para visualização dos dados.

O processo de seleção de características e uma boa classificação estão fortemente relacionados. Um seletor de características ideal deve produzir uma representação que torna o trabalho do classificador trivial [3]. Entretanto, algumas

questões devem ser levadas em consideração no processo de seleção de características, tais como quantas e quais são as melhores características a serem selecionadas.

Este trabalho propõe um algoritmo com o objetivo de solucionar essas questões: quantas e quais. Nesse sentido, respectivamente, a quantidade e qualidade de características selecionadas para o processo de classificação. Procura-se, assim, produzir um classificador otimizado, a fim de gerar resultados ótimos de desempenho. O algoritmo apresentado neste trabalho é intitulado Algoritmo de Otimização por Seleção de Características (AOSC).

O método escolhido para seleção das melhores características do AOSC é o ReliefF [4]. Ele atua junto ao classificador *Support Vector Machine* (SVM) [5]. O critério de desempenho do classificador em análise é a chamada *Area Under ROC* (*Receiver Operating Characteristic*) [6], traduzida como área abaixo da curva ROC. A topologia do AOSC para a otimização é baseada em algoritmos genéticos [7].

A busca pela otimização na seleção de características entregue a um classificador, especialmente o SVM, é discutido na literatura. Huang e Wang [8] lembram que a escolha de quais características entregues ao SVM é de grande importância para uma boa classificação. Diversos trabalhos lidam com a otimização do processo de seleção de características por meio de algoritmos genéticos [9, 10, 11]. Recentemente, [12] propôs o algoritmo MCSO. Este procura otimizar a seleção de características junto ao SVM, utilizando técnicas de computação evolucionária. Porém, diferente do AOSC, o algoritmo MCSO não utiliza o método ReliefF como extrator de características.

A qualidade de características selecionadas pelo AOSC se dá pela escolha do número de vizinhos mais próximos analisados pelo ReliefF, contrapondo o valor estipulado por [13]. Já a quantidade de características selecionadas pelo AOSC indica a dimensão de trabalho do classificador SVM.

Os conceitos dos métodos e métricas abordados no AOSC são brevemente comentados na Seção II. Apresentamos na Seção III o algoritmo AOSC. Na Seção IV são expostos testes que comprovam a melhoria no desempenho do classificador SVM ao utilizar o AOSC. Os resultados obtidos são ainda comparados com os resultados de outros três algoritmos. Por fim, na Seção V, apresentamos as conclusões.

II. METODOLOGIA

A. Seleção de Características – Método ReliefF

Um método difundido para a seleção de características é o Relief, proposto em [14]. Segundo [15], a ideia principal deste algoritmo é a estimação de quanto uma determinada característica ajuda na distinção de classes levando em consideração amostras vizinhas.

Diversas variações do método original são encontradas na literatura. Muitas destas buscam otimizar o método para seleção de características. Dentre elas, destaca-se: ReliefF [4], RRelieff [16] e IRelief [17, 18].

Conforme citado anteriormente, o método abordado no AOSC é o ReliefF. Segundo [2] este método foi concebido para problemas multiclases e é baseado no algoritmo de K-Vizinhos Mais Próximos (KNN, do inglês *K-Nearest Neighbors*).

Embora [13] afirmarem que o valor de 10 vizinhos mais próximos pode ser seguramente adotado para a maioria dos casos ao se trabalhar com o ReliefF, neste trabalho, tratamos o número de vizinhos considerados, como uma variável de busca no processo de otimização. Ele é responsável pela qualidade das amostras selecionadas para o processo de classificação.

B. Classificador – Support Vector Machine

A técnica de classificação por *Support Vector Machine* (SVM) surgiu com a apresentação do algoritmo do hiperplano de margem máxima [5]. Este algoritmo conseguiu promover a classificação de amostras ao criar uma fronteira linear entre amostras de classes distintas. Boser, Guyon e Vapnik [19] em pesquisas com a intenção de encontrar formas de separação de dados não lineares proporcionaram, com o uso das funções Kernel, a transformação do espaço de entrada em um espaço de dimensão maior, onde os dados são linearmente separáveis. Três anos depois, [20] apresentaram uma proposta com a modificação que permitiu a introdução de variáveis de folga construindo assim hiperplanos de separação com margens flexíveis.

A Fig. 1 apresenta resumidamente a técnica SVM pela busca do hiperplano ótimo para separação de duas classes. Nota-se que em ambos os quadros (a e b) existe uma separação das classes representadas pelas amostras em círculos ou quadrados, onde a linha tracejada representa o hiperplano de separação. Porém, no quadro (b) se tem a otimização desta barreira de separação. Desta forma, novos dados terão maior probabilidade de serem classificados corretamente. O SVM trata justamente de otimizar este hiperplano de separação. Vale ressaltar que o exemplo da Fig. 1 apresenta um problema de apenas duas dimensões, no entanto, vários problemas práticos costumam envolver um número elevado de dimensões, o que dificulta a determinação de um hiperplano classificador.

A dimensão de trabalho é a outra variável de busca no processo de otimização utilizado pelo AOSC. Ressalta-se que cada característica selecionada representa aumento na dimensão de trabalho, logo, ao selecionar a quantidade de características, o AOSC está selecionando a dimensão de trabalho do classificador SVM.

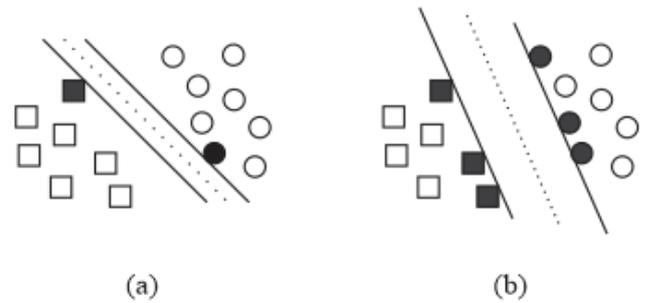


Fig. 1. Support Vector Machine (SVM)

C. Critério de desempenho – Área abaixo da curva ROC

Muitos trabalhos na área de reconhecimento de padrões tem a análise de desempenho de seus algoritmos baseada na acurácia dos seus classificadores. Entretanto, sabe-se que a acurácia pode mascarar taxas de erros de classificação quando se trabalha com classes desbalanceadas [21].

Outra forma de avaliar o desempenho de um classificador é por meio da curva ROC [6]. As curvas ROC são baseadas em duas informações retiradas da tabela de confusão: sensibilidade e especificidade. Os gráficos para as curvas ROC são bidimensionais, nos quais o eixo das ordenadas plota-se a sensibilidade e no eixo das abscissas a especificidade [22]. A curva ROC de um classificador ideal possui o formato da função *Heaviside* (*Heaviside step function*) [21].

Com o objetivo de evitar os problemas típicos associados com o simples uso da acurácia, a área abaixo da curva ROC, conhecida como AUC (*Area Under the ROC Curve*), foi adotada como critério de desempenho do AOSC. A AUC tem valor máximo igual a 1; quanto mais próximo deste valor, melhor é o desempenho do classificador. Evitam-se, assim, más interpretações em bases de dados desbalanceadas.

D. Topologia do algoritmo - Algoritmo Genético

Conforme mencionado o AOSC tem sua topologia como de um algoritmo genético. Os algoritmos genéticos foram propostos por John Holland (1975) e aperfeiçoados por Holland e seus alunos na Universidade de Michigan na década subsequente [7]. Estes algoritmos são técnicas que simulam o processo de evolução natural. Denominadas gerações, suas iterações geram através de processos de seleção, elitismo e operações de cruzamento e mutação, novos indivíduos mais adaptados ao meio [23]. Indivíduos são formados por variáveis do problema, ou cromossomos que influenciam na função objetivo.

Na próxima seção, com a descrição do AOSC, melhor entendimento sobre algoritmos genéticos e sua aplicabilidade é absorvido.

III. ALGORITMO AOSC

O AOSC é representado pelo fluxograma exposto na Fig. 2. O bloco inicial indica a separação da base de dados em três conjuntos de amostras denominados: treinamento, avaliação e validação.

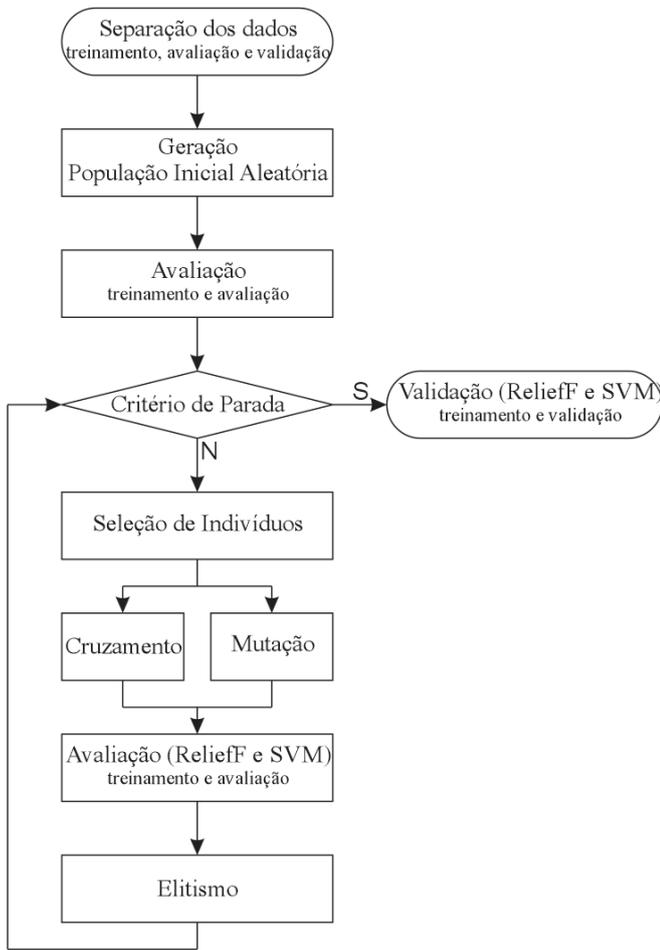


Fig. 2. Algoritmo de Otimização por Seleção de Características

As amostras separadas no conjunto de treinamento são utilizadas durante todo o AOSC. Já as amostras do conjunto de avaliação participam apenas do processo de busca das variáveis durante a otimização. As amostras do conjunto de validação participam apenas do processo final de classificação para validação de desempenho do algoritmo.

A divisão da base de dados é proposta de forma que 49% das amostras pertençam ao conjunto de treinamento, 21% das amostras pertençam ao conjunto de avaliação e 30% das amostras pertençam ao conjunto de validação. A escolha destes valores assegura que, mesmo durante a evolução do algoritmo, 30% das amostras envolvidas sejam destinadas a avaliação do classificador SVM, assim como no processo final de validação.

Como se deseja otimizar a saída do classificador pelo número de características selecionadas e pelo número de vizinhos mais próximos analisados pelo algoritmo ReliefF, estas duas grandezas compõem cada um dos indivíduos do algoritmo genético. Uma população inicial aleatória com 20 indivíduos é criada. Apesar de [24] indicar o valor mínimo de 50 indivíduos para um bom desempenho de um algoritmo genético, o valor reduzido de 20 indivíduos proporcionou bons resultados com menor tempo de processamento.

A população inicial criada aleatoriamente é avaliada no próximo passo do AOSC. O processo de avaliação acontece do

seguinte modo: cada indivíduo serve de argumento para a extração de características no método ReliefF. Em outras palavras, os indivíduos indicam quantos vizinhos mais próximos serão considerados pelo ReliefF e quantas características serão entregues ao classificador SVM.

As características ranqueadas e selecionadas a partir do conjunto de amostras de treinamento são entregues a etapa de treino do SVM. Após a etapa de treino é feita a classificação do conjunto de amostras de avaliação, com o SVM. Conforme exposto, é calculada a AUC, para cada um dos indivíduos, como critério de desempenho.

A partir deste momento, o algoritmo entra no processo de evolução. Inicialmente, verifica-se se algum critério de parada foi atingido. Definiu-se como critério de parada o número máximo de 20 gerações ou a involução do algoritmo caracterizado somente por atingir o valor máximo da AUC. O valor de 20 gerações foi assim determinado por testes que mostraram a estabilização de resultados.

O processo de evolução começa com a seleção de indivíduos pelo método torneio. No método torneio, dentro de um par de indivíduos selecionados aleatoriamente, somente aquele que produziu maior AUC segue para a etapa de cruzamento ou mutação. Os indivíduos selecionados são divididos: 85% destes indivíduos devem ser cruzados enquanto o restante sofre o processo de mutação.

O processo de cruzamento é realizado por uma média alterada entre as variáveis de dois indivíduos selecionados aleatoriamente. O operador de cruzamento é exposto em (1):

$$x_a = \frac{x_a + x_b}{2 + \alpha} \quad (1)$$

Sendo x_a a variável a ser alterada, x_b a variável de um indivíduo selecionado aleatoriamente e α um valor aleatório entre -0,05 e +0,05. Um novo valor de α é considerado a cada operação de cruzamento.

Já no processo de mutação, um valor aleatório modifica as variáveis do indivíduo selecionado. O operador de mutação segue a equação apresentada por [25], modificado e aqui exposto em (2):

$$x_a = 0.15 \times \beta \times x_R \quad (2)$$

Sendo x_a a variável a ser alterada; β um número aleatório a cada operação com distribuição gaussiana, média zero e variância um; e x_R a diferença entre os máximos e mínimos da variável em processo de mutação.

Ressaltamos que as variáveis trabalhadas são: o número de vizinhos mais próximos analisados pelo ReliefF e o número de características selecionadas ao SVM. Foi definido o intervalo de vizinhos mais próximo entre 1 e 40. Lembramos que comumente adota-se o valor de 10 vizinhos [13]. Já o intervalo do número de características selecionadas depende da base de dados em

estudo. O valor mínimo foi definido como 1, independente da base de dados em estudo, e o valor máximo igual ao número de características da base de dados em estudo.

Após o processo de cruzamento e mutação, a nova população formada é avaliada conforme descrito anteriormente. Por fim, aplica-se o elitismo: compara-se a população anterior com a nova população, indivíduo por indivíduo. São selecionados apenas os indivíduos que produziram a maior AUC com os conjuntos de treinamento e avaliação.

Este processo é repetido até que um dos critérios de parada seja atingido. Quando isto ocorre, seleciona-se o melhor indivíduo da população. Lembramos que o melhor indivíduo é aquele composto pela melhor combinação do número de características selecionadas com o número de vizinhos mais próximos analisados pelo algoritmo ReliefF que produziu maior AUC.

Este indivíduo irá ranquear e selecionar as melhores características no processo de treinamento do classificador, utilizando o próprio conjunto de amostras de treinamento. Para a validação, após treinado o classificador com o melhor dos indivíduos, faz-se a classificação no conjunto de amostras de validação pelo SVM.

IV. TESTES E ANÁLISE DE RESULTADOS

Para validação do AOSC uma série de testes foi realizada. O algoritmo foi parametrizado para o processo de classificação da base de dados Breast Cancer Wisconsin [26], denominada neste trabalho como Câncer. Esta base de dados consiste em amostras de tumores de câncer de mama das classes maligno e benigno. Cada amostra possui 30 características. Algumas destas características são: área, perímetro, simetria e concavidade de células de tumores amostrados.

A parametrização do algoritmo genético envolve o modo de divisão dos conjuntos de amostras, o número de gerações para evolução, o número de indivíduos criados e a divisão entre os indivíduos para as etapas de cruzamento ou mutação.

Para verificar a robustez do AOSC, outra base de dados, não usada para parametrização do algoritmo genético, também serviu para teste. Esta outra base, Pima Indians Diabetes [26], nomeada neste trabalho como Diabetes, classifica pacientes que apresentam sinais de diabetes segundo critérios da Organização Mundial de Saúde. A base dispõe amostras com 8 características, cada.

O primeiro teste comparou o resultado do AOSC com dois outros algoritmos. O primeiro deles não possui o processo de extração e seleção de características, ou seja, não utiliza o método ReliefF. Os conjuntos de amostras de treinamento e avaliação formaram uma única base de dados que serviu para o treinamento do classificador SVM. Este, por sua vez, foi testado com o conjunto de amostras de validação. Nomeou-se este algoritmo por K0.

No segundo algoritmo para comparação, adotou-se o valor fixo de K no método ReliefF. Utilizou-se K igual a 10, conforme indicado por [13]. Entretanto, o número de características sele-

cionadas foi variado no intervalo de 1 até a quantidade máxima de características das amostras do banco de dados em teste. Selecionou-se o número de características que apresentou melhor AUC após treinamento e classificação pelo SVM. Nomeou-se este algoritmo por K10.

Por se tratar de um método heurístico, o AOSC foi avaliado em uma batelada de 33 execuções de forma a analisar os resultados de valor médio, desvio padrão, valor mínimo e valor máximo para cada uma das bases de dados testadas.

Os resultados deste primeiro teste são expostos na Tabela I. Os valores apresentados referem-se a AUC obtida após a classificação das amostras de validação. Como os algoritmos K0 e K10 não apresentam variação de resultados somente o valor médio da AUC, obtido por estes algoritmos, é analisado na Tabela I.

TABELA I. EVOLUÇÃO DE RESULTADOS DA AREA UNDER THE ROC CURVE (AUC)

	K0	K10	AOSC			
	Valor Médio	Valor Médio	Valor Médio	Desvio Padrão	Valor Máximo	Valor Mínimo
Cânc.	0,9695	0,9825	0,9929	0,0037	0,9956	0,9825
Diab.	0,7810	0,7906	0,7956	0,0111	0,8121	0,7804

A Tabela I mostra que o AOSC melhorou os resultados perante os algoritmos sem otimização. Para a base de dados Câncer, utilizada para parametrização do AOSC, verificou-se que o valor médio da AUC obtido é superior aos outros dois algoritmos. O AOSC apresenta baixo desvio padrão e valor máximo acima de 0,9950. Salienta-se ainda que o valor mínimo apresentado pelo AOSC é exatamente o resultado obtido sem a otimização, utilizando o valor de K fixo e pré-determinado de acordo com [13]. Neste sentido, em nenhuma execução o AOSC apresentou piora de desempenho para esta base de dados se comparado aos demais algoritmos sem otimização. Comprova-se assim a eficiência do AOSC no processo de otimização para a base de parametrização.

Para a base de dados Diabetes, testada para verificação da robustez do algoritmo, o AOSC também apresentou valor médio da AUC superior aos valores obtidos nos algoritmos sem otimização. O desvio padrão encontra-se mais elevado.

Outra comparação posta ao AOSC foi perante a métodos de otimização já encontrados na literatura. Comparou-se o valor de acurácia obtido com os algoritmos MOBJ [27], DEANN [28] e NSS [29]. Estes algoritmos utilizam o treinamento multiobjetivo de redes neurais artificiais baseado em coordenadas esféricas [29]. Os mesmos possuem maior complexidade de classificação se comparados ao classificador SVM empregado neste trabalho.

Os métodos de obtenção dos resultados no AOSC foi exatamente como nos outros métodos, descritos em [29]. Por este motivo, apesar de utilizar a AUC como critério de desempenho do classificador, os resultados obtidos foram comparados por meio da acurácia de classificação. Os resultados são expostos na Tabela II.

TABELA II. ACURÁCIA OBTIDA PELO AOSC E OUTROS MÉTODOS

		AOSC	MOBJ	DEANN	NSS
Câncer	<i>Md</i>	0,9895	0,9882	0,9512	0,9819
	<i>Dp</i>	0,0037	0,0000	0,0028	0,0033
	<i>Mx</i>	0,9942	0,9882	0,9529	0,9883
	<i>Mn</i>	0,9825	0,9882	0,9471	0,9766
Diabetes	<i>Md</i>	0,7951	0,7491	0,7783	0,7765
	<i>Dp</i>	0,0524	0,0036	0,0065	0,0062
	<i>Mx</i>	0,8519	0,7522	0,7913	0,7826
	<i>Mn</i>	0,7037	0,7435	0,7696	0,7609
<i>Md</i> – Valor médio <i>Dp</i> – Desvio padrão <i>Mx</i> – Valor máximo <i>Mn</i> – Valor mínimo					

Ao analisar a Tabela II, percebemos a eficiência do AOSC perante os demais métodos ao verificar que para a base de parametrização, Câncer, obteve-se valor médio e valor máximo de acurácia superior aos demais métodos, apesar de leve aumento do desvio padrão. Apenas em comparação a um dos métodos obteve-se piora no valor mínimo encontrado.

Para a base de dados Diabetes, tem-se a mesma conclusão sobre a melhora no valor médio e valor máximo obtidos, porém o elevado desvio padrão mostra uma indesejável inconstância no resultado final. Piorou-se o valor mínimo.

Como o AOSC tem sua topologia baseada em algoritmos genéticos e estes podem ser parametrizados de forma a diminuir a variância dos resultados, acredita-se que é possível conseguir melhor constância no resultado obtido sem prejudicar seu desempenho médio. Uma alteração simples é o aumento de gerações para o critério de parada. Ressalta-se que esta modificação acarretará no aumento do tempo de processamento.

V. CONCLUSÃO

Após a parametrização e testes realizados foi possível concluir que o algoritmo proposto neste trabalho, AOSC, mostrou-se eficaz no que diz respeito ao critério de desempenho utilizado para otimizar a performance de um classificador por modificação no método de seleção de características. Sendo esta modificação dada pelo número de K-Vizinhos Mais Próximos considerados pelo ReliefF em conjunto ao número de características selecionadas. Testes mostraram melhoria na classificação do SVM após a introdução do processo de otimização proposto.

Percebeu-se, ainda, que a generalização proposta por [13] sobre o número de vizinhos mais próximos analisados no método ReliefF não se mostrou como uma boa alternativa. Este valor pode ser alterado, para valores superiores ou inferiores ao proposto, em combinações com o número de características selecionadas. Com esta combinação, produziu-se melhores resultados de desempenho.

Ao comparar com outros métodos, o AOSC mostrou bons resultados para a base de dados utilizada em sua parametrização. Para uma base de dados diferente, o mesmo forneceu aumento no desvio padrão de resposta, mas ainda assim produziu valor

médio e valor máximo de acurácia superior aos demais métodos comparados.

Este trabalho trata apenas da proposta inicial de AOSC. Otimização dos seus parâmetros de configuração podem ser feitos para melhor generalização. Testes comparativos com outros algoritmos baseados em computação evolucionária, como o MSCO [12], podem ajudar em sua melhor validação.

AGRADECIMENTO

O presente trabalho foi realizado com o apoio financeiro da CAPES – Brasil.

BIBLIOGRAFIA

- [1] S. Theodoridis e K. Koutroubas, *Pattern Recognition*, 4th ed., San Diego: Academic Press, 2009.
- [2] I. Guyon e A. Elisseeff, “An Introduction to Feature Extraction,” em *Feature Extraction: Foundations and Applications*, New York, Springer, 2006.
- [3] R. O. Duda, P. E. Hart e D. G. Stork, *Pattern Classification*, New York: Wiley, 2000.
- [4] I. Kononenko, E. Simec e M. Robnik-Sikonja, “Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF,” *Applied Intelligence*, vol. 7, nº 1, pp. 39-55, 1997.
- [5] V. Vapnik e A. Lerner, “Pattern recognition using generalized portait method,” *Automation and Remote Control*, nº 24, pp. 774-780, 1963.
- [6] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, 2006.
- [7] M. Mitchell, *An introduction to genetic algorithms*, Cambridge: MIT Press, 2002.
- [8] C.-L. Huang e C.-J. Wang, “A GA-based feature selection and parameters optimization for support vector machines,” *Expert Systems with Applications*, nº 31, pp. 231-240, 2006.
- [9] J. Yang e V. Honavar, “Feature subset selection using a genetic algorithm,” *IEEE Intelligent Systems*, vol. 13, nº 2, pp. 44-49, 1998.
- [10] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn e A. K. Jain, “Dimensionality Reduction Using Genetic Algorithms,” *IEEE Transactions on Evolutionary Computation*, vol. 4, nº 2, pp. 164-171, 2000.
- [11] S. Salcedo-Sanz, M. Prado-Cumplido e Pérez-Cruz, “Feature Selection via Genetic Optimization,” *Artificial Neural Networks — ICANN 2002 (Lecture Notes in Computer Science)*, vol. 2415, pp. 547-522, 2002.
- [12] K.-C. Lin, Y.-H. Huang, J. C. Hung e Y.-T. Lin, “Feature Selection and Parameter Optimization of Support Vector Machines Based on Modified Cat Swarm Optimization,” *International Journal of Distributed Sensor Networks*, vol. 2015, 2015.

- [13] M. Robnik-Sikonja e I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, pp. 23-69, 2003.
- [14] K. Kira e L. A. Rendell, "A practical approach to feature selection," em *International Conference on Machine Learning*, Aberdeen, 1992.
- [15] I. Kononenko, "Estimating Attributes: Analysis and Extensions of RELIEF," *Proceedings of European Conference on Machine Learning*, pp. 171-182, 1994.
- [16] M. Robnik-Sikonja e I. Kononeko, "An adaptation of Relief for attribute estimation in regression," em *14th International Conference on Machine Learning*, 1997.
- [17] Y. Sun e L. Jian, "Iterative RELIEF for Feature Weighting," em *23^a International Conference on Machine Learning*, Pittsburgh, 2006.
- [18] Y. Sun, "Iterative RELIEF for Feature Weighting: Algorithms, Theories, and Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, n^o 6, Junho 2007.
- [19] B. E. Boser, I. M. Guyon e V. N. Vapnik, "A training algorithm for optimal margin classifiers," em *Annual Conference on Computational Learning Theory*, Pittsburgh, 1992.
- [20] C. Cortes e V. Vapnik, "Support vector networks," *Machine Learning*, n^o 20, pp. 273-297, 1995.
- [21] C. L. d. Castro e A. P. Braga, "Aprendizado supervisionado com conjuntos de dados desbalanceados," *Revista Controle & Automação*, vol. 22, n^o 5, 2011.
- [22] A. C. d. S. Braga, "Curvas ROC: aspectos funcionais e aplicações," Braga, 2000.
- [23] M. Srinivas e L. M. Patnaik, "Genetic algorithms: a survey," *Computer*, vol. 27, n^o 6, pp. 17-26, 1994.
- [24] R. Linden, *Algoritmos Genéticos*, 2^a ed., Rio de Janeiro: Brasport, 2008.
- [25] R. A. Meneguim, "Análise da Sensibilidade de Soluções em Otimização Através de Elipsóides Mínimos," Belo Horizonte, 1999.
- [26] M. Lichman, "UCI repository of machine learning databases," University of California, 2013. [Online]. Available: <http://archive.ics.uci.edu/ml/>. [Acesso em junho 2015].
- [27] R. d. A. Teixeira, A. P. Braga, R. H. C. Takahashi e R. R. Saldanha, "Improving generalization of mlps with multi-objective optimization," *Neurocomputing*, vol. 35, pp. 189-194, 2000.
- [28] H. P. Rocha, C. L. Castro e A. P. Braga, "Seleção de modelos neurais utilizando evolução diferencial através do controle de erro e norma do vetor de pesos," em *X Congresso Brasileiro de Inteligência*, 2011.
- [29] H. P. Rocha, M. A. Costa e A. P. Braga, "Treinamento Multi-Objetivo de Redes Neurais Artificiais Baseado em Coordenadas Esféricas," em *1st BRICS Countries*