

Redes Neurais aplicadas na predição de valores genéticos

Gabi Nunes Silva¹

Isabela de Castro Sant'Anna²

Rafael Simões Tomaz³

Cosme Damião Cruz²

1 Introdução

Nos programas de melhoramento genético, os pesquisadores almejam selecionar genótipos superiores e recomendar variedades por meio de metodologias eficazes de seleção (CRUZ, 2010). Para tal fim, é usual que os melhoristas lancem mão das técnicas de genética estatística existentes na literatura, e que envolvem métodos de experimentação e análises estatísticas (RESENDE, 2007).

A problemática vista nestas metodologias é que, por diversas vezes, os efeitos ambientais se expressam fortemente nas características de interesse econômico. Esses ruídos ambientais influenciam o fenótipo dos indivíduos, e conseqüentemente, fazem com que os experimentos sejam avaliados com baixa confiabilidade de predição do valor fenotípico. Isso força o melhorista a indicar superioridade genética e predizer ganhos com base somente na média fenotípica das repetições dos experimentos.

Considerando o grande potencial de aplicação nas diversas áreas do melhoramento apresentado pelas Redes Neurais Artificiais (GIANOLA *et al.*, 2011; NASCIMENTO *et al.*, 2013; VENTURA *et al.*, 2012), e visando aumentar a acurácia na predição, o presente trabalho tem como objetivo avaliar a eficiência das Redes Neurais Artificiais como abordagem alternativa para predição de genótipos superiores a partir de informações adicionais, além da média fenotípica dos genótipos e de dados simulados.

2 Material e Métodos

¹Departamento de Estatística – Universidade Federal de Viçosa. e-mail: gabi.silva@ufv.br

²Departamento de Biologia Geral – Universidade Federal de Viçosa

³Departamento de Zootecnia – Universidade Federal de Viçosa

O conjunto de dados utilizado foi obtido por meio do aplicativo computacional GENES (CRUZ, 2013) a partir do procedimento de simulação de ensaios sob o delineamento de blocos casualizados com 100 genótipos e seis blocos, para herdabilidades de 40 %, 50%, 55%, 65% e 70% média 100 e coeficiente de variação experimental igual a 15%.

Para fins de treinamento das Redes Neurais Artificiais, foi utilizado um arquivo ampliado do conjunto de dados inicial, com 5000 genótipos e os mesmos valores de média herdabilidade e coeficiente de variação do arquivo contemporâneo.

A arquitetura da rede utilizada foi a *Multilayer Perceptron* (MLP), com três camadas ocultas, variação de um a sete neurônios por camada e número máximo de 2000 iterações, totalizando 9261 arquiteturas investigadas. O algoritmo de treinamento adotado foi o *trainlm*, e as funções de ativação investigadas foram a tangente hiperbólica (*tansig*) e a função sigmoideal (*logsig*), representadas pelas equações (1) e (2), respectivamente.

$$\varphi(x) = \tanh\left(\frac{x}{2}\right) = \frac{1 - e^{-x}}{1 + e^x} \quad (1)$$

e

$$\varphi(x) = \frac{1}{1 + e^{-ax}} \quad (2)$$

Como estratégia de treinamento, adotou-se o valor genético verdadeiro como saída desejada, e como entrada, foi considerada a média fenotípica, acrescida de outras sete informações estatísticas de grande relevância ao pesquisador tal como descrito a seguir: x_1 : a média fenotípica; x_2 : o máximo de cada genótipo nos b blocos; x_3 : o mínimo de cada genótipo nos b blocos; x_4 : o desvio padrão de cada genótipo; x_5 : o coeficiente de variação para cada genótipo; x_6 : o valor da soma de rank nos b blocos para cada genótipo; x_7 : o valor da estatística P_1 , adaptada de Lin & Bins (1988) para estudos de adaptabilidade e estabilidade, dada por $\sum_{j=1}^b \frac{(X_{ij} - M_j)^2}{2b}$, sendo X_{ij} , o valor fenotípico do i -ésimo genótipo no j -ésimo bloco; M_j , a resposta máxima observada entre todos os genótipos no bloco j ; e b , o número de blocos do experimento e x_8 : os valores discretos, codificados de 1 a 6, estabelecidos a partir da transformação da média fenotípica considerando a sua variação em torno da média nos limites, inferior e superior, tomados como um, dois ou três desvio-padrão.

Para avaliar a eficiência da Rede Neural, procedeu-se à etapa denominada validação, na qual foram gerados 120 arquivos, também obtidos por simulação, e com os mesmos valores de média, variância e covariâncias do arquivo original.

Finalmente, o critério de parada se baseou no número de vezes em que a predição dada pela rede superou a média fenotípica, e a acurácia foi representada pela correlação entre a resposta da rede e o valor genético verdadeiro.

3 Resultados e Discussões

A fim de avaliar a eficiência preditiva da rede, considerou-se como acerto o número de vezes que o quadrado da correlação entre a resposta obtida pela RNA foi superior à herdabilidade da característica, que é o critério usualmente utilizado para a seleção em programas de melhoramento e que pode ser obtido por meio de Modelos Lineares.

Em todos os cinco cenários estudados, as Redes Neurais Artificiais demonstraram superioridade quando comparadas com o Método dos Mínimos Quadrados utilizado em Modelos Lineares (Tabela 1).

Tabela 1. Médias das estimativas das acurácias obtidas por meio das redes (RNA) versus modelos lineares (ML), assim como a média da superioridade da rede (Sup.) para cada cenário.

Acurácias (%)	$h^2 = 40$	$h^2 = 50$	$h^2 = 55$	$h^2 = 65$	$h^2 = 70$
ML	0.6325	0.7071	0.7416	0.8062	0.8366
RNA	0.6411	0.7160	0.7432	0.8128	0.8403
Sup.(%)	1.3697	1.2587	0.2157	0.8187	0.4423

Os resultados obtidos demonstram o grande potencial das Redes Neurais Artificiais como metodologia alternativa de predição nos programas de melhoramento. Demonstraram ainda que a utilização de dados simulados representa grande aplicabilidade quando se deseja gerar ou replicar informações representativas de um genótipo.

A tabela 2 apresenta as coincidências de classificação dos genótipos selecionados e rejeitados quando utilizadas as RNA's e a metodologia de ML.

Em todos os cinco cenários, as Redes Neurais Artificiais proporcionaram coincidências médias superiores às proporcionadas pela metodologia de Modelos Lineares, demonstrando que as predições da RNA proporcionaram uma melhor discriminação dos genótipos tomando por base o valor genético verdadeiro.

Tabela 2. Média percentual de coincidência de genótipos selecionados (Selec.) e rejeitados (Rej.) por meio de previsão através da RNA e dos modelos lineares (ML),

		Coincidência (%)				
		h ² = 40	h ² = 50	h ² = 55	h ² = 65	h ² = 70
Selec. (5%)	ML	30.17	40.83	42.67	52	50.67
	RNA	34.17	44	42.67	52.33	52.33
	ML-RNA	69.67	85.5	89.67	93	91.83
Selec. (10%)	ML	38.58	44.58	46.08	53.42	57.92
	RNA	39.83	47.33	45.92	53.67	58.17
	ML-RNA	80.92	89.33	88.33	95.58	94.92
Rej. (20%)	ML	51.29	56.29	59	63.63	67.58
	RNA	52.33	56.46	59.29	64.38	67.67
	ML-RNA	89.63	97.04	93.58	96.04	98.25
Rej. (10%)	ML	42.5	46.42	48.92	52.42	54.58
	RNA	46	47.75	49.5	51.92	53.75
	ML-RNA	81.87	83.75	94.17	92.5	93.5

As porcentagens de coincidência da rede foram superiores a 69% quando comparadas aos genótipos selecionados pelo método convencional, o que demonstrou que os mesmos genótipos foram selecionados por ambas as abordagens.

4 Conclusões

As Redes Neurais foram eficientes ao predizer valores genéticos em ensaios simulados segundo o delineamento de blocos casualizados, e ainda, demonstraram superioridade quando comparadas aos métodos tradicionais de seleção pois levam em conta estatísticas subutilizadas nos processos de predição.

5 Agradecimentos

Fapemig, Capes e CNPq pelo apoio financeiro.

6 Referências

- [1] CRUZ, C.D. GENES - A software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum**. v.35, n.3, p.271-276, 2013.
- [2] CRUZ, C. D. **Princípios de Genética Quantitativa**. 2. ed. Viçosa: UFV. 2010. 394p.
- [3] GIANOLA, D.; OKUT, H.; WEIGEL, K. A.; ROSA, G. J. M. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. **BMC Genetics**. p.12-87, 2011. Disponível em <http://www.biomedcentral.com/1471-2156/12/87>. Acesso em: 13 de dezembro de 2013.
- [4] NASCIMENTO, M.; PETERNELLI, L. A.; CRUZ, C. D.; NASCIMENTO, ANA CAROLINA CAMPANA; FERREIRA, R. P.; BHERING, L. L.; SALGADO, C. C. **Artificial neural networks for adaptability and stability evaluation in alfafa genotypes**. *Crop Breeding and Applied Biotechnology*, v.12, p.152-156, 2013.
- [5] RESENDE, M. D. V. **Matemática e Estatística na Análise de Experimentos e no Melhoramento Genético**. Colombo: Embrapa Florestas. 2007. 362p.
- [6] VENTURA, R.; SILVA, M.; MEDEIROS, T.; DIONELLO, N.; MADALENA, F.; FRIDRICH, A.; VALENTE, B.; SANTOS, G.; FREITAS, L.; WENCESLAU, R. Use of artificial neural networks in breeding values prediction for weight at 205 days in Tabapuã beef cattle. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**,v.64, n.2, p.411-418, 2012.