



Redes neurais artificiais na predição do valor genético em experimentos em blocos ao acaso

Prelecionista: Leonardo de Azevedo Peixoto

Orientador: Leonardo Lopes Bhering

Resumo

O sucesso de um programa de melhoramento genético depende da acurácia da predição do valor genético a partir de valores fenotípicos. As estimativas de herdabilidade podem orientar um melhorista durante as três fases principais de um programa de melhoramento: i) criação da variabilidade genética; ii) seleção de indivíduos superiores na população segregante; iii) utilização dos indivíduos selecionados no programa. Assim, tem sido usual e indispensável a quantificação da herdabilidade que é a medida do quadrado da correlação entre o valor genético e a média fenotípica. A bioinformática refere-se a aplicação de técnicas computacionais para análises biológicas. A rede neural artificial (RNAs) é uma técnica de bioinformática que utiliza aproximação bayesiana. RNAs são modelos computacionais que tentam imitar a função do cérebro humano e o sistema biológico neural de uma forma simples. As redes neurais tentam encontrar a melhor função não linear para explicar determinado fenômeno. Elas podem aprender funções complexas relacionadas entre os dados de entrada e saída, que não podem ser estimados pelo pesquisador. As redes neurais tem sido utilizadas por inúmeros autores para classificação de imagens em sensoriamento remoto (Aitkenhead and Aalders, 2008), análise de diversidade genética (Barbosa et al., 2011), identificação de genótipos superiores (Mugnai et al., 2008) e predição do valor genético em animais (Ventura et al., 2012). No entanto não existe relato na literatura sobre utilização das RNAs para predição do valor genético em experimentos balanceados em plantas. Portanto, o objetivo deste trabalho foi avaliar a eficiência da utilização das redes neurais artificiais na predição do valor genético em experimentos em blocos casualizados, comparando a eficiência do valor de rede com o a eficiência do valor fenotípico, dado pela correlação entre o valor de rede e o valor genotípico e a correlação entre o valor fenotípico e o valor de rede, e ainda definir uma estrutura de rede ideal para a predição do valor genético.

Foram simulados experimentos em blocos casualizados com seis repetições, 150 e 200 genótipos. Para cada experimentos foram simuladas 5 variáveis com diferentes médias (20, 40, 60, 80 e 100). As populações de treinamento foram compostas por 6 blocos com 5000 genótipos mantendo as características da população simulada. As populações de validação foram compostas por 6 blocos com 150 ou 200 genótipos. Para cada experimento variou-se a herdabilidade (10, 20, 30 e 40) e o coeficiente de variação (5 e 10%). Utilizou-se 100 populações de validação em cada tipologia de rede. A eficácia das RNAs foi avaliada comparando a correlação do valor de rede com o valor genético e o valor fenotípico com o valor genético. As funções de ativação da rede utilizadas foram purelin, Logsig e Tansig.

Observou-se que para a predição do valor genético as redes neurais foram superiores a utilização do valor fenotípico. Em 70% das redes simuladas o ganho foi acima de 80%, do valor de rede em relação ao valor fenotípico. Portanto, a utilização das redes neurais para predição do valor genético é melhor que a utilização do valor fenotípico, pois a rede consegue diminuir o efeito ambiental (ruído) sobre um determinado experimento. Com isso, o valor de rede se aproxima mais do valor genético que o valor fenotípico.

Observou-se que não houve diferenças entre as populações simuladas com diferentes herdabilidade (10, 20, 30 e 40) e CV (5 e 10%). Portanto mesmo que no experimento o efeito ambiental seja alto, o valor de rede consegue se aproximar do valor genético. As RNAs foram eficientes para a predição do valor genético para herdabilidade baixas (10 e 20%) e moderadas (30 e 40%). Portanto, a utilização das redes em características qualitativas (alta herdabilidade) e quantitativas (baixa herdabilidade) pode ser realizada.

A correlação entre o valor genético e o valor de rede foi maior que a correlação entre o valor fenotípico e o valor de rede em todas as características. A diferença entre estas correlações variou de 0,64 a 10,3%.

Foi observado diferenças entre o número de neurônios nas camadas e as funções de ativação. O número de neurônios na camada 1 variou de 2 a 10, na camada 2 de 0 a 8 e na camada 3 de 2 a 8. Na camada 2 foi utilizado 20 neurônios no processo de treinamento e validação.