

Sociedade Brasileira de Química (SBQ)

Aplicação de redes neurais artificiais na classificação e identificação de cultivares de soja por região de plantio.

Olívio F. Galão (PQ)^{1*}, Dionísio Borsato (PQ)¹, Jurandir P. Pinto (PQ)¹, Marcos V. R. Pina (IC)¹, Jesui V. Visentainer (PQ)², Mercedes C. Panizzi (PQ)³, Rodrigo S. Leite (PQ)³. galao@uel.br.

¹Depto. de Química da Universidade Estadual de Londrina. C.P. 6001. ²Depto. de Química da Universidade Estadual de Maringá. ³Embrapa Soja de Londrina-PR.

Palavras Chave: perceptron de múltiplas camadas, soja transgênica e reamostragem.

Introdução

A soja é uma fonte importante de proteínas, óleo além de vitaminas do complexo B. Porém, o que mais interessa ao mercado é o teor de óleo e de proteína, mas os aspectos antinutricionais e outros de interesse à saúde humana, como a presença de frutooligossacarídeos e fitosteróis devem ser considerados. A solução de problemas de classificação através de sistemas automatizados, na maioria dos casos, é de extrema complexidade principalmente se os padrões forem descritos por um grande número de variáveis independentes. As redes neurais artificiais (RNA) são um conjunto de técnicas baseadas em princípios estatísticos, que vem crescentemente ganhando espaço no reconhecimento e classificação de padrões. As RNA são extremamente versáteis para realizar o mapeamento de relações complexas e não lineares entre múltiplas variáveis de entrada e saída.¹

O objetivo do trabalho foi avaliar o desempenho de uma rede neural artificial, do tipo MLP, na identificação e classificação de amostras de sementes de soja, convencionais e transgênicas, cultivadas na região de Ponta Grossa e de Londrina.

Resultados e Discussão

Neste trabalho, foram estudadas 20 variedades de soja desenvolvidas pela Embrapa-soja de Londrina e recomendadas para plantio na região Centro-Sul do Brasil para safra 2008/2009, sendo 6 cultivares transgênicas e 14 convencionais. Essas 20 variedades foram plantadas simultaneamente em Londrina e Ponta Grossa, com clima, terreno e temperatura média diferentes recebendo igual tratamento de adubação e irrigação. Para os dois casos estudados foi utilizada uma rede do tipo perceptron de múltiplas camadas contendo uma camada oculta de neurônios.

As amostras foram divididas em três partes sendo que a primeira consistiu o conjunto de treinamento da rede, e foi formada com 70% das amostras. A segunda, chamada de teste, foi constituída por 15% e a terceira, também constituída por 15% das amostras, foi chamada de validação sendo escolhidas de forma aleatória. Foi utilizada a técnica de reamostragem gerando maior número de dados

aleatórios. Na classificação por região, a rede utilizada apresentou um acerto de 100% no treinamento, na classificação e no teste. A partir da rede treinada foi possível estipular uma ordem de importância para as variáveis de entrada sendo que o teor de C14:00, de cálcio, zinco, C24:00, campesterol e manganês foram as variáveis identificadas como as seis mais importantes, isto é, as que mais influenciaram na classificação por região de plantio. O teor de C18:2n6, C18:1n7, estigmasterol, ácido fítico, óleo e o teor de β -fitosterol, as menos importantes.

Na segunda análise a primeira e a segunda saída foi para a soja convencional e transgênica plantada em Londrina, a terceira e quarta para a soja convencional e transgênica plantada em Ponta Grossa. Para este caso foi utilizado o módulo automático do programa Statística v.9. Os limites da MLP para o número de neurônios foi de no mínimo 7 e máximo de 21. Foram treinadas 20 redes e dentre as melhores a que apresentou maior acerto no treinamento, no teste e na validação foi aquela com 14 neurônios na camada oculta isto é, a MLP28:14:4.

A partir da rede treinada foi possível estipular uma ordem de importância para as variáveis de entrada sendo que agora, o teor de zinco, manganês, C14:0, proteína, cálcio e de C24:0 foram identificadas como as mais importantes na classificação das amostras

Conclusões

A utilização de redes neurais artificiais do tipo perceptron de múltiplas camadas mostrou ser uma ferramenta útil para identificar variedades de soja cultivadas em duas regiões do Paraná já que foi possível reconhecer 100% das amostras analisadas quando se comparou as duas regiões. Para o banco de dados contendo informações sobre sojas transgênicas e convencionais foi obtido um desempenho de 94,43% no treinamento da rede, 83,33% no teste e 100% na validação.

Agradecimentos

A Embrapa Soja, UEM, UEL e CNPq.

¹ Bishop, C.M. Neural Networks for Pattern Recognition, 1ed. Oxford, 1995