



The use of multilayer perceptron artificial neural networks for the classification of ethanol samples by commercialization region

Erica Signori Romagnoli, Livia Ramazzoti Chanan Silva, Karina Gomes Angilelli, Bruna Aparecida Denobi Ferreira, Aline Regina Walkoff and Dionisio Borsato*

Departamento de Química, Universidade Estadual de Londrina, Rodovia Celso Garcia Cid, Pr-445 Km 380, Cx. Postal 10011, 86057-970, Londrina, Paraná, Brazil. *Author for correspondence. E-mail: dborsato@uel.br

ABSTRACT. Samples of automotive ethanol, marketed in the northern and eastern regions of the state of Paraná, Brazil, underwent physical and chemical tests. Rates were assessed by Multilayer Perceptron (MLP) neural network for classification. For network training, two hundred epochs, a 0.05 learning rate and a random subdivision of samples in three groups with 70 for training, 15 for test and 15% for validation were employed. Sixty networks were trained from three different initializations. Three networks, one at each start-up, were highlighted and the one with the best performance presented 8 neurons in the hidden layer, with 95 accuracy training, 96 in the test and 96% in validation. The most important variables in classifications, identified by the network, occurred in the following order: alcohol content, density, pH and electrical conductivity. Application of MLP segmented ethanol samples and identified the commercialization regions.

Keywords: biofuel, *backpropagation*, hidden layer, training.

Utilização de Redes Neurais Artificiais Perceptron de Múltiplas Camadas para classificação de amostras de etanol com a região de comercialização

RESUMO. Amostras de álcool etílico automotivo, comercializadas nas regiões Norte e Leste do Estado do Paraná, foram submetidas a ensaios físico-químicos e os valores tabulados foram apresentados à rede neural do tipo Perceptron de múltiplas camadas (MLP) para classificação. Para o treinamento das redes foram utilizadas 200 épocas, uma taxa de aprendizagem de 0.05 e uma subdivisão das amostras, de forma aleatória, em três grupos sendo 70 para treinamento, 15 para teste e 15% para validação. Foram treinadas 60 redes obtidas a partir de três inicializações diferentes. Três redes (uma de cada inicialização) foram destacadas, e aquela com melhor desempenho apresentou oito neurônios na camada oculta com 95 de acerto no treinamento, 96 no teste e 96% na validação. As variáveis mais importantes na classificação, identificadas pela rede, seguiram a seguinte ordem: teor alcoólico, pH, massa específica e condutividade elétrica. A aplicação da rede MLP permitiu a segmentação das amostras de álcool e, com isso, identificar as regiões de comercialização.

Palavras-chave: biocombustível, retropropagação, camada oculta, treinamento.

Introduction

Advances in research on the production of new fuels from renewable sources to reduce toxic gas emission levels into the atmosphere has underscored ethanol as the main biofuel in Brazil. Ethanol is environmentally important since it reduces carbon dioxide emissions. In fact, parts of the gas result from ethanol combustion absorbed by large plantations (Silva, Damasceno, Silva, Madruga, & Santana, 2008; Spacino et al., 2013).

Although hydrated ethanol is mainly produced from sugar cane as raw material, many other sources, such as corn and beet, may be used. Depending on the region or even the municipality,

ethanol production may have physical and chemical parameters differentiated by conductivity, pH and others (Spacino et al., 2013). Various aspects on these characteristics may be studied since they emerge from different interests, comprising distillation results, municipality of origin, separation of different sugar cane crops, comparison of yields and production year for productivity (Silva et al., 2008; Spacino et al., 2013).

According to Bona, Silva, Borsato, and Bassoli (2012), the Artificial Neural Network (ANNs) is one of the study tools which has gained great importance and has been successful in targeting samples. There are several types of neural networks

such as Multilayer Perceptron (MLP), radial basis networks, Self-Organizing Maps (SOM) among others (Haykin, 2001; Bishop, 2007).

ANNs reproduce the logical operations that the brain performs in various tasks. In recent decades, this tool has highlighted pattern recognition and classification (Borsato et al., 2009; Nobrega, Bona, & Yamashita, 2013; Link, Lemes, Marquetti, Scholz, & Bona, 2014). In addition, it may be applied in several areas, such as health (Read et al., 2010; Karelina et al., 2015), food (Debska & Guzowska-Swider, 2011) and engineering (Haykin, 2001; Vuković & Miljković, 2015; Kosic, 2015).

Supervised training is carried out through *backpropagation* algorithm based on learning by error correction (Haykin, 2001; Bishop, 2007).

Basically, the learning of MLP by *backpropagation* occurs when the set of pairs named input and output performs two steps among the different layers of the network: a step forward called propagation and a step backward called *backpropagation*. Specifically, the actual response of the network is subtracted from the expected response to produce an error signal. The error signal is back-propagated through the network, adjusting synaptic weights to make the actual response of the network move closer to the expected response by minimizing the error (Bishop, 2007). When the response does not match with what was expected by the network, the procedure is repeated several times until the input-output set achieves accuracy (Haykin, 2001; Borsato et al., 2009).

Considering the commercial importance of ethanol and the effectiveness of ANN technique, the application of MLP-type ANN for the classification of ethanol samples commercialized in two regions of the state of Paraná, Brazil, has been proposed.

Material and methods

Samples of hydrated ethanol

The two hundred and four samples of commercial hydrated ethanol (112 from the eastern region and 92 from the northern region of the state of Paraná) used for the application of MLP underwent alcohol content, density, pH and electrical conductivity tests.

Alcohol content and Density

The determination of density and alcohol contents follows ASTM D4052-11 (American Society for Testing and Materials [ASTM], 2011), NBR 5992 standards (Associação Brasileira de Normas Técnicas [ABNT], 2008).

pH

The pH of ethanol fuel was determined with a MD-20 Digimed equipment, according to NBR 10891 (Associação Brasileira de Normas Técnicas [ABNT], 2013).

Electrical conductivity

Electrical conductivity was determined by Digimed DM-31 equipment, according to ASTM D1125-11 (American Society for Testing and Materials [ASTM], 2014).

ANN

The classification module of Statistica 9.0 was employed for MLP-type neural networks and for automatic segmentation. Networks were trained with 70% of the samples for the training group and 15% for test and validation. The choice of samples in each group was performed randomly and learning rate was maintained at 0.05. A categorical variable region was chosen, whilst alcohol content, density, electrical conductivity and pH were selected as continuous variables.

The number of hidden layer was 1; the number of neurons comprised was in the range from 1 to 10; maximum number of epochs was 200. Selected error functions were SOS (Sum of Squares), Cross Entropy; the activation of hidden layer were Identity, Logistic, Hyperbolic Tangent, Exponential; the functions for activation of the output layer were Identity, Logistic, Hyperbolic Tangent, Exponential, and Softmax. The employed *backpropagation* algorithm was BFGS, opted for 20 networks training, and the 5 best performing were selected by software.

Results and discussion

Figure 1 shows the rates for density (kg m^{-3}), alcohol content ($\text{g } 100^{-1} \text{ g}$), pH and electrical conductivity ($\mu\text{S m}^{-1}$) of the 204 samples analyzed, separated by commercialization regions. Horizontal lines indicate the boundaries of the compliance parameters. Results showed that all samples were within the parameters set for marketing.

Table 1 shows maximum and minimum, mean and standard deviation of each parameter used for training, test and validation of the network (MLP 4-8-2), with the best performance. Electrical conductivity and alcohol content were the compliance parameters respectively with the highest and lowest standard deviation.

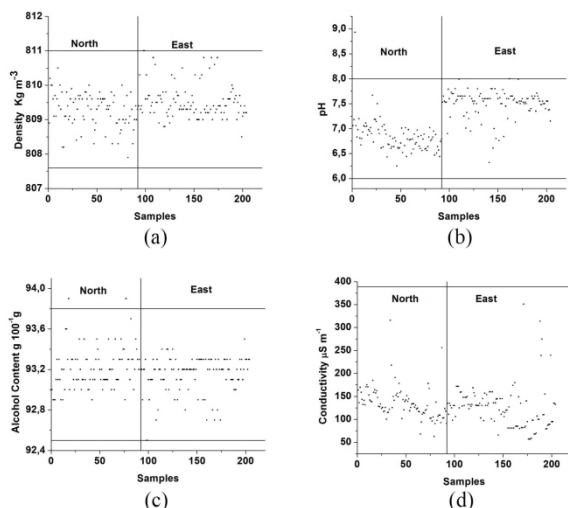


Figure 1. Data of (a) density, (b) pH, (c) alcohol content and (d) electrical conductivity for ethanol samples.

Table 1. Statistics of parameters used for training, test and validation of the network employed.

	Density (Kg m ⁻³)	Alcohol Content (g100g ⁻¹)	pH	Electrical Conductivity (μS m ⁻¹)
Training				
Minimum	808.2000	93.90000	6.250000	57.0000
Maximum	811.0000	95.50000	8.000000	351.0000
Mean	809.4160	93.18194	7.200556	131.7236
Stand. D.	0.5153	0.20162	0.444275	131.7236
Test				
Minimum	806.6000	92.80000	6.590000	57.0000
Maximum	810.5000	93.50000	8.930000	170.000
Mean	809.2733	93.18000	7.173667	119.000
Stand. D.	0.6817	0.16274	0.488301	32.2707
Validation				
Minimum	807.9000	92.70000	6.640000	68.0000
Maximum	810.8000	93.70000	7.710000	275.0000
Mean	809.5100	93.13667	7.260667	136.0500
Stand. D.	0.5708	0.24280	0.449225	22.1685
General				
Minimum	806.6000	92.50000	6.250000	57.0000
Maximum	811.0000	93.90000	8.930000	351.0000
Mean	809.4088	93.17500	7.205441	130.6309
Stand. D.	0.5588	0.19805	0.440123	41.9046

Figure 2 demonstrates each compliance parameter with other input variables. When density rates were related to alcohol content, the behavior was close to linear, albeit with $R^2 = 0.73$. The relationship of parameters to each other showed a typical nonlinear behavior with few scattered samples. The nonlinear relationship among the parameter rates showed that ANNs may be applied to the case in question, since they are inherently nonlinear (Ritter, 1995).

The 204 samples tested were classified by a MLP-type neural network. The chosen neural networks comprised an input layer with one neuron for each variable, a hidden layer responsible for the separation of standards and an output layer with the decision taken by the hidden neurons.

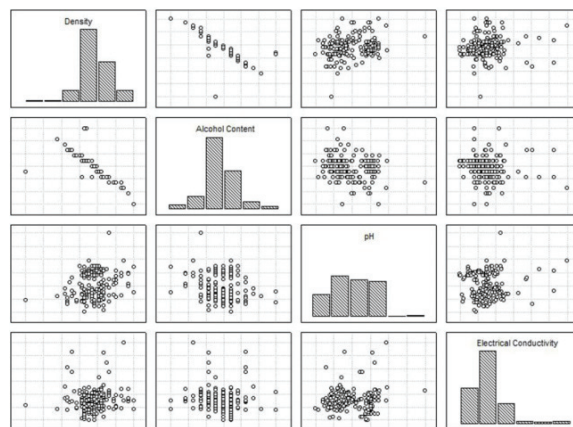


Figure 2. Matrix of graphs with parameters and their behavior.

In the training of samples, the automatic module Statistica 9.0 used activation functions identity, logistic, hyperbolic tangent and exponential type for neurons of the hidden layer and the output. Twenty networks were tested and the 5 best performances were highlighted for every start-up of the program.

Since they act characteristically as detectors, the hidden neurons had an important role in the operation of a network Perceptron learning by *backpropagation*. As the learning process advances, the hidden neurons gradually start discovering the peculiarities that characterize the training data (Haykin, 2001; Bishop, 2007).

The number of epochs and neurons in the hidden layer cannot be too high because when a neural network learns too many input-output examples, it may end up memorizing the training data. This phenomenon is known as adjustment or overtraining and makes the network lose its generalization capacity (Haykin, 2001; Bishop, 2007). Furthermore, according to Haykin (2001), the lower the learning rate parameter, the lower are the variations in synaptic weights of interaction to another network, and the softer is the trajectory of the weight space.

On the other hand, if the learning rate is very high, the modifications will result in major synaptic weights, that may render the network unstable. Therefore, in this case, a learning rate of 0.05 was applied, a maximum number of epochs equal to 200 and between 1 and 10 neurons were chosen in the hidden layer for training the MLP.

The samples and analyzed parameters presented to the network were subdivided randomly into three groups: the first consisted of the training set comprising 70%; the second group, called test, comprised 15% of samples; and the third, called validation, also comprised 15%. The second and third groups, which were not present during training, aimed at validating and verifying the

generalization capacity of the trained network (Haykin, 2001; Borsato et al., 2009; Link et al., 2014).

Error was determined at each epoch of training and the information was used to adjust the weights to reduce the error until stabilization. Figure 3 shows the number of epochs used for training the network with the highest performance, revealing that the network needed only 60 epochs to achieve stability.

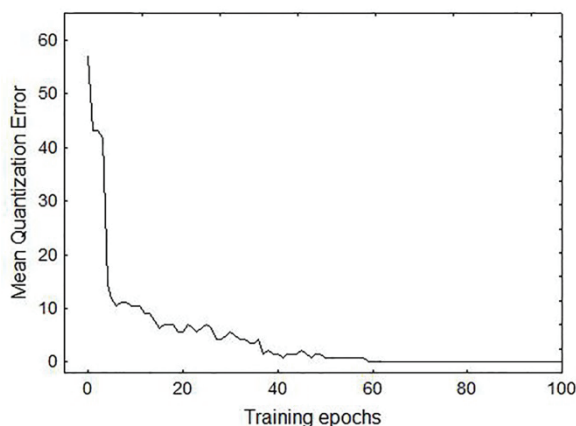


Figure 3. Representation of error obtained related to the number of training epochs of the network MLP 4-8-2.

As two main variability sources of the network were initialization and sampling, three initializations were used, totaling 60 networks. The 15 networks with the best performance, chosen by the computer program, presented a percentage of accuracy ranging between 88 and 95% for training, between 83 and 96% for test, and between 93 and 100% for validation, which were higher than those obtained by Anderson and Smith (2002) who applied ANNs in the differentiation of coffee samples by geographical origin.

In each initialization, a network was highlighted when the accuracy rate was compared with the other. Table 2 presents the three selected networks with their respective percentage of correct answers. Two of the selected networks had 8 neurons in the hidden layer. As each initialization used a random choice, different algorithms for training, error function, activation of hidden layer and output alter the hit rates. Table 2 presents the accuracy rates of the networks chosen for training (Tr), test (T) and validation (Val), as well as the characteristics of each algorithm (Al) used, including the error function (EF), the activation functions of the hidden layer (HA) and the output layer (OA).

Table 2. Accuracy percentages of selected MLPs and their characteristics.

Network	Tr (%)	T. (%)	Val (%)	Al	E.F.	H.A.	O.A.
1MLP 4-8-2	94	90	100	BFGS	SOS	Tanh	Identity
2MLP 4-8-2	95	96	96	BFGS	Entropy	Logistic	Softmax
3MLP 4-7-2	92	90	96	BFGS	Entropy	Exponential	Softmax

According to Table 2, the best performance among the networks stood out with 8 neurons in the hidden layer, and the best of 2MLP 4-8-2 with a 95 hit rate for training, 96 for test and 96% for validation. The 1MLP 4-8-2 presented 94 for training, 90 for test and 100% for validation. The third network selected showed 7 neurons in the hidden layer, with hit rates of 92 for training, 90 for test and 96% for validation. Tukey's test applied to the averages rates for training, test and validation percentages of the 15 selected networks revealed a significant difference ($p_{\max} = 0.009$) when compared to rates of the selected network (1MLP 4-8-2). The percentage of accuracy responses for the classification of samples, MLP network, according to the commercialization region, was significantly higher (Anderson & Smith, 2002).

In the case of 1MLP 4-8-2, only 11 out of 204 samples used including training, test and validation, were not classified correctly. Since there were only 9 and 15 errors respectively for 2MLP 4-8-2 and 3MLP 4-7-2, the above shows that the network 2MLP 4-8-2 had the highest number of correct responses. In fact, it was the network with the best performance in the classification of ethanol samples.

Among the algorithms used for training, the software automatic module selected the BFGS (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970), a quasi-Newton method of the second order, very efficient with very fast convergence, although a high memory is required to store the Hessian matrix (Statistica, 2009; Borsato, Pina, Spacino, Scholz, & Androcioli, 2011).

Figure 4 shows the representation of the network 2MLP 4-8-2, with 4 input parameters, 8 neurons in the hidden layer and 2 regions analyzed, whereas 3MLP 4-7-2 had 4 input parameters, 7 neurons in the hidden layer and 2 regions analyzed. In the hidden layer, neurons with more intense tones were more activated by the network in search of the target response.

From the selected networks, an order of importance could be stipulated for input parameters; in the case of network 1MLP 4-8-2, the alcohol content compliance parameter was characterized as the most important, followed by pH, density and electrical conductivity, in this order. For network 2MLP 4-8-2, the most important variable was alcohol content followed by the density, pH and

electrical conductivity. Finally, in network 3MLP 4-7-2, the order of importance was density, alcohol content, electrical conductivity and pH, with the lowest percentage of correct responses (Table 2).

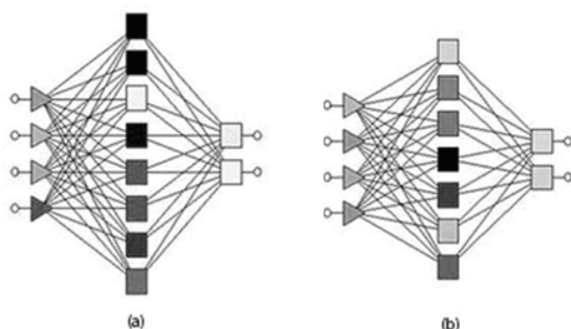


Figure 4. Representation of MLPs in (a) 2MLP 4-8-2 and (b) 3MLP 4-7-2.

This order of importance was given by the network's sensitivity analysis, in which the program calculates the sum of squared residuals or error rate for the classification of each network for the model when one of the compliance parameters is taken. Further, relationships are established between the full model, which includes the parameter and when it is deleted. The parameters of relevance order may be established from these data, or rather, those with higher rates are the most important for the network classification. Table 3 shows the sensitivity analysis of the selected networks (Statistica, 2009).

Table 3. Sensitivity analysis of selected MLP.

Networks	pH	Alcohol Content	Density	Electrical Conductivity
1MLP 4-8-2	4.887	8.584	2.991	1.218
2MLP 4-8-2	3.352	4.694	3.494	1.469
3MLP 4-7-2	0.323	2.618	3.633	2.080

Conclusion

MLP-type ANN proved to be a suitable tool for the classification of ethanol samples according to their commercialization region, whilst input variables of the sensitivity analysis reveal that the alcohol contents, followed by pH and density, were the input variables of the sensitivity analysis. Alcohol contents, followed by pH and density, were instrumental for the identification of samples with importance of compliance parameters for segmentation. The features used for training the network, the learning rate, epochs number, training algorithm, and number of hidden neurons were also effective for differentiating the samples.

References

- American Society for Testing and Materials. (2011). *ASTM D4052-11: Standard test method for density, relative density, and API gravity of liquids by digital density meter*. West Conshohocken, PA: ASTM International.
- American Society for Testing and Materials. (2014). *ASTM D1125-14: Standard test methods for electrical conductivity and resistivity of water*. West Conshohocken, PA: ASTM International.
- Anderson, K. A., & Smith, B. W. (2002). Chemical profiling to differentiate geographic growing origins of coffee. *Journal of Agricultural and Food Chemistry*, 50(7), 2068-2075.
- Associação Brasileira de Normas Técnicas. (2008). *NBR 5992: Alcool etílico e suas misturas com água - Determinação da massa específica e do teor alcoólico - Método do densímetro de vidro*. Rio de Janeiro, RJ: ABNT.
- Associação Brasileira de Normas Técnicas. (2013). *NBR 10891: Etanol hidratado combustível — Determinação do pH — Método potenciométrico*. Rio de Janeiro, RJ: ABNT.
- Bishop, C. M. (2007). *Neural networks for pattern recognition*. New York City, NK: Oxford University Press.
- Bona, E., Silva, R. S. S. F., Borsato, D., & Bassoli, D. G. (2012). Self-organizing maps as a chemometric tool for aromatic pattern recognition of soluble coffee. *Acta Scientiarum. Technology*, 34(1), 111-119.
- Borsato, D., Moreira, I., Nobrega, M. M., Moreira, M. B., Dias, G. H., Silva, R. S. S. F., & Bona, E. (2009). Aplicação de redes neurais artificiais na identificação de gasolinas adulteradas comercializadas na região de Londrina - Paraná. *Química Nova*, 32(9), 2328-2332.
- Borsato, D., Pina, M. V. R., Spacino, K. R., Scholz, M. B. S., & Androcioli, A. (2011). Application of artificial neural networks in the geographical identification of coffee samples. *European Food Research & Technology*, 233(3), 533-543.
- Broyden, C. G. (1970). The convergence of a class of double rank minimization algorithms: 2. The new algorithm. *Journal of Applied Mathematics*, 6(3), 222-231.
- Debska, B., & Guzowska-Swider, B. (2011). Application of artificial neural network in food classification. *Analytica Chimica Acta*, 705(1-2), 283-291.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13(3), 317-322.
- Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Journal of Computational Mathematics*, 24(109), 23-26.
- Haykin, S. (2001). *Redes neurais: princípios e práticas*. Porto Alegre, RS: Bookman.
- Karelina, K., Liu, Y., Alzate-Correa, D., Wheaton, K. L., Hoyt, K. R., Arthur, J. S. C., & Obrietan, K. (2015). Mitogen and stress-activated kinases 1/2 regulate ischemia-induced hippocampal progenitor cell proliferation and neurogenesis. *Neuroscience*, 285(1), 292-302.

- Kosic, D. (2015). Fast clustered radial basis function network as an adaptive predictive controller. *Neural Networks*, 63(1), 79-86.
- Link, J. V., Lemes, A. L. G., Marquetti, I., Scholz, M. B. S., & Bona, E. (2014). Geographical and genotypic segmentation of Arabica coffee using self-organizing maps. *Food Research International*, 59(1), 1-7.
- Nobrega, M. M., Bona, E., & Yamashita, F. (2013). An artificial neural network model for the prediction of mechanical and barrier properties of biodegradable films. *Materials Science & Engineering C, Biomimetic Materials, Sensors and Systems*, 33(7), 4331-4336.
- Read, S. J., Monroe, B. M., Brownstein, A. L., Yang, Y., Chopra, G., & Miller, L. C. (2010). A neural network model of the structure and dynamics of human personality. *Psychological Review*, 117(1), 61-92.
- Ritter, H. (1995). Self-organizing feature maps: Kohonen maps. In M. A. Arbib (Ed.), *The handbook of brain theory and neural network* (p. 846-851). Cambridge, UK: MIT Press.
- Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Journal of Computational Mathematics*, 24(111), 647-656.
- Silva, J. A., Damasceno, B. P. G. L., Silva, F. L. H., Madruga, M. S., & Santana, D. P. (2008). Aplicação da metodologia de planejamento fatorial e análise de superfícies de resposta para otimização da fermentação alcoólica. *Química Nova*, 31(5), 1073-1077.
- Spacino, K. R., Silva, H. C., Angilelli, K. G., Silva, E. T., Moreira, I., Mesquita, M. V., & Borsato, D. (2013). Using self-organizing maps as a chemometric tool for alcohol classification by distillery. *International Journal of Environment and Bioenergy*, 8(1), 1-11.
- Statistica. (2009). *Graphics software. Statistica for Windows, version 9*. Tulsa, OK: Statistica.
- Vuković, N., & Miljković, Z. (2015). Robust sequential learning of feedforward neural networks in the presence of heavy-tailed noise. *Neural Networks*, 63(1), 31-47.

Received on April 29, 2015.

Accepted on October 19, 2015.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.