Journal of Thermal Analysis and Calorimetry, Vol. 68 (2002) 893–899

NEURAL NETWORKS AS A TOOL FOR QUALITY ASSESSMENT OF SOYBEAN AND RAPESEED OILS ANALYZED BY THERMOGRAVIMETRY*

M. Wesołowski and B. Suchacz

Department of Analytical Chemistry, Medical University of Gdańsk, Al. Gen. Hallera 107, 80-416 Gdańsk, Poland

Abstract

The quality of commercial vegetable oils is usually evaluated via chemical parameters such as density, refractive index, saponification, iodine and acid values. In this paper, the applicability of thermal parameters for the quality assessment of vegetable oils is proposed. In order to achieve this goal, different back-propagation neural network architectures were trained, using chemical and thermal parameters as inputs. To avoid any accidental correlation due to the random initialization of the weights, each topology was repeated three times and three networks were chosen, with 5-3-2, 8-5-2 and 13-6-2 structures. The error function sum square error (SSE) was used as the criterion for finalization of the learning process. A model was developed for the correct classification of oils with regard to their type and quality.

Keywords: artificial neural networks, back-propagation of error, chemical parameters, quality assessment, thermal parameters, vegetable oils

Introduction

The quality of vegetable oils has traditionally been assessed via values of chemical parameters (density, refractive index, saponification, iodine and acid values) [1]. During recent years, however, the application of thermoanalytical methods for the study of vegetable oils has begun [2]. Although relationships have been demonstrated between the temperatures of successive mass losses and the values of chemical parameters, few publications have dealt with characterization of the quality of vegetable oils in terms of thermal parameters [3–5]. The present research was therefore intended to find thermal parameters capable of defining the quality of vegetable oils. In an attempt to solve this problem, it was decided to use artificial neural networks (ANN)s [6, 7].

The application of ANNs in analytical chemistry is receiving increasing attention in connection with data analysis problems [6–8]. This is mainly because they are applicable in all situations in which a relationship exists between predictor variables

* It was presented at CCTA 8 Zakopane, Poland

1418–2874/2002/ $ 5.00

Akadémiai Kiadó, Budapest
Kluwer Academic Publishers, Dordrecht

(inputs) and predicted variables (outputs), even when the relationship itself is very complex. Thus where problems of prediction, classification or control appear, ANNs prove to be very helpful.

Since ANNs do not differ from other data-handling applications as concerns the importance of the proper selection of data, the essential problems of data analysis do not vary as the method by which they are approached is changed. If the initial data are not representative we can encounter two situations. At best the ANN model is compromised, and at worst it is simply useless.

The aim of the present study was to create an ANN with the ability to recognize a type of vegetable oil and to evaluate its quality on the basis of chemical and thermal analysis. Moreover, it was intended to verify the utility of thermal analysis for quality evaluation and to establish whether the quality of an oil can be defined by using only thermal parameters, with omission of the generally accepted chemical parameters. We also set out to examine whether certain chemical parameters can be replaced by thermal parameters in order to estimate the quality of vegetable oil.

Experimental

Materials and methods

One hundred and one samples of commercially distributed soybean and rapeseed oils were analyzed. The chemical parameters (density, refractive index, saponification, iodine and acid values) were determined according to official Polish methods. The thermal parameters were obtained with an OD-103 derivatograph (MOM, Hungary). The thermal decomposition temperatures (T_0 and T_{100}) were read from the TG and DTG curves, whereas the temperatures of 1, 5, 15, 30, 50 and 75% mass losses (T_1, T_5, T_{15}, T_{30}, T_{50} and T_{75}) were read only from the TG curves. Detailed descriptions of the above-mentioned analytical procedures, with their results, were published elsewhere [4, 5].

Calculations

As this was mainly a classification problem, it was decided to use multilayer feedforward models. To create ANNs, Statistica Neural Network (Statsoft, Inc.) software [9] with chemical and thermal parameters as inputs was applied. As shown in Fig. 1, the ANNs consisted of three layers: input, output and hidden layers. A unit in one layer was connected to all units in the next one. The signals flowed from the first hidden layer (inputs), forward through hidden nodes, where a weighted sum of inputs was computed and passed through a sigmoid activation function, and the result was finally presented to the output layer. This process is called feedforward [10].

A proper weight setting is not known beforehand and hence the weights are initially given a random value. The process of updating the weights to a correct set of values is called training or learning, which is mostly achieved by means of a back-propagation (BP) algorithm.

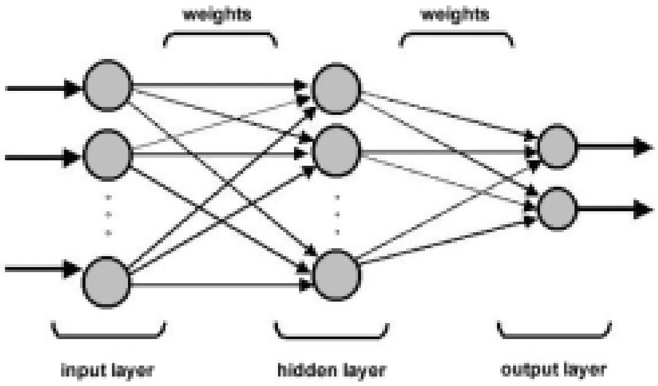

Fig. 1 Scheme of feedforward multilayer neural network

BP is a generalization of the least mean squared algorithm that modifies the ANN weight to minimize the mean squared error between the desired and actual outputs of the ANN. BP uses supervised learning in which the ANN is trained by using data for which inputs and also desired outputs are known [6–8, 10].

The adaptation of the weights is given by

$$\Delta w_{ji}(n+1)=\eta\delta_j o_i+\alpha\Delta w_{ji}(n)$$

in which η is the learning rate, α is the momentum and o_i is the output of unit i. For the hidden and output layers, δ_j is the activation function used, but additionally for the output layer, it is the difference between the desired and the obtained output [11].

The error function sum square error (*SSE*) was used as the criterion for finalization of the learning process. This could be obtained from the following equation:

$$SSE=\sum_j (t_j-o_j)^2$$

where t_j is the desired output and o_j is the actual output. For the representation of the class to which the input object belonged (output), a sequence of binary values (0 and 1) was taken.

The learning rate η and momentum α were set to 0.3 and 0.1, respectively. The transfer function of the input layer was given by $f(x)=x$. The input layer therefore operated as a flowthrough unit. For the hidden layer and the output layer, the output, o_j of unit j was a sigmoid function of the net_j to unit j, given by

$$o_j=f(net_j)=\frac{1}{1+e^{-(net_j+\theta_j)}}$$

in which θ_j is a threshold (bias) influencing the horizontal offset of sigmoid and *net* is given by

$$net_j=\sum_i w_{ji}o_i$$

The configuration of an ANN still remains a difficult problem. It involves the choice of ANN architecture (a number of inputs and outputs; and the way in which the outputs of the neurons are connected to the other neurons), error criterion, learning algorithm, activation function, the number of hidden layers and how many hidden units per layer there should be. When these ANN components were specified, the ANN was presented with the test data and training began. During the training process, the performance of the ANN was measured not only on the training set, but also on the test set. When the lowest root mean squared error was reached, the training was terminated.

Results and discussion

The results of chemical analysis (density, refractive index, saponification, iodine and acid values) and of thermal analysis (T_0, T_1, T_5, T_{15}, T_{30}, T_{50}, T_{75} and T_{100}) were used as input variables. The cases in the data set were randomly divided into two subsets: one for training the ANNs (81 samples) and one for independent testing (20 samples). The procedure employed in the study is presented schematically in Fig. 2.

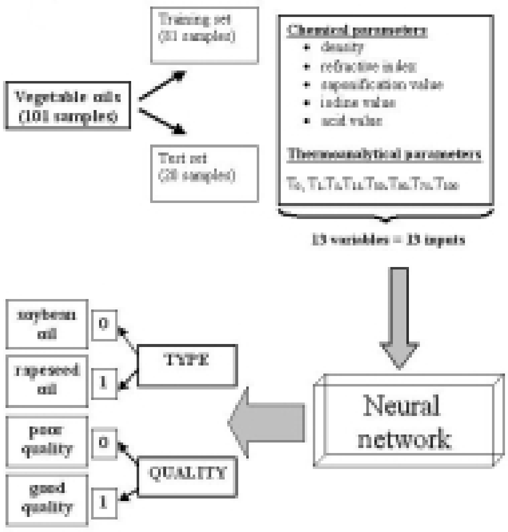

Fig. 2 The procedure employed in the study

In an effort to determine the optimum number of hidden nodes, a series of different architectures were used in which the nodes were altered from 1 to 13. To avoid any accidental correlation due to the random initialization of the weights, each topology was repeated three times and three ANNs were chosen, with 5-3-2, 8-5-2 and 13-6-2 structures.

A model was developed for the correct classification of oils with regard to their type and quality. A smaller number of hidden units (3) and reduction of the initial input variables (5) caused the ANN to fit the training and test sets better, with relatively small errors of 0.0956 and 0.0779.

The ANN with a full set of 13 variables and 6 hidden nodes also performed very well, with errors 0.1117 and 0.0001.

The ANN with only thermoanalytical parameters as input variables could not classify oils property. Its error for the training and test sets were the highest, at 0.2614 and 0.2211.

The percentages of false quality and type classification of oil samples on the basis of the different inputs and the architectures of the ANNs are shown in Table 1. The results obtained prove that the thermal parameters alone are insufficient for quality assessment. While the other ANNs had the ability to classify oils according to their type, the 8-5-2 ANN could not even handle this problem, as demonstrated by the high percentage of misclassified oil samples.

Table 1 Percentages of false quality and type classification of oil samples for different inputs and architectures of networks

	ANN with chemical parameters as inputs		ANN with thermal parameters as inputs		ANN with chemical and thermal parameters as inputs	
	5-3-2		8-5-2		13-6-2	
	training set	testing set	training set	testing set	training set	testing set
	Quality classification					
Incorrectly classified	0	5	7.41	0	2.47	0
Unclassified	4.93	15	27.16	35	2.47	0
	Type classification					
Incorrectly classified	0	0	1.23	20	0	0
Unclassified	0	0	13.58	10	0	0

ANN=artificial neural network

Table 2 relates to the sensitivity analysis for the training sets. It presents information on the relative importance of the variables used in the ANNs employed in the research. The Error is the most important sensitivity figure. This indicates the performance of the ANN if the given variable is omitted. Important variables have a high error, indicating that the ANN performance deteriorates considerably if they are not present. The Ratio reports the ratio between the Error and the error of the ANN if all variables are used. If the Ratio is one or lower, then removal of the variable either has no effect on the performance of the ANN, or actually enhances it. The Rank represents the relative importance of the parameters in predicting the output of the ANN. The higher the number, the more the variable contributes to the classification [9].

Table 2 Sensitivity analysis of artificial neural network architectures used in the study

Parameters	ANN with chemical parameters as inputs			ANN with thermal parameters as inputs			ANN with chemical and thermal parameters as inputs		
	5-3-2			8-5-2			13-6-2		
	error	ratio	rank	error	ratio	rank	error	ratio	rank
Density	0.1453	1.52	5				0.1116	0.99	13
Refractive index	0.2169	2.27	4				0.1606	1.44	5
Saponification value	0.2352	2.46	3				0.1641	1.47	4
Iodine value	0.3129	3.27	2				0.2367	2.12	2
Acid value	0.4212	4.40	1				0.4111	3.68	1
T_0				0.3853	1.47	5	0.1363	1.22	7
T_1				0.6173	2.36	1	0.1523	1.36	6
T_5				0.3934	1.50	4	0.1135	1.01	11
T_{15}				0.5245	2.01	2	0.1226	1.10	8
T_{30}				0.4499	1.72	3	0.1119	1.00	12
T_{50}				0.2807	1.07	8	0.1213	1.09	9
T_{75}				0.3051	1.17	7	0.1158	1.04	10
T_{100}				0.3833	1.47	6	0.1980	1.77	3

Although the sensitivity analysis does not rate the importance of variables in modelling in an absolute manner, in practice it is extremely useful. Identifying key variables that are always of high sensitivity is often possible when a number of ANN architectures are studied.

The results of sensitivity analysis show that in the ANNs with 5-3-2 and 13-6-2 topologies the least important parameter is density, because of its low rank and error values. Accordingly, it can be omitted, especially in the ANN with 13 parameters used as inputs. In spite of the fact that the 8-5-2 ANN was unable to classify oils properly, the thermal parameters of least importance were T_{50} and T_{75}, and their elimination would not affect the performance of the ANN. Sensitivity analysis of the ANN with chemical and thermal parameters implies that T_0, T_1 or T_{100} can be used as alternative variables to density in quality evaluation.

Conclusions

The advantages of ANNs for quality assessment were examined in this study. It is shown that an appropriate selection of the architectures of the ANN permits a quite accurate evaluation of oil quality.

The results are encouraging and confirm the possibility of employing thermal parameters to determine the quality of oils. However, the research is still in the preliminary stages and there are still a number of improvements to be made, especially in selecting the most important variables with which ANNs could be used more effectively and which would provide functionally to the commercial assessment of the quality of vegetable oils.

References

1 S. Krauze, Zarys nauk o środkach żywności (An overview of food science), PZWL, Warszawa 1975.
2 Vademecum of applications in thermal analysis, Setram Groupe, Caluire, France 1995.
3 M. Wesołowski, J. Thermal Anal., 32 (1987) 1781.
4 M. Wesołowski and J. Erecińska, Thermochim. Acta, 323 (1998) 137.
5 M. Wesołowski, Fat. Sci. Technol., 10 (1993) 377.
6 J. Zupan and J. Gastaiger, Neural networks for chemists, VCH, Weinheim 1993.
7 T. B. Blank and S. D. Brown, Anal. Chim. Acta, 277 (1993) 273.
8 G. Kateman and J. R. M. Smits, Anal. Chim. Acta, 277 (1993) 179.
9 Statistica Neural Networks (manual), Statsoft Inc., USA 1998.
10 C. Bishop, Neural Networks for Pattern Recognition, University Press, Oxford 1995.
11 J. R. M. Smits, W. J. Melssen, M. W. J. Derksen and G. Kateman, Anal. Chim. Acta, 284 (1993) 91.