ASN, Vol 3, No 1, Pages 19-24, 2016

DE GRUYTER OPEN

Acta Scientifica Naturalis

Former Annual of Konstantin Preslavsky University – Chemistry, Physics, Biology, Geography
Journal homepage: http://www.shu.bg

Received: 30.10.2015
Accepted: 11.03.2016

Upon the opportunity to apply ART2 Neural Network for clusterization of biodiesel fuels

T. Petkov[1], Z. Mustafa[1], S. Sotirov[1], R. Milina[1], M. Moskovkina[2]
[1]*University "Prof. Dr. Assen Zlatarov" Bourgas, Bulgaria*
[2]*Konstantin Preslavsky University of Shumen, Shumen, Bulgaria*

e-mail: todor_petkov@btu.bg

Abstract*: A chemometric approach using artificial neural network for clusterization of biodiesels was developed. It is based on artificial ART2 neural network. Gas chromatography (GC) and Gas Chromatography – mass spectrometry (GC-MS) were used for quantitative and qualitative analysis of biodiesels, produced from different feedstocks, and FAME (fatty acid methyl esters) profiles were determined. Totally 96 analytical results for 7 different classes of biofuel plants: sunflower, rapeseed, corn, soybean, palm, peanut, "unknown" were used as objects. The analysis of biodiesels showed the content of five major FAME (C16:0, C18:0, C18:1, C18:2, C18:3) and those components were used like inputs in the model. After training with 6 samples, for which the origin was known, ANN was verified and tested with ninety "unknown" samples. The present research demonstrated the successful application of neural network for recognition of biodiesels according to their feedstock which give information upon their properties and handling.*

Keywords*: biodiesel fuel, fatty acid methyl esters, gas chromatography, vegetable oil, artificial neural networks.*

Introduction

Artificial neural networks [1] are inspired by the biological neural network of humans and animals. They can be used for recognition and identifying the things in places where known methods are not capable of solving problems. During the years lots of neural networks [2-4] were developed and each one of them can be used to solve specific tasks. Each artificial neural network is composed of simple processing elements called neurons (Figure 1) [1, 5] connected with each other to arrange in a layer. Depending of the type of the neural network the neurons in a layer apply different calculations and there are two types of learning known as supervised and unsupervised.

In supervised learning [6] each input sample has target where target is the correct output of the network thus in the learning procedure each input is associated with the correct output. When the network is learned it is ready for test and the result of testing is according to the input – output pairs.

In unsupervised learning [7] there are not targets and the network learns according to the input samples only. The goal is to group those samples that have common characteristics this grouping of samples is also known as clustering and it is achieved through the learning trial. Unsupervised networks are based on competition where one neuron wins and it will learn its weights in accordance to the vector.

19 Corresponding author: todor_petkov@btu.bg

DOI: 10.1515/asn-2016-0003

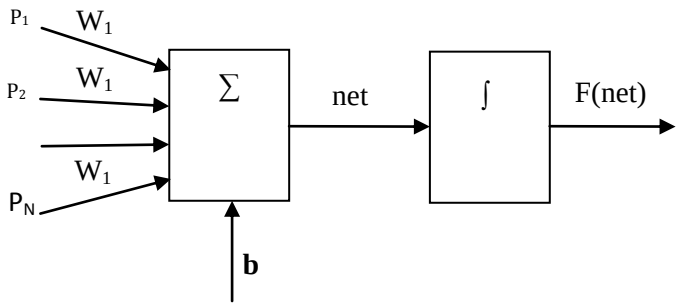

Figure. 1. Artificial neuron

During the years there were lots of papers devoted to the application of the neural networks to different problems [2, 3]. The meaning of recognition is important in many areas because of the fact that there might not be necessarily to perform some of the experiments, here is shown a method of clusterization of biodiesel fuels.

The goal of this work was to group biodiesels from different feedstocks by a method based on neural network using their gas chromatographic FAME profiles. To the aim some available literature data on FAME profiles of biodiesels from several types of oils were used [8-11]. In the investigations also our data of FAME of Bulgarian biodiesels and samples produced in our laboratory by transesterification of plant oils were utilized. Only those samples meeting the requirement of EN 14214 [12] were included in our study. The analysis of the samples was performed using gas chromatography (GC) and mass spectrometry (MS).

Experimental

Samples

Samples № 17, 28, 47, 53 and 60 to 83 (Table 1) were bought from the market. Three biodiesels (samples 8, 48 and 80) were laboratory synthesized by methanol transesterification of sunflower, rapeseed and palm oils respectively.

Gas chromatography (GC)

GC analyses were performed on a GC system Agilent Technologies 7890A equipped with FID [13].

Artificial Neural Networks (ANN)

In the first step the network has to be learned and after that we can use for the recognitions and for predictions of the properties of the materials. Figure 2 shows in abbreviated notation of ART2 [14] neural network.

20 Corresponding author: todor_petkov@btu.bg

DOI: 10.1515/asn-2016-0003

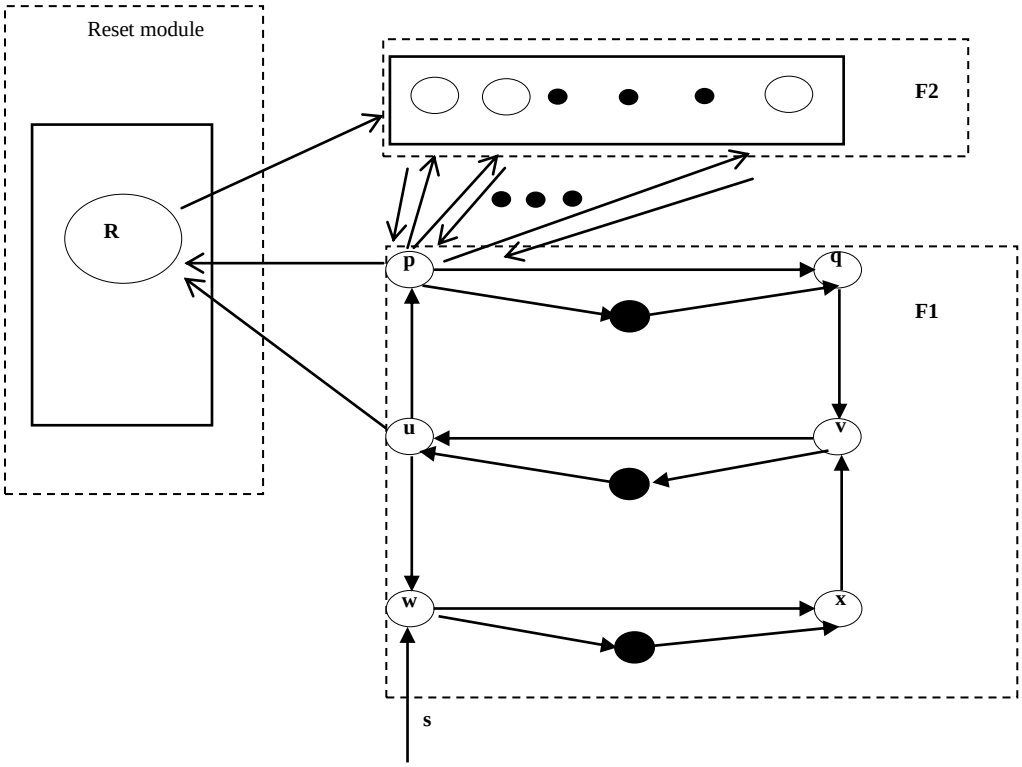

Figure. 2. ART2 neural network

21 Corresponding author: todor_petkov@btu.bg

DOI: 10.1515/asn-2016-0003

Table 1. FAME (percentage levels) of biodiesel samples

Sample №	Recognition Cluster	C16:0	C18:0	C18:1	C18:2	C18:3	Sample №	Recognition Cluster	C16:0	C18:0	C18:1	C18:2	C18:3
1	1	6,3	3,9	20,8	67,8	0,2	46	2	4,1	1,6	55,7	17,8	7,6
2	1	6,1	4,1	24	64,3	0,1	47	2	4,5	1,6	60,8	17,7	7,9
3	1	7,2	4,5	25,6	61,1	0,2	48	2	4,6	1,8	55,9	17,7	5,3
4	1	6,2	4,1	26,8	61,4	0,1	49	2	4,8	2.0	60,4	23,3	9,5
5	1	6,1	4,0	25,2	63,2	0,1	50	2	4	2,5	61,2	20,6	11,7
6	1	6,1	4,1	26,7	61,0	0,6	51	3	10,7	1,8	30,2	55,5	0,9
7	1	6,2	4,0	25,8	62,2	0,4	52	3	10	2,3	36,9	47,1	2,2
8	1	6,0	3,8	29,6	57,7	1,2	53	3	11,5	1,9	26,6	58,7	0,8
9	1	6,5	2,8	23,2	66,1	0,1	54	3	10,1	1,6	26,9	59,9	0,7
10	1	7,0	3,4	23,2	64,9	0,1	55	3	11,8	2,1	35,5	46,6	1,4
11	1	6,0	4,0	28,3	60,1	0,1	56	3	12,3	1,7	25,3	60.0	0,7
12	1	6,1	4,1	25,1	63,5	0,5	57	3	12,9	1,7	25,3	59,5	0,6
13	1	6,3	4,1	27,4	61,4	0,7	58	4	11,4	4,1	23,5	53,5	6,6
14	1	6,2	4,1	27,1	61,8	0,7	59	4	10,9	3,6	26.0	52,6	5,5
15	1	6,2	4,2	27,0	60,9	0,5	60	4	10,5	4,2	25,5	52,0	7,8
16	1	6,2	4,0	28,3	59,7	0,9	61	4	10	4,2	24,9	53,2	6,9
17	1	5,6	4,2	25,7	58,9	1,7	62	4	10,4	4,2	25,9	50,8	7,5
18	1	6,4	3,7	25,8	63,5	0,5	63	4	10,5	4,2	24,4	52,1	7,5
19	1	6,8	4,3	26,4	60,5	1,3	64	4	10,5	4,3	24,6	53,1	7,6
20	1	6,0	4,4	27,1	60,9	0,9	65	4	10,2	4,0	23,1	55,1	7,1
21	1	6,4	4,8	25,3	61,8	1,0	66	4	10,9	3,8	27,2	49,5	6,4
22	1	5,9	4,5	24,1	61,7	0,9	67	4	11,9	3,8	25,7	52,7	5,8
23	1	6,6	4,1	29,9	57,8	1,3	68	4	9,7	3,9	25,1	54,2	5,9
24	1	6,6	4,7	24,5	62,8	0,3	69	3	13,9	2,1	23,2	56,2	4,3
25	1	6,4	4,4	24,4	63,7	0,4	70	4	10,0	4,0	23,0	51,0	7,0
26	1	6,0	3,0	17,0	74,0	0	71	4	11,9	4,1	23,2	54,2	6,3
27	1	6,4	2,9	17,7	72,9	0	72	4	11,7	3,8	22	52,6	7,7
28	1	6,2	3,7	25,2	63,2	0,2	73	4	11,3	3,6	24,9	53	6,1
29	1	6,1	3,1	30	57,9	0,5	74	4	11,6	3,9	23,7	53,8	5,9
30	1	6,4	3,6	21,7	66,3	1,5	75	4	11,2	4	21,9	53,8	7,3
31	1	6,2	3,7	25,2	63,1	0,2	76	5	42,4	4,2	40,9	10	0,3
32	1	6,4	3,1	27,4	62,1	0,1	77	5	42,5	4,2	41,3	9,5	0,3
33	2	4,1	1,6	62,2	20,6	8,7	78	5	36,7	6,6	46,1	8,6	0,3
34	2	5,1	2,3	55,9	27,4	6,8	79	5	46,1	4,3	38,4	10,0	0,2
35	2	4,8	1,8	62,6	20,0	9,5	80	5	45,0	5,0	40,0	10,0	0
36	2	5,5	1,7	59	21,3	9,3	81	6	10,3	2,8	47,6	31,5	0,5
37	2	5,1	1,9	59,2	22,3	9,3	82	6	8	1,8	53,3	28,4	0,3
38	2	4,8	1,9	61,6	20,9	8,0	83	6	9,7	3,4	59,3	20,5	0,1
39	2	5,4	2,0	53,2	28,9	7,3	84	6	10	3,3	60	21,3	0,2
40	2	5,7	2,1	54,6	26,8	8,0	85	6	9,6	3,3	57,7	20,7	0,2
41	2	6,2	2,2	52,2	29	8,0	86	2	5,1	2,1	46	34,5	4,9
42	2	4,2	1,6	59,5	21,5	8,4	87	1	5,5	3,5	22,2	54,6	0,1
43	2	3,5	0,9	64	22	8,1	88	0	5,7	2,9	38,4	45,9	2,5
44	2	3,5	0,9	54,1	22,3	8,0	89	0	5,6	2,9	34,8	47,0	2,0
45	2	4,1	1,8	58,6	22,2	13,3	90	1	8,5	3,4	31,1	48,1	0,1

22 Corresponding author: todor_petkov@btu.bg

DOI: 10.1515/asn-2016-0003

ART2 is designed to perform operation over continuous valued input vectors or binary input vectors that have noise. Basically the network is consists of two layers composed of neurons that are fully connected with set of weights also known as bottom-up and top-down and one orienting sub system. First layer consists of three sub layers of neurons each sub layer supports combination of normalization of the vector and suppression of the noise. Second layer is competitive one and the neuron with maximum value will win therefore its weights will learn according to the vector. The orienting sub system takes design whether the neuron with maximum value responds to the criteria that we are inserted or not. The equations according to [15] describing this operation are:

Step 1. Initialize parameters:

- a, b, Θ, c, d, e, α, ρ, n, m, b_{ij}, t_{ji}

where: a, b – fixed weights in the F1 layer,.Θ - noise suppression parameter,.c – fixed weight used in testing for reset,.d – activation of winning F2 unit,.e – small parameter using preventing division by zero,.α – learning rate,.ρ – vigilance threshold,.n – number of input units (F1 layer), m – number of cluster units (F2 layer), b_{ij} – initial bottom-up weights typical values

Step 2. For each input vector "s "do steps 3-11.

Step 3. Update F1 unit activation: $u_i = 0,\ x_i = \frac{w_i}{e + \|w\|}, q_i = 0$

$w_i = s_i,\ p_i = 0, v_i = f(x)_i$

The activation function is $f(x) = \begin{cases} x \text{ if } x \geq \Theta \\ 0 \text{ if } x < \Theta \end{cases}$

Update F_1 activations again

$u_i = \frac{v_i}{e + \|v\|},\ x_i = \frac{w_i}{e + \|w\|},\ q_i = \frac{p_i}{e + \|p\|}$

$w_i = s_i + a * u_i,\ p_i = u_i,\ v_i = f(x)_i + bf(q)_i$

Step 4. Compute the signals to F2 units: $y_j = \sum_j b_j * p_i$

Step 5. While reset is true, do steps 6-7

Step 6. Find F2 unit with largest signal. (Define J such that $y_J \geq y_j$ for j=1, 2…m.)

Step 7. Check for reset: $u_i = \frac{v_i}{e + \|v\|},\ r_i = \frac{u_i + cp_i}{e + \|u\| + c\|p\|},\ p_i = u_i + dt_{Ji}$

If $\|r\| < p - e$ then y_J=-1 (inhibit J) (Reset is true; repeat step 5)

If $\|r\| \geq p - e$ Reset is false; proceed to step 8

Step 8 Update weights for winning unit J

$t_{Ji} = \alpha d u_i + \{1 + \alpha d(d-1)\} t_{Ji} \qquad b_{iJ} = \alpha d u_i + \{1 + \alpha d(d-1)\} b_{iJ}$

Results and Discussion

The ANN model was built using the biodiesels for which the origin was known. After training with 6 samples from each group (Table 2), the neural network was verified and tested with 90 samples. Samples from number 1 to 85 are regarded as unknown samples and information about them were used for verification of the network. In order to test the network samples from number 86 to 90 were used. Below the target and conditions of training and results obtained are given.

The number of neurons (clusters) in the second layer is six, where each cluster represents the type of biofuel plant. In case that the network returns 0 output means that none of the clusters responds to the vector. In order to normalize the input data we multiplied columns C16:0 by 4, C18:0 by 10 and C18:3 by 10

Table 2 Normalized Learning vectors

class	C16:0	C18:0	C18:1	C18:2	C18:3
1	24,8	39	27,1	59,9	13
2	18	17	64,9	18,6	83
3	47,2	21	27,4	57,7	6
4	38,4	35	30,3	49,2	59
5	170	44	40,5	10,1	2
6	45,6	24	48,3	32	9

In Table 3 the parameters used for the training of Neural Network are shown.

23 Corresponding author: todor_petkov@btu.bg

DOI: 10.1515/asn-2016-0003

Table 3. Parameters for the training of Neural Network

a	b	Θ	c	d	e	α	ρ	t_{ji}	b_{ij}	n	m
10	10	0,45	0,1	0,9	0,1	0,6	0,8157	0	4,4721	5	6

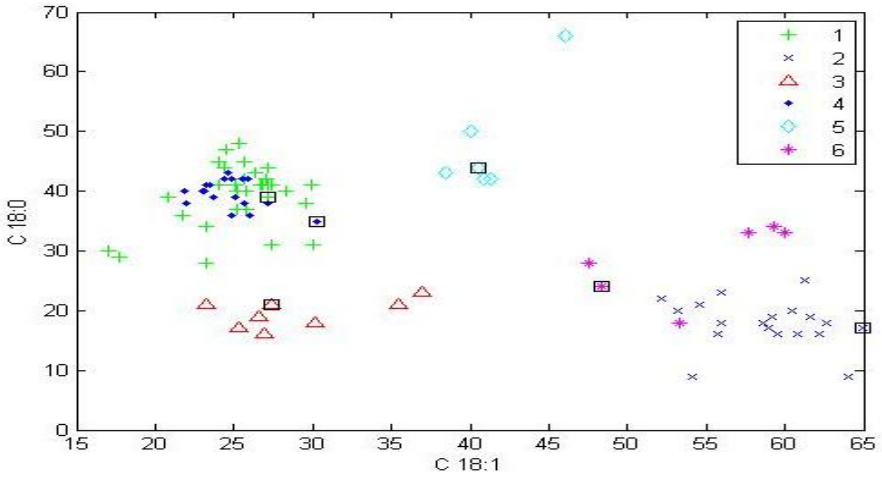

Fig. 3. Representation of the clusters and test vectors according to C18:0 and C18:1

For testing of ANN samples 86 to 90 were used (Table 1) with the same five inputs and the real outputs are given in Table 4.

Table 4. Representation of the Euclidean distance between clusters and "unknown" samples

Sample №	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
86	1.0402	**0.3630**	1.0402	0.7401	1.0402	0.7792
87	**0.3654**	1.0468	0.6846	0.7595	1.0468	1.0468
88	0.6820	0.8369	0.7154	0.7805	1.0374	0.7954
89	0.6614	0.8542	0.6975	0.7672	1.0386	0.8165
90	**0.5135**	1.0475	0.5149	0.8130	0.7847	0.8756

The classes of the samples are marked with symbols: "+"; "x"; "◊";"♦";"Δ";"*" in Figure 3 where each symbol represents different class. Since we use two columns C18:0 and C18:1 to draw the graphic some of the vectors seem that belongs to different clusters. In Table 4 the Euclidean distance between each unknown sample and clusters are shown.

In previous paper [13] it was developed neural network of type multilayer perceptron for classification of biodiesels. The network was trained by 85 samples and tested with eight "unknown" samples the components (C16:0, C18:0, C18:1, C18:2, C18:3) were used like inputs in the model. In this paper ART2 neural network was used for clustering the types of biodiesels. The network was trained by 6 samples, verified by 85 and tested by five unknown samples the same components (C16:0, C18:0, C18:1, C18:2, C18:3) were used like inputs in the model. It can be concluded that ART2 NN needs less training examples than multilayer perception (MLP) NN and if there is lack of data it can be used for better recognition than MLP. One of the advantages that ART2 NN also characterizes is that it shows the samples that does not hit any of the clusters.

Conclusion

The present research demonstrates the successful application of ART2 neural network for recognition of biodiesels according to their feedstock. The proposed model was created on the base of content of five major components: esters of C16:0, C18:0, C18:1, C18:2, C18:3 acids in the FAME profiles of biodiesel from vegetable oils origin. Those components were used like inputs in the model. Totally 96 analytical results for 7 different classes of biofuel plants: sunflower, rapeseed, corn, soybean, palm, peanut, "unknown" were used as objects. The model was trained with 6 samples, for which the origin was known then it was verified with 85

24 Corresponding author: todor_petkov@btu.bg

DOI: 10.1515/asn-2016-0003

and tested with five "unknown" samples. The obtained predictions correlate well with the available information of the samples. The quality of the prediction of mixed samples composition cannot be confirmed since there is no exact information from the producers about oils used. We anticipate further application of the model to recognize of biodiesels, produced from mixed oils.

The overview of predicted results indicates that the proposed model is of significant value for the determination of unknown biodiesels and could be implemented as an efficient method that enables the prediction of the raw material. The latter provides information upon the properties and handling of biodiesel fuels.

References

[1]. Hagan, M.; Demuth, H.; Beale, M., Neural Network Design, Boston, MA: PWS Publishing, **1996**.

[2]. Rocabruno-Valdez, C.I.; Hernandez, J.A., Artificial neural network models to predict density, dynamic viscosity and cetane number of biodiesel, *Fuel*, **2015**, *147*, 9–17.

[3]. Pandey, M.; Pandey, A. K., Application of chemometric analysis and Self Organizing Map-ANN source receptor modeling for metal speciation in river sediment, *Environmental Pollution*, **2015**, *204*, 64–73.

[4]. Oonk, S.; Spifker, J., A supervised machine-learning approach towards geochemical predictive modelling in archaeology, *Journal of archaeological science*, **2015**, *59*, 80-88,.

[5]. Haykin, S., Neural Networks: A Comprehensive Foundation, NY: Macmillan, **1994**.

[6]. Adam, A.; Ibrahim, Z., A two-step supervised learning ANN for imbalanced dataset problems, *International Journal of Innovative Computing*, **2012**, 8.

[7]. Rojas, R., Unsupervised learning and Clustering algorithms, *Springer*, **1996**, 109-131.

[8]. Ramos, M.J.; Fernandez, C.M.; Casas, A.; Rodriguez, L.; Perez, A., Influence of fatty acid composition of raw materials on biodiesel properties, *Bioresource Technology*, **2009**, *100*, 261–268.

[9]. Moser, B.R.; Vaughn, S.F., Efficacy of fatty acid profile as a tool for screening feedstocks for biodiesel production, *Biomass and bioenergy*, **2012**, *37*,31–41.

[10]. Brodnjak-Voncina, D.; Kodba, Z.C.; Novic, M., Multivariate data analysis in classification of vegetable oils characterized by the content of fatty acids, *Chemom. Intell. Lab. Syst.*, **2005**, *75*, 31-43.

[11]. Giakoumis, E.G., A statistical investigation of biodiesel physical and chemical properties, and their correlation with the degree of unsaturation, *Renewable Energy*, **2013**, *50*, 858-878.

[12]. EN 14214:**2008** + A1, Automotive fuels – FAME for diesel engines – Requirements and test methods.

[13]. Mustafa, Z.; Surchev, S.; Milina, R.; Sotirov, S., A contribution to the recognition of biodiesel fuels according to their FAME profiles by the ANN, *Petroleum & Coal*, **2015**, *57(1,)* 40-47.

[14]. Grossberg, G. A. Carpenter ART2: Self-organization of stable category recognition codes of analog input patterns, *Boston University, Center for Adaptive systems*, **1987,** *23*, 4919-30.

[15]. Fausett, L., Fundamentals of Neural Networks; Architecture algorithms and applications, **1993**.

 Corresponding author: todor_petkov@btu.bg

DOI: 10.1515/asn-2016-0003