Revue Roumaine de Chimie 1995, 40(6), 567—574

ARTIFICIAL NEURAL NETWORKS APPLICATIONS. PART 1

ESTIMATION OF THE TOTAL π-ELECTRON ENERGY OF BENZONOID HYDROCARBONS

OVIDIU IVANCIUC

Polytechnic University of Bucharest, Faculty of Chemistry, Department of Organic Chemistry, Splaiul Independenţei 313, 77206 Bucharest, Roumania

Received December 7, 1993

An approximate method that allows effcient calculation of the total π-electron energy (E_π) of benzenoid hydrocarbonsi(BH) is presented. Artificial neural networks (ANN) are used for modeling the relationship between topological features of BH and E_π. The ANN was trained by a test calculation on 265 benzenoid hydrocarbons with up to 7 benzenoid rings. Numerical calculations demonstrate effectiveness of the present approach. It is found that E_π is well approximated by only two topological features : the number of carbon atoms (N) and the number of Kekulé structures (K).

INTRODUCTION

The Hückel total π-electron energy (E_π) is one of the most important topological properties of conjugated molecules. A great part of the work toward elucidation of the stability of conjugated alternant molecules on the ground of chemical topology has been devoted to the derivation and numerical verification of various approximate formulae that express E_π as a function of carbon atoms (N), carbon-carbon bonds (M) and Kekulé structures (K).

Such an approximation to E_π for benzenoid hydrocarbons (BH) was derived by Cioslowski [1] (equation (1)), enabling one to partition E_π into a term corresponding to the energy due to formation of carbon-carbon bonds, and a second term due to cyclic conjugation represented by K. Other mathematical models for E_π have been reported. [2–8]

$$E_\pi \simeq 0.79848\ (2MN)^{1/2} + 0.13853\ N\ K^{2/N} \qquad (1)$$

The goal of the present study is to investigate the applicability of artificial neural networks (ANN), a methodology of nonlinear modeling, to express the relationship between topological features of BH (N, M, and K) and E_π calculated within the framework of the simple tight-binding Hückel MO model.

The advantage gained through the use of ANN is due to the unique capacity of ANN both to learn the underlying rules inherent in a training set of data and to generalize using those rules to a new set of data.[9-15] Recent research efforts have focused on exploring the use of neural networks for predicting ^{13}C-NMR chemical shifts, estimating aqueous solubilities, classifying mass spectra, identifying ^{1}H-NMR-spectra, modeling structure-activity relationships.

Artificial neural networks are tipically composed of interconnected units, which serve as model neurons. The function of the synapse is modeled by a modifiable weight, which is associated with each connection.

The commonest type of ANN consists of three layers of units : a layer of input units is connected to a layer of hidden units, which is connected to a layer of output units. This type of ANN is termed a multi-layer perceptron (MLP) network. The activity of the input units represents the raw information that is fed into the network. The activity of each hidden unit is determined by the activities of the input units and the weights on the connections between the input and hidden units. Similarly, the behaviour of the output units depends on the activity of the hidden units and the weights between the hidden and output units.

Learning is the way a network builds an internal representation of its environment. This environment consists of a set of training patterns, in our case the set of BH characterized by their topological features and E_π. The functionality of the network depends on the learning rule and architecture.

During the learning phase, the connection weights are adjusted step-by-step, leading to correlations building up between input data and the corresponding target data. The network's knowledge is distributed among all network connections.

RESULTS AND DISCUSSION

A major problem with multilayer ANN is determining how many hidden layer neurons should be for a given application. The number can vary widely and depends on such factors as the correlation structure between dependent and independent variables and their number, as well as the number of patterns in the training set. If too few hidden neurons are included then the network will not train very well and may not be able to provide a very satisfactory input-output mapping. However, if too many neurons are used, in addition to the extra time it will take to train, the final network will have a tendency to memorise the training data.

The ANN used in the present study was a MLP with three layers, trained with the backpropagation algorithm; the transfer function was the hyperbolic tangent, which was found to perform better than the usual sigmoid function. Other specifications for the networks used in simulations are presented in Table 1.

Table 1

Artificial neural network specifications

Type of neural network	Multi-layer perceptron
Learning algorithm	Backpropagation
Bias neuron	Yes
Learning set presentation	Random
Learning rate	0.01
Momentum	0.8
Transfer function	Tanh
No. input neurons	3 (NN1)/2 (NN2)
No. hidden neurons	variable
No. output neurons	1
Input scaling (min/max)	−0.9/0.9
Output scaling (min/max)	−0.9/0.9
Initial weights scaling (min/max)	−0.1/0.1

The quality of ANN output was assessed by three statistical variables : the mean square error (MSE) in scaled units, and the standard deviation s and the correlation coefficient r of the linear correlation between $E_{\pi\ HMO}$ and $E_{\pi\ ANN}$ of the type $E_{\pi\ HMO} = a + b\ E_{\pi\ ANN}$. MSE is defined by the following expression :

$$MSE = \frac{\sum (y_{observed} - y_{predicted})^2}{P \cdot O}$$

where P is the number of patterns in the test set and O is the number of neurons in the output layer. High-quality ANN predictions should have MSE and s close to zero, and r close to unity.

Besides these three statistical indices, we will use $\Delta E_{\pi\ av}$, defined as $\Delta E_{\pi\ av} = \sum |E_{\pi\ ANN} - E_{\pi\ HMO}|/P$, and $\Delta E_{\pi\ max}$, the greatest difference between $E_{\pi\ ANN}$ and $E_{\pi\ HMO}$ for the whole set of P patterns.

The number of nodes in the hidden layer was selected on the basis of empirical trials, in which ANN with different number of hidden neurons are trained to predict the E_π for BH.

The training was done by presenting the three selected topological features (N, M and K) for a set of 265 BH, containing all BH with 2 benzenoid rings (naphthalene) up to BH with 7 benzenoid rings, forming the training set TS1. The set of BH was presented randomly, each example being presented the same number of times. Five ANN were generated, with the number of hidden neurons between 1 and 5. The training was terminated after 5000 complete cycles, and the results obtained in the evaluation of E_π are presented in Table 2.

It can be seen from Table 2 that the networks give almost identical results in terms of MSE, r and s for all hidden layer sizes between 3 and 5 neurons. The performance of these networks begins to decline as the hidden layer size is reduced to 2 neurons.

Table 2

Statistical results in estimating E_π for benzenoid hydrocarbons with the network NN1 trained for 5000 cycles with the training set TS1

No. of hidden neurons	MSE ·10⁵	s	r	$\Delta E_{\pi\ av}$	$\Delta E_{\pi\ max}$
1	120.56	0.493	0.9945	0.419	5.087
2	4.67	0.110	0.9997	0.068	1.036
3	3.48	0.091	0.9998	0.066	0.851
4	3.27	0.092	0.9998	0.060	0.831
5	3.14	0.089	0.9998	0.061	0.811

On the other hand, although $\Delta E_{\pi\ av}$ is small, in all cases the largest error ($\Delta E_{\pi\ max}$) is exhibited by naphthalene, the only BH with 2 benzenoid rings. This situation happens when a pattern is not well represented in the training set and the solution is to present that pattern to the network more than one time in a cycle.

In a second trial naphthalene was presented for five times in the training set, making a total of 269 patterns and representing the training set TS2. Again five ANN with 1 to 5 hidden neurons were trained for 5000 cycles; the results of the simulations are presented in Table 3.

Table 3

Statistical results in estimating E_π for benzenoid hydrocarbons with the network NN1 trained for 5000 cycles with the training set TS2

No. of hidden neurons	MSE ·10⁵	s	r	$\Delta E_{\pi\ av}$	$\Delta E_{\pi\ max}$
1	134.36	0.580	0.9924	0.498	3.120
2	5.71	0.120	0.9996	0.081	−0.777
3	1.72	0.064	0.9999	0.050	−0.449
4	2.54	0.079	0.9999	0.053	−0.503
5	2.21	0.073	0.9999	0.057	−0.541

As one expects from the results in the first trial, for 3 to 5 neurons in the hidden layer there is a dramatic improvement in the E_π predicted by the ANN. Although in this case the E_π of naphthalene is well reproduced, the largest deviation is exhibited by the two BH with three benzenoid rings : anthracene and phenathrene. Due to this situation, in a third trial the first three BH were presented more than one time : three times for naphthalene, and two times anthracene and phenanthrene, forming the training set TS3 with 269 patterns. A set of five ANN were trained for 5000 cycles and the corresponding results are presented in Table 4. The general statistics are not very different from those reported in Table 3, but the E_π for the first three BH is well predicted. Due to this situation, this kind of training set was used in the subsequent computations.

Table 4

Statistical results in estimating E_π for benzenoid hydrocarbons with the network NN1 trained for 5000 cycles with the training set TS3

No. of hidden neurons	MSE $\cdot 10^5$	s	r	$\Delta E_{\pi\ av}$	$\Delta E_{\pi\ max}$
1	121.28	0.541	0.9934	0.464	3.746
2	5.39	0.110	0.9997	0.082	0.699
3	1.47	0.060	0.9999	0.047	−0.315
4	2.58	0.075	0.9999	0.054	0.396
5	2.10	0.072	0.9999	0.059	−0.393

The results in Table 4 indicate that 3 nodes in the hidden layer is a reasonable number to use; using more units does not improve the result significantly.

A network (denoted by NN1), with 3 hidden neurons was trained for a period of 100000 cycles, when there was no further decrease in overall error. The MSE decreased to $0.411 \cdot 10^{-5}$ and the correlation between the HMO E_π and NN1 E_π is given by the equation:

$$E_{\pi\ HMO} = 0.030742 + 0.999026\ E_{\pi\ NN1} \qquad (2)$$

$$n = 265 \quad s = 0.0315 \quad r = 0.999978$$

which shows the excellent prediction obtained with the ANN.

For comparison, a similar correlation between HMO E_π and E_π computed by equation (1) gives the following equation:

$$E_{\pi\ HMO} = 0.106235 + 0.998055\ E_{\pi\ eq.1} \qquad (3)$$

$$n = 265 \quad s = 0.0458 \quad r = 0.999953$$

with statistical indices somewhat lower than in equation (2). The major difference between the two methods to estimate the E_π is that the first one uses a well defined function to fit the data, while the neural network performs a model-free mapping of the topological features of BH to predict the E_π.

Table 5 presents the topological invariants, the HMO E_π, the NN1 predicted E_π and the error for six BH depicted in Fig. 1. The results indicate that the network NN1 predicts the E_π of BH with very high precision, on the basis of very simple topological invariants: N, M and K.

Although it is generally considered that all three topological characteristics (N, M and K) are needed in order to predict the E_π of BH, we tested the situation in which only two such invariants were used as independent variables.

A set of three layer ANN were simulated, with two input neurons representing either the pair of variables (N, K) or (M, K), and one output unit representing E_π. The number of hidden neurons was set between 1 and 8, and the training was performed for 5000 cycles. The results of these simulations, reported in Table 6, show that the ANN with input

Table 5

Topological invariants, HMD E_π, NN1 E_π and the error in predicting E_π with the network NN1 for the six benzenoid hydrocarbons in Fig.1

Benzenoid hydrocarbon	N	M	K	$E_{\pi\ HMO}$	$E_{\pi\ NN1}$	ΔE_π^a
1	26	31	16	36.5169	36.5131	−0.0038
2	30	36	25	42.2536	42.2606	0.0070
3	30	36	30	42.3594	42.3595	0.0001
4	30	36	34	42.4370	42.4364	−0.0006
5	30	36	34	42.4367	42.4364	−0.0003
6	28	34	21	39.6901	39.6906	0.0005

$^a\Delta E_\pi = E_{\pi\ NN1} - E_{\pi\ HMO}$

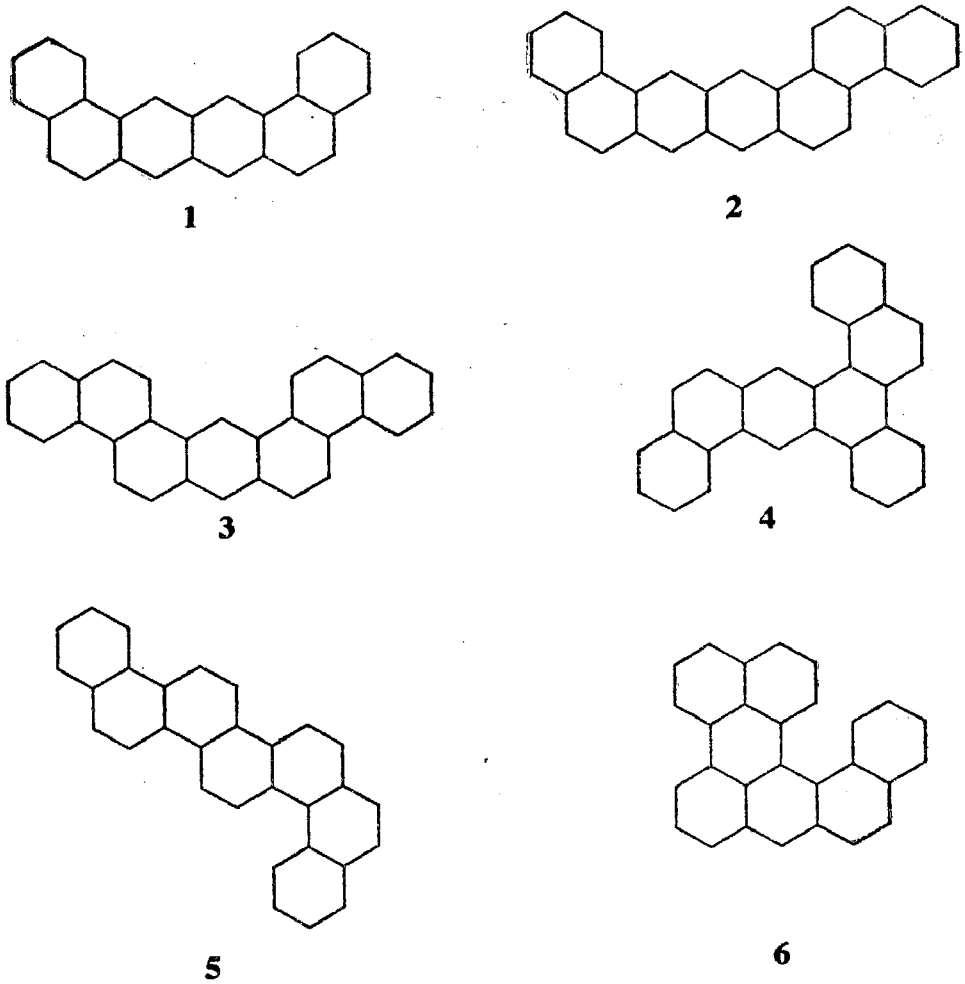

Fig. 1. — Benzenoid hydrocarbons whose topological features, $E_{\pi\ HMO}$ and $E_{\pi\ NN1}$ are presented in Table 5.

parameters (N, K) performs better than the ANN with independent variables (M, K). Also, from the same table one can conclude that when the number of hidden neurons is increased to 3 there is a dramatic improvement in the prediction of E_π, with little variation for higher number of hidden neurons.

Table 6

Statistical results in estimating E_π for benzenoid hydrocarbons with the network NN2 trained for 5000 cycles with the training set TS3

No. of hidden neurons	Input data			
	N, K		M, K	
	MSE $\cdot 10^4$	s	MSE $\cdot 10^4$	s
1	35.594	0.662	63.847	0.669
2	1.300	0.168	5.684	0.279
3	1.256	0.156	5.434	0.281
4	1.299	0.156	5.262	0.278
5	1.284	0.155	4.932	0.271
6	1.301	0.156	5.156	0.277
7	1.347	0.158	5.137	0.276
8	1.302	0.156	4.938	0.272

Following this finding, an ANN with 3 hidden neurons was chosen as the one that gave the best balance between fitting the training data and a small number of adjustable parameters. The two input neurons are one for the number of carbon atoms N, and the other one for the number of Kekulé structures K. This ANN, denoted NN2, was trained for 100000 cycles, when there was no further improvement in the prediction of E_π. The final MSE was $4.533 \cdot 10^{-5}$, and the correlation between the $E_{\pi\,\mathrm{HMO}}$ and $E_{\pi\,\mathrm{NN2}}$ is given by the equation :

$$E_{\pi\,\mathrm{HMO}} = 0.000876 + 0.999939\ E_{\pi\,\mathrm{NN2}} \qquad (4)$$

$$n = 265 \quad s = 0.1085 \quad r = 0.999735$$

This equation reflects the fact that only two topological invariants, N and K, give a very good estimation of E_π when used in a neural network model.

The standard deviation of equation (4) is greater than in equation (2). This fact is a consequence of the fact that NN2 has only 13 connections (adjustable parameters), while NN1 has 16 connections. Of course, the prediction of NN2 model could be improved by using a higher number of hidden neurons, but in this study we intended to maintain the size of the network as small as possible.

There are two advantages of adopting networks with a small number of hidden units. Firstly, the efficiency of each node increases and consequently the time of the computer simulation is significantly reduced. Secondly, and more importantly, the network can generalize the input patterns better, and this results in superior predictive power. On the other hand, an ANN with insufficient hidden units will not be able to extract all the rellevant correlations between topological descriptors of BH and E_π.

The L20 O (Leave—20%—Out of data) cross-validation was applied in order to test the prediction ability of the ANN. The best results for the prediction were obtained by the (N, M, K) network, with r=0.9984 and s=0.270, while for the (N, K) network we obtained

r=0.9980 and s=0.297, and for the (M, K) network r=0.9979 and s=0.302. In conclusion, all three networks have a high predictive ability for the E_π of benzenoid hydrocarbons, with a small advantage for the first one.

CONCLUSIONS

We have demonstrated that the approach of neural networks offers a simple novel tool potentially well suited for estimation of the total π-electron energy of benzenoid hydrocarbons. It was found that E_π is well approximated by only two topological features: the number of carbon atoms and the number of Kekulé structures. The results presented here indicate that neural networks can give valuable predictions. On the other hand, ANN approaches do not replace other forms of computing predictions, but they promise to be a useful alternative tool.

ACKNOWLEDGEMENT. We thank the Ministry of Research and Technology for financial support of this research under Grant 469.

REFERENCES

1 J. Cioslowski, *Top. Curr. Chem.*, **1990**, *153*, 85—99.
2 I. Gutman, *MATCH*, **1991**, *26*, 123—135.
3 J. Cioslowski, *Theor. Chim. Acta*, **1985**, *68*, 315—319.
4 J. Cioslowski, *Z. Naturforsch.*, **1985**, *40a*, 1167—1168.
5 J. Cioslowski, *Int. J. Quantum Chem.*, **1987**, *31*, 581—590.
6 J. Cioslowski and I. Gutman, *Z. Naturforsch.*, **1986**, *41a*, 861—865.
7 I. Gutman, *J. Chem. Soc. Faraday Trans.*, **1990**, *86*, 3373—3375.
8 I. Gutman and G. G. Hall, *Int. J. Quantum Chem.*, **1992**, *41*, 667—672.
9 P. D. Wasserman, "Neural Computing", Van Nostrand Reinhold, New York, 1989.
10 R. Hecht-Nielsen, "Neurocomputing", Addison-Wesley Publishing Company, Redwood City, CA, 1990.
11 P. K. Simson, "Artificial Neural Systems: Foundations, Paradigms, Applications and Implementations", Pergamon Press, New York, 1990.
12 J. Hertz, A. Krogh and R. G. Palmer, "Introduction to the Theory of Neural Computation", Addison-Wesley Publishing Company, Redwood City, CA, 1991.
13 M. E. Lacy, *Tetrahedron Comput. Methodol.*, **1990**, *3*, 119—128.
14 J. Zupan and J. Gasteiger, *Anal. Chim. Acta*, **1991**, *248*, 1—30.
15 D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Nature*, **1986**, *323*, 533—536.