ChemNet: A Novel Neural Network Based Method for Graph/Property Mapping

Dmitry B. Kireev†

Institute of Physiologically Active Compounds of Russian Academy of Sciences, 142432, Chernogolovka, Moscow Region, Russia, and Laboratoire de Chimiométrie, Université d'Orléans, 45067, B.P. 6759, B.P. 6759, Orleans Cedex 2, France

Received January 3, 1994⊗

ChemNet is a new method for mapping molecular properties. The input of ChemNet consists of two-dimensional matrices of variable sizes, instead of the sets of molecular descriptors used by the conventional mapping methods. The two-dimensional matrices considered in this study are graph distance matrices. The diagonal elements of the matrices are atomic properties. ChemNet uses these matrices to form the topology of the artificial neural network. Each molecule of a training set corresponds to a single network configuration. The weighted connections of the networks are adjusted, using the "backprop" procedure. The original background of the method and details of the current realization are presented. Examples of how ChemNet learns topological and physicochemical molecular properties demonstrate the practical use of the method.

INTRODUCTION

Artificial neural networks[1−4] (ANN) constitute a recently emerged data processing technology. Over the last decade chemistry became a field of their wide application.[5−8] The feedforward neural networks trained by "backprop" procedure (BPN)[9] have been most widely used by chemists. ANN first appeared as an attempt to understand and emulate the brain information processing, but they now have moved far away from their original destination. However ANN keep their principal original property, i.e., the parallel distributed processing.

Quantitative structure−activity relationships (QSAR) have recently become the field, in which ANN are being intensively used. QSAR have been based upon linear models from the very beginning. Yet two principal shortcomings of the linear models should be stressed. The first is their poor flexibility. A linear method will not reveal the relationship between an activity and input data if this relationship is nonlinear. The second shortcoming is the necessity to represent the input data by numerical vectors of a constant dimensionality, which leads to a loss of useful information and/or introduces noise. Representations of the chemical information, which involve molecular graphs, three-dimensional (3D) models, or electron density distributions, are more natural for the chemical thinking. All above presentations may be adequatly rewritten in numerical form by means of two-dimensional **molecular matrices**, such as connectivity matrix, distance matrix, bond order matrix, density matrix, etc.

Contrary to the linear methods ANN are rather flexible. However the necessity of matrix/vector transformation is inherent to the existing ANN. The input data for ANN should still be represented by vectors of a constant dimensionality. The use of molecular matrixes as an input of ANN should obviously solve the problem. In the recent reviews,[2−4] one can find an ANN method avoiding conventional molecular descriptors. The approach of Elrod et al.[11,12] concerns manipulations with a connectivity matrix in order to obtain the optimal vector representation. However this method can only be applied to a limited number of chemical problems because of the strict requirements to a training set. The molecules of a training set should contain a relatively large common substructure. In fact the vector representation of the input data still remains in this method. It only reaches the optimal matrix/vector transformation based on structural similarity of molecules within a training set.

ChemNet is a novel ANN based method for QSAR. The method aims at the activity mapping and uses the molecular matrices as input. The main feature of the method is the dynamic network topology, instead of the rigid topology of the conventional BPN. The input information not only occupies the input layer but also forms the network topology.

The goal of this paper is twofold: (i) to present the methodological basis of ChemNet and (ii) to demonstrate the efficiency of this method on some model training sets. Topological invariants and chromatographic data were used as target properties.

METHOD

Background. ChemNet substantially differs from the conventional BPN in that it can accept input information in the form of molecular matrices. BPN starts learning with predefined network topology. Each input layer node of BPN corresponds to a single molecular descriptor. ChemNet does not require to predefine the network topology. It creates its own unique network topology for each molecule. Each node in each layer corresponds to a single atom with the exception of the bias nodes. Thus, even during a single epoch, the network changes the number of nodes. Each node is connected to all the nodes of the neighboring layers. Since nodes correspond to atoms, connections correspond to relations betwen the corresponding atoms. If the relations within given pairs of atoms are the same, the corresponding weighted connections are identical. A totality of the identical weighted connections corresponds to a single adjustable parameter. In this work the relations used are the topological interatomic distances. Thus, the number of adjustable

† Present address: Laboratoire de Chimiométrie, Université d'Orléans, 45067, B.P. 6759, Orleans Cedex 2, France.

⊗ Abstract published in *Advance ACS Abstracts*, January 15, 1995.

0095-2338/95/1635-0175$09.00/0 © 1995 American Chemical Society

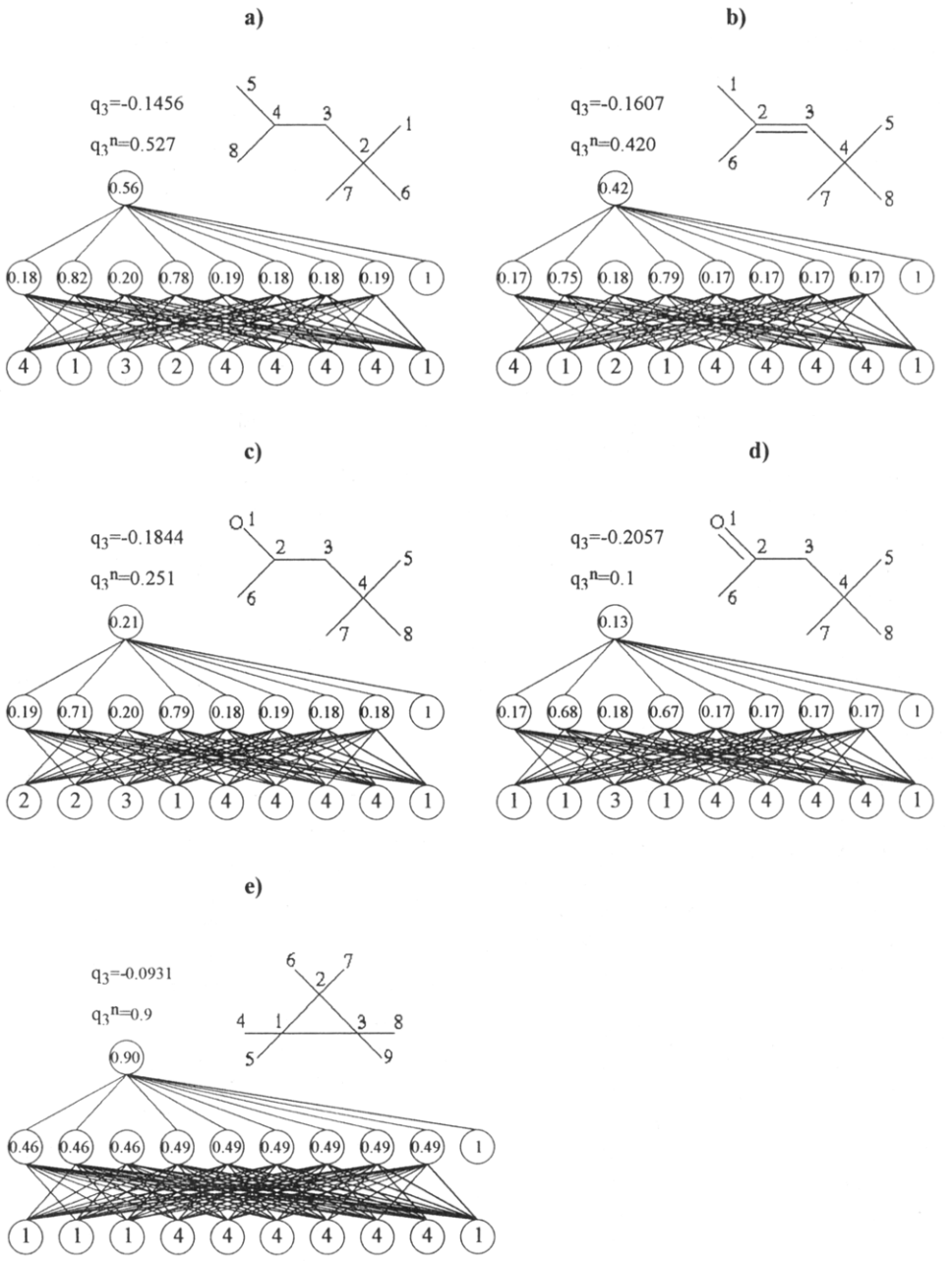

Figure 1. The compounds, constituting the model training set are shown. The normalized charges on the third node of each graph (q_n^3) serve as target properties. The values, shown inside the network nodes in the input layer, are the numbers of attached hydrogens, increased by one. The values, shown in the other layers were calculated by trained ChemNet. The correlation between actual q_n^3 and the calculated output values is quite significant ($r = 0.99$).

parameters between two neighboring layers is equal to the number of the equivalence classes for the interatomic distances observed within the corresponding molecules.

The Chemnet network output is invariant to the numbering of the atoms in a molecule. Indeed the topological distance between any pair of atoms does not depend on the ordinal numbers of the atoms. Each node, therefore, will always be connected to the same set of the neigboring layer nodes by the same set of weighted connections. Thus, the value of each output (hidden) layer node is the atomic invariant derived by ChemNet. Since the node in a molecular graph with its entire environment is unique (with the exception of the automorphic nodes in the same molecule), the atomic invariant represents a molecular invariant as well. ChemNet adjusts the weights of connections in order to fit these molecular invariants to target properties.

A model training set of five molecules is used below to illustrate the principles of ChemNet. Figure 1a–e shows the molecules with corresponding networks. The output values were fitted to the atomic charges on a third atom of each molecule. In terms of graph theory these nodes are the graph centroids. The charges were calculated by the AM1[15] quantum chemical method. They were normalized using the formula

$$q_c^n = 0.1 + 0.8\,\frac{q_c - q_c^{\min}}{q_c^{\max} - q_c^{\min}}$$

where q_c is the AM1 atomic charge, $q_c^{\min}$ and $q_c^{\max}$ are the minimal and maximal charge values among the molecules. Each network consists of three layers. The number of nodes

in each layer is equal to the number of atoms in a corresponding molecule plus the bias node.

The distances from 1 to 4 are present in these molecules. Four adjustable parameters between input and hidden layers correspond to them. One more adjustable parameter corresponds to the connections between the bias node of the input layer and the nodes of the hidden layer. A total of five adjustable parameters appear between the input and hidden layers. Compound **5**, however, does not contain the distance of 4, and it does not influence the evolution of this parameter. Each output layer contains only one node. This only node corresponds to the third atom of each molecule. The distances of 1 and 2 connect the third atoms to others. These distances correspond to two adjustable parameters. And one more parameter corresponds to the connection of the hidden layer bias node to the output node. A total of three adjustable parameters appears between the hidden and output layers. Thus, the entire network counts eight adjustable parameters. In BPN a single connection corresponds to a single adjustable parameter. In the ChemNet network the number of adjustable parameters depends, in a complicated manner, on the constituent molecules of the training set.

The atomic properties used as input are the numbers of attached hydrogens (NH) increased by one. The values were increased to avoid zero values, disabling the tertiary atoms. The NH values indirectly contain the information about the hybridization type and the element number. They should, therefore, be useful for learning rather simple properties such as atomic charges, chemical shifts, etc.

The network corresponding to the molecular graph of 2,2,4-trimethyl-pentane **1** is shown in Figure 2 in details. Figure 2a contains the entire network. Figure 2b–e shows only the identical weighted connections. For example, Figure 1b shows the connections, corresponding to the interatomic distance of 1.

The state v_i^k of each ith node in the kth layer is calculated as follows

$$v_i^k = \frac{1}{1 + e^{-\varphi_i^k}}; \qquad \varphi_i^k = \sum_j w_{ij,n}^{kk-1} v_j^{k-1} - \theta^{k-1} \tag{1}$$

where $w_{ij,n}^{kk-1}$ is the weight of the connection between the ith node in the kth layer and the jth node in the $k-1$ layer in the nth network. One has to keep in mind that some of the w_{ij}^{kk-1} in the exponent are identical and the w_{ij}^{kk-1} are generally different in the different networks of the training set. An error function is

$$E = \sum_n E_n = \frac{1}{2} \sum_n (A_n - v_{o,n}^O)^2 \tag{2}$$

where E_n is a contribution of the nth molecule to the error function, $v_{o,n}^O$ estimates the target property A_n. Thus, the delta rule has the following form:

$$w_{ij}^{kk-1,e+1} = w_{ij}^{kk-1,e} - \gamma \frac{\partial E}{\partial w_{ij}^{kk-1}} \tag{3}$$

where $w_{ij}^{kk-1,e+1}$ and $w_{ij}^{kk-1,e}$ are the weights of the connections between the ith node in the kth layer and the jth node

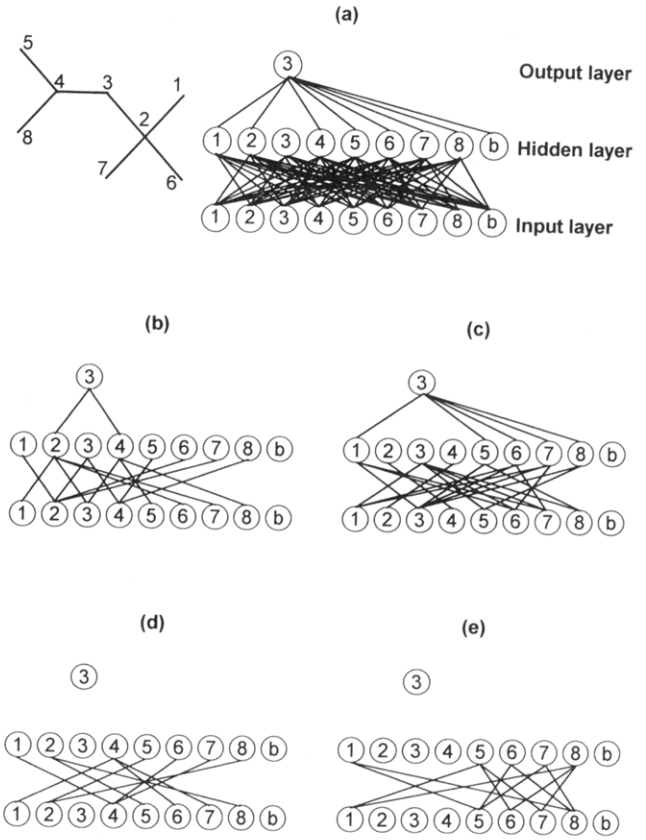

Figure 2. ChemNet, corresponding to the molecular graph of 2,2,4-trimethylpentane **1** (a) and the fragments of the network, composed of the connections, corresponding to the following topological interatomic distances: $D_{ij} = 1$ (b), $D_{ij} = 2$ (c), $D_{ij} = 3$ (d), and $D_{ij} = 4$ (e).

in the $(k+1)$th layer during the eth and the $(e+1)$th epochs of the learning, γ is the learning rate, and

$$\frac{\partial E}{\partial w_{ij}^{kk-1}} = \sum_n \frac{\partial E_n}{\partial w_{ij}^{kk-1}} \tag{4}$$

where $\partial E_n / \partial w_{ij}^{kk-1}$ should be defined recursively depending on k. In the case of a three layer network the gradient for a connection between output and hidden layers (w^{OH}) is

$$\frac{\partial E_n}{\partial w_{ij}^{OH}} = \frac{\partial E_n}{\partial v_{o,n}^O} \frac{\partial v_{o,n}^O}{\partial w_{oi}^{OH}} = (A_n - v_{o,n}^O) v_{o,n}^O (1 - v_{o,n}^O) \tag{5a}$$

where $v_{o,n}^O$ is the value of the output node in the output layer for the nth molecule, A_n is the target property value of the target property value of the nth molecule, w_{oi}^{OH} is the weight of the connection between the ith node in the hidden layer and the oth (output) node in the output layer. And, for the connection between hidden and input layers (w^{HI}), the gradient is

$$\frac{\partial E_n}{\partial w_{ij}^{HI}} = \frac{\partial E_n}{\partial v_{o,n}^O} \frac{\partial v_{o,n}^O}{\partial w_{ij}^{HI}} = (A_n - v_{o,n}^O) v_{o,n}^O (1 - v_{o,n}^O) w_{oi}^{OH} v_{i,n}^H (1 - v_{i,n}^H) v_{j,n}^I \tag{5b}$$

where w_{ij}^{HI} is the weight of the connection between the jth node in the hidden layer and the ith node in the output layer. These derivatives can be calculated separately for each

Table 1. Set of 27 Alkanes and Their Actual, Estimated, and Predicted Values of teh Wiener Index (W), the Connectivity Index (χ^2), and the Index of Merrifield–Simmons (MS)[a]

no.	compound	W	W_{est}	W_{pred}	χ^2	χ^2	χ^2	MS	MS_{est}	MS_{pred}
1	*n*-hexane	35	38.2	38.6	1.707	2.10	1.69	3.045	3.25	3.22
2	2-methylpentane	32	38.4	38.6	2.183	2.43	2.04	3.136	3.31	3.32
3	3-methylpentane	31	34.5	35.1	1.922	2.44	2.12	3.091	3.27	3.27
4	2,3-dimethylbutane	29	34.5	35.0	2.488	2.44	2.11	3.178	3.27	3.27
5	2,2-dimethylbutane	28	30.8	31.1	2.914	2.45	2.16	3.258	3.24	3.23
6	*n*-heptane	56	60.6	61.7	2.061	2.30	2.01	3.526	3.65	3.66
7	2-methylhexane	52	55.0	55.0	2.536	2.65	2.47	3.611	3.71	3.73
8	3-methylhexane	50	50.0	51.2	2.302	2.66	2.57	3.584	3.67	3.68
9	2,3-dimethylpentane	46	47.3	47.4	2.630	3.04	2.98	3.714	3.78	3.80
10	2,4-dimethylpentane	48	51.3	51.5	3.023	3.03	2.84	3.638	3.74	3.75
11	2,2-dimethylpentane	46	51.3	51.4	3.311	3.03	2.74	3.761	3.78	3.75
12	3,3-dimethylpentane	44	43.0	42.8	2.871	3.05	3.26	3.679	3.70	3.71
13	2,2,3-trimethylpentane	42	43.0	43.6	3.521	3.05	3.11	3.784	3.70	3.72
14	*n*-octane	84	79.0	76.1	2.414	2.37	2.5	4.007	3.94	3.92
15	2-methylheptane	79	81.0	80.6	2.890	2.52	2.17	4.094	4.03	4.00
16	3-methylheptane	76	76.2	76.9	2.656	2.89	2.85	4.060	4.07	4.11
17	4-methylheptane	75	73.3	72.8	2.683	2.89	3.14	4.078	4.04	4.05
18	2,3-dimethylhexane	70	66.8	65.9	3.010	3.27	3.47	4.127	4.09	4.10
19	2,4-dimethylhexane	71	70.0	68.7	3.143	3.26	3.36	4.159	4.12	4.13
20	2,5-dimethylhexane	74	76.2	75.9	3.365	2.89	2.70	4.174	4.07	4.07
21	2,3,4-trimethylpentane	65	60.9	61.1	3.347	3.60	3.71	4.190	4.14	4.14
22	2-methyl-3-ethylpentane	67	60.9	60.8	2.821	3.60	3.90	4.111	4.14	4.16
23	2,2-dimethylhexane	71	70.0	69.4	3.664	3.26	3.35	4.234	4.12	4.12
24	3,3-dimethylhexane	67	62.9	62.5	3.268	3.27	3.59	4.190	4.06	4.06
25	2,2,3-trimethylpentane	63	60.9	61.1	3.675	3.60	3.71	4.344	4.16	4.18
26	2,2,4-trimethylpentane	66	63.8	64.0	4.159	3.60	3.78	4.249	4.14	4.15
27	2,2,3,3-tetramethylbutane	58	57.4	57.8	4.500	3.61	3.89	4.382	4.11	4.12

[a] $Index_{est}$ means that the value was estimated by the model, trained on the whole data set. $Index_{pred}$ means that the value was predicted for a compound by the model, obtained with this compound, exluded from the training set.

connection, but, as some of these connections are identical, one more summation (over identical connections) should be carried out for each $\partial E_n/\partial w_{ij}^{kk-1}$.

The prediction of a target property for a new molecule is made in two stages. First ChemNet constructs the network for the molecule. Then values of the hidden and output nodes should be calculated by eq 1, substituting $w_{ij,n}^{kk-1}$ by the corresponding weights from the relevant model. The new molecule may contain longer distances than those existing within the molecules of the training set. This will mean that the model does not account for these distant interatomic perturbations. However, the influence of distant atoms does not substantially contribute to the properties of an atom in a molecule.

Software Details. The current version of ChemNet runs under Windows 3.1. The input to the program is the set of the connectivity matrixes created by the built-in molecular editor ChemPro.[16] The diagonal elements are optional and may be set to (i) 1, (ii) the numbers of hydrogens attached, or (iii) atomic electronegativities by Pauling.[17] When starting a session the program constructs the set of distance matrixes and then the set of the networks. All the data have dynamic structures, which allows treatment of large amounts of information. The output is the database of models (the model means the set of fitted parameters). The program has an easy-to-use interface for storing and retrieving the models.

RESULTS AND DISCUSSION

It was noted above that ChemNet currently uses the distance matrices of molecular graphs as input. Thus, one can expect the method to be able to model the properties which depend on molecular topology. As biological activity depends on many factors, it is not a good test for a new method. To be sure that the test is pure enough, three molecular graph invariants[18] were chosen as target properties: the Wiener index (W),[19,20] the second order connectivity index (χ^2),[21] and the index of Merrifield and Simmons (MS).[22]

	W	χ^2	MS
W	1.00		
χ^2	0.356	1.00	
MS	0.854	0.754	1.00

Figure 3. Correlation table for the Wiener index (W), connectivity index (χ^2), and the index of Merrifield–Simmons (MS).

Table 2. Statistics and Cross-Validation Criterions (r, r^2, cv-r^2; and SD) and Number of Adjustable parameters (NP) for Mapping of Topological Properties (W, χ^2, and MS) for the 27 Alkanes

target property	r	r^2	cv-r^2	SD	NP
W	0.986	0.972	0.964	2.408	13
χ^2	0.815	0.664	0.598	0.288	13
MS	0.987	0.974	0.960	0.062	13

The series containing all acyclic linear and branched C_6–C_8 alkanes (Table 1) was examined. Predictive ability was evaluated using the cross-validation procedure[23] with five validation groups. The quality of models was estimated using (i) the conventional correlation coefficient (r) between the actual and calculated values, (ii) the cross-validated correlation coefficient (cv-r) between the actual values and the values, which were calculated for the compounds excluded from the training set, and (iii) standard deviation of the error (SD). All the criteria are collected in Table 2 The results for the W and MS indexes are quite good (cv-r^2 = 0.964 and cv-r^2 = 0.960, respectively), while the model of the connectivity index (χ^2) is rather poor, compared to

the others. The model might probably be improved by the choice of an appropriate activation function (studies with various functions are in progress).

After the above examination we subjected the method to a more serious test. The chromatographic behavior was studied for a wide variety of hydrocarbons, including alkanes, cycloalkanes, *n*-alkenes, branched alkenes, cycloalkenes, and aromatic hydrocarbons (Table 3). The experimental Kovats indices obtained in stationary phases which consist of unsulfonated Porapak Q (UnS)[24] were used for the analysis. In ref 24 these data were followed by the boiling points which were modeled by us as well. The chromatographic behavior of separate sets of alkanes and alkenes was previously studied by means of the DARC topological system.[25] The application of a neural network to retention data in gas chromatography was also presented recently[26] with a homogeneous set of substituted phenols. However, this study is the first that involves learning on such a diverse training set.

The numbers of hydrogens attached were used as input. ChemNet has successfully learned the UnS retention data as well as boiling points showing high statistic and cross-validation criteria (Table 4). One should note that both predicted and estimated values have been calculated with the same quality. The significant robustness of the model can be explained both by the small number of parameters and by the chemical sensitivity of the discussed model. For a set of DHFR inhibitors[10] it was found that in order to have a robust model, i.e., to avoid "memorizing", BPN needs to have a ratio of objects to connections (ϱ) larger than 1.8. However, a good model usually requires ϱ to be smaller than 2.2. In this study ϱ was equal to 2.1 for the first training set (Table 1) and 2.8 for the second one (Table 3). Both r and cv-r asymptotically grew in the course of learning. Such a behavior of r and cv-r shows that the number of parameters can be increased without loss of robustness. Introducing additional parameters is a chemical problem concerning the concepts of compound description (various approaches to that will be presented in a coming study). Finally it has been studied how the ChemNet behavior depends on the number of hidden layers. It was established from numerous learnings that changing this number from 1 to 2 resulted in rising of the correlation coefficient by approximately 0.01. Setting the number of hidden layers to three and more did not improve the models compared to those obtained with two hidden layers.

CONCLUSIONS

ChemNet is a novel neural network method. It was developed for the processing of chemical information. ChemNet uses the backprop optimization procedure, but it principally differs, however, from the BPN method proposed originally by Rumelhart and McClelland.[9] The main advantage of ChemNet (in chemical applications), compared to Rumelhart's method, is its ability to process training sets containing objects of variable sizes. This is important for chemistry because it allows the use of nontransformed molecular matrices (e.g., distance matrices, bond order matrices, density matrices, etc) as input.

The modeling of topological and physicochemical properties has shown high statistic and predictive characteristics. The compounds involved in this study are rather varied. The training sets include cyclic and acyclic alkanes, alkenes, and aromatic hydrocarbons.

One has to notice, however, that the properties studied are relatively simple. To make the new method highly practical, one should try it with more complex target properties, such as chemical reactivity or biological activity. ChemNet is rather promising in this respect as it allows manipulation of chemical information of any level of complexity without its reduction.

Table 3. Actual, Estimated, and Predicted Kovats Retention Indices in Gas-Solid Chromatography on Unsulfonated Porapak Q (UnS) and the Boiling Points (BP) of the 38 Alkanes, Alkenes, and Aromatic Hydrocarbons

no.	compound	UnS	UnS_{est}	UnS_{pred}	BP	BP_{est}	BP_{pred}
1	1-pentene	494	524	527	30.0	44.4	47.2
2	2-pentene	499	524	523	36.4	44.4	45.2
3	2-methyl-2-butene	504	521	518	38.6	43.2	42.0
4	1,3-pentadiene	519	524	523	42.0	44.4	44.1
5	cyclopentene	515	524	529	44.2	44.4	44.1
6	cyclopentane	526	524	527	49.3	44.4	47.2
7	3,3-dimethyl-1-butene	550	583	591	41.2	59.5	66.1
8	4-methyl-1-pentene	577	590	587	53.9	62.3	62.1
9	3-methyl-2-pentene	605	586	584	70.4	60.9	60.4
10	2-methyl-2-pentene	597	590	593	67.3	62.3	62.4
11	1-hexene	596	628	631	63.5	78.9	80.0
12	2-hexene	609	628	627	67.9	78.9	78.9
13	methylcyclopentane	611	586	584	79.8	60.9	59.8
14	benzene	622	628	627	80.1	78.9	79.0
15	cyclohexane	631	628	631	80.7	78.9	79.0
16	cyclohexene	637	628	631	83.0	78.9	80.0
17	2,3,3-trimethyl-1-butene	664	661	670	77.9	79.9	85.5
18	2,4-dimethylpentane	664	669	669	80.5	82.9	83.2
19	2,2,3-trimethylbutane	672	661	659	80.9	79.9	79.1
20	1-heptene	694	746	744	93.6	115.0	115.0
21	toluene	726	705	706	110.6	98.2	97.5
22	cycloheptane	755	746	745	118.8	115.0	114.0
23	methylencyclohexane	718	705	704	102.5	98.2	96.7
24	2,2,4-trimethylpentane	742	747	749	99.8	103.0	104.0
25	1,4-dimethylcyclohexane	801	808	802	121.0	129.0	130.0
26	1-octene	798	838	839	121.3	134.0	135.0
27	ethylbenzene	820	777	779	136.2	117.0	117.0
28	*p*-xylene	834	808	808	128.4	139.0	129.0
29	*m*-xylene	836	831	831	139.0	138.0	139.0
30	*o*-xylene	852	838	829	144.4	134.0	134.0
31	cyclooctane	885	838	829	151.1	134.0	134.0
32	cycloheptene	746	746	748	–	–	–
33	1,3-cyclohexadiene	631	628	627	115.0	115.0	115.0
34	1,4-cyclohexadiene	647	628	625	80.5	78.9	78.9
35	cyclooctene	871	838	837	86.9	78.9	77.8
36	1,2-dimethylcyclohexane	799	838	829	123.4	134.0	134.0
37	2-methyl-1-pentene	592	590	593	62.1	62.3	62.4

Table 4. Statistics Criterions (r, r^2, cv-r^2, and SD) and Number of Adjustable Parameters (ND) for the Mapping of Physicochemical Properties (UnS and BP) for the 38 Alkanes, Alkenes, and Aromatic Hydrocarbons

target property	r	r^2	cv-r^2	SD	NP
UnS	0.981	0.962	0.962	18.96	13
BP	0.957	0.916	0.910	8.23	13

ACKNOWLEDGMENT

I wish to thank Mr. Igor Lisyanskii for the molecular editor ChemPro which was presented to me for permanent use within the ChemNet program. I have highly appreciated the participation of Prof. J. R. Chretien in the testing of the method by chromatographic data and the discussing of the manuscript. I have also very much valued help of Prof. O. A. Raevsky. I should like to acknowledge contribution of Prof. J. Gasteiger to the presentation of the method. My thanks go to Prof. N. S. Zefirov and Dr. I. I. Baskin for the

discussion of the method's background and to the referees who have helped much to improve the manuscript.

REFERENCES AND NOTES

(1) Simpson, P. K. *Artificial Neural Systems: Foundations, Paradigmes, Applications and Implementations*; Pergamon Press: New York, 1990.
(2) Zupan, J.; Gasteiger, J. *Neural networks for Chemists—An Introduction*; VCH Publishers: 1993.
(3) Gasteiger, J.; Zupan, J. Neural Networks in Chemistry. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 503−527.
(4) Maggiora, G. M.; Elrod, D. W.; Trenary, R. G. Computational Neural Nets as Model Free Mapping Devices. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 732−741.
(5) Rose, V. S.; Croall, I. F.; MacFie, H. J. An Application of Unsupervised Neural Network Methodology (Kohonen-Topology Preserving Mapping) to QSAR Analysis. *Quant. Struct.-Act. Relat.* **1991**, *10*, 6−15.
(6) Chastrette, M.; de Saint Laumer, J. Y. Structure-odor Relationships Using Neural Networks. *Eur. J. Med. Chem.* **1991**, *26*, 829−833.
(7) Weinstein, J. N.; Kohn, K. W.; Grever, M. R.; Viswanadhan, V. K.; Rubinstein, L. V.; Monks,; A. P.; Scudiero, D. A.; Welch, L.; Koutsokos, A. D.; Chiausa, A. J.; Paull, K. D. Neural Computing in Cancer Drug Development: Predicting Mechanisms of Action. *Science* **1992**, *258*, 447−451.
(8) Tetko, I. V.; Luik, A. I.; Poda, G. I. Applications of Neural Networks in Structure−Activity Relationships of Small Number of Molecules. *J. Med. Chem.* **1993**, *36*, 811−814.
(9) Rumelhart, D. E.; McClelland, J. L. *Parallel Distributed Processing*; MIT Press: Cambridge, MA, 1986.
(10) Andrea, T. A., Kalayeh, H. Applications of Neural Networks in Quantitative Structure−Activity Relationships of Dihydrofolate Reductase Inhibitors. *J. Med. Chem.* **1991**, *34*, 2824−2836.
(11) Elrod, D. W.; Maggiora, G. M.; Trenary, R. G. Applications of Neural Networks in Chemistry 1. Prediction of Electrophilic Aromatic Substitution Reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 477−484.
(12) Elrod, D. W.; Maggiora, G. M.; Trenary, R. G. Applications of Neural Networks in Chemistry 2. A General Connectivity Representation for the Prediction in Regiochemistry. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 477−484.
(13) Kvasnicka, V. An Applications of Neural Networks in Chemistry. Prediction of ^{13}C NMR Chemical Shifts. *J. Math. Chem.* **1991**, *6*, 63−76.
(14) Kvasnicka, V.; Pospichal, J. Applications of Neural Networks in Chemistry. Prediction of Product Distribution of Nitration in Series of Monosubstituted Benzenes. *J. Mol. Struct. (THEOCHEM)* **1991**, *235*, 227−242.
(15) Dewar, M. J. S.; Zoebich, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(16) The molecular editor ChemPro was developed by Dr. I. Lisyanski, Institute of Physical Activatiag Computing, Chernogolovka, Russia.
(17) Pauling, L. *Chimie Générale*; Dunod: Paris, 1960; p 231.
(18) Rouvray, D. H. Should We Have Designs on Topological Indexes? *Stud. Phys. Theor. Chem.* **1983**, *28* (*Chem. Appl. Topol. Graph Theory*), 159−177.
(19) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.
(20) All topological indexes were calculated by program EMMA elaborated in Moscow State University (runs under MS DOS).
(21) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
(22) Merrifield, R. E.; Simmons, H. E. Finite Point-set topology and molecular structure *Stud. Phys. Theor. Chem.* **1983**, *28* (Chem. Appl. Topol. Graph. Theory), 1−16.
(23) Cramer, R. D.; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18−25.
(24) Hirsh, R. F.; Gaydosh, R. J.; Chretien, J. R. Factor Analysis of Trends in Selectivity in Gas-Solid Chromatography on Cation Exchange Resins. *Anal. Chem.* **1980**, *52*, 723−728.
(25) Dubois, J. E.; Chretien, J. R. Data Analysis Methodology: the DARC Topological System. *J. Chromatogr. Sci.* **1974**, *12*, 811−821.
(26) Peterson, K. L. Counter-Propagation Neural Networks in Modeling and Prediction of Kovats Indexes for Substituted Phenols. *Anal. Chem.* **1992**, *64*, 379−386.

CI9403200