NEURAL NETWORKS APPLIED TO PHARMACEUTICAL PROBLEMS. I. METHOD AND APPLICATION TO DECISION MAKING

Tomoo AOYAMA,[a] Yuji SUZUKI,[b] and Hiroshi ICHIKAWA*,[b]

Hitachi Computer Engineering Co., Ltd.,[a] Horiyamashita, Hadano Kanagawa 259-13, Japan and
Hoshi College of Pharmacy,[b] Shinagawa, Tokyo 142, Japan

Neural networks, which are also called perceptrons or multi-layer networks, were found to be useful tools in decision making. The model study showed that the predictions by the neural network were better than those by the linear learning machine and cluster analysis.

KEYWORDS neural network; perceptron; parallel-distributed processing unit; decision making; pharmaceutical problem; diagnosis

Recently data-processing methods called parallel-distributed processing (PDP) have been applied in various fields.[1] The characteristics of PDP have been found to be suitable for the data processing in which the relationship between the cause and its results cannot be exactly defined. One of the typical PDP's is the neural network which is a computer-based system derived from a simplified concept of the brain in which a number of nodes, called processing elements or neurons, are interconnected in a netlike structure. The characteristics of PDP's operation strongly suggest their applications to the data processing in biology-related responses and reactions such as diagnosis based on clinical data and creation of new drugs based on the structure-activity relationship. Our first report deals here with the general theory of the neural network used in our series of studies and an example in decision making as a model study for the application to clinical diagnosis. This may be the first example of application of the neural network to pharmaceutical problems and the results show that the neural network is promising in this and related fields.

Shown in Fig. 1 is the neural network: the circles are neurons which are actually variables taking values ranging from 0 to 1. The weights interconnecting neurons can take either positive or negative values. The number of the layers is arbitrary and generally consists of $\underline{n}$ layers. The data are input to A and are output from B.

The value of a neuron (O_j) at the $\underline{n}$th layer can be expressed by Eq. 1,

$$O_j = 1/[1 + \exp(-\alpha y_j)] \equiv f(y_j), \qquad y_j = (\Sigma W_{ij}x_i) - \theta_j \tag{1}$$

where x_i is one of the values of the neurons at the $\underline{n}$-1 layer; W_{ij}, an element of the weight matrix, expresses the weight value between neurons, i and j; θ_j is a characteristic value for neuron j; α is a parameter which expresses the non-linearity of the neuron. When the values of the neurons of each layer are renewed by feeding data into A, the all values expressed by Eq. 1 are synchronously changed.

Given N neurons at the first layer. A set of the input data can be expressed by a vector with N elements for N neurons which is, here, called an "input pattern." Likewise, the output data can also be regarded as a vector and be called an "output pattern." The vector which is compared with an output pattern to obtain the fixed W_{ij} is called a "training pattern" (t_j). The training is carried out according to the following equations.

$$W_{ij} = -d_j x_i \varepsilon \quad (2), \qquad d_j = (O_j - t_j)f'(y_j) \quad (3a), \qquad d_j = (W'_{j1}d'_1)f'(y_j) \quad (3b)$$

Here, ε is a parameter which determines the shift for correction in recursive cycles. Equation 3a is used only for the correction of the last (output) layer and 3b for other layers where W'_{j1} and d'_1 at the $\underline{n}$th layer are W_{ij} and d_j at the $\underline{n}$+1 layer, respectively. The function f' in Eq. 3 is,

$$f'(y_j) = f(y_j)[1 - f(y_i)]\alpha \quad (4)$$

where both ε and α can be set to be independent of the layer. The recursive iteration is carried out until E,

$$E = \Sigma(O_j - t_j)^2 \quad (5)$$

becomes small enough (typically less than 0.01). Even when that M sets of the input and training patterns are given, all of the output patterns can possibly be made close enough to the training patterns by the iteration through Eqs. 1 and 2. Then the neural network can classify the input patterns into M groups.

As a model study, we adopted the well-studied data and compared the ability of the neural network with coventional methods.[2] We used a three-layer network. Its structure and parameters are shown in Table I. Shown in Table II are the endo/exo conformations and the relative ^{13}C-NMR chemical shifts in the derivatives of norbornene.[3] The compound numbers are the same as those in the literature. In accordance with the former studies,[2] we used 25 (No. 1 to 25) out of 38 data as training data. First, in order to obtain W_{ij}, the training patterns were fed recursively until each W_{ij} settled down. Then, the network was used to answer new problems, i. e., the remaining 13 samples were fed to the network to see what answer the network gave. Each neuron was set to have a value ranging from ca. 0 to 1; the data of ^{13}C-NMR chemical shifts were, therefore, rescaled to the region between ca. 0.1 and 1.0 by the following equation,

$$\tilde{x}_i = (x_i - x_{min} + 0.1)/(x_{max} - x_{min} + 0.1) \quad (6)$$

where x_{min} and x_{max} are the minimun and maximum data. The reason not to give 0 as the least value is to avoid a situation where x_i being 0, W_{ij} may remain unchanged. The endo and exo conformations were expressed by the vectors (0,1,1,0) and (1,0,0,1), respectively, where instead of 2-element vectors, we used 4-element vectors simply to check whether or not the network operates appropriately. Although the neurons at the first layer may take continuous values between 0 and 1, those of the last layer are required to take discrete values 0 or 1. For this requirement, α was forced to change in the second and third layers. The parameter θ was set to be zero as is typically so and ε was so determined as to give a good convergence in the iteration.

The upper part of Table II shows input data (unscaled) and the output patterns for 25 compounds which fixed W_{ij}. The number of iterations was 60. The final output patterns do not show completely discrete values, which shows the limit of the resolution ability of the network. The lower part of the table shows the answers (endo or exo) obtained by the network for the untrained 13 compounds. The elements of the output patterns for untrained samples deviate considerably from 0 or 1 because the weight matrix was not determined for them, and it may be seen that the more the input data depart from the training patterns, the more the output data deviate from the

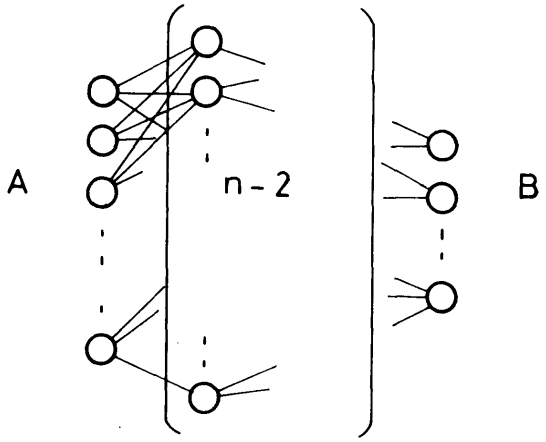

Fig. 1. n-Layer Neural Networks

Table I. Structure and Parameters of the Network

	Neurons	α	ε	θ
Layer 1	7			
Layer 2	14	1.5	0.2	0.0
Layer 3	4	3.0	0.2	0.0

Table II. Relative ^{13}C-NMR Chemical Shifts and Output Patterns by the Neural Network

No.	endo/exo	C_1	C_2	C_3	C_4	C_5	C_6	C_7	Output pattern				Decision
1	exo	6.7	6.7	10.1	0.5	0.2	-1.1	-3.7	0.97	0.02	0.02	0.97	
2	exo	8.9	25.3	12.4	-0.4	-1.2	-3.1	-4.4	1.00	0.00	0.00	1.00	
3	exo	7.7	44.3	12.3	-1.0	-1.3	-5.2	-4.4	1.00	0.00	0.00	1.00	
4	exo	4.6	16.7	4.4	-0.2	-0.3	-1.0	-1.8	0.94	0.06	0.06	0.94	
5	exo	1.8	15.1	4.4	-0.2	0.2	-0.7	-3.3	0.99	0.01	0.01	0.94	
6	exo	5.7	3.0	2.6	-0.5	-0.4	0.7	-3.5	0.99	0.01	0.01	0.94	
7	exo	6.1	5.9	10.6	0.6	0.2	0.2	-3.7	0.99	0.01	0.01	0.99	
8	exo	6.5	6.3	10.4	0.3	-0.8	-0.1	-3.5	0.99	0.01	0.01	0.99	
9	exo	6.5	7.5	9.5	0.5	1.7	0.7	-3.8	0.97	0.03	0.03	0.97	
10	exo	7.8	47.0	11.7	-1.3	3.9	-2.7	-3.2	0.99	0.01	0.01	0.99	
11	exo	6.9	6.4	10.1	0.7	-1.2	0.1	-3.9	1.00	0.00	0.00	1.00	
12	exo	5.6	4.9	7.0	0.2	-1.1	0.2	-3.9	1.00	0.00	0.00	1.00	
13	exo	2.5	42.5	11.9	-0.8	-1.1	-2.4	1.4	0.95	0.05	0.05	0.95	
14	endo	5.4	4.5	10.6	1.4	0.5	-7.7	0.2	0.00	1.00	1.00	0.00	
15	endo	6.8	23.3	10.5	1.2	0.6	-9.5	0.3	0.00	1.00	1.00	0.00	
16	endo	6.3	42.4	9.5	0.9	0.2	-9.7	-0.9	0.03	0.96	0.96	0.03	
17	endo	4.2	16.2	2.1	0.9	-0.6	-4.8	1.9	0.00	1.00	1.00	0.00	
18	endo	1.7	12.8	4.0	0.4	0.2	-7.2	1.4	0.00	1.00	1.00	0.00	
19	endo	4.7	3.1	2.2	0.3	1.3	-6.5	-0.6	0.00	1.00	1.00	0.00	
20	endo	4.7	5.3	9.2	1.3	-0.4	-6.5	1.4	0.00	1.00	1.00	0.00	
21	endo	4.6	11.5	8.9	-0.1	0.8	0.4	1.8	0.08	0.93	0.93	0.08	
22	endo	5.6	7.5	8.7	1.4	1.7	-3.0	1.7	0.00	1.00	1.00	0.00	
23	endo	7.1	47.8	13.3	2.2	3.6	-3.4	0.3	0.01	0.99	0.99	0.01	
24	endo	4.1	4.2	7.0	0.7	0.5	-7.4	0.0	0.00	1.00	1.00	0.00	
25	endo	3.2	40.2	10.4	-0.5	0.0	-10.3	3.1	0.00	1.00	1.00	0.00	
26	exo	5.5	1.0	6.3	-0.3	-1.5	-1.6	-1.3	0.83	0.18	0.18	0.83	exo
27	endo	3.4	0.1	5.5	0.2	-0.7	-4.9	0.0	0.02	0.98	0.98	0.02	endo
28	exo	5.1	16.4	4.2	-0.4	-1.1	-1.4	-2.1	0.98	0.02	0.02	0.98	exo
29	endo	4.0	15.9	2.2	0.7	-0.7	-5.0	1.7	0.00	1.00	1.00	0.00	endo
30	exo	6.6	7.0	10.1	0.2	-1.2	0.5	-3.7	1.00	0.00	0.00	1.00	exo
31	endo	6.0	8.4	11.2	-0.1	0.7	-1.5	-1.6	0.74	0.24	0.24	0.74	exo
32	exo	6.3	7.2	9.8	0.7	-0.1	0.8	-3.5	0.99	0.01	0.01	0.99	exo
33	endo	5.1	4.8	8.4	1.1	-0.1	-7.3	1.6	0.00	1.00	1.00	0.00	endo
34	exo	1.9	17.1	5.2	-0.1	0.9	0.9	-3.4	0.99	0.01	0.01	0.99	exo
35	endo	2.3	18.3	5.0	0.3	1.3	-2.9	1.4	0.01	0.99	0.99	0.01	endo
36	endo	5.1	4.0	8.4	1.1	0.2	-7.7	1.6	0.00	1.00	1.00	0.00	endo
37	exo	2.9	30.3	13.4	-0.5	-2.1	-0.7	2.0	0.94	0.06	0.06	0.94	exo
38	endo	3.7	29.8	10.8	-1.6	-1.1	-9.0	2.2	0.04	0.96	0.96	0.04	endo

typical output patterns. However, all output patterns are close to (1,0,0,1) or (0,1,1,0) and are apparently different from meaningless (0,1,0,1), etc. The network decisions are shown in the last column of the table and the rate for the correct answer was 12/13 = 0.923, which is better than the previous methods by the linear learning machine and the cluster analysis (both 11/13 = 0.846).

Quick and correct decision making based on ambiguous data is always required in clinical fields. As our model study shows, if W_{ij} for a series of symptons and clinical data in a specified disease is established, a quick diagnosis may be obtained using a small computer such as a personal computer.

REFERENCES

1) Eg. D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, "Parallel Distributed Processing Exploration in the Microstructure of Cognition," the MIT Press, Cambridge, Mass., Vol. 1 and 2 (1986).
2) S. Sasaki, H. Abe, Y. Takahashi, T. Takayama, and Y. Miyashita,"Introduction to Pattern Recognition for Chemists," Tokyo Kagaku Dojin, Tokyo, Japan, Chap. 1 (1984) (in Japanese).
3) B. R. Kowalski, "Chemometrics: Theory and Applications," ACS Symp. Ser. 53, Am. Chem. Soc., Washington, D. C., p243 (1977).

(Received July 7, 1989)