

ニューラルネットワークによる構造活性相関

青山智夫

日立コンピュータエンジニアリング

鈴木雄二 市川 紘

星薬科大学 物理教室

階層型ニューラルネットワークを用いて薬物の生理活性の強さを分子構造データから計算する方法を研究した。ニューラルネットワークに対し、既知の前記2種類のデータを入力、教師パターンとして学習を行い、学習後薬理作用を調べたい化合物の分子構造データを入力し、出力パターンからその生理活性の強弱および化学構造と生理活性相関をネットワークの出力から判定する。この方法は従来の統計的な方法と比較すると誤差の多いデータのときはより正確な分類結果が得られる。しかし相関については検討しなければならない点がある。

Neural Networks Applied to  
Structure-Activity Relationship

Tomoo Aoyama

Hitachi Computer Engineering Co.Ltd.(Hadano, Kanagawa, 259-13, Japan)

Yuji Suzuki, Hiroshi Ichikawa

Hoshi College of Pharmacy

This paper reports a general method of the study in the structure-activity-relationship. We studied the relationship between activities and structural data of the drug compounds, using the multi-layer neural network. The network was supervised by the training data which consisted of the activities and structural data, and trained by the backpropagation algorithms. After learning, untrained data was analyzed by the network, i.e., the category and correlation were calculated and the results were examined. It was found that the neural network had an ability superior in classification to conventional methods, and that there was a correlation between the coefficients in the neural network and those of the multi-regression analysis.

## 1. 序

望ましい医薬品を効率よく開発するため、薬物の生理活性の種類と強度を化学構造から推定する研究が行われている。推定的手法として従来から統計学および多変量解析が用いられていたが、最近新しい考え方に基づく方法たとえば学習機械法、シンプレックス法、k-最近隣法、Simple Classification Algorithm法、Adaptive Least Squares(ALS)法などが提唱されている。<sup>1~4)</sup>

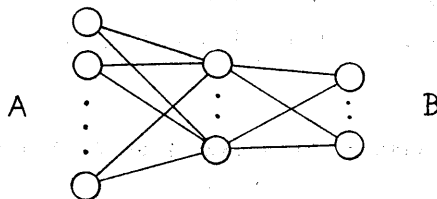
ALS等の方法は情報科学の観点ではパターン認識技術の一つである。すなわち化合物のスペクトルデータ、化学構造データを一種のパターンとして見て、生理活性データを別種のパターンとしてそれらの間の関係を解析している。

情報処理の分野ではパターン認識に関して種々の方法が提示されているが最近階層型ニューラルネットワークの方法が注目されている。階層型ニューラルネットワーク法<sup>5)</sup>は学習機能の他、因果関係がアルゴリズムで明確に記述出来ない処理に有効である。また入力データに誤差が含まれていてもパターン認識機能が損なわれる割合が少ない。この特徴は薬物の生理活性計算によく適合する。

## 2. 理論

### 2.1 学習

階層型ネットワークの一例を図1にしめす。



○印はニューロンと呼ばれる素子で、0~1の間の値をとる。ニューロン間を結合している直線はニューロン間の結合強度を示し $-\infty \sim +\infty$ の値をとる。ニューロン層の層数は一般にはnであって、データはA側のニューロン層から入力されB側のニューロン層に出力される。

1個のニューロンの動作は一般に式(1)で表される。

$$O_j = \beta / \{1 + \exp(-\alpha y_j)\} + (1 - \beta) y_j$$
$$y_j = \left( \sum W_{ij} x_i \right) - \theta_j \quad (1)$$

ここで $x_i$ はニューロンjの属するニューロン層の1つ前のニューロン層に属するニューロンiの値である。 $W_{ij}$ はニューロンi, jを結合している結合荷重の値である。 $\theta_j$ はニューロンjに固有の値である。パラメータ $\alpha$ でニューロン層の非線形動作をあらわし、 $\beta$ でsigmoid関数と線形関数の混合の割合を調整する。

データがA側から入力されて各層のニューロンの値を更新するとき、(1)式であらわされる動作が同期して行われる。A側から見て第1層(以下入力層)のニューロンの個数をNとすると、入力データはN個のニューロンの値のベクトルで表される。このベクトルのことを入力パターンという。B側からみて第1層(以下出力層)のニューロンの個数をMとすると、出力データはM個のニューロンの値を要素とするベクトルで表される。これを出力パターンという。

式(1)の $W_{ij}$ を計算するために出力層に提示するベクトルデータを教師パターンという。 $W_{ij}$ は出力パターンと教師パターンが一致するように次の式で計算される。

$$W_{ij} = -d_j x_j \varepsilon \quad (2)$$

$$\begin{cases} d_j = (o_j - t_j) f'(y_j) & (3a) \end{cases}$$

$$\begin{cases} d_j = (\sum W'_{j\ell} d_\ell) f'(y_j) & (3b) \end{cases}$$

ここで $\varepsilon$ は $W_{ij}$ の1回の修正量を決めるパラメータで、 $t_j$ は教師パターンである。(3a)は出力層の $W_{ij}$ 計算に用いる。(3b)はそれ以外の層の $W_{ij}$ 計算に用いる。 $W'$ は $W$ を修正する層より1つだけ出力層に近い層の $W$ 値である。 $d'$ も同様である。 $f'$ は(1)式の場合、

$$f'(y_j) = f(y_j) \{1 - f(y_j)\} \alpha \beta + (1 - \beta) \quad (4)$$

である。 $\varepsilon, \alpha, \beta$ とも各ニューロン層について異なる値を与えることが出来る。

式(2)は、

$$E = \sum (o_j - t_j)^2 \quad (5)$$

が十分小さくなるまで繰り返す( $E < 0.01$ )。この繰り返し計算をバックプロパゲーションという。

以上の計算によって入力パターン $\{x_j\}$ を階層型ネットワークに与えた時、出力パターンとして教師パターン $\{t_j\}$ に十分類似した出力パターン $\{o_j\}$ が得られる。さらにZ組の入力パターン、教師パターンがあるとき式(1,2)の対を繰り返し行うことによって、通常の場合はZ個の出力パターンを教師パターンに十分な精度で一致させる事が出来る。そのとき階層型ネットワークは入力パターンをZ個のカテゴリに分類する作用をもつ。

## 2.2 reforming

階層型ニューラルネットワークでは情報がニューロン間の結合の強さとして保持される。ニューロン間の結合は教師パターンを用いたバックプロパゲーション法によって求められる。このニューロン間の結合を求めるフェイズをlearningという。learningの反対の動作として、ニューロン間の結合の強さの絶対値を小さくする動作を考える。

この動作をネットワークに行なわせると学習した情報の一部が失われる。以下この動作をerasingという。erasingは次式で行う。

$$W_{ij} = W_{ij} - \text{sgn}(W_{ij}) \delta \{1 - \delta(W_{ij})\} \quad (6)$$

ここで  $\delta$  は  $|W_{ij}|$  が  $\zeta$  より小さい時 1、 $\zeta$  を超える時 0 となる関数である。

与えられた M セットの入力データで learning を行った後 erasing を行うと、情報は蓄積されつづいて失われる。この相反する動作はニューロン間の結合で均一に起こるのではない。ある結合では学習効果が大きく情報が蓄積されていく。他の結合では失う事が多く情報が消えていく。すなわちニューロン間の結合に蓄積された情報の再構成がおこる。このフェイズを reforming という。

ニューロン層に十分な数のニューロンが存在する時、reforming が可能である。

Z 組の入力、教師データがあるとき、式 (2,3) のバックプロパゲーション操作を各パターンについて順次行い、これを learning cycle とする。1 回の learning cycle 後式 (5) の erasing 処理を行う。これを reforming cycle という。

ここで  $\zeta$  は  $\delta$  の 1/10 程度の値を初期値とし、reforming cycle 中 E が変化しないように可変値をとる。reforming が成立し得ないような少数のニューロンを用いたネットワークの場合は  $\zeta \rightarrow 0$  となりバックプロパゲーションになる。

### 2.3 rescaling

ニューラルネットワークに入力するためにデータを下式により (L, H) 区間にスケーリングする必要がある。

$$x'_i = (x_i - B) g \quad (7a)$$

$$B = x_{\max} - H(x_{\max} - x_{\min})(H-L)^{-1} \quad (7b)$$

$$g = (H-L)(x_{\max} - x_{\min})^{-1} \quad (7c)$$

添え字 i はパターン番号をしめす。通常  $L=0$ ,  $H=1$  であるが、 $L=0.1$ ,  $H=1$  又は  $L=0.2$ ,  $H=0.8$  とすることもある。

入力データが  $\{x_{k,i}\}$  で添え字 K が分子構造データで  $x_K^p, x_L^q$  の関係に意味があるとき、

$$x_M = x_K^p \otimes x_L^q \quad (8)$$

$\{x_{M,i}\}$  を入力データに加えてスケーリングする。 $\otimes$  は四則演算等を表す。

$\{x_{K,i}\}$  を p 乗することで生体反応の非線形性を考慮している。

$p=0$  すなわち 1 ベクトルを  $\{x_{k,i}\}$  に追加し、各層のニューロンの  $\theta$  を 0、かつ出力層のニューロン数を 1、 $\beta=0$  としたネットワークは重回帰分析との関係を考察する場合重要である。この型のニューラルネットワークを MR 型のネットワークという。

## 2. 4 偏相関係数の計算

MR型のニューラルネットワークを用いて、入力データと出力データとの相関を次のようにして計算した。

1.  $\{t_i\}$  と  $[\{1\}, \{x_{k(\neq L), i}\}]$  でニューラルネットに学習させ、計算された  $\{o_i\}$  と  $\{t_i\}$  の差を  $\{\Delta_i\}$  とする。
2.  $\{x_{k, i}\}$  と  $[\{t_i\}, \{x_{k'(\neq L, k), i}\}, \{1\}]$  でニューラルネットの学習を行い、ニューラルネットから計算される  $\{o_i\}$  と  $\{x_{k, i}\}$  の差を  $\{d_i\}$  とする。
3.  $\{\Delta_i\}$  と  $\{d_i\}$  の相関係数を求める。

1~3の操作を各 $\beta$ 値について行う。 $\alpha$ は $\sim 1$ 程度でよい。 $\beta \rightarrow 0$ の極限では重回帰分析とニューラルネットの偏相関係数は一致する。

第2層で $\beta = 0$ としたMR型のニューラルネットでは

$$h_j = \sum W_{ij}^{(1,2)} x_i \quad (9a)$$

$$o_k = \sum W_{jk}^{(2,3)} h_j = \sum \sum W_{jk}^{(2,3)} W_{ij}^{(1,2)} x_i = \sum W_{ki}^o x_i \quad (9b)$$

という関係で $x_i, o_i$ が関係づけられている。第2層のニューロンを非線形とすると、次の $V_{ijk}$ という量が相関に関することになる。

$$V_{ijk} = W_{ij}^{(1,2)} \{ \beta^{(2)} H_j + (1 - \beta^{(2)}) \} \\ \times W_{jk}^{(2,3)} \{ \beta^{(3)} O_k + (1 - \beta^{(3)}) \} \quad (10)$$

ここで $W^{(1,2)}, W^{(2,3)}$ はそれぞれ(1, 2)層間, (2, 3)層間のweight-matrixである。 $\beta^{(2)}, \beta^{(3)}$ はそれぞれ第2, 3層のニューロンの動作を規定する $\beta$ 値である。 $H_j, O_k$ は第2, 3層のニューロン値である。 $V_{ijk}$ はjという第2層のニューロンを介して入力層のi番目のニューロンの情報が出力層のk番目のニューロンに伝播する度合いに関係する。 $V_{ijk}$ はn層のニューラルネットに容易に拡張出来る。jは任意性があるので、総計をとった

$$U_{ik} = \sum_j V_{ijk} \quad (11)$$

が有意な量である。 $U_{ik}$ は入出力間のニューロン値の関係を表している。M個の入力データがあるとき、次の平均量 $\bar{U}_{ik}$ を

$$\bar{U}_{ik} = M^{-1} \sum_I U_{ik}(I) \quad (12)$$

で定義する。ここで $U_{ik}(I)$ はI番のパターンに対するUである。 $\tilde{U}_k(\max) = \max \{ |U_{ik}| \}$ として、 $\tilde{U}(\max) = 1$ となるように規格化すると便利である。

以下 $\tilde{U}$ をcorrelation matrixという。

### 3. 数値実験 1

6) 文献に示されている38種のノルボルナン誘導体データのうち25データを階層型ニューラルネットワークに学習させた。すなわち各化合物と対応する基本骨格化合物との<sup>13</sup>C-NMR化学シフト値の差[ppm] 7種とendo,exo配向をベクトル表現したデータをネットワークに与えた。ニューロンの値は(0~1)区間で定義されるため、<sup>13</sup>C-NMRデータは(0.090901-1)の間になるようにスケーリングした。

endo,exoのベクトル表現は(0,1,1,0),(1,0,0,1)とした。4要素の冗長なベクトル表現を用いた理由はニューラルネットワークの結果をこの冗長性を利用して確認するためである。

ニューラルネットワークのニューロン数は第1~3層でそれぞれ7,14,4個である。は50 cycle以前では0.2それ以後では0.1とした。

表1aにlearning cycleだけで求めたE値、第3層の2個のニューロンの出力パターンを示す。表1bにはlearning, reforming cycleを50回ずつ交互に繰り返した時の結果を示す。2種類の方法で求めたニューラルネットワークは同じレベルのデータ分類機能を持っている。

表 1 a

表 1 b

	E 値	出力 パターン		E 値	出力パターン		判定	
1	0.011	0.945	0.054	0.017	0.931	0.066	↑ exo	
2	0.000	0.996	0.003	0.001	0.983	0.015		
3	0.000	0.998	0.001	0.000	0.986	0.013		
4	0.045	0.894	0.107	0.045	0.893	0.106		
5	0.000	0.963	0.035	0.000	0.958	0.040		
6	0.001	0.980	0.020	0.003	0.969	0.029		
7	0.004	0.966	0.032	0.007	0.956	0.042		
8	0.001	0.982	0.017	0.003	0.968	0.030		
9	0.012	0.942	0.056	0.014	0.937	0.060		
10	0.000	0.984	0.016	0.004	0.967	0.032		
11	0.000	0.985	0.014	0.002	0.974	0.025		
12	0.000	0.986	0.013	0.002	0.975	0.023		
13	0.041	0.900	0.097	0.067	0.877	0.131		↓ exo
14	0.000	0.008	0.991	0.002	0.024	0.975	↓ endo	
15	0.000	0.010	0.989	0.002	0.025	0.974	↑ endo	
16	0.023	0.081	0.924	0.028	0.084	0.915		
17	0.000	0.012	0.987	0.004	0.031	0.968		
18	0.000	0.009	0.991	0.002	0.025	0.974		
19	0.000	0.014	0.986	0.004	0.033	0.965		
20	0.000	0.008	0.991	0.002	0.024	0.975		
21	0.095	0.151	0.845	0.113	0.165	0.830		
22	0.000	0.007	0.992	0.002	0.026	0.972		
23	0.000	0.033	0.966	0.017	0.065	0.933		
24	0.000	0.011	0.988	0.003	0.028	0.970		
25	0.000	0.013	0.986	0.002	0.024	0.976		↓ endo

それぞれのネットワークのニューロン間の結合強度を図 2,3に示す。

図 2

図 3

1, 2 層間

	1	2	3	4	5	6	7	8	9	0	1	2	3	4
1	-1	1	0	0	1	0	1	0	0	0	0	0	0	0
2	-1	-1	1	1	2	5	-1	0	-3	0	1	0	1	-3
3	0	-1	0	1	0	1	-1	0	-1	0	0	0	1	0
4	3	2	-1	-4	-3	-4	2	-1	4	-3	-1	-2	-2	3
5	1	2	-2	-1	-1	-5	2	-1	2	-1	-1	-1	-2	3
6	-2	-4	2	4	2	7	-1	0	-5	3	1	2	2	-5
7	4	6	-3	-6	-4	-9	2	-1	8	-2	0	-3	-3	7

	1	2	3	4	5	6	7	8	9	0	1	2	3	4
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	5	0	0	-2	0	0	0	0	-1
3	0	0	0	0	0	1	0	0	0	0	0	0	0	0
4	0	0	0	0	0	-4	0	0	2	0	0	0	0	1
5	0	0	0	0	0	-4	0	0	2	0	0	0	0	1
6	0	0	0	1	0	8	0	0	-5	0	0	0	0	-3
7	0	1	0	-2	0	-9	0	0	6	0	0	0	0	5

2, 3 層間

1	-4	-3	2	4	3	9	-2	1	-7	3	1	3	2	-6
2	2	4	-3	-6	-3	-9	3	0	6	-2	0	-2	3	6
3	3	5	-1	-6	-4	-7	2	-1	6	-3	-1	-2	3	7
4	-5	-5	3	4	3	9	-2	0	-6	3	2	3	3	-5

1	0	0	0	0	0	9	0	0	-5	0	0	0	0	-3
2	0	0	0	-1	0	-9	0	0	5	0	0	0	0	3
3	0	0	0	-2	0	-9	0	0	4	0	0	0	0	4
4	0	0	0	1	0	9	0	0	-5	0	0	0	0	-3

図 2、3 では行列要素の絶対値の大きさを 0~9 の整数で表現している。

learning cycle だけで求めたニューラルネットワークは情報が全ニューロン間に分散しているが、learning, reforming cycle を繰り返したネットワークでは情報が局在化する。図 3 より第 2 層の 1,3,5,7,8,10~13 番目のニューロンは分類作用に寄与していないことが明らかである。これらのニューロンは第 3 層のニューロンのしきい値として作用する。

主に分類作用に寄与するニューロンは第 2 層の 6,9 番目のニューロンである。これらのニューロンからそれぞれ  $(a, -a, -a, a)$ ,  $(-b, b, b, -b)$  型の結合によって第 3 層の 4 個のニューロンに情報が伝播する。第 3 層のニューロンはその情報を結合して endo, exo を識別する。第 2 層 6,9 番目のニューロンと最も強く結合している第 1 層のニューロンは 6,7 番目のニューロンであって、ノルボルナン誘導体の  $C_6$ ,  $C_7$  位の情報が endo/exo 配向性を決定していることを示す。これは Fisher 比から予想されている結果と一致する。

以上のようにして作成した 2 種類のニューラルネットワークにまだ学習させていない 13 種のノルボルナン誘導体の測定データを入力し、出力パターンから endo/exo の配向を判定した。結果を表 2a, b に示す。

表 2 a

	1	2	3	4	
26	0.793	0.211	0.208	0.796	exo
27	0.048	0.951	0.952	0.049	endo
28	0.953	0.048	0.046	0.953	exo
29	0.016	0.984	0.984	0.015	endo
30	0.991	0.008	0.008	0.991	exo
31	0.738	0.259	0.251	0.744	exo*
32	0.971	0.027	0.027	0.972	exo
33	0.006	0.994	0.994	0.005	endo
34	0.977	0.021	0.021	0.978	exo
35	0.029	0.970	0.971	0.029	endo
36	0.005	0.995	0.995	0.004	endo
37	0.889	0.106	0.117	0.886	exo
38	0.106	0.898	0.894	0.098	endo

表 2 b

	1	2	3	4	
26	0.753	0.243	0.241	0.753	exo
27	0.062	0.934	0.932	0.064	endo
28	0.939	0.060	0.061	0.938	exo
29	0.034	0.964	0.964	0.035	endo
30	0.979	0.019	0.019	0.979	exo
31	0.692	0.302	0.299	0.692	exo*
32	0.962	0.036	0.036	0.962	exo
33	0.021	0.978	0.978	0.021	endo
34	0.971	0.027	0.027	0.971	exo
35	0.053	0.945	0.944	0.054	endo
36	0.019	0.980	0.980	0.020	endo
37	0.867	0.141	0.148	0.864	exo
38	0.070	0.931	0.931	0.070	endo

\* : error

ネットワークの出力ニューロン値が 0,1 から離れば離れるほど学習した25種のデータとの相違が大きい事を示す。表 2a,bの結果はendo/exoの判定が出来ない程ではない。判定結果を表 2 中に示した。

表 2 中の31番目のデータの判定結果は実測と異なる。このデータのニューロンの出力値は実測と最も大きい隔たりを示した。表 2 では未学習データの認識率は 92%で、従来の方法の 85%を上回る。

## 4. 数値実験 2

重回帰分析による偏相関係数、ニューラルネットワークによるcorrelation matrixの関連を薬品の収率(y)、主原料(x<sub>1</sub>)、反応温度(x<sub>2</sub>)、反応圧力(x<sub>3</sub>)、副原料(x<sub>4</sub>)に関するデータを用いて調べた。

ニューラルネットのパラメータは $\beta^{(3)} = 0$ ,  $\beta^{(2)} = 1$ である。

表 3

	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	y	y(M.R)	y(N.N)
1	53.0	153.2	43.7	10.6	92.36	92.54	92.50
2	34.8	120.4	27.9	10.8	75.06	75.54	75.59
3	47.3	135.4	37.2	9.4	81.54	81.59	81.73
4	47.1	135.7	34.4	7.6	78.25	78.73	78.96
5	56.5	146.9	41.3	11.9	86.94	87.77	88.33
6	57.9	149.0	44.2	10.1	89.67	88.27	88.61
7	43.9	131.9	34.8	10.2	81.19	80.66	80.79
8	25.7	107.2	20.5	11.0	69.12	68.97	68.94
9	68.1	156.6	57.3	9.3	93.67	94.29	94.27
10	57.2	145.6	46.3	10.1	88.63	88.07	88.23



このデータでは決定係数が0.993であって、重回帰分析とニューラルネットの y 計算値はほとんど同じである。

y と x の偏相関係数、correlation matrix を表 4 に示す。

表 4

		x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>
重回帰分析	偏相関係数	-0.77	0.94	0.86	0.82
ニューラル ネットワーク	偏相関係数	-0.78	0.94	0.86	0.82 ( $\beta=0.2$ )
		-0.54	0.90	0.50	0.15 ( $\beta=1.0$ )
	correlation matrix	-0.61	1.00	0.74	0.16

y と x<sub>4</sub> のケースでは、 $\beta^{(2)}$  を大きくしていくとニューラルネットの偏相関係数と重回帰分析のそれとは大きく異なるようになる。この原因は x<sub>4</sub> の変化をニューロンの非線形動作のため { x<sub>1,2,3</sub> } で表されることにある。

ニューラルネットの特徴は決定係数が小さく、重回帰分析では実験を十分に再現できない時に表れる。表 5 は別の種類の薬品製造工程のデータである。

表 5

	y (obs)	y (M.R)	y (N.N)
1	110.9	90.1	110.9
2	84.6	91.3	84.6
3	81.7	82.3	81.6
4	74.5	71.0	74.6
5	79.8	84.2	79.8

R<sup>2</sup> = 0.369

以下 略

変数は 6 種あって、それぞれの偏相関係数と correlation matrix は次のようになった

偏相関係数	0.13	-0.04	0.51	0.33	-0.10	-0.24
correlation matrix	0.25	-0.07	0.90	0.46	0.19	-0.24

この例ではニューラルネットワークの計算で実測をほぼ完全に再現している。偏相関係数と correlation matrix の対応も良い。

## 5 結論

ニューラルネットワークを用いて、原因と思われる誤差の含まれるデータから実測値を再現出来る可能性がある。これはニューラルネットの強い非線形 *f i t t i n g* 作用による。この機能はデータを分類する場合に有用である。

ニューラルネットでも偏相関係数を求めたり、その簡約法として *correlation matrix* を計算出来る。しかしニューラルネットには強い非線形 *f i t t i n g* 作用があるので、その結果を鵜呑みにしてはいけない。パラメータ  $\beta$  を 0 に近づけて、相関係数が大きく変動しないことを確かめる必要がある。

## 6. 文献

- 1) "Drug Design", Vol.1-10, E.J.Ariens,Ed., Academic Press, New York.
- 2) "Chemometrics", M.A.Sharaf, D.L.Illmann, and B.R.Kowalski, Ed., John Wiley & Sons, New York(1986).
- 3) S.Wold and M.Sjostrom, "Chemometrics:Theory and Applications", ACS Sym.Ser. No.52, B.R.Kowalski,Ed., Am.Chgem.Soc., Washington,D.C., 1977,p.243.
- 4) I.Moriguchi and K.Komatsu, Chem.Pharm.Bull.,25,2800(1977)
- 5) 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).
- 6) 佐々木慎一、阿部英次、高橋由雅、高山千代蔵、宮下芳勝, 「化学者のためのパターン認識序説」東京化学同人 (1984), p.7