6

Forced Information for Information-Theoretic Competitive Learning

Ryotaro Kamimura
IT Education Center, Information Technology Center, Tokai University, Japan

1. Introduction

We have proposed a new information-theoretic approach to competitive learning [1], [2], [3], [4], [5]. The information-theoretic method is a very flexible type of competitive learning, compared with conventional competitive learning. However, some problems have been pointed out concerning the information-theoretic method, for example, slow convergence. In this paper, we propose a new computational method to accelerate a process of information maximization. In addition, an information loss is introduced to detect the salient features of input patterns.

Competitive learning is one of the most important techniques in neural networks with many problems such as the dead neuron problem [6], [7]. Thus, many methods have been proposed to solve those problems, for example, conscience learning [8], frequency-sensitive learning [9], rival penalized competitive learning [10], lotto-type competitive learning [11] and entropy maximization [12]. We have so far developed information-theoretic competitive learning to solve those fundamental problems of competitive learning. In the information-theoretic learning, no dead neurons can be produced, because entropy of competitive units must be maximized. In addition, experimental results have shown that final connection weights are relatively independent of initial conditions.

However, one of the major problems is that it is sometimes slow in increasing information. As a problem becomes more complex, heavier computation is needed. Without solving this problem, it is impossible for the information-theoretic method to be applied to practical problems. To overcome this problem, we propose a new type of computational method to accelerate a process of information maximization. In this method, information is supposed to be maximized or sufficiently high at the beginning of learning. This supposed maximum information forces networks to converge to stable points very rapidly. This supposed maximum information is obtained by using the ordinary winner-take-all algorithm. Thus, this method is one in which the winter-takeall is combined with a process of information maximization.

We also present a new approach to detect the importance of a given variable, that is, information loss. Information loss is difference between information with all variables and information without a variable, and is used to represent the importance of a given variable. Forced information with information loss can be used to extract main features of input patterns. Connection weights can be interpreted as the main characteristics of classified groups. On the other hand, information loss is used to extract the features on which input

Source: Machine Learning, Book edited by: Abdelhamid Mellouk and Abdennacer Chebira,
ISBN 978-3-902613-56-1, pp. 450, February 2009, I-Tech, Vienna, Austria

patterns or groups are classified. Thus, forced information and information loss has a possibility to show clearly main features of input patterns.
In Section 2, we present how to compute forced information as well as how to compute information loss. In Section 3 and 4, we present experimental results on a simple symmetric and Senate problem to show that one epoch is enough to reach stable points. In Section 5, we present experimental results on a student survey. In this section, we try to show that learning is accelerated more than sixty times faster, and explicit representations can be obtained.

2. Information maximization

We consider information content stored in competitive unit activation patterns. For this purpose, let us define information to be stored in a neural system. Information stored in a system is represented by decrease in uncertainty [13]. Uncertainty decrease, that is, information I, is defined by

$$I = -\sum_{\forall j} p(j) \log p(j) + \sum_{\forall s} \sum_{\forall j} p(s) p(j \mid s) \log p(j \mid s), \tag{1}$$

where $p(j)$, $p(s)$ and $p(j|s)$ denote the probability of firing of the jth unit, the probability of the sth input pattern and the conditional probability of the jth unit, given the sth input pattern, respectively. When the conditional probability $p(j|s)$ is independent of the occurrence of the sth input pattern, that is, $p(j|s) = p(j)$, mutual information becomes zero.

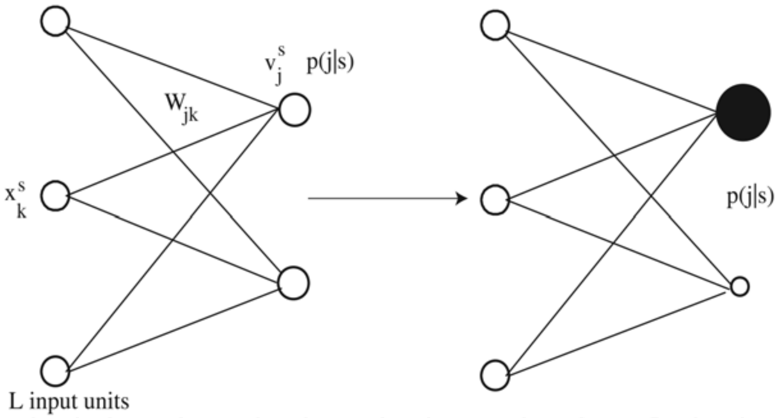


Fig. 1. A single-layered network architecture for information maximization.

Let us present update rules to maximize information content. As shown in Figure 2, a network is composed of input units x_k^s and competitive units v_j^s. We used as the output function the inverse of the square of the Euclidean distance between connection weights and outputs for facilitating the derivation. Thus, distance is defined by

$$d_j^s = \sum_{k=1}^{L} (x_k^s - w_{jk})^2. \tag{2}$$

An output from the jth competitive unit can be computed by

$$v_j^s = \frac{1}{d_j^s}, \tag{3}$$

where L is the number of input units, and w_{jk} denote connections from the kth input unit to the jth competitive unit. The output is increased as connection weights are closer to input patterns.

The conditional probability $p(j \mid s)$ is computed by

$$p(j \mid s) = \frac{v_j^s}{\sum_{m=1}^{M} v_m^s}, \tag{4}$$

where M denotes the number of competitive units. Since input patterns are supposed to be uniformly given to networks, the probability of the jth competitive unit is computed by

$$p(j) = \frac{1}{S}\sum_{s=1}^{S} p(j \mid s). \tag{5}$$

Information I is computed by

$$I = -\sum_{j=1}^{M} p(j)\log p(j) + \frac{1}{S}\sum_{s=1}^{S}\sum_{j=1}^{M} p(j \mid s)\log p(j \mid s). \tag{6}$$

Differentiating information with respect to input-competitive connections w_{jk}, we have

$$\begin{aligned}\Delta w_{jk} = & -\beta\sum_{s=1}^{S}\left(\log p(j) - \sum_{m=1}^{M} p(m \mid s)\log p(m)\right) Q_{jk}^s \\ & +\beta\sum_{s=1}^{S}\left(\log p(j \mid s) - \sum_{m=1}^{M} p(m \mid s)\log p(m \mid s)\right) Q_{jk}^s,\end{aligned} \tag{7}$$

where β is the learning parameter, and

$$Q_{jk}^s = \frac{1}{S} p(j \mid s) v_j^s (x_k^s - w_{jk}). \tag{8}$$

3. Maximum information-forced learning

One of the major shortcomings of information-theoretic competitive learning is that it is sometimes very slow in increasing information content to a sufficiently large level. We here present how to accelerate learning by supposing that information is already maximized before learning. Thus, we have a conditional probability $p(j \mid s)$ such that the probability is set to ϵ for a winner, and $1 - \epsilon$ for all the other units. We here suppose that ϵ ranges between zero and unity. For example, supposing that information is almost maximized with two competitive units, and this means that a conditional probability is close to unity, and all the other probabilities are close to zero. Thus, we should take the parameter ϵ as a value close to unity, say, 0.9. In this case, all the other cases are set to 0.1. Weights are updated so as to maximize usual information content. The conditional probability $p(j \mid s)$ is computed by

$$p(j \mid s) = \frac{v_j^s}{\sum_{m=1}^{M} v_m^s}, \tag{9}$$

where M denotes the number of competitive units.

$$p(j \mid s) = \begin{cases} \epsilon & \text{for a winner} \\ \frac{1-\epsilon}{M-1} & \text{otherwise} \end{cases} \tag{10}$$

At this place, we suppose that information is already close to a maximum value. This means that if the jth unit is a winner, the probability of the jth unit should be as large as possible, and close to unity, while all the other units' firing rates should be as small as possible.

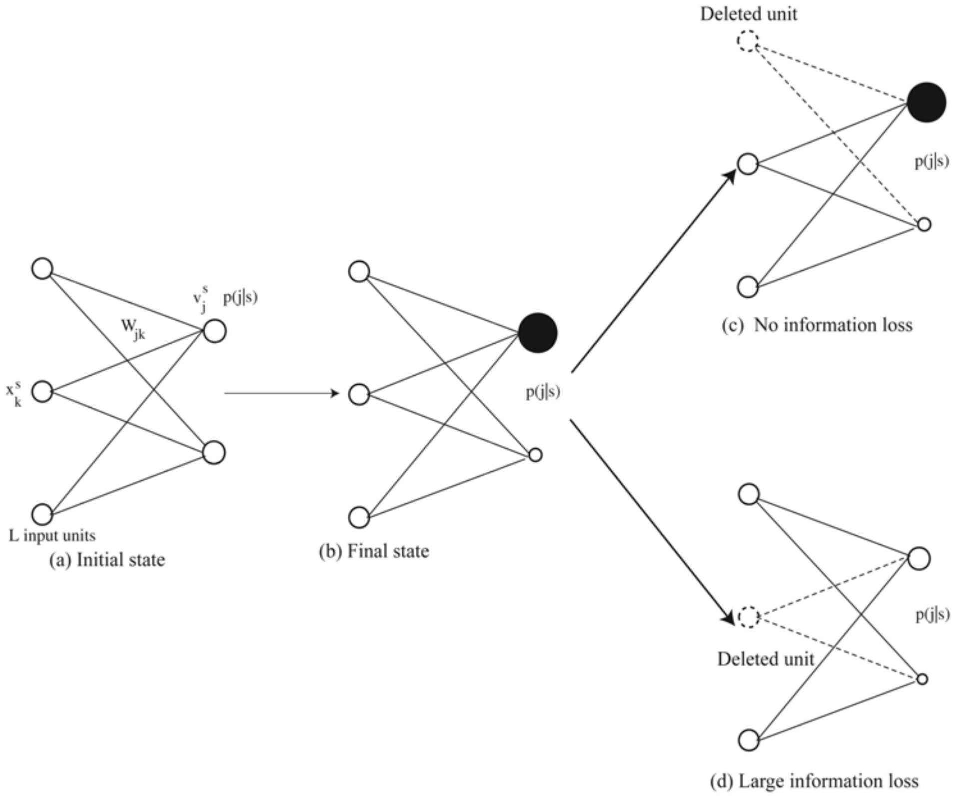


Fig. 2. A single-layered network architecture for information maximization.

This forced information is a method to include the winner-take-all algorithm inside information maximization. As already mentioned, the winner-take-all is a realization of forced information maximization, because information is supposed to be maximized.

4. Information loss

We now define information when a neuron is damaged by some reasons. In this case, distance without the *m*th unit is defined by

$$d_{jm}^s = \sum_{k \neq m} (x_k^s - w_{jk})^2, \tag{11}$$

where summation is over all input units except the *m*th unit. The output without the *m*th unit is defined by

$$v_{jm}^s = \frac{1}{d_{jm}^s}. \tag{12}$$

The normalized output is computed by

$$p^m(j \mid s) = \frac{v_{jm}^s}{\sum_{l=1}^M v_{lm}^s}. \tag{13}$$

Now, let us define mutual information without the *m*th input unit by

$$I_m = -\sum_{j=1}^M p^m(j) \log p^m(j) + \sum_{s=1}^S \sum_{j=1}^M p(s) p^m(j \mid s) \log p^m(j \mid s), \tag{14}$$

where p_m and $p_m(j \mid s)$ denote a probability and a conditional probability, given the *s*th input pattern. Information loss is defined by difference between original mutual information with full units and connections and mutual information without a unit. Thus, we have information loss

$$IL_m = I - I_m. \tag{15}$$

For each competitive unit, we compute conditional mutual information for each competitive unit.
For this, we transform mutual information as follows.

$$I = \sum_{s=1}^S \sum_{j=1}^M p(s) p(j \mid s) \log \frac{p(j \mid s)}{p(j)}. \tag{16}$$

Conditional mutual information for each competitive unit is defined by

$$I_j = \sum_{s=1}^S p(s) p(j \mid s) \log \frac{p(j \mid s)}{p(j)}. \tag{17}$$

Thus, conditional information loss is defined by

$$IL_{jm} = I_j - I_{jm} \tag{18}$$

We have the following relation:

$$IL_m = \sum_{j=1}^{M} IL_{jm} \tag{19}$$

5. Experiment No.1: symmetric data

In this experiment, we try to show that symmetric data can easily be classified by forced information. Figure 3 shows a network architecture where six input patterns are given into input units. These input patterns can naturally be classified into two classes. Figure 4 shows

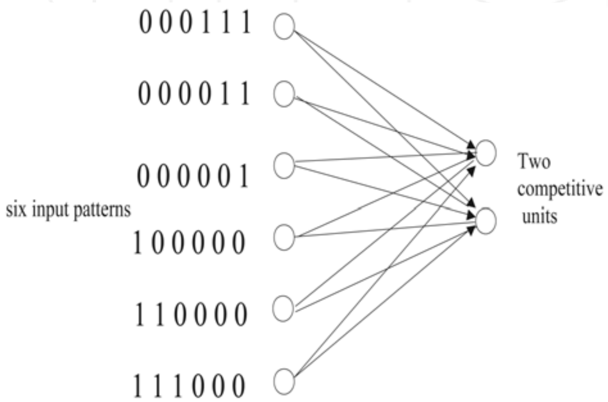


Fig. 3. A network architecture for the artificial data.

No. Party	1 D	2 D	3 D	4 D	5 D	6 D	7 D	8 D	9 R	10 R	11 R	12 R	13 R	14 R	15 R
1	0	1	0	1	1	1	1	0.5	0	0	1	1	1	1	0
2	0.5	0	0	0	0	0	0	0	1	0	0	0	1	0	1
3	0	0.5	1	1	1	1	1	1	1	0.5	1	1	0.5	1	1
4	1	0.5	1	1	1	1	0	0	0	0	0	0	0	1	0
5	1	0.5	1	1	1	1	1	0	0	0.5	0	0	0	1	1
6	1	0.5	0	0	0	0	0	0	1	0.5	0	1	1	0	1
7	1	1	1	0.5	1	1	1	1	0	0	1	1	1	1	1
8	0	0	0.5	0	0	0	0	0	1	1	1	0	0	0	1
9	1	1	0.5	1	1	1	0.5	1	0	0	0	0	0	1	1
10	1	1	1	1	1	1	0	1	0	0	0	0	0	1	0
11	1	1	1	1	1	1	0	1	0	0	1	0	0	1	0
12	1	0.5	1	1	1	1	1	1	0	0.5	0	0	0	1	0
13	0	1	0	0	0	0	0	1	1	0.5	0.5	0	1	0	0
14	0	0	0	0	0	0	0	0	0	0.5	0	0	1	0	0
15	0	0	0	0	0	0	0	0	1	0.5	1	0	1	0	1
16	0	0.5	0	0	0.5	0	0.5	0	0.5	0.5	1	1	1	0	1
17	0	0	0	0	0	0	0	0	0.5	0.5	1	0	1	0	0
18	1	1	1	1	1	1	1	1	1	1	0.5	0	1	1	0
19	1	1	1	1	0	0	0	1	0	0	0.5	0	0	1	0

Table 1: U.S. congressmen by their voting attitude on 19 environmental bills. The first 8 congressmen are Republicans, while the latter 7 (from 9 to 15) congressmen are Democrats. In the table, 1, 0 and 0.5 represent *yes*, *no* and undecided, respectively.

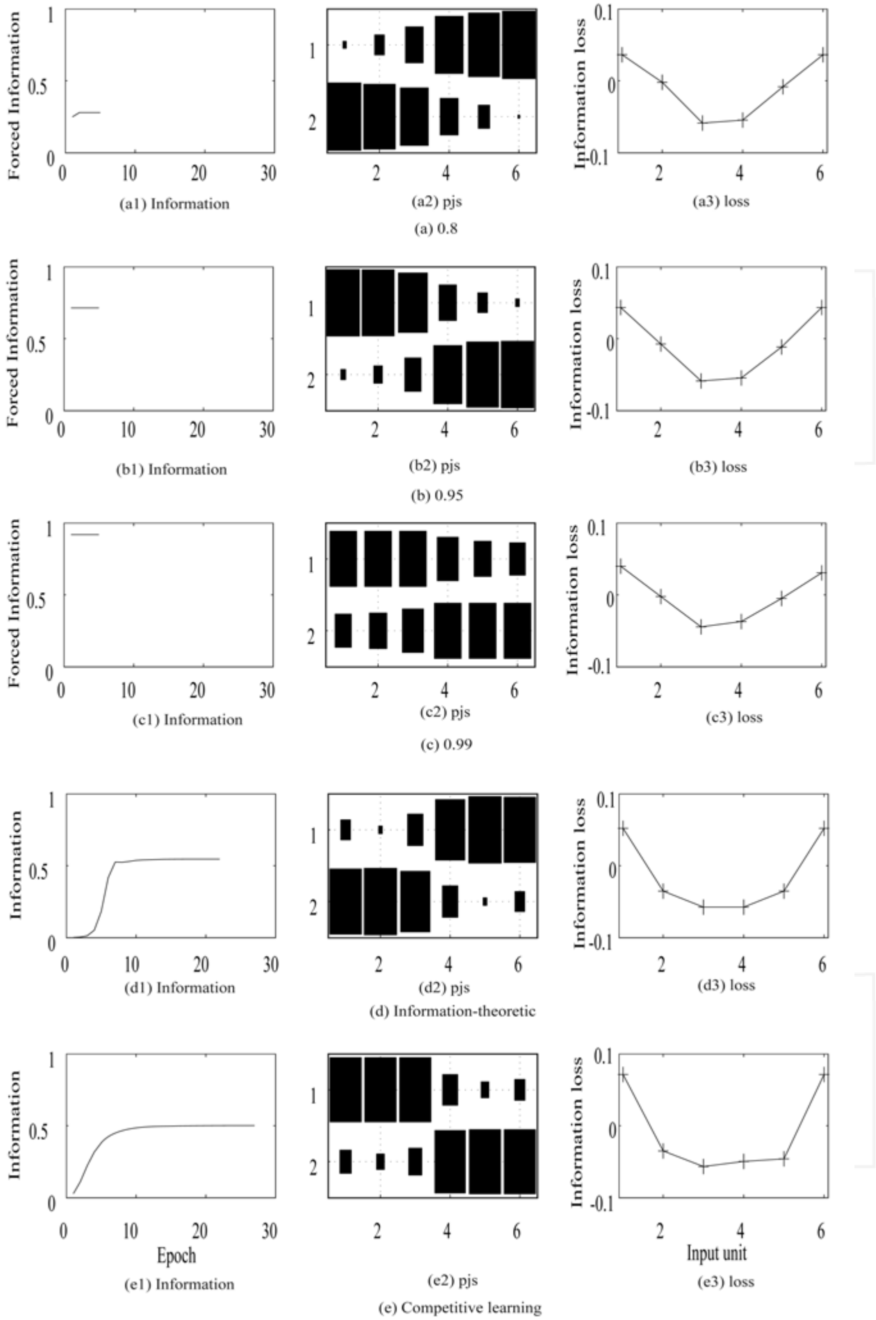


Fig. 4. Information, forced information, probabilities and information losses for the artificial data.

information, forced information, probabilities and information losses for the symmetric data. When the constant ϵ is set to 0.8, information reaches a stable point with eight epochs. When the constant is increased to 0.95, just one epoch is enough to reach that point. However, when information is further increased to 0.99, information reaches easily a stable point, but obtained probabilities show rather ambiguous patterns. Compared with forced information, information-theoretic learning needs more than 20 epochs and as many as 30 epochs are needed by competitive learning. We could obtain almost same probabilities $p(j|s)$ except ϵ = 0.99. For the information loss, the first and the sixth input patterns show large information loss, that is, important. This represents quite well symmetric input patterns.

6. Experiment No.2: senate problem

Table 1 shows the data of U.S. congressmen by their voting attitude on 19 environmental bills ??. The first 8 congressmen are Republicans, while the latter 7 (from 9 to 15) congressmen are Democrats. In the table, 1, 0 and 0.5 represent *yes*, *no* and undecided. Figure 5 shows information, forced information and information loss for the senate problem. When the constant ϵ is set to 0.8, information reaches a stable point with eight epochs. When the constant is increased to 0.95, just one epoch is enough to reach that point. However, when information is further increased to 0.99, obtained probabilities show rather ambiguous patterns. Compared with forced information, information-theoretic learning needs more than 25 epochs and as many as 15 epochs are needed by competitive learning. In addition, in almost all cases, the information loss shows the same pattern. The tenth, eleventh and twelfth input unit take large losses, meaning that these units play very important roles in learning. By examining Table 1, we can see that these units surely divide input patterns into two classes. Thus, the information captures the features in input patterns quite well.

7. Experiment 3: student survey

7.1 Two groups analysis

In the third experiment, we report an experimental result on a student survey. We did student survey about what subjects they are interested in. The number of students was 580, and the number of variables (questionnaires) was 58. Figure 6 shows a network architecture with two competitive units. The number of input units is 58 units, corresponding to 58 items such as *computer*, *internet* and so on. The students must respond to these items with four scales.

In the previous information-theoretic model, when the number of competitive units is large, it is sometimes impossible to attain the appropriate level of information. Figure 7 shows information as a function of the number of epochs. By using simple information maximization, we need as many as 500 epochs to be stabilized. On the other hand, by forced information, we need just eight epochs to finish learning. Almost same representations could be obtained. Thus, we can say that forced information maximization can accelerate learning almost seven times faster than the ordinary information maximization.

Figure 8 shows connection weights for two groups analysis. The first group represents a group with a higher interest in the items. The numbers of students in these groups are 256 and 324.

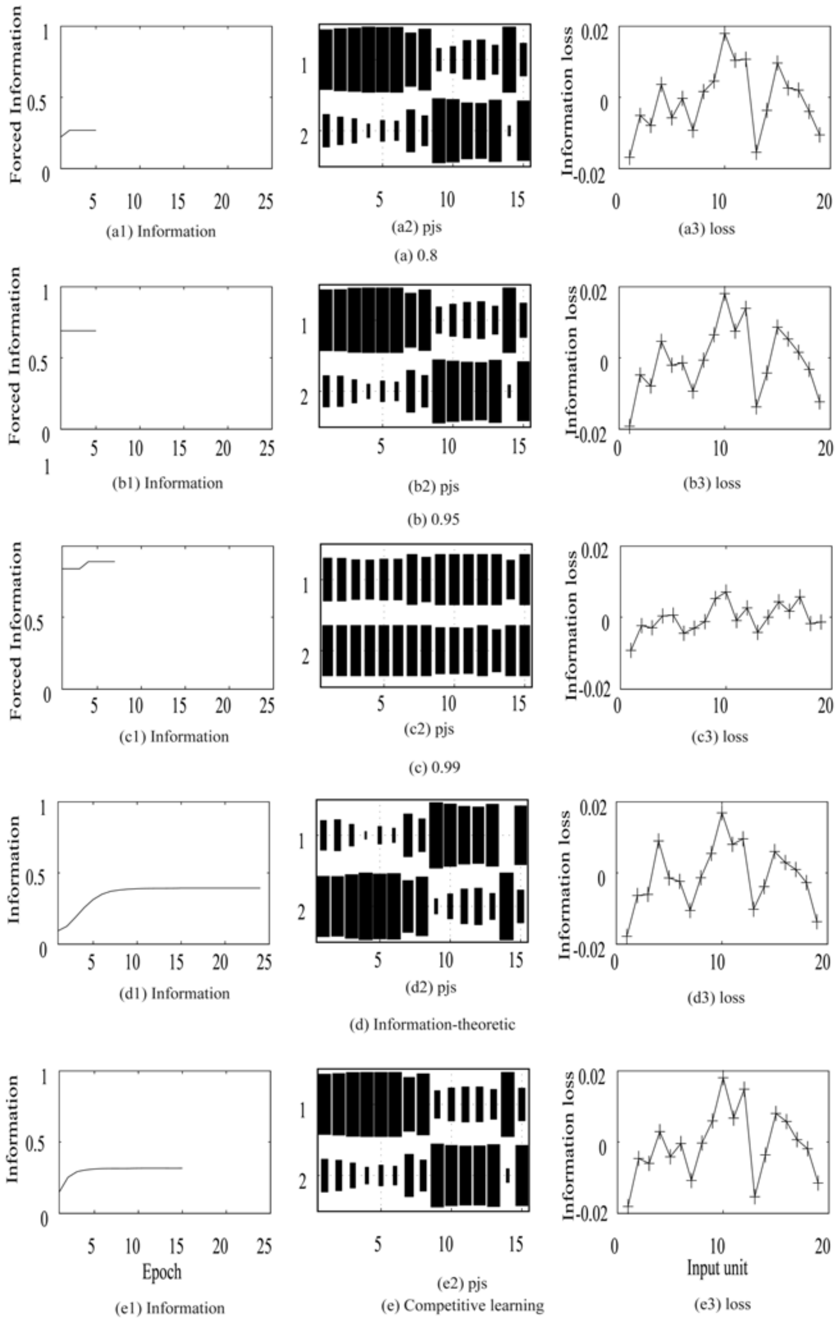


Fig. 5. Information, forced information, probabilities and information loss for the senate problem.

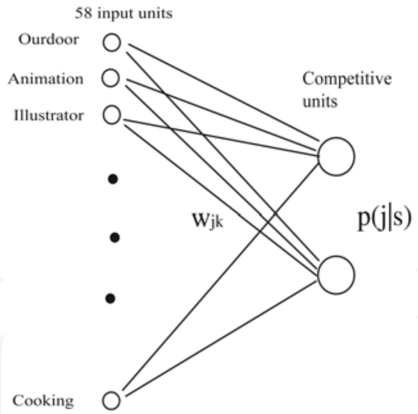


Fig. 6. Network architecture for a student analysis.

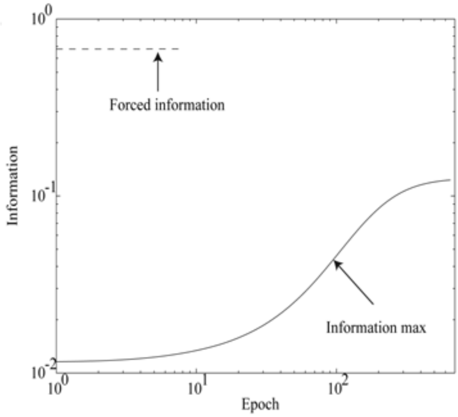


Fig. 7. Information and forced information as a function of the number of epochs by information-theoretic and forced-information method.

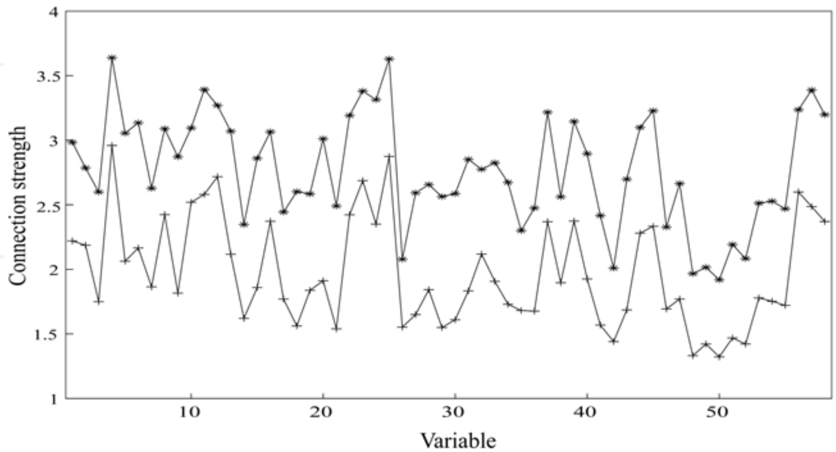


Fig. 8. Connection weights for two groups analysis.

This means that the method can classify 580 students by the magnitude of connection weights. Because connection weights try to imitate input patterns directly, we can see that two competitive units show students with high interest and low interest in the items in the questionnaire.
Table 2 represents the ranking of items for a group with a high interest in the items. As can be seen in the table, students respond highly to *internet* and *computer*, because we did this survey for the classes of information technology. Except these items, the majority is related to the so-called *entertainment* such as *music*, *travel*, *movie*. In addition, these students have some interest in *human relations* as well as *qualification*. On the other hand, these students have little interest in traditional and academic sciences such as *physics* and *mathematics*. Table 3 represents the ranking of items for a group with a low interest in the items. Except the difference of the strength, this group is similar to the first group. That is, students in this gropup respond highly to *internet* and *computer*, and they have keen interest in *entertainment*. On the other hand, these students have little interest in traditional and academic sciences such as *physics* and *mathematics*. Table 4 shows the information loss for the two groups. As can be seen in the table, two groups are separated by items such as *multimedia*, *business*. Especially, many terms concerning *business* appear in the table. This means that two groups are separated mainly based upon *business*. The most important thing to differentiate two groups is whether students have some interest in *buisiness* or *multimedia*. Let us see what the information loss represents in actual cases. Figure 9 shows the information loss (a) and difference between two connection weights (b). As can be seen in the figure, two figures are quite similar to each other. Only difference is the magnitude of two measures. Table 5 shows the ranking of items by difference between two connection weights. As can be seen in the table, the items in the list is quite similar to those in information loss. This means that the information loss in this case is based upon difference between two connection weights.

No.	No.(Figure)	Strength	Item
1	4	3.639	Internet
2	25	3.630	Music
3	11	3.393	Computer
4	57	3.389	Travel
5	23	3.381	Movie
6	24	3.314	Visual media
7	12	3.269	Sport
8	56	3.236	Comic
9	45	3.229	Human relations
10	37	3.217	Qualification
49	14	2.347	Trading
50	46	2.327	Mathematics
51	35	2.299	Archeology
52	51	2.193	Statistics
53	52	2.083	Physics
54	26	2.078	Chemistry
55	49	2.015	Earth science
56	42	2.009	Craftwork
57	48	1.966	Shipping
58	50	1.918	Railroad

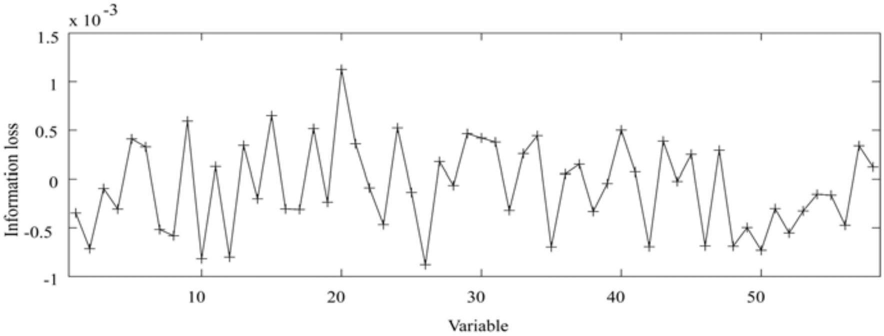
Table 2. Ranking of items for a group of students who responded to items with a low level of interest.

No.	No.(Figure)	Strength	Item
1	4	2.959	Internet
2	25	2.874	Music
3	12	2.717	Sport
4	23	2.687	Movie
5	56	2.599	Comic
6	11	2.580	Computer
7	10	2.519	Game
8	57	2.486	Travel
9	8	2.423	Entertainment
10	22	2.422	Eating and drinking
49	18	1.561	Marketing
50	26	1.552	Chemistry
51	29	1.549	Management sciences
52	21	1.539	Exchange
53	51	1.468	Statistics
54	42	1.440	Craftwork
55	52	1.421	Physics
56	49	1.421	Earth science
57	48	1.331	Shipping
58	50	1.321	Railroad

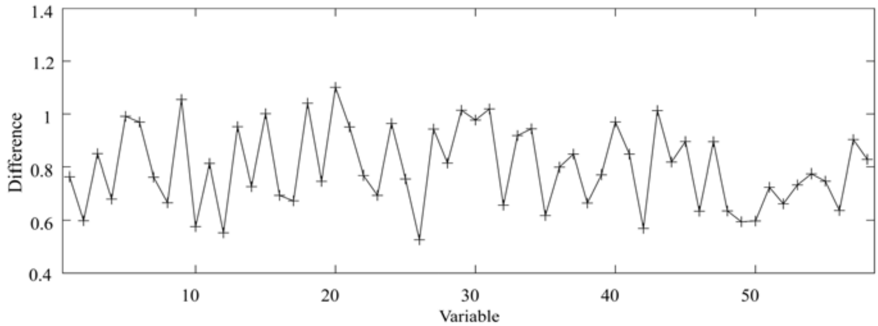
Table 3. Ranking of items for a group of students who responded to items with a low level of interest.

No.	No.(Figure)	Strength	Item
1	20	1.125	Multimedia
2	15	0.649	Buisiness
3	9	0.594	Creator
4	24	0.524	Visual Media
5	18	0.516	Marketing
6	40	0.501	Photograph
7	29	0.464	Business management
8	34	0.444	Publicity
9	30	0.420	Economics
10	5	0.410	Internet business
49	8	-0.581	Entertainment
50	46	-0.687	Mathematics
51	48	-0.690	Shipping
52	42	-0.695	Craftwork
53	35	-0.698	Archeology
54	2	-0.714	Animation
55	50	-0.732	Railroad
56	12	-0.802	Sport
57	10	-0.818	Game
58	26	-0.880	Chemistry

Table 4. Ranking of information loss for two groups analysis ($\times 10^{-3}$).



(a) Information loss



(b) Difference between two connection weights

Fig. 9. Information loss (a) and difference between two connection weights ($w2k - w1k$) (b).

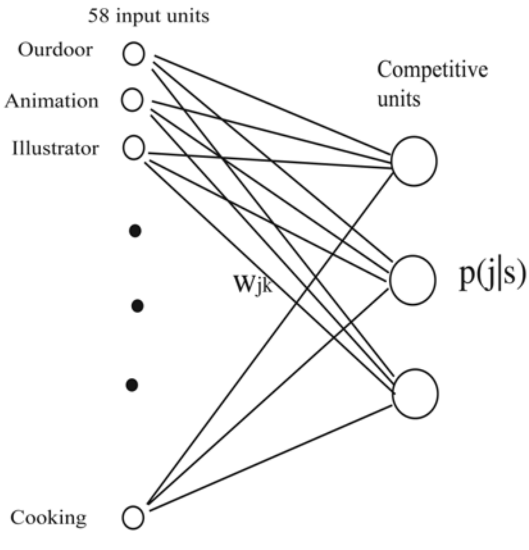


Fig. 10. Network architecture for three groups analysis.

No.	No.(Figure)	Strength	Item
1	20	1.100	Multimedia
2	9	1.055	Creator
3	18	1.040	Marketing
4	31	1.020	Arts
5	29	1.014	Business management
6	43	1.014	Information sciences
7	15	1.001	Business
8	5	0.991	Internet business
9	30	0.977	Economics
10	40	0.971	Photograph
49	48	0.636	Shipping
50	46	0.634	Mathematics
51	35	0.618	Archeology
52	2	0.598	Animation
53	50	0.597	Railroad
54	49	0.595	Earth science
55	10	0.575	Game
56	42	0.569	Craftwork
57	12	0.552	Sport
58	26	0.526	Chemistry

Table 5. Difference between two groups of students.

7.2 Three groups analysis

We increase the number of competitive units from two to three units as shown in Figure 10. Figure 11 shows connection weights for three groups. The third group detected at this time shows the lowest values of connection weights. The numbers of the first, the second and the third groups are 216, 341 and 23. Thus, the third group represents only a fraction of the data. Table 6 shows connection weights for students with strong interest in the items. Similar to a case with two groups, we can see that students have much interest in *entertainment*. Table 7 shows connection weights with moderate interest in the items. In the list, *qualification* and *human relations* disappear, and all the items expcet *computer* and *internet* are related to *entertainment*. Table 8 shows connection weights for the third group with low interest in the items. Though the scores are much lower than the other groups, this group also shows keen interest in *entertainment*. Table 9 shows conditional information losses for the first competitive unit. Table 10 shows information losses for the second competitive unit. Both tables show the same patterns of items in which business-related terms such as *economics*, *stock* show high values of information losses. Table 11shows a table of items for the third competitive units. Though the strength of information losses is small, more practical items such as *cooking* are detected.

7.3 Results by the principal component analysis

Figure 12 shows the contribution rates of principal components. As can be seen in the figure, the first principal component play a very important role in this case. Thus, we interpret the first principal component. Table 12 shows the ranking of items for the first principal component.

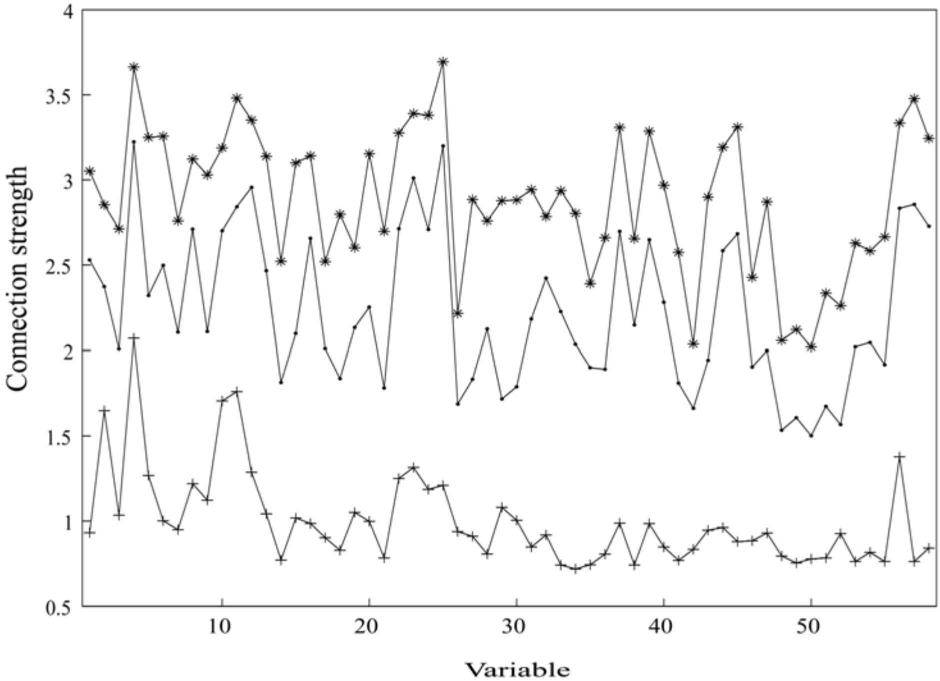


Fig. 11. Connection weights for three group analysis.

No.	No.(Figure)	Strength	Item
1	25	3.694	Music
2	4	3.663	Internet
3	11	3.481	Computer
4	57	3.478	Travel
5	23	3.390	Movie
6	24	3.379	Visual media
7	12	3.352	Sport
8	56	3.334	Comic
9	45	3.311	Human relations
10	37	3.309	Qualification
49	17	2.520	Volentier
50	46	2.427	Mathematics
51	35	2.391	Archeology
52	51	2.335	Statistics
53	52	2.261	Physics
54	26	2.216	Chemistry
55	49	2.122	Earth science
56	48	2.059	Shipping
57	42	2.037	Craftwork
58	50	2.019	Railroad

Table 6. Connection weights for students with strong interest in those items.

No.	No.(Figure)	Strength	Item
1	4	3.224	Internet
2	25	3.200	Music
3	23	3.012	Movie
4	12	2.956	Sport
5	57	2.856	Travel
6	11	2.843	Computer
7	56	2.834	Comic
8	58	2.727	Cooking
9	22	2.713	Eating and drinking
10	8	2.710	Entertainment
49	30	1.789	Economics
50	21	1.781	Exchange
51	29	1.717	Business management
52	26	1.687	Chemistry
53	51	1.674	Statistics
54	42	1.661	Craftwork
55	49	1.607	Earth science
56	52	1.566	Physics
57	48	1.533	Shipping
58	50	1.501	Railroad

Table 7. Connection weights for students with moderate interest in those items.

No.	No.(Figure)	Strength	Item
1	4	2.071	Internet
2	11	1.760	Computer
3	10	1.705	Game
4	2	1.648	Animation
5	56	1.377	Comic
6	23	1.314	Movie
7	12	1.286	Sport
8	5	1.267	Internet
9	22	1.250	Eating and drinking
10	8	1.219	Entertainment
49	14	0.770	Trading
50	41	0.769	Sociology
51	57	0.762	Travel
52	53	0.762	Literature
53	55	0.762	Law
54	49	0.754	Earth science
55	35	0.743	Archeology
56	33	0.741	Language
57	38	0.741	Bicycle
58	34	0.718	Publicity

Table 8. Connection weights for students with low interest in those items.

No.	No.(Figure)	Strength	Item
1	29	0.559	Business management
2	30	0.541	Economics
3	15	0.538	Business
4	20	0.510	Multimedia
5	27	0.445	Stock
6	18	0.400	Marketing
7	21	0.377	Exchange
8	9	0.334	Creator
9	43	0.320	Information sciences
10	47	0.314	Politics
49	56	-0.333	Comic
50	35	-0.353	Archeology
51	26	-0.355	Chemistry
52	4	-0.362	Internet business
53	42	-0.397	Craftwork
54	2	-0.422	Animation
55	8	-0.428	Entertainment
56	10	-0.429	Game
57	12	-0.435	Sport
58	23	-0.446	Movie

Table 9. Information loss No.1($\times 10^{-3}$).

No.	No.(Figure)	Strength	Item
1	30	0.335	Economics
2	15	0.317	Business
3	20	0.271	Multimedia
4	29	0.270	Business management
5	27	0.266	Stock
6	21	0.228	Exchange
7	18	0.227	Marketing
8	47	0.213	Politics
9	43	0.186	Information sciences
10	9	0.162	Creator
49	48	-0.226	Shipping
50	23	-0.229	Movie
51	50	-0.241	Railroad
52	26	-0.258	Chemistry
53	12	-0.271	Sport
54	42	-0.272	Craftwork
55	8	-0.289	Entertainment
56	10	-0.329	Game
57	2	-0.336	Animation
58	4	-0.392	Internet

Table 10. Information loss No.2($\times 10^{-3}$).

No.	No.(Figure)	Strength	Item
1	57	0.384	Travel
2	45	0.258	Human relations
3	34	0.249	Publicity
4	58	0.235	Cooking
5	33	0.211	Language
6	25	0.192	Music
7	44	0.176	Psychology
8	37	0.175	Qualification
9	40	0.175	Photograph
10	20	0.161	Multimedia
49	52	-0.218	Physics
50	3	-0.223	Illustrator
51	56	-0.234	Comic
52	26	-0.243	Chemistry
53	42	-0.268	Craftwork
54	29	-0.382	Business management
55	11	-0.447	Computer
56	4	-0.466	Internet business
57	10	-0.493	Game
58	2	-0.569	Animation

Table 11. Information loss No.3($\times 10^{-3}$).

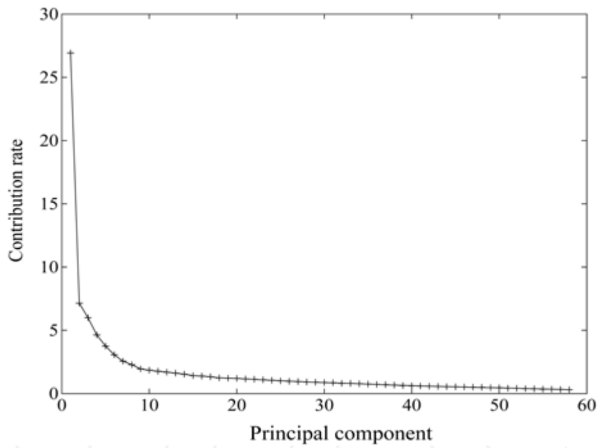


Fig. 12. Contribution rates for 58 variables.

The ranking seems to be quite similar to that obtained by the information loss. This means that the principal component seems to represent the main features by which different groups can be separated. On the other hand, connection weights by forced information represent the absolute magnitude of students' interest in the subjects. In forced-information maximization, we can see information loss as well as connection weights. The connection weights represent the absolute value of the importance. On the other hand, the information loss represents difference between several groups. This is a kind of relative importance of variables, because the importance of a variable in one group is measured in a relation to the other group.

No.	No.(Figure)	Strength	Item
1	20	0.1589	Multimedia
2	30	0.1539	Economics
3	15	0.1537	Business
4	29	0.1536	Business management
5	27	0.1514	Stock
6	18	0.1512	Marketing
7	34	0.1492	Publicity
8	47	0.1492	Politics
9	31	0.1482	Arts
10	33	0.1477	Language
49	52	0.1150	Physics
50	8	0.1115	Entertainment
51	49	0.1094	Illustrator
52	48	0.1085	Shipping
53	10	0.1070	Game
54	4	0.1049	Internet
55	42	0.1028	Craftwork
56	50	0.0986	Railroad
57	26	0.0981	Chemistry
58	2	0.0922	Animation

Table 12. The first principal component.

8. Conclusion

In this paper, we have proposed a new computational method to accelerate a process of information maximization. Information-theoretic competitive learning has been introduced to solve the fundamental problems of conventional competitive learning such as the dead neuron problem, dependency on initial conditions and so on. Though information theoretic competitive learning has demonstrated much better performance in solving these problems, we have observed that sometimes learning is very slow, especially when problems become very complex. To overcome this slow convergence, we have introduced forced information maximization. In this method, information is supposed to be maximized before learning. By using the WTA algorithm, we have introduced forced information in information-theoretic competitive learning. We have applied the method to several problems. In all problems, we have seen that learning is much accelerated, and for the student survey case, networks converge more than seventy times faster. Though we need to explore the exact mechanism of forced information maximization, the computational method proposed in this paper enables information theoretic learning to be applied to more large-scale problems.

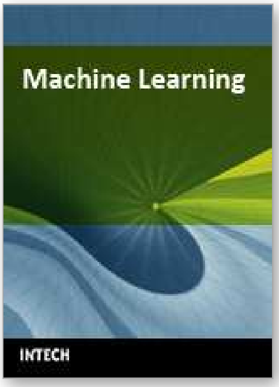
9. Acknowledgment

The author is very grateful to Mitali Das for her valuable comments.

10. References

[1] R. Kamimura, T. Kamimura, and O. Uchida, "Flexible feature discovery and structural information," *Connection Science*, vol. 13, no. 4, pp. 323–347, 2001.

[2] R. Kamimura, T. Kamimura, and H. Takeuchi, "Greedy information acquisition algorithm: A new information theoretic approach to dynamic information acquisition in neural networks," *Connection Science*, vol. 14, no. 2, pp. 137–162, 2002.
[3] R. Kamimura, "Information theoretic competitive learning in self-adaptive multi-layered networks," *Connection Science*, vol. 13, no. 4, pp. 323–347, 2003.
[4] R. Kamimura, "Information-theoretic competitive learning with inverse euclidean distance," *Neural Processing Letters*, vol. 18, pp. 163–184, 2003.
[5] R. Kamimura, "Unifying cost and information in information-theoretic competitive learning," *Neural Networks*, vol. 18, pp. 711–718, 2006.
[6] D. E. Rumelhart and D. Zipser, "Feature discovery by competitive learning," in *Parallel Distributed Processing* (D. E. Rumelhart and G. E. H. et al., eds.), vol. 1, pp. 151–193, Cambridge: MIT Press, 1986.
[7] S. Grossberg, "Competitive learning: from interactive activation to adaptive resonance," *Cognitive Science*, vol. 11, pp. 23–63, 1987.
[8] D. DeSieno, "Adding a conscience to competitive learning," in *Proceedings of IEEE International Conference on Neural Networks*, (San Diego), pp. 117–124, IEEE, 1988.
[9] S. C. Ahalt, A. K. Krishnamurthy, P. Chen, and D. E. Melton, "Competitive learning algorithms for vector quantization," *Neural Networks*, vol. 3, pp. 277–290, 1990.
[10] L. Xu, "Rival penalized competitive learning for clustering analysis, RBF net, and curve detection," *IEEE Transaction on Neural Networks*, vol. 4, no. 4, pp. 636–649, 1993.
[11] A. Luk and S. Lien, "Properties of the generalized lotto-type competitive learning," in *Proceedings of International conference on neural information processing*, (San Mateo: CA), pp. 1180–1185, Morgan Kaufmann Publishers, 2000.
[12] M. M. V. Hulle, "The formation of topographic maps that maximize the average mutual information of the output responses to noiseless input signals," *Neural Computation*, vol. 9, no. 3, pp. 595–606, 1997.
[13] L. L. Gatlin, *Information Theory and Living Systems*. Columbia University Press, 1972.



Machine Learning
Edited by Abdelhamid Mellouk and Abdennacer Chebira

ISBN 978-953-7619-56-1
Hard cover, 450 pages
Publisher InTech
Published online 01, January, 2009
Published in print edition January, 2009

Machine Learning can be defined in various ways related to a scientific domain concerned with the design and development of theoretical and implementation tools that allow building systems with some Human Like intelligent behavior. Machine learning addresses more specifically the ability to improve automatically through experience.