Method for Improving the Reliability of Learning Machine†

Toshio NAKAGAWA

Shionogi Research Laboratory, Shionogi & Co., Ltd.,
Fukushima-ku, Osaka 553

Received June 3, 1974

Optimization of learning machine prediction has been attempted by the addition of a "width" parameter proposed by Wangen *et al.* Predictions were made for the production of phenazine compounds, the utilizations of D-sorbitol, of *n*-hexadecane, and of histamine by pseudomonads. In all but one cases examined, the "stability" of prediction was much improved when a high positive value was assigned to the parameter. A slight increase was often found in the hit-to-miss ratio, but the χ^2 test revealed that it was statistically insignificant.

At applying a learning machine, everyone desires to get 1) an optimum and 2) stable result 3) based upon data as few as possible. In previous works of this series, a standard method¹⁾ has been applied to actual cases²⁾ accompanied with hit-to-miss ratios higher than those given by experienced microbiologists,³⁾ and then a procedure has been worked out for minimizing the number of data that should be collected prior to the prediction.⁴⁾ The present paper describes a procedure by which the "stability" or "constancy" of prediction is much improved. A slight increase found often in the hit-to-miss ratio has turned out to be statistically insignificant.

PROCEDURE

The learning-machine prediction is based on the assumption that the presence or absence of a certain feature (feature in question) in a microorganism is decided by the value of S which is determined from the presence or absence of n other features of the same organism.

$$S = w_1x_1 + w_2x_2 + \cdots + w_nx_n + w_0x_0 \quad (1)$$

where x_i ($i=1\sim n$): the number denoting the presence or absence of the i th feature (1 for presence, 0 for absence), w_i ($i=1\sim n$): the weight assigned to the i th feature, w_0 : a constant, $x_0\equiv 1$.

Provided the $n+1$ weights, w_i 's ($i=0\sim n$), are known, one can calculate S for any microorganism, for which the n features are known. The feature in

question is present when $S\geq 0$, and absent when $S<0$. The weights w_i 's can be determined with the calculation scheme shown in Fig. 1, if we have a large number of microorganisms, for each of which the presence or absence of every feature (the feature in question and the other n features) is known.

In the standard method of learning machine, δ is equated to zero, and therefore, the part enclosed with the broken line is omitted. When the "requirement"— $S\geq 0$ if the feature in question is present, and, $S<0$ if the feature in question is absent—is not satisfied, corrections must be applied to the temporary w_i 's appearing on the stage. The rules of correction are: a) When $S>0$ whereas the feature in question is absent, or, $S<0$ whereas the feature in question is present, then calculate $\text{Cor}=2S/(\sum_{i=1}^n x_i)$ and subtract Cor from w_i with respect to every i for which $x_i=1$. b) When $S=0$ whereas the feature in question is absent, then decrease w_0 by one unit. Some examples have been shown in Part I of this series of works.¹⁾

In the present work, the "requirement" is modified into a more general expression as described in the legend of Fig. 1. The rules of correction are thence altered as follows: a) When $S>-\delta$ whereas the feature in question is absent, then calculate $\text{Cor}=2(S+\delta)/(\sum_{i=1}^n x_i)$, and, when $S<\delta$ whereas the feature in question is present, then calculate $\text{Cor}=2(S-\delta)/(\sum_{i=1}^n x_i)$; and in both cases, subtract Cor from w_i with respect to every i for which $x_i=1$. b) When $S=-\delta$ whereas the feature in question is absent, then decrease w_0 by one unit.

Wangen *et al.* have claimed that such a modification in determining w_i values will improve the hit-to-miss ratio of prediction compared with the standard method.⁵⁾ One of the purposes of the present work is to see how the hit-to-miss ratio varies with an increase

† Prediction of Microbial Characters by a Computerized Learning Machine, Part V. See reference 4).

FIG. 1. Flow of Calculation Procedure in a Computer.

* $S \geq \delta$ if the feature in question is present, and, $S < -\delta$ if the feature in question is absent.

** See text for the method of correction.

in δ value and thus to decide whether Wangen's claim is correct or not. The scheme shown in Fig. 1 gives rise to different sets of final weights not only by the variation of δ but also by the variations in the arrangement of microorganisms and/or in the provisional w_i 's set at the start of calculation, and therefore, some fluctuation is brought about in the prediction result by these latter variations too. Another purpose of the work is to demonstrate that the assignment of a large positive value to δ diminishes the fluctuation thus bringing increased stability in the prediction.

RESULTS AND DISCUSSION

Likewise as in previous works,^{2,4)} 214 strains of pseudomonads were picked up from a publication of Stanier *et al.*,⁶⁾ for which the presence or absence of 94 features was known. These

strains were divided into two groups of equal size, A and B. For all 107 strains of B group, the presence or absence of each of the four features—the production of phenazine compound and the utilizations of D-sorbitol, of *n*-hexadecane, and of histamine—was “predicted” by using the 94 weights (w_0 and 93 weights corresponding to 93 features other than the feature to be predicted) obtained by treating the data of A group with the above-mentioned procedure. Microorganisms of A group were arranged in two different orders, “classification order” and “random order,” and furthermore, two provisional values, +1,000,000 and −1,000,000, were assigned to each w_i at the start of calculation. In this way, four sets of weights were obtained for an individual value of δ . As for

TABLE I. UTILIZATION OF *n*-HEXADECANE BY B GROUP MICROORGANISMS
(RANDOM, +1,000,000)

Mispredicted members at

 $\delta=0$: 2, 5, 9, 12, 13, 14, 102, 104, 106, 107 $\delta=12,000,000$: 2, 5, 9, 13, 106, 107

δ	Nr. of cycles	$\delta/\text{largest } S $	NE	PA	EAF	EPF
0	29		10	.91	1.68	1.53
1,000,000	31	0.028	7	.93	1.78	1.64
2,000,000	38	0.057	7	.93	1.78	1.64
4,000,000	44	0.094	6	.94	1.86	1.68
8,000,000	49	0.104	6	.94	1.86	1.68
12,000,000	50	0.107	6	.94	1.86	1.68

TABLE II. UTILIZATION OF D-SORBITOL BY B GROUP MICROORGANISMS
(RANDOM, -1,000,000)

Mispredicted members at

 $\delta=$: 41, 43, 49, 52, 72, 77 $\delta=12,000,000$: 27, 37, 39, 41, 43, 72, 77

δ	Nr. of cycles	$\delta/\text{largest } S $	NE	PA	EAF	EPF
0	8		6	.94	1.89	1.87
1,000,000	11	0.051	3	.97	1.92	1.96
2,000,000	17	0.087	5	.95	1.89	1.90
4,000,000	34	0.110	7	.93	1.85	1.86
8,000,000	43	0.124	7	.93	1.85	1.86
12,000,000	46	0.124	7	.93	1.85	1.86

the members of A group, the prediction was made by exchanging the roles of A and B.

Table I shows an example of prediction results, in which Wangen's clasim is satisfied. The prediction was made for the utilization of *n*-hexadecane by microorganisms belonging to B group. The weights were obtained by arranging the members of A group in the random order and by assigning +1,000,000 as the starting weight. The results was evaluated on four measures; NE=the number of erred (miss) predictions among total 107 predictions, PA*=the number of correct predictions/107, EAF=evaluation based on actual feature, EPF=evaluation based on predicted feature. In the last three measures, a larger value means a better result as detailed in reference 1).

In this rather special example, an augmentation in δ , a) postulates a larger number of calculation cycles, which is roughly pro-

* hit-to-miss ratio = $(107 - \text{NE})/\text{NE} = \text{PA}/(1 - \text{PA})$.

portional to the calculation time, b) increases the ratio of " $\delta/\text{the largest absolute value of } S$," which seems to converge to a definite value, and c) brings a better result. Although the tendencies a) and b) are generally true, c) is not so, as recognized from Table II, where it should also be noted that the mispredicted members are considerably different between $\delta=0$ and $\delta=12,000,000$.

Table III summarizes the numbers of erred predictions at these two δ 's in all systems examined. Two characteristics may be pointed out. 1) The number of erred predictions does not always decrease with an increase in δ . Although the grand total of erred predictions decreased from 347 to 335 in line with the increase of δ from zero to twelve millions, the value of χ^2 revealed that the decrease was statistically insignificant. 2) The dispersion of the number of erred predictions is smaller at $\delta=12,000,000$ than at $\delta=0$; a larger δ seems to

TABLE III. NUMBER OF ERRED PREDICTIONS (before excluding noise features)

Weights are obtained from			Production of phenazine compounds		Utilization of D-sorbitol		Utilization of <i>n</i> -hexadecane		Utilization of histamine ^{a)}	
			$\delta=0$	12×10^6	$\delta=0$	12×10^6	$\delta=0$	12×10^6	$\delta=0$	12×10^6
A	Classn.	+	8	10	5	6	8	8	18	19
A	Classn.	—	9	9	4	6	5	7	14	19
A	Random	+	8	9	6	7	10 ^{a)}	6 ^{a)}	16	17
A	Random	—	6	9	6 ^{b)}	7 ^{b)}	7	7	13	17
B	Classn.	+	9	9	9	6	8	8	19	16
B	Classn.	—	10	10	14	6	11	8	22	16
B	Random	+	9	12	7	5	14	11	18	19
B	Random	—	11	11	6	5	11	11	26	19

Grand total of erred predictions: 347 at $\delta=0$, 335 at $\delta=12,000,000$.

$\chi^2=0.23$

^{a)} cf. Table I.

^{b)} cf. Table II.

^{c)} Because of the small computer available, the prediction was made previously by using the weights obtained at a stage of imperfect recognition.^{1,2)} Calculation was continued in the present work until perfect recognition was attained with a larger computer, FACOM 270-20.

TABLE IV. NUMBER OF MICROORGANISMS PREDICTED DIFFERENTLY BY FOUR SETS OF WEIGHTS (before excluding noise features)

Group predicted		Production of phenazine compounds		Utilization of D-sorbitol		Utilization of <i>n</i> -hexadecane		Utilization of histamine ^{a)}	
		$\delta=0$	12×10^6	$\delta=0$	12×10^6	$\delta=0$	12×10^6	$\delta=0$	12×10^6
B		7	1	11	1	12	2	17	2
(χ^2)		(4.67*)		(8.83**)		(7.64**)		(13.00**)	
A		11	9	22	1	14	3	21	3
(χ^2)		(0.22)		(21.48**)		(7.73**)		(15.21**)	

Grand total of differently predicted microorganisms: 115 at $\delta=0$, 22 at $\delta=12,000,000$.

$\chi^2=68.62^{**}$.

* significant (95)%, ** highly significant (99%).

^{a)} cf. footnote ^{c)} in Table III.

TABLE V. NUMBER OF ERRED PREDICTIONS (after excluding noise features)

Weights are obtained from			Production of phenazine compounds		Utilization of D-sorbitol		Utilization of <i>n</i> -hexadecane	
			NF	$\delta=0$ 6×10^6	NF	$\delta=0$ 6×10^6	NF	$\delta=0$ 6×10^6
A	Classn.	+		12 10		4 5		6 7
A	Classn.	—		11 10		6 6		7 7
			21		20		11	
A	Random	+		11 11		5 8		7 7
A	Random	—		11 11		8 8		7 7
B	Classn.	+		10 9		6 5		10 8
B	Classn.	—		9 9		8 5		11 8
			26		18		17	
B	Random	+		7 9		5 7		10 9
B	Random	—		10 9		7 7		9 9

NF; Number of features remained ultimately.

Grand total of erred predictions: 197 at $\delta=0$, 191 at $\delta=6,000,000$.

$\chi^2=0.10$.

TABLE VI. NUMBER OF MICROORGANISMS PREDICTED DIFFERENTLY BY FOUR SETS OF WEIGHTS (after excluding noise features)

Group predicted	Production of Phenazine compounds			Utilization of D-sorbitol			Utilization of <i>n</i> -hexadecane		
	NF	$\delta=0$	6×10^6	NF	$\delta=0$	6×10^6	NF	$\delta=0$	6×10^6
B	21	3	3	20	8	3	11	4	0
(χ^2)		(0.00)			(2.40)			(4.08*)	
A	26	5	0	18	12	2	17	5	1
(χ^2)		(5.12*)			(7.64**)			(2.74)	

NF; Number of features remained ultimately.

Grand total of differently predicted microorganisms: 37 at $\delta=0$, 9 at $\delta=6,000,000$. $\chi^2=17.68$.**

* significant (95%), ** highly significant (99%).

increase the stability of prediction. The latter characteristic is confirmed by Table IV that lists up the numbers of strains for which the four sets of weights gave different predictions (some said "present" while the others said "absent"). It may be safely concluded that different arrangements of strains and/or different sets of starting weights do not affect the prediction notably if a large δ is used. The χ^2 test certified highly significant differences between the results obtained from the two δ values.

Exclusion of "noise" features enables one to

perform the prediction based on a decreased number of features.⁴⁾ Even in this case, the above-mentioned conclusions hold true as deduced from Tables V and VI.

REFERENCES

- 1) T. Nakagawa, *Agr. Biol. Chem.*, **35**, 1880 (1971).
- 2) E. Masuo and T. Nakagawa, *ibid.*, **35**, 1885 (1971).
- 3) E. Masuo and T. Nakagawa, *ibid.*, **36**, 1433 (1972).
- 4) T. Nakagawa and E. Masuo, *ibid.*, **36**, 2459 (1972).
- 5) L. E. Wangen, N. M. Frew and T. L. Isenhour, *Anal. Chem.*, **43**, 845 (1971).
- 6) R. Y. Stanier, N. M. Palleroni and M. Doudoroff, *J. Gen. Microbiol.*, **43**, 159 (1966).