論　　文

Short-term Prediction of a Test Score Using the BP Method with Takens' Embedding Theorem in a Neural Network

ターケンスの埋め込み定理を組み込んだニューラルネットワークのBP法を用いたテストスコアの短期予測

次世代教育学部教育経営学科
木戸　和彦
KIDO, Kazuhiko
Department of Management for Education
Faculty of Education for Future Generations

早稲田大学基幹理工学部
福田　衣里
FUKUDA, Eri
Waseda University
School of Fundamental Science and Engineering

電気通信大学情報理工学部
橋本　信一
HASHIMOTO, Shinichi
The University of Electro-Communications
Faculty of Informatics and Engineering

秋田県立大学総合科学教育研究センター
岡崎　弘信
OKAZAKI, Hironobu
Akita Prefectural University
Research and Education Center

Abstract：At the fifty-first national conference of the Japan Association for Language Education and Technology (LET), we reported the possibility of using a neural network to make short-term predictions of language test scores if they are in a chaotic time-series. At that time, one possible area for further development was increasing the accuracy of the prediction. In our presentation, we will discuss the utility of improving accuracy by applying Takens' theorem to the short-term chaotic time-series prediction using the well-known Backpropagation Method.

Keywords：Neural Network, Short-term Prediction, Takens' Embedding Theorem, Test Score, e-Learning

1．Introduction

The Central Council for Education Board called for maintenance and improvement of the quality of education, emphasizing the idea of learning outcomes in their 2008 proposal, *Towards the building of undergraduate education*. Learning outcomes are statements that describe what learners are expected to know, understand, practice, and perform upon completion of study. When considering them in the context of language education, without a sound method of predicting results of language test scores in the short term, learning outcomes as a result of year-long education cannot be guaranteed. Therefore, in order to establish a guideline for the long-term goals of education, we examined whether it is possible to predict language test scores by a two-dimensional, chaotic time-series short term prediction method. The results were presented at the fifty-first national convention of Language Education and Technology held in 2011. Also, these results were published in *Short-term chaotic time-series prediction of language test scores adopting the backpropagation algorithm* (Kido et al, 2016) [1]. Nevertheless, there was a problem found in the study: refinement of prediction accuracy.

To address this issue, the backpropagation (BP) method was adopted to examine its effectiveness in predicting the scores of the sixth test based on the scores of the previous five tests.

In the current study, to further improve the accuracy of prediction, Taken's embedding theorem was incorporated, which was suggested in *Study on chaotic time-series short prediction: Is it possible* *to predict typhoon movements*? (Kido, 2002) [2] and *Short-term prediction of a test score using the BP method with Takens' embedding theorem in a neural network* (Kido, 1999) [3]. The method, results, effectiveness, and necessity will be discussed from a pedagogical point of view.

2. Literature review

In recent years, progress in language testing research has been remarkable, and statistical analysis is now a norm in the field. Many of the studies focus on test validity or the interpretation of the data (Shimizu, 2005 [5]; Shimizu et al., 2003 [4]). Few of them explore test score prediction. Meanwhile, outside the field of testing, short-term, chaotic time-series method using a neural network has been examined in various studies. For example, Ogasawara et al. (2009) [6] used it in bio-information research, Sakai et al. (2009) [7] for predicting the variation of stock prices. In addition, the authors of the current study described the method in the study on the prediction of typhoon movements (Kido, 2002) [2].

The purpose of this article is to develop a method for predicting language test scores.

3. Conventional prediction methods and the newly adopted method

Gram-Schmidt orthogonalization and tessellation are a few of the examples of short-term prediction methods. Many short-term prediction methods often fail to predict due to how the parameters are decided, or calculation time drastically increases if there are too many dimensions. In the field of linguistics, for instance, Watanabe (1983) attempted to predict test scores using regression analysis; however, the participants' test scores were treated as a batch of data. In other words, a method to predict an individual learner's test score has not yet been established. Furthermore, another difficulty with prediction is that a large amount of data which contains a number of variables is required.

Another method of prediction is to utilize the characteristics of time-series data. Namely, a deterministic prediction method which only uses past time-series data for analysis (Sugihara et al., 1990) [11]. In this paper, a neural network was adopted for two reasons in particular. First, it causes fewer problems and secondly, it is commonly used in predictive studies.

3.1 Three-layer hierarchical neural network and its validity

In this section, we describe the structure of the neural network used in the current study. We also explain reasons for using a three-layer neural network and its validity. In Figure 1, *i* indicates an input value for the input layer; *h* indicates an output value for the hidden layer (= an input value for the output layer); *o* indicates an output value for the output layer; *w* indicates a combined load between the input and hidden layers; *v* indicates a combined load between the hidden and the output layers.

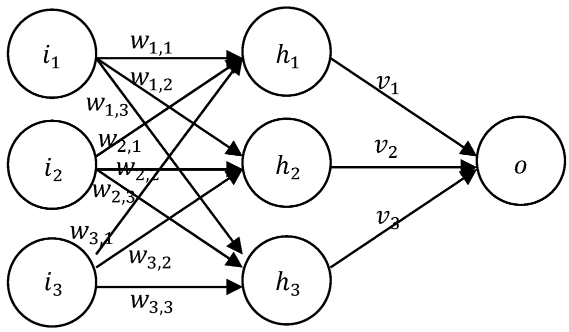

Figure 1. A Three-Layer Neural Network

This neural network contains three nodes in the input and the hidden layers respectively and one node in the output layer: a three-layer hierarchical neural network. The input and the output values for each layer are as follows:

(Input value for the input layer)

$$i_k = x_k \qquad (k = 1,2,3) \qquad \cdots\cdots\cdots \ (1)$$

(Output value for the hidden layer)

$$h_k = f\left(\bar{h}_k\right)$$

$$\bar{h}_k = \sum_{k=1}^{3} i_k w_{k,j} \qquad (k,j = 1,2,3) \qquad \cdots\cdots\cdots \ (2)$$

(Output value for the output layer)

$$o = f(\bar{o})$$

$$\bar{o} = \sum_{k=1}^{3} h_k v_k \qquad (k = 1,2,3) \qquad \cdots\cdots\cdots \text{(3)}$$

The function f indicates the sigmoid function, and when it is one-dimensional, it is defined by Formula 4.

(Sigmoid function)

$$f(y) = \frac{1}{1 + e^{-y}} \qquad \cdots\cdots\cdots \text{(4)}$$

(Error function)

Moreover, the error function for training is shown as Formula 5, and t indicates a teacher.

$$E = \frac{1}{2} \sum_{k=1}^{3} (t_k - o_k)^2 \quad (k = 1,2,3) \qquad \cdots\cdots\cdots \text{(5)}$$

Since its establishment, the hierarchical neural network has been applied in a variety of fields. One well-known difficulty related to this system, is deciding the number of nodes in the hidden layer. If there are not enough number of nodes in the hidden layer, training does not end; whereas, if there are too many, training speed and versatility might diminish. Nevertheless, it is difficult to decide the optimal number of nodes at the initial stage. Thus, in the current study, the optimal number of nodes was determined based on the method employed in Ataka et al. (2005) [12]. Utilizing this method, we identified that the optimal number of hidden layers in this study was three.

３．２　The short-term prediction method using the BP method for a neural network

BP method is one of the training methods for the hierarchical neural network. In this method, a perceptron, an architecture of a neural network connecting the input and the hidden layers as well as the hidden and output layers, is extended. By providing the output layer with a correct answer as a teacher, an error between the teacher and the output value is obtained to modify the weight. Since this process proceeds backwards, it is called the backpropagation method. In addition, a steepest descend method is used as a training algorithm. The following is an example of the application of the method.

We will predict x_i to x_{11} in a set of time-series data defined as $x_i\ (i = 1,2,\cdots,10)$. Firstly, we train the neural network indicated in Figure 1 with the following pattern of data (See Figure 2).

	Input value		Teacher
Case 1：	x_1, x_2, x_3	→	x_4
Case 2：	x_2, x_3, x_4	→	x_5
⋮	⋮		⋮
Case 7：	x_7, x_8, x_9	→	x_{10}

Figure ２． Learning Process

After all the patterns are learned, the combined load is recorded. Then as a predictive process, the same neural network calculates once with the input values x_8, x_9, x_{10} and the recorded combined load. The obtained result is defined as the predictive value x_{11} (See Figure 3).

This is the short-term prediction method using the BP method for a neural network.

	Input value		Predicted value
Prediction process：	x_8, x_9, x_{10}	→	x_{11}

Figure 3. Predicting Process

３．３　The short-term prediction method using the BP method incorporating Taken's embedding theorem

Firstly, Taken's embedding theorem is applied to the original time-series data and plugged in into the three-dimensional state space reconstruction. Then the neural network (See Figure 1) learns the following pattern (in the case where the number of input nodes is three) (See Figure 4)

	Input value		Teacher
Case 1：	$[x_1, x_2, x_3], [x_2, x_3, x_4], [x_3, x_4, x_5]$	→	x_6
Case 2：	$[x_2, x_3, x_4], [x_3, x_4, x_5], [x_4, x_5, x_6]$	→	x_7
⋮	⋮		⋮
Case 5：	$[x_5, x_6, x_7], [x_6, x_7, x_8], [x_7, x_8, x_9]$	→	x_{10}

Figure ４． Learning Process

After all the patterns are learned, the combined load is recorded. Then as a predictive process, the same neural network calculates once with the input values $[x_6, x_7, x_8], [x_7, x_8, x_9], [x_8, x_9, x_{10}]$ and the recorded combined load. The obtained result is defined as the predictive value x_{11} (See Figure 5).

	Input value	Predicted value
Prediction process：	$[x_6, x_7, x_8], [x_7, x_8, x_9], [x_8, x_9, x_{10}]$ →	x_{11}

Figure 5. Predicting Process

This is the short-term prediction method using the BP method incorporating Taken's embedding theorem.

3.4 Criteria for confirming the chaotic time-series data

The premise for the current study is that the original time-series data exhibits chaotic behaviors. There are various criteria to determine whether or not the original time-series data exhibits chaotic behavior. Among them, some of the most common criteria are the largest Lyapunov exponent and the correlation function, and these criteria were used in this study.

4. Application of the method to the pedagogical data and the results of simulation

In this section, we will describe the real data to which the aforementioned methods were applied. The set of data used in this study was the scores of six English listening tests (100 points maximum) collected from 28 students registered in a one-year course in 2006 (See Table 1). Firstly, the data was examined with the following formula to judge whether it demonstrates chaotic behaviors.

4.1 The calculation result of largest Lyapunov exponent

Table 2 shows that all the largest Lyapunov

Table 1. English Listening Test Scores

Student	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
1	64.3	61.4	75.4	59.9	65.0	68.4
2	53.2	56.3	49.8	45.8	61.7	53.2
3	56.1	54.0	49.8	52.8	57.5	53.8
4	45.0	44.7	45.3	51.9	57.9	59.1
5	48.0	38.6	36.0	33.5	59.3	53.8
6	69.6	62.8	56.2	57.5	71.5	69.6
7	42.7	47.9	34.0	30.2	49.5	40.4
8	52.0	58.1	53.2	51.9	69.2	67.3
9	48.5	53.0	45.3	33.5	61.7	45.0
10	49.1	61.9	51.2	51.9	66.8	60.2
11	63.2	57.2	60.1	60.4	68.2	67.3
12	54.4	56.7	65.5	56.1	69.6	68.4
13	64.9	66.0	71.9	63.7	72.4	69.6
14	69.6	70.2	75.4	74.5	79.4	80.1
15	61.4	58.1	49.8	57.5	63.6	64.9
16	49.1	52.1	44.3	43.9	50.9	56.7
17	53.2	44.7	52.7	52.4	61.2	59.6
18	59.6	59.5	55.7	58.0	71.0	69.6
19	55.6	64.7	61.6	65.1	72.9	70.8
20	59.6	52.6	51.7	52.8	59.3	59.6
21	48.5	44.7	46.8	36.8	44.4	50.3
22	37.4	42.8	44.3	44.3	60.7	56.7
23	50.9	50.7	50.7	49.5	60.7	57.9
24	52.6	51.2	52.2	51.9	52.3	53.2
25	46.8	56.7	48.3	49.1	52.8	54.4
26	48.0	50.7	47.8	46.7	55.1	48.5
27	38.6	44.7	46.3	41.0	52.8	62.6
28	69.6	67.9	65.0	55.2	72.0	76.0
Mean	54.0	54.6	53.1	51.0	62.1	60.6

exponents except that of Student 22 are positive values. This result suggested that the test scores used in this study exhibited chaotic behaviors.

4.2 The calculation result of the correlation function

Based on the calculation result of the correlational function (See Table 3), we identified that the 85%

Table 2. The Largest Lyapunov Exponent

Student	The largest Lyapunov exponent	Student	The largest Lyapunov exponent
1	0.526777	16	0.48986
2	0.473776	17	0.324598
3	0.540556	18	1.090894
4	0.880534	19	0.341931
5	0.290761	20	0.356958
6	0.38764	21	0.430501
7	0.341695	22	−0.00674
8	0.400145	23	0.646066
9	0.383764	24	0.606264
10	0.258038	25	0.283972
11	0.4029	26	0.481385
12	0.565411	27	0.388079
13	0.691242	28	0.633351
14	0.815682	Mean	0.482947
15	0.496487		

Table 3. Correlation Function Values

Student	Correlation Function Values	Range (−1.0∼1.0)	Range (−2.0∼2.0)
1	1.2		○
2	1.1		○
3	2.1		
4	0.7	○	
5	1.1		○
6	1.8		○
7	0.6	○	
8	0.6	○	
9	0.7	○	
10	0.0	○	
11	2.9		
12	1.0	○	
13	2.0		○
14	1.3		○
15	1.1		○
16	0.7	○	
17	1.8		○
18	1.3		○
19	0.2	○	
20	2.9		
21	1.8		○
22	0.4	○	
23	1.6		○
24	18.1		
25	1.1		○
26	1.2		○
27	0.4	○	
28	1.8		○
Frequency		35.7%	85.7%

of the language test scores demonstrated chaotic behaviors.

5. Comparison of the simulation results

Considering the calculation results of the largest Lyapunov exponents (See Table 3) and the correlational function (Table 1), we confirmed that the data in Table 1 exhibits chaotic behaviors. Accordingly, we simulated the two prediction methods: the BP method for a neural network and the BP method incorporating Taken's embedding theorem. Then the results were compared.

5.1 The result of the short-term prediction by the BP method

Firstly, the result of the predicted test scores of Test 6 is shown in Table 4, which was calculated with the scores of Test 1 to 5 as input values.

Next, the significance of the results was analyzed by single regression analysis on the measured and predicted values.

Figure 6 shows the correlation coefficient between the measured and the predicted values and the regression line.

The correlation coefficient between the measured and predicted values was r = 0.8196. Considering the common statistical criteria (limit value according to the number of samples) (KogoLab, 2004), a strong correlation was found between the measured and predicted values.

Table 4. Comparison of Test Scores between Predicted Values and Measurement Records

Student	Measured value of test 6	Predicted value of test 6
1	68.4	65.1
2	53.2	62.1
3	53.8	57.6
4	59.1	58.0
5	53.8	59.1
6	69.6	71.2
7	40.4	49.7
8	67.3	69.8
9	45.0	63.1
10	60.2	67.1
11	67.3	68.3
12	68.4	69.2
13	69.6	72.7
14	80.1	79.4
15	64.9	63.3
16	56.7	51.0
17	59.6	61.3
18	69.6	71.3
19	70.8	72.9
20	59.6	59.0
21	50.3	43.5
22	56.7	61.0
23	57.9	60.8
24	53.2	52.3
25	54.4	52.9
26	48.5	55.2
27	62.6	53.5
28	76.0	72.6
Mean	60.6	62.3

Table 5. The result of the BP method incorporating the Taken's embedding theorem

Student	Measured value of test 6	Predicted value of test 6
1	68.4	68.7
2	53.2	53.3
3	53.8	57.8
4	59.1	61.7
5	53.8	58.8
6	69.6	68.0
7	40.4	45.8
8	67.3	66.2
9	45.0	61.2
10	60.2	63.4
11	67.3	67.7
12	68.4	67.8
13	69.6	69.0
14	80.1	79.9
15	64.9	64.4
16	56.7	50.8
17	59.6	60.3
18	69.6	70.1
19	70.8	73.5
20	59.6	62.0
21	50.3	44.6
22	56.7	58.5
23	57.9	60.8
24	53.2	55.6
25	54.4	54.3
26	48.5	53.9
27	62.6	54.7
28	76.0	72.6
Mean	60.6	61.6

5.2 The result of the BP method incorporating Taken's embedding theorem

Next, the prediction result obtained from the BP method incorporating Taken's embedding theorem is shown in Table 5. The significance of the result was analyzed by single regression analysis on the measured and predicted values.

Figure 7 demonstrates the correlation coefficient between the measured and predicted values and the regression line.

The correlation coefficient between the measured and predicted values was r = 0.8410. Considering the common statistical criteria (limit value according to the number of samples) (KogoLab, 2004), a strong correlation was found between the measured and predicted values.

The following are the results of the least square error between the measured and predicted values presented earlier in Table 4 and 5.

The least square error between the measured and predicted values calculated by the BP method was 31.36. On the other hand, the least square

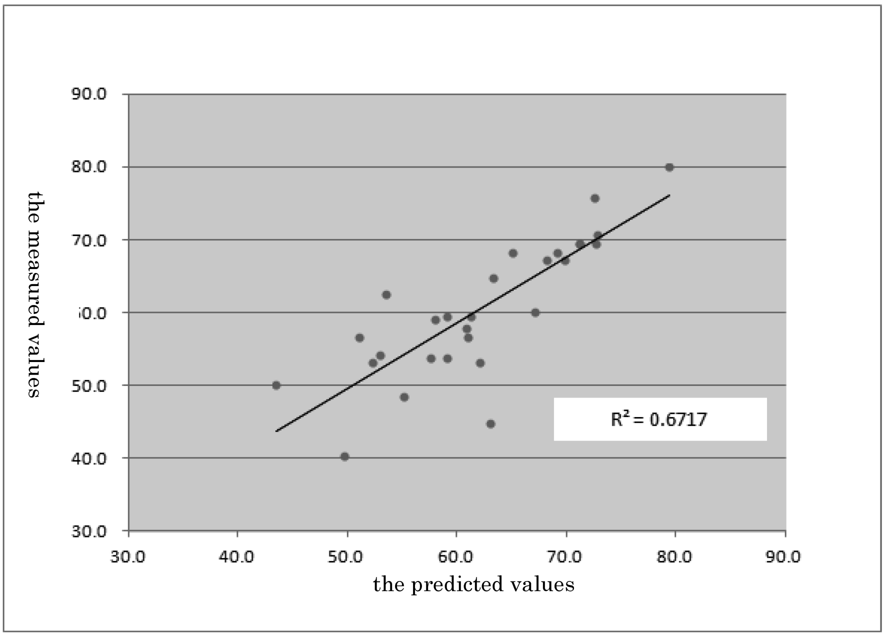

Figure 6. The Result of the Short-Timer Prediction by the BP Method

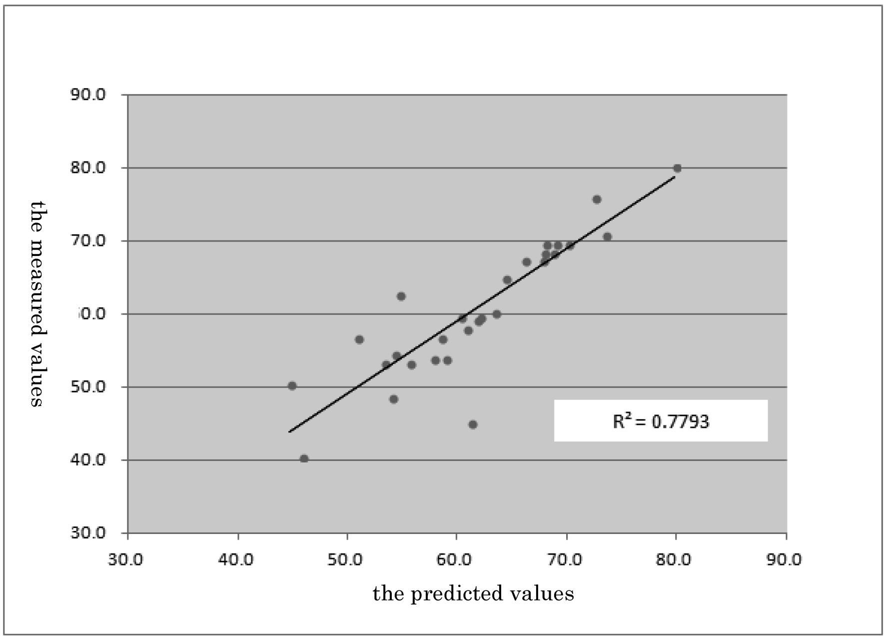

Figure 7. The Result of the BP Method Incorporating the Taken's Embedding Theorem

error between the measured and the predicted values calculated by the BP method incorporating Taken's embedding theorem was 19.81. Therefore, the prediction accuracy was higher with the BP method incorporating Taken's embedding theorem.

6. Conclusion

In the current study, we examined whether it is possible to predict test scores in a short-time period by using the two-dimensional, chaotic time-series short-term prediction method.

By analyzing the largest Lyapunov exponent and the correlational function, we found that the data exhibited chaotic behaviors. Also, because there was a strong correlation between the measured and the predicted values, we conclude that the short-term prediction of the language test scores is possible using the two-dimensional, chaotic time-series short-term prediction method.

In addition, although satisfactory results were gained from both methods, the BP method incorporating Taken's embedding theorem was more accurate than the conventional BP method in prediction.

7. Future research

A problem we need to address in future research is the number of samples. In this study, we used test scores of 28 students. However, in order to use this method in practice, we have to confirm if the method is still valid with a hundred or more students' data.

Furthermore, another problem from a pedagogical perspective is the treatment of the data which does not exhibit chaotic behaviors as observed in Student 22's data in this study. If all the test takers' scores cannot be predicted, this method cannot be used in the actual classroom. One possible solution to this issue is to use an additional method to predict test scores, as the number of cases which do not show chaotic behaviors is limited.

Also, in order to give students advice such as "You should study harder because the predicted score is low" or "Your current study method might not suit you," consistency between the predicted and the measured values is more important than the prediction accuracy. We need to resolve this issue by improving prediction accuracy or employing an additional method. Therefore, future research will aim at finding solutions for the problems above.

Acknowledgement

This research is partially supported by Grant-in-Aid for Scientific Research (25282061) by the Japan Society for the Promotion of Science (JSPS).

References

[1] Kido, K., Fukuda, E., Hashimoto, S., Okazaki, H. (2016). *Short-term Chaotic Time-Series Prediction of Language Test Scores Adopting the Backpropagation Algorithm*. *Journal of Japan e-Learning, 16,* 65-72.

[2] Kido, K. (2002). *Study on chaotic time-series short prediction: Is it possible to predict typhoon movements*? *RIMS Kokyuroku, 1252,* 147-153.

[3] Kido, K. (1999). *Short-term prediction on chaotic time-series using neuro-computing*. *Proceedings of the First Western Pacific and Third Australia-Japan Workshop*, 278-284.

[4] Shimizu, Y., Kimura, S., Sugino, N., Yamakawa, K., Ohba, H., Nakano, M. (2003). *Eigo bunpo noryoku hyojun test no datosei, shinraisei no kensho to shin eigo* [The validity and reliability of standardized English tests and compilation of "Measure of English Grammar (MEG)"]. *Policy Science*, *10*(3), 59-68.

[5] Shimizu, Y. (2005). *Sokutei ni okeru datosei no rikai no tameni: gengo test no kihongainen toshite* [To understand validity in prediction: As a basic concept of language testing]. *Ritsumeikan Studies in Language and Culture*, *16*(4), 241-254.

[6] Ogasawara, T., & Inoue, H. (2009). *Fukugo neural network ni yoru koritsutekina chaos jikeiretsukei yosoku* [Efficient hybrid neural network for chaotic time-series prediction]. *Forum on Information Technology*, *8*(2), 597-598.

[7] Sakai, S., & Takahama, T. (2009). *Taika wo donyu shita differential evolution ni yori saitekika sareta neural network ni yoru kabuka*

yosoku [stock market prediction through neural networks optimized by differential evolution with degeneration]. *RIMS Kokyuroku, 1636*, 79–87.

[8] Jimenez, J., Moreno, J. A., & Ruggeri, G. J. (1992). Forecasting on chaotic time series: A local optimal linear-reconstruction method, *Phys. Rev. A, 45*(6), 3553–3558.

[9] Mizukami, Y., Nishimori, T., Okamoto, J., & Aihara, K. (1995). *Gram-Schmidt chokkoukaho wo mochiita ketteironteki yosokushuhou ni motozuku yokujitsu saidai denryoku yoso* [Forecasting daily peak load by a deterministic prediction method with Gram-Schmidt orthonormalization]. *The Transactions of the Institute of Electrical Engineers of Japan, C115*(6), 792–797.

[10] Watanabe, H. (1983). *Shohyohon ni okeru tokuten yoso no tameno kaiki gijutsu no hikaku* [Comparison of regression techniques for the prediction of test scores from small samples]. *Research Bulletin, 4*, 1–17.

[11] Sugihara, G., & May, M. R. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature, 344*, 734–741.

[12] Ataka, S., & Okamoto, Y. (2005). *Kaiso gata neural network ni okeru gakushu algorithm no kairyo* [Improvement of learning algorithms in multi-layer neural networks]. *OIU Journal of International Studies, 19*(1), 73–85.

[13] KogoLab, Limit value by the number of samples, (2018-10-4), [http://kogolab.chillout.jp/elearn/icecream/chap3/sec4.html].

[14] Shimojo, T. (1992). *Simulation butsurigaku 6 chaos rikigaku nyumon* [Simulation Physics 6: Introduction to chaotic dynamics]. Tokyo: Kindai Kagakusha.