Time Series Forecasting with Deep Learning

Takaomi Hirata, Yamaguchi University, v003we@yamaguchi-u.ac.jp
Takashi Kuremoto, Yamaguchi University, Japan, wu@yamaguchi-u.ac.jp
Shingo Mabu, Yamaguchi University, Japan, mabu@yamaguchi-u.ac.jp
Masanao Obayashi, Yamaguchi University, Japan, m.obayas@yamaguchi-u.ac.jp
Kunikazu Kobayashi, Aichi Prefectural University, Japan, kobayashi@ist.aichi-pu.ac.jp

Abstract: Mathematical models, which aimed on finding characteristic of time series data and forecasting the future values, have been studied since many years ago, and they have been bringing tremendous benefits for science and economy. In this study, a novel time series model is proposed using a Deep Belief Net (DBN). The DBN is composed by a Restricted Boltzmann Machine (RBM) and a Multi-Layered Perceptron (MLP) which are artificial neural network models. To decide the optimal size of the model for different time series data, Particle Swarm Optimization (PSO) is adopted in the proposed method. To confirm the effectiveness of the proposed model, chaotic time series data and exchange rate time series data were used in the forecasting experiments. It was shown that the prediction precision of proposed method was higher than using conventional neural network models such as MLP and DBN with RBMs.

Key-Words: *time series forecasting, deep belief net (DBN), restricted Boltzmann machine (RBM), multi-layered perceptron (MLP), chaos*

1. Introduction

A time series is a data string to be observed in a temporal change in a certain phenomenon. To analyze and to predict time series data such as foreign currency exchanged rate, sunspot, etc. can bring immeasurable benefits to our life. Historically, linear prediction systems such as ARIMA model and nonlinear predictors such as artificial neural networks (ANNs) have been proposed and showed their efficiency in the field of time series forecasting.

Since 2006, studies of “Deep Learning” by “Deep Belief Nets (DBNs)”, which is constructed using plural restricted Boltzmann machines (RBMs), have become popular [1]. DBNs have been active in many areas such as clustering, image compression, pattern recognition, big data processing, etc. DBNs can extract high order features, and compress the high-dimensional data into low dimensional data. Conventional neural networks for the purpose of time series prediction are multi-layered perceptron (MLP), radial basis function network (RBFN), recurrent neural networks (RNNs) and so on. Meanwhile, a prediction system using a DBN composed by multiple RBMs has been proposed recently [2]. However, when the DBN using RBMs to predict the future value of time series, it may lack satisfied expression capability because the number of units of the hidden layer of the output RBM is only one. Also, it is considered that there is a limit to approximate the non-linear function by the linear sum of the outputs of units of the adjacent layer.

In this paper, we propose a novel time series prediction method by a DBN with one or several RBMs, which extracting the characteristics of the original time series data, and a MLP to approximate the time series data according to the features extracted by RBMs. Furthermore, to determine the optimum network structure for each input time series data, the particle swarm optimization (PSO) [3] is used. Two chaotic time series data, i.e., Lorenz chaos and Henon map, and the time series data of foreign currency exchange rate of

USD/JPY [4] were used in the prediction experiments. The performance comparison between the conventional DBN, MLP and RBFN was executed and the results showed the priority of the proposed method.

2. Artificial Neural networks (ANN)

ANNs are able to be excellent function approximators which represent the high information processing ability of the human brain with mathematical models, and they are applied to a wide range of scientific research fields such as pattern recognition, control engineering, forecasting, and so on.

2.1 Multi-Layered Perceptron (MLP)

MLP is a feed-forward neural network with a multi-layered structure, i.e., usually it is composed of the input layer, the hidden layer and the output layer (Figure. 1). Units (neurons) in the input layer receive the input signal from the external, and propagate it to the units in the hidden layer with synaptic connections between the units of layers. The outputs of units in the hidden layer are given by an active function $f(x)$, and they are fed forward to the units in the output layer of MLP. Units in the output layer show their outputs according to an active function $f(z)$ as same as units in the hidden layer. Generally, the active function is a nonlinear function, e.g., a sigmoid function shown in equation (1).

$$f(x) = \frac{1}{1+\exp(-x/\varepsilon)} \qquad \varepsilon\text{: gradient} \qquad (1)$$

The error back propagation method (BP) [5] is well-known to modify the synaptic weights of connections between units in different layers. Details of BP are omitted here.

2.2 Radial Basis Function Network (RBFN)

RBFN is a three-layer feed-forward neural network which has the same structure as MLP (e.g., Fig. 1). In this network generally Gaussian function $\varphi(x)$ is used as the active function for units in the hidden layer as given by equation (2), and a linear function given by equation (3) is used as the active function of units in the output layer.

$$\varphi_j(x) = \exp\left(-\frac{(x-x_j)^2}{\sigma_j^2}\right) \qquad (2)$$

$$\hat{y}(x) = \sum_j^J w_j \varphi_j(x) \qquad (3)$$

where j=1,2,…,J is the number of units in the hidden layer, $\boldsymbol{x_j}$ is the center vector of jth unit in the hidden layer, σ_j is the dispersion of the jth unit in the hidden layer, w_j is a synaptic weight of the connection between the jth unit in the hidden layer and the unit in the output layer. BP method is also used for the learning of RBFN generally.

Prediction of chaotic time series by RBFN was done by Casdagli in 1989 [6].

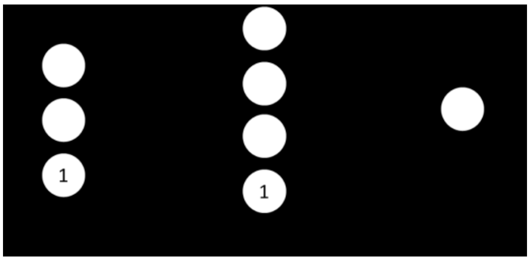


Figure 1 A Structure of MLP

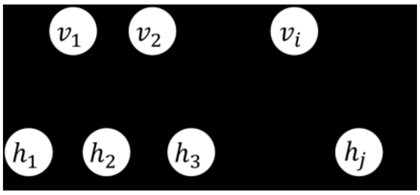


Figure 2 A Structure of RBM

2.3 Restricted Boltzmann Machine (RBM)

RBM [1] is a Boltzmann machine consisting of two layers of the visible layer and the hidden layer, no connections between units in the same layer (See Figure 2). It is possible to extract features and compress the high-dimensional data to low-dimensional data by performing an unsupervised learning algorithm. Each unit v_i in the visible layer has a symmetric connection weights w_{ij} with the unit h_j in the hidden layer. Synaptic connections between the units in the different layers are bi-directional, so the stimulation between units is in both directions. Unit v_i of the visible layer has a bias b_i, and h_j in the hidden layer with b_j. All units of RBM stochastically output 1 or 0, according to probabilities with the sigmoid function.

$$p(h_j = 1 \mid v) = \frac{1}{1 + \exp(-b_j - \sum_i v_i w_{ij})} \quad (4)$$

$$p(v_i = 1 \mid h) = \frac{1}{1 + \exp(-b_i - \sum_j h_j w_{ij})} \quad (5)$$

Using a learning rule which modifies the weights of connections, RBM network can reach a convergent state by observing its energy function:

$$E(v, h) = -\sum_i b_i v_i - \sum_j b_j h_j - \sum_{i,j} v_i h_j w_{ij} \quad (6)$$

Details of the learning algorithm of RBM can be found in [1] [2].

2.5 Deep Belief Nets (DBNs)

DBN is a multi-layered neural network which is usually composed by multiple RBMs [1]. DBN can extract the high order (abstract) features of high-dimensional data by the multiple ANNs, so it is also called "deep learning", and applied to many fields such as dimensionality reduction, image compression, pattern recognition, time series forecasting, and so on in the recent years.

In [1], Hinton & Salakhutdinov proposed an effective training algorithm for DBN, that is after the unsupervised learning of RBMs (A former RBM's hidden layer is used as the visible layer of the adjacent RBM), BP learning is adopted to perform the fine adjustment of the entire network.

In [2], Kuremoto et al. proposed a DBN prediction system with two RBMs. The DBN showed its priority to forecasting chaotic time series and benchmark data comparing to the conventional MLP predictor and linear model ARIMA. However, because the output of DBN is given by one unit in the hidden layer of the deepest RBM, and it is the linear function (summation) with the weighted output of adjacent units, the representation ability to the nonlinear time series may be limited.

In this paper, it is proposed to conFig. a new DBN prediction system using RBMs and a MLP. Figure 3 shows the structure of the proposed time series prediction system with one RBM and one MLP. To train this novel DBN, the unsupervised learning of RBM(s) is performed at first, then the output of RBM, i.e., the units of hidden layer of RBM are used as the input units of MLP. BP is used for the modification of connection weights of MLP. So the proposed DBN can be considered as a kind of pre-trained MLP indeed.

As a predictor, several historical data, e.g., $x(t-\tau), x(t-2\tau), \ldots$ are used as the input to DBN, and the output of DBN $\hat{y}(t)$, where $\tau = 1, 2, \ldots T$ is a delay constant.

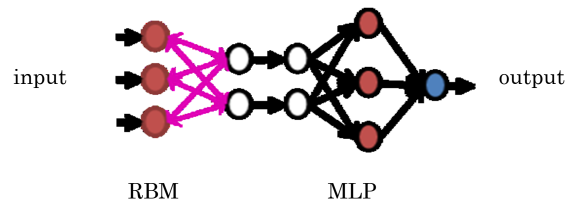


Figure 3 A structure of DBN with RBM and MLP

2.6 DBN optimization with Particle Swarm Optimization (PSO)

PSO proposed by Kennedy & Eberhart [3] is an evolutionary optimization algorithm using the swarm intelligence, that is, it mimics the behavior of a flock of birds, a school of fish, etc, to find the optimal solution of optimization problems. In PSO, a plurality of individuals (particles) form a group (swarm), to search the solution space randomly while sharing information with each other. It searches for the optimum solution in consideration of the previous best solutions of each individual and whole group. Conventionally, PSO is used to decide the learning rates and the number of units of RBMs in DBN successfully [2]. In this study, the learning rates of RBM and MLP, the optimal structure, i.e., the number of units in the visible layer and the hidden layer of RBM, and the number of units in the hidden layer of MLP, are also decided by PSO algorithm. The evaluation function of PSO is the prediction error (mean squared error between the output of DBN and teacher data).

3. Experiments and Results

To investigate the prediction performance of the proposed method, experiments were executed using a chaotic time series data, Henon map, and real data of foreign currency exchange rate (USD/JPY) in [4]. Conventional ANNs such as MLP, RBFN and DBN with RBMs were used in the comparison of the proposed method. One step ahead forecasting (short-term prediction) was used, and the prediction error to test samples of time series was calculated by the mean square error (MSE):

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y(t) - \hat{y}(t))^2 \tag{7}$$

where n is the number of time series data of training or predction, $\hat{y}(t)$ is the output of DBN, $y(t) = x(t)$ is the teach data of training samples, or target data of test samples.

Table 1 Parameters used by ANNs for Henon map data

Method	Structure	Learning rate
RBFN [6]	3-11-1	0.024
MLP [5]	2-6-1	0.1
DBN [2]	2-16, 16-1	0.025, 0.082, 0.04 (BP)
Proposed method (RBM+MLP)	2-9, 9-1	0.1, 0.1
Proposed method (2 RBMs+MLP)	2-11, 11-20, 20-15-1	0.1, 0.1, 0.084

Table 2 Training and prediction error (MSE) of Henon map ($\times 10^{-5}$)

Method	Training	Evaluation	Prediction
RBFN [6]	**0.1**	**0.1**	**0.07**
MLP [5]	28.8	19.2	7.9
DBN [2]	391.2	135.8	91.8
Proposed method (RBM+MLP)	13.5	6.7	3.6
Proposed method (2 RBMs+MLP)	13.0	7.2	3.1

3.2 Henon map

Henon map with two variables is given by equation (8).

$$\begin{cases} x(t+1) = 1 - ax(t)^2 + y(t) \\ y(t+1) = bx(t) \end{cases} \tag{8}$$

where parameters were $a = 1.4$, $b = 0.3$.

Figure 4 shows the chaotic attractor of Henon map, and Figure 5 shows the time series data of $x(t)$ and predictor's output. There were 600 data ($x(t)$, t=1, 2, …, 600) used as training samples (teacher data), 200 data (t=601, 602, …, 800) used as evaluation samples (unknown data), and 200 data (t=801, 802, …, 1000) used as test samples (unknown data). The learning rates and the number of units of ANNs were searched by PSO (20 particles in 3 to 5-dimensional space), and optimized parameters are listed in Table 1.

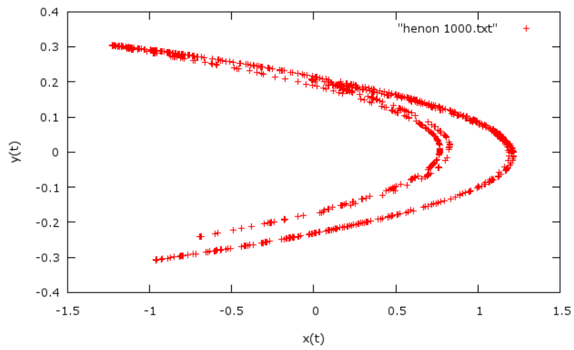
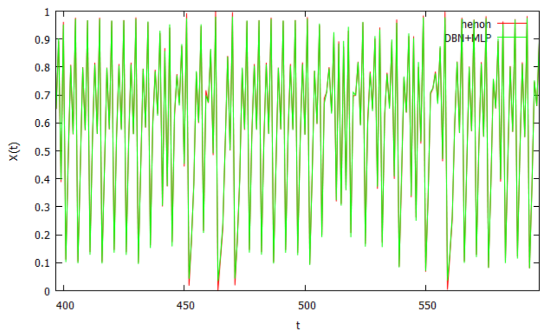
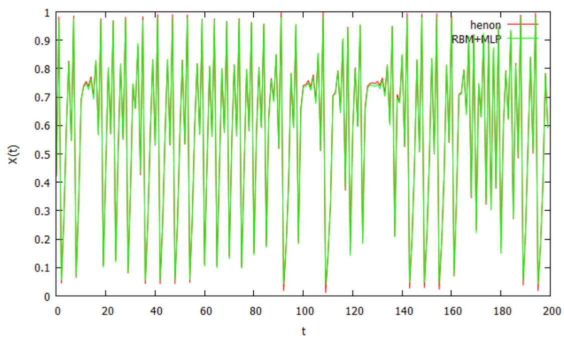


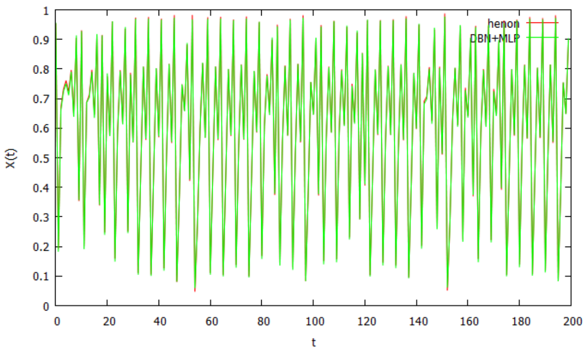
Figure 4 The chaotic attractor of Henon map



(a) Learing using training samples (teacher data)



(b) Prediction results using evaluation data (unknown data)



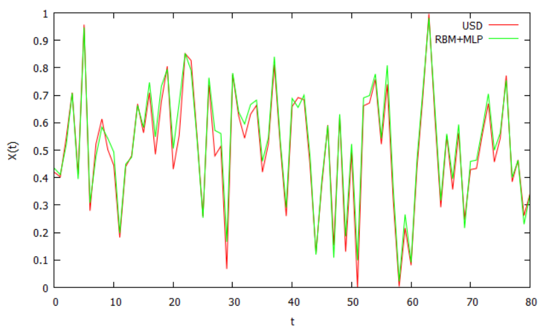
(c) Prediction result using test samples (unknown data)

Figure 5 Prediction results of Henon map by the proposed DBN (2 RBMs with one MLP)

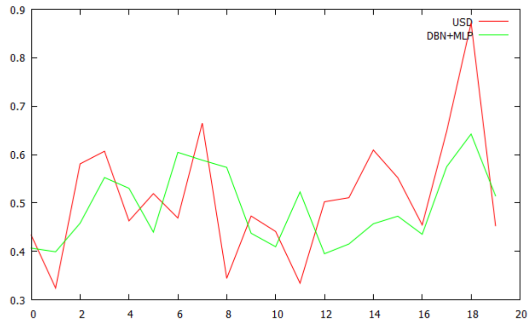
The forecasting precision of different methods using the time series data of Henon map is compared in Table 2. RBFN [6] showed the highest precisions in all cases of training, evaluation, and prediction. The proposed method, a DBN with 2 RBMs and 1 MLP showed its higher prediction performance than the conventional methods MLP [5] and DBN [2].

3.3 Foreign Currency Exchange Rate Prediction

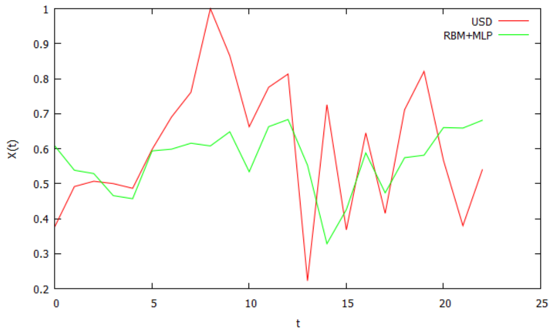
A time series of real data, JPY/USD exchange rate (One US dollar is exchanged to *x* Japanese Yen, monthly average value) presented by Mizuho Bank online [4] was used in the prediction experiment (normalized). Similar to the chaotic time series data prediction experiment described former, the first 100 data were used as training sample (teacher data), next 20 data (unknown data, i.e., not used to modify the DBN) were used to evaluate the learning performance avoiding over fitting, and the last 23 data were test samples (unknown data). The forecasting precision of different methods for the exchange rate is compared in Table 3.



(a) Prediction result to training sample (teacher data)



(b) Prediction result to evaluation samples (unknown data)



(c) Prediction result to test samples (unknown data)

Figure 6 Training and prediction results of foreign currency exchange rate (JPY/USD) with the proposed method (2 RBMs with 1 MLP)

From Table 3, we can confirm that the proposed method one RBM with MLP had the highest training precision (1.4×10^{-3}), and prediction precision (34.5×10^{-3}). Meanwhile, when the training stopped at the convergence of MSE of evaluation samples, the proposed method 2 MLPs with 1 MLP marked the best performance (10.6×10^{-3}).

Table 3 Training and prediction results of exchange rate ($\times10^{-3}$) by different methods

Method	Training	Evaluation	Prediction
RBFN [6]	43.4	13.5	41.6
MLP [5]	38.8	15.9	38.9
DBN [2]	42.0	14.1	39.3
Proposed method (RBM+MLP)	**1.4**	12.1	**34.5**
Proposed method (2 RBMs+MLP)	17.8	**10.6**	42.7

3.4 Discussion of the Optimization by PSO

As described in Section 2.6, particle swarm optimization (PSO) [3] was used to find the learning rates of all ANNs and the number of units in all layers except the output layer. In Figure 7, it is shown that the unit number of the visible layer, the hidden layer of RBM and the unit number of hidden layer of MLP changed with the iterations of PSO. Learning rates of RBM and BP learning rules were shown in Figure 8, and the decreasing of the training error (MSE) by the optimization process can be confirmed in Figure 9. These investigations showed the effectiveness of the PSO adopted in DBN training.

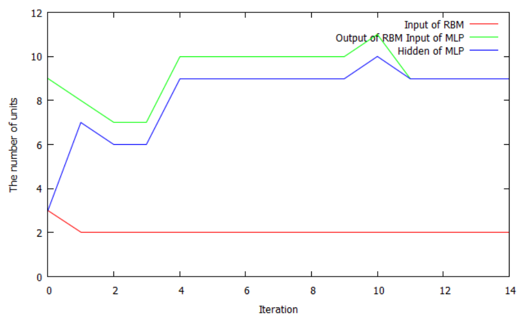


Figure 7 The number of units of DBN (RBM + MLP) changed in the iterations of PSO

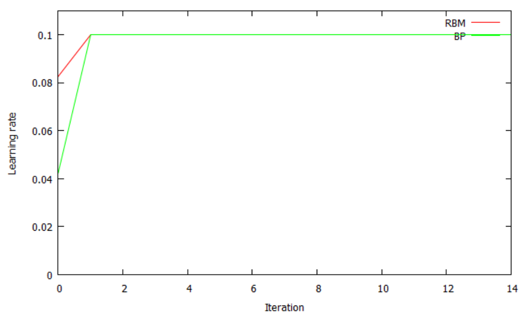


Figure 8 The learning rates of RBM and MLP changed in the iterations of PSO

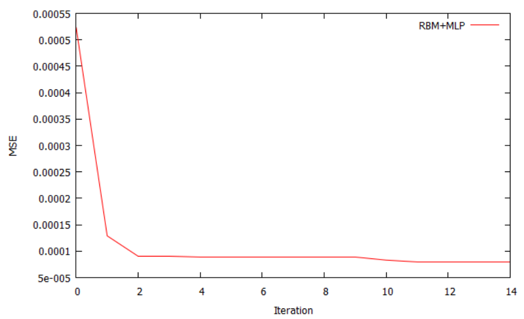


Figure 9 The training error of DBN (RBM + MLP) changed in the iterations of PSO

4. Conclusion

A novel time series prediction system which is a kind of deep belief net (DBN) composed by one or RBM more RBMs and MLP was proposed. The essential proposal was that RBMs extracted the features of the time series data by the unsupervised learning algorithm, and the learning results of RBMs were input to MLP which has the higher nonlinear representation ability than the linear unit of the conventional DBN's output. The proposed system worked well in the prediction experiments using chaotic time series data and real data. As the RBFN performed the highest prediction precision to chaotic time series data, DBNs with RBMs and RBFN are also expected to be built as a novel prediction system in the future.

ACKNOWLEDGMENT

This work was supported by Grant-in-Aid for Scientific Research (JSPS No. 25330287, and No. 26330254).

References:

[1] G. E. Hinton, and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", Science, Vol.313, 2006, pp. 504-507.

[2] T. Kuremoto, S. Kimura, K. Kobayashi, M. Obayashi, "Time series forecasting using a deep belief network with restricted Boltzmann machines", Neurocomputing, Vol.137, No.5, Aug. 2014, pp.47–56.

[3] J. Kennedy and R. C. Eberhart, "Particle swarm optimization", Proceedings of the IEEE International Conference on Neural Networks, IEEE Press, 1995, pp.1942-1948.

[4] USD/JPY data on the website of Mizuho Bank, http://www.mizuhobank.co.jp/rate/market/historical.html, 2014.

[5] E. E. Rumelhart, G. E. Hinton, R. J. Williams, "Learning representations by back-propagating errors", *Nature,* Vol. 323, pp.533–536, 1986.

[6] M. Casdagli: "Nonlinear prediction of chaotic time series", Physica D, Vol.35, pp.335-356, 1989.