

Adaptive Demand-Forecasting Approach based on
Principal Components Time-series
(an application of data-mining technique to detection of market movement)

Toshio Sugihara

Abstract

In this study, an adaptive demand-forecasting approach adopting the data-mining technique which detects the correlation between the target variable and other related elements, is proposed. With the inclusion of the scheme of the time-series analysis based on the state-space approach, this approach has two characteristic points. One is the state-space which is formed by principal components composed of various market variables. Another is the self-organization of the state-space using a neural network. This approach is applied to the two cases of the demand-forecasting. Some comparisons of forecasting accuracy (extrapolation test) with this approach and non-self-organizing models (AR, etc) are used to evaluate the effectiveness of the proposed approach. Consequently, we achieved significantly higher accuracy using this approach compared to the other approaches.

Keywords: Adaptive demand-forecasting; Data-mining Approach, (Extended) Kalman Filter; Principal Component Analysis; Neural Networks

1. Introduction

In the causes of unpredictable market movement in nowadays and logistics requirements with JIT system, demand-forecasting, one of the core functions of marketing, has been required to be improved on the performances. This is shown by the fact that demand-forecasting is positioned as the center of SCP (Supply Chain Planning) which is the planning function of SCM (supply Chain Management). In the area of SCM which is the eminent innovation of management, SCP is the total control program which transverses from production, logistics to marketing. Therefore, in the published SCP packages, the performances of demand-forecasting are improved distinctly[1]. The main requirements to these demand-forecasting are considered the next two points.

- Real-time response at any time
- Dynamic acquisition of the market movement relating the target product.

In trying to satisfy these requirements, demand-forecasting had been influenced by the various approaches of the economic time-series analysis[2, 3]. In nowadays, not only linear models, but nonlinear models (for example, neural network) are applied[4, 5]. Almost of these approaches process the single time-series which is the sales or inventory, etc of the target variable. In the case of the approach using state-space, the state-space is built mainly with past lag parameters of the single time-series[6]. The market unpredictability is increasing according to the varieties of products and consumer's behaviors. The approach of including the correlation with the target product

and the related market movement is indispensable in demand-forecasting approach. In recent days, the data-mining approach which based on the data warehouse, has made startling progress [7, 8, 9]. In demand-forecasting including market movement, the approach which takes in data-mining technique is considered to be effective approach.

Above stated, in this study, an adaptive and multi-variable prediction scheme which the target product and the market movement are predicted simultaneously, is proposed. The market movement is represented as the state-space which is built with the target variable and the principal components time-series of the related variables. The state-space is renewed at each observed time by the self-organized mechanism using neural network. And for the aim at real-time response of this demand-forecasting, total demand-forecasting scheme is based on the framework of the Kalman Filter.

2. Data-mining

Data-mining is not always defined as a confirmed and unified concept. However, it is described generally as the "detection of valuable and nontrivial information from a large-scale database" , and its concept is mainly based on the detection of knowledge [4, 7]. However, data-mining is not the inspection of hypotheses or the verification of established hypotheses, but should be "the automatic detection of new facts and relations from a plain database". According to this definition, data-mining is considered to be built on a large-scale database such as a data warehouse. The data-mining approaches presented now and their applications are shown in Table 1.

[Table 1. Data-mining approaches and their applications [4]]

Data-mining function	Algorithm	Applications
correlation	statistics, ensemble theory	market basket analysis
classification	decision tree, neural network	target marketing, quality control, risk evaluation
modeling	linear/nonlinear regression,	customer ordering, price modeling,
time-series prediction	curve fitting, neural network statistical ARMA model,	process control sales-forecasting, rate-forecasting
time-series pattern	Box-Jenkins, neural network statistics, ensemble theory	inventory control time-series market basket analysis

In Table 1, as the tools for data-mining, some statistical methods, curve-fitting methods, regression analysis, neural network and genetic algorithm are listed. The contents in Table 1 appear to indicate that "data-mining is not the inspection of statistical hypotheses or the verification of established hypotheses", but this means that data-mining is nothing but a tool which automatically detects new facts or relations from data warehouses. Therefore, data-mining is a tool not of inspection but of detection.

3. Basic aspects of the proposed demand-forecasting

3.1. Extraction of market movement and principal components time-series

The market movement is lied in data warehouse and it is considered that the time varying of the stored data represents the market movement. Therefore, for the extraction of the market movement, not only single time-series of the target product but also multi time-series related with it should be processed as integrated. But these time-series are considered strongly correlated with each other, and should not be handled individually. For that reason, the principal components time-series which are integrated from the original multi time-series by principal components analysis (PCA), can represent the market movement[10]. The adopted principal components time-series are selected by the proportion of eigenvalue. Consequently, the adoption of principal components time-series may be considered as one of the Data-mining tools.

By applying principal component analysis, the time-series of N original variables are integrated in the time-series of M (M < N) principal components. The most variations based on the original market variables are represented as the movement of common factors, namely market movement is represented by far fewer variables than in the original method.

PCA of time-series composes principal components time-series from original time-series, and the time-series of the principal components are independent of each other. At any time, the j-th component is presented as

$$p_j = \sum_{i=1}^N f_{ji}x_i \quad (1)$$

Here,

index of the principal components : j = 1, 2,, M

index of the original variables : i= 1, 2,, N

and

$$\sum_{i=1}^N f_{ji}^2 = 1 \quad (2)$$

$$\sum_{i=1}^N f_{ji}f_{ji} = 0 \quad (j \neq 1) \quad (3)$$

The number of principal components adopted is determined by the proportion of the accumulated eigenvalue, which occupies generally about 80-90 percent. The principal components are selected in the order of high to low eigenvalue. f_j is an eigenvector obtained as the linear combination of the original variables. f_j is determined at each time (with constant time period) , as shown by formula (1), the time-series of a principal component can be generated over the entire period.

The time-series of a principal component is symbolized as follows.

$$\{ p_{kj} : k = 1, 2, \dots, n : j = 1, 2, \dots, M \} \quad (4)$$

Here,

index of the principal components : j = 1, 2,, M

index of the time : k = 1, 2,, n

Generally, M is far smaller than N, and $\{ p_j : j = 1, 2, \dots, M \}$ is considered to present the common variance of the original variables. Considering the independence of each component from other components, the time-series of the principal component is assumed also to be independent. Therefore, the time-series of a principal component can be the extracted market movement.

3.2. Total scheme of prediction

For the online prediction scheme, the Kalman Filter is adopted as the fundamental framework of prediction in this work [11, 12]. In the field of linear system theory, the

Kalman Filter is the rational estimation / prediction approach to the sequential online time-series. Also in the nonlinear field, the Kalman Filter is formulated as the Extended Kalman Filter. Its concept is the multi input / output control system and multi variables are estimated / predicted simultaneously.

The basic equations for the Kalman Filter are shown as equations (5) and (6) as below. The Kalman Filter consists of the state vector and the measurement vector involving multiple variables.

$$\mathbf{x}_k = \mathbf{A}_{k,k-1} \mathbf{x}_{k-1} + \mathbf{C}_k \mathbf{u}_k + \mathbf{v}_k \quad (5)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k \quad (6)$$

\mathbf{x}_k : state vector $\mathbf{A}_{k,k-1}$: state transition matrix

\mathbf{C}_k : control matrix \mathbf{u}_k : control vector \mathbf{v}_k : state noise vector

\mathbf{y}_k : measurement vector \mathbf{H}_k : measurement matrix

\mathbf{w}_k : measurement noise vector

The state noise and the measurement noise are both assumed to be white at each time and independent of each other. Fig.1 shows the total processing steps of the Kalman filter. The estimation is processed on the basis of maximum likelihood method and the prediction is gained as the output of the dynamical state-transition from input of its estimation value. In the Kalman Filter, the chain of these process (from Step 1 to Step 7) is repeated. The predicted / estimated value is outputted in Step 3 / Step 6.

The transition matrix updates the state-space and the measurement matrix transforms the state variables to the measurement variables. In this study, the state variables are constructed with the target variable and the principal components. Therefore, the state space is noted as below.

$$\{p_{kj} : k=1, 2, \dots, n : j=1, 2, \dots, M\}$$

$$\{y_k : k=1, 2, \dots, n\}$$

As showed below, the transition matrix is renewed by self organized mechanism using neural network, the measurement variables are the same as the state space variables. From these premise,

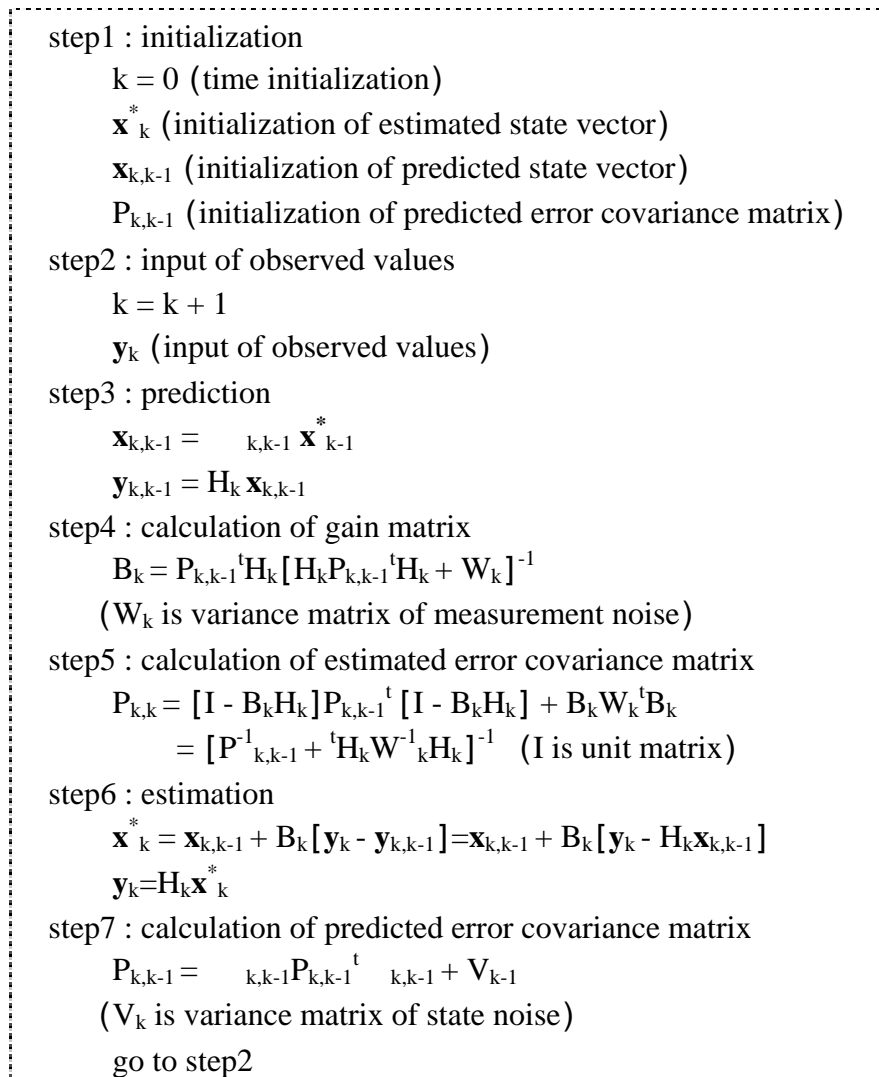
(5) and (6) are formulated as below.

$$\mathbf{x}_k = \mathbf{A}_{k,k-1}^* \mathbf{x}_{k-1} + \mathbf{v}_k \quad (7)$$

$$\mathbf{y}_k = \mathbf{I} \mathbf{x}_k + \mathbf{w}_k \quad (8)$$

$\mathbf{A}_{k,k-1}^*$ is self-organized transition matrix by neural network and I is unit matrix.

[Fig1 . The chain of Kalman Filter processing steps]



3.3. Acquisition of the structural market change to the prediction mechanism

As above mentioned, the structural market change is buried in stored data. In this study, the prediction mechanism which intends to acquire the market movement is considered to be driven by the effective Data-mining tool using neural network.

The essential points in forming the Kalman Filter are the determinations of the state transition matrix and measurement matrix. In forming the Kalman Filter in physical systems, the dynamics of state and observation systems are represented as obvious equations which give the expressions of them. In particular, the state transition matrix presents the dynamics of system from a time to the next time, and therefore, plays the main role in the estimation of the system dynamics. Otherwise, in economic / management field, it appears to be impossible to elucidate the system dynamics deductively, the components of the state transition matrix are obtained as the coefficients of regression equations applied to the preceding time-series. In such a case, the elements are constants that cannot cope with the varying system. Considering the theoretical conditions of the Kalman Filter, when the time variation of the transition matrix is not guaranteed, this formulation cannot be applied to the non-stationary

state [13].

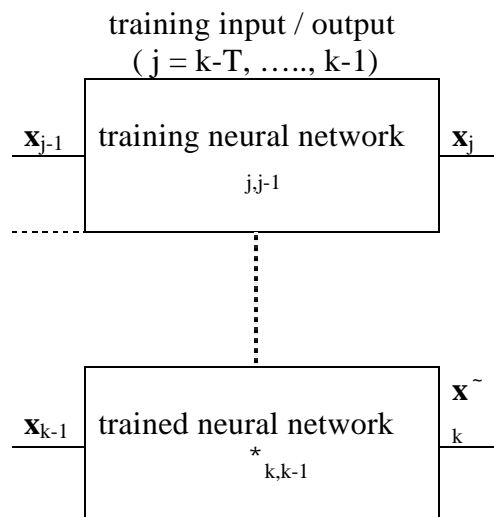
To solve this problem, the self-organization of state space has been proposed [14]. Each element of $\mathbf{x}_{k,k-1}$ is calculated based on self-organization at the time of update of state variables and its relations. In this paper, we apply a method which uses neural network processing of nonlinear relations [8, 9]. The applied neural network is structured type, and is trained using the sets of neighboring time-series as training data at constant interval. At each cycle of the Kalman Filter processing, the state transition matrix is renewed by the above-mentioned self-organized process. Here, for the generation of the state transition matrix, training data are set as follows.

input set of training data $\{\mathbf{x}_j : j = k - T, \dots, k - 1\}$ (T: training interval)

output set of training data $\{\mathbf{x}_j : j = k - T + 1, \dots, k\}$ (T: training interval)

By this training process, the learned transition matrix $\mathbf{x}_{k,k-1}^*$ is generated and the prediction vector $\{\mathbf{x}_{k+1}\}$ is obtained by inputting the vector $\{\mathbf{x}_k\}$ into the trained neural network. This procedure is shown as Fig 2.

[Fig 2. training / prediction process of transition matrix]



4. Scheme of proposed demand-forecasting approach

4.1. Total picture

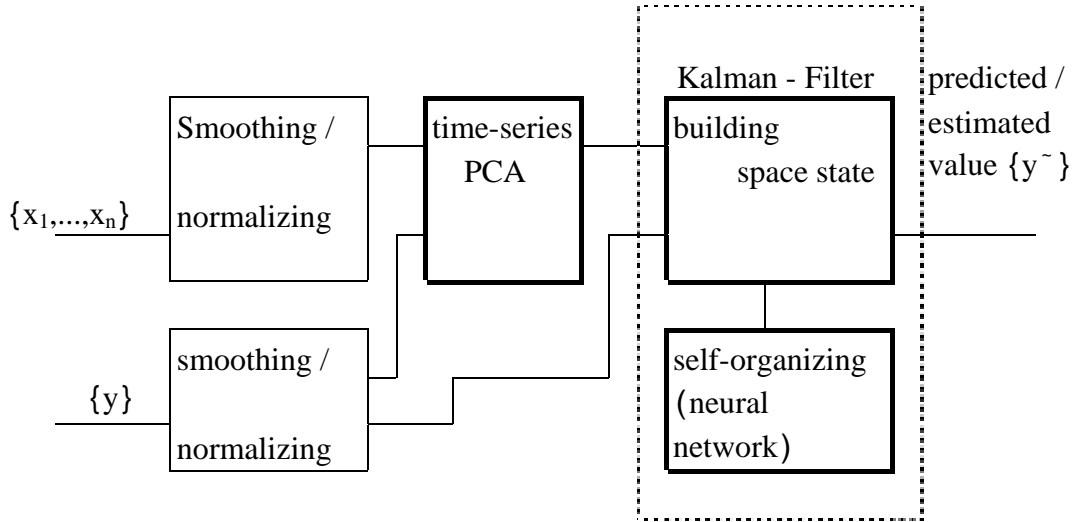
Total picture of this demand-forecasting system is shown as Fig 3. The input time-series are as follows.

$\{y\}$: target variable time-series

$\{x_1, x_2, \dots, x_n\}$: relating variables time-series

These are seasonal adjusted and pre-processed by smoothing and normalization. The state space is built with $(M + 1)$ variables. The output variables are the estimation / prediction value of these variables.

[Fig 3. Total picture]



4.2. Procedure of the Kalman Filter

The procedure of the Kalman Filter which is the framework of this demand-forecasting approach is built as follows. For the acquisition of market movement, the processing time period is divided into two parts. The former period is named "Initial period" and the latter is named "Prediction period". In "Initial period", the elements of transition matrix are formed with regression coefficients corresponding to the update of each variable. In "Prediction period", they are formed with the ratio of output / input by neural network corresponding to each variable. As above mentioned, the state space variables are the same as the measurement variables.

< measurement time-series >

principal variable time-series : $\{p_{kj} : k=1, 2, \dots, n : j=1, 2, \dots, M\}$

target variable time-series : $\{y_k : k=1, 2, \dots, n\}$

< state space variables >

$${}^t(p_{k1}, p_{k2}, \dots, p_{kM}, y_k) \quad (t \text{ shows transposed}) \quad (9)$$

< measurement variables >

$$\text{the same as the state space variables} \quad (10)$$

< transition matrix >

[Initial period]

The principal variable is independent at each other and regression formula of each variable is shown as below.

$$p_{kj} = a_{jj}p_{kj-1} + v_{kj} \quad (j = 1, 2, \dots, M) \quad (11)$$

a_{jj} : regression coefficient, v_{kj} : regression error

The regression formula of target variable is formed as follow.

$$y_k = \sum_{j=1}^{M+1} a_{M+1j} p_{kj} + v_{kM+1} \quad (12)$$

a_{M+1j} : regression coefficient, v_{kM+1} : regression error

Consequently, transition matrix is formed as below.

$${}_{k,k-1} = \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & & 0 \\ 0 & & a_{MM} & 0 \\ a_{M+11} & a_{M+12} & & a_{M+1M+1} \end{bmatrix} \quad (13)$$

[Prediction period]

The elements of the transition matrix are formed with the ratio of output / input by neural network corresponding to the each variable.

$${}_{k,k-1}^* = \begin{bmatrix} \tilde{p}_{k1} / p_{k-11} & 0 & 0 & 0 \\ 0 & \tilde{p}_{k2} / p_{k-12} & 0 & 0 \\ 0 & 0 & & 0 \\ 0 & 0 & & 0 \\ 0 & 0 & 0 & \tilde{y}_k / y_{k-1} \end{bmatrix} \quad (14)$$

< measurement matrix >

The measurement matrix is unit matrix and its row is (M+1), column is (M+1).

< state noise and measurement noise >

[state noise]

State noise is generated at the update of state space and it is equivalent to the prediction error. Namely, in "Initial period", the regression error is used and in "Prediction period", the convergence error of neural network is used.

[measurement noise]

Different from physical time-series, it is difficult to understand the meaning of measurement noise of economic time-series. In this study, the measurement noise is generated as the difference between observed series and its curve-fitted series. This operation is applied to both the principal components time-series and the target variable time-series.

5. Case study

5.1. Case study(1)

As the case, the demand-forecasting of Passenger New Car Sales (monthly time-series) is adopted. Here, the time-series is seasonally adjusted, and the trend leading series is the target series for demand-forecasting. The processing period is from 1991 to 2000 and sample size is 120. The prediction period is 24 month(1999 - 2000). Five time-series are related to the target series, and are listed as follows. They are all seasonally adjusted, and the period and sample size are the same as the target series.

- y : Passenger New Car Sales
- x₁ : Private Final Consumption Expenditure
- x₂ : Money Supply (M2 + CD)
- x₃ : Stock Average
- x₄ : Consumer Price Indices
- x₅ : Building Construction Starts.

After PCA of the 6 time-series above, each eigenvalue and accumulated proportion are obtained as follows.

	eigenvalue	accumulated proportion
first PC (p_1)	0.027	0.722
second PC (p_2)	0.005	0.860
third PC (p_3)	0.004	0.955

We use the first PC, second PC, and third PC which make up 95 percents of the accumulated proportion. The formation of the Kalman Filter in this case is as follows.

(a) state variables

first principal component --- third principal component : $\{ p_1, p_2, p_3 \}$

passenger new car sales : $\{ y \}$

Three models are defined as combinations of the above-stated state variables.

model A : $\{ p_1, y \}$

model B : $\{ p_1, p_2, y \}$

model C : $\{ p_1, p_2, p_3, y \}$

(b) measurement variables

equal to state variables in each model

(c) state transition matrix

Elements of state transition matrix are coefficients estimated by regression in the initial period (1991 ~ 1998, 8 years and 96 samples).

In the prediction period (1999 ~ 2000), the predicted values are obtained as output of the neural network, the input of which is the last observation. Here, the neural network is trained using samples of time-series for the preceding 2 years. The degree of convergence in training is about 10^{-3} .

(d) measurement matrix

The measurement matrix transforms the state vector to the measurement vector.

(e) state noise

In the initial period, each state noise is defined as the root mean square error of regression equations according to models A, B and C.

In the prediction period, each state noise is defined as the mean square error of network training. This is calculated and updated in the prediction cycle of Kalman Filter processing.

(f) measurement noise

Measurement noises are defined as the root mean square error of the moving average of $\{ p_1, p_2, p_3, y \}$. These are used both in the initial and prediction periods.

The evaluation of each model is shown in Table 2. The prediction error is calculated as a root mean square error (RMSE).

[Table 2 Prediction accuracy of each model]

model	model A	model B	model C
prediction error	0.0030	0.0036	0.0032

To verify the efficiency of this approach, the prediction accuracy compared to those of other approaches is evaluated (RMSE).

[Table 3. Prediction accuracy of each model]

state variables \ state space	$\{y, x_1, x_2, x_3, x_4, x_5\}$	$\{y, p_1\}$
non self- organized	0.0041	0.0032
self – organized	0.0044	0.0030

AR Model (order 5) : 0.0040 (Estimation)

5.2. Case study (2)

Next the case, the demand-forecasting of Air Conditioners Sales (monthly time-series) is adopted. Here, the time-series is seasonally adjusted, and the trend leading series is the target series for demand-forecasting. The processing period is from 1991 to 1999 and sample size is 108. The prediction period is 24 month (1998 - 1999). Three time-series are related to the target series, and are listed as follows. They are all seasonally adjusted, and the period and sample size are the same as in the target series.

y : Air Conditioner Sales

x_1 : Index of Industrial Production

x_2 : Consumer Price Indices

x_3 : Householders.

After PCA of the 4 time-series above, each eigenvalue and accumulated proportion are obtained as follows.

	eigenvalue	accumulated proportion
first PC (p_1)	0.024	0.747
second PC (p_2)	0.006	0.945
third PC (p_3)	0.002	0.992

To verify the efficiency of this approach, the prediction / estimation accuracy compared to those of other approaches is evaluated (RMSE).

[Table 4. Prediction accuracy of each model]

state variables \ state space	$\{y, x_1, x_2, x_3\}$	$\{y, p_1\}$
non self- organized	0.0081	0.0152
self – organized	0.0074	0.0066

AR Model (order 3) : 0.0082 (Estimation)

Based on Table 2, 3 and 4 the following is revealed.

- The effect of adopting self-organization of state-space on the prediction accuracy, is heavily weighted. The prediction accuracy of non self-organized state space which has the transition matrix determined throughout the entire prediction period, is inferior to those of self-organized cases.

- In the approach using principal components, the use of one components is optimal. This indicates that increasing the number of principal components does not always result in higher accuracy and the optimal the number of components representing market variance exists.

6. Conclusions and discussions

In this study, we tried to define state-space formed by principal components. The results presented in Section 5 indicated that better prediction accuracy than that obtained other approach (AR and stationary Kalman Filter approach) was achieved, and it was considered that adopting of principal components was effective. The principal components which were integrated from the original variables, therefore represent independent market movement. It was considered that the formation of state-space based on independent common market movement is more effective than that based on original market variables.

Based on these results, the following are concluded.

- This approach is essentially a data-mining approach. For this reason, the adopted common market movement can be integrated from any arbitrarily selected market variable.

- The state-space which is formed with non-interactive market movement, is more effective than that is formed with original market variables.

- The optimal number of principal components which form state-space, exists. Namely, adopting too few or too many principal components reduces the accuracy.

With regard to this last point, further study on the optimal selection of principal components is needed. The relative contribution of the variance of target time-series to the total variance should be studied simultaneously.

On the premise of real-time prediction, this prediction approach is based on the framework of the Kalman Filter. Properly speaking, because of the convergence processing of neural network, this approach is not strictly real-time processing. However, the convergence cycle is about 20 - 30 iterations sufficiently up to the efficient convergence limit, and computer has sufficient CPU speed to process. This approach can be nearly real-time processing. In addition to these points, a neural network which has a feedback mechanism is considered to be a valuable tool for self-organization. a feedback neural network will be evaluated with respect to not only the self-organization method of state-space transition but also a direct method for demand-forecasting in real-time processing.

References

1. T. Miyazaki, " SCM and OR", Communications of the Operations Research Society of Japan, Operations Research, Vol. 44, No. 6, pp.300-303, 1999.
2. P. J. Brockwell, R. A. Davis "Time Series : Theory and Methods", 2nd Edition, Springer-Verlag, New York, 1991.
3. C. W. J. Granger "Forecasting In Business And Economics", Academic Press, 1989.
4. J. P. Bigus, " Data Mining with Neural networks", McGraw-Hill, 1996.
5. H. Araki, I. Kimura, I. Arizono, H. Ohta " Demand Forecasting based on Differences of Demands via Neural Networks" J. JIMA, No. 2, pp.59-68, 1996
6. G. Kitagawa, W. Gersch " A smoothness priors-state space modeling of time-series with trend and seasonality", Journal of American Statistical Association, Vol. 79, No. 386, pp.378-389. 1984
7. M.J.A. Berry, G. Linoff " Data Mining Techniques for Marketing, Sales and Customer Support", John Wiley, 1997
8. C. G. Dasgupta, G. S. Dindens, S. Ghose " Computing the predictive performance of a neural network model with some traditional market response models", International Journal of Forecasting, vol.10, pp. 235-244, 1994
9. S. Dutta, S. Shekhar, W.Y. Wong " Decision support in non-conservative domains: generalization with neural networks", Decision Support Systems, Vol.11, No.5, pp.527-544, 1994
10. T. Kariya " The New Development of Econometric Analysis", Iwanami, 1994
11. M. Aoki " State Space Modeling of Time Series" Springer-Verlag, 1987
12. T. Sugihara " Economic time-series analysis adopting adaptive models" Kougaku Toshiyo, 1996
13. G. Kitagawa, W. Gersch " A smoothness priors time-varying AR coefficient modeling of non-stationary covariance time series", IEEE Transactions on Automatic Control, Vol. AC-30, No.1, pp.48-56, 1985
14. T. Sugihara, H. Tsubone " Demand Forecasting System adopting to market movements for Supply Chain planning", Proceedings of 6-th. International Symposium of Logistics, pp.113-116, 2001