A Method to Search ARX Model Orders and Its Application to Sales Dynamics Analysis

Kenta Fukata
kenta-f@ar.sanken.osaka-u.ac.jp

Takashi Washio
washio@ar.sanken.osaka-u.ac.jp

Hiroshi Motoda
The Institute of Scientific and Industrial Research, Osaka University
8-1, Mihogaoka, Ibarakishi, Osaka, 567-0047, Japan

Abstract

An Auto-Regressive eXogenous input (ARX) model has been widely used in engineering fields to model dynamic response of a system to exogenous factors. A difficulty in this modeling is the determination of an appropriate model complexity, i.e., orders, for given data. In this paper, we develop a new and practical approach to determine the appropriate orders. Moreover, we apply the developed technique to a real marketing data, and analyze dynamic response character of sales amount to advertisement and sales promotion. In marketing study, static response of sales to some exogenous factors such as advertisement and sales promotion have been analyzed. However, if we can model dynamic response of sales to exogenous factors, more precise strategies of the sales to reduce the risk of the item stock management and increase the associated profit can be designed.

1 Introduction

In this paper, an approach is proposed to model the dynamics of an objective system based on the temporal behaviors observed from the system under some exogenous influences. One of the representative and quantitative dynamics modeling approaches which have been often used in the time series analysis is Auto-Regressive eXogenous input (ARX) modeling [1]. Similarly to the many other empirical modeling approaches based on given time series data, a main issue in the modeling is to determine an appropriate model complexity well capturing the dynamic structure of the objective system. In terms of the ARX modeling, its complexity is defined by the model orders which are the finite numbers of the past consecutive quantitative states to take into account under a constant time interval sampling. This issue has been explored in many aspects including the indices of AIC[2], BIC[3] and MDL[4]. However, these works assume that the time series data is observed from the system without any uncertainty and distortions such as sensing accuracy limit, discretization errors from analog to digital information and unexpected biases not following any statistical expectations and distributions. This assumption strongly limits the applicability of these criteria to determine the model complexity in many practical modeling problems. This issue on the use of the information criteria in practical situations have been partially discussed, and some ideas to use the difference of the indices between two candidate models such as ΔAIC[5] has been proposed. However, the selection of the appropriate model order based on the indices still remains within some artistry.

Another important issue which has not been explored in the selection of the appropriate ARX model orders is the efficient and complete search algorithm of the orders having the optimal index value. In conventional approaches, a simple line search is used in which the index of each model order is exhaustively computed up to a limit order. However, in case that an ARX model has multiple exogenous inputs, the number of the parameters to define the model orders is also more than one. Because the number of the parameter combinations of the ARX model to be explored within some order limit is exponential to the number of the parameters, the search of the optimal model orders becomes easily intractable under the increase of the exogenous inputs. In this paper, we propose a novel information criterion named ΔAIC* which is an extension of the ΔAIC, and further propose an efficient and complete search algorithm of the parameter combination of the optimal model orders under the criterion.

Another objective of this paper is to apply the proposed novel approach of the ARX modeling to the real world investigation in the marketing area and to demonstrate its feasibility in the investigation. The example investigation we demonstrate in this paper is the analysis on the dynamic response of the sales amount of an item to its TV advertisement and instore sales promotion. The proposed approach has been applied to a time series data consisting of an objective item sales amount, two exogenous variables of the TV advertisement amount and the instore promotion strength. If we can appropriately grasp the dynamic contributions of each exogenous factor to the objective item sales amount through the analysis, the efficiencies of the advertisement and the promotion can be easily evaluated, and their suitable strategies can be precisely designed.

2 State of the Art

2.1 Auto-Regressive eXogenous input model

An Auto-Regressive eXogenous input (ARX) model is a linear recurrence equation to relate the current value of an objective variable $x(s)$ with its past finite time series and the past finite time series of the other exogenous input variables y_g $(g = 1, ..., h)$ as follows.

$$\begin{aligned} x(s) = \sum_{i=1}^{p} a_i x(s-i) + \sum_{j_1=1}^{q_1} b_{j_1} y_1(s - j_1 - k_1) + \\ ... + \sum_{j_h=1}^{q_h} b_{j_h} y_h(s - j_h - k_h) + e(s), \end{aligned} \tag{1}$$

where s is a current time step, a_i the contribution coefficient of an i-step past value of the objective variable to its current value, b_{j_g} the contribution coefficient of the j-step past value of an exogenous input variable y_g, k_g the time lag of the propagation delay of the exogenous input variable, and p, q_g $(g = 1, ..., h)$ the model order parameters which define the finite and maximum time steps of the contributions from the objective and the exogenous variables. In addition, let $\hat{x}(s)$ be a prediction of $x(s)$ and $e(s) = x(s) - \hat{x}(s)$ their prediction error. The model coefficients a_i(i=1,...,p) and b_{j_g} (j_g =1,...,q_g, g =1,...,h) are determined by the least square principle on the variance of the prediction error $e(s)$ over a given time series data. The combination of the time lags k_g $(g = 1, ..., h)$ which are integers is determined by a greedy method to search the combination which provides less least square prediction error on the combination lattice. The model orders, *i.e.*, the parameter values of p, q_g $(g = 1, ..., h)$, are conventionally determined by the AIC index as explained in the next subsection.

2.2 Conventional order determination

The selection of the appropriate orders of the ARX model is crucial, and it has been performed by using AIC (Akaike information criterion) in the conventional and standard approach. AIC is an information measure to evaluate the difference between the actual probability distribution of the value $x(s)$ and that of the predicted value $\hat{x}(s)$. AIC can be defined by the following Eq.(2) as the measure of the difference between these two probability distributions based on Kullback-Leibler quantity of information.

$$AIC = N \log(\hat{\sigma}_M^2) + 2|M|, \tag{2}$$

where N stands for the total number of data and $\hat{\sigma}_M^2$ the variance of the model prediction error e(s). Moreover, $|M|$ stands for the total number of coefficients in the ARX model where $M = [p, q_g \ (g = 1, ..., h)]$. The smaller value indicates that the estimated distribution function is closer to the true distribution function.

An important issue on this AIC is the limitation in its practical use due to its strong assumption on the linearity of the objective system and the absence of observation error. If the objective time series data is observed from a linear system without any observation error, AIC should have a clear bottom on a model complexity. Thus the optimal model complexity is uniquely determined by the bottom of AIC. However, the AIC curve does not follow the ideal case when some nonlinearity of the system and some observation error exist, and does not show any clear bottom in many cases. This happens since the errors induced by the nonlinearity and the observation are incorporated in the evaluation of AIC as if they are some meaningful errors. Accordingly, more practical measure to determine an appropriate model order must be established.

3 Proposal of ΔAIC*

ΔAIC[5] which takes the difference of the AIC between consecutive model orders provides a criteria to determine some appropriate order of the Auto-Regressive (AR) model which has only a unique order parameter. In the principle of ΔAIC, a model order which shows a significant decrease of the AIC value is selected as an appropriate order instead of the minimum value of the AIC. This is because the significant decrease of the AIC value may not occur when the model incorporates the errors, but may occur when it incorporates major characteristics of the objective system. This principle can be similarly applied to the models having multiple model order parameters such as the ARX model. However, to our best knowledge, no approach had addressed the application of this principle to the case of the multiple model order parameters.

In this paper, we extend the ΔAIC to ΔAIC* to address the above issue. Given a model order parameter vector $M = [p, q_1, q_2, ..., q_h]$ and the AIC under M as AIC_M, ΔAIC* is defined as the minimum difference of AIC_M from AIC_{M_p} and $AIC_{M_{q_g}}$ $(g = 1, ..., h)$ where $M_p = [p-1, q_1, q_2, ..., q_h]$ and $M_{q_g} = [p-1, q_1, ..., q_g-1, ..., q_h]$.

More formally, ΔAIC* is described as follows.

$$\Delta AIC* = \max(\Delta AIC_p, \max_{g=1,...,h}(\Delta AIC_{q_g})), \quad (3)$$

where

$$\begin{array}{l}\Delta AIC_p = AIC_M - AIC_{M_p}, \\ \Delta AIC_{q_g} = AIC_M - AIC_{M_{q_g}}.\end{array} \quad (4)$$

Note that the minimum difference between two AICs corresponds to the maximum value of the difference since their values are always negative. By definition, ΔAIC* stands for the least improvement of the AIC under a unit extension of the model complexity. On the other hand, the best model order is considered to be the order which provide the maximum improvement of the AIC similarly to the principle of the ΔAIC. Accordingly, the order providing the maximum improvement of the AIC should be selected as the appropriate model order by using the ΔAIC*. This is done by seeking the model order parameter vector M providing the minimum value of the ΔAIC* due to its negativeness. This approach enables to select the model order to achieve the maximum value of the least improvement of the AIC among the order parameter changes. The purpose of this strategy is to discover the model where its any simplification certainly and significantly reduces the accuracy of the model beyond the errors. When an order parameter is zero, ΔAIC* including the parameter can not be computed since the further simple model for the parameter does not exist. In this case, ΔAIC* is evaluated by Eq.(3) while excluding the zero order parameters since the variables corresponding to the zero order parameters are not included in the model.

From Eq.(4), the AIC for the model simpler by one order parameter is as follows.

$$AIC_{M'} = N\log(\hat{\sigma}^2_{M'}) + 2(|M| - 1), \quad (5)$$

where $\hat{\sigma}^2_{M'}$ is the variance of the prediction error $e(s)$ by the simpler model. Accordingly, the concrete formula of ΔAIC* in Eq.(3) which is the difference between Eq.(4) and (5) is represented as

$$\Delta AIC* = N\log\frac{\hat{\sigma}^2_M}{\min_{g=1,...,h}(\hat{\sigma}^2_{M_p}, \hat{\sigma}^2_{M_{q_g}})} + 2. \quad (6)$$

Instead of the original definition of ΔAIC* in Eq.(3), this formula is used for the computation of ΔAIC* and the search for the optimal ARX model.

4 Search for Optimal Model Order

The simplest and complete way to search the optimal model order vector M by ΔAIC* under a given time series data is the thorough search by using loops for all order parameters. However, the computational complexity of this algorithm is $O(L^{h+1})$ where L is the upper limit of the order to search, and hence the computation becomes intractable when the number of the exogenous variables and/or the upper limit of the order are large.

For practically efficient search of the optimal M based on ΔAIC*, we introduce A* search[6],[7]. A* search uses a lower bound $\underline{f}$ of an objective function f to minimize instead of the objective function itself. This f is ΔAIC* in our context. Starting from an initial model order vector $M_{min} = [1, ..., 1]$ where all parameters are one, the algorithm evaluates $f_1 = $ ΔAIC*(M_{min}) for the vector, further increment one of the element of M_{min} as M^+_{min}, and evaluate the lower bound $\underline{f_2} = \underline{\Delta}$AIC*($M^+_{min}$). If $\underline{f_2} > f_1$, this fact implies that no M deduced by the further increments of M^+_{min} does not have the smaller value than f_1, and hence the depth first search beyond the M^+_{min} is pruned. This pruning principle is recursively applied at every step to evaluate the model order vector and its corresponding model. As easily understood by this explanation, A* search is complete, *i.e.*, not to miss the optimal solution. The main issue of the A* search is to design an efficient lower bound $\underline{f}$ which is close to the actual value of f. From Eq.(6), ΔAIC* is the minimum when the ratio of the variances of the model prediction error is the minimum. As the variance monotonically decreases when any element in M increases, the ratio of the variance under $M_{max} = [L, ..., L]$ over the variance under the current M is the lower bound of the ratio, where L is the upper limit of the order to search. This derives the following lower bound of the ΔAIC*(M).

$$\underline{\Delta} AIC* = N\log\frac{\hat{\sigma}^2_{M_{max}}}{\hat{\sigma}^2_M} + 2. \quad (7)$$

Figure 1 shows the algorithm of this A* search where M_{opt} is the optimal model order vector and ΔAIC*(M_{opt}) the ΔAIC* under M_{opt}. The final output of this algorithm are these M_{opt} and ΔAIC*(M_{opt}).

The lower boundary of ΔAIC* given by Eq.(7) sometimes too small to efficiently prune the search space, since it is based on the error variance $\hat{\sigma}^2_{M_{max}}$ which is minimum within the search space. To obtain more efficient search performance which is not complete but sufficiently practical, we introduce a heuristic measure for the pruning as follows.

$$\underline{\Delta} AIC*^{1/n} = N\log\left(\frac{\hat{\sigma}^2_{M_{max}}}{\hat{\sigma}^2_M}\right)^{1/n} + 2. \quad (8)$$

Because the ratio of the variances always lies in $[0, 1]$, its n-root is closer to 1, and hence Eq.(8) gives a larger value than Eq.(7). Though this change does not ensure the lower boundary property of the measure, the larger value of the measure enables tighter pruning which increase the search efficiency.

Main

(1) Given $M = M_{min}$ and M_{max}.

(2) Compute ΔAIC*(M) and $\underline{\Delta}$AIC*(M).

(3) Let $M_{opt} = M$ and ΔAIC*(M_{opt})=ΔAIC*(M).

(4) If ΔAIC*(M) $\geq$ $\underline{\Delta}$AIC*(M) then
[M_{opt},ΔAIC*(M_{opt})]=
A*(M,M_{max},M_{opt},ΔAIC*(M_{opt}),1)

(5) end

Function [M_{opt},ΔAIC*(M_{opt})]=
A*(M,M_{max},M_{opt},ΔAIC*(M_{opt}),gs)

(1) for $g = gs$ to h

(2) Let $q_g = q_g + 1$.

(3) Compute ΔAIC*(M) and $\underline{\Delta}$AIC*(M).

(4) If ΔAIC*(M) $\leq$ ΔAIC*(M_{opt})
then $M_{opt} = M$ and ΔAIC*(M_{opt})=ΔAIC*(M).

(5) If ($r_g < L$) and (ΔAIC*(M) $\geq$ $\underline{\Delta}$AIC*(M))
then [M_{opt},ΔAIC*(M_{opt})]=
A*(M,M_{max},M_{opt},ΔAIC*(M_{opt}),g)

(6) end

Figure 1. A* search for optimal model order

5 Performance Evaluation

The performance of the proposed approach to determine the ARX model order is evaluated by using some artificial data sets. The data sets are generated by the following semi-ARX system containing quadratic nonlinear terms.

$$x(s) = \sum_{i=1}^{p} a_i x(s-i) + \sum_{i_2=1}^{p} a_{i_2} x^2(s-i) + \sum_{j_1=1}^{q_1} b_{j_1} y_1(s-j_1) + \sum_{j_2=1}^{q_2} b_{j_2} y_2(s-j_2) + e(s) \quad (9)$$

Two order vectors $M = [p = 3, q_1 = 3, q_2 = 4]$ and $M = [p = 6, q_1 = 7, q_2 = 6]$ are used. The coefficients of a_i, b_{j_1}, b_{j_2} has been determined by a design method of Infinite-duration Impulse Response (IIR) filter while ensuring the stability of the system[8][9]. The coefficients of the nonlinear terms a_{i_2} are set to be very small values comparing with a_i for the stability, and they are a_{i_2}=$a_i/20$ for $M = [p = 3, q_1 = 3, q_2 = 4]$ and a_{i_2}=$a_i/500$ for $M = [p = 6, q_1 = 7, q_2 = 6]$. When the data sets containing only linear dynamics are generated, all coefficients a_{i_2} are set to be 0. The total time steps for the data generation is $N = 10000$, and the time series of the input variables $y_1(s)$ and $y_2(s)$ are chosen to be a unit stepwise form or Gaussian noise having unit variance depending on the required conditions of the evaluation. Furthermore, the objective variable $x(s)$ generated by this system is distorted by adding Gaussian noises having various relative amplitudes in terms of the standard deviation of $x(s)$.

The performance to identify ARX model orders under various conditions of the data is compared between the standard AIC and our ΔAIC*. The upper limit of each order parameter for the search is set to be $L = 9$. The identified orders for the system having the order parameters $M = [p = 3, q_1 = 3, q_2 = 4]$ under step/Gaussian inputs of both $y_1(s)$ and $y_2(s)$, linear/nonlinear dynamics and various noise distortion levels are shown in Table 1 and 2 for the AIC and the ΔAIC* respectively. The results for the case of $M = [p = 6, q_1 = 7, q_2 = 6]$ are shown in Table 3 and 4. Table 1 and 3 indicate that the orders determined by the AIC tends to be significantly larger than the true order parameters when the noise and/or the nonlinearity are large. In contrast, Table 2 and 4 indicate that our approach using ΔAIC* provides almost same or slightly lower orders comparing with the true order parameters in case of Gaussian inputs. The reason of the lower order estimations in case of the step inputs is that the step inputs contain mainly low frequency signal components which tend not to affect the higher order terms of the system. In short summary, ΔAIC* derives better results than AIC for the data containing much errors.

Table 1. Orders by AIC for $M = [3, 3, 4]$

Inputs	a_{i_2}	Noise 0%	Noise 5%	Noise 20%	Noise 50%
Step	0	[3,3,4]	[9,3,4]	[9,3,3]	[9,3,2]
Gaussian	0	[3,4,5]	[9,8,9]	[9,9,9]	[9,9,9]
Gaussian	$a_i/20$	[6,8,8]	[3,5,1]	[9,9,9]	[9,9,9]

Table 2. Orders by ΔAIC* for $M = [3, 3, 4]$

Inputs	a_{i_2}	Noise 0%	Noise 5%	Noise 20%	Noise 50%
Step	0	[1,1,1]	[1,1,1]	[1,1,1]	[1,1,1]
Gaussian	0	[1,2,3]	[1,2,3]	[1,2,3]	[2,2,3]
Gaussian	$a_i/20$	[1,2,3]	[1,2,3]	[1,2,3]	[3,2,3]

Table 3. Orders by AIC for $M = [6, 7, 6]$

Inputs	a_{i_2}	Noise 0%	Noise 5%	Noise 20%	Noise 50%
Step	0	[6,5,7]	[9,2,4]	[9,2,2]	[9,1,0]
Gaussian	0	[7,8,8]	[9,9,9]	[9,8,9]	[9,9,9]
Gaussian	$a_i/500$	[8,9,9]	[9,9,9]	[8,8,9]	[9,9,9]

Table 4. Orders by $\triangle$AIC* for $M = [6, 7, 6]$

Inputs	a_{i_2}	Noise 0%	Noise 5%	Noise 20%	Noise 50%
Step	0	[4,1,1]	[3,1,1]	[2,1,1]	[2,1,1]
Gaussian	0	[5,7,7]	[5,3,8]	[3,4,4]	[5,5,5]
Gaussian	$a_i/500$	[6,3,8]	[2,3,9]	[3,4,4]	[5,5,5]

6 Analysis on Real Marketing Data

6.1 Objective data

The data was acquired through a marketing investigation. From March, 1st to June, 30th in a year, the daily sales amount of a confectionery item in a store belonging to a supermarket chain have been recorded. During the period, TV advertisements of the item were broadcasted from April, 13th to May, 3rd. In additions, the store actively promoted the item sales by placing the items on a main rack significantly exposed to the customers in some weeks including the TV advertisement period. Let the objective variable $x(s)$ be the daily sales amount of the item, the exogenous input variables $y_1(s)$ and $y_2(s)$ the daily amount of the TV advertisement and the instore sales promotion respectively. $y_1(s)$ is measured by an index named Gross Rating Point (GRP) representing the amount of the TV advertisement exposed to audiences[10]. This is evaluated by the sum of the audience rating at the times when the TV advertisements are broadcasted. GRP was constantly around 100 from April, 15th to 26th when the TV advertisements were the most actively broadcasted. $y_2(s)$ is 1 during the instore sales promotion in the store and 0 otherwise.

6.2 Performance on efficiency and accuracy

Because the $\triangle$AIC* value changes in complex manners for practical data, and the search process heavily depends on the value, the computation time for the search has been evaluated by using this real data. Table 5 shows the number of search steps and the search time of the thorough searches, the A* search using Eq.(7) and the heuristic searches using Eq.(8) with $n = 2, ..., 5$ under the upper order limit $L = 9$. Because the loop based thorough search uses a simple pointer management, it is faster than the A* search under this upper order limit. However, they becomes almost identical 630sec under $L = 15$ since the loop based thorough search has high computational complexity $O(L^{h+1})$ as mentioned earlier. Accordingly, the A* search is advantageous for the large scale problems in terms of the number of input variables and the upper order limits. The heuristic searches based on Eq.(8) search the solution far faster than the A* search. Even we apply the 4th-rooted ratio, the optimal solution can be found. Though the 5th-rooted ratio can not derive the optimum, the resulted solution is not very far from the optimum. In this regard, the A* search and its associated heuristic searches are very advantageous for the practical use.

Table 5. Search steps and times for sales data.

Algorithm	Steps	Time (sec)	p, q_1, q_2, k_1 and k_2
Thorough by loops	1000	20	1,4,1,7,0
Thorough by recursions	729	99	1,4,1,7,0
A* search	458	61	1,4,1,7,0
Search by 2nd-root	231	39	1,4,1,7,0
Search by 3rd-root	107	19	1,4,1,7,0
Search by 4th-root	54	10	1,4,1,7,0
Search by 5th-root	42	8	2,6,1,9,0

6.3 Discussion on analysis result

The impulse responses of the optimal ARX model obtained in the former subsection have been investigated to understand the dynamic relation of the sales amount of the item with the TV advertisement and the instore sales promotion. The impulse responses are a response of the sales amount under the virtual TV advertisement of a unit GRP for a day and a response of the sales amount under virtual instore sales promotion for a day. They can be estimated by introducing an impulse time series to each input variable into the ARX model. For example, the impulse response of the sales amount to the TV advertisement is derived by adding the time series of $y_1(1) = 0$,...,$y_1(s-1) = 0$,$y_1(s) = 1$,$y_1(s+1) = 0$,...,$y_1(n) = 0$ to the input of the ARX model.

Figure 2 represents the impulse responses for both the instore sales promotion and the TV advertisement. In both cases in the figure, the unit impulse is introduced on the 60th day from the beginning. Based on the upper response, the effect of one day instore promotion on the sales amount is about $+1000$yen. The lower response indicates that the effect of a unit GRP advertisement on the sales amount is $\pm 10 \sim 15$yen, and its delay is almost 9days. This time delay indicates the time interval required to impress the item among the customers by the TV advertisement. Because the standard input amplitude of the GRP is 100, the actual response of the sales amount is around $\pm 1000 \sim 1500$yen which is comparable with the effect of the instore sales promotion while the response to the instore sales promotion does not include any time delay. This is because the instore sales promotion promptly impress the item onsite. The negative response of the sales to the GRP is not consistent with our background knowledge. This occurred by the characteristics of the ARX modeling where the effects of the input variables can not be decomposed perfectly within the finite number of the data samples. This effect is called as cross talk among inputs. Even under this cross talk problem, however, the quantitative amplitudes and time delays of the responses to the input variables can be approximately known through the analysis, and the information can be used for the detailed marketing analysis and its associated marketing strategy planning. Figure 3 shows the impulse response to the TV advertisement of the ARX model derived by using the standard AIC. Though the true consumer behaviors are hardly known, it is reasonable under background knowledge in marketing field to consider that the TV advertisement affects sales amount after a certain period as mentioned earlier. Due to the selection of an excessively high model order by the AIC, the response in Figure 3 des not match with this behavior of the item sales amount, and is not interpreted by the marketing domain knowledge. Based on these observations, the performance of our AIC* approach to the ARX modeling is superior to the standard AIC under practically noisy and erroneous modeling situations.

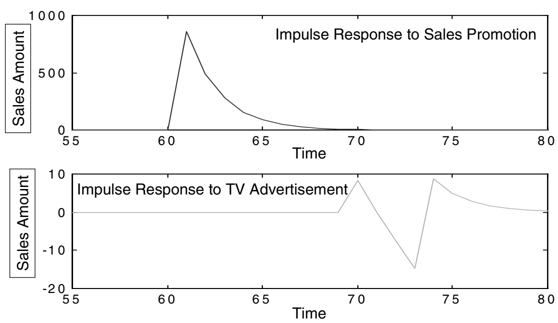

Figure 2. By ΔAIC* based ARX model.

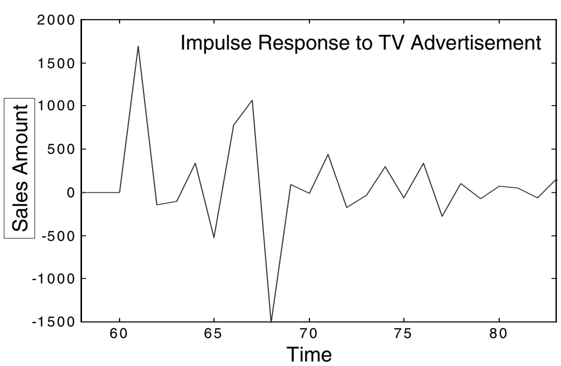

Figure 3. By AIC based ARX model.

7 Conclusion

In this paper, we proposed a novel measure named ΔAIC* to overcome the current limitations to determine the model complexity of the ARX model. In addition, we proposed some efficient complete/heuristic search algorithms to determine the optimal combinations of the model order parameters. Through the empirical evaluations, ΔAIC* is confirmed to suggest appropriate model complexity under practical conditions, and the algorithms are confirmed to derive the optimal or the semi-optimal model complexity in high efficiency. The proposed approaches are expected to provide a novel measure for the analysis of the dynamic behaviors of the customer sales in marketing.

Acknowledgment

This research was partially supported by Japan MEXT Grant-in-Aid for Scientific Research on Priority Areas, Cyber Infrastructure for the Information-explosion Era, 18049052-A01-0008, 2006.

References

[1] L. Ljung, System Identification Theory For the User, pp,71-73 (1987).

[2] H. Akaike, Fitting autoregressive models for prediction, Ann. Inst. Statist. Math. Vol.21, 243-247 (1969).

[3] J. Kuha, AIC and BIC: Comparisons of Assumptions and Performance, Sociological Methods & Research, Vol.33, No.2, 188-229 (2004).

[4] P. D. GrAunwald, I. J. Myung, and M. A. Pitt (Eds.), Advances in Minimum Description Length: Theory and Applications. MIT Press. (2004).

[5] H. Akaike, Information theory and an extension of the maximum likelihood principle. Proc. 2nd International Symposium on Information Theory (B. N. Petrov and F. Csaki eds.) Akademiai Kiado, Budapest, 267-281 (1973).

[6] P.E. Hart, N.J. Nilsson and B. Raphael, A Formal Basis for the Heuristic Determination of Minimum Cost Paths, IEEE Transactions on Systems Science and Cybernetics SSC4 (2): pp.100-107 (1968).

[7] N. J. Nilsson, Principles of Artificial Intelligence, Palo Alto, California: Tioga Publishing Company (1980).

[8] T.P. Krauss, L. Shure, J.N. Little, Signal Processing TOOLBOOX For Use with MATLAB (1994).

[9] L. Ljung, System Identification TOOLBOOX For Use with MATLA, (1996).

[10] R. Damani, C. Damani, D. Farbo and J. Linton, Online Marketing, Imano plc, London (2006).