Engineering

Industrial & Management Engineering fields

Okayama University *Year* 1996

Identification of nonlinear systems with missing data using stochastic neural network

Masahiro Tanaka
Okayama University

Proceedings of the 35th
Conference on Decision and Control
Kobe, Japan • December 1996

Identification of Nonlinear Systems with Missing Data Using Stochastic Neural Network

Masahiro Tanaka
Department of Information Technology, Okayama University
Tsushima-naka 3-1-1, Okayama 700, Japan
tanaka@mathpro.it.okayama-u.ac.jp

Abstract

In this paper, nonlinear identification is dealt with by using Gaussian sum distribution. Recently this model is called stochastic neural networks. By using the stochastic model, it is possible to estimate the output and also the missing elements in the input vector within the framework of conditional estimation. The model parameters can be estimated by using the EM algorithm. By interpolating the unknown elements, we don't have to discard the vectors including the missing elements.

1 Introduction

In the identification of dynamical systems, the number of incomplete tuples of input-output data become huge even if only a small amount of missing data occur. Thus, we may lose a lot of information if we discard incomplete tuples. Here we take an approach to interpolate the missing elements and use it for the parameter estimation of the system.

Let the output of the system be $y(k)$ and input vector $z(k)$, both subject to missing data. A vector $z(k)$ may include the past values of the output and the input vectors. Suppose y and z are jointly distributed random variables. Then with the minimum variance criterion

$$E_{y,z}[\|y-\hat{y}(z)\|^2] \le E_{y,z}[\|y-v(z)\|^2] \quad \forall v(\cdot) \tag{1}$$

we have

$$\hat{y}(z) = E[y|z] = \int y p(y|z) dy \tag{2}$$

where $p(y|z)$ is the conditional probability density function. By using the joint density function, we have

$$p(y|z) = \frac{p(y,z)}{p(z)} = \frac{p(y,z)}{\int p(y,z)dy}$$

Hence

$$\hat{y} = \frac{\int y p(y,z)dy}{\int p(y,z)dy} \tag{3}$$

Therefore, it suffices to estimate the joint probability density function in case the data complete.

Streit and Luginbuhl [5] developed the EM algorithm for learning the parameters of Gaussian mixture distribution for pattern classification problem. Parzen window [3] is a very simple method for estimating probability densities, but basically it has to have the same number of kernel functions as the learning data. Although Gaussian mixture model has much smaller number of kernel functions than parzen window, it has some difficulty in estimating the model parameters. EM algorithm is a promising approach for this problem.

In this paper, the minimum variance estimator will be derived by Gaussian mixture model. The estimate of missing variable is also derived by the same framework. Finally we propose to use the interpolation and parameter estimation alternatively for the system identification with missing elements.

2 Model

In the following, we use $x = [y^T \ z^T]^T$. The PDF of $x \in \mathbf{R}^n$ is defined by $g(x)$. The Gaussian mixture model is described by

$$g(x) = \sum_{i=1}^{G} \alpha_i p_i(x|\mu_i, \Sigma) \tag{4}$$

$$\mathrm{p_i(x|\mu_i,\Sigma)} = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-(\mathrm{x}-\mu_i)^T\Sigma^{-1}(\mathrm{x}-\mu_i)/2\right] \tag{5}$$

where α_i is the weight coefficient and $\alpha_i \ge 0$, $\sum_i \alpha_i = 1$. The learning elements in this model are $\{(\alpha_i, \mu_i), i = 1, \cdots, G\}$ and Σ.

3 EM Algorithm

If the vectors $\{x(1), x(2), \cdots, x(T)\}$ are known, it is possible to use the parameter estimation scheme that was given in [5]. The algorithm is as follows. The estimation of $w_i'(x(k))$ corresponds to the E-Step and

0-7803-3590-2/96 $5.00 © 1996 IEEE

the rest correspond to M-Step.

$$w_l'(x(k)) = \frac{\pi_l' \exp[-(x(k)-\mu_l')\Sigma^{-1'}(x(k)-\mu_l')/2]}{\sum_{i=1}^{G} \pi_i' \exp[-(x(k)-\mu_i')\Sigma^{-1'}(x(k)-\mu_i')/2]}$$

$$\pi_i = \frac{1}{T}\sum_k w_i'(x(k))$$

$$\Sigma = \frac{1}{T}\sum_{i=1}^{G}\sum_{k=1}^{T} w_i'(x(k))(x(k)-\mu_i)(x(k)-\mu_i)^T$$

$$\mu_i = \frac{\sum_{k=1}^{T} w_i'(x(k))x(k)}{\sum_{k=1}^{T} w_i'(x(k))}$$

4 Estimation of Missing Data

Substituting eqns. (4) and (5) into equation (3), we have

$$E[y|z] = \frac{\sum_i \alpha_i \int y g_i(y,z)dy}{\sum_i \alpha_i \int g_i(y,z)dy} \tag{6}$$

Suppose y is the output. Then equation (6) gives the output estimate. Suppose now y is the missing value. Then it gives the interpolation. Here we deal with both of them in the same manner. By some matrix manipulations, we have

$$\int g_i(y,z)dy = \frac{1}{(2\pi)^{(n-m)/2}}\frac{1}{|\Lambda_{zz}|^{1/2}}\exp\left\{-\frac{1}{2}(z-\bar{z}_i)^T\Lambda_{zz}^{-1}(z-\bar{z}_i)\right\} \tag{7}$$

$$\int y g_i(y,z)dy = \left(\bar{y}_i + \Lambda_{yz}\Lambda_{zz}^{-1}(z-\bar{z}_i)\right)\times \tag{8}$$

$$\times\frac{1}{(2\pi)^{(n-m)/2}}\frac{1}{|\Lambda_{zz}|^{1/2}}\exp\left\{-\tfrac{1}{2}(z-\bar{z}_i)^T\Lambda_{zz}^{-1}(z-\bar{z}_i)\right\}$$

where the parameters in the right hand side are the current estimates by the EM algorithm.

In the missing data case, we estimate it by equation (6) where y is the vector of missing elements gathered to the top. The transformation matrix is expressed $H(k)$. This expression is dependent on k because the transformation matrix $H(k)$ is k-dependent.

5 Numerical example

The plant model we use is given by ([2])

$$y(k+1) = \frac{5y(k)y(k-1)}{1+y^2(k)+y^2(k-1)+y^2(k-2)} + u(k) + 0.8u(k-1)$$

where the input was given by the uniform distribution in $(-0.5, 0.5)$. The amount of training data is 100 tuples, where

$$x(k) = [y(k+3)\ y(k+2)\ y(k+1)\ y(k)\ u(k+2)\ u(k+1)]'$$

and 10% of missing data was assumed, which were generated randomly. G is 20.

The estimate of the output was also subject to missing values. The missing values were supposed to be 15% of the input/output data. The number of training data instants whose missing elements are $0 \sim 6$ are

Table 1: Missing data instants in training data

# of missing elements	0	1	2	3	4	5	6
# of instants	52	31	14	3	0	0	0

Without our estimation-identification scheme, we could use only 52 instants for identification which may yield a very poor result.

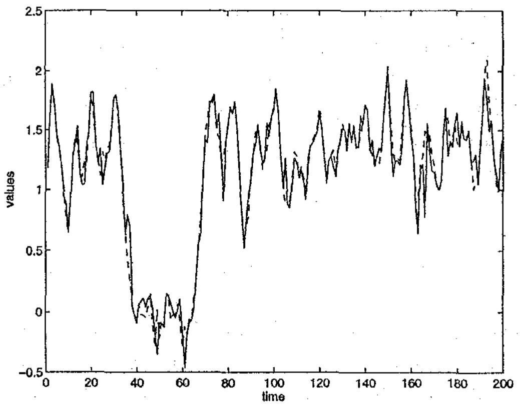

Figure 1: Output(dashdot) and the estimate(solid)

References

[1] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc.*, Vol. B-39, 1977, pp. 1-38.

[2] K. S. Narendra and K. Parthasarathy, "Identification and Control of Dynamical Systems Using Neural Networks," *IEEE Trans. Neural Networks*, Vol.1, No.1, 1990, pp. 4-27.

[3] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, Vol. 33, 1962, pp. 1065-1076.

[4] D. F. Specht, "A general regression neural network, *IEEE Trans. Neural Networks*, Vol. 2, No.6, 1991, pp. 568-576.

[5] R. L. Streit and T. E. Luginbuhl, "Maximum likelihood training of probabilistic neural networks," *IEEE Trans. Neural Networks*, Vol. 5, No. 5, 1994, pp. 764-783.