

*Research Article*

# Incomplete Time Series Prediction Using Max-Margin Classification of Data with Absent Features

**Shang Zhaowei,<sup>1</sup> Zhang Lingfeng,<sup>1</sup> Ma Shangjun,<sup>2</sup> Fang Bin,<sup>1</sup> and Zhang Taiping<sup>1</sup>**

<sup>1</sup> College of Computer Science, University of Chongqing, Chongqing 400030, China

<sup>2</sup> School of Mechatronic Engineering, Northwestern Polytechnical University, Xi'an 710072, China

Correspondence should be addressed to Shang Zhaowei, szw@cqu.edu.cn

Received 18 February 2010; Revised 24 March 2010; Accepted 20 April 2010

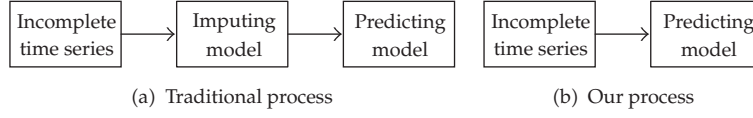
Academic Editor: Ming Li

Copyright © 2010 Shang Zhaowei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper discusses the prediction of time series with missing data. A novel forecast model is proposed based on max-margin classification of data with absent features. The issue of modeling incomplete time series is considered as classification of data with absent features. We employ the optimal hyperplane of classification to predict the future values. Compared with traditional predicting process of incomplete time series, our method solves the problem directly rather than fills the missing data in advance. In addition, we introduce an imputation method to estimate the missing data in the history series. Experimental results validate the effectiveness of our model in both prediction and imputation.

## 1. Introduction

The subjects of time series prediction have sparked considerable research activities, ranging from short-range-dependent series to long-range-dependent series [1–4], from conventional time series to fractal time series [5, 6]. Traditional predicting technologies are targeted for complete time series, such as Neural networks (NNs) [7], and Support Vector Regression (SVR) [8], and so forth. However, the time series we encountered in real life often contain missing data due to malfunctioned sensors, human factors, and other reasons. When dealing with prediction of incomplete time series, traditional process consists of two steps. The first step is to recover the incomplete time series by an imputation model, and the second step is to estimate the predicting model as complete time series. This process is shown in Figure 1(a). It may consume large number of calculations, and bring deviation for inaccurate imputation.



**Figure 1:** Two distinct processes of incomplete time series prediction.

In this paper, we propose a novel predicting model built directly by incomplete time series, which is shown in Figure 1(b).

The issue of modeling incomplete time series is interpreted as classification of data with missing features in this paper. We use the optimal hyperplane of the classification to determine the prediction values. A similar approach has been applied to prediction of complete data [9]. In addition, our model is also used as an imputation method, which means estimating the missing data of history series. There have been several works carried out on the imputation of missing data. In the following, these methods are separated into two groups.

- (1) *Statistical Methods*. Examples include Maximum likelihood (ML) algorithm [10], expectation maximization (EM) algorithm [11], Multiple Imputation [12], and so forth. Based on statistical theory, ML and EM algorithms need to know the distribution model of data and often have higher computational complexity. Multiple Imputation, which imputes the missing data  $M$  times, assumes that the missing data are Missing At Random (MAR).
- (2) *Machine Learning Methods*. Examples include mean, K-Nearest Neighbor (KNN) [13], Varies Windows Similarity Measure (VWSM) [14] and Regional Gradient Guided Bootstrapping (RGGB) [15], and so forth. KNN algorithm finds a plausible value of missing data by measuring the distance. VWSM algorithm fills the missing data by complete subsequences which are in the similar cycles. RGGB algorithm imputes the missing data by estimating the slopes of the imputing boundary regions. From the results of [15], RGGB algorithm outperforms other traditional imputation methods, such as MI and VWSM. However, it cannot be used to analyze dataset with high fluctuations.

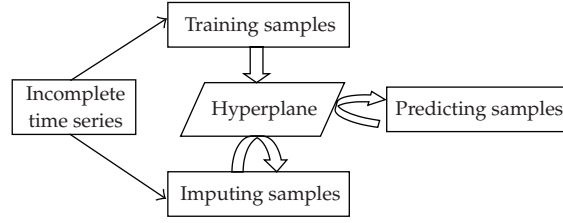
Compared with traditional imputation methods, different samples can be selected to calculate the absent data in the history series using our model, which ensures the imputing accuracy of missing data.

The rest of this paper is organized as follows. Section 2 introduces the establishment of our model. The theory of max-margin classification of data with absent features is reviewed briefly in Section 3. The solution and algorithm of our model are discussed in Section 4. Section 5 follows with the experiments, in which the prediction and imputation performance of our model are tested in detail. Finally, conclusions are presented in Section 6.

## 2. Presentation of Our Model

We start by formalizing the problem of incomplete time series. Assume that a time series with missing data is given as

$$\{x_1, x_2, ?_3, \dots, x_d, x_{d+1}, ?_{d+2}, \dots, x_{q-1}, ?_q, ?_{q+1}, \dots, x_n\}, \quad (2.1)$$



**Figure 2:** The implementation process of our model.

where “?” represents the missing data. The sample set  $X = \{X_1, \dots, X_d \mid X_{d+1}\}$  of the incomplete time series can be formulated as

$$X = \begin{pmatrix} x_1 & x_2 & \cdots & x_d & | & x_{d+1} \\ x_2 & ?_3 & \cdots & x_{d+1} & | & ?_{d+2} \\ ?_3 & x_4 & \cdots & ?_{d+2} & | & x_{d+3} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ x_{n-d} & x_{n-d+1} & \cdots & x_{n-1} & | & x_n \end{pmatrix}, \quad (2.2)$$

where  $d$  denotes the embedding dimension.

Predicting technologies usually establish regression models by  $X$ , where  $X_{d+1}$  acts as the predicting target. In order to predict the value of  $x_{k+d}$ ,  $\{x_k, x_{k+1}, \dots, x_{k+d-1}\}$  must be the input data of the model.

The implementation process of our model starts by dividing the sample set  $X$  into two parts: training set  $Xt$  and imputing set  $Xm$ . The predicting targets of the training samples in  $Xt$  are existing values, while those of the imputing samples in  $Xm$  are missing values. Training set is used to construct our predicting model. The role of imputing set is to estimate the missing values.

We construct two classes of incomplete data  $C_1$  and  $C_2$  by  $Xt$ , which can be expressed as

$$\begin{aligned} C_1 &= \{Xt_1, \dots, Xt_{p-1}, Xt_p, \dots, Xt_{d+1} + \varepsilon\}, \\ C_2 &= \{Xt_1, \dots, Xt_{p-1}, Xt_p, \dots, Xt_{d+1} - \varepsilon\}, \end{aligned} \quad (2.3)$$

where  $\varepsilon$  is the fitting error. Being the optimal hyperplane of classification of  $C_1$  and  $C_2$ ,  $H$  is obtained by the theory of max-margin classification of data with missing features. Predicting samples must fall on the hyperplane determined by the training set for a small  $\varepsilon$ ; thus the prediction values can be calculated by  $H$ .

This model can also be used to predict the missing data of incomplete time series. The imputing samples, taken from the sample set, also fall on the hyperplane. Therefore the missing values can be estimated in the same way as the prediction values. The implementation process of our model is shown in Figure 2.

In the process of imputation, each missing data can be estimated by all the samples containing it in  $X$ , not just the imputing sample in  $X_m$ . Assume that  $x_i$  is absent; the number of different samples we can use to compute the value of  $x_i$  is

$$N_i = \begin{cases} i, & 1 \leq i \leq d, \\ d + 1, & d + 1 \leq i \leq n - d, \\ n - i + 1, & n - d + 1 \leq i \leq n, \end{cases} \quad (2.4)$$

where  $N_i$  is equal to the frequency of  $x_i$  in  $X$ .

### 3. Max-Margin Classification of Data with Missing Features

In the previous discussion, the issue of modeling incomplete time series is interpreted as classification of data with missing features. In this section, we review the theory of max-margin classification of data with missing features proposed by Chechik [16].

Assume a set of samples  $x_1, \dots, x_n$  with missing features.  $y_i$  denotes the binary class label of  $x_i$ , and  $y_i \in \{-1, 1\}$ . Each sample is characterized by a subset of features from a full set  $\mathbb{F}$ .

The problem of classification can be interpreted as to find an optimal hyperplane with the max-margin framework. In the case of classification of incomplete data, the instance margin treating the margin of each instance in its own relevant subspace is defined as

$$\rho_i(w) = \frac{y_i(w^{(i)}x_i + b)}{\|w^{(i)}\|}, \quad (3.1)$$

where  $w^{(i)}$  is a vector obtained by taking the entries of  $w$  that are related to  $x_i$ . Considering the geometric margin to be the minimum over all instance margins, it comes to an optimization problem

$$\max_w \left( \min_i \frac{y_i(w^{(i)}x_i + b)}{\|w^{(i)}\|} \right). \quad (3.2)$$

Define the scaling coefficients  $s_i = \|w^{(i)}\|/\|w\|$ , and rewrite (3.2) as

$$\max_w \left( \min_i \frac{y_i(w^{(i)}x_i + b)}{\|w^{(i)}\|} \right) = \max_w \frac{1}{\|w\|} \left( \min_i \frac{y_i(w^{(i)}x_i + b)}{s_i} \right), \quad s_i = \frac{\|w^{(i)}\|}{\|w\|}. \quad (3.3)$$

For a given set of  $s_i$ , we can solve a constrained optimization problem

$$\begin{aligned} \max_{\alpha \in \mathbb{R}^n} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \frac{y_i}{s_i} \langle x_i, x_j \rangle \frac{y_j}{s_j} \alpha_j \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \quad i = 1, \dots, n, \\ & \sum_{i=1}^n \alpha_i \frac{y_i}{s_i} = 0, \end{aligned} \quad (3.4)$$

where the inner product  $\langle \cdot, \cdot \rangle$  is taken only over features that are valid for both  $x_i$  and  $x_j$ . The nonlinear classification is solved by using kernels. Thus we obtain the optimal separating hyperplane of classification of data with missing features, which is expressed as

$$\sum_{i=1}^n \alpha_i \frac{y_i}{s_i} K(x, x_i) + b = 0, \quad (3.5)$$

where  $b$  is set as in Support Vector Machines [17, 18].

## 4. Solution and Algorithm

In our model, the hyperplane of classification of data with missing data is used to compute the estimation values. Both predicting samples and imputing samples satisfy (3.5). In this section we introduce the solution and algorithm of our model.

### 4.1. Analytical Solution

Suppose a test sample  $x' = \{x'_1, x'_2, \dots, x'_d, x'_{d+1}\}$ , where  $x'_{d+1}$  is the value to be estimated. In this paper we use the kernel function  $K'(x_i, x_j) = (\langle x_i, x_j \rangle_F + 1)^2$ . Replacing the kernel of (3.5), we obtain

$$\sum_{i=1}^n \alpha_i \frac{y_i}{s_i} (\langle x', x_i \rangle_{\mathbb{F}} + 1)^2 + b = 0. \quad (4.1)$$

The simplification of (4.1) is

$$\begin{aligned} & \sum_{i=1}^n \alpha_i \frac{y_i}{s_i} x_{i,d+1}^2 \cdot x_{d+1}^2 + 2 \sum_{i=1}^n \alpha_i \frac{y_i}{s_i} \left( \sum_{j=1}^d x_{i,j} \cdot x'_j + 1 \right) x_{i,d+1} \cdot x'_{d+1} \\ & + \sum_{i=1}^n \alpha_i \frac{y_i}{s_i} \left( \sum_{j=1}^d x_{i,j} \cdot x'_j + 1 \right)^2 + b = 0, \end{aligned} \quad (4.2)$$

where the product operator “.” is taken only over features that are valid for both  $x_i$  and  $x'$ . Equation (4.2) is a quadratic equation of  $x'_{d+1}$ , and can be solved easily.

#### 4.2. Numerical Solution

Sometimes, analytical solution is meaningless or nonexistent. We need to get numerical solution of our model by iterative algorithms [19, 20]. Still take  $x'$  as an example, supposing we use Newton method, the object function of our model can be represented as

$$F(x'_{d+1}) = \sum_{i=1}^n \alpha_i \frac{y_i}{s_i} K(x', x_i) + b. \quad (4.3)$$

The iterative equation of  $x'_{d+1}$  can be expressed as

$$x'^{(i+1)}_{d+1} = x'^{(i)}_{d+1} - \frac{F(x'^{(i)}_{d+1})}{F'(x'^{(i)}_{d+1})}. \quad (4.4)$$

Therefore, the estimation values are calculated by our model effectively. Numerical solution is more complicated, but applicable in every case.

In conclusion, we have introduced the establishment and solution of our model. The key idea is to first identify a hyperplane of classification of data with missing features by incomplete time series. Then, the hyperplane is used to calculate the estimation values in predicting and imputing samples. Figure 3 provides the algorithms of our model for prediction and imputation.

### 5. Experiments

To check the validity of our model, four experiments are conducted in this section. Firstly, the prediction performance of our model is evaluated in test A. Given that conventional imputation methods usually perform distinctly when incomplete time series are missing discretely and continuously, we examine the imputation performance of our model in two missing modes in test B and test C, respectively. The performance of our model is compared with that of RGGGB and other two classical imputation methods: Mean and KNN. Finally, we verify the prediction performance of incomplete time series imputed by different models in test D.

The time series used in the experiments are Mackey-Glass time series and Henon time series. Mackey-Glass time series is generated by the chaotic equation

$$\frac{dx}{dt} = -b^*x(t) + \frac{a^*x(t-\tau)}{1+x^{10}(t-\tau)}, \quad (5.1)$$

```

Input:
  X/*sample set of incomplete time series*/
  Xp/*predicting set*/
Variables:
  Xp(i)/*the ith predicting sample*/
  xp(i)/*the ith prediction values*/
Output:
  xp/*prediction values*/
BEGIN
  (1) Construct training set Xt by X.
  (2) Create C1 and C2 by Xt.
  (3) Calculate the optimal separating hyperplane H of C1 and C2.
  (4) for each Xp(i) {
  (5) Compute xp(i) by solution (A) or solution (B).
  }
END

```

(a) Incomplete time series prediction algorithm

```

Input:
  X/*sample set of incomplete time series*/
Variables:
  Xm(i)/*the ith imputing sample*/
  xm(i)/*the ith imputation values*/
Output:
  xm/*imputation values*/
BEGIN
  (1) Construct training set Xt by X.
  (2) Create C1 and C2 by Xt.
  (3) Calculate the optimal separating hyperplane H of C1 and C2.
  (4) for each Xm(i) {
  (5) Compute xm(i) by solution (A) or solution (B).
  (6) Replace the corresponding missing data in Xm by xm(i).
  }
END

```

(b) Incomplete time series imputation algorithm

**Figure 3:** The algorithms of our model for prediction and imputation.

where parameter  $\tau$  is set to 17,  $a = 0.2$ , and  $b = 0.1$ . The interval of  $t$  is 5. Henon time series is generated by the nonlinear equation

$$x(n+1) = 1 - 1.4x(n)^2 + 0.3x(n). \quad (5.2)$$

By contrast, Henon time series has a higher volatility. The dimension of the sample set  $d = 15$ . Parameters of our model are  $\varepsilon = 0.001$  and  $C = 100$ . The value of  $K$  in KNN is set to 5.

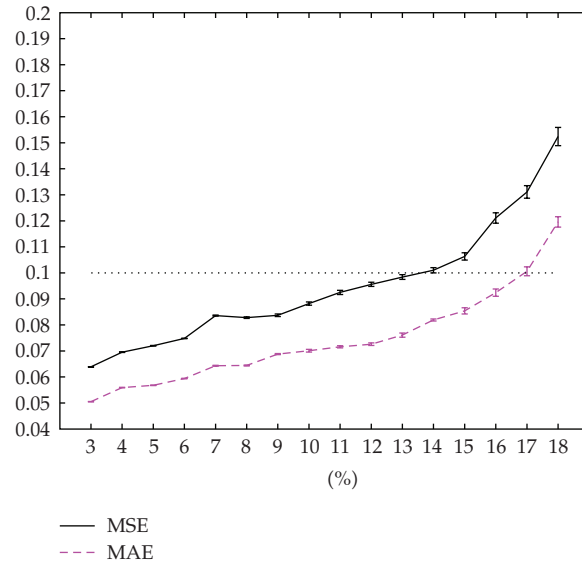


Figure 4: Prediction results of our model in Mackey-Glass time series.

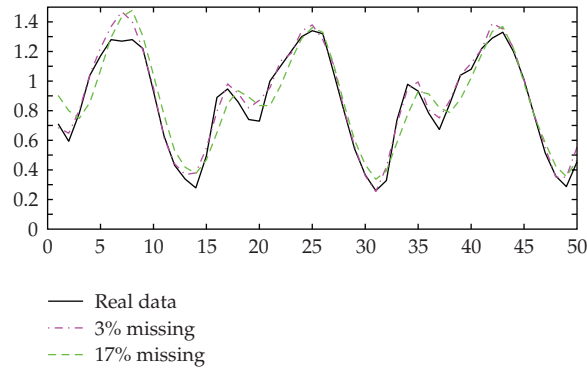


Figure 5: Prediction performance of our model in Mackey-Glass time series.

MSE (Mean Squared Error) and MAE (Mean Absolute Error MAE) are used to evaluate the performance of the experiments. All the results are obtained by repeating the algorithms 10 times.

### 5.1. Prediction of Incomplete Time Series

In this test, continuous 115 data of Mackey-Glass time series with the missing level from 3% to 18% are used to construct the initial sample set, and the next 65 data are for testing the prediction performance of our model. The prediction results are shown in Figure 4.

From Figure 4 we can see that, with the increase of the missing level, larger deviations of the prediction results occur inevitably due to the decrease of the number of training samples. However, in practice, we can use an acceptable limit of error as the basis for judgment. For example, set the acceptable limit of error  $MAE_t = 0.1$ , which equals to the



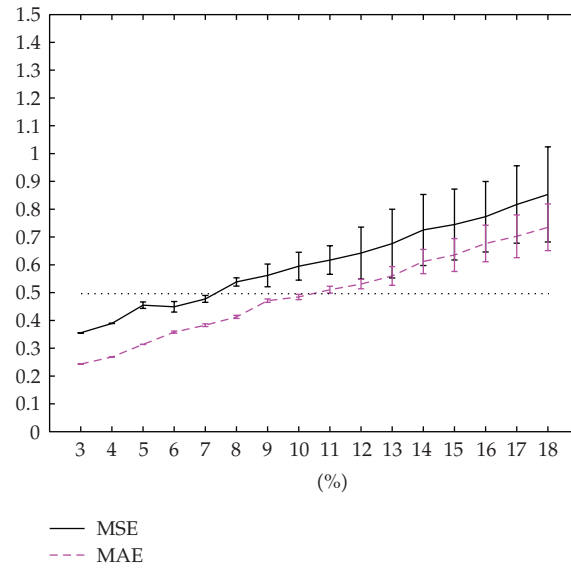


Figure 6: Prediction results of our model in Henon time series.

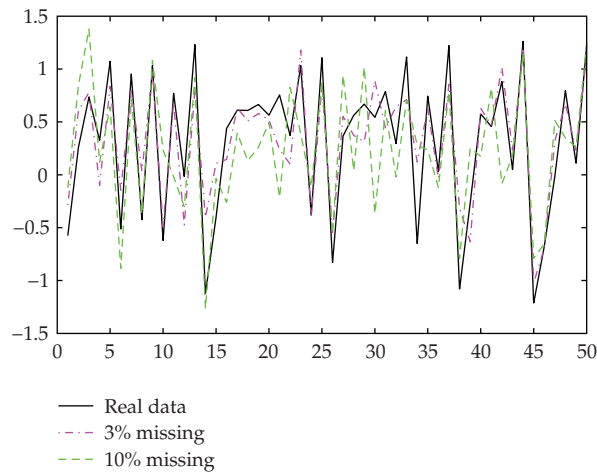


Figure 7: Prediction performance of our model in Henon time series.

minimum scale of Mackey-Glass time series. Thus our model performs well even when the missing level reaches 17%. Figure 5 shows an example of prediction performance of our model in Mackey-Glass time series with the missing level of 3% and 17%. From Figure 5, our model predicts the future time series roughly when 17% of the history data are absent. Compared with the performance of missing level at 3%, only some details are missing. Similar results are obtained in Henon time series, which are shown in Figures 6 and 7.

**Table 1:** Imputation results of Mackey-Glass time series with discrete missing data.

| Missing level | Our    |        | Mean   |        | KNN    |        | RGGB   |        |
|---------------|--------|--------|--------|--------|--------|--------|--------|--------|
|               | MSE    | MAE    | MSE    | MAE    | MSE    | MAE    | MSE    | MAE    |
| 4–6%          | 0.0716 | 0.0626 | 0.0628 | 0.0545 | 0.0728 | 0.0602 | 0.0692 | 0.0594 |
| 7–9%          | 0.0783 | 0.0744 | 0.0809 | 0.0657 | 0.0823 | 0.0612 | 0.0840 | 0.0771 |
| 10–12%        | 0.0912 | 0.0865 | 0.0969 | 0.0833 | 0.0795 | 0.0669 | 0.0979 | 0.0857 |
| 13–15%        | 0.1064 | 0.0933 | 0.1010 | 0.0836 | 0.1086 | 0.1050 | 0.1242 | 0.1098 |

**Table 2:** Imputation results of Henon time series with discrete missing data.

| Missing level | Our    |        | Mean   |        | KNN    |        | RGGB   |        |
|---------------|--------|--------|--------|--------|--------|--------|--------|--------|
|               | MSE    | MAE    | MSE    | MAE    | MSE    | MAE    | MSE    | MAE    |
| 4–6%          | 0.2385 | 0.1999 | 0.8352 | 0.6791 | 0.4624 | 0.4602 | 0.6683 | 0.5592 |
| 7–9%          | 0.4223 | 0.3120 | 0.9954 | 0.8148 | 0.6707 | 0.5534 | 0.7888 | 0.6156 |
| 10–12%        | 0.6461 | 0.4422 | 1.3951 | 1.2638 | 0.6746 | 0.6193 | 0.8321 | 0.6378 |
| 13–15%        | 0.6787 | 0.5261 | 1.4655 | 1.3114 | 0.7868 | 0.6467 | 1.2035 | 0.7639 |

## 5.2. Imputation of Incomplete Time Series with Discrete Missing Data

The continuous 115 data of Mackey-Glass time series and Henon time series with discrete missing data are used as the experimental data in this test. The imputation results of different models are shown in Tables 1 and 2.

From Tables 1 and 2 we can see that, the imputation performance of our model is similar to that of KNN and RGGB in Mackey-Glass time series. However, in Henon time series our model outperforms other three methods at every missing level. An example of imputation performance of our model over Henon time series with the missing level of 10% is shown in Figure 8.

Figure 8 shows that our model imputes most of the missing data in Henon time series effectively. Compared with other methods, the performance of our model is not sensitive to the fluctuation of time series.

## 5.3. Imputation of Incomplete Time Series with Continuous Missing Data

We evaluate the performance of different imputation methods by incomplete time series with continuous missing data in the same way. Set the maximum length of continuous missing data  $l = 5$ . The imputation results are shown in Tables 3 and 4.

Tables 3 and 4 indicate that our model outperforms other three methods in both Mackey-Glass time series and Henon time series when the data are missing continuously. Compared with Tables 1 and 2, no significant difference is observed between the two missing modes in our model, while other methods perform better in the former. Figure 9 shows an example of imputation performance of our model over Mackey-Glass time series with the missing level of 10%.

There are three sets of continuous missing data in Figure 9. Our method imputes the first two effectively. Based on the above observation, we conclude that our method performs better than other traditional technologies of imputation.

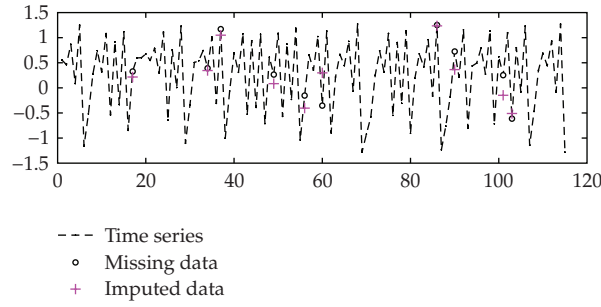


Figure 8: Imputation performance of our model in Henon time series with the missing level of 10%.

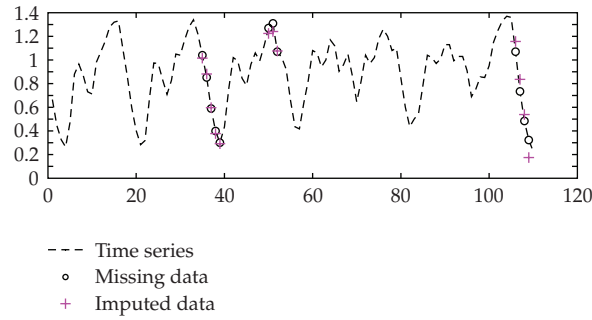


Figure 9: Imputation performance of our model in Mackey-Glass time series with the missing level of 10%.

Table 3: Imputation results of Mackey-Glass time series with continuous missing data.

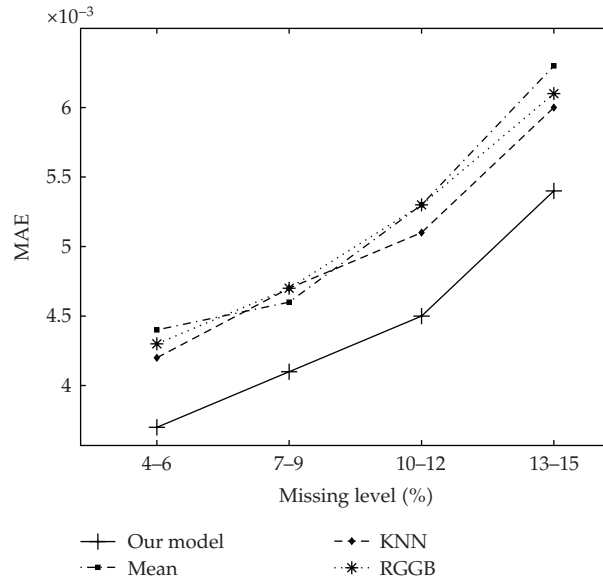
| Missing level | Our    |        | Mean   |        | KNN    |        | RGGB   |        |
|---------------|--------|--------|--------|--------|--------|--------|--------|--------|
|               | MSE    | MAE    | MSE    | MAE    | MSE    | MAE    | MSE    | MAE    |
| 4–6%          | 0.0741 | 0.0604 | 0.1044 | 0.0941 | 0.0846 | 0.0676 | 0.1296 | 0.1013 |
| 7–9%          | 0.0831 | 0.0691 | 0.1718 | 0.1374 | 0.1299 | 0.1081 | 0.1531 | 0.1320 |
| 10–12%        | 0.0832 | 0.0748 | 0.1791 | 0.1508 | 0.1091 | 0.0853 | 0.2020 | 0.1553 |
| 13–15%        | 0.1082 | 0.0814 | 0.1951 | 0.1526 | 0.1347 | 0.1119 | 0.2220 | 0.1843 |

Table 4: Imputation results of Henon time series with continuous missing data.

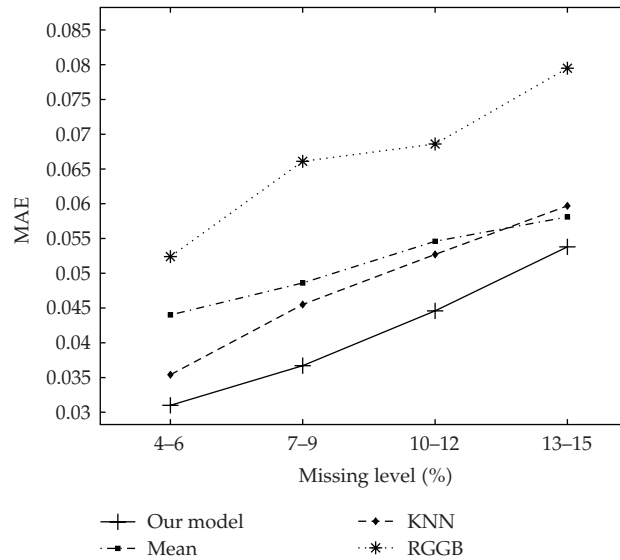
| Missing level | Our    |        | Mean   |        | KNN    |        | RGGB   |        |
|---------------|--------|--------|--------|--------|--------|--------|--------|--------|
|               | MSE    | MAE    | MSE    | MAE    | MSE    | MAE    | MSE    | MAE    |
| 4–6%          | 0.2409 | 0.1966 | 0.9633 | 0.8316 | 0.5449 | 0.4247 | 0.7294 | 0.5216 |
| 7–9%          | 0.3945 | 0.3391 | 1.0269 | 0.8957 | 0.5450 | 0.5022 | 0.8275 | 0.5777 |
| 10–12%        | 0.5472 | 0.4904 | 1.1343 | 0.9802 | 0.6777 | 0.5297 | 1.1462 | 0.7090 |
| 13–15%        | 0.8340 | 0.6190 | 1.3937 | 1.3101 | 0.7895 | 0.6711 | 1.2137 | 0.7657 |

### 5.4. Prediction after Imputation

The prediction performance of incomplete time series imputed by different models in test B and test C is evaluated in this test. We also use the next 65 data to test the prediction



**Figure 10:** Prediction results in Mackey-Glass time series imputed by different methods.



**Figure 11:** Prediction results in Henon time series imputed by different methods

performance. The error-tolerant BP algorithm is used to build the predicting model. The prediction results are shown in Figures 10 and 11.

From Figures 10 and 11 we can see that, the prediction performance in Mackey-Glass time series and Henon time series imputed by our model are both superior to that of imputed by other imputation algorithms.

## 6. Conclusions

Learning and prediction of incomplete data are still pervasive problems, although extensive studies have been conducted to improve the efficiency of data acquisition and transmission [21, 22]. We have proposed a new prediction model for incomplete time series. Experiments conducted in this paper confirm that our model can be successfully applied to prediction of incomplete time series with a missing level below than that of acceptable error limit. In addition, the imputation performance of our model is superior to that of other imputation methods, and insensitive to the fluctuation of time series. Future work may focus on applications of the model in some relevant fields [23, 24] and real-life problems.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China under the project Grants nos. 60573125, 90820306, and 60873264. The authors would like to thank the anonymous reviewers in MPE for helpful suggestions and corrections.

## References

- [1] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*, Springer Texts in Statistics, Springer-Verlag, New York, NY, USA, 2000.
- [2] M. Li and J.-Y. Li, "On the predictability of long-range dependent series," *Mathematical Problems in Engineering*, vol. 2010, Article ID 397454, 9 pages, 2010.
- [3] E. G. Bakhoun and C. Toma, "Dynamical aspects of macroscopic and quantum transitions due to coherence function and time series events," *Mathematical Problems in Engineering*, vol. 2010, Article ID 428903, 13 pages, 2010.
- [4] M. Li and W. Zhao, "Variance bound of ACF estimation of one block of fGn with LRD," *Mathematical Problems in Engineering*, vol. 2010, Article ID 560429, 14 pages, 2010.
- [5] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, Prentice Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1994.
- [6] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.
- [7] V. R. Vemuri and R. D. Rogers, *Artificial Neural Networks: Forecasting Time Series*, IEEE Computer Society Press, Los Alamitos, Calif, USA, 1993.
- [8] L. J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1506–1518, 2003.
- [9] Y. Ning, L. Zuopeng, D. Yisheng, and W. Huoli, "SVM nonlinear regression algorithm," *Computer Engineering*, vol. 31, no. 10, pp. 19–21, 2005.
- [10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [11] Z. Ghahramani and M. I. Jordan, "Supervised learning from incomplete data via an EM approach," in *Advances in Neural Information Processing Systems (NIPS 6)*, pp. 120–127, Morgan Kaufman, San Francisco, Calif, USA, 1994.
- [12] D. B. Rubin, "Multiple Imputation after 18+ years," *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 473–489, 1996.
- [13] I. Wasito and B. Mirkin, "Nearest neighbour approach in the least-squares data imputation algorithms," *Information Sciences*, vol. 169, no. 1-2, pp. 1–25, 2005.
- [14] S. Chiewchanwattana, C. Lursinsap, and C.-H. H. Chu, "Imputing incomplete time-series data based on varied-window similarity measure of data sequences," *Pattern Recognition Letters*, vol. 28, no. 9, pp. 1091–1103, 2007.
- [15] S. Prasomphan, C. Lursinsap, and S. Chiewchanwattana, "Imputing time series data by regional-gradient-guided bootstrapping algorithm," in *Proceedings of the 9th International Symposium on*

- Communications and Information Technology (ISCIT '09)*, pp. 163–168, Incheon, South Korea, September 2009.
- [16] G. Chechik, G. Heitz, G. Elidan, P. Abbeel, and D. Koller, “Max-margin classification of data with absent features,” *Journal of Machine Learning Research*, vol. 9, pp. 1–21, 2008.
  - [17] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
  - [18] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization Optimization and Beyond*, MIT Press, Cambridge, Mass, USA, 2002.
  - [19] W.-S. Chen, B. Pan, B. Fang, M. Li, and J. Tang, “Incremental nonnegative matrix factorization for face recognition,” *Mathematical Problems in Engineering*, vol. 2008, Article ID 410674, 17 pages, 2008.
  - [20] L. Q. Qi and J. Sun, “A nonsmooth version of Newton’s method,” *Mathematical Programming*, vol. 58, no. 1–3, pp. 353–367, 1993.
  - [21] S. Y. Chen, Y. F. Li, and J. Zhang, “Vision processing for realtime 3-D data acquisition based on coded structured light,” *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 167–176, 2008.
  - [22] J. A. C. Bingham, “Multicarrier modulation for data transmission: an idea whose time has come,” *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, 1990.
  - [23] M. Li and W. Zhao, “Representation of a stochastic traffic bound,” *IEEE Transactions on Parallel and Distributed Systems*. In press.
  - [24] G. Mattioli, M. Scalia, and C. Cattani, “Analysis of large amplitude pulses in short time intervals: application to neuron interactions,” *Mathematical Problems in Engineering*. In press.