Optimal Selection of Parameters for Nonuniform Embedding of Chaotic Time Series Using Ant Colony Optimization

Meie Shen, Wei-Neng Chen, *Member, IEEE*, Jun Zhang, *Senior Member, IEEE*, Henry Shu-Hung Chung, *Senior Member, IEEE*, and Okyay Kaynak, *Fellow, IEEE*

***Abstract*—The optimal selection of parameters for time-delay embedding is crucial to the analysis and the forecasting of chaotic time series. Although various parameter selection techniques have been developed for conventional uniform embedding methods, the study of parameter selection for nonuniform embedding is progressed at a slow pace. In nonuniform embedding, which enables different dimensions to have different time delays, the selection of time delays for different dimensions presents a difficult optimization problem with combinatorial explosion. To solve this problem efficiently, this paper proposes an ant colony optimization (ACO) approach. Taking advantage of the characteristic of incremental solution construction of the ACO, the proposed ACO for nonuniform embedding (ACO-NE) divides the solution construction procedure into two phases, i.e., selection of embedding dimension and selection of time delays. In this way, both the embedding dimension and the time delays can be optimized, along with the search process of the algorithm. To accelerate search speed, we extract useful information from the original time series to define heuristics to guide the search direction of ants. Three geometry- or model-based criteria are used to test the performance of the algorithm. The optimal embeddings found by the algorithm are also applied in time-series forecasting. Experimental results show that the ACO-NE is able to yield good embedding solutions from both the viewpoints of optimization performance and prediction accuracy.**

***Index Terms*—Ant colony optimization (ACO), attractor embedding, nonuniform embedding, phase space reconstruction, time series.**

I. Introduction

A TIME series is a chronological sequence of data points that are measured on a particular variable at successive times. Due to the broad application of time series, time-series analysis and forecasting have attracted a considerable amount of attention during the last decades [1]–[4]. Usually, time series are driven by some underlying dynamical systems, in which the state changes with time as a function of its current state [5]. As the dynamical system evolves, the set of attracting states forms an attractor. The intrinsic dynamical systems of many real-world observed time series are usually nonlinear [6]. The states of a nonlinear dynamical system may evolve to a chaotic attractor, exhibiting a highly irregular geometrical pattern and becoming very sensitive to initial conditions. As such, the analysis of chaotic time series has become a challenge in many fields of science and engineering [7], [8].

One of the most important techniques for investigation of chaotic time series is time-delay embedding [1], [9]. The technique is theoretically supported by the celebrated embedding theorem of Takens [10], [11]. The theorem showed that, for a time series of scalar observations of a dynamical system, it is possible to reconstruct its underlying dynamics in the phase space using a time-delay embedding with a sufficiently large dimension. To correctly reconstruct the attractor, a key issue is to find suitable embedding parameters, i.e., the embedding dimension m and the time delay τ. Various methods for the optimal selection of embedding parameters have been proposed [12]–[18]. However, even with these methods, it is found that no single method outperforms all others in all situations [1], [6], and the research on proposing improved methods for phase space reconstruction continues.

In the aforementioned methods, the time delay in each dimension is assumed to uniformly grow by the same value τ. Therefore, this kind of embedding methods is called *uniform embedding*. Although uniform embedding is simple and effective for classic chaotic dynamical systems such as the Lorenz and Rossler systems, Judd and Mees [19] pointed out that uniform embedding has difficulties in dealing with the time series with multiple periodicities. To overcome the deficiency of uniform embedding, they proposed the *nonuniform embedding* technique [19]. Different from the conventional uniform embedding methods, nonuniform embedding allows different dimensions to use different time delays. As a result, it manages to deal with the time series with multiple time scales and provides more accurate predictions [19]–[24].

Although nonuniform embedding is flexible and effective, it brings in a new problem. That is, not a single time-delay

Manuscript received February 2, 2012; revised June 24, 2012 and August 30, 2012; accepted September 7, 2012. Date of publication October 23, 2012; date of current version April 16, 2013. This work was supported in part by the National Natural Science Foundation of China under Grant U0835002 and Grant 61070004 and in part by the Scholarship Award for Excellent Doctoral Student granted by the Ministry of Education, China. This paper was recommended by Associate Editor H. Takagi. *(Corresponding author: W.-N. Chen)*.

M. Shen is with the School of Computer Science, Beijing Information Science and Technology University, Beijing 100081, China.

W.-N. Chen and J. Zhang are with the Department of Computer Science, Sun Yat-sen University, Guangzhou 510275, China (e-mail: chenwn3@mail.sysu.edu.cn).

H. S.-H. Chung is with the City University of Hong Kong, Kowloon, Hong Kong.

O. Kaynak is with the Department of Electrical and Electronic Engineering, Bogazici University, Istanbul 34342, Turkey.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCB.2012.2219859

value τ but a group of time-delay values τ_i $(i = 2, 3, \ldots, m)$ has to be optimally selected. Supposing the maximum value of a time delay is $\tau_{\max}$, the number of all possible combinations for $(\tau_2, \tau_3, \ldots, \tau_m)$ is $(\tau_{\max})^{m-1}$. In other words, the selection of time delays in nonuniform embedding is a problem where the solution space exponentially grows with the increase in the embedding dimension m. In this case, the traditional methods for uniform embedding fail to work. By now, there is still lack of standard methods for selecting dimension and time delays for nonuniform embedding. In [22], a minimum description length (MDL) method was proposed to evaluate the performance of an embedding, and a simple binary genetic algorithm (GA) was developed. Another GA approach was designed by Vitrano and Povinelli [23], but the performance of these algorithms is still not satisfactory. For the problem of time-series forecasting with fuzzy inference systems, some specialized deterministic or stochastic methods for selecting embedding dimension and time delays also exist [5], [8], [24].

In order to find suitable embedding parameters for nonuniform embedding efficiently, this paper aims to develop an ant colony optimization (ACO) approach. The ACO is a swarm intelligence method proposed by Dorigo [25], [26] in inspiration of the foraging behavior of ant colonies. Since its proposal in the early 1990s, ACO has evolved into an important optimization technique and has been successfully applied to a wide range of combinatorial optimization problems [27]–[31].

The proposed ACO for nonuniform embedding, which is termed as ACO-NE, takes advantages of the characteristics of the ACO to find suitable embedding parameters. First, the ACO has the characteristic of incremental solution construction [32], i.e., it builds solutions to the problem by adding components to the partial solutions step by step incrementally. Based on this characteristic, the ACO-NE divides the solution construction procedure of the ACO into two phases, i.e., selection of embedding dimension and selection of time delays. Ants in the ACO-NE first choose a suitable embedding dimension and then select a set of time delays according to the dimension number to build a complete solution. In this way, both the embedding dimension and the time delays can be optimized together in the proposed ACO-NE. Second, the ACO enables the use of problem-based heuristics to facilitate the search process [33]. Thereby, instead of selecting the parameters in a completely random way, it is possible to extract useful information from the original time series to provide guidance for the selection of the embedding dimension and the time delays. In the ACO-NE, a heuristic is defined to favor smaller dimensions for embedding dimension selection, and some nearest-neighborhood-based heuristics are developed for the selection of time delays.

Although several criteria have been proposed in the literature for evaluating the performance of a nonuniform embedding [23], [24], so far, none of these criteria is widely used in the research on nonuniform embedding as a unified standard criterion. Therefore, to test the performance of the ACO-NE, three criteria are applied as the fitness function of the algorithm, including the minimal neighborhood distance (MND) criterion, the minimal false nearest neighborhood (MFNN) criterion, and the MDL criterion. In addition, the optimal embeddings found by the proposed algorithm are applied in time-series forecasting. In the experiment, the ACO-NE is compared with the traditional uniform embedding method and the GA approaches [22], [23]. Experimental results show that the ACO-NE is able to yield good embedding solutions from both the viewpoints of optimization performance and prediction accuracy. In addition, it is found that the ACO-NE can always find solutions with lower dimensions, which is more convenient for time-series analysis. These results demonstrate the effectiveness of the proposed approach.

The rest of this paper is organized as follows. In Section II, the background of time-delay embedding is introduced. Section III proposes the ACO-NE algorithm. Experimental results are given in Section IV. The conclusion finally comes in Section V.

II. Background

A. Uniform Time-Delay Embedding

Let $\{x(t)\}_{t=1}^{N} = \{x(1), x(2), \ldots, x(N)\}$ be a scalar time series of N observations. The uniform time-delay embedding method is to find an embedding dimension m and a time delay τ to obtain a set of vectors, i.e.,

$$v_t = (x(t), x(t-\tau), x(t-2\tau), \ldots, x\left(t-(m-1)\tau\right)), \quad t = N, N-1, \ldots \tag{1}$$

For a properly selected τ, if the number of observations N and the embedding dimension m are sufficiently large, Takens theorem [10] guarantees that the trajectory of v_t in the phase space is topologically equivalent to the original dynamical system of the scalar time series. In this sense, we can reconstruct the dynamics of the time series in the phase space based on the set of vectors v_t and use it to forecast the future variable values of the time series.

B. Nonuniform Time-Delay Embedding

Although the uniform embedding technique is simple and effective, its performance becomes poor for complex time series with multiple periodicities [19]–[22]. Judd and Mees [19] suggested that nonuniform embedding is a better and more general approach for the reconstruction of dynamics of chaotic time series. Different from uniform embedding, the nonuniform embedding technique uses different time delays in different dimensions. Suppose the time delay for the jth dimension is τ_j, vector v_t reconstructed in the phase space is

$$v_t = (x(t), x(t-\tau_2), x(t-\tau_3), \ldots, x(t-\tau_m)). \tag{2}$$

In this way, the nonuniform embedding is able to overcome the deficiency of uniform embedding.

In nonuniform embedding, there is a combinatorial explosion of possible settings for $(\tau_2, \tau_3, \ldots, \tau_m)$ as the embedding dimension m increases. Therefore, traditional methods for selecting m and τ in uniform embedding are not applicable. Several works have been done in recent years to tackle this intractable parameter optimization problem. In [22], a GA approach was proposed. The algorithm works by specifying a

maximum embedding window $\tau_{\max}$ and encoding the solution as a binary string with the length of $\tau_{\max}$. The main deficiency of this approach is that its performance is seriously affected by the maximum embedding window $\tau_{\max}$. If the time series requires a large $\tau_{\max}$, the GA usually produces solutions with a large embedding dimension, which are impractical for time-series forecasting. Another GA approach for the embedding parameter selection problem was proposed by Vitrano and Povinelli [23]. The algorithm fixes the length of a chromosome as a predefined maximum embedding dimension. Since the encoding length of a chromosome is fixed but only some of the dimensions are used, the efficiency of the algorithm is not satisfactory.

III. ACO-NE Algorithm

A. Basic Idea of the Algorithm

The basic idea of the ACO is to simulate the foraging behavior of ant colonies. When a group of ants set out from their nest to search for a path to food, they use a special kind of chemical termed pheromone to communicate with each other. Once the ants discover a path to food, they deposit pheromones on the path to record their previous search experience. By sensing pheromones on the path, an ant can follow the trails of other ants to find the food source. As this process continues, the shortest path to the food gradually accumulates more and more pheromones, attracting more and more ants to choose. In this way, although the capability of a single ant is limited, the whole ant colony is able to cooperate to find the best path to the food. Inspired by such social behavior of ants, the ACO was proposed in the early 1990s [25] and has become an important swarm intelligence technique.

Based on the framework of the ACO, an ACO-NE algorithm is designed for the optimal selection of parameters for nonuniform embedding. In particular, the ACO-NE has two main characteristics. First, as the considered problem has two kinds of parameters, i.e., the embedding dimension and the time delays, we divide the procedure of solution construction in the ACO-NE into two stages, i.e., selection of embedding dimension and selection of time delays. In this way, both of these two kinds of parameters can be optimized together in the ACO-NE. Second, as the framework of ACO enables the use of problem-based heuristics to facilitate the search process of the algorithm, we define an effective heuristic for ACO-NE by extracting useful information from the original time series. In order to evaluate the performance of nonuniform embeddings, we use three kinds of criteria as the objective functions of the ACO-NE algorithm, namely, the MND, MFNN, and MDL criteria. The criteria will be described in Section III-D.

Overall, the ACO-NE algorithm has the following steps:

Step 1) Initialization—The parameters of the algorithm are initialized.

Step 2) Solution Construction—During each generation, a group of ants set out to find suitable embedding parameters for the nonuniform time-delay embedding problem.

Step 3) Evaluation—The solutions built by ants are evaluated based on a fitness function. In the proposed approach, we test three different fitness functions of the considered problem, namely, the MND, MFNN, and MDL criteria.

Step 4) Pheromone Management—The pheromone values are updated based on the local pheromone update rule and the global pheromone update rule.

Step 5) Terminal Test—If a predefined number of iterations have been processed, the algorithm ends. Otherwise, the algorithm continues to run steps 2–4 iteratively.

B. Solution Construction

In each iteration of the ACO-NE, a group of M ants is dispatched to build M solutions to the problem. We denote the solution built by the ith ant as

$$S(i) = \big(m(i)|\tau_1(i), \tau_2(i), \tau_3(i), \ldots, \tau_{m(i)}(i)\big) \tag{3}$$

where $m(i)$ is the embedding dimension found by the ith ant, and $\tau_1(i), \tau_2(i), \tau_3(i), \ldots, \tau_{m(i)}(i)$ are the time delays. (Note that the value of $\tau_1(i)$ is 0 as we usually use $x(t)$ as the first dimension of the reconstructed vector in the phase space.)

As the dimension number $m(i)$ is a variable to be optimized, the number of variables in the solution $S(i)$ is unknown. Thus, fixing the encoding length of $S(i)$ is not suitable. In order to optimize both the embedding dimension and the time delays together, the ACO-NE divides the solution construction procedure into two phases. First, the ant selects the embedding dimension $m(i)$ based on some pheromones and heuristics. Then, the ant selects $m(i) - 1$ time delays from the embedding window to compose the time-delay vector $(\tau_2(i), \tau_3(i), \ldots, \tau_{m(i)}(i))$.

More specifically, the procedure of solution construction is described in detail as follows.

1) Selection of Embedding Dimension: To construct an embedding, the ant first needs to determine the embedding dimension $m(i)$. Since a large embedding dimension usually causes much higher computational burden in time-series analysis and forecasting [23], building an embedding with a very large dimension is meaningless. Therefore, it is reasonable to limit the embedding dimension to a maximum value MaxD. The selection of dimension is actually to assign $m(i)$ to a value from $\{2, 3, \ldots, \text{MaxD}\}$.

In the ACO-NE, the embedding dimension is selected using the roulette wheel selection (RWS) scheme [25]. The selection scheme is implemented as follows.

First, for every possible choice k ($k \in \{2, 3, \ldots, \text{MaxD}\}$), the selection scheme evaluates probability $P(m(i) = k)$ of setting the embedding dimension $m(i)$ as k. Probability $P(m(i) = k)$ is evaluated based on pheromone and heuristic values as

$$P\left(m(i) = k\right) = \begin{cases} \dfrac{pd_k \cdot \frac{1}{\sqrt{k}}}{\sum_{l=2}^{\text{MaxD}} pd_l \frac{1}{\sqrt{l}}}, & \text{if } 2 \le k \le \text{MaxD} \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Here, pd_k is the pheromone value of setting $m(i) = k$. In the ACO, pheromone can be regarded as the previous search experience of ants. The pheromone values are updated in every

iteration of the ACO algorithm, which will be introduced in detail in Section III-E. A large pheromone value pd_k means that the previous search experience of ants regards $m(i) = k$ as a good choice and vice versa. The heuristic is some problem-based information, which is helpful for guiding the search direction of ants. Because an embedding with a smaller embedding dimension is more practical and convenient for time-series analysis, in (4), the heuristic value is set to $1/\sqrt{k}$ to favor smaller dimension numbers. Based on (4), a smaller dimension number k with a larger pheromone value pd_k would have a larger probability $P(m(i) = k)$.

After the probabilities of all possible k ($k \in \{2, 3, \ldots, \mathrm{MaxD}\}$) values have been evaluated, the RWS scheme randomly chooses a dimension from $\{2, 3, \ldots, \mathrm{MaxD}\}$ based on their probabilities.

2) Selection of Time Delays: After determining the embedding dimension $m(i)$, the ant selects $m(i) - 1$ time delays to complete the construction of an embedding. The time delays should satisfy

$$1 \le \tau_2(i) < \tau_3(i) < \cdots < \tau_{m(i)}(i) \le \tau_{\max} < N \tag{5}$$

where $\tau_{\max}$ is the upper bound of the embedding window, and N is the size of the time series. Usually, $\tau_{\max}$ can be arbitrarily set to a sufficiently large number that satisfies $\tau_{\max} > m\tau$ [22], where m and τ are the embedding dimension and the time delay yielded by the conventional uniform embedding methods such as false nearest neighborhood (FNN) [12] and average mutual information (AMI) [16]. To select $m(i) - 1$ time delays, the ant repeats the following steps for $m(i) - 1$ times:

Step a) For $k = 1, 2, \ldots, \tau_{\max}$, the value of $\mathrm{pt}_k \mathrm{ht}_k^{\beta}$ is evaluated. Here, pt_k is the pheromone of selecting k as a time delay, ht_k is the heuristic of selecting k as a time delay, and β is a parameter of the ACO algorithm to weigh the importance of heuristics. As mentioned before, in the ACO, pheromone represents the previous search experience of ants, and heuristics are some problem-based useful information.

Step b) The time delay is selected by

$$\text{time\_delay} = \begin{cases} \arg\max_{k=1,2,\ldots\tau_{\max}} \\ \quad \mathrm{pt}_k \mathrm{ht}_k^{\beta} \cdot \Omega(k), & \text{if ran} < q_0 \\ \text{using the RWS scheme,} & \text{otherwise} \end{cases} \tag{6}$$

where function $\Omega()$ is defined as

$$\Omega(k) = \begin{cases} 0, & \text{if } k \text{ has been selected as a time delay} \\ 1, & \text{otherwise} \end{cases} \tag{7}$$

and $q_0 \in (0, 1)$ is a parameter that is the probability of directly choosing the delay with the largest value of $\mathrm{pt}_k \mathrm{ht}_k^{\beta}$.

To perform (6), a random number ran uniformly distributed in (0,1) is generated and is compared with q_0. If ran $< q_0$, the time delay that maximizes the value of $\mathrm{pt}_k \mathrm{ht}_k^{\beta}$ is selected.

Otherwise, the RWS scheme for time-delay selection is applied. The RWS is to select a time delay based on the probabilities defined in (8). For $k = 1, 2, \ldots, \tau_{\max}$, the probability of selecting k as a time delay is given by

$$P(\tau_j = k) = \frac{\mathrm{pt}_k \mathrm{ht}_k^{\beta} \cdot \Omega(k)}{\sum_{l=1}^{\tau_{\max}} \mathrm{pt}_l \mathrm{ht}_l^{\beta} \cdot \Omega(l)}. \tag{8}$$

After running these steps for $m(i) - 1$ times, $m(i) - 1$ different time delays are finally selected. We reorder these selected time delays in ascending order and assign them to $\tau_2(i), \tau_3(i), \ldots, \tau_{m(i)}(i)$. In this way, all embedding parameters are set by the ant, and the time delays are guaranteed to satisfy $1 \le \tau_2(i) < \tau_3(i) < \cdots < \tau_{m(i)}(i) \le \tau_{\max} < N$.

C. Definition of Heuristics

One unique characteristic of the ACO is that it enables the use of heuristics to accelerate the search. Therefore, we can extract some useful problem-based information to define heuristics and use them in the selection schemes of the ACO-NE.

In the scheme (4) for the selection of an embedding dimension, as a smaller dimension is more convenient for time-series analysis, the reciprocal of the square root of the candidate dimension is defined as the heuristic to favor smaller embedding dimensions.

On the other hand, for the selection of time delays in (6)–(8), we need to study the characteristics of the time series in-depth to extract meaningful and useful heuristics. In the proposed algorithm, we define two types of heuristics, i.e., the two-dimensional-based heuristic (2-D heuristic) and the three-dimensional-based heuristic (3-D heuristic). The underlying ideas of these two heuristics are derived from the concept of nearest neighborhood [12].

For $k = 1, 2, \ldots, \tau_{\max}$, the 2-D heuristic of selecting k is given in (9), shown at the bottom of the next page.

The equation actually shows that heuristic ht_k is in inverse proportion to the square of the distance between vector $(x(N), x(N-k))$ and its nearest neighborhood vector $(x(N-l), x(N-l-k))$ in the 2-D phase space. In addition, to make the embedding more convenient, short time delays are more preferred [22]. Therefore, the $1/\sqrt{k}$ item is added to the heuristic to favor short time delays. In fact, if $(x(N), x(N-k))$ has a very close neighborhood vector, it implies that k may possibly be a promising time delay that can lead to a good embedding in terms of geometry. Therefore, this heuristic encourages ants to select the delays that can find closer nearest neighborhood points of $(x(N), x(N-k))$.

We also need to notice that the nearest neighborhood of $(x(N), x(N-k))$ in the 2-D phase space may be a false neighborhood point. In other words, $(x(N), x(N-k))$ and its nearest neighborhood point in the 2-D phase space may not be close to each other anymore when the embedding is unfolded to the phase space with higher dimension. In this case, the 2-D heuristic defined in (9), shown at the bottom of the next page, may induce a misleading effect. To reduce such effect, we also extend the 2-D heuristic to a 3-D version given in (10), shown at the bottom of the next page.

In (10), heuristic ht_k is in inverse proportion to the square of the distance between vector $(x(N), x(N-k), x(N-2k))$ and its nearest neighborhood vector $(x(N-l), x(N-l-k), x(N-l-2k))$ in the 3-D phase space. As we consider one more dimension, the chance of the emergence of false neighborhood points is reduced, and thus, the 3-D heuristic is more reliable. On the other hand, the computational cost of the 3-D heuristic is also higher than the 2-D heuristic. In general, we find that the 3-D heuristic is already enough to accelerate the search speed of the ACO-NE algorithm. Thus, we do not need to define the heuristics in higher dimensions.

Finally, it is important to note that the heuristics defined in (9) and (10) are all static. In other words, once the values of these heuristics are initialized, the values remain unchanged during the whole process of the algorithm. Therefore, we only need to initialize the heuristics once at the beginning of the algorithm. In this sense, the evaluation of heuristics is not a time-consuming task for the proposed ACO-NE.

D. Evaluation

Evaluating the performance of the solutions built by the algorithm is an important step in swarm intelligence techniques. However, so far, there is still lack of a commonly used criterion for evaluating the performance of a nonuniform embedding. In the literature, the existing criteria can be divided into two main types, i.e., geometry-based criteria and model-based criteria. Geometry-based criteria consider the geometrical characteristic of the reconstructed embedding in the phase space and the representatives include the FNN [12] and the Euclidean-distance-based method [23]. Model-based criteria consider the embedding as a model, and the typical model-based criterion is the MDL [19], [20]. In order to test the performance of the proposed algorithm in a more comprehensive way, we test these three criteria in our proposed ACO-NE algorithm, respectively.

1) MND: The MND is a geometry-based criterion defined in this paper using the Euclidean distances between the vectors and their nearest neighborhoods in the phase space. Given a solution $S(i) = (m(i)|0, \tau_2(i), \tau_3(i), \ldots, \tau_{m(i)}(i))$, to evaluate its performance using the MND criterion, we first need to reconstruct the vectors in the phase space, i.e.,

$$v_t(i) = \left(x(t), x\left(t-\tau_2(i)\right), x\left(t-\tau_3(i)\right), \ldots, x\left(t-\tau_{m(i)}(i)\right)\right), \quad t = N, N-1, \ldots, 1+\tau_{m(i)}(i). \tag{11}$$

Then, for each vector $v_t(i)$, we find its nearest neighborhood and denote the square of the distance between $v_t(i)$ and its nearest neighborhood as $\mathrm{ND}(v_t(i))$. $\mathrm{ND}(v_t(i))$ can be formulated as

$$\mathrm{ND}\left(v_t(i)\right) = \min_{l=N,N-1,\ldots,t+1,t-1,\ldots,1+\tau_{m(i)}(i)} \left([x(t)-x(l)]^2 + \sum_{j=2}^{m(i)} \left[x\left(t-\tau_j(i)\right) - x\left(l-\tau_j(i)\right)\right]^2 \right). \tag{12}$$

The fitness of $S(i)$ in terms of the MND criterion is defined as

$$f_{\mathrm{MND}}\left(S(i)\right) = \frac{\sum_{t=1+\tau_{m(i)}(i)}^{N} \mathrm{ND}\left(v_t(i)\right)}{N-\tau_{m(i)}(i)} \cdot \frac{1}{\sqrt{m(i)}}. \tag{13}$$

In (13), $N-\tau_{m(i)}(i)$ is the number of vectors in the phase space. The first factor in the right item of (13) means the mean nearest neighborhood distance averaged on all vectors. Obviously, the smaller the mean nearest neighborhood distance is, in terms of geometry, the better the embedding will be. Because a large embedding dimension usually induces a larger nearest neighborhood distance, in order to make the criterion fair for the solutions with different embedding dimensions, we divide the mean nearest neighborhood distance by $\sqrt{m(i)}$ in (13).

2) MFNN: FNN [12] is an important geometry-based technique for determining the embedding dimension of traditional uniform time-delay embedding. In this paper, we further use the technique to define the MFNN criterion for evaluating the performance of a nonuniform embedding.

In traditional uniform embedding, we assume that the nearest neighborhood of vector $v_i(t) = (x(t), x(t-\tau), \ldots, x(t-(m-1)\tau))$ in the m-dimensional phase space is $(x(l), x(l-\tau), \ldots, x(l-(m-1)\tau))$ and the distance between these two vectors is $\mathrm{ND}(v_t(i))$. In the $m+1$-dimensional phase space, the vectors are extended to $(x(t), x(t-\tau), \ldots, x(t-m\tau))$ and $(x(l), x(l-\tau), \ldots, x(l-m\tau))$, respectively. If

$$\frac{|x(t-m\tau) - x(l-m\tau)|}{\mathrm{ND}\left(v_t(i)\right)} > T \tag{14}$$

$$\mathrm{ht}_k = \frac{1}{\sqrt{k}} \cdot \frac{1}{\min_{l=1,2,\ldots,N-k-1} \left(\sum_{a=0}^{1} [x(N-ak) - x(N-l-ak)]^2 \right)}. \tag{9}$$

$$\mathrm{ht}_k = \frac{1}{\sqrt{k}} \cdot \frac{1}{\min_{l=1,2,\ldots,N-2k-1} \left(\sum_{a=0}^{2} [x(N-ak) - x(N-l-ak)]^2 \right)}. \tag{10}$$

we consider $(x(l), x(l-\tau), \ldots, x(l-m\tau))$ is a FNN of $(x(t), x(t-\tau), \ldots, x(t-m\tau))$. Here, T is a threshold, and Abarbanel [1] suggested that $T = 15$ is suitable for a variety of chaotic systems.

To use the definition of FNN in nonuniform embedding, supposing the nearest neighborhood of $(x(t), x(t-\tau_2), \ldots, x(t-\tau_m))$ in the m-dimensional phase space is $(x(l), x(l-\tau_2), \ldots, x(l-\tau_m))$, we extend the vectors in the $m+1$-dimensional phase space as $(x(t), x(t-\tau_2), \ldots, x(t-\tau_m), x(t-\tau_m-\tau_2))$ and $(x(l), x(l-\tau_2), \ldots, x(l-\tau_m), x(l-\tau_m-\tau_2))$. Similarly, if

$$\frac{|x(t-\tau_m-\tau_2) - x(l-\tau_m-\tau_2)|}{\mathrm{ND}\,(v_t(i))} > T \tag{15}$$

we consider that $(x(l), x(l-\tau_2), \ldots, x(l-\tau_m), x(l-\tau_m-\tau_2))$ is a FNN of $(x(t), x(t-\tau_2), \ldots, x(t-\tau_m), x(t-\tau_m-\tau_2))$.

Based on the aforementioned definition of FNNs in nonuniform embedding, to evaluate the fitness of solution $S(i) = (m(i), \tau_2(i), \tau_3(i), \ldots, \tau_{m(i)}(i))$ in terms of the MFNN criterion, we first reconstruct the vectors in the phase space and compute the nearest neighborhoods of all vectors following (11) and (12). Then, we use (15) to find out the FNNs. The fitness is then defined as

$$f_{\mathrm{MFNN}}\,(S(i)) = \frac{\text{number of FNNs}}{N - \tau_{m(i)}(i)}. \tag{16}$$

In other words, the fitness is given by the percentage of FNNs. As the emergence of FNNs implies that the embedding fails to unfold the time series in the phase space, we consider the solution with a smaller percentage of FNNs as a better solution.

3) MDL: The MDL principle proposed by Rissanen [34] is an important concept in information theory. The idea behind the MDL principle is that we can use the regularity in a given set of data to construct a model and use the model to compress the data. In this sense, the best model is the one that describes the set of data in the most concise way. Judd and Mees [19] introduced the MDL principle to evaluate the performance of a model for a time series. In terms of this principle, if a model has the most concise description of a time series (i.e., the data size required for the description of model parameters and prediction errors is the smallest), the model is considered to be the best for the time series.

Based on this concept, Small and Tse [20], [22] set up an approximation formula for the description length of an embedding based on the class of local constant models and showed that the optimal embedding is the one that minimizes

$$\frac{d}{2}\ln\left[\frac{1}{d}\sum_{j=1}^{d}(x(j)-\bar{x})\right] + d + \mathrm{DL}(d) + \frac{N-d}{2}\ln\left[\frac{1}{N-d}\sum_{j=d+1}^{N} e(j)^2\right] + \mathrm{DL}(P) \tag{17}$$

where d is the length of the actual embedding window of the embedding, $\bar{x}$ is the mean of the data within the embedding window, $\mathrm{DL}(d) = \lceil \log d \rceil + \lceil \log \lceil \log d \rceil \rceil + \cdots + 0$ is the description length of an integer d, N is the number of data in the time series, $e(j)$ is the prediction error of $x(j)$, and $\mathrm{DL}(P)$ is the description length of the parameters in the model. Given the embedding parameter $(\tau_2, \tau_3, \ldots, \tau_m)$, we can obtain

$$d = \tau_m \tag{18}$$

$$\bar{x} = \frac{1}{\tau_m}\sum_{j=1}^{\tau_m} x(j). \tag{19}$$

In addition, according to [22], for the class of local constant models, we have $\mathrm{DL}(P) = 0$ and $e(j+1) = x(j+1) - x(l+1)$, where $l \in \{1, 2, \ldots, N\} \setminus \{j\}$ is the one that minimizes $|x(j) - x(l)|$.

E. Pheromone Management

During the process of the ACO-NE, pheromones are frequently updated to record the previous search experience of artificial ants.

First, at the beginning, all pheromones (including the pheromones for embedding dimension selection and the pheromones for time-delay selection) are initialized by

$$\mathrm{pd}_k = p_{\mathrm{initial}}, \quad k = 2, 3, \ldots, \mathrm{MaxD} \tag{20}$$

$$\mathrm{pt}_l = p_{\mathrm{initial}}, \quad l = 1, 2, \ldots, \tau_{\max} \tag{21}$$

where p_{initial} is the initial value of pheromones. In the proposed algorithm, we set $p_{\mathrm{initial}} = 1$.

Second, after the ith artificial ant selects an embedding dimension $m(i)$ based on (4), the local pheromone update rule for pd given in (22) is applied, i.e.,

$$\mathrm{pd}_{m(i)} = (1-\rho)\cdot \mathrm{pd}_{m(i)} + \rho \cdot p_{\mathrm{initial}}. \tag{22}$$

Here, $\rho \in (0, 1)$ is a parameter of the algorithm. Similarly, after the ith artificial ant selects a time delay $\tau_l(i)$ based on (6), the local pheromone update rule for pt given in (23) is applied, i.e.,

$$\mathrm{pt}\tau_l(i) = (1-\rho)\cdot \mathrm{pt}\tau_l(i) + \rho \cdot p_{\mathrm{initial}}. \tag{23}$$

In fact, the function of the local pheromone update rules (22) and (23) is to reduce the pheromones on the selected components so that the chances for the following ants to choose a different unexplored component are increased. In other words, the local pheromone update rule is designed to increase the search diversity of the algorithm.

Finally, at the end of each iteration, after all ants in the colony have completed their solutions, additional pheromones are added to the components of the best-so-far solution. Note that the performance of a solution is evaluated by one of the criteria defined in Section III-D. Supposing that $(m(i)|0, \tau_2(i), \tau_3(i), \ldots, \tau_{m(i)}(i))$ is the best-so-far solution, the pheromones on $m(i)$ and $\tau_l(i)$ $(l = 2, 3, \ldots, m(i))$ are updated by the global pheromone update rule as follows:

$$\mathrm{pd}_{m(i)} = (1-\rho)\cdot \mathrm{pd}_{m(i)} + \rho \cdot \mathrm{MaxD} \tag{24}$$

$$\mathrm{pt}\tau_l(i) = (1-\rho)\cdot \mathrm{pt}\tau_l(i) + \rho \cdot \tau_{\max}. \tag{25}$$

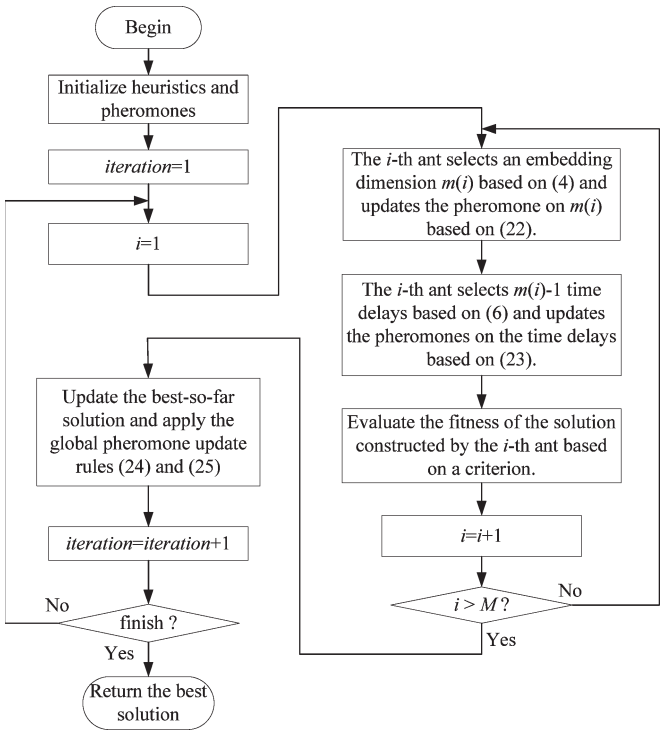

Fig. 1. Flowchart of the proposed ACO-NE algorithm.

The function of the global update rule is to increase the pheromones on the components of the best-so-far solution to make them more attractive.

Based on the aforementioned procedures, the overall flowchart of the proposed ACO-NE algorithm is summarized in Fig. 1.

IV. Experimental Studies

A. Configuration of the Proposed Algorithm

The proposed ACO-NE has the following parameters: the number of artificial ants M, parameter β in (6) and (8) to weigh the importance of heuristics, parameter q_0 in (6), and the pheromone updating rate ρ in (22)–(25). In the experiment, these parameters are set as $M = 10$, $\beta = 1$, $q_0 = 0.5$, and $\rho = 0.1$. Configurations $M = 10$ and $\rho = 0.1$ are common settings for the ACO [26]. For the configurations of β and q_0, we study them empirically on the Mackey–Glass time series. The Mackey–Glass time series is a well-known chaotic time series [35] and is expressed as

$$\frac{dx}{dt} = \frac{0.2x(t-\lambda)}{1+x^{10}(t-\lambda)} - 0.1x(t). \tag{26}$$

We adopt the Mackey–Glass time series with $\lambda = 17$ in the experiment. The length of the time series is $N = 1000$. The Mackey–Glass time series within the time window [0,1000] is plotted in Fig. 2(a).

We first fix $M = 10$, $q_0 = 0.5$, and $\rho = 0.1$, and test the performance of the proposed ACO-NE with $\beta = \{1, 2, 5, 10, 20\}$ using the MDL criterion and the 3-D heuristic on the Mackey–Glass time series. For each version, 20 independent trials are run. In each run, 100 iterations are executed. The

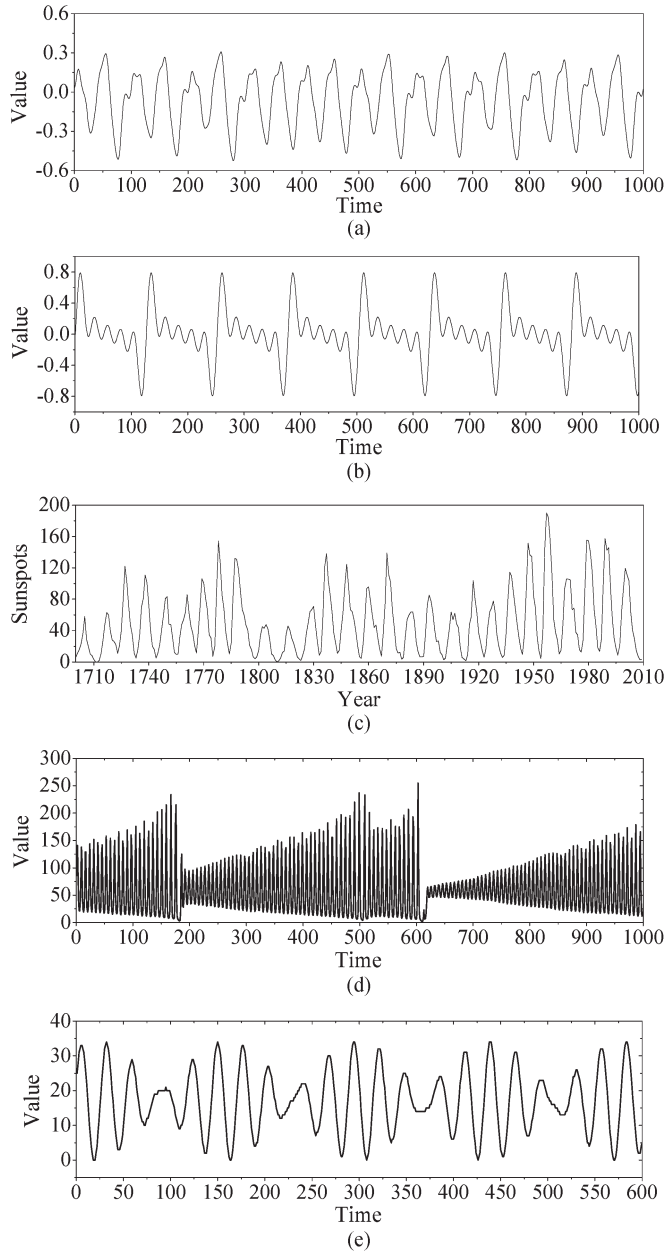

Fig. 2. Time series used in experimental studies. (a) Mackey–Glass time series. (b) Periodic time series. (c) Sunspot time series from 1700 to 2010. (d) Santa Fe Laser time series. (e) Daily-brightness time series of a variable star.

average results found in the 20 runs are showed in Fig. 3(a). From the figure, it can be seen that $\beta = 1$ achieves the best performance. As β increases, the heuristic information becomes overweighed, and the search experiences (i.e., the pheromone values) of ants are undervalued in the solution construction procedure, leading to poorer performance.

We also fix $M = 10$, $\beta = 1$, and $\rho = 0.1$, and test the performance of the ACO-NE with $q_0 = \{0, 0.1, \ldots, 1.0\}$. The results are shown in Fig. 3(b). With a large q_0 value, ants are more likely to choose the time delays with the largest value of $\mathrm{pt}_k \mathrm{ht}_k^\beta$ according to (6). In other words, ants use a more aggressive selection strategy with a larger q_0. In the extreme case $q_0 = 1.0$, ants always choose the time delay with the largest $\mathrm{pt}_k \mathrm{ht}_k^\beta$ value directly, and thus, the algorithm somewhat degenerates to a

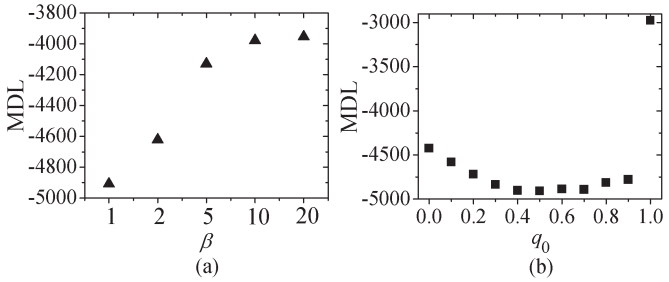

Fig. 3. Empirical studies of parameters β and q_0. Analysis of (a) β and (b) q_0.

TABLE I
COMPARISON OF DIFFERENT HEURISTICS BASED ON TWO-SAMPLE t-TEST

Mackey-Glass				
	Heuristic 2D	Heuristic 3D	No Heuristic	3D/Random
Heuristic 2D		**19.571#**	-1.371	**22.327#**
Heuristic 3D	**-19.571***		**-16.301***	**2.127#**
No Heuristic	-1.371	**16.301#**		**19.102#**
3D/Random	**-22.327***	**-2.127***	**-19.102***	
Mackey-Glass with Noise				
	Heuristic 2D	Heuristic 3D	No Heuristic	3D/Random
Heuristic 2D		**12.312#**	0.392	**12.117#**
Heuristic 3D	**-12.312***		**-9.185***	-0.307
No Heuristic	0.392	**9.185#**		**9.082#**
3D/Random	**-12.117***	0.307	**-9.082***	

* The results obtained by the ACO-NE version of the corresponding row are significantly smaller (better) than those obtained by the ACO-NE version of the corresponding column at the 0.05 level.
# The results obtained by the ACO-NE version of the corresponding row are significantly larger (worse) than those obtained by the ACO-NE version of the corresponding column at the 0.05 level.

greedy search algorithm. Therefore, the performance is very poor. On the other hand, if q_0 is small, ants are prone to use the RWS selection scheme. In this case, the search diversity of ants is increased, but the convergence speed becomes slow. Overall, we find that $q_0 = 0.5$ is better to balance the search exploration and exploitation for the ACO-NE.

One unique characteristic of the ACO is that it enables the use of heuristics. Therefore, we define the 2-D and 3-D heuristics in (9) and (10) based on the geometry topology of the embedding to accelerate the search speed of the ACO-NE. To test if the heuristics can really improve the performance, we also test the algorithm on the Mackey–Glass time series. In addition, we also add Gaussian noise with a standard deviation of 5% of that of the data to the time series to compose the Mackey–Glass time series with noise. The performance of the ACO-NE versions with the 2-D heuristic, the 3-D heuristic, the 3-D/random heuristic, or without using heuristics is compared. Here, the 3-D/random heuristic means that artificial ants either use the 3-D heuristic or do not use any heuristics at random to choose time delays. In each single run of the algorithm, 100 iterations are executed. For each version, 20 independent trials are run based on the MDL criterion, and the best result achieved by the algorithm in each run is recorded.

Table I compares the performance of different versions in terms of two-sample t-tests, and the t-values are tabulated. It can be seen that the difference between the performance of the algorithm with the 2-D heuristic and the one without using heuristics is not significant. This is because, in the 2-D phase space, the proportion of FNNs is too large. This phenomenon induces a misleading effect and causes poor performance. On the other hand, when the dimension of the phase space is extended to three, the proportion of FNNs becomes much smaller, and thus, the 3-D heuristic is much more meaningful and manages to yield significantly better results. To reduce the misleading effect of FNNs, we further develop the 3-D/random scheme that enables artificial ants to use the 3-D heuristic randomly. In this scheme, the influence of the 3-D heuristic is occasionally relaxed so that the time delays with small heuristic values also have some probabilities to be chosen. Therefore, the misleading effect can be further reduced, and the search diversity of the algorithm is increased. Experimental results given in Table I reveal that the 3-D/random scheme can obtain slightly better performance than purely using the 3-D heuristic. Overall, these results demonstrate that the proposed 3-D heuristic and the 3-D/random scheme are effective to improve the performance of the ACO-NE.

B. Performance of Different Criteria

In Section III-D, three criteria are set up as the fitness functions of the ACO-NE. Here, we further study the performance of the algorithm with different criteria.

One of the major goals of time-delay embedding is to forecast chaotic time series. Therefore, in this section, the performance of different criteria is studied in terms of time-series prediction accuracy. To forecast chaotic time series based on a time-delay embedding, we apply the local approximation method proposed in [36]. The reason for using the local approximation method is that it is a classic chaotic time-series forecasting technique based on time-delay embedding. In particular, it directly uses the geometry topology of the time-delay embedding to forecast time series. Thus, the performance of forecasting is seriously affected by the structure of the embedding. In the local approximation method, given an embedding $(m|0, \tau_2, \tau_3, \ldots, \tau_m)$, predicting the value of $x(t+1)$ based on $\{x(1), x(2), \ldots, x(t)\}$ is given by the steps below.

Step 1) Reconstruct the vectors in the phase space based on the time-delay embedding.

Step 2) Evaluate the distances between vector $v_t = (x(t), x(t-\tau_2), x(t-\tau_3), \ldots, x(t-\tau_m))$ and all the other vectors.

Step 3) Find n closest vectors from v_t and denote them as

$$\begin{aligned} v_{t1} &= (x(t_1), x(t_1-\tau_2), x(t_1-\tau_3), \ldots, x(t_1-\tau_m)) \\ v_{t2} &= (x(t_2), x(t_2-\tau_2), x(t_2-\tau_3), \ldots, x(t_2-\tau_m)) \\ &\cdots\cdots \\ v_{tn} &= (x(t_n), x(t_n-\tau_2), x(t_n-\tau_3), \ldots, x(t_n-\tau_m)) . \end{aligned} \tag{27}$$

The distances between v_t and $v_{t_1}, v_{t_2}, \ldots, v_{t_n}$ are denoted as $d_{t_1}, d_{t_2}, \ldots, d_{t_n}$, respectively. In the experiment, we set $n = 10$.

Step 4) The value of $x(t+1)$ is estimated by

$$\widehat{x}(t+1) = \frac{\sum_{j=1}^{n} x(t_j+1)\frac{d_{\max}-d_{t_j}}{d_{\max}-d_{\min}}}{\sum_{j=1}^{n}\frac{d_{\max}-d_{t_j}}{d_{\max}-d_{\min}}} \tag{28}$$

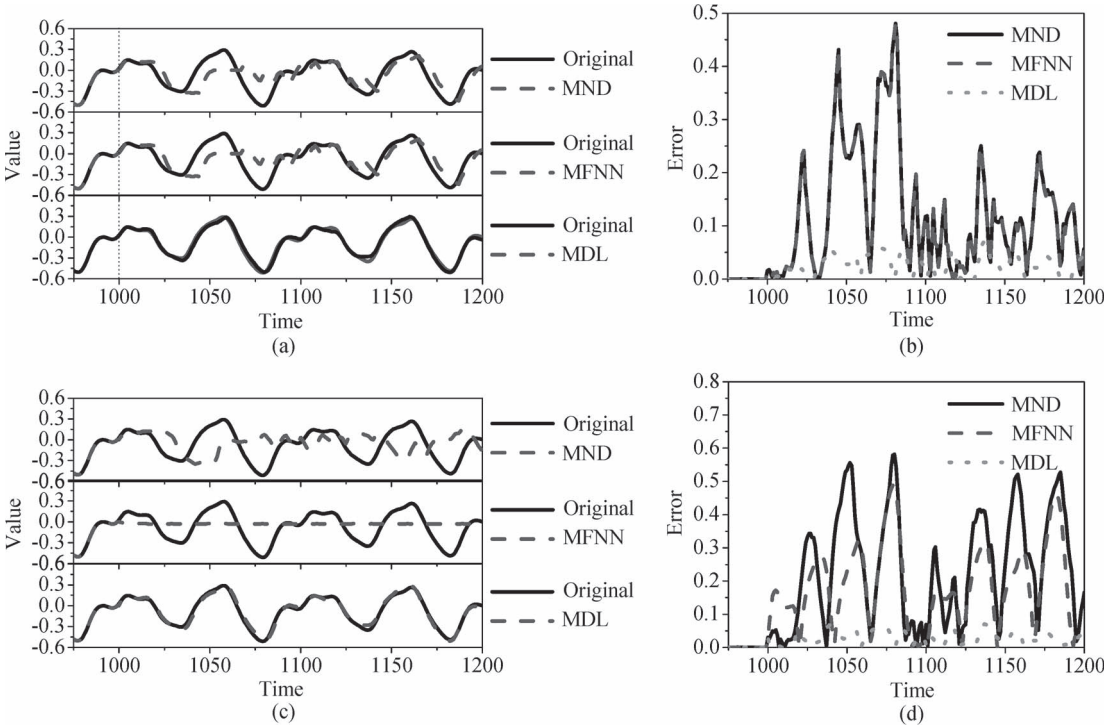

Fig. 4. Comparison of different evaluation criteria. (a) Forecasting of the Mackey–Glass time series. (b) Prediction errors in the Mackey–Glass time series. (c) Forecasting of the Mackey–Glass time series with noise. (d) Prediction errors in the Mackey–Glass time series with noise.

where $d_{\max} = \max_{j=1,2,\ldots,n} d_{t_j}$, and $d_{\min} = \min_{j=1,2,\ldots,n} d_{t_j}$.

In the experiment, the ACO-NE versions with different criteria are tested on the Mackey–Glass time series with or without noise. The data length of the time series for training the embedding parameters is $N = 1000$. The forecasting window is $t = 1001, 1002, \ldots, 1200$. Fig. 4 illustrates the prediction errors of the solutions built based on these three criteria. Obviously, the MDL criterion has the best prediction performance. Compared with the geometry-based criteria, the MDL criterion includes the estimation of prediction errors in its formulation (17). Thus, it has the best performance on the time-series forecasting. According to these results, in the following parts of this section about experimental studies, we use the MDL criterion for both the proposed ACO-NE algorithm and the existing GA approaches in the comparison.

C. Comparison With Other Approaches

1) Experimental Settings: To further demonstrate the performance of the proposed algorithm, we compare the ACO-NE with some existing approaches. The experimental studies are in two aspects. First, we compare the proposed ACO-NE with the GA approaches in [22] and [23] in terms of optimization performance based on the MDL criterion. Then, we compare the ACO-NE with the GA approaches and the traditional uniform embedding technique in terms of prediction performance.

In the experiment, the parameters of the GAs are configured as follows: crossover rate $px = 0.7$, mutation rate $pm = 0.1$, and the population size is 20. These parameters are commonly used in the classic simple GA. The parameters of the ACO-NE are set based on Section IV-A. For the traditional uniform embedding technique, the time delay τ is yielded by the AMI method [16], and the embedding dimension m is determined by the FNN method [12]. The experiments are run on the following six time series:

1) *The Mackey–Glass time series*

   The Mackey–Glass time series has been defined in (26). In the experiment, data $\{x(t)\}_{t=1}^{1000}$ will be used as the training set, and the forecasting window is $t = 1001, 1002, \ldots, 1200$.

2) *The Mackey–Glass time series with noise*

   The Mackey–Glass time series with noise is yielded by adding a Gaussian noise with a standard deviation of 5% of that of the data to the original Mackey–Glass time series. Similarly, the training set is $\{x(t)\}_{t=1}^{1000}$, and the forecasting window is $t = 1001, 1002, \ldots, 1200$.

3) *The periodic time series*

   The periodic time series under investigation is

$$x(t) = \frac{1}{\omega} \sum_{i=a}^{b} \sin it \tag{29}$$

   where $\omega = 5$, $a = 1$, and $b = 5$. The time series is plotted in Fig. 2(b). In the experiment, the training set is $\{x(t)\}_{t=1}^{1000}$, and the forecasting window is $t = 1001, 1002, \ldots, 1200$.

4) *The sunspot time series*

   The sunspot time series is a time series of the numbers of sunspots in May from 1700 to 2009. The sunspot time series is illustrated in Fig. 2(c). The time series exhibits quasi-periodicity, and we usually call the periodicity as sunspot cycles. In the experiment, we use the sunspot data

TABLE II
COMPARISON OF THE OPTIMIZATION PERFORMANCE OF GAS AND ACO-NE IN TERMS OF TWO-SAMPLE t-TEST

	M-G	M-G noise	periodic	sunspot	laser	star brightness
$\tau_{\max}$	30	40	30	70	40	40
ACO-NE – GA[23]	**-3.210#**	**-7.063#**	**-5.995#**	**-6.786#**	**-30.58#**	**-4.682#**
ACO-NE – GA[24]	**-2.01#**	**-1.46**	**-1.95**	**0.23**	**-6.213#**	**-2.767#**
Dimension of ACO-NE	3	3	3	4	4	6
Dimension of GA[23]	5	7	5	30	10	9
Dimension of GA[24]	2	2	3	4	5	8

The objective function values obtained by the ACO-NE are significantly smaller (better) than the GA at the 0.05 level

TABLE III
BEST EMBEDDINGS FOUND BY ACO-NE, GAS, AND THE TRADITIONAL UNIFORM EMBEDDING METHOD

	M-G	M-G noise	periodic	sunspot	laser	star brightness
ACO-NE	(3 | 0, 1, 17)	(3 | 0, 3, 17)	(3 | 0, 2, 4)	(4 | 0, 1, 4, 5)	(4 | 0, 6, 10, 12)	(6 | 0, 1, 10, 11, 12, 24)
GA[23]	(5 | 0, 1, 2, 17, 18)	(7 | 0, 1, 2, 4, 17, 19, 21)	(5 | 0, 1, 2, 3, 4)	(30 | 0, 3, 4, 5, 7, 8, 9, 15, 18, 20, 21, 26, 28, 30, 31, 33, 36, 37, 40, 41, 42, 44, 49, 50, 54, 55, 57, 58, 59, 60)	(10 | 0, 1, 2, 5, 6, 8, 12, 22, 27, 29)	(9 | 0, 1, 6, 7, 9, 11, 12, 13, 24)
GA[24]	(2 | 0, 17)	(2 | 0, 17)	(3 | 0, 1, 2)	(4 | 0, 1, 4, 6)	(5 | 0, 1, 4, 7, 11)	(8 | 0, 1, 2, 11, 12, 13, 14, 15)
Uniform	m=3, τ=7	m=4, τ=9	m=5, τ=5	m=4, τ=16	m=7, τ=2	m=6, τ=4

The embeddings are in the form of $(m \mid \tau_1, \tau_2, \tau_3, \cdots, \tau_m)$ where m is the embedding dimension and τ_i(i=1,2,…, m) are the time delays.

from 1700 to 1960 as the training set, and the forecasting window is 1961–2009.

5) *The Santa Fe laser time series*

The Santa Fe laser time series derived in [37] is a univariate time record measured on an 81.5-μm 14NH3 laser. This time series is shown in Fig. 2(d). In the experiment, we use the training set $\{x(t)\}_{t=1}^{800}$, and the forecasting window is $t = 801, 802, \ldots, 1000$.

6) *The daily-brightness time series of a variable star time series*

This time series derived from the Web site http://robjhyndman.com/TSDL/ is the daily brightness of a variable star on 600 successive midnights. This time series is shown in Fig. 2(e). In the experiment, we use the training set $\{x(t)\}_{t=1}^{500}$, and the forecasting window is $t = 501, 502, \ldots, 600$.

2) Comparison in Terms of Optimization Performance: First, we compare the performance of the ACO-NE and the GAs in [22] and [23] from the viewpoint of the optimization results. Both algorithms are run on the training sets of the aforementioned four time series based on the MDL criterion, and the best result obtained in each single run is recorded. In each run of these algorithms, 1000 solutions are generated. For each algorithm, 20 independent runs are executed. The maximum embedding window $\tau_{\max}$ used for each time series are defined in Table II. Small [22] suggested that $\tau_{\max}$ should satisfy $\tau_{\max} > \tau m$, where τ and m are the time delay and the embedding dimension, respectively, for uniform embedding. The values of τ and m can be determined by the AMI and FNN methods. Therefore, the value of $\tau_{\max}$ in Table II is given by the smallest integer that is larger than $m\tau$ and is divisible by ten. For the ACO-NE, the maximum embedding dimension MaxD is set equal to $\tau_{\max}$.

Table II compares the optimization performance of the ACO-NE and the GAs in terms of t-tests. According to the t-values, in all six cases, the ACO-NE achieves significantly better results than the GA proposed in [22]. The ACO-NE also yields significantly better results than the GA proposed in [23] on three instances. The results produced by the ACO-NE generally have smaller objective function values in terms of the MDL criterion given in (17). In addition, it is notable that, in all six cases, the results produced by the ACO-NE have smaller embedding dimensions than those produced by the GA in [22]. Since we define heuristics to prefer smaller dimensions and smaller time delays for the proposed ACO-NE, artificial ants tend to build solutions with relatively small embedding dimensions and short embedding windows. Therefore, according to Table II, the results found by the ACO-NE are not only better optimized but also more practical for time-series analysis.

Overall, these results demonstrate that the proposed algorithm is effective in terms of optimization performance.

3) Comparison in Terms of Prediction Performance: To further demonstrate the performance of the proposed algorithm, we compare the ACO-NE, the GAs, and the traditional uniform embedding approaches in terms of prediction errors. The best embeddings found by the ACO-NE, the GAs, and the uniform embeddings built by the AMI and FNN methods are adopted in chaotic time-series forecasting based on the forecasting method given in Section IV-B. The embeddings used in the comparison are given in Table III, and the prediction results are shown in Fig. 5 and Table IV. The embeddings given in Table III are in the form of $(m|\tau_1, \tau_2, \tau_3, \ldots, \tau_m)$, where m is the embedding dimension, and $\tau_i (i = 1, 2, \ldots, m)$ are the time delays. As we use $x(t)$ as the first dimension of the reconstructed vector in the phase space, we usually have $\tau_1 = 0$.

According to the results, it can be seen that the embeddings found by the ACO-NE are the most stable in terms of prediction errors. In the six-time series, the ACO-NE yields the smallest prediction errors on three instances. We rank the four approaches considered in the comparison study on each time series according to their average prediction errors. The

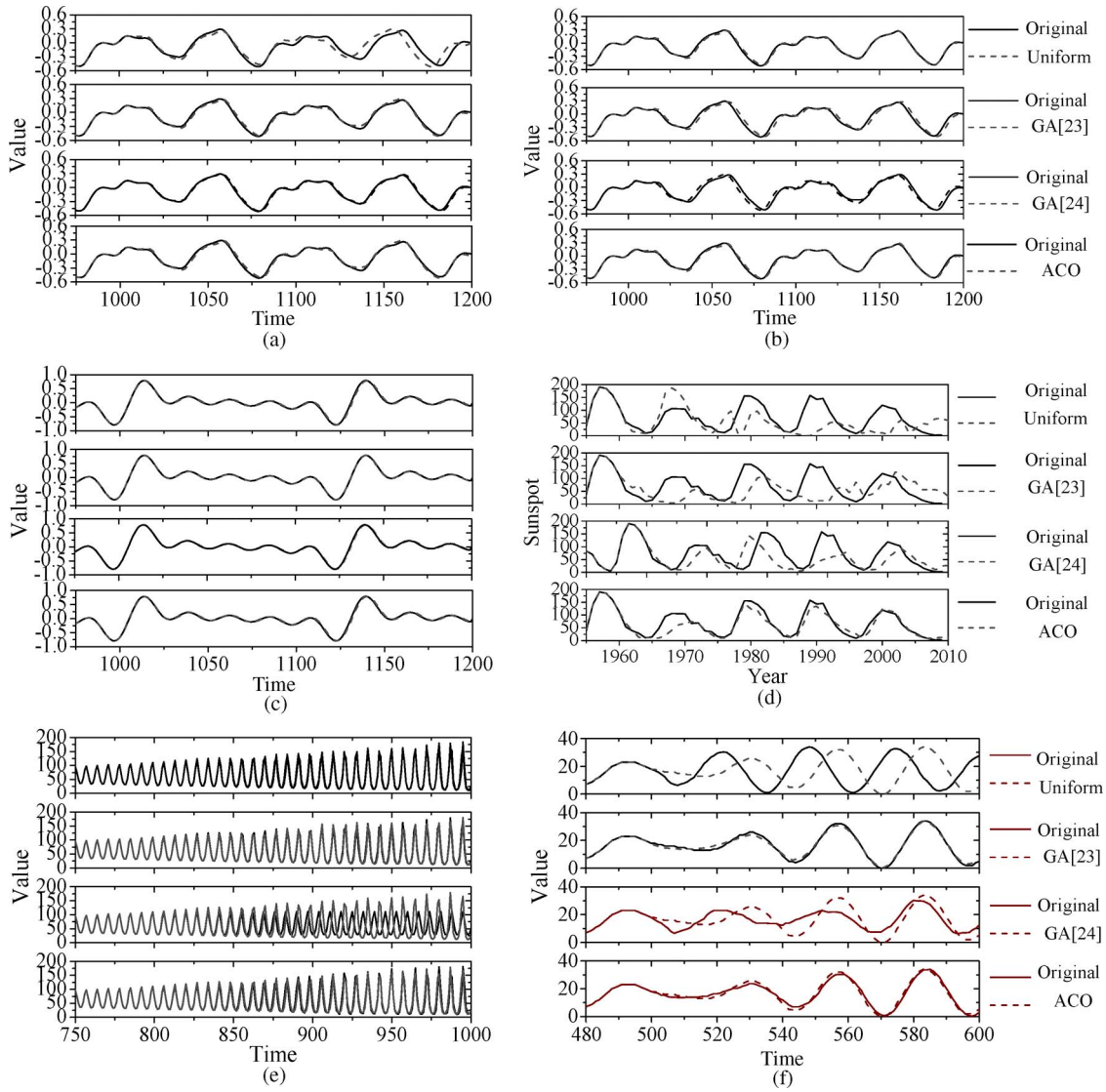

Fig. 5. Comparison of the prediction results of the traditional uniform embedding and the GA and ACO-NE algorithms. (a) Forecast on Mackey–Glass time series. (b) Forecast on Mackey–Glass time series with noise. (c) Forecast on the periodic time series. (d) Forecast on the sunspot time series. (e) Forecast on the Santa Fe laser time series. (f) Forecast on star-brightness time series.

TABLE IV
COMPARISON OF PREDICTION ERRORS

	M-G	r	M-G noise	r
ACO-NE	$2.58\times10^{-2}\pm1.24\times10^{-3}$	3	$2.78\times10^{-2}\pm1.28\times10^{-3}$	2
GA[23]	$2.56\times10^{-2}\pm1.25\times10^{-3}$	2	$4.58\times10^{-2}\pm2.35\times10^{-3}$	3
GA[24]	$\mathbf{1.86\times10^{-2}\pm1.42\times10^{-2}}$	1	$5.37\times10^{-2}\pm1.83\times10^{-2}$	4
Uniform	$7.71\times10^{-2}\pm4.89\times10^{-3}$	4	$\mathbf{1.97\times10^{-2}\pm1.04\times10^{-3}}$	1
	periodic		sunspot	
ACO-NE	$\mathbf{1.97\times10^{-2}\pm1.39\times10^{-3}}$	1	**15.11±2.35**	1
GA[23]	$1.98\times10^{-2}\pm1.39\times10^{-3}$	3	45.3±5.71	3
GA[24]	$1.98\times10^{-2}\pm1.76\times10^{-3}$	4	35.24±31.85	2
Uniform	$\mathbf{1.97\times10^{-2}\pm1.39\times10^{-3}}$	1	47.8±5.40	4
	laser		star brightness	
ACO-NE	11.00±10.50	**3**	**1.616±0.800**	1
GA[23]	**6.444±7.374**	1	1.713±0.882	2
GA[24]	35.27±34.12	4	6.011±3.198	3
Uniform	6.709±8.126	2	14.17±8.34	4

ranking of the proposed ACO-NE on these six instances is 1.83, which is the best among the four approaches. Overall, the proposed ACO-NE manages to produce well-performed and simple embeddings in general. These results reveal that the ACO-NE is also effective in producing practical embeddings for time-series analysis and forecasting.

V. CONCLUSION

An ACO algorithm has been proposed for the optimal selection of embedding parameters in nonuniform embedding for chaotic time series. The main advantages of the proposed algorithm is that it provides a flexible way for optimizing the embedding dimension and the time delays together, and enables the use of heuristics to accelerate the search speed. Experimental results reveal that the proposed algorithm is able to find simple and well-performed embeddings for time-series analysis.

In future research, it is interesting to consider the embedding and modeling problems together and to build the ACO approach to find optimal embedding parameters and select the optimal model [38], [39]. For the more complex variable embedding problem [19], the ACO may provide a new and effective way for finding suitable basis functions and lag vectors. It is also promising to apply the ACO-NE in the phase space reconstruction problem of multidimensional chaotic time series.

ACKNOWLEDGMENT

The authors would like to thank the associate editor and reviewers for their valuable comments and suggestions.

REFERENCES

[1] H. Abarbanel, *Analysis of Observed Chaotic Data*. New York: Springer-Verlag, 1996.
[2] R. Sitte and J. Sitte, "Analysis of the predictive ability of time delay neural networks applied to the S&P 500 time series," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 30, no. 4, pp. 568–572, Nov. 2000.
[3] S.-T. Li and Y.-C. Cheng, "A stochastic HMM-based forecasting model for fuzzy time series," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1255–1266, Oct. 2010.
[4] K. Huang and H.-K. Yu, "Ratio-based lengths of intervals to improve fuzzy time series forecasting," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 2, pp. 328–340, Apr. 2006.
[5] Y. Manabe and B. Chakraborty, "A novel approach for estimation of optimal embedding parameters of nonlinear time series by structural learning of neural network," *Neurocomputing*, vol. 70, no. 7-9, pp. 1360–1371, Mar. 2007.
[6] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[7] L. Chen and G. Chen, "Fuzzy modeling, prediction, and control of uncertain chaotic systems based on time series," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 47, no. 10, pp. 1527–1531, Oct. 2000.
[8] M. Ragulskis and K. Lukoseviciute, "Non-uniform attractor embedding for time series forecasting by fuzzy inference systems," *Neurocomputing*, vol. 72, no. 10-12, pp. 2618–2626, Jun. 2009.
[9] V. Varadan, H. Leung, and É. Bossé, "Dynamical model reconstruction and accurate prediction of power-pool time series," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 1, pp. 327–336, Feb. 2006.
[10] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Lecture Notes in Mathematicians*. New York: Springer-Verlag, 1981.
[11] K. Alligood, T. Sauer, and J. A. Yorke, *Chaos: An Introduction to Dynamical Systems*. New York: Springer-Verlag, 1997.
[12] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A*, vol. 45, no. 6, pp. 3403–3411, Mar. 1992.
[13] P. Grassberger, "An optimized box-assisted algorithm for fractal dimensions," *Phys. Lett. A*, vol. 148, no. 1/2, pp. 63–68, Aug. 1990.
[14] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Phys. D, Nonlin. Phenom.*, vol. 110, no. 1/2, pp. 43–50, Dec. 1997.
[15] T. Schreiber, "Interdisciplinary applications of nonlinear time series methods," *Phys. Rep.*, vol. 308, no. 1, pp. 1–64, Jan. 1999.
[16] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A*, vol. 33, no. 2, pp. 1134–1140, Feb. 1986.
[17] N. Xie and H. Leung, "Reconstruction of piecewise chaotic dynamic using a genetic algorithm multiple model approach," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 6, pp. 1210–1222, Jun. 2004.
[18] H. S. Kim, R. Eykholt, and J. D. Salas, "Nonlinear dynamics, delay times, and embedding windows," *Phys. D, Nonlin. Phenom.*, vol. 127, no. 1/2, pp. 48–60, Mar. 1999.
[19] K. Judd and A. Mees, "Embedding as a modeling problem," *Phys. D, Nonlin. Phenom.*, vol. 120, no. 3/4, pp. 273–286, Sep. 1998.
[20] M. Small and C. K. Tse, "Optimal selection of embedding parameters for time series modeling," in *Proc. Eur. Conf. Circuits Theory Des., Eur. Circuit Soc. Inst. Elect. Electron. Eng.*, Krakow, Poland, 2003, pp. 1–4.
[21] M. Small and C. K. Tse, "Optimal embedding parameters: A modeling paradigm," *Phys. D, Nonlin. Phenom.*, vol. 194, no. 3/4, pp. 283–296, Jul. 2004.
[22] M. Small. (2003, December). Optimal time delay embedding for nonlinear time series modeling. arXiv:nlin/0312011v1. Retrieve Nov. 23, 2010. [Online]. Available: http://arxiv.org/abs/nlin.CD/0312011/
[23] J. B. Vitrano and R. J. Povinelli, "Selecting dimensions and delay values for a time-delay embedding using a genetic algorithm," in *Proc. Genetic Evol. Comput. Conf.*, 2001, pp. 1423–1430.
[24] K. Lukoseviciute and M. Ragulskis, "Evolutionary algorithms for the selection of time lags for time series forecasting by fuzzy inference systems," *Neurocomputing*, vol. 73, no. 10-12, pp. 2077–2088, Jun. 2010.
[25] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 26, no. 1, pp. 29–41, Feb. 1996.
[26] M. Dorigo and L. M. Gambardella, "Ant colony system: A cooperative learning approach to travelling salesman problem," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 53–66, Apr. 1997.
[27] W.-N. Chen and J. Zhang, "An ant colony optimization approach to a Grid workflow scheduling problem with various QoS requirements," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 39, no. 1, pp. 29–43, Jan. 2009.
[28] X.-M. Hu, J. Zhang, H. S.-H. Chung, O. Liu, and J. Xiao, "An intelligent testing system embedded with an ant colony optimization based test composition method," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 39, no. 6, pp. 659–669, Nov. 2009.
[29] Z.-H. Zhan, J. Zhang, and Y. Li, "An efficient ant colony system based on receding horizon control for the aircraft arrival sequencing and scheduling problem," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 399–412, Jun. 2010.
[30] W.-N. Chen, J. Zhang, and H. S-H. Chung, "Optimizing discounted cash flows in project scheduling–an ant colony optimization approach," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 1, pp. 64–77, Jan. 2010.
[31] Y. Lin, J. Zhang, H. S.-H. Chung, W. H. Ip, Y. Li, and Y.-H. Shi, "An ant colony optimization approach for maximizing the lifetime of heterogeneous wireless sensor networks," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 3, pp. 408–420, May 2012.
[32] X.-M. Hu, J. Zhang, Y. Li, and H. S.-H. Chung, "SamACO: Variable sampling ant colony optimization algorithm for continuous optimization," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 6, pp. 1555–1566, Dec. 2010.
[33] M. Dorigo and T. Stützle, *Ant Colony Optimization*. Cambridge, MA: MIT Press, 2004.
[34] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Ann. Statist.*, vol. 11, no. 2, pp. 416–431, Jun. 1983.
[35] E. Ott, T. Sauer, and J. A. York, *Coping With Chaos: Analysis of Chaotic Data and the Exploitation of Chaotic Systems*. Hoboken, NJ: Wiley, 1994, pp. 1–13.
[36] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Phys. Rev. Lett.*, vol. 59, no. 8, pp. 845–848, Aug. 1987.
[37] U. Huebner, N. B. Abraham, and C. O. Weiss, "Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared NH3 laser," *Phys. Rev. A*, vol. 40, no. 11, pp. 6354–6365, Dec. 1989.
[38] T. Nakamura and M. Small, "Nonlinear dynamical system identification with dynamic noise and observational noise," *Phys. D, Nonlin. Phenom.*, vol. 223, no. 1, pp. 54–68, Nov. 2006.
[39] K. Judd, M. Small, and A. I. Mees, "Achieving good nonlinear models: Keep it simple, vary the embedding, and get the dynamics right," in *Nonlinear Dynamics and Statistics*, A. I. Mees, Ed. Cambridge, MA: Birkhauser Boston, 2001, pp. 65–80.

Meie Shen received the B.S. degree in industrial automation from Huazhong University of Science and Technology, Wuhan, China, in 1986 and the M.S. degree in automatic control from the Chinese Academy of Sciences, Shenyang, China, in 1989.

Currently, she is an Associate Professor with the School of Computer Science, Beijing Information Science and Technology University, Beijing, China. Her research interests include intelligent algorithms and automatic control theory and application.

Wei-Neng Chen (S'07–M'12) received the B.S. degree in network engineering and the Ph.D. degree in computer application technology from Sun Yat-sen University, Guangzhou, China, in 2006 and 2012, respectively.

He is currently a Lecturer with the Department of Computer Science, Sun Yat-sen University. His research interests include evolutionary computation and its applications on financial optimization, operations research, and software engineering.

Jun Zhang (M'02–SM'08) received the Ph.D. degree in electrical engineering from the City University of Hong Kong, Kowloon, Hong Kong, in 2002.

From 2003 to 2004, he was a Brain Korean 21 Postdoctoral Fellow with the Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, Daejeon, Korea. Since 2004, he has been with Sun Yat-Sen University, Guangzhou, China, where he is currently a Professor and the Ph.D. Supervisor with the Department of Computer Science. He has authored seven research books and book chapters, and over 100 technical papers in his research areas. His research interests include computational intelligence, data mining, wireless sensor networks, operations research, and power electronic circuits.

Dr. Zhang was a recipient of China National Funds for Distinguished Young Scientists from the National Natural Science Foundation of China in 2011 and the First-Grade Award in Natural Science Research from the Ministry of Education, China, in 2009. He is currently an Associate Editor of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, and the *IEEE Computational Intelligence Magazine*. He is the founding and current Chair of the IEEE Guangzhou Subsection and IEEE Beijing (Guangzhou) Section Computational Intelligence Society Chapters.

Henry Shu-Hung Chung (M'95–SM'03) received the B.Eng. and Ph.D. degrees in electrical engineering from Hong Kong Polytechnic University, Kowloon, Hong Kong, in 1991 and 1994, respectively.

Since 1995, he has been with the City University of Hong Kong, Kowloon. He is currently the Associate Dean with the College of Science and Engineering and a Professor with the Department of Electronic Engineering. He has authored seven research book chapters and over 280 technical papers including 130 refereed journal papers in his research areas. He is the holder of 16 patents. His current research interests include time- and frequency-domain analyses of power electronic circuits, switched-capacitor-based converters, random switching techniques, control methods, digital audio amplifiers, soft-switching converters, electronic ballast design, and circuit design with evolutionary computation techniques.

Dr. Chung is currently an Associate Editor of the IEEE TRANSACTIONS ON POWER ELECTRONICS and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-PART II.

Okyay Kaynak (M'80–SM'90–F'03) received the B.Sc. (first-class honors) and Ph.D. degrees in electronic and electrical engineering from the University of Birmingham, Birmingham, U.K., in 1969 and 1972, respectively.

From 1972 to 1979, he held various positions in the industry. In 1979, he was with the Department of Electrical and Electronics Engineering, Bogazici University, Istanbul, Turkey, where he is currently a Full Professor, holding the United Nations Educational, Scientific, and Cultural Organization Chair on Mechatronics. He has held long-term (near to or more than a year) Visiting Professor/Scholar positions with various institutions in Japan, Germany, the U.S., and Singapore. His current research interests include intelligent control and mechatronics. He has authored three books and is the editor of five books. In addition, he has authored or coauthored close to 350 papers that have appeared in various journals and conference proceedings.

Dr. Kaynak is active in international organizations, has served on many committees of the IEEE, and was the President of the IEEE Industrial Electronics Society from 2002 to 2003. Currently, he is on the IEEE Member and Geographic Activities Board.