Determining Reconstruction Variables and Parameters on Improved Multi-Objective Immune Algorithm

Wei Wang¹, Ying Li *,², Hui Tao ³, ¹ and Xiao-Ping Ma¹
1. School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, China
2. School of Information Engineering, Jiao Zuo University, Jiaozuo, China
3. School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo, China
Email: wangwei83727@163.com, liying19771224@163.com, taohui@hpu.edu.cn and Xpma@cumt.edu.cn
*Corresponding author

Abstract—Aimed at multi-dimension of evaluating reconstruction effect and the correlation between reconstruction variables and reconstruction parameters, improved multi-objective immune algorithm (IMOIA) was proposed to simultaneously determine reconstruction variables and parameters in multivariate time series phase space. In IMOIA, antibody coding was binary strings cascade of multi-parameter, antibody evaluating was three functions, and a new population evaluating S- is presented. Immune operators were designed: overall clone operator to improve speed, mutation operator with multiple changing probability on gene loci and S enhanced local and global search capability to ensure algorithm convergence. Clonal Selection operator on Rank and Crowding-distance to ensure solutions uniformity, repetitive antibodies determination and deletion operator to ensure solutions diversity. Use our method to two benchmarking chaos system, simulation results showed that our method can simultaneously solved multiple combinations of optimal reconstruction variables and parameters and had obvious superiority to other intelligent optimization algorithms and determining reconstruction parameter methods.

Index Terms—Reconstruction Variables; Reconstruction Parameters; Multivariate Time Series; Chaos; Multi-Objective; Immune Algorithm

I. INTRODUCTION

For chaotic systems, long enough univariate time series can reconstruct original power system under appropriate time-delay and embedding dimension based F. taken embedding theory [1]. But in practice, any univariate time series is not sufficient to reconstruct original system. For instance, to Lorenz system, x coordinate could reconstruct original power system, while z coordinate could not reconstruct well [2]. In fact, for practical complex systems, multivariate time series are able to measure, and using multivariate series to reconstruct original power system has obvious advantages to make up for data length limited and noise effect, relatively univariate reconstruction [3, 4]. Generally, in multivariate time series phase space reconstruction, first reconstruction variables are determined, and then reconstruction parameters (time-delay and embedding dimension) are selected [5]. To determine reconstruction variables, conventional variable screening methods can not take into account the temporal redundancy among variables, and calculation of statistical dependences among time series is on selected reconstruction parameters in advance [6]. While reconstruction parameters selection must on fixed reconstruction variables in advance [7-9], and this would inevitably lead to algorithm contradiction and results inconsistent. Using intelligent optimization algorithm, reconstruction variable and reconstruction parameters can be coded uniformly and identified simultaneously. In [10, 11], under prediction error minimum principle, we successfully determined the reconstruction variable and reconstruction parameters on genetic algorithm in multivariate time series phase space reconstruction. However, there are many different evaluation criteria to reconstruction effect, such as false nearest neighbor in [7], average prediction error in [8] and conditional entropy in [9]. In practice, we expect achieve better prediction effect under less reconstruction variables and embedding dimensions. Therefore, multi-objective intelligence algorithm would be adopted to determine reconstructed variables and reconstruction parameters in the paper.

Artificial immune optimization algorithm, as an intelligent method to imitate biological functions of immune system, is characterized by high success rate and good individual diversity compared to other intelligent algorithm [12]. Among immune algorithm, Clonal Selection algorithm, is a combination of deterministic and stochastic choice based on Clonal Selection theory of biological immune system evolutionary process, has been widely used recently [13]. In addition, multi-objective optimization algorithm requires population fast approach and evenly distribute in the true Pareto Front-end in evolutionary process [14]. So this paper, considered practical problems, focused on solutions convergence and distribution in multi-objective optimization, would improve immune optimization algorithm on immune clonal selection algorithm to determine reconstruction...
variables and reconstruction parameters in multivariate phase space reconstruction.

II. FUNDAMENTAL THEOREMS

A. Phase Space Reconstruction of Multivariate Time Series

1) Multivariate Phase Space Reconstruction

Given a time series \( \{ x_{i,j} \}_{i=1}^N \) \( (i = 1, 2, \cdots, I) \),

\[
\{ x_{i,j} \}_{i=1}^N = \{ x_{1,i}, \cdots, x_{n,i}, \cdots, x_{2,i}, \cdots, x_{2,n}, \cdots, x_{1,n}, \cdots, x_{I,n} \}
\]

(1)

According to multivariate time series reconstruction theory of [5], state variable \( X_n \) is constructed as follows:

\[
x_{n} = (x_{1,n}, x_{1,n-\tau}, \cdots, x_{1,n-(m_1-1)\tau}, x_{2,n}, x_{2,n-\tau}, \cdots, x_{1,n-(m_1-1)\tau}, \cdots,
\]

\[
= (x_{1,n}, x_{1,n-\tau}, \cdots, x_{1,n-(m_1-1)\tau})
\]

\[
n = N_0, N_0 + 1, \cdots, N, \quad N_0 = \max \{ (m_1 - 1)\tau + 1 \}
\]

where \( \tau \) and \( m_1(i=1, 2, \cdots, I) \) are the time delay and embedding dimension of the i-th variate \( x_{i,n} \), respectively. Similar to F, takes delay embedding theorem, as long as \( m_1 \) or \( m = \sum_{i=1}^{I} m_i \) is large enough, there must be a mapping \( F : R^n \rightarrow R^n \), making:

\[
x_{n+1} = F(x_{n})
\]

(3)

Or Multivariate function mapping \( F_i : R^n \rightarrow R \), making:

\[
x_{i,n+1} = F_i(x_{n})
\]

(4)

Now, the evolution \( x_{i} \rightarrow x_{i+1} \) or \( x_{i,n+1} \) in state space reflects the evolution of unknown original power system, that is to say, the geometric characteristics of chaotic attractor in original power system are equivalent to that of reconstructed m-dimensional state space.

2) Determining Reconstructed Variable and Reconstruction Parameters

In multivariate time series phase space reconstruction, there are two key issues. One is reconstruction variables determining. Observed multivariate would be some correlation and some time series may reflect the same characteristics in complex systems, and redundant variables take as reconstructed variables would cause information redundancy and increase computing complexity, so redundant variables must be eliminate, and the time series which reflect power system evolution be screen to involve into phase space reconstruction. The other key is reconstruction parameters selecting, because reconstruction parameters determine phase space reconstruction quality that embedding dimension decides state space complexity and time delay decides which historical data involved into reconstruction.

3) Reconstruction Variables Determining

Traditional variable screening methods, such as principal component analysis, factor analysis and other methods, can not consider the temporal redundancy among variables. To multivariable time series obtained by actually observing, we can study statistical dependencies between one variable and the remaining variables, and then select mutually independent variables as reconstructed variable. The methods measuring time series statistics depend including orthogonal correlation function method, mutual information method, and associated integration method. Orthogonally correlation function method, as a simple method, only can measure linear dependence but can not measure nonlinear dependence; mutual information method can better reflect the coupling between two variables to use histogram to solve the distribution density function, but computes complexly; Correlation integral method, a good improvement of mutual information method, can well reflect statistical dependence between variables and calculate simply [6], but computing correlation integral though calculating different phase points distances in reconstructed state space on the basis of known reconstruction parameters.

4) Reconstruction Parameters Selecting

In general, first univariate time series methods is used to select time delay, and then false nearest neighbor method [7, 15], prediction error minimum method [8], or condition entropy extending dimension [9] to select the embedding dimension in multivariate phase space reconstruction. Kugiumtzis pointed out that there was a strong association with the selection of embedding dimension and time delay [16], and should take into account the mutual influence between reconstruction parameters and select all reconstruction parameters, simultaneously.

In summary, reconstruction variables determination is based on fixed reconstruction parameters in advance and reconstruction parameters determination on the basis of the selected reconstruction variables in multivariate phase space reconstruction. This will inevitably lead to algorithm contradictions and inconsistent results. In practice, to finite-length data with noise, the more reconstruction variables number is and the larger embedding dimension is, the better the prediction affect is. But we expect to minimize embedding dimension based on the smallest forecast error, for the decrease of embedding dimension will reduce prediction model complexity and improve forecast rapid. In addition, forecasting effect will be improved when as much as possible monitoring variables are adopted which more fully information describing system, but we need more monitoring equipment and will increase production cost. In short, we expect achieve better prediction effect under less reconstruction variables and embedding dimensions. Therefore, that multi-objective optimization algorithm is used to determine reconstruction variables and reconstruction parameters in the paper, not only can avoid inconsistent results of conventional algorithm, but also can meet the multiple targets of reconstruction effect.

B. Multi-objective Optimization Algorithm and Its Basic Concepts

Generally, multi-objective optimization problem
(minimization problem) can be expressed as [17]

\[
\min F(\mathbf{y}) = (f_1(\mathbf{y}), f_2(\mathbf{y}), \ldots, f_m(\mathbf{y}))^T \\
\text{s.t.} \quad g_i(\mathbf{y}) \leq 0, i = 1, 2, \ldots, R \\
\quad h_j(\mathbf{y}) = 0, q = 1, 2, \ldots, Q
\]

(5)

In (5), \( \mathbf{y} = (y_1, y_2, \ldots, y_n)^T \in \mathbb{Y} \) is the \( n \) dimension decision vectors, and \( \mathbb{Y} \) is the \( n \)-dimension decision space; \( F(\mathbf{y}) \) is the \( m_0 \) dimension decision target, \( g_i(\mathbf{y}) \leq 0, i = 1, 2, \ldots, R \) is the definition of inequality constraints and \( h_j(\mathbf{y}) \leq 0, q = 1, 2, \ldots, Q \) is the definition of equality constraints. Multi-objective optimization algorithm basic concepts include Dominate, Rank, Front-end and Crowding- distance.

**Dominate.** In multi-objective optimization problem, if and only if \( \forall f_i(\mathbf{y}_1) < f_i(\mathbf{y}_2), (i = 1, 2, \ldots, m_0) \) and \( \forall f_i(\mathbf{y}_2) \leq f_i(\mathbf{y}_1), (i = 1, 2, \ldots, m_0) \) define \( \mathbf{y}_1 \) dominate \( \mathbf{y}_2 \), or \( \mathbf{y}_2 \) is dominated by \( \mathbf{y}_1 \), describe \( \mathbf{y}_1 \gg \mathbf{y}_2 \).

**Rank and Front-end.** When \( \mathbf{y}_1 \) dominates \( \mathbf{y}_2 \), the Rank value of \( \mathbf{y}_1 \) is smaller than that of \( \mathbf{y}_2 \). While \( \mathbf{y}_1 \) and \( \mathbf{y}_2 \) don’t dominate each other, their Rank is same. Individuals whose Rank value is 1 are in the first Front-end, and individuals whose Rank is 2 are in the second Front-end, and so on. Obviously, in current population, the population in the first Front-end are completely free from dominated, and population in the second Front-end are dominated by the first Front-end population.

**Pareto optimal solution set.** In decision space, If there is not a \( \mathbf{y}^* \) making \( \mathbf{y}^* \gg \mathbf{y}, \mathbf{y} \) is Pareto optimal solution and all Pareto solutions construct Pareto optimal solution set. From the definition of Rank and Front-end, it can be seen that the individuals in the first Front-end constitute Pareto optimal solution set.

Multi-objective intelligent optimization algorithms are becoming focus of evolutionary computation research. Although Multi-objective immune algorithm (MOIA) has obtained a better effect than other multi-objective algorithms (MOAs) [14, 18], MOIA still need to design some efficient operator in order to better solve multi-objective optimization problem. So this paper, focused on solutions’ convergence and distribution, would improve MOIA on immune Clonal Selection algorithm to determine the reconstruction variables and reconstruction parameters in multivariate phase space reconstruction.

**C. Immune Clonal Selection Algorithm**

Optimizes problem corresponding to antigen, feasible solutions corresponding to antibodies, immune Clonal Selection algorithm formed by abstracting mathematical evolutionary optimization process from evolutionary chain in biological immune response (antibody group \( \rightarrow \) immune selection \( \rightarrow \) cell clone \( \rightarrow \) somatic hypermutation \( \rightarrow \) generate new antibody \( \rightarrow \) New antibody group) [13].

The main steps of basic immune clonal selection are as follows.

(1) **Initial antibodies generation.** Design appropriate antibody encoding rule according to optimization problem characteristic, and on the rule generate initial antibody group according to prior knowledge.

(2) **Antibodies evaluation.** Calculate antibody affinity and antibody concentration and evaluate each antibody’s and antibody group’s qualities.

(3) **Cloning operation operator.** Use cloning operation operator including immune clone, mutation, cloning inhibition, population update to imitate various immune operation operator in immune response.

**Immune clone operator.** Generate a new population according to individual affinity results calculated by cloning copying. Clone multiple can be related to individual affinity (fitness) or set to a constant.

**Mutation operator.** Obtain mutated antibody group by some hypermutation of the clone population. For binary coding, hypermutation means negating some gene coding with certain probability. For real coding, uniform mutation, gaussian mutation or random mutation strategy can be adopted.

**Clone inhibition operator.** Choice the antibodies with high affinity from mutated antibody group.

**Population update.** Random generate new antibodies to replace the individuals population with low affinity.

**III. DETERMINING RECONSTRUCTION VARIABLES AND PARAMETERS ON MOIA**

Given the interdependence of determining reconstruction variables and reconstruction parameters and multiple evaluation of optimization results, This paper, immune clonal algorithm for the basic framework, would improve MOIA to ensure population quickly approaching and evenly distributed in the true Pareto Front-end, and use improved multi-objective immune algorithm (IMOIA) determine reconstruction variables and reconstruction parameters in the multivariate reconstruction.

**A. Antibody Coding**

In general, embedding dimension \( m \) is a positive integer. In order to simplify optimization parameter encoding, embedding dimension \( m \) is set as an integer greater than or equal to zero by expanding \( m \) range, and define that \( m = 0 \) represents corresponding monitoring variable \( x_i \) does not participate in phase space reconstruction, while \( m > 0 \) represents \( x_i \) participate in reconstruction and embedding dimension is \( m \). Thus, only embedding dimension coding can simultaneously show reconstructed variable and embedding dimension, and the encodings of embedding dimension \( m \) and time delay \( \tau \) can express reconstruction variables and reconstruction parameters (including the embedding dimension and delay time). For \( m \) and \( \tau \), binary coded is used, and to select the combination of the best reconstruction variables and their corresponding embedding dimension and time delay, a multi-parameter
cascade encoding used, so antibody coding composes of \(2I\) binary string cascade.

\[
y_i(t=1,2,...\cdot 2I) = (y_{i,1},\ldots, y_{i,J}) \quad (6)
\]

In (6), \(y_i(t=1,2,...\cdot 2I)\) expresses the binary string encoding of embedding dimension \(m\). \(y_{i,t}, (i=1,2,...\cdot I)\) expresses the binary string encoding of time delay \(\tau\), \(g_{i,t}\) only is 0 or 1, \(J_i\) is the length of binary string \(y_i(t=1,2,...2I)\), and antibody length is \(L = \sum_{i=1}^{2I} J_i\).

When antibody decoding to reconstruction parameters, embedding dimension is the decimal number directly converted from corresponding binary string and delay time is the decimal number converted from corresponding binary string plus 1 because the fact that \(m_i\) is an integer greater than or equal to zero and \(\tau_i\) is a positive integer.

\[
m_i = \sum_{j=1}^{I} 2^{m_i-j} \quad (7)
\]

\[
\tau_i = \sum_{j=1}^{I} 2^{m_i-j} + 1 \quad (8)
\]

When \(m_i = 0\), the corresponding monitoring variable \(x_i\) does not participate into reconstructing and time delay \(\tau_i\) makes no sense. While \(m_i > 0\), \(x_i\) participates into reconstructing, and corresponding embedding dimension and time delay are \(m_i\) and \(\tau_i\), respectively.

**B. Designing Affinity Function and Evaluating Antibody and Population**

1) **Affinity Function Design**

Given the goal of multivariable phase space reconstruction is accurately predicting multiple time series, the first evaluation index is chaotic prediction error \(f_1(y_i)\). In practice, for data containing noise, the more reconstruction variable and the greater embedding dimension, the better prediction results is. However, we want to minimize embedding dimension, thereby reduce prediction model complexity and improve prediction rapid, so the second evaluation index \(f_2(y_i)\) is the proportional function of embedding dimension. In addition, when using as much as possible variables to reconstruct, the more variables information maybe fully describe system and prediction effect would improve, but the more monitoring measure is needed. That is to say, under ensuring prediction effect, the less reconstruction number is the better. Therefore, the third evaluation index \(f_3(y_i)\) is the proportional function of reconstruction variables number. In summary, the optimization objective is

\[
\min f(y_i) = \min(f_1(y_i), f_2(y_i), f_3(y_i)) \quad (9)
\]

Define \(f_1(y_i)\). Given computation complexity and prediction accuracy, Local prediction method is adopted. Reference [18] in the case of single variable time series, improved local average weighted prediction model shown in (10) is simple, but its prediction error is smaller than zero-order approximation model.

\[
\hat{x}_{i,n+1} = \frac{\sum_{k=1}^{K} x_{i,n+1-k} d_{n+k} - d_{n+k}}{\sum_{k=1}^{K} d_{n+k} - d_{n+k}} \quad (10)
\]

In (10), \(K\) is the number of nearest neighbors. \(X_{a_i}\) is the nearest neighbors of \(X_N\), \(d_k = |X_N - X_{a_i}|\).

In order to overcome prediction effect contingency, the error function takes the average of \(P\) points prediction error. Define root mean square error of 1-step predict of \(M\) variables as error function \(E(m_1,m_2,\ldots,m_M, \tau_1, \tau_2, \ldots, \tau_M)\).

\[
E(m_1,m_2,\ldots,m_M, \tau_1, \tau_2, \ldots, \tau_M) = \frac{\sqrt{\sum_{i=1}^{M} \sum_{j=1}^{P} (\hat{x}_{i,n+1} - x_{i,n+1})^2}}{P} \quad (11)
\]

\(E\) value depends entirely on embedding dimension \(m_1,m_2,\ldots,m_M\) and time delay \(\tau_1, \tau_2, \ldots, \tau_M\), therefore define the first evaluation function \(f_1(y_i)\) is as follow.

\[
f_1(y_i) = \frac{1}{MP} \sum_{i=1}^{M} \sum_{j=1}^{P} (\hat{x}_{i,n+1} - x_{i,n+1})^2 \quad (12)
\]

Define \(f_2(y_i)\).

\[
f_2(y_i) = K_m^* \sum_{i=1}^{M} m_i \quad (13)
\]

\(f_2(y_i)\) is the proportional function of embedding dimension \(K_m^*\), a scale factor, is set according to forecast error value and embedding dimension of actual system to ensure the dimension unity of the different objective function.

Define \(f_3(y_i)\).

\[
f_3(y_i) = K_r^* \sum_{i=1}^{M} i(m_i > 0) \quad (14)
\]

Similarly, \(f_3(y_i)\) is the proportional function of reconstruction variables number \(K_r^*\) as a scaling factor, is set according to prediction error, embedding dimension and the number of variables in actual system.

2) **Antibody and Population Evaluation**

Different from traditional single-objective optimization to evaluate antibody by affinity function and concentration, in multi-objective optimization, first calculate antibody multiple affinity functions, next non-dominated sort all antibodies in the population to different Front-end, then calculate the crowding distance of individuals in the same Front-end, finally comprehensively evaluating antibody depends on its Rank value and crowding distance in current generation.
The smaller the antibody Rank, the higher antibody non-domination degree is and its performance is more excellent; the greater the Crowding-distance, the greater the distance between the antibody and its adjacent antibody is in the same Rank and the better antibody diversity and spacing is.

At present, domestic and overseas scholars have put forward a variety of measurement methods, which included approximation measure, broad measure and spacing measure, etc., to evaluate multi-objective evolutionary algorithm by comparing optimization solution set. Among them, spacing measure (S) proposed by Ziltzer [20], can measure solutions distribution of multi-objective optimization. This paper would promote spacing measure concept and improve its calculation, then use S to describe antibody distribution of n-generation population accordingly describing the process of multi-objective optimization.

Reference [20], spacing measures S of n-generation population are defined as follows:

\[
S = \left[ \frac{1}{N_A} \sum_{i=1}^{N_A} (d_i' - \bar{d}')^2 \right]^{0.5}
\]

In (15), \(N_A\) is population size, \(d_i'\) is the distance between \(i\)-th antibody and its nearest antibody. Assuming that the objective space dimension is \(m_o\), the expression of \(d_i'\) is:

\[
d_i' = \min_j \sqrt{(f_1(y_i) - f_1(y_j))^2 + (f_2(y_i) - f_2(y_j))^2 + \cdots + (f_{m_o}(y_i) - f_{m_o}(y_j))^2}
\]

\[j \neq i, \; i, \; j = 1, 2, \ldots, N_A\] 

(16)

\[
\bar{d}' = \frac{1}{N_A} \sum_{i=1}^{N_A} d_i'
\]

(17)

According to (15) ~ (17), S can accurately provide population distribution information. the greater S is, the worse the antibodies distribution uniformity is. But S is less suitable for practical application because of high computation complexity caused by calculating \(d_i'\).

Crowding-distance \(d_i'\), similar to \(d_i'\), is the indicator to describe individual uniformed and diversity, for \(d_i'\) represents the distance between the \(i\)-th individual and its nearest neighbor, and \(d_i'\) represents the distance between two individuals adjacent before and after \(i\)-th individual; However, the \(d_i'\) computation complexity of is more simple than \(d_i'\). So, we use Crowding distance \(d_i'\) to substitute \(d_i'\), algorithm performance would be improved.

In addition, for the individuals in the same Front-end, it is meaningful to compare their Crowding distance, while for the individuals in different Front-end, the important evaluation indicator is Rank describing antibodies Non-dominated characteristics except Crowding-distance. Therefore adjust original definition of spacing measure, redefine S as follows.

\[
S = \sum_{d' \neq \bar{d}'} \left[ \frac{1}{n_{d'}} \sum_{i} (d_i' - \bar{d}')^2 \right]^{0.5}
\]

\[
\bar{d}' = \frac{1}{n_{d'}} \sum_{i} d_i'
\]

(18)

In (18), \(d_i'\) denotes Rank value of the \(i\)-th individual, and \(d_{max}'\) is maximum of all \(d_i'\), \(n_{d'}\) is the number of antibodies which Rank value are \(d_i'\). \(d_i'\) is Crowding-distance.

From (18) we can see that the more Non-dominated antibodies in population and the more consensus crowded distance, the smaller S value is and the better population uniformity and diversity are.

C. Designing Immune Operators

The immune operator of IMOIA including immune clone operator, immune mutation operator and immune selection operation, would be introduced in turn below.

1) Overall Immune Clone Operator

In multi-objective optimization, there would be conflict among different targets, for example, a solution may be the best in a target, but may be the worst in other target, and make a series of solutions can’t simple mutual compare. In view of this, the paper use Overall immune operator that clone multiple is designed a constant, then each antibody has the same clone proportion \(K_c\).

The overall cloning operator is conducive to the broader distribution of Pareto Front through expanding space to generate new antibody population and achieve global search. And, relative traditional clone with different clone multiple in accordance with each antibody affinity, the algorithm operation speed would be improved for don’t need to compute clone multiples by sorting antibody mutual dominant in overall cloning.

2) Mutation Operator with Multiple Changing Probabilities

According to evolution process of immune algorithm, there are only two operations to produce new antibodies: population update and mutation operator. Population update, a means to maintain antibodies diversity, randomly generates a new antibody to replace the antibody with low affinity in the original antibody group; Mutation operator performs mutation operation on cloning results to generate new potential antibodies and achieve local search. so mutation operator largely determine the algorithm performance of local search. In Standard artificial immune algorithm, the mutation operator with strong blindness, which mutation probability is generally set at a constant and no direct relationship with gene location and evolution algebra, is an important reason to cause less local search ability.
Different loci gene mutation has different impact on optimized parameter, and therefore, also has different effect on algorithm search capability. For example, with a four-digit binary string representing an embedding dimension, when the lowest loci genes mutate, corresponding embedding dimension has smaller change – only 1-dimension, while the highest loci genes mutate, corresponding embedding dimension has larger change – 8-dimension. Obviously, high-loci gene mutation is conducive to maintain population diversity and improve global search ability, while low-loci gene variation is conducive to enhance the algorithm search capabilities at a local scale. Therefore, individual different genes would be divided into high-loci gene and low-loci gene to ensure the genes in different loci have different probability. The operation to divide individual Binary sub-string with length of $y_i$ into high-loci and low-loci gene is as follows: Set the constant $k \in (0, 1)$, the low-loci genes locate in $0 \sim \text{round}(kJ_i) - 1$ and the high-loci genes locate in $\text{round}(kJ_i) \sim J_i - 1$.

In addition, when algorithm evolves into a certain degree, traditional mutation operator is difficult to gain due to the lack of local search ability. Srinvas proposed adaptive mutation probabilities with evolution algebra [21], Li Liangmin also proposed a mutation strategy based on Binary-coding loci and evolution algebra that set change mutation probabilities related to both variation position and variation algebra to individual different loci [22].

However, optimization process is not always exact correspondence with evolution effect because immune algorithm is a stochastic optimization algorithm, and when mutation probabilities updating with evolution algebra, it does not guarantee that mutation probabilities update closely associated evolution process. Recalling S definition of (15), $S$ is an indicator of evolution effect describing population distribution and uniformity. In order to make algorithm evolution effect more clearly reflect in mutation probabilities update, we designed mutation operator with multiple changing probability related with but only gene loci and also multi-objective evolution effect - spacing indicator $S$.

$$
\begin{align*}
\alpha_{mh} &= a_1 + \frac{b_1}{1 + e^{\alpha(S-S_0)}} \\
\alpha_{ml} &= a_2 + \frac{b_2}{1 + e^{\beta(S-S_0)}}
\end{align*}
$$

$\alpha_{mh}$ and $\alpha_{ml}$ are the mutation probability of high-loci gene and low-loci gene. $S$ is spacing indicator of the $i$-th generation population. $a_1$, $b_1$, $a_2$, $b_2$, $\alpha$ and $S_0$ are all constants set in advance. $a_1$, $b_1$, $a_2$, $b_2$ determine mutation probability range, and $\alpha$, $S_0$ determine vary speed of mutation probability.

This paper proposed mutation operator with changing probability mutation adjusted according to gene location and evolution process. From (20), it is easy to see that in early evolution stage, optimization effect is usually poor and corresponding population spacing $S$ is large, so $\alpha_{mh}$ is close to $a_1 + b_1$, but $\alpha_{ml}$ close to $a_2$, give high-loci genes large mutation probability in order to maintain population diversity and improve global search capability; When evolution proceeding, multi-objective optimization effect (population spacing) improves and $S$ becomes gradually smaller, $\alpha_{mh}$ becomes smaller while $\alpha_{ml}$ becomes larger; When evolution proceeding to a certain degree, $S$ becomes very small until it is close to 0, $\alpha_{ml}$ is close to $a_1$, but $\alpha_{mh}$ close to $a_1 + b_1$, reduce the mutation probability of high-loci gene to lessen destroyed probability of high-quality antibodies, while increase the probability mutation of low-loci gene to enhance local search ability of the algorithm to guarantee solution convergence. In this way, mutation probability update directly relates to evolution process and algorithm local and global search capability are both enhanced.

3) Clonal Selection Operation on Rank Value and Crowding-distance

After overall clone operation, population size is huge. Clonal Selection operation, the inverse operation of immune cloning operation, would select the best antibodies from the population after clone and mutation operation, and control population size to ensure operation speed. Unlike traditional selection method on affinity, our Clonal Selection operator specifically designed for multi-objective optimization features: whether an antibody can be selected into the next generation depends on its Rank value and Crowding-distance in current generation.

The basic idea of Clonal Selection operation on Rank value and Crowding-distance is: assign all antibodies to different front-end by non-dominated sorting, and then calculate individuals crowding distance in the same front-end, next select non-dominant individuals in accordance with the Rank value from low to high, remove part of the individuals with smaller Crowding-distance in larger Rank value when the individuals number is more than $NA$ (pre-set population size). Our Clonal Selection operation first will choose out the individuals with higher degree dominance to ensure solutions convergence, then select individuals with large crowding distance, thus to ensure distribution uniformity of the obtained solutions.

D. Judging and Deleting Repetition Antibody

After clone operation, the antibody population has high repetition probability for clone operation expanding solution space. Though mutation operation generates some new antibodies, there are still a large number of repetition antibodies. Clonal Selection operation must select non-dominated antibodies with better diversity, so during selection operation the same antibodies in population need to remove.

Different antibody encoding (corresponding reconstruction parameters) may have the same fitness function value, Therefore, determine whether the antibody is repeated should based on antibody encoding. When two antibody encodings are identical, the two antibodies must be repeated. In addition, even if the antibody encodings are not exactly the same, state space reconstruction effect may also be identical. For instance,
when Embedding dimension  $m_i$ is 0 or 1, even if the delay time is different, reconstruction effects is identical and the value of delay time without significance, because $m_i=0$ indicates the variable $x_i$ does not participate in reconstruction and $m_i=1$ indicates $x_i$ participate in reconstruction and Corresponding state variable is the reconstructed variable itself. Strictly speaking, determine whether the two antibodies are identical need to judge whether the corresponding reconstructed state variables are same. However, the calculation of comparing state variables is more complex than that of directly comparison antibody encodings. Accordingly, during Clonal Selection operation, in order to reduce computation complexity, first remove a large amounts of antibodies with identical encoding; next Non-dominated sort the remaining antibodies; then judge whether the antibodies with same Rank value are repeated and delete repetition antibodies; Afterwards, calculate Crowding distance of different antibodies; finally select antibodies on Rank value and crowding distance.

E. Algorithm Flow

Based on the above design, IMOIA flow to determine reconstruction variable and reconstruction parameters is showed in Fig. 1 and the main steps are as follows.

1) Randomly generate initial antibody population according to multi-parameters cascade encoding.

2) Calculate multi-objective function $f(y) = (f_1(y), f_2(y), f_3(y))$ of each antibody and spacing index $S$ of initial population.

3) Execute immune operation, including overall immune clone operator, mutation operator with multiple changing probabilities, deleting repetition antibody, Clonal Selection operation on Rank value and Crowding-distance.

4) Record optimal antibody and calculate population spacing index $S$.

5) Update antibody population. In order to maintain population diversity, randomly generate 10% new antibodies to replace the individuals with largest Rank value and less crowding distance to explore new feasible solution.

IV. EXAMPLES

A. Data Sources

Two benchmark chaos system as example would verify the effectiveness of improved multi-objective immune algorithm to optimize reconstruction variables and reconstruction parameters of chaotic system in matlab 2009b.

1) Lorenz Chaos System

Lorenz equations are as follows [5].

\[
\begin{align*}
x_1' &= \sigma(x_2 - x_1) \\
x_2' &= x_1(r - x_3) - x_2 \\
x_3' &= x_1x_2 - bx_3
\end{align*}
\]

In (21), $\sigma = 10, \ r = 28, \ b = 8/3$. Initial states are $x_{10} = 15.34, \ x_{20} = 13.68$ and $x_{30} = 37.91$. 12200 samples of $x_1, x_2, x_3$ are computed using fourth-order Runge-Kutta integral method when integration step equals 0.04. In order to reduce the influence of transient state, the previous 10000 samples are removed, and in remained 2200 samples, 2000 samples serve as train samples, the middle 100 samples as test samples and the last 100 as prediction samples.

2) The Coupling System of Two Different Rossler Equations

\[
\begin{align*}
x_1' &= \sigma(x_2 - x_1) \\
x_2' &= x_1(r - x_3) - x_2 \\
x_3' &= x_1x_2 - bx_3 \\
x_4' &= -\omega_1x_4 - x_6 + \epsilon(x_5 - x_4) \\
x_5' &= \omega_2x_4 + 0.15x_5 \\
x_6' &= 0.2 + x_6(x_4 - 10)
\end{align*}
\]

where $\omega_1 = 0.99, \ x_1, x_2, x_3$ represents one Rossler system; $\omega_2 = 0.95, \ x_4, x_5, x_6$ represents another Rossler system. Coupling coefficient $\epsilon = 0.05$ or $\epsilon = 0.50$ shows intermittent delayed state or complete synchronization state of coupling system. Initial states are $x_{10} = 0.1, \ x_{20} = 0.2, \ x_{30} = 0.3, \ x_{40} = x_{50} = 0$ and $x_{60} = 15$ [5].
and their reconstruction variables, embedding dimension length in engineering practice. Non-dominated solutions historical data sample as 500 to imitate the limited data. Further, the simulations were executed with shortening reconstruction parameters combination. Because variable \(x_{(500)}\) is NMRSE under 500 historical samples.

mapminmax actual system with noise, then adding 5% noises to with 2000 samples and no noise. Taking into account the influence of transient state, the previous 10000 samples integration step equals 0.02. In order to reduce the difficulty and prediction error requirement.

B. IMOIS Parameters Setting

Owing to the different variables number and value range, many parameter settings of the above two cases is not same and specific settings are given in TABLE I.

C. Results and Analysis

To illustrate the proposed method effectiveness, first we would optimize the reconstruction variables and reconstruction parameters of Lorenz system under different data length and noise level. During simulation, we first solve the non-inferior solutions in historical data with 2000 samples and no noise. Taking into account the actual system with noise, then adding 5% noises to original data, the Pareto solutions were solved again. Further, the simulations were executed with shortening historical data sample as 500 to imitate the limited data length in engineering practice. Non-dominated solutions and their reconstruction variables, embedding dimension and prediction error are shown in TABLE II–III. TABLE II is the Optimize results of Lorenz system without noise and TABLE III is the Optimize results with 5% noise. NMRSE is normalized root-mean-square error using mapminmax as normalized function. NMRSE (2000) is NMRSE under 2000 historical samples, and NMRSE (500) is NMRSE under 500 historical samples.

As can be seen from above two tables, IMOIA can get multiple solutions of reconstruction variables and reconstruction parameters combination. Because variable \(x_3\) has little correlation with \(x_1\), and \(x_3\), as a redundant variable, participating into phase space reconstruction will not improve \(x_1\) prediction effect. \(x_3\) isn’t selected to participate into reconstruction in the non-inferior solutions. This explains our IMOIA has chosen suitable reconstruction variables. Moreover, the optimization results can obtain as small as possible prediction error under the smaller reconstruction variables number and embedding dimension, which shows we achieved optimal reconstruction parameters. In short, our IMOIA can optimize reconstruction variables and reconstruction parameters (embedding dimension and delay time), simultaneously.

To further study the influence of variable number on prediction error, we find the results of two variables \(x_1, x_2\) reconstruction is superior to single variable \(x_1\) under the same embedding dimension \(m\) and regardless whether or not containing, because two variables contain more system information which can better describe the system. In engineering, the optimal combination of reconstruction variables and parameters can be selected from non-inferior solutions according to variable monitoring difficulty and prediction error requirement.

In addition, to observe predicted effect of different data length, we find the prediction error obviously increases with the decrease of the historical data length, and under the same \(m\), when containing noise, the prediction error of double variable reconstruction using 500 historical data is less than that of univariate reconstruction forecast using 2000 historical data, while without noise, the two prediction results have little difference. This shows multivariate reconstruction prediction can overcome noise influence and make up for data length inadequacy to some extent, thus the prediction precision has improved relative to single variable reconstruction.

Comparison TABLE II and TABLE III, we can conclude that under ideal condition of system without noise, as \(m or I\) increase in a certain range, the prediction error reaches the minimum, but continue increasing or \(I\), prediction effect becomes worse. Under the condition of system containing noise, the \(m\) increased significantly corresponding to the minimum prediction error. In
summary, prediction effect can be improved significantly with the increase of $m$ or $I$.

Similarly, we can observe the optimal combination of reconstruction parameters and variables of Rossler coupling systems, and under 2000 historical samples additional 5% noise, reconstruction variables number $I$ embedding dimension $m$ NRMSE of non-inferior solutions corresponding were shown in TABLE IV. In the Table, the maximum of $I$ is 4, $x_3$ and $x_6$ are redundant variable, not selected to participate in reconstruction, which shows that IMOA can effectively select reconstruction variables. At the same time, our algorithm obtained dozen of assemblies of embedding dimension $m$ and delay time $\tau_i$ in which less $m$ and $I$ embedding dimension achieved lower forecast error. Namely, IMOA can optimize to variables and reconstruction parameters (embedding dimension and delay time). For actual engineering data, we can select a solution from obtained non-inferior solution set, according to prediction error, variable monitoring difficulty, model complexity, then reconstruct phase space the basis of the solution, in order to solve geometric invariants of chaos system and realized chaos prediction.

In addition, in order to illustrate further the excellent performance of IMOA presented in this paper, multi-objective genetic algorithm (NSGA-II), multi-objective particle swarm algorithm (MOPSA), traditional multi-objective immune algorithm (TMOIA) and our algorithm (IMOIA) were adopted to above two systems with 500 training samples, and through 10 times simulation under different initial conditions the mean of optimization success rate, simulation time, and evolution generation obtained ultimate non-inferior solutions are showed in TABLE V. From TABLE V, we can conclude that our method has distinct advantages in all indicators relative to TMOIA, and has obvious advantages in optimization success rate and evolution algebra although a small increase in simulation time relative to NSGA-II and MOPSA. On one hand, overall immune cloning operation expands search scope and reduces computational complexity, immune mutation operation with multi-variation probability on population spacing is conducive to global search in initial evolution stage and local search in later evolution stage, and vast repeat antibodies were removed in Clonal Selection process to ensure population diversity. Therefore, the optimization success probability and iterative generation have obvious advantages. However, that the actual population size is greater than the genetic algorithm and particle swarm algorithm in each evolutionary process increases computing time.

Finally, to compare our method to traditional methods determining reconstruction parameters, ours method, false nearest neighbor method (FNN) [8], conditional entropy expansion dimensional method (CEED) [9] or information entropy optimization (IEO) [23] were adopted to select $m$ and $\tau_i$ of Lorenz system, then according to the selected $m$ and $\tau_i$ reconstruct phase space, At last use three different prediction methods including average local linear prediction method (ALLP), BP neural network (BPNN), RBF neural network(RBFNN) [24, 25] forecast the last 100 samples, obtained the reconstruction parameters and prediction error (NRMSE) are shown in TABLE VI.

In TABLE VI, different prediction method has different NRMSE that BPNN and RBFNN have better predict effect than ALLP. Contrast different reconstruction parameters determining methods, we can see that whatever prediction method is used under the
condition of same variables number, relative to CEED or IEO) our method has obvious advantages on embedding dimension $m$ and prediction error (NRMSE); Relative to IFNN, our method has slightly larger of $m$, but much less prediction error.

V. CONCLUSIONS

Aimed at the correlation of determining reconstruction variables and reconstruction parameters, and the multi-dimension of evaluating prediction effect, improved multi-objective immune algorithm (IMOIA) was proposed to simultaneously determine reconstruction variables and reconstruction parameters in multivariate time series phase space.

(1) In IMOIA, antibody coding was binary strings cascade of multi-parameter, antibody evaluating was three functions, and a new population evaluating $S$ is presented.

(2) Immune operators were designed: overall clone operator improved speed; mutation operator with multiple changing probability which adjusted on gene loci and $S$ enhanced local and global search capability to ensure algorithm convergence; Clonal Selection operator on Rank and Crowding-distance ensured solutions uniformity; determine and delete operation to repetitive antibodies ensure solutions diversity.

(3) Simulation results showed that our method could simultaneously solve multiple combinations of optimal reconstruction variables and parameters and had obvious superiority to other intelligent optimization algorithms and traditional reconstruction parameter determining methods.

ACKNOWLEDGMENT

The authors are grateful to the National Natural Science Council of China and the Natural Council of Jiangsu Province for their financial supports (No. 60974126 and BK2009094).

REFERENCES


