ARTICLE IN PRESS

Swarm and Evolutionary Computation xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

journal homepage: www.elsevier.com/locate/swevo

Improved gene expression programming to solve the inverse problem for ordinary differential equations

Kangshun Li[a], Yan Chen[a,*], Wei Li[a,b], Jun He[c], Yu Xue[d]

[a] South China Agricultural University, College of Mathematics and Informatics, Guangzhou, Guangdong 510642, China
[b] Jiangxi University of Science and Technology, School of Information Engineering, Ganzhou 341000, China
[c] Aberystwyth University, Department of Computer Science Aberystwyth, Ceredigion, UK
[d] Nanjing University of Information Science and Technology, School of Computer and Software, Nanjing, Jiangsu, China

ARTICLE INFO

Keywords:
Gene expression programming
System of ordinary differential equations
Inverse problem
Runge-Kutta algorithm

ABSTRACT

Many complex systems in the real world evolve with time. These dynamic systems are often modeled by ordinary differential equations in mathematics. The inverse problem of ordinary differential equations is to convert the observed data of a physical system into a mathematical model in terms of ordinary differential equations. Then the model may be used to predict the future behavior of the physical system being modeled. Genetic programming has been taken as a solver of this inverse problem. Similar to genetic programming, gene expression programming could do the same job since it has a similar ability of establishing the model of ordinary differential systems. Nevertheless, such research is seldom studied before. This paper is one of the first attempts to apply gene expression programming for solving the inverse problem of ordinary differential equations. Based on a statistic observation of traditional gene expression programming, an improvement is made in our algorithm, that is, genetic operators should act more often on the dominant part of genes than on the recessive part. This may help maintain population diversity and also speed up the convergence of the algorithm. Experiments show that this improved algorithm performs much better than genetic programming and traditional gene expression programming in terms of running time and prediction precision.

1. Introduction

There are many complex systems or non-linear phenomena varying with the time in the real world. Such systems are called dynamic systems, including weather change, population increase, disease diffusion and so on. In order to predict the development trend of such dynamic systems, it is often required to establish their mathematical models, that is, to establish the functional relationship or changing trend among variables of the systems. It is difficult to find the functional relations among variables in complicated changing processes, but it is still possible to find out the change rate or differential coefficients of some variables, and then to model them by ordinary differential equations (ODEs). If there are more than one unknown functions, we need to establish a group of ordinary differential equations (ODEs). Through the ODEs model of a physical system, it is possible to learn the development trend of the system and apply the prediction in the real world.

The problem of converting observed data of a physical system into a mathematical model in terms of differential equations is known as the inverse problem [1,2] of differential equations [3–6]. For instance, if we have previous data of a stock market, we may create an ODEs model for the stock market using previous data and then predict the development trend of the stock market. The inverse problem of differential equations plays an important role in many areas from scientific experiments to stock markets. However, given observed data, it is not an easy task to create models of ODEs for complex dynamical systems, because these problems are very complicated and usually belong to non-linear systems, so it is difficult to determine the structure of ODEs and parameters in ODEs in order to create a correct model.

In this paper, an improved Gene Expression Programming (GEP) is put forward to solve the inverse problems of ordinary differential equations. GEP is a kind of evolutionary algorithms based on genome and phenomena and referred to the gene expression rule in the genetics [7,8]. It intends to combine the advantages of both GP and GA [9]. Unlike GP where an individual is expressed in the form of a tree, an individual in GEP is represented by the Isometric linear symbols. GEP [10] has been successfully applied in problem solving [7], combinatorial optimization [11], real parameter optimization [12], evolving and modeling the functional parameters [13], classification [14,15], event selection in high energy physics [16].

* Corresponding author.
E-mail address: cheny@scau.edu.cn (Y. Chen).

http://dx.doi.org/10.1016/j.swevo.2017.07.005
Received 31 December 2016; Received in revised form 19 June 2017; Accepted 6 July 2017
2210-6502/

Choosing GEP is based on several reasons. First, an GEP algorithm adopts a multi-gene structure, where each gene stands for an ODE and each chromosome for a group of differential equations. This is different from traditional algorithms in which an individual cannot be used to represent a group of ODEs directly. Secondly, previous experiments show that GEP algorithms have a better prediction effect in the shorter time and its time cost is so stable that it is seldom influenced by the complexity of dynamical systems. In addition, an improvement is made in our GEP algorithm. It is more suitable for studying the inverse problems of ordinary ODEs than the traditional ones because more genetic operations are centered at the dominant segment of the gene and fewer genetic operations are centered at the recessive segment [17].

The remainder of this paper is organized as follows: Section 2 introduces inverse problems of ODEs. Section 3 presents an improved GEP algorithm for solving the inverse problems of ODEs. Section 4 gives computer experiment results. Section 5 concludes the whole paper.

2. Inverse problems for ordinary differential equations

A dynamic system is represented by n correlated functions: $x_1(t), x_2(t), \ldots, x_n(t)$ where t denotes time. The system has a series of observed data collected at times $t_j = t_0 + j \times \Delta t$, $(j = 0, 1, \ldots, m-1)$, where t_0 represents the starting time, Δt the time increment, and $x_i(t_j)$ the observed value of x_i at the time t_j. Write the observation data in a matrix form:

$$\mathbf{X}_m := \begin{pmatrix} x_1(t_0), & x_2(t_0), & \ldots, & x_n(t_0) \\ x_1(t_1), & x_2(t_1), & \ldots, & x_n(t_1) \\ \ldots, & \ldots, & \ldots, & \cdots \\ x_1(t_{m-1}), & x_2(t_0), & \ldots, & x_n(t_0) \end{pmatrix}. \tag{1}$$

Denote

$$\mathbf{x}(t) := [x_1(t), x_2(t), \ldots, x_n(t)]^T, \tag{2}$$

$$\mathbf{f}(\mathbf{x}, t) := [f_1(\mathbf{x}, t), f_2(\mathbf{x}, t), \ldots, f_n(\mathbf{x}, t)]^T \tag{3}$$

where $f_j(\mathbf{x}, t) = f_j(x_1(t), x_2(t), \ldots, x_n(t), t)$ $(j = 1, 2, \ldots, n)$ is a composite function of several elementary functions involving of $x_i (i = 1, \ldots, n)$ and t. Let $\mathcal{F}$ denote the set of all possible composite functions.

A system of ordinary differential equations (ODEs) in the form of

$$\frac{dx_i(t)}{dt} = f_i(x_1, \ldots, x_n, t), \quad i = 1, \ldots, n \tag{4}$$

can be written in the vector form

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}, t). \tag{5}$$

The goal of the inverse problem of ODEs is to find a mathematical model which is represented by a system of ODEs

$$\frac{d\mathbf{x}^*(t)}{dt} = \mathbf{f}(\mathbf{x}^*, t) \tag{6}$$

such that

$$\min\{\| \mathbf{X}_m^* - \mathbf{X}_m \|; \mathbf{f} \in \mathcal{F}\} \tag{7}$$

where the matrix norm

$$\| \mathbf{X}_m^* - \mathbf{X}_m \| := \sqrt{\sum_{j=0}^{m-1} \sum_{i=1}^{n} (x_i^*(t_j) - x_i(t_j))^2} \tag{8}$$

The above the matrix norm represents the difference between the observed data and the corresponding values derived from the ODEs model.

Then we may use the obtained ODEs (6) to predicate the future trend of the system. The above problem is called the inverse problem of ODEs.

Different approaches have applied to solving the inverse problem of ODEs. Linear modeling, such as Autoregressive model, Moving Average model, Auto- regressive Moving Average model, are simple and popular [18–20]. However, there exist several restrictions for linear models. Firstly, they are linear models so that they can not represent non-linear dynamical systems. Secondly, identification and estimation of linear models requires strong mathematical knowledge and expertise, which often lacks in practice. Finally, once a model is established, it is not easy to constantly adjust the structure and parameters of the model based on updated observation data.

Another simple modeling approach is to take a form of differential equations which are pre-selected by experience, and then a numerical method is used to determine the variables [21]. However, how to pre-select the right differential equation model is a difficult task, especially for the differential equations whose number of variables increases.

Evolutionary modeling [22–25] has been successfully used in studying the inverse problems of ordinary differential equations. Current evolutionary modeling are mainly based on Genetic Programming [22,25–27] where an equation is represented in the form of tree. A hybrid evolutionary methods are proposed for evolving ODEs, for example, to predict small-time scale traffic measurements data in [28], which uses tree model to evolve but its speed is also slow. ECSID [29] found good models for linear pendulum, non-linear pendulum with friction, coupled mass-spring, and linear circuit. But the difference between model found by ECSID and original model becomes large when the model becomes complex. GEP-SWPM is proposed in [30], and the prediction is based on several generations of data before, so it is seriously affected by the noise. A new methods of GEP was proposed in [13,31], which is very effective for identifying parameter functions. But it is based on the assumed model and doesn't provide a common solution and it can't be extended to most situations.

In this paper, we proposed an improved GEP algorithm for solving the inverse problem of ODEs.

3. Improved GEP for the inverse problem of ODEs

3.1. Gene representation in GEP algorithm

The genetic codes of GEP is the isometric linear symbols (GEP chromosome). Each chromosome can be composed of several genes. GEP gene consists of a head and a tail, where the former may contain both the functional symbols and termination symbols, while the latter only has the terminal symbols. For example, * + − aQ* + aababbbaab is a legal gene, of which, * stands for the multiplication operation, Q the square root operation, the segment without underline belongs to the head, while the underlined segment is the tail. Fig. 1 shows the expression of the gene in the form of a tree.

For each problem, the length of the tail t is a function of the length of the head h and the number of arguments of the function with the

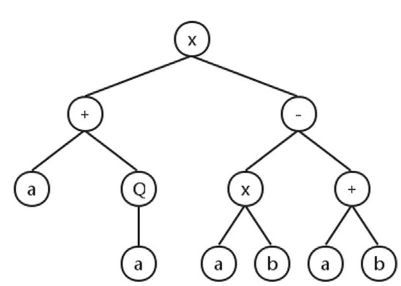

Fig. 1. Expression tree 1.