Information Sciences 178 (2008) 4453–4468

Contents lists available at ScienceDirect

Information Sciences

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

Inference of differential equation models by genetic programming

Hitoshi Iba *

Department of Frontier Informatics, Graduate School of Frontier Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku 113-0033, Japan

ARTICLE INFO

Keywords:
Genetic programming
Ordinary differential equations
Genome informatics

ABSTRACT

This paper describes an evolutionary method for identifying a causal model from the observed time-series data. We use a system of ordinary differential equations (ODEs) as the causal model. This approach is known to be useful for practical applications, e.g., bioinformatics, chemical reaction models, control theory, etc. To explore the search space more effectively in the course of evolution, the right-hand sides of ODEs are inferred by genetic programming (GP) and the least mean square (LMS) method is used along with the ordinary GP. We apply our method to several target tasks and empirically show how successfully GP infers the systems of ODEs. We also describe an extension of the approach to the inference of differential equation systems with transcendental functions.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Ordinary differential equations (ODEs) are one of the easiest media for modeling complex systems, where basic differential relationships are known between the system components. Solving a set of differential equations to produce their equivalent functions (for obtaining useful time-series data) is relatively easy. On the other hand, the inverse problem, i.e., the inference of the system of ODE from the observed time-series data, is not necessarily easy, although very important for many fields. This is because the appropriate form of the ODE (i.e., the order and terms) is not known in advance.

In this paper, we deal with an arbitrary form in the right-hand side of the system of ODEs to allow for flexibility of the model. More precisely, we consider the following general form:

$$\frac{dX_i}{dt} = f_i(X_1, X_2, \ldots, X_n) \quad (i = 1, 2, \ldots, n), \tag{1}$$

where X_i is the state variable and n is the number of the observable components.

For the sake of identifying the system, we use genetic programming (GP) to evolve the ODEs from the observed time series. Although GP is effective in finding a suitable structure of the purported solution, it is sometimes difficult to optimize the parameters, such as constants or coefficients of the polynomials. This is because the ordinary GP searches for them simply by combining randomly generated constants. To avoid this difficulty, we introduce the least mean square (LMS) method.

There have been several studies for identifying differential equation models by means of EAs (evolutionary algorithms). For instance, GP was used to find a function in a symbolic form, which satisfies the differential equation and initial conditions [9]. Cao et al. used hybrid evolutionary modeling algorithms [4]. The main idea was to embed GA in GP, where GP was employed to discover and optimize the structure of a model, while GA was used to optimize its parameters, i.e., coefficients. GP was applied [3] to approximate several ODEs from the domain of ecological modeling, e.g., Lotka–Volterra and logistic equations. They showed that the GP-based approach introduced numerous advantages over the most available modeling

* Tel.: +81 3 5841 7424; fax: +81 3 5841 6070.
E-mail address: iba@iba.k.u-tokyo.ac.jp
URL: http://www.iba.k.u-tokyo.ac.jp/

0020-0255/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2008.07.029

methods. Grammatical evolution was also used for the sake of solving ordinary and partial differential equations [20]. In our previous researches [16,17], we proposed another integrated scheme, in which the least mean square (LMS) method is used along with GP. In that scheme, some individuals were created by the LMS method at some intervals of generations and they replaced the worst individuals in the population.

In this paper, we extend our previous approach, in order to achieve the inference of the ODEs more effectively. More precisely, we empirically show the following points:

- The success in the acquisition of ODEs, which are close to the observed time series.
- The inference of the exact equation form, i.e., the exact causal relationship.
- The effectiveness of the LMS method.
- The superiority of our approach over the previous methods.

The rest of this paper is organized as follows. In Section 2, we describe the details of our method, i.e., how GP and LMS methods are integrated to work in the course of evolution. Three examples are used to examine the effectiveness of our method and experimental results are shown in Section 3. The extension of this approach is described in Section 4 for the inference problem of a differential equation system including transcendental functions. Results are discussed in Section 5 and we draw conclusions in Section 6.

2. Integration of GP and LMS

We use GP to identify a causal model in the form of the system of ODEs. Though GP is capable of finding a desirable structure effectively, it cannot always be effective in finding the proper coefficients because GP uses the combination of randomly selected ones. We have chosen the least mean square method (LMS) to tackle this deficiency of the ordinary GP. For this purpose, coefficients are not included in the terminal set of a GP individual tree. The coefficients of each term of a GP tree are calculated by the LMS method. A GP individual is represented by a tree and a table of coefficients.

2.1. Inference of the form of equations using GP

We use GP to identify the form of the system of differential equations. For this purpose, we encode right-hand sides of ODEs into a GP individual. Each individual contains a set of n trees, i.e., an n-tuple of trees $(f_1, \ldots, f_n)$. For example, consider the two trees in Fig. 1. This shows the following system of ODEs:

$$\begin{cases} \dot{X}_1 = aX_1X_2^2 + b, \\ \dot{X}_2 = cX_1X_2 + dX_2, \end{cases} \tag{2}$$

where the coefficients a, b, c, d, are derived by LMS (described later in this paper). Note that the constant term b is added to the right-hand side of the first equation, because of the constant terminal, i.e., 1. Thus, each equation uses a distinct program. A GP individual maintains multiple branches, each of which serves as the right-hand side of a differential equation.

Crossover operations are restricted to corresponding branch pairs. Actually, each tree, i.e., each right-hand side of the ODE system, is evolved independently in parallel.

2.2. Optimization of models using LMS method

Coefficients of a GP individual are derived by the LMS method described below. Assume that we want to acquire the approximate expression in the following form:

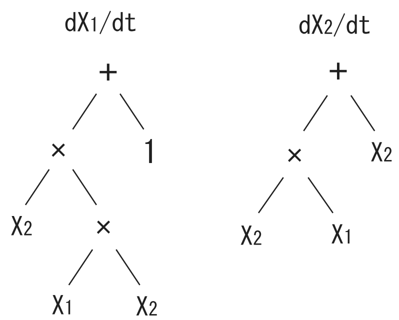

Fig. 1. Example of a GP individual.

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات