



Artificial bee colony programming for symbolic regression

Dervis Karaboga, Celal Ozturk*, Nurhan Karaboga, Beyza Gorkemli

Erciyes University, Engineering Faculty, Intelligent Systems Research Group, Kayseri, Turkiye

ARTICLE INFO

Article history:

Received 21 November 2011
 Received in revised form 25 April 2012
 Accepted 4 May 2012
 Available online 15 May 2012

Keywords:

Genetic programming
 Symbolic regression
 Artificial bee colony algorithm
 Artificial bee colony programming

ABSTRACT

Artificial bee colony algorithm simulating the intelligent foraging behavior of honey bee swarms is one of the most popular swarm based optimization algorithms. It has been introduced in 2005 and applied in several fields to solve different problems up to date. In this paper, an artificial bee colony algorithm, called as Artificial Bee Colony Programming (ABCP), is described for the first time as a new method on symbolic regression which is a very important practical problem. Symbolic regression is a process of obtaining a mathematical model using given finite sampling of values of independent variables and associated values of dependent variables. In this work, a set of symbolic regression benchmark problems are solved using artificial bee colony programming and then its performance is compared with the very well-known method evolving computer programs, genetic programming. The simulation results indicate that the proposed method is very feasible and robust on the considered test problems of symbolic regression.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Symbolic regression, a process of evolving summary expressions for available data by analyzing and modeling numeric multi-variate data sets, is used when some data of unknown process are obtained. Unlike traditional linear and nonlinear regression methods that fit parameters to an equation of a given form, symbolic regression tries to form mathematical equations by searching the parameters and the form of equations [41]. In other words, symbolic regression method searches nonlinear equation form and its parameters simultaneously for an addressed modeling problem. It attempts to derive a mathematical function to describe the relation between dependent and independent variables. The problem of symbolic regression is an optimization problem the purpose of which is finding the best combination of variables, symbols, and coefficients to develop an optimum model satisfying a set of fitness cases. Moreover, task of symbolic regression is to identify the variables (inputs) in the data that affect the changes in the important control variables of interest (outputs), to express these effects in mathematical models and to analyze the quality and generality of the constructed relationships [30].

Evolutionary computing (EC) techniques are being successfully applied to numerous problems, including optimization, symbolic regression, automatic programming, signal processing, bioinformatics, social systems, etc. [12]. A computational intelligence technique used to discover mathematical expressions belongs to the broad class of evolutionary computation techniques in which evolution models and evolutionary operators are used. More precisely, evolutionary computing techniques maintain a population of structures (expressions) and evolve them according to rules of selection and search operators. Evolutionary computing is the whole name for a range of problem-solving techniques [6] such as evolutionary programming [14], evolution strategies [39,42], genetic algorithms [17], differential evolution [44], genetic programming

* Corresponding author.

E-mail addresses: karaboga@erciyes.edu.tr (D. Karaboga), celal@erciyes.edu.tr (C. Ozturk), nurhan_k@erciyes.edu.tr (N. Karaboga), bgorkemli@erciyes.edu.tr (B. Gorkemli).

[31], and also swarm based algorithms such as artificial immune system [10], particle swarm optimization [29], ant colony optimization [8], honey-bees optimization [1,2,4], and artificial bee colony optimization [21].

Genetic programming (GP) is the most popular technique used in symbolic regression and was invented by Cramer in 1985 [9] and developed by Koza in 1992 [31] lead to known as standard GP. GP can be accepted as an extended version of genetic algorithms (GAs), where the main differences between them are the representation of the structure and the meaning of representation. In GP, the programs are expressed as parse trees instead of code lines and the individual population members are variable-length character strings that encode possible solutions [5]. So, GP is a kind of genetic algorithm, which relies on operating on a population of parse trees and selecting a group of improved parse trees according to their fitness and also having main operators: crossover, mutation, permutation. It should be noticed that most GP applications use only tree crossover as the genetic operator. In all cases, the solution must be expressed by using an entire parse tree [15].

GP has captured much attention and has been applied to solve many practical problems, in the last decade [37]. Within the basic form of GP, there is much interest in the research to the development of semantic-aware variants of GP [34] which are using operators that respect the semantics of programs or solutions and performing better search with higher-level constructs. Despite the growing attention to semantic-aware forms of GP, there are other different improved versions of genetic programming, for example, linear genetic programming [18], cartesian genetic programming [35], gene expression programming, etc. Gene expression programming (GEP), an extension of GP and GAs, was firstly introduced in [13]. The main difference between the GA, GP, and GEP stands in the nature of the individuals: in GAs the individuals are linear strings of fixed length (chromosomes); in GP the individuals are nonlinear entities of different sizes and shapes (parse trees); and in GEP the individuals are encoded as linear strings of fixed length (the genome or chromosomes) which are afterwards expressed as nonlinear entities of different sizes and shapes (i.e., simple diagram representations or expression trees) [13]. Prefix Gene Expression Programming (PGE) is an adaptation of GEP and it also does not have separate genotypes (genetic makeup) and phenotypes (body and behavior), like GEP [32].

Differential evolution (DE) algorithm is similar to genetic algorithm in terms of used evolutionary operators. It employs crossover, mutation and selection operators as used in GA. While GA relies on crossover in constructing better solutions, DE relies on both crossover and mutation operations. DE algorithm uses mutation operation as a search mechanism and selection operation to direct the search toward the prospective regions in the search space. In [7], DE algorithm is evolved with prefix gene expression programming (PGE) to approximate the values of leaf nodes by changing the tree structure.

Swarm intelligence is an artificial intelligence technique involving studies of collective behaviors in decentralized natural artificial or artificial systems. Swarm based algorithms have been shown to outperform other algorithms in many important applications, such as optimization, machine learning, computer security, data mining, clustering, pattern recognition, and function approximation. Despite developing in many areas, there is few research on using swarm based algorithms in symbolic regression or automatic programming. Artificial immune system, inspired by the human immune system, is used as optimal search engine in [36], where called immune programming. Clone selection algorithm [11], a kind of artificial immune system algorithm, is proposed for symbolic regression by using dynamic fitness functions in [20]. Ant colony optimization is applied for symbolic regression in [33,40,43] as building and modifying expressions using tree structures by ants. There are studies of hybridization of the genetic programming with ant colony optimization or particle swarm optimization. In [3], ant colonies are combined with grammar-guided genetic programming and in [38], the particle swarm optimization is explored for the possibility of evolving optimal force generating equations to control the particles using genetic programming.

In this paper, an artificial bee colony (ABC) algorithm, which was originally introduced by Karaboga based on the foraging behavior of honey bees for numerical optimization problems [22], is described to solve symbolic regression benchmark problems. There has recently been growing interest in applications of ABC algorithm to many complex problems of the real world. This is the first study of ABC based method evolving functions on symbolic regression, which is called as ABC programming (ABCP). The performance of the proposed programming approach is compared with the results of GP approach which is well-known and most widely used symbolic regression technique. The organization of the paper is as follows: ABC algorithm is described in Section 2, ABCP is introduced in Section 3, and experiments and results are presented and discussed in Section 4. The paper is concluded in Section 5 with remarking the future work.

2. Artificial bee colony algorithm

Artificial bee colony algorithm that simulates the intelligent foraging behavior of honey bee swarms was introduced by Karaboga in [21]. The ABC algorithm was tested on unconstrained and constrained problems in [23,24] by comparing the performance of the algorithm with those of other well-known evolutionary computing such as genetic algorithm, differential evolution, and particle swarm optimization. ABC algorithm was used in machine learning on classification by training neural networks in [25] and by clustering techniques in [26]. The comparisons demonstrated that the performance of the ABC algorithm is competitive to other population-based algorithm with advantages of having simple structure and employing fewer control parameters. The basic version of the artificial bee colony algorithm has only one control parameter called limit apart from the common control parameters of the population-based algorithms such as population size (colony size (SN)) and maximum generation number (maximum cycle number (MCN)). Detailed pseudo-code of the basic ABC algorithm is as given in [26]:

متن کامل مقاله

دریافت فوری ←

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