Solver-Based Fitness Function for the Data-Driven Evolutionary Discovery of Partial Differential Equations

Mikhail Maslyaev
Nature Systems Simulation Laboratory
ITMO University
Saint-Petersburg, 197101, Russia
mikemaslyaev@itmo.ru

Alexander Hvatov
Nature Systems Simulation Laboratory
ITMO University
Saint-Petersburg, 197101, Russia
alex_hvatov@itmo.ru

***Abstract*—Partial differential equations provide accurate models for many physical processes, although their derivation can be challenging, requiring a fundamental understanding of the modeled system. This challenge can be circumvented with the data-driven algorithms that obtain the governing equation only using observational data. One of the tools commonly used in search of the differential equation is the evolutionary optimization algorithm. In this paper, we seek to improve the existing evolutionary approach to data-driven partial differential equation discovery by introducing a more reliable method of evaluating the quality of proposed structures, based on the inclusion of the automated algorithm of partial differential equations solving. In terms of evolutionary algorithms, we want to check whether the more computationally challenging fitness function represented by the equation solver gives the sufficient resulting solution quality increase with respect to the more simple one. The approach includes a computationally expensive equation solver compared with the baseline method, which utilized equation discrepancy to define the fitness function for a candidate structure in terms of algorithm convergence and required computational resources on the synthetic data obtained from the solution of the Korteweg-de Vries equation.**

***Index Terms*—equation discovery, partial differential equation, fitness function selection, data-driven modelling**

I. Introduction

Differential equations are commonly used as mathematical models for continuous physical processes. They describe the interdependence of various derivatives of a studied multivariate function. In addition, differential equations can be used for system state predictions. By integrating the governing partial differential equation with specific initial and correctly stated boundary conditions, the state of the modeled process in the future can be obtained.

The task of partial differential equation derivation for a specific problem in the past involved examining conservation laws that can be applied to the system and using analysis and variational principles to extract the equation from the known properties of the system. While these methods are widely used to derive the equations, they demand significant preliminary study of the process, which in some cases can not be held due to low comprehension of the system. Occasionally, regardless of the dynamical system understanding, the researchers must collect measurements as a preliminary part of the analysis. Thus, the data-driven equation derivation can be introduced as an alternative. Data-driven algorithms can develop a model for a process using measurements data and a few assumptions about the constructed model structure (i.e., the model will take the form of a partial differential equation).

One of the contemporary approaches to data-driven PDE discovery is based on evolutionary algorithms (EA). Here, the objective of EA is the construction of a model in the form of a partial differential equation from a selected set of elementary constituents and with specific conditions that have the best performance on the input data. One of the issues linked with applying evolutionary operators to such ambiguously stated tasks is selecting an objective (fitness) function that shall be optimized during the problem solution.

The computation complexity of the fitness function is a classical problem that is considered for classical optimization benchmark problems [1]. However, in the equation discovery case, the function landscape cannot be known a priori. Thus, only empirical conclusions on the influence of the hardness of computational complexity of a fitness function may be obtained. In this paper, two possible approaches to compute fitness functions for equation discovery evolutionary algorithm are considered. First is the computationally hard complete equation solution. Equation solution should be a more noise-resistant approach to evaluating the quality of evolutionary algorithm candidates due to the independence of a noisy data differentiation error. The ability to handle noisy data is crucial for modeling real-world physical systems since instrumental observational data gathering is usually imposed by various disturbances. Moreover, the solver approach allows considering boundary values, which are not in the scope of previously used computationally more straightforward fitness calculation technique. It involved evaluating differential operator discrepancy from zero and performing poorly on noisy data.

The paper is structured as follows: Section II is dedicated to the analysis of existing approaches of data-driven discovery of