Data-driven modelling of LTI systems using symbolic regression

Data-Driven Modelling of LTI systems using Symbolic Regression

D. Khandelwal, R. Tóth, P.M.J Van den Hof
Control Systems Group
Department of Electrical Engineering
Eindhoven University of Technology
P.O. Box 513, 5600 MB, Eindhoven, The Netherlands
Email: D.Khandelwal@tue.nl

1 Introduction

The aim of this project is to automate the task of data-driven identification of dynamical systems. The underlying goal is to develop an identification tool that models a physical system without distinguishing between classes of systems such as linear, nonlinear or possibly even hybrid systems. Such an identification tool would be able to mine data generated by the system to infer such structural knowledge, without relying on the expertise of a skilled user. This will allow researchers to shift their focus back from the modelling task to the actual utilization of the model. Such a research objective requires the identification technique to employ tools that are not targeted towards nuanced modelling tasks, but remain applicable for a very broad range of systems. Hence, we seek to develop a new framework for system identification that uses generic tools.

2 A hybrid evolutionary algorithm for identification of LTI systems

Evolutionary computation techniques are a class of nature-inspired meta-heuristic algorithms that seek to *evolve* optimal solutions to a given problem, based on a user-defined fitness function. Evolutionary algorithms stand out as a candidate computational technique for the the stated research objective due to their non-dependence on the problem domain. This characteristic stems from the fact that these techniques find solutions by validating many candidate solutions rather than estimating/computing a single optimal solution.

We approach to the task of developing a framework for *structure-free* identification by making a distinction between the two aspects of the underlying computational challenge:

1. Global search - This refers to the search of appropriate model structure that can sufficiently capture the dynamic relations present in the data.
2. Local search - This refers to the search of an optimal model within a chosen model structure. This essentially boils down to parameter estimation for any chosen model structure.

The multi-faceted nature of the challenge necessitates the utilization of a computational tool that can address these two tasks efficiently. To approach this problem, we develop a hybrid algorithm (also known as Memetic Algorithm (MA) in evolutionary literature), based on Genetic Programming (GP) and Evolutionary Strategies (ES), with each component of the algorithm tuned to address the two challenges separately.

3 Simulation Results

The proposed MA was applied to the identification of a discrete-time model of an SISO LTI system. Prediction Error Method (PEM) was used as a benchmark to compare the simulation results. The data-generating system was chosen to be a Box-Jenkins model. To this extent, the MA was tuned to evolve 4 polynomials $B(q^{-1}), F(q^{-1}), C(q^{-1})$ and $D(q^{-1})$, that are interepreted as:

$$y(k) = \frac{B(q^{-1})}{F(q^{-1})}u(k) + \frac{C(q^{-1})}{D(q^{-1})}e(k), \qquad (1)$$

where, y, u and e are the output, input and white noise process respectively, and q is the time-shift operator.

The two identification routines were applied to data generated from the simulation example in a 100-run Monte Carlo simulation. The average squared error computed on validation data-sets over the 100 runs is depicted in Figure 1. The figure shows that average prediction error obtained from the MA is comparable to that of PEM, thereby validating the proposed concept of the MA as an optimization tool for system identification.

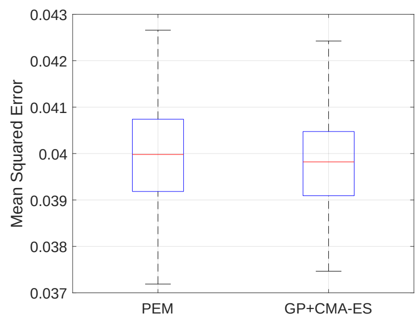

Figure 1: Symbolic regression illustrated.