Measurement 114 (2018) 122–131

Contents lists available at ScienceDirect

Measurement

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

An application of evolutionary system identification algorithm in modelling of energy production system

Yuhao Huang[a], Liang Gao[b], Zhang Yi[c], Kang Tai[d], P. Kalita[e], Paweena Prapainainar[f], Akhil Garg[a,*]

[a] Department of Mechatronics Engineering, Shantou University, Shantou 515063, China
[b] The State Key Laboratory of Digital Manufacturing Equipment and Technology, 1037 Luoyu Road, Huazhong University of Science and Technology, Wuhan 430074, China
[c] Department of Civil and Earth Resources Engineering, Graduate School of Engineering, Kyoto University, Japan
[d] School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore
[e] Centre for Energy, Indian Institute of Technology Guwahati, Guwahati 781 039, India
[f] Department of Chemical Engineering, Faculty of Engineer, Kasetsart University, Bangkok 10900, Thailand

A R T I C L E I N F O

Keywords:
System identification
Modelling methods
Genetic programming
Fuel cell
Energy system

A B S T R A C T

The present work introduces the literature review on System Identification (SI) by classifying it into several fields. The review summarizes the need of evolutionary SI method that automates the model structure selection and its parameter evaluation based on only the system data. In this context, the evolutionary SI approach of genetic programming (GP) is applied in modeling and optimization of cleaner energy system such as direct methanol fuel cell. The functional response of the power density of the fuel cell with respect to input conditions is selected based on the minimum training error. Further, an experimental data is used to validate the robustness of the formulated GP model. The analysis based on 2-D and 3-D parametric procedure is further conducted to reveals insights into functioning of the fuel cell. The pareto front obtained from optimization of model reveals that the operating temperature of 64.5 °C, methanol flow rate of 28.04 mL/min and methanol concentration of 0.29 M are the optimum settings for achieving the maximum power density of 7.36 mW/cm^2 for DMFC.

1. Introduction

The art of building mathematical models from the input–output data obtained from the System is known as System Identification. The main purpose of building the models from the system is to understand its behavior and predicts its performance for fault diagnosis. The term modeling has been extensively used in the field of SI. Basically, several types of models, systems and methods can be studied under various fields of SI as shown in Fig. A1 in Appendix A [1]. The systems modelled can be the manufacturing processes, cleaner energy production systems such as fuel cell or such as those involving the study of mechanical and thermal properties of graphene and carbon nanotubes or the stock market and weather phenomenon, etc. [1]. Among these processes, the energy storage systems such as fuel cells and batteries are the potential ones because they reduce the environmental burden arising due to toxic gases emitted by industries and by transportation vehicles. The system complexity and its functioning depends on the multiple set of input–output variables attributed to it. The cost incurred in optimization of system behavior is high because it is not economical to measure the data. The main concept behind analyzing the obtained data from the system is to unravel the useful information that is responsible for its long term efficient performance before the system is replaced. Given with the multi-functional nature of the systems, the need of modeling and optimization has been strengthened to understand its complex behavior.

The work described in this manuscript is divided into seven sections. Section 2 discusses the models and modeling methods classified under various fields of SI. Section 3 discusses the alternative methodology (evolutionary approach of Genetic programming (GP)) suggested in the latest trends in era of SI. Section 4 illustrates the application of GP in modeling of the fuel cell system. Section 5 discussed the model formulation and statistical fit on the experimental data. Section 6 provides the uncertainty analysis, 2-D and 3-D surface analysis and optimization results of GP model. Finally, Section 7 concludes the study with recommendations and novelty in work.

2. Literature review on SI

This section describes the various types of models, systems and methods being studied in SI. The methods such as those based on

* Corresponding author.
E-mail address: akhil@stu.edu.cn (A. Garg).

http://dx.doi.org/10.1016/j.measurement.2017.09.009
Received 13 April 2017; Received in revised form 29 August 2017; Accepted 11 September 2017
Available online 18 September 2017
0263-2241/

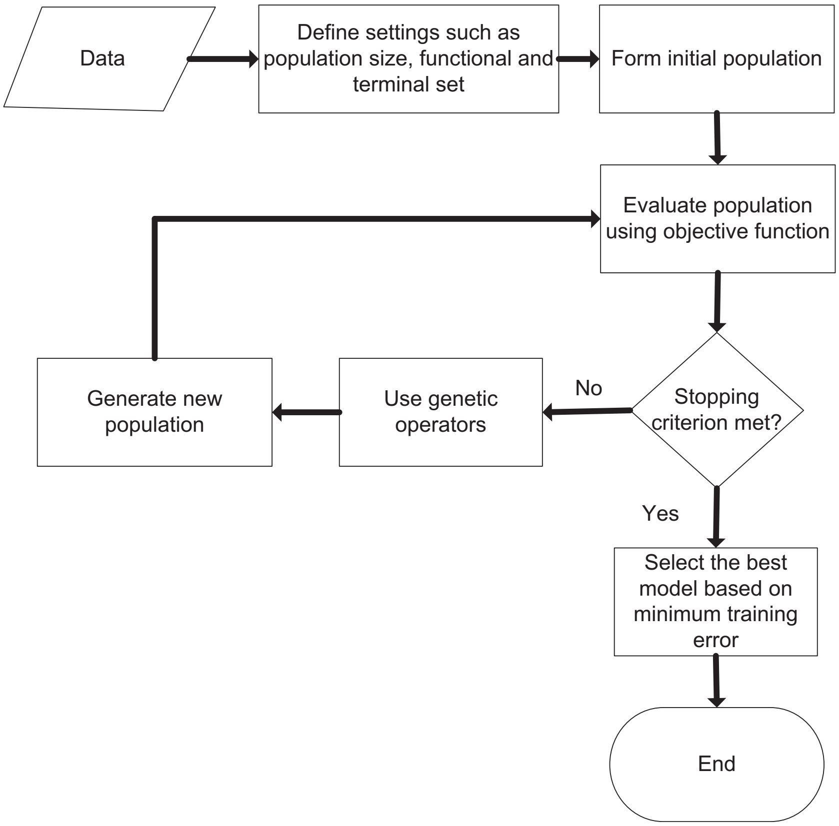

Fig. 1. Steps for Implementing GP approach.

Table 1
Three operating conditions used in the experimental study.

No.	Inputs for DMFC	Unit	Minimum, maximum value for DMFC
1.	Methanol flow rate (mL/min)	mL/min	0.5, 54.5
2.	Methanol conc. (M)	M	0.25, 3.25
3.	Operating temperature	$°C$	26,74

statistics, finite element analysis or artificial intelligence can be used to formulate models. Models are also formed using Analysis of Variance (ANOVA) or the hypothesis tests. These models however serve common purpose to understand the complex systems behavior. Models build on only the given input–output continuous data are known as regression models. The methods such as regression analysis, response surface methodology (RSM), partial least square regression, genetic programming (GP), artificial neural network (ANN), fuzzy logic (FL), M5- prime (M5′), support vector regression (SVR), adaptive neuro-fuzzy inference systems (ANFIS), etc. can be applied [2–7] to formulate these models. The models build must not only accurate predict the system output but shall also satisfy the system constraints. This implies that the models should not provide non-permissible values beyond the operating range of the system. For optimization of systems, industry experts generally demand simpler, accurate and explicit models, which can be easily coded into system [1].

Under various classifications of SI [8,9], several types of Systems, models as well as modeling methods can be studied. For example, modeling methods can be studied under the three categories of modeling: grey box, white box and black box [10]. In grey box modeling, the model structure (polynomial, differential equations) is assumed and then its coefficients are computed using the standard optimization methods. In white box modeling, the models are formulated based on the first/principle laws. For example, newtons law and gravitational law. In black box modeling, the model form and system behavior is completely unknown. The methods such as neural network, genetic programming can be used to model the system from the given data automatically.

Another way of classifying SI is into linear, nonlinear and evolutionary [8]. Due to multi-functional nature of the systems, these systems behave non-linearly. The experts mainly used the methods classified under non-linear SI such as the kernel method, linear regression and FFT for modeling the systems behavior [11–13]. However, these methods require the extensive computation of the set of coefficients, which is deemed not reliable and induces ambiguity in the predictive ability of the model [14,15]. To counter this problem, the evolutionary system identification approach of GP can be used to model the system output explicitly from only the given input–output data [16]. Third route of studying the methods, models and systems is by classifying the SI into various fields such as statistics, econometrics, machine learning (ML), statistical learning theory, statistical process control, chemometrics, etc. [9]. In brief, the modeling methods are categorized mainly into Statistical methods (regression analysis, RSM, partial least square regression, etc.) [4]. However, these are based on the assumptions such as the pre-definition of form of a model, residuals properties, etc. which may not be always reliable to model the system in presence of higher non-linearity. On the other hand, the methods classified under CI comprises of GP, ANN, M5′, SVR, ANFIS, etc. [17–19].

It is well learned that by merging features of methods from several fields of SI, an efficient hybrid CI methods such as GA-FL, GA-ANN, FL-