Identification of Switched Models in Non-Stationary Time Series based on Coordinate-Descent and Genetic Algorithm*

Mohammad Gorji Sefidmazgi, Mina Moradi Kordmahalleh, and Abdollah Homaifar
Department of Electrical Engineering, North Carolina A&T State University
Greensboro, NC, USA
{mgorjise, mmoradik}@aggies.ncat.edu, homaifar@ncat.edu

ABSTRACT

Time series analysis is an important research topic in science and engineering. Real-world time series are usually non-stationary with time-varying parameters. Identification of non-stationary time series with a switched model includes finding the switch times and model parameters in each cluster. This problem is a non-convex optimization with equality constraints. Conventional identification methods suffer from restrictive statistical assumptions about the data or switch times, locality of solution, and computational complexity particularly for longer time series. In this paper, a novel coordinate-descent algorithm with the genetic algorithm (GA) and statistical inference is developed. In the evolutionary process, innovative types of crossover and mutation are proposed to improve exploration and exploitation capabilities of the GA, and fitness of the individuals are calculated by the maximum likelihood or least-mean-square.

CCS Concepts

•**Mathematics of computing** → **Time series analysis;** •**Information systems** → *Clustering;* •**Computing methodologies** → Genetic algorithms;

Keywords

System Identification, Time Series Analysis, Genetic Algorithm, Regression, Maximum Likelihood

1. INTRODUCTION

Complex systems usually exhibit nonlinear deterministic or stochastic characteristics, and their generated time series are modeled in order to provide insight into the system behavior [5]. Common modeling approaches based on statistical inference only work for modeling of time series with time-invariant parameters. However, real-world time series in climate, economy, biology, and engineering are generally non-stationary, i.e. their statistical properties change with time. An important model of non-stationary time series is the switched model, where the time series has some clusters (segments). Each cluster has its own distinct and time-invariant model parameters, and there is a switching mechanism between the clusters. Identification of switched models requires determining the switch times among the clusters and the mathematical model of each cluster. A time series with T data points has approximately T^W distinct placements of W switches. The mathematical model is considered as a function of time (such as a polynomial or differential equation) or a probability density function (such as a Gaussian distribution).

Several identification approaches exist in literature including the hypothesis tests, Bayesian inference, Hidden Markov Models, convex relaxation and brute force search. The major issues in aforementioned approaches are locality of solution, restrictive statistical assumptions and high volume of computations (especially for longer time series)[2]. In this paper, a new method is proposed which combines the GA and statistical inference under a coordinate-descent algorithm. In this method, each individual in the GA represents the switch times and the cluster sequence in the time series. For generation of offsprings, innovative crossover and mutation operators are proposed to evolve the switch times and the cluster sequences in the population. The cluster parameters and error of identification are determined by the least-mean-square/maximum likelihood.

2. IDENTIFICATION METHOD

Let $x(t) \in \mathbb{R}^n$ be a multidimensional non-stationary time series over $t = \{1, \ldots, T\}$ with C clusters. The model of the time series in each cluster could be a function of time and other inputs, i.e. $x(t) = f(x(t-1), ..., x(t-p), u(t-1), ...$ $, u(t-q), t, \theta_c)$ or a probability density function $P(X = x(t)) = f(x(t)|u(t), \theta_c)$. Here, f is the model of the time series, and θ_c is the set of parameters in the c-th cluster. For these two case, the *distance function* $d(x(t), \theta_c)$ between the time series at time t and the model of the c-th cluster is respectively defined using Euclidean distance $\|x(t) - f(x(t-1), ..., x(t-p), u(t-1), ..., u(t-q), t, \theta_c\|^2$ or negative log-likelihood function $\ell(f(x(t)|u(t), t, \theta_c))$. Then, the identification problem is defined as a minimization problem as $\min_{\mu,\theta} \sum_{t=1}^{T} \mu_c(t) d(x(t), \theta_c)$. Here, $\mu_c(t) \in \{0, 1\}$ is the *cluster membership function* indicating whether the data at time t belongs to the c-th cluster or not, and $\sum_{c=1}^{C} \mu_c(t) = 1$. The above mentioned optimization is a non-convex mixed-integer optimization with two sets of unknown parameters. One solution to the problem is to use the coordinate-descent technique that solves the problem in two iterative steps with respect to $\mu_c(t)$ and θ_c. In order to make the problem well-posed, a constraint is added to the problem formulation such

*This work is supported by the Expeditions in Computing by the National Science Foundation under Award Number: CCF-1029731: Expedition in Computing: Understanding Climate Change. Contribution of the third author has been supported by the Air Force Research Laboratory and OSD under agreement number FA8750-15-2-0116.

GECCO '15 July 11-15, 2015, Madrid, Spain

ACM ISBN 978-1-4503-3488-4/15/07.
DOI: http://dx.doi.org/10.1145/2739482.2764687

that the number of switches is known equals e.g. W. For real time series problem with unknown number of the clusters and the switches, the optimal value of these parameters can be estimated by *Bayesian Information Criterion* [4].

Each Individual represent W switch times among the C clusters, and $W+1$ corresponding indices of the clusters sequence in a non-stationary time series with the length of T. The switch times are integer numbers from the set of $\{1, \dots T\}$ and the sequence indices indicate the name of the clusters from the set of $\{1, \dots C\}$. A sample individual is shown in Fig.1 where the length of time series is $T = 250$ and there are six switches among the three clusters. The clusters 1, 2 and 3 are activated at $t \in \{1, \dots, 9\} \cup \{40, \dots, 69\} \cup \{150, \dots, 200\}$, $t \in \{10, \dots, 39\} \cup \{95, \dots, 149\} \cup \{201, \dots, 250\}$ and $t \in \{70, \dots, 94\}$ respectively. It is worth noting that the sequences with consecutive identical clusters (such as {1, 2, 2, 1, 3, 2, 1} are not valid. In an ordered sequence, the first and the second loci are always 1 and 2 respectively. The other loci of the sequence might be any indices from the set of $\{1, \dots, C\}$ with no consecutive equal indices.

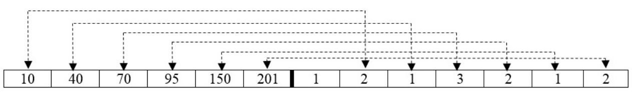
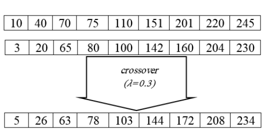
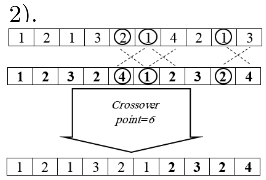


Figure 1: A sample individual with three clusters and six switches.

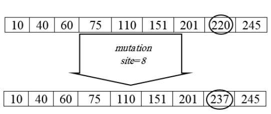
Since the proposed individual represents the set of switch times as well as the cluster sequences, the crossover and mutation operators should be applied on both sets. We generate a random number $\alpha \in [0, 1]$ ($\beta \in [0, 1]$), and if $\alpha \geq 0.5$ ($\beta \geq 0.5$), the crossover (mutation) is applied on the switch times. If $\alpha < 0.5$ ($\beta < 0.5$), the crossover (mutation) is applied on the cluster sequence. The crossover and mutation on the switch time are similar to real-valued crossover, with an additional *floor* operator. The crossover/mutation on the cluster sequence, generate a new sequence without consecutive identical clusters (Fig. 2).



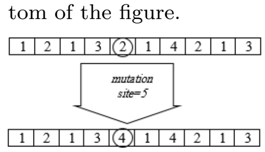
(a) The crossover is applied on the switch times of the parents to find a new set of switch times.



(b) The possible crossover sites are marked with circle. For $z = 6$, the crossover child is shown at the bottom of the figure.



(c) Mutation on the switch times, mutation site $z = 8$ is chosen randomly and the switch times is mutated to another value in the range of $\{202, \dots, 244\}$.



(d) Mutation on sequence of the clusters: the mutation site $z = 5$ (marked with circle) is randomly selected. The sequence value is mutated to 4.

Figure 2: Crossover and Mutation operator on the switch times and the sequence of individuals with $W = 9$, $C = 4$ and $T = 250$

The next step is to calculate the fintness of each individual. According to each individual $\mu_c(t) = 1$, if $x(t)$ belongs to the cluster c at time t. Otherwise $\mu_c(t) = 0$. Then, the cluster parameters θ_c should be obtained by solving an unconstrained optimization as $\min_{\theta_c} \sum_{t=1}^{T} \mu_c(t) d(x(t), \theta_c)$ for $c \in \{1, \dots, C\}$. Since the parameters of the C clusters are independent, this optimization is converted to C separate optimizations. Depending on the mathematical model of the clusters, each optimization is solved either by the least-mean-square or the maximum likelihood [1]. After finding the cluster parameters, error of identification can be calculated using the cost function. In this paper, we implemented the steady-state GA called Crowding/Replace Worth (CD/RW) [3] which uses a hybrid replacement scheme to replace the individuals with poor fitness and low contribution to diversity. We generate a random populations of individuals, apply crossover/mutation operators. For each individual, the corresponding $\mu_c(t)$ and θ_c are estimated and the identification error is assumed as the fitness.

Fig. 3 shows a time series with Gaussian distributions in each of five clusters, having eight switch times. The GA converges to an individual in the form of $\{701, 1501, 1999, 2401, 2700, 3502, 4000, 4500, 1, 2, 3, 2, 1, 2, 1, 4, 1\}$. In this example, the size of the search space to find the eight switch times is about $5000^8 \simeq 4 \times 10^{29}$. Thus, the brute force search algorithms are not efficient here. Additionally, solving this problem with methods based on convex optimizations [4] requires solving a linear programming with size about 50000 which shows memory error in our PC. In contrast, the proposed approach based on the GA finds a very good solution in a reasonable time (about 10 minutes).

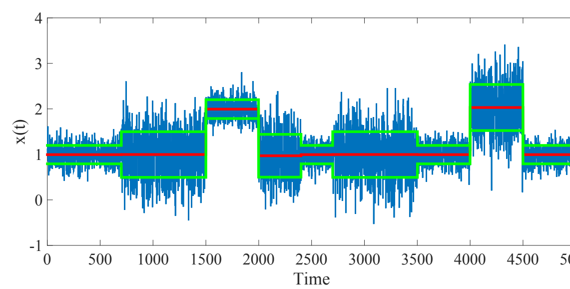


Figure 3: The time series $x(t)$ with five clusters and eight switches. The *means* and *means±standard deviations* detected by algorithm are shown in red and green lines.

3. REFERENCES

[1] M. Gorji Sefidmazgi, M. Moradi Kordmahalleh, A. Homaifar, and A. Karimoddini. A finite element based method for identification of switched linear systems. In *American Control Conference (ACC)*, pages 2644–2649, 2014.

[2] M. Gorji Sefidmazgi, M. Moradi Kordmahalleh, A. Homaifar, and S. Liess. Change detection in linear trend of temperature over US 1900-2012. In *Fourth International Workshop on Climate Informatics*, 2014.

[3] M. Lozano, F. Herrera, and J. R. Cano. Replacement strategies to preserve useful diversity in steady-state genetic algorithms. *Information Sciences*, 178(23):4421 – 4433, 2008.

[4] P. Metzner, L. Putzig, and I. Horenko. Analysis of persistent nonstationary time series and applications. *Communications in Applied Mathematics and Computational Science*, 7(2):175–229, 2012.

[5] M. Moradi Kordmahalleh, M. Gorji Sefidmazgi, A. Homaifar, D. B. KC, and A. Guiseppi-Elie. Time-series forecasting with evolvable partially connected artificial neural network. In *Proceedings of the 2014 Conference Companion on Genetic and Evolutionary Computation Companion*, GECCO Comp '14, pages 79–80. ACM, 2014.