A Constructive Competitive Regression Method for Analysis and Modeling of Non-Stationary Time Series

Slobodan Vucetic and Zoran Obradovic
School of Electrical Engineering and Computer Science
Washington State University, Pullman, WA 99164-2752, USA

Abstract. A simple constructive learning algorithm is proposed for regime discovery in non-stationary time series. The approach is based on competition among regression models aided by averaging of their squared residuals over neighboring data points and incremental introduction of additional models when needed. A case study on California's deregulated power market discovered that 4 regimes existed in market clearing price for April 1998 – June 1999 period, which could be helpful information for better understanding of the market. Discovered local regression models outperformed a global modeling approach indicating that the proposed method can be used in practice for forecasting applications.

1. Introduction

Time series modeling techniques typically assume stationary signal with fixed data distribution [1]. However, real-life time series are often non-stationary with data distribution changing over time. Sometimes non-stationary series can be converted to a semi-stationary series through an appropriate transformation [2], but this is not always possible. Piece-wise stationary time series consisting of time intervals that could be considered stationary are common in practice. Union of all intervals of a piece-wise stationary time series with approximately the same stationary distribution is called a regime. In general, the number of regimes and their switching dynamics is not known in advance. Fitting a global regression model to such data would fail to capture the underlying relationships and would result in learning an averaged behavior of its regimes [3]. Therefore, an analysis of piece-wise stationary time series should consist of two objectives: (1) discovering existing regimes and their positions within a time series and (2) analyzing each of the regimes separately. One of the promising solutions are gated experts [3] that partition domain into a number of subsets and learn separate models on each of them. The drawback of gated experts is that partitioning is based solely on input variables, which depends on an implicit assumption that domains for regimes do not overlap and that switching between regimes depends on input rather than on hidden variables.

Here, we propose an algorithm for regimes discovery in a time-series that is based on a competition among separate regression models for the parts of a given time series. The proposed regime discovery algorithm starts from learning a global regression model on whole time series and continues by adding new regression models to competition until each model becomes an expert for one of the existing regimes. Using moving average filtering of the competing models, the algorithm is able to discover regimes longer then half the filter length.

As an illustration, the algorithm is employed for an analysis of the electricity price time series corresponding to California's recently deregulated power market [4].

2. Method

For a non-stationary time series $\{y_t\}$ consisting of a number of regimes, data generating process (DGP) for regime R_i can be represented as $y_t = E_i[y_t|\mathbf{x}_t] + \varepsilon_t$, $t\in R_i$, where $\mathbf{x}_t$ is a vector consisting of lags of y_t, exogenous variables, or their combination, and ε_t is error representing the unexplained variance.

The proposed algorithm is based on the competition between regression models for disjoint intervals of the time series and the following notation is used for its explanation. A data set representing a time series of length T is defined as $S_0 = \{(\mathbf{x}_t, y_t), t\in[1, T]\}$. If time interval [1, T], is partitioned into L disjoint subsets T_i, i=1, …L, we denote the corresponding data subsets as S_i, i=1, …L, where $S_i = \{(\mathbf{x}_t, y_t), t\in T_i\}$, and the whole partition as $\mathbb{S}_L = \{S_1, \ldots S_L\}$. If separate regression models, $y_i(\mathbf{x}, \beta)$, i=1, …L, are built on each of the subsets S_i, i=1, …L, we define the *composite error* as mean squared error (MSE) of composite prediction, $\hat{y}_t$, where $\hat{y}_t = y_i(\mathbf{x}_t, \boldsymbol{\beta})$, if $t\in T_i$.

Before applying the regime discovery algorithm three choices have to be made. The first is the choice of input variables, $\mathbf{x}$, and the structure of regression models used for the regime discovery, $y(\mathbf{x}; \beta)$,where β is the vector of parameters. This choice should allow fair approximation of $E_i[y_t|\mathbf{x}_t]$, and any prior knowledge about DGP would be useful for this. The second, discussed more in Section 3, is the size of a moving average filter used for averaging the squared residuals of employed regression models defined as

$fe_t = \frac{1}{2M+1}\sum_{i=-M}^{M} e_{t+i}^{2}$, where $\{e_t\}$ and $\{fe_t\}$ are the time series of regression errors and filtered squared errors, and 2M+1 is the size of moving average filter. The third choice is parameter α controlling the algorithm stopping criterion, where $0<\alpha<1$.

The competitive regression scheme starts from a global regression model and, if needed, adds new models in a tree-like manner until the improvement in composite error is smaller than the predetermined threshold. It consists of the following steps:

- Train a single regression model on global data set S_0 and calculate its MSE, mse_1.
- Initially split S_0 into $\mathbb{S}_2^0 = \{S_{1,2}^0, S_{2,2}^0\}$, where $S_{1,2}^0$ and $S_{2,2}^0$ are disjoint and of equal size.
- Modify $\mathbb{S}_2^0$ through the Competition procedure, and calculate composite error mse_2 achieved by the obtained partition $\mathbb{S}_2 = \{S_{1,2}, S_{2,2}\}$.
- Terminate the algorithm if ratio mse_2/mse_1 is larger than α. No regimes are discovered.
- L=2.

repeat

  for i = 1 to L

   - Split $S_{i,L}$ into two disjoint subsets of the same sizes, $S_{i,L}'$ and $S_{i,L}''$, to obtain the initial partition $\mathbb{S}_{L+1,i}^0 = \{ \mathbb{S}_L/S_{i,L}, S_{i,L}', S_{i,L}''\}$.
   - Modify $\mathbb{S}_{L+1,i}^0$ through the Competition procedure and calculate composite error $mse_{L+1,i}$ achieved by the obtained partition $\mathbb{S}_{L+1,i} = \{S_{1,L+1,i}, \ldots S_{L+1,L+1,i}\}$.

  end

  - Out of L partitions, $\mathbb{S}_{L+1,i}$, i = 1, …L, choose the partition achieving the smallest composite error, $mse_{L+1,i}$, to represent the new partition, $\mathbb{S}_{L+1} = \{S_{1,L+1}, \ldots S_{L+1,L+1}\}$, and save its composite error as mse_{L+1}.
  - $L \leftarrow L + 1$

until the ratio mse_L/mse_{L-1} is smaller than α

Output: L−1 'regimes' defined by partition $\mathbb{S}_{L-1}$ and their corresponding models.

Starting from the initial partition $\mathbb{S}_L^0$ a collection of L regression models is used to obtain a modified partition to L subsets with minimized composite error. This is achieved through the following Competition procedure:

---

repeat

  - Train L models on L subsets, S_i, i=1, …L, of the partition $\mathbb{S}_L^n$.
  - Filter squared errors of all models on whole time series with moving average filter of size 2M+1.
  - If the regression function $y_i(\mathbf{x}, \beta)$ gives the smallest filtered error on $(\mathbf{x}_t, y_t)$, assign this point to subset S_i. Form a new partition, $\mathbb{S}_L^{n+1}$, by reassigning all data in such a way.
  - $n \leftarrow n+1$

until less than p% of data points change subsets.

It is expected that each of the L−1 discovered 'regimes' corresponds to the exactly one of the K true regimes existing in the time-series. Important aspects of the proposed procedure proving its capabilities of converging to the regimes are discussed in Section 3.

3. Analysis for two regimes

In this section we provide an insight into the performance of the competition procedure by analyzing its behavior for two regimes (K=2), assuming that errors of ideal models for regimes R_1 and R_2 are normally distributed with mean zero and variances σ_1^2 and σ_2^2, respectively. First, the algorithm is analyzed without the error filtering, and later it is shown that error filtering can significantly improve its performance.

For two models (L=2), determining the regression function with minimum squared error on a data point $(\mathbf{x}_t, y_t)$ is equivalent to determining the sign of $\Delta e_t^2 = e_{t,2}^2 - e_{t,1}^2$, where $\{e_{t,1}\}$ and $\{e_{t,2}\}$ are the errors on the whole time series of the two competing regression models, $y_1(\mathbf{x}, \beta)$ and $y_2(\mathbf{x}, \beta)$, built on disjoint subsets S_1 and S_2, respectively. A positive sign of Δe_t^2 means assigning a data point to subset S_1 and negative to subset S_2. The following proposition gives the expression for a conditional distribution of Δe_t^2 on the two regimes, R_1 and R_2. Here, it is assumed, without loss of generality, that $mse_{1,1}<mse_{2,1}$ and $mse_{2,2}<mse_{1,2}$, where $mse_{i,j}$ is defined as MSE of $y_i(\mathbf{x}, \beta)$ on regime R_j and can be expressed as

$$mse_{i,j} = E_{D_j}[(y - E_j(y|\mathbf{x}))^2] + E_{D_j}[(y_i(\mathbf{x},\beta) - E_j(y|\mathbf{x}))^2]$$

with D_j denoting the domain for regime R_j, i, j =1, 2.

Proposition 1.$\{\Delta e_t^2 \,|\, \mathbf{x}_t, t\in R_1\} \sim \mathcal{N}(A(\mathbf{x}_t)+\delta_1,\ 4A(\mathbf{x}_t)\sigma_1^2)$, where δ_1 is a small positive number, and $A(\mathbf{x}_t) = (y_1(\mathbf{x}_t, \beta) - y_2(\mathbf{x}_t, \beta))^2$. Similarly, $\{\Delta e_t^2 \,|\, \mathbf{x}_t, t\in R_2\} \sim \mathcal{N}(-A(\mathbf{x}_t)-\delta_2,\ 4A(\mathbf{x}_t)\sigma_2^2)$, where δ_2 is a small positive number.

Proof. Omitted for lack of space.

The following proposition proves the convergence of the Competition procedure, where $P_{i,j}(n)$ denotes the fraction of data points from regime R_j in subset S_i at iteration n of the Competition procedure.

Proposition 2. If $P_{j,j}(n_0+1) > P_{j,j}(n_0)$ for some n_0, then $P_{j,j}(n)$ is a strictly increasing function for $n>n_0$.

Proof sketch. $E_{D_j}[(y_j(\mathbf{x},\beta) - E_j(y|\mathbf{x}))^2]$ decreases from n_0th to (n_0+1)th iteration, which in turn leads to increase of $A(\mathbf{x}_t)$. Consequently, the probabilities $P\{\Delta e_t^2 > 0 \,|\, \mathbf{x}_t, t\in R_1\}$ and $P\{\Delta e_t^2 < 0 \,|\, \mathbf{x}_t, t\in R_2\}$ increase, leading to the increase of $P_{j,j}$ for the following iterations.

However, $P_{j,j}(n)$ does not reach 1 in the limit (when n→∞), since $P\{\Delta e_t^2 > 0 \mid \mathbf{x}_t,\ t \in R_1\} < 1$. The bound on $P_{j,j}(n)$ for n→∞ depends on the variances σ_1^2 and σ_2^2, and the difference between DGPs of regimes R_1 and R_2, as shown by the following proposition.

Proposition 3. When n→∞, the probability of correctly assigning data points is upper bounded as

$$P_{j,j}(n) < \int_{D_j} P\{Z(\mathbf{x}_t) > 0\} p_j(\mathbf{x}_t) d\mathbf{x}_t \text{ , where}$$

$Z(\mathbf{x}_t) \sim \mathcal{N}((E_1[y_t \mid \mathbf{x}_t] - E_2[y_t \mid \mathbf{x}_t])^2,\ 4\sigma_j^2 (E_1[y_t \mid \mathbf{x}_t] - E_2[y_t \mid \mathbf{x}_t])^2)$.

Proof sketch. Assuming a perfect separation at some iteration, $y_1(\mathbf{x}, \beta) = E_1[y \mid \mathbf{x}]$ and $y_2(\mathbf{x}, \beta) = E_2[y \mid \mathbf{x}]$, the probability of a correct classification, $P_{j,j}$, at the next iteration depends on distribution given by Proposition 1.

A moving average filtering of squared errors, $\{e_{t,1}^2\}$ and $\{e_{t,2}^2\}$ is equivalent to moving average filtering of $\{\Delta e_t^2\}$ leading to $\{\Delta fe_t^2\}$. It improves the Competition procedure as given by following proposition and its corollary.

Proposition 4. If the error terms $\{e_{t,1}^2\}$ and $\{e_{t,2}^2\}$ do not have serial correlation, and the expected length between regime switching, N, is much larger than the moving average filter length, 2M+1, then

$$E[\Delta fe_t^2 \mid \mathbf{x}_t, t \in R_j] = A + \delta_j \text{ , and}$$

$$Var[\Delta fe_t^2 \mid \mathbf{x}_t, t \in R_j] = 4A\sigma_j^2 / (2M+1) \text{ , j=1,2.}$$

Proof. A consequence of Proposition 1.

Corollary. Error filtering significantly improves the separation between regimes, and speeds up convergence of the Competition procedure by increasing the upper bound on $P_{j,j}(n)$, n→∞.

Proof. Direct consequence of decreased conditional variance and Proposition 3.

4. Experiments

Choice of error filter length. To illustrate the influence of different regime lengths, N, and filter sizes, 2M+1, on the separation between regimes, a Monte Carlo experiment has been performed on time series with two regimes. Three cases with regime lengths of N=10, 100, 1000 were examined, with M ranging from 0 to 200, A=1, and σ_1^2=4. Results shown in Figure 1 demonstrate that error filtering can be successfully used for discrimination between regimes. It could be seen that for sufficiently long regimes (N=100, 1000) a wide range of moving average filter sizes significantly improves classification as compared to non-filtering case (M=0). However, the length of the filter larger then twice the regime length, (2M+1>2N) diminishes the ability of the algorithm to discriminate between the regimes. In such a case the algorithm would terminate after the initial step, reporting that no regimes have been found.

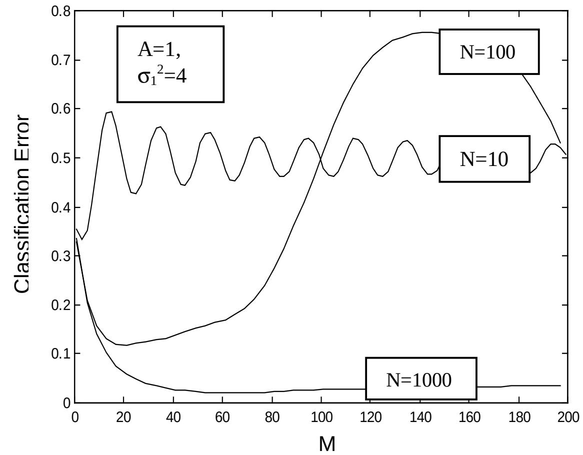

Figure 1. *Influence of filter size 2M+1 for different regime switching dynamics (N=10, 100, 1000)*

Analysis of California electricity market. To illustrate the performance of the proposed algorithm, 10,224 hour-by-hour data points in the period April 1, 1998 - May 31, 1999, on day-ahead unconstrained market clearing prices $\{p_t\}$ of California's deregulated power market, and day-ahead forecasted load for California known at the time of bidding $\{l_t\}$ shown in Figure 2, have been used to analyze the behavior of the market. The relationship between price and generated power for thermal power generators is usually expressed as a quadratic function, and so we have decided to model the relationship between $\{p_t\}$ and $\{l_t\}$ as $p_t = y(l_t, \beta) = \beta_0 + \beta_1 l_t + \beta_2 l_t^2 + \beta_3 l_t^3$, and used this simple functional form in our algorithm. We were interested in discovering potential seasonal changes in the load-price relationship, and hence we set the size of moving average filter to one week with 2M=168. Stopping criterion, α, has been set to 0.95.

Competition results starting from 2 and incrementing to 3 and 4 models are shown as partitions $\mathbb{S}_2$, $\mathbb{S}_3$ and $\mathbb{S}_4$ in Figure 3. According to the evolution of composite error (Table 1) the algorithm stops at $\mathbb{S}_5$ discovering 4 regimes given by $\mathbb{S}_4$. Regression functions corresponding to discovered regimes are shown in

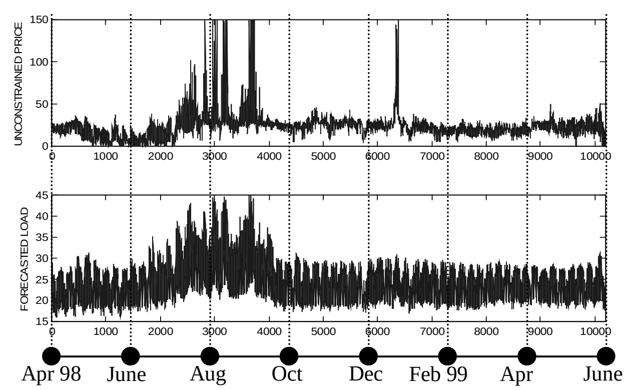

Figure 2. *Hourly unconstrained prices and forecasted load*

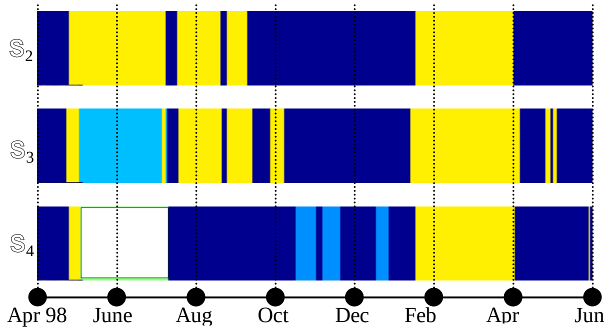

Figure 3. *Partitions $\mathbb{S}_2$, $\mathbb{S}_3$ and $\mathbb{S}_4$*

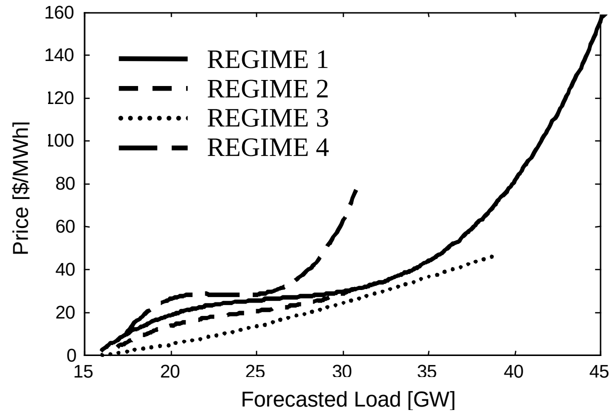

Figure 4. *Regression functions for 4 regimes*

Figure 4, with further analysis provided in Table 2. Regression function 1 was comparable to the general model on its corresponding regime (53% of total data), but functions 2, 3 and 4 performed much better on regimes 2, 3 and 4 as compared to using the global model (which was no better than mean predictor on regime 3 and 4). It could be seen that regression functions 1-4 are significantly different, indicating that identified subsets can be considered as regimes existing in load-price relationship during the 14 observed months. For example, relatively low prices existed during late spring of 1998, which could be explained by generation pressure of hydro-generating units.

Table 1. *Evolution of composite error*

Partition	**Global**	$\mathbb{S}_2$	$\mathbb{S}_3$	$\mathbb{S}_4$	$\mathbb{S}_5$
MSE	91.8	75.2	66.3	61.9	60.4

Table 2. *Comparison of global model and 4 models defined for partition $\mathbb{S}_4$*

Regime	Size	regime MSE	regime R^2	global MSE	global R^2
1	5370	77.9	0.77	116.1	0.75
2	2061	6.8	0.78	13.7	0.56
3	1622	29.1	0.72	80.2	-0.08
4	1146	135.5	0.33	284.9	-0.21

5. Conclusions

In this paper a constructive competitive regression scheme for analyzing and modeling non-stationary time series has been proposed. The analysis of the algorithm for two regimes provided useful insight into its behavior by showing that the Competitive procedure is able to converge to regimes, and that its convergence and final separation can be greatly improved by error filtering of squared errors.

The regime discovery algorithm is not restrictive to the choice of regression model and its complexity, and can range from linear and nonlinear parametric models to neural networks. There is significant difference between these models, and the algorithm in its presented form is more suited to parametric models characterized by simple learning process and estimation of prediction capabilities. Slight modifications might be needed to account for neural networks specifics. Such an effort should be rewarded by the ability to discover non-linear relationships and regimes, and it is the subject of the further research.

The proposed algorithm can be useful for on-line prediction of non-stationary time series by providing an intelligent partition of available data set into subsets with similar behavior. Starting from a partition obtained by the proposed algorithm, representationally more powerful regression functions such as neural networks can be applied to learn relationships separately on each of the regimes. As a result, a library of models can be established and used for on-line non-stationary time series prediction. Extending our previous work [5], this can be achieved by evaluating the library models on data from recent history to identify the most appropriate model for the current prediction.

References

[1] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, *Time Series Analysis: Forecasting and Control*, Prentice-Hall Inc., Englewood Cliffs, 1994.

[2] R. Davidson, J. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993.

[3] A.S.Weigend, M. Mangeas, and A. N. Srivastava, Nonlinear Gated Experts for Time Series: Discovering Regimes and Avoiding Overfitting, *Int. J. of Neural Systems, 6, 373-399*, 1995.

[4] T. De la Torre, J. W. Feltes, T.S.R. Gomez, and H. M. Merrill, Deregulation, Privatization, and Competition: Transmission Planning Under Uncertainty, *IEEE Trans. on Power Systems, 14, 460-165*, 1999.

[5] R. Drossu, and Z. Obradovic, Regime Signaling Techniques for Non-stationary Time Series Forecasting, *J. of Computational Intelligence in Finance, 4, 7-15*, 1996.