Performance Evaluation of Some Change Detection and Data Segmentation Algorithms

THEODOR D. POPESCU
National Institute for R&D in Informatics
Research Department
8-10 Averescu Avenue, 011455 Bucharest
ROMANIA
pope@ici.ro

DANIELA DOINA CIOBOATA
National Institute for R&D in Mechatronics
and Measurement Technique
6-8 Pantelimon Road, 02163147 Bucharest
ROMANIA
cioboatadoina@yahoo.com

Abstract: The problem of change detection and data segmentation has received considerable attention during the last two decade in a research context and appears to be the central issue in various application areas. The following techniques are investigated in the paper for their performance evaluation: filtering techniques with a whiteness test, techniques based on sliding windows and distance measures and maximum likelihood techniques for data segmentation. The used model will be the simplest extension of linear regression models to data with abruptly changing properties, or piecewise linearizations of non-linear models. Finally, some Monte-Carlo simulations for change detection and data segmentation are presented, to evaluate the performance of these algorithms in a number of cases.

Key–Words: Change detection, segmentation, filtering, maximum likelihood, distance measures, Monte-Carlo simulation.

1 Introduction

The problem of change detection and data segmentation has gained considerable attention during the last two decades in a research context and appears to be the central issue in various application areas. In this framework, the problem of segmentation between "homogenous" parts of the data (or detection of changes in the data) arises more or less explicitly.

The proposed problem formulation assumes the off-line or batch-wise data processing, although the solution is sequential in data and an on-line data processing can be used. The following techniques are investigated in the paper: filtering techniques with a whiteness test, techniques based on sliding windows and distance measures, and maximum likelihood techniques for data segmentation.

The presented approaches have been evaluated by Monte-Carlo simulation for change detection in the mean of a signal, change detection in the parameters of an autoregressive model with exogenous variable (ARX) model and for analysis of the robustness properties of the algorithms. Based on the obtained results it can be noted that the performances of the maximum a posteriori probability (MAP) method are superior to the other investigated approaches.

2 Problem Formulation

We introduce now the general change detection and segmentation problem for linear regression model with piecewise constant parameters. The goal is to find a sequence of time indices $k^n = k_1, k_2, \ldots, k_n$, where both the number n and the locations k_i are unknown, such that a linear regression model with piecewise constant parameters,

$$y_t = \phi_t^T \theta(i) + e_t, \;\; E(e_t^2) = \lambda(i) R_t \qquad (1)$$

when $k_{i-1} < t \leq k_i$ is a good description of the observed signal y_t. Here $\theta(i)$ is the d-dimensional parameter vector in segment i, ϕ_t is the regressor and k_i denotes the change times. The noise e_t is assumed to be Gaussian with variance $\lambda(i)R_t$, where $\lambda(i)$ is a possibly segment dependent scaling of the noise and R_t is the nominal covariance matrix of the noise. We can think of λ either as a scaling of the noise variance or variance itself (R_t = 1). Neither $\theta(i)$ or $\lambda(i)$ are known. The Gaussian assumption on the noise is a standard one, partly because it gives analytical expressions and partly because it has proven to work well in practice. We will assume R_t to be known and the scaling as a possibly unknown parameter. The model (1) is referred to as changing regression, because it changes between regression models. Its im-

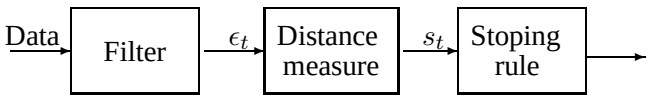

Figure 1: Change detection based on a whiteness test for filter residuals

portant feature is that the jumps divide the measurements into a number of independent segments, since the parameter vectors in different segments are independent.

The assumption on the regression models in (1) is not too restrictive since many stationary processes encountered in practice can be closely approximated by such models. The identification and parameters estimation methods represent only tools to perform change detection and segmentation. Good and precise models offers high performance in these schemes, but also biased parametric models can be used for change detection and segmentation. This bias decreases, but does not annihilate the performance of the detection and segmentation procedures. Some important cases of the model (1) are the changing mean model, the autoregressive (AR) model, the autoregressive model with exogenous variable (ARX) and finite impulse response (FIR) model, etc, where ϕ_t has different expressions.

3 Approaches

Several methods for change detection and data segmentation have been suggested earlier, see e.g. [1], [2], [3], among others. They typically employ multiple detection algorithms, hidden Markov models, explicit management of multiple model, AFMM (adaptive forgetting by multiple models), or formulate the segmentation problem as a least-squares problem with sum-of-norm regularization over the state parameter jumps. We present in the following three class of approaches for change detection and segmentation.

3.1 Change detection based on filtering

One useful approach for change detection consists in filtering of the observed data through a known or identified AR filter, and in looking for changes in the residual signal of innovations, $\{\epsilon_t\}$. Actually, the use of cusum techniques based upon the innovations (one-step prediction errors) ,$\{\epsilon_t\}$, or the squared innovations, $\{\epsilon_t^2\}$, is a standard approach for change detection in AR models. Such a technique, based upon $\{\epsilon_t^2\}$ is based upon the fact that, before the change $E(\epsilon_t^2) = \sigma_1$ and thus: $E(\epsilon_t^2/\sigma_1 - 1) = 0$.

To conclude, statistical whiteness tests can be used to test if the residuals are white noise as they should be if there is no change. Fig. 1 shows the basic structure, where the filter residuals are transformed to a *distance measure*, that measures the deviation from the no-change hypothesis. The *stopping rule* decides whether the deviation is significant or not. The most natural distances are the following, [1], change in the mean $s_t = \epsilon_t$, change in variance $s_t = \epsilon_t^2 - \lambda$, with λ known, change in correlation $s_t = \epsilon_t y_{t-k}$ or $s_t = \epsilon_t u_{t-k}$ for some k, and change in sign correlation, $s_t = \text{sign}(\epsilon_t \epsilon_{t-1})$.

The main problem in statistical change detection is to decide what "large" are these distances. Many change detection algorithms can be recast into the problem of deciding on the following two hypotheses:

$$H_0 : E(s_t) = 0, \qquad H_1 : E(s_t) > 0$$

where s_t is a *distance measure*. A stopping rule is essentially achieved by low-pass filtering s_t and comparing this value to a threshold. Below, two such low-pass filters are given:

- The CUmulative SUM (CUSUM) test of Page:

  $$g_t = \max(g_{t-1} + s_t - \nu, 0), \quad \text{change if} \quad g_t > h$$

  The *drift parameter* ν influences the low-pass effect, and the *threshold* h (and also ν) influences the performance of the detector.

- The Geometric Moving Average (GMA) test:

  $$g_t = \lambda g_{t-1} + (1 - \lambda) s_t, \qquad \text{change if} \quad g_t > h.$$

  Here, the forgetting factor λ is used to tune the low-pass effect, and the threshold h is used to tune the performance of the detector. Using no forgetting at all (λ =0), corresponds to thresholding directly, which is one option.

It seems that classical approach consisting in testing how much the sequence of innovations , $\{\epsilon_t\}$ is far from hypothesis "zero-mean white noise" is not sufficient for change detection in practice.

3.2 Change detection based on sliding windows and distance measures

The main idea underlying this approach consists in comparison of two models: a model (M_2), based on data from a sliding window of size L ($y_{t-L+1}, \ldots, y_t$) is compared to a model (M_1) based on all data or a substantially larger sliding window ($y_1, y_2, \ldots, y_t$), [2]:

$$\text{Data} : \underbrace{y_1, y_2, \ldots, y_{t-L+1}, \ldots, y_t}_{M_1}$$
$$\text{Data} : y_1, y_2, \ldots, \underbrace{y_{t-L+1}, \ldots, y_t}_{M_2}$$

If the model based on the larger data window gives larger residuals

$$\|\epsilon_t^1\| > \|\epsilon_t^2\| \tag{2}$$

then a change is detected. The problem here is to choose a norm that corresponds to a relevant statistical measure. Some norms that have been proposed are:

- The Generalized Likelihood Ratio (GLR).
- The divergence test.
- Change in spectral distance. There are many methods to measure the distance between two spectra. One approach would be to compare the spectral distance of two models.

These criteria provide an s_t to be put into a stopping rule for instance, the CUSUM test. The choice of window size L is very critical here. On the one hand, a large value is need to get an accurate model in the sliding window and, on the other hand, a small value is needed to get quick detection.

Concerning the distance functions presented above, we will give in the following their expressions.

In Basseville and Benveniste, [2], two different test statistics for the case of two different models are given. A straightforward extension of the generalized likelihood ratio test leads to:

$$d_{GLR} = L \log \frac{\sigma_1}{\sigma_2} + \frac{(y_t - \phi_t^T \theta_1)^2}{\sigma_1} - \frac{(y_t - \phi_t^T \theta_2)^2}{\sigma_2} \tag{3}$$

This test statistic will be referred as Brandt's GLR test.

To measure the distance between two models, any norm can be used. So, the Kullback discrimination information, [4], between two probability density functions p_1 and p_2 is defined as:

$$I(1,2) = \int p_1(x) \log \frac{p_1(x)}{p_2(x)} dx \geq 0 \tag{4}$$

In the special case of Gaussian distribution

$$p_i(x) = N(\hat{\theta}_i, P_i)$$

we get

$$I(1,2) = \frac{1}{2}\text{tr}(P_2^{-1}P_1 - I) + \frac{1}{2}(\hat{\theta}_1 - \hat{\theta}_2)^T P_2^{-1}(\hat{\theta}_1 - \hat{\theta}_2) -$$
$$-\frac{1}{2}\log\left(\frac{\det P_1}{\det P_2}\right) \tag{5}$$

The Kullback information is not a norm (it is not symmetric) and is not suitable as a distance measure. Instead, Kullback divergence is used:

$$V(1,2) = I(1,2) + I(2,1) \geq 0 \tag{6}$$

From Kullback divergence, the divergence test can be derived and it equals:

$$\begin{aligned} d_{DIV} &= L\left(\frac{\sigma_1}{\sigma_2} - 1\right) + \left(1 + \frac{\sigma_1}{\sigma_2}\right)\frac{(y_t - \phi_t^T\theta_1)^2}{\sigma_1} - \\ &- 2\frac{(y_t - \phi_t^T\theta_1)(y_t - \phi_t^T\theta_2)}{\sigma_2} \end{aligned} \tag{7}$$

The corresponding algorithm will be called the divergence test. d_{GLR} and d_{DIV} start to grow when a jump has occurred, and again the task of the stopping rule is to decide whether the growth is significant.

3.3 Maximum a posteriori probability estimator

We use in the following the general segmentation problem for linear regression model with piecewise constant parameters. As we mentioned before, in segmentation the goal is to find a sequence of time indices $k^n = k_1, k_2, \ldots, k_n$, where both the number n and the locations k_i are unknown, such that a linear regression model with piecewise constant parameters (1).

One way to guarantee that the best possible solution is found is to consider all possible segmentation k^n, estimate one linear regression model in each segment, and then choose the particular k^n that minimizes an optimality criteria:

$$\widehat{k^n} = \arg \min_{n \geq 1, 0 < k_1 < \ldots < k_n = N} V(k^n) \tag{8}$$

For the measurements in the ith segment, that is $y_{k_{i-1}+1}, \ldots y_{k_i} = y_{k_{i-1}+1}^{k_i}$, the least square estimate and its covariance matrix are denoted:

$$\hat{\theta}(i) = P(i) \sum_{t=k_{i-1}+1}^{k_i} \phi_t R_t^{-1} y_t, \tag{9}$$

$$P(i) = \left(\sum_{t=k_{i-1}+1}^{k_i} \phi_t R_t^{-1} \phi_t^T\right)^{-1}. \tag{10}$$

The following quantities, V - the sum of squared residuals, D - $-\log\det$ of the covariance matrix P and N - the number of data in each segment, are given by

$$V(i) = \sum_{t=k_{i-1}+1}^{k_i} (y_t - \phi_t^T \hat{\theta}(i))^T R_t^{-1} (y_t - \phi_t^T \hat{\theta}(i)) \quad (11)$$

$$D(i) = -\log\det P(i) \quad (12)$$

$$N(i) = k_i - k_{i-1} \quad (13)$$

and represent sufficient statistics in each segment. The data and quantities used in segmentation procedure are shown in Table 1.

Table 1: Data and quantities used in MAP procedure

Data	$y_1, y_2, \ldots, y_{k_1}$	...	$y_{k_{n-1}+1}, \ldots, y_{k_n}$
Segm.	Segment 1	...	Segment n
LS est.	$\hat{\theta}(1), P(1)$	...	$\hat{\theta}(n), P(n)$
Stat.	$V(1), D(1), N(1)$	...	$V(n), D(n), N(n)$

Note that the segmentation k^n has $n-1$ degrees of freedom. Two types of optimality criteria have been mainly proposed in this field: statistical criteria (Maximum Likelihood (ML) or Maximum A posteriori Probability estimate (MAP)) and information based criteria. The main problem in segmentation is the dimensionality. The number of segmentations k^n is 2^N (can be a change or no change at each time instant).

We give now the conceptual description of the Maximum a posteriori probability (MAP) estimator, [1], for the data and quantities given in Table 1:

1. Examine every possible segmentation, parameterized in the number of jumps n and jump times k^n, separately.

2. For each segmentation, compute the best models in each segment parameterized in the least square estimates $\hat{\theta}(i)$ and their covariance matrices $P(i)$.

3. Compute in each segment the sum of squared prediction errors $V(i)$ and $D(i) = -\log\det P(i)$.

4. The MAP estimate of the model structure for the three different assumptions on noise scaling ((i) known $\lambda(i) = \lambda_0$, (ii) unknown but constant $\lambda(i) = \lambda$ and (iii) unknown and changing $\lambda(i)$ is given by the following equations, [1]:

$$\widehat{k^n} = \arg\min_{k^n,n} \sum_{i=1}^{n} (D(i) + V(i)) + 2n \log\frac{1-q}{q}$$

$$\widehat{k^n} = \arg\min_{k^n,n} \sum_{i=1}^{n} D(i) + (Np - nd - 2) \times \times \log\sum_{i=1}^{n} \frac{V(i)}{Np - nd - 4} + 2n\log\frac{1-q}{q}$$

$$\widehat{k^n} = \arg\min_{k^n,n} \sum_{i=1}^{n} (D(i) + (N(i)p - d - 2) \times \times \log\frac{V(i)}{N(i)p - d - 4}) + 2n\log\frac{1-q}{q}$$

respectively. The user can assume a fixed probability q of jump at each new time instant.

Computing the exact likelihood is computationally intractable because of the exponential complexity. Some algorithms implementing recursive local search techniques and numerical searches based on dynamic programming or MCMC (Markov Chain Monte Carlo) techniques are given in [1].

4 Simulation Results

We present in the following some simulation results obtained by Monte Carlo simulation for different signal models including only models with changes in the mean, ARX model parameters, and robustness evaluation when the following techniques have been investigated: filtering techniques with a whiteness test, using the Cumulative Sum (CUSUM), and Geometric Moving Average (GMA), techniques based on sliding windows and distance measures, using Generalized Likelihood Ratio (GLR) and Divergence Test (DIV), and MAP estimator with unknown and constant noise scaling, when MCMC numerical search procedure has been used.

4.1 Change detection in the mean of a signal

The results are obtained by Monte-Carlo simulation, for 1000 noise realizations, in the case of a piecewise constant mean model, having the following structure:

$$y_t = \theta(i) + e_t \quad (14)$$

where e_t is a random sequence of zero mean and variance $E(e_t^2) = \lambda\sigma_t^2$, with σ_t^2 = 1, for each experiment. The model parameters are given in Table 2. The change detection results for different level of the noise, λ, are given in Table 3 for filtering approach, in Table 4 for sliding windows approach and in Table 5,

for MAP approach with unknown and constant noise scaling and MCMC algorithm (with a value of jump probability, q = 0.3), respectively; the results represent the number of the real change instants detected (0 delay in detection). The problem of change detection in the mean of a signal is used in "local approach", [2], where the general detection problem is reduced to a classical problem of monitoring the mean of a Gaussian vector variable.

Table 2: Mean model parameters

i	1	2
Segment i	1-100	101-200
$\theta(i)$	1.0	2.0

Table 3: Change detection in mean: Filtering

Exp.	λ	GMA	CUSUM
1	0.01	806	976
2	0.04	615	685
3	0.09	603	603
4	0.25	555	560

Table 4: Change detection in mean: Sliding windows

Exp.	λ	Brand GLR	DIV
1	0.01	986	918
2	0.04	902	764
3	0.09	787	674
4	0.25	333	337

Table 5: Change detection in mean: MAP

Exp.	λ	MAP
1	0.01	1000
2	0.04	990
3	0.09	878
4	0.25	637

4.2 Change detection in parameters of an ARX model

.

The ARX model used has the following structure:

$$\begin{aligned} y_t &= -a_1(i) * y_{t-1} - a_2(i) * y_{t-2} + \\ &+ b_1(i) * u_{t-1} + b_2(i) * u_{t-2} + e_t \end{aligned} \quad (15)$$

where e_t is a random sequence of zero mean and variance $E(e_t^2) = \lambda\sigma_t^2$, with σ_t^2 = 1, for each experiment.

Table 6: ARX(2,2,1) model parameters

i	1	2	3
Segment i	1-100	101-200	201-300
$a_1(i)$	1.5	1.5	1.5
$a_2(i)$	0.8	0.8	0.6
$b_1(i)$	2.	2.	2.
$b_2(i)$	0.5	1.	1.

The model parameters are given in Table 6; the input u_t was a random signal of mean zero and variance 1.

As in the previous case, the experiments were performed, for constant value of λ in each experiment. The change detection results for different level of the noise, λ, are given in Table 7, for filtering approach, in Table 8 for sliding windows approach and in Table 9, for MAP approach with unknown and constant noise scaling and MCMC algorithm (with a value of jump probability, q = 0.3), respectively; the results are given under the the form of the number of the real change instants detected (0 delay in detection).

Table 7: Change detection in ARX(2,2,1): Filtering

Exp.	λ	GMA		CUSUM	
		1st ch.	2nd ch.	1st ch.	2nd ch.
1	0.01	313	398	585	719
2	0.04	247	292	298	385
3	0.09	176	222	189	248
4	0.25	164	168	180	169

Table 8: Change detection in ARX(2,2,1): Sliding windows

Exp.	λ	Brand GLR		DIV	
		1st ch.	2nd ch.	1st ch.	2nd ch.
1	0.01	158	199	195	184
2	0.04	93	136	118	176
3	0.09	42	111	66	119
4	0.25	18	65	26	57

Table 9: Change detection in ARX(2,2,1): MAP

Exp.	λ	MAP	
		1st ch.	2nd ch.
1	0.01	729	834
2	0.04	512	638
3	0.09	327	490
4	0.25	173	227

4.3 Robustness evaluation

The robustness of the algorithms to the model structure, has been tested, also by Monte Carlo simulation,

for an underestimated model order ARX (1,1,1) and for an overestimated model order ARX(3,3,1). It resulted that the performances are affected, when the model order is underestimated, for all methods, especially when the noise level increases. If the model order is overestimated, the performances are not affected to a great extend for MAP segmentation procedure. The results are given in Table 10, Table 11 and Table 12, for filtering, sliding windows and MAP approach with unknown and constant noise scaling and MCMC algorithm (q = 0.3), respectively. As in the previous cases, the results are given under the form of the number of the real change instants detected (0 delay in detection). So, when no information on the model order is available, the best solution is to use an high-order model to perform signal segmentation.

Table 10: Change detection with ARX(3,3,1): Filtering

Exp.	λ	GMA		CUSUM	
		1st ch.	2nd ch.	1st ch.	2nd ch.
1	0.01	321	400	559	725
2	0.04	227	253	286	357
3	0.09	152	173	182	214
4	0.25	116	142	138	170

Table 11: Change detection with ARX(3,3,1): Sliding windows

Exp.	λ	Brand GLR		DIV	
		1st ch.	2nd ch.	1st ch.	2nd ch.
1	0.01	71	146	100	132
2	0.04	48	108	678	138
3	0.09	23	111	40	100
4	0.25	11	56	14	64

Table 12: Change detection with ARX(3,3,1): MAP

Exp.	λ	MAP	
		1st ch.	2nd ch.
1	0.01	739	812
2	0.04	484	606
3	0.09	275	422
4	0.25	132	214

5 Conclusions

The paper gives the conceptual description of some change detection and data segmentation algorithms based on filtering, sliding windows and likelihood techniques and evaluates their performances by Monte-Carlo simulation. Based on the obtained results it can be noted that the performances of the MAP segmentation approach are superior to the other approaches investigated, but with the price of the computation effort. The performances of the filtering, sliding windows and distance measures approaches depend to a great extend of the choosing of the design parameters ν and h. The single design parameter in MAP segmentation is q, the probability that the system jumps at each sample. Other design parameters (see [6]) could be chosen without problems. Some practical applications of the presented approaches in seismic signal processing and vibration signal analysis are given in [6] and [5], [7], respectively.

Acknowledgements: The research has been supported by Executive Agency for Higher Education, Research, Development and Innovation Funding (UEFISCDI), Bucharest, Romania, under Contract PN-II-PT-PCCA-2013-4-1671, Grant 268-2014.

References:

[1] F. Gustafsson, *Adaptive Filtering and Change Detection*,Willey, NJ, 2001.

[2] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*, Information and system science series, Prentice Hall, Englewood Cliffs, NJ. 1993.

[3] H. Ohlson, L. Ljung, S. Boyd, Segmentation ARX-models using sum-of-norms regularization, *Automatica*, vol. 46 pp. 1107-1111, 2010.

[4] K. S. Kumamaru, K. S. Sagara and T. Söderstrom, *Some statistical methods for fault diagnosis in dynamical systems*, in R. Patton, P. Frank and R. Clark, Editors, Fault diagnosis in dynamic systems - Theory and applications, pp. 439–476, Prentice Hall International, London, UK, 1989.

[5] Th. D. Popescu, Blind separation of vibration signals and source change detection - Application to machine monitoring, *Applied Mathematical Modelling*, vol. 34, pp. 3408-3421, 2010.

[6] Th. D. Popescu, Signal segmentation using changing regression models with application in seismic engineering, *Digital Signal Processing* , vol. 24, pp. 14-26, 2014.

[7] Th. D. Popescu, B. Dumirascu, An application of Rényi entropy segmentation in fault detection of rotating machinery, *Proc. of the 16th International Conference on Research and Education in Mechatronics (REM2015)*, R. Biesenbach, A. Weinert (Eds.), Bochum, Germany, November 18-20, pp. 228-295, 2015.