HAL
open science

A regression model with a hidden logistic process for signal parametrization

Faicel Chamroukhi, Allou Samé, Gérard Govaert, Patrice Aknin

To cite this version:

Faicel Chamroukhi, Allou Samé, Gérard Govaert, Patrice Aknin. A regression model with a hidden logistic process for signal parametrization. 17th European Symposium on Artificial Neural Networks, Advances in Computational Intelligence and Learning, Apr 2009, Bruges, Belgium. pp.1-6. hal-00447806

HAL Id: hal-00447806

https://hal.science/hal-00447806v1

Submitted on 15 Jan 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A regression model with a hidden logistic process for signal parametrization

F. Chamroukhi[1,2], A. Samé[1], G. Govaert[2] and P. Aknin[1]

1- French National Institute for Transport and Safety Research (INRETS)
2 Rue de la Butte Verte, 93166 Noisy-Le-Grand Cedex (France)

2- Compiègne University of Technology
HEUDIASYC Laboratory, UMR CNRS 6599
BP 20529, 60205 Compiègne Cedex (France)

Abstract. A new approach for signal parametrization, which consists of a specific regression model incorporating a discrete hidden logistic process, is proposed. The model parameters are estimated by the maximum likelihood method performed by a dedicated Expectation Maximization (EM) algorithm. The parameters of the hidden logistic process, in the inner loop of the EM algorithm, are estimated using a multi-class Iterative Reweighted Least-Squares (IRLS) algorithm. An experimental study using simulated and real data reveals good performances of the proposed approach.

1 Introduction

In the context of the predictive maintenance of the French railway switches (or points) which enable trains to be guided from one track to another at a railway junction, we have been brought to parameterize switch operations signals representing the electrical power consumed during a point operation (see figure 1). The final objective is to exploit these parameters for the identification of incipient faults.

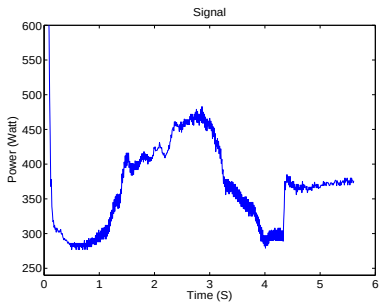

Fig. 1: Example of the electrical power consumed during a point operation

The method we propose to characterize signals is based on a regression model incorporating a discrete hidden process allowing abrupt or smooth switchings between various regression models. This approach has a connection with the switching regression model introduced by Quandt and Ramsey [9] and is very

linked to the Mixture of Experts (ME) model introduced by Jordan and Jacobs [7] by the using of a time-dependant logistic transition function. The ME model, as discussed in [10], uses a conditional mixture modeling where the model parameters are estimated by the Expectation Maximization (EM) algorithm [1][3]. Other alternative approaches are based on Hidden Markov Models in a context of regression [6]. A dedicated EM algorithm including a multi-class Iterative Reweighted Least-Squares (IRLS) algorithm [8] is proposed to estimate the model parameters.

Section 2 introduces the proposed model and section 3 describes the parameters estimation via the EM algorithm. The fourth section is devoted to the experimental study using simulated data and real data.

2 Regression model with a hidden logistic process

2.1 The global regression model

We represent a signal by the random sequence $\mathbf{x} = (x_1, ..., x_n)$ of n real observations, where x_i is observed at time t_i. This sample is assumed to be generated by the following regression model with a discrete hidden logistic process $\mathbf{z} = (z_1, \ldots, z_n)$, where $z_i \in \{1, \ldots, K\}$:

$$x_i = \boldsymbol{\beta}_{z_i}^T \boldsymbol{r}_i + \varepsilon_i \quad ; \quad i = 1, \ldots, n \tag{1}$$

In this model, $\boldsymbol{\beta}_{z_i}$ is the $(p+1)$-dimensional coefficients vector of a p degree polynomial, $\boldsymbol{r}_i = (1, t_i, \ldots, (t_i)^p)^T$ is the time dependant $(p+1)$-dimensional covariate vector associated to the parameter $\boldsymbol{\beta}_{z_i}$ and the ε_i are independent random variables distributed according to a Gaussian distribution with zero mean and variance $\sigma_{z_i}^2$.

2.2 The hidden logistic process

This section defines the probability distribution of the process $\mathbf{z} = (z_1, \ldots, z_n)$ that allows the switching from one regression model to another. The proposed hidden logistic process supposes that the variables z_i, given the vector $\boldsymbol{t} = (t_1, \ldots, t_n)$, are generated independently according to the multinomial distribution $\mathcal{M}(1, \pi_{i1}(\mathbf{w}), \ldots, \pi_{iK}(\mathbf{w}))$, where

$$\pi_{ik}(\mathbf{w}) = p(z_i = k; \mathbf{w}) = \frac{\exp\left(\boldsymbol{w}_k^T \boldsymbol{v}_i\right)}{\sum_{\ell=1}^{K} \exp\left(\boldsymbol{w}_\ell^T \boldsymbol{v}_i\right)} \tag{2}$$

is the logistic transformation of a linear function of the time-dependant covariate $\boldsymbol{v}_i = (1, t_i, \ldots, (t_i)^q)^T$, $\boldsymbol{w}_k = (\boldsymbol{w}_{k0}, \ldots, \boldsymbol{w}_{kq})^T$ is the $(q+1)$-dimensional coefficients vector associated to the covariate $\boldsymbol{v}_i$ and $\mathbf{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K)$. Thus, given the vector $\boldsymbol{t} = (t_1, \ldots, t_n)$, the distribution of $\mathbf{z}$ can be written as:

$$p(\mathbf{z}; \mathbf{w}) = \prod_{i=1}^{n} \prod_{k=1}^{K} \left(\frac{\exp\left(\boldsymbol{w}_k^T \boldsymbol{v}_i\right)}{\sum_{\ell=1}^{K} \exp\left(\boldsymbol{w}_\ell^T \boldsymbol{v}_i\right)} \right)^{z_{ik}} , \tag{3}$$

where $z_{ik} = 1$ if $z_i = k$ i.e when x_i is generated by the kth regression model, and 0 otherwise.

The pertinence of the logistic transformation in terms of flexibility of transition can be illustrated through simple examples with $K = 2$ components. As it can be shown in figure 2 (left), the dimension q of $\boldsymbol{w}_k$ controls the number of changes in the temporal variation of $\pi_{ik}(\mathbf{w})$. More particularly, if the goal is to segment the signals into convex homogenous parts, the dimension q of $\boldsymbol{w}_k$ must be set to 1. The quality of transitions and the change time point are controlled by the components values of the vector $\boldsymbol{w}_k$ (see figures 2 (middle) and (right)).

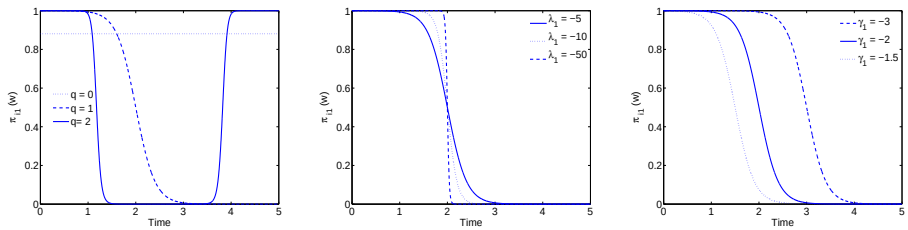

Fig. 2: Variation of $\pi_{i1}(\mathbf{w})$ over time for (left) different values of the dimension q of $\boldsymbol{w}_k$, (middle) different values of $\lambda_k = \boldsymbol{w}_{k1}$ and (right) different values of $\gamma_k = \frac{\boldsymbol{w}_{k0}}{\boldsymbol{w}_{k1}}$ with $q = 1$.

3 Parameter estimation

From the model given by equation (1), it can be proved that the random variable x_i is distributed according to the normal mixture density

$$p(x_i; \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_{ik}(\mathbf{w}) \mathcal{N}\big(x_i; \boldsymbol{\beta}_k^T \boldsymbol{r}_i, \sigma_k^2\big) \ , \tag{4}$$

where $\boldsymbol{\theta} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2)$ is the parameter vector to be estimated. The parameter $\boldsymbol{\theta}$ is estimated by the maximum likelihood method. As in the classic regression models we assume that, given $\boldsymbol{t} = (t_1, \ldots, t_n)$, the ε_i are independent. This also involves the independence of x_i $(i = 1, \ldots, n)$. The log-likelihood of $\boldsymbol{\theta}$ is then written as:

$$L(\boldsymbol{\theta}; \mathbf{x}) = \log \prod_{i=1}^{n} p(x_i; \boldsymbol{\theta}) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \pi_{ik}(\mathbf{w}) \mathcal{N}\big(x_i; \boldsymbol{\beta}_k^T \boldsymbol{r}_i, \sigma_k^2\big) \tag{5}$$

Since the direct maximization of this likelihood is not straightforward, we use the Expectation Maximization (EM) algorithm [1][3] to perform the maximization.

3.1 The dedicated EM algorithm

The proposed EM algorithm starts from an initial parameter $\boldsymbol{\theta}^{(0)}$ and alternates the two following steps until convergence:

E Step (Expectation): This step consists of computing the expectation of the complete log-likelihood $\log p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta})$, given the observations and the current value $\boldsymbol{\theta}^{(m)}$ of the parameter $\boldsymbol{\theta}$ (m being the current iteration):

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)}) &= E[\log p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}) | \mathbf{x}; \boldsymbol{\theta}^{(m)}] \\ &= \sum_{i=1}^{n} \sum_{k=1}^{K} t_{ik}^{(m)} \log(\pi_{ik}(\mathbf{w}) \mathcal{N}(x_i; \boldsymbol{\beta}_k^T \boldsymbol{r}_i, \sigma_k^2)) \ , \end{aligned} \tag{6}$$

where $t_{ik}^{(m)} = p(z_{ik} = 1 | x_i; \boldsymbol{\theta}^{(m)}) = \frac{\pi_{ik}(\mathbf{w}^{(m)}) \mathcal{N}(x_i; \boldsymbol{\beta}_k^{T(m)} \boldsymbol{r}_i, \sigma_k^{2(m)})}{\sum_{\ell=1}^{K} \pi_{i\ell}(\mathbf{w}^{(m)}) \mathcal{N}(x_i; \boldsymbol{\beta}_\ell^{T(m)} \boldsymbol{r}_i, \sigma_\ell^{2(m)})}$ is the posterior probability that x_i originates from the kth regression model. As shown in the expression of Q, this step simply requires the computation of $t_{ik}^{(m)}$.

M step (Maximization): In this step, the value of the parameter $\boldsymbol{\theta}$ is updated by computing the parameter $\boldsymbol{\theta}^{(m+1)}$ maximizing the expectation Q with respect to $\boldsymbol{\theta}$. The maximization of Q can be performed by separately maximizing

$$Q_1(\mathbf{w}) = \sum_{i=1}^{n} \sum_{k=1}^{K} t_{ik}^{(m)} \log \pi_{ik}(\mathbf{w}) \quad \text{and} \quad Q_2(\boldsymbol{\beta}_k, \sigma_k^2) = \sum_{i=1}^{n} t_{ik}^{(m)} \log \mathcal{N}(x_i; \boldsymbol{\beta}_k^T \boldsymbol{r}_i, \sigma_k^2) \tag{7}$$

The maximization of Q_1 with respect to $\mathbf{w}$ is a multinomial logistic regression problem weighted by the $t_{ik}^{(m)}$. We use a multi-class Iterative Reweighted Least Squares (IRLS) algorithm [8][2][5] to solve it. Maximizing Q_2 with respect to $(\boldsymbol{\beta}_k, \sigma_k^2)$ consists of analytically solving a weighted least-squares problem.

3.2 Denoising and segmenting a signal

In addition to providing a signal parametrization, the proposed approach can be used to denoise and segment signals. The denoised signal can be approximated by the expectation

$$E(x_i; \hat{\boldsymbol{\theta}}) = \int_{\mathbb{R}} x_i p(x_i; \hat{\boldsymbol{\theta}}) dx_i = \sum_{k=1}^{K} \pi_{ik}(\hat{\mathbf{w}}) \hat{\boldsymbol{\beta}}_k^T \boldsymbol{r}_i \ , \quad \forall i = 1, \ldots, n \tag{8}$$

where $\hat{\boldsymbol{\theta}} = (\hat{\mathbf{w}}, \hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_K, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_K^2)$ is the parameters vector obtained at the convergence of the algorithm. On the other hand, a signal segmentation can also be deduced by computing the estimated label $\hat{z}_i$ of x_i: $\hat{z}_i = \arg \max_{1 \leq k \leq K} \pi_{ik}(\widehat{\mathbf{w}})$.

4 Experiments

This section is devoted to the evaluation of the proposed algorithm using simulated and real data sets. Two evaluation criteria are used in the simulations: the misclassification rate between the simulated partition and the estimated partition and the euclidian distance between the denoised simulated signal and the estimated denoised signal normalized by the sample size n. The proposed approach is compared to the piecewise regression approach [11] .

4.1 Simulated signals

Each signal is generated according to the regression model with a hidden logistic process defined by eq (1). The number of states of the hidden variable is fixed to $K = 3$ and the order of regression is set to $p = 2$. The order of the logistic regression is fixed to $q = 1$ what guarantees a segmentation into convex intervals. We consider that all signals are observed over 5 seconds. For each size n we generate 20 samples. The values of assessment criteria are averaged over the 20 samples. Figure 3 (left) shows the misclassification rate obtained by the two approaches in relation to the sample size n. It can be observed that the proposed approach is more stable for a few number of observations. Figure 3 (right) shows the results obtained by the two approaches in terms of signal denoising. It can be observed that the proposed approach provides a more accurate denoising of the signal compared to the piecewise regression approach. For the proposed model, the optimal values of (K, p) has also been estimated by computing the Bayesian Information Criterion (BIC) [4] for k varying from 2 to $K_{max} = 8$ and p varying from 0 to $p_{max} = 6$. The simulated model, corresponding to $K = 3$ and $p = 2$, has been chosen with the maximum percentage of 85%.

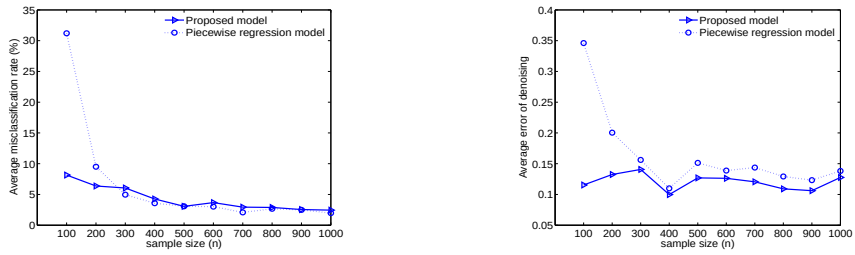

Fig. 3: Average values of misclassification rate (left) and error of denoising (right) in relation to the sample size n obtained with the proposed approach (triangle) and the piecewise regression approach (circle).

4.2 Real signals

This section presents the results obtained by the proposed model for signals of switch points operations. One situation corresponding to a signal with a critical defect is presented. The number of the regressive components is chosen in accordance with the number of the electromechanical phases of a switch points operation ($K = 5$). The value of q has been set to 1, what guarantees a segmentation into convex intervals, and the degree of the polynomial regression has been set to 3 which is adapted to the different regimes in the signals. Figure 4 (left) shows the original signal and the denoised signal (given by equation (8)). Figure 4 (middle) shows the variation of the proportions π_{ik} over time. It can be observed that these probabilities are very closed to 1 when the kth regressive model seems to be the most faithful to the original signal. The five regressive components involved in the signal are shown in figure 4 (right).

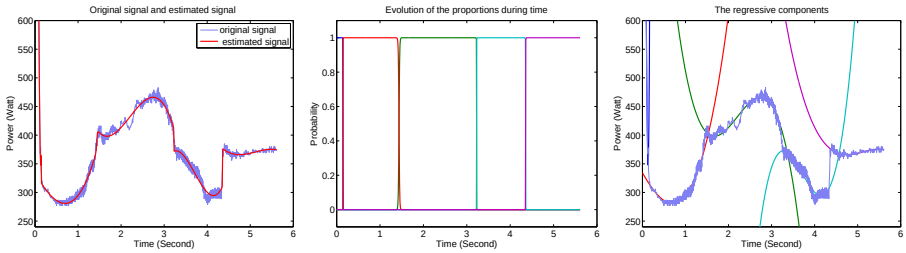

Fig. 4: Results obtained for a signal with defect

5 Conclusion

In this paper a new approach for signals parametrization, in the context of the railway switch mechanism monitoring, has been proposed. This approach is based on a regression model incorporating a discrete hidden logistic process. The logistic probability function, used for the hidden variables, allows for smooth or abrupt switchings between polynomial regressive components over time. In addition to signals parametrization, an accurate denoising and segmentation of signals can be derived from the proposed model.

References

[1] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society,* B, 39(1): 1-38, 1977.

[2] B. Krishnapuram, L. Carin, M.A.T. Figueiredo and A.J. Hartemink, Sparse multinomial logistic regression: fast algorithms and generalization bounds, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 27(6): 957-968, June 2005.

[3] G. J. McLachlan and T. Krishnan, *The EM algorithm and extensions,* Wiley series in probability and statistics, New York, 1997.

[4] G. Schwarz, Estimating the dimension of a model, *Annals of Statistics,* 6: 461-464, 1978.

[5] K. Chen, L. Xu and H. Chi, Improved learning algorithms for Mixture of Experts in multi-class classification, *IEEE Transactions on Neural Networks,* 12(9): 1229-1252, November 1999.

[6] M. Fridman, Hidden Markov Model Regression, Technical Report, Institute of mathematics, University of Minnesota, December 1993.

[7] M. I. Jordan and R. A. Jacobs, Hierarchical Mixtures of Experts and the EM algorithm, *Neural Computation,* 6: 181-214, 1994.

[8] P. Green. Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some robust and resistant alternatives, *Journal of the Royal Statistical Society,* B, 46(2): 149-192, 1984.

[9] R. E. Quandt and and J. B. Ramsey, Estimating mixtures of normal distributions and switching regressions, *Journal of the American Statistical Association,* 73(364): 730-752, 1978.

[10] S. R. Waterhouse, *Classification and regression using Mixtures of Experts,* PhD thesis, Department of Engineering, Cambridge University, 1997.

[11] V. E. McGee and W. T. Carleton, Piecewise regression, *Journal of the American Statistical Association*, 65, 1109-1124, 1970.