

SEGMENTATION OF TIME SERIES DATA USING GAUSSIAN MIXTURE MODELS

Antti Nurkkala

Thermal and Flow Engineering Laboratory

Department of Chemical Engineering

Åbo Akademi University

anurkkal@abo.fi

Switching system identification

- Switching systems (also called hybrid systems) are dynamical systems sharing both continuous and discrete dynamics.
- They can be modeled using models whose parameter values will change seldom in time.
- Example:

$$\begin{aligned} y(t) = & \sum_{i=1}^{n_y} a_i(\delta_t)(t-i)y(t-i) \\ & + \dots \sum_{i=1}^{n_{u_1}} b_{1i}(\delta_t)(t-i)u_1(t-i) \\ & + \dots \sum_{i=1}^{n_{u_k}} b_{ki}(\delta_t)(t-i)u_k(t-i), \end{aligned}$$

where y is the output, u the input and $\delta_t \in \{1, \dots, m\}$ the mode of the system at time t , n_y, n_{ui} are the model orders for the variables and a_i 's and b_i 's the parameters of the model that will change over time according to system mode δ_t .

- The identification problem is to find simultaneously the data segments in which the different parameter combinations of the models are valid, the suitable values for the parameters and possible orders of the models.
- In model or signal segmentation the model parameters are assumed to be piecewise constant but also to undergo changes that are more or less rare, i.e.
$$\theta(t) = \theta, t_i < t \leq t_{i+1}.$$
- Addressed by e.g. the following techniques: switching Markov regression, multiple model technique (multiple simultaneously developing Kalman filters), Algebraic and clustering based identification techniques.
- Applications: signal analysis, for example speech and seismic data, fault detection and diagnosis.

The regularization method

- Ohlsson et al. (2010) proposed a technique for segmentation of data set using linear ARX-models. The method is in sum-of-norms regularization programming form:

$$\min_{\theta(t)} \sum_{t=1}^N \left\| y(t) - \varphi^T(t)\theta(t) \right\|^2 + \lambda \sum_{t=2}^N \|\theta(t) - \theta(t-1)\|_{\text{reg}}, \quad (1)$$

where $y(t)$ contain the system output , $\varphi^T(t)$ the regressor variables and $\theta(t)$ the model parameters at time t , λ is the regularization variable and N the size of the data such as l_1 - or l_p -norm.

- Ohlsson et al. proposed also an iterative refinement procedure for reducing the number of the found segments.

- Good features of the above method:
 - Convex optimization problem – Efficient to solve and global solution guaranteed.
 - Global method: handles the entire data set as a whole – not so sensitive to the noise of the data that may affect the parameter identification of other methods.
 - Sum-of-norms regularization form favors solutions in which many of the regularized variables come out as exactly zero in the solution: the parameters will change infrequently.
 - Only one design parameter: λ that controls the trade-off between the model fit and the number of parameter changes.

Gaussian mixture models

- Belong to class of kernel density estimators.
- Are widely used to model complex densities that could not be modeled e.g. by using a single Gaussian kernel model.
- Are of form:

$$f(x) = \sum_{i=1}^K w_i \phi_i(x, \mu_i, \Sigma_i) \quad (2)$$

where ϕ_i is the (multivariate) gaussian function with mean μ_i and covariance matrix Σ_i , w_i is the weight of the function and K is the number of kernels in GMM.

- Are nonparametric methods: have flexible structure as compared to many parametric regression models, such as linear models.

Nonlinear segmentation

- The idea in the study is to find segmentation for time series data using nonlinear GMMs in the regularization method (1).
- First, a suitable GMM is to be searched and secondly, the weights of the GMM (see (2)) are to be parametrized over time in (1).
- The problem (1) is formulated such that $\varphi^T(t) = (\phi_1(x(t)), \dots, \phi_K(x(t)))$ contains the values calculated for each of the Gaussian kernel functions ϕ_i of the found GMM at $x(t)$ and $\theta(t)$ the weights for the Gaussians. Thus, actually the heights of the kernels of a proper single GMM are actually controlled over time in (1).

Finding a suitable GMM

- The task is to find a GMM to be used in (1) such that use of it will give good modeling results.
- One approach could be to first construct a Gaussian kernel around each of the input data point $x(t)$ in the training data set and to then reduce the number of kernels by merging kernels that are similar enough.
- One such approach employing this principle is the iterative pairwise replacement algorithm (IPRA) which was introduced by Scott and Szewczyk (2001) and which is widely used in finding proper mixture model.

IPRA

- The method is based on the idea that a density function could be approximated by a mixture model having small number of components without losing significant information.
- Begins from a GMM with a kernel built around each of the points in the data which is being modeled and the equal variances and weights:

$$f(x) = \sum_{i=1}^N \frac{1}{N} \phi(x, x_i, \Sigma).$$

Here, $\Sigma = h^2 \cdot I$, h being the initial bandwidth length for the kernels and I the identity matrix.

- Merges iteratively the most similar pair of kernels in a sense of some metric until there is only a single kernel left. When merging the kernels, the weights, means and variances are also updated.

- The choice of bandwidth h is crucial for the method. If h is being selected too small, it results to an overfitting GMM with very small covariances. This will also lead to numerical problems when solving (1), since most of the values of the model kernel functions will be zero at $x(t)$.
- Too big values for h will lead to oversmoothing model.
- Scott and Szewczyk (2001) proposed a rule for finding an optimal bandwidth:

$$h_* = \hat{\sigma}_i \cdot N^{(-1/(d+4))},$$

where $\hat{\sigma}_i$ is the standard deviation estimated for the i 'th component of the input vector, N is the size of the data set and d the dimension of the problem.

- It seems that the value given by this rule is too small for the modeling the blast furnace data. Thus, actually the value $h = 70 \cdot h_*$ is (probably finally) being used. This value has been found empirically based on the produced predictions when using the found segmental models.

- Similarly to Sung (2004), the Hellinger metric between two Gaussians is being used:

$$H(w_i, w_j, \phi_i, \phi_j) = \sqrt{w_i w_j} (1 - \int \sqrt{\phi_i \phi_j} dx).$$

- It is sufficient to calculate the distance for pairs of kernels that are defined by a minimum spanning tree (MST) of the set of kernels. This would require $O(N \cdot \log N)$ operations. In the study, the MST is being defined using well-known Prim's algorithm.
- The method of moments (MoM) parameter updating rules are used when merging the components:

$$\begin{aligned} w &= w_i + w_j, \\ \mu &= \frac{w_i}{w} \mu_i + \frac{w_j}{w} \mu_j, \\ \Sigma &= \frac{w_i}{w} \Sigma_i + \frac{w_j}{w} \Sigma_j + \frac{w_i w_j}{w} (\mu_i - \mu_j)(\mu_i - \mu_j)^\top. \end{aligned}$$

- Features of IPRA:

- MoM parameter updates are efficiently calculable as compared to e.g. well-known maximum likelihood estimation (MLE) updates.
- IPRA does not take regression capability of the model into account when building GMMs. It only reduces the number of similar components in the model.

Application of the method: hot metal silicon prediction

- The above nonlinear segmentation method is being validated in a benchmark problem from industry. It is applied in finding suitable nonlinear models for prediction of silicon concentration in pig iron produced in a blast furnace.
- The best model during prediction of the system outputs is defined using the exponential weighting criterion:

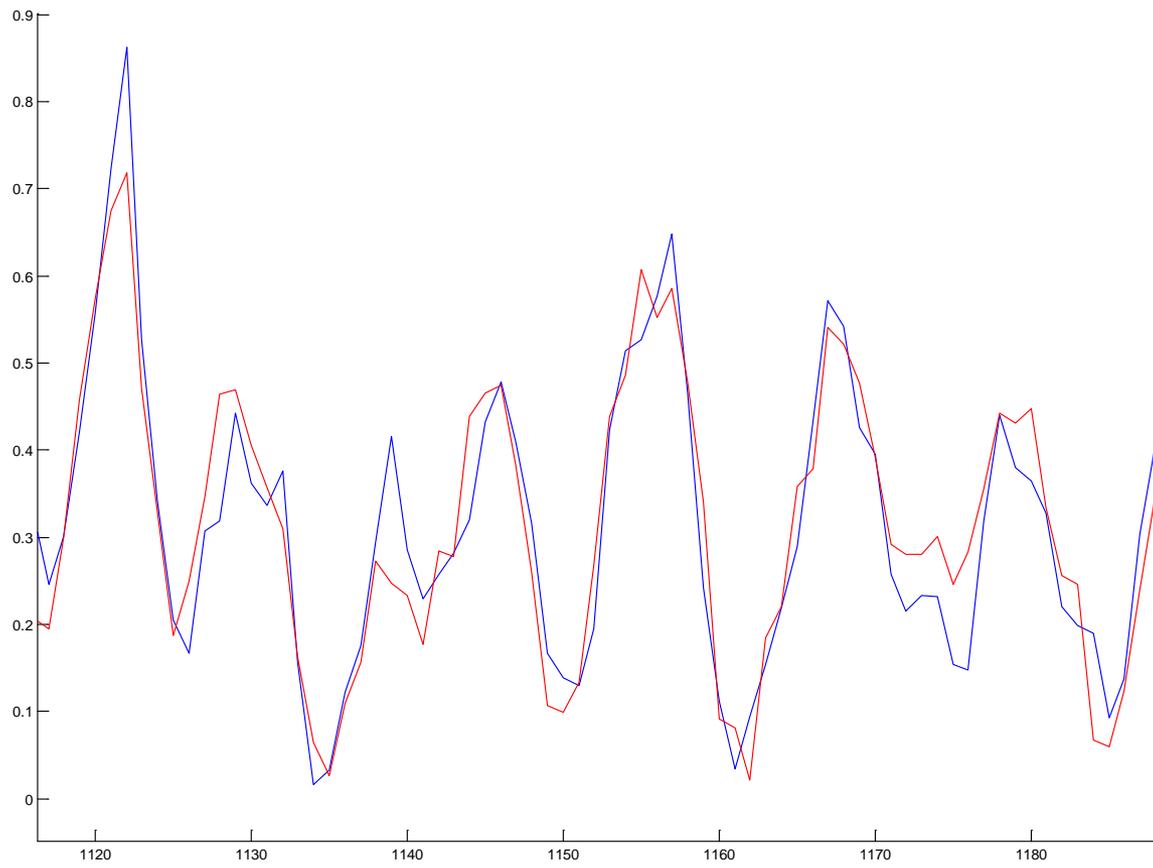
$$\epsilon(t) = \sum_{k=0}^{J-1} \lambda^{k+1} (y(t-k) - \hat{y}(t-k))^2,$$

where λ , $0 < \lambda < 1$, is the weighting factor.

- The prediction procedure is the following:
 1. Use the above segmentation method for finding predictive GMMs that are valid in different local regions of the training data.
 2. Use the models that were found on the previous step of the procedure in prediction of the future values of the system. The proper model at each time instant during prediction is being selected using the exponential weighting strategy.
- A proper value for the bandwidth h together with the desired number of kernels in GMM is to be given to the IPRA. A correct value for the weight λ in the exponential strategy is to be defined, also.
- In the study, the values $K = 50$ and $\lambda = 0.99$ are being used. The choices are based on empirical studies of the method.

- Data for a period of five months of operation collected from a blast furnace with unit volume of approximately 1300 m^3 is being used in modeling. Thus, data consists of at total 3490 hourly recorded samples.
- Data consists of at total 16 input variables and a single output variable. The input variables are partly controllable but partly uncontrollable, being results of measurements or calculated quantities of the process.
- The data is being divided to two sets of equal length: the training and the prediction data sets.

- The prediction results are seen to be superior to the predictions of obtained by using many other existing modeling techniques, for example to the neural network modeling (Nurkkala, 2011) and linear modeling technique.
- The level of the predicted silicon values is close to level of the silicon. Also, the changes of the directions of the curves of true and estimated values of the output correspond strongly each other.
- The good level of the predictions is obtained for the whole prediction period due to the use of segmented models found from the training data. This result is verified by the fact that when using a single GMM the overall level of the predictions is remarkably worse.



References

1. Ohlsson H.; Ljung L.; Boyd S.; Segmentation of ARX-models using sum-of-norms regularization, *Automatica*. **2010**, 46, 1107.
2. Scott, D.; Szewczyk.; From Kernels to Mixtures, *Technometrics*, **2001**, 42, 323.
3. Guang Sung H.; Gaussian Mixture Regression and Classification, Doctoral thesis, Rice University, Houston, Texas, **2004**.
4. Nurkkala, A.; Pettersson, F; Saxén; H. Nonlinear Modeling Method Applied to Prediction of Hot Metal Silicon in the Ironmaking Blast Furnace. *Ind. & Eng. Chem. Res.* **2011**, 50, 9236.