REALISATION OF MARKOV CHAIN'S PARAMETER REDUCTION ALGORITHM IN MATLAB ENVIRONMENT

Hofreiter Milan & Garajayewa Gunca A.[*]
Department of Instrumentation and Control Engineering,
Faculty of Mechanical Engineering, Czech Technical University in Prague

Abstract: The contribution concerns with the problems related to approximate identification of stochastic systems modelled by Markov chains. Although Markov chains are easily identifiable and adaptable their use is restricted because of extremely large dimension of the sufficient statistic. The paper refers about possible way, which helps to overcome this drawback. The proposed algorithm for dimensionality reduction in Markov chains is based on kernel smoothing technique. The applicability of the suggested methodology is presented in the Matlab programming environment.

Key words: Markov, chains, probability, reduction, kernel, identification, prediction.

1 Introduction

Most processes met in practice are uncertain in the sense that it is not possible to determine exactly the future output values of the process. (Controlled) Markov chain is an important class of universal black-box models suitable for the description of non-linear stochastic systems. When using the most general parameterisation their estimation as well as control design is simple. High dimensionality is their only but significant drawback.

Large variety of methodologies has been proposed to solve the dimensionality problem in parameterisation of MC (Kárný, M., et al., 1994; Pavelková, L., 1994; Hofreiter, M., 1997; Valečková, M., Kárný, M. & Sutanto, E.L., 2001).

In this paper we introduce the approach for dimensionality reduction, which uses the kernel smoothing technique (Hofreiter, M. & Garajayewa, G.A., 2000). The proposed algorithm from the measured input-output data estimates parameters of transition probability matrix (TPM), which relate to the values of the so-called *regression vector*. These parameters are preserved and then used for the prediction of output signal. During the prediction time we (almost) always meet the situation, when the *unknown* regression vectors (which were not estimated before) have been occurred. In such a case the algorithm determines from the estimated TPM the set of neighbouring regression vectors, which are in close proximity to the measured unknown one. Then, according to the detailed description of their transition probabilities and distance information, the algorithm evaluates the resulting transition probability for the measured regression vector. By this way, the unknown row (regression vector) of TPM is estimated and the output prediction is determined. The result of this solution is considerable parameter reduction, which at least helps to overcome the mentioned above disadvantage of Markov chains. Furthermore, the proposed algorithm has been applied for one-step-ahead prediction of real ECG (Electro Cardio Gram) signal that is a basic instrument for a diagnosis of heart diseases in cardiology. Achieved results of this application confirm the feasibility of the proposed algorithm.

2 Preliminaries

In this section, we sum up some basic relations on uncertainty whose application leads to solutions to all identification and control problems we deal with. Given three uncertain quantities a, b and c, hold the following three fundamental relations.

- **Marginalization.**

$$p(b \mid c) = \int p(a,b \mid c)da \tag{1}$$

where $p(\cdot \mid \cdot)$ represents conditional probability.

- **Chain rule.**

$$p(a,b \mid c) = p(a \mid b,c) \cdot p(b \mid c) \tag{2}$$

- **Bayes formula.**

$$p(a \mid b,c) = \frac{p(b \mid a,c) \cdot p(a \mid c)}{\int p(b \mid a,c) \cdot p(a \mid c)da} \propto p(b \mid a,c) \cdot p(a \mid c) \tag{3}$$

where the symbol $\propto$ stands for *proportionality*, i.e. equality up to a normalizing factor.

The integral taken over a^* is converted into a sum whenever a is a discrete-valued quantity.

Bayes formula is the heart of Bayesian statistics. It can be derived from the above two relations.

3 Markov Chain

Markov chains describe a very general class of dynamic systems, which are naturally discrete or treated by a digital technology. Identification of a Markov chain is shown below for a single discrete output $y(\tau)$ and discrete input $u(\tau)$ (Hofreiter, M., 1996)

$$y(\tau) \in S_y = \{1,\ 2,\ ...,\ N_y\}, \quad N_y < \infty$$
$$u(\tau) \in S_u = \{1,\ 2,\ ...,\ N_u\}, \quad N_u < \infty$$

Markov chain is a probabilistic model describing by the transition probability matrix the relationships of the discrete output $y(\tau)$ to the finite past input-output history $d^{(\tau-1)}$, stored in a discrete valued *regression vector* $z(\tau) = [z_1(\tau),\ z_2(\tau),\ ...,\ z_{\rho_z}(\tau)]$.

Here $\tau = 1,\ 2,\ 3,\ ...,\ t$ is discrete time

$y(\tau)$ is ρ_y-dimensional vector of system outputs

$u(\tau)$ is ρ_u-dimensional vector of system inputs

$d^{(\tau-1)} = [d(1),\ d(2),\ ...,\ d(t-1)]$,

$d(\tau-1) = [y(\tau-1);\ u(\tau-1)]$ is ρ_d-dimensional vector of measured data $(\rho_d = \rho_y + \rho_u)$,

$z_i(\tau) \in S_{z_i} = \{1,\ 2,\ ...,\ N_{z_i}\},\ N_{z_i} < \infty$

$z \in S_z = S_{z_1} \times S_{z_2} \times \ ... \ \times S_{z_{\rho_z}}$, ($\times$ denotes Cartesian product).

3.1 System Model

By a *system model* M we shall mean any mathematical model, which for the time period required through a finite set of parameters defines the set of conditional probability distribution

$$p(y(t) \mid u(t),\ d^{(t-1)},\ M), \quad t = 1,2,3,\ ... \tag{4}$$

If the model structure is known and we need to estimate a finite set of unknown parameters θ within a given model structure then M in (4) can be replaced by θ.

Let us suppose that the output $y(t)$ depends on a limited amount of past data through ρ_{Mz}-dimensional statistic $z(\ \cdot\ )$. The statistic $z(\ \cdot\ )$ (regression vector) maps

$$z(\ \cdot\ ): d^{(t-1)},\ \ u(t)\ \rightarrow z(t) \tag{5}$$

That means that the structure of the model is determined by the structure of the statistic ${}_{M}z(\ \cdot\ )$ and it holds

$$p(y(t)\,|\,u(t), d^{(t-1)},\ \theta) = p(y(t)\,|\,z(t),\ \theta) \tag{6}$$

3.2 System Identification

The problem of identification is to find a mathematical model of a given real system. If a mathematical model expressing input-output description is considered in the form (4) then the identification can be done in two steps (Peterka,V., I981):

I. Estimation of the system model, i.e. calculation the posterior probability distribution $p(\theta\,|\ d^{(t)})$.

Under the assumption (6) we may derive the formula for calculation $p(\theta\,|\,d^{(t)})$ using the rules (1) and (2)

$$p(M\,|\,d^{(t)}) \propto p(M)\cdot\prod_{\tau=1}^{t} p(y(\tau)\,|\,{}_{M}z(t),\ M), \tag{7}$$

where $p(\theta)$ is a prior probability.

II. Prediction of the output i.e. calculation of the posterior probability distribution $p(y(t+1)\,|\,u(t+1),\ d^{(t)})$.

With respect to the integration rule (1) and the product rule (2) under the assumption (6) we may derive

$$p(y(t+1)\,|\,u(t+1),\ d^{(t)}) = \int p(y(t+1)\,|\,z(t+1),\ \theta)\cdot p(\theta\,|\,d^{(t)})d\theta \tag{8}$$

Equation (6) holds for Markov chains as well. If the transition table is taken as an unknown parameter, then it is possible to estimate it from the known past history of the process via Bayesian statistics (Gao, H. & Kárný, M., I996; Pavelková,V., 1994).

In this case

$$\theta(z, y) = p(y(t) = y\,|\,z(t) = z,\ \theta) \tag{9}$$

and the matrix θ is given by $\theta = \left[\theta(z,\ y)\right]$.

Restrictions on ${}_{i}\theta$ resulting from the definition of the conditional probability are:

$$\theta(z,\ y) \geq 0 \qquad \sum_{y\in S_y}\theta(z,\ y) = 1 \qquad \text{for}\ \ \forall z \in S_z \tag{10}$$

Two important pieces of information can be derived for the structure of Markov chain using the relations (7) and (8):

I. Estimation of the parameters (the identification of probability table can be done for different z independently for each row of the probability table)

$$p(\theta(z,\ 1),\ ...,\ \theta(z,\ N_y)\,|\,d^{(t)}) \propto \prod_{y\in S_y}(\theta(z,\ y))^{(n_{z,y}(t)\ -1)} \tag{11}$$

II. Prediction of the output for a fixed t, fixed past history $d^{(t)}$ and given $z = z(t+1)$

$$p(y(t+1) = y \mid u(t+1),\ d^{(t)}) = \left. \frac{n_{z,y}(t)}{t_z(t)} \right|_{z=z(t+1)} \tag{12}$$

where $n_{z,y}(t) = n^1_{z,y}(t) + n^0_{z,y}(t)$, $t_z(t) = t^1_z(t) + t^0_z(t)$, $t^1_z(t) = \sum_{y \in S_y} n^1_{z,y}(t)$, $t^0_z(t) = \sum_{y \in S_y} n^0_{z,y}(t)$, $t^1_{z,y}(t)$ is the number of events $y(\tau) = y,\ z(\tau) = z$ for $\tau \le t$, $n^0_{z,y}(t)$ may be interpreted as the number of outputs with value y following the regression vector with value z observed before the identification start at a time $t = 0$.

It is obvious from the previous that we may evaluate the actual matrix $n(t) = [n_{z,y}(t)]$ recursively

$$n_{z,y}(t) = n_{z,y}(t-1) + \delta_{z,z(t)} \cdot \delta_{y,y(t)} \quad \text{for } \forall y \in S_y,\ \forall z \in S_z \tag{13}$$

where the Kronecker symbol $\delta_{a,b}$ is defined by

$$\delta_{a.b} = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} \tag{14}$$

That means we may determine probability distribution of parameters and the predictive probability distribution for the model structure. Using (11) and (12) we obtain for evaluation the posterior probability distribution about Markov chain structure:

$$p(\theta \mid d^{(t)}) \propto \prod_{\tau=1}^{t} \left(\left. \frac{n_{z,y}(\tau-1)}{t_z(\tau-1)} \right|_{z=z(\tau),\ y=y(\tau)} \right) \tag{15}$$

4 Approximate Identification of Markov Chains

The previous relations for estimation of the parameters and prediction of the output show that the estimation of Markov chain consists just in simple counting. However, as the dimension of the sufficient statistic is extremely large even for medium dimension of a regression model and small cardinality of data-value sets the applicability of Markov chains is restricted.

In real cases, it is not possible to assume, that the transition probability matrix will be known for all possible previous states and current inputs, which define the values of the regression vector and therefore a complete model of the system cannot be obtained in straightforward way. Nevertheless, if the parameters of the transition probability matrix are known for the regression vectors, which are in close proximity to the measured one, then it is possible to estimate unknown parameters of the measured regression vector through a kernel smoother.

Kernel smoother uses an explicitly defined set of local weights defined by the kernel to produce the estimate at each target value. Usually a kernel smoother uses weights that decrease in a smooth fashion as one moves away from the target point (Hastil & Tibshirani, 1997; Härdle, 1990). The weight given to the *j*-th point in producing the estimate at x_0 is defined by

$$w_{0j} = \frac{c_0}{\lambda} \cdot d\left(\left| \frac{x_0 - x_j}{\lambda} \right| \right) \tag{16}$$

where $d(l)$ is an even function decreasing in $|l|$. The parameter λ is the window-width, also known as the bandwidth and the constant c_0 is usually chosen so that the weights sum to unity. Epanechnikov kernel (Epanechnikov, V., 1969) ranks among popular kernels

$$d(l) = \begin{cases} 2/4 \cdot (1 - l^2) & \text{for } |l| \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

We may use this technique for estimation unknown parameters $\theta(r, y)$; $y \in S_y$ if some parameters $\theta({}^j z, y)$; $y \in S_y$, ${}^j z \in S_r \equiv \{ {}^j z : | r - {}^j z | < \lambda \}$, $j = 1, 2, ..., q$; $q > 0$ are known. In the previous relation was used the following notation:

r is the for the first time observed value of the regression vector,

q is the number of the known rows of the transition probability matrix θ with the corresponding values of the regression vector in the set S_r,

${}^j z$ is the *j*-th value of the regression vector,

$|v|$ denotes the absolute value of v,

λ is chosen bandwidth,

In such a case, the suggested algorithm estimates the parameters $\theta(r, y)$; $y \in S_y$ according to the following relation

$$\theta(r, y) = \sum_{j=1}^{q} w_{r,j} \cdot \theta({}^j z, y), \quad y \in S_y \tag{18}$$

where $w_{r,j} = c_0 \cdot \left(1 - \frac{\| r - {}^j z \|^2}{\lambda^2} \right)$

c_0 is chosen so that $\sum_{j=1}^{q} w_{r,j} = 1$, and $\| v \|$ denotes Euclidean norm of the vector v.

Described algorithm for output prediction does not require to know all parameters of the transition probability matrix θ and therefore it preserves only parameters of the transition probability table relating to the values of the regression vector that has occurred by the actual time which radically reduces memory demands.

5 Application

Described in the section 4 dimensionality reduction methodology using kernel smoothing technique was applied for one-step ahead prediction of real ECG (Electro Cardio Gram) signal that is a basic instrument for a diagnosis of heart diseases in cardiology. In Figure 1 is shown the fragment of ECG-output signal, which was used to illustrate the application of the suggested methodology. Sampling interval was 0.003 s.

It is should be noted that we consider the *autonomous* system. In this case $d(t) = y(t)$ and instead of (4) we have the set of conditional distributions

$$p(y(t) \,|\, y(t-1),\ \theta), \quad t = 1, 2, \ ... \tag{19}$$

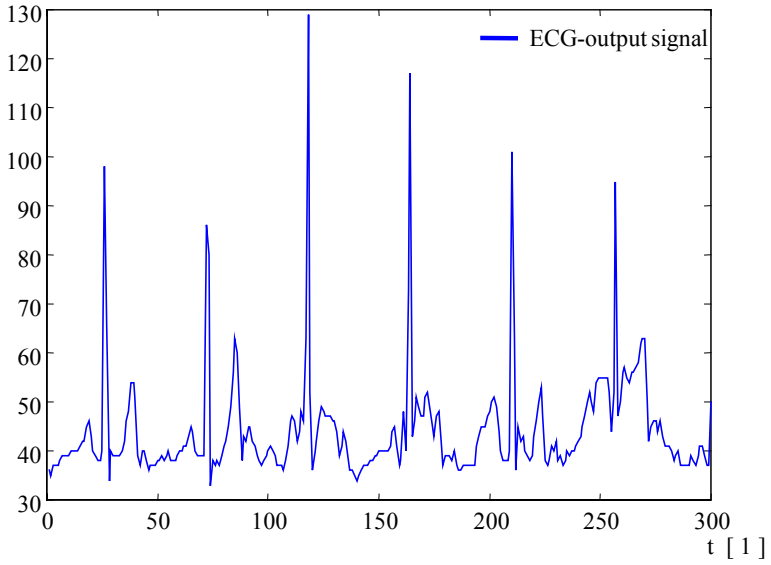

Figure 1 - Course of ECG output signal; t - discrete time

Our task was to estimate Markov chain model and predict the output signal. Then, after applying the algorithm for parameter reduction in Markov chains, make a comparison of results and show the improvement of the prediction.

As we wanted to use Markov chains for modeling and output prediction, the output value interval was divided into 34 parts. The discretized output set is thus $S_y = \{1, 2, 3, \ldots, 34\}$. The structure of regression vector was determined $z(t) = [\, y(t-1) \quad y(t-2) \,], \quad t = 1, 2, \ldots$.

Figure 2 demonstrates the course of output $y(t)$ and the output prediction determined by the expected value $Ey(t)$ of the output $y(t)$ derived from the Markov model with the regression vector z. Results of this prediction were received before we applied suggested algorithm for the parameter reduction.

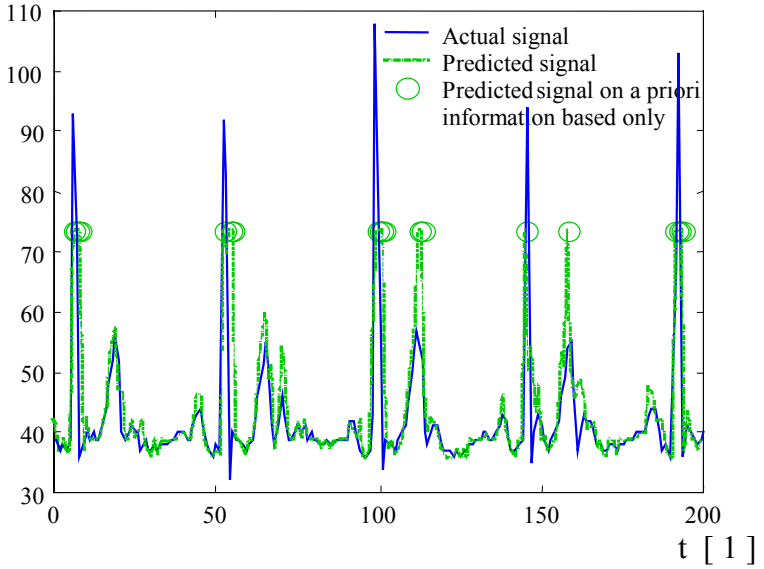

Figure 2 - Course of actual and predicted signals.

Situation before applying algorithm for dimensionality reduction

Number of unknown regression vectors, which occurred during the output prediction, was seventeen. In the Figure they are marked by symbol "○". It is seen, that existence of unknown regression vectors cause inaccurate output prediction.

Improved results of the output prediction were received after we applied new suggested method, which is described in previous section. It is evident from Figure 3, that mentioned previously algorithm is able to determine the set of neighbouring regression vectors, which are in close proximity to the measured unknown regression vector and according to the detailed description of their transition probabilities and distance information, the algorithm evaluates the resulting transition probability for the measured regression vector. By this way, the unknown rows of transition probability table are estimated and the output prediction is determined and improved.

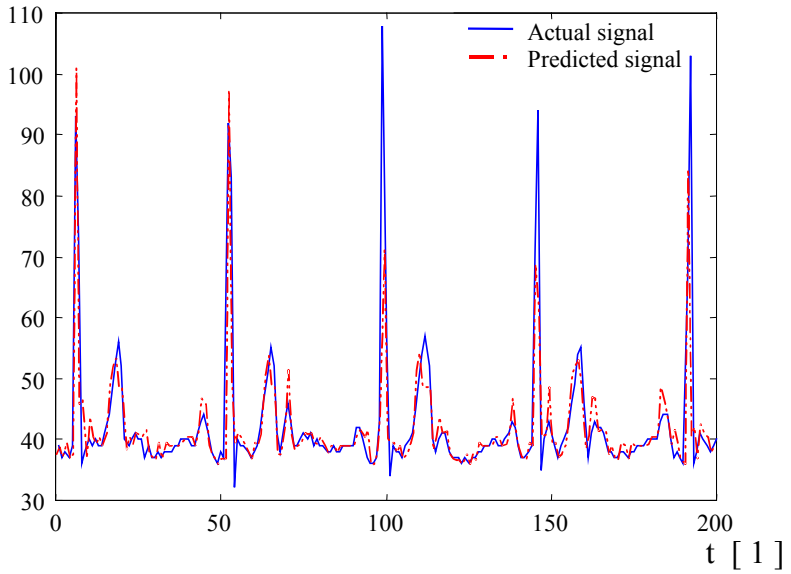

Figure 3 - Course of output and predicted signals.

The situation after applying algorithm for dimensionality reduction

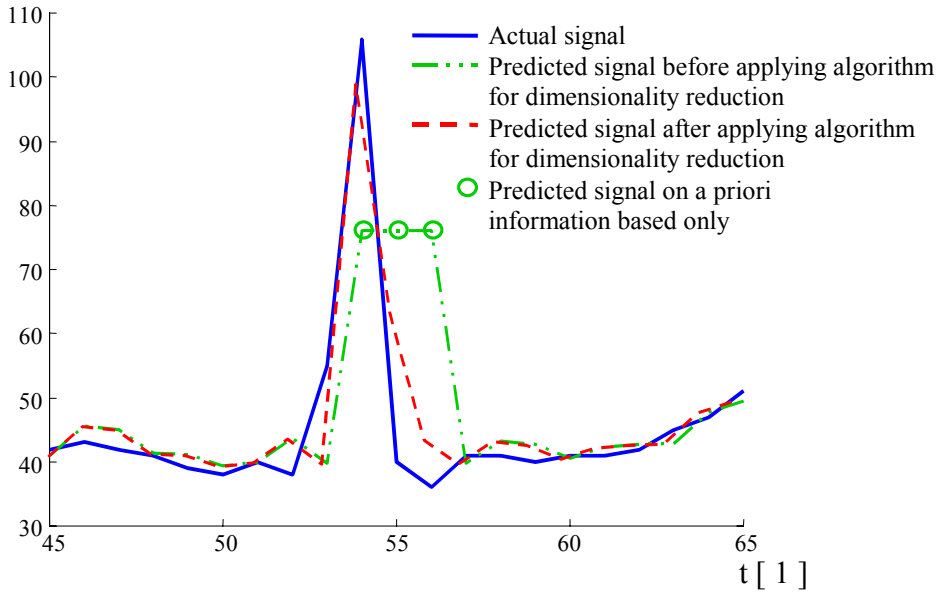

Figure 4 - Enlarged window time interval 45 - 65 s of Figure 3

In the Figure 4 is illustrated output prediction, where the time interval [45, 65] was enlarged on purpose to accentuate the quality of prediction using mentioned algorithm.

To show the computation precision, the criterion of Mean Absolute Deviation (MAD) was chosen:

$$MAD = \frac{1}{t}\sum_{t=1}^{T} \left| y(t) - Ey(t) \right| ,$$

where *Ey(t)* is the output prediction determined by the expected value of the output *y(t)*. Figure 5 demonstrates the typical course of the MAD for algorithm with (the curve *a*) and without (the curve *b*) considering the neighboring regression vectors for output prediction.

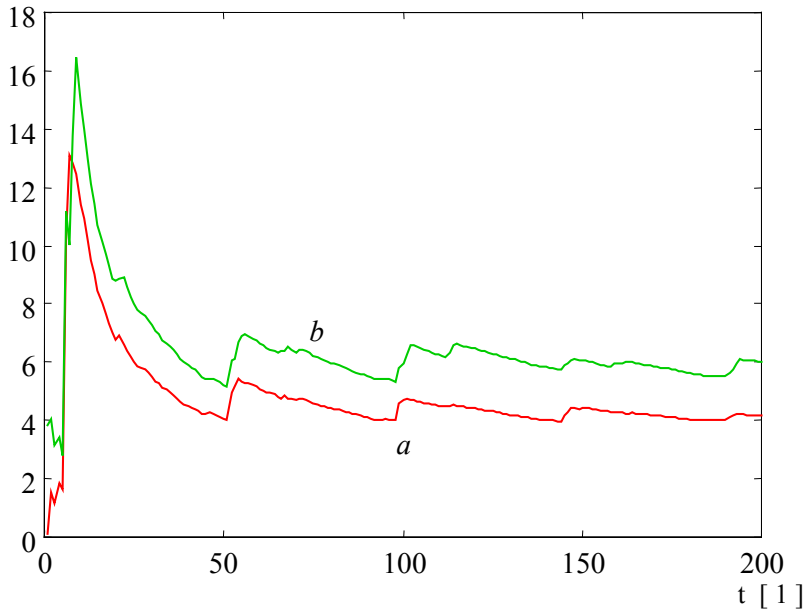

Figure 5 - Course of the MAD for algorithm with (the curve a) and without (the curve b) considering the neighboring regression vectors

6 Conclusion

This contribution develops an approximate prediction methodology in order to combat the curse of dimensionality inherent in Markov chains. For this purpose Kernel smoothing technique has been used. The proposed algorithm from the measured input-output data estimates parameters of transition probability matrix (TPM), which relate to the values of the regression vector. These parameters are preserved and then used for the prediction of output signal. In case the unknown regression vectors have been occurred, algorithm determines from the estimated TPM the set of neighbouring regression vectors, which are in close proximity to the measured unknown one. Then, according to the detailed description of their transition probabilities and distance information, the algorithm evaluates the resulting transition probability for the measured regression vector. By this way, the unknown rows (regression vectors) of TPM are estimated and the output prediction is determined. Furthermore, the proposed algorithm has been applied for one-step-ahead prediction of real ECG (Electro Cardio Gram) signal that is a basic instrument for a diagnosis of heart diseases in cardiology. Achieved results of this application confirm the feasibility of the proposed algorithm.

Acknowledgement: The presented research has been supported by the Research Project No. J04/98: 212200008 Development of Integrated Methods and Tools in Mechanical Engineering.

7 References

Epanechnikov, V. *Nonparametric estimates of a multivariate probability density* Theory of Probability and Applications, 14, 153 – 8, 1969

Gao, H. & Kárný, M. *Can Markov chains supplement adaptive AR predictors?* In: Preprints of the 2nd European IEEE Workshop on Computer Intensive Methods in Control and Signal Processing, Praha, 1996, pp.246-249

Gao, H. *Bayesian Approximation On-line Identification and Adaptive Control of Large-state-space and High-order Markov Chains* Dissertation, Prague, 1998

Härdle, W. *Applied nonparametric regression* Cambridge University Press, 1990, ISBN 0-521-42950-1

Hastil, T.J. & Tibshirani, R.J. *Generalized Additive Models* Chapman & Hall, London, 1997, ISBN 0-412-343908

Hofreiter, M. *Estimation of the Markov chain order* In: Proceedings of conference Process Control ŘIP 96, University of Pardubice, Horni Bečva, 1996, pp.109-112

Hofreiter, M. *Approximate identification of continuous system* Proceedings of Workshop 1997, Czech Technical University, Prague, 1997, pp.223-224

Hofreiter, M. *Identification of Markov Chains Using Discretization with Decomposition.* Proceedings of 8th International DAAAM Symposium 23-25th October 1997. Croatia: ICCU, Dubrovnik, 1997, pp.129–130. ISBN 3-901509-04-6.

Hofreiter, M. & Garajayewa, G.A. *Parameter Reduction in Markov Chains* Summaries 2, ICPR-16, Prague, 29 July – 3 August 2001, p.139

Kárný, M., Halousková, A. & Zornigová, L. *On pooling of expert opinions* In: Preprints SYSID'94, Kopenhagen, 1994, Vol.2. pp.477-482

Pavelková, L. *Approximate Identification of Markov Chains.* Preprints of the European IEEE Workshop CMP´94. Prague: UTIA AVCR, 1994, pp.335–340.

Peterka,V. *Bayesian system identification*. In: Eykhoff P.(ed). Trends and Progress in System Identification, Pergamon Press, Oxford, 1981, pp.239-304

Valečková, M., Kárný, M. & Sutanto, E.L. Bayesian M-T Clustering for Reduced Parameterisation of Markov Chains Used for Non-linear Adaptive Elements Automatica, vol. 37, 2001, pp. 1071-1078

Information about author:

Surname	First Name	Institution	E-mail & telephone
GARAJAYEWA	GUNCA	Department of Instrumentation and Control Engineering, Faculty of Mechanical Engineering, Czech Technical University in Prague *Technicka 4, 16607, Praha-6*	*guncha_g@hotmail.com* +420-2-24352896